

DASPack: Controlled Data Compression for Distributed Acoustic Sensing

Aleix Seguí¹, Arantza Ugalde², Andreas Fichtner¹, Sergi Ventosa²,
Josep Ramon Morros³

¹ *ETH Zürich, Zürich, Switzerland*

² *Institute of Marine Sciences, CSIC, Barcelona, Spain*

³ *Universitat Politècnica de Catalunya (UPC), Barcelona, Spain*

SUMMARY

We present DASPack, a high-performance, open-source compression tool specifically designed for distributed acoustic sensing (DAS) data. As DAS becomes a key technology for real-time, high-density, and long-range monitoring in fields such as geophysics, infrastructure surveillance, and environmental sensing, the volume of collected data is rapidly increasing. Large-scale DAS deployments already generate hundreds of terabytes and are expected to increase in the coming years, making long-term storage a major challenge. Despite this urgent need, few compression methods have proven to be both practical and scalable in real-world scenarios. DASPack is a fully operational solution that consistently outperforms existing techniques for DAS data. It enables both controlled lossy and lossless compression by allowing users to choose the maximum absolute difference per datum between the original and compressed data. The compression pipeline combines wavelet transforms, linear predictive coding, and entropy coding to optimise efficiency. Our method achieves up to $3\times$ file size reductions for strain and strain rate data in lossless mode across diverse datasets. In lossy mode, compression improves to $6\times$ with near-perfect sig-

nal fidelity, and up to $10\times$ is reached with acceptable signal degradation. It delivers fast throughput ($100\text{--}200\text{ MB s}^{-1}$ using a single-thread and up to 750 MB s^{-1} using 8-threads), enabling real-time deployment even under high data rates. We validated its performance on 15 datasets from a variety of acquisition environments, demonstrating its speed, robustness, and broad applicability. DASPack provides a practical foundation for long-term, sustainable DAS data management in large-scale monitoring networks.

Key words: Distributed acoustic sensing – Image processing – Wavelet transform – Time-series analysis – Probability distributions

1 INTRODUCTION

Distributed acoustic sensing (DAS), turning standard fibre-optic cables into dense arrays of strain and strain rate sensors, is a widely used technique for high-density and large-scale monitoring (e.g., Hartog 2017; Zhan 2020; Lindsey & Martin 2021). DAS has grown in popularity across geophysics, seismology, and engineering. While early applications included infrastructure monitoring (e.g., Owen et al. 2012; Hill 2015), DAS has been successfully applied in borehole seismic imaging (e.g., Mateeva et al. 2014; Daley et al. 2016) and near-surface exploration and monitoring (e.g., Daley et al. 2013; Martin et al. 2017; Tuinstra et al. 2024). More recently, it has opened new opportunities for seismic observation in settings where conventional instruments are difficult to deploy, such as urban areas (e.g., Lindsey et al. 2017; Biondi et al. 2017; Ajo-Franklin et al. 2019; Spica et al. 2020; Smolinski et al. 2024), ocean floors (e.g., Lior et al. 2021; Ugalde et al. 2021; Igel et al. 2024), active volcanoes (e.g., Jousset et al. 2022; Klaasen et al. 2022, 2023), and the cryosphere (e.g., Booth et al. 2020; Walter et al. 2020a; Hudson et al. 2021; Fichtner et al. 2025). DAS has also enabled the detection of dynamic processes in unstable slopes (e.g., Ouellet et al. 2024; Kang et al. 2024).

Yet, the scientific gains come with a concerning data management challenge: for example, a single interrogator recording 5,000 channels at 1 kHz using 4-byte floating-points produces nearly 1.5 TB per day, and multi-interrogator campaigns can exceed hundreds of terabytes per month; a number that is likely to increase in the near future. The large data generation rates also entail impractical data transfer times. In many cases, users resort to physically transporting storage disks, which becomes increasingly unfeasible as the projects expand.

This work is published at *Geophys. J. Int.*. Please visit the original publication at

<https://doi.org/10.1093/gji/ggaf397>.

Long-term archiving initiatives are similarly limited and face expensive storage challenges (NSF SAGE Data Services 2024).

We propose using data compression as an effective solution to address these issues. In particular, we study a controlled data compression approach, where data can be reconstructed after rounding at a desired decimal value. While lossy compression can achieve greater size reduction, it may discard critical information, limiting the value of the data for precise scientific analyses. Hence, a controlled compression method, including lossless mode, is imperative for applications that demand complete data integrity (Seguí et al. 2024; Bozzi 2025).

Existing approaches often compress each channel independently. Dong et al. (2022) proposed a lossless integer compression method, H5TurboPFor, that provides both speed and efficiency by combining a preprocessing step with a general one-dimensional integer compressor. Wu et al. (2023) employed a 2-D redundancy exploitation strategy using one-step channel differences, achieving approximately a 50% reduction in data size. General-purpose dictionary-based compression algorithms, such as GZIP and Zstd, are also widely used, but are known to typically offer lower compression efficiency (Dong et al. 2022). ZFP (Lindstrom 2014) is another method used in the DAS community, which focuses on floating points and supports both lossless and lossy compression of multidimensional arrays, making it suitable for DAS applications (Issah & Martin 2024). We evaluate their performance in our results section.

Given that DAS data naturally forms a 2-D structure (time and distance, where each channel corresponds to a specific position along the fibre), image compression approaches are also appealing. However, their direct application is challenging: the time and distance domains in DAS data exhibit highly unbalanced correlations, unlike the patterns found in natural images. Moreover, compression speeds optimised for image data can be too slow for DAS applications, which typically involve much larger data volumes. In the previous example, 5,000 channels at 1 kHz can generate data at a rate about 20 MB s^{-1} . Even so, methods like JPEG2000 (Taubman & Marcellin 2012), based on wavelet transforms and advanced entropy coding, have been explored in various fields for 2-D data compression. Additionally, wavelets have been known to be effective for compression of seismic data (Vassiliou & Wickerhouser 1997; Sebai et al. 2024). Therefore, we include JPEG2000 in our comparative study.

Recent studies have explored deep-learning-based compression for DAS. Ni et al. (2024) used a lightweight model to achieve reasonable wavefield reconstruction from $5\times$ subsampling. Chen et al. (2025) achieved up to $8\times$ lossy compression using a vision transformer on 10-second DAS tiles, which can be further enhanced through downsampling techniques to

achieve greater reductions in data size. However, learned compression methods currently face significant limitations. These include a lack of generalisability across datasets, inability to guarantee explicit error bounds, and the tendency to introduce data-driven reconstruction artifacts that can inadvertently regenerate patterns present in the training data. Moreover, such models are computationally intensive, often requiring resources that prevent real-time deployment, especially on CPU-based systems. Only lightweight variants with moderate compression power can operate at reasonable speed. As a result, these approaches remain impractical for many real-world scenarios in their current form.

In this paper, we present DASPack, a method that combines a user-defined data degradation step with a lossless compression pipeline to achieve effective compression. The lossless compression pipeline begins with wavelet decomposition, which separates the signals in frequency and wavenumber (i.e., spatial frequency) components. A linear prediction filter is then applied to exploit redundancy in both the temporal and spatial dimensions. This results in a decorrelated, low-variance signal that is well-suited for effective entropy coding, which encodes the data into the shortest possible bit sequences based on symbol probability distribution. All steps in the pipeline are fully reversible, allowing to compress losslessly if desired.

Our approach seeks to boost compression performance compared to the most widely used generalist methods and specialised alternatives, while maintaining fast processing speed, to be able to handle large DAS data volumes seamlessly. We provide a fast, real-time, and fully open-source implementation of this pipeline under the name DASPack. The software is developed in Rust (Klabnik & Nichols 2019), a programming language designed to build reliable and efficient software. This choice allows for fine grain control over computational resources while ensuring memory safety, making DASPack both fast and robust. Furthermore, DASPack can be accessed directly from Python, and the data is stored in HDF5 format.

2 DATA

Datasets from diverse sources were compiled for a thorough benchmarking of the compression methods. Recorded data types include strain (measured in nanostrain ($\text{n}\epsilon$), equivalent to nanometres per metre or nm m^{-1}) and strain rate (measured in nanostrain per second ($\text{n}\epsilon \text{ s}^{-1}$), equivalent to nanometres per metre per second or $\text{nm m}^{-1} \text{ s}^{-1}$). Table 1 summarises the datasets used in this study. Datasets 1-8 were recorded using a Silixa iDAS interrogator, storing data with 16-bit integers. Dataset 9 was recorded with an OptaSense ODH3 unit storing 32-bit integers, while datasets 10-15 were collected using an Aragon Photonics HDAS interrogator and stored 32-bit floating-points. The datasets span a wide range of environments,

Table 1. Datasets used in the study. An index (Id.) is assigned for referencing through the text and figures. Datasets 1-8 recorded nanostrain rate ($\text{nm m}^{-1} \text{s}^{-1}$) data while datasets 9-15 recorded nanostrain (nm m^{-1}). For published datasets, the corresponding journal publication is provided in the Reference column. Three datasets (Ids. 4, 7 and 12) do not have a journal publication: Id. 4 was acquired in November 2023 during a volcanic crisis in Iceland, using a ~ 8 km long telecommunication cable extending from the Blue Lagoon to the town of Grindavík. Id. 7 was acquired in April 2022 using a 1 km long fibre-optic cable deployed in an old irrigation tunnel near Guía de Isora, which extends into the mountain towards the peak of El Teide. Id. 12 was acquired between February and May 2022 along ~ 60 km of the Balalink submarine telecommunications cable, connecting Valencia with Palma de Mallorca.

Id.	Location	Environment	Reference	Sampling Rate [Hz]	Channel Spacing (Gauge Length) [m]	Num. of Channels
1	VdIS (Switzerland)	Landslides	Aichele et al. (2025)	400	2 (8)	2176
2	Athens (Greece)	Urban	Tian et al. (2022)	200	2 (8)	7424
3	Rhonegletscher (Switz.)	Glacier	Walter et al. (2020b)	500	4 (8)	576
4	Blue Lagoon (Iceland)	Volcanic	–	500	4 (8)	512
5	Susch (Switzerland)	Landslides	Edme et al. (2023)	100	2 (8)	5120
6	Fagradalsfjall (Iceland)	Volcanic	Klaasen et al. (2021b)	400	2 (8)	2496
7	Tenerife (Spain)	Volcanic	–	200	2 (8)	1280
8	Zurisee (Switzerland)	Lake	Rychen et al. (2023)	10 000	1 (8)	2496
9	Mt. Meager (Canada)	Volcanic	Klaasen et al. (2021a)	1 000	8 (7)	375
10	Alboran Sea (Spain)	Submarine	Ugalde et al. (2023a)	100	10 (10)	5984
11	Gran Canaria (Spain)	Submarine	Ugalde et al. (2023b)	50	10 (10)	7008
12	Valencia (Spain)	Submarine	–	250	10 (10)	5983
13	Vinaroz (Spain)	Submarine	Villasenor et al. (2023)	100	10 (10)	3488
14	Tenerife (Spain)	Volcanic	Ugalde et al. (2021)	100	10 (10)	2176
15	La Palma (Spain)	Volcanic	Villasenor et al. (2021)	100	10 (10)	832

including urban settings, landslide-prone areas, glaciers, volcanic zones, lakes, and offshore regions. This diversity ensures that DASPack is evaluated under realistic and varied signal conditions. The datasets also differ in key acquisition parameters such as signal type (strain and strain rate), data type (integers and floating-points), sampling rate (ranging from 50 Hz to 10,000 Hz), and channel spacing. The number of channels defines the distance extent and varies from a few hundred to above 7,000, allowing to test the scalability of the compression pipeline.

This work is published at *Geophys. J. Int.*. Please visit the original publication at

<https://doi.org/10.1093/gji/ggaf397>.

3 METHODS

3.1 DASPack Overview

DASPack compresses DAS records in two decoupled stages that separate scientific requirements (how much error is acceptable) from engineering optimisation (how to store the data with minimal bits). The stages are:

- (i) *Controlled Degradation*. Data amplitudes are first mapped from their raw representation (integers or floating points) to signed integers by rounding the data to a user-defined resolution. This mapping fixes the absolute reconstruction error so the user can choose from mathematically lossless storage to more aggressive, but explicitly quantified, data reduction.
- (ii) *Lossless Compression*. The integer array is passed through a sequence of perfectly invertible transforms that progressively lower its entropy and a final coding step that creates an optimised binary file. Each transform is exactly reversible, so the only potential information loss is the user-specified rounding at the first step.

3.2 Controlled degradation

Lossy compression techniques used in image and audio processing reduce data size by discarding perceptually insignificant information. These methods are tailored to human perception and often rely on subjective quality metrics. In scientific applications subjectivity is unsuitable, and the reference low seismic background noise model (Peterson 1993), that could potentially be used in seismometer and accelerometer recordings as an alternative, is not appropriate for strain and strain rate DAS recordings. For this reason, DASPack adopts a principled, user-defined approach to degradation based on fixed-accuracy quantisation.

In DASPack, controlled degradation is implemented as a single, explicit rounding step prior to lossless compression. This fixed-accuracy approach maps the original data to signed integers, effectively redefining the resolution of the analogue-to-digital conversion post-acquisition,

$$d_{\text{rounded}} = \text{round} \left(\frac{d_{\text{orig}}}{\Delta} \right) \cdot \Delta, \quad (1)$$

where d_{orig} is the raw datum and Δ is the chosen step size, leading to a maximum error introduced by compression of $\frac{\Delta}{2}$. This quantisation ensures that the reconstruction error is strictly bounded and interpretable.

Unlike filtering or decimation, quantisation preserves the original signal bandwidth, maintaining both temporal and spatial resolution. For smaller Δ , the rounding error tends to be

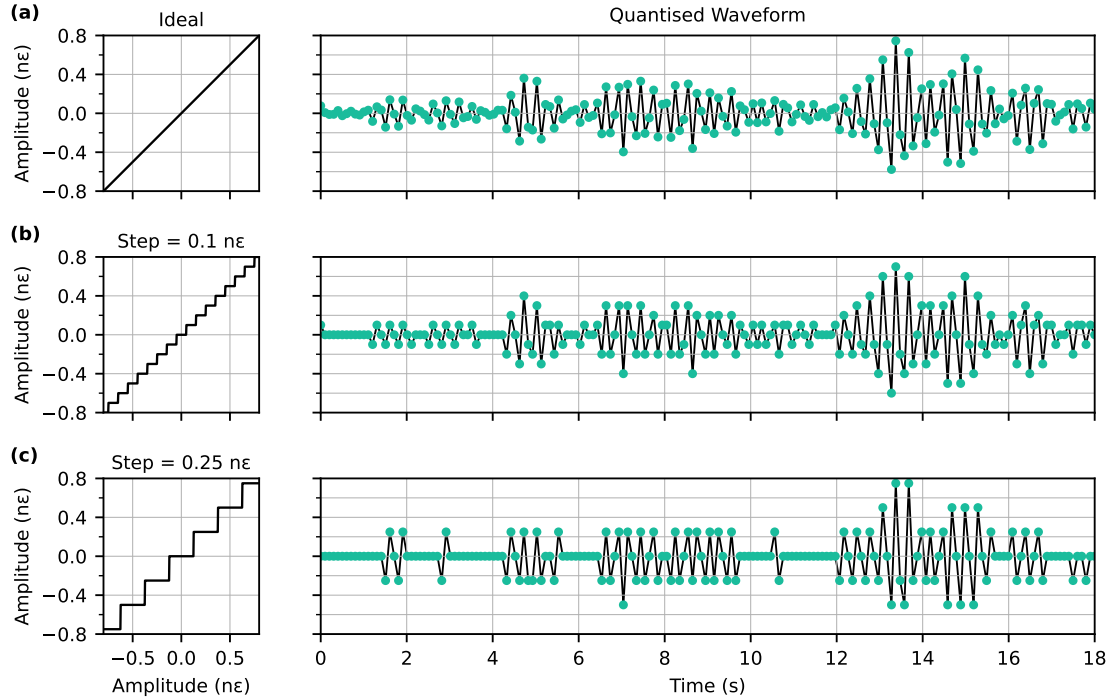


Figure 1. Controlled degradation of the amplitude in nanostrain ($\text{n}\epsilon$) at different quantisation steps. The data points are represented as coloured dots, and are joined via a black line. The continuous, original signal is represented in (a), (b) shows a quantisation step of $0.1 \text{ n}\epsilon$ and (c) is $0.25 \text{ n}\epsilon$. The waveform is extracted from dataset 10 (Alboran Sea).

uniform and uncorrelated across time and channels, resembling additive white noise (Widrow & Kollár 2008).

Fig. 1 exemplifies the effect on an earthquake waveform. The degradation removes all signals that have an amplitude smaller than the quantisation step. We observe that, for this particular example, at $0.1 \text{ n}\epsilon$ the very low amplitude signal is diminished, and at a very aggressive step of $0.25 \text{ n}\epsilon$, significant information is lost.

3.3 Compression Pipeline

DASPack's main contribution is the compression pipeline that integrates several complementary techniques, each serving a different role to achieve effective compression. The presented method is designed for integer arithmetic and the degradation step is responsible for converting data into tractable integers.

The proposed compression pipeline is depicted in Fig. 2, which shows a reversible process, including both compression and decompression steps. The steps are illustrated using an ex-

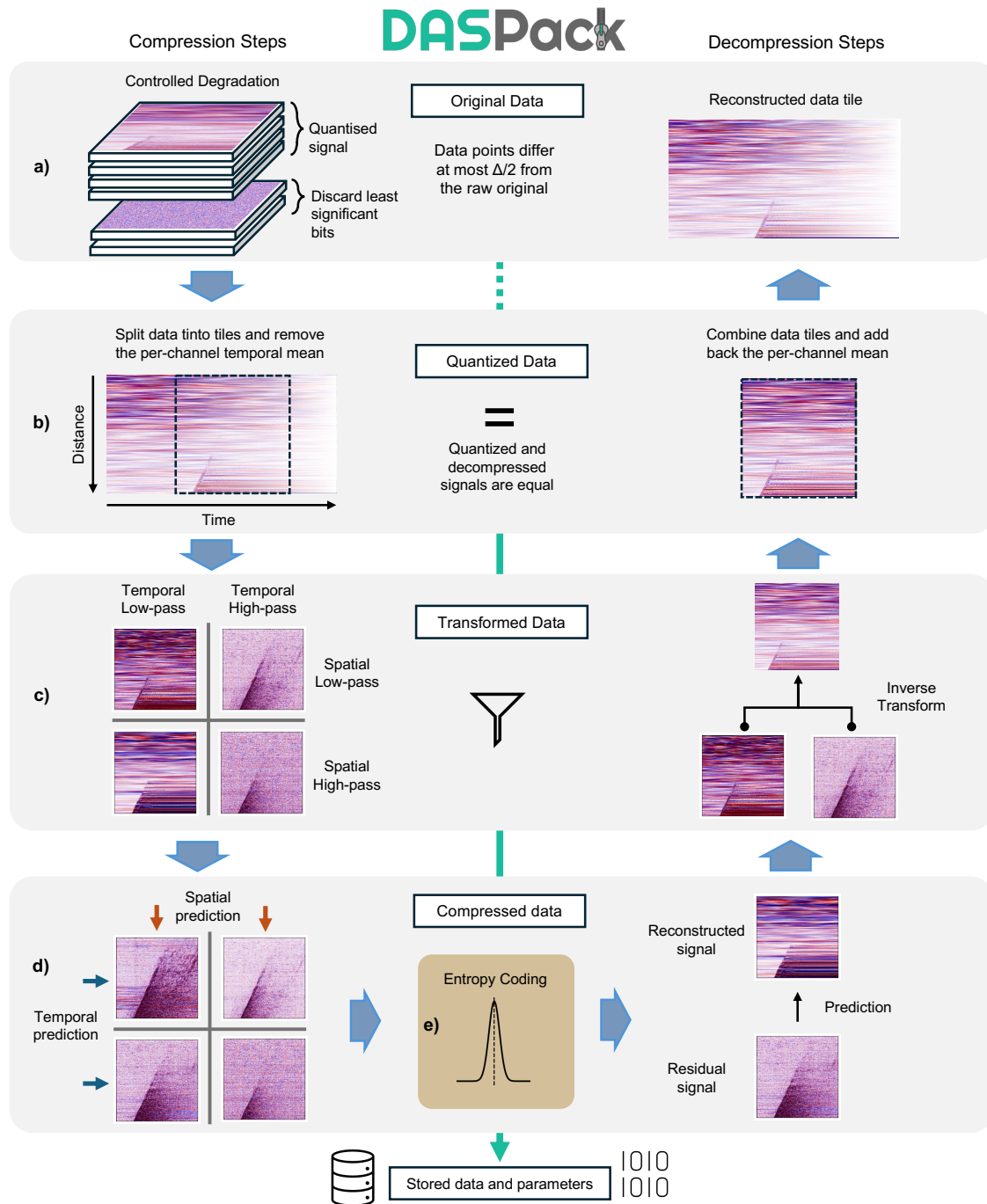


Figure 2. Compression and decompression pipelines. (a) Data is transformed to integers and optionally quantised via a user-defined quantisation step. (b) Data is first split into non-overlapping tiles that are processed separately and the temporal mean is removed per-channel. (c) A wavelet transform is optionally applied in both (time and space) dimensions. (d) A linear prediction filter is optionally applied with a fixed number of parameters in both dimensions in every wavelet subband. (e) Data is entropy coded to obtain a bit sequence. Step (a) is irreversible but tuneable; steps (b), (c), (d) and (e) are all reversible, ensuring lossless compression.

ample from a submarine fibre-optic cable from Vinaroz (dataset 13) that is deployed at very shallow depth (less than 60 m) and contains a seismic event of magnitude $mbLg$ 1.6 occurred on 20 July 2023 at 11:49:43 UTC. The stored data is also indicated, consisting of compressed 2-D arrays and side information, such as the predictive model coefficients. Below, we detail each step of the DASPack compression pipeline:

(a) First, the raw data is converted to integers according to the desired absolute accuracy (Fig. 2a), defined as the maximum absolute difference between the compressed and original data. This step becomes irreversible (lossy) if the original data is in floating-point format, or if the quantisation step exceeds the resolution of the raw integer data. The same degradation is applied over all input data, avoiding discontinuities or jumps.

(b) Next, the data is divided into non-overlapping windows (or tiles), which are processed independently (Fig. 2b). The optimal size of each tile depends on various data characteristics, such as temporal and spatial resolution or signal entropy. This design also allows for parallel processing of data. To improve the effectiveness of subsequent steps, the temporal mean is subtracted from each channel and stored as side information.

(c) Then, a discrete wavelet transform is optionally applied along both the time and distance dimensions (Fig. 2c). The wavelet transform decomposes data in several frequency and wavenumber bands along the time and distance dimensions, respectively, without increasing the total number of samples and allowing afterward a perfect reconstruction of the original signal (Jacques et al. 2011).

(d) A linear predictive coding (LPC) model is computed separately for each row (temporal dimension) and each column (spatial dimension) of the low-pass wavelet subbands (Fig. 2d). The LPC prediction captures the dominant low-frequency components of the signal, which typically account for most of its energy. As a result, the residual (the difference between the prediction and the actual signal) has lower variance, improving the effectiveness of subsequent step (entropy coding). The optimal, discretised prediction coefficients are applied during compression and stored as side information to ensure exact reconstruction during decompression.

(e) Finally, the residual signal is then converted into a compact bitstream using entropy coding and stored on disk, along with the associated optimal parameters (Fig. 2e). The decompression algorithm reverses the above steps. Integer operations ensure a mathematically exact reconstruction up until degradation (back from (e) to (b)). The only difference between the decompressed data and the original is the controlled degradation in step (a).

Following this summary, we now describe the different parts in more detail.

This work is published at *Geophys. J. Int.*. Please visit the original publication at

<https://doi.org/10.1093/gji/ggaf397>.

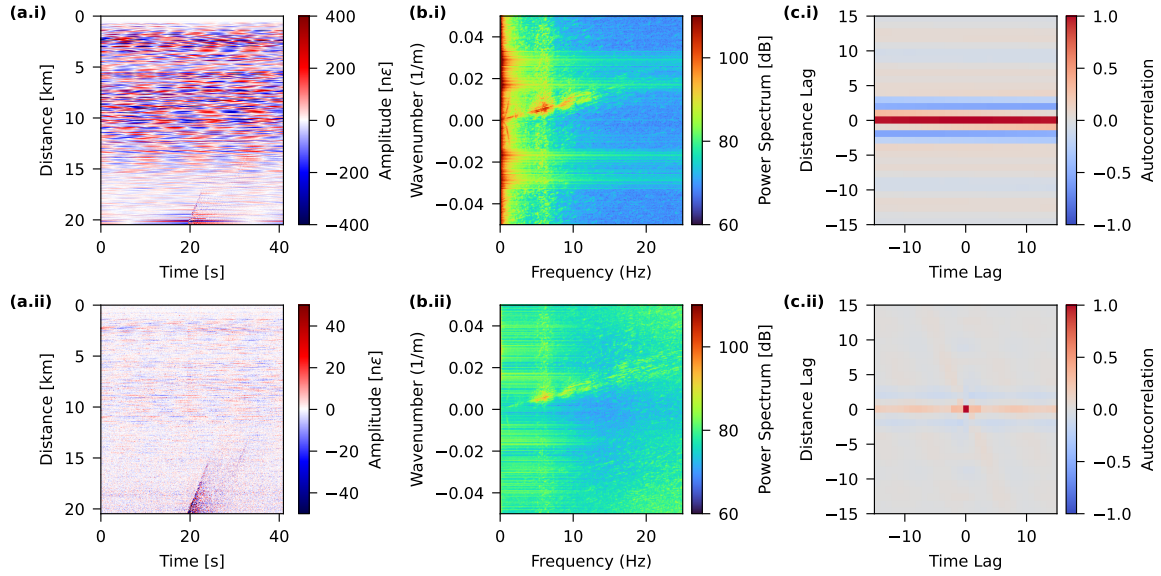


Figure 3. (i) Raw original and (ii) LPC-filtered data example, showing strain from the dataset 13 (Vinaroz), see Table 1. Column (a) shows the strain values, (b) the power spectrum in the f - k transform and (c) the 2-D autocorrelation function.

3.4 Discrete Wavelet Transform

A wavelet basis is a set of functions used to decompose signals into components that capture both frequency and location information (Mallat 2009). To enable perfect reconstruction of a signal from its components, a complete orthogonal (or biorthogonal) basis of wavelet functions is required. Discrete wavelets act a carefully designed filter bank that iteratively decompose a signal into low-frequency (approximation) and high-frequency (detail) components, without increasing the total number of samples. We employ the LGT 5/3 wavelet (Le Gall & Tabatabai 1988), which is a biorthogonal wavelet commonly used in image compression, particularly in the JPEG2000 standard (Taubman & Marcellin 2012). This wavelet is implemented through the lifting scheme (Daubechies & Sweldens 1998), as done in JPEG2000, for computational efficiency and exact integer arithmetic.

3.5 Linear Predictive Coding

DAS signals exhibit strong temporal and spatial correlations that can be leveraged to improve compression. One effective method for exploiting these redundancies is linear predictive coding (LPC), which is conceptually equivalent to a linear prediction-error filter. In LPC, each signal is decomposed into two components: a predictable part that can be approximated as a linear combination of past values, and a residual part that captures the unpredicted variations. This

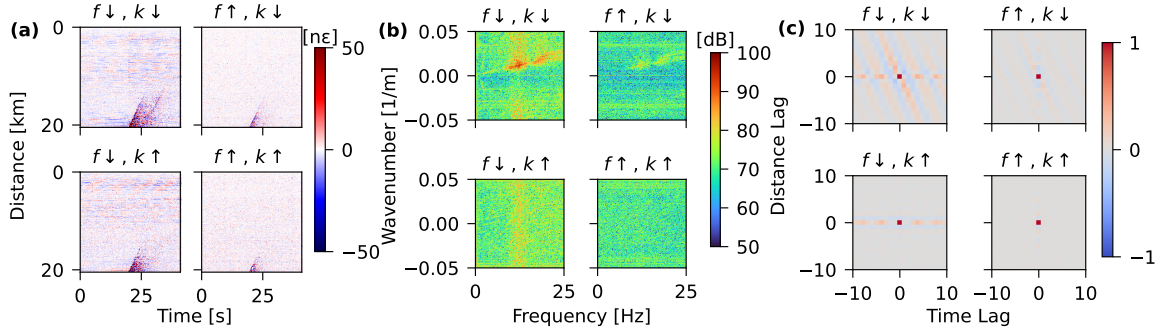


Figure 4. Example of wavelet-transformed and LPC-filtered strain data from dataset 13 (Vinaroz). For every subfigure, the four subbands obtained after the applying the wavelet transform in both dimensions are shown. Left plots correspond to frequency low-pass and right plots, to frequency high-pass. Top plots are wavenumber low-pass and bottom plots, wavenumber high-pass (a) Strain values. (b) f - k power spectrum. (c) 2-D autocorrelation function.

decomposition is reversible via integer arithmetic, making it suitable for lossless compression. While it does not attempt to model the full physical complexity of DAS signals, it provides a compact representation that enhances compressibility. Specifically, for each signal $c_i(t)$ we define:

$$c_i(t) = \sum_{k=1}^K a_k c_i(t - k). \quad (2)$$

The coefficients $\{a_1, a_2, \dots, a_K\}$ are chosen to minimise the mean squared error between the actual sample $c_i(t)$ and its linear prediction. We compute the optimal coefficients using the Levinson-Durbin recursion algorithm (Durbin 1960), which leverages the Toeplitz form of the autocorrelation matrix estimated using the unbiased correlation estimator. The mean removal performed in the preprocessing step improves the model's fit by eliminating low-frequency bias in the signal. In practice, a separate LPC filter is first fitted and applied to each channel along the time dimension; the same procedure is then applied along the spatial dimension to the resulting residuals.

Fig. 3 depicts the effect of LPC on DAS data, using the example event from Vinaroz presented in section 3.3. The top row shows the raw strain data, its frequency-wavenumber (f - k) transform, and its autocorrelation function. The bottom row shows the corresponding outputs after LPC filtering. By comparing the plots in panel (i), we observe how large-amplitude, long-period components are captured by the linear filter. In panel (ii) the low-frequency features are largely removed, and in panel (iii), the decorrelation effect is evident, with the autocorrelation significantly reduced in the LPC-filtered signal.

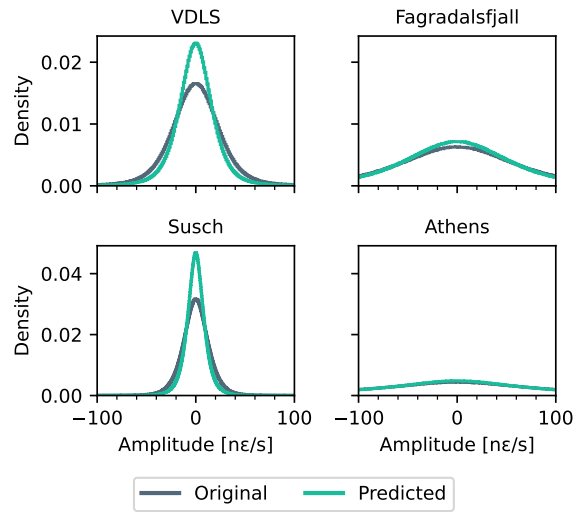


Figure 5. Density histograms of raw original, and predicted (wavelet-transformed and LPC-filtered) data, for datasets 1, 2, 5 and 8 (see Table 1). The variance is reduced when the models are able to predict the signals. The lower the variance, the more compressible the data source is. The lower variance in the predicted data reveals how the processing contributes to better compression.

To demonstrate the effect of LPC within the pipeline, Fig. 4 shows the four wavelet subbands for the same event, obtained after applying the wavelet transform in both the time and space dimensions, followed by LPC filtering. Notably, the low-frequency subbands exhibit stronger correlation, while the high-frequency components appear more uniform and less correlated.

Surface gravity-wave signals observed in shallow ocean-bottom DAS recordings are typically very strong (see, e.g., Romanowicz et al., 2023; Shi et al., 2025), and can even obscure signals from nearby moderate earthquakes. However, the statistical properties of these waves tend to evolve slowly over time, and their velocity dispersion, under the linear approximation, is primarily controlled by ocean depth. These characteristics allow LPC to effectively track the signal evolution using only a few coefficients, significantly reducing the residual energy that needs to be stored after prediction.

3.6 Entropy Coding

Entropy coding is a lossless data compression technique that removes redundancy from a signal by assigning shorter code-words to more probable symbols. It can be applied to any stream of data, regardless of its origin. The goal is to approach the theoretical limit of compressibility as defined by the Shannon’s source coding theorem (Shannon 1948), which relates directly to

the entropy of the signal. Entropy coding is a core component of high-performing lossless and lossy compression schemes.

A fundamental concept in entropy coding is variable-length coding, where symbols are represented by bit-strings of varying lengths. By assigning shorter codewords to more frequently occurring symbols (and longer codewords to less frequent ones), a more efficient coding (i.e., better compression) can be achieved.

A widely used method for entropy coding is arithmetic coding (Witten et al. 1987). Arithmetic coding encodes the entire message as a single number between 0 and 1, allowing fine granularity in probability modelling and achieving near-optimal compression for signals with fixed distributions.

Examining the residuals obtained through our processing reveals a Gaussian-like distribution (see Fig. 5). For a (finely) quantised Gaussian random variable with variance σ^2 , its entropy can be approximated by the differential entropy $H \approx \frac{1}{2} \log(2\pi e \sigma^2)$ (Cover & Thomas 2005). Fig. 5 also shows how the distribution of LPC-filtered data exhibit lower variance in several example datasets, effectively improving compression performance.

DASPack leverages an existing implementation of arithmetic coding by Bamler (2022). A Gaussian distribution is fit for each one-dimensional channel, and the corresponding mean and variance are stored as parameters. To efficiently encode tail values and outliers, we use k -exponential Golomb codes, which provide a compact, variable-length representation for large integers (Richardson 2010).

4 RESULTS

4.1 Compression efficiency and speed

We evaluate different compression algorithms by measuring the processing time and the compressed file size over 5-min excerpts randomly selected within each of the 15 datasets in Table 1. We repeat this process for different quantisation steps, ranging from lossless mode to moderate degradation.

Compression efficiency is evaluated in terms of the compression factor, defined as the ratio: $c_f = \frac{\text{Original Size}}{\text{Compressed Size}}$. This metric indicates the number of times the compressed dataset fits within the original size. It can also be interpreted as a measure of storage savings and as a proxy for file transfer time speed-up.

Figs. 6 and 7 present a comparative analysis of the trade-off between compression speed (throughput in MB s^{-1}) and efficiency (compression factor) for five compression methods

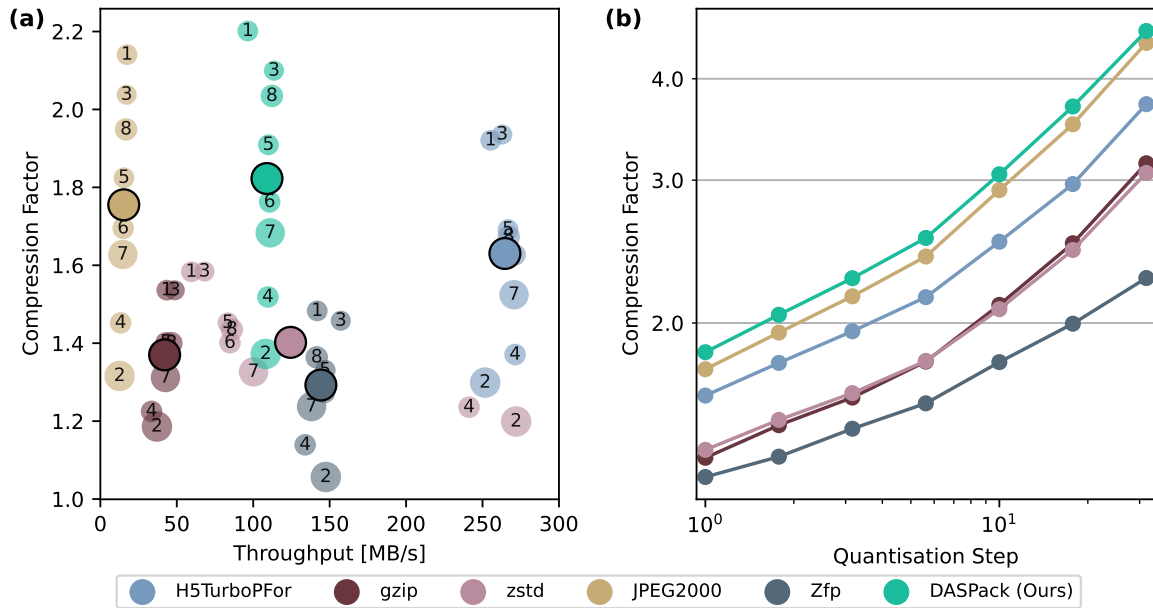


Figure 6. Compression factor of different methods for strain rate datasets (1-8). (a) Compression against throughput at lossless compression mode. The size of the points indicates the variance of the numbered dataset. The circle without a number is the centroid of the points, which is plotted to illustrate the trends. (b) Average compression across datasets at different quantisation steps. Our method, DASPack, beats all other methods in compression efficiency for all datasets, and competes in compression speed with the fastest methods.

(gzip, zstd, JPEG2000, H5TurboPFor, Zfp and DASPack) at different accuracies. Experiments were run in a conventional desktop computer (Apple M1 Pro CPU) in single-thread mode. Fig. 6 focuses on strain rate datasets 1-8, all recorded as 16-bit integers, while Fig. 7 focuses on strain datasets (9-15), recorded as 32-bit integers or floating-points. Because H5TurboPFor only works on 16-bit data, 32-bit integers are interpreted as two 16-bit values for this particular method. We fixed DASPack's parameters to square tiles of 2000 samples per side and two LPC coefficients, and implemented the steps using 32-bit integer operations. Strain data generally spans a higher dynamic range than strain rate due to the presence of high-amplitude, long period signals, hence requiring more bits to store raw data. Yet, for this same reason, strain data exhibit higher temporal correlation, hence achieving higher compression ratios.

Figures 6a and 7a reveal the performance of different methods for each individual dataset at a given fixed-accuracy (lossless for strain rate, $\Delta = 10^{-1.5}$ nε for strain). Overall, gzip and zstd achieve moderate compression factors but exhibit relatively low throughput, generally below 100 MB/s. H5TurboPFor and Zfp prioritise speed, often exceeding 200 MB/s, and H5TurboPFor maintains high compression factors for strain rate data. Zfp performs poorly in

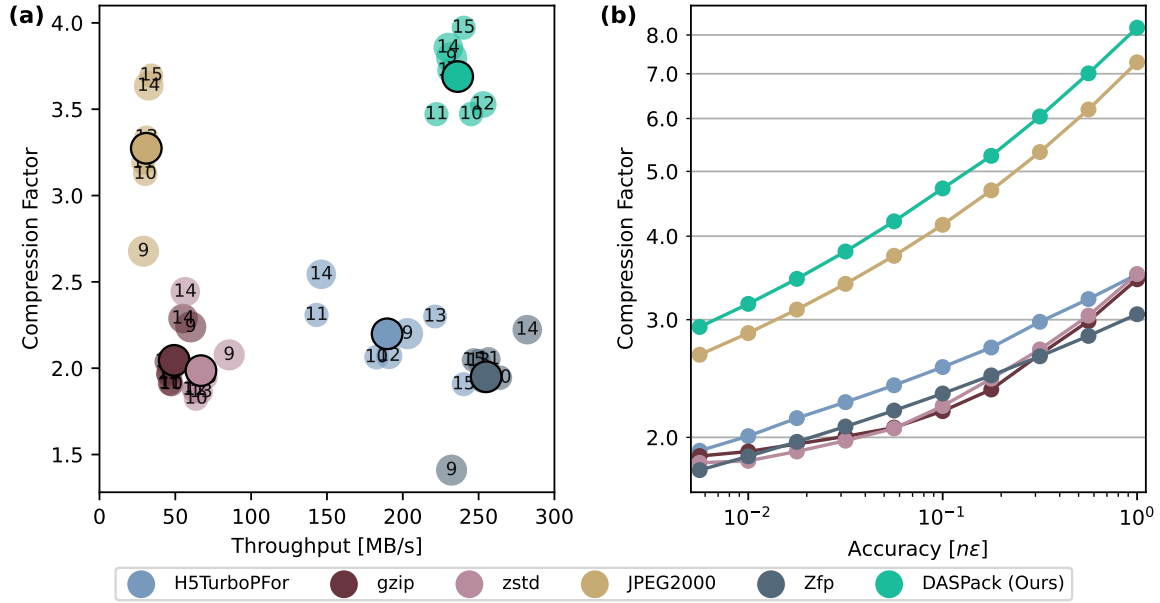


Figure 7. Compression factor of different methods for strain datasets (9-15). (a) Compression against throughput at a quantisation step $\Delta = 10^{-1.5} n\epsilon$. The size of the points indicates the variance of the numbered dataset. The circle without a number is the centroid of the points, which is plotted to illustrate the trends. (b) Average compression across datasets at different nanostrain accuracies. Our method, DASPack, beats all other methods in compression efficiency for all datasets, and is also the fastest in compression speed for strain data.

strain rate because it is a method optimised for floating-points; but the performance is only moderately better for strain data. JPEG2000 achieves higher compression factor, but is the slowest method, achieving 10-20 MB s⁻¹. DASPack, by contrast, offers both high throughput (up to 250 MB s⁻¹ for strain, 125 MB s⁻¹ for strain rate) and the strongest compression performance. It is important to note that compression factors are dataset-dependent and should not be taken as absolute numbers. While different datasets may yield higher or lower compression factors, the consistent trend is that DASPack outperforms alternative methods across all cases considered. DASPack consistently occupies a favourable region of the plot, offering high compression ratios at speeds that match or exceed leading alternatives. This makes it a strong choice for applications requiring both data reduction and fast, scalable performance.

Figures 6b and 7b demonstrate how the compression factor evolves as accuracy constraints (the quantisation step, in nano-strain rate and nano-strain units) are relaxed. As expected, all methods achieve increasingly higher compression factors as the allowed tolerance increases, since relaxed accuracy requirements directly translate into smaller compressed file sizes. For

strain datasets, the compression is shown for varying accuracy in nanostrain units. In contrast, for strain rate data, the effective resolution varies with the sampling rate: since strain rate is computed by dividing phase change by the sampling interval, the quantisation step in strain rate increases with sampling rate. As a result, higher sampling rates yield coarser strain rate resolution. To allow consistent comparison across strain rate datasets, we express the quantisation step relative to the raw 16-bit integer values, where a step of 1 corresponds to lossless compression and a step of 10 represents a one-digit reduction in absolute precision.

JPEG2000 and DASPack provide consistently superior compression performance at higher tolerances, reaching compression factors beyond $8\times$. Even more relaxed accuracies (such as $1\text{ n}\epsilon$ or a 10^2 quantisation step) can result in up to $10\times$ compression for some datasets. In certain applications or datasets, this may offer a viable trade-off without compromising critical signal content.

The results indicate that DASPack’s fixed-accuracy compression approach is not only able to effectively leverage lossy compression, but also highly competitive when precision constraints are relaxed, making it a robust choice for applications that require significant data reduction with controlled accuracy.

4.2 Application examples

DASPack is suited to every DAS application by appropriately tuning the quantisation step to balance data fidelity and compression efficiency. To demonstrate its performance, this section presents several examples of compressed real events with varying levels of degradation. Each example also includes error maps, defined as the difference between the original and compressed signals.

Fig. 8 displays unfiltered strain rate data from an ML 3.4 regional earthquake recorded in Athens (dataset 2), with a dynamic range of $\pm 25\text{ n}\epsilon\text{ s}^{-1}$. Two quantisation steps are applied: $\Delta = 2\text{ n}\epsilon\text{ s}^{-1}$ and $\Delta = 8\text{ n}\epsilon\text{ s}^{-1}$. Results show that with low step values, the quantised trace is virtually indistinguishable from the original, indicating negligible error. In contrast, at high quantisation levels, low-amplitude features disappear, and only the largest peaks remain, indicating information loss, a trend that is likewise observed in the frequency domain. This is confirmed by the error map, which is nearly blank for $2\text{ n}\epsilon\text{ s}^{-1}$, while the $8\text{ n}\epsilon\text{ s}^{-1}$ case reveals systematic residuals, especially for low-amplitude arrivals. Compression factors increase accordingly, from $6.1\times$ to $9.7\times$. The completely lossless mode achieves a compression factor of $2.8\times$ on the same data, demonstrating the importance of allowing slight data degradation.

Fig. 9 presents filtered strain data from a magnitude *mbLg* 3.1 earthquake recorded in the Alboran Sea (dataset 10). This example illustrates how DASPack performs when the signal amplitude is very small (only about ± 2.5 n ϵ across the entire event). In this case, quantisation steps of $\Delta = 0.1$ n ϵ and $\Delta = 1$ n ϵ result in compression factors of $6.6\times$ and $14.4\times$, respectively. At the finer step, the quantised waveform closely matches the original, with negligible error. At 1 n ϵ , the error map shows residuals below 0.5 n ϵ that do not exhibit any visual patterns, indicating that most of the signal energy survives even at this aggressive compression level.

Similar patterns are observed in Fig. 10, which shows filtered strain data from a quarry blast recorded during the Vinaroz experiment (dataset 13), with amplitudes reaching about ± 10 n ϵ . Quantisation steps of $\Delta = 0.1$ n ϵ and $\Delta = 1$ n ϵ result in compression factors of $5.7\times$ and $11.6\times$, respectively. In this example, the error is not distributed uniformly: offshore channels (beyond 10 km) retain nearly all signal information even at the higher quantisation level, while the onshore section (within the first 10 km) records lower amplitudes and therefore exhibits higher residuals in the error maps.

Finally, Fig. 11 extends the examples to a high-amplitude (± 100 n ϵ), 4.9 *mbLg* earthquake recorded during the 2021 La Palma eruption (dataset 15). This case confirms that DASPack maintains high fidelity at moderate quantisation. For example, $\Delta = 4$ n ϵ results in a compression factor of $10.9\times$. In this higher degradation case, the error appears to be mostly uncorrelated, with higher values along the high-amplitude front.

5 DISCUSSION

5.1 Compression efficiency

DASPack contributes to DAS data management by providing sustained, real-time throughput while simultaneously delivering the highest compression factors across a broad range of acquisition settings. Its performance advantage is most pronounced for strain data, where the combination of wavelet decomposition and linear predictive coding (LPC) removes the slowly varying, large-amplitude components that dominate the signal variance. Because strain is typically stored in 32-bit words to accommodate this broad dynamic range, both the entropy reduction and the arithmetic-coding efficiency are amplified, yielding compression factors that exceed $8\times$. strain rate, in contrast, is acquired with narrower dynamic range and usually quantised at 16 bits; here the algorithm still outperforms specialised competitors with smaller margins, confirming that the same pipeline remains effective even when the predictable, low-frequency portion of the spectrum is reduced. This difference in performance arises not from

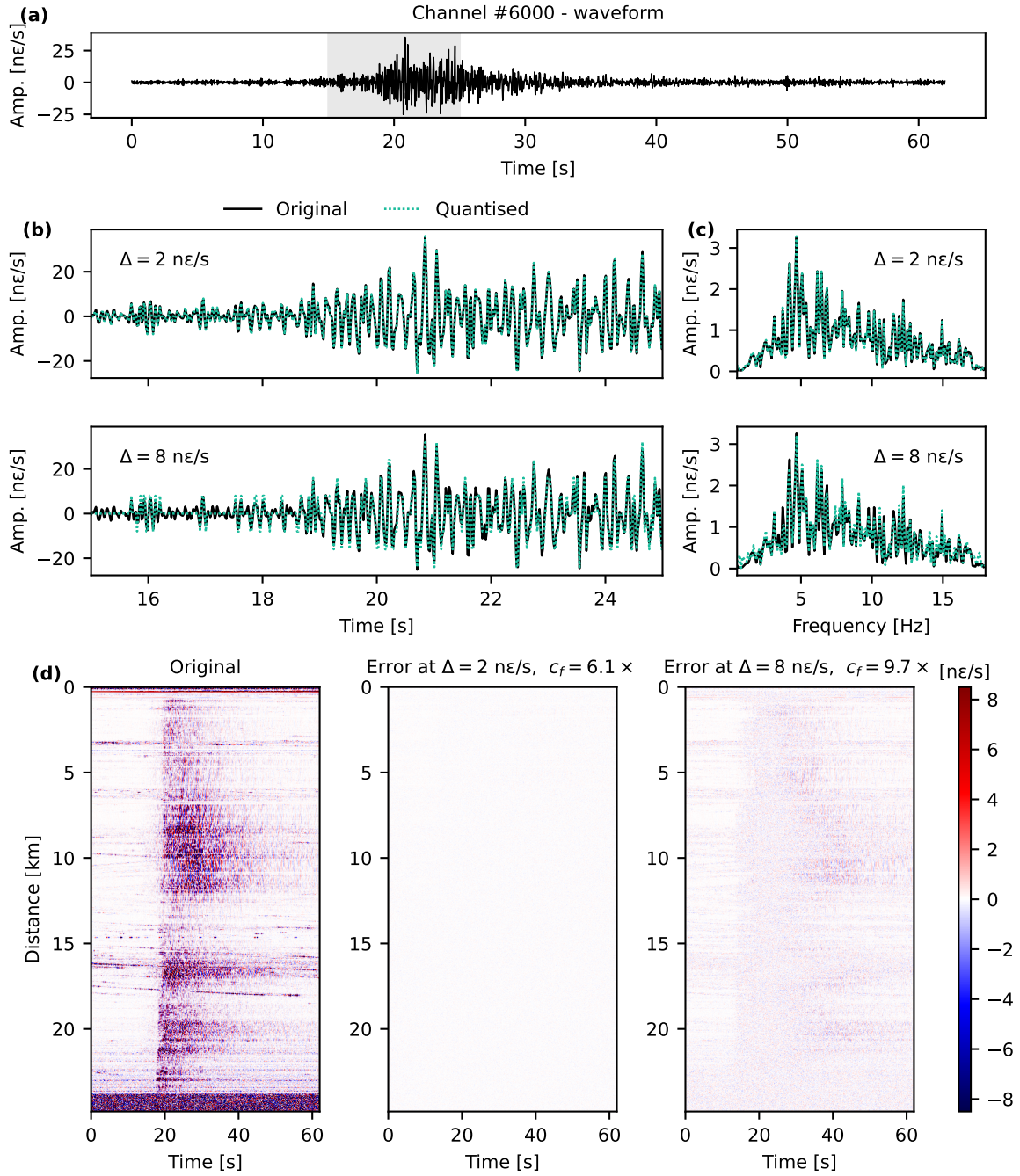


Figure 8. A regional earthquake of magnitude ML 3.4 occurred 48 km northwest of the interrogator on 23 September 2021 at 07:06:54 UTC during the Athens experiment (dataset 2). The figure compares unfiltered raw signals with quantised versions using two degradation levels ($\Delta = 2 \text{ n}\epsilon \text{ s}^{-1}$ and $\Delta = 8 \text{ n}\epsilon \text{ s}^{-1}$), corresponding to compression factors of $c_f = 6.1\times$ and $c_f = 9.7\times$. (a) Waveform from channel 6000 (at 6000 m). (b) Zoomed view of the waveform. (c) Discrete Fourier Transform of the zoomed window. (d) 2-D strain map of the original signal (left) and error maps between the original and degraded signals (middle and right).

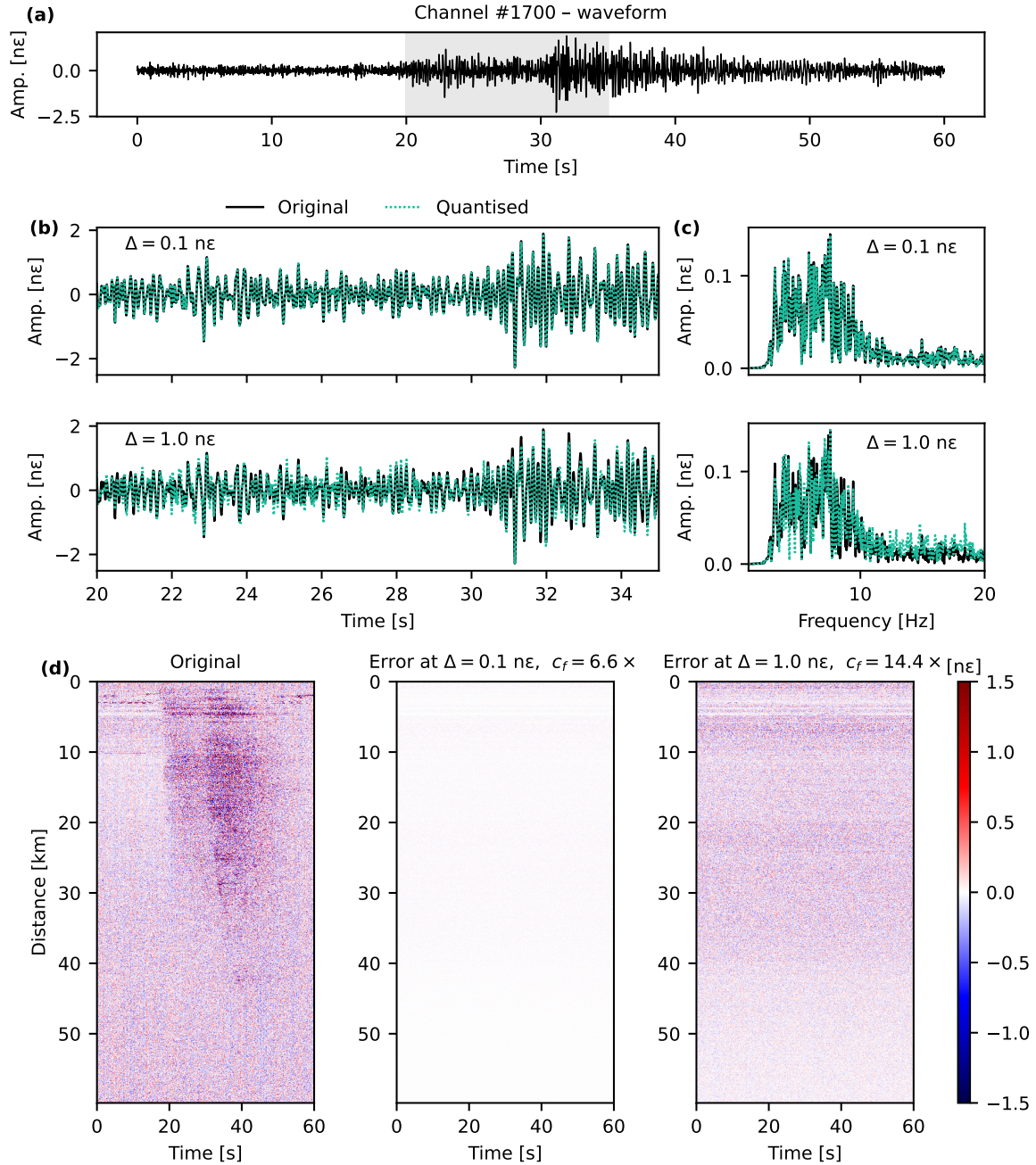


Figure 9. An onshore earthquake of magnitude $mbLg$ 3.1 occurred 97 km northeast of the interrogator on 5 October 2023 at 21:40:04 UTC and was recorded on a submarine fibre-optic cable, in the Alboran Sea (dataset 10). The figure compares filtered raw signals with quantised versions using two degradation levels ($\Delta = 0.1 \text{ n}\epsilon$ and $\Delta = 1 \text{ n}\epsilon$), corresponding to compression factors of $c_f = 6.6\times$ and $c_f = 14.4\times$. (a) Waveform from channel 1700 (at 17,000 m). (b) Zoomed view of the waveform. (c) Discrete Fourier Transform of the zoomed window. (d) 2-D strain map of the original signal (left) and error maps between the original and degraded signals (middle and right).

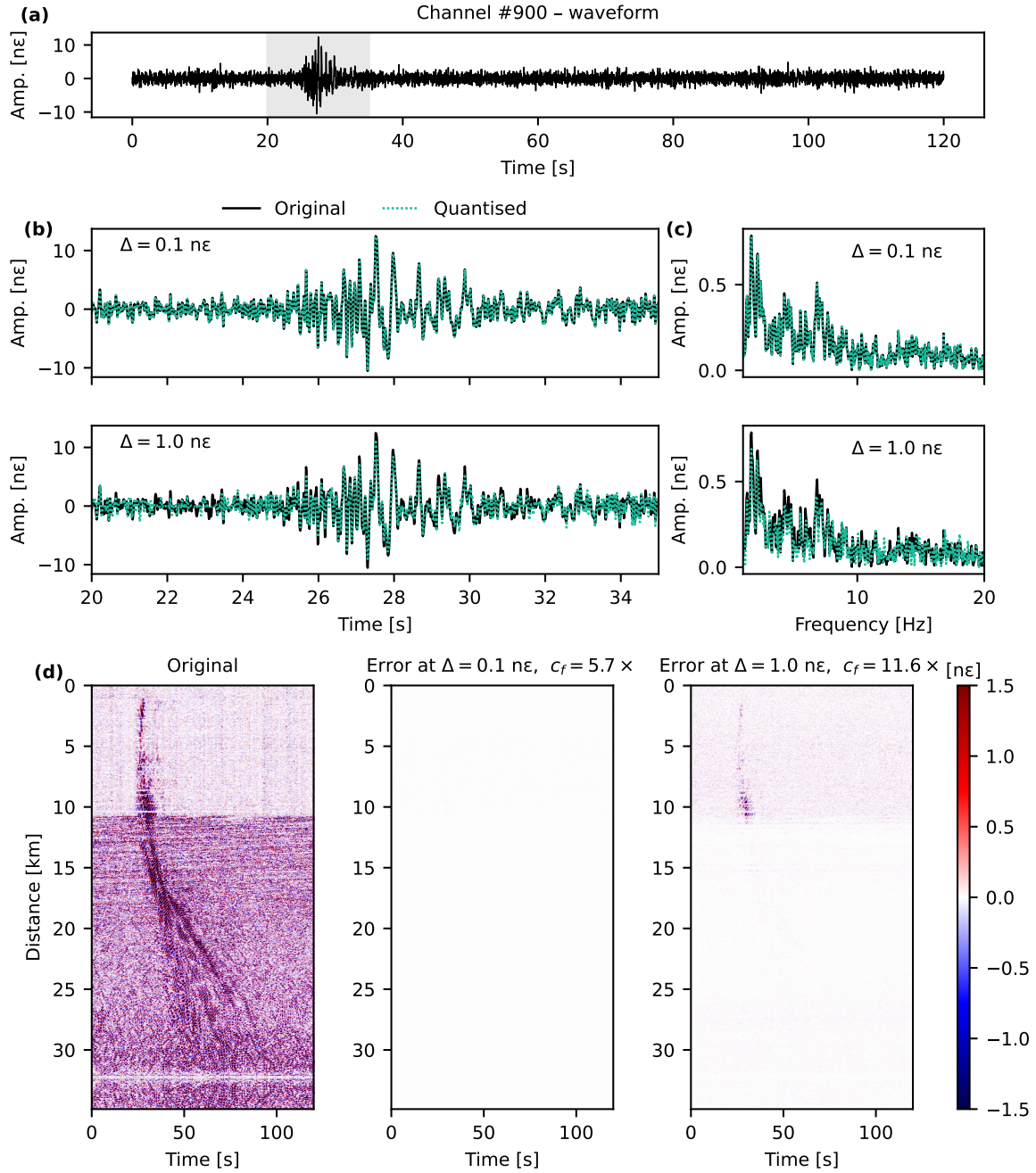


Figure 10. A quarry blast occurred 26 km southeast of the interrogator on 9 June 2023 at 10:26:30 UTC and was recorded during the Vinaroz experiment (dataset 13). The figure compares filtered raw signals with quantised versions using two degradation levels ($\Delta = 0.1 \text{ n}\epsilon$ and $\Delta = 1 \text{ n}\epsilon$), corresponding to compression factors of $c_f = 5.7\times$ and $c_f = 11.6\times$. (a) Waveform from channel 900 (at 9,000 m). (b) Zoomed view of the waveform. (c) Discrete Fourier Transform of the zoomed window. (d) 2-D strain map of the original signal (left) and error maps between the original and degraded signals (middle and right).

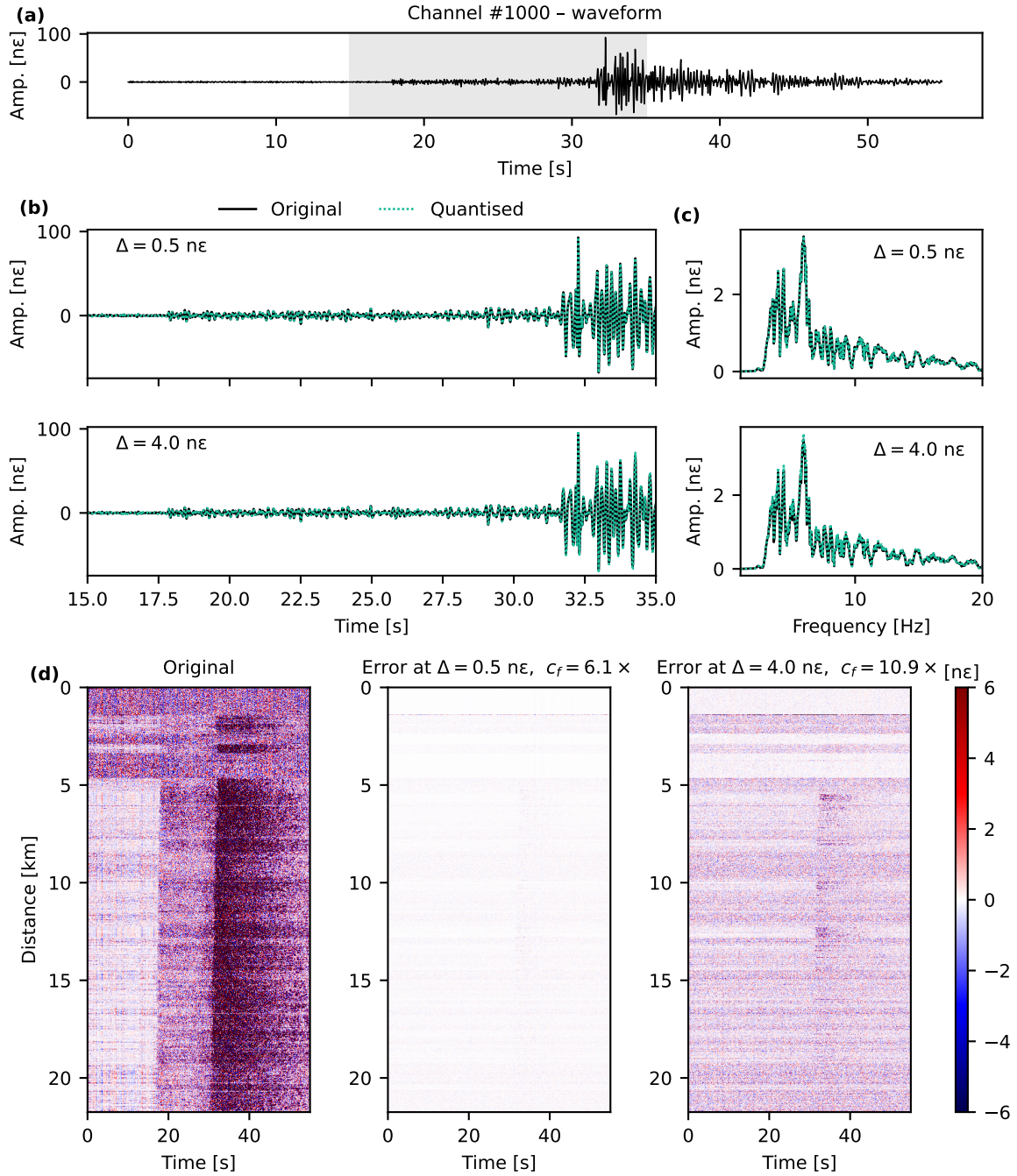


Figure 11. An earthquake of magnitude $mbLg$ 4.9 occurred 148 km west from the interrogator on 26 October 2021 at 16:25:37 UTC during the La Palma volcanic eruption and was recorded at the Teide Astronomical Observatory (dataset 14). The figure compares filtered raw signals with quantised versions using two degradation levels ($\Delta = 0.1 \text{ n}\epsilon$ and $\Delta = 4 \text{ n}\epsilon$), corresponding to compression factors of $c_f = 6.1\times$ and $c_f = 10.9\times$. (a) Waveform from channel 1000 (at 10,000 m). (b) Zoomed view of the waveform. (c) Discrete Fourier Transform of the zoomed window. (d) 2-D strain map of the original signal (left) and error maps between the original and degraded signals (middle and right).

the compression method itself but from the statistical properties of the input data: as a temporal derivative of strain, strain rate naturally exhibits lower amplitude variability and reduced temporal coherence, which limits the predictive power of both wavelet and LPC stages and reduces overall compressibility.

A central design choice in DASPack is the explicit separation of the scientific decision from engineering optimisation. By restricting every lossy decision into a single, user-controlled quantisation step, the method offers a guarantee that the reconstruction error is strictly bounded and readily interpretable as an absolute amplitude. This contrasts with perceptual or signal-adaptive codecs in which loss is distributed across many internal stages and is therefore difficult to predict. The fixed-accuracy paradigm enables practitioners to tailor the degradation to the specific use case: a researcher interested in microseismic events may keep the step below $0.05 \text{ n}\epsilon$ to preserve the smallest arrivals, whereas monitoring traffic-induced vibration along a bridge may tolerate $0.1 \text{ n}\epsilon$ and enjoy a four-fold reduction in storage cost.

5.2 Compression speed and real-time capability

For a given interrogator, the raw data generation rate is

$$R = N_{\text{ch}} f_s b \quad [\text{bytes s}^{-1}], \quad (3)$$

where N_{ch} is the number of channels, f_s the sampling rate (samples s^{-1}), and b the datum size in bytes (2 for int16, 4 for int32 and float32). A compressor with sustained throughput T_{method} can operate in real time whenever $R \leq T_{\text{method}}$. We therefore plot in Fig. 12 the level curves $R = T_{\text{method}}$ for each algorithm, using the average T_{method} values measured in section 4.1. The visual test is simple: if a point falls below a curve, that algorithm can compress the dataset on-the-fly. Otherwise, the raw stream outruns the compressor.

All datasets fall below the DASPack curve, confirming that DASPack can handle even the most demanding case: Zurisee (dataset 8), which involved an underwater application with a 10 kHz sampling rate across 2,496 channels, aimed at recording human and animal activity. JPEG2000, Zstd and Gzip fail for this extreme. DASPack, Zfp and H5TurboPfor are observed to be able to compress all configurations in real time.

In practice, DASPack uses 32-bit integer operations and achieves a sustained compression speed of approximately 250 MB s^{-1} for strain data when running in single-threaded mode. For strain rate data, typically stored as 16-bit integers, performance is about half, around 125 MB s^{-1} . Thanks to independent per-tile processing, DASPack supports parallelism. When configured with 8 threads on an Apple M1 Pro CPU, the average compression speed across our

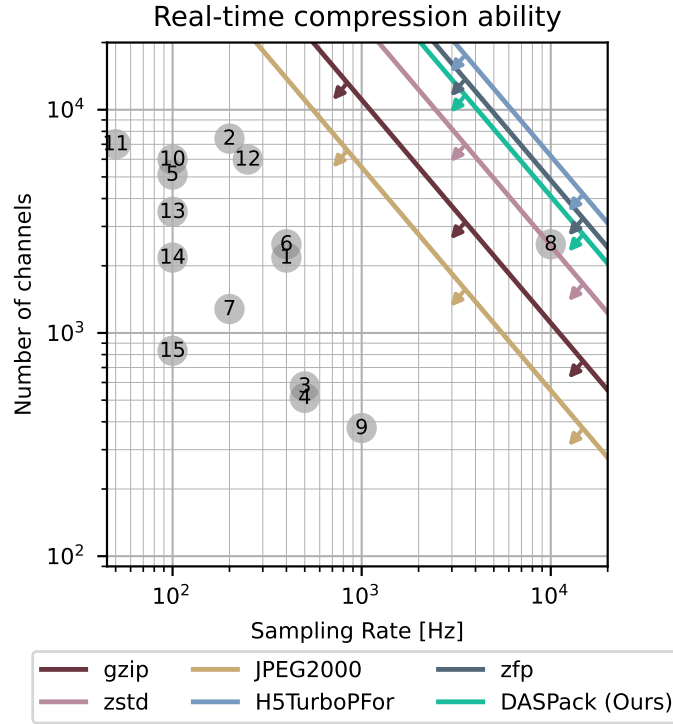


Figure 12. Real-time compression envelope. Each marker corresponds to one of the 15 datasets (Table 1) and is positioned by its number of channels (x-axis) and sampling rate (y-axis). Solid curves are the real-time limits of five compression methods: for any point below a curve the method can keep up with the raw data stream, because $R \leq T_{\text{method}}$ (see equation 3). Datasets lying above a curve would overwhelm that compressor.

datasets increases to roughly 750 MB s^{-1} . This throughput exceeds the write speed of standard hard disk drives (HDDs), which typically peak around 200 MB s^{-1} , making storage, rather than computing, the limiting factor in many setups. To fully leverage DASPack’s compression speed, faster storage such as solid-state drives (SSDs), which can sustain several gigabytes per second, is recommended. Clusters with a large number of nodes and drive bays offering high read and write speeds further exploit parallelism, enabling highly efficient compression and decompression.

5.3 Understanding the effect of quantisation

The controlled degradation step in DASPack is based on fixed-accuracy quantisation, which has well-understood statistical properties. For sufficiently small quantisation steps, the rounding error behaves like additive uniform noise with zero mean and known variance (Widrow & Kollár 2008). This behaviour explains why the resulting error is often modeled as white

noise, and was experimentally observed in Figures 8–11b,d. However, unlike truly independent noise, quantisation can lead to repeated values in low-amplitude segments. For example, entire portions of a trace being rounded to zero. At very coarse quantisation levels, this may introduce visible artifacts, even if the overall frequency content remains largely preserved.

Figures 8–11c illustrate that the frequency spectrum is barely modified by quantisation, and only the very low-amplitude parts of the signal are suppressed. Indeed, interpreted as random noise, the quantisation error converges to Gaussian behaviour in the frequency domain (Oppenheim & Schaffer 2009).

Another perspective on quantisation is to view it as a uniform degradation of the signal-to-noise ratio (SNR). If the error behaves like additive uniform noise, its power increases proportionally to the square of the quantisation step (Widrow & Kollár 2008). This is consistent with the results in Figure 13, where the SNR was computed as the ratio of signal and noise window powers. To prevent spurious vanishing errors caused by large quantisation steps rounding entire traces to zero, dithering was added during the SNR calculation, which helps recover idealised behaviour of quantisation noise (Widrow & Kollár 2008). Importantly, dithering was introduced solely during SNR evaluation to obtain meaningful estimates at coarse quantisation levels; it is not part of the DASPack method itself. The measured SNR follows closely the expected noise law, providing an additional validation of the theoretical model. This analysis demonstrates that quantisation in DASPack is not only interpretable as a bounded reconstruction error, but can also be understood within the familiar framework of additive noise, with clear implications for the effective resolution and dynamic range of compressed DAS signals.

While the statistical interpretation of quantisation is useful, it does not by itself capture the full implications for downstream seismic analysis. DAS is applied to a wide variety of problems, each of which may respond differently to small compression errors. For instance, typical processing tasks include (i) detecting an event through STA/LTA or template matching, (ii) picking traveltimes using manual, automatic, or array-based approaches, (iii) estimating dispersion curves from ambient noise correlations, or (iv) measuring amplitude decay to infer attenuation. Each of these workflows involves distinct sensitivities to low-amplitude signals, phase accuracy, or spectral content. A superficial treatment of all these aspects would risk being incomplete or even misleading. For this reason, we deliberately refrain from providing application-specific guidelines in this paper. Instead, we emphasise that a rigorous evaluation of the effect of controlled degradation on the accuracy of diverse seismological analyses deserves a dedicated study.

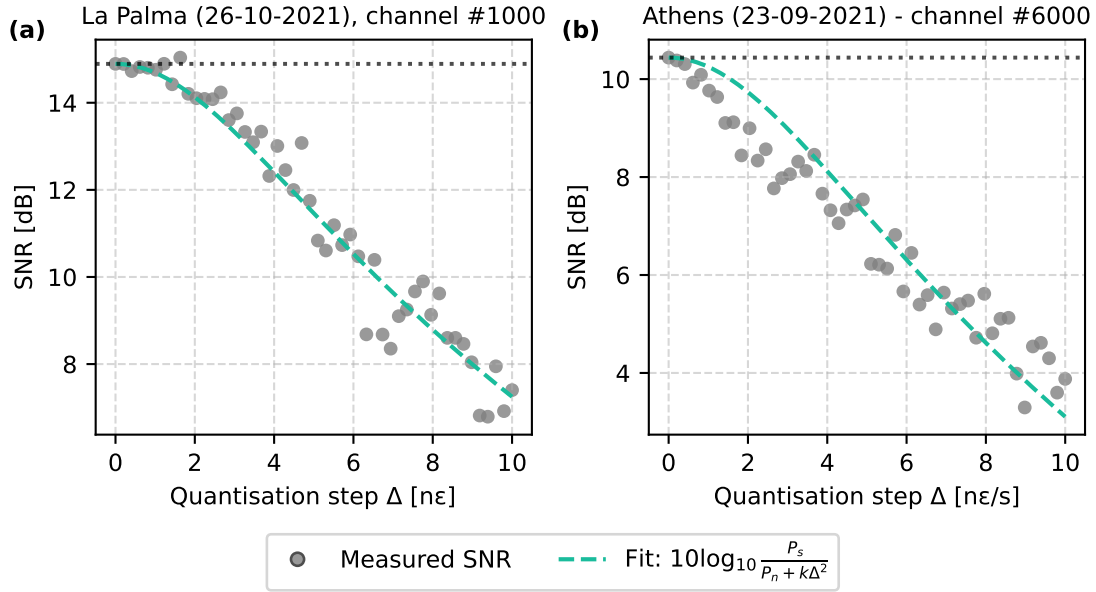


Figure 13. Signal-to-noise ratio (SNR) as a function of quantisation step Δ for two DAS datasets. (a) La Palma eruption (26 October 2021, channel #1000), also shown in Fig. 11 and (b) Athens array (23 September 2021, channel #6000), also shown in Fig. 8. Grey markers denote measured SNR values, while dashed green lines show fitted theoretical predictions of the form $10\log_{10} \frac{P_s}{P_n + k\Delta^2}$. The fitted constants are $k = 0.013$ ($R^2 = 0.95$) for La Palma and $k = 0.034$ ($R^2 = 0.87$) for Athens. For comparison, ideal i.i.d. uniform quantisation noise would correspond to $k = 1/12 \approx 0.083$. Dithering was applied during the SNR calculation to mitigate artificial vanishing errors at large Δ .

5.4 Analysis of DASPack’s lossless pipeline

The wavelet front-end contributes primarily by aligning the data with the locality assumption of LPC. DAS records exhibit strong temporal correlation that is well captured by low-order predictors, especially after the spatial and temporal low-pass information has been separated from the high-pass detail. The 5/3 bi-orthogonal wavelet was selected because it allows an integer-to-integer lifting implementation, guaranteeing perfect reversibility. Although cascading the transform in multiple steps was tested, a single step already removed most inter-sample redundancy; additional levels did not improve compression efficiency.

Linear predictive coding plays a complementary role by flattening the residual spectrum. Figures 3 and 4 show that, after LPC, the residuals better approximate white noise, a condition under which arithmetic coding approaches the Shannon limit. Prior to this, subtracting the temporal mean (stored as side information) was found to improve stationarity and thus yield a more effective model. In contrast, removing the spatial mean did not provide any noticeable gain in compression. The order of the predictor is kept small (typically two or three in time

and one in space) to bound the side information that must be stored with each tile. Larger numbers of coefficients were not found to improve compression efficiency and, in fact, tend to degrade it due to the increased volume of side information. Since these coefficients are stored independently for each tile, smaller tile sizes lead to more frequent coefficient storage, which increases overhead and hinders overall compression. However, excessively large tiles may span signals with varying statistical properties, reducing the effectiveness of the linear prediction. A balance between these effects was found by using square tiles of 1000 or 2000 samples per side, but the optimal choice depends on the data.

Beyond raw performance, DASPack addresses practicalities that often hinder adoption of research prototypes. The core is written in Rust, a language that combines speed with memory-safety guarantees. The code interface is designed to integrate seamlessly with existing seismic workflows and can be combined with widely used formats such as HDF5 and Zarr. Python bindings expose the compressor directly to popular data-analysis environments, so that researchers can read and write compressed files with little effort. Default parameters were established for the processing steps, accommodating the vast majority of cases and resulting in only marginal variations in compression efficiency, conveniently relieving users from the need to adjust them.

5.5 Combination with sampling methods

Beyond entropy-based compression, decimation and subsampling are widely used signal processing techniques that reduce data volume by lowering the temporal or spatial sampling rate. These operations can be combined with DASPack to achieve further reductions in storage requirements, provided that the reduced sampling rate preserves the bandwidth of interest. A more advanced alternative is compressed sensing (CS), which exploits the sparsity of many DAS wavefields. By leveraging this sparsity, CS allows signals to be sampled at rates below the Nyquist limit while still enabling accurate reconstruction. This makes CS particularly well-suited for application-specific DAS deployments where only a subset of information (e.g., source localisation or detection of transient events) is required. For example, Wang et al. (2025) presented a DAS system achieving a 40-fold lossy data reduction by combining envelope extraction with CS, while preserving accurate event localisation. Although CS is not universally applicable due to its reliance on signal sparsity and the complexity of reconstruction, it is a powerful tool in targeted deployments. Importantly, DASPack is complementary to such approaches: it can be applied downstream in lossless mode after CS or other degradation techniques, providing additional entropy-based compression without further information

loss. This integration is especially attractive for environments where storage and transmission resources are constrained.

5.6 Comparison and potential combination with deep learning methods

Recent studies have proposed deep learning for DAS data reduction, primarily through subsampling with reconstruction (Ni et al. 2024) and transformer-based autoencoders (Chen et al. 2025). While effective in extracting higher-order structure, these approaches fundamentally differ from DASPack. Ni et al. (2024) approximate missing traces from downsampled data, whereas Chen et al. (2025) encode DAS tiles into latent features, storing decoder weights for signal regeneration. Both strategies produce approximations without explicit error bounds, and may obscure low-amplitude or transient events. In contrast, DASPack provides deterministic guarantees: either exact lossless or fixed-accuracy lossy compression with strictly bounded error.

The computational trade-offs are also substantial. Chen et al. (2025) report ~ 30 minutes of GPU training for a one-minute DAS segment, with transfer learning partially mitigating the cost. Ni et al. (2024) achieved 51 seconds of A100 GPU compute to reconstruct a 10-minute record, and ~ 80 reconstructions per second on a Raspberry Pi for 1,000 channels. This is equivalent to a $\sim 3\times$ real-time throughput at 25 Hz sampling, with poor scaling for larger deployments. In comparison, DASPack compresses and decompresses in real time on standard CPUs, achieving sustained throughput up to 200 MB s^{-1} using one thread, without requiring GPUs or model storage. This is equivalent to real-time sampling throughput of about 50,000 Hz for the same example of 1,000 channels.

Importantly, downsampling or approximation should not be confused with entropy-based compression. Subsampling discards samples and interpolates missing information, while DASPack re-encodes the full signal bandwidth with user-defined error control. This distinction implies complementarity: DASPack can be applied prior to or after learned reconstruction, providing reliable entropy reduction while leaving room for interpolation-based inpainting. Likewise, DASPack's controlled quantisation could act as a regulariser for neural models, suppressing noise below application thresholds and potentially accelerating training convergence.

5.7 Applications of DASPack

Applications of DASPack extend across the spectrum of DAS research and industry practice. In earthquake seismology, the ability to record multi-month campaigns with significantly reduced storage opens the door to monitoring arrays spanning hundreds of kilometres of fibre

with reduced disk budgets. Environmental sensing applications, such as glacier dynamics, landslide detection or traffic-induced urban vibration benefit similarly because the relevant signals can sustain different levels of quantisation without compromising scientific utility.

The fixed-accuracy compression approach allows practitioners to choose a quantisation step that is safely below the minimum signal amplitude of interest. Figs. 8, 9, 10 and 11 demonstrate this principle: near-perfect signal fidelity is achieved at moderate quantisation steps, achieving up to $6\times$ compression factors.

Controlled compression is well suited to domains that produce large data volumes where only a small fraction contains transient events. With automatic event detection and phase picking, adaptive compression strategies can be applied: data windows with strong signals are stored at high fidelity, while quieter periods are compressed more aggressively. This approach can reduce storage use without compromising the integrity of relevant data.

Finally, because of its high throughput and low hardware requirements, DASPack can be deployed directly on edge devices or interrogators, enabling real-time compression of DAS data at the source before transmission or storage. This is particularly valuable for remote or bandwidth-limited deployments, where GSM, satellite or intermittent connectivity makes raw data handling challenging.

6 CONCLUSIONS

We have presented DASPack, a high-performance, open-source compression tool specifically tailored for DAS data. DASPack addresses the growing challenge of managing the massive data volumes generated by DAS deployments, offering a unique combination of controlled, user-defined degradation and an efficient, fully reversible lossless compression pipeline. Its fixed-accuracy quantisation approach allows users to make informed decisions about data fidelity while ensuring optimal compression.

Across a diverse set of 15 real-world DAS datasets, DASPack consistently outperformed both general-purpose and domain-specific compression algorithms in terms of compression efficiency and processing speed. File size reductions of up to $3\times$ were achieved in lossless mode, with up to $6\times$ compression realised with minimal information loss, and $10\times$ compression attained with acceptable degradation. It demonstrated real-time capability even under demanding conditions, such as high channel counts and high sampling rates. The tool supports a broad range of acquisition scenarios and integrates easily into existing workflows via HDF5 support and Python bindings.

DASPack also enables efficient on-site compression, making it suitable for deployment in

This work is published at *Geophys. J. Int.*. Please visit the original publication at

<https://doi.org/10.1093/gji/ggaf397>.

remote or bandwidth-limited environments. Its principled design empowers users to balance storage savings with scientific accuracy, providing a practical and scalable solution for long-term DAS data management. We believe DASPack can play an important role in future DAS deployments, supporting applications across seismology, environmental sensing, infrastructure monitoring, and beyond.

ACKNOWLEDGMENTS

We thank Professor Francesco Grigoli and an anonymous reviewer for the valuable feedback that helped improve the manuscript.

The first author thanks Professor Cesar R. Ranero for his support during the realisation of the work.

This work was supported in part by the "Severo Ochoa Centre of Excellence" program (CEX2019-000928-S-21-2); the European Union NextGenerationEU/PRTR Program under projects PSI (PLEC2021-007875), TREMORS (CPP2021-008869), the CSIC contract SAFE with Telxius Cable España, S. L. (ref. 20233069) and the Spanish Ministry of Science, Innovation and Universities under DeeLight project (PID2020-117142GB-I00).

Canalink, IslaLink, Telxius, ENAGAS, Teide and Roque de los Muchachos Astronomical Observatories facilitated fibre-optic cable access. Aragon Photonics provided the HDAS interrogator used in some of the experiments.

DATA AVAILABILITY

DASPack is openly available at <https://github.com/asleix/daspack>, including examples to use the software. We used an existing entropy coding implementation named Constriction, <https://github.com/bamler-lab/constriction>. For published datasets listed in Table 1, details on open data access can be found in the corresponding publications. Unpublished datasets are available from the respective authors upon reasonable request. Data segments used in Figs. 8–11 are available at <https://doi.org/10.20350/digitalCSIC/17586> (Ugalde & Fitchner 2025).

REFERENCES

- Aichele, J., Simeon, A., van Herwijnen, A., Volpi, M., Sovilla, B., Huguenin, P., Gaume, J., Fitchner, A., Pérez, C., & Edme, P., 2025. In situ snow avalanche monitoring and characterization, in

This work is published at *Geophys. J. Int.*. Please visit the original publication at

<https://doi.org/10.1093/gji/ggaf397>.

- EGU General Assembly 2025*, no. EGU25-7061 in EGU General Assembly Conference Abstracts, Copernicus Meetings, Vienna, Austria, doi: 10.5194/egusphere-egu25-7061.
- Ajo-Franklin, J. B., Dou, S., Lindsey, N. J., Monga, I., Tracy, C., Robertson, M., Rodriguez Tribaldos, V., Ulrich, C., Freifeld, B., Daley, T., & Li, X., 2019. Distributed acoustic sensing using dark fiber for near-surface characterisation and broadband seismic event detection, *Sci. Rep.*, **9**, doi:10.1038/s41598-018-36675-8.
- Bamler, R., 2022. Understanding entropy coding with asymmetric numeral systems (ans): a statistician's perspective, *arXiv preprint arXiv:2201.01741*.
- Biondi, B., Martin, E., Cole, S., Karrenbach, M., & Lindsey, N., 2017. Earthquake analysis using data recorded by the Stanford DAS array, *SEG Expanded Abstracts*, **2017**, 2752–1756.
- Booth, A. D., Christoffersen, P., Schoonman, C., Clarke, A., Hubbard, B., Law, R., Doyle, S. H., Chudley, T. R., & Chalari, A., 2020. Distributed Acoustic Sensing of seismic properties in a borehole drilled on a fast-flowing Greenlandic outlet glacier, *Geophys. Res. Lett.*, **47**, doi:10.1029/2020GL088148.
- Bozzi, E., 2025. *The Potential of Distributed Acoustic Sensing Technology for Seismological Monitoring*, Phd thesis, University of Milano-Bicocca, Doctoral Thesis, PhD Cycle 37, Academic Year 2023/2024.
- 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.
- Cover, T. M. & Thomas, J. A., 2005. *Gaussian Channel*, chap. 9, pp. 261–299, John Wiley & Sons, Ltd, doi: 10.1002/047174882X.ch9.
- Daley, T. M., Pevzner, R., Shulakova, V., Kashikar, S., Miller, D. E., Goetz, J., & ans S. Lueth, J. H., 2013. Field testing of fiber-optic distributed acoustic sensing (DAS) for subsurface seismic monitoring, *The Leading Edge*, **June 2013**, 936–942.
- Daley, T. M., Miller, D. E., Dodds, K., Cook, P., & Freifeld, B. M., 2016. Field testing of modular borehole monitoring with simultaneous distributed acoustic sensing and geophone vertical seismic profiles at Citronelle, Alabama, *Geophys. Prosp.*, **64**, 1318–1334.
- Daubechies, I. & Sweldens, W., 1998. Factoring wavelet transforms into lifting steps, *Journal of Fourier Analysis and Applications*, **4**(3), 247–269, doi: 10.1007/BF02476026.
- Dong, B., Popescu, A., Tribaldos, V. R., Byna, S., Ajo-Franklin, J., & Wu, K., 2022. Real-time and post-hoc compression for data from distributed acoustic sensing, *Computers & Geosciences*, **166**, 105181, doi: doi.org/10.1016/j.cageo.2022.105181.
- Durbin, J., 1960. The fitting of time-series models, *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, **28**(3), 233–244.
- Edme, P., Paitz, P., Walter, F., van Herwijnen, A., & Fichtner, A., 2023. Fiber-optic detection of snow avalanches using telecommunication infrastructure.
- Fichtner, A., Hofstede, C., Kennett, B. L. N., Svensson, A., Westhoff, J., Walter, F., Ampuero,

- J., Cook, E., Zigone, D., Jansen, D., & Eisen, O., 2025. Hidden cascades of seismic ice stream deformation, *Science*, **387**, 858–864, doi: 10.1126/science.adp8094.
- Hartog, A., 2017. *An introduction to distributed optical fibre sensors*, CRC Press, Boca Raton.
- Hill, D., 2015. Distributed Acoustic Sensing (DAS): Theory and applications, *Frontiers in Optics*, **2015**, doi:10.1364/FIO.2015.FTh4E.1.
- 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, *J. Geophys. Res.*, **126**, doi:10.1029/2020JB021493.
- Igel, J., Klaasen, S., Noe, S., Nomikou, P., Karantzaos, K., & Fichtner, A., 2024. Challenges in submarine fiber-optic earthquake monitoring, *ESS Open Archive*, doi: 10.22541/essoar.171691177.74747140/v1.
- 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.
- Jacques, L., Duval, L., Chaux, C., & Peyré, G., 2011. A panorama on multiscale geometric representations, intertwining spatial, directional and frequency selectivity, *Signal Processing*, **91**(12), 2699–2730, doi: 10.1016/j.sigpro.2011.04.025.
- Jousset, P., Currenti, G., Schwarz, B., Chalari, A., Tilmann, F., Reinsch, T., Zuccarello, L., Privitera, E., & Krwaczek, C. M., 2022. Fibre optic distributed acoustic sensing of volcanic events, *Nat. Comm.*, **13**, doi:10.1038/s41467-022-29184-w.
- Kang, J., Walter, F., Paitz, P., Aichele, J., Edme, P., Meier, L., & Fichtner, A., 2024. LAutomatic monitoring of rock-slope failures using distributed acoustic sensing and semi-supervised learning, *Geophys. Res. Lett.*, **51**, doi: 10.1029/2024GL110672.
- Klaasen, S., Paitz, P., Lindner, N., Dettmer, J., & Fichtner, A., 2021a. Distributed Acoustic Sensing in Volcano-Glacial Environments—Mount Meager, British Columbia, *Journal of Geophysical Research: Solid Earth*, **126**(11), e2021JB022358, doi: 10.1029/2021JB022358, e2021JB022358 2021JB022358.
- Klaasen, S., Thrastarson, S., Jonsdottir, K., Cubuk Sabuncu, Y., Paitz, P., & Fichtner, A., 2021b. Distributed Acoustic Sensing on Volcanoes: Grimsvotn and Fagradalsfjall, Iceland, in *AGU Fall Meeting Abstracts*, vol. 2021, pp. V23B–07.
- Klaasen, S., Thrastarson, S., Fichtner, A., Cubuk-Sabuncu, Y., & Jonsdottir, K., 2022. Sensing Iceland’s most active volcano with a ”buried hair”, *EOS*, **103**, doi:10.1029/2022EO220007.
- Klaasen, S., Thrastarson, S., Cubuk-Sabuncu, Y., Jonsdottir, K., Gebraad, L., Paitz, P., & Fichtner, A., 2023. Subglacial volcano monitoring with fiber-optic sensing: Grímsvötn, Iceland, *Volcanica*, **6**, doi:10.30909/vol.06.02.301311.
- Klabnik, S. & Nichols, C., 2019. The rust programming language, Accessed: 2025-04-02.
- Le Gall, D. & Tabatabai, A., 1988. Sub-band coding of digital images using symmetric short kernel filters and arithmetic coding techniques, in *ICASSP-88., International Conference on Acoustics, Speech, and Signal Processing*, pp. 761–764 vol.2, doi: 10.1109/ICASSP.1988.196696.

- Lindsey, N. J. & Martin, E., 2021. Fiber-optic seismology, *Ann. Rev. Earth Planet. Sci.*, **49**, 309–336.
- Lindsey, N. J., Martin, E. R., Dreger, D. S., Freifeld, B., Cole, S., James, S. R., Biondi, B. L., & Ajo-Franklin, J. B., 2017. Fiber-optic network observations of earthquake wavefields, *Geophys. Res. Lett.*, **44**, 11792–11799.
- Lindstrom, P., 2014. Fixed-rate compressed floating-point arrays, *IEEE Transactions on Visualization and Computer Graphics*, **20**(12), 2674–2683, doi: 10.1109/TVCG.2014.2346458.
- Lior, I., Sladen, A., Rivet, D., Ampuero, J.-P., Hello, Y., Becerill, C., Martins, H. F., Lamare, P., Jestin, C., Tsagkli, S., & Markou, C., 2021. On the detection capabilities of underwater Distributed Acoustic Sensing, *J. Geophys. Res.*, p. doi:10.1029/2020JB020925.
- Mallat, S., 2009. Chapter 1 - sparse representations, in *A Wavelet Tour of Signal Processing (Third Edition)*, pp. 1–31, ed. Mallat, S., Academic Press, Boston, third edition edition., doi: 10.1016/B978-0-12-374370-1.00005-7.
- Martin, E. R., Castillo, C. M., Cole, S., Sawasdee, P. S., Yuan, S., Clapp, R., Karrenbach, M., & Biondi, B. L., 2017. Seismic monitoring leveraging existing telecom infrastructure at the SDASA: Active, passive, and ambient-noise analysis, *The Leading Edge*, **36**, 1025–1031.
- Mateeva, A., Lopez, J., Potters, H., Mestayer, J., Cox, B., Kiyashchenko, D., Wills, P., Grandi, S., Kuvshinov, B., Berlang, W., Yang, Z., & Detomo, R., 2014. Distributed acoustic sensing for reservoir monitoring with vertical seismic profiling, *Geophys. Prosp.*, **62**, 679–692.
- Ni, Y., Denolle, M. A., Shi, Q., Lipovsky, B. P., Pan, S., & Kutz, J. N., 2024. 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: doi.org/10.1029/2024JH000247, e2024JH000247 2024JH000247.
- NSF SAGE Data Services, 2024. NSF SAGE Data Services, Accessed: 2025-02-20.
- Oppenheim, A. V. & Schaffer, R. W., 2009. *Discrete-Time Signal Processing*, Prentice Hall, 3rd edition.
- Ouellet, S. M., Dettmer, J., Lato, M. J., Cole, S., Hutchinson, D. J., Karrenbach, M., Dashwood, B., Chambers, J. E., & Crickmore, R., 2024. Previously hidden landslide processes revealed using distributed acoustic sensing with nanostrain-rate sensitivity, *Nature Communications*, **15**(1), 6239, doi: 10.1038/s41467-024-50604-6.
- Owen, A., Duckworth, G., & Worsley, J., 2012. Fibre-optic distributed acoustic sensing for border monitoring, in *2012 European Intelligence and Security Informatics Conference*, p. doi:10.1109/EISIC.2012.59.
- Peterson, J. R., 1993. Observations and modeling of seismic background noise, *Survey, U.S. Geological*, doi: 10.3133/ofr93322, Report.
- Richardson, I. E., 2010. *The H.264 Advanced Video Compression Standard*, Wiley Publishing, 2nd edition.
- Romanowicz, B., Allen, R., Brekke, K., Chen, L., Gou, Y., Henson, I., Marty, J., Neuhauser, D.,

- Pardini, B., Taira, T., Thompson, S., Zhang, J., & Zuzlewski, S., 2023. SeaFOAM: A Year-Long DAS Deployment in Monterey Bay, California, *Seismological Research Letters*, **94**(5), 2348–2359, doi: 10.1785/0220230047.
- Rychen, J., Paitz, P., Edme, P., Smolinski, K., Brackenhoff, J., & Fichtner, A., 2023. Test experiments with distributed acoustic sensing and hydrophone arrays for locating underwater sounds., doi: 10.5281/zenodo.7886409.
- 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., & Ventosa, S., 2024. Inter-channel lossless data compression for distributed acoustic sensing, in *Galileo conference: Fibre Optic Sensing in Geosciences*, doi: 10.5194/egusphere-gc12-fibreoptic-5.
- Shannon, C. E., 1948. A mathematical theory of communication, *The Bell System Technical Journal*, **27**(3), 379–423, doi: 10.1002/j.1538-7305.1948.tb01338.x.
- Shi, Q., Williams, E. F., Lipovsky, B. P., Denolle, M. A., Wilcock, W. S. D., Kelley, D. S., & Schoedl, K., 2025. Multiplexed Distributed Acoustic Sensing Offshore Central Oregon, *Seismological Research Letters*, **96**(2A), 784–800, doi: 10.1785/0220240460.
- Smolinski, K., Bowden, D., Paitz, P., Kugler, F., & Fichtner, A., 2024. Shallow subsurface imaging using challenging urban DAS data, *Seis. Res. Lett.*, **2024**, doi: 10.1785/0220240052.
- Spica, Z. J., Pertou, M., Martin, E. R., Beroza, B. C., & Biondi, B., 2020. Urban seismic site characterization by fiber-optic seismology, *J. Geophys. Res.*, **125**, doi:10.1029/2019JB018656.
- Taubman, D. & Marcellin, M., 2012. *JPEG2000 Image Compression Fundamentals, Standards and Practice: Image Compression Fundamentals, Standards and Practice*, Springer Science & Business Media.
- Tian, L., Bowden, D., Smolinski, K. T., Obermann, A., Wang, Y., Lentas, K., Melis, N. S., & Fichtner, A., 2022. Urban shallow subsurface imaging and hazard assessment with distributed acoustic sensing in Athens, Greece, AGU Fall Meeting 2022; Conference Location: Chicago, IL, USA; Conference Date: December 12-16, 2022; Poster Abstract.
- Tuinstra, K., Grigoli, F., Lanza, F., Rinaldi, A. P., Fichtner, A., & Wiemer, S., 2024. Locating clustered seismicity using Distance Geometry Solvers: applications for sparse and single-borehole DAS networks, *Geophys. J. Int.*, **238**, 661–680, doi: 10.1093/gji/ggae168.
- Ugalde, A. & Fichtner, A., 2025. DASPack: Controlled Data Compression for Distributed Acoustic Sensing [Dataset], DIGITAL.CSIC, doi: 10.20350/digitalCSIC/17586.
- Ugalde, A., Becerril, C., Villaseñor, A., Ranero, C. R., Fernández-Ruiz, M. R., Martin-Lopez, S., González-Herráez, M., & Martins, H. F., 2021. Noise levels and signals observed on submarine fibers in the Canary Islands using DAS, *Seismological Research Letters*, **93**(1), 351–363, doi: 10.1785/0220210049.
- Ugalde, A., Bartolome, R., Cabieces, R., Grevemeyer, I., Villaseñor, A., Estrada, F., Palomino, D.,

- Cubas, M., Neri, A., & Latorre, H., 2023a. SAFE: OBS and DAS acquisition off the coast of Estepona (Spain), International Federation of Digital Seismograph Networks, doi: 10.7914/ZGE4-KD84.
- Ugalde, A., Latorre, H., Vidal, P., Martins, H. F., Martin-Lopez, S., & Gonzalez-Herraez, M., 2023b. Canary Islands seismic monitoring with Distributed Acoustic Sensing (CANDAS), International Federation of Digital Seismograph Networks, doi: 10.7914/73K1-1369.
- Vassiliou, A. A. & Wickerhouser, M. V., 1997. Comparison of wavelet image coding schemes for seismic data compression, in *Wavelet Applications in Signal and Image Processing V*, vol. 3169, pp. 118 – 126, International Society for Optics and Photonics, SPIE, doi: 10.1117/12.279684.
- Villasenor, A., Vidal Moreno, P. J., Gonzalez-Herraez, M., & Muñoz Tuñon, C., 2021. DAS deployments in the Instituto de Astrofísica de Canarias observatories to monitor the 2021 Cumbre Vieja eruption, doi: 10.7914/SN/9G.2021.
- Villasenor, A., Ugalde, A., & Ranero, C., 2023. Vinaroz submarine cable DAS deployment, International Federation of Digital Seismograph Networks, doi: 10.7914/8CTX-8J82.
- Walter, F., Gräff, D., Lindner, F., Paitz, P., Köpfli, M., Chmiel, M., & Fichtner, A., 2020a. Distributed Acoustic Sensing of microseismic sources and wave propagation in glaciated terrain, *Nat. Comm.*, **11**, doi: 10.1038/s41467-020-15824.
- Walter, F., Gräff, D., Lindner, F., Paitz, P., Köpfli, M., Chmiel, M., & Fichtner, A., 2020b. Distributed acoustic sensing of microseismic sources and wave propagation in glaciated terrain, *Nature Communications*, **11**(1), 2436, doi: 10.1038/s41467-020-15824-6.
- Wang, Y., Jiang, J., Liu, K., Zhang, M., Wang, S., Xu, T., Zhang, X., Ding, Z., & Liu, T., 2025. Ultra compression-ratio distributed acoustic sensing system via combination of envelope extraction and compressed sensing, *Opt. Express*, **33**(5), 9854–9864, doi: 10.1364/OE.554324.
- Widrow, B. & Kollár, I., 2008. *Quantization Noise: Roundoff Error in Digital Computation, Signal Processing, Control, and Communications*, Cambridge University Press.
- Witten, I. H., Neal, R. M., & Cleary, J. G., 1987. Arithmetic coding for data compression, *Commun. ACM*, **30**(6), 520–540, doi: 10.1145/214762.214771.
- Wu, C., Chen, J., Yan, G., & He, Z., 2023. A lossless data compression method for distributed acoustic sensors, in *28th International Conference on Optical Fiber Sensors*, p. W4.20, Optica Publishing Group, doi: 10.1364/OFS.2023.W4.20.
- Zhan, Z., 2020. Distributed acoustic sensing turns fiber-optic cables into sensitive seismic antennas, *Seis. Res. Lett.*, **91**, 1–15.