Geophysical Journal International

Geophys. J. Int. (2019) **216,** 1–18
Advance Access publication 2018 September 20
GJI Seismology

doi: 10.1093/gji/ggy390

Accurate determination of *P*-wave backazimuth and slowness parameters by sparsity-constrained seismic array analysis

Jing Hu,[1] Haijiang Zhang[1] and Haiying Yu[2]

[1] *Laboratory of Seismology and Physics of Earth's Interior, School of Earth and Space Sciences, University of Science and Technology of China, Hefei, Anhui 230026, China. E-mail: zhang11@ustc.edu.cn, haiyingyu@hotmail.com*
[2] *Shanghai Earthquake Agency, Shanghai 200062, China*

Accepted 2018 September 18. Received 2018 September 14; in original form 2018 February 13

SUMMARY

Based on seismic array, the conventional frequency–wavenumber method, or the FK method, is generally used to determine the backazimuth and slowness parameters for the plane wave arriving at the seismic array. However, at low frequencies the FK method results in low-resolution estimation of both parameters. To overcome this issue, we have developed a new seismic array analysis method based on sparsity-constrained (SC) inversion to better determine the backazimuth and slowness parameters. The SC method takes advantage of the sparsity of the incoming wave front in the domain of slowness and backazimuth. We have tested the proposed SC method using synthetic data calculated for Shanghai Earthquake Agency Array (SEAA) from one single source and two sources, respectively. Compared to the FK method, the proposed SC method has a much higher resolution in estimating the backazimuth and slowness parameters. Furthermore, the SC method can more easily separate multiple sources and estimate the two parameters. We also compare the proposed SC method with another advanced seismic array analysis method CLEAN-PSF, which can find the position and strength of point sources. Overall, our proposed SC method performs better than or at least at the comparable level as the CLEAN-PSF method. We further demonstrate the effectiveness of the new method on the 2008 M_w 8.0 Wenchuan earthquake data recorded on the SEAA. The deviation between theoretic backazimuth and optimal solution is approximately 4°, indicating the capability of the proposed SC method for reliably estimating the backazimuth of teleseismic event based on a seismic array.

Key words: Array analysis; Backazimuth and slowness; P-wave; Sparsity constraint;.

1 INTRODUCTION

An array consisting of many sensors in a certain configuration can be used to better detect and characterize incoming wavefields from distant sources, which has been used in many areas such as seismology, astronomy and acoustics (Johnson & Dudgeon 1992). In seismology, conventional beamforming (CBF) method is widely used to detect weak signals and to determine backazimuth and apparent slowness parameters of the incoming wave front across seismic array (Rost & Thomas 2002, 2009; Schweitzer *et al.* 2002). The backazimuth is an important parameter for constraining source location in combination with arrival times and slowness estimation (Schweitzer 2001; Oye & Roth 2003; Bayer *et al.* 2012; Gibbons *et al.* 2015; Chen *et al.* 2016). In addition, array beamforming can also be used to track the rupturing process of large earthquakes (Spudich & Cranswick 1984; Goldstein & Archuleta 1991; Huang 2001) and identify and locate secondary sources due to seismic scattering (Gupta *et al.* 1990; Frost *et al.* 2013).

In general, both polarization analysis (Jurkevics 1988) and CBF can be used to determine the backazimuth of the incoming wavefield. In comparison, the latter method is generally superior to the former one (Harris 1990; Kvaerna & Ringdal 1992) in that the latter can also obtain the apparent slowness of an incoming wavefield. CBF is robust for determining the parameters of backazimuth and apparent slowness of the incoming signal, however, these parameters have poor resolution at low frequencies (Frost *et al.* 2013; Gal *et al.* 2016). As a result, it has poor capability of separating multiple signals with different directions (Gal *et al.* 2016). Recently, Frost *et al.* (2013) applied the F-statistic (Selby 2011, 2013) to the array beam to increase its signal-to-noise ratio (SNR), allowing the analysis of small amplitude arrivals and improving the resolution in measuring slowness and backazimuth parameters. Gal *et al.* (2016) proposed using the CLEAN-PSF algorithm to improve the resolution of CBF by iteratively deconvolution of contributions from individual sources from beampower of the array. This algorithm has been successfully applied to ambient noise analysis (Gal *et al.* 2016).

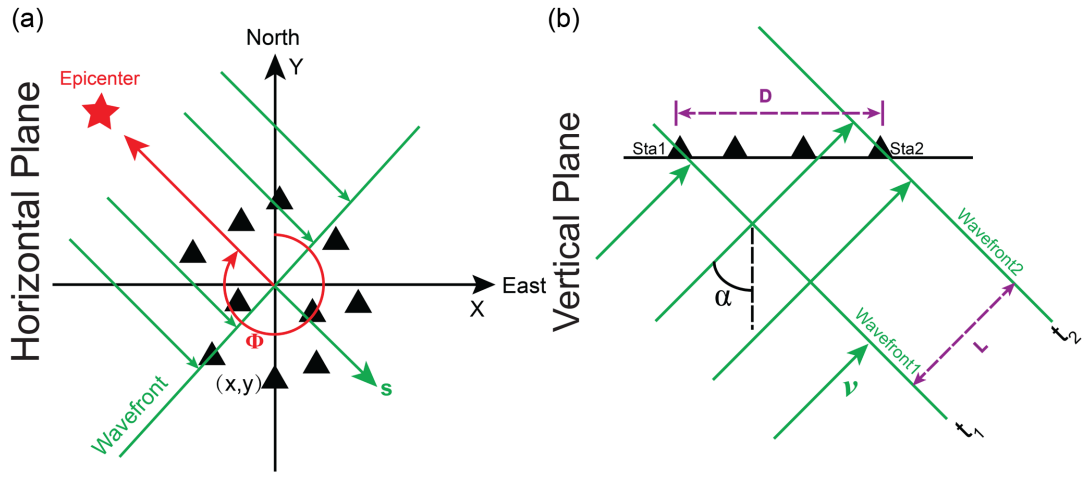

Figure 1. A sketch of an incident plane wave from the view in the horizontal plane (a) and vertical plane (b). (a) A plane wave arriving with backazimuth Φ (in degree) and slowness s (in s km^{-1}). (b) A plane wave crossing station Sta1 at time t_1 and station Sta2 at time t_2 with incident angle α and velocity v (km s^{-1}). D denotes the distance between Sta1 and Sta2, and L marks the distance of the two wave fronts. Black triangles are seismic stations and red star is the earthquake epicentre. The relationship of slowness s and velocity v is $s = \sin(\alpha)/\upsilon$.

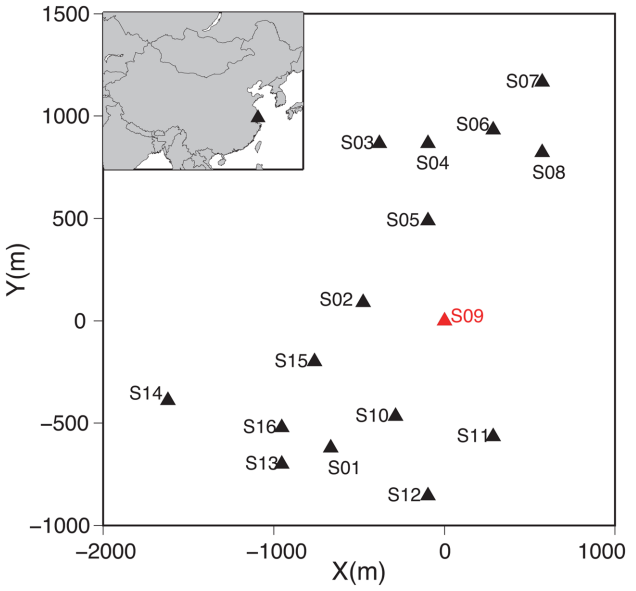

Figure 2. The distribution of substations (triangles) of SEAA. The reference station S09 is denoted in red, which is at (31.1°N, 121.1°E).

In this study, we aim at improving the resolution in estimating backazimuth and apparent slowness parameters of the incoming signal based on the concept of compressive sensing (CS) (Donoho 2006; Candès & Wakin 2008; Xenaki *et al.* 2014). The application of CS in seismology is not new and there have been several applications. Yao *et al.* (2011) utilized CS to determine the earthquake rupturing process in higher resolution. A novel waveform fitting method based on CS was proposed in Rodriguez *et al.* (2012) for estimation of seismic source parameters.Mun *et al.* (2015) proposed an approach based on CS to resolve higher mode Rayleigh wave dispersion curves from seismic data. Aharchaou & Levander (2016) proposed a new high-resolution linear Radon transform based on CS and applied it to extract signals of interest embedded in teleseismic wavefields. In these applications, they assume that the sources

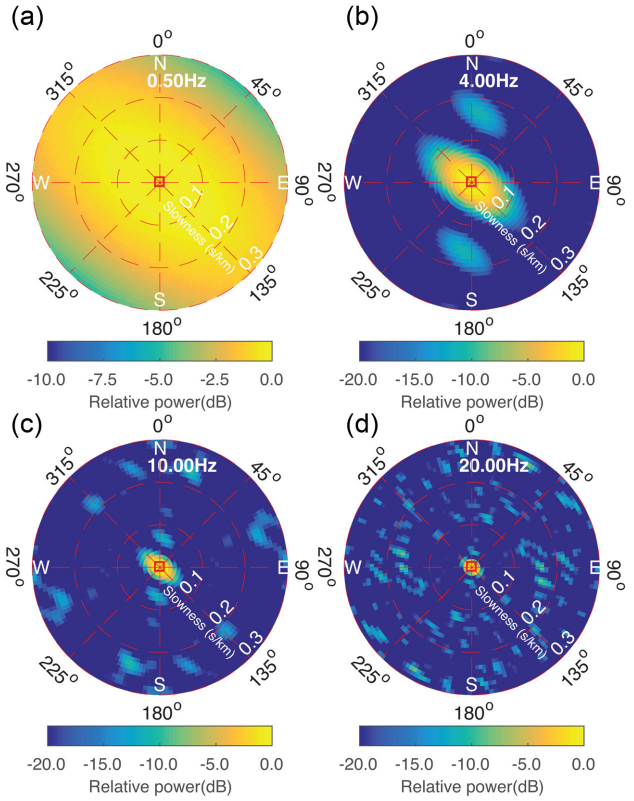

Figure 3. Polar diagrams of array response functions at different frequencies for SEAA. (a) 0.5 Hz, (b) 4 Hz, (c) 10 Hz and (d) 20 Hz. For the polar diagram, the radial axis denotes the slowness from 0 to 0.3 s km^{-1} and the tangential axis denotes the backazimuth (clockwise from north) from 0° to 360°. The power is represented with dB unit. The red box denotes the true point with a backazimuth of 0° and a slowness of 0 s km^{-1}.

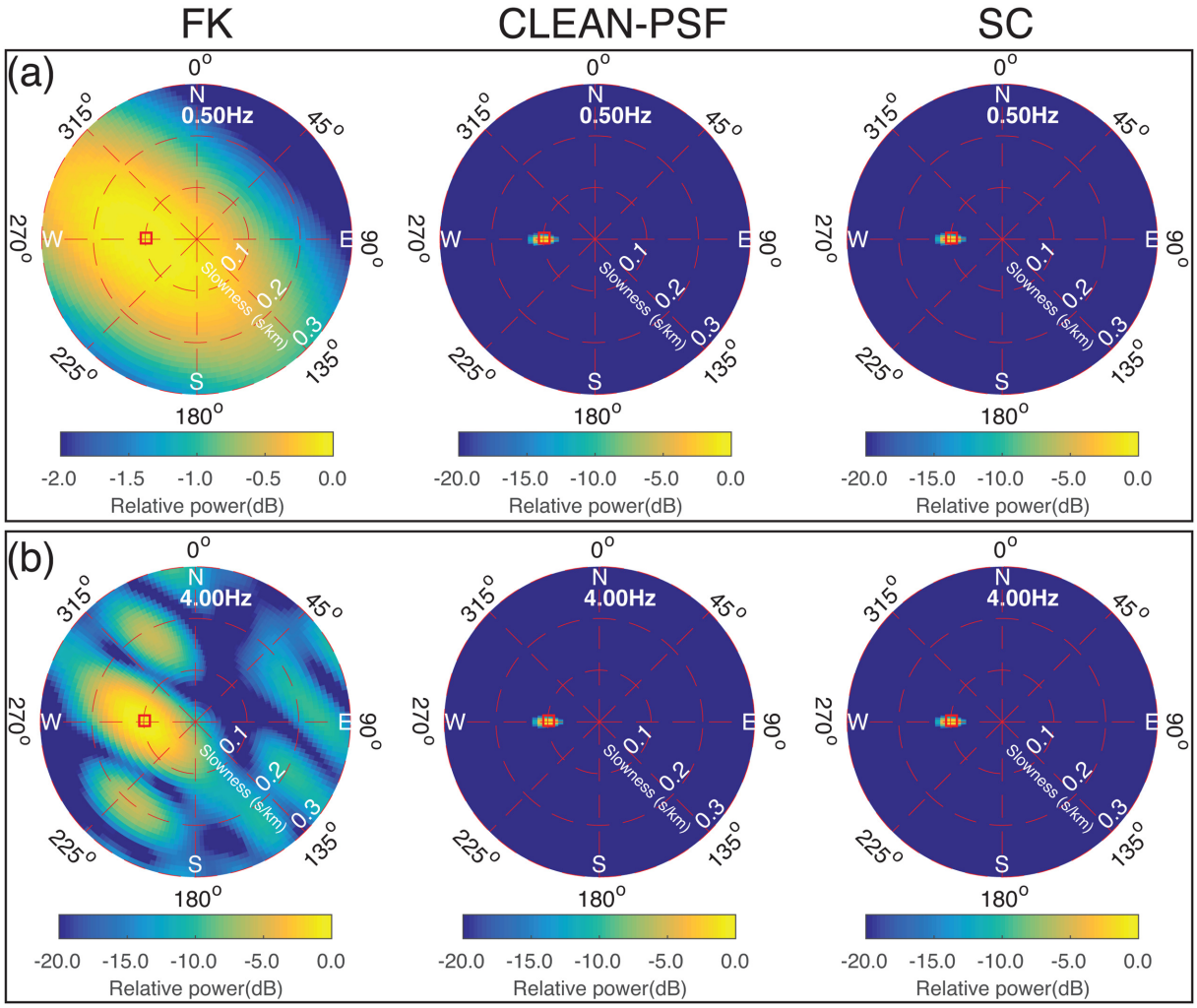

Figure 4. Comparison of polar diagrams for the FK, CLEAN-PSF and SC methods at frequencies of 0.5 and 4 Hz in the case that 10 per cent random noise is added to the spectrum of one single source. Panels (a) and (b) are for 0.5 and 4 Hz, respectively. Red square in the polar diagram denotes true model parameters. For visualization, we replace non-zeros in CLEAN-PSF and SC results with a Gaussian kernel that occupies 6 × 6 grid points in the polar diagram.

or model parameters are sparse and thus a sparsity-constrained (SC) inversion can be applied to this study.

In the following sections, first we will introduce the CBF method or the FK method, the CLEAN-PSF method and SC array analysis method, respectively. Second, with a synthetic data set based on Shanghai Earthquake Agency Array (SEAA), we compare these three array analysis methods for estimating backazimuth and slowness parameters. Third, we will further compare these three methods on three other seismic arrays with different station configurations and apertures. Finally, we test the effectiveness of the proposed method on estimating the backazimuth of the 2008 M_w 8.0 Wenchuan earthquake based on SEAA.

2 CONVENTIONAL ARRAY BEAMFORMING IN THE FREQUENCY DOMAIN

A seismic wave front, which can be assumed as a plane wave if the source is far away from the seismic array, arrives at the seismic array consisting of N substations with slowness s (s km^{-1}) and backazimuth Φ (Fig. 1). For simplicity, 2-D seismic array is taken into account. $\mathbf{r}_i = [x_i\ y_i]$ represents the position vector of the ith station. The origin of the coordinate system is set at a reference station, or one of the substations of the seismic array. For the data recorded at the substations, we define $\mathbf{d} = [d_1(t), \ldots, d_i(t), \ldots, d_N(t)]$ as a 2-D data matrix with the size of $n_s \times N$, where n_s is the number of data points and $d_i(t)$ denotes the data at ith station.

In the frequency domain, for a given frequency f_0, the beampower E for the array is calculated using the following equation:

$$E = \left| \frac{1}{N} \sum_{i=1}^{N} D_i(f_0)\, \mathrm{e}^{-j2\pi f_0 \tau_i} \right|^2, \tag{1}$$

where j denotes the imaginary unit, f_0 denotes the frequency at which the beampower is computed and $D_i(f_0)$ denotes the Fourier spectrum of $d_i(t)$ at frequency f_0. The delay time τ_i in the phase shift term $\mathrm{e}^{-j2\pi f_0 \tau_i}$ at the ith station is the differential arrival time of the incoming signal at ith station and the reference station, which

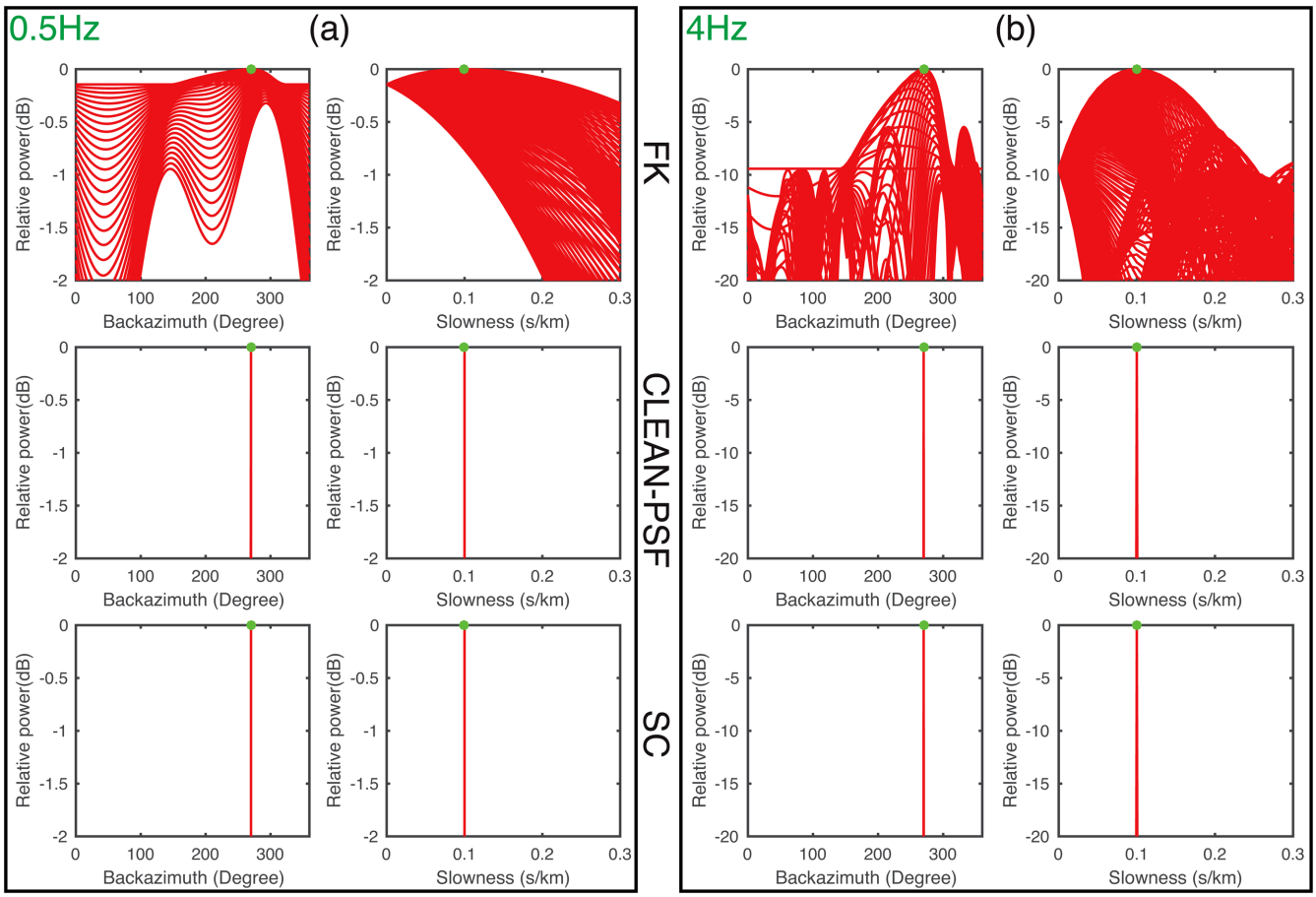

Figure 5. Comparison of cross-sections of backazimuth and slowness parameters for the FK, CLEAN-PSF and SC array analysis algorithms at frequencies 0.5 Hz (a) and 4 Hz (b). In panels (a) and (b), the left-hand plots show cross-sections through the backazimuth for a sweep of slowness values and the right-hand plots show cross-sections through the slowness for a sweep of backazimuth parameters in the polar diagram. The green dots in cross-sections are true model parameters.

is computed as follows (Schweitzer *et al.* 2002):

$$\tau_i = -(x_i \sin\Phi + y_i \cos\Phi)\, s. \tag{2}$$

Essentially, the beampower is obtained by summing the phase-shifted data over all stations and then squaring the average. In practice, we discretize the parameters of slowness and backazimuth into M 2-D grids with $M = S_{\rm slw} \times B_{\rm baz}$, where $S_{\rm slw}$ and $B_{\rm baz}$ denote the number of grid points in the radial slowness direction and the number of grid points in the tangential backazimuth direction, respectively. For a small frequency bin around f_0, the beampower is averaged over the frequency bin. By scanning each grid node in the slowness and backazimuth domain, we can determine the optimal slowness and backazimuth parameters for the incoming signal at which the beampower is maximum.

The array response function (ARF) is defined as follows (Rost & Thomas 2002):

$$|A(\mathbf{k}-\mathbf{k}_0)| = \left| \frac{1}{N} \sum_{i=1}^{N} e^{j\mathbf{r}_i(\mathbf{k}-\mathbf{k}_0)f_0} \right| \tag{3}$$

where $\mathbf{k}_0 = 2\pi f_0 s[\sin\Phi\ \cos\Phi]^T$ is the target wavenumber vector, $\mathbf{k} = 2\pi f_0 s'[\sin\Phi'\ \cos\Phi']^T$ is the scanning wavenumber, Φ' is the scanning backazimuth and s' is the scanning slowness. From eq. (3), we can see that the ARF is a function of the wavenumber $\mathbf{k}_0$, which is related to the observed signal and the array geometry. For a given array geometry, the resolution of ARF is poor in the low-frequency band, whereas there are many local peaks in ARF for the high-frequency band. In general, larger aperture seismic arrays are used to analyse low-frequency signals while smaller aperture arrays are used for relatively high-frequency signals (e.g. Johnson & Dudgeon 1992; Rost & Thmoas 2002). Although the CBF is robust to noise, it suffers from low resolution and it is difficult to separate multiple incoming signals with different backazimuth or slowness.

3 SEISMIC ARRAY ANALYSIS USING THE CLEAN-PSF ALGORITHM IN THE FREQUENCY DOMAIN

One way to improve over conventional array beamforming is to use the CLEAN-PSF algorithm (Gal *et al.* 2016), which originally is a technique widely used to remove side lobes of bright stars from energy maps obtained with telescopes in radio astronomy (Högbom 1974). This CLEAN-PSF algorithm assumes that the energy map is constructed by a collection of point sources. It can find the position and strength of all point sources by first finding

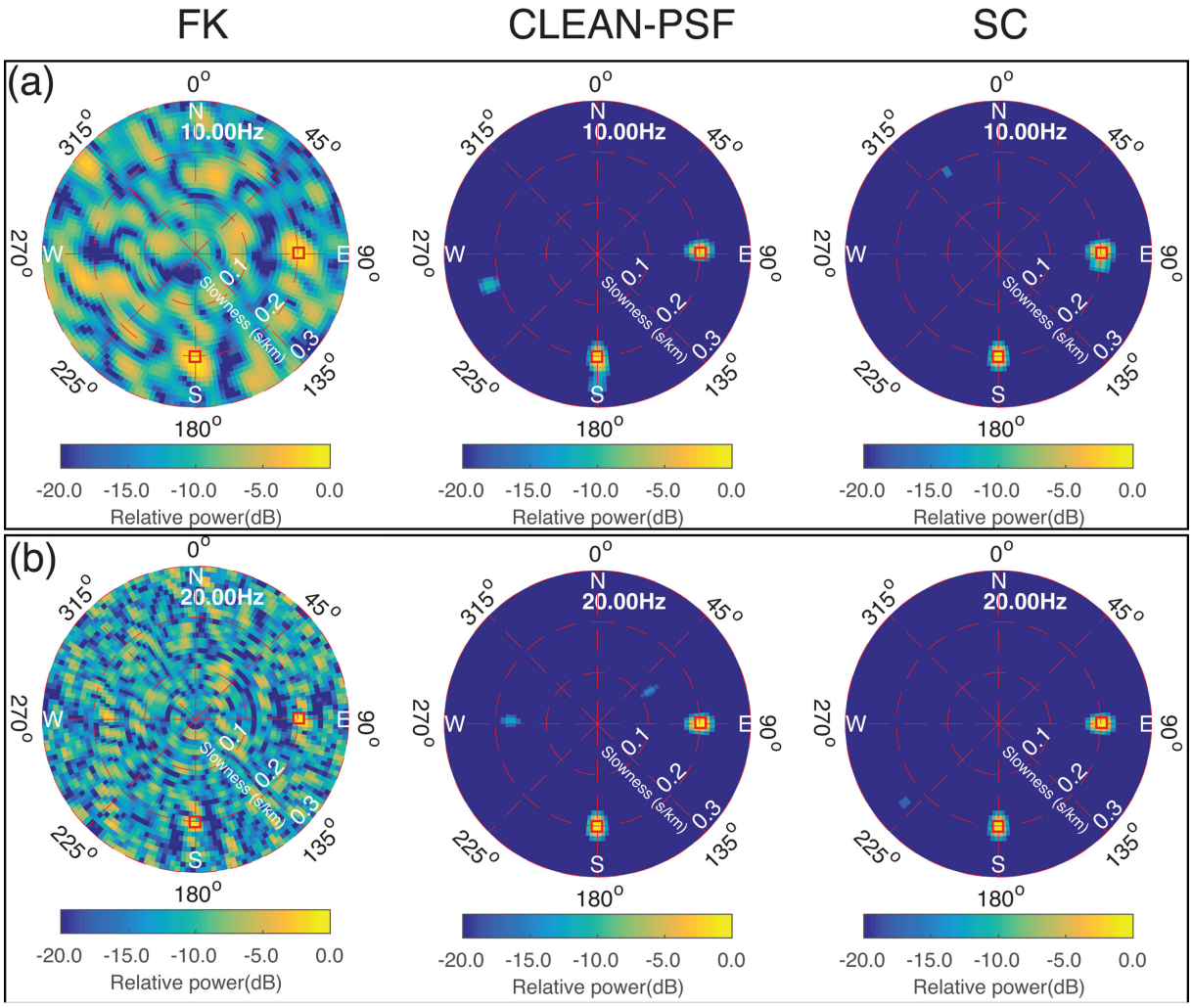

Figure 6. Comparison of the polar diagrams for the FK, CLEAN-PSF and SC array analysis algorithms at different frequencies in the case that 10 per cent random noise is added to the spectrum from two sources. Panels (a) and (b) show results for 0.5 and 4 Hz, respectively. Red square in the polar diagram denotes true model parameters.

the strongest sources and then iteratively subtracting their contributions [point spread function (PSF) or ARF] from the energy map. The application to the seismic ambient noise analysis based on an array produces better results than conventional array beamforming method (Gal *et al.* 2016). Here, we first introduce some basics of the CLEAN-PSF algorithm and then its implementation procedure. For details of the CLEAN-PSF algorithm, refer to Sijtsma (2007).

The beampower $P(\mathbf{k}_0)$ at $\mathbf{k}_0 = 2\pi f_0 s[\sin\Phi \cos\Phi]^T$ can be represented in another way,

$$P(\mathbf{k}_0) = \mathbf{w}(\mathbf{k}_0)^+ \mathbf{C}(f_0)\mathbf{w}(\mathbf{k}_0), \tag{4}$$

where $\mathbf{C}(f_0)$ is cross-spectral matrix, $\mathbf{w}(\mathbf{k}_0)$ is a weight vector and $+$ is the conjugate transpose. The cross-spectral matrix $\mathbf{C}(f_0)$ is defined as

$$\mathbf{C}(f_0) = \mathbf{S}(f_0)\mathbf{S}(f_0)^+, \tag{5}$$

where $\boldsymbol{S}(f_0) = [D_1(f_0), \ldots, D_i(f_0), \ldots, D_N(f_0)]$ is the spectrum vector at N stations at the frequency f_0. The diagonal elements C_{ii} are real values and represent estimates of the power spectral density of the signal on station i, while off-diagonal elements C_{ij} are complex values and give estimates of the cross-spectral power density between stations i and j. Note that $\mathbf{C}(f_0)$ is Hermitian. The weight vector $\mathbf{w}(\mathbf{k}_0)$ is usually constructed by the normalized steering vector to achieve the maximum beamforming power as follows:

$$\mathbf{w}(\mathbf{k}_0) = \frac{\mathbf{a}(\mathbf{k}_0)}{\sqrt{\mathbf{a}(\mathbf{k}_0)^+ \mathbf{a}(\mathbf{k}_0)}}, \tag{6}$$

where $\mathbf{a}(\mathbf{k}_0)$ is steering vector $[e^{j\mathbf{k}_0\mathbf{r}_1}, \ldots, e^{j\mathbf{k}_0\mathbf{r}_i}, \ldots, e^{j\mathbf{k}_0\mathbf{r}_N}]^T$, which is different from the common steering vector because of the phase shift calculated by backazimuth instead of azimuth.

Following Sijtsma (2007) and Gal *et al.* (2016), the CLEAN-PSF algorithm removes phase information associated with the strongest source directly from the cross-spectral matrix. The implementation of CLEAN-PSF algorithm is as follows:

(1) Calculate the beampower (or the dirty map) of the data received by the array using eq. (4) by scanning the backazimuth and slowness domain (with the dimension of $S_{\mathrm{slw}} \times B_{\mathrm{baz}}$). Set the clean

Lorem ipsum

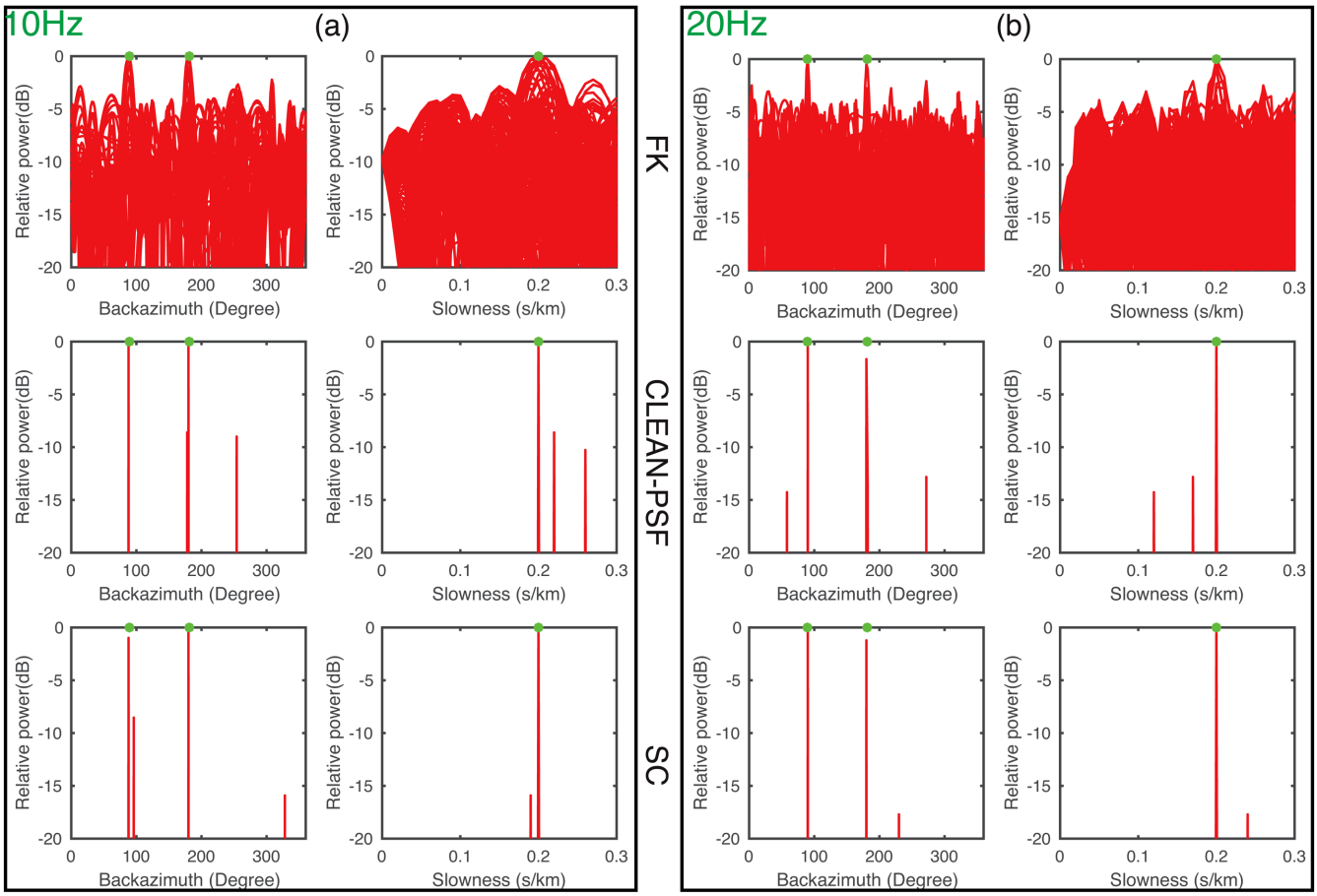

Figure 7. Comparison of cross-sections through polar diagram at different slowness and backazimuth values for frequencies of 10 Hz (a) and 20 Hz (b) using the FK, CLEAN-PSF and SC algorithms in the case of two sources. The green dots in cross-sections indicate true model parameters.

map with zeros and iter $=$ 0. Note that the dimension of the dirty map is the same as the clean map.

(2) Find the peak $P_{\max}^{\text{iter}}$ of the dirty map and the weight vector $\mathbf{w}_{\max}$ associated with $P_{\max}^{\text{iter}}$.

(3) Subtract the appropriately scaled PSF multiplied with the strong source strength from the cross-spectral matrix via $\mathbf{C}(f_0)^{\text{iter}+1} = \mathbf{C}(f_0)^{\text{iter}} - \eta P_{\max}^{\text{iter}} \mathbf{w}_{\max} \mathbf{w}_{\max}^{+}$.

(4) Place the detected strong source in the clean map with the scaled strength.

(5) Construct new dirty map by $P(\mathbf{k}_0)^{\text{iter}+1} = \mathbf{w}(\mathbf{k}_0)^{+} \mathbf{C}(f_0)^{\text{iter}+1} \mathbf{w}(\mathbf{k}_0)$.

(6) While $||\mathbf{C}^{\text{iter}+1}|| \leq ||\mathbf{C}^{\text{iter}}||$, set iter $=$ iter $+ 1$ and repeat steps (2)–(6).

After several iterations, the above deconvolution operation can find multiple sources from the original dirty map and can thus improve the resolution. Here η ($0 < \eta \leq 1$) is the loop gain, which defines the fraction of the beam considered at each iteration. If small loop gain is given, the algorithm converges slowly. However, if the loop gain is too large, the algorithm will miss some weaker sources that could be overshaded by the side lobes.

For comparison, next we will introduce the SC array analysis method that improves the resolution on estimating slowness and backazimuth parameters and has the ability to separate multiple source signals.

4 ESTIMATING BACKAZIMUTH AND SLOWNESS PARAMETERS BASED ON SPARSITY-CONSTRAINED ARRAY ANALYSIS

The same as the CBF, we represent the model of slowness and backazimuth by $M = S_{\text{slw}} \times B_{\text{baz}}$ 2-D grid nodes. In the parametrized model, each grid point can be treated as a source for emitting the plane wave with different slowness and backazimuth parameters. With the observed array data, the problem to find the optimal backazimuth and slowness information of the incoming signal is equivalent to the source localization problem (Yao *et al.* 2011). Source spectrum at each grid point can be denoted by $\mathbf{X}(f) = [x_1(f), x_2(f), \ldots, x_M(f)]^T$, where f is the frequency. The ith source arrives at the nth station with a delay of τ_{ni}, which is the function of backazimuth and slowness as well as the station position relative to the reference station. The station record can be treated as the linear combination of plane waves from all potential sources with correct phase shifts. In the frequency domain, the data

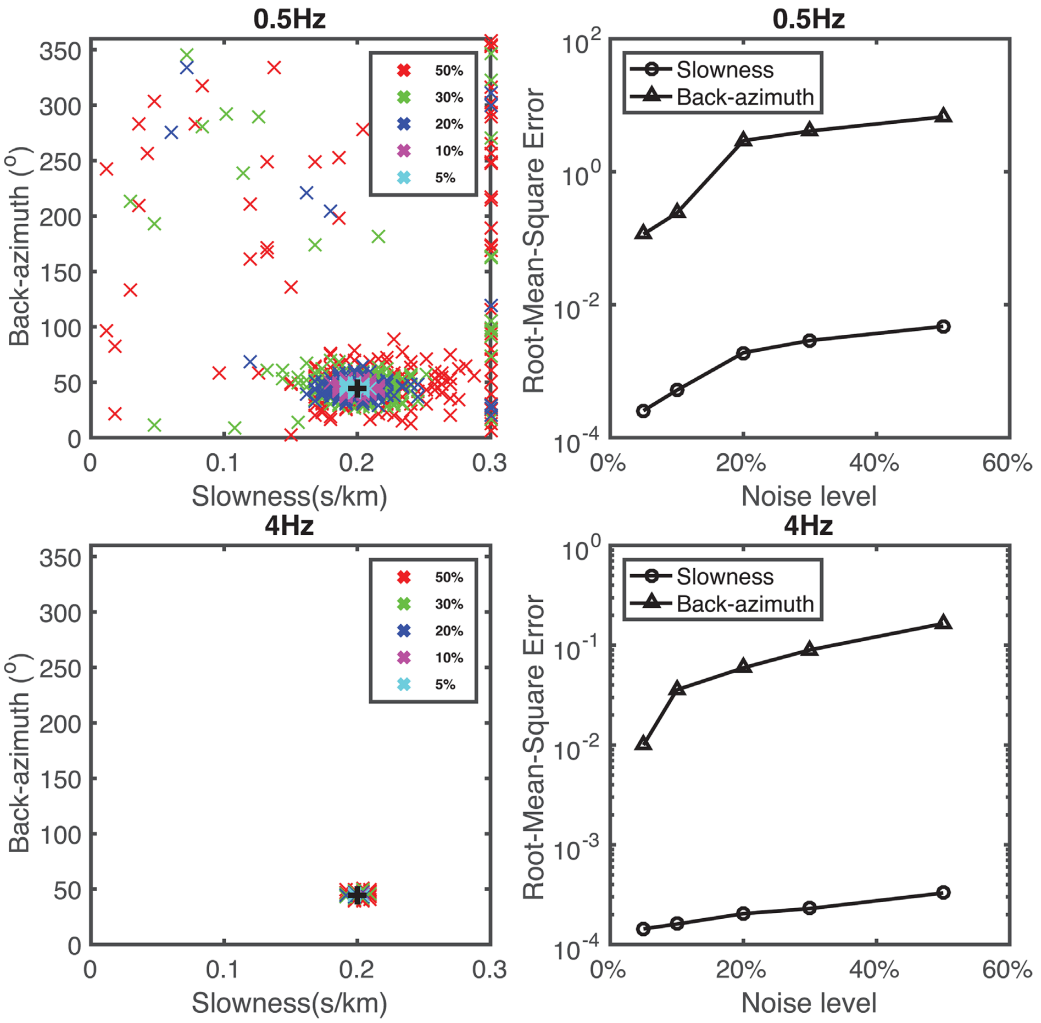

Figure 8. Comparison of the SC array analysis results with different noise levels of 5–50 per cent at frequencies 0.5 and 4 Hz. Left-hand panels: scattering plots of slowness and backazimuth parameters inverted at different noise levels and frequencies. Right-hand panels: root-mean-square (RMS) errors of slowness and backazimuth parameters with respect to true values at different noise levels and frequencies. For each noise level, 200 inversions are performed. Thick black + denotes true model parameters.

spectrum at the nth station can be calculated as follows:

$$d_n(f) = \sum_{i=1}^{M} x_i(f)\, e^{-j2\pi\tau_{ni}}, \tag{7}$$

where $d_n(f)$ is the displacement or velocity spectrum recorded by the nth station.

For an array of N substations, we can write system of equations of eq. (7) into the matrix form with Gaussian distributed white noise as follows:

$$\mathbf{d}(f) = \mathbf{G}(f)\,\mathbf{X}(f) + \mathbf{E}(f), \tag{8}$$

where $\mathbf{G}(f)$ is called the measurement matrix with the element of $G_{nm}(f) = e^{-j2\pi f\tau_{nm}}$, τ_{nm} is the delay time from the mth grid node to the nth station and can be computed using eq. (2), and $\mathbf{E}(f)$ is the Gaussian white noise spectrum.

In general, the number of grid nodes or the number of potential sources, M, is much larger than the number of measurements, N. As a result, there are many possible solutions to satisfy the linear system eq. (8) because the system is underdetermined (Aster *et al.* 2013). To better determine the linear system represented by eq. (8), we can constrain the source spectrum vector $\mathbf{X}(f)$ to be sparse. The sparsity can be measured by the L_0 norm. With the sparsity constraint, the inverse problem can be represented as follows:

$$\begin{aligned} \mathbf{X}_{\text{sol}}(f) &= \text{argmin}\,||\mathbf{X}(f)||_0 \\ &\quad s.t. \quad ||\mathbf{d}(f) - \mathbf{G}(f)\,\mathbf{X}(f)||_2 < \epsilon, \end{aligned} \tag{9}$$

where ϵ is the specified tolerance for noise $\mathbf{E}$.

The L_0-minimization problem is a non-deterministic polynomial hard problem and non-convex, thus is computationally expensive and is likely unstable when there is noise (Baraniuk 2007; Boyd & Vandenberghe 2004). For this reason, the L_0-minimization is often solved by L_1-minimization, which is convex and can thus be solved by convex function optimization. The L_1-minimization for finding the optimal source spectrum is as follows:

$$\begin{aligned} \mathbf{X}_{\text{sol}}(f) &= \text{argmin}\,||\mathbf{X}(f)||_1 \\ &\quad s.t. \quad ||\mathbf{d}(f) - \mathbf{G}(f)\,\mathbf{X}(f)||_2 < \epsilon, \end{aligned} \tag{10}$$

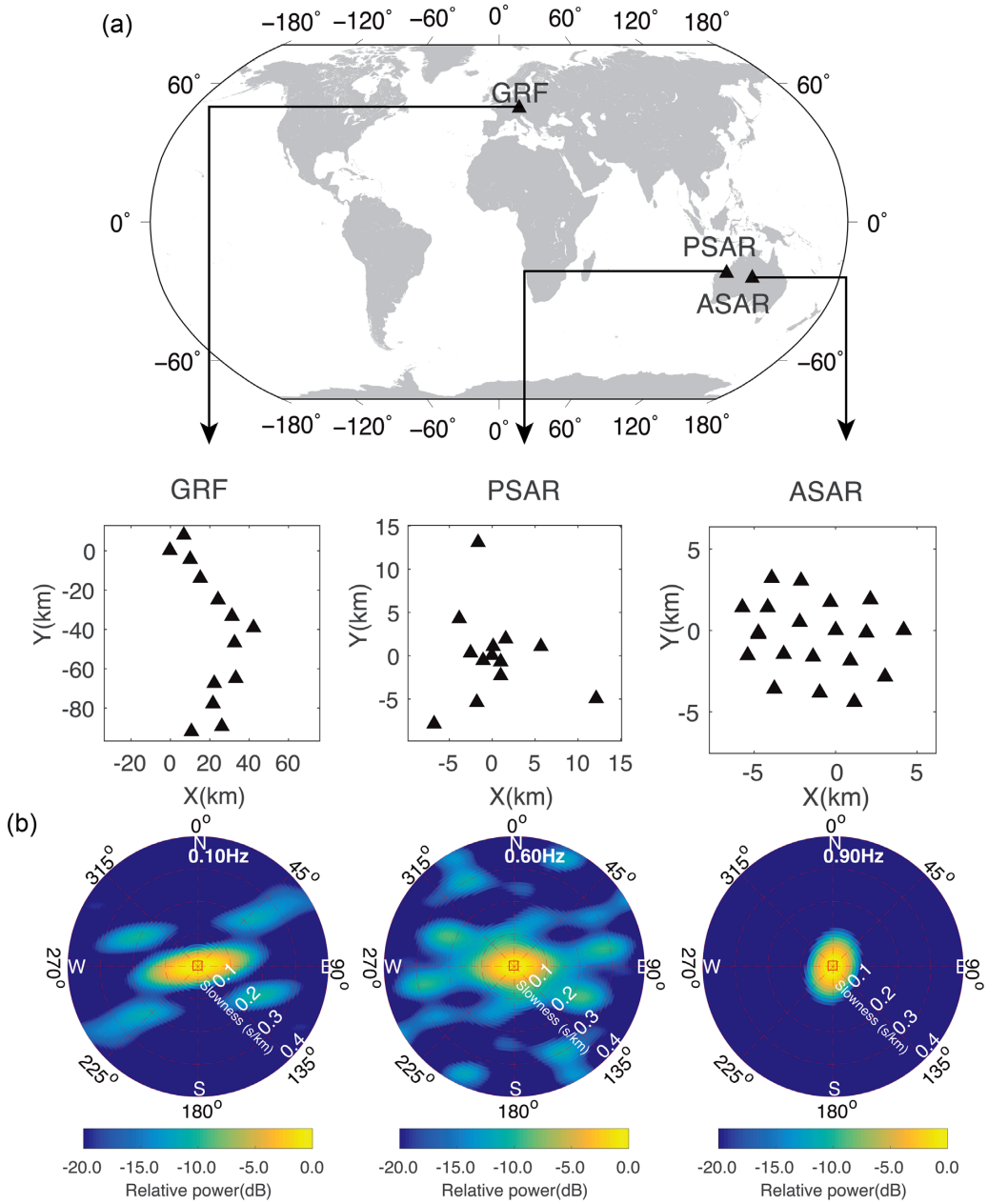

Figure 9. Array configurations and response functions for three different aperture arrays of GRF, PSAR and ASAR. (a) Geographical locations and station configurations of three arrays. (b) Array response functions for three arrays at the corresponding optimal frequencies.

where $||\mathbf{X}(f)||_1 = \sum_{i=1}^{M} |x_i(f)|$. Here, we adopt Orthogonal Matching Pursuit to solve the L_1-minimization problem in eq. (10) (Tropp & Gilbert 2007), which has the advantage of easy implementation, fast convergence and low complexity.

5 SYNTHETIC TESTS OF THREE ARRAY ANALYSIS ALGORITHMS BASED ON SEAA

We first test the proposed method by using the synthetic data created with SEAA. SEAA consists of 16 substations with station spacing about 600 m (Fig. 2). These substations are irregularly distributed and span an aperture of ∼3 km. By selecting station S09 as a reference station, the ARF is calculated for the incoming plane wave with different frequencies of 0.5, 4, 10 and 20 Hz, respectively (Fig. 3). It can be seen that at low frequency (e.g. 0.5 Hz), the main peak of the ARF is not sharp, indicating the array resolution is poor. In comparison, at high frequency (e.g. 20 Hz), the ARF has sharper main peak but also has many local peaks.

We first test the performance of the proposed method for the case of one source. Assuming an incoming plane wave signal with backazimuth of 270° and slowness of 0.1 s km^{-1}, we test the SC method at frequencies of 0.5 and 4 Hz. For the synthetic test, we first generate the theoretic spectrum based on eq. (1). Then we add

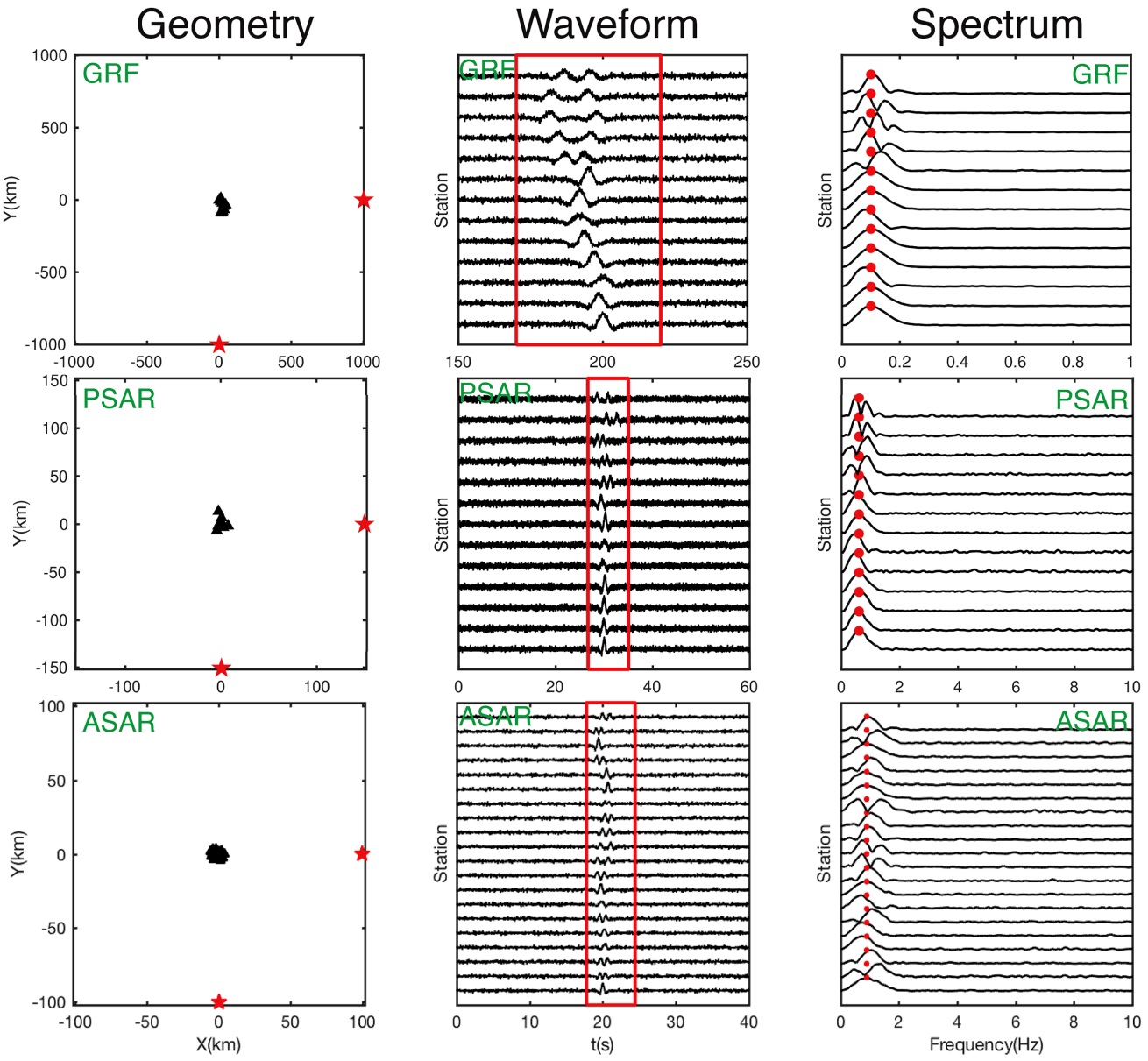

Figure 10. Synthetic waveforms and their spectra for three arrays of GRF, PSAR and ASAR. Left-hand panels: configurations of two sources (red stars) and array stations (black triangles) for three arrays. Middle panels: synthetic waveforms with real noise added. The SNR is 1.25. Red box denotes the time window used for selecting signals for spectra analysis. Right-hand panels: frequency spectra of selected noisy signals. Red dots denote the selected frequencies that are optimal frequencies for each array.

Gaussian distributed random noise, which has the amplitude of 10 per cent of the data amplitude and randomly shifted phases, to the data in the frequency domain. For the inversion, the slowness varies from 0 to 0.3 s km^{-1} with an interval of 0.01 s km^{-1}, and the backazimuth varies from 0° to 360° with an interval of 2°.

In the case of one source, we compare the performances of the FK, CLEAN-PSF and our proposed SC array analysis algorithms. At 0.5 Hz, all FK, CLEAN-PSF and SC methods are capable of recovering the true backazimuth and slowness parameters (Fig. 4a), but both of the CLEAN-PSF and SC methods have much higher resolution in resolving the two parameters than the FK method (Fig. 5a). This is because at 0.5 Hz the SEAA has very poor resolution (Fig. 3) and as a result the polar diagram of the beamforming power from the FK method does not show any clear peak. In comparison, at 4 Hz the FK method has higher resolution than that in the case of 0.5 Hz (Figs 4a and b), as indicated by a sharper main peak in the ARF (Fig. 3b). In comparison, the CLEAN-PSF and SC methods still have higher resolution in resolving both slowness and backazimuth parameters (Figs 4 and 5).

We also test the performance of the SC method when there are two sources with the same amplitude in the slowness and backazimuth domain (Fig. 6). Two sources have the same slowness of 0.2 km s^{-1} but with different backazimuths of 90° and 180°, respectively. The parametrization of the model space is the same as that in the case of one source. We also add Gaussian distributed

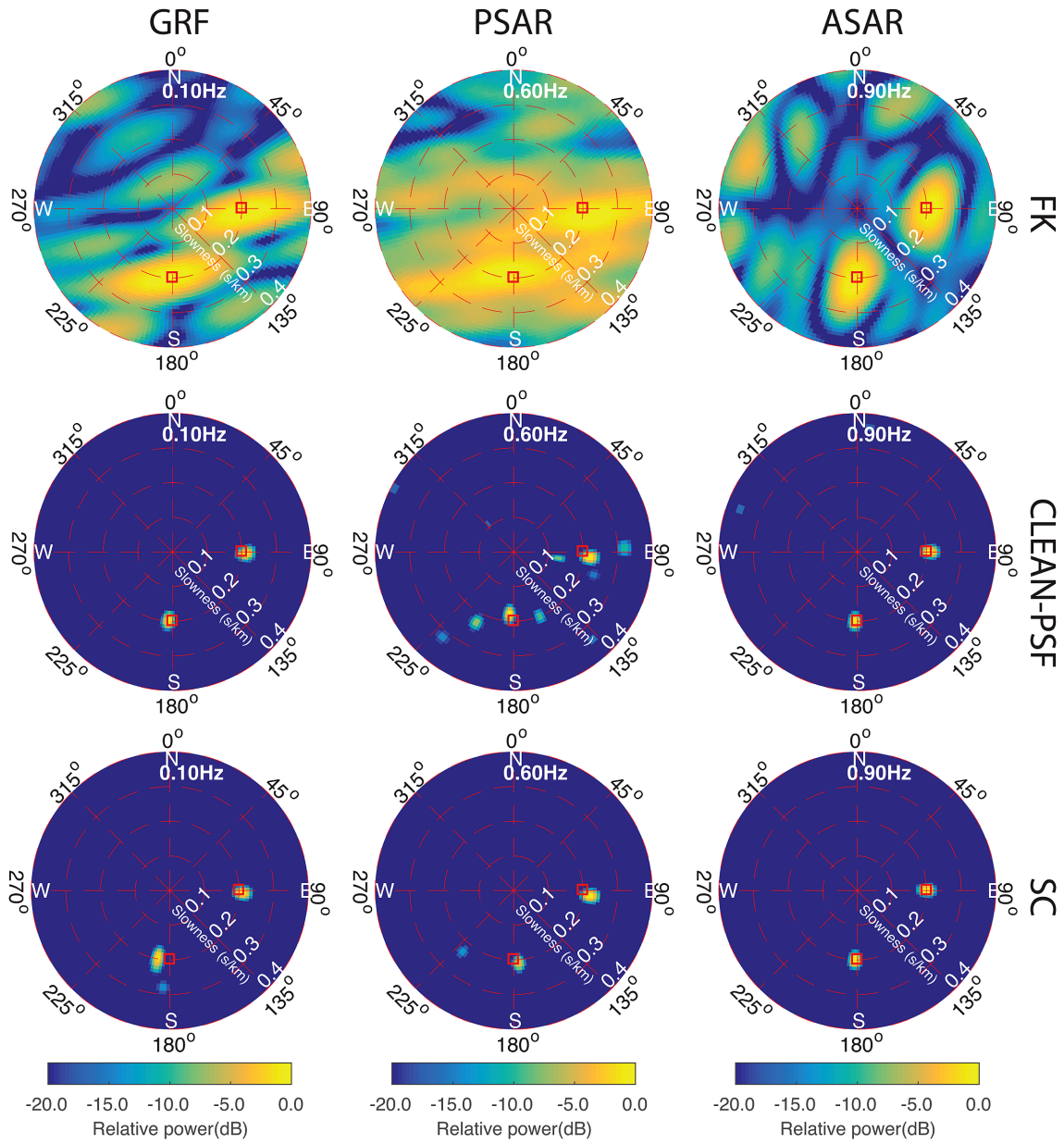

Figure 11. Comparison of the polar diagrams at optimum frequencies for three seismic arrays of GRF, PSAR and ASAR from the FK, CLEAN-PSF and SC array analysis algorithms in the case of two sources.

random noise at the same level as the case of one source. At frequencies of 10 and 20 Hz, the FK method can resolve correct parameters of backazimuth and slowness for two sources. However, there are many local peaks in the ARF, which is a challenge to recognize correct source parameters when there is no *a priori* information about sources (Fig. 6). In comparison, the SC and CLEAN-PSF array analysis methods clearly result in two strong peaks in the polar diagram (Fig. 6). This can be further seen from cross-sections through the backazimuth and slowness for a sweep of slowness and backazimuth values, respectively (Fig. 7). It shows that for the FK method there are many local peaks around the target model parameters. In comparison, the CLEAN-PSF and SC algorithms produce much fewer local peaks. Furthermore, compared to the CLEAN-PSF algorithm, our proposed SC method does a better job and there are almost no local peaks around true parameters. This indicates that the SC method performs the best among the three methods. By choosing an appropriate threshold for the SC method, we could find true solutions corresponding to largest peaks.

To test the robustness of the proposed SC method, we add different levels of noise (5, 10, 20, 30 and 50 per cent) to the synthetic data at frequencies of 0.5 and 4 Hz for one source associated with parameters of slowness of 0.2 s km^{-1} and backazimuth of 45°. Then we perform Monte Carlo simulations for 200 times and analyse the statistics of all the obtained results (Fig. 8). It can be seen that at 0.5 Hz when the noise level is higher, the root-mean-square (RMS) differences between inverted and true model parameters of backazimuth and slowness are also greater. For the backazimuth, the RMS

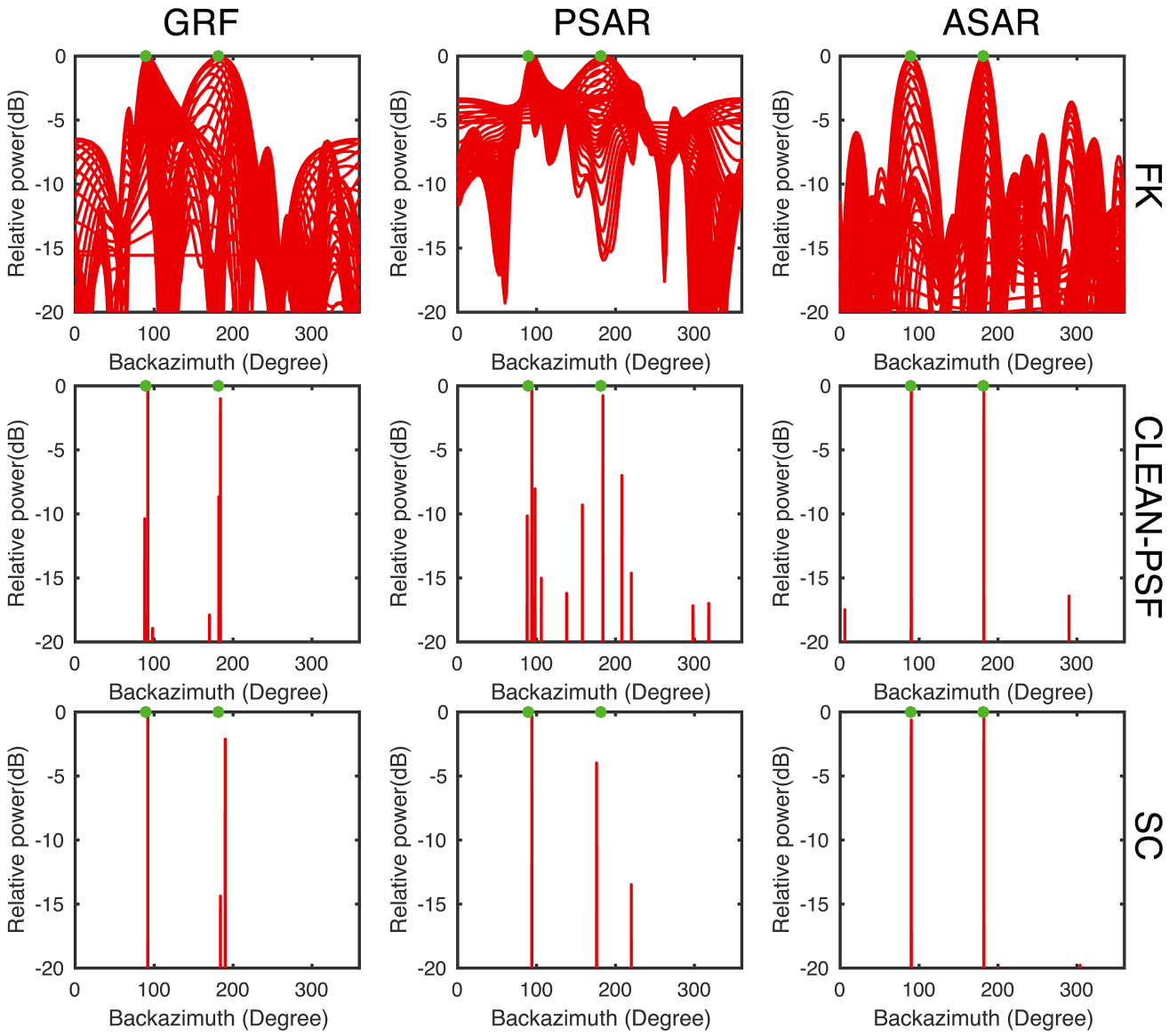

Figure 12. Comparison of cross-sections through polar diagrams in Fig. 11 at different slowness values for three arrays using the FK, CLEAN-PSF and SC array analysis algorithms. Green dots mark true backazimuths.

difference value is about 6.4° at 50 per cent noise level, 3.2° at 30 per cent noise level and about 0.1° at 5 per cent noise level. It can be seen that at lower frequencies, there could be relatively large uncertainties in the estimated model parameters. However, when the noise level is relatively low, the proposed SC method can still give reasonably good estimation of the model parameters. This indicates that the proposed SC method can be used to obtain the backazimuth and slowness parameters of relatively low-frequency teleseismic event by the small-aperture seismic array. In comparison, for the signal at higher frequency of 4 Hz, the proposed SC method has the ability to estimate model parameters of slowness and backazimuth even when the noise level is up to 50 per cent (Fig. 8). This test demonstrates that our proposed SC method has the ability to robustly estimate the slowness and backazimuth parameters at high noise levels.

6 COMPARISON OF THE FK, CLEAN-PSF AND SC ARRAY ANALYSIS ALGORITHMS ON OTHER ARRAYS WITH DIFFERENT APERTURES

In addition to SEAA, we also test the performances of three array analysis algorithms on three other arrays with different station configurations and apertures in the case of two sources. These three arrays include the Gräfenberg array (GRF), which consists of 13 broadband stations and extends ∼100 km north–south and ∼40 km, the Pilbara Seismological Array (PSAR) in northwestern Australia, which has an aperture of ∼25 km and is established based on the spiral-arm concept and the Alice Springs Array (ASAR), which has an aperture of ∼10 km and consists of dense stations (Rost & Thomas 2002; Kennett *et al.* 2015; Gal *et al.* 2016). Fig. 9 shows the station configurations and the ARFs at the corresponding optimal

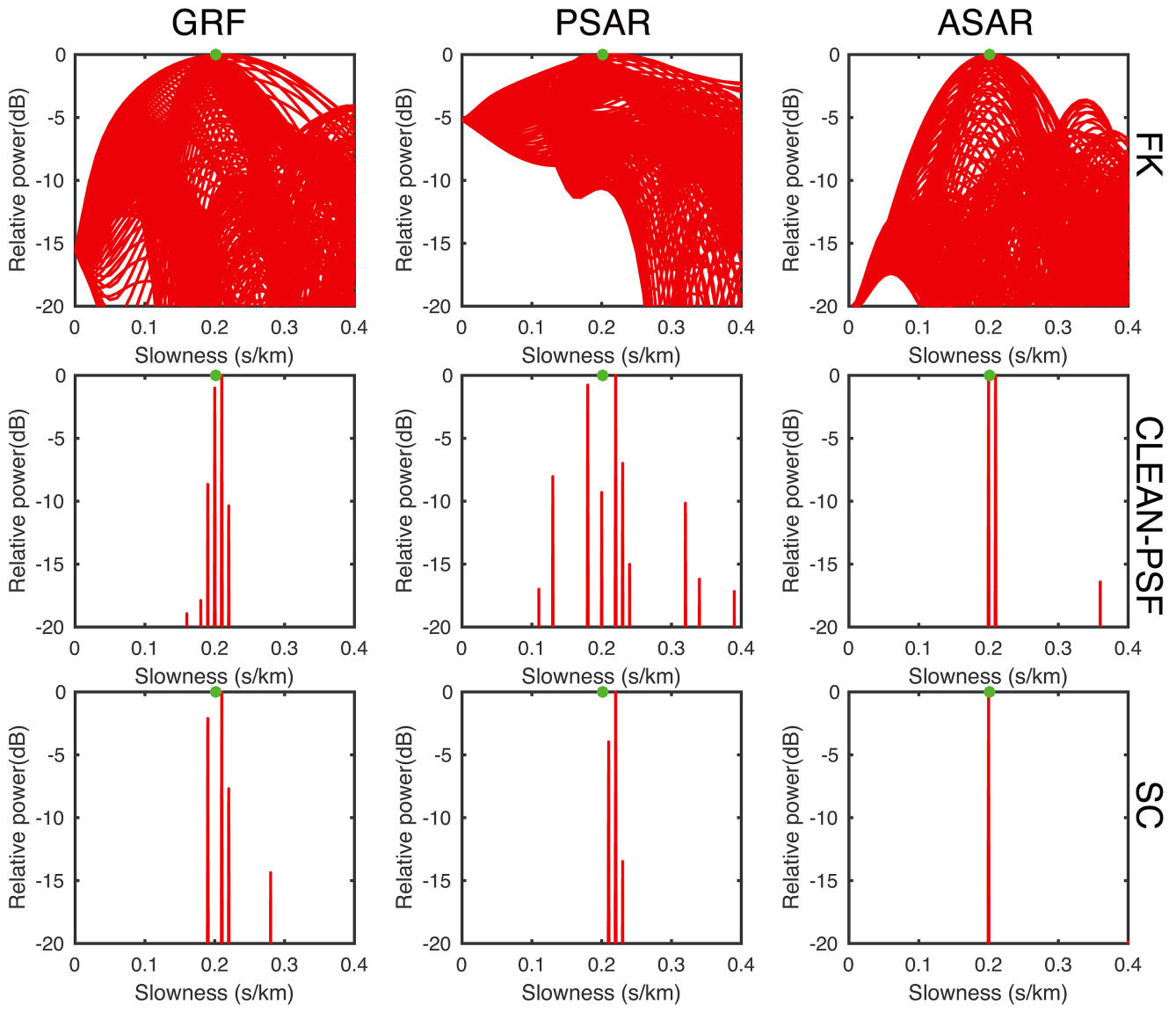

Figure 13. Comparison of cross-sections through polar diagrams in Fig. 11 at different backazimuth values for three arrays using the FK, CLEAN-PSF and SC array analysis algorithms. Green dots mark true slowness values.

frequencies for three arrays. The same as the test case shown in Fig. 6, we test two sources having the same slowness of 0.2 s km^{-1} but with different backazimuths of 90° and 180°, respectively. The sources are Ricker wavelets with different central frequencies of 0.1 Hz for GRF, 0.6 Hz for PSAR and 0.9 Hz for ASAR, respectively. Data are then added with real seismic noise recorded by SEAA with an SNR of 1.25 (Fig. 10). The power and origin time of two sources are the same.

It can be seen that the larger the aperture of the array, the more sensitive is the array to lower frequency signals (Fig. 9). At optimum frequencies, the FK method can determine backazimuth and slowness values close to true values with small biases for three seismic arrays (Figs 11–13). Among three arrays, ASAR performs the best because there are more substations and they are distributed more evenly in space. However, the resolution of the FK method is very poor for all three arrays. Overall, CLEAN-PSF and SC array analysis algorithms perform at similar levels and are much better than the FK algorithm in resolution for three arrays with different station configurations and apertures. In comparison, there are fewer local peaks in the polar diagram from the SC algorithm than the CLEAN-PSF algorithm, suggesting overall the SC algorithm performs better. It is also noted that for seismic arrays GRF and PSAR, the SC algorithm cannot retrieve two sources in equal powers although true sources have the same powers. But for the array ASAR, the SC algorithm can resolve two sources in higher fidelity. Actually, compared to arrays GRF and PSAR, all three methods perform better on array ASAR, which is most likely due to even spatial distribution of array substations.

7 APPLICATION TO THE 2008 M_w 8.0 WENCHUAN EARTHQUAKE, CHINA

Here, we apply the proposed SC array analysis method to the 2008 M_w 8.0 Wenchuan earthquake data recorded by the SEAA, which

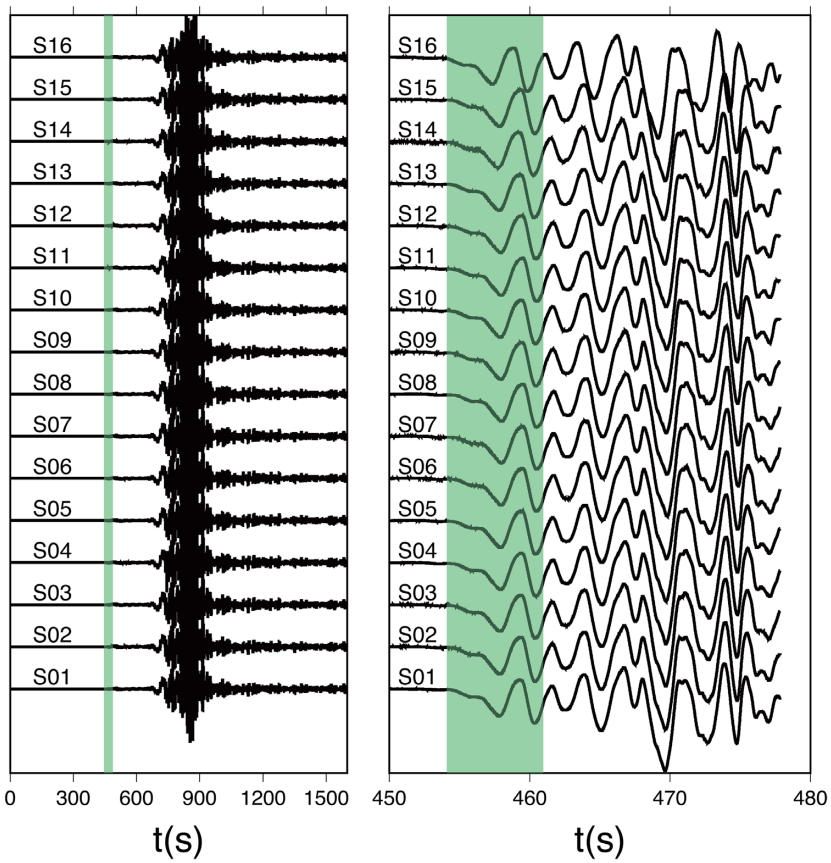

Figure 14. Vertical component waveforms of the 2008 M_w 8.0 Wenchuan earthquake recorded at 16 substations of SEAA. Left-hand panel: record of 1600 s. Right-hand panel: zoom-in record around the *P*-wave arrival. The green shaded area (7 s) delineates the time window used in the SC method.

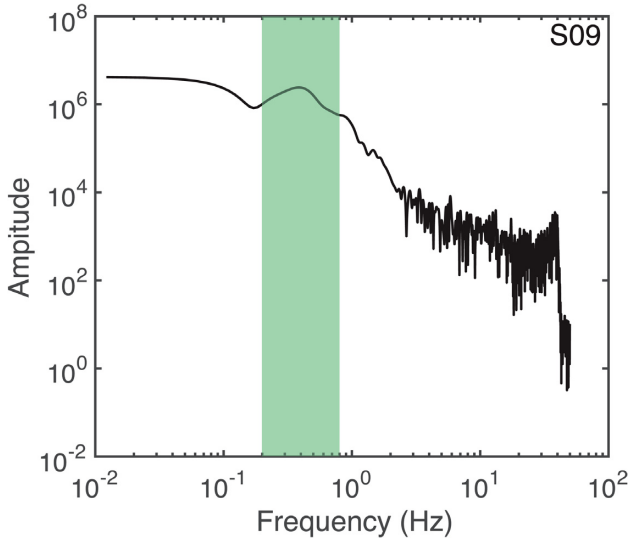

Figure 15. The frequency spectrum of the selected waveform around first P-arrival recorded at reference station S09. The green shaded area denotes the frequency band 0.2–0.8 Hz in which the energy is dominant.

is about 1700 km away. Based on its epicentre at latitude 31.01°N and longitude 103.42°E, we can theoretically estimate the backazimuth value of about 274.2°. Before applying the SC method, we first remove the instrument response and then the mean from each recorded seismic waveforms by array stations. The Hamming window of 7 s is applied to the data starting from the *P*-wave arrival before taking the Fourier transform for avoiding the energy leakage. Fig. 14 shows the recorded waveforms at the SEAA for the 2008 Wenchuan earthquake. From the zoomed-in waveforms around the first *P* arrivals, it can be seen they are very similar at array stations.

For the recorded waveforms from 2008 Wenchuan earthquake, the dominant frequency range is 0.2–0.8 Hz (Fig. 15). For this reason, we choose spectra at the central frequency of 0.5 Hz for the array analysis. As seen in Fig. 3, for the SEAA the ARF has very low resolution at 0.5 Hz. As a result, theoretically it lacks the ability to reliably determine the backazimuth of the 2008 Wenchuan earthquake if the FK method is used. The actual application of the FK method to the Wenchuan earthquake data is consistent with the theoretical prediction. Indeed, the FK method cannot reliably estimate the backazimuth of the 2008 Wenchuan earthquake when using the SEAA because there is no obvious peak in the polar diagram (Fig. 16). Numerically, the FK method can still give an estimate of backazimuth by searching for the maximum beamforming value, which corresponds to the backazimuth of ∼266°. In comparison, the SC method gives a high-resolution estimate of the backazimuth (∼268°), which is very close to the theoretical one (the final-row plots in Fig. 16). For the CLEAN-PSF algorithm, because it first extracts the highest peak from the FK map in the first iteration, in the case of one source it gives the same solution as the FK method. For the apparent slowness parameter, the SC method estimates it to

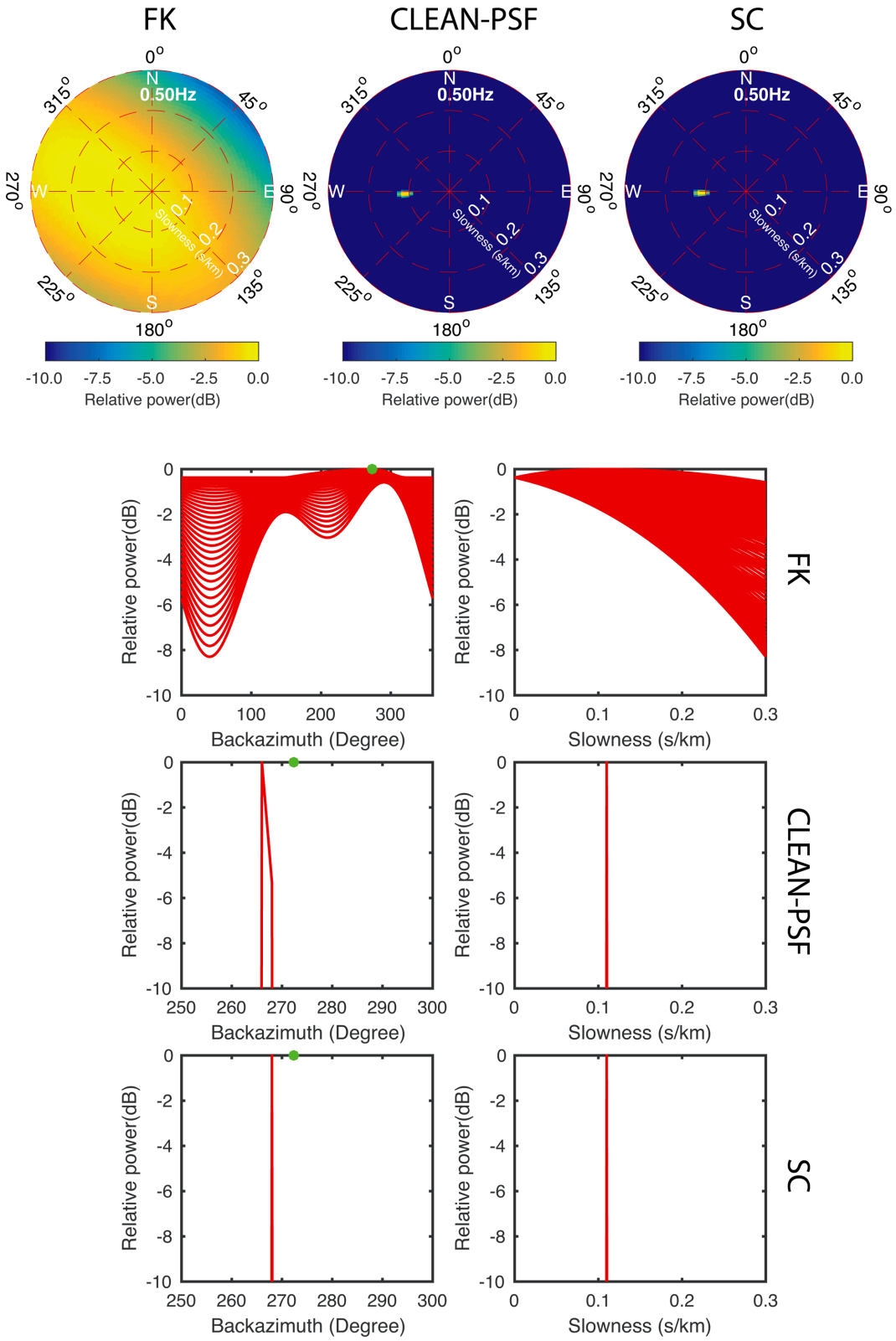

Figure 16. Comparison of three array analysis methods of the FK, CLEAN-PSF and SC for the 2008 Wenchuan earthquake recorded by the SEAA at central frequency of 0.5 Hz. The first-row plots from left to right show the polar diagrams from the FK, CLEAN-PSF and SC methods, respectively. From second to fourth row, it shows cross-sections through polar diagram through slowness (left) and backazimuth (right) from the FK, CLEAN-PSF and the proposed SC methods, respectively. The green dot denotes the theoretical backazimuth.

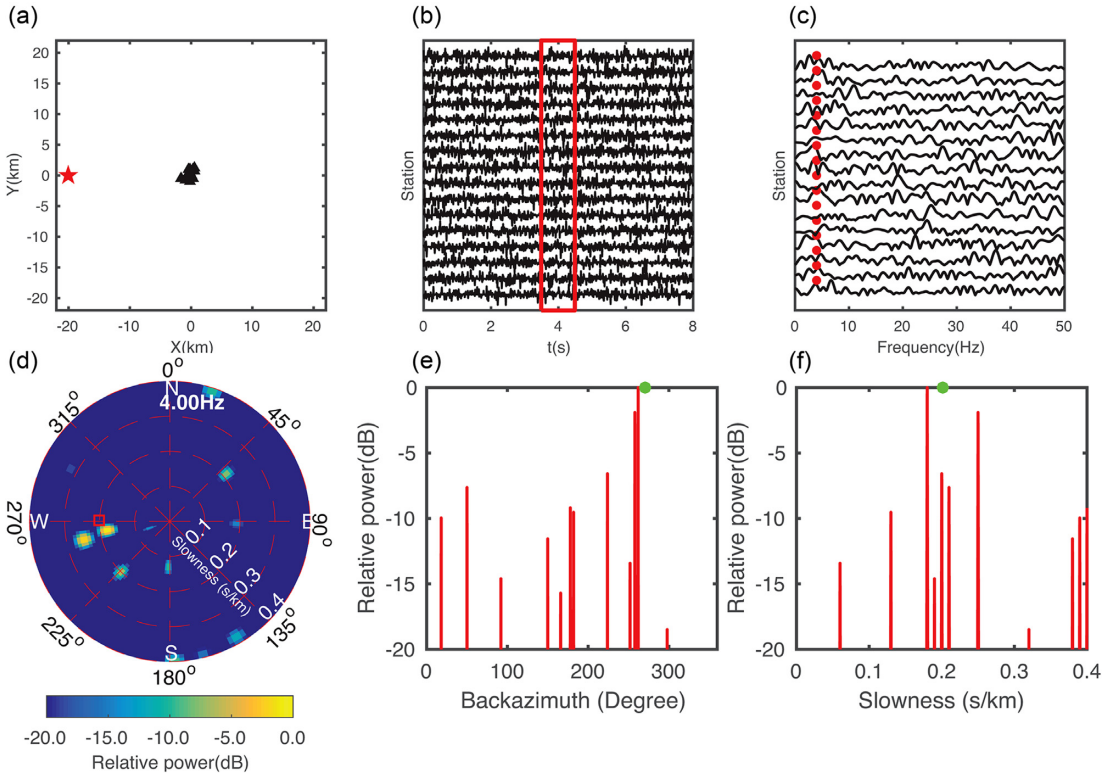

Figure 17. Synthetic test of the SC method in the case of extremely low SNR of 0.25 using the SEAA. (a) Source (red) and seismic array (black triangles) geometry for generating synthetic waveforms. (b) Synthetic noisy waveforms on array substations. Red box denotes the time window for spectra analysis in (c). (c) Frequency spectra of waveforms enclosed in the red box in panel (b). Red dots denote the optimal frequencies for the array. (d) Polar diagram from the SC array analysis. Red square marks the true source parameters. (e and f) show cross-sections through the polar diagram in panel (d) at different slowness and backazimuth values.

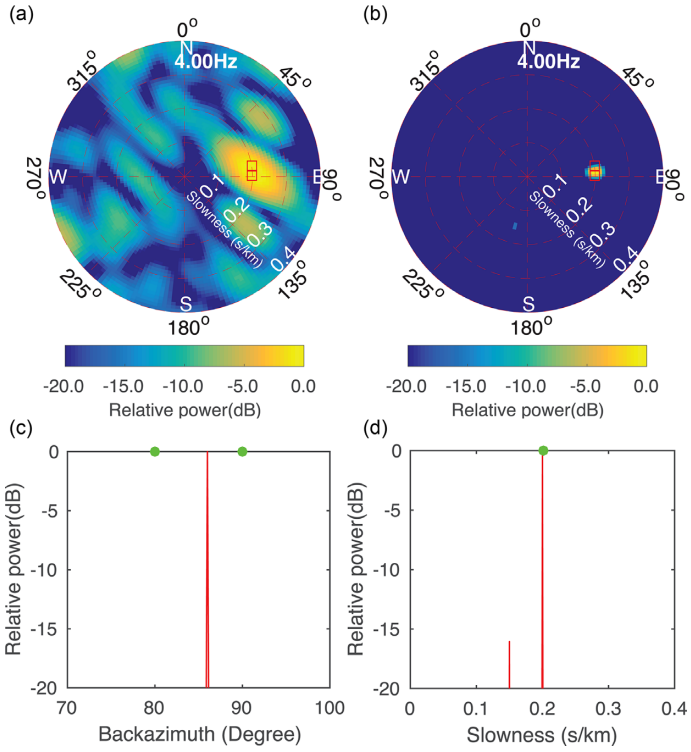

Figure 18. Synthetic test of the SC array analysis method on two very close sources. (a) Polar diagram for the FK method. (b) Polar diagram for the SC method. Two red squares denote the two sources with the same slowness of 0.2 s km^{-1} and slight different backazimuths of 80° and 90°. (c and d) are the cross-sections through the polar diagram in panel (b) at different slowness and backazimuth values. Green dots indicate the true source parameters.

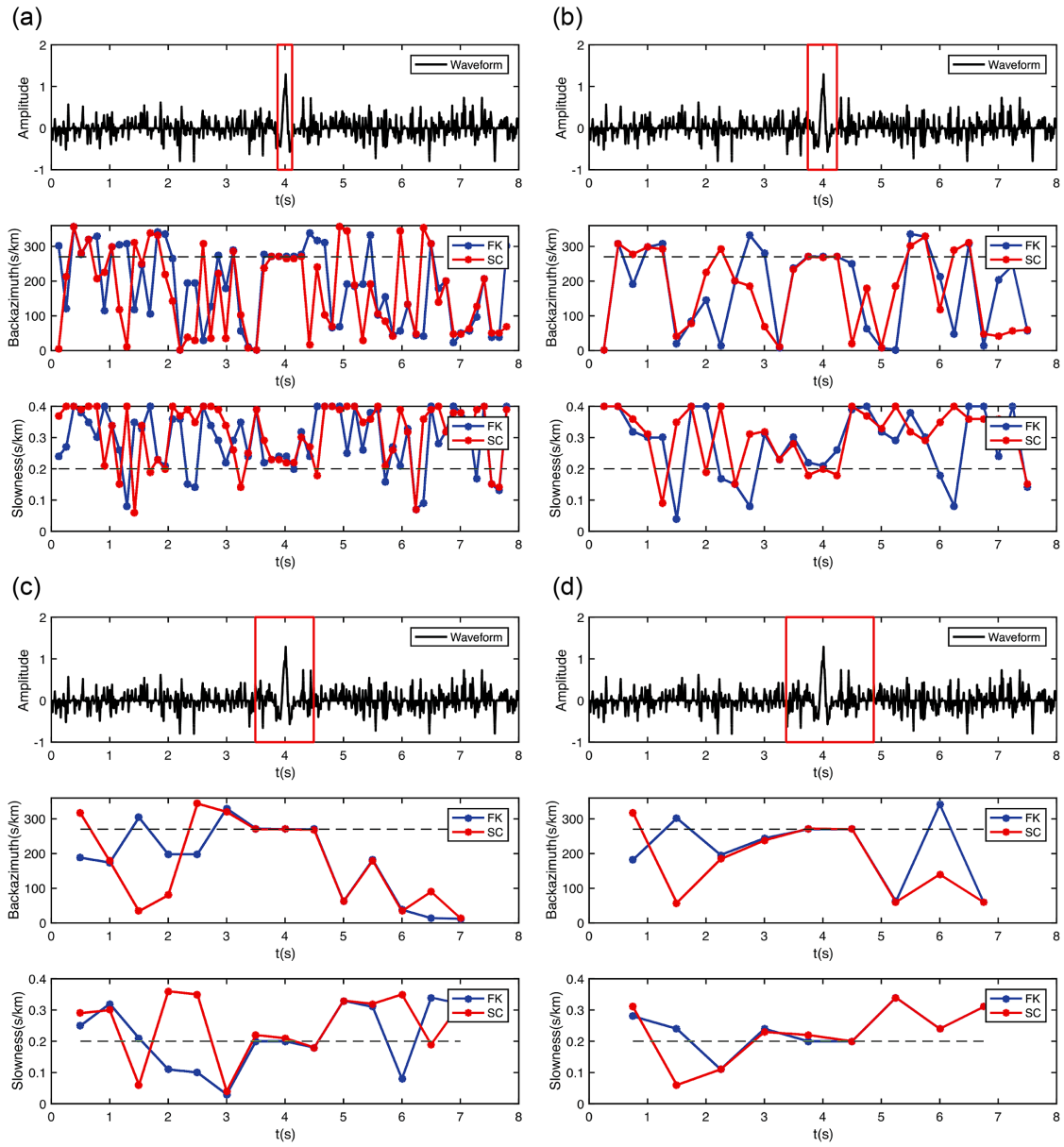

Figure 19. Synthetic test of sliding time window for estimating source parameters with different window sizes by the FK and SC array analysis methods. The time window lengths are (a) 1*T*, (b) 2*T*, (c) 4*T* and (d) 5*T*, respectively. In each panel, the top plot shows a noisy trace recorded by a station of the SEAA, with the red box marking the time window for array analysis. The middle and bottom plots show the estimates of backazimuth and slowness parameters by continuously sliding the time window using the FK (blue) and SC (red) methods.

be around 0.11 s km^{-1}, corresponding to the apparent velocity of 9.1 km s^{-1}, while the FK method cannot reliably give the estimation of the slowness value (the second-row plots in Fig. 16). For this real data application, both of CLEAN-PSF and SC methods give much higher resolution results over the FK method, and the SC method is slightly better in estimating the backazimuth.

8 DISCUSSIONS

We have tested the proposed SC array analysis algorithm when the signal is contaminated by different levels of noise (Fig. 8). The tests show that the SC method can still reliably estimate parameters of backazimuth and slowness even when the noise level reaches up to 50 per cent. To further test the ability of the proposed SC method in the case of extremely low SNR, we generate noisy waveforms with an SNR of 0.25 (Fig. 17). The source is located 20 km away from the substation S09 of the SEAA with a slowness of 0.2 s km^{-1} and a backazimuth of 270° (Fig. 17). The source is a Ricker wavelet with a central frequency of 4 Hz. It can be seen that even the SNR is extremely low, the peak value in the polar diagram obtained from the proposed SC method can still give a good estimate of slowness and backazimuth parameters (Fig. 17). However, there are many local peaks in the polar diagram that are away from the real solution. This indicates that in an extremely noisy environment or if the signal is too weak, we should be cautious of picking peaks in the polar diagram. The best practice would be only choosing the largest peak, which could very likely correspond to the true solution.

We also test the ability of the proposed SC method to discern individual sources with the SEAA if they are very close in slowness

and backazimuth parameters. For this test, we construct two sources with the same slowness of 0.2 s km^{-1} but slightly different backazimuths of 80° and 90° (Fig. 18). In this case, no noise is added to the data. At 4 Hz, both the FK and SC methods cannot separate two nearby sources. The resulted parameter of backazimuth is close to the average of backazimuth parameters for two sources and the slowness is close to the true value. This test indicates that if sources have very close slowness and backazimuth parameters, it may be equivalent that these sources average into one source and the SC method could only determine the average source parameters.

The proposed SC array analysis method, similar to other array analysis methods, requires to select a time window for analysis. In case we do not know when the signal arrives at the array, generally a time window is used to continuously slide through the recorded data for detecting signals. Here we test how the time window length affects the array analysis results of the proposed SC method and how the sliding time window strategy works with the SC method. For comparison, we also apply the FK method on the same data set. For the time window length, we use the signal period T to represent it. We use four different time window sizes of 1T, 2T, 4T and 5T, and slide the time window along the noisy waveform with an overlap of half of the window size (Fig. 19).

In this case, the distribution of stations and source as well as the central frequency of the source wavelet are the same as the case shown in Fig. 17, but SNR is 1.25. It can be seen that for different time window sizes, the estimation of backazimuth is accurate when the signal appears in the time window (Fig. 19). However, for the time window length of 1T, the slowness estimation has slightly higher errors. With the increase of the time window length, the slowness estimation is more accurate. This test shows that the time window length should be around 2T in order to have a reliable estimation of slowness and backazimuth. However, when the time window length is too large, the time resolution for detecting signals is worse. This test also shows that for sliding time window along the data, it is possible to detect when the signals appear by checking if both parameters of slowness and backazimuth are stable for the duration of the time window (Fig. 19).

9 CONCLUSIONS

In this study, we have developed a new method to use seismic array to determine parameters of backazimuth and slowness of seismic event based on SC inversion. Compared to the conventional FK method used for seismic array analysis, the proposed SC method has much higher resolution, even at low frequencies. We have validated the SC method based on both synthetic and real data recorded by the SEAA, as well as synthetic data recorded on three other arrays with different station configurations and apertures. The synthetic tests show that the SC method is able to better separate multiple sources than the FK method, and can determine source parameters even in the cases of low SNR. Compared to another advanced array analysis method CLEAN-PSF, our newly developed SC method performs better with fewer local peaks around true peaks in the polar diagram. The application of the SC method to the 2008 Wenchuan earthquake with the SEAA further demonstrates its capability to determine slowness and backazimuth parameters even at low frequencies, where the FK method fails to do so. Our proposed SC method can be effectively used to extend the capability of seismic array for detecting signals from continuous recorded data and reliably determining backazimuth and slowness parameters of seismic events at both low and high frequencies.

ACKNOWLEDGEMENTS

We are very grateful for the comments from two anonymous reviewers that help to greatly improve the paper. This research is supported by the Shanghai Science and Technology Project under grant number 14231202600.

REFERENCES

Aharchaou, M. & Levander, A., 2016. A compressive sensing approach to the high-resolution linear Radon transform: application on teleseismic wavefields, *Geophys. J. Int.,* **207**(2), 811–822.

Aster, R.C., Borchers, B. & Thurber, C.H., 2013, Parameter Estimation and Inverse Problems, 2nd Edition, Academic Press

Baraniuk, R.G., 2007. Compressive sensing [lecture notes], *IEEE Signal Process. Mag.,* **24**(4), 118–121.

Bayer, B., Kind, R., Hoffmann, M. & Yuan, X., 2012. Tracking unilateral earthquake rupture by *P*-wave polarization analysis, *Geophys. J. Int.,* **188**(3), 1141–1153.

Boyd, S. & Vandenberghe, L., 2004. *Convex Optimization,* Cambridge Univ. Press.

Candès, E.J. & Wakin, M.B., 2008. An introduction to compressive sampling, *IEEE Signal Process. Mag.,* **25**(2), 21–30.

Chen, Y., Zhang, H., Miao, Y., Zhang, Y. & Liu, Q., 2016. Backazimuth constrained double-difference seismic location and tomography for downhole microseismic monitoring, *Phys. Earth planet. Inter.,* **264,** 35–46.

Donoho, D.L., 2006. Compressed sensing, *IEEE Trans. Inform. Theory,* **52,** 1289–1306.

Frost, D.A., Rost, S., Selby, N.D. & Stuart, G.W., 2013. Detection of a tall ridge at the core–mantle boundary from scattered PKP energy, *Geophys. J. Int.,* **195**(1), 558–574.

Gal, M., Reading, A.M., Ellingsen, S.P., Koper, K.D., Burlacu, R. & Gibbons, S.J., 2016. Deconvolution enhanced direction of arrival estimation using one-and three-component seismic arrays applied to ocean induced microseisms, *Geophys. J. Int.,* **206**(1), 345–359.

Gibbons, S.J., Kværna, T. & Mykkeltveit, S., 2015. Could the IMS infrasound stations support a global network of small aperture seismic arrays? *Seismol. Res. Lett.,* **86**(4), 1148–1159.

Goldstein, P. & Archuleta, R.J., 1991. Deterministic frequency-wavenumber methods and direct measurements of rupture propagation during earthquakes using a dense array: data analysis, *J. geophys. Res.,* **96**(B4), 6187–6198.

Gupta, I.N., Lynnes, C.S., McElfresh, T.W. & Wagner, R.A., 1990. FK analysis of NORESS array and single station data to identify sources of near-receiver and near-source scattering, *Bull. seism. Soc. Am.,* **80**(6B), 2227–2241.

Harris, D.B., 1990. Comparison of the direction estimation performance of high-frequency seismic arrays and three-component stations, *Bull. seism. Soc. Am.,* **80**(6B), 1951–1968.

Huang, B.S., 2001. Evidence for azimuthal and temporal variations of the rupture propagation of the 1999 Chi-Chi, Taiwan Earthquake from dense seismic array observations, *Geophys. Res. Lett.,* **28**(17), 3377–3380.

Högbom, J.A., 1974. Aperture synthesis with a non-regular distribution of interferometer baselines, *Astron. Astrophys. Suppl. Ser.,* **15,** 417.

Johnson, D.H. & Dudgeon, D.E., 1992. *Array Signal Processing: Concepts and Techniques,* Simon & Schuster.

Jurkevics, A., 1988, Polarization analysis of three-component array data. , *Bull. Seism. Soc. Am.,* **78,** 1725–1743.

Kennett, B.L.N., Stipčević, J. & Gorbatov, A., 2015. Spiral-arm seismic arrays, *Bull. seism. Soc. Am.,* **105**(4), 2109–2116.

Kvaerna, T. & Ringdal, F., 1992. Integrated array and three-component processing using a seismic microarray, *Bull. seism. Soc. Am.,* **82**(2), 870–882.

Mun, S., Bao, Y. & Li, H., 2015.Generation of Rayleigh-wave dispersion images from multichannel seismic data using sparse signal reconstruction, *Geophys. J. Int.,* **203**(2), 818–827.

Oye, V. & Roth, M., 2003. Automated seismic event location for hydrocarbon reservoirs, *Comput. Geosci.,* **29**(7), 851–863.

Rodriguez, I.V., Sacchi, M. & Gu, Y.J., 2012. Simultaneous recovery of origin time, hypocentre location and seismic moment tensor using sparse representation theory, *Geophys. J. Int.,* **188,** 1188–1202.

Rost, S. & Thomas, C., 2002. Array seismology: methods and applications, *Rev. Geophys.,* **40,** 1008.

Rost, S. &Th omas, C., 2009. Improving seismic resolution through array processing techniques, *Surv. Geophys.,* **30**(4-5), 271–299.

Schweitzer, J., 2001. HYPOSAT—an enhanced routine to locate seismic events, in *Monitoring the Comprehensive Nuclear-Test-Ban Treaty: Source Location,* pp. 277–289, eds Ringdal, F. & Kennett, B.L.N., Springer.

Schweitzer, J., Fyen, J., Mykkeltveit, S., Gibbons, S.J., Pirli, M., Kühn, D. & Kværna, T., 2002. Seismic arrays, in *New Manual of Seismological Observatory Practice 2 (NMSOP-2),* pp. 31–32, ed. Bormann, P., GeoForschungsZentrum Potsdam, IASPEI.

Selby, N.D., 2011. Improved teleseismic signal detection at small-aperture arrays, *Bull. seism. Soc. Am.,* **101**(4), 1563–1575.

Selby, N.D., 2013. A multiple-filter F detector method for medium-aperture seismometer arrays, *Geophys. J. Int.,* **192**(3), 1189–1195.

Sijtsma, P., 2007. CLEAN based on spatial source coherence, *Int. J. Aeroacoust.,* **6**(4), 357–374.

Spudich, P. & Cranswick, E., 1984. Direct observation of rupture propagation during the 1979 Imperial Valley earthquake using a short baseline accelerometer array, *Bull. seism. Soc. Am.,* **74**(6), 2083–2114.

Tropp, J.A. & Gilbert, A.C., 2007. Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inf. Theory,* **53**(12), 4655–4666.

Xenaki, A., Gerstoft, P. & Mosegaard, K., 2014. Compressive beamforming, *J. acoust. Soc. Am.,* **136**(1), 260–271.

Yao, H., Gerstoft, P., Shearer, P.M. & Mecklenbräuker, C., 2011. Compressive sensing of the Tohoku-Oki Mw 9.0 earthquake: frequency-dependent rupture modes, *Geophys. Res. Lett.,* **38**(20).