Geophysical Journal International

Geophys. J. Int. (2016) **206,** 1831–1846
Advance Access publication 2016 June 29
GJI Marine geosciences and applied geophysics

doi: 10.1093/gji/ggw246

Wavelet-based multicomponent matching pursuit trace interpolation

Jihun Choi,[1] Joongmoo Byun,[1] Soon Jee Seol[1] and Young Kim[2]

[1] *Department of Earth Resources and Environmental Engineering, Hanyang University,* 222 *Wangsimni-Ro, Seongdong-gu,* 133-791 *Seoul, Korea. E-mail: jbyun@hanyang.ac.kr*
[2] *YK Geophysics LLC,* 174 *Elk Pass Road, Sequim, WA* 98382*, USA*

Accepted 2016 June 27. Received 2016 June 27; in original form 2015 July 29

SUMMARY

Typically, seismic data are sparsely and irregularly sampled due to limitations in the survey environment and these cause problems for key seismic processing steps such as surface-related multiple elimination or wave-equation-based migration. Various interpolation techniques have been developed to alleviate the problems caused by sparse and irregular sampling. Among many interpolation techniques, matching pursuit interpolation is a robust tool to interpolate the regularly sampled data with large receiver separation such as crossline data in marine seismic acquisition when both pressure and particle velocity data are used. Multicomponent matching pursuit methods generally used the sinusoidal basis function, which have shown to be effective for interpolating multicomponent marine seismic data in the crossline direction. In this paper, we report the use of wavelet basis functions which further enhances the performance of matching pursuit methods for de-aliasing than sinusoidal basis functions. We also found that the range of the peak wavenumber of the wavelet is critical to the stability of the interpolation results and the de-aliasing performance and that the range should be determined based on Nyquist criteria. In addition, we reduced the computational cost by adopting the inner product of the wavelet and the input data to find the parameters of the wavelet basis function instead of using L-2 norm minimization. Using synthetic data, we illustrate that for aliased data, wavelet-based matching pursuit interpolation yields more stable results than sinusoidal function-based one when we use not only pressure data only but also both pressure and particle velocity together.

Key words: Fourier analysis; Wavelet transform; Spatial analysis.

1 INTRODUCTION

Typically, seismic data are sparsely and irregularly sampled because of the high cost and field conditions in the exploration environment, such as large sail line and streamer intervals and feathering or existing production facilities. Spatial aliasing and irregularity in the data complicate seismic processing. Therefore, many interpolation techniques have been developed to prevent these problems. Among the interpolation techniques, POCS (Projection Onto Convex Sets) method (Abma & Kabir 2006; Kim *et al*. 2015), MWNI (Minimum Weighted Norm Interpolation) method (Liu & Sacchi 2004; Naghizadeh & Sacchi 2010) and ALFT (Anti-Leakage Fourier Transform) method (Schonewille *et al*. 2009; Xu *et al*. 2010) are robust tools for the irregularly sampled data. In the irregularly sampled data, the energy of the spectral components of the signal can leak at different wavenumber and the aliased replicas do not focus in the spectrum. Thus, the interpolation can be performed by reducing the leakage. However, these methods are difficult to directly apply to relatively regular data which are severely aliased such as crossline direction of marine seismic data. To interpolate the spatially aliased data, the assumption that seismic trace can be expressed by linear events has been conventionally used (Spitz 1991). This method is performed using the prediction filter obtained from the traces. In addition, most previous methods are for interpolating the data of a single physical quantity—pressure in marine seismic data. Recently, multicomponent matching pursuit methods which use pressure and its gradient data simultaneously have been developed to overcome the aliasing problem (e.g. Vassallo *et al*. 2010). In these methods, since the pressure and its gradient data together provide more information than either pressure or gradient alone, the linearity assumption does not need anymore.

The concept of matching pursuit was introduced by Mallat & Zhang (1993) in the signal-processing field. This method is considered as the CLEAN method in the astronomy field (Thiebaut & Roques 2005; Lachowicz & Done 2010). Matching pursuit is an iterative method that decomposes the original data into a set of basis functions. There are many different types of basis functions, including sinusoidal function, wavelet and curvelet. After performing the

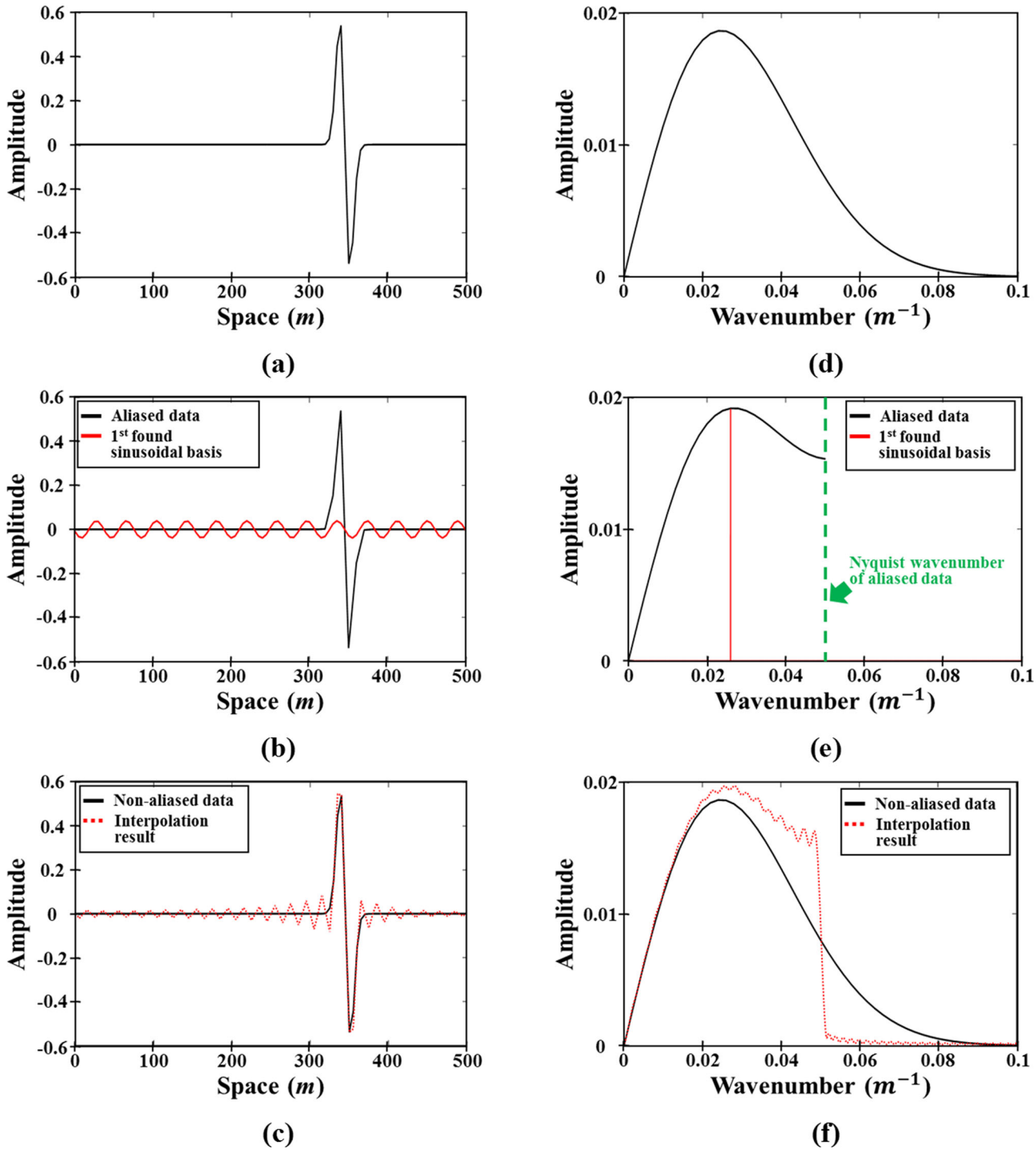

Figure 1. Simple data to test the aliasing effect of the sinusoidal basis. Views (a–c) are expressed in the space domain and (d–f) are expressed in the wavenumber domain: (a) and (d) are non-aliased data which consist of the 1st derivative Gaussian wavelet; (b) and (e) are aliased data decimated from (a) and the 1st found sinusoidal wavelet from aliased data; (c) and (f) are interpolation results using the sinusoidal function. The green dotted line in (e) indicates the Nyquist wavenumber of aliased data. The aliasing effect is represented by the summation of below and above the Nyquist energy in the Fourier domain. Since the sinusoidal functions form the basis of the Fourier series, it is transformed to impulse function in the Fourier domain. Therefore, it cannot distinguish the non-aliased energy and the aliased energy and interpolation result using sinusoidal basis cannot reconstruct the high-wavenumber energy over the Nyquist wavenumber.

matching pursuit process, discrete original data can be represented by continuous functions consisting of optimized basis functions. Then it is possible to interpolate the data at any locations from the optimized basis functions. Özbek *et al*. (2009) first applied this concept to seismic interpolation using sinusoidal basis functions for single component data. However, the matching pursuit method applied to the single component data is still susceptible to spatially aliased data. To alleviate the spatial aliasing problem, Vassallo *et al*.

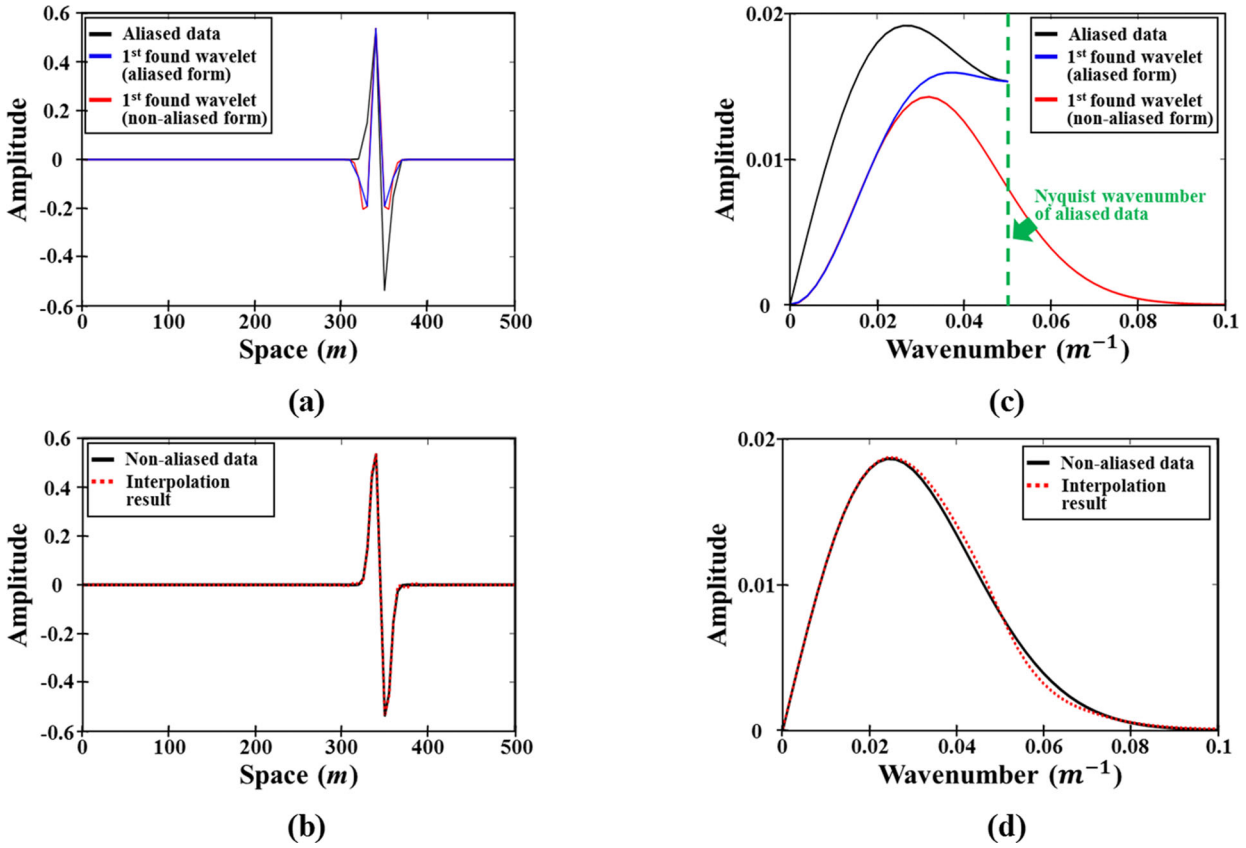

Figure 2. Test of the de-aliasing performance of the wavelet basis using data in Fig. 1. Views (a,b) are expressed in the space domain and (c,d) are expressed in the wavenumber domain: (a) and (c) are aliased data decimated from Fig. 1(a) and 1st found Ricker wavelet from aliased data; (b) and (d) are interpolation results using the wavelet. The green dotted line in (c) indicates the Nyquist wavenumber of aliased data. Since the wavelet has wavenumber bandwidth, aliased form of wavelet which is discretized for matching with aliased input data has the below and the above the Nyquist energy as shown in blue line of (c). However, through the non-aliased form (red line in c) made by the equation of wavelet, it can reconstruct high-wavenumber energy over the Nyquist wavenumber.

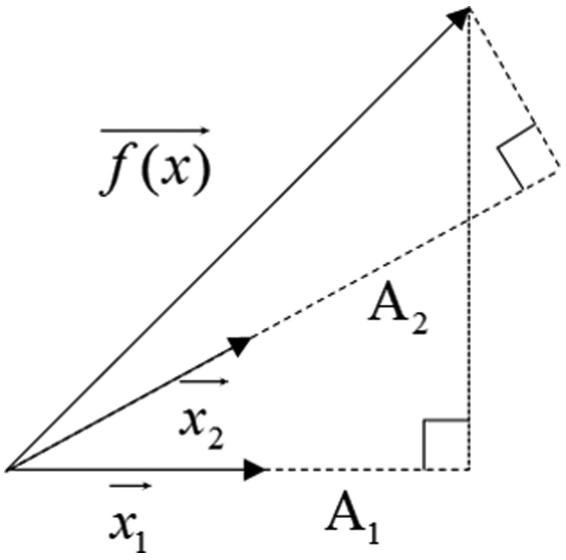

Figure 3. Inner product of the seismic data $f(\vec{x})$ and the discretized basis function vectors $\vec{x_j}$ ($j = 1, 2, \ldots, m$). The basis function closer to the seismic data has a higher inner product value. In addition, inner product values become the coefficient of the basis function.

(2010) simultaneously used pressure and its crossline gradient data, obtained from a marine seismic survey, as input data. Kamil *et al.* (2014) applied a moveout operator to reduce aliasing from higher-order aliased shallow events. These two methods provided improved interpolation result for spatially aliased data by using multicomponent data including pressure and particle velocity. In the matching pursuit method, it is desirable to use a basis function that can be best optimized for the input data. Thus far, matching pursuit methods have been developed using only sinusoidal basis functions for seismic trace interpolation. However, since wavelets have a range of wavenumbers and at the same time a limited spatial extent, they can better represent seismic data, which typically consist of localized events with a finite spatial bandwidth.

In this paper, we present a wavelet-based matching pursuit method for interpolating spatially aliased data without a large gap. In Sections 2 and 3, we describe the wavelet-based matching pursuit method for the single component data and a strategy for reducing the computational cost. Then, we extend our wavelet-based method to multicomponent data to further improve de-aliasing performance in Section 4. Next, we present how to find the optimal parameter range for successful interpolation using the wavelet-based method.

2 WAVELET-BASED MATCHING PURSUIT METHOD FOR TRACE INTERPOLATION

In the matching pursuit interpolation method, the signal $f(x)$ is expressed by the summation of basis functions, $g(x; \theta_i)$:

$$f(x) = \sum_{i=1}^{\infty} a_i g(x; \theta_i), \tag{1}$$

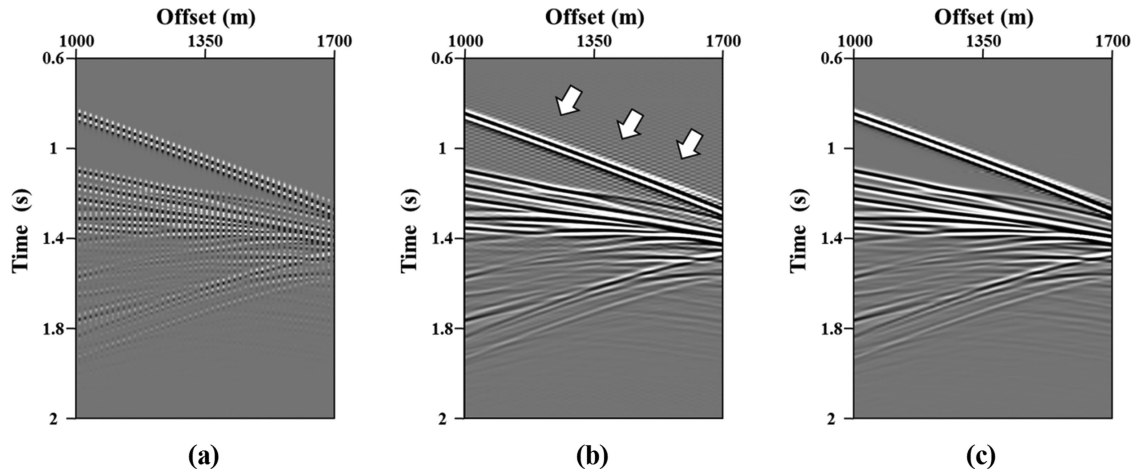

Figure 4. The comparison of interpolation results using the pressure data: (a) spatially aliased seismic data (trace interval = 15 m); (b,c) interpolation results (trace interval = 5 m) by the sinusoidal-based matching pursuit method and the wavelet-based matching pursuit method, respectively. In the case of the sinusoidal-based method, the artificial noise caused by aliasing is shown by the white arrows. In contrast, the wavelet-based method reconstructed the data with far less artefacts.

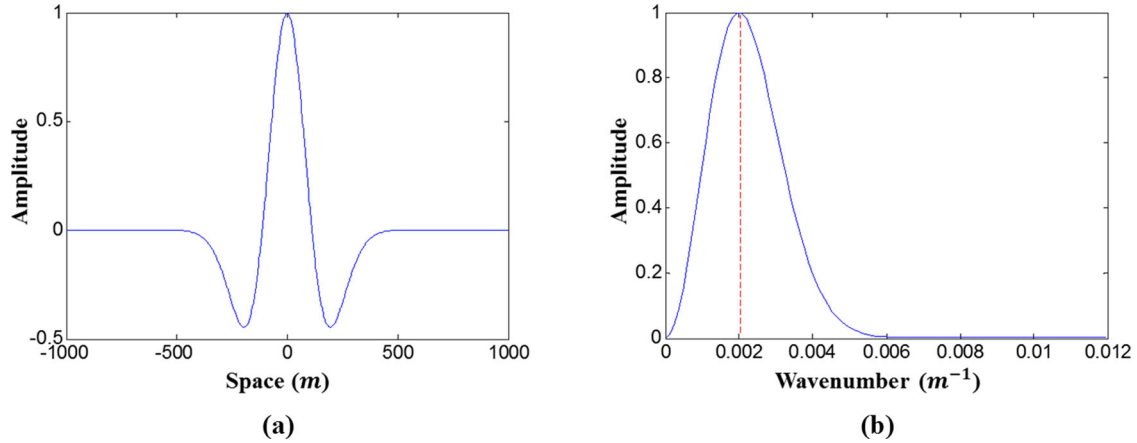

Figure 5. Ricker wavelet with a peak wavenumber of 0.002 m^{-1}: (a) in space domain; (b) its amplitude spectrum in the wavenumber domain.

where θ_i is the parameter set and a_i is the coefficient of each basis function that will be determined to match the signal at the points where the signal was measured. Once we determined the parameter set and the coefficients of the basis functions, we can compute the value of $f(x)$ at any arbitrary point.

First, the parameter set of the basis function closest to the input signal is calculated. Then, the basis function with the calculated parameter set is added to the output signal, initially set to zero, and the residual signal is obtained by subtracting the basis function from the input signal. Next, the residual signal is used as the new input signal and the same process is iteratively performed until the residual signal becomes less than a certain criterion. In the nth iteration, the above process can be expressed as

$$R_n f(x) = R_{n-1} f(x) - a_n g(x;\theta_n),$$
$$f_n(x) = f_{n-1}(x) + a_n g(x;\theta_n), \tag{2}$$

where $R_n f(x)$ indicates the residual data after the nth iteration; this is used as the input data for the $(n+1)$th iteration, and $f_n(x)$ is the output data after the nth iteration.

In previous matching pursuit interpolation, sinusoidal functions are used as basis functions. However, sinusoidal basis functions are susceptible to aliasing when only use the single component. Since the aliased energy overlaps the non-aliased energy in the Fourier domain, the sinusoidal basis function cannot separate these two energies (Fig. 1e). Fig. 1 shows the aliasing effect of the sinusoidal basis. In the Fig. 1, we use the 1st derivative Gaussian wavelet to make the input data. Figs 1(b) and (e) show the aliased data and the first found sinusoidal basis. As shown in these figures, since the sinusoidal functions form the basis of the Fourier series, found basis function is represented to impulse function in the wavenumber domain. Therefore, sinusoidal basis cannot reconstruct the high-wavenumber energy over the Nyquist wavenumber. This aspect is also shown in the interpolation result (Figs 1c and f). On the other hand, if we use a basis function which has bandwidth over the Nyquist wavenumber and non-zero values in its wavenumber range, the reconstruction of aliased energy is possible (Vaidyanathan 2001). Thus, wavelet can be solution to alleviate the aliasing problem because it has those properties. Fig. 2 shows the de-aliasing performance of the wavelet basis. In this figure, we use the Ricker wavelet as the basis function of matching pursuit method. In the matching pursuit process, in order to match the wavelet to aliased data, the wavelet is converted to aliased discrete form. Since the wavelet has wavenumber bandwidth, its aliased form includes not only below but also above the Nyquist energy like blue line of Figs 2(a) and (c). However, we already know its non-aliased

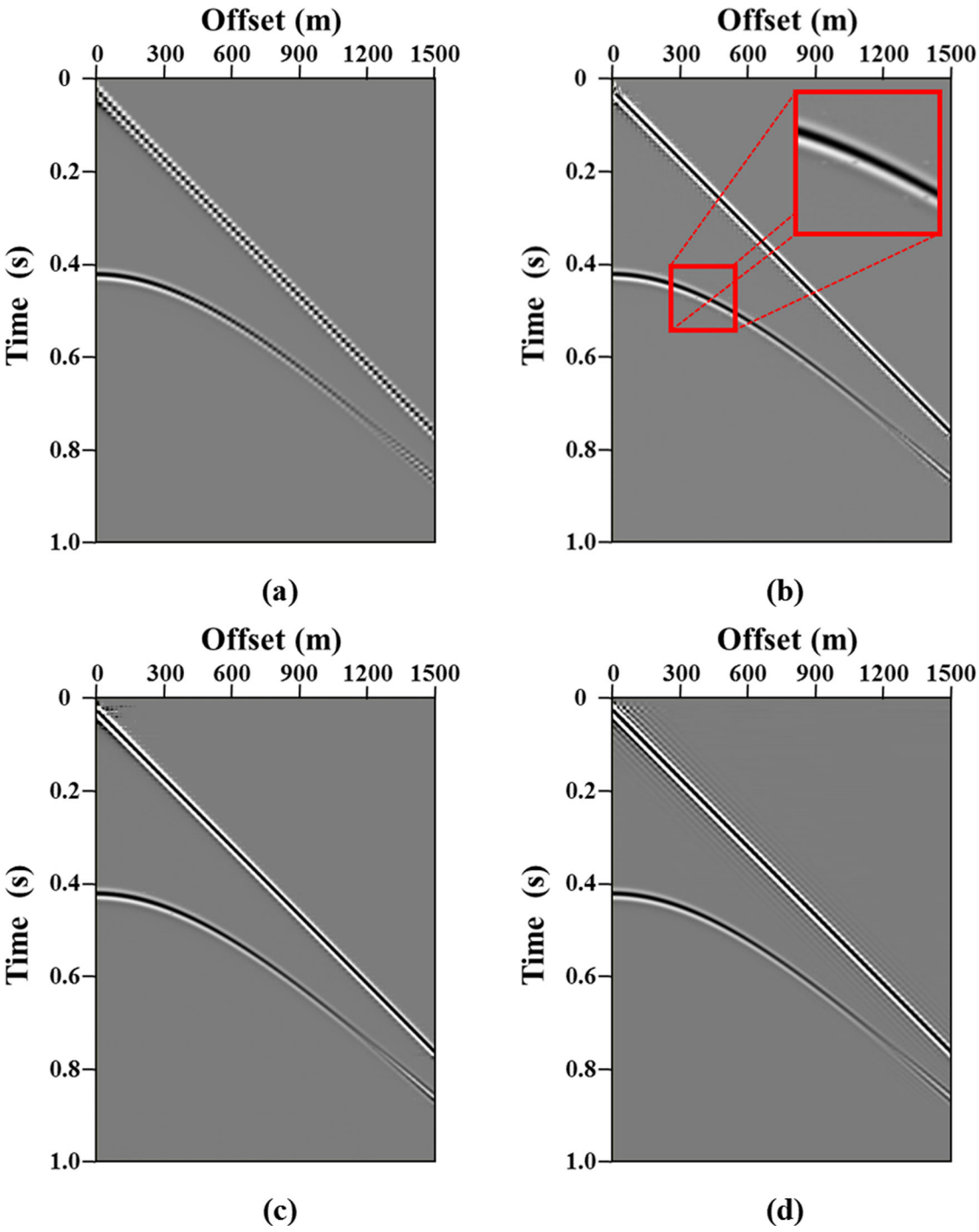

Figure 6. Comparison of the wavelet-based multicomponent matching pursuit interpolation results depending on parameter, $n_{\text{grid,multi}}$: (a) spatially aliased input data; (b) interpolated data using $n_{\text{grid,multi}} = 1$. The red box shows an enlarged image of the artificial noise; (c) interpolated data using $n_{\text{grid,multi}} = 2$; (d) interpolated data using $n_{\text{grid,multi}} = 6$.

continuous form through its equation, we can separate aliased and non-aliased energies when its aliased form has the uniqueness as shown in red line of Fig. 2(c). Therefore, the interpolation result of the wavelet shown in the Figs 2(b) and (d) are almost identical to the reference data. Wavelet basis functions have another advantage over sinusoidal basis functions in that the wavelet has a limited spatial extent. In case of the sinusoidal function, since the improperly found basis function creates artificial noise widely, the time-space windowing strategy has to be carefully applied (Naghizadeh & Innanen 2011). However, careful windowing is not required when we use wavelet basis function.

There are various wavelet functions that can be used as basis functions such as the Ricker, Morlet and Ormsby wavelets. In this study, we used the Ricker wavelet which has a simple equation to apply. Its equation consists of a peak wavenumber, which is related to duration, and a spatial shift. Since Morlet or Ormsby wavelets which have more parameters than the Ricker wavelet, the process of finding the basis function matching the input signal becomes more

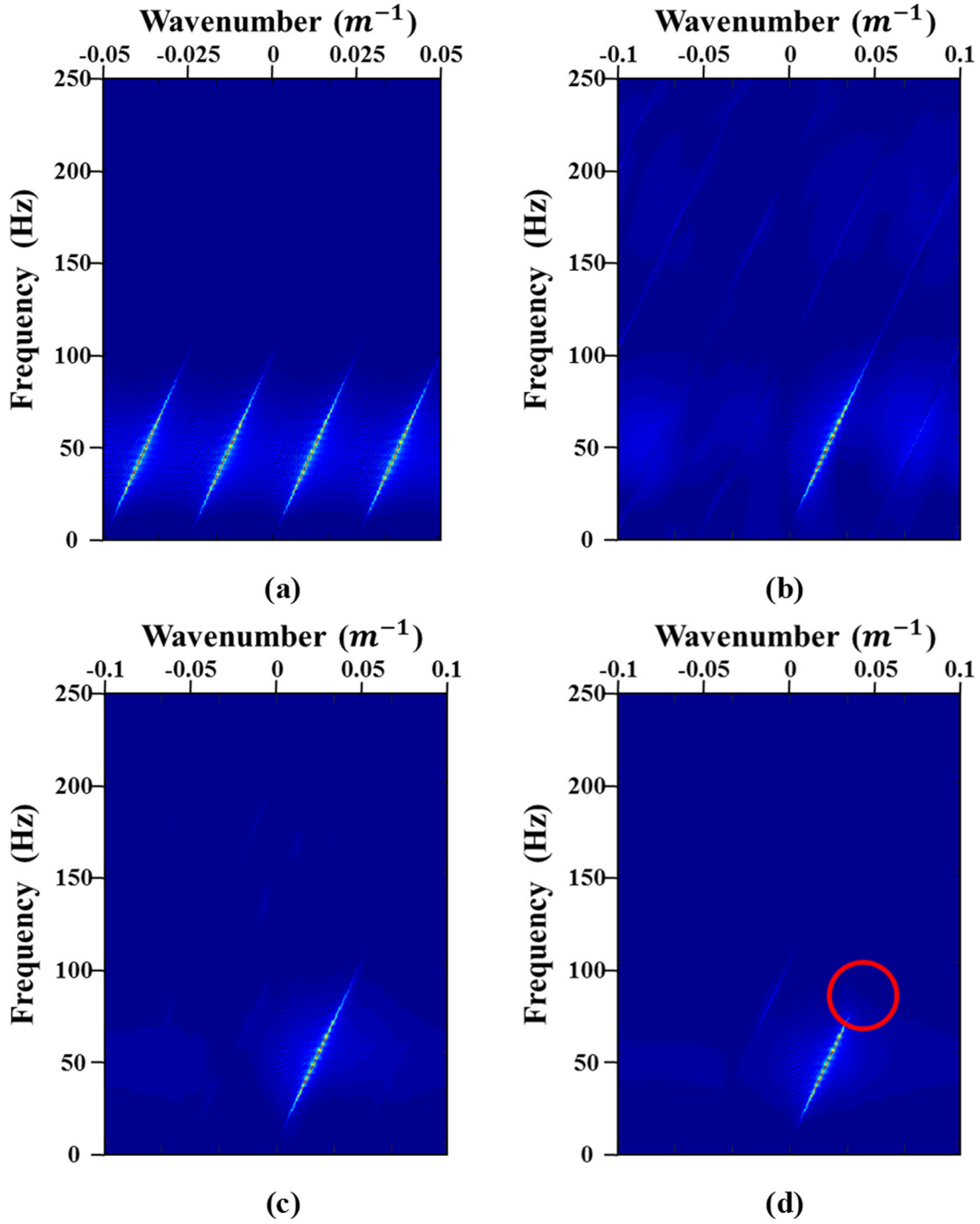

Figure 7. Comparison of the *f–k* spectrums of the input data and interpolation results, depending on parameter $n_{\text{grid,multi}}$ shown in Fig. 6: (a) spatial aliased input data; (b) interpolated data using $n_{\text{grid,multi}} = 1$; (c) interpolated data using $p\ n_{\text{grid,multi}} = 2$; (d) interpolated data using $n_{\text{grid,multi}} = 6$. In this case, high-wavenumber energy that cannot be reconstructed is shown in the red circle.

time consuming. The equation of Ricker wavelet is expressed as follows:

$$g(x;\theta_i) = (1 - 2\pi^2 {k_{\text{maj},i}}^2 (x - \tau_{xi})^2) \exp(-\pi^2 {k_{\text{maj},i}}^2 (x - \tau_{xi})^2), \tag{3}$$

where $k_{\text{maj},i}$ is the peak wavenumber of the ith Ricker wavelet and τ_{xi} is its spatial shift in the x-direction.

3 STRATEGY FOR REDUCING THE COMPUTATIONAL COST

To find the coefficient of the sinusoidal basis function, L-2 norm minimization is typically used in matching pursuit interpolation methods (Özbek *et al*. 2009; Vassallo *et al*. 2010). In these methods, the number of parameters that has to be determined is three (two amplitudes and a wavenumber). Among these parameters, amplitudes are simply calculated from the determined wavenumber.

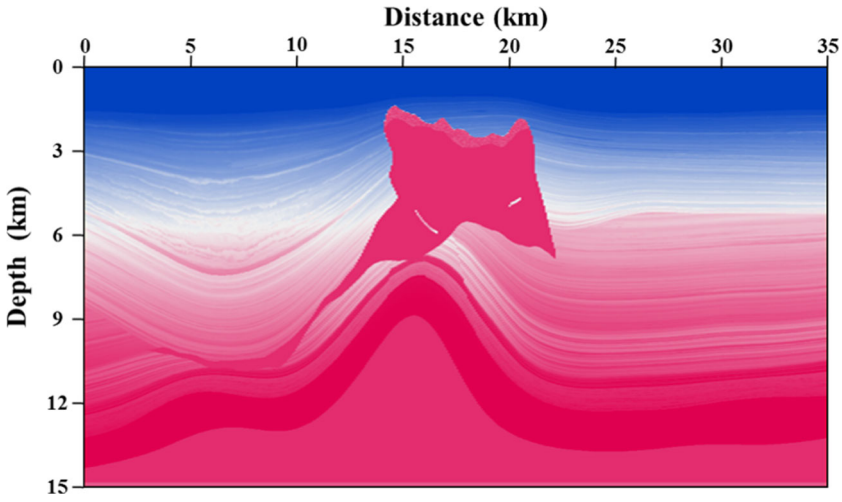

Figure 8. A 2-D vertical section at 20 km in the y-direction, extracted from SEAM (SEG Advanced Modeling Program) Phase 1 velocity model.

Likewise, the number of parameters of a wavelet basis function is three (amplitude, peak wavenumber and spatial shift) but the amplitude has to be calculated by other two parameters. Thus, since the number of parameters that has to be previously determined in the wavelet is greater than that of the sinusoidal function, we utilize a different method from L-2 norm minimization. We use the inner product of the wavelet and the input data obtained along the spatial direction, to find the parameters, following Mallat & Zhang (1993). First, we make wavelet basis functions corresponding to the peak wavenumber, $k_{\mathrm{maj},i}$, and spatial shift τ_{xi}. Second, the wavelet basis functions are discretized at the same locations where the input seismic traces are located, to compute the inner product. A group of discretized basis functions is called a 'dictionary'. Next, each discretized basis function in the dictionary is normalized by the amplitude of the vector (discretized basis function vector) whose components are sample values of the discretized basis function. After these processes, we calculate the coefficient of the basis functions using the inner product. The inner product value of the seismic data, $\overrightarrow{f(x)}$, and the discretized basis function vectors, $\overrightarrow{x_j}$ $(j = 1, 2, \ldots, m)$, indicate the goodness of fit of each discretized basis function to the seismic data, as shown in Fig. 3. Therefore, when the best-fit discretized basis function for the seismic data is found, the parameters of the selected basis function (wave number and spatial shift) can be obtained from the dictionary. In addition, the coefficient of the basis function can be estimated from the value of the inner product, because the discretized basis function vector used in the inner product corresponds to the unit vector. Thus, the coefficient of the optimized discretized basis function, a_n, in the nth iteration becomes

$$a_n = \langle R_{n-1} f(x), \overline{x_n} \rangle, \tag{4}$$

where $R_{n-1}f(x)$ means the residual data in the previous iteration, $\overline{x_n}$ means optimized discretized basis function in the nth step and $\langle \cdot, \cdot \rangle$ represents the inner product operator. Then, using eq. (4), eq. (2) becomes

$$R_n f(x) = R_{n-1} f(x) - \langle R_{n-1} f(x), \overline{x_n} \rangle \overline{x_n},$$
$$f_n(x) = f_{n-1}(x) + \langle R_{n-1} f(x), \overline{x_n} \rangle \overline{x_n}. \tag{5}$$

To interpolate seismic data, however, eq. (5) must be expressed by the continuous basis function, $g(x; \theta_n)$, instead of the discretized basis function $\overline{x_n}$. This discretized basis function does not have the same amplitude as the continuous basis function because of normalization. Therefore, the normalization value must be compensated for as follows:

$$R_n f(x) = R_{n-1} f(x) - \langle R_{n-1} f(x), \overline{x_n} \rangle \frac{g(x; \theta_n)}{h_n},$$
$$f_n(x) = f_{n-1}(x) + \langle R_{n-1} f(x), \overline{x_n} \rangle \frac{g(x; \theta_n)}{h_n}, \tag{6}$$

where h_n is the amplitude ratio of the continuous basis function to the normalized basis function.

To find the best-fit basis function in each iteration, all basis functions in the dictionary must be conducted with the inner product of the residual input data from the previous iteration. Using eqs (4) and (5), the inner product value in the next iteration can be calculated as follows:

$$a^{\mathrm{Dic}}_{j,n+1} = \left\langle R_{n-1} f(x) - a_n \overline{x_n}, \overline{x^{\mathrm{Dic}}_{j,n+1}} \right\rangle, \quad (j = 1, 2, 3, \ldots) \tag{7}$$

where $a^{\mathrm{Dic}}_{j,n+1}$ and $\overline{x^{\mathrm{Dic}}_{j,n+1}}$ are inner product values and whole basis functions in the dictionary in the next iteration, respectively. Because the inner product has linearity, eq. (7) can be converted into

$$a^{\mathrm{Dic}}_{j,n+1} = \left\langle R_{n-1} f(x), \overline{x^{\mathrm{Dic}}_{j,n+1}} \right\rangle - \left\langle a_n \overline{x_n}, \overline{x^{\mathrm{Dic}}_{j,n+1}} \right\rangle. \tag{8}$$

Because basis functions in the dictionary do not change by iteration, the subscript of the basis functions in the dictionary, $\overline{x^{\mathrm{Dic}}_{j,n+1}}$, that represent the iteration, $n + 1$, can be ignored. Therefore, the equation finally becomes

$$a^{\mathrm{Dic}}_{j,n+1} = a^{\mathrm{Dic}}_{j,n} - \left\langle a_n \overline{x_n}, \overline{x^{\mathrm{Dic}}_{j}} \right\rangle. \tag{9}$$

Using eq. (9), we can calculate the inner product value in the next iteration more efficiently. Because the duration of the wavelet in the space domain is limited, many inner product values of the previously found basis function, and the basis functions in the dictionary, become zero. Thus, we can reduce the computational cost by skipping the unnecessary calculation.

Fig. 4 compares the interpolation results using the pressure data between the sinusoidal-based matching pursuit interpolation method and the wavelet-based one. Panel (a) shows the spatially aliased seismic data. Using a sinusoidal basis function, when the matching pursuit decomposition process is carried out for spatially aliased seismic data, the aliased energy of the seismic data cannot

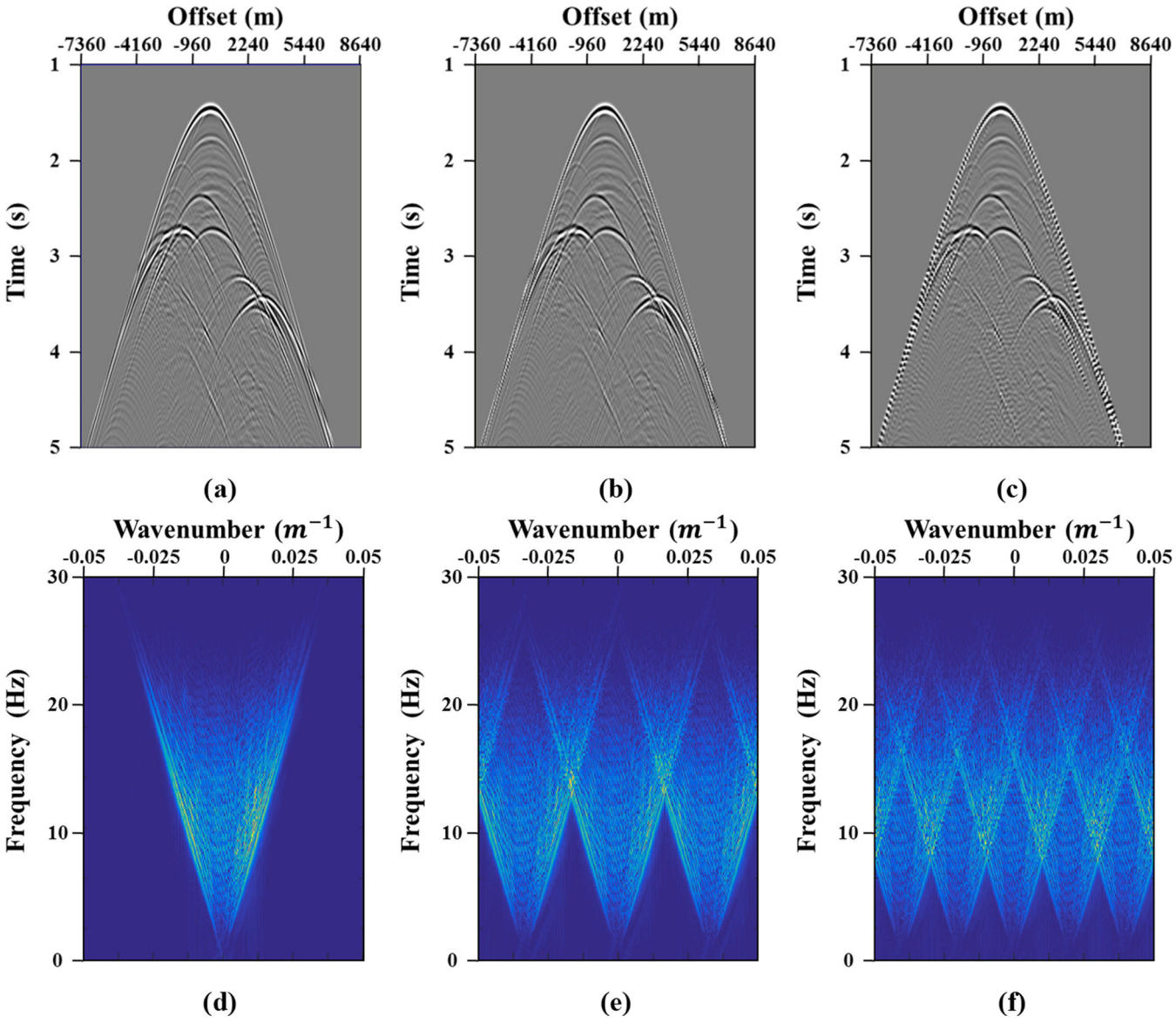

Figure 9. Decimated input data to test the performance of the wavelet-based multicomponent matching pursuit interpolation method. Views (a–c) are expressed in the *t*–*y* domain and (d–f) are expressed in the *f*–*k* domain: (a) and (d) are original data (trace interval = 20 m); (b) and (e) are decimated data sampled with 60 m spacing and its *f*–*k* spectrum, respectively; (c) and (f) are decimated data sampled with 100 m spacing and its *f*–*k* spectrum, respectively.

be properly distinguished. As a result, improperly found basis functions made artificial noise, as shown by the white arrows in panel (b). In contrast, the wavelets reduced the artificial noise caused by aliased energy. In addition, since the duration of the wavelet in the space domain is limited, the remaining aliased energy has less influence on other events. Fig. 4(c) shows the improved interpolation result and illustrates the strength of the wavelet basis function.

4 WAVELET-BASED INTERPOLATION BY MATCHING PURSUIT USING MULTICOMPONENT STREAMER DATA

Although wavelet-based matching pursuit yields better interpolation results than sinusoidal-based matching pursuit when seismic data are aliased, it is still limited when seismic data are severely aliased in the spatial direction. To alleviate this problem, we followed the idea of using multicomponent data (Vassallo *et al*. 2010; Kamil *et al*. 2014). Vassallo *et al*. (2010) applied the sinusoidal-based matching pursuit method to pressure and its crossline gradient to improve de-aliasing performance. Because the aliasing problem is related to the uniqueness problem when sampled data are expressed by any basis function, additional data (e.g. the crossline gradient) can also help wavelet-based matching pursuit to overcome the aliasing problem. Therefore, to improve de-aliasing performance, this concept is applied to the wavelet-based matching pursuit method developed in this study. In the case of a marine survey, because there is no shear wave, the equation of motion can be expressed,

$$\frac{\partial P}{\partial y} = -\rho \frac{\partial V_y}{\partial t}, \tag{10}$$

where P is the pressure, V_y is the particle velocity in crossline direction, y is the crossline distance and ρ is density. The differentiation of pressure in eq. (10) can be expressed as the sum of basis functions

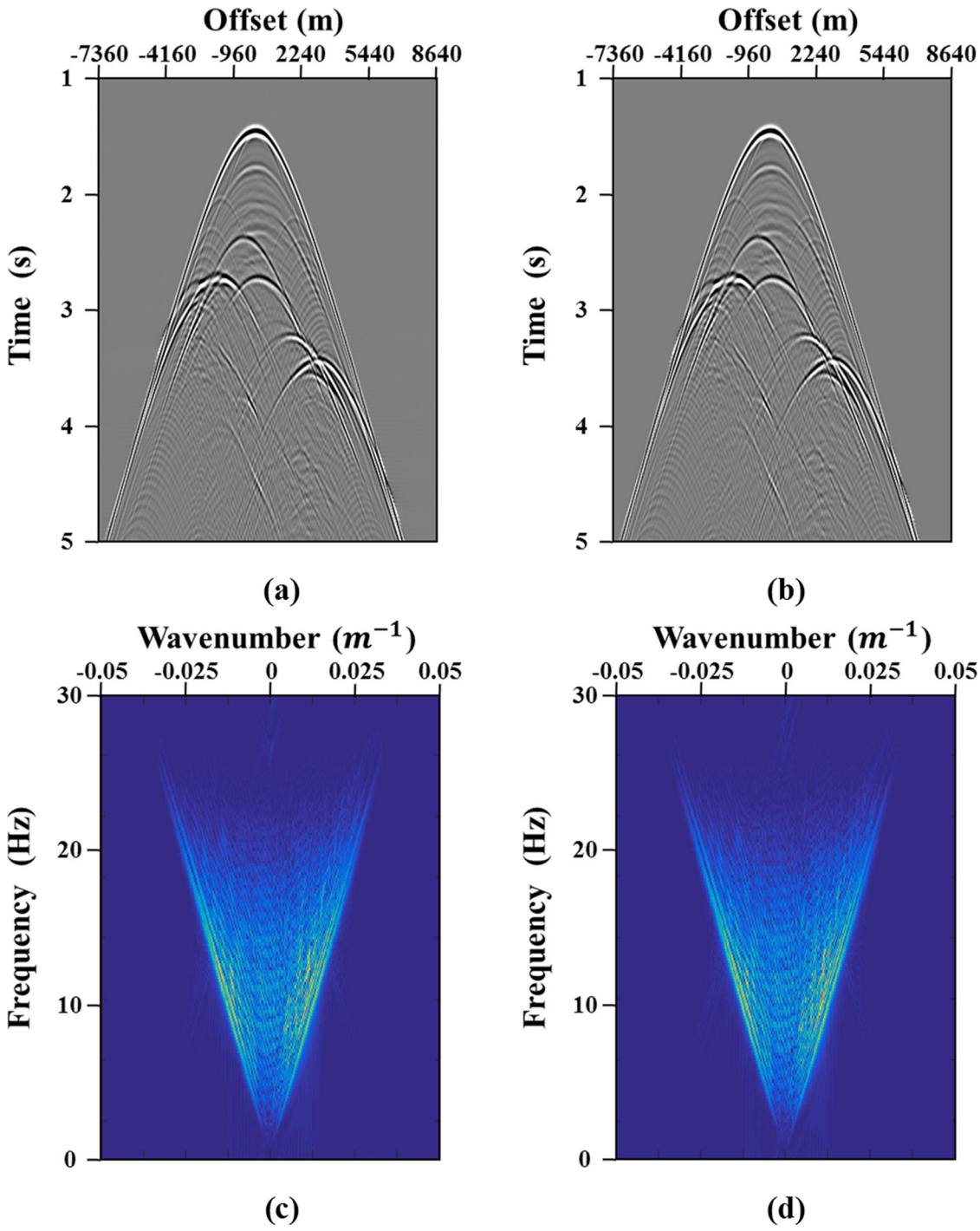

Figure 10. Comparison of the sinusoidal-based multicomponent matching pursuit interpolation result and the wavelet-based result for decimated data with 60 m spacing: (a,c) Sinusoidal-based interpolation result (trace interval = 20 m) and its *f–k* spectrum, respectively; (b,d) Wavelet-based interpolation result (trace interval = 20 m) and its *f–k* spectrum, respectively.

that are expressed as the partial differentiation of $g(y;\theta_i)$ like eq. (1) as follows:

$$\frac{\partial f(y)}{\partial y} = \sum_{i=1}^{\infty} a_i \frac{\partial g(y;\theta_i)}{\partial y}. \tag{11}$$

When we use the Ricker wavelet, described in eq. (3), as a basis function, the basis function matched with the crossline gradient data becomes

$$\frac{\partial g(y,\theta_i)}{\partial y} = (-2\pi^2 k_i^2 (y-\tau_{yi}))(3-2\pi^2 k_i^2 (y-\tau_{yi})^2) \times \exp(-\pi^2 k_i^2 (y-\tau_{yi})^2). \tag{12}$$

In addition, we have to consider the balancing between pressure and its derivative because they have different energy levels. We used the weighting factor suggested by Vassallo *et al*. (2010). This optimal

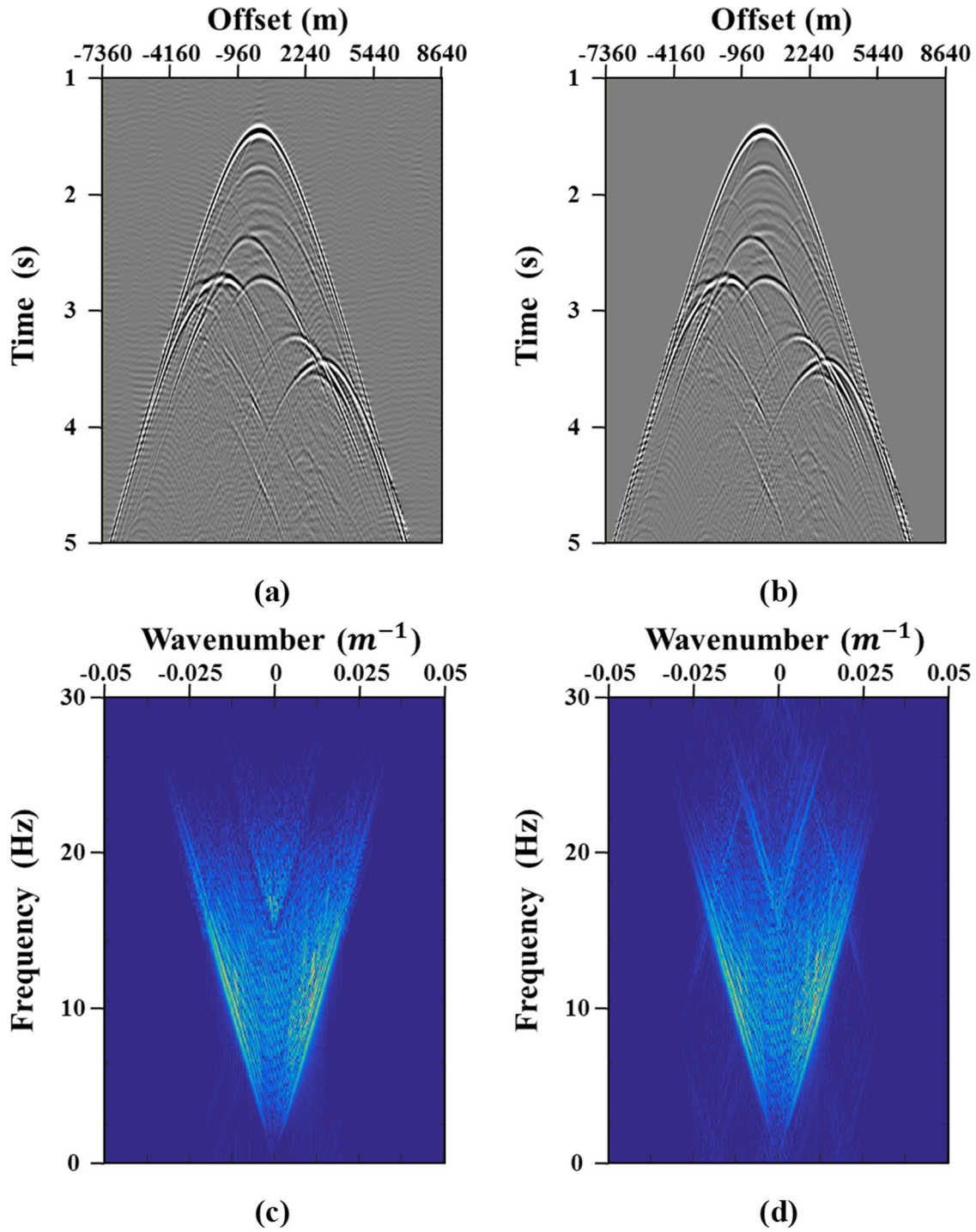

Figure 11. Comparison of the sinusoidal-based multicomponent matching pursuit interpolation result and the wavelet-based result for the decimated data with 100 m spacing. (a) and (c) present the sinusoidal-based interpolation result (trace interval = 20 m) and its *f–k* spectrum, respectively. (b) and (d) present the wavelet-based interpolation result (trace interval = 20 m) and its *f–k* spectrum, respectively.

weighting factor was determined based on the energy ratio of two data sets and the signal-to-noise ratio (SNR):

$$\lambda \propto \sqrt{E\left[\frac{P^2}{P_y^2}\right]\frac{\mathrm{SNR}(P_y)}{\mathrm{SNR}(P)}}. \tag{13}$$

Next, to use two components simultaneously in the inner product, the row vectors of the input data and the basis functions of the dictionary must be composed of two components as follows

$$f(y) = [P(y) \quad \lambda V_y(y)],$$
$$g'(y;\theta_i) = \left[g(y;\theta_i) \quad \lambda\frac{\partial g(y;\theta_i)}{\partial y}\right], \tag{14}$$

where $g'(y;\theta_i)$ is the rearranged basis function. The subsequent processes are similar to the single component case. First, we find the parameters of the best-fit basis function using the inner product

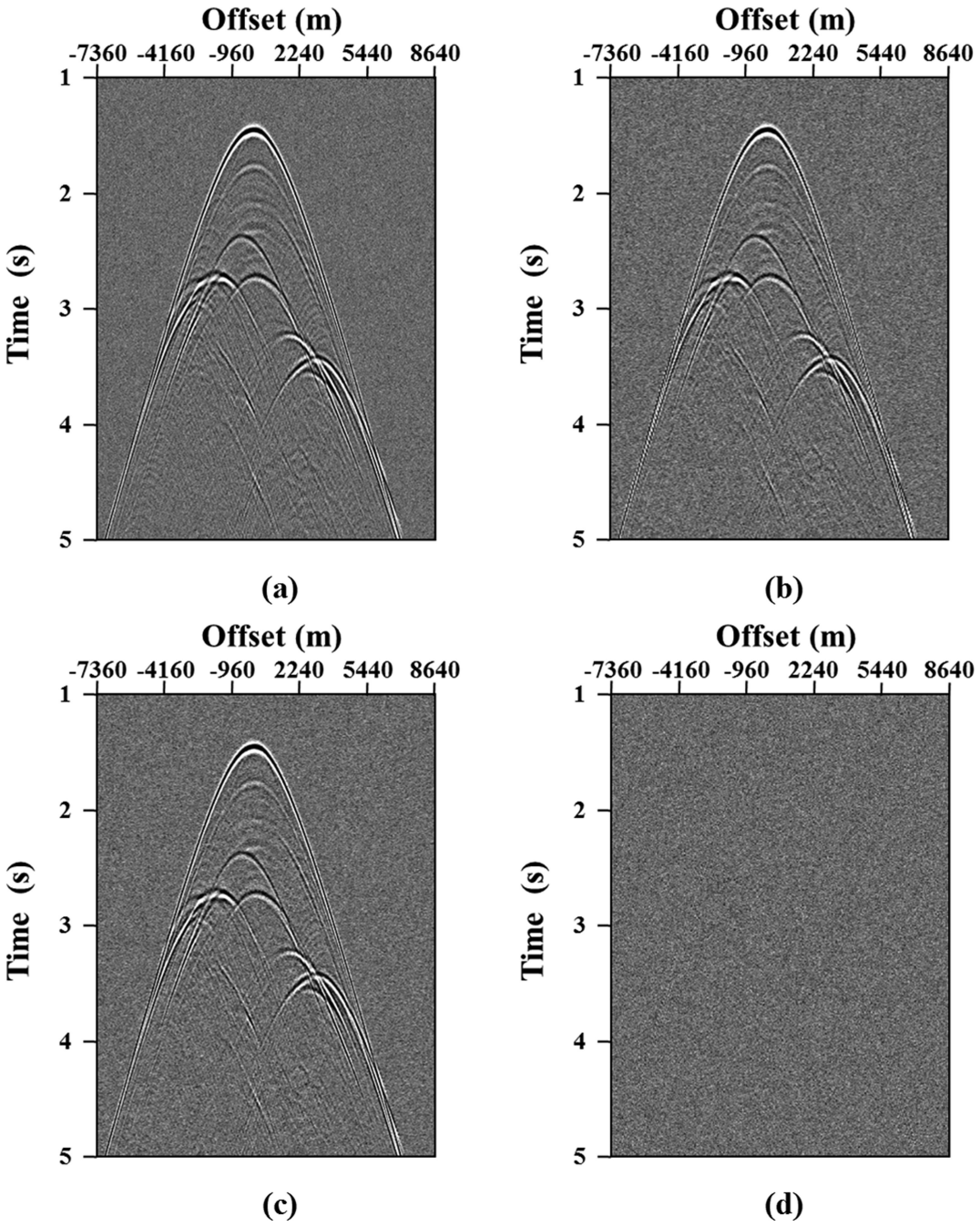

Figure 12. Interpolation result for noisy data: (a) original reference data (trace interval = 20 m); (b) decimated input data (trace interval = 60 m); (c) interpolation result of wavelet-based multicomponent matching pursuit method (trace interval = 20 m); (d) differences between the reference data and the interpolation result.

of the input data $f(y)$ and the normalized basis function of $g'(y;\theta_i)$. Then, the residual data and the output data are calculated iteratively while the energy of the residual data sufficiently converges to zero. Finally, the amplitude of the crossline particle velocity data ($V_y(y)$) is restored, considering the weighting factor λ.

5 CONSTRUCTION OF BASIS FUNCTION DICTIONARY

When we make the dictionary of the basis functions, wavelets with various spatial shifts and peak wavenumbers are used. In this process, the ranges of the two parameters are important because they are related to the decomposition capability. In the case of spatial shift, the maximum value has to be determined to cover the whole distance of the seismic data in the trace interpolation. Then, the range of the spatial shift is easily set from 0 to the maximum distance of the seismic data. In contrast, the maximum value of the peak wavenumber should be decided carefully. Fig. 5 shows the Ricker wavelet with a peak wavenumber of 0.002 m^{-1} in the space domain and its amplitude spectrum in the wavenumber domain. The peak wavenumber of the wavelet is related to its duration in the space domain. The duration is inversely proportional to the peak wavenumber. In addition, the maximum wavenumber of the wavelet

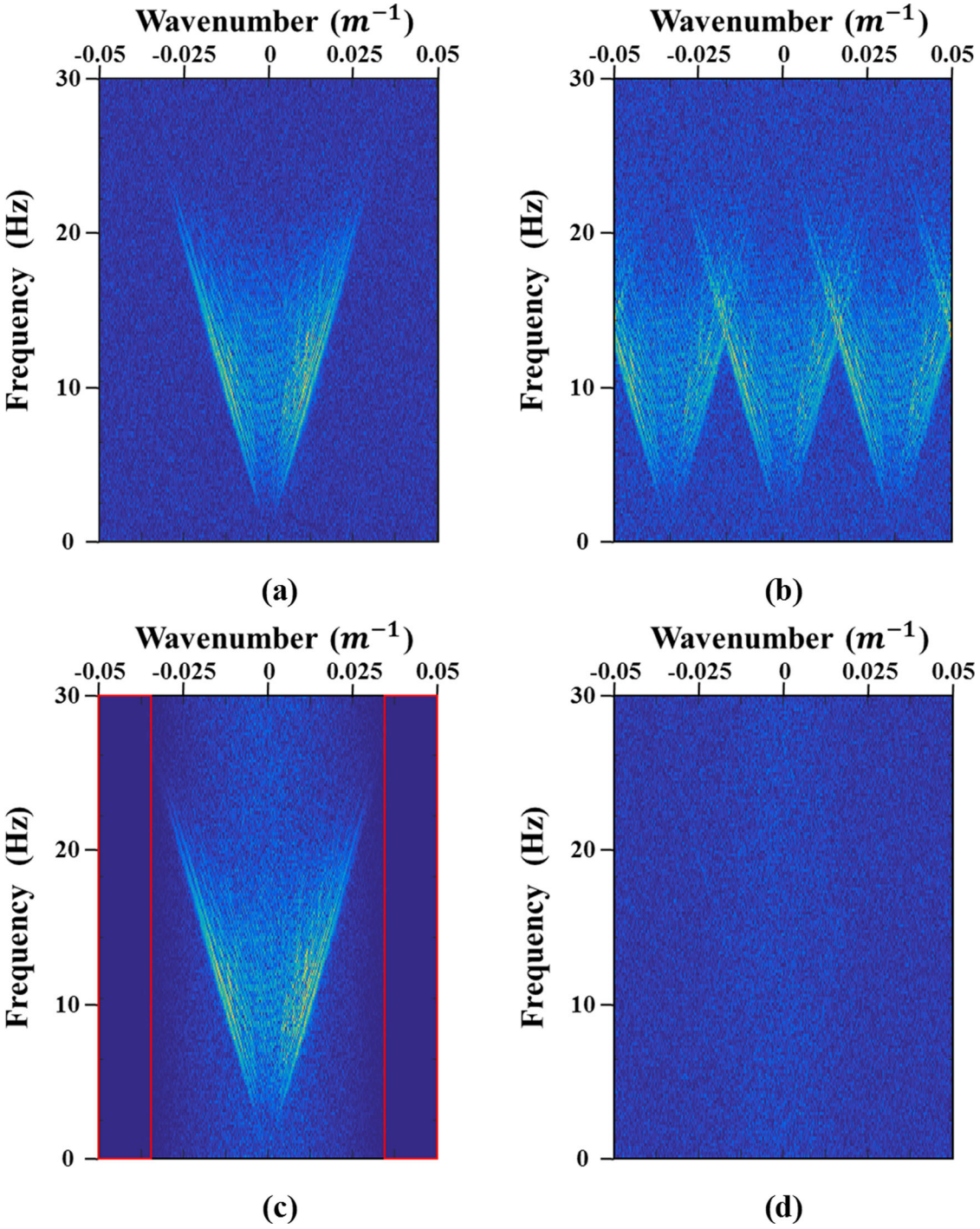

Figure 13. The *f*–*k* spectrum of interpolation result for noisy data shown in Fig. 12: (a) original reference data; (b) decimated input data; (c) interpolation result of wavelet-based multicomponent matching pursuit method; (d) differences between the reference data and the interpolation result.

is proportional to its peak wavenumber. Therefore, if the maximum peak wavenumber is too low, the basis functions of the dictionary cannot cover the maximum wavenumber of the input data. In contrast, if the maximum peak wavenumber is too high, an artificial error can occur because the discretized basis function vector used in the inner product has an insufficient number of samples to properly represent the continuous basis functions with small durations in the space domain. Thus, the optimal value of the maximum peak wavenumber for the dictionary is the peak wavenumber with which wavelet in the space domain can be represent with the minimum number of the samples. At this time, the sampling interval in the space domain is the trace interval of the input data.

When the peak wavenumber of the Ricker wavelet is k_{maj}, its bandwidth in the wavenumber domain ranges from 0 to $3k_{\mathrm{maj}}$ and its duration ranges from $-0.7797/k_{\mathrm{maj}}$ to $0.7797/k_{\mathrm{maj}}$ in the space domain (Ryan 1994). To find the optimal value of the maximum peak wavenumber, $k_{\mathrm{maj,max}}$, we suggest the following condition:

$$\frac{2 \cdot 0.7797}{k_{\mathrm{maj,\ max}}} \geq n_{\mathrm{grid}}\, dy, \tag{15}$$

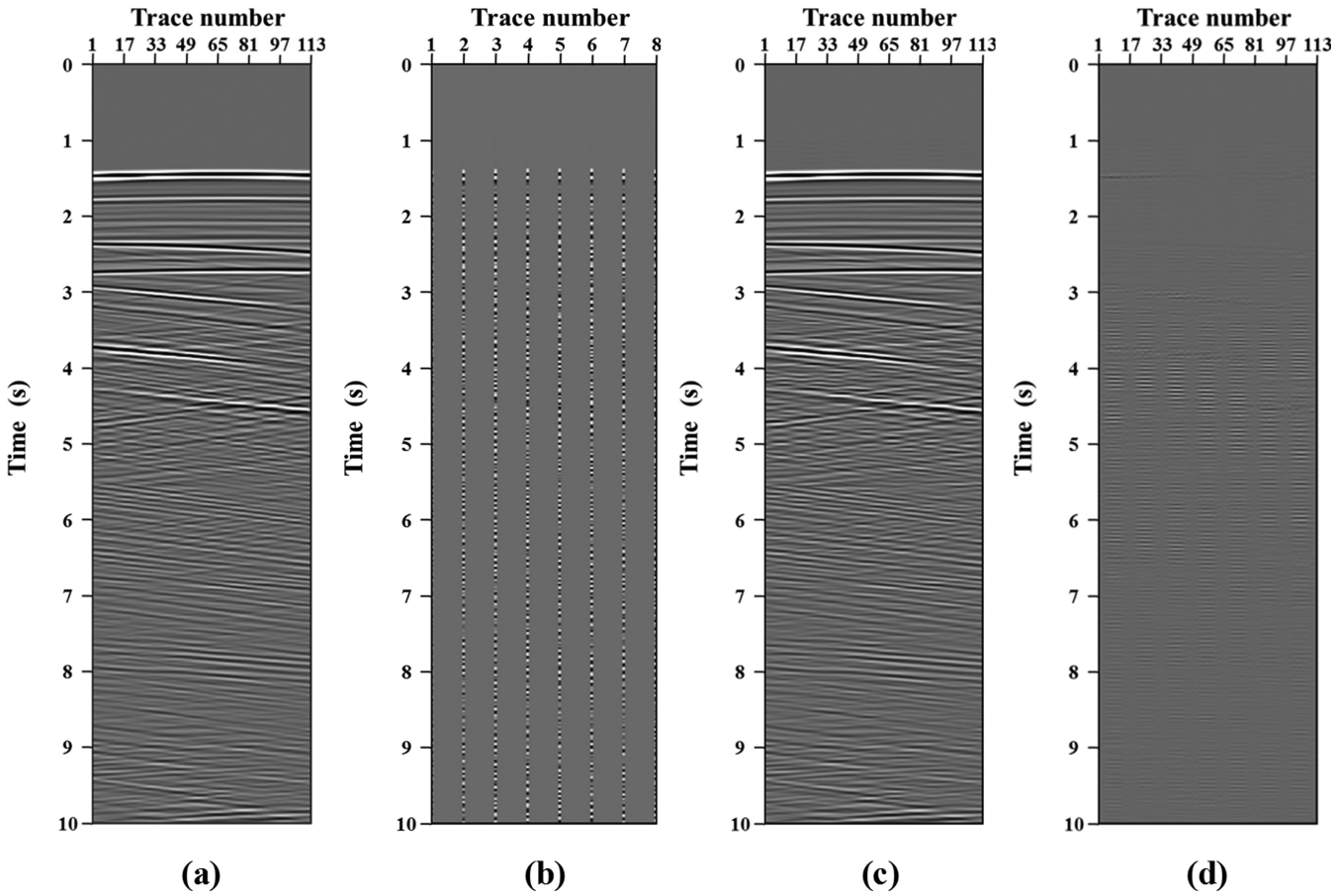

Figure 14. Interpolation result of the wavelet-based multicomponent matching pursuit method using only eight traces: (a) original reference data; (b) decimated input data; (c) interpolated data; (d) the difference between (a) and (c).

where n_{grid} and dy are the minimum number of samples of the wavelet basis function vector that can uniquely represent the continuous wavelet and the trace interval of the input data, respectively.

Eq. (15) can be rearranged as follows:

$$\frac{1.5594}{n_{\text{grid}}\, dy} \geq k_{\text{maj, max}}. \tag{16}$$

On the other hand, our method uses not only the pressure but also the particle velocity. In this case, multichannel sampling halves the required sampling rate to avoid aliasing theoretically (Linden 1959; Vassallo *et al*. 2010). Therefore, $n_{\text{grid,multi}}$, the minimum number of samples of the wavelet basis function vector uniquely representing the continuous wavelet when using both pressure and particle velocity data will be half of n_{grid} when using either pressure or gradient alone.

To find the optimal $n_{\text{grid,multi}}$ value of the Ricker wavelet, we analyse the interpolation result depending on $n_{\text{grid,multi}}$. Fig. 6 compares the interpolation results depending on the parameter $n_{\text{grid,multi}}$ in the t–y domain, and Fig. 7 shows their f–k spectrums. The trace intervals of the input decimated data and the reconstructed data are 20 and 5 m, respectively. Fig. 6(a) shows the spatially aliased input data, and Fig. 7(a) clearly demonstrates the aliasing effect in the f–k spectrum. The interpolation result when $n_{\text{grid,multi}} = 1$ is shown in Fig. 6(b). When $n_{\text{grid,multi}} = 1$, the maximum peak wavenumber, $k_{\text{maj,max}}$, becomes 0.07797 m^{-1} according to eq. (16). As shown in the red box of Fig. 6(b), artificial noises are created. From the result, one sample ($n_{\text{grid,multi}} = 1$) is insufficient to delineate the Ricker wavelet and the basis functions in the dictionary cannot decompose the aliased data perfectly due to the uniqueness problem. When $n_{\text{grid,multi}}$ and $k_{\text{maj,max}}$ are 2 and 0.038985 m^{-1}, respectively, a successful interpolation result can be obtained (Figs 6c and 7c). That means that two samples are sufficient to delineate the wavelet. Figs 6(d) and 7(d) shows the interpolation result when $n_{\text{grid,multi}}$ and $k_{\text{maj,max}}$ are 6 and 0.012995 m^{-1}, respectively. The matching pursuit decomposition process is stably carried out because $n_{\text{grid,multi}}$ is sufficient. However, the used $k_{\text{maj,max}}$ is too small to recover the high wavenumber component over 1.5 times the Nyquist wavenumber of the input data as shown in the red circle of Fig. 7(d). Consequently, for the Ricker wavelet, 2 is the optimal $n_{\text{grid,multi}}$. In other words, continuous form of Ricker wavelet can be delineated with only 2 samples when we simultaneously use the pressure and the particle velocity data. Therefore, eq. (16) finally becomes

$$\frac{1.5594}{2dy} \geq k_{\text{maj,max}}, \tag{17}$$

and this condition can be applied to data with any trace interval, dy, as long as the Ricker wavelets are used as basis functions.

As described in above, $k_{\text{maj,max}}$ is related to de-aliasing performance of wavelet. In case of the Ricker wavelet, since the bandwidth of the Ricker wavelet in the wavenumber domain ranges from 0 to $3k_{\text{maj}}$, the de-aliasing capability of multicomponent wavelet-based matching pursuit method is up to 4.6 times of Nyquist wavenumber of input data.

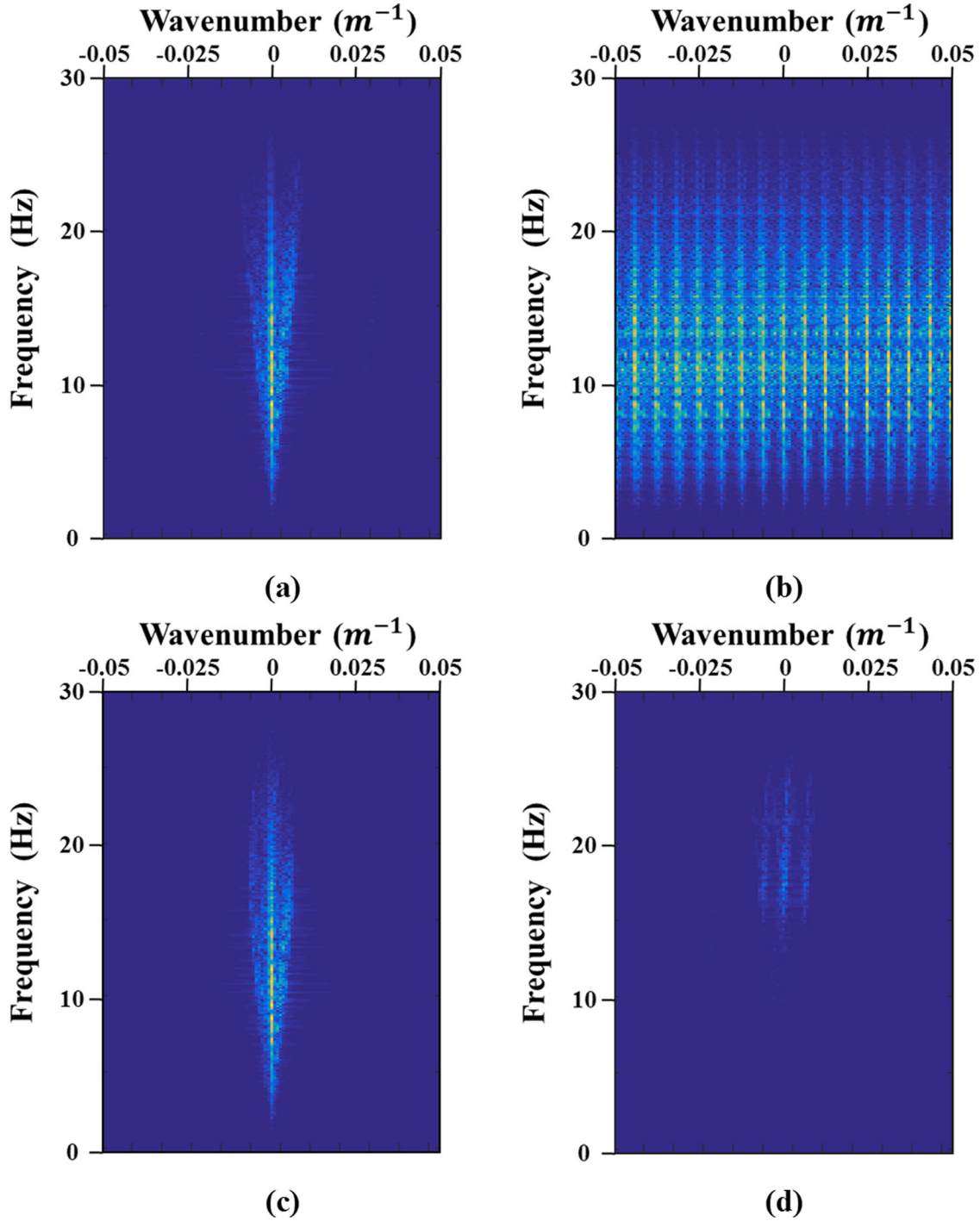

Figure 15. The *f–k* spectrum of interpolation result of the wavelet-based multicomponent matching pursuit method when the number of traces in the input data is only eight shown in Fig. 14: (a) original reference data (6.25 m trace interval); (b) decimated input data (100 m trace interval); (c) interpolated data (6.25 m trace interval); (d) the differences between (a) and (c).

6 NUMERICAL EXAMPLE (SEAM VELOCITY MODEL)

To test the performance of the wavelet-based multicomponent matching pursuit method, we generated pressure and particle velocity synthetic seismic data using 2-D finite difference modelling (Han *et al*. 2012) with the SEAM (SEG Advanced Modeling Program) Phase 1 velocity model. Because the SEAM model was originally a 3-D model, we extracted a 2-D vertical section at 20 km in the *y*-direction, as shown in Fig. 8, to generate a 2-D synthetic data set. In the generated synthetic data, the main frequency of the source was 10 Hz and the receiver interval was 20 m (Fig. 9a). Then, to create the spatially aliased input data, pressure and crossline particle velocity data were reduced through decimation to 60 m and 100 m spacing, respectively (Figs 9b and c). The decimated data with 60 m spacing were aliased a little over twice and the decimated data with 100 m spacing were aliased about three times, as shown in Figs 9(e) and (f). The interpolation results by sinusoidal-based and wavelet-based

matching pursuit methods for 60 m spacing data are presented in Fig. 10. In the case that the data were aliased about twice, both methods show interpolation results without artefacts in the time-space domain (Figs 10a and b) and the frequency-wavenumber domain (Figs 10c and d). To demonstrate the superior de-aliasing performance of the wavelet-based method, we applied both methods to decimated data with 100 m spacing, which were aliased three times. For producing better result with the sinusoidal-based method, we limit the range of the sinusoidal basis function for each frequency so that the interpolation can be applied only to seismic events with the slopes less than that of the direct wave following Vassallo *et al*. (2010). In the wavelet-based method, we use linear weighting with which the wavelet basis functions with lower peak wavenumbers are found prior to those with higher peak wavenumbers to yield the stable result. The interpolation results of both methods are not perfect as shown in Figs 11(c) and (d). However, the wavelet-based method (Fig. 11b) showed a more stable interpolation result than the sinusoidal-based method (Fig. 11a) in the time-space domain. The improperly found sinusoidal basis function affected the entire data, whereas the improperly found wavelet basis function affected only a limited area, as shown in Figs 11(a) and (b).

To investigate the effects of noise on the wavelet-based matching pursuit method, we added Gaussian random noise to pressure and crossline particle velocity synthetic data (SNR = 10), as shown in Fig. 12(a). Then these data were decimated with 60 m spacing, which can be successfully reconstructed using the wavelet-based method in noise-free case (Fig. 12b). To analyse the effect of noise depending on the frequency and the wavenumber, Gaussian random noise, which is not band limited, was added separately to the pressure and particle velocity data. The interpolation result of the noise-added data and the differences between the reference and interpolated data are presented in Figs 12(c) and (d), respectively. It seemed to be well interpolated, like the noise-free case. However, because the noise affected the inner product value which measures the goodness of fit of the basis function in the matching pursuit decomposition process, events with lower amplitude than the noise could not be found. Except for events with lower amplitudes, all events were successfully interpolated, as confirmed by the results shown in Figs 12(c) and (d). This aspect is also shown in the *f–k* spectrum using the rectangles with red colour (Fig. 13). In the Fig. 13(c), random noises in the red rectangles which have higher wavenumber values than wavenumber bandwidth of main signal are disappeared.

Finally, we tested the method with very few spatially sampled data points, because crossline data are typically sparse and few spatially sampled in marine seismic acquisition. Figs 14 and 15 show the interpolation result using only eight traces and its *f–k* spectrum, respectively. Fig. 14(a) shows the reference undecimated data (6.25 m spacing) which are generated by using the same velocity model as Fig. 9(a), and Fig. 14(b) shows decimated input data (100 m spacing). Although there were only eight traces with 100 m spacing, a typical streamer separation, in the input data, we successfully interpolated the input data such that the output looked almost the same as the reference data. The differences in time trace and in *f–k* spectrum between the reference and interpolated data are shown in Figs 14(d) and 15(d), respectively.

7 CONCLUSIONS AND DISCUSSION

We have developed a wavelet-based matching pursuit interpolation method that improves the de-aliasing capability of matching pursuit interpolation. The wavelet basis function offers important advantages over the sinusoidal basis function for interpolating aliased data. A wavelet better recognizes aliased wavenumbers due to its broader bandwidth than a sinusoid which has a single wavenumber. Furthermore, any misfit wavelet does not affect other locations due to its limited spatial range. To further improve the de-aliasing performance, we simultaneously used pressure and crossline particle velocity in the wavelet-based matching pursuit method. As a result, we reconstructed the energy at over twice the Nyquist rate and obtained more stable interpolation results. However, because the wavelet basis function has more parameters than the sinusoidal basis function, it is more time-consuming to determine the parameters by minimizing L-2 norm. To reduce the computational cost, we used the inner product instead of L-2 norm minimization to find the coefficient of the basis function. To ensure good performance of the wavelet-based method, the parameter range of the wavelets in the dictionary must be properly selected, because it is related to the uniqueness problem in the matching pursuit decomposition process. We formulated the optimal parameter range according to the bandwidth in the wavenumber domain and the duration in the space domain. The optimal parameter range varies depending on the type of the wavelet.

In this paper, we interpolated the 2-D seismic data since the main purpose of our method is interpolation of cross-line direction in the marine seismic data which are relatively regular and severely aliased. For this case, 1-D wavelet basis function was sufficient. However, in order to perform the interpolation of the higher dimensional, wavelet basis is also extended to higher dimension. However, since the higher dimensional wavelet basis has more parameters to find, the computational cost is also increased. Therefore, a technic which reduces the computational cost is required for higher dimensional interpolation.

ACKNOWLEDGEMENTS

This work was supported by the Human Resources Program in Energy Technology of the Korea Institute of Energy Technology Evaluation and Planning (KETEP) granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (No. 20134010200520) and the Korea Meteorological Administration Research and Development Program under grant KMIPA (KMIPA2015-7012).

REFERENCES

Abma, R. & Kabir, N., 2006. 3D interpolation of irregular data with a POCS algorithm, *Geophysics,* **71**(6), E91–E97.

Han, B., Seol, S.J. & Byun, J., 2012. Elastic modelling in tilted transversely isotropic media with convolutional perfectly matched layer boundary conditions, *Explor. Geophys.,* **43,** 77–86.

Kamil, A.I., Vassallo, M., Brouwer, W., Nichols, D., Cowman, M. & Özbek, A., 2014. Joint crossline reconstruction and 3D deghosting of shallow seismic events from multimeasurement streamer data, in *76th EAGE Conference and Exhibition Abstracts*, doi:10.3997/2214-4609.20141450.

Kim, B., Jeong, S. & Byun, J., 2015. Trace interpolation for irregularly asmpled seismic data using curvelet-transform-based projection onto convex sets algorithm in the frequency-wavenumber domain, *J. Appl. Geophys.,* **118,** 1–14.

Lachowicz, P. & Done, C., 2010. Quasi-periodic oscillations under wavelet microscope: the application of Matching Pursuit algorithm, *Astron. Astrophys.,* **515,** A65, doi:10.1051/0004-6361/200913144.

Linden, D.A., 1959. A discussion of sampling theorems, in *Proc. Inst. Radio Eng.,* Vol. 47, pp. 1219–1226.

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

Mallat, S.G. & Zhang, Z., 1993. Matching pursuits with time-frequency dictionaries, *IEEE Trans. Signal Process.,* **41**(12), 3397–3415.

Naghizadeh, M. & Innanen, A., 2011. Seismic data interpolation using a fast generalized Fourier transform, *Geophysics,* **76**(1), V1–V10.

Naghizadeh, M. & Sacchi, M.D., 2010. On sampling functions and Fourier reconstruction methods, *Geophysics,* **75**(6), WB137–WB151.

Özbek, A., Özdemir, K. & Vassallo, M., 2009. Interpolation by matching pursuit, in *79th SEG Annual Conference Expanded Abstracts,* Vol. 28, pp. 3254–3257.

Ryan, H., 1994. Ricker, Ormsby, Klauder, Butterworth – a choice of wavelets, *CSEG Recorder,* **19,** 8–9.

Schonewille, M., Klaedtke, A. & Vigner, A., 2009. Anti-alias anti-leakage Fourier transform, in *79th SEG Annual Conference Expanded Abstracts,* Vol. 28, pp. 3249–3253.

Spitz, S., 1991. Seismic trace interpolation in the *F-X* domain, *Geophysics,* **56**(6), 785–794.

Thiebaut, C. & Roques, S., 2005. Time-scale and time-frequency analyses of irregularly sampled astronomical time series, *EURASIP J. Appl. Signal Process.,* **15,** 2486–2499.

Vaidyanathan, P.P., 2001. Generalizations of the sampling theorem: Seven decades after Nyquist, *IEEE Trans. Circuits Syst. I,* **48**(9), 1094–1109.

Vassallo, M., Özbek, A., Özdemir, K. & Eggenberger, K., 2010. Crossline wavefield reconstruction from multicomponent streamer data: Part 1 – Interpolation by matching pursuit using pressure and its crossline gradient, *Geophysics,* **75**(6), WB53–WB67.

Xu, S., Zhang, Y. & Lambaré, G., 2010. Antileakage Fourier transform for seismic data regularization in higher dimensions, *Geophysics,* **75**(6), WB113–WB120.