Coherent wavefield reconstruction improves event location with dense seismic arrays

Lei Li[1] and Benjamin Schwarz[2]

[1]Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring, Central South University
[2]GFZ German Research Centre for Geosciences

May 2, 2023

Abstract

We propose a novel reconstruction method for coherent wavefields recorded by dense seismic arrays. The inherent spatio-temporal coherence in collocated time series is quantified by means of the semblance norm. Using field data recorded by the 1800-station LASSO array and realistic simulations, we demonstrate that the proposed method can reconstruct the wavefields well and produce more coherent and regularized waveform data with high signal-to-noise ratio. We further examine the effectiveness of the reconstructed and enhanced data with stacking-based seismic location. The comparison of imaging results for two synthetic and four field events demonstrates the superiority of reconstructed waveforms regarding source energy focusing and imaging resolution. Polarity-uncorrected traces of reconstructed waveforms produce high-resolution source images, and the corresponding short-term average to long-term average traces yield more stable source images with lower imaging resolution, suggesting the method's applicability to a wide range of common imaging and monitoring tasks.

Coherent wavefield reconstruction improves event location with dense seismic arrays

Lei Li[1,2,3] and Benjamin Schwarz[4*]

[1]Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring (Central South University), Ministry of Education, 410083 Changsha, China.

[2]Hunan Key Laboratory of Nonferrous Resources and Geological Hazard Exploration, 410083 Changsha, China.

[3]School of Geosciences and Info-Physics, Central South University, 410083 Changsha, China.

[4]GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany.

*Corresponding author: Benjamin Schwarz (bschwarz@gfz-potsdam.de)

Key Points:

- A novel reconstruction method for coherent seismic wavefields is proposed
- Application to the dense LASSO array in Oklahoma reveals improved signal-to-noise ratio, coherence, and regularity of the data
- Reconstructed wavefields allow for high-resolution source images with polarity-uncorrected waveforms

Abstract

We propose a novel reconstruction method for coherent wavefields recorded by dense seismic arrays. The inherent spatio-temporal coherence in collocated time series is quantified by means of the semblance norm. Using field data recorded by the 1800-station LASSO array and realistic simulations, we demonstrate that the proposed method can reconstruct the wavefields well and produce more coherent and regularized waveform data with high signal-to-noise ratio. We further examine the effectiveness of the reconstructed and enhanced data with stacking-based seismic location. The comparison of imaging results for two synthetic and four field events demonstrates the superiority of reconstructed waveforms regarding source energy focusing and imaging resolution. Polarity-uncorrected traces of reconstructed waveforms produce high-resolution source images, and the corresponding short-term average to long-term average traces yield more stable source images with lower imaging resolution, suggesting the method's applicability to a wide range of common imaging and monitoring tasks.

Plain language summary

Dense seismic arrays are now used as a regular tool for seismic surveys at multiple scales, mainly due to the popularity of portable digital seismometers. However, the irregularity of seismic station coverage imposes challenges on array-based processing techniques. Here, we propose a technique that quantifies and utilizes the spatio-temporal coherence of dense seismic array recordings to arrive at regularized and enhanced data that can improve subsequent reconstructions. We apply the method to a dense local network consisting of about 1800 stations, and demonstrate the superiority of the reconstructed data regarding the signal-to-noise ratio and waveform coherence. We further validate the reconstructed data by source imaging and the enhanced waveforms produce cleaner and more focused source images. Reconstruction methods can further maximize the potential of dense arrays and benefit array-based seismic investigations, to achieve refined source and structure characterization.

1. Introduction

Since the establishment of array seismology about two decades ago, seismic arrays and accompanying methods and applications have gained significant progress. Though originating from detecting nuclear explosions, seismic arrays are now utilized as a regular and even standardized tool for seismic monitoring ranging from large natural earthquakes to exploration and engineering scales (Hansen & Schmandt, 2015; K. L. Li et al., 2017; Furumura & Maeda, 2021; Lei Li et al., 2022). Currently, more and more temporary and/or permanent dense arrays are deployed on regional and smaller reservoir scales to achieve high-quality earthquake catalogs, efficient seismic hazards assessment, and dynamic characterizations of related engineering activities. With densely sampled wavefields, seismic arrays have promoted the development of array-based processing techniques, such as backprojection and beamforming, and yielded more detailed reconstructions of seismic sources and subsurface structures (e.g., Rost & Thomas, 2002; Gibbons & Ringdal, 2006; Karplus & Schmandt, 2018; Zefeng Li et al., 2018).

Dense arrays can offer high-quality and even regularized full wavefields, which share similarities with recorded waveforms of controlled active sources in reflection seismology. Though the purposes may vary for seismic monitoring at different scales, the well-established summation-based techniques, such as seismic migration, which originate from exploration seismology are becoming routinely applicable in passive seismology when the station coverage is sufficiently dense. Dense arrays were shown to yield data that can be used for directly constructing energy images of seismic sources, even for triggered and/or induced microearthquakes with low signal-to-noise ratios (SNRs) (Kao & Shan, 2004; Steiner et al., 2008; Lei Li, Tan, et al., 2020). Waveform stacking and reverse-time imaging methods are successfully applied to automatically detect and locate microseismic events monitored with sparse and dense arrays, respectively (Hansen & Schmandt, 2015; K. L. Li et al., 2017).

One type of modern arrays is the extremely dense local network with a large number (Large-N) of sensors. For example, the Large-N array consisting of 5200 sensors in the City of Long Beach has been utilized to recover high-resolution 3D shear velocity structure (Lin et al., 2013) and improve the detection capacity of earthquake monitoring networks in urban areas (Yang et al., 2022). However, due to varying surface and deployment conditions and the relatively high cost of the array deployment, the data quality of dense arrays might not always be as good as expected. In this case, waveform reconstruction techniques can make use of waveform similarities across neighboring stations to enhance the data quality or estimate the response at intermediate or nearby locations where no stations had been deployed (see e.g., Chen et al., 2019). Modern large-N station deployments, for the first time, promise to allow for a complete and un-aliased reconstruction of the full seismic wavefield.

Building on the coherence of waveforms across neighboring stations, summation-based techniques originating in controlled-source seismology can be used for targeted and surgical data preconditioning and improved (micro-) seismic monitoring. Here we demonstrate the applicability and effectiveness of such a framework for enhanced seismic source location for the Large-N Seismic Survey in Oklahoma (LASSO; Dougherty et al., 2019). Backed up by numerical simulations it is found that noise levels can be effectively reduced and that for favorable sampling conditions, the effective number of stations can be increased manifold, thereby allowing for a targeted reconfiguration of the array for improving waveform-based event location accuracy.

2. Methodology

In exploration seismology, summation-based coherence analysis is known to be noise-robust and was shown to be a powerful and expressive foundation for process automation (Jäger et al., 2001), effective data preconditioning (Höcht et al., 2009), weak wavefield separation (Schwarz, 2019) and data-driven velocity inversion (Duveneck, 2004; Bauer et al. 2017, Diekmann et al., 2019).We quantify spatiotemporal wavefield coherence by means of the semblance norm S (Neidell & Taner, 1971), which can be viewed as the ratio of the stack (beam) energy and the total energy considered in a data window of interest,

$$S(x_0, y_0, t_0) = \frac{1}{n} \frac{\sum_{\delta t}\left[\sum_{i=1}^{n} D(x_0+\Delta x_i, y_0+\Delta y_i, t_0+\Delta t_i)\right]^2}{\sum_{\delta t}\sum_{i=1}^{n} D^2(x_0+\Delta x_i, y_0+\Delta y_i, t_0+\Delta t_i)}, \qquad (1)$$

where D refers to the data amplitude and (x_0, y_0) to the reconstruction point within the array. The lateral separation of the i-th neighboring station is denoted by $(\Delta x_i, \Delta y_i)$, and

$$\Delta t_i = p_x \Delta x_i + p_y \Delta y_i + \frac{1}{2}\left(H_{xx}\Delta x_i^2 + 2H_{xy}\Delta x_i \Delta y_i + H_{yy}\Delta y_i^2\right) \quad (2)$$

is a second-order approximation of traveltime moveout caused by an elliptical wavefront observed in the local vicinity of the considered reference location (x_0, y_0). In this local approximation, the tilt of the wavefront is characterized by the horizontal slowness vector (p_x, p_y) and wavefront curvature is governed by the elements of the Hessian of the traveltime H_{xx}, H_{xy}, and H_{yy} (e.g., Bortfeld, 1989; Diekmann et al., 2019). In order to not mis-associate different phases, the time window δt should be chosen reasonably small and, in practice, often corresponds to the pre-dominant signal period of interest. Because S takes only values between 0 (not at all coherent) to 1 (perfectly coherent), it lends itself well for optimization.

To ensure reliable convergence even for very low SNRs, we employ a differential evolution global optimizer to locally maximize S for every considered reconstruction point (e.g., Das & Suganthan, 2011). In order to increase robustness and reduce computational costs, we further assume local spherical symmetry of the wavefront, which corresponds to coinciding Hessian components in x and y direction. As an extension of conventional beamforming (e.g., Rost & Thomas, 2002), the optimization of S leads to a data-derived estimate of the local horizontal slowness vector of the emerging wavefront and its curvature radius, which in turn allows for the reconstruction of the data amplitude $\underline{D}$ via

$$\underline{D}(x_0, y_0, t_0) = \frac{1}{n} S(x_0, y_0, t_0) \sum_{i=1}^{n} D(x_0 + \Delta x_i, y_0 + \Delta y_i, t_0 + \Delta t_i)\ . \quad (3)$$

In equation (3), weighting by the semblance has a noise-suppression effect comparable to phase-weighted stacking (Schimmel & Paulssen, 1997) without leading to undesired distortions of the waveforms. For mere data enhancement purposes, coherence analysis and wavefield reconstruction is performed for every actual station location within the array, whereas a departure from the actual station geometry results in the construction of a new, imaginary station response. Thus, the technique can also be used to inter- and extrapolate fields and reconfigure the overall array geometry.

To further assess the improvement of data enhancement by the proposed framework, we test it with stacking-based seismic source location using synthetic and field waveforms from a Large-N array. Stacking-based location methods, as a modern but well-established methodology, have been widely used to automatically detect and locate seismic events at local and regional scales (Grigoli et al., 2013; Shi et al., 2019; Yang et al., 2022). The methods share the essence of weighted backprojection/stacking of wavefields with Kirchhoff migration (Esmersoy & Miller, 1989), and the source location can be easily picked from the energy-focused image. In general, only primary phases are used due to the dominant energy, though all subsequent phases are usable in theory. Diffraction stacking (DS) is the most common operator which simply summarizes waveforms from individual stations along a theoretical traveltime moveout curve (Baker et al., 2005; Gajewski et al., 2007; Ishii et al., 2005; Kao & Shan, 2004). The formula of the DS method reads as follows,

$$M_{DS}(x, t_s) = \sum_{i=1}^{N} CF^i(t)\delta\left[t - \left(t_s + t_{i,x}\right)\right], \quad (4)$$

where $M_{DS}(x, t_s)$ is the stacking value, x denotes the source coordinates, t_s denotes the source origin time, $CF^i(t)$ is the characteristic function (CF) of the waveform recorded at station i, $\delta[t - (t_s + t_{i,x})]$ is the DS operator, where δ is the Dirac delta function and $t_{i,x}$ is the theoretical travel time from station i to the source x. Through simple mathematical derivation and testing, we know that the basic imaging patterns of DS are deformed circular arc and spherical surface intersections under general 2D and 3D models, respectively (L. Li et al., 2018). For surface monitoring, there is an inherent depth-origin time tradeoff when determining the source locations by searching for the maximum imaging values.

3. Dataset

The LASSO experiment, led by the USGS, is a recent and notable example of a Large-N dense array involving more than 1800 single-component nodal seismometers with 500 Hz sampling rate, covering a 25 km by 32 km area of active saltwater disposal in northern Oklahoma (Figure. 1, Dougherty et al., 2019). The seismometers were buried in shallow holes along county roads with a spacing of ~400 m, yielding a dense and regularized pattern which is ideal for wavefield reconstruction and regularization. The LASSO array operated for nearly a month from April to May of 2016 and recorded more than 3000 events (Peña Castro et al., 2019), including the 112 local events from the Oklahoma Geological Survey (OGS) earthquake catalog. Several recent studies have shown the suitability of the array for analyzing the spatiotemporal clustering of seismicity (Cochran et al., 2020), source spectral properties (Kemna et al., 2020), near-source radiation patterns (Trugman et al., 2021), and the leaking modes from ambient noise (Zhengbo Li et al., 2022). Two events from the OGS catalog beneath the array have already been used to demonstrate the effectiveness of DS-based seismic location methods at regional scale (Lei Li, Xie, et al., 2020).

In this work, we use the array to investigate the performance of wavefield reconstruction and its advantages for seismic source location. We select four field events from both the OGS catalog and the extended catalog from previous studies (Dougherty et al., 2019; Peña Castro et al., 2019; Trugman et al., 2021). Two events from the OGS catalog denoted by the numbers 24021 and 23183 were located near the western margin and outside of the array, respectively. The other two events from the extended catalog have a magnitude of M_L 0.08 and M_L 2.03 and were located near the central area of the array. In the following, we refer to these four events by ‘event 24021’, ‘event 23183’, ‘event ML008’, and ‘event ML203’, respectively.

To further validate the reliability of the proposed workflow, we also conduct numerical simulations to mimic the field-recorded waveforms of event 24021 and event ML203 under controlled and reproducible conditions. The finite-difference method is adopted to generate the synthetic waveforms with double-couple source mechanisms in an isotropic layered model (Rubinstein et al., 2018; L. Li et al., 2021). Real noise recorded by the LASSO array is added to the synthetic waveforms to simulate realistic conditions with varying SNR levels (see supplementary information S1).

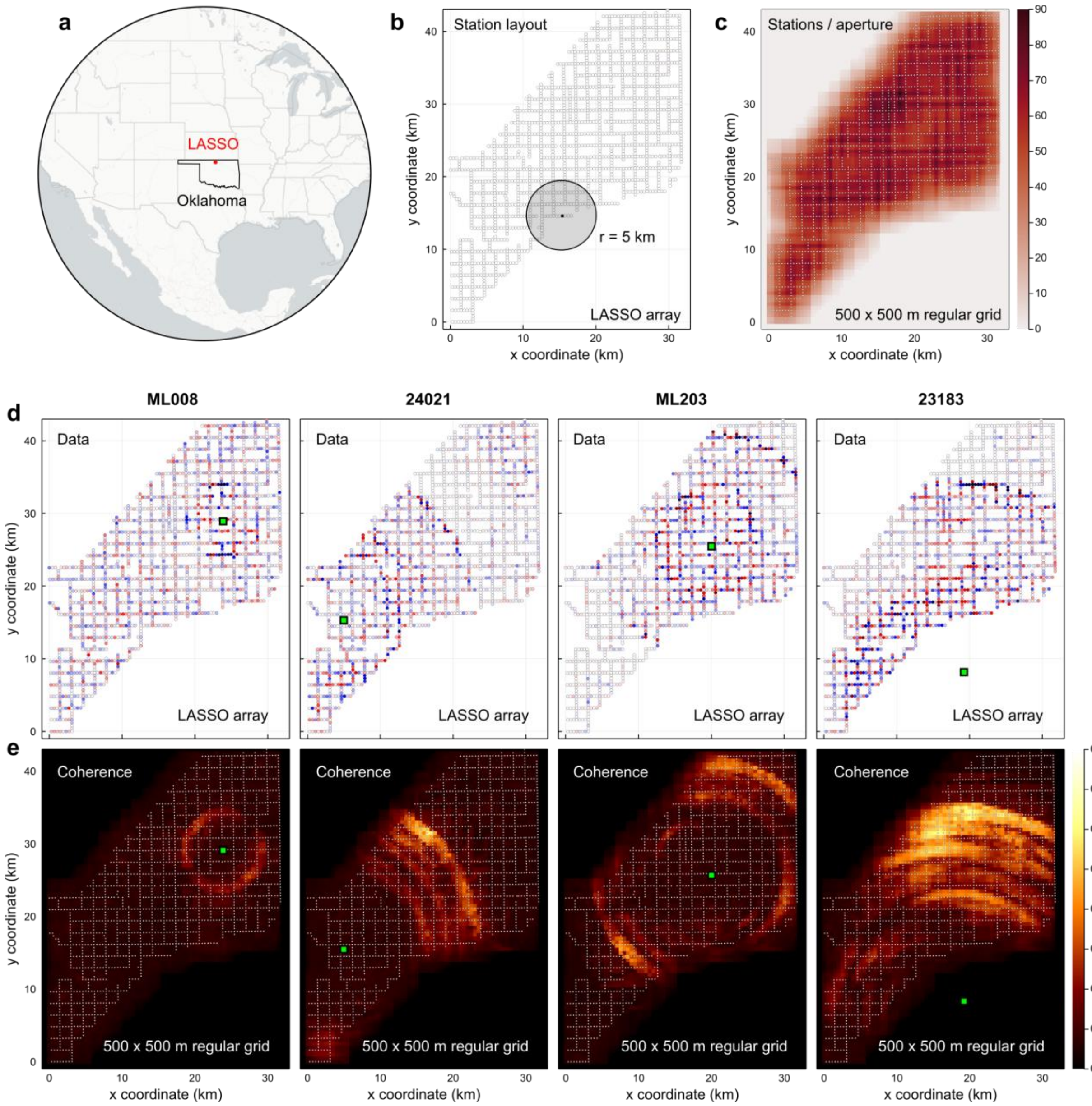


Figure 1. The location and layout of the LASSO array. (a) the LASSO array is located in Oklahoma, North America; (b) the station layout of the array, the small circles denote the positions of the ~1800 stations, the single large circle denotes the local aperture of 5 km that is mainly discussed in this work; (c) the discretization of the reconstruction grid is 500 m in x and y direction (trace density maps for apertures of 3 km and 7 km can be found in Figure S1). The reference point (x, y) = (0, 0) corresponds to x= 579000 and y= 4051000 in the UTM coordinate system. (d) Temporal snapshots of the raw waveform data recorded for events ML008, 24021, ML203, and 23183 (from left to right), respectively. (e) Semblance estimated locally on the regular grid. The green markers denote the cataloged epicenter locations.

4. Results

4.1 Waveform Coherence Analysis

Wavefield snapshots of events recorded with the LASSO array indicate that the dense station coverage allows to infer the directionality and curvature of wavefronts. Despite the favorable station configuration, however, waveform coherence is either compromised by the presence of noise sources, or by coda complexity indicating the presence of structure-related secondary sources in the near subsurface. In Figure 1(d), such snapshots are displayed for events ML008, 24021, ML203, and 23183, respectively. While for very small magnitudes, the primary signal threatens to drown in the ambient noise field (compare e.g., event ML008), stronger events allow for the discrimination of different phases, independent of whether the sources were located directly underneath or outside of the array. Figure 1(e) shows the distribution of waveform coherence, represented by the semblance coefficient (equation (1)) evaluated on a dense regular grid with an increment of 500 m in x and y direction, respectively. For all four considered events, first arrivals appear as prominent coherent signatures that follow the primary wavefront traversing the LASSO array. A comparison with the raw waveform data displayed in (d) indicates that especially later-arriving phases of lower amplitude become more easily recognizable in the positive-definite semblance map (e). The semblance can be directly utilized to enhance the waveform consistency and SNR (compare Figures S2-S5). Accompanying the regular semblance grid are fields of local estimates of the horizontal slope vector (p_x, p_y) as a by-product of the optimization procedure. The estimated slope and semblance fields allow for a multitude of applications, including but not limited to wavefront-tomographic inversion (Diekmann, et al. 2019).

First and foremost, following equation (3), these quantities enable the reconstruction of the recorded seismic waveforms by performing coherence-weighted local directional summation (averaging) within circular apertures of 5 km radius. Following this strategy, folds of up to 90 can be reached in the central parts of the array and a natural tapering occurs once the spatial grid leaves the array (compare Figure 1(c)). In Figure 2, results of coherent wavefield reconstruction of event ML203 on the 500 x 500 m spatial grid are compared with the respective raw waveform data that went into the analysis. To more systematically investigate the noise suppression performance of the method, three realistic synthetic reproductions of the event are shown alongside the field data recordings and their reconstruction. The three synthetic versions of ML203 are characterized by different noise levels, covering the noise-free case and SNRs of ~1 and less than 1, respectively. For all three synthetic realizations and the original field data, the reconstruction preserves coherent energy and suppresses incoherent noise, resulting in cleaner, more continuous, and more resolved datasets that can subsequently be used for improved source imaging. Although this is not fully verified yet, in the field data case, circular shapes in the northern part of the primary wavefront (compare the fourth column in Figure 2(b)) might indicate the presence of a secondary source – possibly related to distinct lateral change in structure – that could have been excited by event ML203.

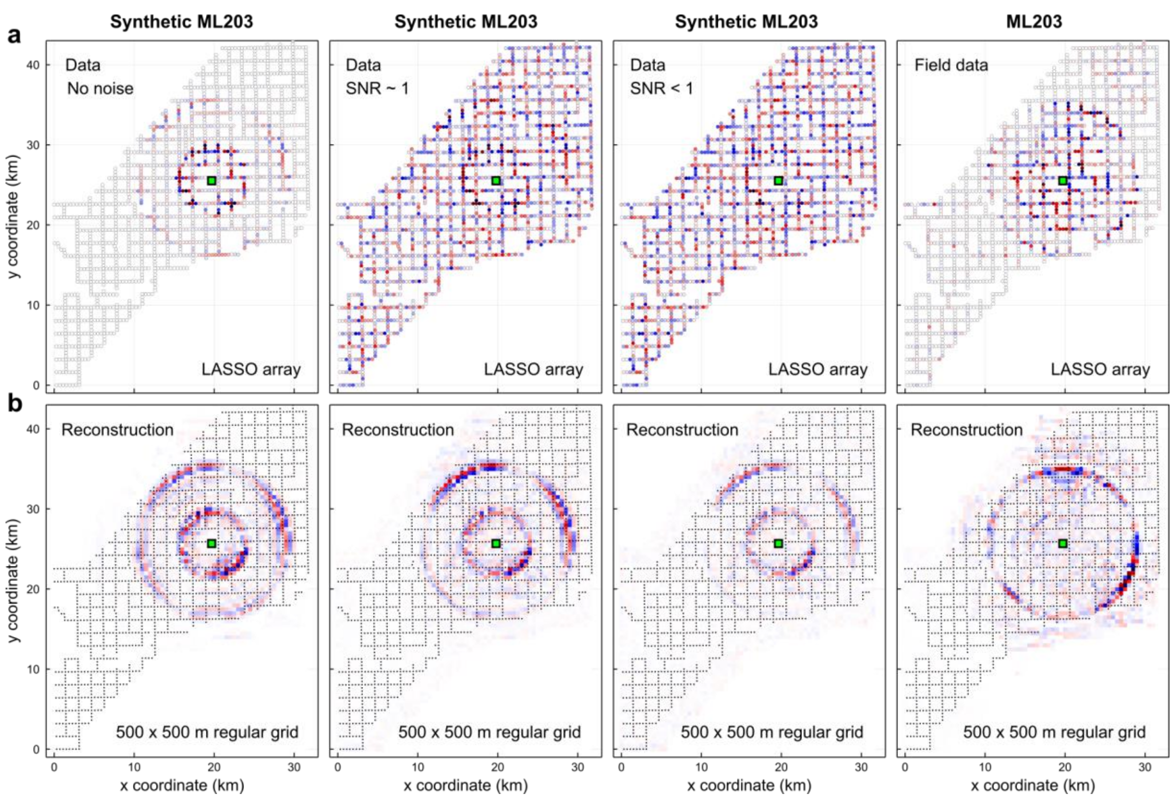


Figure 2. Coherent wavefield reconstruction performed on the same 500 m x 500 m spatially regular grid as for the semblance optimization, whose results are displayed in Figure 2(b). Displayed are 4 different versions of event ML203 – noise-free synthetics, synthetics with SNR ~1, synthetics with SNR <1, and the actual field data recordings (from left to right). Like in Figure 2, the cataloged event location and the station locations in (b) are denoted by a green marker and black dots, respectively. Results for all the four field events using different apertures and time series plots can be found in Figures S3, S4 and S5).

4.2 Stacking-based Seismic Source Images

In the following, we consider P-waves only when stacking the source energy using equation (3), since only vertical components are available in field data and the P-wave, as the primary phase, is not contaminated by subsequent phases.

Figure 3 shows the source imaging results of STA/LTA traces for the synthetic and field event ML203, directly corresponding to Figure 2. The relatively large number of stations in this dense array can tolerate a certain amount of noise, while the reconstructed waveforms produce higher imaging resolution even though the raw waveforms are contaminated by high-level noise. As observed and discussed in Section 4.1, the reconstructed waveforms exhibit a higher coherence and SNR, which naturally produce stacking-based source images with more coherent energy concentration and higher imaging resolution. The horizontal locations are basically consistent with reference values from the catalog and/or previous studies (indicated as white circles). The slightly biased depth locations are mainly resulting from the combined effects of velocity uncertainty and the inherent depth-origin time tradeoff. The dense surface array also yields higher imaging resolution in the horizontal direction than in the depth direction.

Figure 4 shows the source imaging results of raw waveforms for the two controlled simulations of events ML203 and 24021. After wavefield reconstruction, the imaging results show only few and comparably weak secondary peaks and fewer artifacts. The improvements in Figure 4(a) indicate that while trace summation carried out during reconstruction leads to decreased noise levels not only in the reconstructed domain, but also in the subsequently formed image. Figure 4(b) on the other hand illustrates the positive impact the regularizing and interpolating capabilities of the reconstruction have on waveform-based high-resolution source imaging. In summary, the reconstructed waveforms from this dense array enable high-resolution source images with polarity-uncorrected waveforms. The results for two other field events (ML008 and 23183) can be found in Figures S6 and S7. The highly consistent imaging results of the same event using different input waveforms further demonstrate the reliability of the proposed reconstruction strategy.

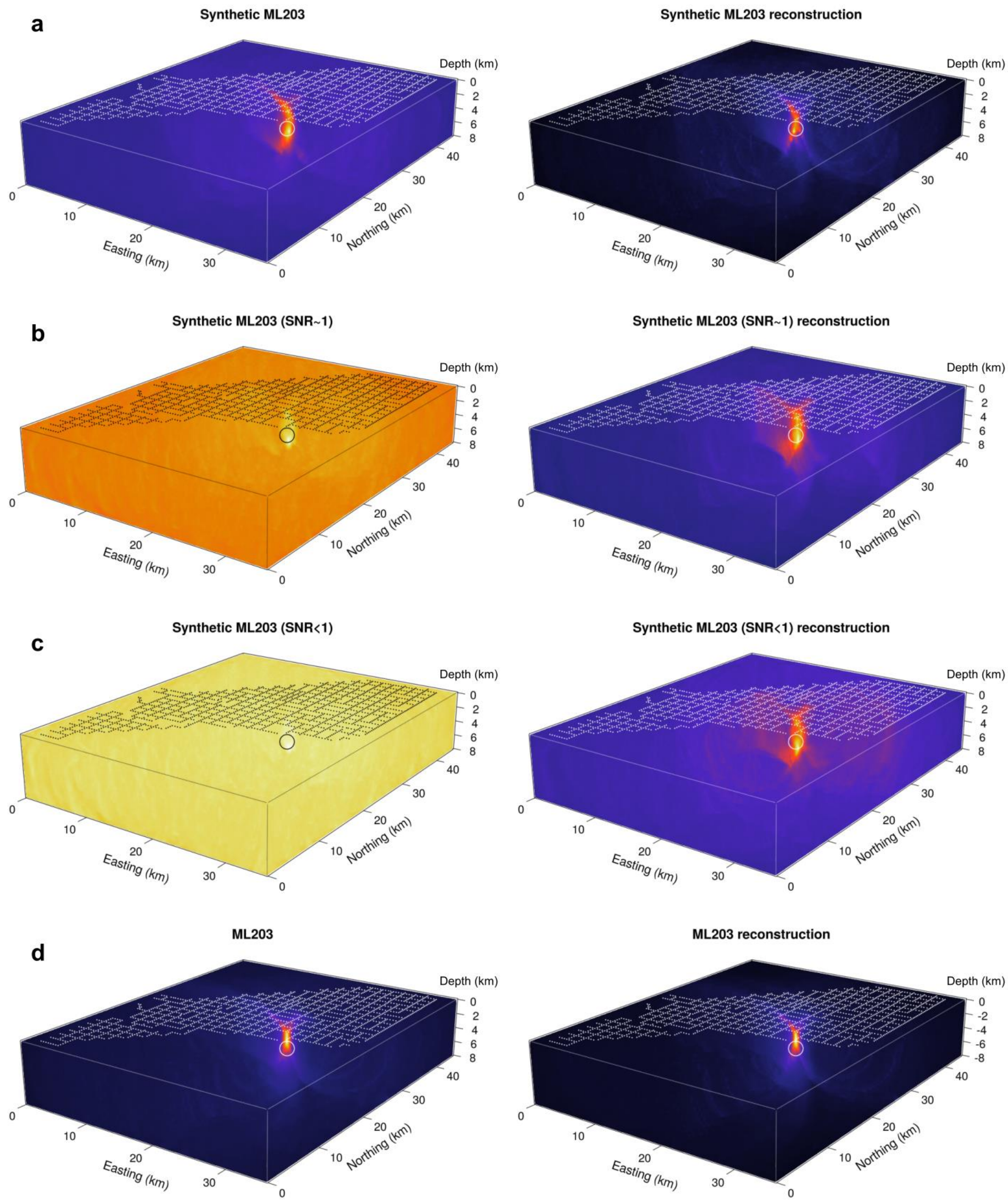


Figure 3. The source imaging results of STA/LTA traces for the synthetic and field event ML203. The left column corresponds to the result of using unreconstructed waveforms, and the right column images result from the use of wavefield reconstruction. (a) to (c) correspond to results gained with realistic synthetic waveforms with different noise levels applied, (d) shows the results of the field event ML203. Reference locations from the catalog and/or previous studies are indicated by a circle.

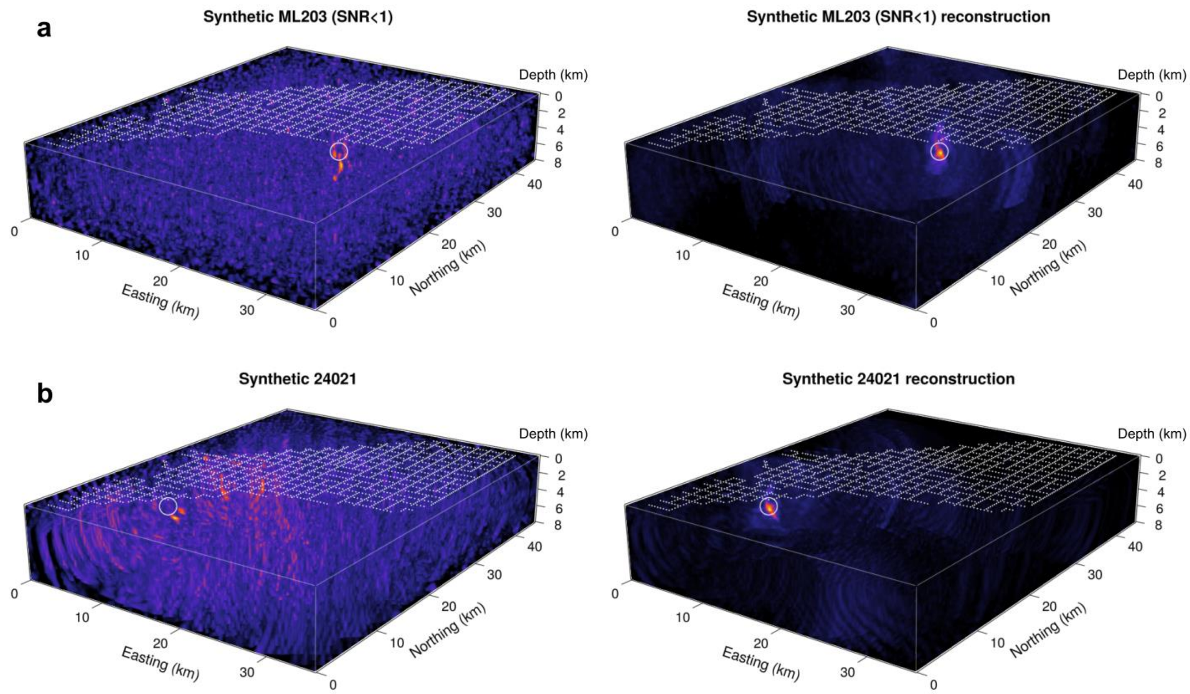


Figure 4. The source imaging results of raw traces for the synthetic event ML203 (a) and synthetic event 24021 (b). The left column corresponds to the result of unreconstructed waveforms, and the right column images made use of wavefield reconstruction. Real locations are indicated as white circles. Whereas in (a), severe noise contamination of the raw data leads to a strong noise footprint in the image, the strong artifacts in the raw-data image in (b) result from insufficient spatial sampling of the LASSO array in the vicinity of the source.

Compared with polarity-uncorrected raw traces, the corresponding stacked values are smeared more severely and the imaging resolution is lower for STA/LTA traces mainly due to the lower waveform resolution (compare Figures S6 and S7). Since the STA/LTA function further improves the SNR by flattening the waveforms and suppressing phase information, the respective imaging profiles are cleaner but exhibit stronger footprints surrounding the source area, suggesting prominent energy concentration along the isochronous surfaces. Consequently, the results of STA/LTA traces involving different input waveforms are relatively stable and the contrast of their imaging resolution is gentler. It is worth noting that using original polarity-uncorrected waveforms to stack the source energy involves the possibility of blurring the inferred source locations, if the combined effects of source mechanism and source-receiver geometry accidentally yield severely destructive results (e.g., Zhebel & Eisner, 2015). Alternatively, with more coherent and regular waveform records, we can obtain reliable source images and location estimates with even fewer traces, and thus, help to lower computational demands for source location and other subsequent processing tasks.

5. Discussion

While the noise suppression and regularization capabilities of coherent wavefield reconstruction could be successfully demonstrated for the LASSO array, it needs to be stated that not all array configurations are equally suitable for such a reconstruction. Especially very directional array configurations, such as the typically T-shaped array geometries introduced for nuclear test monitoring make a 3D spatio-temporal reconstruction of the wavefield more challenging in practice. However, if only coherent data enhancement is concerned, 2D projections of wavefields can still be successfully regularized and enhanced with reduced (2D) moveout operators and directional (2D) apertures. One prominent use case of such 2D projections is fiber-optic distributed acoustic sensing (DAS), where strain or strain rate is recorded along the local axis of buried fiber-optic cables (e.g., Lindsey et al., 2017; Jousset et al., 2018). In the context of volcano monitoring, coherent data enhancement techniques were demonstrated to improve the sensitivity and overall data quality of DAS arrays (Jousset et al., 2022). No matter whether 2D or 3D applications are concerned, the appropriate choice of local spatial aperture dimensions are crucial for the success of the method. There exists a natural trade-off between fold (the more neighboring traces are included the better are the noise suppression capabilities) and wavefield complexity (the smaller the aperture, the more wavefield complexity can be honored). So, while a simple circular aperture radius of 5 km led to reasonable results for LASSO, more complicated and possibly spatially varying aperture dimensions should be utilized under less favorable conditions (compare supplementary Figures S2 and S3).

We present the imaging results for all considered events to further demonstrate the effectiveness of waveform reconstruction, but the source-receiver geometry is not the only dominant factor for stacking-based methods, other factors like velocity model and stacking operators also directly affect the final imaging results. Besides, this work only presents a qualitative evaluation of the imaging resolution. A detailed quantitative analysis of the imaging resolution is out of the scope of the current study, and we are aware that a comprehensive uncertainty evaluation of source imaging remains challenging so far. Though we only investigate the performance of coherent wavefield reconstruction with stacking-based seismic location, we believe the proposed method can introduce benefits for other array-based techniques and other scenarios associated with dense seismic arrays. For example, time reverse techniques rely on dense and regular wavefields to produce coherent and focused imaging results (e.g., Werner & Saenger, 2018). We believe reconstructed wavefields have good potential to achieve improved subsurface structures and source characterizations by enhancing the imaging profiles. Besides, the enhanced and regularized wavefields can be used for various subsequent seismic processing, such as constraining velocity tomography and seismic migration, detecting small earthquakes, and facilitating source mechanism inversion.

6. Conclusions

We proposed a novel reconstruction method for coherent seismic wavefields with dense arrays. The spatiotemporal wavefield coherence embedded in the dense seismic array is quantified and utilized to reconstruct the wavefields. The summation-based techniques enable the method adapted to weak events with low SNRs. Application to both realistic and field seismic events recorded by the dense LASSO array in Oklahoma reveals improved SNR, data coherence and regularity. Results of events at different locations demonstrate the applicability of the proposed method in waveform reconstruction to general source distributions. To further examine the merits of the proposed method, we tested the reconstructed waveforms using stacking-based location and compared the imaging results with those of unreconstructed waveforms. Reconstructed waveforms produce better location results accounting for the SNR and resolution of the images, due to their higher SNRs and data coherence than unreconstructed waveforms.

Data Availability Statement

Field waveform data in this study are archived by the IRIS-DMC and are available for public download using the PH5 Web Services Interface under network code 2A (http://service.iris.edu/ph5ws/dataselect/docs/1/builder/, last accessed March 2023). Synthetic waveforms are generated with the open-source *FDwave3D* package (L. Li et al., 2021). The reconstruction and source imaging results are freely accessible in an open repository (L. Li & Schwarz, 2023). The easily reproducible reconstruction algorithm was written in the high-performance programming language Julia (Bezanson et al., 2017) and semblance maximization was accomplished with Julias global optimization library *BlackBoxOptim.jl* (Feldt & Stukalov, 2018). Visualizations (images and animations) were generated with the Julia plotting API *Plots.jl* and *Makie.jl* – a flexible open-source plotting library for the Julia language (Danisch & Krumbiegel, 2021).

Acknowledgements

We thank the LASSO deployment volunteers for collecting the data and making them openly available to the research community. Lei Li acknowledges the support from Natural Science Foundation for Excellent Young Scholars of Hunan Province, China (Grant No. 2022JJ20057), Central South University Innovation-Driven Research Programme (Grant No. 2023CXQD064), and National Natural Science Foundation of China (Grant No. 42004115). Benjamin Schwarz is grateful to Karin Sigloch, Tarje Nissen-Meyer, Charlotte Krawczyk, Christian Haberland and Michael Weber for insightful discussions.

References

Baker, T., Granat, R., & Clayton, R. W. (2005). Real-time earthquake location using Kirchhoff reconstruction. *Bulletin of the Seismological Society of America*, *95*(2), 699–707.

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM review*, *59*(1), 65-98. https://doi.org/10.1137/1410006

Bauer, A., Schwarz, B., & Gajewski, D. (2017). Utilizing diffractions in wavefront tomography. *Geophysics*, *82*(2), R65-R73.

Bortfeld, R. (1989). Geometrical ray theory: Rays and traveltimes in seismic systems (second-order approximations of the traveltimes). *Geophysics*, *54*(3), 342-349.

Chen, Y., Bai, M., & Chen, Y. (2019). Obtaining free USArray data by multi-dimensional seismic reconstruction. *Nature Communications*, *10*(1), 4434. https://doi.org/10.1038/s41467-019-12405-0

Christ, S., Schwabeneder, D., & Rackauckas, C. (2022). Plots..jl – a user extendable plotting API for the julia programming language. *arXiv*. https://doi.org/10.48550/arXiv.2204.08775

Cochran, E. S., Wickham-Piotrowski, A., Kemna, K. B., Harrington, R. M., Dougherty, S. L., & Peña Castro, A. F. (2020). Minimal Clustering of Injection-Induced Earthquakes Observed with a Large-n Seismic Array. *Bulletin of the Seismological Society of America*, *110*(5), 2005–2017. https://doi.org/10.1785/0120200101

Das, S., & Suganthan, P. N. (2011). Differential Evolution: A Survey of the State-of-the-Art. *IEEE Transactions on Evolutionary Computation*, *15*(1), 4–31. https://doi.org/10.1109/tevc.2010.2059031

Danisch, S., & Krumbiegel, J. (2021). Makie.jl: Flexible high-performance data visualization for Julia. *Journal of Open Source Software*, *6*(65), 3349. https://doi.org/10.21105/joss.03349

Diekmann, L., Schwarz, B., Bauer, A., & Gajewski, D. (2019). Source localization and joint velocity model building using wavefront attributes. *Geophysical Journal International*, *219*(2), 995-1007.

Dougherty, S. L., Cochran, E. S., & Harrington, R. M. (2019). The LArge-n Seismic Survey in Oklahoma (LASSO) Experiment. *Seismological Research Letters*, *90*(5), 2051–2057. https://doi.org/10.1785/0220190094

Duveneck, E. (2004). Velocity model estimation with data-derived wavefront attributes. *Geophysics*, *69*(1), 265-274.

Esmersoy, C., & Miller, D. (1989). Backprojection versus backpropagation in multidimensional linearized inversion. *GEOPHYSICS*, *54*(7), 921–926. https://doi.org/10.1190/1.1442722

Feldt, R., & Stukalov, A. (2018). BlackBoxOptim.jl: Black-box optimization for Julia (Version 0.6.2) [Software]. https://github.com/robertfeldt/BlackBoxOptim.jl

Furumura, T., & Maeda, T. (2021). High-resolution source imaging based on time-reversal wave propagation simulations using assimilated dense seismic records. *Geophysical Journal International*, *225*(1), 140–157. https://doi.org/10.1093/gji/ggaa586

Gajewski, D., Anikiev, D., Kashtan, B., & Tessmer, E. (2007). Localization of seismic events by diffraction stacking. In *SEG Technical Program Expanded Abstracts 2007* (pp. 1287–1291). Society of Exploration Geophysicists.

Gibbons, S. J., & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, *165*(1), 149–166. https://doi.org/10.1111/j.1365-246X.2006.02865.x

Grigoli, F., Cesca, S., Vassallo, M., & Dahm, T. (2013). Automated Seismic Event Location by Travel-Time Stacking: An Application to Mining Induced Seismicity. *Seismological Research Letters*, *84*(4), 666–677.

Hansen, S. M., & Schmandt, B. (2015). Automated detection and location of microseismicity at Mount St. Helens with a large-N geophone array. *Geophysical Research Letters*, *42*(18), 7390–7397. https://doi.org/10.1002/2015GL064848

Höcht, G., Ricarte, P., Bergler, S., & Landa, E. (2009). Operator-oriented CRS interpolation. *Geophysical Prospecting*, *57*(6), 957-979.

Ishii, M., Shearer, P. M., Houston, H., & Vidale, J. E. (2005). Extent, duration and speed of the 2004 Sumatra–Andaman earthquake imaged by the Hi-Net array. *Nature*, *435*(7044), 933–936. https://doi.org/10.1038/nature03675

Jäger, R., Mann, J., Höcht, G., & Hubral, P. (2001). Common-reflection-surface stack: Image and attributes. *Geophysics*, *66*(1), 97-109.

Jousset, P., Reinsch, T., Ryberg, T., Blanck, H., Clarke, A., Aghayev, R., et al. (2018). Dynamic strain determination using fibre-optic cables allows imaging of seismological and structural features. *Nature Communications*, *9*(1), 2509.

Jousset, P., Currenti, G., Schwarz, B., Chalari, A., Tilmann, F., Reinsch, T., et al. (2022). Fibre optic distributed acoustic sensing of volcanic events. Nature Communications, 13(1), 1753.

Kao, H., & Shan, S. J. (2004). The source-scanning algorithm: Mapping the distribution of seismic sources in time and space. *Geophysical Journal International*, *157*(2), 589–594.

Karplus, M., & Schmandt, B. (2018). Preface to the Focus Section on Geophone Array Seismology. *Seismological Research Letters*, *89*(5), 1597–1600. https://doi.org/10.1785/0220180212

Kemna, K. B., Peña Castro, A. F., Harrington, R. M., & Cochran, E. S. (2020). Using a Large- *n* Seismic Array to Explore the Robustness of Spectral Estimations. *Geophysical Research Letters*, *47*(21), e2020GL089342. https://doi.org/10.1029/2020GL089342

Li, K. L., Sadeghisorkhani, H., Sgattoni, G., Gudmundsson, O., & Roberts, R. (2017). Locating tremor using stacked products of correlations. *Geophysical Research Letters*, *44*(7), 3156–3164.

Li, L., Becker, D., Chen, H., Wang, X., & Gajewski, D. (2018). A systematic analysis of correlation-based seismic location methods. *Geophysical Journal International*, *212*(1), 659–678. https://doi.org/10.1093/gji/ggx436

Li, L., & Schwarz, B. (2023). LASSO coherent seismic wavefield reconstruction and source imaging [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7842090

Li, L., Tan, J., Zhang, D., Malkoti, A., Abakumov, I., & Xie, Y. (2021). FDwave3D: A MATLAB solver for the 3D anisotropic wave equation using the finite-difference method. *Computational Geosciences*, *25*, 1565–1578.

Li, Lei, Xie, Y., & Tan, J. (2020). Application of Waveform Stacking Methods for Seismic Location at Multiple Scales. *Energies*, *13*(18), 4729. https://doi.org/10.3390/en13184729

Li, Lei, Tan, J., Schwarz, B., Staněk, F., Poiata, N., Shi, P., et al. (2020). Recent advances and challenges of waveform-based seismic location methods at multiple scales. *Reviews of Geophysics*, *58*(1), e2019RG000667. https://doi.org/10.1029/2019RG000667

Li, Lei, Wong, W., Schwarz, B., & Lau, T. (2022). Seismology Perspectives on Integrated, Coordinated, Open, Networked (ICON) Science. *Earth and Space Science*, *9*, e2021EA002109.

Li, Zefeng, Peng, Z., Hollis, D., Zhu, L., & McClellan, J. (2018). High-resolution seismic event detection using local similarity for Large-N arrays. *Scientific Reports*, *8*(1), 1646. https://doi.org/10.1038/s41598-018-19728-w

Li, Zhengbo, Shi, C., Ren, H., & Chen, X. (2022). Multiple Leaking Mode Dispersion Observations and Applications From Ambient Noise Cross-Correlation in Oklahoma. *Geophysical Research Letters*, *49*(1), e2021GL096032. https://doi.org/10.1029/2021GL096032

Lin, F.-C., Li, D., Clayton, R. W., & Hollis, D. (2013). High-resolution 3D shallow crustal structure in Long Beach, California: Application of ambient noise tomography on a dense seismic array. *Geophysics*, *78*(4), Q45–Q56. https://doi.org/10.1190/geo2012-0453.1

Lindsey, N. J., Martin, E. R., Dreger, D. S., Freifeld, B., Cole, S., James, S. R., et al. (2017). Fiber-optic network observations of earthquake wavefields. *Geophysical Research Letters*, *44*(23), 11-792.

Neidell, N. S., & Taner, M. T. (1971). Semblance and other coherency measures for multichannel data. *Geophysics*, *36*(3), 482-497.

Peña Castro, A. F., Dougherty, S. L., Harrington, R. M., & Cochran, E. S. (2019). Delayed Dynamic Triggering of Disposal-Induced Earthquakes Observed by a Dense Array in Northern Oklahoma. *Journal of Geophysical Research: Solid Earth*, *124*(4), 3766–3781. https://doi.org/10.1029/2018JB017150

Rost, S., & Thomas, C. (2002). Array Seismology: Methods and Applications. *Reviews Of Geophysics*, *40*(3), 2–1.

Rubinstein, J. L., Ellsworth, W. L., & Dougherty, S. L. (2018). The 2013–2016 Induced Earthquakes in Harper and Sumner Counties, Southern Kansas. *Bulletin of the Seismological Society of America*, *108*(2), 674–689. https://doi.org/10.1785/0120170209

Schimmel, M., & Paulssen, H. (1997). Noise reduction and detection of weak, coherent signals through phase-weighted stacks. *Geophysical Journal International*, *130*(2), 497-505.

Schwarz, B. (2019). Coherent wavefield subtraction for diffraction separation. *Geophysics*, *84*(3), V157-V168.

Shi, P., Angus, D., Rost, S., Nowacki, A., & Yuan, S. (2019). Automated seismic waveform location using multichannel coherency migration (MCM)–I: theory. *Geophysical Journal International*, *216*(3), 1842–1866. https://doi.org/10.1093/gji/ggy132

Steiner, B., Saenger, E. H., & Schmalholz, S. M. (2008). Time reverse modeling of low-frequency microtremors: Application to hydrocarbon reservoir localization. *Geophysical Research Letters*, *35*, L03307. https://doi.org/10.1029/2007GL032097

Trugman, D. T., Chu, S. X., & Tsai, V. C. (2021). Earthquake Source Complexity Controls the Frequency Dependence of Near-Source Radiation Patterns. *Geophysical Research Letters*, *48*, e2021GL095022. https://doi.org/10.1029/2021GL095022

Werner, C., & Saenger, E. H. (2018). Obtaining reliable source locations with time reverse imaging: limits to array design, velocity models and signal-to-noise ratios. *Solid Earth*, *9*(6), 1487–1505. https://doi.org/10.5194/se-9-1487-2018

Yang, L., Liu, X., Zhu, W., Zhao, L., & Beroza, G. C. (2022). Toward improved urban earthquake monitoring through deep-learning-based noise suppression. *Science Advances*, *8*(15), eabl3564. https://doi.org/10.1126/sciadv.abl3564

Zhebel, O., & Eisner, L. (2015). Simultaneous microseismic event localization and source mechanism determination. *Geophysics*, *80*(1), KS1–KS9.

Geophysical Research Letters

Supporting Information for

Coherent wavefield reconstruction improves event location with dense seismic arrays

Lei Li[1,2,3] and Benjamin Schwarz[4*]

1. Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring (Central South University), Ministry of Education, 410083 Changsha, China.
2. Hunan Key Laboratory of Nonferrous Resources and Geological Hazard Exploration, 410083 Changsha, China.
3. School of Geosciences and Info-Physics, Central South University, 410083 Changsha, China.
4. GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany.

Contents of this file

Other Supplementary Material for this manuscript includes the following:

Movies S1 to S5

Introduction

The text, figures and table included in this document are designed to complement and support the analysis presented in the main text.

Text S1: Generation of realistic waveform data

To generate realistic synthetic data, we use the same model parameters of event ML203 and event 24021, including the velocity model, source parameters, and station geometry, to simulate the waveforms, and then add corresponding field noise from each station to respective traces of synthetic waveforms. The noise is created and defined by the following equation (Staněk et al., 2014) :

$$Noise = \frac{A_{noise}}{mean(|A_{noise}|)} \cdot mean(|A_{synthetic}|) \cdot NL \, , \qquad \text{(S1)}$$

where A_{noise} is the noise retrieved from field waveforms on each trace at a given time window before the first arrival, $A_{synthetic}$ is the computed amplitudes for the given model parameters, NL means the level or intensity of the noise, which approximates the reversed value of the SNR. The selected noise amplitudes are extended recurrently to match the noise-free synthetic amplitudes. In this work, we select the first 400 samples of noise amplitudes from each trace to resemble the field noise and set the noise level as 0, 5, and 10, denoting them as noise-free, SNR~1, and SNR<1, respectively. The $Noise$ is then added to the original synthetic amplitudes to obtain realistic waveforms with field noise:

$$A_{real} = A_{synthetic} + Noise \, . \tag{S2}$$

The noisy synthetic waveforms are normalized trace by trace before entering the reconstruction and location workflow.

References

Staněk, F., Eisner, L., & Jan Moser, T. (2014). Stability of source mechanisms inverted from P-wave amplitude microseismic monitoring data acquired at the surface: Stability of source mechanisms inverted from P-wave amplitude data. *Geophysical Prospecting*, *62*(3), 475–490. https://doi.org/10.1111/1365-2478.12107

Figure S1. Number of traces in different apertures

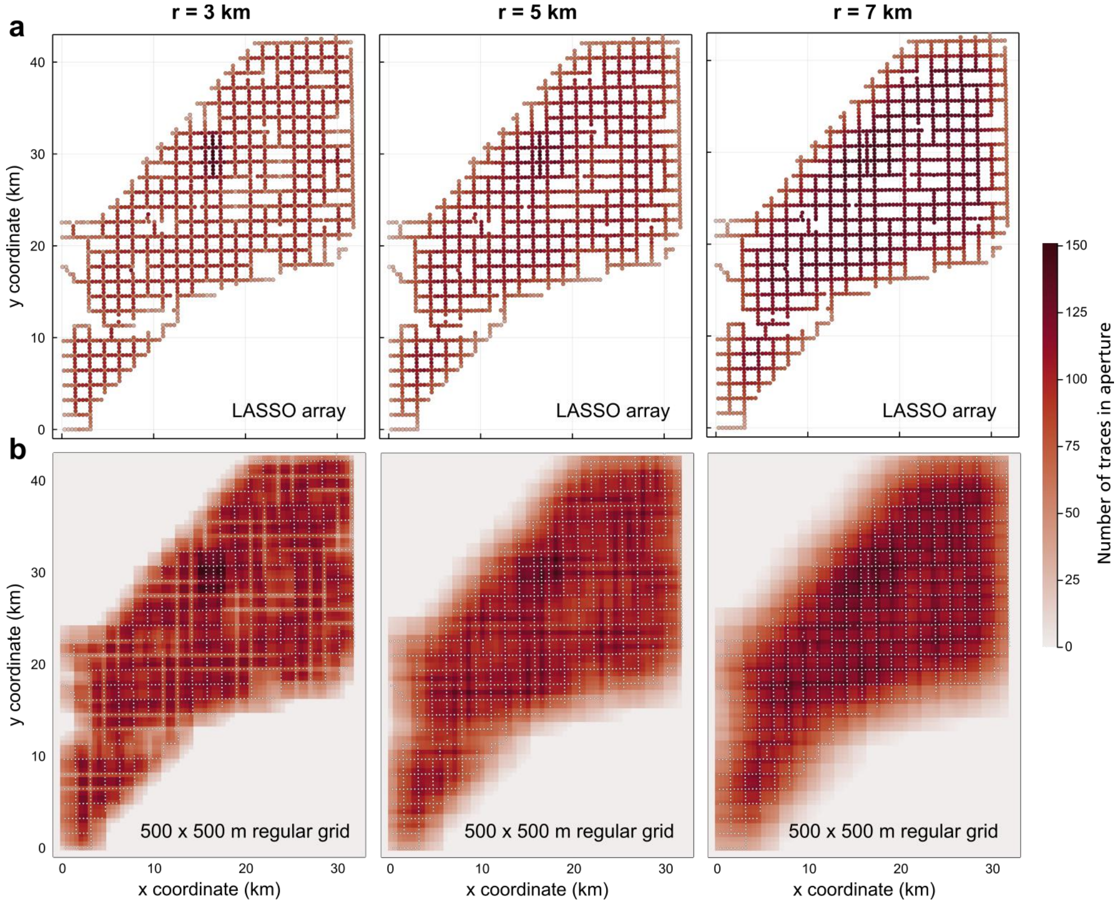


Figure S2. Coherence of the four field events using different apertures

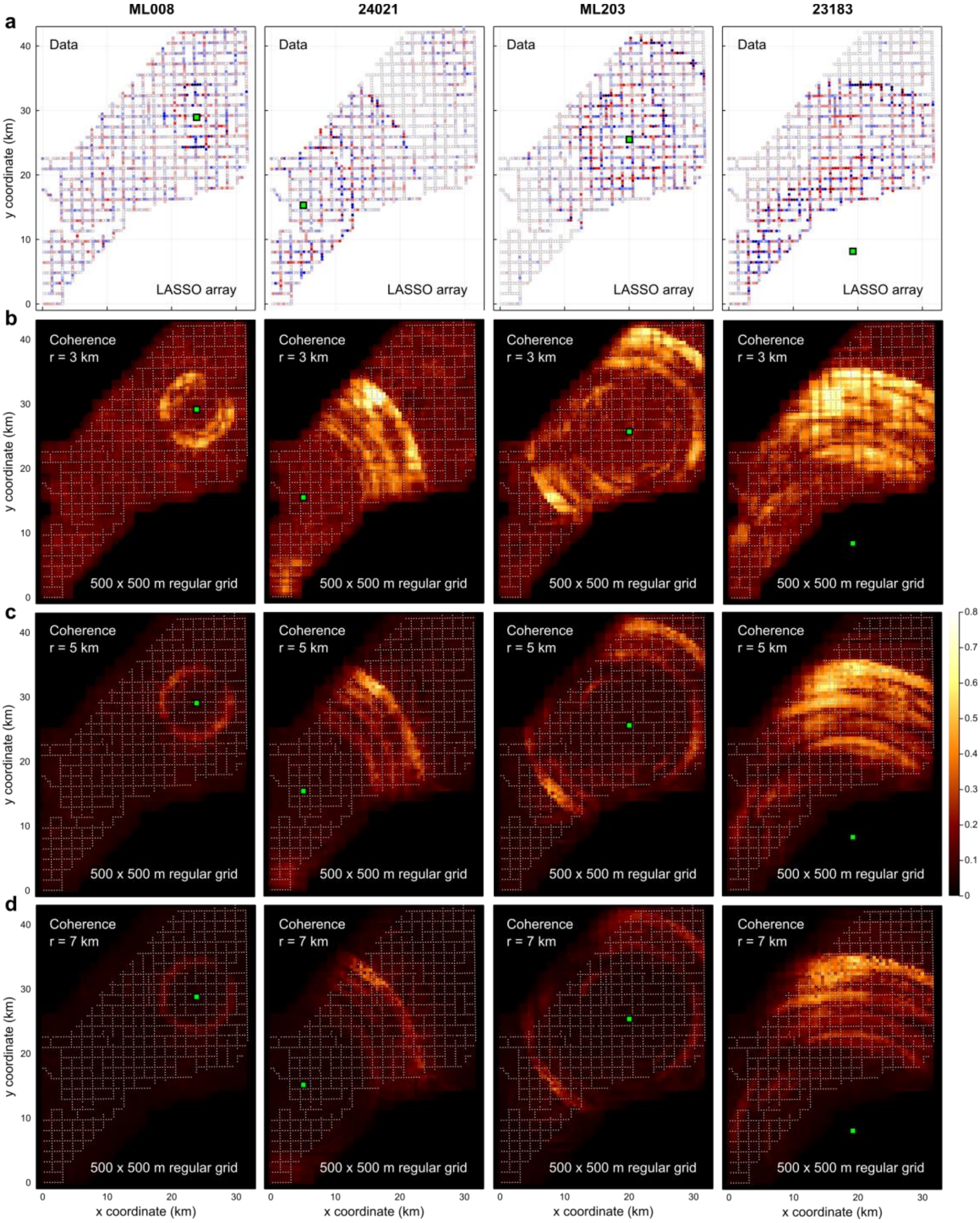


Figure S3. Reconstructed wavefields of the four field events using different apertures

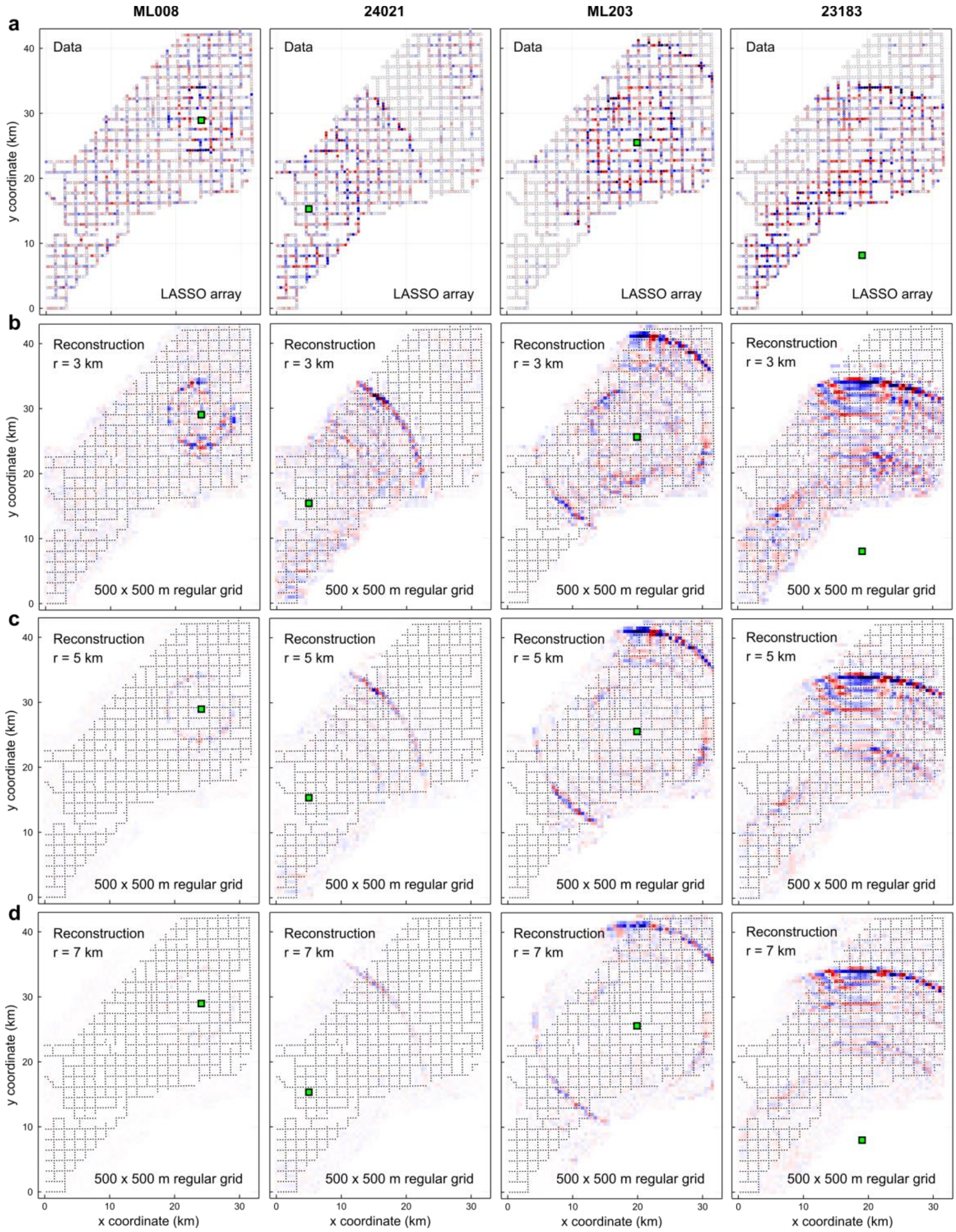


Figure S4. Unreconstructed and reconstructed waveforms of the four field events. (a) event ML008, (b) event 24021, (c) event ML203, (d) event 23183. The left column corresponds to raw traces of unreconstructed waveforms, and the right column are those of reconstructed waveforms. Only portions (50 traces) of the seismograms are shown.

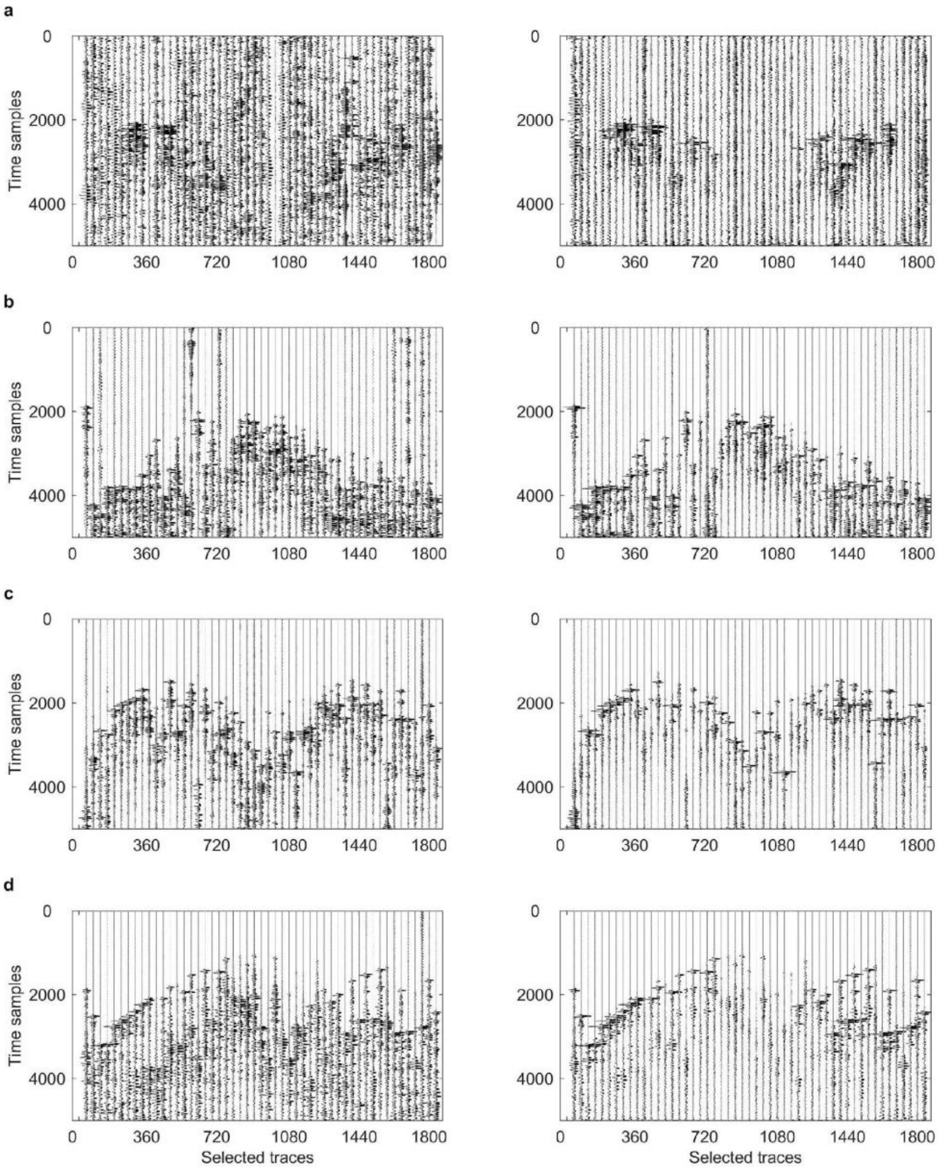


Figure S5. STA/LTA traces of the unreconstructed raw waveforms and reconstructed waveforms of the synthetic and field event ML203. The left column corresponds to STA/LTA traces of unreconstructed waveforms, and the right column are those of reconstructed waveforms. (a) to (c) correspond to different noise levels, (d) corresponds to the field event ML203. Only portions (about 50 traces) of the seismograms are shown.

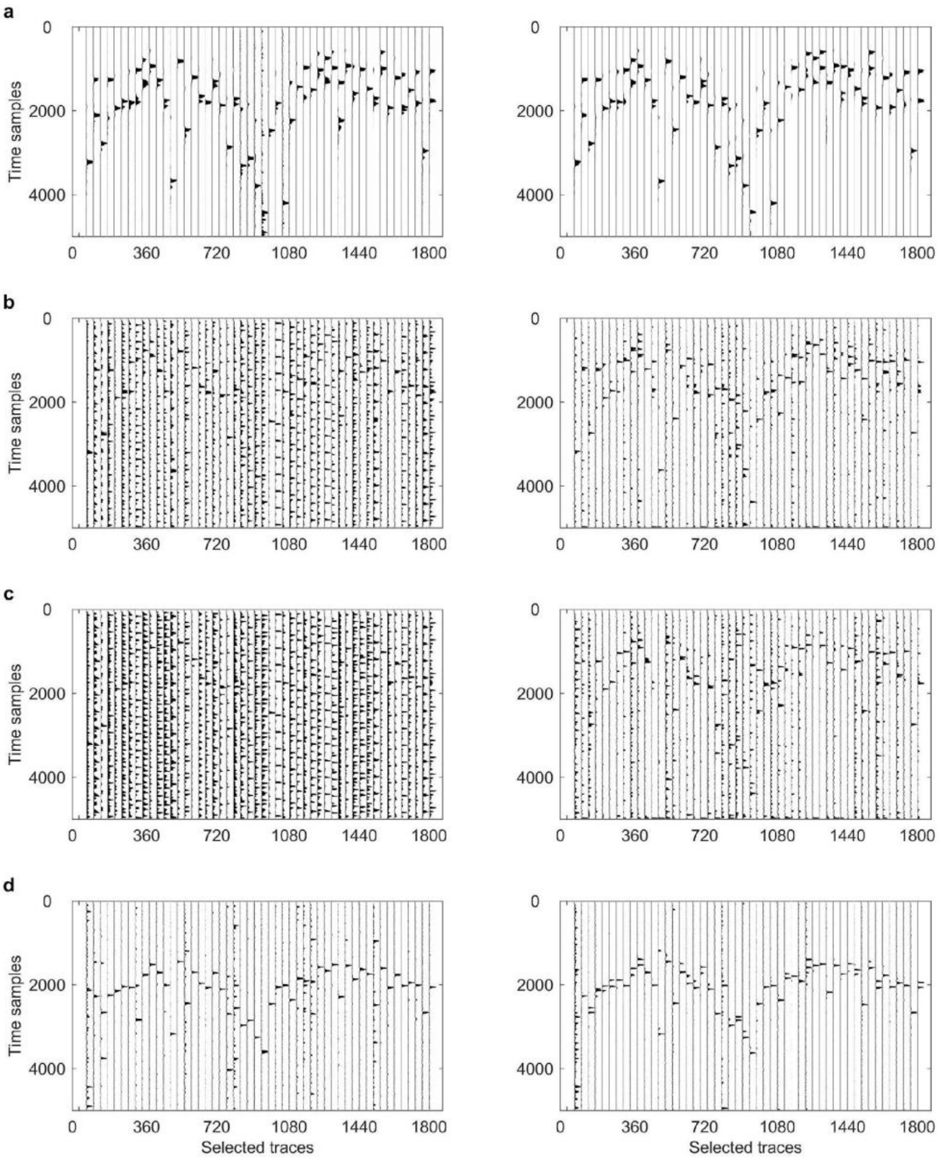


Figure S6. The source imaging results for the four field events using raw traces. (a) event ML008, (b) event 24021, (c) event ML203, (d) event 23183. The left column corresponds to the result of unreconstructed waveforms, and the right column images result from reconstructed waveforms. Reference locations from the catalog and/or previous studies are indicated as white circles. Please note the degradation of the depth resolution of event 24021 and event 23183, mainly caused by the array geometry and the properties of the chosen stacking operator. The horizontal resolution of the images, on the other hand, remains consistently high.

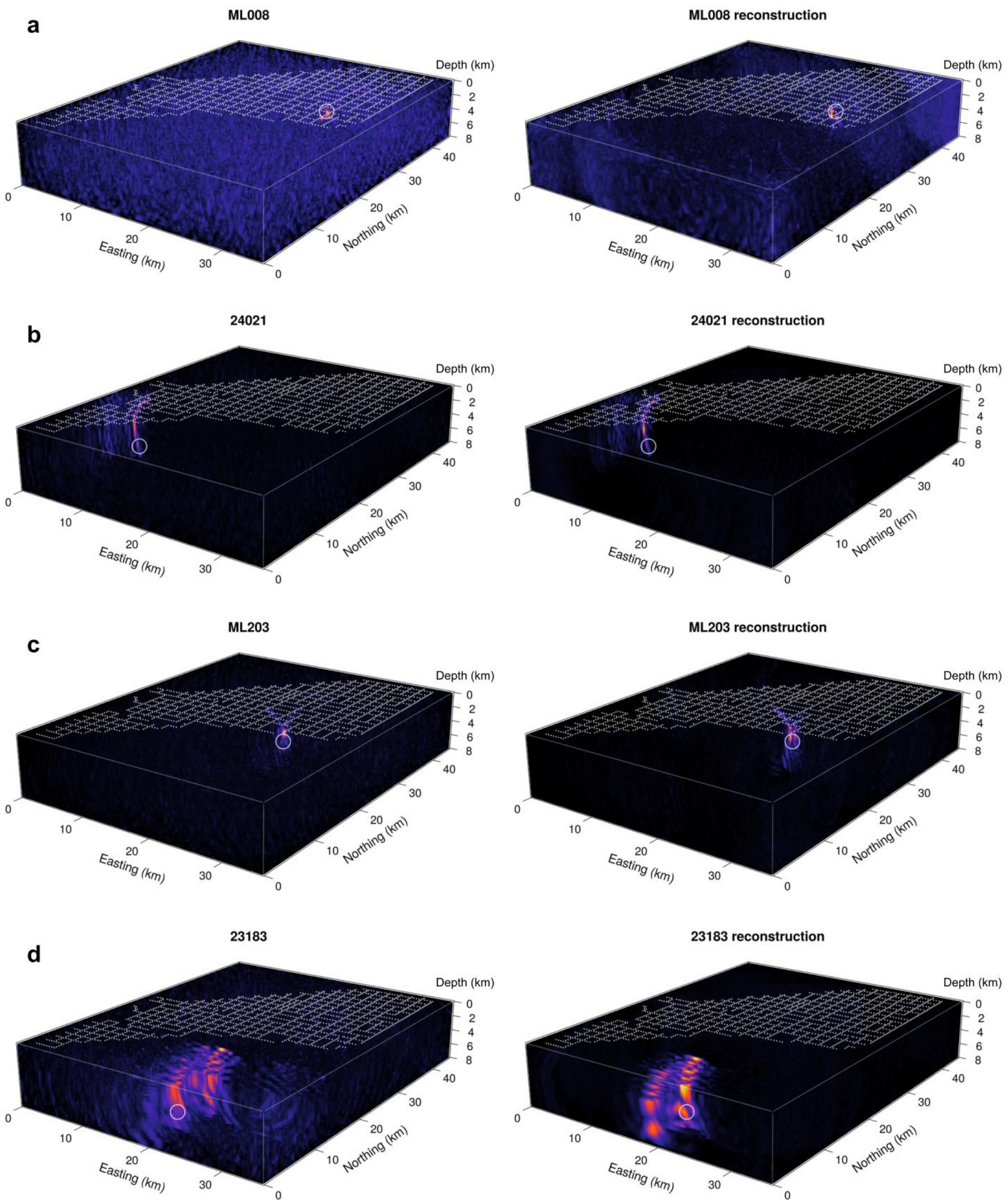
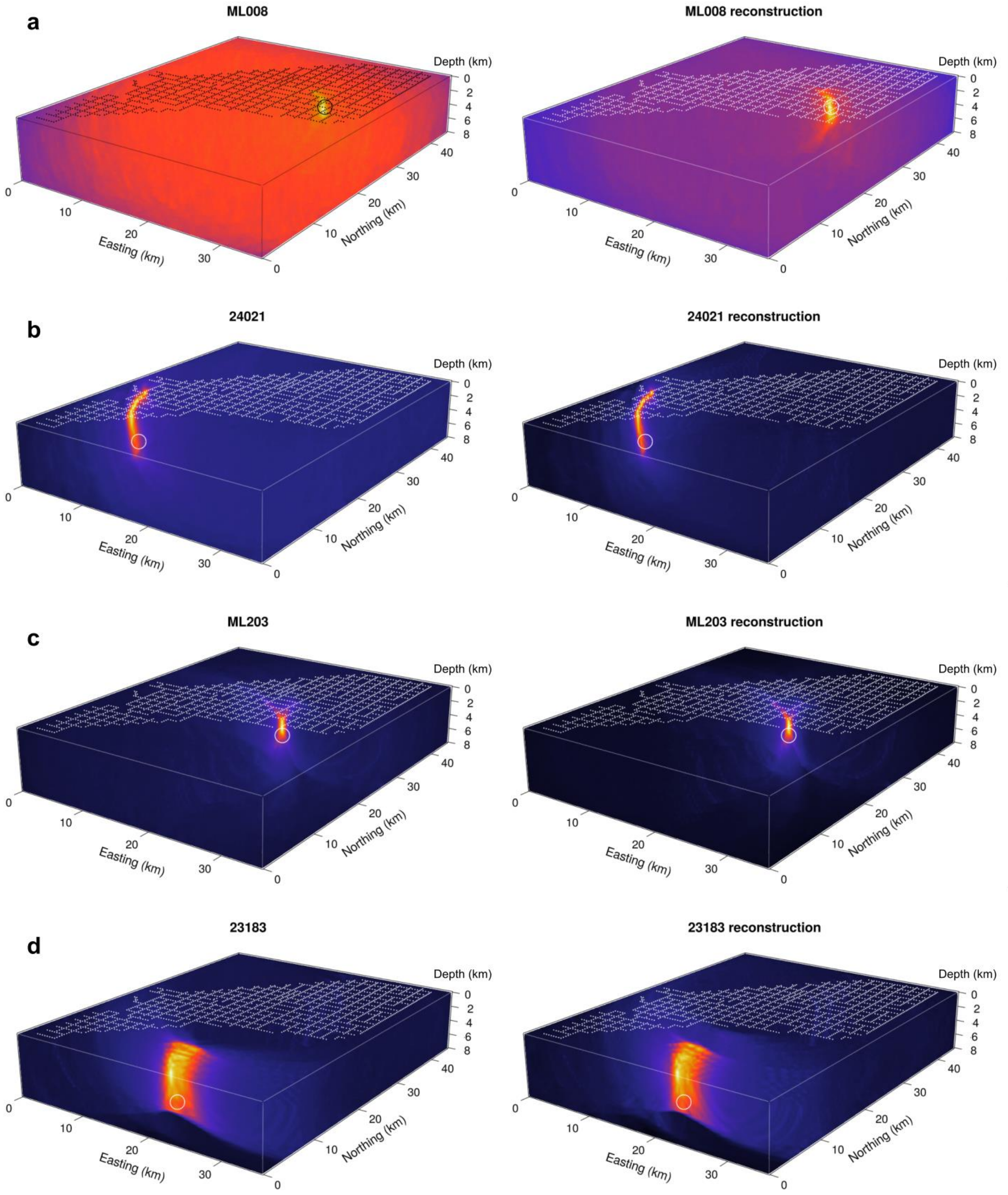


Figure S7. The source imaging results for the four field events using STA/LTA traces. (a) event ML008, (b) event 24021, (c) event ML203, (d) event 23183. The left column corresponds to the result of unreconstructed waveforms, and the right column images result from reconstructed waveforms. Reference locations from the catalog and/or previous studies are indicated by a circle.



Movies S1 to S5: Synchronized animations of the raw data (left), its reconstruction (center) and the estimated waveform coherence (right) for the four field events and the synthetic event ML203 (SNR<1).