

## **3-D data Interpolation and denoising by adaptive weighting rank-reduction method using singular spectrum analysis algorithm**

Farzaneh Bayati, Daniel Trad

### **ABSTRACT**

A difficult challenge in seismic processing and imaging is to address insufficient and irregular sampling. Most processing algorithms require well-sampled data, which involves small sampling intervals with a regular distribution. Recently, rank reduction methods are used in seismic processing algorithms. These methods are based on rank reduction of the trajectory matrices using truncated SVD. Estimation of the rank of the Hankel matrix depends on the number of the plane waves; however, when it comes to more complicated data, the rank reduction method may fail or give poor results as a consequence of curved events not having a small rank (sparse) representation. To satisfy the plane wave assumption for the rank reduction method, one can utilize local windows to assume that events are plane waves. The rank reduction method requires the number of events as the rank parameter. This number defines the minimum rank selected in each step. In this paper, we first propose a method, which selects the rank automatically in each window by finding the maximum ratio of the energy between two singular values. The method may select a large rank to get the best result for very high curved events, which leads to remain residual errors. To overcome the residual errors, then we apply a weighting operator on the selected singular values to minimize the effect of noise projection on the signal projection. We test the efficiency of the proposed method by applying it to both synthetic and real seismic data.

### **INTRODUCTION**

Seismic surveys are designed to keep a consistent grid of sources and receivers. In typical real-life scenarios, regularly sampled seismic surveys are uncommon or rare because of logistic obstacles or economic constraints. These limitations make the large shot and receiver sampling intervals, which produce poorly irregularly sampled seismic data along with spatial coordinates with gaps without recorded traces. Seismic reconstruction methods can be divided into four main classes: signal processing-based methods, wave equation-based methods, machine learning-based methods, and rank reduction-based methods.

Most of the methods in the signal processing-based category are multidimensional and use prediction filters (Abma and Claerbout, 1995; Spitz, 1991; Porsani, 1999), transform domains such as Fourier transform (Sacchi et al., 1998; Liu and Sacchi, 2004; Trad, 2009), Radon transform (Sacchi and Ulrych, 1995; Trad et al., 2002), or Curvelet transform (Herrmann, 2008). There are Some hybrid techniques using a combination of Fourier transform with prediction errors filters (Naghizadeh and Sacchi, 2010) as a way to improve interpolation beyond aliasing.

Wave equation-based algorithms execute an implicit migration de-migration pair. Stolt (2002) introduced a mapping and reconstruction algorithm that regularized a data set when removing acquisition footprints. A finite-difference offset continuation filter was proposed by (Fomel, 2003).

In recent years, machine learning has attracted much attention in geophysical studies. Deep

learning (DL) has had a great influence on signal and image processing. However, Machine learning for geophysical problems is new and needs to develop more. Jin et al. (2018) introduced deep residual networks (ResNets) for denoising. Wang et al. (2018) proposed a ResNets block designing to interpolate anti-aliased seismic data. Siahkoochi et al. (2018) proposed Generative Adversarial Networks (GANs) to reconstruct heavily sub-sampled seismic data by giving up linearity and using an adaptive non-linear model. Convolutional Neural Networks (CNNs) is a novel strategy for the reconstruction of missing traces in pre-stack seismic images (Mandelli et al., 2018). A new framework for training Artificial Neural Networks (ANNs) presented by Mikhailiuk and Faul (2018) to restore corrupted multidimensional seismic signals. Wang et al. (2019) utilized an eight-layer residual learning networks (ResNets) with a better back-propagation property for interpolation. Shi et al. (2020) proposed a regularization method that parameterizes the model with a random-initialized convolutional neural network (CNN). A method based on imaging techniques from single-image super-resolution (SISR) proposed by Greiner et al. (2019) that consists of a deep (CNN) and a periodic resampling layer for upscaling to the non-aliased wavefield. In rank reduction-based methods, the linear events in a clean seismic data set are low rank in the time or frequency domain. However, noise and missing traces increase the rank of data (Trickett, 2008). In the Fourier domain, The rank reduction algorithms reduce the rank of Hankel/Toeplitz matrices generated from the frequency slices. The singular spectrum analysis (SSA) method proposed by (Sacchi, 2009; Trickett and Burroughs, 2009; Oropeza and Sacchi, 2011) works by rank-reduction of the Hankel matrix with an iterative algorithm in the frequency domain. Gao et al. (2013) extended the SSA method to higher seismic data dimensions and called it a multichannel singular spectrum (MSSA). Interpolating regularly missing traces with a de-aliased MSSA method was proposed by Naghizadeh and Sacchi (2013). Kreimer et al. (2013) developed the algorithm for pre-stack data via low-rank tensor completion. Kreimer and Sacchi (2012) used the higher-order SVD for rank-reduction of the pre-stack seismic data tensor. Ely et al. (2015) used a statistical test to control the complexity of regularization by tuning a single regularization parameter. Kumar et al. (2015) developed rank-reduction techniques using matrix completion (MC) to seismic data reconstruction. Carozzi and Sacchi (2021) proposed a method called (I-MSSA) that interpolates seismic data in its exact data coordinates that overcome the problem of vertical errors in rough binning.

One of the necessary assumptions of the rank-reduction-based method techniques is that the Hankel matrix generated from the clean and complete seismic data containing plane waves is the low-rank data, where its rank is equal to the number of the plane waves. Regarding this assumption, the gaps and noise increase the rank. However, dealing with the complex structures of the real data makes the rank selection difficult and inconvenient. One approach to deal with this problem is to apply the algorithm to the local window. However, in most cases, it is not easy to find the proper window size because it is hard to decide if the structure in the local window is linear or not. Moreover, it is hard to approximate the rank of each window. Choosing the wrong rank will lead to a failure because the overestimation of rank remains a significant residual and underestimation of it will cause random noise and distort the signal. In this paper, we will apply a method that selects rank automatically for each local window. The method was proposed by Wu and Bai (2018) for 2D data. We apply the proposed rank selection criterion to both synthetic 3D data and demonstrate its successful performance. The strategy is to choose the second cutoff in the singular value spectrum of the block Hankel matrix. In the presence of the noise when the signal to noise

ratio is low, the reconstructed data tends to contain a significant amount of residual noise when using the TSVD. This amount of residual noise is because of the noise subset projection on the signal component projection. Nadakuditi (2014) introduced an algorithm that provides a weighted approximation and focuses on the reduction and thresholding form of the optimal weights, on the singular values. Their method mitigates the effect of rank overestimation. If the rank is properly estimated, the algorithm will better estimate weak subspace components of the signal matrix. Chen et al. (2016) introduced a damping operator that shrinks the singular values containing significant particles of residual noise. Their method in the reconstruction of highly noisy incomplete 5-D data set reached better results than the Cadzow rank-reduction method (Trickett and Burroughs, 2009; Oropeza and Sacchi, 2011).

We test the efficiency of the proposed method by applying it to both synthetic and real seismic data and compare it with the damping operator proposed by Chen et al. (2016).

### Background

Consider  $\mathbf{S}(t, x, y)$  a block of 3-D seismic data in the  $t - x - y$  domain on  $N_t$  by  $N_x$  by  $N_y$  samples. ( $t = 1, \dots, N_t$ ), ( $x = 1, \dots, N_x$ ), ( $y = 1, \dots, N_y$ ). In the frequency domain, the data are represented as  $\mathbf{S}(\omega, x, y)$  and ( $\omega = 1, \dots, N_\omega$ ). Each frequency slice of the data at a given frequency  $\omega_m$ , can be represented by the following matrix:

$$\mathbf{S}(\omega_m) = \begin{pmatrix} S(1, 1) & S(1, 2) & \dots & S(1, N_x) \\ S(2, 1) & S(2, 2) & \dots & S(2, N_x) \\ \vdots & \vdots & \ddots & \vdots \\ S(N_y, 1) & S(N_y, 2) & \dots & S(N_y, N_x) \end{pmatrix}. \quad (1)$$

To avoid notational confusion, let's ignore the argument  $\omega_m$ . Then, construct a Hankel matrix from each inline of  $\mathbf{S}$ ; for the inline  $i$ th the Hankel matrix in the  $m$ th frequency slice would be:

$$\mathbf{H}_i = \begin{pmatrix} S(i, 1) & S(i, 2) & \dots & S(i, l) \\ S(i, 2) & S(i, 3) & \dots & S(i, l+1) \\ \vdots & \vdots & \ddots & \vdots \\ S(i, N_x - l + 1) & S(i, N_x - l + 2) & \dots & S(i, N_x) \end{pmatrix}. \quad (2)$$

At this step, we have  $N_y$  Hankel matrices of each inline. Then multichannel singular spectrum analysis (MSSA) constructs a block Hankel matrix  $\mathbf{M}$  from Hankel matrices  $\mathbf{H}_i$ :

$$\mathbf{M} = \begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_n \\ \mathbf{H}_2 & \mathbf{H}_3 & \dots & \mathbf{H}_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}_{N_y-n+1} & \mathbf{H}_{N_y-n+2} & \dots & \mathbf{H}_{N_y} \end{pmatrix}. \quad (3)$$

The size of  $\mathbf{M}$  is  $I \times J$ , where  $I = (N_x - m + 1)(N_y - n + 1)$  and  $J = mn$ . The integers  $m$  and  $n$  are chosen to make the Hankel matrices of  $\mathbf{H}$  and the block Hankel matrix of  $\mathbf{M}$  square matrices or close to square. As we see, including the two spatial dimensions of the 3-D cube for each frequency can make the block Hankel matrix very large.

In seismic data, the observed data can be indicated as  $S_{obs} = R(S_0 + \eta)$ . Where  $S_{obs}$  is observed seismic data,  $S_0$  indicates the full noiseless data,  $\eta$  represents the random noise and the residuals, and  $R$  indicates the sampling matrix formed of zeros and ones. Using the SSA algorithm, we can write the Hankel matrix of observed data as:

$$M = P + N, \quad (4)$$

where  $P$  represents low rank Hankel matrix of the desired signal and  $N$  denotes the noise component which includes the gaps as well. In SSA, decomposition of the Hankel matrix in the rank-reduction step using TSVD will be as follow :

$$M = U \Sigma V^H. \quad (5)$$

If one knows the desired rank of the Hankel matrix, the estimated signal desired signal is recovered by:

$$\hat{M} = U_k \Sigma_k V_k^H, \quad (6)$$

where  $\hat{M}$  is the estimated signal, and  $k$  is the predefined rank of the Hankel matrix equals to the number of the linear events in each local window.

Let's investigate the effect of linear events on the singular values in each frequency. Figure 1 (a) shows the cube of 3-D data. Figure 1 (b) indicates an inline of the data, and figure 1 (c) represents a slice of data in cross-line direction.

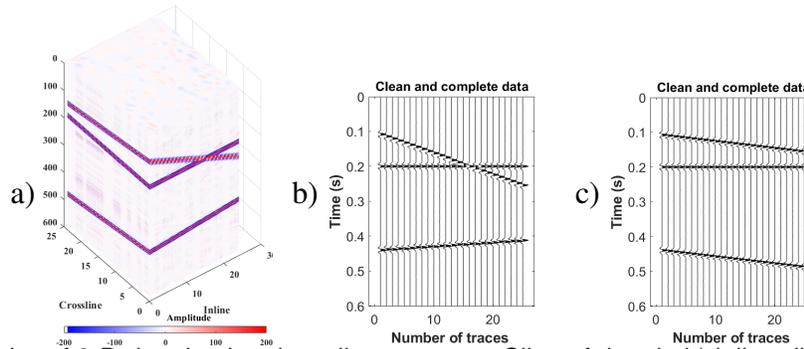


FIG. 1. a) Cube of 3-D data having three linear events. Slice of data in b) inline direction, c) cross-line direction.

We investigate the singular value distribution of the data by converting it to the  $(f, x, y)$  domain. Next is generating the block Hankel matrix in each frequency slice. The size of the generated block Hankel matrix in each constant frequency would be  $256 \times 256$ . Figure 2 (a) shows the block Hankel matrix for the frequency slice of 20 Hz. Figure 2 (b) displays the block Hankel matrix generated in the frequency slice of 60 Hz. According to figures 2 (a) and (b), it is clear that the block Hankel matrix for the frequency 20 Hz is smoother than the one for the frequency 50 Hz. It signifies that the block Hankel matrices for the higher frequencies require a higher rank than those in the lower frequencies to recover.

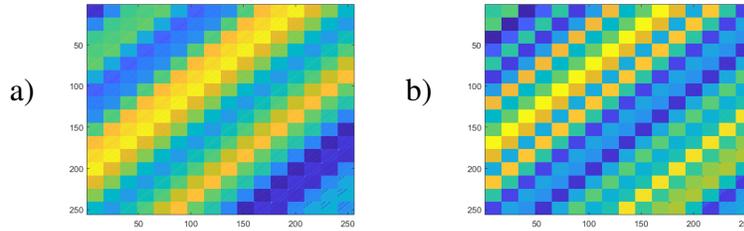


FIG. 2. Block Hankel matrix for frequency slice of a) 20 Hz, b) 50 Hz.

The singular values spectrum of the data’s block Hankel matrices for its frequency range is presented in figure 3. Figure 3 (b) represents a zoom into the first 15 singular values spectrum. This figure illustrates how the number of nonzero singular values of the block Hankel matrix at low frequencies is equal to the number of the linear events but increases with frequency.

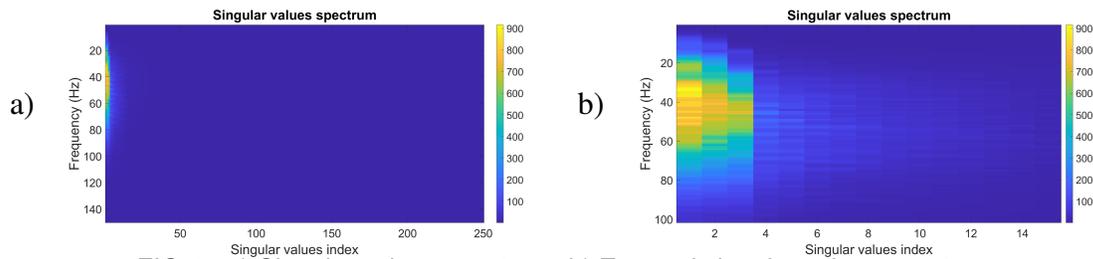


FIG. 3. a) Singular values spectrum, b) Zoomed singular values spectrum.

Similarly, Figure 4 (a) represents the bar plot of the normalized singular values for the frequency slice of 20 Hz. Figure 4 (b) shows the zoomed image for the first 20 singular values. Figure 4 (c) shows the bar plot of the singular values for the frequency slice of 50 Hz. Figure 4 (d) is the zoomed image for the first 20 singular values. From the figures, one can see that for the frequency of 20 Hz the maximum difference of the energy between two adjacent singular values occurs in the third singular value, and it is the same as the number of the linear events. However, for the frequency slice of 50 Hz, there is a second group of nonzero singular values. It means that we need to keep more singular values to recover all the frequencies completely.

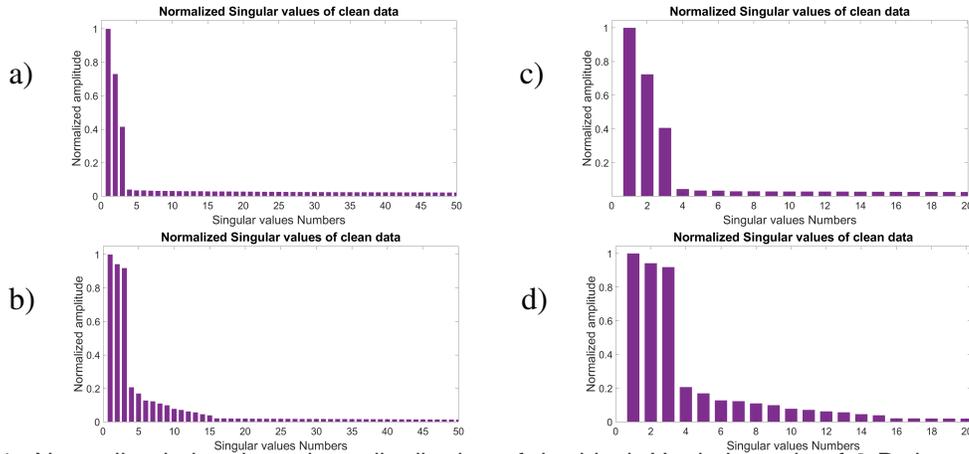


FIG. 4. Normalized singular values distribution of the block Hankel matrix of 3-D data with three linear events in the frequency slice a) 20 Hz, b) 50 Hz. The first 20 singular values of the same data for frequency slice of c) 20 Hz, d) 50 Hz.

To analyze the effect of each singular value on the power of each frequency for the data presented in figure 1, data are decomposed into its singular matrices and recovered with each singular value from one to 15 separately. Then, the power per frequency of the recovered data with each singular value is calculated. In Figure 5 the red line represents the graph of frequency power per frequency and the blue line shows the estimate of the mean normalized frequency, of the power spectrum. Figure 5 (a) is the result of recovering data with the first singular value. Figure 5 (b) shows the power per frequency for the recovered data with the second singular value. Figure 5 (c) is the power per frequency for the recovered data including just the third singular value. The estimated mean frequency for these first three singular values is 77 Hz. Figures 5 (d), (e), and (f) represent the power per frequency for the recovered data including just the 4th, 5th, and 6th singular values, respectively. The estimated mean frequency for these singular values is 90 Hz. Figure 5g), h), and i) indicates the power per frequency for the recovered data including just the 7th, 8th and 9th singular values, respectively. The estimated mean frequency for these singular values is 88 Hz. Figure 5 (j), (k), and (l) shows the power per frequency for the recovered data including just the 10th, 11th and 12th singular values, respectively. It is clear that the most power of the useful signal is recovered with the first three singular values. Nevertheless, the shift in the mean frequency from 77 Hz to 90 Hz states that there is a leakage for the higher frequencies in the data. We conclude from this test that despite the remaining residual errors in the presence of the additive noise, we need to choose the rank of the block Hankel matrix greater than the number of linear events in each processing window of frequency and space.

To find the best rank of data that minimizes the residual errors, we tested different kinds of data containing various numbers of events with different slopes and amplitudes. The best result is when we choose the rank from the second cutoff of singular values instead of the first cutoff .

In all tested data, it is possible to see the first cutoff with a distinct change of the energy of singular values. However, sometimes there is no abrupt change in the amplitude of the singular values for the second cutoff, especially in the presence of a high level of random noise. Finding the second cutoff could be a challenge when the quality of the signal is poor. With a careful look at where the first and the second cutoff in clean and complete

data happen, we conclude that there is a linear relationship between the first and second cutoff. The best estimate for the second cutoff can be estimated as follows:

$$\mathcal{T}(\Sigma, k) = \max_i \frac{\sigma_i^2}{\sigma_{i+1}^2}, \quad (7)$$

$$\mathcal{B}(\Sigma, \tilde{k}) = 3 \times \mathcal{T}(\Sigma, k), \quad (8)$$

where  $\sigma_i$  is the  $i$ th singular value of the Hankel matrix in each frequency.  $\mathcal{T}(\Sigma, k)$  indicates the operator that finds the rank at the point where the two following singular values become more scattered.  $\mathcal{B}(\Sigma, \tilde{k})$  indicates the second cutoff of the singular values spectrum of the block of Hankel matrix in each frequency slice.  $\tilde{k}$  can be introduced as the optimal rank of the block Hankel matrix that minimizes the Frobenius-norm difference between the approximated and the exact signal components.

Substituting equation 8 to the rank-reduction step will give us:

$$\tilde{\mathbf{M}} = \mathbf{U}_{\tilde{k}} \mathcal{B}(\Sigma, \tilde{k}) \mathbf{V}_{\tilde{k}}^H, \quad (9)$$

Applying equation 9 on the rank-reduction step of MSSA, the rank reduced block Hankel matrix is reconstructed. The next step is averaging anti-diagonals of the recovered block Hankel matrix. It recovers the signal in the Fourier domain for each frequency slice. Applying the iterative algorithm will reconstruct the amplitude of missing traces. Selecting the rank using equation 9 makes the estimated signal contains all the signal components. One of the advantages of this rank-reduction method is that it is adaptive and data-driven, and there is no need to set any parameter for the rank-reduction step in each processing window. Moreover, choosing the second cutoff as a substitute for the first cutoff lets the high frequencies recover completely. However, selecting large ranks leads to remain more residual errors in the recovered data. But, applying the weighting operator reduces the effect of noise on the projected signal components and improves data results by recovering higher frequencies.

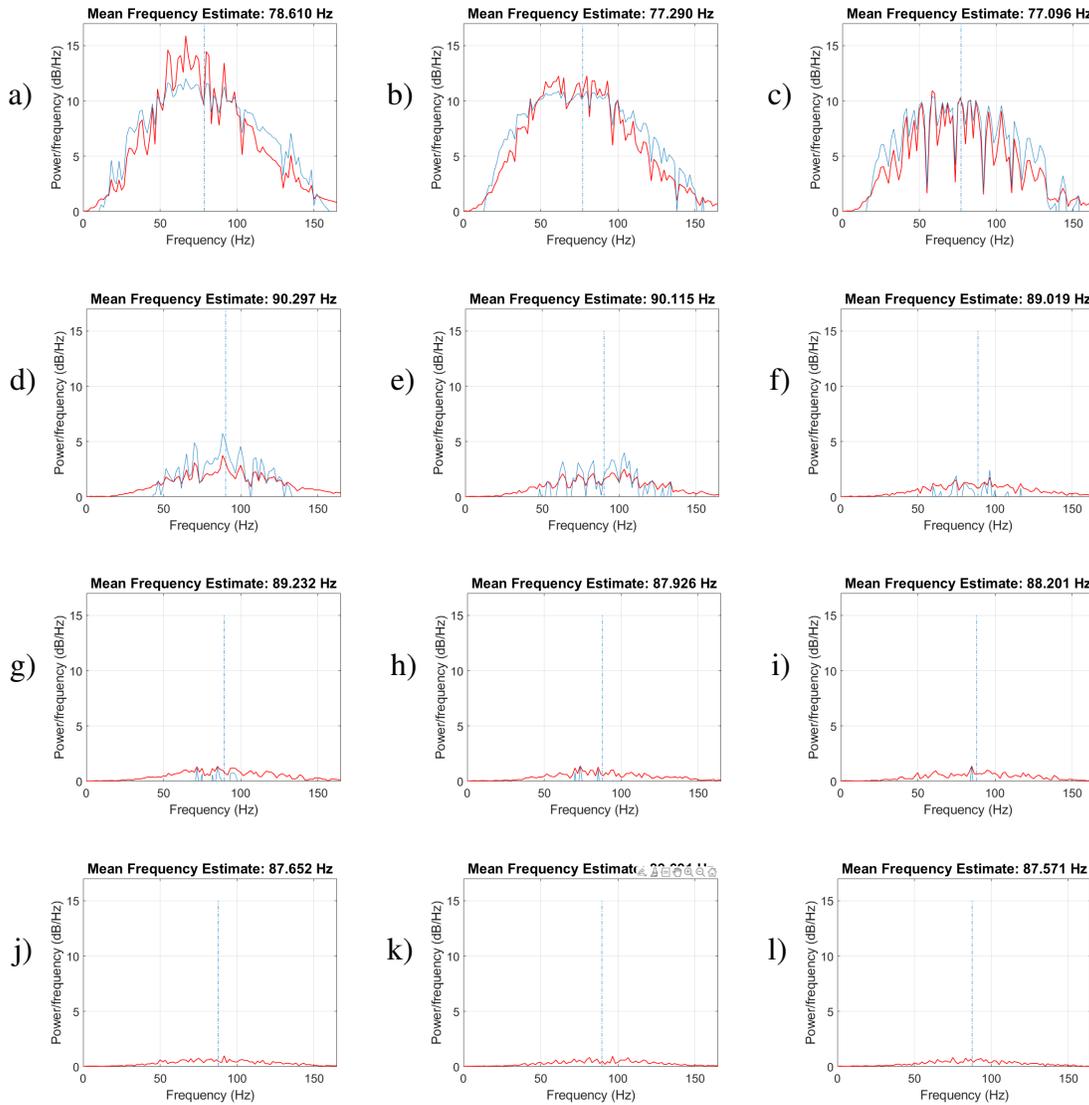


FIG. 5. Analysis of the frequency power for each singular value. The red line shows power per frequency, and the blue line indicates the estimated mean frequency. a) Power per frequency of data recovered by the first singular value. b), c), d), e) f), g), h), l), j), k), and l) power per frequency recovered by the 2nd, 3rd, 4th, 5th, 6th, 7th, 8th, 9th, 10th, 11th, and 12th singular value, respectively.

We are looking for a weighting operator  $\hat{\mathbf{W}}$ , to adjust the singular values of  $\tilde{\mathbf{M}}$  to calculate the best estimation of the desired signal. Nadakuditi (2014) proposed an algorithm for low-rank matrix denoising that can be summarized as follow:

- 1:  $\mathbf{M}_{n \times m}$  is the signal plus noise Hankel matrix.
- 2:  $k$  is the best effective rank that can represent the signal,
- 3: for  $i = 1 : k$

$$\text{compute } \hat{w}_i = \left( -\frac{2}{\sigma_i} \frac{\mathcal{D}(\sigma_i; \Sigma)}{\mathcal{D}'(\sigma_i; \Sigma)} \right), \quad (10)$$

- 4: end for loop,
- 5: compute results as  $\hat{\mathbf{M}} = \sum_{i=1}^k \hat{w}_i \sigma_i u_i v_i^H$

In this algorithm  $\mathcal{D}(\sigma; \Sigma)$  is computed as:

$$\mathcal{D}(\sigma; \Sigma) = \frac{1}{k} \text{Tr}(\sigma(\sigma^2 \mathbf{I} - \Sigma \Sigma^H)^{-1}) \frac{1}{k} \text{Tr}(\sigma(\sigma^2 \mathbf{I} - \Sigma^H \Sigma)^{-1}) = \left[ \frac{1}{k} \text{Tr}(\sigma(\sigma^2 \mathbf{I} - \Sigma^2)^{-1}) \right]^2, \quad (11)$$

where  $\mathcal{D}$  represents  $D$ -transform, and  $\text{Tr}(\cdot)$  denotes the trace operator of the input. The  $\mathcal{D}'$  represents the derivative of  $\mathcal{D}$  with respect to  $\sigma$ :

$$\begin{aligned} \mathcal{D}'(\sigma; \Sigma) &= 2 \left[ \frac{1}{k} \text{Tr}(\sigma(\sigma^2 \mathbf{I} - \Sigma^2)^{-1}) \right] \\ &= \left[ \frac{1}{k} \text{Tr}((\sigma^2 \mathbf{I} - \Sigma^2)^{-1} - 2\sigma(\sigma^2 \mathbf{I} - \Sigma^2)^{-2}\sigma) \right] \\ &= \frac{2}{k^2} \left[ \text{Tr}(\sigma(\sigma^2 \mathbf{I} - \Sigma^2)^{-1}) \right] \left[ \text{Tr}((\sigma^2 \mathbf{I} - \Sigma^2)^{-1} - 2\sigma^2(\sigma^2 \mathbf{I} - \Sigma^2)^{-2}) \right], \end{aligned} \quad (12)$$

The  $D$ -transform describes how the distribution of the singular values of the sum of the independent matrices is related to the distribution of the singular values of the individual matrices (Benaych-Georges, 2009). Benaych-Georges and Nadakuditi (2012) indicate that the principal singular values and vectors of a large matrix can be set apart as the singular values of the signal matrix and the  $D$ -transform of the limiting noise-only singular value distribution. From Equation 10 the weighting operator  $\hat{\mathbf{W}}$  can be written as:

$$\hat{\mathbf{W}} = \text{diag}(\hat{w}_1, \hat{w}_2, \dots, \hat{w}_k). \quad (13)$$

We can substitute the weighting algorithm obtained from equation 13 in equation 9 to enhance the results of the rank-reduction step as:

$$\hat{\mathbf{M}} = \mathbf{U}_{\tilde{k}} \hat{\mathbf{W}} \mathcal{B}(\Sigma, \tilde{k}) \mathbf{V}_{\tilde{k}}^H, \quad (14)$$

where  $\hat{\mathbf{M}}$  indicates the reduced rank block Hankel matrix. Missing traces can be interpolated completely by applying the iterative algorithm. This Adaptive Weighting Rank-Reduction (AWRR) method leads the way that sorts out the rank of the block Hankel matrix automatically while denoising data.

The weighting operator satisfies the equation below (Nadakuditi, 2014):

$$\left\| \tilde{\mathbf{M}} - \mathbf{P} \right\| \leq \epsilon. \quad (15)$$

To understand the effect of the weighting operator on the singular values of the block Hankel matrices, the predefined rank-reduction method (TRR) and the WRR method are applied to a block Hankel matrix.

Figures 6 (a) shows clean and complete 3-D synthetic data having four linear events, figure 6 (b) indicates same 3-D data with  $SNR = 2$  and 51% missing traces. Figures 6 (c) and 6 (d) show the slices of data in inline directions for cube of Figures 6 (a) and 6 (b), respectively. Figures 6 (e) and 6 (f) are the slices of the cubes of Figures 6 (a) and 6 (b) in crossline direction, respectively.

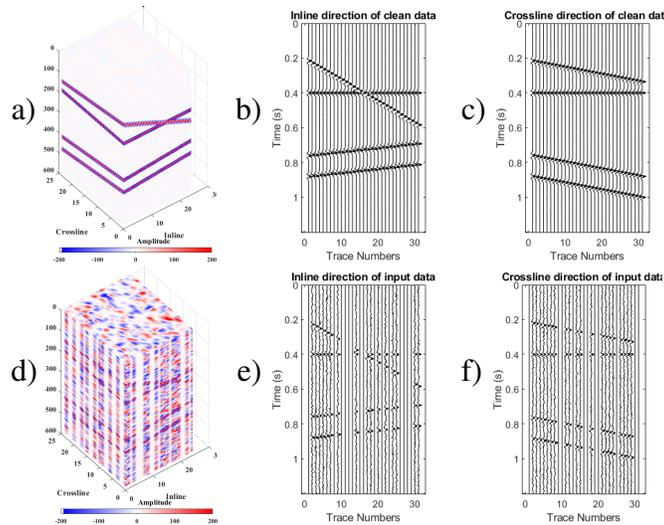


FIG. 6. a) Cube of 3-D synthetic clean and complete data, b) and c) inline and cross-line sections of cube (a), respectively. d) Cube of 3-D synthetic data with  $SNR = 2$  and 51% missing traces, e) and f) inline and cross-line sections of cube (d), respectively.

Figure 7 shows the Hankel matrices of data in figure 6 for the frequency slice of 20 Hz. Figure 7 (a) is the block Hankel matrix of clean and complete data. Figure 7 (b) corresponds to the block Hankel matrix of data with  $SNR = 2$  and 51% missing traces. Figure 7 (c) shows the block Hankel matrix after applying the predefined rank=12 after ten iterations. Figure 7 (d) indicates the block Hankel matrix after applying the weighting operator to adjust the singular values of the data using TSVD with predefined rank=12 after ten iterations. We can see from the figure 7 that the block Hankel matrix of the WRR method is smoother than the result of applying the predefined rank=12 after ten iterations.

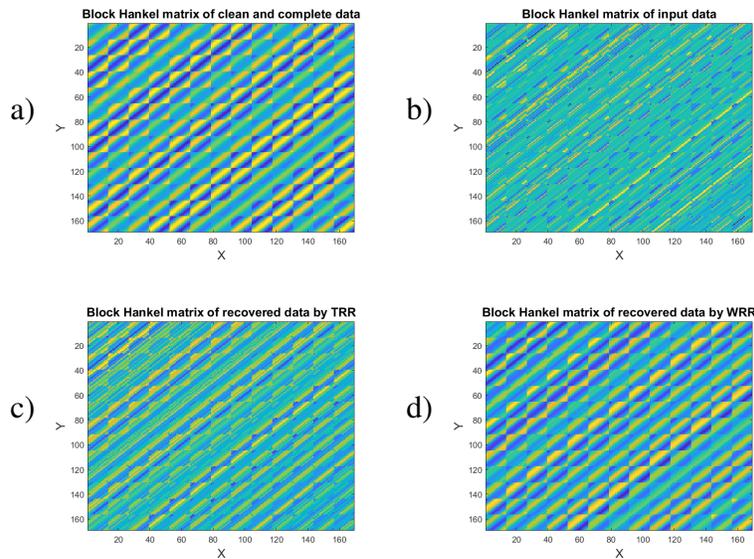


FIG. 7. Block Hankel matrices for frequency of 20 Hz for a) clean and complete data, b) data with  $SNR = 2$  and 51% missing traces, c) recovered data by TRR rank-reduction after ten iterations, d) recovered data by WRR after ten iterations.

Figure 8 corresponds to the first 25 singular values spectrum of the block Hankel matrices of figure 6. Figure 8 (a) represents the singular values spectrum of the block Hankel matrix of clean and complete data for the frequency of 20 Hz, the first 3 singular values indicate the useful signal that relates to the coherent events in the data. Figure 8 (b) relates to the singular values spectra of the noisy and incomplete data, figure 8 (c) shows the singular values spectra of the noisy and incomplete data after applying TSVD with rank=12. Figure 8 (d) indicates the singular values spectra of the noisy and incomplete data after applying TSVD with rank=12 and the weighting operator. We can see that the singular spectrum of the data after applying the weighting operator is much closer to the one relates to the desired signal in figure 8 (a).

Figure 9 corresponds to the block Hankel matrices of the same data in the constant frequency slice of 50 Hz. Figure 9 (a) shows the block Hankel matrix of the clean and complete data, figure 9 (b) is the block Hankel matrix of data with  $SNR = 2$  where 51% of traces are killed. Figure 9 (c) displays the recovered block Hankel matrix after applying TRR and rank=4 after ten iterations. Figure 9 (d) indicates the block Hankel matrix after applying the WRR with predefined rank=4. Figure 9 (e) represents the recovered block Hankel matrix of the ARR method, and figure 9 (f) is the block Hankel matrix recovered by the weighting ARR (AWRR) method. Generally, the results of AWRR and WRR are smoother than the output of TRR and ARR, and the results of ARR and AWRR include more details than TRR and WRR.

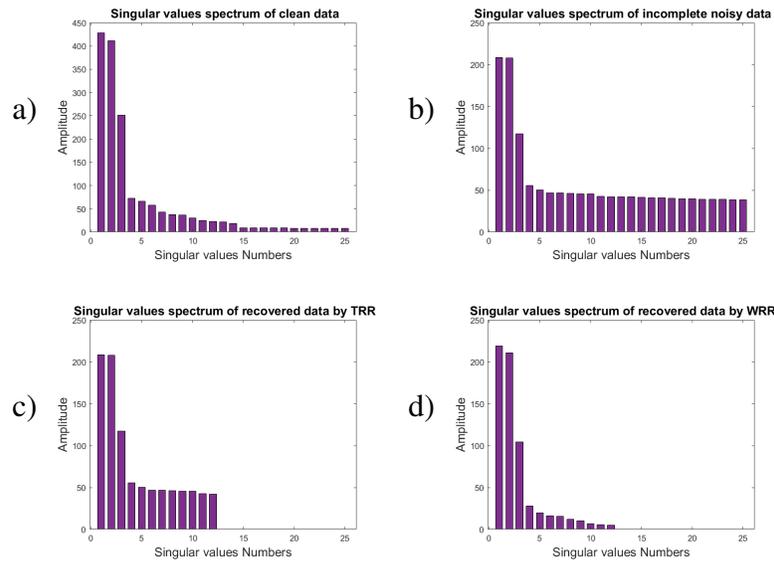


FIG. 8. The first 25 singular values spectrum of the block Hankel matrices for frequency of 20 Hz for a) clean and complete data, b) data with  $SNR = 2$  and 51% missing traces, c) The first 12 singular value of the data, d) singular values spectrum of the data after applying the weighting operator.

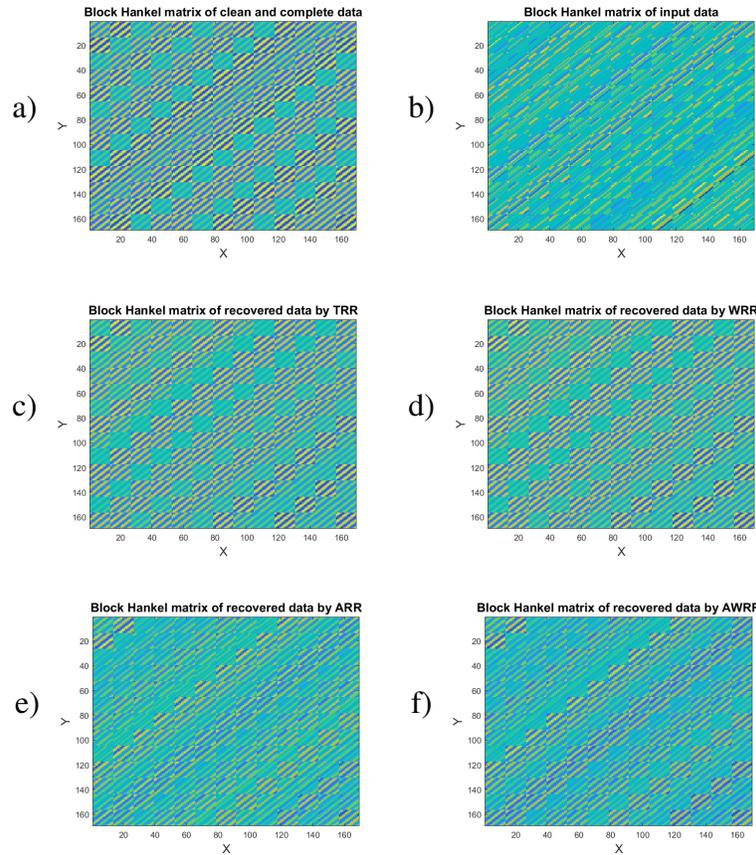


FIG. 9. Block Hankel matrices for frequency of 50 Hz for a) clean and complete data, b) data with  $SNR = 2$  and 50% missing traces, c) recovered block Hankel matrix by TRR rank=4 after ten iterations, d) recovered data by WRR after ten iterations. e) Recovered data by applying the ARR method after ten iterations, d) recovered data after applying AWRR after ten iterations.

Figure 10 corresponds to the first 25 singular values spectrum of the block Hankel matrices of figure 9. Figure 10 (a) represents the singular values spectrum of the block Hankel matrix of clean and complete data for the frequency of 50 Hz. The abrupt drop of energy of the third and the fourth singular value relates to the useful signal. However, there are still nonzero singular values that are related to the useful signal. Figure 10 (b) is the singular values spectra of the block Hankel matrix of noisy and incomplete data. Figure 10 (c) shows the singular values spectrum of the noisy and incomplete data after TRR with rank=4. Figure 10 (d) displays the singular values spectrum of the noisy and incomplete data after applying WRR with rank=4. Figure 10 (e) depicts the singular values spectrum after applying ARR. Figure 10 (d) indicates the singular values spectrum of the noisy and incomplete data after implementation of the AWRR method. From the figures, one can see that the result of AWRR is more comparable to the one relates to the desired signal in Figure 10 (a).

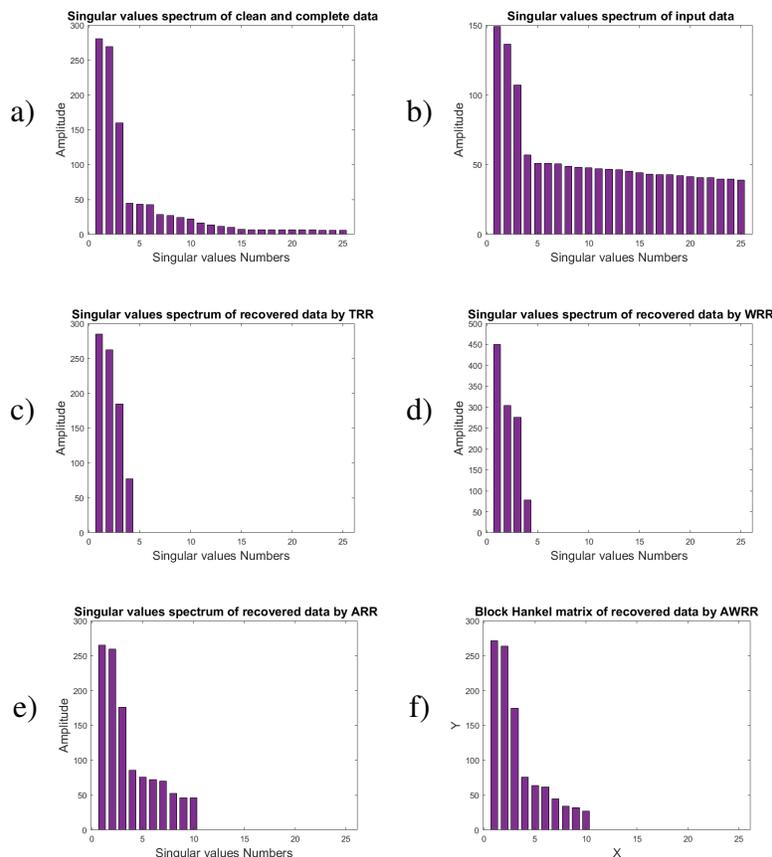


FIG. 10. The first 25 singular values spectrum of the block Hankel matrices for frequency slice of 50 Hz for, a) clean and complete data, b) data with  $SNR = 2$  and 50% killed traces, c) recovered data by TRR, d) recovered data by WRR, e) recovered data by ARR, f) recovered data by AWRR.

## Results and discussions

Several evaluations are presented in this section to estimate the proficiency of the different methods of rank-reduction for the MSSA algorithm. First, the methods are tested on a synthetic shot gather containing nine hyperbolic events with different curvatures. Then the efficiency of the methods on a shot gather of a 3-D field data set is tested.

### Synthetic data

Figure 11 shows the result of applying the methods of rank-reduction to a synthetic shot gather holding nine hyperbolic events with different curvatures. The first test is a cube of 100 inline and 11 crosslines with  $SNR = 2$  and 60% decimated traces. We choose the local window with 23 traces in the inline direction and 11 in the crossline direction for each method and half a window overlapping in each direction. We set the number of iterations constant for all of the methods. The frequency range of the MSSA algorithm is 1 to 100 Hz. Furthermore, the rank of TRR and WRR is set to 9. Figure 11 (a) is the desired data arranged into a 2-D matrix. Figure 11 (b) shows input data with  $SNR = 2$  and 60% missing traces. The results of applying TRR, ARR, WRR, and AWRR methods are shown in figures 11 (c), (e), (g), and (i), respectively. Figures 11 (d), (f), (h), and (j) are residual

errors of figures 11 (c), (e), (g), and (i), respectively. All four methods have recovered the amplitude of the signals. However, TRR remains severe residual errors, the result from ARR causes significant residual noise and AWRR is much cleaner than TRR, ARR and WRR. The input data quality is calculated  $QF = -1.44$  dB. The output quality factor of the TRR, ARR, WRR, and AWRR methods are  $QF = 5.23, 6.58, 7.56, 8.12$ , dB, respectively.

For the stability analyses of each algorithm, we run them under different levels of gaps percentage with 20 several realizations of sampling ratio for a given  $SNR = 2$ . The other parameters are set constant for each run. Figure 12 shows the graph of the interpolation quality factor for each method versus the gap ratio for  $SNR = 2$ . The length of the error bars is the standard deviation of the interpolation quality factor, and each coloured lines connect the mean value of the interpolation quality factor for each method at each gap ratio. As we can see from figure 12, the input quality factor decreases with increasing gap ratio also the output quality factor decreases with increasing gap ratios. Moreover, the output quality factor curve of the AWRR method is always above the other curves.

Table 1 indicates the values of the mean and standard deviation of the interpolation quality of each method for each gap percent. We can see that the average value of the AWRR quality factor is higher than the other methods. Comparing the discussed methods, AWRR has lower standard deviation values that indicate the values tend to be close to the mean value, especially when the data is more sparse.

In the next experiment, the stability analysis is repeated to test the sensitivities of each method to the additive noise. The test is performed by changing the level of signal to noise ratio with 20 different realizations of SNR for a gap ratio = 50%. We set the other parameters constant for each run. In figure 13 we see that the proposed method (AWRR) outperforms the other methods even in the poor quality of the signal.

Table 2 represents the values of the mean and standard deviation of the output quality factor of each method for each signal to noise ratio inputs.

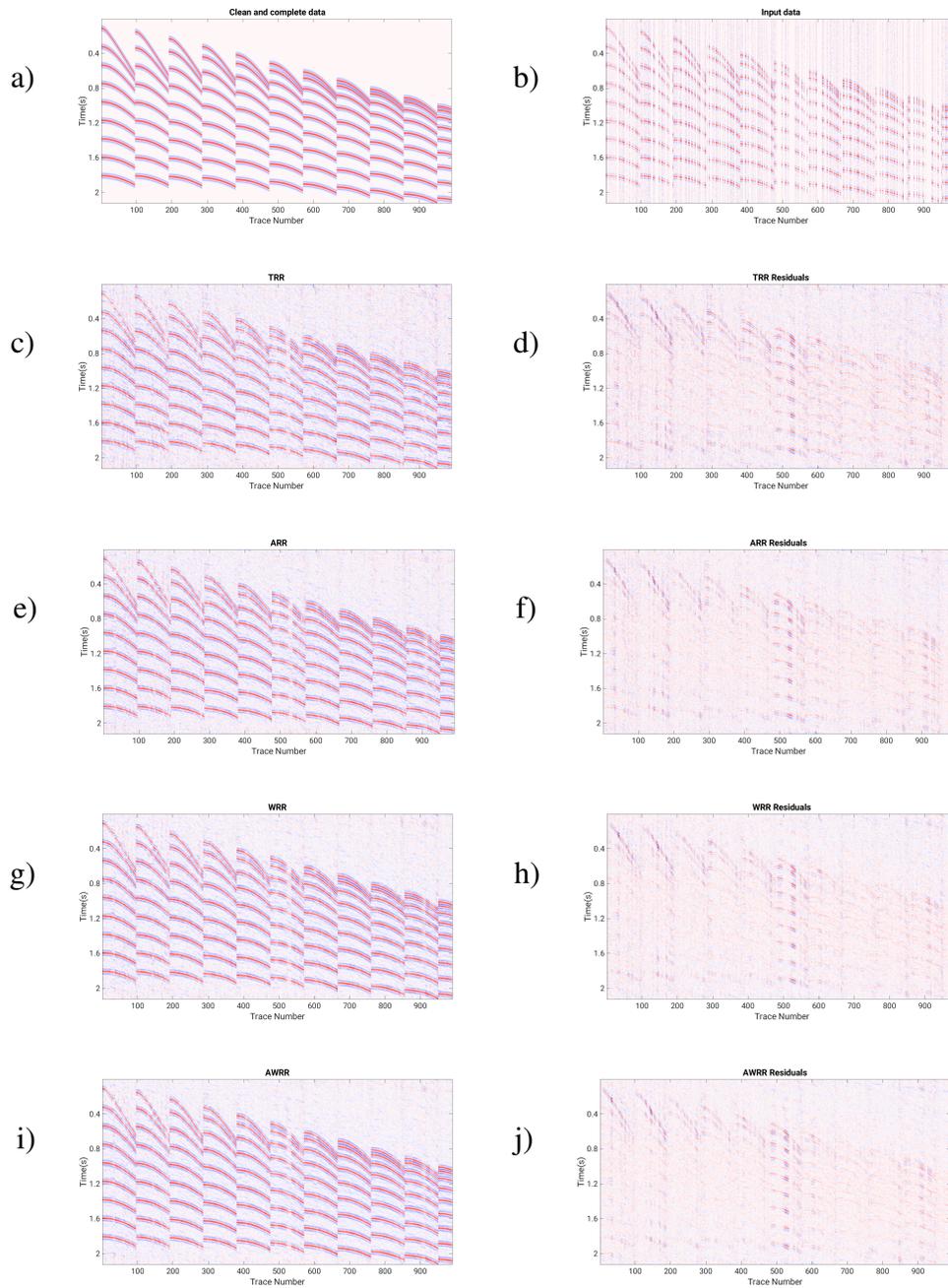


FIG. 11. Comparison of different methods of rank-reduction. a) Clean and complete data, b) input data with  $SNR = 2$  and 60% missing traces, c), e), g), and i) interpolated data by TRR, ARR, WRR, and AWRR, respectively. d), f), h), and j) residual errors of TRR, ARR, WRR, and AWRR, respectively.

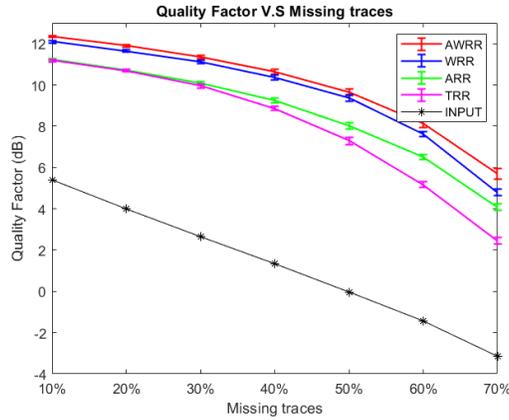


FIG. 12. The mean and standard error of the quality of reconstruction versus gap ratio for different methods. These results are obtained by running each method on 20 realizations of different data set for different gap ratios and rank=9 for the predefined rank methods.

Table 1. Mean and standard deviation of the quality of interpolation for each method for different gap ratio inputs.

Gap Ratio	Output Quality Factor (dB)							
	AWRR		WRR		ARR		TRR	
	MEAN (dB)	STD	MEAN (dB)	STD	MEAN (dB)	STD	MEAN (dB)	STD
10	12.34	0.04	12.10	0.04	11.22	0.06	11.17	0.04
20	11.90	0.05	11.64	0.05	10.71	0.09	10.69	0.06
30	11.36	0.05	11.12	0.05	10.08	0.08	8.86	0.08
40	10.45	0.07	10.35	0.07	9.24	0.10	8.90	0.11
50	9.64	0.16	9.38	0.20	8.01	0.17	7.28	0.20
60	8.13	0.13	7.61	0.13	6.51	0.14	5.16	0.20
70	5.70	0.16	4.78	0.21	4.10	0.21	2.46	0.25

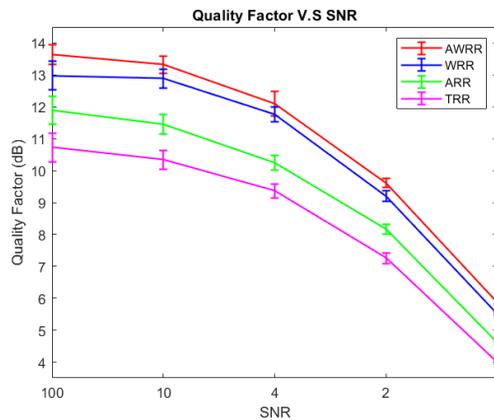


FIG. 13. The mean and standard error of the reconstruction's quality versus SNR for different methods. These results are obtained by running each algorithm on 20 realizations of the different data set for a signal-to-noise ratio with the gap ratio = 50%, and rank=9 for the predefined rank methods.

Table 2. Mean and standard deviation of the quality of interpolation for each method for different SNR inputs.

Input SNR	Output Quality Factor (dB)							
	AWRR		WRR		ARR		TRR	
	MEAN (dB)	STD	MEAN (dB)	STD	MEAN (dB)	STD	MEAN (dB)	STD
<b>100</b>	13.64	0.41	12.98	0.41	11.90	0.44	10.72	0.30
<b>10</b>	13.32	0.42	12.89	0.42	11.44	0.30	10.34	0.28
<b>4</b>	12.10	0.26	11.77	0.28	10.25	0.23	9.36	0.38
<b>2</b>	9.61	0.26	9.20	0.26	8.16	0.16	7.26	0.15
<b>1</b>	5.88	0.09	5.51	0.09	4.61	0.05	3.40	0.07

### Optimally damped rank-reduction method

We studied the performance of the AWRR algorithm comparing to the predefined rank-reduction method in interpolation and denoising of 3-D data. In this section, we compare AWRR as an adaptive and data-driven rank-reduction method with a "nearly-adaptive" method proposed by Chen et al. (2016). Their technique first derives the weighting operator, then based on the original weighting operator a damping operator is obtained to decrease the residual noise after the optimal weighting. In this method, the weighting operator is adjusted with an optimal damping operator. Moreover, regardless of the number of the linear events, the rank is selected larger. This "nearly adaptive" algorithm is less sensitive to the rank parameter also reduces the residual noise in the rank-reduction step. Due to this advantage, one can choose a large rank and preserve the signal perfectly while remaining fewer residual errors. The algorithm can be summarized as below.

The optimally damped rank-reduction (ORR) method considers a predefined rank =  $D$ , which is usually selected large enough to deal with complicated data and local windows. Then the estimated signal is calculated as follows:

$$\hat{\mathbf{M}} = \mathbf{U}_D \Sigma_D \mathbf{V}_D^H, \quad (16)$$

where  $D$  is the selected rank. There is a damping factor as below:

$$\nu = \hat{\delta}^Q (\Sigma_D)^{-Q}, \quad (17)$$

where  $\hat{\delta}$  represents the maximum element of  $\Sigma_D$ , and  $Q$  is the damping factor.

Using the weighting operator as  $\hat{\mathbf{W}} = \text{diag}(\hat{w}_1, \hat{w}_2, \dots, \hat{w}_D)$ , and the damping operator from equation 17 substituting in equation 16, the estimated signal applying the ORR method will be:

$$\hat{\mathbf{M}} = \mathbf{U}_D (\mathbf{I} - \nu) \hat{\mathbf{W}} \Sigma_D \mathbf{V}_D^H, \quad (18)$$

where  $\mathbf{I}$  is a unit matrix,  $\hat{\mathbf{M}}$  denotes the recovered block Hankel matrix by the ORR method. By using the iterative algorithm, data can be interpolated.

The method is insensitive to the rank so, one can choose a large rank and preserve the signal completely with remaining fewer residual errors. One of the drawbacks of this method is its computational cost that makes the algorithm less desirable, particularly for data with more than two spatial dimensions. Because computing the SVD using a large rank can be extremely time-consuming for large-scale problems.

We test the sensitivity of the ORR method assuming that we do not have prior information about the structural complexity of the data and thus the proper rank for data. We evaluate

the efficiency of the ORR by selecting the rank=7 to 18. Figure 14 shows the diagram of the interpolation quality with respect to the selected rank for the range of 7 to 18. By increasing the selected rank the output quality factor is increasing. The quality factor reaches its maximum value for the selected rank=12 while the curve drops for the ranks larger than 12. However, the quality factor is in the range of 9 to 9.2 (dB) for the chosen rank 10-15. Accordingly, the method is not fully insensitive to the rank unless we select the large rank in a reasonable range.

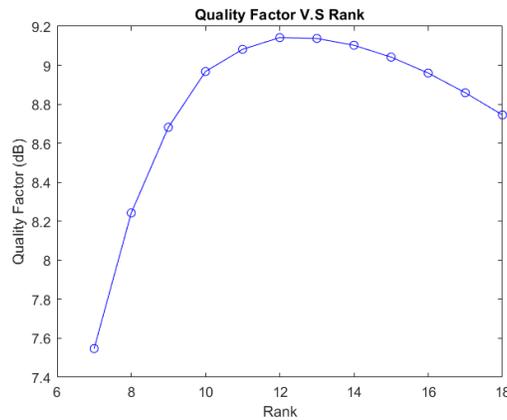


FIG. 14. Diagram of the output Quality of interpolation using ORR with respect to the selected rank.

To compare the efficiency of AWRR and ORR, a data set with additive random noise  $SNR = 2$  is tested for different percentages of missing traces. The selected rank for the ORR method is chosen 12. The parameters such as the number of iterations and the processing window size are selected the same for both methods. Figure 15 shows the output quality factor of AWRR and the ORR method with respect to the missing traces percentage. The blue line with the ‘o’ marker is the result of the ORR method and the red line with the ‘\*’ marker is the result of the AWRR method.

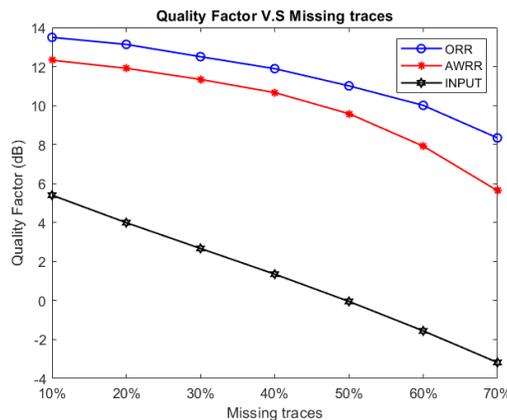


FIG. 15. The quality of interpolation versus gap percent for ORR, and AWRR.

Figure 16 refers to the result of applying ORR and AWRR to the data in figure 11. The local processing window is set to 23 traces in the inline direction and 11 in the crossline

direction for each method, with half a window overlapping in each direction. Figure 16 (a) is the initial data before noise contamination and killing traces arranged into a 2-D matrix. Figure 16 (a) shows input data with  $SNR = 2$  and 60% missing traces. Figure 11 (c) is the result of applying ORR, figure 16 (e) displays the result of applying AWRR. 16 (d) and (f) indicate residual error results subtracting the clean and complete data from the recovered data of figure 11 (c) and (e), respectively. We can see that both methods are successful in recovering data; however, the residual errors of the ORR method are less than the AWRR method. The input data quality factor is  $QF = -1.44 \text{ dB}$ . The outputs quality factor for the ORR and AWRR methods are  $QF = 11.12$  and  $9.44 \text{ dB}$ , respectively. When zooming the data, the comparison of the two methods is more noticeable. Figure 17 corresponds to the zooming area of the time 0.4 (s) to 10.2 (s), and the trace number 400 to 460 of data in figure 17. We can see that both ORR and AWRR are very powerful in reconstructing the data; however, the ORR's result is smoother than AWRR.

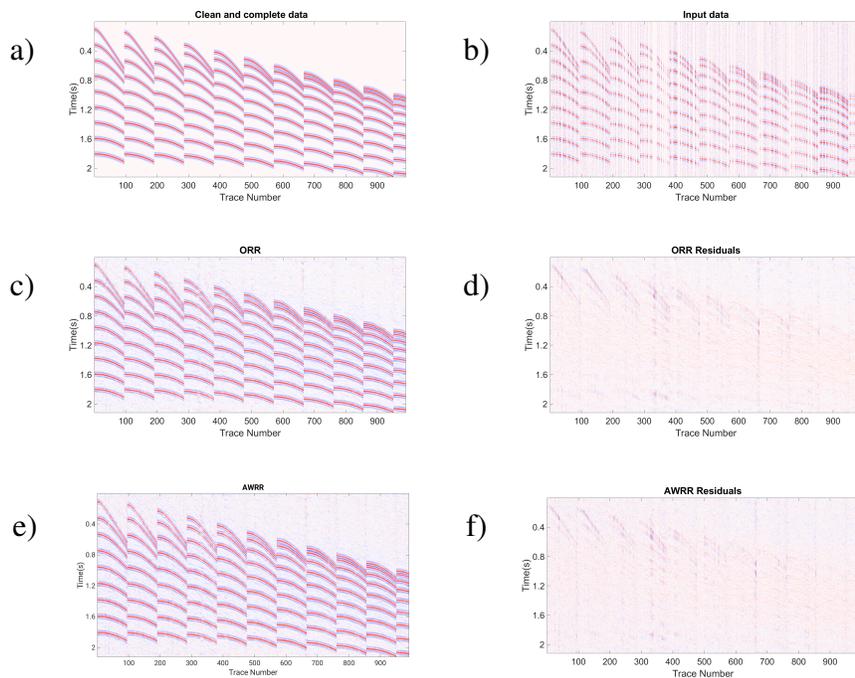


FIG. 16. Comparison of the interpolation results for ORR method. a) Clean and complete data, b) data with  $SNR = 2$  and 50% missing traces, c) recovered data by ORR (output  $QF = 11.12 \text{ dB}$ ), d) residual errors for ORR, e) interpolated data by AWRR (output  $QF = 9.44 \text{ dB}$ ), f) residual errors for AWRR.

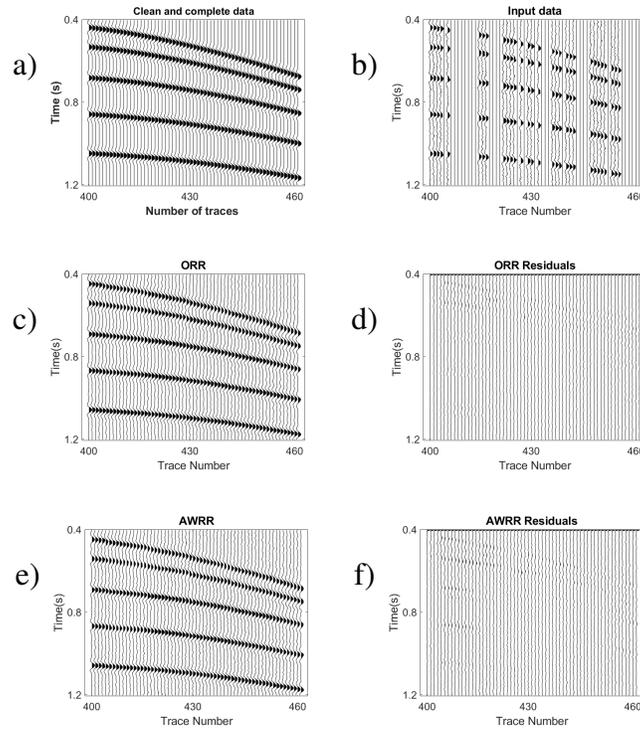


FIG. 17. Zoomed images of figure 16 from the time 0.4 s to 1.2 s and trace number 400 to 460. a) Clean and complete data, b) data with  $SNR = 2$  and 50% missing traces, c) recovered data by ORR, d) residual errors for ORR, e) interpolated data by AWRR, f) residual errors for AWRR.

### Real field data

This experiment tests the efficiency of the methods on a shot gathers of a 3-D field data set. It is easy to inspect traces in shot/receiver gather displays for poor receivers or any bad shots, which are the logistic constraints during the seismic survey. Regarding the input data, the best results are obtained with NMO-corrected data (Trad, 2009). However, in this experiment, we applied the MSSA interpolation before NMO correction to see the effect of the algorithm on the curvature.

Figure 18 refers to the acquisition coordinates of the shot gather. The red star represents the location of the shot, and the black dot indicates the receiver's location. Figure 18 (a) shows the initial distribution of the traces in a shot gather. For this experiment, 41% of the traces are killed. Figure 18 (b) illustrates the geometry of input traces after killing 41% of them.

This test is applying the MSSA algorithm using the AWRR and TRR methods in the rank-reduction step. Figure 19 (a) shows the input cube of data, which presents the missing traces. Figure 19 (b) shows the result of the interpolation using TRR in the rank-reduction step, and figure 19 (c) shows the result after using AWRR. For both tests, the number of iterations and processing window remain unchanged. For the TRR approach, the predefined rank is set to rank = 10. Figure 20 represents the cube of data arranged into a 2-D matrix. Figure 20 (a) is the input data. Figure 19 (b) represents the result of applying TRR. Figure 20 (c) shows the result of applying AWRR. Figure 21 corresponds to a patch of data from the time 1.35 (s) to 1.75 (s) and the trace number 160 to 190. Figure 21 represents the cube of data rearranged in a 2-D matrix. Figure 21 (a) is the input data. Figure 21 (b) represents

the result of applying TRR. Figure 21 (c) shows the result of applying AWRR.

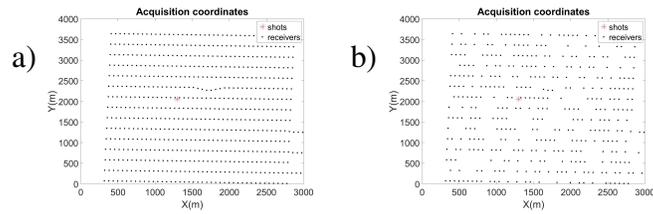


FIG. 18. Geometry of the field data. (a) Initial distribution of the traces. (b) Distribution of traces after killing 41% of them.

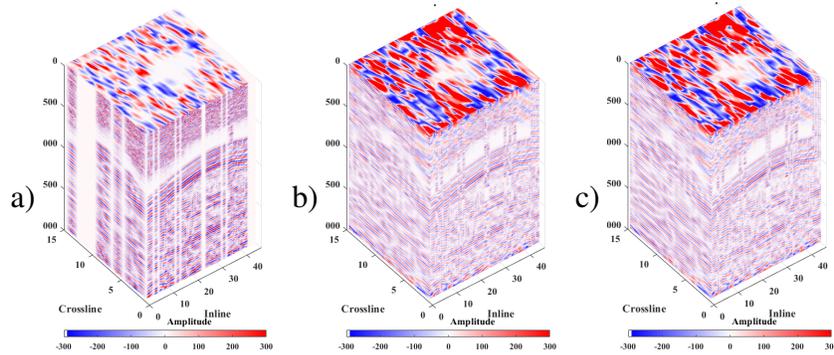


FIG. 19. Cube of 3-D field data. a) The input data. b) Interpolation result of applying TRR. c) Interpolation result of applying AWRR.

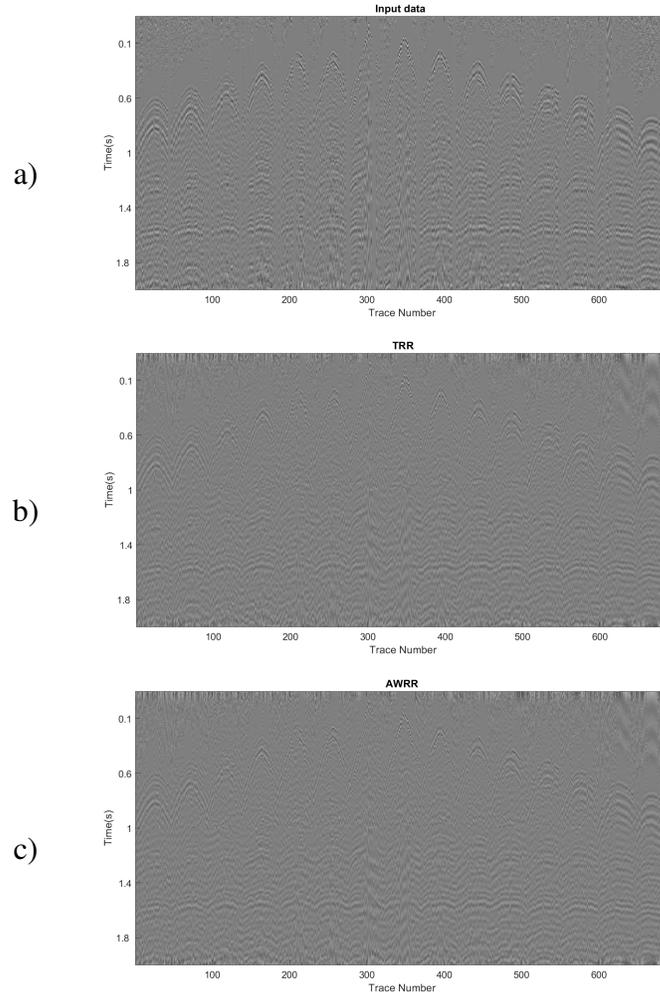


FIG. 20. a) Input real data, b) interpolation result of applying TRR approach, c) interpolation result of applying AWRR method.

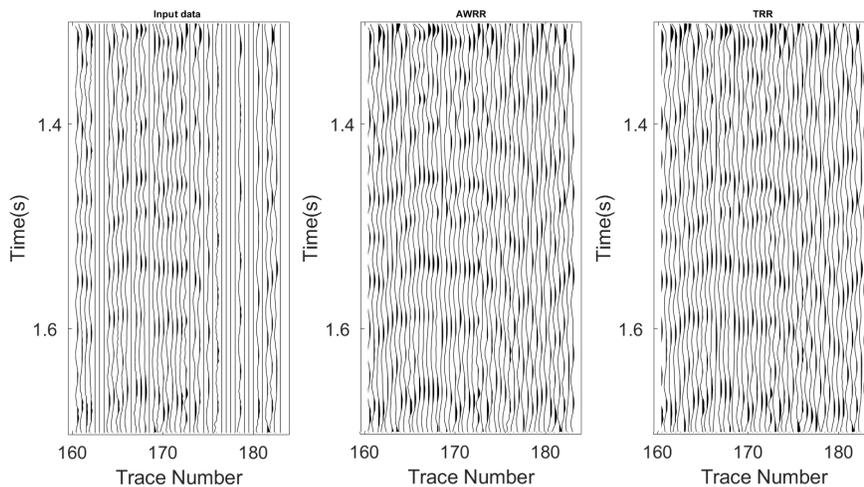


FIG. 21. Zooming area of figure 20 from time 1.35 (s) to 1.75 (s) and trace number 160 to 190. a) input data, b) TRR result, c) AWRR result

## CONCLUSIONS

We compared different methods of rank-reduction in the MSSA algorithm to interpolate and denoise 3-D seismic data. We proposed the AWRR method, which is a parameter-free method for selecting the proper rank. AWRR is based on a cascade weighting operator combined with the adaptive rank-reduction method. It is adaptive in selecting the rank of the Hankel matrices or the block Hankel matrices. Moreover, because of the weighting operator, AWRR can cause less residual noise.

AWRR is compared with the other rank-reduction methods the results showed that AWRR is more effective than the other methods especially in the case of low signal-to-noise ratio and when the data is sparse. The recovered block Hankel matrices in the frequency slices using AWRR are smoother and containing more details than the other rank-reduction methods.

The AWRR method also was compared with a “nearly adaptive” rank reduction method called ORR. The results showed better quality of interpolation for ORR compared to WRR. However, the ORR method is not completely data-driven, and it is a computationally expensive algorithm.

## ACKNOWLEDGMENTS

We thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 543578-19.

## REFERENCES

- Abma, R., and Claerbout, J., 1995, Lateral prediction for noise attenuation by tx and fx techniques: *Geophysics*, **60**, No. 6, 1887–1896.
- Benaych-Georges, F., 2009, Rectangular random matrices, related convolution: *Probability Theory and Related Fields*, **144**, No. 3-4, 471–515.
- Benaych-Georges, F., and Nadakuditi, R. R., 2012, The singular values and vectors of low rank perturbations of large rectangular random matrices: *Journal of Multivariate Analysis*, **111**, 120–135.
- Carozzi, F., and Sacchi, M. D., 2021, Interpolated multichannel singular spectrum analysis: A reconstruction method that honors true trace coordinates: *Geophysics*, **86**, No. 1, V55–V70.
- Chen, Y., Zhang, D., Jin, Z., Chen, X., Zu, S., Huang, W., and Gan, S., 2016, Simultaneous denoising and reconstruction of 5-d seismic data via damped rank-reduction method: *Geophysical Journal International*, **206**, No. 3, 1695–1717.
- Ely, G., Aeron, S., Hao, N., and Kilmer, M. E., 2015, 5d seismic data completion and denoising using a novel class of tensor decompositions: *Geophysics*, **80**, No. 4, V83–V95.
- Fomel, S., 2003, Seismic reflection data interpolation with differential offset and shot continuation: *Geophysics*, **68**, No. 2, 733–744.
- Gao, J., Sacchi, M. D., and Chen, X., 2013, A fast reduced-rank interpolation method for prestack seismic volumes that depend on four spatial dimensions: *Geophysics*, **78**, No. 1, V21–V30.
- Greiner, T. L., Kolbjørnsen, O., Lie, J. E., Nilsen, E. H., Evensen, A. K., and Gelius, L., 2019, Cross-streamer wavefield interpolation using deep convolutional neural network, *in* SEG Technical Program Expanded Abstracts 2019, Society of Exploration Geophysicists, 2207–2211.
- Herrmann, F. J., 2008, Curvelet-domain matched filtering, *in* 2008 SEG Annual Meeting, OnePetro.

- Jin, Y., Wu, X., Chen, J., Han, Z., and Hu, W., 2018, Seismic data denoising by deep-residual networks, *in* SEG Technical Program Expanded Abstracts 2018, Society of Exploration Geophysicists, 4593–4597.
- Kreimer, N., and Sacchi, M. D., 2012, A tensor higher-order singular value decomposition for prestack seismic data noise reduction and interpolation: *Geophysics*, **77**, No. 3, V113–V122.
- Kreimer, N., Stanton, A., and Sacchi, M. D., 2013, Tensor completion based on nuclear norm minimization for 5d seismic data reconstruction: *Geophysics*, **78**, No. 6, V273–V284.
- Kumar, R., Da Silva, C., Akalin, O., Aravkin, A. Y., Mansour, H., Recht, B., and Herrmann, F. J., 2015, Efficient matrix completion for seismic data reconstruction: *Geophysics*, **80**, No. 5, V97–V114.
- Liu, B., and Sacchi, M. D., 2004, Minimum weighted norm interpolation of seismic records: *Geophysics*, **69**, No. 6, 1560–1568.
- Mandelli, S., Borra, F., Lipari, V., Bestagini, P., Sarti, A., and Tubaro, S., 2018, Seismic data interpolation through convolutional autoencoder, *in* SEG Technical Program Expanded Abstracts 2018, Society of Exploration Geophysicists, 4101–4105.
- Mikhailiuk, A., and Faul, A., 2018, Deep learning applied to seismic data interpolation, *in* 80th EAGE Conference and Exhibition 2018, vol. 2018, European Association of Geoscientists & Engineers, 1–5.
- Nadakuditi, R. R., 2014, Optshrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage: *IEEE Transactions on Information Theory*, **60**, No. 5, 3002–3018.
- Naghizadeh, M., and Sacchi, M., 2013, Multidimensional de-aliased cadzow reconstruction of seismic records: *Geophysics*, **78**, No. 1, A1–A5.
- Naghizadeh, M., and Sacchi, M. D., 2010, Robust reconstruction of aliased data using autoregressive spectral estimates: *Geophysical prospecting*, **58**, No. 6, 1049–1062.
- Oropeza, V., and Sacchi, M., 2011, Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis: *Geophysics*, **76**, No. 3, V25–V32.
- Porsani, M. J., 1999, Seismic trace interpolation using half-step prediction filters: *Geophysics*, **64**, No. 5, 1461–1467.
- Sacchi, M., 2009, Fx singular spectrum analysis: Cspg cseg cwls convention: Abstracts, Calgary, Alberta, Canada, 392–395.
- Sacchi, M. D., and Ulrych, T. J., 1995, Model re-weighted least-squares radon operators, *in* SEG Technical Program Expanded Abstracts 1995, Society of Exploration Geophysicists, 616–618.
- Sacchi, M. D., Ulrych, T. J., and Walker, C. J., 1998, Interpolation and extrapolation using a high-resolution discrete fourier transform: *IEEE Transactions on Signal Processing*, **46**, No. 1, 31–38.
- Shi, Y., Wu, X., and Fomel, S., 2020, Deep learning parameterization for geophysical inverse problems, *in* SEG 2019 Workshop: Mathematical Geophysics: Traditional vs Learning, Beijing, China, 5-7 November 2019, Society of Exploration Geophysicists, 36–40.
- Siahkoobi, A., Kumar, R., and Herrmann, F., 2018, Seismic data reconstruction with generative adversarial networks, *in* 80th EAGE Conference and Exhibition 2018, vol. 2018, European Association of Geoscientists & Engineers, 1–5.
- Spitz, S., 1991, Seismic trace interpolation in the fx domain: *Geophysics*, **56**, No. 6, 785–794.
- Stolt, R. H., 2002, Seismic data mapping and reconstruction: *Geophysics*, **67**, No. 3, 890–908.
- Trad, D., 2009, Five-dimensional interpolation: Recovering from acquisition constraints: *Geophysics*, **74**, No. 6, V123–V132.

- Trad, D. O., Ulrych, T. J., and Sacchi, M. D., 2002, Accurate interpolation with high-resolution time-variant radon transforms: *Geophysics*, **67**, No. 2, 644–656.
- Trickett, S., 2008, F-xy cadzow noise suppression, *in* SEG Technical Program Expanded Abstracts 2008, Society of Exploration Geophysicists, 2586–2590.
- Trickett, S., and Burroughs, L., 2009, Prestack rank-reducing noise suppression: Theory, *in* SEG Technical Program Expanded Abstracts 2009, Society of Exploration Geophysicists, 3332–3336.
- Wang, B., Zhang, N., Lu, W., and Wang, J., 2019, Deep-learning-based seismic data interpolation: A preliminary result: *Geophysics*, **84**, No. 1, V11–V20.
- Wang, B., Zhang, N., Lu, W., Zhang, P., and Geng, J., 2018, Seismic data interpolation using deep learning based residual networks, *in* 80th EAGE Conference and Exhibition 2018, vol. 2018, European Association of Geoscientists & Engineers, 1–5.
- Wu, J., and Bai, M., 2018, Adaptive rank-reduction method for seismic data reconstruction: *Journal of Geophysics and Engineering*, **15**, No. 4, 1688–1703.