DASPy: A Python Toolbox for DAS Seismology

Minzhe Hu[1] and Zefeng Li*[1,2]

Abstract

Distributed acoustic sensing (DAS) has emerged as a novel technology in geophysics, owing to its high-sensing density, cost effectiveness, and adaptability to extreme environments. Nonetheless, DAS differs from traditional seismic acquisition technologies in many aspects: big data volume, equidistant sensing, measurement of axial strain (strain rate), and noise characteristics. These differences make DAS data processing challenging for new hands. To lower the bar of DAS data processing, we develop an open-source Python toolbox called DASPy, which encompasses classic seismic data processing techniques, including preprocessing, filter, spectrum analysis, and visualization, and specialized algorithms for DAS applications, including denoising, waveform decomposition, channel attribute analysis, and strain–velocity conversion. Using openly available DAS data as examples, this article makes an overview and tutorial on the eight modules in DASPy to illustrate the algorithms and practical applications. We anticipate DASPy to provide convenience for researchers unfamiliar with DAS data and help facilitate the rapid growth of DAS seismology.

Cite this article as Hu, M., and Z. Li (2024). DASPy: A Python Toolbox for DAS Seismology, *Seismol. Res. Lett.* **95**, 3055–3066, doi: 10.1785/0220240124.

Introduction

Distributed acoustic sensing (DAS) is an emerging vibration monitoring technology increasingly utilized in geophysics. It converts fiber-optic cables into an ultradense seismic array with meter-scale spacing and a frequency range of 0.01–100 kHz. DAS recovers axial strain or strain rate along the fiber-optic cable by measuring the subtle optical phase shift of backscattered light within the fiber (Zhan, 2019; Lindsey and Martin, 2021). Over recent years, it has been demonstrated useful in many seismological applications such as earthquake monitoring (Lindsey *et al.*, 2017; Li and Zhan, 2018; Li *et al.*, 2021; Nayak *et al.*, 2021; Zeng *et al.*, 2022), source property estimate (Chen, 2023; Li, Kim, *et al.*, 2023; Li, Zhu, *et al.*, 2023), subsurface imaging (Dou *et al.*, 2017; Ajo-Franklin *et al.*, 2019; Cheng *et al.*, 2021; Luo *et al.*, 2021; Nayak and Ajo-Franklin, 2021; Yang, Atterholt, *et al.*, 2022), fault-zone detection (Jousset *et al.*, 2018; Lindsey *et al.*, 2019; Atterholt, Zhan, and Yang, 2022; Yang, Zhan, *et al.*, 2022), and urban seismology (Lindsey, Yuan, *et al.*, 2020; Wang *et al.*, 2021; Zhu *et al.*, 2021). It has also been applied broadly outside seismology, such as volcanology (Nishimura *et al.*, 2021; Jousset *et al.*, 2022), oceanography (Sladen *et al.*, 2019; Williams *et al.*, 2019, 2022; Xiao *et al.*, 2022; Lin *et al.*, 2024), glaciology (Walter *et al.*, 2020; Hudson *et al.*, 2021), marine biology (Bouffaut *et al.*, 2022; Landrø *et al.*, 2022; Rørstadbotnen *et al.*, 2023; Wilcock *et al.*, 2023), and meteorology (Zhu and Stensrud, 2019; Hong *et al.*, 2024).

DAS produces data gathers with regular spacing, similar to exploration seismic data. Hence, one may process DAS data with software for exploration data, such as Seismic Unix (Cohen and Stockwell, 2008) and Madagascar (see Data and Resources). However, compared to conventional seismic arrays in earthquake seismology, DAS differs in several key aspects, especially the voluminous data and uniaxial measurement of strain or strain rate (Zhan, 2019; Li, 2021; Lindsey and Martin, 2021). The noise composition of DAS tends to be more complex due to its different self-noise, common-mode noise, and traffic noise for often along-road fibers (Bakku, 2015; Costa *et al.*, 2019; Zhirnov *et al.*, 2019; Lindsey, Rademacher, and Ajo-Franklin, 2020). These differences often require different processing techniques from those for conventional seismometers, making it challenging for researchers newly using DAS data.

Inspired by the success of ObsPy for conventional seismic data processing (Beyreuther *et al.*, 2010), we believe that a new Python processing package specifically designed for DAS data could facilitate the development of DAS seismology. We notice that an ongoing project, called DASCore, is developing a

1. Laboratory of Seismology and Physics of Earth's Interior, School of Earth and Space Sciences, University of Science and Technology of China, Hefei, China, https://orcid.org/0000-0003-3080-1426 (MH); https://orcid.org/0000-0003-4405-8872 (ZL);
2. Mengcheng National Geophysical Observatory, University of Science and Technology of China, Mengcheng, China

*Corresponding author: zefengli@ustc.edu.cn

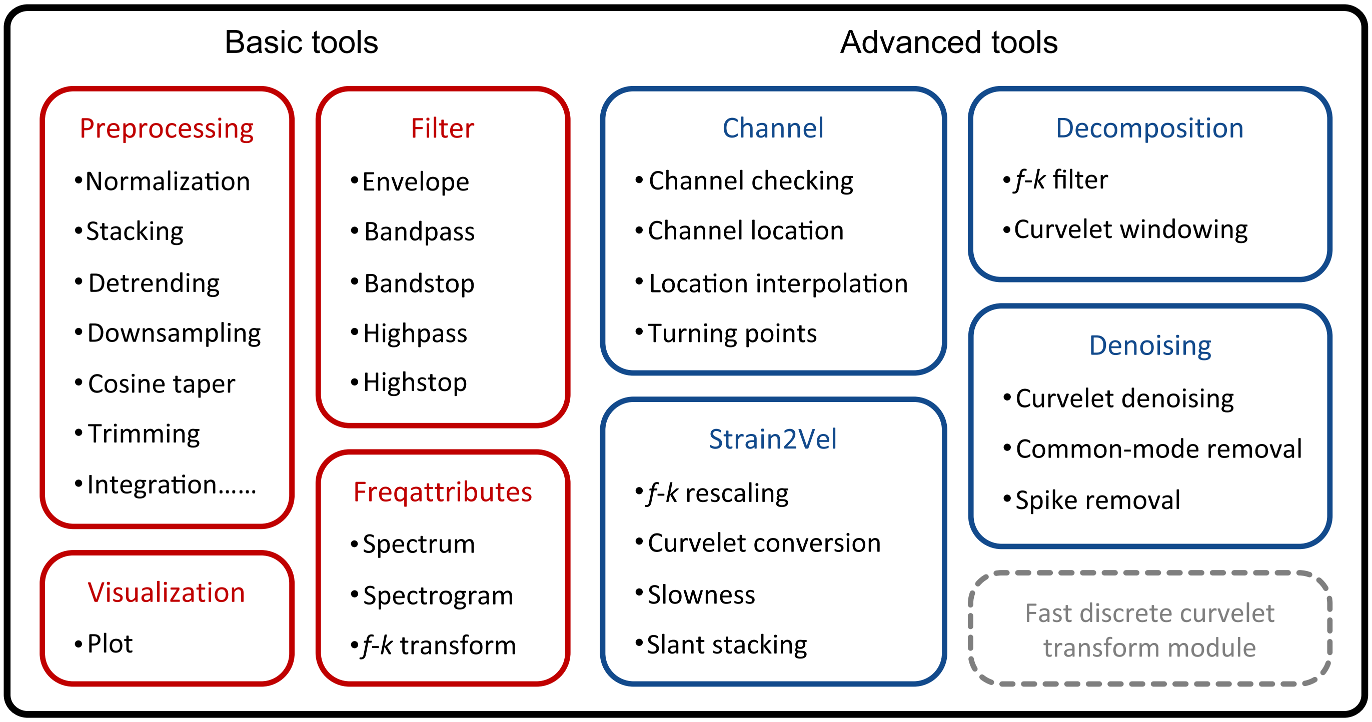


Figure 1. Main structure of DASPy toolbox. Each block indicates a module composed of multiple user-facing functions. The modules for basic tools are shown in red boxes, and modules for advanced tools are shown in blue boxes. The module within the gray dotted box is specifically built for discrete fast curvelet transforms. The color version of this figure is available only in the electronic edition.

Python package for reading and writing, visualization, and basic processing of DAS data (Chambers *et al.*, 2024). In this study, in addition to the functionalities offered by DASCore, we aim to provide a wider diversity of practical data processing tools dedicated for DAS applications. This new open-source Python package is named DASPy and comprises two primary components: a set of basic tools including modules for preprocessing, filtering, frequency attributes, and visualization; and another set of advanced tools including modules for channel analysis, waveform decomposition, denoising, and strain–velocity conversion (Fig. 1). As follows, we showcase the key functionalities using various publicly available datasets (Fig. 2) and ensure that the experiments can be easily replicated by readers.

Basic Tools

Classic processing techniques

Typical seismic data processing includes filtering, frequency attribute analysis, and certain preprocessing methods. We wrap these techniques for 2D DAS data, eliminating the need for iterating over channels. For example, the Python code below band-pass filters the data from the RAPID (A community test of distributed acoustic sensing on the ocean observatories initiative regional cabled array) dataset (Wilcock and Ocean Observatories Initiative, 2023; see Data and Resources; Fig. 2a) between 15 and 27 Hz and yields a spectrogram averaged over 100 channels and a frequency–wavenumber (*f-k*) spectrum (Fig. 3). This dataset was collected offshore central Oregon and recorded various signals including fin whale calls, northeast Pacific blue whale A and B calls, and ship noises (Wilcock *et al.*, 2023).

```
>>> from daspy.basic_tools.filter import bandpass
>>> from daspy.basic_tools.freqattributes import spectrogram, fk_transform
>>>
>>> data_filtered = bandpass(data, fs, 15, 27, detrend=True, taper=0.04)
>>> spec, f1, t = spectrogram(data[4880-50:4880+50], fs=fs,
                              nperseg=256, noverlap=246,
                              nfft=1024, detrend=True)
>>> fk, f2, k = fk_transform(data, dx, fs)
```

Visualization

The function, plot, can be used to visualize 2D DAS data. It offers a number of optional parameters to accommodate the users' requirements for plotting a variety of data types, such as waveforms, spectra, spectrograms, and *f-k* spectra. Below is the Python code for visualizing the data in the previous example: unfiltered and filtered data, the spectrogram, and the *f-k* spectrum (Fig. 3). The band-pass filtered waveform reveals high-frequency fin whale calls, with amplitudes approximately four to five orders of magnitude lower than the ocean wave signals (Fig. 3b). The spectrogram demonstrates the sequential production of high- and low-frequency calls by the fin whale (Fig. 3c). The *f-k* spectrum reveals an apparent velocity of this acoustic signal exceeding 1400 m/s along the axial direction of the optical cable (Fig. 3d).

```
>>> from daspy.basic_tools.visualization import plot
>>>
>>> fig, ax = plt.subplots(4, 1, figsize=(7,8))
>>> plot(data, dx=dx, fs=fs, ax=ax[0], transpose=True,
       x0=xmin*dx, xlabel=False, colorbar_label='Strain')
>>> plot(data_filtered, dx=dx, fs=fs, ax=ax[1],
         transpose=True, x0=xmin*dx, xlabel=False,
         colorbar_label='Strain')
>>>plot(Zxx, fs=fs, obj='spectrogram', ax=ax[2], f=f1, t=t,
         vmin=2e-8, vmax=3e-6, ylim=[0, 40])
>>> plot(fk, obj='fk', ax=ax[3], f=f2, k=k, xlim=[-0.025,
         0.025], ylim=[0, 40], vmin=0.05, vmax=0.2)
>>> plt.tight_layout(pad=0.5)
>>> plt.savefig('figure3.pdf', dpi=800)
```

Advanced Tools

Channel analysis

DAS channels have equidistant spacing, but the location of each channel is often unknown and requires tap tests. Besides, the linearity and ground coupling of the fibers often need to be taken care of. We develop three functions for channel location and quality analysis: channel location interpolation, turning point detection, and low-quality channel checking.

Channel location interpolation for DAS is calculated using two types of inputs: points with known channel numbers, and optional fiber spatial track points without channel numbers. Points with known-channel numbers are typically acquired through tap tests and are often sparse. The spatial fiber track points are used to constrain the array geometry. They are optional but dense track points are particularly useful for accurate location interpolation. Figure 2a shows examples of the two DAS arrays of the RAPID dataset (Wilcock and Ocean Observatories Initiative, 2023). In DASPy, we implemented the interpolation method used by the RAPID team, in which interpolation is performed after the coordinates are projected to the Universal Transverse Mercator co-ordinate system.

The turning point detection function determines the points where the fiber strike varies noticeably based on the given channel co-ordinates or based on waveform coherency across neighboring channels. The application of the co-ordinate-based detection function to Brady's geothermal field DAS array (University of Wisconsin, 2016a; see Data and Resources; Fig. 2c) produces results consistent with those of Piana Agostinetti *et al.* (2022). As cross-channel waveform coherency is not only affected by the fiber strike angle, but also controlled by other factors including the quality of the backscattered light, coupling conditions and small-scale scatterers at different locations, its results could be less stable than those of co-ordinate-based computations assuming the coordinates are accurate. However, when the coordinates are unavailable or inaccurate, inference from cross-channel waveform coherency could be an alternative.

The channel quality checking function detects segments with obvious poor coupling (e.g., zip-tied loops of telecommunication cables) by identifying outliers of waveform energy along the fiber. It fits the waveform energy (the square of the amplitudes) variations with channels by a high-order polynomial and removes the fitted polynomial from the data. A threshold of four times of standard deviation below the median is set to identify the outliers. We assume that poor coupling tends to be spatially continuous. Hence, isolated normal values among a group zof outliers would be identified as bad channels and vice versa. Using this function, we assess the channel quality of Ridgecrest DAS (Fig. 2d) with 15 s of traffic noise (Atterholt, 2021; see Data and Resources, Fig. 4). Our waveform-based detection results are generally consistent with the hand-picked results of Atterholt, Zhan, *et al.* (2022; Fig. 4b–f), except for the initial segment that was identified from a priori knowledge during field installation. It is noteworthy that the spikes (Fig. 4a) do not significantly influence the low-quality channel detection because we use a robust fit for the waveform energy (abnormal points are excluded from fitting).

Data denoising

As aforementioned, DAS data are often mixed with complex types of noise. DASPy integrates functions for the removal of typical noise types, including spike noise (Bakku, 2015), common-mode noise (Lindsey, Rademacher, and Ajo-Franklin, 2020), stochastic noise (Costa *et al.*, 2019), and coherent noise. DASPy constructs a denoising module that incorporates three methods that take advantage of different noise properties.

Spikes are unusually large amplitudes (Fig. 5a) and could be caused by laser frequency drift or laser noise (Zhirnov *et al.*, 2019). The spike removal function first applies the across-channel median filter and then the across-time median

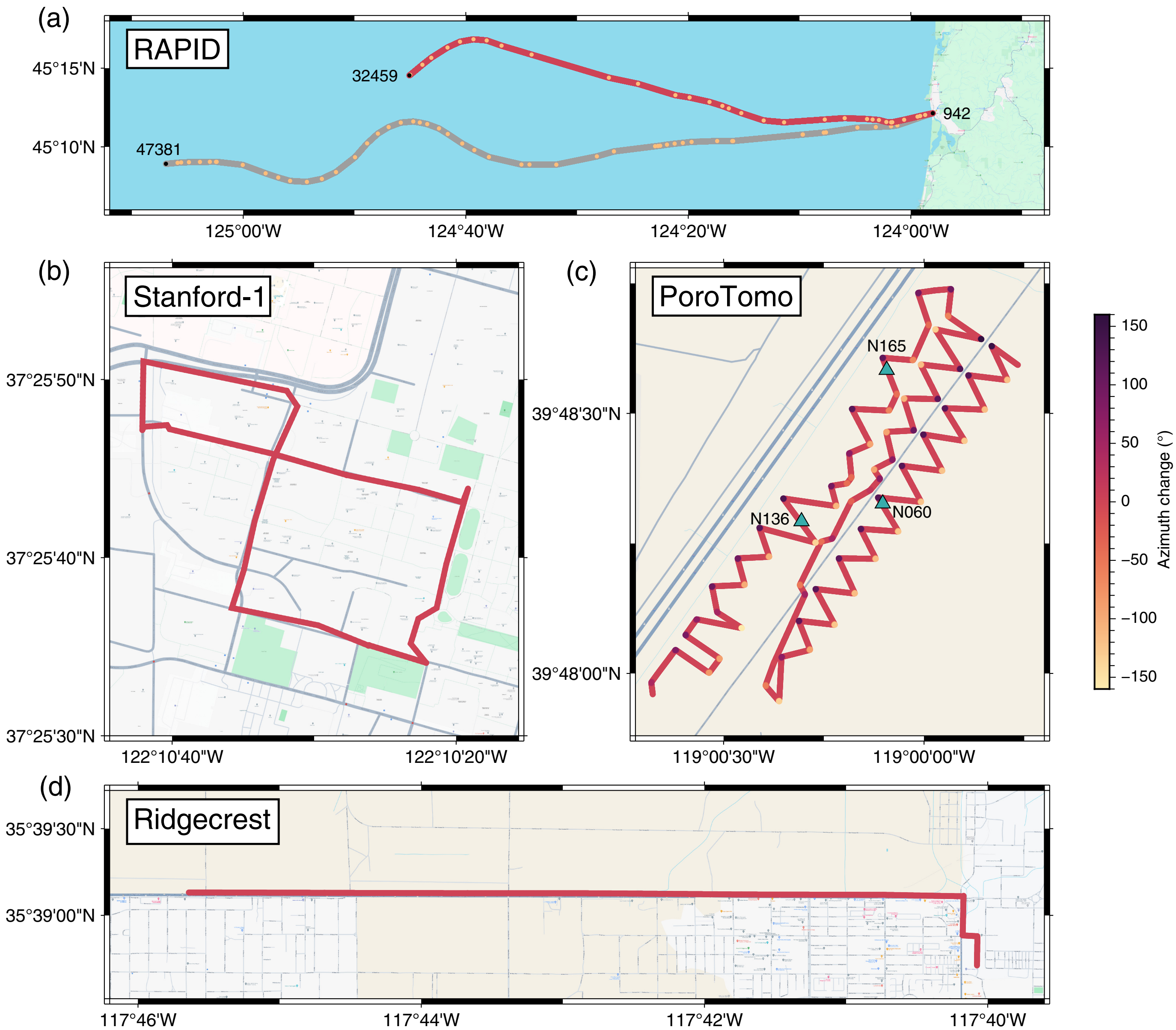


Figure 2. Geometry of the distributed acoustic sensing (DAS) arrays which data we used for testing. (a) RAPID DAS arrays that land at Pacific City, Oregon (Wilcock and Ocean Observatories Initiative, 2023). The red line indicates the array that we utilized for our test (referring to the north cable here), which is the same for panels (b) and (d). The gray line indicates the south cable, which data are not used. The black dots represent three points along the cable with known coordinates and channel numbers, whereas the orange dots represent those with known coordinates but unknown channel number. (b) Stanford campus array in California (Biondi *et al.*, 2017; Martin *et al.*, 2017). (c) Brady's geothermal field DAS array (University of Wisconsin, 2016b) and three collocated geophone stations (University of Wisconsin, 2016c) in Nevada. The color of the DAS cable indicates the azimuth change of the cable before and after the corresponding channel. (d) DAS arrays started after the 2019 M_w 7.1 Ridgecrest earthquake, California (Li *et al.*, 2021; Atterholt, Zhan, *et al.*, 2022). The color version of this figure is available only in the electronic edition.

filter to generate a median map from the absolute amplitudes. Points with amplitudes exceeding a predefined threshold of the median map are identified as spikes. All spikes are subsequently substituted with interpolated values from adjacent channels. The spike removal function is validated using an earthquake waveform recorded by the Stanford-1 DAS array (Figs. 2b, 5a,b; Biondi *et al.*, 2017; Martin *et al.*, 2017).

Common-mode noise, also known as in-phase noise, is generated by vibrations of the optoelectronic system and arises on all channels simultaneously (Fig. 5d). DASPy employs spatial averaging of waveforms to obtain common-mode noise. Subsequently, we compute the correlation coefficient with the channel record and the common-mode noise, multiply the common-mode noise by the coefficient, and subtract it from the channel record. We evaluate the common-mode noise removal algorithm using a segment of offshore channels

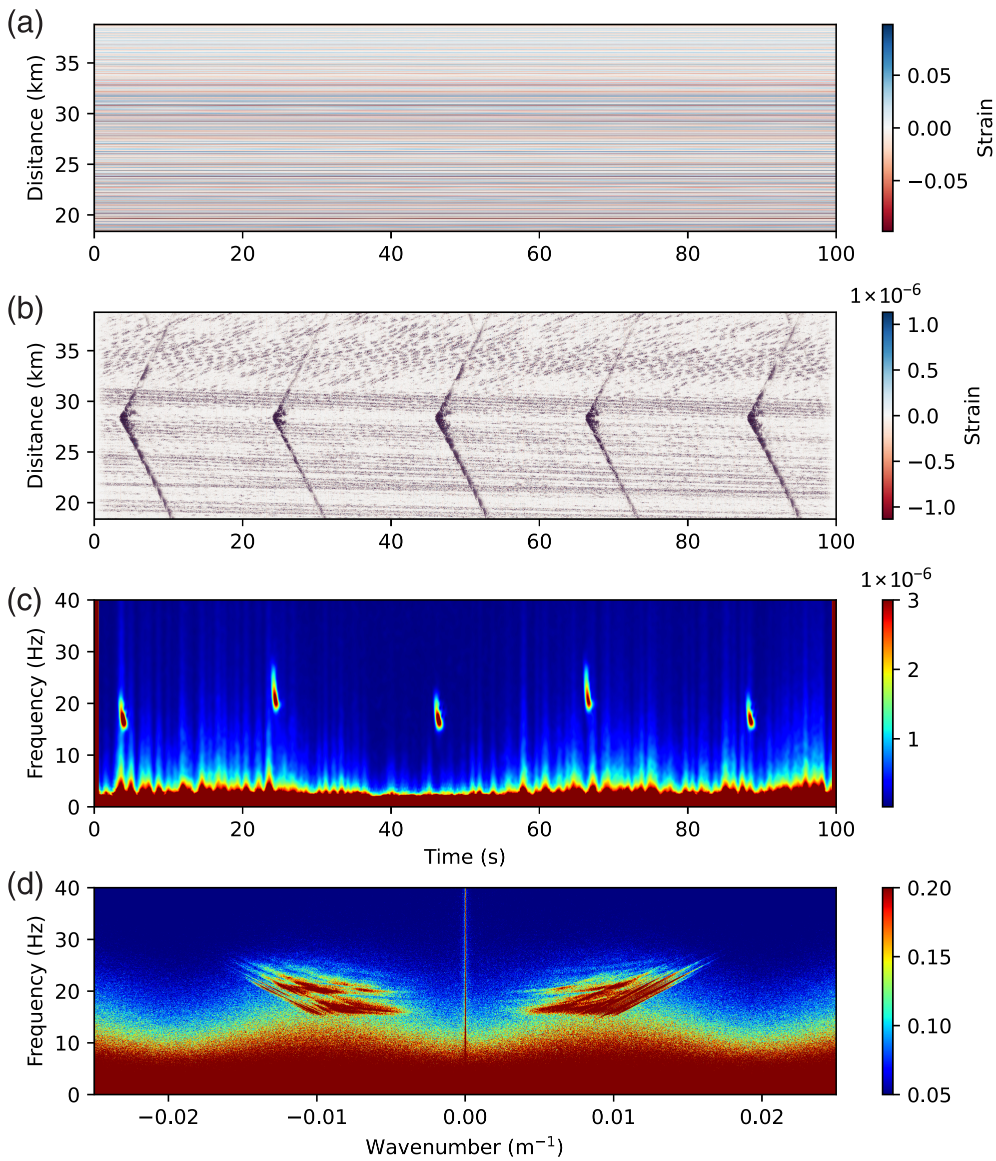


Figure 3. Demonstration of signal processing and visualization. (a) Original strain recording for 100 s beginning on 4 November 2021 01:59:02 UT, recorded by the Optasense interrogator channel 9000–19,000 on north ocean-bottom cable from RAPID dataset (Wilcock and Ocean Observatories Initiative, 2023). (b) Filter to 15–27 Hz, following Wilcock *et al.* (2023). (c) Spectrogram averaged over 100 channels. (d) Frequency–wavenumber (*f-k*) spectrum calculated from 2D fast Fourier transform. The color version of this figure is available only in the electronic edition.

represent the object position, scale, and angle. The curvelet denoising function uses a silent DAS recording to estimate stochastic noise. After FDCT, the amplitude of the curvelet coefficients is used as an empirical threshold. By default, DASPy employs a soft threshold to remove stochastic noise in the curvelet domain. We apply curvelet denoising to the spike-removed waveform of Stanford-1 DAS (Biondi *et al.*, 2017; Martin *et al.*, 2017; Figs. 2b, 5b), resulting in a notable reduction in stochastic noise before the arrival of *P* waves (Fig. 5c).

Coherent noise can be defined as any coherent signal that is not of interest. For example, for studies on an earthquake, a traffic signal is coherent noise; for studies on traffic footprints, an earthquake signal is coherent noise. Coherent noise can be removed by applying velocity screening in either the curvelet transform or the *f-k* transform. In this case, coherent noise removal is treated as wavefield decomposition based on apparent velocity, which will be elaborated upon in the subsequent section.

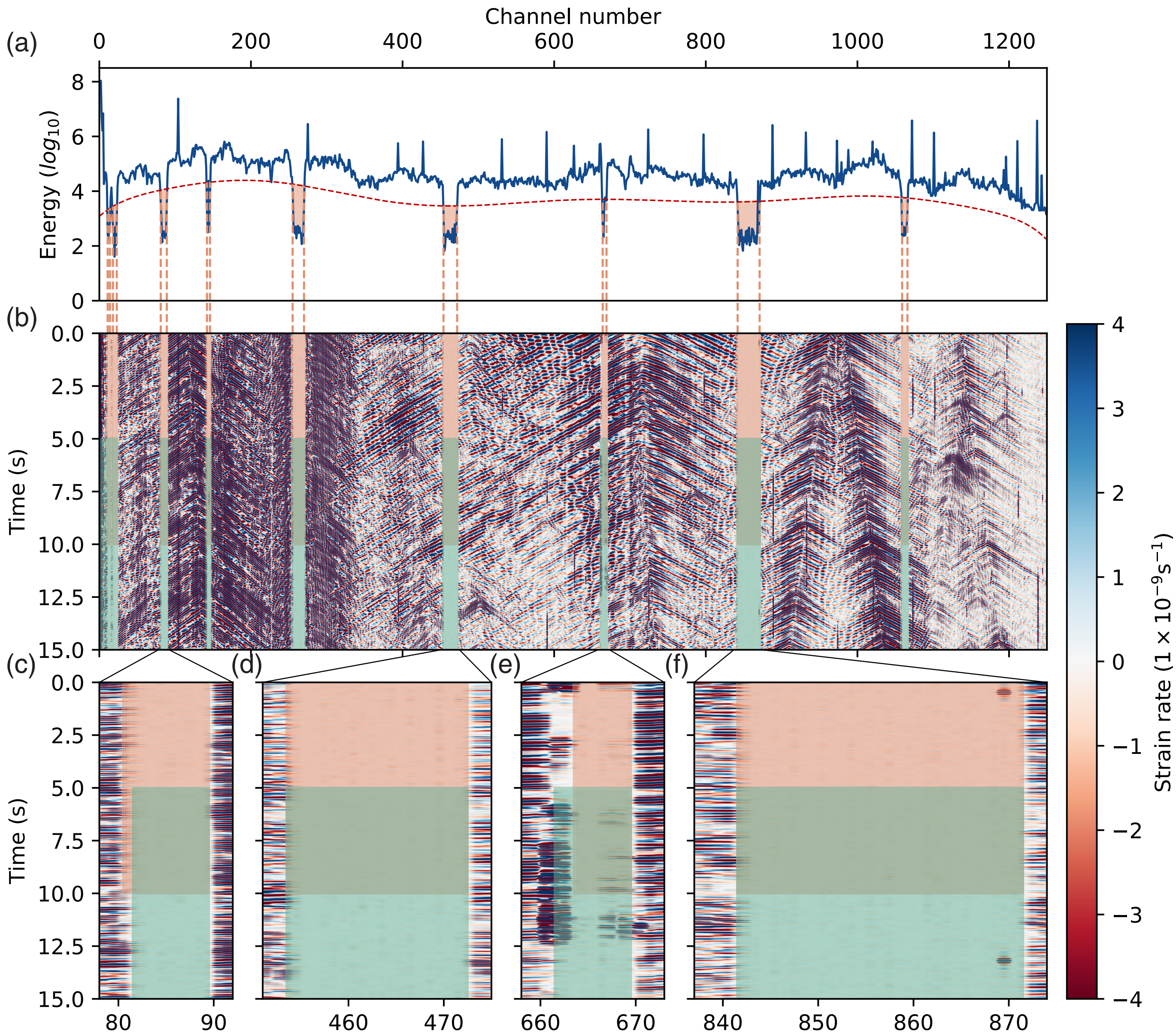
Wavefield decomposition

Image processing techniques, such as the 2D fast Fourier transform (e.g., *f-k* transform in DAS data processing) and FDCT, have been widely used in the decomposition of 2D DAS wavefields, such as the separation of seismic signals and traffic noise and the separation of direct seismic waves and locally scattered seismic waves (Atterholt, Zhan, *et al.*, 2022; Williams *et al.*, 2022). DASPy integrates the *f-k* filtering and curvelet windowing techniques in the decomposition module.

Each point within the *f-k* domain corresponds to a specific apparent velocity. In wavefield decomposition, the *f-k* filter method employs a velocity threshold for separation, followed by an inverse transform to produce low-speed and high-speed

of the RAPID dataset (Wilcock and Ocean Observatories Initiative, 2023; Fig. 2a). The processing effectively mitigates the common-mode noise (Fig. 5d,e).

The inherent stochastic noise in DAS data is primarily caused by instrumental deficiencies such as sampling error and phase noise (Costa *et al.*, 2019). The fast discrete curvelet transform (FDCT; Candès and Donoho, 2004; Candès *et al.*, 2006) is used to obtain an effective nonadaptive sparse representation of the regular-spaced DAS seismic data and remove stochastic noise (Atterholt, Zhan, *et al.*, 2022). The basi functions of curvelet transform are defined as polar wedges in the *f-k* domain and



waveforms. Analogously, the curvelet basis utilized by the curvelet transform is wedges on the *f-k* domain, with specific velocity ranges. The curvelet window technique separates the curvelet coefficients of the curvelet basis with different velocities. Therefore, the effects of these two techniques are nearly identical, which can be clearly determined in the *f-k* domain of the separated waveforms. Both wavefield decomposition techniques are evaluated on stripping traffic noise from seismic waveform from the Ridgecrest DAS array (Li *et al.*, 2021; Fig. 2d). The results show that both techniques effectively enhance the signal-to-noise ratio without significant difference (Fig. 6).

Figure 4. Bad channel detection of the DAS array near Ridgecrest, California. (a) Energy curve (blue line) and thresholds (red dotted line) for bad channel detection. (b) DAS recording of 15 s traffic noise (Atterholt, 2021) used for bad channel detection. Orange areas indicate bad channels detected by our function, whereas green areas are bad channels picked by Atterholt, Zhan, *et al.* (2022). (c–f) Zoom-in plot of four parts of the DAS recording. (c) Channel 81 and (e) channels 662 and 663 are identified differently by our function (Atterholt, Zhan, *et al.*, 2022). The color version of this figure is available only in the electronic edition.

Conversion to ground motions

DAS measures strain or strain rate, in contrast to ground-motion velocity and displacement used in typical seismology studies. Strain and strain rate can be converted to particle velocity and acceleration by multiplying apparent phase velocity. The difficulty of such conversion lies in the accurate estimation of apparent phase velocity of every wiggle. DASPy integrates three methodologies for converting strain (strain rate) into ground-motion velocity: *f-k* rescaling (Lindsey, Rademacher, and

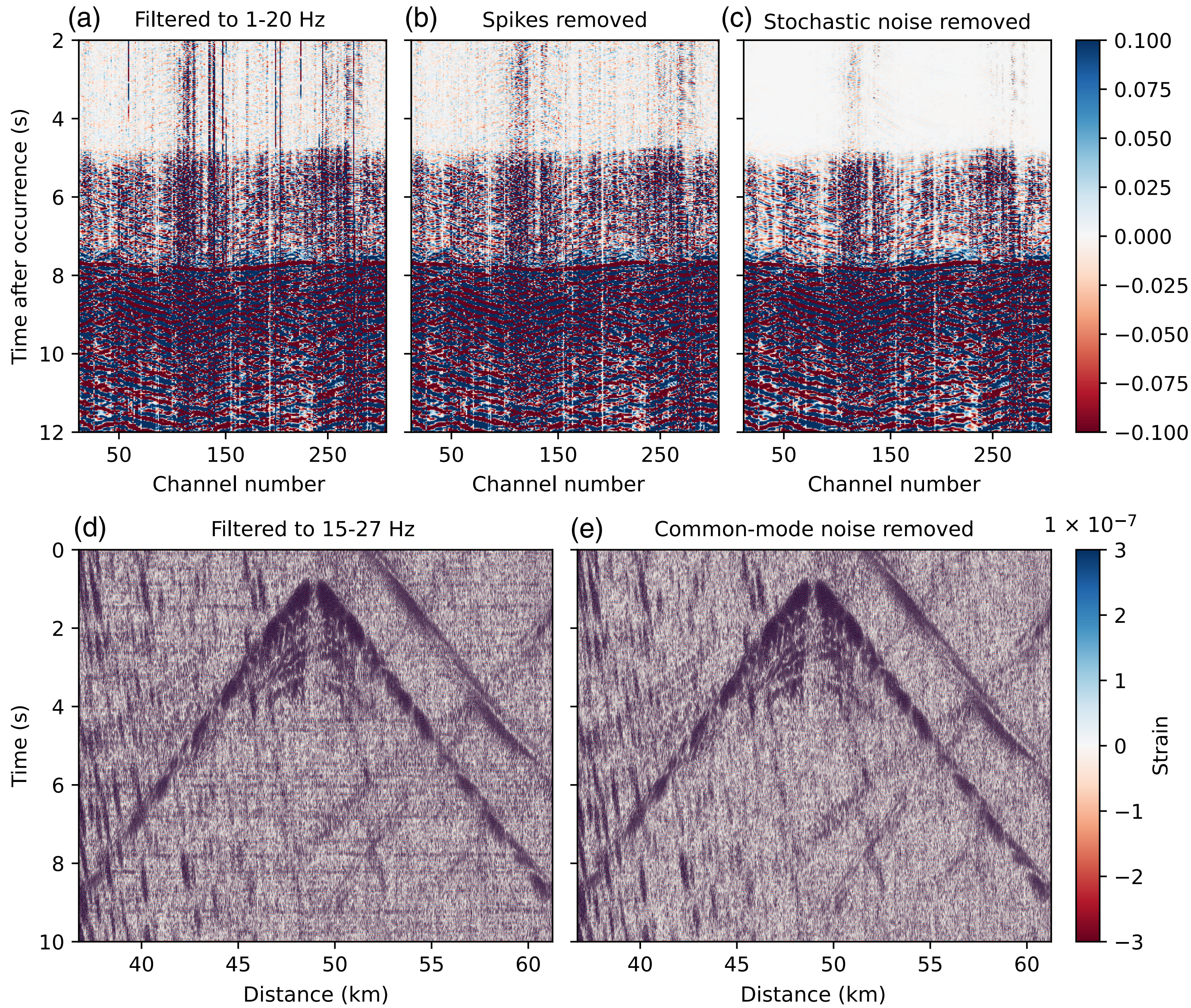
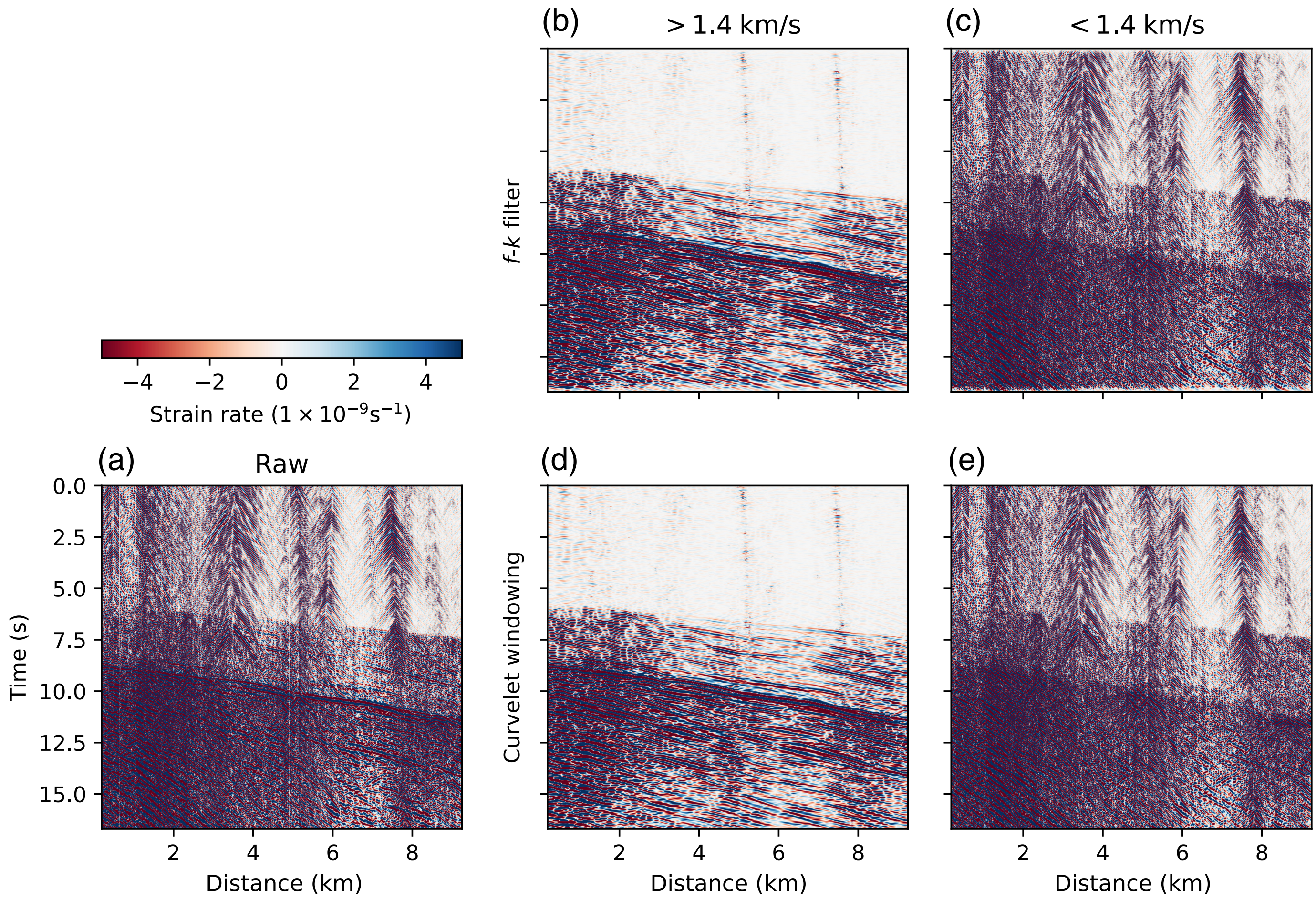


Figure 5. Cases of wavefield denoising. (a) Waveforms of an M_D 2.8 earthquake (see Data and Resources) recorded by Stanford-1 DAS array (Biondi *et al.*, 2017; Martin *et al.*, 2017). Bad channels are removed and band-pass filter to 1–20 Hz. (b) Waveforms with spikes removed based on panel (a). (c) Waveforms with stochastic noise removed by curvelet transform based on panel (b). (d) Strain recording filtered to 15–27 Hz for 10 s beginning on 4 November 2021 01:59:22 UT, recorded by the Optasense interrogator on north ocean-bottom cable from RAPID dataset (Wilcock and Ocean Observatories Initiative, 2023). (e) Waveforms with common-mode noise removed based on panel (d). The color version of this figure is available only in the electronic edition.

Ajo-Franklin, 2020), curvelet transform (Yang, Atterholt, *et al.*, 2022), and time-domain slowness determination (Lior *et al.*, 2021). The *f-k* rescaling method is implemented by multiplying each point in the *f-k* domain by its corresponding apparent velocity (slop in the *f-k* domain). Similarly, the basisfunctions of the curvelet transform, which are defined in the *f-k* domain, also correspond to varying velocity ranges. The curvelet transform conversion method multiplies each curvelet coefficient by the median velocity of its basis function. The coefficients of the largest scale basis functions, which represents waves with all velocity ($-\infty$ to $+\infty$), is set to zero. The time-domain slowness determination method obtains the apparent velocity at each time step by searching for the maximal semblance.

These three methods are tested using an M_L 4.3 earthquake recorded by a collocated DAS and seismometer array in the Brady Hot Springs (University of Wisconsin, 2016bc; see Data and Resources; Fig. 3c), following Wang *et al.* (2018). We define a nodal geophone and a DAS channel for which distance is less than 5 m as a geophone-channel pair. Among 238 geophones and 8621 DAS channels, we match a total of 344 geophone-channel pairs. For each geophone-channel pair, we find the corresponding linear DAS segment (Fig. 2c) and rotate the three-component geophone recording



to the axial fiber direction. The original DAS strain-rate recordings are integrated to strain in the time domain, and converted to velocity using *f-k* rescaling, curvelet transform, and time-domain slowness determination methods (Fig. 7). We correct the DAS data timing (−1.048 s) using the Global Positioning System timing of nodal seismometers, and cross correlate the waveforms of each geophone-channel pair with time shift less than ±0.01 s. All waveforms are bandpass filtered to 1–5 Hz.

We evaluate the cross-correlation coefficient between the converted DAS velocity and the rotated geophone velocity. For all 344 geophone-channel pairs, 104, 71, and 0 pairs obtain cross-correlation coefficients greater than 0.7 after *f-k* rescaling, curvelet transformation, and time-domain slowness determination, respectively. For this particular case, the curvelet transform and the time-domain slowness determination have limitations. Most linear segments consist of about 100 channels, which is not quite enough for curvelet transform at larger scales. The largest scale curvelet coefficients, which are set to zero, miss more details, resulting in smaller amplitudes of the converted waveforms (Fig. 7). As for time-domain slowness determination methods, the assumption of monochromatic wavefields makes it difficult to recover the complex shallow surface scattered waves and earthquake coda waves.

Figure 6. An example of wavefield decomposition. (a) Waveforms of an M_L 2.6 earthquake (see Data and Resources) recorded by Ridgecrest DAS array (Li *et al.*, 2021), with spikes removed. (b) Waveforms with an *f-k* filter to retain energy with an apparent velocity >1.4 km/s (cosine tapered from 1.2–1.6 km/s). (c) Waveforms with an *f-k* filter to retain energy with an apparent velocity <1.4 km/s (cosine tapered from 1.2 to 1.6 km/s). (d) Waveforms with curvelet windowing to retain energy with an apparent velocity >1.4 km/s. (e) Waveforms with curvelet windowing to retain energy with an apparent velocity <1.4 km/s. The color version of this figure is available only in the electronic edition.

Discussion and Conclusions

DASPy aims to offer a user-friendly, integrated Python toolkit that facilitates the analysis and processing of DAS data. Overall, the toolkit includes "basic tools" of preprocessing, filtering, spectrum analysis, and visualization techniques and "advanced tools" of channel attribute analysis, noise removal, wavefield decomposition, and strain–velocity conversion.

DASPy can read and write a variety of DAS file formats, including .h5, .segy/.sgy, .tdms, and .pkl (used for storing daspy. Section instances). These formats are often required by other open-source software. For example, PhaseNet-DAS (Zhu *et al.*, 2023) take input in .h5 or .segy format. DASPy also supports reading DAS-Research Coordination Network (RCN) format

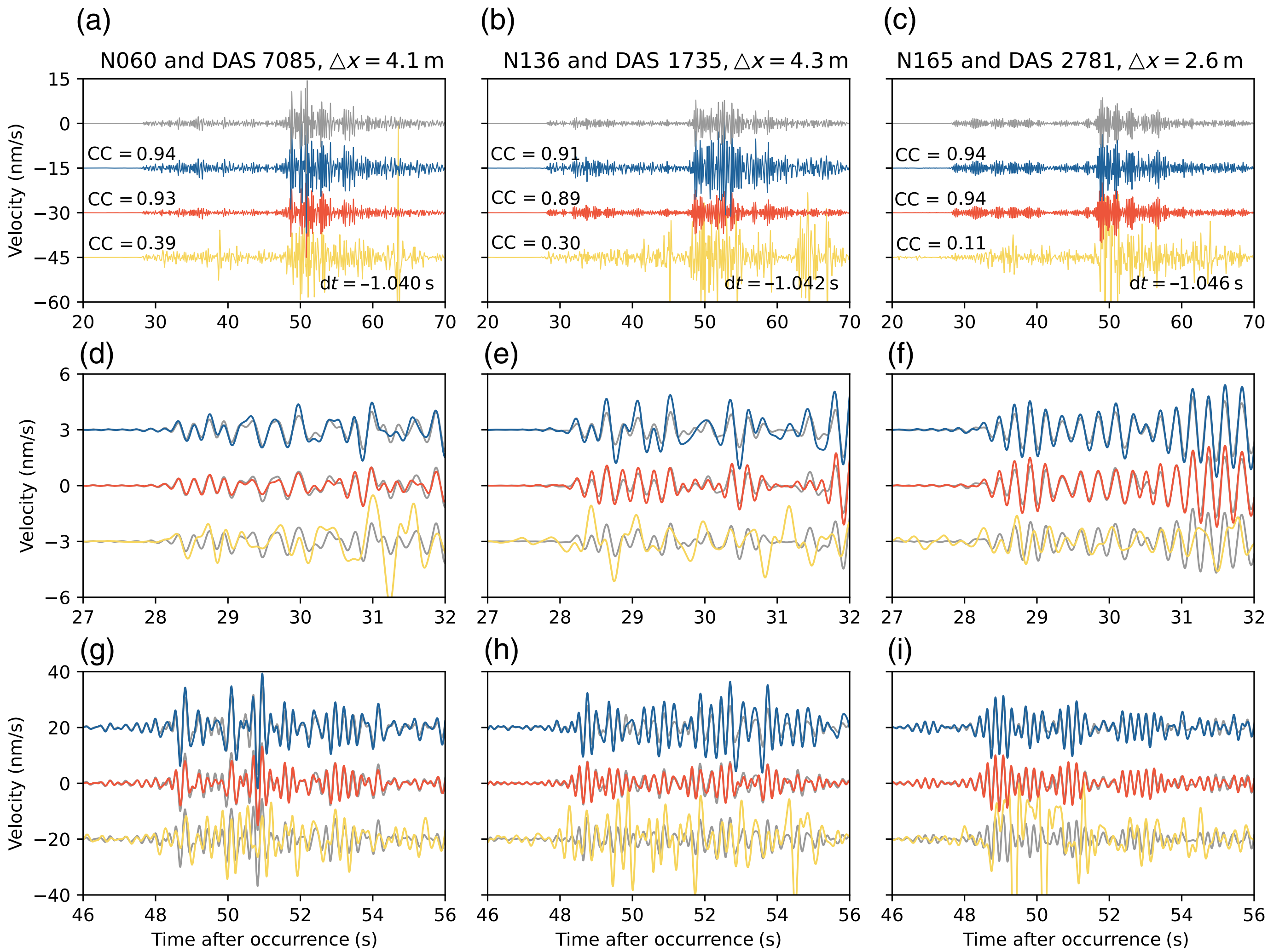


Figure 7. Conversion from strain to velocity by three methods of an M_L 4.3 Hawthorne earthquake (see Data and Resources) recorded by Brady's geothermal field DAS array. (a–c) Rotated geophone velocity (gray), and velocity converted from integrated DAS strain by *f*-*k* rescaling (blue), curvelet transform (red), and time-domain slowness determination (yellow), same as following. All waveforms are band-pass filtered to 1–5 Hz. (d–f) Zoom-in window for *P* arrival of panels (a–c). (g–i) Zoom-in window for *S* arrival of panels (a–c). CC, cross-correlation. The color version of this figure is available only in the electronic edition.

as a daspy. Section instance that inherits the attributes from source DAS-RCN files (Lai *et al.*, 2024). One may note that DASCore supports more reading formats than DASPy. In addition, ObsPy provides IO support for almost all traditional seismological formats, such as Seismic Analysis Code and MiniSEED. Therefore, we provide methods (Section.to_obspy_stream, Section.to_dascore_patch, Section.from_obspy_stream, and Section.from_dascore_patch) for mutual transformation between daspy. Section instances and ObsPy's Stream instances (Beyreuther *et al.*, 2010) and DASCore's Patch instances (Chambers *et al.*, 2024). These conversion functions allow smooth data flow between ObsPy, DASCore, and DASPy.

DASPy operates in the form of functions, which are designed to accommodate as many optional parameters as possible, and with sensible default values. All functions within DASPy are implemented as methods of the daspy. Section class. This approach is advantageous in that data attributes are stored within the class and avoid the need for manual entry. Calling functions and using methods of daspy. Section class are functionally equivalent, providing flexibility to suit users' needs.

Moreover, DASPy is currently programmed in pure Python for ease of use and modification but in some cases computational efficiency is compromised. Consequently, processing continuous data with a large number of channels and/or a high-sampling rate could take a long time. As an example, downsampling a 30 s waveform recorded at 1000 Hz by a 10,000-channel DAS array takes 12.08 s. Therefore, we suggest that users consider implementing central processing unit parallelization when undertaking large tasks. Future development of DASPy could include exploring the potential of shared libraries to replace computationally intensive functions.

With aforementioned designs, DASPy can be easily incorporated into the data up- and downstream tasks. The following is an example code snippet that combines DASPy and ObsPy

```
>>> from daspy import read
>>> from obspy.signal.trigger import classic_sta_lta,
trigger_onset
>>>
>>> sec = (
>>>   read('raw_data.h5')
>>>   .spike_removal()
>>>   .downsampling(xint=10, tint=10)
>>>   .fk_filter(fmin=1, fmax=15, vmin=2000)
>>> )
>>> sec.plot()
>>> sec.save('preprocessed_data.h5')
>>> onsets = []
>>> for d in sec.data:
>>>   cft = classic_sta_lta(d, nsta=int(0.5*sec.fs),
nlta=int(5*sec.fs))
>>>   onsets.append(trigger_onset(cft, thres1=5,
thres2=4))
```

(Beyreuther *et al.*, 2010), for a typical task phase picking in earthquake monitoring:

The code reads in DAS data into an instance of daspy. Section, removes spike noise, performs a 10-fold downsampling in both distance dimension (stacking every 10 channels into one) and time dimension (after an automatic low-pass filter), separates signal with frequency of 1–15 Hz and apparent velocity less than 2000 m/s using *f-k* filter. Subsequently, the preprocessed data are visualized, saved, and fed into ObsPy to compute the short-term average/long-term average (Allen, 1982) and to generate triggered picks.

As shown in the previous example, we envision that users can take advantage of DASPy to develop advanced packages developing new functions and/or modules (such as earthquake monitoring, ambient noise imaging, and traffic detection algorithms). We welcome users to contribute to the improvement and expansion of the DASPy project. In addition, to foster a community of compatible packages, we add instructions for potential developers about how to contribute to the DASPy platform (see Data and Resources). Developers are recommended to fork the DASPy repository on GitHub (see Data and Resources) and submit their modifications and additions through pull requests.

Data and Resources

The RAPID dataset is openly available at http://piweb.ooirsn.uw.edu/das/. The traffic signals recorded by the Ridgecrest distributed acoustic sensing (DAS) can be downloaded from https://data.caltech.edu/records/31emd-wmv98. The Stanford DAS-1 dataset from PubDAS is accessible via the link https://app.globus.org/file-manager?origin_id=706e304c-5def-11ec-9b5c-f9dfb1abb183&origin_path=%2F. The earthquake waveforms recorded by Brady's Geothermal Field DAS and seismometer array are available at https://gdr.openei.org/submissions/848 and https://gdr.openei.org/submissions/846. The DASPy Python package is open-source and available at https://github.com/HMZ-03/DASPy/. We include tutorials in both English and Chinese (https://daspy-tutorial.readthedocs.io/en/latest/, https://daspy-tutorial-cn.readthedocs.io/zh-cn/latest/) and a Jupyter notebook for quick use (https://github.com/HMZ-03/DASPy/blob/main/document/example.ipynb). The software package Madagascar is available at http://www.reproducibility.org. The geothermal data repository (GDR) is available at https://gdr.openei.org/submissions/829. The Caltech data library is available at https://data.caltech.edu/records/1955. Instructions for potential developers about how to contribute to the DASPy platform is available at https://github.com/HMZ-03/DASPy/blob/main/CONTRIBUTING.md. The information about an M_D 2.8 earthquake is available at https://earthquake.usgs.gov/earthquakes/eventpage/nc73940346/executive, an M_L 2.6 earthquake is available at https://earthquake.usgs.gov/earthquakes/eventpage/ci38972328/executive, and an M_L 4.3 is available at https://earthquake.usgs.gov/earthquakes/eventpage/nn00536374. All websites were last accessed in July 2024.

Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest recorded.

Acknowledgments

The authors appreciate the constructive comments from Editor-in-Chief Allison Bent, the anonymous Associate Editor, and two anonymous reviewers. The authors are grateful to Bin Luo for publishing code for reading multiple distributed acoustic sensing (DAS) data formats (https://github.com/RobbinLuo/das-toolkit/blob/main/DasTools/DasPrep.py, last accessed July 2024) and to Zhongwen Zhan for the earthquake data from the Ridgecrest DAS array (Fig. 6). The authors thank Feng Cheng, Xiangfang Zeng, and Jiaxuan Li for their helpful discussions on the project and Yanlan Hu and Jun Zhu for their support on coding. This research is supported by the National Key R&D Program of China (Number 2022YFC3005602) and Anhui Provincial Key R&D Program (Number 2022m07020002).

References

Ajo-Franklin, J. B., S. Dou, N. J. Lindsey, I. Monga, C. Tracy, M. Robertson, V. Rodriguez Tribaldos, C. Ulrich, B. Freifeld, T. Daley, *et al.* (2019). Distributed acoustic sensing using dark fiber for near-surface characterization and broadband seismic event detection, *Sci. Rep.* **9,** no. 1, 1–14, doi: 10.1038/s41598-018-36675-8.

Allen, R. (1982). Automatic phase pickers: Their present use and future prospects, *Bull. Seismol. Soc. Am.* **72,** no. 6B, 225–242.

Atterholt, J. (2021). Earthquake waveforms from Curvelet-Denoising paper (Data Supplement) (1.0) [Data set], *CaltechDATA*, doi: 10.22002/D1.1955.

Atterholt, J., Z. Zhan, Z. Shen, and Z. Li (2022). A unified wavefield-partitioning approach for distributed acoustic sensing, *Geophys. J. Int.* **228,** no. 2, 1410–1418, doi: 10.1093/gji/ggab407.

Atterholt, J., Z. Zhan, and Y. Yang (2022). Fault zone imaging with distributed acoustic sensing: Body-to-surface wave scattering, *J. Geophys. Res.* **127,** no. 6, 1–17, doi: 10.1029/2022jb024329.

Bakku, S. K. (2015). Fracture characterization from seismic measurements in a borehole, *Ph.D. Thesis*, Massachusetts Institute of Technology.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). ObsPy: A python toolbox for seismology, *Seismol. Res. Lett.* **81,** no. 3, 530–533, doi: 10.1785/gssrl.81.3.530.

Biondi, B., E. Martin, S. Cole, M. Karrenbach, and N. Lindsey (2017). Earthquakes analysis using data recorded by the Stanford DAS array, *SEG Technical Program Expanded Abstracts 2017*, 2752–2756, doi: 10.1190/segam2017-17745041.1.

Bouffaut, L., K. Taweesintananon, H. J. Kriesell, R. A. Rørstadbotnen, J. R. Potter, M. Landrø, S. E. Johansen, J. K. Brenne, A. Haukanes, O. Schjelderup, *et al.* (2022). Eavesdropping at the speed of light: Distributed acoustic sensing of baleen whales in the Arctic, *Front. Mar. Sci.* **9,** 901348, doi: 10.3389/fmars.2022.901348.

Candès, E., L. Demanet, D. Donoho, and L. Ying (2006). Fast discrete curvelet transforms, *Multiscale Model. Simul.* **5,** no. 3, 861–899, doi: 10.1137/05064182X.

Candès, E. J., and D. L. Donoho (2004). New tight frames of curvelets and optimal representations of objects with piecewise C2 singularities, *Commun. Pure Appl. Math.* **57,** no. 2, 0219–0266, doi: 10.1002/cpa.10116.

Chambers, D., G. Jin, A. Tourei, A. H. S. Issah, A. Lellouch, E. Martin, D. Zhu, A. Girard, S. Yuan, T. Cullison, *et al.* (2024). DASCore: A Python library for distributed fiber optic sensing, doi: 10.31223/X5B978.

Chen, X. (2023). Source parameter analysis using distributed acoustic sensing—An example with the PoroTomo array, *Geophys. J. Int.* **233,** 2208–2214.

Cheng, F., B. Chi, N. J. Lindsey, T. C. Dawe, and J. B. Ajo-Franklin (2021). Utilizing distributed acoustic sensing and ocean bottom fiber optic cables for submarine structural characterization, *Sci. Rep.* **11,** no. 1, 1–14, doi: 10.1038/s41598-021-84845-y.

Cohen, J. K., and J. W. Stockwell (2008). CWP/SU: Seismic Un*x: an open source software package for seismic research and processing, *Center for Wave Phenomena*, Colorado School of Mines, 40 pp.

Costa, L., H. F. Martins, S. Martín-López, M. R. Fernández-Ruiz, and M. González-Herráez (2019). Fully distributed optical fiber strain sensor with 10–12 $\varepsilon/\sqrt{\mathrm{Hz}}$ sensitivity, *J. Lightwave Technol.* **37,** no. 18, 4487–4495, doi: 10.1109/JLT.2019.2904560.

Dou, S., N. Lindsey, A. M. Wagner, T. M. Daley, B. Freifeld, M. Robertson, J. Peterson, C. Ulrich, E. R. Martin, and J. B. Ajo-Franklin (2017). Distributed acoustic sensing for seismic monitoring of the near surface: A traffic-noise interferometry case study, *Sci. Rep.* **7,** no. 1, 1–12, doi: 10.1038/s41598-017-11986-4.

Hong, H., B. Wang, G. Lu, X. Li, Q. Ge, A. Xie, Y. Wu, X. Qiu, and J. Chen (2024). Tracking lightning through 3D thunder source location with distributed acoustic sensing, *J. Geophys. Res.* **129,** no. 4, 1–13, doi: 10.1029/2023JD038882.

Hudson, T. S., A. F. Baird, J. M. Kendall, S. K. Kufner, A. M. Brisbourne, A. M. Smith, A. Butcher, A. Chalari, and A. Clarke (2021). Distributed acoustic sensing (DAS) for natural microseismicity studies: A case study from Antarctica, *J. Geophys. Res.* **126,** no. 7, 1–19, doi: 10.1029/2020JB021493.

Jousset, P., G. Currenti, B. Schwarz, A. Chalari, F. Tilmann, T. Reinsch, L. Zuccarello, E. Privitera, and C. M. Krawczyk (2022). Fibre optic distributed acoustic sensing of volcanic events, *Nat. Commun.* doi: 10.1038/s41467-022-29184-w.

Jousset, P., T. Reinsch, T. Ryberg, H. Blanck, A. Clarke, R. Aghayev, G. P. Hersir, J. Henninges, M. Weber, and C. M. Krawczyk (2018). Dynamic strain determination using fibre-optic cables allows imaging of seismological and structural features, *Nat. Commun.* **13,** no. 1, doi: 10.1038/s41467-022-29184-w.

Lai, V. H., K. M. Hodgkinson, R. W. Porritt, and R. Mellors (2024). Toward a metadata standard for distributed acoustic sensing (DAS) data collection, *Seismol. Res. Lett.* **95,** no. 3, 1986–1999, doi: 10.1785/0220230325.

Landrø, M., L. Bouffaut, H. J. Kriesell, J. R. Potter, R. A. Rørstadbotnen, K. Taweesintananon, S. E. Johansen, J. K. Brenne, A. Haukanes, O. Schjelderup, *et al.* (2022). Sensing whales, storms, ships and earthquakes using an Arctic fibre optic cable, *Sci. Rep.* **12,** no. 1, 1–10, doi: 10.1038/s41598-022-23606-x.

Li, J., T. Kim, N. Lapusta, E. Biondi, and Z. Zhan (2023). The break of earthquake asperities imaged by distributed acoustic sensing, *Nature* **620,** doi: 10.1038/s41586-023-06227-w.

Li, J., W. Zhu, E. Biondi, and Z. Zhan (2023). Earthquake focal mechanisms with distributed acoustic sensing, *Nat. Commun.* **14,** no. 1, 4181, doi: 10.1038/s41467-023-39639-3.

Li, Z. (2021). Recent advances in earthquake monitoring i: Ongoing revolution of seismic instrumentation, *Earthq. Sci.* **34,** no. 2, 177–188, doi: 10.29382/eqs-2021-0011.

Li, Z., and Z. Zhan (2018). Pushing the limit of earthquake detection with distributed acoustic sensing and template matching: A case study at the Brady geothermal field, *Geophys. J. Int.* **215,** no. 3, 1583–1593, doi: 10.1093/gji/ggy359.

Li, Z., Z. Shen, Y. Yang, E. Williams, X. Wang, and Z. Zhan (2021). Rapid response to the 2019 Ridgecrest earthquake with distributed acoustic sensing, *AGU Adv.* **2,** no. 2, doi: 10.1029/2021av000395.

Lin, J., S. Fang, R. He, Q. Tang, F. Qu, B. Wang, and W. Xu (2024). Monitoring ocean currents during the passage of Typhoon Muifa using optical-fiber distributed acoustic sensing, *Nat. Commun.* **15,** no. 1, doi: 10.1038/s41467-024-45412-x.

Lindsey, N. J., and E. R. Martin (2021). Fiber-optic seismology, *Annu. Rev. Earth Planet. Sci.* **49,** 309–338.

Lindsey, N. J., T. Craig Dawe, and J. B. Ajo-Franklin (2019). Illuminating seafloor faults and ocean dynamics with dark fiber distributed acoustic sensing, *Science* **366,** no. 6469, 1103–1107, doi: 10.1126/science.aay5881.

Lindsey, N. J., E. R. Martin, D. S. Dreger, B. Freifeld, S. Cole, S. R. James, B. L. Biondi, and J. B. Ajo-Franklin (2017). Fiber-optic network observations of earthquake wavefields, *Geophys. Res. Lett.* **44,** no. 23, 11,792–11,799, doi: 10.1002/2017GL075722.

Lindsey, N. J., H. Rademacher, and J. B. Ajo-Franklin (2020). On the broadband instrument response of fiber-optic DAS arrays, *J. Geophys. Res.* **125,** no. 2, 1–16, doi: 10.1029/2019JB018145.

Lindsey, N. J., S. Yuan, A. Lellouch, L. Gualtieri, T. Lecocq, and B. Biondi (2020). City-scale dark fiber DAS measurements of infrastructure use during the COVID-19 pandemic, *Geophys. Res. Lett.* **47,** no. 16, 1–8, doi: 10.1029/2020GL089931.

Lior, I., A. Sladen, D. Mercerat, J. P. Ampuero, D. Rivet, and S. Sambolian (2021). Strain to ground motion conversion of distributed acoustic sensing data for earthquake magnitude and stress drop determination, *Solid Earth* **12,** no. 6, 1421–1442, doi: 10.5194/se-12-1421-2021.

Luo, B., W. Trainor-Guitton, E. Bozdag, L. LaFlame, S. Cole, and M. Karrenbach (2021). Horizontally orthogonal distributed acoustic sensing array for earthquake- and ambient-noise-based multichannel analysis of surface waves, *Geophys. J. Int.* **222,** no. 3, 2147–2161, doi: 10.1093/GJI/GGAA293.

Martin, E. R., C. M. Castillo, S. Cole, P. S. Sawasdee, S. Yuan, R. Clapp, M. Karrenbach, and B. L. Biondi (2017). Seismic monitoring leveraging existing telecom infrastructure at the SDASA: Active, passive, and ambient-noise analysis, *The Leading Edge* **36,** 1025–1031, doi: 10.1190/tle36121025.1.

Nayak, A., and J. Ajo-Franklin (2021). Measurement of surface-wave phase-velocity dispersion on mixed inertial seismometer—Distributed acoustic sensing seismic noise cross-correlations, *Bull. Seismol. Soc. Am.* **111,** no. 6, 3432–3450, doi: 10.1785/0120210028.

Nayak, A., J. Ajo-Franklin, and Imperial Valley Dark Fiber Team (2021). Distributed acoustic sensing using dark fiber for array detection of regional earthquakes, *Seismol. Res. Lett.* **92,** no. 4, 2441–2452, doi: 10.1785/0220200416.

Nishimura, T., K. Emoto, H. Nakahara, S. Miura, M. Yamamoto, S. Sugimura, A. Ishikawa, and T. Kimura (2021). Source location of volcanic earthquakes and subsurface characterization using fiber-optic cable and distributed acoustic sensing system, *Sci. Rep.* **11,** no. 1, 1–12, doi: 10.1038/s41598-021-85621-8.

Piana Agostinetti, N., A. Villa, and G. Saccorotti (2022). Distributed acoustic sensing as a tool for subsurface mapping and seismic event monitoring: A proof of concept, *Solid Earth* **13,** no. 2, 449–468, doi: 10.5194/se-13-449-2022.

Rørstadbotnen, R. A., J. Eidsvik, L. Bouffaut, M. Landrø, J. Potter, K. Taweesintananon, S. Johansen, F. Storevik, J. Jacobsen, O. Schjelderup, *et al.* (2023). Simultaneous tracking of multiple whales using two fiber-optic cables in the Arctic, *Front. Mar. Sci.* **10,** 1–15, doi: 10.3389/fmars.2023.1130898.

Sladen, A., D. Rivet, J. P. Ampuero, L. De Barros, Y. Hello, G. Calbris, and P. Lamare (2019). Distributed sensing of earthquakes and ocean-solid Earth interactions on seafloor telecom cables, *Nat. Commun.* **10,** no. 1, 1–8, doi: 10.1038/s41467-019-13793-z.

University of Wisconsin (2016a). Brady's geothermal field DAS earthquake data [Data set], University of Wisconsin, doi: 10.15121/1334285.

University of Wisconsin (2016b). Brady's geothermal field nodal seismometer earthquake data [Data set], University of Wisconsin, doi: 10.15121/1334284.

University of Wisconsin (2016c). Brady's geothermal field DAS and DTS surface and borehole array metadata [Data set], University of Wisconsin, doi: 10.15121/1261907.

Walter, F., D. Gräff, F. Lindner, P. Paitz, M. Köpfli, M. Chmiel, and A. Fichtner (2020). Distributed acoustic sensing of microseismic sources and wave propagation in glaciated terrain, *Nat. Commun.* **11,** no. 1, doi: 10.1038/s41467-020-15824-6.

Wang, H. F., X. Zeng, D. E. Miller, D. Fratta, K. L. Feigl, C. H. Thurber, and R. J. Mellors (2018). Ground motion response to an ML 4.3 earthquake using co-located distributed acoustic sensing and seismometer arrays, *Geophys. J. Int.* **213,** no. 3, 2020–2036, doi: 10.1093/GJI/GGY102.

Wang, X., Z. Zhan, E. F. Williams, M. G. Herráez, H. F. Martins, and M. Karrenbach (2021). Ground vibrations recorded by fiber-optic cables reveal traffic response to COVID-19 lockdown measures in Pasadena, California, *Commun. Earth Environ.* **2,** no. 1, 1–9, doi: 10.1038/s43247-021-00234-3.

Wilcock, W. S. D., and Ocean Observatories Initiative (2023). Rapid: A community test of distributed acoustic sensing on the ocean observatories initiative regional cabled array [Data set], *Ocean Observatories Initiative*, doi: 10.58046/5J60-FJ89.

Wilcock, W. S. D., S. Abadi, and B. P. Lipovsky (2023). Distributed acoustic sensing recordings of low-frequency whale calls and ship noise offshore Central Oregon, *JASA Express Lett.* **3,** no. 2, 026002, doi: 10.1121/10.0017104.

Williams, E. F., M. R. Fernández-Ruiz, R. Magalhaes, R. Vanthillo, Z. Zhan, M. González-Herráez, and H. F. Martins (2019). Distributed sensing of microseisms and teleseisms with submarine dark fibers, *Nat. Commun.* **10,** no. 1, 1–11, doi: 10.1038/s41467-019-13262-7.

Williams, E. F., Z. Zhan, H. F. Martins, M. R. Fernández-Ruiz, S. Martín-López, M. González-Herráez, and J. Callies (2022). Surface gravity wave interferometry and ocean current monitoring with ocean-bottom DAS, *J. Geophys. Res.* **127,** no. 5, 1–27, doi: 10.1029/2021JC018375.

Xiao, H., B. Gaite, and S. Ruiz-Barajas (2022). Locating the precise sources of high-frequency microseisms using distributed acoustic sensing, *Geophys. Res. Lett.* doi: 10.1029/2022GL099292.

Yang, Y., J. W. Atterholt, Z. Shen, J. B. Muir, E. F. Williams, and Z. Zhan (2022). Sub-kilometer correlation between near-surface structure and ground motion measured with distributed acoustic sensing, *Geophys. Res. Lett.* **49,** no. 1, doi: 10.1029/2021GL096503.

Yang, Y., Z. Zhan, Z. Shen, and J. Atterholt (2022). Fault zone imaging with distributed acoustic sensing: Surface-to-surface wave scattering, *J. Geophys. Res.* **127,** no. 6, doi: 10.1029/2022jb024329.

Zeng, X., F. Bao, C. H. Thurber, R. Lin, S. Wang, Z. Song, and L. Han (2022). Turning a telecom fiber-optic cable into an ultradense seismic array for rapid postearthquake response in an urban area, *Seismol. Res. Lett.* **93,** no. 2A, 853–865, doi: 10.1785/0220210183.

Zhan, Z. (2019). Distributed acoustic sensing turns fiber-optic cables into sensitive seismic antennas, *Seismol. Res. Lett.* **91,** no. 1, 1–15, doi: 10.1785/0220190112.

Zhirnov, A. A., K. V. Stepanov, A. O. Chernutsky, A. K. Fedorov, E. T. Nesterov, C. Svelto, A. B. Pnev, and V. E. Karasik (2019). Influence of laser frequency drift in phase-sensitive optical time-domain reflectometry, *Opt. Spectrosc.* **127,** doi: 10.1134/S0030400X1910031X.

Zhu, T., and D. J. Stensrud (2019). Characterizing thunder-induced ground motions using fiber-optic distributed acoustic sensing array, *J. Geophys. Res.* **124,** no. 23, 12,810–12,823, doi: 10.1029/2019JD031453.

Zhu, T., J. Shen, and E. R. Martin (2021). Sensing earth and environment dynamics by telecommunication fiber-optic sensors: An urban experiment in Pennsylvania, USA, *Solid Earth* **12,** no. 1, 219–235, doi: 10.5194/se-12-219-2021.

Zhu, W., E. Biondi, J. Li, J. Yin, Z. E. Ross, and Z. Zhan (2023). Seismic arrival-time picking on distributed acoustic sensing data using semi-supervised learning, *Nat. Commun.* **14,** no. 1, 1–17, doi: 10.1038/s41467-023-43355-3.

Manuscript received 29 March 2024
Published online 26 July 2024