Supplementary material for a **Research Article** in SEISMICA

Near real-time channel selection for Distributed Acoustic Sensing technology (SM)

Emanuele Bozzi * [1], **Giulio Pascucci** [1], **Giacomo Rapagnani** [1], **Gian Maria Bocchini** [2], **Rebecca Harrington** [2], **Arantza Ugalde** [3], **Gilberto Saccorotti** [4], **Francesco Grigoli** [1]

[1]Department of Earth Sciences, University of Pisa, Pisa, Italy, [2]Institute of Geosciences, Ruhr University Bochum, Bochum, Germany, [3]Department of Marine Geosciences, Institut de Ciències del Mar, CSIC, 08003 Barcelona, Spain, [4]Istituto Nazionale di Geofisica e Vulcanologia, Pisa, Italy

Author contributions: *Conceptualization*: Emanuele Bozzi, Giulio Pascucci, Francesco Grigoli, Gilberto Saccorotti. *Methodology*: Emanuele Bozzi, Giulio Pascucci, Francesco Grigoli. *Software*: Emanuele Bozzi, Giulio Pascucci. *Formal Analysis*: Emanuele Bozzi, Giulio Pascucci. *Resources*: Rebecca Harrington, Arantza Ugalde. *Writing - Original draft*: Emanuele Bozzi. *Writing - Review & Editing*: Giulio Pascucci, Giacomo Rapagnani, Gian Maria Bocchini, Rebecca Harrington, Arantza Ugalde, Gilberto Saccorotti, Francesco Grigoli. *Visualization*: Emanuele Bozzi. *Supervision*: Francesco Grigoli. *Funding acquisition*: Francesco Grigoli.

Abstract Supplemental material for the manuscript "Near real-time channel selection for Distributed Acoustic Sensing technology".

References

Afanasiev, M., Boehm, C., van Driel, M., Krischer, L., Rietmann, M., May, D. A., Knepley, M. G., and Fichtner, A. Modular and flexible spectral-element waveform modelling in two and three dimensions. *Geophysical Journal International*, 216(3):1675–1692, Nov. 2018. doi: 10.1093/gji/ggy469.

Bocchini, G. M., Bozzi, E., Roth, M. P., Gaviano, S., Pascucci, G., Grigoli, F., Biondi, E., Sokos, E., and Harrington, R. M. Earthquake Catalog and Continuous Waveforms from a Two-Week Distributed Acoustic Sensing experiment on Kefalonia Island, Greece. preprint, Dec. 2025. doi: 10.22541/essoar.176460631.19268758/v2.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.

*Corresponding author: emanuele.bozzi@dst.unipi.it

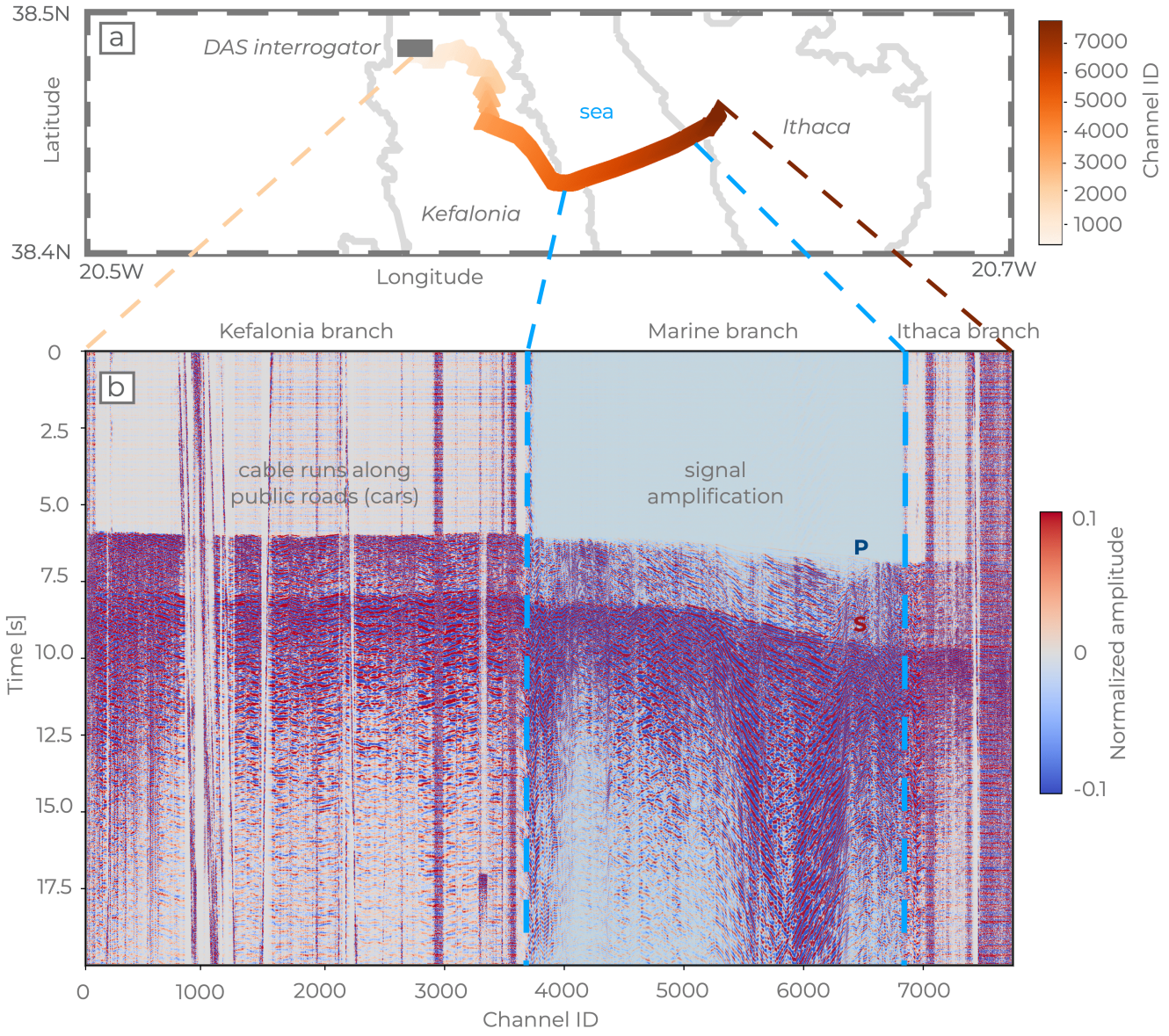

Figure S1 Overview on the Kefalonia DAS array used to test event relocation with selected channels from ORION. (a) Array geometry and channel layout on a simplified map of the Kefalonia-Ithaca area (see Figure 5 for a more detailed map). (b) DAS recording of an earthquake (event ID kef0367 Bocchini et al. (2025)) from the target seismicity cluster north of Kefalonia Island (see Figure 5 for cluster location) recorded by the Kefalonia DAS array. Light blue dotted lines mark the marine portion of the cable, where strong signal amplification is observed. Traces are normalized individually; as a result, energetic S waves in the marine section make P waves appear less clear (see also Figure S5).

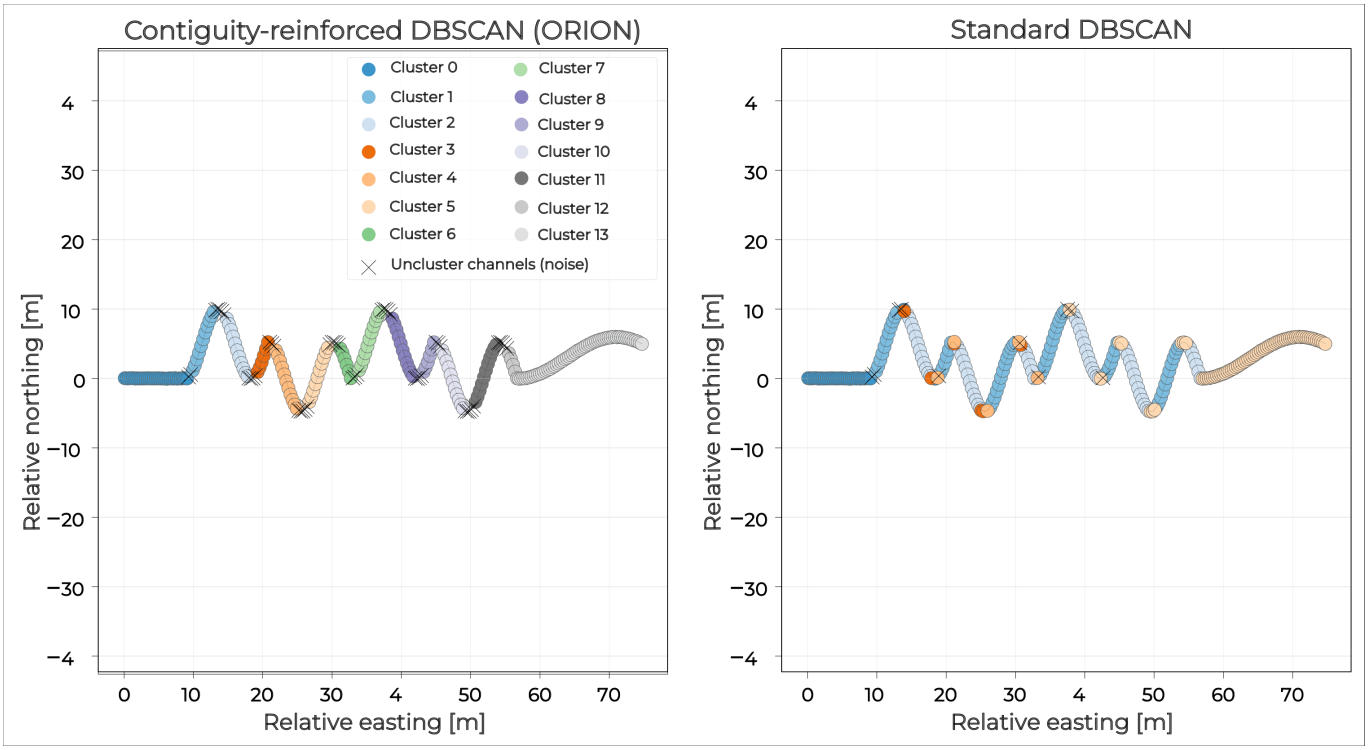

Figure S2 Spatial clustering of a synthetic DAS array geometry using ORION and DBSCAN Ester et al. (1996). Local azimuth is evaluated by the two methods, while ORION additionally evaluates spatial contiguity over a spatial window. Left) ORION identifies 13 clusters, separated by unclustered channels at each location where the cable exhibits a sudden azimuth change. Right) DBSCAN identifies 5 clusters, correctly grouping channels with similar azimuths but without distinguishing between spatially separate segments along the array.

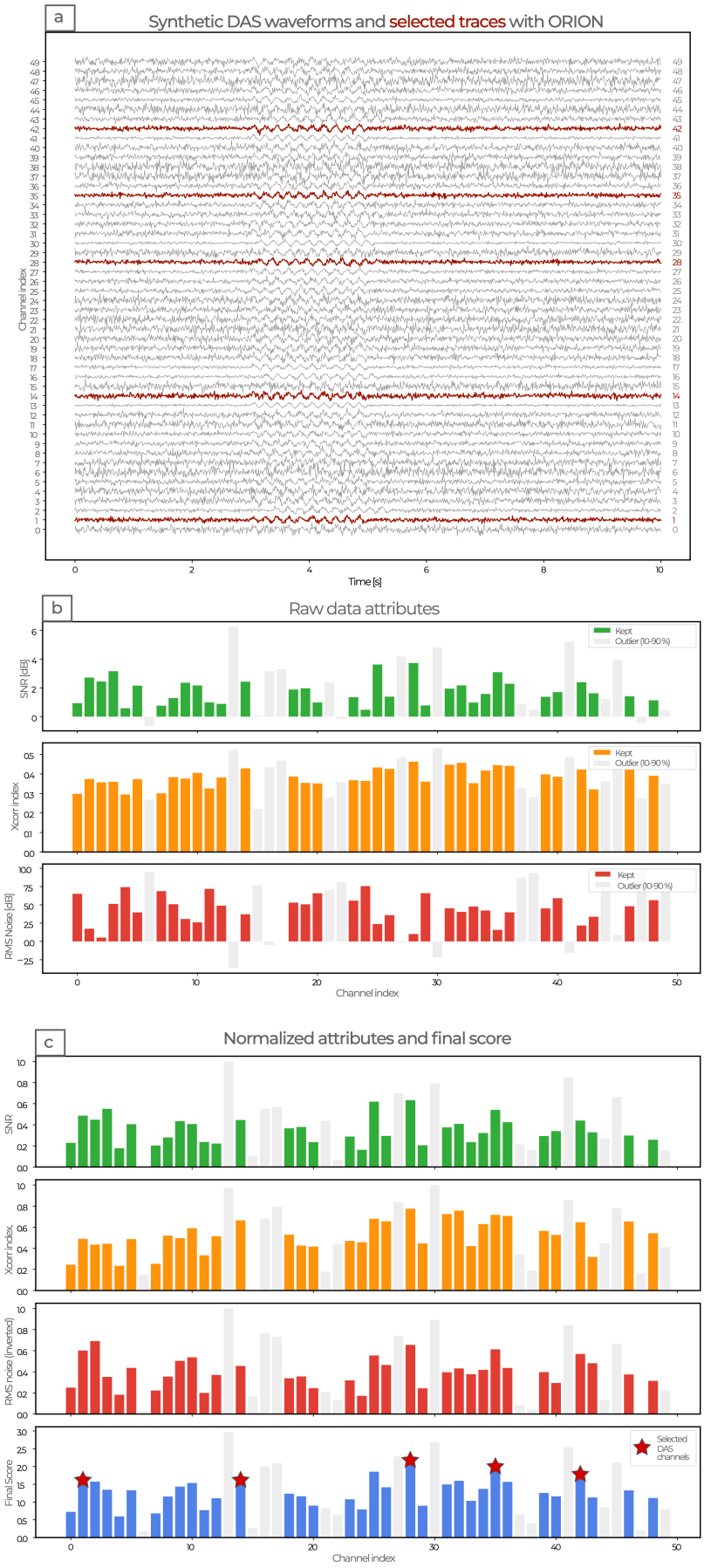

Figure S3 Overview of the waveform-attribute–based selection procedure in ORION. a) Synthetic waveforms simulating different levels of Gaussian noise contamination and temporal shifts of the sinusoidal signal. Red traces highlight the channels automatically selected by ORION. b) Raw waveform-attribute values for each channel, such as SNR, cross-correlation indices, and noise RMS amplitude, with outliers outside the 10th–90th percentile excluded (gray bins). c) Normalized attributes and final score values. Red stars indicate the five selected traces.

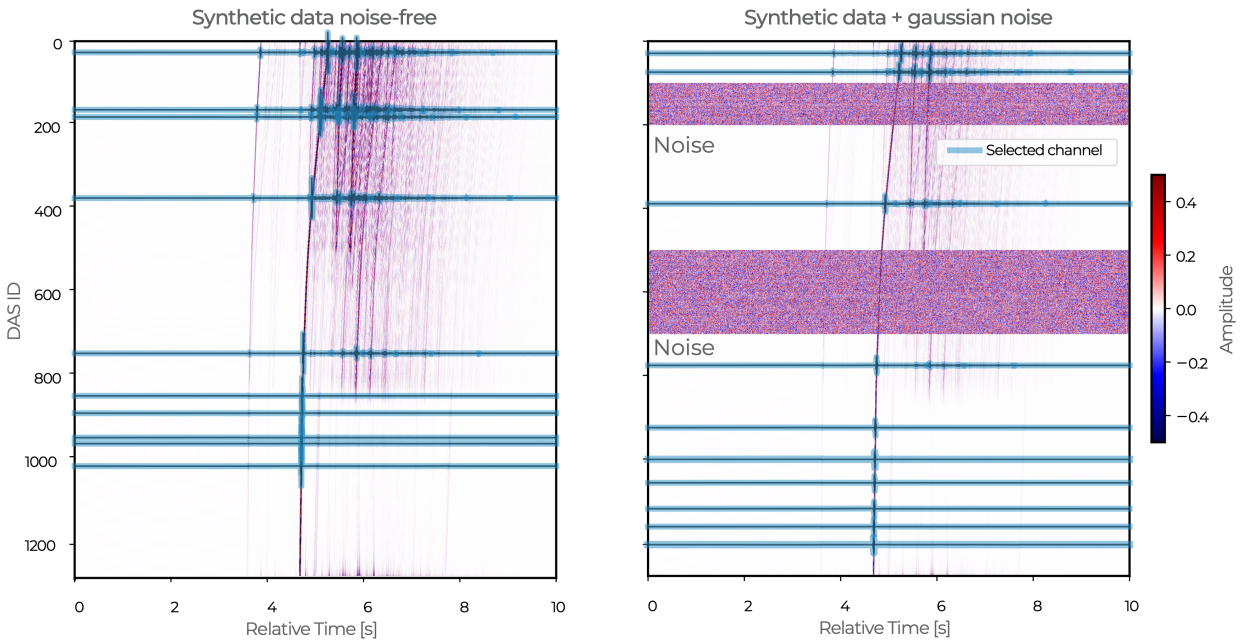

Figure S4 Full-waveform synthetic test of ORION. a) Noise-free synthetic DAS data generated with Salvus (Afanasiev et al., 2018), mimicking an instrumented well and local earthquake event. The 10 traces selected by ORION are highlighted in blue. b) Noise-contaminated synthetic DAS data generated with Salvus (Afanasiev et al., 2018), with the corresponding traces selected by ORION. Note that selected traces concentrate in noise-free areas, but are sufficiently spread along the well.

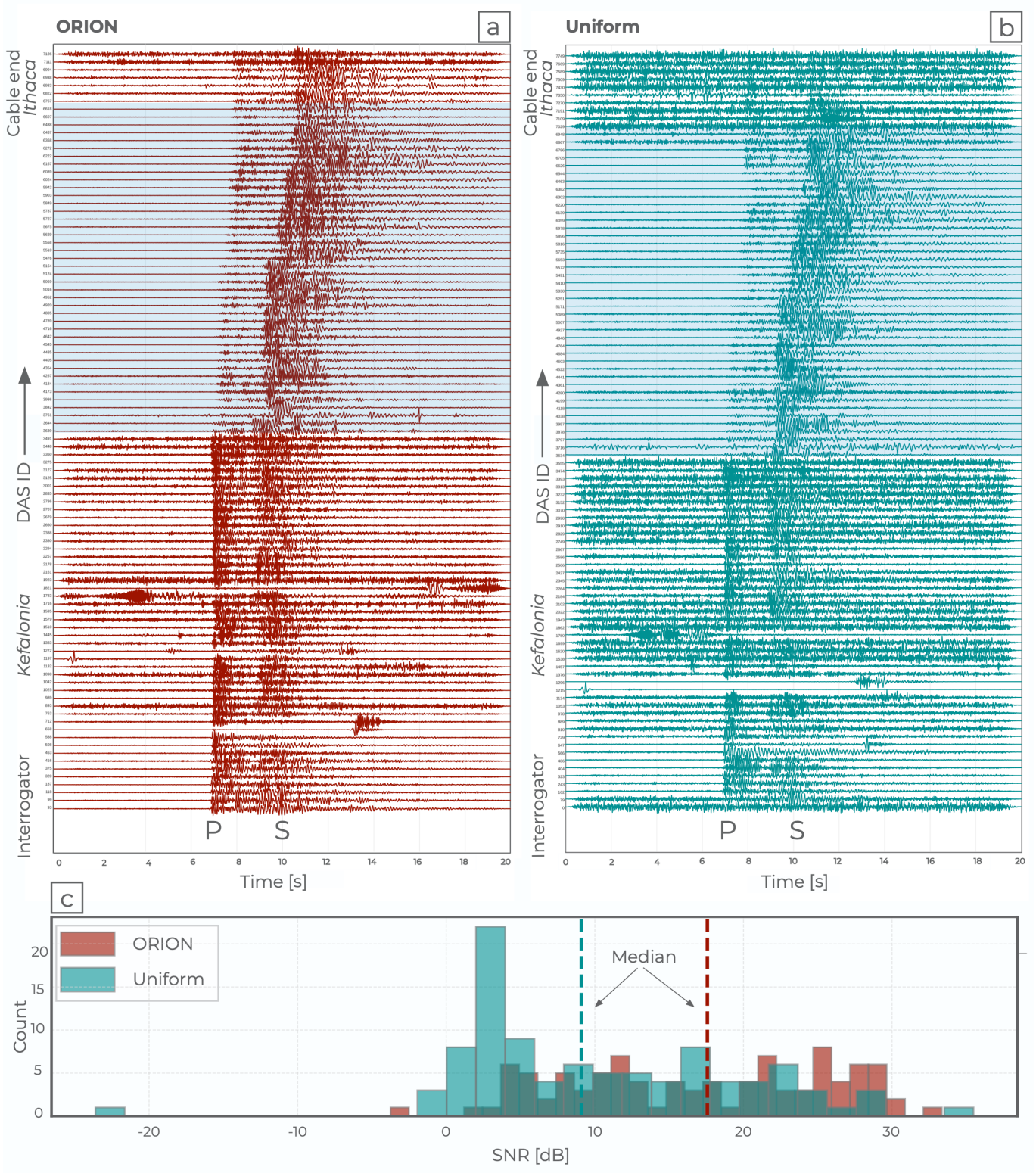

Figure S5 Comparison between ORION-selected DAS channels and a typical fixed-interval (uniform) selection along the cable. a) Automatic DAS channel selection for event ID kef0367 (Bocchini et al., 2025) No fixed final number of channels was imposed; instead, the minimum length of subclusters was set to 50 channels. This results in a reduction by two orders of magnitude compared with the original number of traces (over 7000). b) Selection of an identical number of traces using a fixed channel step along the cable (uniform). The selected DAS channel IDs differ between ORION and the uniform approach. c) Overall, the median SNR is higher for ORION-selected traces, with better channel quality, particularly in terms of the clarity of P- (arrival times around 6-7 s) and S-waveforms (arrival times around 10-11 s) in the land sections (Kefalonia and Ithaca).

	Azuma-Volcano	CANDAS-2	FORESEE	Monterey	POROTOMO	SAFE
Sampling rate [Hz]	200	50	500	100	100 (temporal subsampling from original data)	100
Gauge length [m]	40	10	10	10	10	50
Volume of data (original number of channels)	1364	6587	2120	9993	8649	5981
Band-pass frequency [Hz]	3-20	3-20	3-20	3-20	1-49	3-49
Window SNR-coherence [s]	4	8	8	8	8	8
SNR-coherence-noise weights	1-1-1	1-1-1	1-1-1	1-1-1	1-1-1	1-1-0.3
Azimuth computation chunks	10	10	10	10	5	5
Min number of channels to define a cluster	10	10	10	10	5	5
Neighbour window size	10	10	10	10	5	5
Max number of channels before defining a subsection	50	100	100	50	100	100
Final percentile	1st	1st	1st	1st	1st	1st
Max final number of channels	50	50	50	50	50	50
Spatial sampling reduction factor	~ 27	~ 132	~ 42	~ 200	~ 173	~120

Table S1 Principal features of the DAS datasets used to test ORION on real data (individual events) from six cable deployments. *Gauge-length*: Coherence neighbourhood radius. Each channel is compared with all channels within $\pm$ gauge_length/2 indices. It coincides with the physical gauge length of the DAS acquisition. *Win SNR-coherence (N_w)*: Half-width of the signal window centred on the detected start time, used for both final SNR and coherence computation. Noise window = samples before start time – win/2. *Weights*: Multiplicative weight for the normalised SNR–Coherence–inverted RMS noise component in the combined score. *Azimuth chunk-size*: Window length for azimuth computation. *Min number of channels to define a cluster ($N_{\min}$)*: Minimum neighbours required for a point to be a core point in the spatial sampling step. Points that are not core and not reachable from one are labelled -1 (noise). *Neighbour window size*: Maximum distance allowed when searching neighbours, which prevents distant fiber sections with similar azimuths from merging in the same cluster. *Minimum selected channel distance*: Minimum index separation between any two selected channels, which prevents selection of near-identical adjacent channels from two adjacent clusters. *Max number of channels subsection ($L_{\max}$)* : Sub-samples the full ORION selection by uniform sections if the cluster contains more than this many channels. *Percentile*: Final selection based on the distribution of the normalized final score. *Max final channels*: Maximum number of channels in the final output, selected by score rank across all sections. None keeps all channels passing the percentile threshold.

	Kefalonia
Sampling rate [Hz]	250
Gauge length [m]	10
Volume of data (original number of channels)	7210
Band-pass frequency [Hz]	5-50
Window SNR-coherence [s]	4
SNR-coherence-noise weights	1-1-1
Azimuth computation chunks	50
Min number of channels to define a cluster	50
Neighbour window size	25
Max number of channels before defining a subsection	10 (relatively small, principally to support dense coverage of the marine section, compensating the dense sampling of the onland sections, guaranteed by their complex geometry)
Final percentile	1st
Max final number of channels	None

Table S2 Principal features of the Kefalonia DAS dataset used to test ORION on a seismic sequence. Refer to Table S1 for description of the parameters.