submitted to *Geophys. J. Int.*

Distributed Acoustic Sensing Data Compression for Seismological Applications via Compressive Sensing

Yang Ma[1], Lingsen Meng[1,*], Yen-Yu Lin[2,3,4]

[1]*Department of Earth, Planetary and Space Sciences, University of California Los Angeles, Los Angeles, CA, USA*

[2]*Department of Earth Sciences, National Central University, Taoyuan, Taiwan*

[3]*Earthquake-Disaster & Risk Evaluation and Management Center, National Central University, Taoyuan, Taiwan*

[4]*Graduate Institute of Applied Geology, National Central University, Taoyuan, Taiwan*

[*]*Corresponding author: L. Meng (lsmeng@g.ucla.edu)*

SUMMARY

Distributed Acoustic Sensing (DAS) is an emerging technology that turns existing optical-fiber cables into high-density seismic arrays, generating vast amounts of observational data in various contexts. Consequently, large-scale storage, transmission and processing of DAS data present both challenges and opportunities in seismology. In this study, we propose a data compression and seismic detection workflow based on compressive sensing (CS) and apply it to DAS data from the Taiwan Milun Fault Drilling and All-inclusive Sensing (MiDAS) project. Our algorithm achieves a compression ratio of at least 15, with reconstructed data from specific earthquakes compatible with standard seismological algorithms. Additionally, we migrate the detection algorithm to the compressed domain, enabling quasi-real-time, high-accuracy seismic detection. This approach demonstrates the feasibility of directly processing compressed data, reducing computational burdens in DAS processing. Furthermore, we discuss an empirical criterion for determining the maximum compression ratio of a given signal based on CS. Our workflow is compared with other mature compression algorithms across eight different datasets to demonstrate its advantages, limitations, and applicability. Generally, the CS algorithm is more effective for high-SNR event-oriented DAS applications rather than continuous noise-dominated monitoring. Collectively, these results highlight the potential of the CS algorithm in advancing the development of efficient, user-friendly DAS data products.

Key words: Time-series analysis; Distributed acoustic sensing; Earthquake monitoring

1 INTRODUCTION

Distributed Acoustic Sensing (DAS) is an emerging seismic observation technology that has rapidly developed in recent years. By repurposing existing telecommunication optical cables, DAS creates low-cost, high-density seismic arrays with broad applicability. It measures the phase shift of Rayleigh backscattered waves within the fiber gauge length, thereby capturing strain or strain rate signals along the fiber. Previous studies have shown that DAS provides useful and high-fidelity data for natural earthquake studies (Yu et al. 2019). Compared to traditional seismometers, DAS offers meter-scale sensor spacing and kilohertz sampling rates, enabling the recording of seismic wavefield information with higher resolution and frequency. This makes DAS ideal for applications such as microseismic monitoring (Yu et al. 2024; Hudson et al. 2021; Trabattoni et al. 2022; Lin et al. 2026), earthquake source analysis (Li et al. 2023; Strumia et al. 2024; Yin et al. 2023), subsurface structure tomography, (Yang et al. 2022; Benjumea et al. 2024; Shragge et al. 2021; Atterholt & Zhan 2024) and fault-zone amplification (Ma et al. 2024). Moreover, its ease of deployment and maintenance has extended seismic research into diverse environments, including urban infrastructure (Lin et al. 2025), seabed (Williams et al. 2019, 2022; Mata Flores et al. 2023; Fukushima et al. 2022; Lindsey et al. 2019), volcanic regions (Biondi et al. 2023; Jousset et al. 2022; Biagioli et al. 2024), and glaciers (Yang et al. 2024; Fichtner et al. 2025), demonstrating its broad interdisciplinary potential. Comprehensive reviews of DAS applications are available in recent literature (Zhan 2020; Lindsey & Martin 2021).

However, the high spatiotemporal resolution of DAS systems leads to massive data volumes. A typical DAS deployment, spanning kilometers and operating at kilohertz sampling rates, can generate terabytes of data per day, posing significant challenges for storage, transmission, and analysis. Data compression is thus a critical strategy to address this issue. Various compression methods have been proposed as seismic datasets expand (Sebai et al. 2024). The simplest method is downsampling, which first applies low-pass filtering to remove high-frequency components that could otherwise introduce aliasing and then reduces data volume by retaining samples at longer intervals than the original sampling rates. While effective in reducing volume, downsampling irreversibly discards high-frequency information, and interpolation used for reconstruction fails to exactly recover the original signal.

Another approach is to apply classical encoding-based compression algorithms, such as TurboPFor, which converts the raw data into unsigned integer format and subsequently reduces volume by encoding repeating patterns (Dong et al. 2022). However, such methods primarily exploit statistical redundancy in the data and may not fully capture the strong spatiotemporal correlations and physical structure of DAS recordings. Wu et al. (2023) proposed a two-dimensional lossless compression method by combining inter-channel prediction, intra-channel linear predictive coding, and entropy coding. Although their approach achieved more than 50% space savings and outperformed conventional lossless compressors, the resulting compression ratios remain relatively limited.

More advanced techniques involve coordinate transformations to a new basis, allowing sparse representation, extraction of key features and compression. Typical transformations include wavelet, curvelet and those tailored to seismic data (Spanias et al. 1991; Geng et al. 2020). While similar principles underpin widely used image and video compression algorithms, such as the Joint Photographic Experts Group (JPEG) (Marcellin et al. 2000) and High Efficiency Video Coding (HEVC) (Wien 2015), the non-uniform nature of DAS data limits their direct application, and additional data quality control and selection are often required. Several studies have reviewed and benchmarked these compression strategies. Issah & Martin (2024) compared encoding and transformation compression methods and showed that the preferred method depends on the tolerated compression errors and the scientific objective. Seguí et al. (2026) proposed a high-performance compression framework that integrates both lossy and lossless modes, allowing users to specify compression settings. It also balances efficiency and computational speed, making it promising for future DAS data management.

Recent advances in artificial intelligence have introduced alternative compression approaches. Chen et al. (2025) reported eightfold compression on validation datasets using a vision transformer structure. The efficiency can be further improved through downsampling processes. However, the generalizability of the method and the physical interpretation of its reconstruction errors remain unclear, which limits their application to real-world scenarios.

Compressive sensing (CS) (Donoho 2006), by reformulating compression as an optimization problem, provides a novel method for both data reduction and interpretation of compressed signals. It has been combined with transform-based compression algorithms and executed at the sensor scale (Bai et al. 2020; Bin et al. 2021). Here, we adopt a frequency-domain CS algorithm, which achieves relatively high compression ratios while offering better physical interpretability, providing a more transparent alternative.

Furthermore, most compression approaches still require data reconstruction before scientific use. To overcome this limitation, we propose a workflow based on frequency-domain CS that enables both high-ratio compression and direct seismic detection in the compressed domain. By exploiting the ap-

Table 1. Main fiber sections of the MiDAS array and their corresponding channel indices. Hole B channels 673-773 are used to evaluate the CS-based compression and seismic detection workflow, whereas Surface B channels 100 and 501 are used to evaluate the CS algorithm on ambient-noise data.

Fiber section	Length (m)	Channel index	Type
Surface B, forward	1185	57–347	Surface
Surface B, backward	1185	355–645	Surface
Hole B, downward	497	650–772	Downhole
Hole B, upward	497	772–894	Downhole
Telecommunication cables	1164	900–1185	Surface
Hole A, downward	695	1208–1378	Downhole
Hole A, upward	695	1378–1548	Downhole
Horizontal strain section	102	1559–1584	Surface
Surface A, forward	330	1588–1669	Surface
Surface A, backward	335	1670–1752	Surface

proximate amplitude-preserving behavior of CS measurements, we adapt the conventional STA/LTA seismic detection algorithm to the compressed domain, allowing quasi-real-time monitoring and compression without full reconstruction. We also demonstrate that the reconstructed seismic signals are compatible with standard seismological analyses, including phase picking and array seismology. Finally, we compare our method with other compression algorithms across different datasets to evaluate the advantages, limitations, and applicability of the CS algorithm for DAS data.

2 DATA

The Milun Fault Drilling and All-inclusive Sensing (MiDAS) project is a collaborative seismic observation study conducted by the Institute of Earth Sciences, Academia Sinica, and the Earthquake-Disaster & Risk Evaluation and Management Center, National Central University (Huang et al. 2024). This project deploys a fiber optic network across the Milun Fault, both on the surface and at depth. The network consists of three sections: borehole A on the hanging wall, borehole B on the footwall, and existing telecommunication optical cables in populated urban areas nearby (Table 1). The total length of the cable is approximately 7,500 meters, with 4 meters spacing. The Silixa iDAS interrogator is used for data acquisition, with a 10 meters gauge length and a 1,000 Hz sampling rate. The project generates 300 GB of data per day.

The Milun Fault, an active strike-slip fault, has a high probabilistic seismic hazard in Taiwan (Wang et al. 2016; Chan et al. 2020) with a short recurrence interval (Shyu et al. 2016). It is located near the plate boundary between the Eurasian Plate and the Philippine Sea Plate. In the 2018 M_w 6.4 Hualien earthquake, the rupture initiated on an offshore fault and propagated unilaterally to the south-southwest, triggering the nearly vertical Milun Fault and leading to surface rupture (Lo et al. 2019; Yen et al. 2019; Kuo et al. 2019). On April 3, 2024, an M_w 7.4 mainshock struck offshore of Hualien, followed by a M_L 6.6 aftershock after 10 minutes, causing at least 18 fatalities and over 1,100 injuries, highlighting the region's high seismic risk (Zheng et al. 2024). Fortunately, the DAS network provides high-resolution continuous data during the 2024 Hualien earthquake, offering valuable insights for future studies of fault morphology and seismic rupture processes.

In this study, we first use continuous data from borehole B between 00:00 and 01:00 UTC on the day of the 2024 Hualien earthquake to evaluate the performance of the proposed compression and detection workflow during an active aftershock sequence. To minimize the influence of anthropogenic noise, we retain 100 DAS channels from depths ranging from 100 to 500 meters (channel 673-773). The raw data are preprocessed by demeaning, low-pass filtering at 10 Hz, and normalizing, because the amplitude spectrum of the original data is very weak above 10 Hz. Both compression and detection are then performed on the preprocessed data. We also use 32 days of continuous data from channels 100 and 501 on the surface (Figure 1b), spanning 31 March to 1 May 2024, when seismicity gradually decayed, to evaluate the applicability of the CS algorithm to traffic and ambient noise records. For this dataset, we use the raw data without additional preprocessing. Figure 1 shows the cable layout, highlighting the locations of the selected channels and borehole B, together with the locations of the 18 earthquakes detected by our detection algorithm.

To assess the generalizability of CS beyond the Hualien case study, we compile seven additional datasets reported in the literature. Together with the synthetic Marmousi benchmark, these datasets span a broad range of observation scenarios, including onshore and offshore deployments, active volcanic systems, unstable mountainous areas, and the Greenland ice sheet, exposing the CS algorithm to wavefields generated by earthquakes, tectonic tremors, volcanic activity, ocean waves, icequakes, and landslides under different noise conditions. The datasets also differ in acquisition parameters, such as sampling rates (20–4000 Hz), sensing lengths (1–120 km), channel spacing (1–9.8 m), and gauge lengths (5–78.4 m). Some datasets represent continuous long-term monitoring, whereas others consist of event-oriented recordings. Such variations influence the spatial coherence, bandwidth, and signal complexity of the recorded data, providing a more demanding and representative test for compression performance. Table 2 summarizes the datasets used in this study. Their diversity enables us to evaluate CS under a broad range of realistic recording conditions and to better understand its advantages, limitations, and practical applicability.

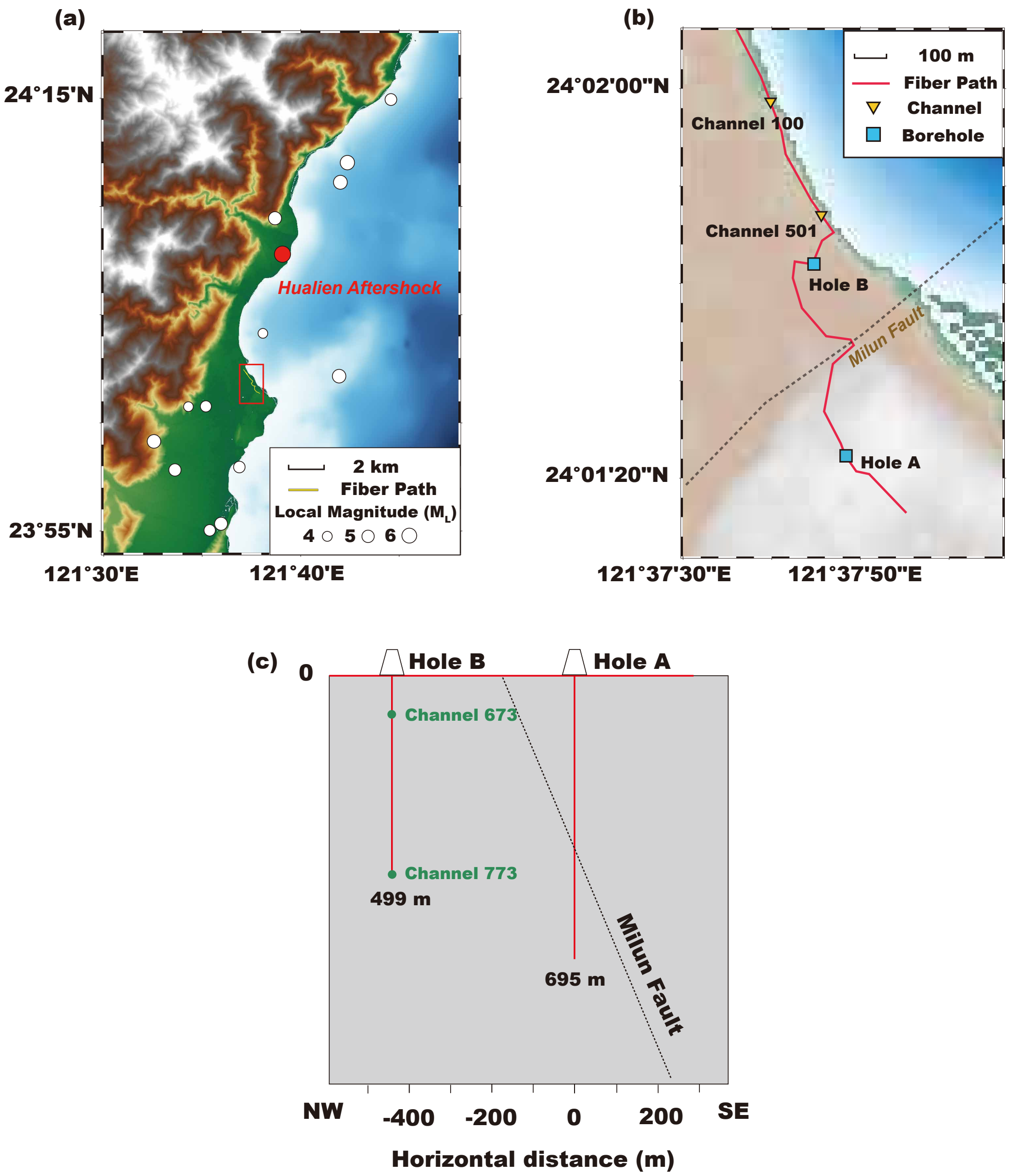

Figure 1. Compressed domain STA/LTA detection results and DAS fiber path. (a) Distribution of detected earthquakes, with the farthest event located 27 km from the DAS deployment. The Hualien aftershock is marked by the red circle, and the DAS fiber path is enclosed in the red box. (b) Schematic diagram of the fiber path, highlighting the two boreholes and the channels used for the ambient noise cross-correlation stacking. Faults are marked with dotted lines. (c) Side view of the boreholes. The red line represents the optical-fiber cable path. The relative positions of hole A and hole B, along with their maximum depths, are indicated. The fault is marked with the dashed line.

Table 2. Summary of the datasets used in this study. The datasets are ordered by decreasing signal-to-noise ratio of the raw data. All datasets except Hualien are published and can be accessed through the corresponding references. These datasets were acquired using different interrogators and vary in channel spacings, sampling rates, and data types (strain or strain rate), providing a benchmark for evaluating the performance of the CS algorithm.

Location	Dominant signal	Sampling rate (Hz)	Reference
Marmousi	Synthetic seismic	4000	Iqbal et al. (2023)
Alaska	Oceanic	25	Ni et al. (2024b)
Hualien	Earthquake	1000	Huang et al. (2024)
Stromboli	Volcanic	200	Biagioli et al. (2024)
Ridgecrest	Earthquake	250	Atterholt (2022)
Noto	Tremors	100	Baba et al. (2025)
Greenland	Icequakes	1000	Fichtner et al. (2025)
Brienz	Landslides	20	Kang et al. (2024)

3 METHODS

3.1 Compressive Sensing

Compressive Sensing (CS) is a signal processing theory introduced in the early 2000s (Donoho 2006). It states that signals with sparse representations can be recovered from fewer measurements than required by the Shannon–Nyquist sampling theorem. This provides a convenient and effective approach to compression in practice, which has been applied across various fields (Rani et al. 2018), such as video processing (Mun & Fowler 2011), communication systems (Liu et al. 2015) and biomedical applications (Tang et al. 2011). Recent studies have shown that the CS algorithm performs well in compression of laboratory DAS recordings (Qu et al. 2019a,b). Building on this, we demonstrate that field-deployed DAS systems can also be effectively compressed using the CS algorithm.

Let the original single-channel DAS signal x be an $N \times 1$ column vector in the time domain. The general coordinate transformation can be expressed as:

$$x = \mathbf{\Psi} s, \tag{1}$$

where Ψ is an $N \times N$ transformation matrix whose columns are basis vectors in the new linear space, and s is an $N \times 1$ column vector representing the signal in this basis. Commonly used transformations include the discrete Fourier transform (DFT), discrete cosine transform (DCT), and discrete wavelet transform (DWT). If there are only K non-zero elements in s, this signal is referred to as a K-sparse signal. A smaller K indicates fewer elements are required to fully represent the original signal in the compressed domain, which reduces the dimensionality of the recovery problem and can accelerate reconstruction. In this study, we observe that DAS records are typically sparse in the frequency domain, and because many standard seismic data processing methods rely on frequency analysis, which facilitates algorithm migration, we select the DFT.

CS theory states that when a signal is sparse in a suitable transform domain, it can be recovered from a small number of linear measurements. Specifically, we construct an $M \times N$ sensing matrix $\boldsymbol{A}$, whose entries are randomly sampled from the standard normal distribution. The compressed signal is obtained:

$$y = \boldsymbol{A}s = \boldsymbol{\Phi\Psi}s = \boldsymbol{\Phi}x, \tag{2}$$

where $\boldsymbol{\Phi}$ denotes the corresponding measurement matrix and is used later for seismic detection in the compressed domain. Here, M is much smaller than N, i.e., $K < M \ll N$, so the compressed signal y is significantly smaller than the original signal x. To reconstruct the original signal with high possibility, the number of compressed measurements M should be greater than the signal sparsity K, which is a necessary condition for the sensing matrix $\boldsymbol{A}$ to satisfy the Restricted Isometry Property (RIP), so that the inverse problem is solvable (Candès et al. 2006).

The schematic diagram of the CS algorithm is shown in Figure 2a. In this study, we use the Orthogonal matching pursuit (OMP) algorithm, a simple and widely used reconstruction method that performs well in most scenarios (Manchanda & Sharma 2020). OMP is an iterative greedy algorithm that reconstructs a signal by selecting the most relevant components at each step based on their correlation with the residual. The compression efficiency is characterized by the compression ratio N/M. We first apply the algorithm to the MiDAS records from a selected window to explore compression performance at various ratios. Finally, we choose a 15-fold compression ratio for processing all the data. The specific details are discussed in the following section.

3.2 Seismic Detection in Compressed Domain

Vibration detection serves as a starting point for integrating the CS algorithm with DAS. Driven by the need for real-time monitoring at specific locations in engineering, some laboratory DAS experiments have used piezoelectric transducers to simulate external vibration sources, and have applied CS-based methods to detect vibration, with promising results (Xiao et al. 2023). Most studies identify vibration sources by tracking changes in the sparsity K estimated during compression, treating abrupt changes in sparsity as indicators of vibration events (Gao et al. 2022). This approach assumes that the frequency components of a vibration are generally absent in the background noise, and it fails to fully exploit the information within the compressed signals. A more refined approach introduces a filter in

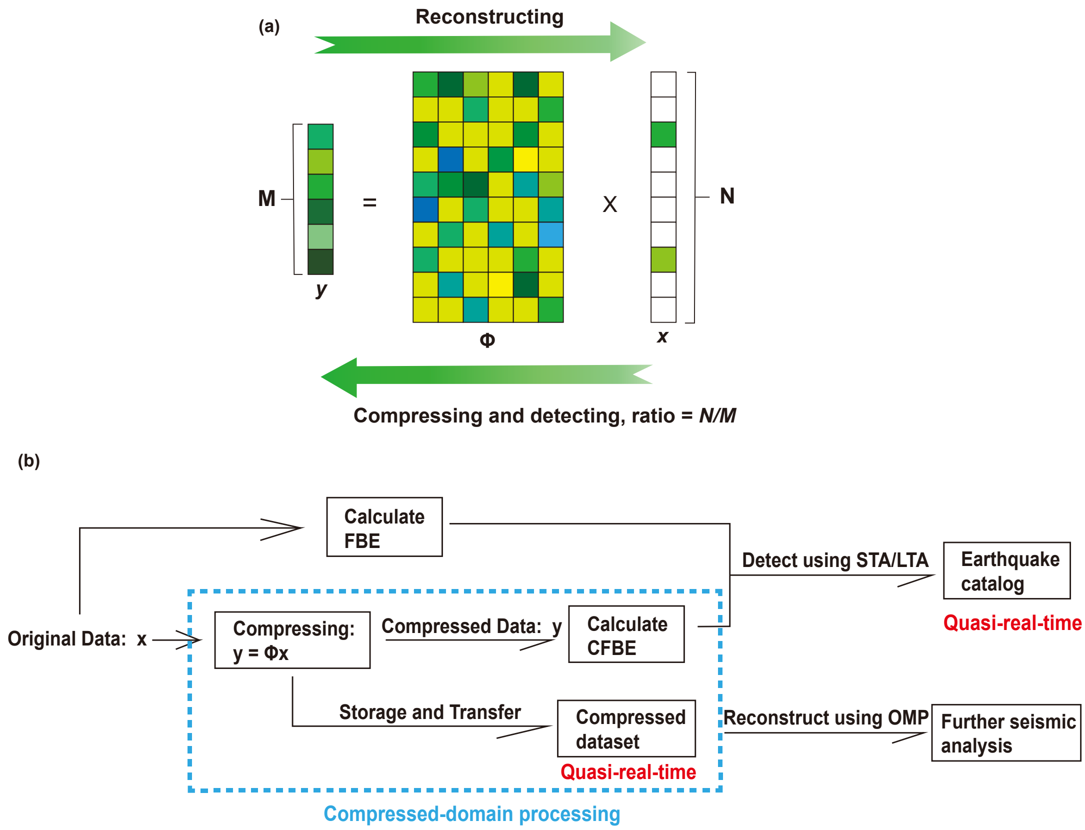

Figure 2. Data compression and seismic detection workflow based on the CS algorithm. (a) Schematic diagram of the CS algorithm. White squares indicate zero values, illustrating the sparsity of the original data x. (b) Workflow diagram proposed in this study.

the compressed domain to decompose the signal into different frequency components. The CS algorithm's nearly amplitude-conserving property ensures that the average energy of the signal in different frequency bands can be calculated within the compressed domain, allowing vibration sources to be identified by monitoring energy variations across these bands (Shen et al. 2024).

In this study, we adopt this design, specifically:

$$FBE\left(\omega\right) = \frac{1}{N}\sum_{i=1}^{N} h_i^2, \tag{3}$$

$$h = f * x = \boldsymbol{F}x \tag{4}$$

Frequency Band Energy (FBE) represents the average energy of the signal h in a given frequency band. Therefore, FBE is a function of frequency ω. In the time domain, h can be obtained either by convolving with a bandpass filter f, or by multiplying the signal vector by the finite-response bandpass filter matrix $\boldsymbol{F}$, represented as an $N \times N$ Toeplitz matrix. Although the latter increases the algorithm's time complexity, it facilitates the design of the filter matrix in the compressed domain. In the compressed domain, we require:

$$\boldsymbol{\Phi F} x = \boldsymbol{F_c \Phi} x = \boldsymbol{F_c} y, \tag{5}$$

where the left side represents compression after filtering, while the right side indicates bandpass filtering of the compressed signal. Here, $\boldsymbol{F_c}$ is the filter matrix in the compressed domain, and the measurement matrix $\boldsymbol{\Phi}$, which satisfies the RIP, preserves amplitude. A straightforward linear algebra derivation yields:

$$\boldsymbol{F_c} = \boldsymbol{\Phi F \Phi^T} \left(\boldsymbol{\Phi \Phi^T}\right)^{-1} \tag{6}$$

This is the bandpass filtering matrix in the compressed domain, thus allowing us to calculate the Compressed Frequency Band Energy (CFBE) directly using compressed signals:

$$h_c = \boldsymbol{F_c \Phi x} = \boldsymbol{F_c y} \tag{7}$$

$$CFBE\left(\omega\right) = \frac{1}{M} \sum_{i=1}^{M} h_{c\_i}^2 \tag{8}$$

where h_c represents the filtered signal in the compressed domain.

Encouragingly, this approach aligns with the classical STA/LTA seismic detection algorithm (Allen 1978), which has been widely performed in early warning systems. The STA/LTA uses two overlapping time windows of different lengths to extract segments of the signal and calculate average energy in each window. The longer window averages seismic signals and ambient noise, representing the background energy, while the shorter window is sensitive to energy fluctuations, typically corresponding to rapid-onset tectonic earthquakes. Building on this, we incorporate long and short time windows into the compression and detection workflow. Long windows are extracted to characterize background energy (FBE) in the time domain. Within each short window, the original signal is compressed using the CS algorithm, and the bandpass filter matrix in the compressed domain is applied to compute CFBE in the same frequency bands. Comparing CFBE with FBE then allows classification of each short window as either containing an event or not. The operations are conducted in both the time domain and the compressed domain to establish that the computed CFBE and FBE are numerically comparable, thereby validating that the CS algorithm preserves amplitude during compression. If only the compressed data are available, it is entirely feasible to directly compute the average energy within short and long windows in the compressed domain, enabling efficient seismic detection. A simplified workflow is shown in Figure 2b.

After testing, we select 10 equally spaced frequency bands between 0.1–10 Hz, a 3-min long window with no overlap and a 10-s short window with a 5-s overlap. An earthquake is considered to be detected when more than 50% of the DAS channels show a CFBE value 1.5 times the FBE value. It is important to note that the choice of parameters highly depends on the array setting, source amplitudes and signal-to-noise ratio. Moreover, the monitoring conditions can be further refined using energy from different frequency bands and channels.

3.3 Determining the Maximum Compression Ratio

Establishing a systematic approach to determine the maximum compression ratio is essential, particularly when applying the CS algorithm to diverse DAS deployments. Thus, we derive an empirical relationship between compression ratio and reconstruction error for DAS data based on the mathematical framework of the CS algorithm. This relationship enables the rapid estimation of the maximum compression ratio for a prescribed reconstruction error threshold.

Following the notation in section 3.1, the number of measurements M and sampling points N define the compression ratio (CR), whereas M depends on the signal sparsity K in the frequency domain. Since DAS data are not strictly sparse, K is defined as the number of spectral amplitudes exceeding a threshold, where the top K frequency components S_K represent the signal, and the remaining components constitute the noise e. We adopt the universal threshold $T = \frac{\sigma}{2}\sqrt{2lnN}$ proposed by Donoho & Johnstone (1994), where σ denotes the standard deviation of the amplitude spectrum. For the random Gaussian sensing matrix $\boldsymbol{A}$ used in this study, Baraniuk (2007) demonstrated that the original signal can be reconstructed with high probability when

$$M \geq C_1 K log_{10}\left(N/K\right), \tag{9}$$

where C_1, the compression number, depends on the RIP of $\boldsymbol{A}$. The parameters C_1, K, and N jointly determine the number of measurements M, which in turn defines the compression ratio (CR). Larger C_1 values correspond to lower compression ratios.

High probability implies acceptable reconstruction accuracy, expressed as

$$\parallel \hat{x} - x \parallel \leq C_2 \parallel e \parallel \tag{10}$$

by Candès & Tao (2006) and adopted in subsequent studies (Wu et al. 2013; Cai & Wang 2011). The parameter C_2, termed the error number, also reflects the RIP characteristics of $\boldsymbol{A}$. Although C_1 and C_2 are mathematically linked through RIP, deriving their exact relationship is impractical (Tillmann & Pfetsch 2014). Therefore, we can only empirically derive an approximate numerical relationship between C_1 and C_2 using real-world data. The results and the specific functional forms are discussed in a later section.

To express the error number C_2 in a more intuitive form, we link it with the lower bound for the correlation coefficient between the reconstructed and original signals (Appendix),

$$\rho \geq 1 - \frac{(1 + C_2)^2}{2\left(SNR - 2\sqrt{SNR} + 1\right)} \tag{11}$$

where $SNR = \mid S_K \mid^2 / \mid e \mid^2$ is the signal-to-noise ratio. Thus, by specifying a target correlation coefficient ρ_0, one can sequentially invert for C_2 and C_1, then calculate the number of measurements M, thereby determining the maximum compression ratio given the signal's SNR and sparsity K. These relationships also indicate that the achievable compression ratio depends primarily on the strength of the dominant frequency components (SNR), and the effective bandwidth, represented by the sparsity level K. It does not depend on the sampling rate itself.

4 RESULTS

4.1 Compression and Quasi Real-Time Detection of the Hualien Aftershock Sequence

In this section, we apply the compression and detection workflow to the Hualien aftershock dataset. We first extract a 90-second time window (00:11:20–00:12:50) containing the M_L 6.6 aftershock to identify the optimal compression ratio using the catalog provided by the Taiwan Seismological and Geophysical Data Management System (GDMS) (Central Weather Administration 2012). We exclude the mainshock data because potential wrapping in the DAS records could affect the detection results. For various compression ratios, we randomly select ten channels for compression and reconstruction, calculating the correlation coefficient between the original and reconstructed signals. Compression is performed using CS, whereas reconstruction is performed using the OMP algorithm, as disscussed in Section 3.1. The results indicate that for compression ratios up to 20, the correlation coefficient exceeds 0.9, with small standard deviation (Figure 3a). With a more conservatively selected compression ratio of 15, the correlation coefficient is 0.9605, preserving nearly all information of the original data. To further assess whether the compression error is acceptable, we compare the time series and amplitude spectra of the first channel, which shows strong agreement at most points. Since DAS data are not strictly sparse, slight noise is introduced (Figure 3b). The original and reconstructed signals for all 100 channels are plotted (Figure 3c, d, f, g), and the reconstruction errors are calculated (Figure 3e, h), demonstrating an error sufficiently small to preserve the applicability of most classical seismic signal processing techniques, which will be discussed in the discussion section. Quantitatively, we calculate the root-mean-square error (RMSE) and peak signal-to-noise ratio (PSNR), where RMSE indicates

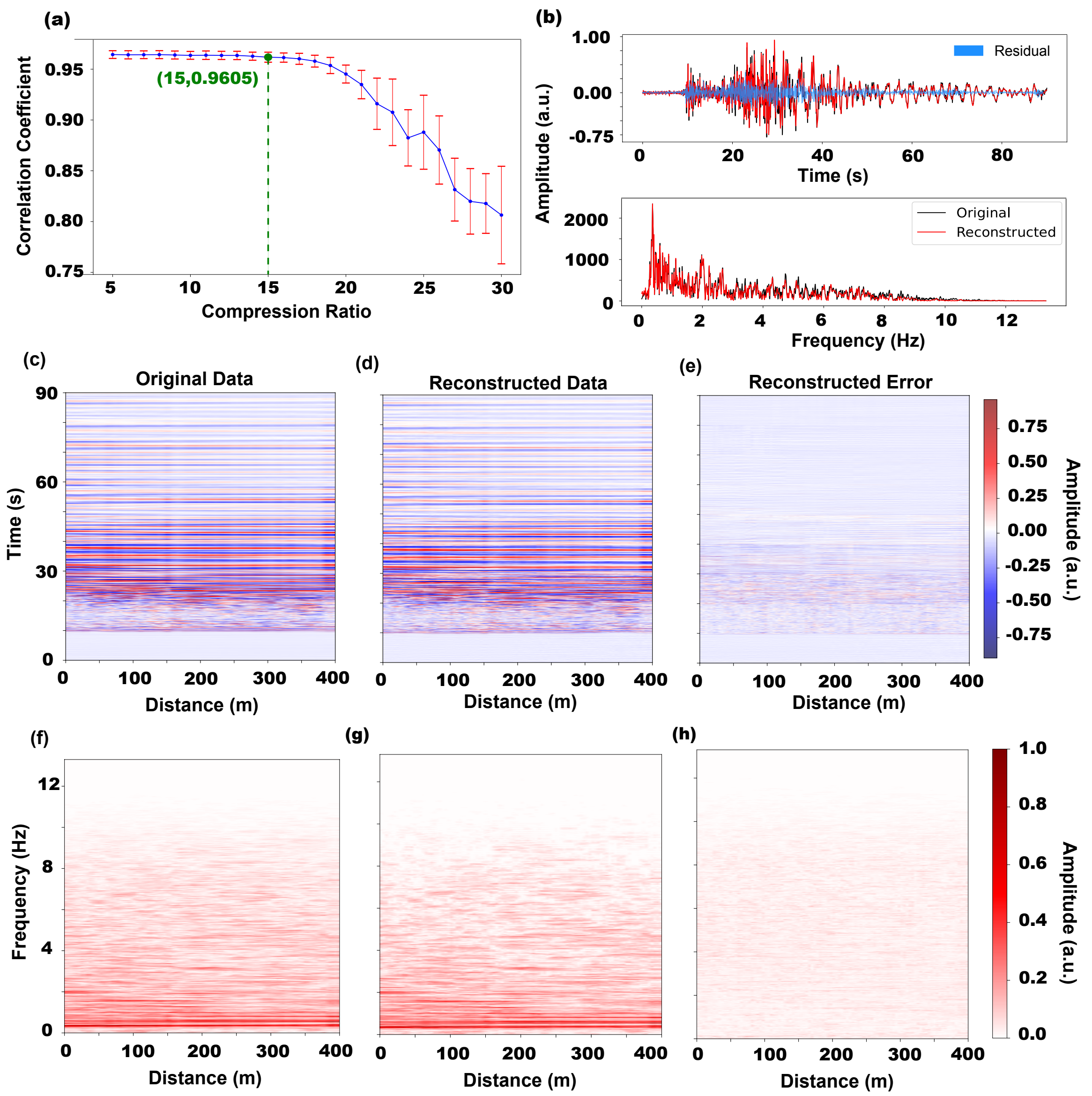

Figure 3. Evaluation of the CS-algorithm compression results. (a) Optimization of the compression ratio, with a final selection of a 15-fold compression ratio, demonstrating strong and stable reconstruction performance. (b) Comparison of the time series and amplitude spectrum between the original and reconstructed data of the Hualien aftershock at the channel located 100 m below the surface (channel 673) in borehole B. The black line represents the original data, while the red line corresponds to the reconstructed data after compression. The blue line represents the residuals. (c) Original data of the largest aftershock from borehole B. (d) Reconstructed data after compression of the data shown in panel (c). (e) Reconstruction error of the compressed data. (f, g, h) Amplitude spectra of the original and reconstructed signals and the reconstruction error.

Table 3. Seismic detection results using the compressed-domain STA/LTA algorithm. The table lists 18 detected earthquakes in chronological order. The Start time and End time columns indicate the time windows automatically extracted by the compressed-domain STA/LTA algorithm. Other information is obtained from the GDMS catalog. Two potentially undocumented events are marked in red; the last detected event has a weak amplitude and is not recorded in the GDMS catalog.

Start time (UTC)	End time (UTC)	Origin time (UTC)	Latitude (°)	Longitude (°)	Epicentral distance (km)	M_L
0:11:20	0:12:50	0:11:26	24.1400	121.6513	12.62	6.56
0:12:50	0:13:10	0:12:54	24.1577	121.6450	14.48	5.28
0:17:40	0:18:00	0:17:37	24.0360	121.6992	7.17	5.48
0:18:35	0:18:45	0:18:27	23.9172	121.5897	12.99	4.60
0:21:10	0:21:25	0:19:15	24.0125	121.5717	6.09	3.92
0:21:50	0:22:10	0:19:33	24.0069	121.6347	4.56	3.94
0:25:15	0:25:30	0:25:16	24.0127	121.5865	4.66	4.48
0:27:10	0:27:25	0:27:02	23.8275	121.5705	23.11	4.91
0:31:10	0:31:40	0:31:14	23.9222	121.5993	12.18	5.25
0:35:35	0:36:25	0:35:37	24.2005	121.7062	20.68	5.79
0:37:05	0:37:20	0:37:04	23.9637	121.5602	10.03	4.96
0:39:45	0:40:00	0:39:41	24.2490	121.7403	27.12	4.75
0:41:05	0:41:25	0:41:04	23.9660	121.6148	7.08	4.69
0:43:55	0:44:20	0:43:55	23.9857	121.5425	10.50	5.48
0:46:45	0:47:15	0:46:44	24.1853	121.7002	18.89	5.61
0:50:50	0:51:15	0:50:46	23.8293	121.5722	22.87	5.52
0:54:20	0:54:35	0:54:09	23.8057	121.6085	24.84	4.93
0:59:10	0:59:20	/	/	/	/	/

the average error level, and PSNR indicates the signal fidelity:

$$RMSE = \sum_{i=1}^{N} \sqrt{\frac{1}{N} \left(I_{original}\left(i\right) - I_{reconstructed}\left(i\right)\right)^2} \tag{12}$$

$$PSNR = 20 * log_{10} \left(\frac{1}{RMSE}\right) \tag{13}$$

where $I_{original}$ and $I_{reconstructed}$ denote the normalized amplitudes of the original and reconstructed signals, respectively, and i iterates over all channels and time samples. The calculated $RMSE$ across all channels is 0.044 in relative amplitude, and the $PSNR$ reaches 27.050 dB, demonstrating minimal reconstruction error and high signal fidelity.

Next, we apply the proposed detection workflow in the compressed domain to the one-hour continuous dataset. The computation, run on a single-core Apple M4 CPU, took 18 minutes—less than the data length—demonstrating the feasibility of real-time compression and seismic detection for edge computing. The compressed data are stored in HDF5 format, a hierarchical scientific data storage

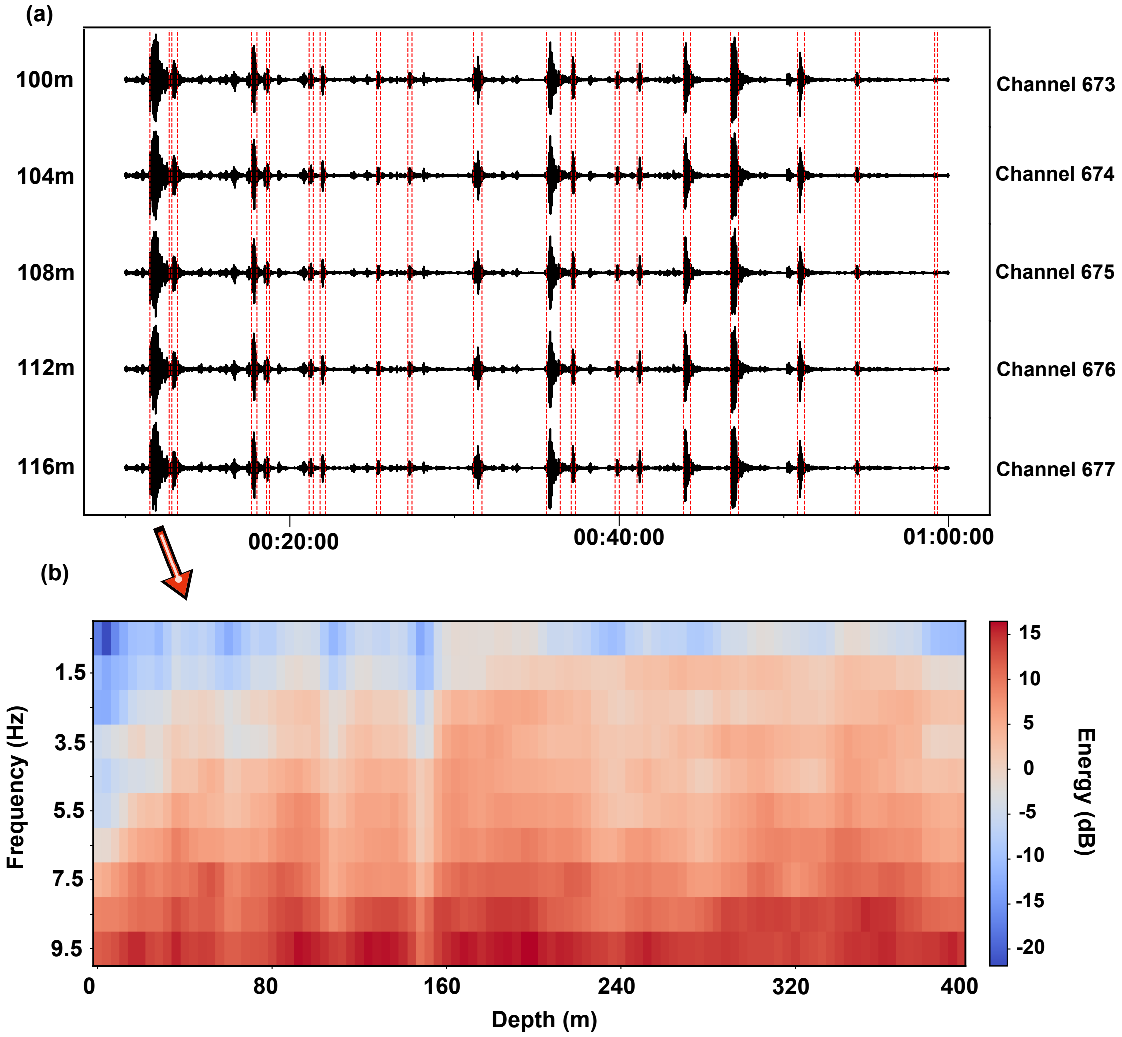

Figure 4. Evaluation of the STA/LTA detection algorithm in the compressed domain. (a) Original data from five adjacent channels during the early morning of April 3, 2024. The red dashed line highlights earthquakes detected by the STA/LTA algorithm in compressed domain. (b) Energy differences across the frequency bands used for detection. The horizontal axis represents the depth (measured from the reference point at 100 m below the surface) of each channel, and the vertical axis denotes the center frequency of each band. The color scale indicates the energy difference between the earthquakes and the background noise for each frequency band.

structure commonly used for DAS data. The original 1440 MB dataset is compressed to 96.8 MB, achieving a compression ratio of 14.88, slightly lower than the theoretical ratio of 15 due to the need to store the sparsity K for reconstruction and the hierarchical metadata in the HDF5 format. The algorithm successfully detects 18 earthquakes, 15 of which are recorded in the Taiwan Geophysical Database Management System (GDMS) catalog with a minimum local magnitude of M_L 4.48.

The event highlighted in bold red exhibits earthquake-like waveforms, but its estimated arrival time based on the GDMS catalog does not align with the STA/LTA detection window. This discrepancy suggests that the signal may correspond to a possible seismic event that is not documented in the GDMS catalog (Table 3). Based on the catalog-provided epicentral coordinates, we map the detected earthquakes, all within 30 km from Borehole B (Figure 1). This indicates that borehole DAS records are particularly suitable for near-field earthquake monitoring. Most detected earthquakes exhibit large amplitudes, while some manually identifiable small earthquakes are missed by the algorithm (Figure 4a), likely due to the selected STA/LTA parameters. Fine-tuning these parameters is necessary for improved microseismic monitoring. For the largest aftershock, energy below 1 Hz is lower than the background across all channels, with this trend extending up to 4 Hz in near-surface channels. This indicates persistent low-frequency noise may arise from wind or ocean currents and relatively high-frequency contamination from anthropogenic activities. In contrast, energy above 5 Hz is significantly elevated across all channels, corresponding to the dominant frequency range of the earthquake signal with high consistency (Figure 4b).

4.2 Generalization Across Diverse Datasets

In this section, we evaluate the generalizability of the CS algorithm across DAS datasets collected in different deployment environments. To avoid an exhaustive search for the optimal compression ratio and make the workflow easier to apply to previously untested datasets, we estimate the lower bound of the maximum compression ratio for each dataset using the procedure described in Method 3.3.

The first step is to establish the relationship between the compression number C_1 and the error number C_2, which is derived from Hualien aftershock data. We calculate the sparsity K of the original data, then invert C_1 for different fixed compression ratios (CR). Each signal is subsequently compressed and reconstructed to calculate C_2 by definition. The results (Figure 5a) show a clear L-curve pattern: when C_1 is small (high compression ratio), C_2 is large, indicating greater reconstruction error. As the compression ratio decreases, reconstruction error rapidly declines but never reaches zero due to measurement noise. For practical purposes, we fit the C_1–C_2 curve,

$$log_{10}C_2 = a + b * arctan\left(c * (log_{10}C_1 - d)\right) + e * log_{10}C_1 \tag{14}$$

where $a = 0.39$, $b = -0.42$, $c = 32.43$, $d = 0.244$, $e = -0.29$. Since the relationship between error number C_2 and the correlation coefficient ρ is derived (equation 11), with a reconstruction correlation threshold of $\rho_0 = 0.9$, the calculated maximum compression ratios for different datasets are shown in Table 4.

The results suggest that SNR is the primary control on the maximum compression ratio and re-

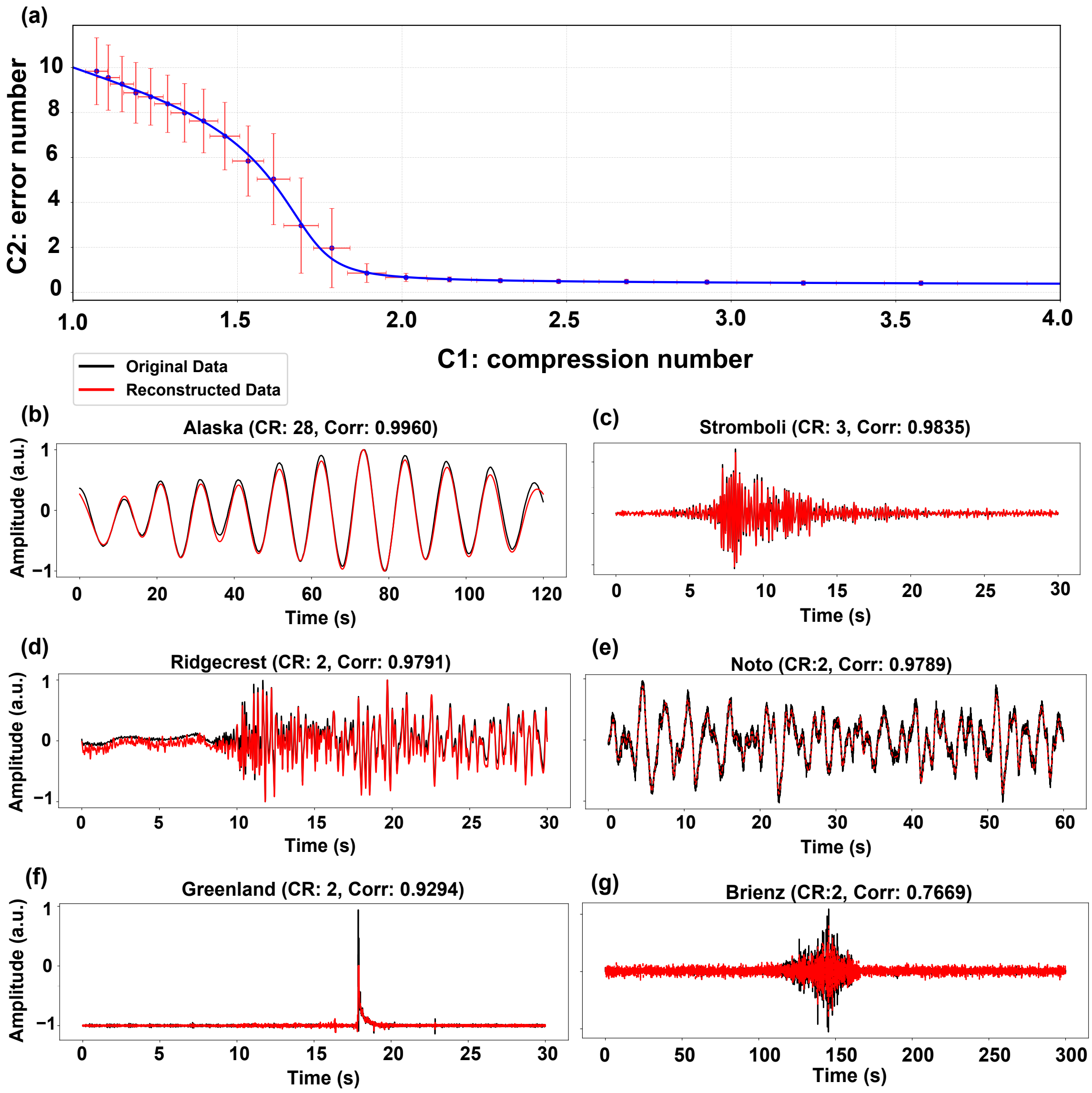

Figure 5. Empirical relationship between C_1 and C_2 and reconstruction results at the maximum compression ratio. (a) The C_1–C_2 curve derived from Hualien aftershock data. Vertical error bars represent variations in reconstruction performance among channels, while horizontal error bars reflect sparsity differences at the same compression ratio. (b)-(g) Comparison between the original and reconstructed single-channel signals for different datasets. The compression ratio and correlation coefficient are shown in panel titles. The panels are ordered by decreasing signal SNR; accordingly, the estimated maximum compression ratio decreases, reconstruction error increases, and the correlation coefficient decreases, consistent with predictions from our empirical relationship.

Table 4. Key parameters used to calculate the maximum compression ratio for each dataset. From left to right, the table lists sample points, sparsity, normalized sparsity, signal-to-noise ratio, error number, compression number, minimum number of measurements, and maximum compression ratio—arranged in the same order as the calculation sequence. Uncertainties in K and SNR reflect inter-channel variability. Datasets are ordered by decreasing signal SNR. For the last three low-SNR datasets, our procedure does not predict a compression ratio capable of achieving a correlation above 0.9, a conservative compression ratio of 2 was used for evaluation.

Dataset	N	K	K/N (%)	SNR	C_2	C_1	$M_{\min}$	$CR_{\max}$
Marmousi	8192	202.35 ± 28.72	2.5	165.45 ± 78.08	4.30	1.65	537	15
Alaska	3000	29.70 ± 5.84	1.0	74.73 ± 46.12	2.42	1.74	104	28
Hualien	90000	1353 ± 55.08	1.5	38.15 ± 2.82	1.32	1.80	4440	20
Stromboli	6000	251.93 ± 27.63	4.2	15.77 ± 2.35	0.33	6.44	2234	3
Ridgecrest	7500	330.23 ± 75.75	4.4	15.24 ± 3.82	0.30	8.65	3874	2
Noto	6000	373.16 ± 360.26	6.2	5.27 ± 4.42	/	/	/	/
Greenland	30000	1254.14 ± 1210.50	4.2	4.61 ± 1.20	/	/	/	/
Brienz	6000	749 ± 154.63	12.5	1.88 ± 0.28	/	/	/	/

construction correlation. High SNR signals are more likely to achieve high compression ratios with low reconstruction error, whereas low SNR ones require more caution. This limitation may be mitigated by applying denoising before CS compression. A secondary control is the normalized sparsity K/N. Lower K/N generally allows higher compression ratios because the signal better satisfies the sparse representation assumption of the CS algorithm. For instance, the Alaska dataset has relatively high SNR and the lowest K/N, allowing the highest compression ratio of 28 despite having been downsampled to 25 Hz. This further indicates that CS compression performance is controlled mainly by the strength of the dominant frequency components and the effective bandwidth, rather than the sampling rate itself. Downsampled signals can be further compressed if they have sufficiently high SNR and sparsity. Conversely, high sampling rate data do not necessarily yield high compression ratios if the dominant signals are weak or broadly distributed in the frequency domain. Therefore, combining denoising methods with CS may provide a promising strategy for achieving higher compression ratios in noisy DAS records.

To further evaluate the compression error of the CS algorithm, we compress and reconstruct all datasets using their estimated maximum compression ratios. For datasets whose SNR is too low for the empirical relationship to predict a reconstruction correlation above 0.9, we use a compression ratio of 2. We then randomly select one channel from each dataset and compare the original and reconstructed signals in Figure 5. The datasets for which the empirical prediction fails still achieve relatively high reconstruction correlations, indicating that our estimate provides a conservative lower

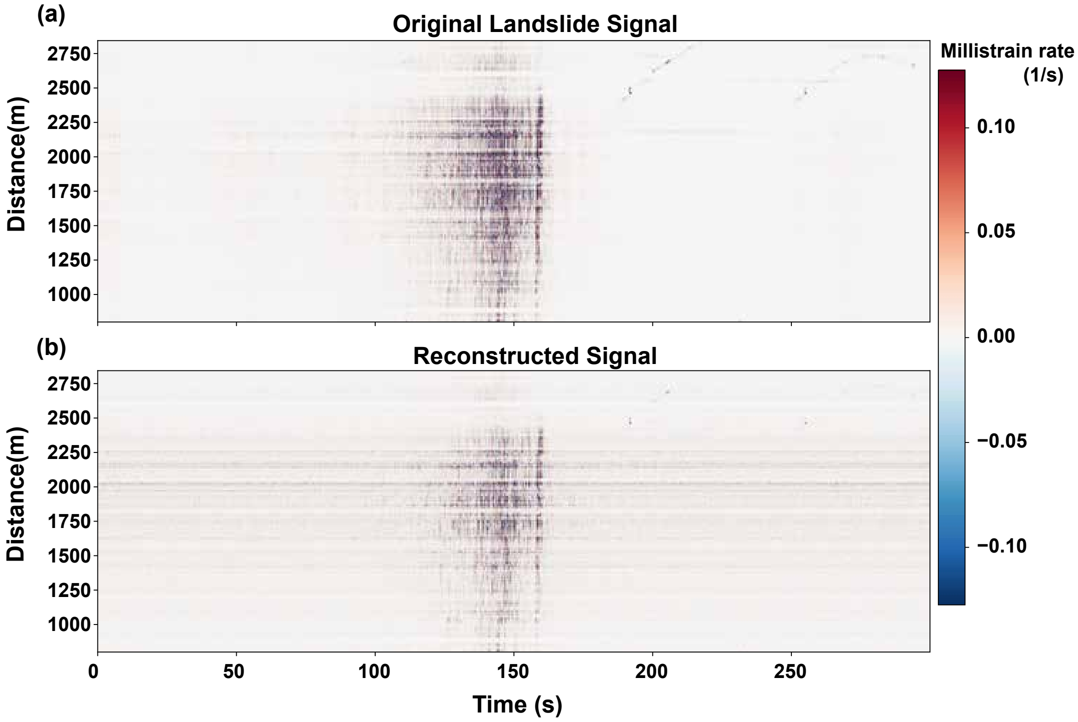

Figure 6. Evaluation of the CS algorithm using a landslide signal recorded by a DAS array in Brienz, Switzerland. The raw data have a low average SNR of 1.88, likely due to strong anthropogenic noise from the nearby cantonal road and railway line (Kang et al. 2024). The low SNR leads to a mean reconstruction correlation of only 0.6634. Nevertheless, the onset and termination of the landslide, as well as the cascading behavior of the bedrock failure, are preserved in the reconstructed data.

bound. The results show that the achievable compression ratio increases and the reconstruction error generally decreases as SNR increases. Compression errors are either noisy fluctuations or the smoothing of high-frequency variations in the original signal (Figure5 d,e). Despite these artifacts, the primary characteristics, including onset, amplitude, and decay are generally preserved. Even for the landslide example, which has the lowest reconstruction correlation of 0.7669, the event onset remains clear when all channels are plotted (Figure 6), indicating that the reconstructed data can still support downstream analyses such as detection and localization. As a result, although CS is inevitably lossy, it does not obscure the dominant signal features, and the reconstructed data remain useful for subsequent scientific analysis.

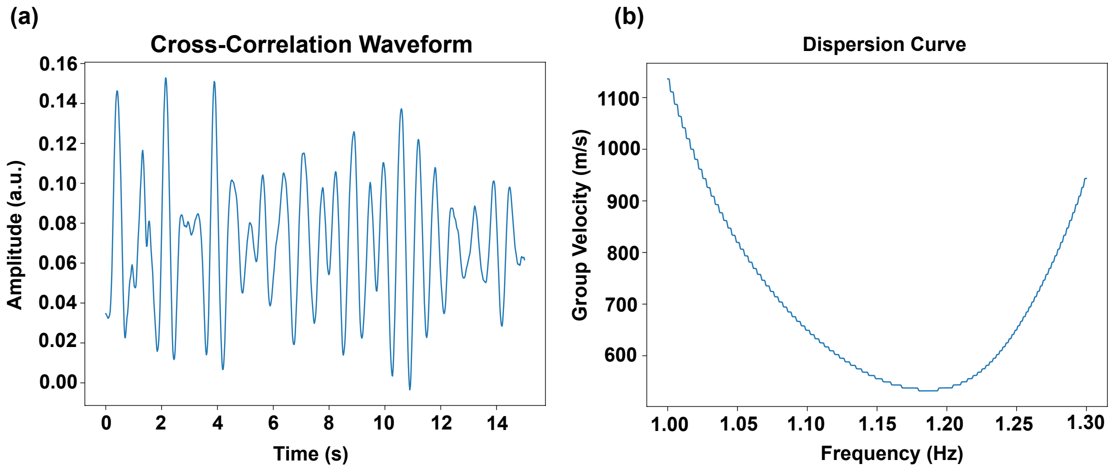

Figure 7. Ambient noise cross-correlation stacking test of DAS continuous data. (a) Cross-correlation stacking of one month of noise data, with the Rayleigh-wave interception. The primary frequency component is around 1 Hz. (b) Group velocity dispersion curve obtained using the narrow-band filtering method. The extremum corresponds to the Airy phase, indicating the frequency component with the slowest attenuation of wave energy.

4.3 Challenges in Compressing Ambient Noise Records

Reconstructing data compressed by the CS algorithm presents challenges in certain scenarios, particularly with ambient noise records. Ambient noise tomography (ANT) is a well-established seismic method that uses cross-correlation stacking of ambient noise to extract Green's functions between stations. Recent studies have demonstrated that ANT can generate high-quality seismic wave traces, enabling S-wave velocity tomography of the local subsurface using DAS data (Yang et al. 2022). This technique offers insights into regional heterogeneity and strong ground motion.

We apply ANT to 32 days of continuous data from channels 100 and 501 of the MiDAS project (March 31 to May 1). Following Bensen et al. (2007), we segment the original data hourly, apply running-absolute-average temporal normalization and spectral whitening, and perform cross-correlation stacking to extract Rayleigh waves (Figure 7a). The group velocity dispersion curve, derived from wave packet arrivals at different frequencies, shows clear positive dispersion with an Airy phase near 1.18 Hz (Figure 7b).

However, reconstructing ambient-noise data proves inefficient. On a single-core Apple M4 CPU, reconstruction takes 0.17 s per channel per second of data, more than twice the 0.08 s/(channel · s) required for event data. This is because ambient noise lacks prominent frequency components, leading to lower sparsity and requiring more reconstruction iterations. Ambient noise records are typically much longer, further increasing the computational cost. Full reconstruction of continuous noise-dominated DAS data is currently impractical on personal computers, limiting our evaluation of CS performance for such records. Future solutions might involve acceleration using large-scale GPU clusters or integration of neural networks for reconstruction (Iqbal et al. 2023). The CS algorithm is therefore more suitable for compressing, transmitting, and reconstructing DAS data for scientific studies of specific events within limited frequency bands.

5 DISCUSSION

5.1 Enhanced Spectrum Reconstruction Capability

The Fourier transform-based CS algorithm compresses data while preserving dominant frequency components. This is useful in complex environments with multiple dominant frequencies, such as bay areas where tectonic earthquakes, ocean currents, and gravity waves coexist. The CS algorithm retains frequency information, suppresses weak noise, and requires no manual intervention. Unlike other methods, it fits the spectrum to reconstruct the original signal, providing accurate spectral reconstruction.

We first test this idea using synthetic data. We generate a 2-second signal at a 1000 Hz sampling rate with two segments: a 5-Hz component in the first second and a 20-Hz component in the second, both with a Gaussian amplitude spectrum and added noise. Compared with a basic method involving 10-Hz low-pass filtering and downsampling, the CS algorithm preserves both dominant frequencies at the same compression ratio, while the simpler method loses information after 1 second (Figure 8a). The CS algorithm accurately maintains the positions, shapes, and amplitudes of the frequency domain components while providing denoising (Figure 8b). Time-frequency analysis further shows that the CS method captures temporal frequency variations (Figure 8c-e).

The tremor records provide another example. Tremors generally have low amplitudes, lack high frequency energy and do not show distinct P-wave arrivals, and are therefore usually obscured by background noise. By exploiting coherence across DAS channels, Baba et al. (2025) identified more than ten tremor events following the 2024 M7.6 Noto earthquake. In these records, ocean current signals dominate the raw data, and the tremor moveout across DAS channels becomes visible after applying the 2–10 Hz bandpass filter (Figure 9a,b). We apply the CS algorithm directly to the raw data without preprocessing. The reconstructed data remain dominated by the same low-frequency ocean-current signals before filtering, while the tremor moveout is clearly recovered after 2–10 Hz filtering (Figure 9c,d). The low amplitude tremor signals can be retained during CS compression. These results highlight the ability of CS to preserve key frequency components automatically, distinguishing it from compression methods that require predefined filtering or signal selection.

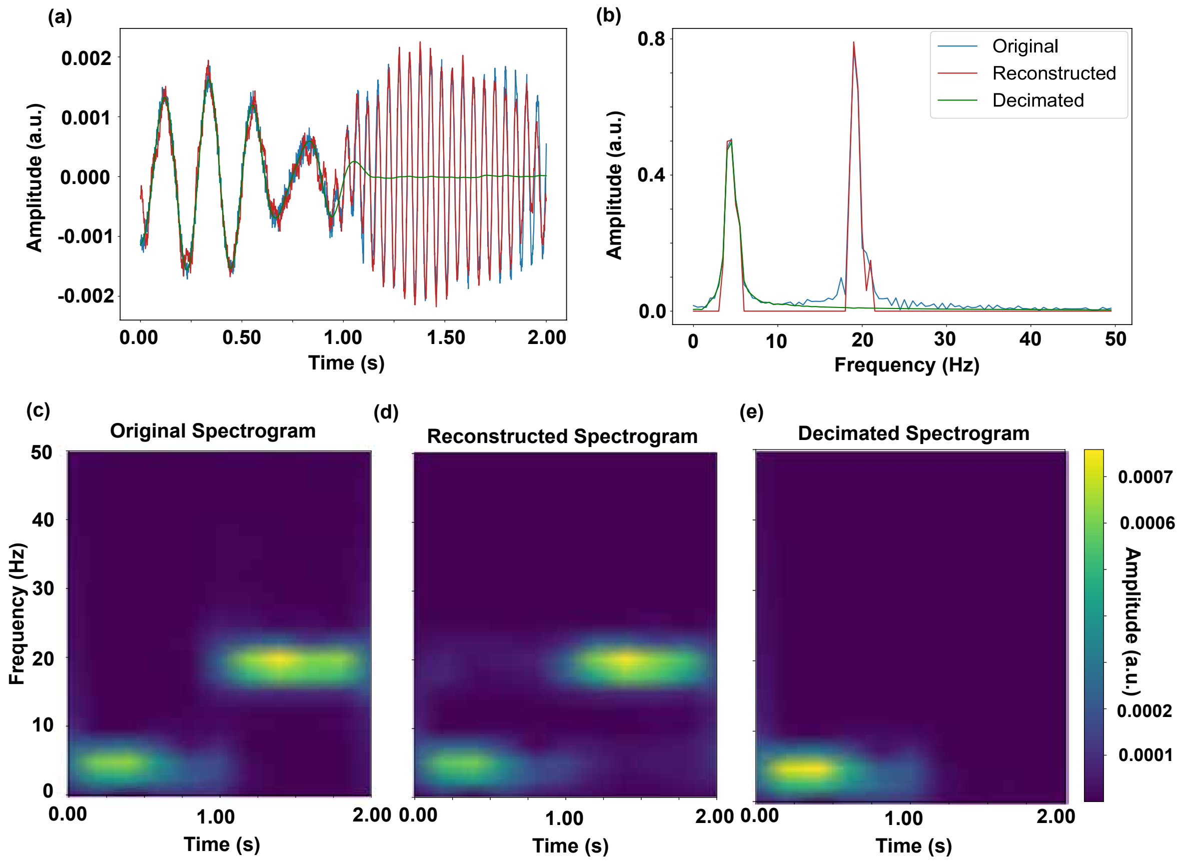

Figure 8. Spectrum reconstruction performance test. (a) Time-domain comparison of synthetic data (blue), reconstructed data (red), and decimated data (green). The reconstructed data preserve high frequencies, while the decimated data loses them after 1 s. (b) Comparison of amplitude spectra. The reconstructed data retains the two main frequencies, while the decimated data loses high-frequency content. (c) Spectrogram of synthetic data. (d) Spectrogram of reconstructed data. (e) Spectrogram of decimated data.

5.2 Compatibility with Classical Seismic Algorithms

To demonstrate that data compressed and reconstructed using the CS algorithm can still support standardized seismic processing, we analyze the time window (00:11:20-00:12:50) containing the largest Hualien aftershock. We apply the CS algorithm for compression and reconstruction and compare the phase-picking and array seismology performance of the original and reconstructed data.

First, we examine the compatibility with phase picking, a fundamental process critical to seismic monitoring, localization, and tomography. With the rapid advancement of artificial intelligence (AI), several automated phase picking models have been developed to cope with the accelerating rate of data generation (Zhu & Beroza 2019). For DAS data, the challenges arise from the two-dimensional data structure and the limited number of manually labeled phases, which complicate training AI models.

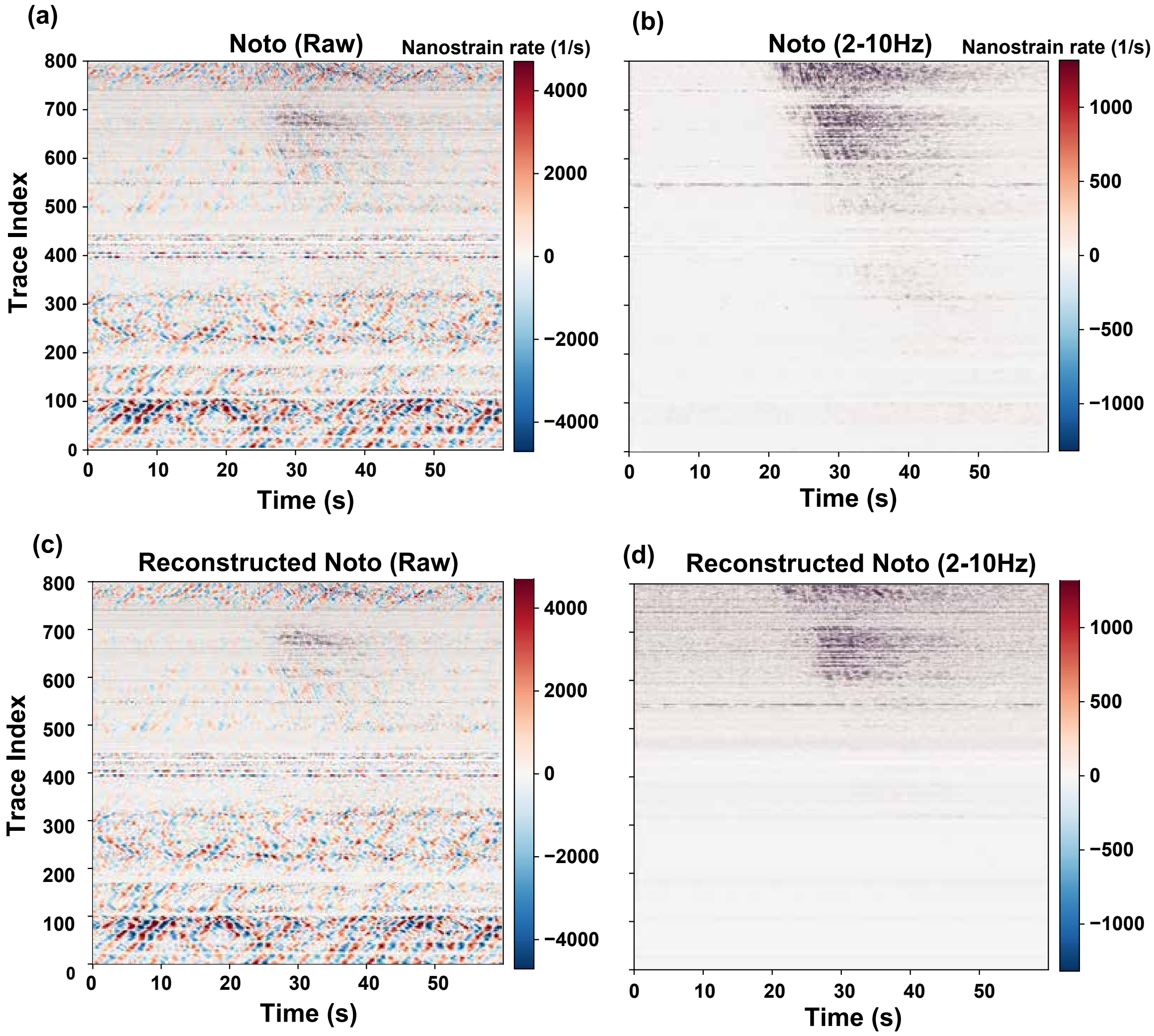

Figure 9. Evaluation of the spectrum reconstruction capability of the CS algorithm using the Noto tremor dataset. (a) Original DAS data, dominated by strong ocean current signals that obscure the tremor. (b) Original data after 2–10 Hz bandpass filtering, which removes the low-frequency ocean-current signals and reveals the tremor moveout. (c) CS-compressed and reconstructed data without filtering, showing ocean current signals that are nearly identical to those in (a). (d) Reconstructed data after 2–10 Hz bandpass filtering. The tremor moveout is clear even with slight noise.

PhaseNet-DAS, which employs a semi-supervised learning framework, addresses these issues and has been effectively applied in Long Valley, CA (Zhu et al. 2023). We use the open-source PhaseNet-DAS v1 model to pick phases from both the original and reconstructed data. Since the model was trained on a dataset sampled at 100 Hz, we downsample the data to 100 Hz for stable prediction results. The P-wave arrivals picked from the reconstructed and original signals show negligible differences (Figure 10), with the maximum error not exceeding 0.05 s. In this case, PhaseNet-DAS does not pick the

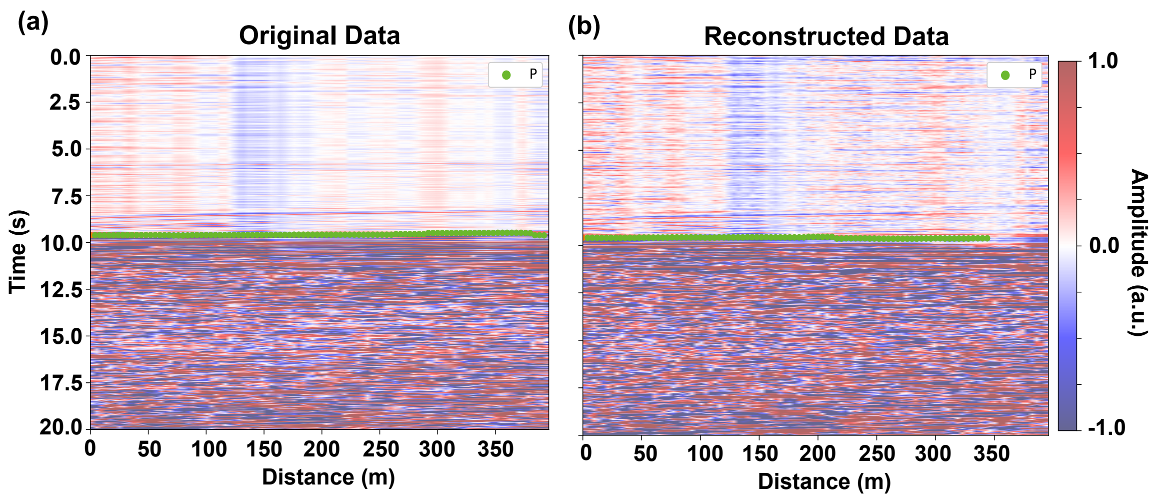

Figure 10. Testing the compatibility of PhaseNet-DAS with the CS algorithm. (a) P-wave arrivals picked by PhaseNet-DAS in the original data. (b) P-wave arrivals picked by PhaseNet-DAS in the reconstructed data, showing nearly identical results to the original data.

S-wave arrival, as the epicentral distance is only 11.69 km, causing the S-wave to be obscured by the P-wave's coda. This scenario may not have been included in the model's training set, leading to reduced generalization performance. In the future, with sufficient continuous data, we can fine-tune the model parameters to enhance phase picking performance for near-field events. This case clearly demonstrates that the reconstructed signals are well-suited for AI-based phase picking models.

Another test is to perform beamforming on CS-reconstructed data, a classical array seismology algorithm used to determine the slowness and azimuth of a given ray. Recent studies have shown that array seismology methods are applicable for DAS data because of their dense spatial sampling (van den Ende & Ampuero 2021; Ni et al. 2024b). Since our data are from a borehole array, it is sensitive only to vertical slowness η. If a reference sensor receives a a signal $s\left(t\right)$, the signal received by the i th sensor is expressed as $x_i\left(t\right) = s\left(t + z_i * \eta\right)$, where z_i is the depth of each sensor relative to the reference point. The time delay varies with different vertical slownesses, and if the vertical slowness is appropriate, the signal associated with a particular ray will be enhanced by stacking. We follow the procedure outlined in Rost & Thomas (2002) and use a sliding window to compute the vespagram, a diagram of the energy content of incoming signals as a function of slowness and time. The first seismic phase appears at 10.6 s, with a vertical slowness of 0.000332 s/m, closely matching the P-wave arrival picked by PhaseNet-DAS, which we interpret as the direct P-wave. The second phase appears at 25.3 s, with a vertical slowness of 0.000642 s/m, slightly greater than that of the P-wave, indicating a lower vertical apparent velocity, consistent with S-wave characteristics. The third phase appears at

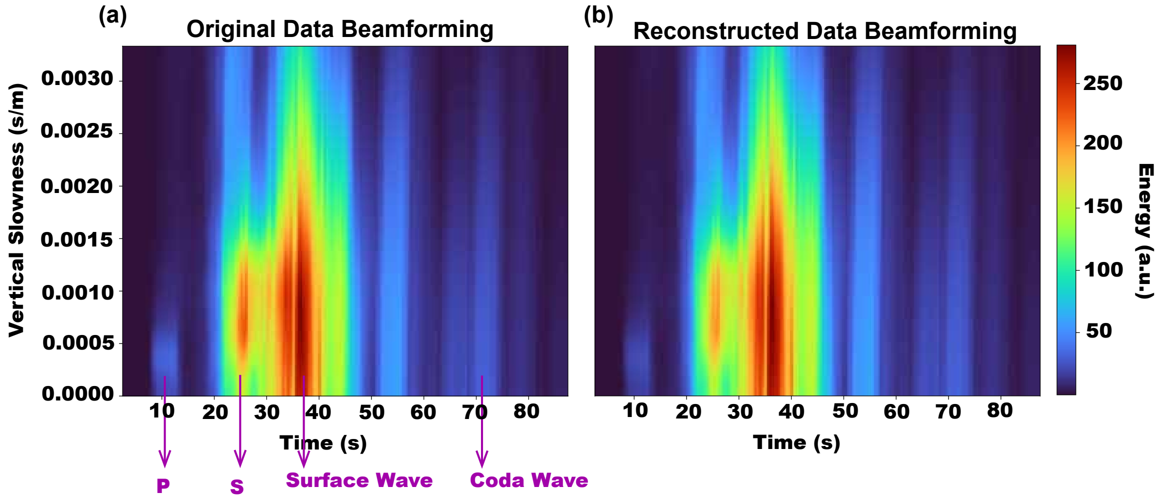

Figure 11. Testing the compatibility of array seismology methods with the CS algorithm. (a) Beamforming using original data, with colors indicating signal energy at different vertical slownesses. Prominent phases, including the direct P wave, S wave, surface wave, and coda wave are labeled. (b) Beamforming using the reconstructed data shows that all phases are also detectable.

36.5 s, and the vertical slowness at the point of maximum energy is close to 0, indicating that nearly all channels receive the signal simultaneously, consistent with surface wave propagation. After 40 s, long-duration, full-slowness signals are recorded. Full slowness implies that the rays detected by the sensors come from various incident angles, which we attribute to coda waves generated by shallow crustal scatterers. Both the original and reconstructed data produce vespagrams that clearly capture these signals (Figure 11), demonstrating the potential of the CS algorithm to preserve data suitable for array seismology processing.

5.3 Comparison with Alternative Compression Algorithms

To comprehensively evaluate the advantages, limitations, and applicability of the CS algorithm, we compare it with several established compression methods, including wavelet compression, gzip and zfp. In the wavelet approach, we use the Haar mother wavelet, and the coefficients are sorted by absolute amplitude, retaining only the largest to control the compression ratio. Gzip is a general-purpose lossless method that compresses data by leveraging repeated patterns. Zfp is a lossy compressor for numerical floating arrays that provides high compression efficiency and allows controlled reconstruction errors.

We perform compression tests on all real-world datasets. For the CS algorithm, the compression ratio is set to the estimated maximum value calculated in Section 4.2. The wavelet compression uses

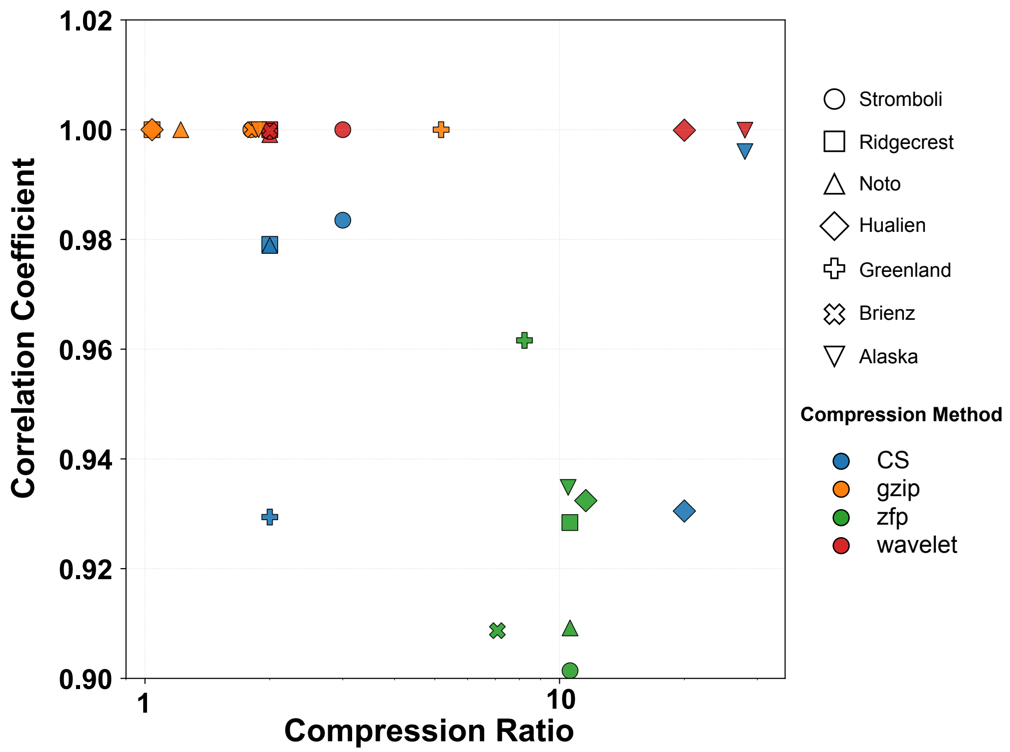

Figure 12. Comparison of different compression algorithms on real-world DAS datasets. Colors indicate different algorithms, and marker shapes indicate different datasets. The horizontal axis shows the compression ratio, with larger values indicating higher compression efficiency. The vertical axis shows the correlation coefficient between the reconstructed and original signals, with larger values indicating lower compression error.

the same compression ratios for direct comparison. Zfp and gzip can determine their achievable compression ratios automatically. Overall, CS achieves higher compression ratios than gzip but lower than zfp. In terms of reconstruction quality, CS produces lower correlations than gzip but higher than zfp (Figure 12). This indicates that CS provides a balance between compression efficiency and compression error, achieving relatively high compression ratios while maintaining sufficiently small losses for downstream analysis.

Wavelet compression is a special case. In most datasets, it achieves higher reconstruction correlations than CS at the same compression ratio, suggesting that DAS data may be sparser in the wavelet domain and highlighting the importance of choosing an appropriate basis representation. However, we do not use the wavelet transform because the frequency domain has a clearer physical meaning and allows amplitude-based algorithms, such as STA/LTA, to be migrated to the compressed domain,

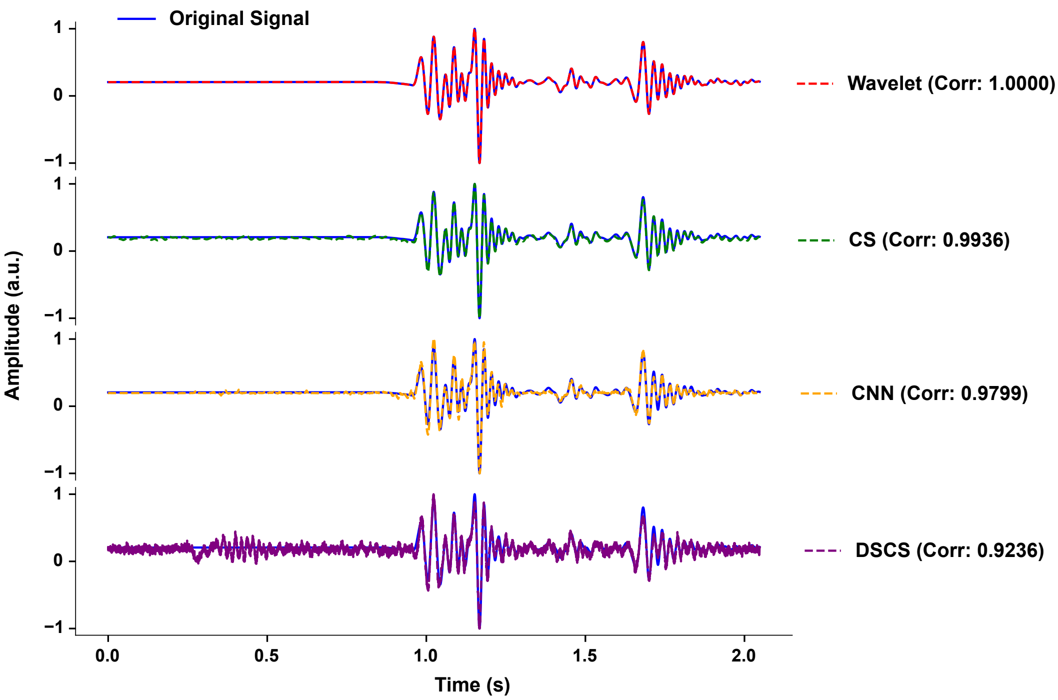

Figure 13. Comparison of different compression algorithms using synthetic data generated with the Marmousi model. From top to bottom: wavelet compression, the CS algorithm, CNN-based compression, and DSCS compression, with correlation coefficients decreasing sequentially.

which is not straightforward in the wavelet domain. Nevertheless, these results indicate that wavelet bases may be advantageous when the primary objective is to achieve higher compression ratios.

We also compare the CS algorithm with advanced artificial-intelligence-based compression methods. We first design an encoder-decoder convolutional neural network (CNN) architecture. The encoder consists of three convolutional layers with max-pooling and ReLU activation, followed by a fully connected layer that maps the input signal to a latent space whose dimensionality controls the compression ratio. The decoder is constructed symmetrically to reconstruct the signal. The second model is Deep Seismic Compressive Sensing (DSCS), a hybrid framework that combines CS with a CNN architecture (Iqbal et al. 2023). In DSCS, the CS-compressed signal is reconstructed using a neural network with encoder, decoder, and skip-connection components. This method has been reported to achieve a 64-fold compression ratio on test data.

We use the dataset from Iqbal et al. (2023), generated from the Marmousi model using a 0–90 Hz Ormsby source wavelet. The dataset contains 1,968 surface traces sampled at 4 kHz. The CS compression ratio is set to 15, following Section 4.2. For a direct comparison, the same compression ratio is used for other methods, and both the CNN and DSCS models are trained on 90% of the dataset.

As shown in Figure 13, all methods achieve high reconstruction accuracy. Wavelet and CS compression perform particularly well, yielding correlation coefficients above 0.99 while requiring simple implementations and short processing times. CNN and DSCS achieve slightly lower accuracy—the former inaccurately reconstructs small-amplitude features, while the latter introduces random noise. Both models require extensive training, and each trained architecture is strictly tied to a prescribed compression ratio. These results indicate that for high SNR data, transform-based methods such as CS and wavelet compression can achieve high compression ratios and are therefore more suitable for practical applications.

To sum up, the CS algorithm provides a balanced and flexible strategy. Relative to conventional compressors, CS achieves a favorable trade-off between compression efficiency and reconstruction error. Compared with AI-based methods, it is simpler to implement and easier to deploy, requires no training, and is not tied to a specific compression ratio. In our tests, it also achieves better reconstruction performance. Unlike wavelet compression, CS allows amplitude-based algorithms to be migrated directly to the compressed domain. Overall, CS provides a simple, interpretable, and adaptable framework for DAS compression and downstream scientific analysis.

5.4 Prospects of CS Algorithms in DAS Data

Further applications of the CS algorithm to DAS data can be explored in several directions. First, leveraging the amplitude-preserving characteristics of the frequency-based CS algorithm could allow more seismological methods to be adapted directly to the compressed domain. This approach eliminates the need for time-consuming data reconstruction and enables simultaneous compression and processing, thereby facilitating edge computing. The compressed domain STA/LTA algorithm proposed in this study demonstrates this feasibility. Further progress can be made by integrating CS with other big-data processing techniques to develop DAS into more user-friendly data products. These efforts may include CS-based on-site computation, AI-driven waveform denoising, and the use of object storage formats to enable cloud computation (Shi et al. 2025; Ni et al. 2024a). In this context, the CS algorithm can serve as a bridge linking on-site and cloud-based dataset.

6 CONCLUSIONS

This study introduces a data compression and seismic detection workflow for DAS based on compressive sensing. To the best of our knowledge, this is the first application of the CS algorithm to field DAS data to achieve promising performance. The results demonstrate that the CS algorithm achieves high compression ratios while preserving both temporal and spectral information, maintaining full compatibility with standard seismological analyses. We further derive an empirical relationship between the compression ratio and reconstruction error for DAS data, validated across multiple datasets. This finding provides a practical basis for broad application of the CS algorithm and assessing the maximum compression ratio in broader scenarios.

Importantly, the approximately amplitude-conserving property of the frequency-based CS algorithm enables the STA/LTA method to be migrated to the compressed domain, opening the possibility of edge computation and even DAS-based early warning systems. This property also motivates the transfer of other energy-based seismological algorithms into the compressed domain, avoiding time-consuming reconstruction. At this stage, we do not suggest that compressive sensing should replace conventional lossless or transform-based compression methods. Instead, it provides an alternative framework when preserving dominant physical features and enabling compressed-domain processing are prioritized.

Finally, given the noise sensitivity of the CS algorithm, an acquisition workflow combining denoising, CS compression, on-site computation, and cloud storage may provide a more efficient pathway to develop more user-friendly DAS data products.

ACKNOWLEDGMENTS

This work is supported by NSF Grant EAR-2506181. We sincerely thank Prof. Brad Lipovsky from University of Washington for his support with data resources and insight into the synthetic tests during the early stages. We sincerely thank Prof. Naveed Iqbal from King Fahd University of Petroleum and Minerals for providing the DSCS training data and source code in the Discussion section.

DATA AVAILABILITY

The earthquake catalog used in this study was developed by the Central Weather Administration of Taiwan(Central Weather Administration 2012), and the seismic data were obtained from the Taiwan MiDAS project. All data and Python scripts used for processing in this study are available on the Zenodo repository (Ma 2025a,b).

DECLARATION

The authors declare that they have no conflicts of interest relevant to this study.

REFERENCES

Allen, R. V., 1978. Automatic earthquake recognition and timing from single traces, *Bulletin of the Seismological Society of America*, **68**(5), 1521–1532, doi: 10.1785/BSSA0680051521.

Atterholt, J., 2022. Distributed acoustic sensing recorded Ridgecrest earthquake aftershocks, CaltechDATA, doi: 10.22002/D1.20038.

Atterholt, J. & Zhan, Z., 2024. Fine-scale Southern California Moho structure uncovered with distributed acoustic sensing, *Science Advances*, **10**(48), eadr3327, doi: 10.1126/sciadv.adr3327.

Baba, S., Araki, E., & Hori, T., 2025. Shallow tremors near the Nankai Trough activated after the M 7.6 Noto Peninsula earthquake, *Geophysical Research Letters*, **52**, e2025GL118973, doi: 10.1029/2025GL118973.

Bai, L., Lu, H., & Liu, Y., 2020. High-efficiency observations: Compressive sensing and recovery of seismic waveform data, *Pure and Applied Geophysics*, **177**(1), 469–485, doi: 10.1007/s00024-018-2070-z.

Baraniuk, R. G., 2007. Compressive sensing [lecture notes], *IEEE Signal Processing Magazine*, **24**(4), 118–121, doi: 10.1109/MSP.2007.4286571.

Benjumea, B., Gaite, B., Schimmel, M., Bohoyo, F., Spica, Z. J., Mancilla, F. D. L., Li, Y., Almendros, J., & Morales, J., 2024. Subsurface imaging in urban areas with ambient noise using DAS and seismometer data sets: Granada, Spain, *Journal of Geophysical Research: Solid Earth*, **129**(11), e2024JB029820, doi: 10.1029/2024JB029820.

Bensen, G. D., Ritzwoller, M. H., Barmin, M. P., Levshin, A. L., Lin, F., Moschetti, M. P., Shapiro, N. M., & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophysical Journal International*, **169**(3), 1239–1260, doi: 10.1111/j.1365-246X.2007.03374.x.

Biagioli, F., Métaxian, J.-P., Stutzmann, E., Ripepe, M., Bernard, P., Trabattoni, A., Longo, R., & Bouin, M.-P., 2024. Array analysis of seismo-volcanic activity with distributed acoustic sensing, *Geophysical Journal International*, **236**(1), 607–620, doi: 10.1093/gji/ggad427.

Bin, K., Luo, S., Zhang, X., Lin, J., & Tong, X., 2021. Compressive data gathering with generative adversarial networks for wireless geophone networks, *IEEE Geoscience and Remote Sensing Letters*, **18**(3), 558–562, doi: 10.1109/LGRS.2020.2978520.

Biondi, E., Zhu, W., Li, J., Williams, E. F., & Zhan, Z., 2023. An upper-crust lid over the Long Valley magma chamber, *Science Advances*, **9**(42), eadi9878, doi: 10.1126/sciadv.adi9878.

Cai, T. T. & Wang, L., 2011. Orthogonal matching pursuit for sparse signal recovery with noise, *IEEE Transactions on Information Theory*, **57**(7), 4680–4688, doi: 10.1109/TIT.2011.2146090.

Candès, E. J. & Tao, T., 2006. Near-optimal signal recovery from random projections: Universal encoding strategies?, *IEEE Transactions on Information Theory*, **52**(12), 5406–5425, doi: 10.1109/TIT.2006.885507.

Candès, E. J., Romberg, J. K., & Tao, T., 2006. Stable signal recovery from incomplete and inaccurate measurements, *Communications on Pure and Applied Mathematics*, **59**(8), 1207–1223, doi: 10.1002/cpa.20124.

Central Weather Administration, 2012. Central Weather Administration Seismographic Network, International Federation of Digital Seismograph Networks, doi: 10.7914/SN/T5.

Chan, C.-H., Ma, K.-F., Shyu, J. B. H., Lee, Y.-T., Wang, Y.-J., Gao, J.-C., Yen, Y.-T., & Rau, R.-J., 2020. Probabilistic seismic hazard assessment for Taiwan: TEM PSHA2020, *Earthquake Spectra*, **36**(1 Suppl.), 137–159, doi: 10.1177/8755293020951587.

Chen, Y., Saad, O. M., Chen, Y., & Savvaidis, A., 2025. Deep learning for seismic data compression in distributed acoustic sensing, *IEEE Transactions on Geoscience and Remote Sensing*, **63**, 1–14, doi: 10.1109/TGRS.2025.3526933.

Dong, B., Popescu, A., Rodríguez Tribaldos, V., Byna, S., Ajo-Franklin, J. B., & Wu, K., 2022. Real-time and post-hoc compression for data from distributed acoustic sensing, *Computers & Geosciences*, **166**, 105181, doi: 10.1016/j.cageo.2022.105181.

Donoho, D. L., 2006. Compressed sensing, *IEEE Transactions on Information Theory*, **52**(4), 1289–1306, doi: 10.1109/TIT.2006.871582.

Donoho, D. L. & Johnstone, I. M., 1994. Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, **81**(3), 425–455, doi: 10.1093/biomet/81.3.425.

Fichtner, A., Hofstede, C., Kennett, B. L. N., Svensson, A., Westhoff, J., Walter, F., Ampuero, J.-P., Cook, E., Zigone, D., Jansen, D., & Eisen, O., 2025. Hidden cascades of seismic ice stream deformation, *Science*, **387**(6736), 858–864, doi: 10.1126/science.adp8094.

Fukushima, S., Shinohara, M., Nishida, K., Takeo, A., Yamada, T., & Yomogida, K., 2022. Detailed S-wave velocity structure of sediment and crust off Sanriku, Japan by a new analysis method for distributed acoustic sensing data using a seafloor cable and seismic interferometry, *Earth, Planets and Space*, **74**(1), 92, doi: 10.1186/s40623-022-01652-z.

Gao, X., Hu, W., Dou, Z., Li, K., & Gong, X., 2022. A method on vibration positioning of Φ-OTDR system based on compressed sensing, *IEEE Sensors Journal*, **22**(16), 16422–16429, doi: 10.1109/JSEN.2022.3191863.

Geng, Z., Wu, X., Fomel, S., & Chen, Y., 2020. Relative time seislet transform, *Geophysics*, **85**(2), V223–V232, doi: 10.1190/geo2019-0212.1.

Huang, H.-H., Ma, K.-F., Wu, E.-S., Cheng, Y.-Z., Lin, C.-J., Ku, C.-S., Su, P.-L., & MiDAS Working Group, 2024. Spatiotemporal monitoring of a frequent-slip fault zone using downhole distributed acoustic sensing at the MiDAS project, in *Distributed Acoustic Sensing in Borehole Geophysics*, pp. 459–475, John Wiley & Sons, Inc., doi: 10.1002/9781394179275.ch26.

Hudson, T. S., Baird, A. F., Kendall, J. M., Kufner, S. K., Brisbourne, A. M., Smith, A. M., Butcher, A., Chalari, A., & Clarke, A., 2021. Distributed acoustic sensing (DAS) for natural microseismicity studies: A case study from Antarctica, *Journal of Geophysical Research: Solid Earth*, **126**(7), e2020JB021493, doi: 10.1029/2020JB021493.

Iqbal, N., Masood, M., Alfarraj, M., & Waheed, U. B., 2023. Deep seismic CS: A deep learning assisted compressive sensing for seismic data, *IEEE Transactions on Geoscience and Remote Sensing*, **61**, 1–9, doi: 10.1109/TGRS.2023.3289917.

Issah, A. H. S. & Martin, E. R., 2024. Impact of lossy compression errors on passive seismic data analyses, *Seismological Research Letters*, **95**(3), 1675–1686, doi: 10.1785/0220230314.

Jousset, P., Currenti, G., Schwarz, B., Chalari, A., Tilmann, F., Reinsch, T., Zuccarello, L., Privitera, E., & Krawczyk, C. M., 2022. Fibre optic distributed acoustic sensing of volcanic events, *Nature Communications*, **13**(1), 1753, doi: 10.1038/s41467-022-29184-w.

Kang, J., Walter, F., Paitz, P., Aichele, J., Edme, P., Meier, L., & Fichtner, A., 2024. Automatic monitoring of rock-slope failures using distributed acoustic sensing and semi-supervised learning, *Geophysical Research Letters*, **51**, e2024GL110672, doi: 10.1029/2024GL110672.

Kuo, Y.-T., Wang, Y., Hollingsworth, J., Huang, S.-Y., Chuang, R. Y., Lu, C.-H., Hsu, Y.-C., Tung, H., Yen, J.-Y., & Chang, C.-P., 2019. Shallow fault rupture of the Milun Fault in the 2018 M 6.4 Hualien earthquake: A high-resolution approach from optical correlation of Pléiades satellite imagery, *Seismological Research Letters*, **90**(1), 97–107, doi: 10.1785/0220180227.

Li, J., Kim, T., Lapusta, N., Biondi, E., & Zhan, Z., 2023. The break of earthquake asperities imaged by distributed acoustic sensing, *Nature*, **620**(7975), 800–806, doi: 10.1038/s41586-023-06227-w.

Lin, C.-R., von Specht, S., Ma, K.-F., Ohrnberger, M., & Cotton, F., 2025. Analysis of saturation effects of distributed acoustic sensing and detection on signal clipping for strong motions, *Geophysical Journal International*, **241**(2), 971–985, doi: 10.1093/gji/ggaf089.

Lin, Y.-Y., Huang, H.-H., Hsieh, M.-C., Kuo, L.-W., Chen, C.-C., Chang, M.-H., Chen, T.-Y., Chen, H.-A., Brown, D., & Wang, S.-J., 2026. Lightning-induced implosive microseismic events at depth detected by downhole distributed acoustic sensing in Taiwan, *Seismological Research Letters*, doi: 10.1785/0220250432, Advance online publication.

Lindsey, N. J. & Martin, E. R., 2021. Fiber-optic seismology, *Annual Review of Earth and Planetary Sciences*, **49**(1), 309–336, doi: 10.1146/annurev-earth-072420-065213.

Lindsey, N. J., Dawe, T. C., & Ajo-Franklin, J. B., 2019. Illuminating seafloor faults and ocean dynamics with dark fiber distributed acoustic sensing, *Science*, **366**(6469), 1103–1107, doi: 10.1126/science.aay5881.

Liu, X.-Y., Zhu, Y., Kong, L., Liu, C., Gu, Y., Vasilakos, A. V., & Wu, M.-Y., 2015. CDC: Compressive data collection for wireless sensor networks, *IEEE Transactions on Parallel and Distributed Systems*, **26**(8), 2188–2197, doi: 10.1109/TPDS.2014.2345257.

Lo, Y.-C., Yue, H., Sun, J., Zhao, L., & Li, M., 2019. The 2018 M_w 6.4 Hualien earthquake: Dynamic slip partitioning reveals the spatial transition from mountain building to subduction, *Earth and Planetary Science Letters*, **524**, 115729, doi: 10.1016/j.epsl.2019.115729.

Ma, K.-F., von Specht, S., Kuo, L.-W., Huang, H.-H., Lin, C.-R., Lin, C.-J., Ku, C.-S., Wu, E.-S., Wang, C.-Y., Chang, W.-Y., & Jousset, P., 2024. Broad-band strain amplification in an asymmetric fault zone observed from borehole optical fiber and core, *Communications Earth & Environment*, **5**(1), 402, doi: 10.1038/s43247-024-01558-6.

Ma, Y., 2025a. Python scripts for compression and earthquake detection based on compressive sensing, Zenodo, doi: 10.5281/zenodo.15242168.

Ma, Y., 2025b. Selected original and compressed data from the Taiwan MiDAS project, Zenodo, doi:

10.5281/zenodo.15241981.

Manchanda, R. & Sharma, K., 2020. A review of reconstruction algorithms in compressive sensing, in *2020 International Conference on Advances in Computing, Communication & Materials (ICACCM)*, pp. 322–325, IEEE, Dehradun, India, doi: 10.1109/ICACCM50413.2020.9212838.

Marcellin, M. W., Gormish, M. J., Bilgin, A., & Boliek, M. P., 2000. An overview of JPEG-2000, in *Proceedings of the 2000 Data Compression Conference (DCC 2000)*, pp. 523–541, IEEE Computer Society, Snowbird, UT, USA, doi: 10.1109/DCC.2000.838192.

Mata Flores, D., Sladen, A., Ampuero, J.-P., Mercerat, E. D., & Rivet, D., 2023. Monitoring deep sea currents with seafloor distributed acoustic sensing, *Earth and Space Science*, **10**(6), e2022EA002723, doi: 10.1029/2022EA002723.

Mun, S. & Fowler, J. E., 2011. Residual reconstruction for block-based compressed sensing of video, in *2011 Data Compression Conference (DCC)*, pp. 183–192, IEEE, Snowbird, UT, USA, doi: 10.1109/DCC.2011.25.

Ni, Y., Denolle, M. A., Fatland, R., Alterman, N., Lipovsky, B. P., & Knuth, F., 2024a. An object storage for distributed acoustic sensing, *Seismological Research Letters*, **95**(1), 499–511, doi: 10.1785/0220230172.

Ni, Y., Denolle, M. A., Shi, Q., Lipovsky, B. P., Pan, S., & Kutz, J. N., 2024b. Wavefield reconstruction of distributed acoustic sensing: Lossy compression, wavefield separation, and edge computing, *Journal of Geophysical Research: Machine Learning and Computation*, **1**(3), e2024JH000247, doi: 10.1029/2024JH000247.

Qu, S., Chang, J., Cong, Z., Chen, H., & Qin, Z., 2019a. Data compression and SNR enhancement with compressive sensing method in phase-sensitive OTDR, *Optics Communications*, **433**, 97–103, doi: 10.1016/j.optcom.2018.09.064.

Qu, S., Liu, Z., Xu, Y., Cong, Z., Zhou, D., Wang, S., Li, Z., Wang, H., & Qin, Z., 2019b. Phase sensitive optical time domain reflectometry based on compressive sensing, *Journal of Lightwave Technology*, **37**(23), 5766–5772, doi: 10.1109/JLT.2019.2938789.

Rani, M., Dhok, S. B., & Deshmukh, R. B., 2018. A systematic review of compressive sensing: Concepts, implementations and applications, *IEEE Access*, **6**, 4875–4894, doi: 10.1109/ACCESS.2018.2793851.

Rost, S. & Thomas, C., 2002. Array seismology: Methods and applications, *Reviews of Geophysics*, **40**(3), 1008, doi: 10.1029/2000RG000100.

Sebai, D., Zouaoui, M., & Ghorbel, F., 2024. Seismic data compression: An overview, *Multimedia Systems*, **30**(1), 38, doi: 10.1007/s00530-023-01233-4.

Seguí, A., Ugalde, A., Fichtner, A., Ventosa, S., & Morros, J. R., 2026. DASPack: Controlled data compression for distributed acoustic sensing, *Geophysical Journal International*, **244**(1), 1–16, doi: 10.1093/gji/ggaf397.

Shen, X., Wu, H., Zhu, K., Liu, H., Li, Y., Zheng, H., Li, J., Shao, L., Shum, P. P., & Lu, C., 2024. Fast and storage-optimized compressed domain vibration detection and classification for distributed acoustic sensing, *Journal of Lightwave Technology*, **42**(1), 493–499, doi: 10.1109/JLT.2023.3308173.

Shi, Q., Denolle, M. A., Ni, Y., Williams, E. F., & You, N., 2025. Denoising offshore distributed acoustic sensing using masked auto-encoders to enhance earthquake detection, *Journal of Geophysical Research:*

Solid Earth, **130**(2), e2024JB029728, doi: 10.1029/2024JB029728.

Shragge, J., Yang, J., Issa, N., Roelens, M., Dentith, M., & Schediwy, S., 2021. Low-frequency ambient distributed acoustic sensing (DAS): Case study from Perth, Australia, *Geophysical Journal International*, **226**(1), 564–581, doi: 10.1093/gji/ggab111.

Shyu, J. B. H., Chuang, Y.-R., Chen, Y.-L., Lee, Y.-R., & Cheng, C.-T., 2016. A new on-land seismogenic structure source database from the Taiwan Earthquake Model (TEM) project for seismic hazard analysis of Taiwan, *Terrestrial, Atmospheric and Oceanic Sciences*, **27**(3), 311, doi: 10.3319/TAO.2015.11.27.02(TEM).

Spanias, A., Jonsson, S., & Stearns, S., 1991. Transform methods for seismic data compression, *IEEE Transactions on Geoscience and Remote Sensing*, **29**(3), 407–416, doi: 10.1109/36.79431.

Strumia, C., Trabattoni, A., Supino, M., Baillet, M., Rivet, D., & Festa, G., 2024. Sensing optical fibers for earthquake source characterization using raw DAS records, *Journal of Geophysical Research: Solid Earth*, **129**(1), e2023JB027860, doi: 10.1029/2023JB027860.

Tang, W., Cao, H., Duan, J., & Wang, Y.-P., 2011. A compressed sensing based approach for subtyping of leukemia from gene expression data, *Journal of Bioinformatics and Computational Biology*, **9**(5), 631–645, doi: 10.1142/S0219720011005689.

Tillmann, A. M. & Pfetsch, M. E., 2014. The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing, *IEEE Transactions on Information Theory*, **60**(2), 1248–1259, doi: 10.1109/TIT.2013.2290112.

Trabattoni, A., Festa, G., Longo, R., Bernard, P., Plantier, G., Zollo, A., & Strollo, A., 2022. Microseismicity monitoring and site characterization with distributed acoustic sensing (DAS): The case of the Irpinia Fault System (Southern Italy), *Journal of Geophysical Research: Solid Earth*, **127**(9), e2022JB024529, doi: 10.1029/2022JB024529.

van den Ende, M. P. A. & Ampuero, J.-P., 2021. Evaluating seismic beamforming capabilities of distributed acoustic sensing arrays, *Solid Earth*, **12**(4), 915–934, doi: 10.5194/se-12-915-2021.

Wang, Y.-J., Lee, Y.-T., Chan, C.-H., & Ma, K.-F., 2016. An investigation of the reliability of the Taiwan Earthquake Model PSHA2015, *Seismological Research Letters*, **87**(6), 1287–1298, doi: 10.1785/0220160085.

Wien, M., 2015. *High efficiency video coding: Coding tools and specification*, Signals and Communication Technology, Springer Berlin Heidelberg, Berlin, Heidelberg, doi: 10.1007/978-3-662-44276-0.

Williams, E. F., Fernández-Ruiz, M. R., Magalhaes, R., Vanthillo, R., Zhan, Z., González-Herráez, M., & Martins, H. F., 2019. Distributed sensing of microseisms and teleseisms with submarine dark fibers, *Nature Communications*, **10**(1), 5778, doi: 10.1038/s41467-019-13262-7.

Williams, E. F., Zhan, Z., Martins, H. F., Fernández-Ruiz, M. R., Martín-López, S., González-Herráez, M., & Callies, J., 2022. Surface gravity wave interferometry and ocean current monitoring with ocean-bottom DAS, *Journal of Geophysical Research: Oceans*, **127**(5), e2021JC018375, doi: 10.1029/2021JC018375.

Wu, C., Chen, J., Yan, G., & He, Z., 2023. A lossless data compression method for distributed acoustic sensors, in *Optical Fiber Sensors (OFS) 2023*, p. W4.20, Optica Publishing Group, doi: 10.1364/OFS.2023.W4.20.

Wu, R., Huang, W., & Chen, D.-R., 2013. The exact support recovery of sparse signals with noise via orthogonal matching pursuit, *IEEE Signal Processing Letters*, **20**(4), 403–406, doi: 10.1109/LSP.2012.2233734.

Xiao, Z., Yuan, X., Zhang, Y., Xu, S., Li, Z., & Huang, Y., 2023. Frequency response enhancement of Φ-OTDR using interval-sweeping pulse equivalent sampling based on compressed sensing, *Journal of Lightwave Technology*, **41**(2), 768–776, doi: 10.1109/JLT.2022.3217346.

Yang, Y., Atterholt, J. W., Shen, Z., Muir, J. B., Williams, E. F., & Zhan, Z., 2022. Sub-kilometer correlation between near-surface structure and ground motion measured with distributed acoustic sensing, *Geophysical Research Letters*, **49**(1), e2021GL096503, doi: 10.1029/2021GL096503.

Yang, Y., Zhan, Z., Karrenbach, M., Reid-McLaughlin, A., Biondi, E., Wiens, D. A., & Aster, R. C., 2024. Characterizing South Pole firn structure with fiber optic sensing, *Geophysical Research Letters*, **51**(13), e2024GL109183, doi: 10.1029/2024GL109183.

Yen, J.-Y., Lu, C.-H., Dorsey, R. J., Kuo-Chen, H., Chang, C.-P., Wang, C.-C., Chuang, R. Y., Kuo, Y.-T., Chiu, C.-Y., Chang, Y.-H., Bovenga, F., & Chang, W.-Y., 2019. Insights into seismogenic deformation during the 2018 Hualien, Taiwan, earthquake sequence from InSAR, GPS, and modeling, *Seismological Research Letters*, **90**(1), 78–87, doi: 10.1785/0220180228.

Yin, J., Zhu, W., Li, J., Biondi, E., Miao, Y., Spica, Z. J., Viens, L., Shinohara, M., Ide, S., Mochizuki, K., Husker, A. L., & Zhan, Z., 2023. Earthquake magnitude with DAS: A transferable data-based scaling relation, *Geophysical Research Letters*, **50**(10), e2023GL103045, doi: 10.1029/2023GL103045.

Yu, C., Zhan, Z., Lindsey, N. J., Ajo-Franklin, J. B., & Robertson, M., 2019. The potential of DAS in teleseismic studies: Insights from the Goldstone experiment, *Geophysical Research Letters*, **46**(3), 1320–1328, doi: 10.1029/2018GL081195.

Yu, P., Zhu, T., Marone, C., Elsworth, D., & Yu, M., 2024. DASEventNet: AI-based microseismic detection on distributed acoustic sensing data from the Utah FORGE Well 16A(78)-32 hydraulic stimulation, *Journal of Geophysical Research: Solid Earth*, **129**(9), e2024JB029102, doi: 10.1029/2024JB029102.

Zhan, Z., 2020. Distributed acoustic sensing turns fiber-optic cables into sensitive seismic antennas, *Seismological Research Letters*, **91**(1), 1–15, doi: 10.1785/0220190112.

Zheng, A., Yu, X., Bai, C., Xu, W., Qian, J., Zhang, W., & Chen, X., 2024. Thrust-dominated unilateral rupture of a blind listric fault associated with the 2024 Hualien earthquake, *Scientific Reports*, **14**(1), 31496, doi: 10.1038/s41598-024-82971-x.

Zhu, W. & Beroza, G. C., 2019. PhaseNet: A deep-neural-network-based seismic arrival time picking method, *Geophysical Journal International*, **216**(1), 261–273, doi: 10.1093/gji/ggy423.

Zhu, W., Biondi, E., Li, J., Yin, J., Ross, Z. E., & Zhan, Z., 2023. Seismic arrival-time picking on distributed acoustic sensing data using semi-supervised learning, *Nature Communications*, **14**(1), 8192, doi: 10.1038/s41467-023-43355-3.

Appendix

We derive the relationship between the correlation coefficient and the signal-to-noise ratio (SNR) of the original signal. Let the original signal be

$$S = S_K + e$$

and the reconstructed signal be

$$\hat{S} = S_K + r$$

where S_K denotes the true signal, e is the observation noise, and r is the reconstruction noise. According to Candès & Tao (2006), the reconstruction error satisfies

$$\|r\| \leq C_2 \|e\|$$

The correlation coefficient is defined as

$$\rho = \frac{\langle S, \hat{S} \rangle}{\|S\| \, \|\hat{S}\|}$$

Because

$$\|\hat{S} - S\| = \|(\hat{S} - S_K) + (S_K - S)\| \leq \|\hat{S} - S_K\| + \|S_K - S\| = \|r\| + \|e\| \leq (1 + C_2)\|e\|$$

squaring both sides yields

$$\|\hat{S} - S\|^2 \leq (1 + C_2)^2 \|e\|^2$$

which is equivalent to

$$\|\hat{S}\|^2 + \|S\|^2 - 2\langle S, \hat{S} \rangle \leq (1 + C_2)^2 \|e\|^2$$

From this inequality, and substituting into the definition of the correlation coefficient, we obtain

$$\rho \geq \frac{1}{2} \left(\frac{\|S\|}{\|\hat{S}\|} + \frac{\|\hat{S}\|}{\|S\|} \right) - \frac{(1 + C_2)^2 \|e\|^2}{2\|S\| \, \|\hat{S}\|}$$

Using the basic inequality $\|S\|/\|\hat{S}\| + \|\hat{S}\|/\|S\| \geq 2$, we obtain

$$\rho \geq 1 - \frac{(1 + C_2)^2 \|e\|^2}{2\|S\|^2}$$

Equality holds when $\|S\| = \|\hat{S}\|$. Since

$$\|S\|^2 = \|S_K + e\|^2 \geq \|S_K\|^2 + \|e\|^2 - 2\|S_K\|\|e\|$$

we further derive

$$\rho \geq 1 - \frac{(1 + C_2)^2}{2(\mathrm{SNR} - 2\sqrt{\mathrm{SNR}} + 1)}$$

where

$$\mathrm{SNR} = \frac{\|S_K\|^2}{\|e\|^2}$$

is the signal-to-noise ratio of the original signal.