Five-dimensional seismic data reconstruction using the optimally damped rank-reduction method[a]

[a]Published in Geophysical Journal International, 222, 1824-1845, (2020)

Yangkang Chen[*], *Min Bai*[*], *Zhe Guan*[†], *Qingchen Zhang*[*], *Mi Zhang*[‡], *and Hang Wang*[*]

ABSTRACT

It is difficult to separate additive random noise from spatially coherent signal using a rank-reduction method that is based on the truncated singular value decomposition (TSVD) operation. This problem is due to the mixture of the signal and the noise subspaces after the TSVD operation. This drawback can be partially conquered using a damped rank reduction method, where the singular values corresponding to effective signals are adjusted via a carefully designed damping operator. The damping operator works most powerfully in the case of a small rank and a small damping factor. However, for complicated seismic data, e.g., multi-channel reflection seismic data containing highly curved events, the rank should be large enough to preserve the details in the data, which makes the damped rank reduction method less effective. In this paper, we develop an optimal damping strategy for adjusting the singular values when a large rank parameter is selected so that the estimated signal can best approximate the exact signal. We first weight the singular values using optimally calculated weights. The weights are theoretically derived by solving an optimization problem that minimizes the Frobenius-norm difference between the approximated signal components and the exact signal components. The damping operator is then derived based on the initial weighting operator to further reduce the residual noise after the optimal weighting. The resulted optimally damped rank reduction method is nearly an adaptive method, i.e., insensitive to the rank parameter. We demonstrate the performance of the proposed method on a group of synthetic and real five-dimensional seismic data.

INTRODUCTION

With the increasing demand for exploration accuracy, wide azimuth seismic (WAZ) exploration technology has received more and more attention and development. The seismic data obtained by WAZ acquisition contains rich wavefield information with high illumination (Chen et al., 2014; Gan et al., 2016; Zu et al., 2017; Kim et al., 2017; Lu and Feng, 2017; Tian et al., 2017; Chen et al., 2017a; Bucha, 2017; Qu et al., 2016;

Chen and Song, 2018; Zhang et al., 2018, 2019; Wang et al., 2018). However, due to the influence of surface environments, the shot, receiver, azimuth and offset of the original data tend to be unevenly distributed, which is adverse to the subsequent processing and interpretation such as migration, pre-stack attribute analysis, reservoir prediction and fluid identification. Seismic data interpolation is such a processing step to regularize the irregularly sampled seismic traces onto regular grids (Fomel, 2003; Chen et al., 2016c; Zu et al., 2016b; Zhou and Li, 2018; Zhang et al., 2016b; Xiang et al., 2016; Chen et al., 2016a; Zhang et al., 2016c,a). In the past decades, a number of methods have been developed to reconstruct the missing seismic traces on regular grids. One widely used seismic reconstruction strategy is to transform the noisy seismic data into different domains in order to concisely represent the signal with a few number of selected components that capture the useful information and recover the missing components. These include the methods based on the Fourier transform (Zhou, 2017; Bai and Wu, 2018), Radon transform (Beylkin, 1987), seislet transform (Gan et al., 2015c,a,b; Liu et al., 2016b; Xue et al., 2017; Zhou et al., 2018a; Zhou and Han, 2018b; Chen and Fomel, 2018; Bai and Wu, 2019), curvelet transform (Ma and Dimet, 2009; Candès et al., 2006; Zu et al., 2015, 2016a), dreamlet transform (Wang et al., 2015), wavelet transform (Rioul and Vetterli, 1991; Gilles, 2013; Xie et al., 2015; Liu et al., 2016a; Mousavi and Langston, 2016), dictionary-learning based adaptive transform (Chen et al., 2016b; Anvari et al., 2017; Siahsar et al., 2017a,b; Wu and Bai, 2018b,d; Zu et al., 2018, 2019). Another type of methods utilize the predictable property of seismic data. In the prediction-based approaches, a prediction error filter is designed such that the predicted data and the existing data have the minimum misfit by solving a least-squares linear inverse problem (Spitz, 1991; Fomel, 2002; Chen et al., 2016c). Considering the aliasing issue in reconstructing regularly sampled seismic data, a prediction error filter is first estimated from the aliasing-free low-frequency components and then applied to aliased high-frequency components (Naghizadeh and Sacchi, 2007). The wave equation based methods can also be used to reconstruct highly incomplete seismic data. These methods connect the seismic record and the subsurface elastic properties via the elastic wave equation. However, these methods depend on a prior information about the subsurface elastic properties and are strictly not data-driven methods. In addition, these methods suffer from losing computational efficiency when applied in the interpolation methods (Ronen, 1987; Canning and Gardner, 1996; Fomel, 2003). A review of the latest methods on reconstructing low-dimensional seismic data is given in Chen et al. (2019).

However, the amplitude and phase variations of the WAZ seismic data are often not accurately recovered by conventional three-dimensional data reconstruction. More accurate reconstruction results can be obtained by simultaneously interpolating in five dimensions of inline, crossline, time, azimuth and offset because sampling along any particular subset of all dimensions is often less than ideal. The pre-stack seismic gathers processed by the five-dimensional (5D) interpolation have higher quality, which can not only improve the imaging accuracy, but also improve the capabilities of fracture prediction and fluid identification. Therefore, a large number of 5D seismic data reconstruction methods and techniques have emerged in the past decade.

The minimum weighted norm interpolation (MWNI) is a constrained inversion algorithm that was successfully applied to 5D seismic data interpolation (Liu and Sacchi, 2004; Trad, 2007, 2009). Trad (2007) employed this algorithm to regularize the data in the inline-crossline-azimuth-offset frequency domain and obtained pre-stack interpolated results for the migration input. This work proved that the results after 5D interpolation help to improve the fidelity of the migration, which laid the foundation for the subsequent promotion of five-dimensional seismic data reconstruction. Downton et al. (2008) confirmed the validity of the MWNI for 5D interpolation and found that the 5D interpolation can preserve the amplitude and improve the signal-to-noise ratio. The MWNI did successfully interpolate sparse data and reduced migration artifacts (Trad, 2009), but it had difficulty to deal regularly missing data with spatial aliasing. Chiu (2012) proposed an anti-aliasing MWNI method to improve the deficiency of conventional MWNI in processing aliased data. In addition, some other multidimensional interpolation algorithms are also available for the five dimensions (Chopra and Marfurt, 2013). Jin (2010) proposed a 5D interpolation method based on a damped least-norm Fourier inversion (DLNFI). Benefiting from the use of nonuniform discrete Fourier transform, DLNFI breaks the limitation in MWNI that the input data must be binned into the regular grid.

Another alternative 5D interpolation method utilizes wavefront attributes such as wavefront curvatures and propagation angles. Xie and Gajewski (2017) proposed a wavefront-attribute-based 5D interpolation (5D WABI) via a global optimization strategy instead of pragmatic search approach, which can take advantage of the wide, rich, or full azimuth acquisitions. Application on a synthetic seismic data have shown that the 5D WABI method is better at preserving diffractions than the damped rank-reduction method but at the expense of significantly lower computational efficiency. In recent years, dictionary learning and machine learning are applied to the reconstruction of 5D simple data (Yu et al., 2015; Jia and Ma, 2017; Jia et al., 2018). Data driven tight frame (DDTF) is a kind of dictionary-learning method, which can simultaneously denoise and interpolate 5D seismic data (Yu et al., 2015). In DDTF, a sparsity-promoting algorithm is used to build the dictionary which can represent the observed data and estimate the complete data. Jia and Ma (2017) combined the DDTF with a classic machine learning method named support vector regression (SVR) to optimize the learning, which obtained better performance than the Gauss SVR method. With the continuous improvement of intelligent methods, learning-based 5D data reconstruction will be a hot research topic in the future.

Rank-reduction based algorithms have great potentials for 5D seismic data interpolation. The basic assumption of these methods is that the fully sampled noise-free seismic data can be characterized as a low-rank matrix or tensor and the rank of the matrix or tensor increases when there are missing traces or noise in the data (Zhou et al., 2018b). The missing traces denote the zero-value seismic traces when the seismic data is binned from irregular station positions to regular grids. By solving a low-rank tensor completion problem via convex optimization, the missing traces can be accurately recovered with the information of all dimensions (Kreimer et al., 2013). In addition, this method has stronger denoising ability than the previous methods.

From other perspectives of rank minimization, Ely et al. (2015) estimated the complete data tensor via tensor singular value decomposition (SVD) and parallel matrix factorization, respectively. Subsequently, Gao et al. (2017) further developed a new and fast low-rank tensor completion method based on parallel square matrix factorization and advocated that reshaping the complete data tensor into almost square or square matrices can improve the reconstruction quality. However, when the signal-to-noise ratio of the observed seismic data is very low, the Cadzow rank-reduction method via truncated singular value decomposition (TSVD) does not achieve reasonable reconstruction result, i.e., the result may still contains a large amount of residual noise. Chen et al. (2016c) proposed a damped rank-reduction method to further suppress the residual noise by introducing a damping operator to the block Hankel matrix after TSVD. In the 5D reconstruction of field seismic data, the damped rank-reduction method achieved better performance than the Cadzow rank-reduction method (Trickett and Burroughs, 2009; Oropeza and Sacchi, 2011). Throughout the paper, the Cadzow rank-reduction method is referred to as the traditional rank-reduction method.

The basic assumption of the rank-reduction methods is that the Hankel matrix formulated from the useful seismic signal is of low rank and its rank is equal to the number of linear/planar events (or dipping components) in the seismic data (Oropeza and Sacchi, 2011; Huang et al., 2016; Siahsar et al., 2017c; Chen et al., 2017b; Zhou and Han, 2018a; Wu and Bai, 2018c,e,a; Bai et al., 2018a,b, 2019; Wu and Bai, 2019). However, in general the real seismic data is composed of nonlinear events, which are more complicated than the linear events. A common strategy for addressing this issue is to implement the algorithm in local patches since the events can be approximately viewed as linear in small patches (Zhang et al., 2017). However, this strategy poses another difficulty of choosing an appropriate rank for each local processing window. A practical implementation of the rank-reduction method may require the predefined rank to be relatively large in order to avoid the damage of useful signal due to inappropriate assumption of the structural complexity. This conservative selection of rank would make the resulted data contain a significant amount of residual noise. In this paper, we develop a relatively adaptive rank-reduction method, by which we can suppress the strong residual noise even when using a large rank. We first calculate a set of optimal weighting coefficients to weight the singular values as a first step by solving an optimization problem that minimizes the Frobenius-norm difference between the approximated signal components and the exact signal components. The rank-reduction method based on the optimal weighting strategy can be further improved when connected with the damped rank-reduction method. The resulted method is named as the optimally damped rank-reduction method, and can potentially provide the best solution to the rank-reduction based high-dimensional seismic reconstruction problem in an adaptive way. We will use comprehensive numerical tests and detailed analysis to demonstrate the superiority of the presented algorithm based on several synthetic examples and a real seismic data.

THEORY

Construction of the block Hankel matrix for 5-D seismic data

Let $D(t, hx, hy, x, y)$ denote the 5-D seismic data in the time domain, and $D(f, hx, hy, x, y)$ be the data in the frequency domain. For notation convenience, we omit f in the following context and use $D_{k1,k2,k3,k4}$ to denote $D(f, hx, hy, x, y)$. The traditional rank-reduction based methods require the construction of a level-four block Hankel matrix for 5-D seismic data to meet the low-rank assumption. A level-four block Hankel matrix means that we treat a series of level-three block Hankel matrix as elements and arrange them into a Hankel matrix. In a similar way, a level-three block Hankel matrix is constructed from a series of level-two block Hankel matrices while a level-two block Hankel matrix is formed from several standard Hankel matrices.

The level-four block Hankel matrix has the following explicit expression:

$$\mathbf{H}^{(4)} = \begin{pmatrix} \mathbf{H}_1^{(3)} & \mathbf{H}_2^{(3)} & \cdots & \mathbf{H}_{X_4-Y_4+1}^{(3)} \\ \mathbf{H}_2^{(3)} & \mathbf{H}_3^{(3)} & \cdots & \mathbf{H}_{X_4-Y_4+2}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{Y_4}^{(3)} & \mathbf{H}_{Y_4+1}^{(3)} & \cdots & \mathbf{H}_{X_4}^{(3)} \end{pmatrix}, \tag{1}$$

where

$$\mathbf{H}_{k4}^{(3)} = \begin{pmatrix} \mathbf{H}_{1,k4}^{(2)} & \mathbf{H}_{2,k4}^{(2)} & \cdots & \mathbf{H}_{X_3-Y_3+1,k4}^{(2)} \\ \mathbf{H}_{2,k4}^{(2)} & \mathbf{H}_{3,k4}^{(2)} & \cdots & \mathbf{H}_{X_3-Y_3+2,k4}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{Y_3,k4}^{(2)} & \mathbf{H}_{Y_3+1,k4}^{(2)} & \cdots & \mathbf{H}_{X_3,k4}^{(2)} \end{pmatrix}, \tag{2}$$

and

$$\mathbf{H}_{k3,k4}^{(2)} = \begin{pmatrix} \mathbf{H}_{1,k3,k4}^{(1)} & \mathbf{H}_{2,k3,k4}^{(1)} & \cdots & \mathbf{H}_{X_2-Y_2+1,k3,k4}^{(1)} \\ \mathbf{H}_{2,k3,k4}^{(1)} & \mathbf{H}_{3,k3,k4}^{(1)} & \cdots & \mathbf{H}_{X_2-Y_2+2,k3,k4}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{Y_2,k3,k4}^{(1)} & \mathbf{H}_{Y_2+1,k3,k4}^{(1)} & \cdots & \mathbf{H}_{X_2,k3,k4}^{(1)} \end{pmatrix}, \tag{3}$$

and

$$\mathbf{H}_{k2,k3,k4}^{(1)} = \begin{pmatrix} D_{1,k2,k3,k4} & D_{2,k2,k3,k4} & \cdots & D_{X_1-Y_1+1,k2,k3,k4} \\ D_{2,k2,k3,k4} & D_{3,k2,k3,k4} & \cdots & D_{X_1-Y_1+2,k2,k3,k4} \\ \vdots & \vdots & \ddots & \vdots \\ D_{Y_1,k2,k3,k4} & D_{Y_1+1,k2,k3,k4} & \cdots & D_{X_1,k2,k3,k4} \end{pmatrix}. \tag{4}$$

In order to make all target matrices (from equation 1 to 4) close to square matrices, parameters Y_i are defined as $\lfloor \frac{X_i}{2} \rfloor + 1$, $i = 1, 2, 3, 4$, where X_i denotes the size of the ith dimension. Here, $\lfloor \cdot \rfloor$ denotes the integer part of an input argument.

The process of transforming a four-dimensional hypercube $D_{k1,k2,k3,k4}$ to the block Hankel matrix $\mathbf{H}^{(4)}$ is referred to as the Hankelization process. We can briefly denote this process as:

$$\mathbf{H}^{(4)} = \mathcal{H}D_{k1,k2,k3,k4}. \tag{5}$$

Another important step in the rank-reduction based method is the rank reduction process, which can be denoted as $\mathcal{F}$.

Reconstructing the missing data aims at solving the following equation:

$$\mathcal{S} \circ \mathbf{M} = \mathbf{M}_0, \tag{6}$$

where $\mathcal{S}$ is a sampling matrix, $\mathbf{M} = \mathbf{H}^{(4)}$, and $\mathbf{M}_0$ denotes the block Hankel matrix with missing entries. $\circ$ denotes element-wise product.

Equation 6 is seriously ill-posed and the low-rank assumption is applied to constrain the model,

$$\begin{aligned} &\min_{\mathbf{M}} \parallel \mathcal{S} \circ \mathbf{M} - \mathbf{M}_0 \parallel_F, \\ &\text{s.t.} \quad \text{rank}(\mathbf{M}) = N. \end{aligned} \tag{7}$$

$\| \cdot \|_F$ denotes the Frobenius norm of an input matrix. The constraint in equation 7 means that we constrain the rank of the block Hankel matrix to be N.

The problem expressed in equation 7 can be solved via the following iterative solver:

$$\mathbf{M}_n = a_n \mathbf{M}_0 + (1 - a_n \mathcal{S}) \circ \mathcal{F}\mathbf{M}_{n-1}, \tag{8}$$

a_n is an iteration-dependent scalar that linearly decreases from $a_1 = 1$ to $a_{n_{max}} = 0$. a_n is used to alleviate the influence of random noise existing in the observed data.

Optimal weighting for rank reduction

The observed data $\mathbf{M}$ can be expressed as:

$$\mathbf{M} = \mathbf{S} + \mathbf{N}, \tag{9}$$

where $\mathbf{S}$ and $\mathbf{N}$ denote the signal and noise components. It is worth noting that for the derivation convenience, we assume the noise component $\mathbf{N}$ to be white. Although not exactly correct for seismic data, the denoising model also works for seismic data with band-limited noise. In the following analysis, we assume $\mathbf{M}$ and $\mathbf{N}$ have full rank and $\mathbf{S}$ has deficient rank. $\mathbf{M}$, $\mathbf{S}$ and $\mathbf{N}$ are all of size $I \times J$.

The singular value decomposition (SVD) of $\mathbf{S}$ can be represented as:

$$\mathbf{S} = [\mathbf{U}_1^S \quad \mathbf{U}_2^S] \begin{bmatrix} \mathbf{\Sigma}_1^S & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2^S \end{bmatrix} \begin{bmatrix} (\mathbf{V}_1^S)^H \\ (\mathbf{V}_2^S)^H \end{bmatrix}. \tag{10}$$

Because of the deficient rank, the matrix $\mathbf{S}$ can be written as:

$$\mathbf{S} = \mathbf{U}_1^S \mathbf{\Sigma}_1^S (\mathbf{V}_1^S)^H. \tag{11}$$

The singular value decomposition (SVD) of $\mathbf{M}$ can be represented as:

$$\mathbf{M} = [\mathbf{U}_1^M \quad \mathbf{U}_2^M] \begin{bmatrix} \mathbf{\Sigma}_1^M & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2^M \end{bmatrix} \begin{bmatrix} (\mathbf{V}_1^M)^H \\ (\mathbf{V}_2^M)^H \end{bmatrix}. \tag{12}$$

The rank-reduction method by TSVD refers to

$$\tilde{\mathbf{M}} = \mathbf{U}_1^M \mathbf{\Sigma}_1^M (\mathbf{V}_1^M)^H. \tag{13}$$

Here, $\tilde{\mathbf{M}}$ is the estimated signal component via TSVD. We assume the rank of $\mathbf{M}$ is N, and thus the size of $\mathbf{U}_1^M$ is $I \times N$. $\mathbf{\Sigma}_1^M$ is of size $N \times N$. $\mathbf{V}_1^M$ is of size $J \times N$.

However, the $\tilde{\mathbf{M}}$ is still a mixture of the signal and noise subspaces. Combining equations 9, 10, and 11, it is easy to derive that (Chen et al., 2016c)

$$\tilde{\mathbf{M}} = \mathbf{S} + \mathbf{U}_1^S (\mathbf{U}_1^S)^H \mathbf{N}, \tag{14}$$

where we can see that the estimated signal component via TSVD is still corrupted by the noise component, which is the projection of the noise component to the signal component.

This problem can be partially alleviated by a singular value thresholding step:

$$\tilde{\tilde{\mathbf{M}}} = \mathbf{U}_1^M \tilde{\mathbf{\Sigma}}_1^M (\mathbf{V}_1^M)^H, \tag{15}$$

where $\tilde{\mathbf{\Sigma}}_1^M$ is the thresholded singular values such that:

$$\tilde{\mathbf{\Sigma}}_1^M = \mathbf{T}(\mathbf{\Sigma}_1^M, \tau). \tag{16}$$

where $\mathbf{T}$ denotes a singular value thresholding operator and τ denotes the threshold. However, defining an optimal threshold is inconvenient and sometimes even difficult. It is because the noise component contributes differently to each singular value and a constant threshold is not plausible to deal with the inhomogeneous noise distribution.

Thus, in this paper, we propose an adaptive weighting algorithm to optimally define the singular values in order to best reconstruct the signal component. We introduce a weighting operator $\mathbf{W}$ to adjust the $\mathbf{\Sigma}_1^M$ after applying SVD to the observed noisy signal. To calculate the optimal weighting operator is equivalent to solving the following optimization problem:

$$\hat{\mathbf{W}} = \arg\min_{\mathbf{W}} \parallel \mathbf{U}_1^S \mathbf{\Sigma}_1^S (\mathbf{V}_1^S)^H - \mathbf{U}_1^M \mathbf{W} \mathbf{\Sigma}_1^M (\mathbf{V}_1^M)^H \parallel_F . \tag{17}$$

By an optimally weighted combination of estimated left and right singular vectors, the optimization problem 17 can yield the optimally adjusted singular values to obtain the closest low-rank estimates of signal component.

The optimal solution for the optimization problem was given in Benaych-Georges and Nadakuditi (2012) and Nadakuditi (2013):

$$\hat{\mathbf{W}} = \mathrm{diag}(\hat{w}_1, \hat{w}_2, \cdots, \hat{w}_N), \tag{18}$$

where

$$\hat{w}_i = \left(-\frac{2}{\sigma_i^M} \frac{D(\sigma_i^M; \mathbf{\Sigma})}{D'(\sigma_i^M; \mathbf{\Sigma})} \right). \tag{19}$$

σ_i^M denotes the ith diagonal entry of $\mathbf{\Sigma}_1^M$ ($\mathbf{\Sigma}_1^M \in \mathbb{R}^{N \times N}$). D denotes the D-transform:

$$\begin{aligned} D(\sigma; \mathbf{\Sigma}) &= \frac{1}{N} \mathrm{Tr}\left(\sigma (\sigma^2 \mathbf{I} - \mathbf{\Sigma}\mathbf{\Sigma}^H)^{-1} \right) \frac{1}{N} \mathrm{Tr}\left(\sigma (\sigma^2 \mathbf{I} - \mathbf{\Sigma}^H\mathbf{\Sigma})^{-1} \right) \\ &= \left[\frac{1}{N} \mathrm{Tr}\left(\sigma (\sigma^2 \mathbf{I} - \mathbf{\Sigma}^2)^{-1} \right) \right]^2 \end{aligned}, \tag{20}$$

where D' is the derivative of D with respect to σ, and $\mathrm{Tr}(\cdot)$ is the trace of the input:

$$\begin{aligned} D'(\sigma; \mathbf{\Sigma}) &= 2 \left[\frac{1}{N} \mathrm{Tr}\left(\sigma (\sigma^2 \mathbf{I} - \mathbf{\Sigma}^2)^{-1} \right) \right] \left[\frac{1}{N} \mathrm{Tr}\left((\sigma^2 \mathbf{I} - \mathbf{\Sigma}^2)^{-1} - 2\sigma (\sigma^2 \mathbf{I} - \mathbf{\Sigma}^2)^{-2} \sigma \right) \right] \\ &= \frac{2}{N^2} \left[\mathrm{Tr}\left(\sigma (\sigma^2 \mathbf{I} - \mathbf{\Sigma}^2)^{-1} \right) \right] \left[\mathrm{Tr}\left((\sigma^2 \mathbf{I} - \mathbf{\Sigma}^2)^{-1} - 2\sigma^2 (\sigma^2 \mathbf{I} - \mathbf{\Sigma}^2)^{-2} \right) \right]. \end{aligned} \tag{21}$$

Using the optimally estimated weighting operator expressed in equation 18, we can expressed the optimally estimated signal component as:

$$\hat{\mathbf{M}} = \mathbf{U}_1^M \hat{\mathbf{W}} \mathbf{\Sigma}_1^M (\mathbf{V}_1^M)^H. \tag{22}$$

The optimal weighting strategy of the singular values is a substitute to directly truncating the singular values as used in the traditional rank-reduction method. An early investigation of the strategy to improve the rank-reduction performance in seismic data denoising and reconstruction is presented in Aharchaou et al. (2017). Besides, there are also a number of alternatives to these two approaches (weighting and truncating) such as automatic rank determination (Gavish and Donoho, 2014; Trickett, 2015) and randomized approaches such as the randomized SVD and randomized QR factorizations (Cheng and Sacchi, 2014).

Optimally damped rank-reduction method

In this section, we will derive a damping operator for the optimally estimated signal components to further reduce the residual noise after rank reduction. Let

$$\mathbf{Q} = [\mathbf{U}_1^Q \quad \mathbf{U}_2^Q] \begin{bmatrix} \mathbf{\Sigma}_1^Q & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2^Q \end{bmatrix} \begin{bmatrix} (\mathbf{V}_1^Q)^H \\ (\mathbf{V}_2^Q)^H \end{bmatrix}. \tag{23}$$

to be an SVD of matrix $\mathbf{Q}$, and let

$$\mathbf{U}_1^Q = \mathbf{U}_1^M, \tag{24}$$

$$\mathbf{\Sigma}_1^Q = \hat{\mathbf{W}}\mathbf{\Sigma}_1^M, \tag{25}$$

$$\mathbf{V}_1^Q = \mathbf{V}_1^M, \tag{26}$$

then equation 22 can be understood as a TSVD of the matrix Q:

$$\tilde{\mathbf{Q}} = \mathbf{U}_1^Q \mathbf{\Sigma}_1^Q \mathbf{V}_1^{QH}. \tag{27}$$

Analogous to equation 14, equation 22 can be re-formulated as

$$\tilde{\mathbf{Q}} = \hat{\mathbf{M}} = \tilde{\mathbf{S}} + \mathbf{U}_1^{\tilde{\mathbf{S}}} \mathbf{U}_1^{\tilde{\mathbf{S}}H} \tilde{\mathbf{N}}, \tag{28}$$

When the rank is sufficiently large, we assume the estimated signal contains all signal components of the originally observed noisy data and contains less noise than the observed data. To further suppress the residual noise in the estimated signal, we re-analyze $\hat{\mathbf{M}}$ in detail. We can express the newly estimated signal as:

$$\tilde{\mathbf{Q}} = \hat{\mathbf{M}} = \mathbf{S} + \mathbf{U}_1^S \mathbf{U}_1^{SH} \tilde{\mathbf{N}}, \tag{29}$$

where $\mathbf{U}_1^S \mathbf{U}_1^{SH} \tilde{\mathbf{N}}$ denotes the residual noise component after the step using equation 22.

Following Chen et al. (2016c) and Huang et al. (2016), $\mathbf{S}$ can be approximated as:

$$\mathbf{S} = \mathbf{U}_1^Q \mathbf{\Sigma}_1^Q \mathbf{T} \left(\mathbf{V}_1^Q \right)^H, \tag{30}$$

$$\mathbf{T} = \mathbf{I} - \mathbf{\Gamma}, \tag{31}$$

where $\mathbf{I}$ is a unit matrix and here we name $\mathbf{T}$ the damping operator. The damping relation $\mathbf{\Gamma}$ is expressed as:

$$\mathbf{\Gamma} \approx \hat{\delta}^N \left(\mathbf{\Sigma}_1^Q \right)^{-N}, \tag{32}$$

where $\hat{\delta}$ denotes the maximum element of $\mathbf{\Sigma}_2^Q$ and N denotes the damping factor.

Considering that $\mathbf{U}_1^Q = \mathbf{U}_1^M$, $\mathbf{\Sigma}_1^Q = \hat{\mathbf{W}}\mathbf{\Sigma}_1^M$, and $\mathbf{V}_1^Q = \mathbf{V}_1^M$, equation 30 can be expressed as:

$$\mathbf{S} = \mathbf{U}_1^M \mathbf{T} \hat{\mathbf{W}} \mathbf{\Sigma}_1^M \left(\mathbf{V}_1^M \right)^H, \tag{33}$$

which is referred to as the optimally damped rank reduction method. The $\mathcal{F}$ can thus be chosen as the operation defined in equation 33, and through iterations we can reconstruct and denoise the 5D frequency-domain seismic data.

There are two main advantages of the optimally damped rank reduction method. First, compared with the traditional and damped rank-reduction methods, the optimally damped rank reduction method is insensitive to the rank parameter, making it

nearly an adaptive method for rank reduction based seismic denoising and reconstruction. This advantage is important because one of the most troublesome problems in processing complicated seismic data is the selection of the rank. A large rank tends to result in significant residual noise while a small rank tends to damage useful signals. This parameterization problem becomes more seriously when the rank reduction method is applied locally in windows. Because of the insensitivity to rank of the proposed method, one can choose a relatively large rank for all complicated datasets or local patches. Secondly, compared with the optimal weighting based rank reduction method, the proposed method can further suppress the noise components that reside mostly in the smaller singular values. The damping operator is data-driven and can adaptively separate signal and noise in the singular value spectrum further after the optimal weighting.

In the rank-reduction methods, construction of the level-four Hankel matrix is very computationally expensive. Recent advances in the rank reduction based methods show that the construction of the block Hankel matrix is not required. These methods exploit the structure of such matrices to avoid explicitly forming these matrices prior to factorization (Lu et al., 2015; Cheng and Sacchi, 2016). When factorizing data using only 1 or 2 spatial dimensions these approaches are not necessary, but moving to 3 or 4 spatial dimensions is not computationally feasible without considering matrix free approaches. In the current stage, we cannot move from the SVD-based method to SVD-free method because the damping operation has not been derived for the SVD-free case. Although it takes a large computational cost and is not very practically for the time being, it is still a promising algorithm. We will keep on investigating the acceleration of the current algorithm and make it computationally feasible in the future.

EXAMPLES

We first apply the proposed optimal rank-reduction method to a 5D synthetic data with linear/planar events, as shown in Figure 1. Figure 1(a) shows the clean synthetic data in a common offset gather. The 3D common offset gather $(t - x - y)$ has been arranged into a 2D matrix. Figure 1(b) shows the noisy data with extremely strong band-limited random noise. The useful signals are almost buried in the strong noise. Figure 1(c) shows the incomplete data, where 75% traces have been randomly decimated. The reconstructed and denoised data using three different methods, i.e., the traditional rank-reduction method (RR), the damped rank-reduction method (DRR), and the optimally damped rank-reduction method (ORR), are shown in Figure 2. All three methods successfully recovered most of the signals. It is clear that the RR method causes significant residual noise, the result from the DRR method is much cleaner than that from the RR method but still contains some residual noise. The proposed ORR method, however, obtains the cleanest result among the three results shown in Figure 2. To evaluate the quality of the three reconstructed data, given the exact solution in Figure 1(a), we calculate the local similarity map between each

reconstructed data and the exact solution. The local similarity is a way to evaluate the similarity between two vectors/matrices/cubes in a local manner so that the evaluation can be revealed immediately from the calculated local similarity maps. The detailed introduction of mathematics behind the local similarity calculation can be found in Chen and Fomel (2015). From the local similarity comparison, it is quite obvious that from the top panel (RR method) to the bottom panel (ORR method), the similarity becomes higher and higher, which indicates that the proposed method obtains the most accurate solution.

To examine the amplitude recovery quality, we select the 22nd traces from the clean data in Figure 1(a) and from the three reconstructed data in Figure 2, and draw them in Figure 4. Figure 4(a) shows the trace-by-trace comparison in the range of the whole trace and Figure 4(b) shows the zoomed trade-by-trace comparison. The zooming part is highlighted by the red transparent rectangle shown in Figure 4(a). From Figure 4 we can see that both RR and DRR methods (blue and red lines) cause strong fluctuations in the front part of the trace, while the proposed method obtains a result (the green line) that is the closest to the exact solution (the black line). The observation is even clearer in the zoomed comparison shown in Figure 4(b). Figure 5 shows the comparison of reconstruction error. The reconstruction error is calculated as the difference between the exact solution (Figure 1(a)) and each reconstructed data shown in Figure 2. From Figure 5 we can observe that both RR and DRR cause significant error and the proposed ORR method causes negligible reconstruction error except for some observable signal energy. The reconstruction error of RR method is higher than the DRR method, mostly because of the stronger residual noise caused by the RR method. Since the reconstruction error for the proposed method is mostly negligible, the error caused by the leakage signal energy becomes more apparent. To compare the leakage signal energy among three methods, we calculate the extra noise section of the RR and DRR methods. The extra noise section refers to the extra noise compared with the reference noise section from the ORR method (Figure 5(c)), thus is calculated as the difference between the error from either RR or DRR method and the error from the ORR method. The extra noise sections corresponding to the RR and DRR methods are shown in Figure 6. It is obvious that by calculating the extra error sections, the leakage useful energy from RR/DRR method and ORR method counteract each other and the resulting sections are mostly random noise. This observation indicates that the energy of the leakage signal for the three methods are nearly the same. We conclude from this test that despite the same damage to useful signals, the proposed method causes the least residual noise and thus obtains the cleanest reconstruction result.

We also compare the level-four block Hankel matrices of different methods. Figures 7a and 7b shows the Hankel matrices for the clean and incomplete data for the frequency slice of 30Hz. Figures 7c and 7d show the zoomed Hankel matrices for the clean and incomplete data. The zooming areas are highlighted by the red rectangles shown in Figures 7a and 7b. Because of the missing traces, the Hankel matrix contains a number of blank areas. Figure 8 shows the Hankel matrices for different methods. The left column correspond to the Hankel matrices after 1 iteration. The right column

correspond to the Hankel matrices after 10 iterations. We only conclude from Figure 8 that after 10 iterations all three methods obtain similar Hankel matrices as the exact solution shown in Figure 7a but we cannot see the difference between different methods clearly. The differences between different methods can be clearly observed from the zoomed Hankel matrices, as shown in Figure 9. It is clear that from the top to bottom in Figure 9, the Hankel matrix becomes smoother and smoother and is closer to the exact solution shown in Figure 7c.

In addition to the local similarity mentioned above, we also use the signal-to-noise ratio (SNR) defined as follows to quantitatively measure the performance (Huang et al., 2016; Chen, 2017; Huang et al., 2017):

$$SNR = 10\log_{10}\frac{\|\mathbf{s}\|_2^2}{\|\mathbf{s}-\hat{\mathbf{s}}\|_2^2}. \tag{34}$$

where $\mathbf{s}$ denotes the vectorized clean data and $\hat{\mathbf{s}}$ denotes the vectorized reconstructed data. In this example, the SNRs of the observed data, the reconstructed results via the RR method, the DRR method, and ORR method are -4.59 dB, 5.51 dB, 9.54 dB, and 11.83 dB, respectively. In this test, we use $rank = 10$, assuming that we do not have a prior information about the structural complexity of the data and thus needing to set a relatively high rank. We then test the performance when the rank is chosen smaller. The reconstructed data when $rank = 5$ are shown in Figure 10. Their corresponding error sections are shown in Figure 11. From Figures 10 and 11, we can reach almost the same conclusion as the last test. In this case, the SNRs of the reconstructed results via the RR method, the DRR method, and ORR method are 8.01 dB, 10.87 dB, and 11.97 dB, respectively. The detailed SNR comparison is shown in Table 1, including the case when $rank = 3$.

To test the sensitivities of each method to input noise level, we vary the variance of the additive random noise from 0.1 to 0.9 and calculate the output SNRs corresponding to different methods and draw the diagrams in Figure 12. The black line corresponds to the input SNR curve. The blue, red, and green lines correspond to the RR, DRR, and ORR methods, respectively. As the noise variance increases, the input SNR decreases, and so do the output SNRs of the three methods. However, the proposed method always outperforms the other two methods. It is worth noting that as the noise variance becomes larger, the differences between the proposed method and the other methods are also larger, which indicates that the proposed method is more effective for stronger noise.

To test the effectiveness of each method in situations with different sampling ratios, we vary the sampling ratio from 10% to 90%, and draw the input and output SNR curves in Figure 13. The colorful lines have the same meaning in this case. As we can see from Figure 13, the input SNR decreases with increasing sampling ratio. The output SNRs increase with increasing sampling ratios. It is clear that the output SNR curve of the proposed method is always above the other curves. The proposed method outperforms other two methods more when sampling ratio is higher.

To test the sensitivities of different methods to the input parameter, i.e., the

predefined rank, we vary the rank from 4 to 15 and draw the SNR curves in Figure 14. Both RR and DRR methods decrease fast as the predefined rank increases while the SNR curve of the proposed ORR method is almost flat and is also above the other two curves. This test indicates that while the RR and DRR are more or less sensitive to the predefined rank, the proposed method is almost parameter-free, which makes the ORR method convenient to use in realistic situations.

To compare the computational cost of different methods, we measure the computing time for the synthetic example with linear events. The computation is done on a MacBook Pro laptop equipped with an Intel Core i7 CPU clocked at 2.5 GHz and 16 GB of RAM. The detailed computing time comparison is shown in Table 2. It shows that the computational cost of the ORR method is higher than the other two methods. The RR and DRR methods have almost the same computational cost while the proposed method costs 2-5 times more than the other two methods.

In order to test the effectiveness of the three methods on data containing curved events and to compare the performances of different methods in this case, we then use the second example to show the performance. Figure 15 shows the clean data, noisy data, and incomplete data with 70% traces randonly missing in a common midpoint gather. Figure 16 shows the reconstructed data using different methods. Because this dataset no longer meet the linear-event assumption of the rank-reduction based methods, we use a relatively higher rank in this example. The top row in Figure 16 shows results when $rank = 12$. The bottom row in Figure 16 shows results when $rank = 24$. It is clear that the three methods also work when there are curving events. The proposed method obtains obviously cleaner result than the other two methods. The SNRs of the incomplete data, data from the RR method, DRR method, and ORR method are 0.59, 14.11, 16.46, and 16.86 dB when $rank = 12$ and are 0.59, 13.88, 16.01, and 16.83 dB when $rank = 24$. We also calculate the local similarity between the exact solution shown in Figure 15a and each reconstructed data and show the similarity cubes in Figure 17. The local similarity corresponding to the proposed method is obviously higher than those from the other two methods, indicating a more accurate reconstruction result using the proposed method.

Finally, we apply the three methods to a 5D field data. We use the data previously used in Chen et al. (2016c). The data have been binned onto a regular grid and a common offset gather of the field data is shown in Figure 18. In Figure 18, the colored stripes are the recorded seismic traces. The white blanks denote the missing traces, which means that we do not observe seismic data in these positions. Because of the difficulty in displaying a 5D dataset, we only show a common midpoint gather here. The 3D common midpoint gather is rearranged into a 2D matrix for a better view. The two transparent colored windows denote two zooming areas for an amplified comparison. In this example, roughly 80% traces are missing from the regular grids. Because of the high ratio of missing traces, the observed seismic traces do not show any spatial coherency. It is difficult to see the waveforms from the raw data. The results from the three aforementioned methods are shown in Figure 19. After 5D reconstruction, the white blanks in the raw data have been filled with seismic traces.

The waveforms become well aligned along the spatial direction. Compared with the raw data, all methods seem to obtain a dramatic improvement on the data quality. It is salient that both DRR and ORR methods obtain much smoother and cleaner results while the traditional RR method obtain a result that is noisier. Because of the strong residual noise in the result from the traditional RR method, the spatial coherency of the seismic events are deteriorated, which may affect the subsequent processing tasks like imaging, inversion, and interpretation. When zooming the data in the two transparent blue and red rectangles in both Figures 18 and 19, the comparison among different methods becomes much clearer. From Figures 20 and 21, we observe that although the DRR method obtain a much smoother result compared with the RR result, as already discussed from Chen et al. (2016c), the ORR method obtains a even smoother result, with energy spatially more correlative. We also extract two constant time slices at $t = 0.32s$ and $t = 0.64s$, respectively, and show them in Figures 22 and 23. The pixels corresponding to the proposed method are obviously smoother from the proposed method. The two reconstructed time slices from the proposed method show obvious shapes of a dome.

Table 1: SNRs comparison in dB for the synthetic example with linear events.

	Incomplete	RR	DRR	ORR
N=3	-4.59	9.96	11.68	12.03
N=5	-4.59	8.01	10.87	11.99
N=10	-4.59	5.51	9.54	11.83

Table 2: Computing time comparison in seconds for the synthetic example with linear events. The computation is done on a MacBook Pro laptop equipped with an Intel Core i7 CPU clocked at 2.5 GHz and 16 GB of RAM.

	RR	DRR	ORR
N=3	246.39	249.63	656.33
N=5	249.37	252.52	923.45
N=10	250.28	255.27	1226.63

Table 3: SNRs comparison in dB for the synthetic example with hyperbolic events.

	Incomplete	RR	DRR	ORR
N=12	0.59	14.11	16.46	16.86
N=24	0.59	13.88	16.01	16.83

CONCLUSIONS

We have introduced a new rank-reduction (RR) method for interpolating and denoising five-dimensional seismic data based on a cascaded optimal weighting and damping operations. The proposed optimally damped rank-reduction (ORR) method can further improve the damped rank-reduction (DRR) method in causing less residual noise and making the result smoother. The proposed ORR method works effectively in various data examples including data with linear/planar events, data with hyperbolic events, and field data. While in the case of low SNR, all rank-reduction methods tend to damage some useful energy, the proposed ORR method does not cause extra damage but can remove more noise than the other rank-reduction methods. The waveforms reconstructed from the proposed method are more similar to the ground-truth solution in the synthetic tests. The block Hankel matrix from the proposed method in each frequency slice is smoother than other rank-reduction method, which accounts for why it is more capable of suppressing residual noise. The proposed method outperforms the other rank-reduction methods more in stronger noise and higher sampling ratios. The effectiveness is almost unchanged for different predefined ranks, meaning that the proposed method is almsot parameter-free. One drawback of the proposed method is that the computational cost is 2-5 times higher than the other rank-reduction methods, which can be potentially tackled in future research.

ACKNOWLEDGEMENTS

We would like to thank Dong Zhang and Weilin Huang for inspiring discussions. We also thank Frederik Simons and two anonymous reviewers for constructive suggestions that greatly improved the manuscript. The research is supported by the starting fund from Zhejiang University.

REFERENCES

Aharchaou, M., J. Anderson, S. Hughes, and J. Bingham, 2017, Singular-spectrum analysis via optimal shrinkage of singular values, *in* SEG Technical Program Expanded Abstracts 2017: Society of Exploration Geophysicists, 4950–4954.

Anvari, R., M. A. N. Siahsar, S. Gholtashi, A. R. Kahoo, and M. Mohammadi, 2017, Seismic random noise attenuation using synchrosqueezed wavelet transform and low-rank signal matrix approximation: IEEE Transactions on Geoscience and Remote Sensing, **55**, no. 11, 6574–6581.

Bai, M., and J. Wu, 2018, Seismic deconvolution using iteartive transform-domain sparse inversion: Journal of Seismic Exploration, **27**, no. 2, 103–116.

———, 2019, Iterative sparse deconvolution using seislet-domain constraint: Journal of Seismic Exploration, **28**, no. 1, 73–88.

Bai, M., J. Wu, J. Xie, and D. Zhang, 2018a, Least-squares reverse time migration

of blended data with low-rank constraint along structural direction: Journal of Seismic Exploration, **27**, no. 1, 29–48.

Bai, M., J. Wu, and H. Zhang, 2019, Iterative deblending of simultaneous-source data using smoothed singular spectrum analysis: Journal of Applied Geophysics, **161**, 261–269.

Bai, M., J. Wu, S. Zu, and W. Chen, 2018b, A structural rank reduction operator for removing artifacts in least-squares reverse time migration: Computers and Geosciences, **117**, 9–20.

Benaych-Georges, F., and R. R. Nadakuditi, 2012, The singular values and vectors of low rank perturbations of large rectangular random matrices: Journal of Multivariate Analysis, **111**, 120–135.

Beylkin, G., 1987, Discrete radon transform: IEEE transactions on acoustics, speech, and signal processing, **35**, no. 2, 162–172.

Bucha, V., 2017, Kirchhoff prestack depth migration in simple models with differently rotated elasticity tensor: Orthorhombic and triclinic anisotropy: Journal of Seismic Exploration, **26**, no. 1, 1–24.

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: SIAM, Multiscale Modeling and Simulation, **5**, 861–899.

Canning, A., and G. H. F. Gardner, 1996, Reducing 3D acquisition footprint for 3D dmo and 3D prestack migration: Geophysics, **63**, 1177–1183.

Chen, G., L. Y. Fu, K. F. Chen, W. Sun, W. Wei, and X. Guan, 2017a, Calculation of the seismic imaging complexity of complex geological structures: Journal of Seismic Exploration, **26**, no. 1, 81–104.

Chen, W., and H. Song, 2018, Automatic noise attenuation based on clustering and empirical wavelet transform: Journal of Applied Geophysics, **159**, 649–665.

Chen, Y., 2017, Fast dictionary learning for noise attenuation of multidimensional seismic data: Geophysical Journal International, **209**, no. 1, 21–31.

Chen, Y., H. Chen, K. Xiang, and X. Chen, 2016a, Geological structure guided well log interpolation for high-fidelity full waveform inversion: Geophysical Journal International, **207**, no. 2, 1313–1331.

———, 2017b, Preserving the discontinuities in least-squares reverse time migration of simultaneous-source data: Geophysics, **82**, no. 3, S185–S196.

Chen, Y., X. Chen, Y. Wang, and S. Zu, 2019, The interpolation of sparse geophysical data: Surveys in Geophysics, **40**, 73–105.

Chen, Y., and S. Fomel, 2015, Random noise attenuation using local signal-and-noise orthogonalization: Geophysics, **80**, no. 6, WD1–WD9.

———, 2018, EMD-seislet transform: Geophysics, **83**, no. 1, A27–A32.

Chen, Y., S. Fomel, and J. Hu, 2014, Iterative deblending of simultaneous-source seismic data using seislet-domain shaping regularization: Geophysics, **79**, no. 5, V179–V189.

Chen, Y., J. Ma, and S. Fomel, 2016b, Double-sparsity dictionary for seismic noise attenuation: Geophysics, **81**, V17–V30.

Chen, Y., D. Zhang, Z. Jin, X. Chen, S. Zu, W. Huang, and S. Gan, 2016c, Simultaneous denoising and reconstruction of 5D seismic data via damped rank-reduction method: Geophysical Journal International, **206**, no. 3, 1695–1717.

Cheng, J., and M. Sacchi, 2014, Fast dual-domain reduced-rank algorithm for 3D deblending via randomized qr decomposition: Geophysics, **81**, no. 1, V89–V101.

———, 2016, Fast and memory-efficient singular spectrum analysis for seismic data reconstruction and denoising, *in* SEG Technical Program Expanded Abstracts 2016: Society of Exploration Geophysicists, 4064–4068.

Chopra, S., and K. J. Marfurt, 2013, Preconditioning seismic data with 5D interpolation for computing geometric attributes: The Leading Edge, **32**, 1456–1460.

Downton, J., B. Durrani, L. Hunt, S. Hadley, and M. Hadley, 2008, 5D interpolation, pstm and avo inversion: SEG Technical Program Expanded Abstracts, 237–241.

Ely, G., S. Aeron, N. Hao, and M. E. Kilmer, 2015, 5D seismic data completion and denoising using a novel class of tensor decompositions: Geophysics, **80**, V83–V95.

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

———, 2003, Seismic reflection data interpolation with differential offset and shot continuation: Geophysics, **68**, 733–744.

Gan, S., S. Wang, Y. Chen, and X. Chen, 2015a, Deblending of distance separated simultaneous-source data using seislet frames in the shot domain: SEG expanded abstracts: 85th Annual international meeting, 65–70.

———, 2015b, Seismic data reconstruction via fast projection onto convex sets in the seislet transform domain: SEG expanded abstracts: 85th Annual international meeting, 3814–3819.

Gan, S., S. Wang, Y. Chen, S. Qu, and S. Zu, 2016, Velocity analysis of simultaneous-source data using high-resolution semblance-coping with the strong noise: Geophysical Journal International, **204**, 768–779.

Gan, S., S. Wang, Y. Chen, Y. Zhang, and Z. Jin, 2015c, Dealiased seismic data interpolation using seislet transform with low-frequency constraint: IEEE Geoscience and Remote Sensing Letters, **12**, no. 10, 2150–2154.

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.

Gavish, M., and D. L. Donoho, 2014, The optimal hard threshold for singular values is 4/sqrt(3): IEEE Transactions on Information Theory, **60**, 5040–5053.

Gilles, J., 2013, Empirical wavelet transform: IEEE Transaction on Signal Process, **61**, 3999–4010.

Huang, W., R. Wang, Y. Chen, H. Li, and S. Gan, 2016, Damped multichannel singular spectrum analysis for 3D random noise attenuation: Geophysics, **81**, no. 4, V261–V270.

Huang, W., R. Wang, S. Zu, and Y. Chen, 2017, Low-frequency noise attenuation in seismic and microseismic data using mathematical morphological filtering: Geophysical Journal International, **211**, 13181340.

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

Jia, Y., S. Yu, and J. Ma, 2018, Intelligent interpolation by monte carlo machine learning: Geophysics, **83**, V83–V97.

Jin, S., 2010, 5D seismic data regularization by a damped least-norm Fourier inver-

sion: Geophysics, **75**, WB103–WB111.

Kim, A., D. Ryu, and W. Ha, 2017, An efficient time-domain full waveform inversion using the excitation amplitude method: Journal of Seismic Exploration, **26**, 481–498.

Kreimer, N., A. Stanton, and M. D. Sacchi, 2013, Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction: Geophysics, **78**, V273–V284.

Liu, B., and M. D. Sacchi, 2004, Minimum weighted norm interpolation of seismic records: Geophysics, **69**, 1560–1568.

Liu, W., S. Cao, and Y. Chen, 2016a, Seismic time-frequency analysis via empirical wavelet transform: IEEE Geoscience and Remote Sensing Letters, **13**, 28–32.

Liu, W., S. Cao, S. Gan, Y. Chen, S. Zu, and Z. Jin, 2016b, One-step slope estimation for dealiased seismic data reconstruction via iterative seislet thresholding: IEEE Geoscience and Remote Sensing Letters, **13**, no. 10, 1462–1466.

Lu, B., and J. Feng, 2017, Coal working face imaging by seismic interferometry-using conveyer belt noise as source: Journal of Seismic Exploration, **26**, 411–432.

Lu, L., W. Xu, and S. Qiao, 2015, A fast svd for multilevel block hankel matrices with minimal memory storage: Numerical Algorithms, **69**, 875–891.

Ma, J., and F.-X. L. Dimet, 2009, Deblurring from highly incomplete measurements for remote sensing: IEEE Transactions on Geoscience and Remote Sensing, **47**, no. 3, 792–802.

Mousavi, S. M., and C. A. Langston, 2016, Hybrid seismic denoising using higher-order statistics and improved wavelet block thresholding: Bulletin of the Seismological Society of America, **106**, no. 4, 1380–1393.

Nadakuditi, R. R., 2013, Optshrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage: IEEE Transactions on Information Theory, **60**, 3002–3018.

Naghizadeh, M., and M. D. Sacchi, 2007, Multistep autoregressive reconstruction of seismic records: Geophysics, **72**, V111–V118.

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

Qu, S., D. Verschuur, and Y. Chen, 2016, Full waveform inversion using an automatic directional total variation constraint: 78th Annual International Conference and Exhibition, EAGE, Extended Abstracts, DOI: 10.3997/2214–4609.201701340.

Rioul, O., and M. Vetterli, 1991, Wavelets and signal processing: IEEE signal processing magazine, **8**, 14–38.

Ronen, J., 1987, Wave-equation trace interpolation: Geophysics, **52**, 973–984.

Siahsar, M. A. N., V. Abolghasemi, and Y. Chen, 2017a, Simultaneous denoising and interpolation of 2D seismic data using data-driven non-negative dictionary learning: Signal Processing, **141**, 309–321.

Siahsar, M. A. N., S. Gholtashi, A. R. Kahoo, W. Chen, and Y. Chen, 2017b, Data-driven multi-task sparse dictionary learning for noise attenuation of 3D seismic data: Geophysics, **82**, no. 6, V385V396.

Siahsar, M. A. N., S. Gholtashi, E. Olyaei, W. Chen, and Y. Chen, 2017c, Simultane-

ous denoising and interpolation of 3D seismic data via damped data-driven optimal singular value shrinkage: IEEE Geoscience and Remote Sensing Letters, **14**, no. 7, 1086–1090.

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

Tian, W., Y. Wang, T. Zhou, C. Xie, Z. Guan, H. Liu, and K. Xu, 2017, Characteristics of high-frequency ultra-acoustic wave spectrum and pore size in low-permeability sandstone: Journal of Seismic Exploration, **26**, 399–410.

Trad, D., 2007, A strategy for wide azimuth land interpolation: SEG Technical Program Expanded Abstracts, 946–950.

———, 2009, Five-dimensional interpolation: Recovering from acquisition constraints: Geophysics, **74**, V123–V132.

Trickett, S., 2015, Preserving signal: Automatic rank determination for noise suppression, *in* SEG Technical Program Expanded Abstracts 2015: Society of Exploration Geophysicists, 4703–4707.

Trickett, S., and L. Burroughs, 2009, Prestack rank reducing noise suppression: Theory: SEG Technical Program Expanded Abstracts, 3332–3336.

Wang, B., R. Wu, X. Chen, and J. Li, 2015, Simultaneous seismic data interpolation and denoising with a new adaptive method based on dreamlet transform: Geophysical Journal International, **201**, 1180–1192.

Wang, Y., H. Zhou, H. Chen, and Y. Chen, 2018, Adaptive stabilization for Q-compensated reverse time migration: Geophysics, **83**, no. 1, S15–S32.

Wu, J., and M. Bai, 2018a, Adaptive rank-reduction method for seismic data reconstruction: Journal of Geophysics and Engineering, **15**, 1688.

———, 2018b, Attenuating seismic noise via incoherent dictionary learning: Journal of Geophysics and Engineering, **15**, 1327.

———, 2018c, Fast principal component analysis for stacking seismic data: Journal of Geophysics and Engineering, **15**, 295–306.

———, 2018d, Incoherent dictionary learning for reducing crosstalk noise in least-squares reverse time migration: Computers and Geosciences, **114**, 11–21.

———, 2018e, Stacking seismic data based on principal component analysis: Journal of Seismic Exploration, **27**, 331–348.

———, 2019, Iterative deblending based on the modified singular spectrum analysis: Journal of Seismic Exploration, **28**, no. 1, 1–20.

Xiang, K., X. Chen, Y. Chen, and H. Chen, 2016, Well log interpolation with geological structure constraint for full waveform inversion: 78th Annual International Conference and Exhibition, EAGE, Extended Abstracts, DOI:10.3997/2214–4609.201601548.

Xie, Q., Y. Chen, G. Zhang, S. Gan, and E. Wang, 2015, Seismic data analysis using synchrosqueezeing wavelet transform - a case study applied to boonsville field: 77th Annual International Conference and Exhibition, EAGE, Extended Abstracts, DOI: 10.3997/2214–4609.201412752.

Xie, Y., and D. Gajewski, 2017, 5-D interpolation with wavefront attributes: Geophysical Journal International, **211**, 897–919.

Xue, Z., H. Zhu, and S. Fomel, 2017, Full-waveform inversion using seislet regular-

ization: Geophysics, **82**, no. 5, A43–A49.

Yu, S., J. Ma, X. Zhang, and M. D. Sacchi, 2015, Interpolation and denoising of high-dimensional seismic data by learning a tight frame: Geophysics, **80**, V119–V132.

Zhang, D., Y. Chen, and S. Gan, 2016a, Iterative reconstruction of 3D seismic data via multiple constraints: 78th Annual International Conference and Exhibition, EAGE, Extended Abstracts, DOI: 10.3997/2214–4609.201601241.

———, 2016b, Multidimensional seismic data reconstruction with multiple constraints: 86th Annual International Meeting, SEG, Expanded Abstracts, 4801–4806.

Zhang, D., Y. Chen, W. Huang, and S. Gan, 2016c, Multi-step reconstruction of 3D seismic data via an improved MSSA algorithm: CPS/SEG Beijing 2016 International Geophysical Conference & Exposition, SEG, Expanded Abstracts, 745–749.

Zhang, D., Y. Zhou, H. Chen, W. Chen, S. Zu, , and Y. Chen, 2017, Hybrid rank-sparsity constraint model for simultaneous reconstruction and denoising of 3D seismic data: Geophysics, **82**, no. 5, V351–V367.

Zhang, Q., W. Mao, and Y. Chen, 2019, Attenuating crosstalk noise of simultaneous-source least-squares reverse time migration with gpu-based excitation-amplitude imaging condition: IEEE Transaction on Geosciences and Remote Sensing, **57**, no. 1, 587–597.

Zhang, Q., W. Mao, H. Zhou, H. Zhang, and Y. Chen, 2018, Hybrid-domain simultaneous-source full waveform inversion without crosstalk noise: Geophysical Journal International, **215**, no. 3, 1659–1681.

Zhou, Y., 2017, A POCS method for iterative deblending constrained by a blending mask: Journal of Applied Geophysics, **138**, 245–254.

Zhou, Y., and C. Han, 2018a, Seismic data restoration based on the grassmannian rank-one update subspace estimation method: Journal of Applied Geophysics, **159**, 731–741.

Zhou, Y., C. Han, and Y. Chi, 2018a, Deblending of simultaneous-source data using iterative seislet frame thresholding based on a robust slope estimation: Journal of Applied Geophysics, **138**, 17–37.

Zhou, Y., and W. Han, 2018b, Multiples attenuation in the presence of blending noise: Journal of Seismic Exploration, **27**, no. 1, 69–88.

Zhou, Y., and S. Li, 2018, Simultaneous deblending and interpolation using structure-oriented filters: Journal of Applied Geophysics, **150**, 230–243.

Zhou, Y., S. Li, D. Zhang, and Y. Chen, 2018b, Seismic noise attenuation using an online subspace tracking algorithm: Geophysical Journal International, **212**, no. 2, 10721097.

Zu, S., H. Zhou, Y. Chen, H. Chen, M. Cao, and C. Xie, 2016a, A marine field trial for iterative deblending of simultaneous sources: 86th Annual International Meeting, SEG, Expanded Abstracts, 113–118.

Zu, S., H. Zhou, Y. Chen, Y. Liu, and S. Qu, 2015, A periodically variational dithering code for improving deblending: SEG expanded abstracts: 85th Annual international meeting, 38–42.

Zu, S., H. Zhou, Y. Chen, X. Pan, S. Gan, D. Zhang, and C. Xie, 2016b, Recovering the most from big gaps using least-squares inversion: 86th Annual International Meeting, SEG, Expanded Abstracts, 4128–4133.

Zu, S., H. Zhou, W. Mao, D. Zhang, C. Li, X. Pan, and Y. Chen, 2017, Iterative deblending of simultaneous-source data using a coherency-pass shaping operator: Geophysical Journal International, **211**, no. 1, 541–557.

Zu, S., H. Zhou, R. Ru, M. Jiang, and Y. Chen, 2019, Dictionary learning based on dip patch selection training for random noise attenuation: Geophysics, **84**, no. 3, V169–V183.

Zu, S., H. Zhou, R. Wu, W. Mao, and Y. Chen, 2018, Hybrid-sparsity constrained dictionary learning for iterative deblending of extremely noisy simultaneous-source data: IEEE Transaction on Geosciences and Remote Sensing, **57**, no. 4, 2249–2262.

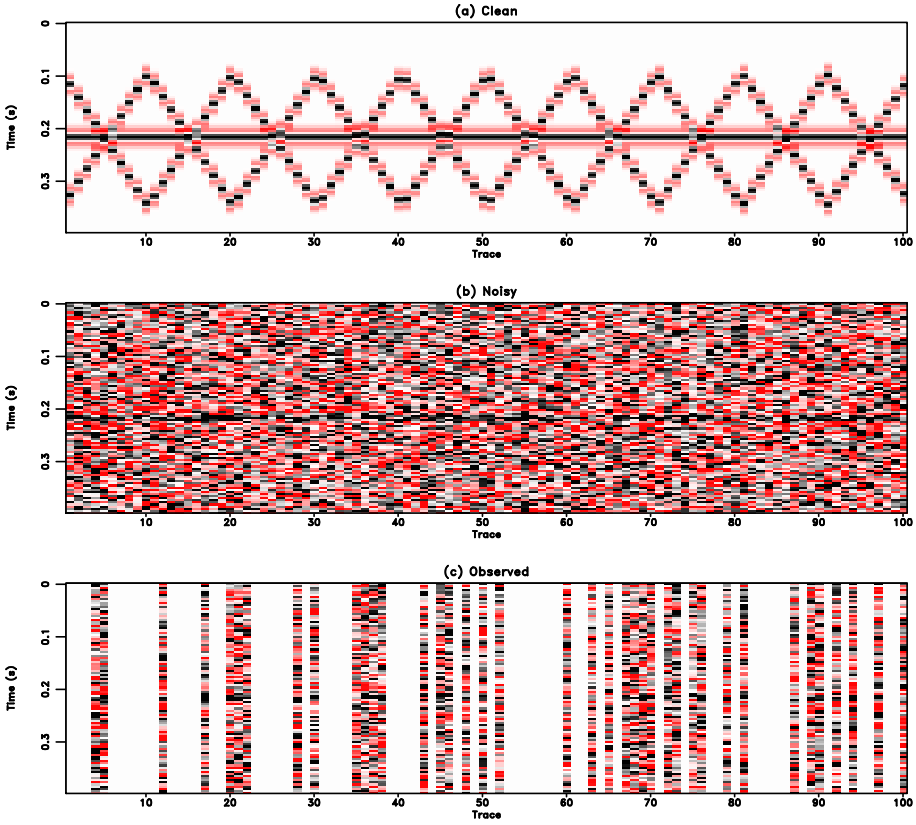

Figure 1: Common offset gather comparison for the synthetic example (reshaped into a 2-D matrix). (a) Clean data. (b) Noisy data. Note that because of the strong random noise, the useful signals are almost buried in the noise. (c) Incomplete data with 70% randomly removed traces. The blanks in (c) indicate where there are missing traces. The strong noise and missing traces make the data quality extremely low.

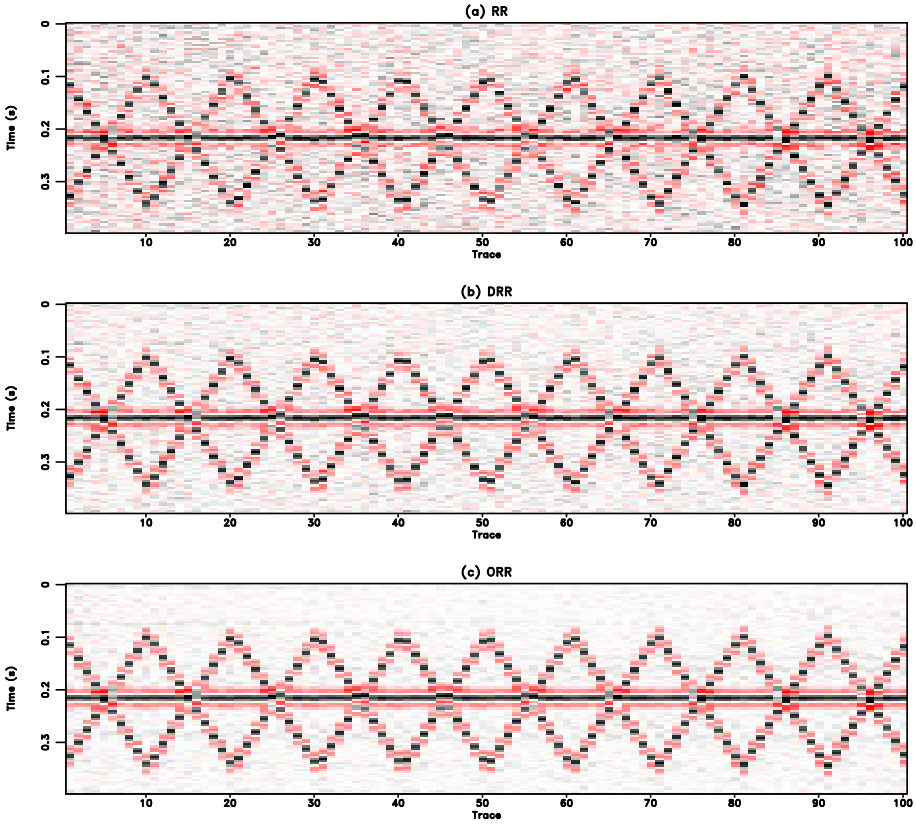

Figure 2: Common offset gather comparison for the synthetic example (reshaped into a 2-D matrix). (a) Reconstructed data using the RR method. (b) Reconstructed data using the DRR method. (c) Reconstructed data using the ORR method. In this case, $rank = 10$.

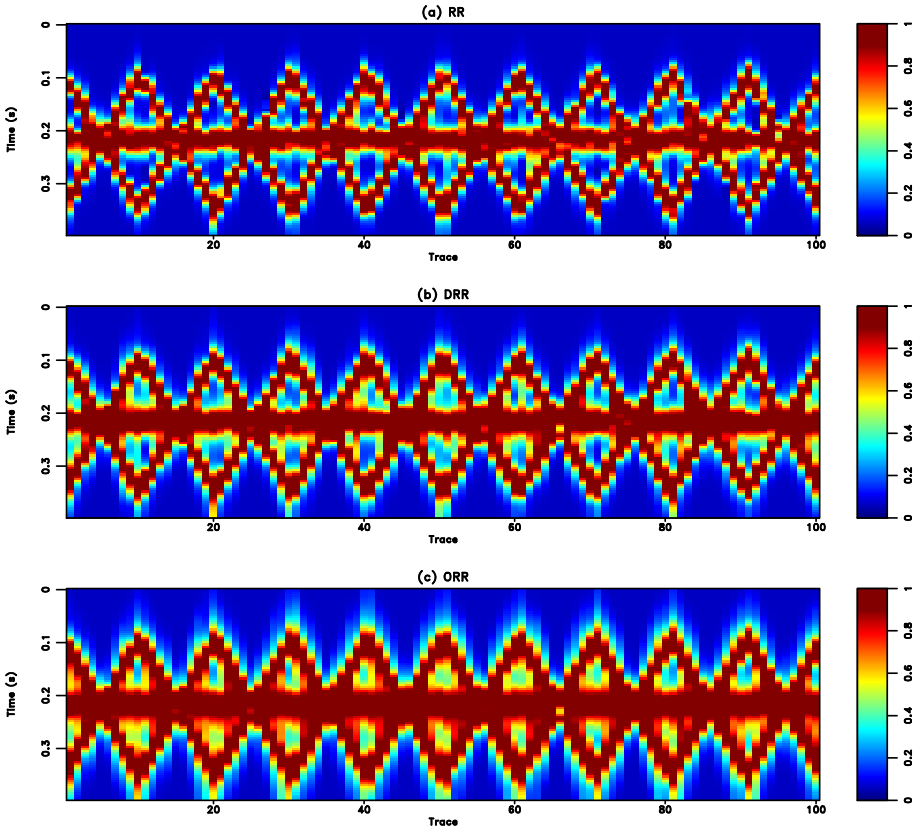

Figure 3: Local similarity comparison for the synthetic example (reshaped into a 2-D matrix). (a) Local similarity using the RR method. (b) Local similarity using the DRR method. (c) Local similarity using the ORR method. It is obvious that the local similarity of the ORR method is much larger than the other two methods, indicating a more accurate reconstruction performance.

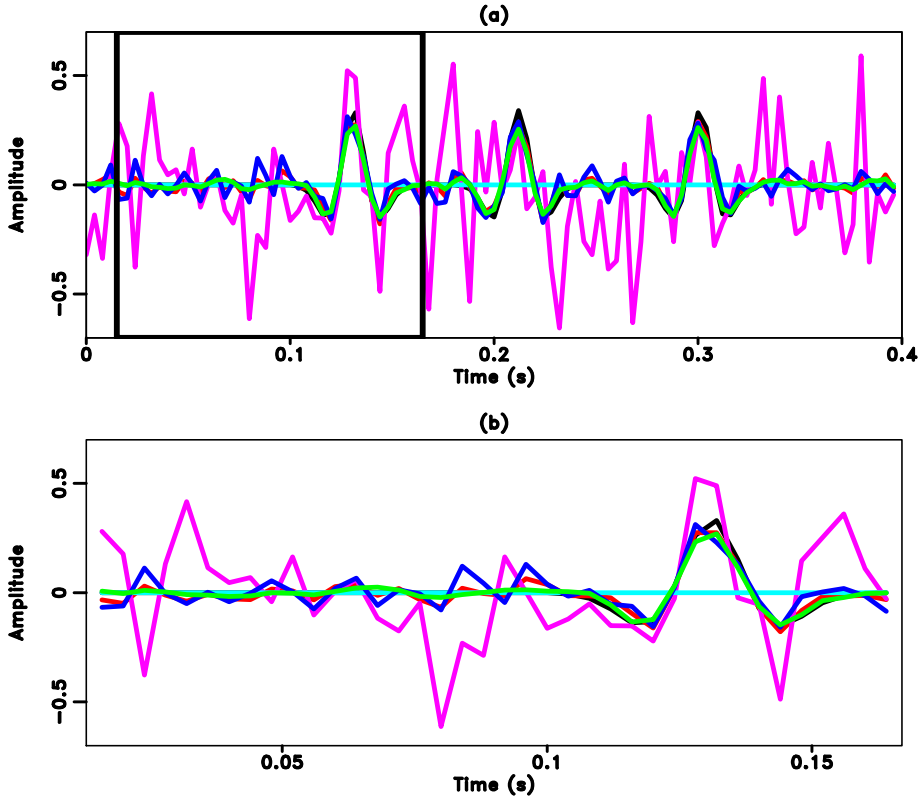

Figure 4: Trace-by-trace comparison. (a) Comparison in the whole trace. The black line denotes the exact solution. Green line denotes the result from the proposed method. Red line denotes the result from the DRR method. Blue line denotes the result from the RR method. (b) Zoomed trace from (a). The transparant red window highlight the zooming area.

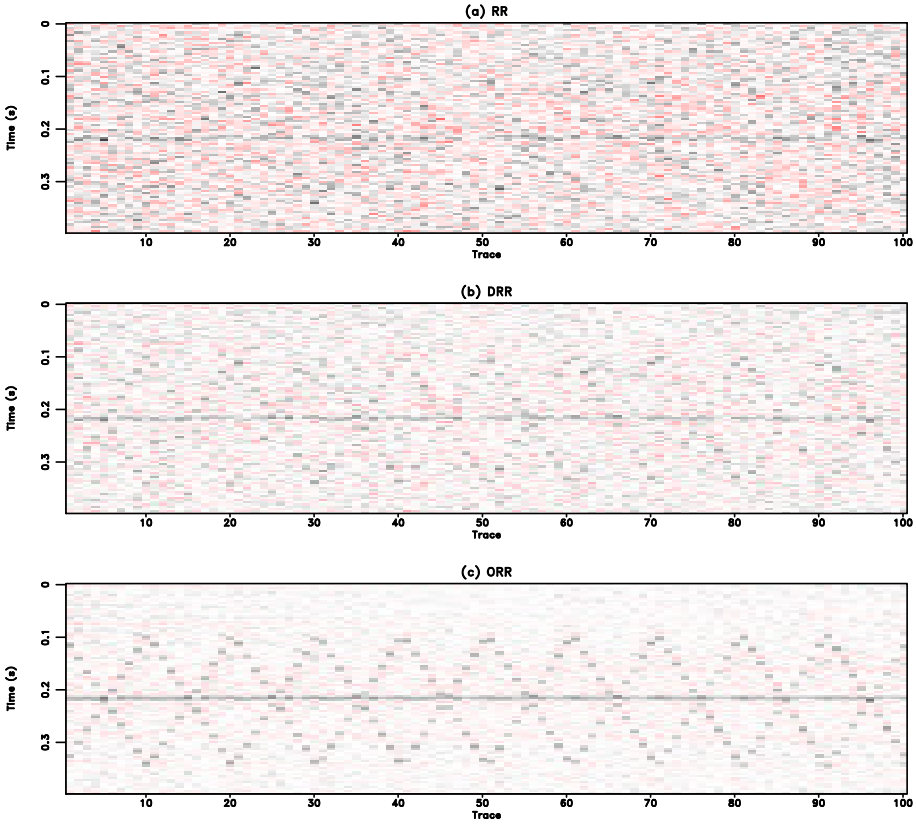

Figure 5: Common offset gather comparison for the synthetic example (reshaped into a 2-D matrix). (a) Reconstruction error using the RR method. (b) Reconstruction error using the DRR method. (c) Reconstruction error using the ORR method. In this case, $rank = 10$.

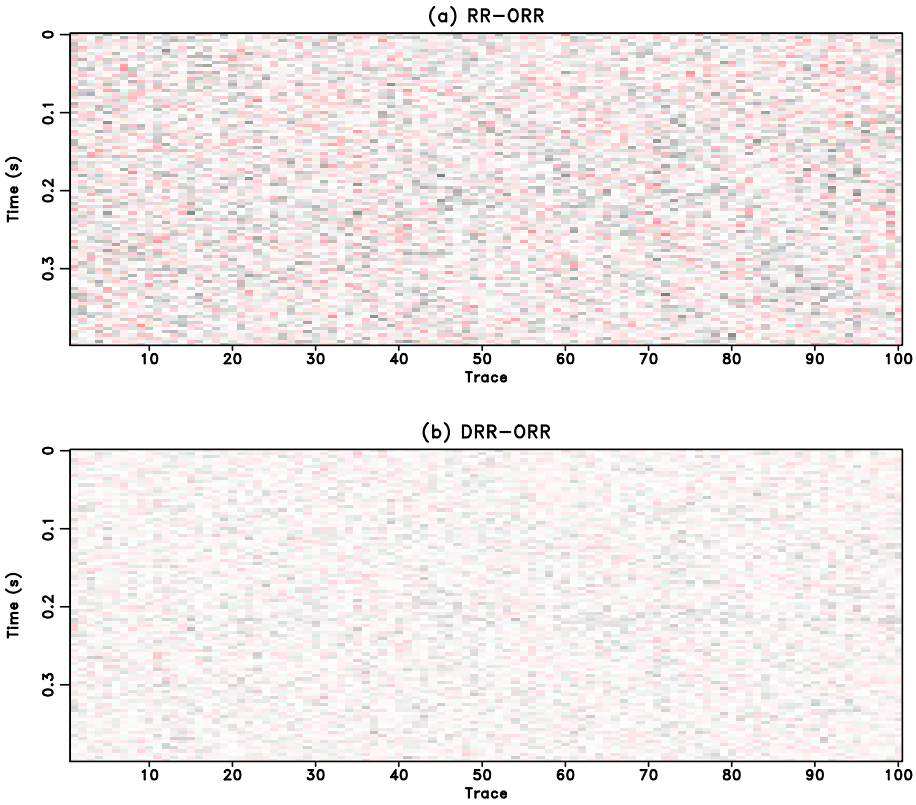

Figure 6: Extra error comparison compared with the ORR method for the synthetic example (reshaped into a 2-D matrix) when $rank = 10$. (a) Extra error caused by RR method (the difference between 5(a) and 5(c)). (b) Extra error caused by RR method (the difference between 5(b) and 5(c)).

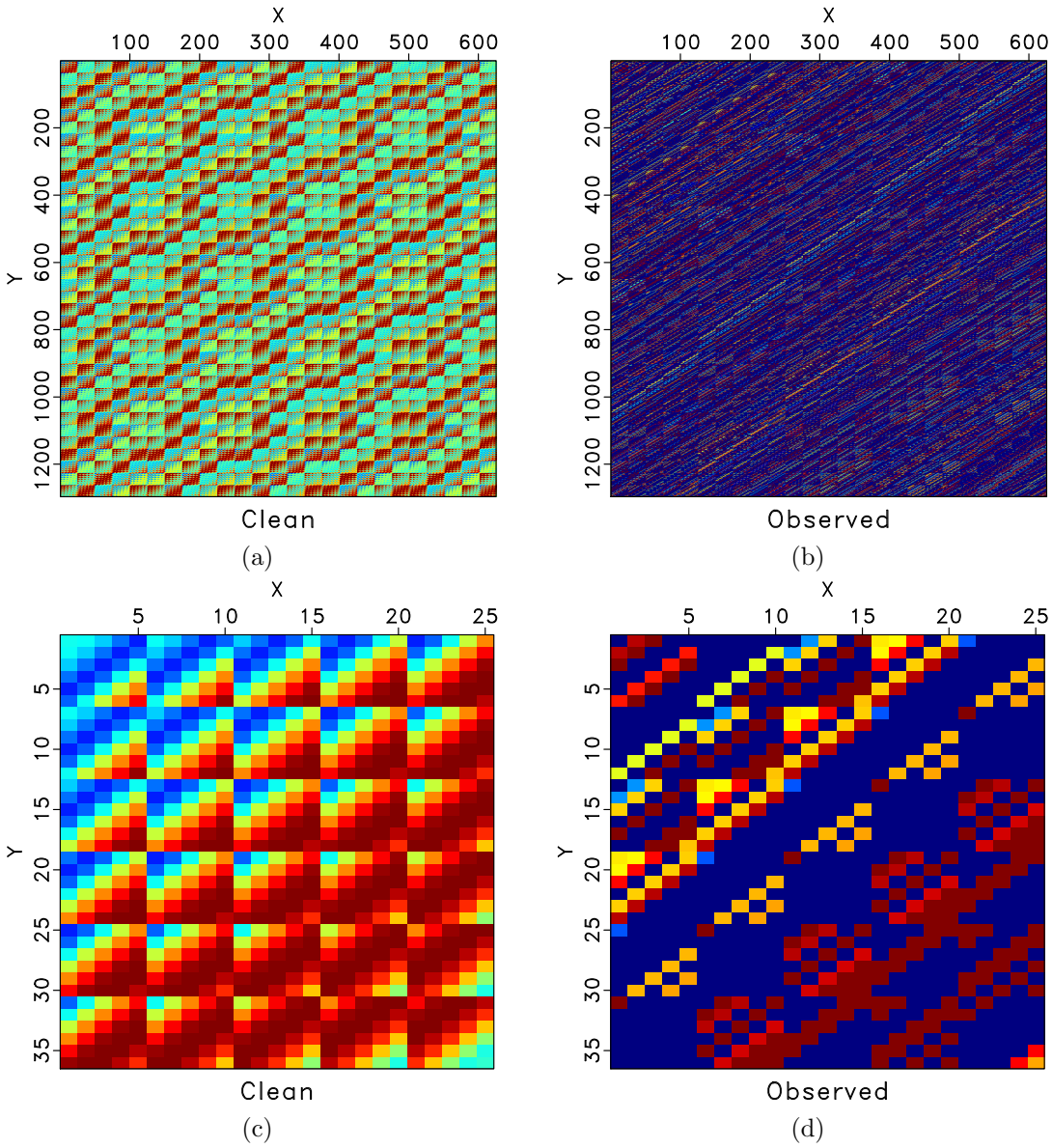

Figure 7: Hankel matrices for the clean and observed data for frequency slice of 30Hz. (a) Hankel matrix of the clean data. (b) Hankel matrix of the observed data. (c) & (d) Zoomed Hankel matrices of (a) & (b). The zooming area is highlighted by the red rectangles.

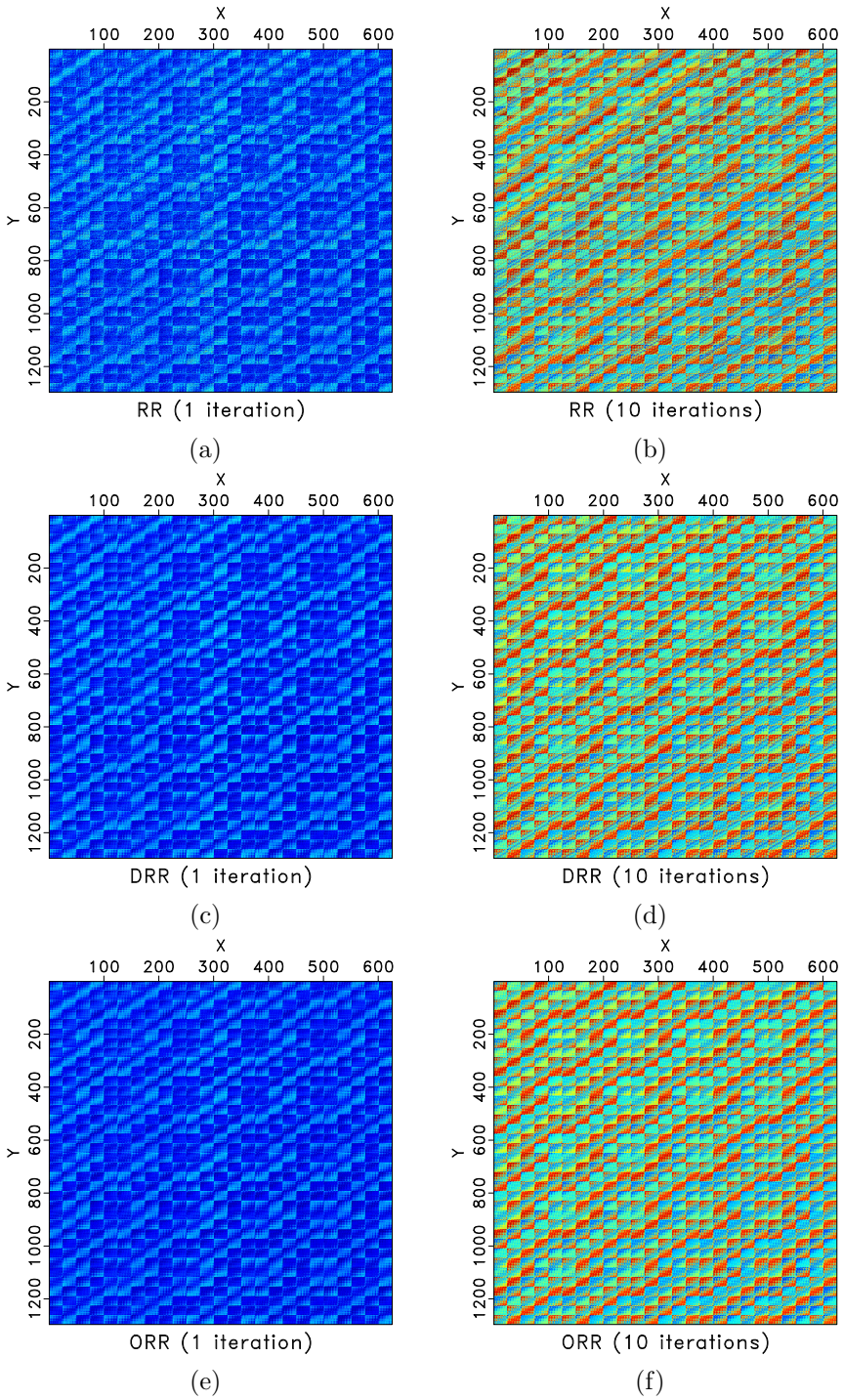

Figure 8: Hankel matrices for different methods for frequency slice of 30Hz. Left column: Hankel matrix after 1 iteration. Right: Hankel matrix after 10 iterations. Top row: Hankel matrices for the RR method. Middle row: Hankel matrix for the DRR method. Bottom row: Hankel matrix for the ORR method.

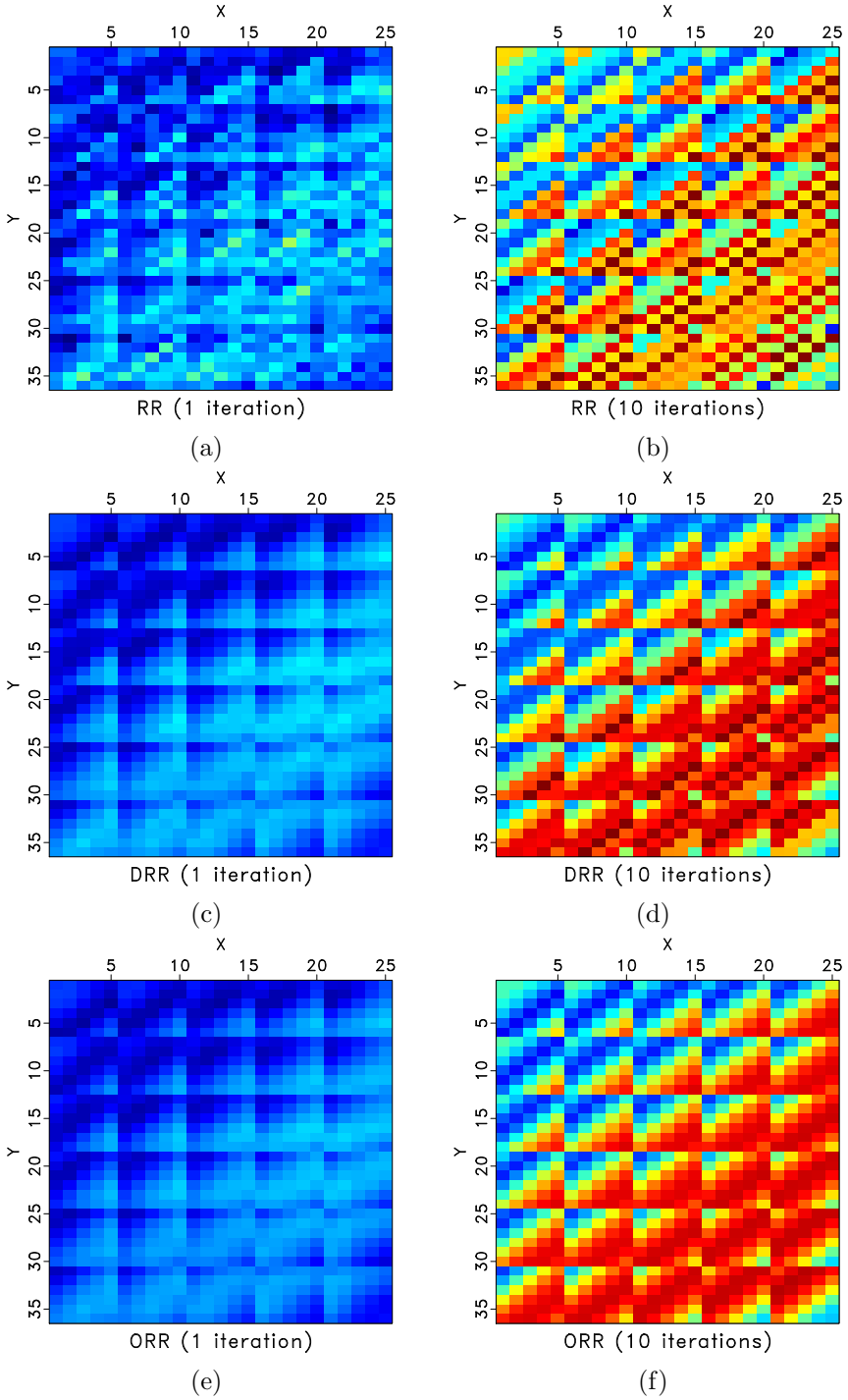

Figure 9: Zoomed Hankel matrices for different methods for frequency slice of 30Hz. Left column: Hankel matrix after 1 iteration. Right: Hankel matrix after 10 iterations. Top row: Hankel matrices for the RR method. Middle row: Hankel matrix for the DRR method. Bottom row: Hankel matrix for the ORR method.

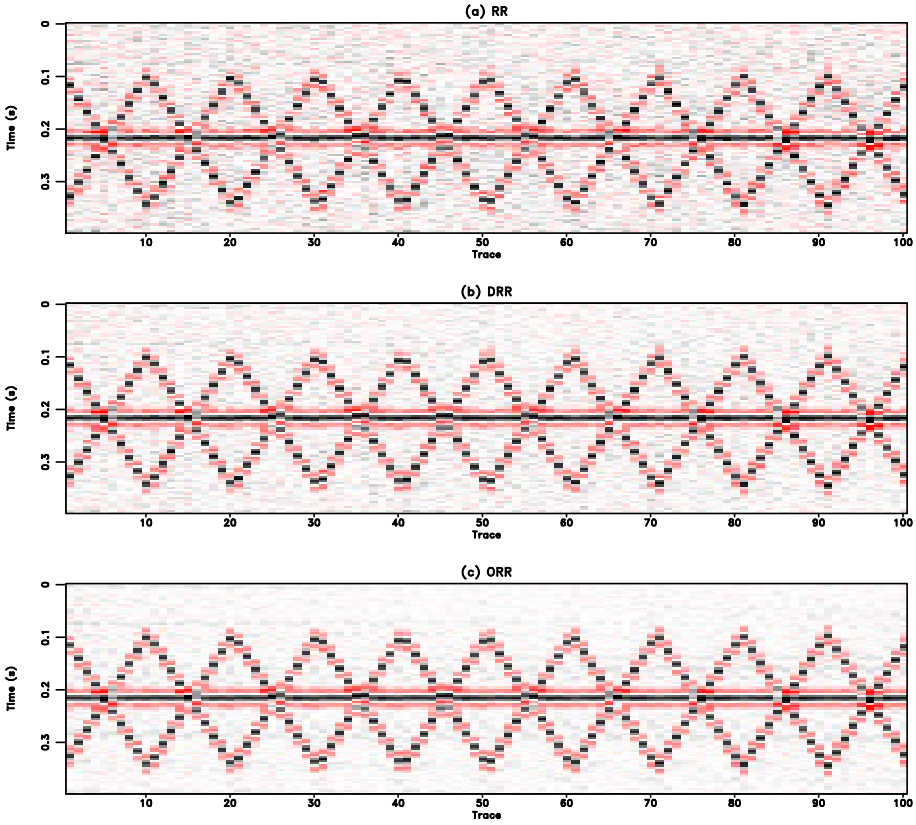

Figure 10: Common offset gather comparison for the synthetic example (reshaped into a 2-D matrix). (a) Reconstructed data using the RR method. (b) Reconstructed data using the DRR method. (c) Reconstructed data using the ORR method. In this case, $rank = 5$.

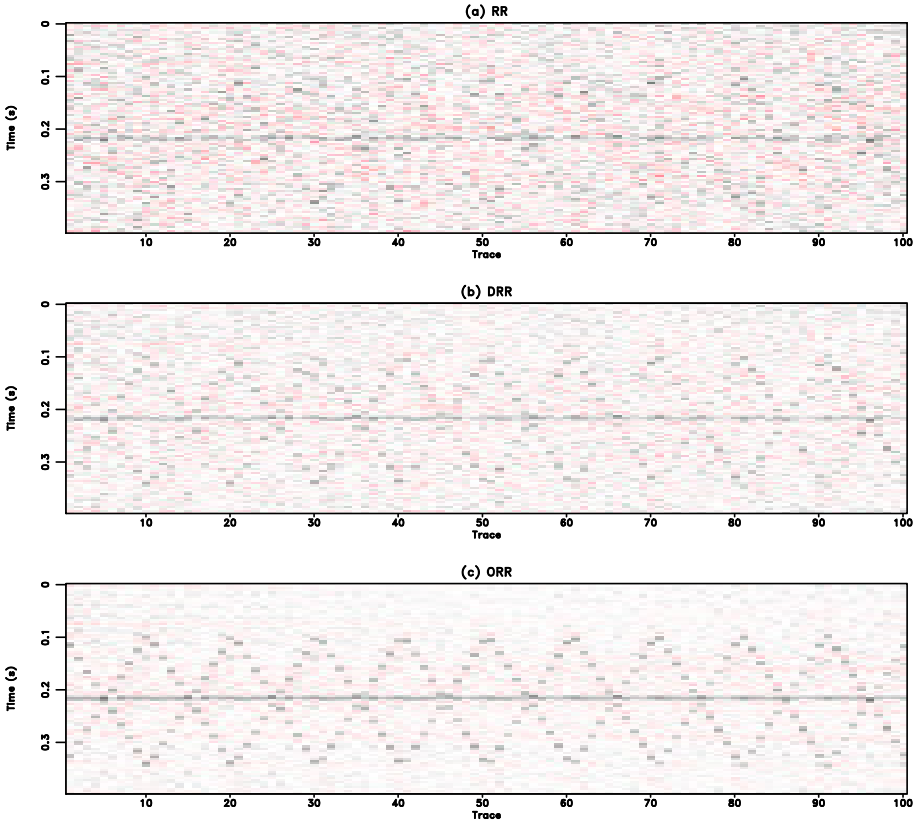

Figure 11: Common offset gather comparison for the synthetic example (reshaped into a 2-D matrix). (a) Reconstruction error using the RR method. (b) Reconstruction error using the DRR method. (c) Reconstruction error using the ORR method. In this case, $rank = 5$.

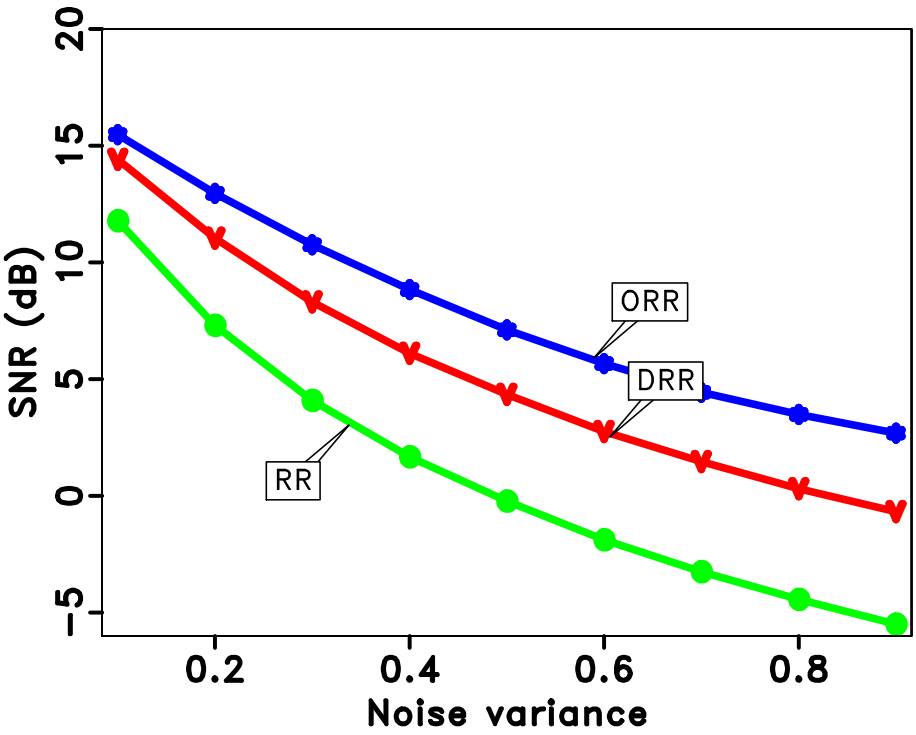

Figure 12: SNR diagrams of the different approaches with respect to the noise level (variance value). Note that the proposed approach outperforms the traditional methods more and more as the noise level increases.

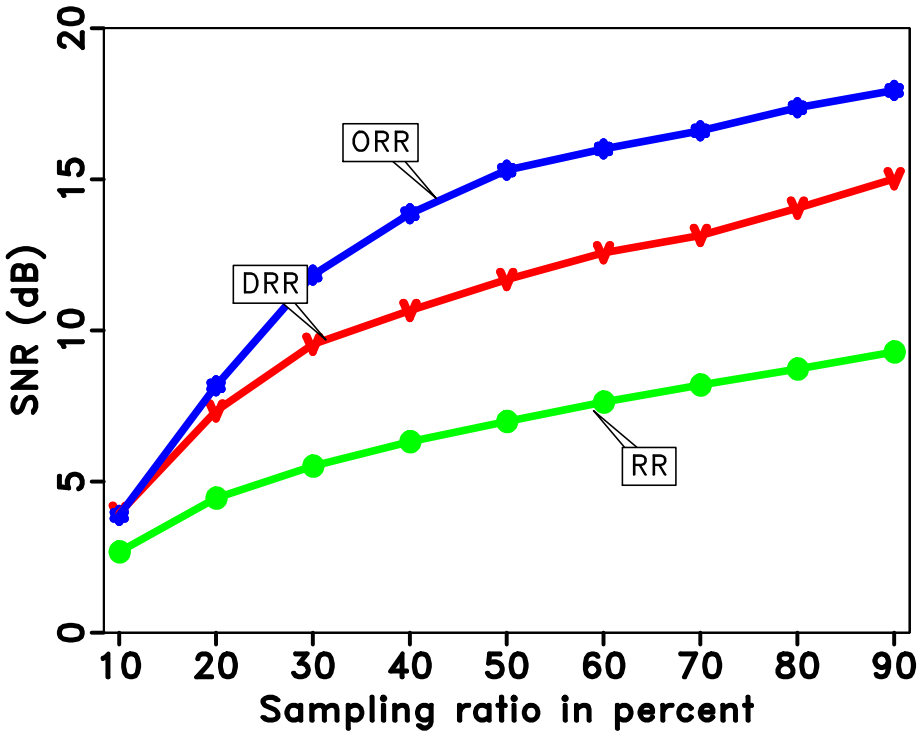

Figure 13: SNR diagrams of the different approaches with respect to the sampling ratio (in percent). It is obvious that the difference between the ORR method and the DRR method becomes larger as the sampling ratio increases. The ORR method always outperforms the other methods for all sampling ratios.

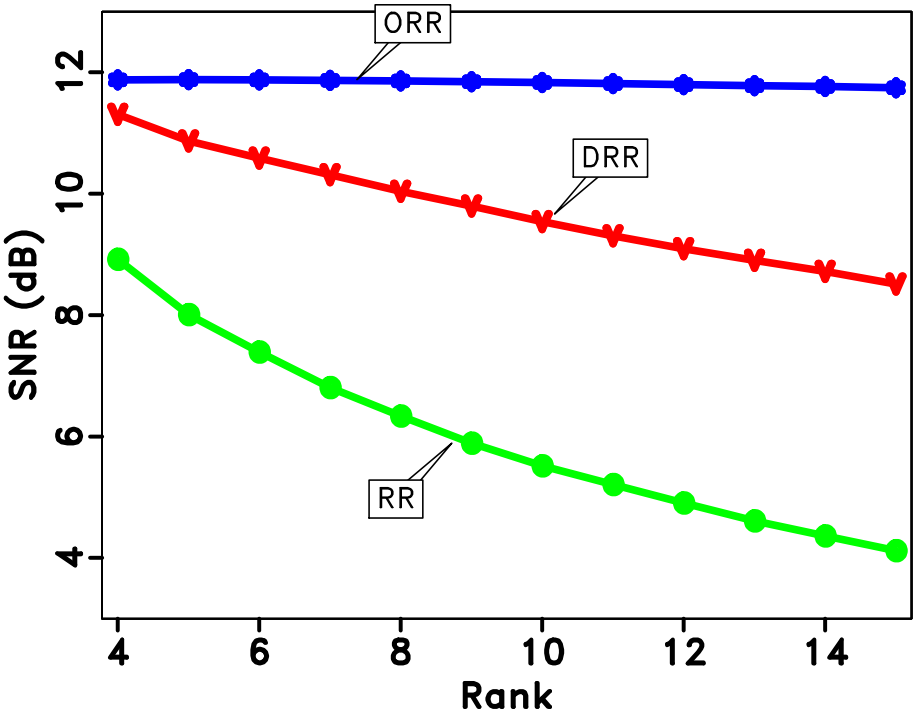

Figure 14: SNR diagrams of the different approaches with respect to the selected rank. The ORR method is obviously much less sensitive to the rank compared with other two methods. This phenomenon indicates that the ORR method can be applied as an adaptive method by setting a relatively large rank.

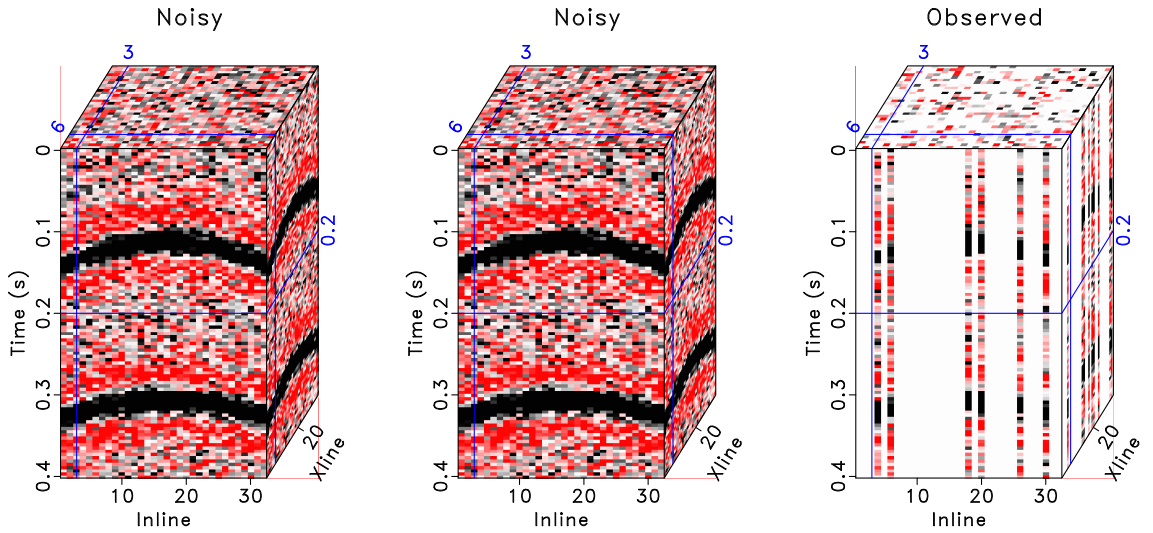

Figure 15: Common midpoint gather comparison for the synthetic example with hyperbolic events. (a) Clean data. (b) Noisy data. (c) Incomplete data. The blanks in (c) indicate where there are missing traces.

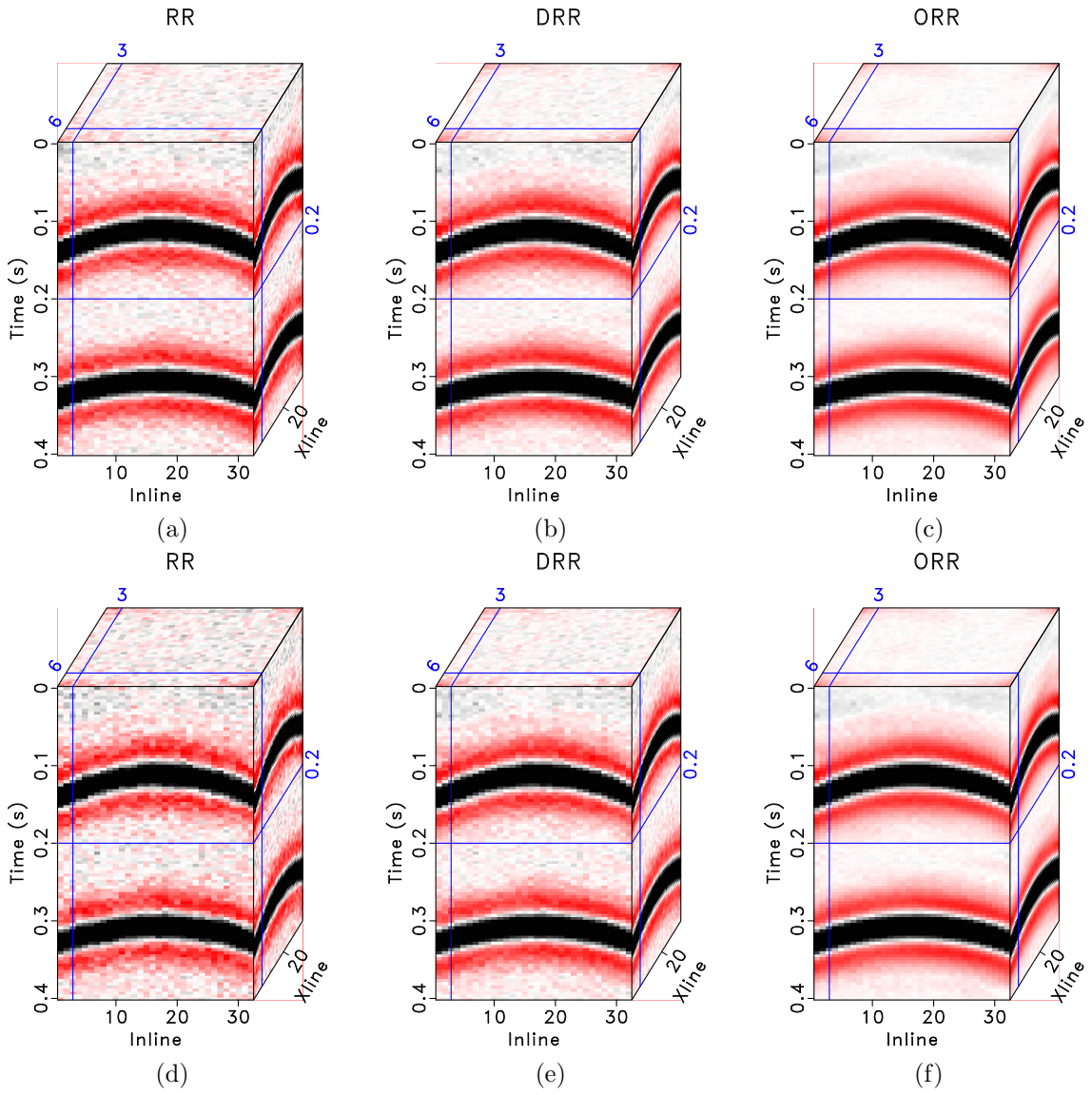

Figure 16: Common midpoint gather comparison for the synthetic example with hyperbolic events. Top row shows results when $rank = 12$. Bottom row shows results when $rank = 24$. Left: Reconstructed data using the RR method. Middle: Reconstructed data using the DRR method. Right Reconstructed data using the ORR method. The SNR comparisons are shown in Table 3.

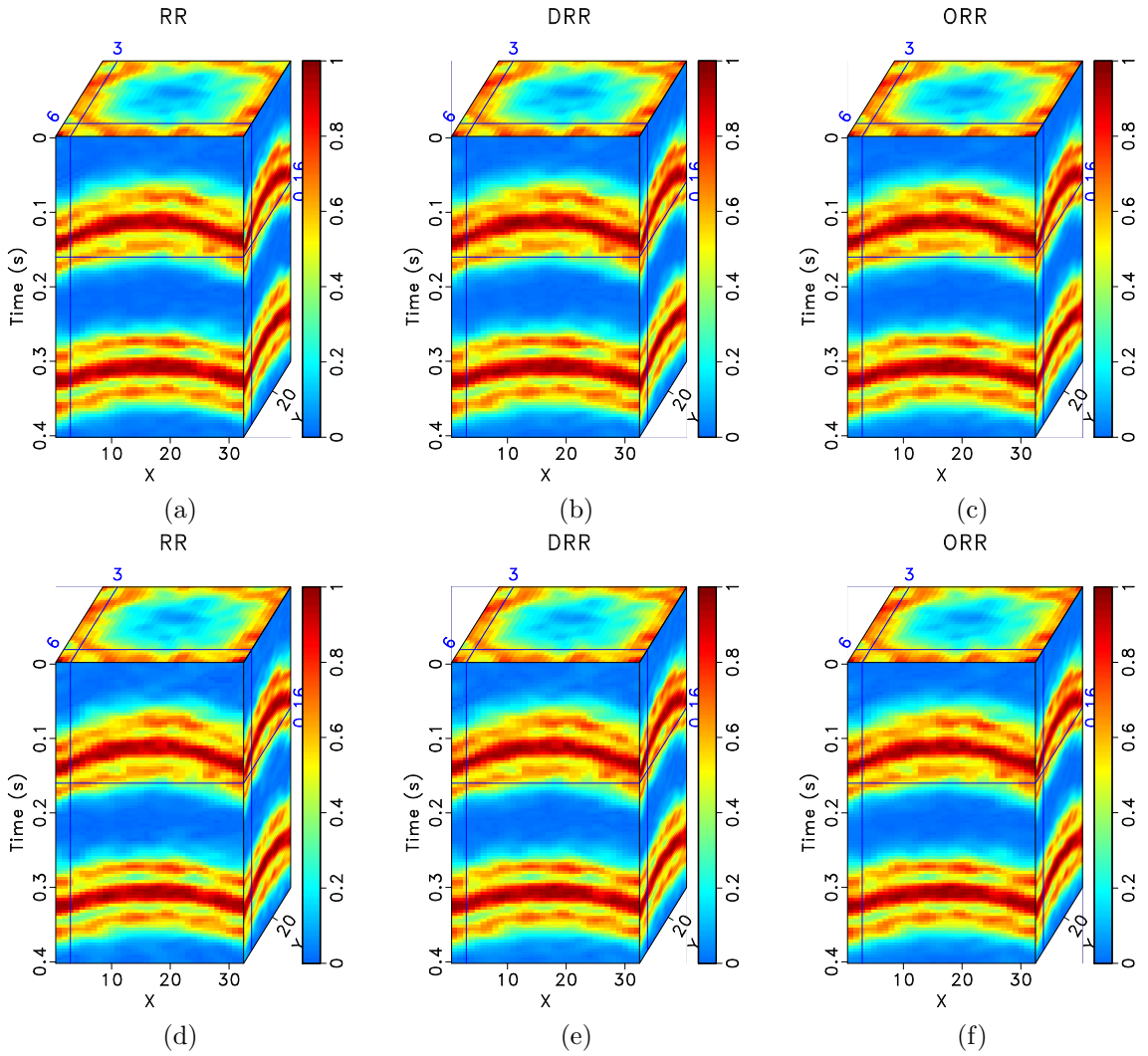

Figure 17: Local similarity comparison for the synthetic example with hyperbolic events. Top row shows results when $rank = 12$. Bottom row shows results when $rank = 24$. Left: Reconstructed data using the RR method. Middle: Reconstructed data using the DRR method. Right Reconstructed data using the ORR method. The introduction of local similarity is given in Chen and Fomel (2015).

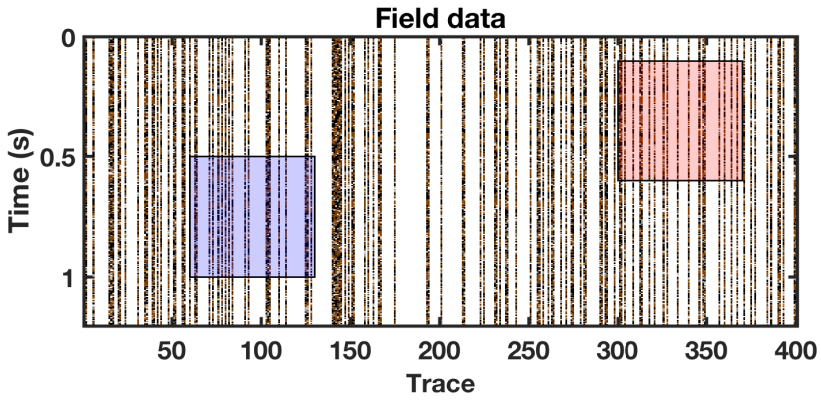

Figure 18: Real data example with a lot of missing traces. Because of the difficulty in display a 5D dataset, only one common midpoint gather is extracted and rearranged into a 2D matrix, and is plotted here. The two transparent colored windows denote two zooming areas for an amplified comparison.

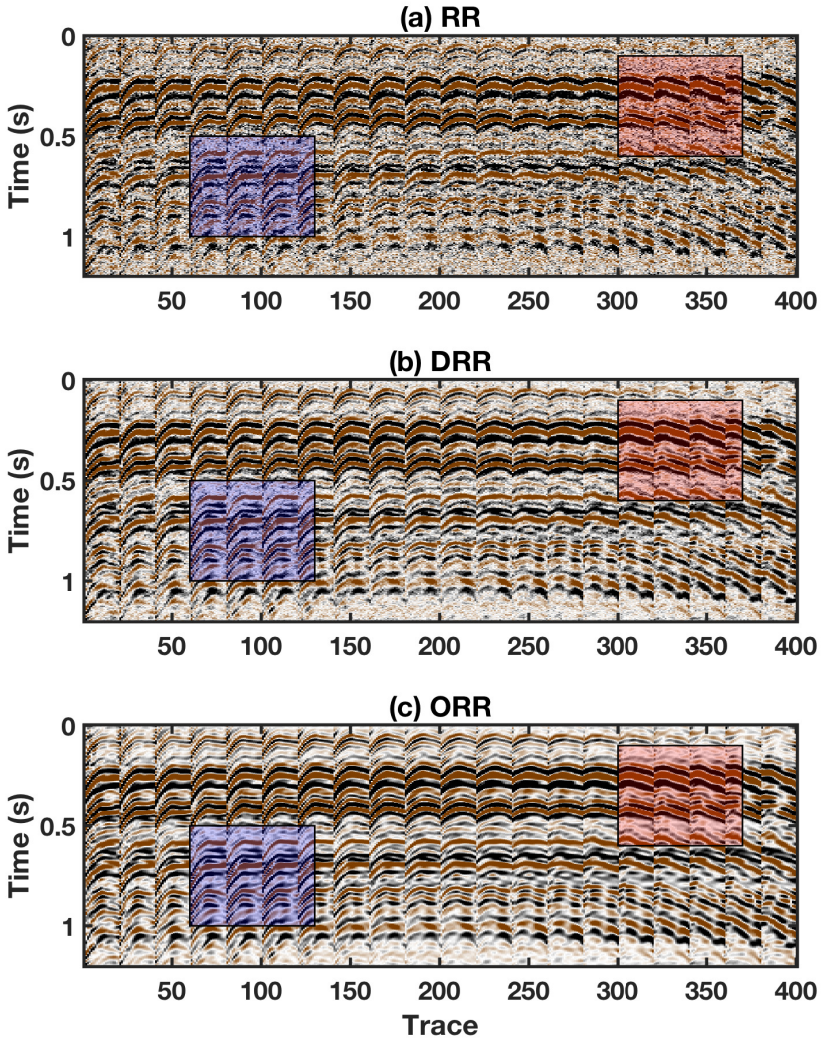

Figure 19: Real data example. (a) Reconstructed data using the RR method. (b) Reconstructed data using the DRR method. (c) Reconstructed data using the ORR method. In this case, $rank = 20$. It is obvious that all three methods obtain dramatic improvement from the raw data. The ORR method obtains cleaner and spatially more coherent seismic events compared with other two methods.

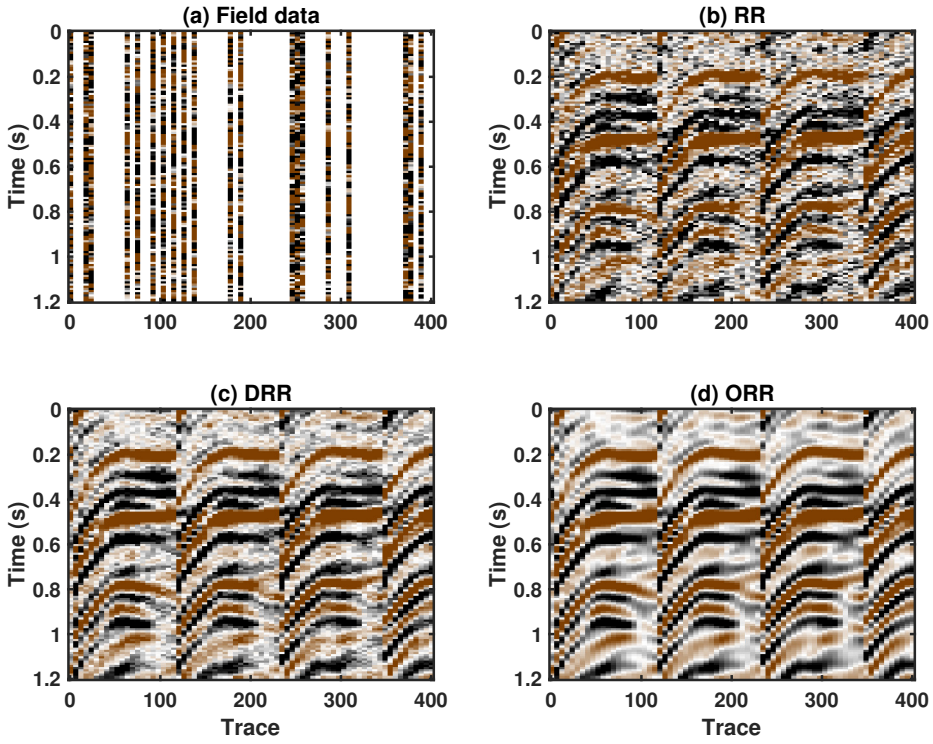

Figure 20: Zoomed comparison for the field data example (the blue frame box). (a) Incomplete data. (b) Reconstructed data using the RR method. (c) Reconstructed data using the DRR method. (d) Reconstructed data using the ORR method.

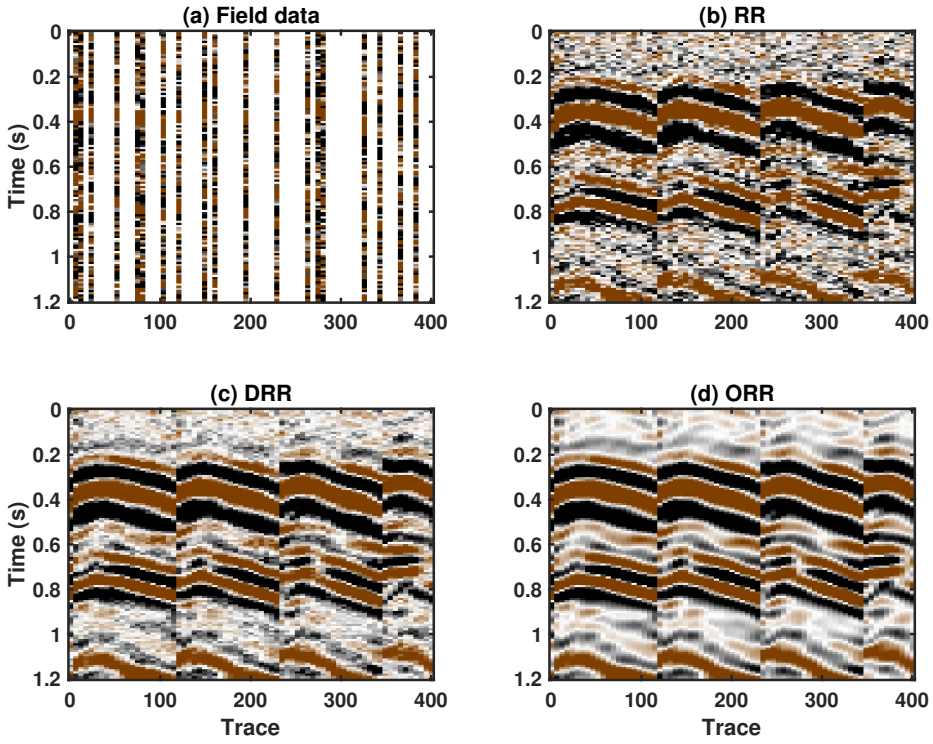

Figure 21: Zoomed comparison for the field data example (the red frame box). (a) Incomplete data. (b) Reconstructed data using the RR method. (c) Reconstructed data using the DRR method. (d) Reconstructed data using the ORR method.

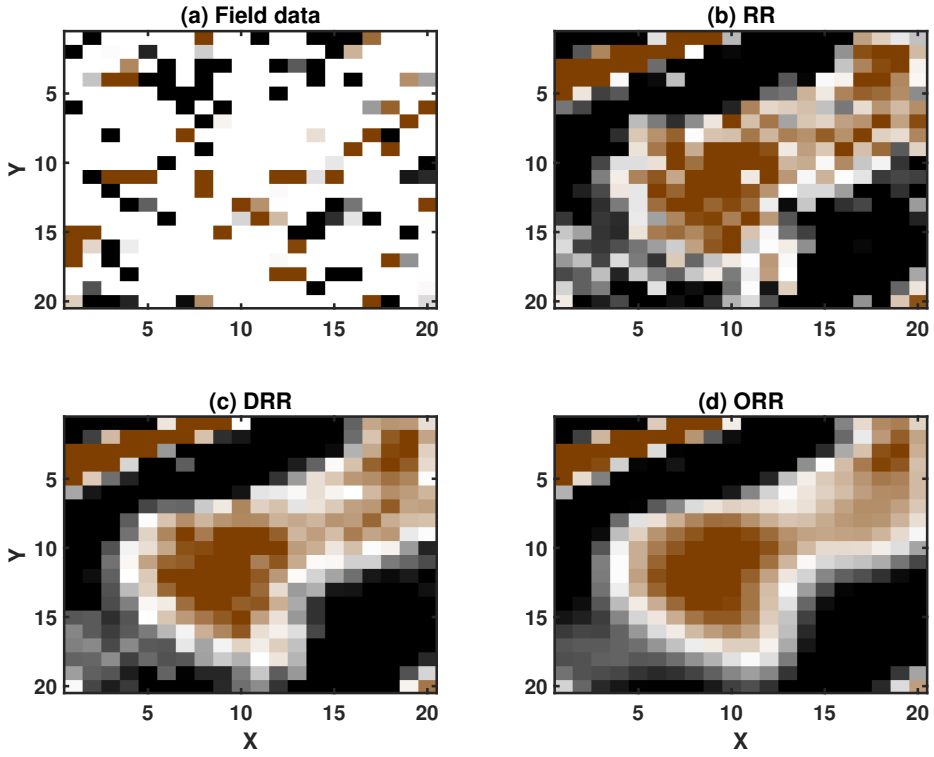

Figure 22: Comparison for constant time slice for the field data example ($t = 0.32$s). (a) Incomplete data. (b) Reconstructed data using the RR method. (c) Reconstructed data using the DRR method. (d) Reconstructed data using the ORR method.

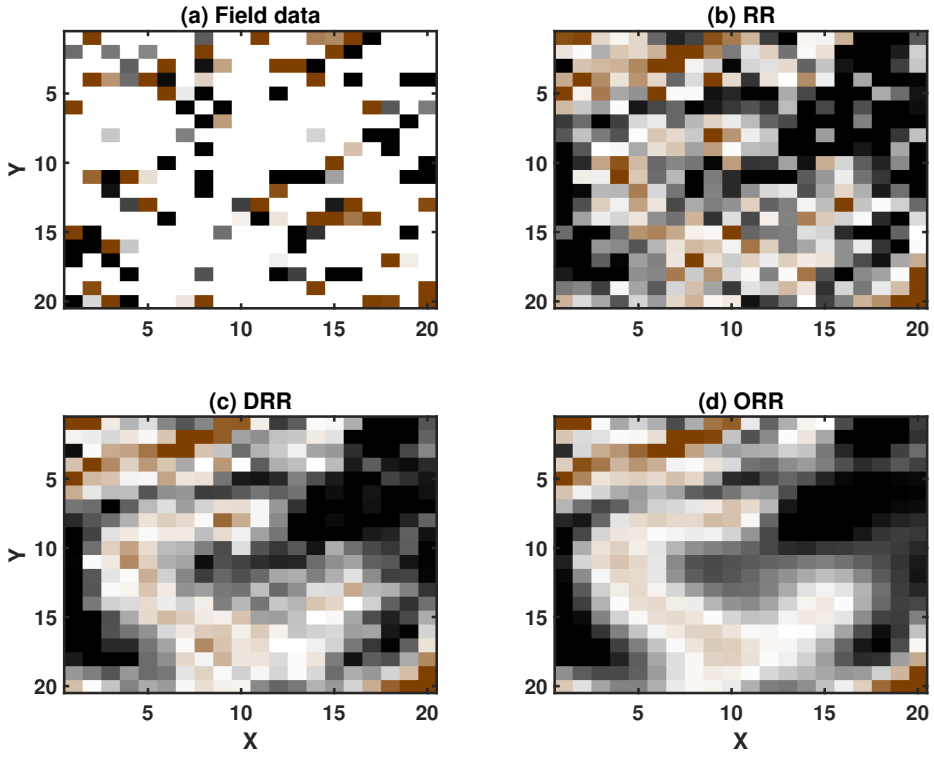

Figure 23: Comparison for constant time slice for the field data example ($t = 0.64$s). (a) Incomplete data. (b) Reconstructed data using the RR method. (c) Reconstructed data using the DRR method. (d) Reconstructed data using the ORR method.