Geophysical Journal International

Geophys. J. Int. (2026) **245,** 1–16
Advance Access publication 2026 February 9
Research Paper

https://doi.org/10.1093/gji/ggag061

DeepSubDAS: an earthquake phase picker from submarine distributed acoustic sensing data

Han Xiao,[1] Frederik Tilmann,[1,2] Martijn van den Ende,[3] Diane Rivet,[3] Afonso Loureiro,[4,5] Takeshi Tsuji,[6] Arantza Ugalde,[7] Qibin Shi[8] and Marine A. Denolle[8]

[1]*Helmholtz Center Potsdam, German GeoResearch Center GFZ, Potsdam,14473, Germany. E-mail: xiaohan@gfz.de*
[2]*Institute for Geological Sciences, Freie Universität Berlin, Berlin,12249, Germany*
[3]*Université Côte d'Azur, CNRS, Observatoire de la Côte d'Azur, IRD, Géoazur, Sophia Antipolis, 06560, France*
[4]*Agência Regional para o Desenvolvimento da Investigação, Tecnologia e Inovação, Funchal, 9020-105, Portugal*
[5]*IDL - Instituto Dom Luiz, Faculdade de Ciências, Universidade de Lisboa, Lisboa, 1749-016, Portugal*
[6]*School of Engineering, The University of Tokyo, Tokyo, 113-8656, Japan*
[7]*Institute of Marine Sciences, ICM-CSIC, Barcelona, 08003, Spain*
[8]*Department of Earth and Space Sciences, University of Washington, Seattle, WA, 98195-131, USA*

Accepted 2026 January 27. Received 2025 December 2; in original form 2025 August 7

SUMMARY

Given the scarcity of seismometers in marine environments, traditional seismology has limited effectiveness in oceanic regions. Submarine Distributed Acoustic Sensing (DAS) systems offer a promising alternative for seismic monitoring in these areas. However, the existing machine learning model trained on land-based DAS data does not perform well with submarine DAS due to differences in noise characteristics, deployment conditions and environmental factors. This study presents a machine learning approach tailored specifically to submarine DAS data to enable automated seismic event detection and *P*- and *S*-wave identification. Leveraging DeepLab v3, a neural network architecture optimized for semantic segmentation, we developed a specialized model to handle the unique challenges of submarine DAS data. Our model was trained and validated on a data set comprising nearly 57 million manually and semi-automatically labelled seismic records from multiple globally distributed submarine sites, providing a robust basis for accurate seismic detection. The model adapts to a variety of deployment scenarios and can process DAS data from cables with different lengths, configurations and channel spacings, making it versatile for various ocean environments. We thus provide an adaptable and efficient tool for automated earthquake analysis of DAS data, which has the potential to enhance real-time earthquake monitoring and tsunami early warning in submarine environments.

Key words: Body waves; Distributed acoustic sensing; Earthquake early warning; Earthquake monitoring and test-ban treaty verification; Seismic instruments.

1 INTRODUCTION

In the last decade, distributed acoustic sensing (DAS) has advanced rapidly, transforming fibre optic cables into dense seismic sensor networks (R. Posey *et al.* 2000). Its high spatial resolution, cost-effectiveness and ease of deployment have expanded its applications across seismology (N.J. Lindsey *et al.* 2017; P. Jousset *et al.* 2018). In addition, the volume of DAS data collected from submarine environments has increased significantly, driven by advances in fibre optic technology and its deployment in underwater settings (A. Sladen *et al.* 2019; N.J. Lindsey *et al.* 2019; B. Romanowicz *et al.* 2023; Z.J. Spica *et al.* 2023; Q. Shi *et al.* 2025). This surge in data presents a unique opportunity to advance earthquake detection and characterization in marine environments (A. Ugalde *et al.* 2021; D. Mata Flores *et al.* 2023; I. Lior *et al.* 2023; C. Strumia *et al.* 2024). Better information on offshore earthquakes is invaluable for studies of fault behaviour, earthquake physics and seismotectonics; this is of particular relevance in subduction zones where often most of the plate interface seismicity occurs offshore and where existing studies frequently suffer from limited backazimuthal coverage. In addition, if submarine DAS data can be processed in

real-time, additional lead time for earthquake and tsunami early warning could be achieved (I. Lior *et al.* 2023; J. Yin *et al.* 2023; M. van den Ende *et al.* 2025).

However, offshore seismic sensing is challenging due to the complex and often uncontrollable underwater environment, where acoustic signals from ocean dynamics overlap with those from tectonic activity. Traditional seismic detection methods for submarine environments, such as the STA / LTA (Short-Term Average / Long-Term Average) trigger algorithm and template matching, face significant challenges when applied to DAS data. The STA/LTA method is highly sensitive to noise and often results in false detections in marine environments due to microseisms, vessel noise, ocean waves and marine life (H. Xiao *et al.* 2022; L. Bouffaut *et al.* 2022; E.F. Williams *et al.* 2023; W.S.D. Wilcock *et al.* 2023; H. Xiao *et al.* 2024; Y. Ni *et al.* 2024), which degrade the signal-to-noise ratio (SNR) of seismic signals (M. Withers *et al.* 1998) and are often also transient. Template matching, while improving detection accuracy, requires pre-defined waveform templates, making it less adaptable to the diverse and complex waveforms observed in submarine DAS recordings (C.E. Yoon *et al.* 2015). At the same time, a nonlinear azimuthal response in DAS must also be accounted for, which can further complicate seismic signal interpretation (R. Posey *et al.* 2000). Denoising before detection, whether in the form of denoising (Q. Shi *et al.* 2025) or wavefield separation (Y. Ni *et al.* 2024), adds complexity to workflows. Furthermore, the shallow marine sediment layer introduces various converted waves, such as basement-converted PS waves, adding another layer of complexity to the signal interpretation (Z.J. Spica *et al.* 2020; A. Trabattoni *et al.* 2024). Meanwhile, the recently developed Kurtosis-Value-Picker (KVP) algorithm may improve seismic event detection and phase picking accuracy in complex DAS and OBS data using multiband kurtosis-based characteristic functions (H. Latorre *et al.* 2025).

Recent advances in machine learning, especially deep learning, have shown promise in addressing these challenges, with studies demonstrating the efficacy of deep learning models in automatically detecting and classifying seismic events (W. Zhu & G.C. Beroza 2018; S.M. Mousavi *et al.* 2019, 2020; W. Zhu *et al.* 2023). However, models trained in terrestrial seismic data often fail (can sometimes be successful) to perform adequately in submarine environments due to different noise characteristics and geological differences (N.A. Ruppert *et al.* 2022; A. Niksejel & M. Zhang 2024; T. Bornstein *et al.* 2024; Fig. S1). Currently, the vast majority of machine learning models are trained on data from traditional three-component seismometers, making them unsuitable for direct application to single-component DAS data.

They are further impractical for large-scale deployment on DAS, either due to their reduced performance or to computational expense (W. Zhu & G.C. Beroza 2018; Q. Shi *et al.* 2025). To our knowledge, currently there is only one model trained on DAS data, PhaseNet DAS, which is based on three terrestrial DAS data sets acquired in California, which limits its applicability in submarine environments (W. Zhu *et al.* 2023). Recognizing these challenges, there is a clear need for specialized models capable of accurately labelling submarine DAS data.

To address these challenges, we propose training a deep learning model specifically on submarine DAS data. Specifically, we adapt the DeepLab model, which is known for its robust performance in image segmentation tasks, as a promising approach for seismic wave detection and classification in submarine environments (L.C. Chen *et al.* 2017). The development of such a model has the potential to improve our understanding of submarine seismic activity and contribute to more effective earthquake monitoring and early warning systems (J. Yin *et al.* 2023).

In this study, we train our model using 57 million DAS records, which were collected from 6314 seismic events across 13 submarine fibre optic cables worldwide. To build a robust training data set, we developed an interactive semi-automated tool to label the *P* and *S* waves in DAS data. This tool improves the accuracy and efficiency of the labelling, creating high-quality manually labelled data essential for adapting DAS models to the complexities of the submarine environment. Through this work, our goal is to advance seismic monitoring capabilities in underwater settings, bridging the gap between terrestrial and underwater DAS applications.

2 DATA AND MODEL

2.1 Curated DAS data set

To ensure that our machine learning model applies to various oceanic environments worldwide, we have made a concerted effort to gather submarine DAS seismic data from multiple global locations. Our study uses DAS data collected from 13 submarine fibre optic cables around the world, covering nearly 92 million seismic records (Fig. 1a). Additionally, we incorporated land-based DAS earthquake records from the Ridgecrest region in California to pre-train our model. The following sections introduce the primary data sources and their key characteristics.

The first data set, which is the largest data source, comes from two optical fibres along Alaska (Fig. 1b; Q. Shi *et al.* 2025). In June 2023, two submarine telecom cables were connected to a single Sintela Onyx v1.0 interrogator unit at the GCI Communication landing station in Homer, Alaska. The two fibre-optic cables include the southern span of the Kodiak Kenai Fiber Link (KKFL-S), oriented north–south, and the east–west oriented TERRA (Terrestrial for Every Rural Region in Alaska) cable. The strain recordings were decimated from a sampling rate of 1.25 kHz to 25 Hz to reduce storage demands. The channel spacing is 9.57 m, with a gauge length of 17.55 m from June to September and 23.93 m from September to December. Each cable spans 81.64 km, with 8531 channels in each DAS data array. During this period, nearly 6000 earthquakes occurred in the vicinity of 200 km, generating close to 80 million single-channel earthquake records (Fig. 1a, Table 1). The earthquakes were detected by the United States Geological Survey Advanced National Seismic System as part of standard operations of the Alaska Earthquake Center network.

Fig. 2(a) shows the distribution of the SNR (db) for all seismic records. The SNR is defined here as the ratio between the maximum absolute amplitude of the seismic signal and the root-mean-square (RMS) amplitude of the noise in the 3 s preceding the event. The signal amplitude is determined based on manually labelled *P* and *S* wave arrival times: we extract a 1 s window centred around each arrival (0.5 s before and after) and use the maximum absolute value within that window. The signal arrival times are identified from manual picks and used directly to define the time window for each phase. This approach ensures

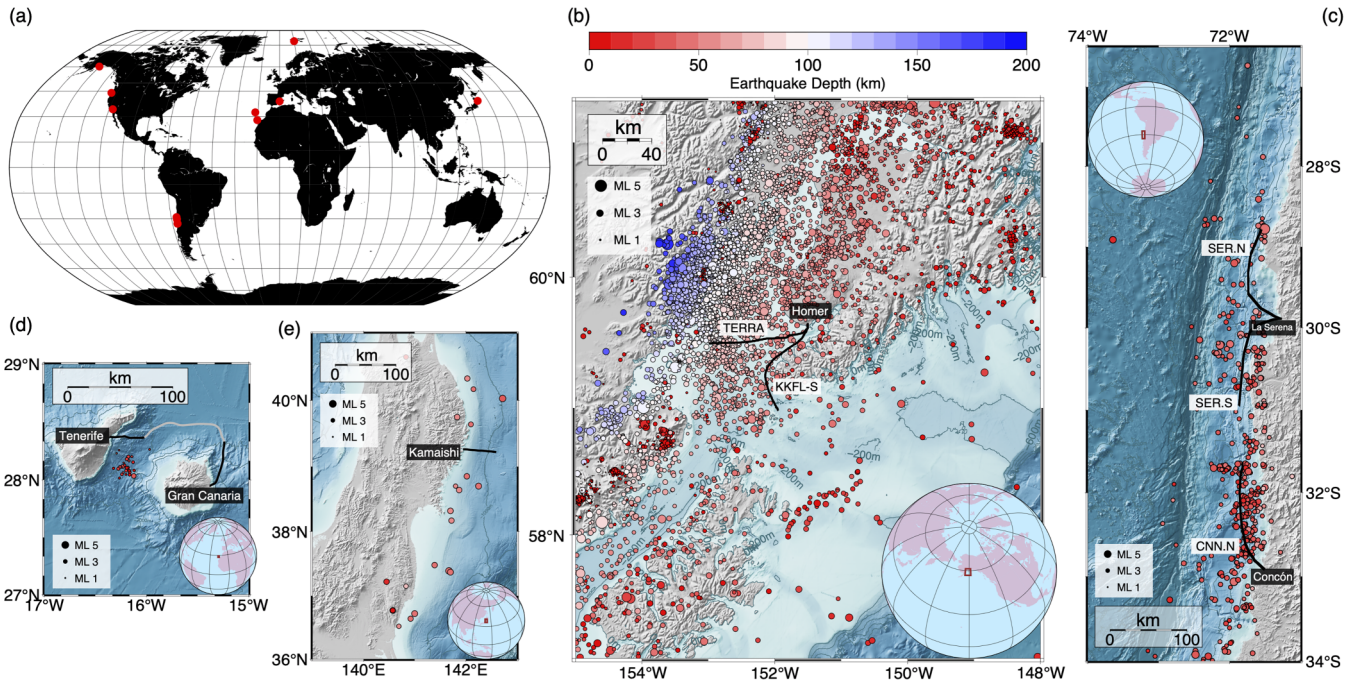

Figure 1. (a) Map showing the locations of all Distributed Acoustic Sensing (DAS) cables used in this study. Each red dot represents the position of a fibre optic cable. (b) The largest data set is from the southern coastal region of Alaska, with two black lines marking the positions of the TERRA and KKFL-S DAS fibres. Earthquake locations are depicted by circles, with colour indicating depth and size representing magnitude. (c) Another primary data set comes from the Chilean coastline, where three cables CCN.N, SER.S and SER.N are shown by black lines alongside earthquake locations. (d) The locations of two DAS cables and corresponding earthquake locations in the Canary Islands. (e) The DAS cable and earthquake locations in Kamaishi, Japan.

Table 1. Summary of DAS data set statistics.

Location	Year	Region	Channel	Event No.	Records	Citation
KKFL-S/Homer	2024	Alaska	8 531	5985	40 M	Q. Shi *et al.* (2025)
TERRA/Homer	2024	Alaska	8 531	5985	40 M	Q. Shi *et al.* (2025)
CCN.N/La Serena	2021	Chile	36 718	41	1.5 M	C. Strumia *et al.* (2024)
CCN.N/La Serena	2023	Chile	15 000	229	3.4 M	M. van den Ende *et al.* (2025)
SER.S/Michilla	2023	Chile	15 000	229	3.4 M	M. van den Ende *et al.* (2025)
SER.N/Michilla	2023	Chile	15 000	229	3.4 M	M. van den Ende *et al.* (2025)
Tenerife	2020	Spain	5 984	28	0.17 M	A. Ugalde *et al.* (2021)
Gran Canaria	2020	Spain	5 984	28	0.17 M	A. Ugalde *et al.* (2021)
Kamaishi	2020	Japan	6 000	23	0.14 M	T. Tsuji *et al.* (2021)
Svalbard	2020	Norway	30 000	3	0.09 M	L. Bouffaut *et al.* (2022)
Valencia	2020	Spain	2 977	2	0.006 M	H. Xiao *et al.* (2022)
Florence	2021	Oregon	3 000	2	0.006 M	H. Xiao *et al.* (2024)
GeoLAB/Madeira	2023	Portugal	11 294	1	0.001 M	A. Loureiro *et al.* (2025)
Total	–	–	116 748	6314 (Unique)	92.0 M	–

that the SNR calculation accurately reflects the observed signal characteristics rather than relying on theoretical traveltime estimates.

As shown in Fig. 2(a), the number of *S* wave records is more than double that of *P* wave records. In addition, the *S* wave generally has a slightly higher SNR compared to the *P* wave, with most *S* wave SNRs around 8, while the *P* wave SNRs are mostly around 6. Fig. 3(a) and Fig. 4(a) present the frequency of earthquake magnitudes and depths. The majority of earthquakes have magnitudes between 1 and 3, with depths less than 150 km.

The second largest data set comes from three DAS recordings along the Chilean coastline (Fig. 1c). In November 2021, DAS data were recorded in Chile using an OptoDAS interrogator unit from Alcatel Submarine Networks. This unit was connected to a submarine fibre optic telecom cable operated by the GTD group, spanning from Concón to La Serena (CCN.N) (C. Strumia *et al.* 2024). The DAS system monitored a 150 km-long segment of this cable. The data were initially recorded with an 8.16 m gauge length, a sampling rate of 625 Hz, and a spatial sampling interval of 1.02 m, averaged over 4.08 m, resulting in a total of 36 718 channels. The data were later downsampled to a 125 Hz sampling rate. In total, we collected 41 earthquakes comprising 1.5 million seismic records.

In a subsequent experiment along the Chilean coast between 2023 and 2024, DAS data were acquired from three fibre optic cables running approximately parallel to the Chilean coastline for most of their span (M. van den Ende *et al.* 2025). The DAS system monitored three 150 km-long segments of these cables

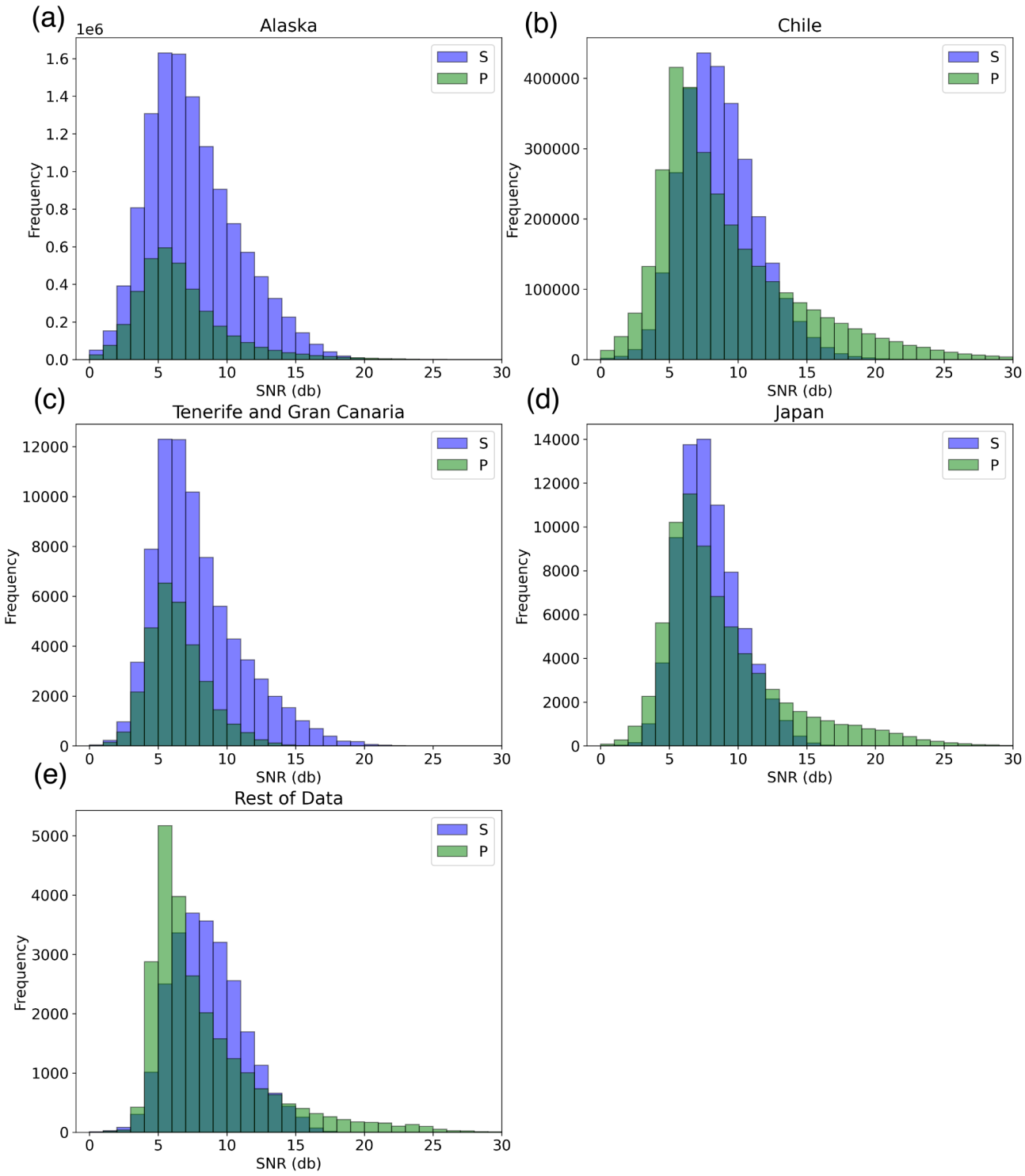

Figure 2. Signal-to-noise ratios (SNR db) of *P* wave and *S* wave recorded by seafloor DAS across various regions. Note that the SNR follows a decibel scale, that is, a value of 20 corresponds to an SNR of 100. (a) The southern coastal region of Alaska. (b) The coastal region of Chile. (c) The area between Tenerife and the Gran Canaria Islands. (d) The offshore region in northeastern Japan. (e) The remaining data in Table 1, excluding data from (a)–(d).

(CCN.N, SER.S and SER.N) (Fig. 1c). The data were recorded with a spatial sampling interval of 10 m and a frequency of 125 Hz. It recorded 229 earthquakes during this period, resulting in 10 million seismic records across the three fibres. As shown in Fig. 2(b), similar to the Alaska data, the *S* wave generally exhibits a higher SNR, with most *S* wave SNRs around 10, while most *P* wave SNRs are around 6. However, *P* wave reaches the highest SNR values, with some exceeding 30. This may be related to our calculation method, as the noise preceding the *P* wave is relatively low, whereas the coda of the *P* wave can affect the SNR estimation for the subsequent *S* wave. Figs 3(b) and 4(b) display the frequency distribution of earthquake magnitudes and

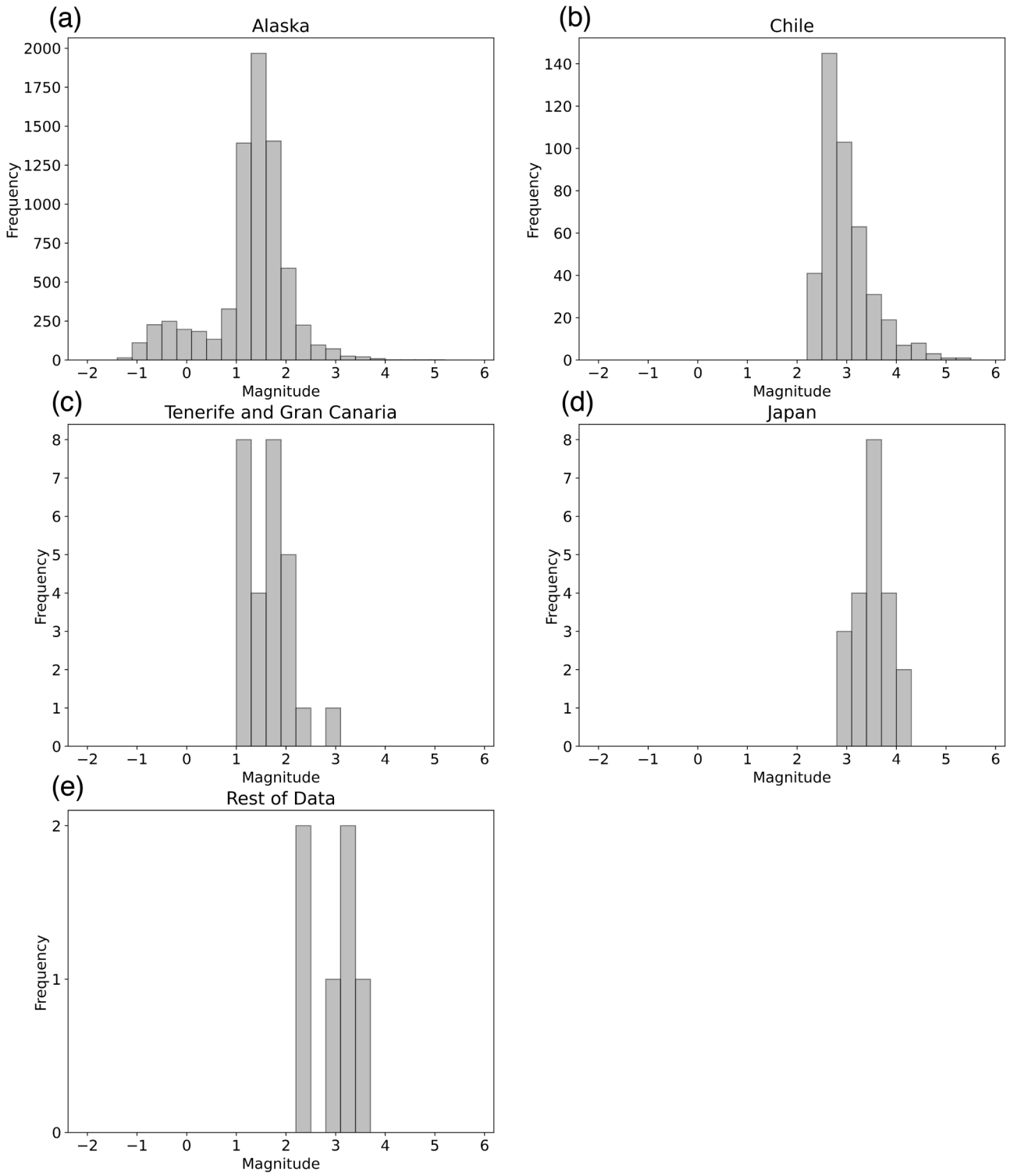

Figure 3. Earthquake local magnitude (ML) distribution. (a) The southern coastal region of Alaska. (b) The coastal region of Chile. (c) The area between Tenerife and the Gran Canaria Islands. (d) The offshore region in northeastern Japan. (e) The remaining data in Table 1, excluding data from (a)–(d).

depths. The majority of earthquakes in these data sets have magnitudes between 2 and 6 and are shallower than 100 km.

The third major data set was collected from an undersea fibre-optic telecommunication cable that links the islands of Tenerife and Gran Canaria. Two DAS interrogators monitored 5984 channels each, with a spatial resolution of 10 m (channel length), covering the first 60 km of fibre starting from the coastline and extending into the sea (A. Ugalde *et al.* 2021). The data were initially recorded with a sampling frequency of 1 kHz and were later downsampled to 50 Hz. The DAS systems recorded 28 earthquakes, comprising 330 064 seismic records. In this data set, the number of *S* wave records exceeds that of *P* wave, and

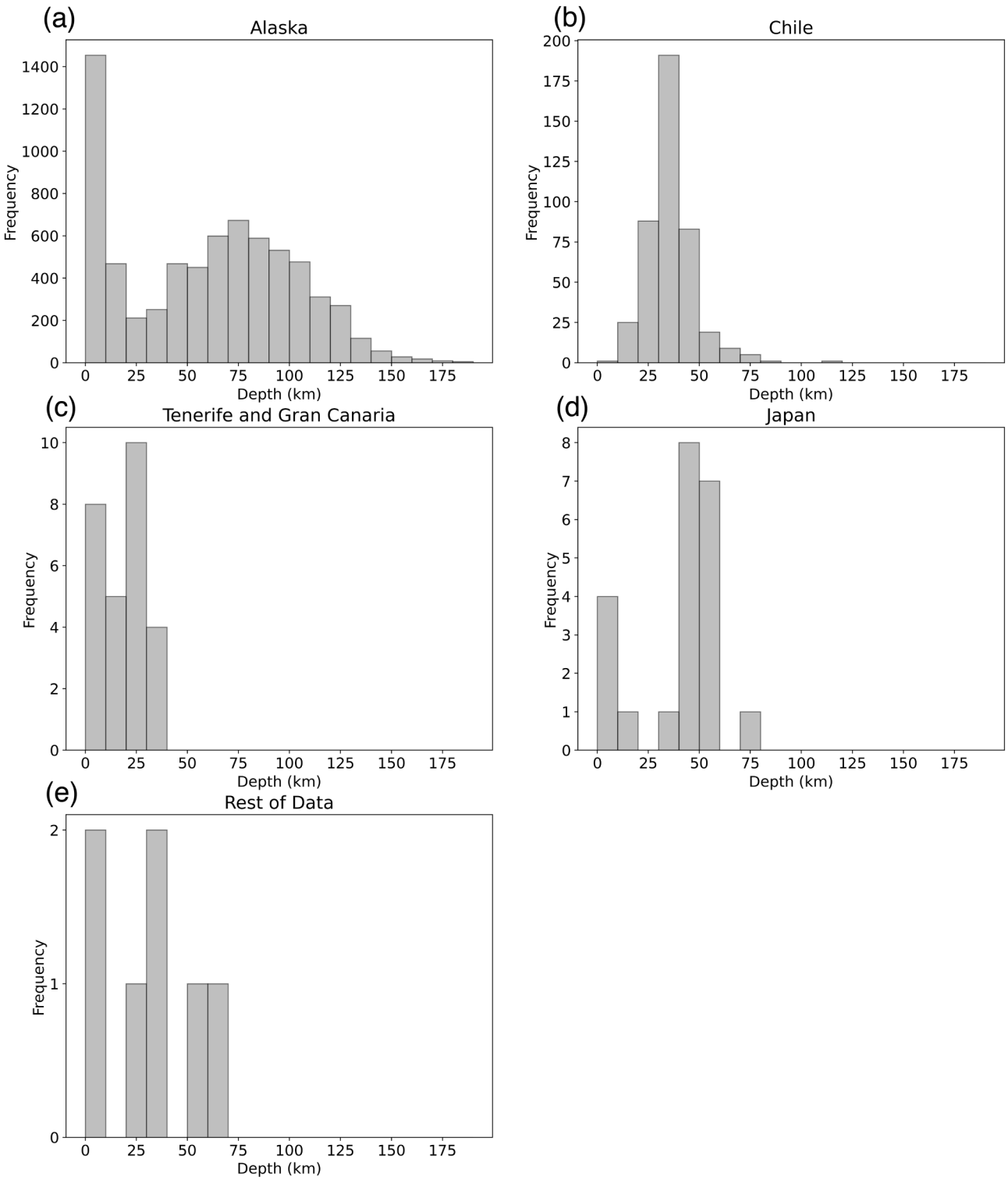

Figure 4. Earthquake depth distribution in the catalogues. (a) The southern coastal region of Alaska. (b) The coastal region of Chile. (c) The area between Tenerife and Gran Canaria Islands. (d) The offshore region in northeastern Japan. (e) The remaining data in Table 1, excluding data from (a)–(d).

the *S* wave also exhibits a significantly higher SNR (Fig. 2c). Similarly, Fig. 3(c) and Fig. 4(c) show the distribution of earthquake magnitudes and depths from these catalogues, with magnitudes ranging from 1 to 3 and depths down to 50 km.

The fourth data set comes from the coast of Japan. We utilized DAS data recorded from the seafloor off the coast of Kamaishi, northeast Japan, collected between 2020 October 10 and November 23 (T. Tsuji *et al.* 2021). In this test survey, the spatial sampling interval was set to 10 m, with a gauge length of

20 m. The cable used in this experiment was 60 km long, providing data from 6000 channels. The sampling rate is 500 Hz in this offshore experiment. In total, it recorded 23 earthquakes, comprising 138 000 seismic records. The SNR of earthquakes in the Japan region is slightly lower than that along the Chilean coast although the magnitude distribution is similar (Figs 3b and d), likely due to the larger epicentral distances of the earthquakes in this area. In general, all the recorded earthquakes are shallow earthquakes 0 to 100 km (Fig. 4d).

We also utilized data from cables located in Svalbard (L. Bouffaut *et al.* 2022), Oregon (H. Xiao *et al.* 2024), and off the coast of Spain in the Mediterranean (H. Xiao *et al.* 2022), as well as data from the GeoLAB fibre in Madeira (A. Loureiro *et al.* 2025). However, since these locations recorded less than five earthquakes each, we only provide summarized descriptions of these data sets in the table. For more information on the seismic data from these locations, please refer to the previously published articles (Table 1).

To enhance our data set and improve the stability of our model, we augmented the data by incorporating real oceanic DAS seismic noise. These noise samples were carefully selected, ensuring that they were recorded during periods without seismic events. We employed the kurtosis function to determine whether the noise contained any seismic signals.

The kurtosis K is calculated using the following formula:

$$K = \frac{\sum (x_i - \bar{x})^4}{n \cdot \sigma^4} - 3,$$

where x_i is the i-th data point in the data set. $\bar{x}$ is the mean of the data set. n is the total number of samples in the data set. σ is the standard deviation of the data set. To maintain the coherence in the seismic waveforms across different channels, we did not add noise to each channel individually. Instead, we added noise with the same array size to the earthquake waveforms in each patch. This method more closely mimics the conditions of our real seismic recordings, ensuring that the augmented data is more representative of actual DAS observations. Notably, these noise samples were sourced from various offshore DAS cables located around the world, further diversifying our data set. Additionally, the noise was normalized, with levels randomly added ranging from 10 per cent to 50 per cent of signal amplitudes. This variability in noise levels enhances the robustness of our model by exposing it to different SNRs, ultimately improving its ability to detect small earthquakes.

2.2 Semi-automatic arrival time pick labelling

Although PhaseNet, which was originally trained on seismometer data, can sometimes produce reasonable phase picks on DAS recordings, its performance is inconsistent. In some cases, the picks generated by PhaseNet exhibit timing errors or fail to detect certain phases, particularly for low SNR or complex wavefield conditions commonly observed in DAS data. Therefore, while seismometer-trained models such as PhaseNet can serve as a useful starting point or reference, their outputs require careful adjustment or manual verification before being used for DAS applications. To ensure the quality and accuracy of our labelled data, we adopted a semi-automatic labelling approach. Initially, we applied PhaseNet (W. Zhu & G.C. Beroza 2018), using the SeisBench (J. Münchmeyer *et al.* 2022; J. Woollam *et al.* 2022) platform, to perform single-channel labelling, leveraging its efficiency in identifying *P* and *S* seismic waves. We applied a bandpass frequency filter in the range of 0.5 to 20 Hz, as well as an FK (frequency–wavenumber) filter with a slowness window corresponding to apparent velocities between 300 and 50 000 m s^{-1}. Since PhaseNet is designed to take a three-component input, but DAS data contains only a single component, we duplicated the single DAS channel across all three input channels. We used the original pre-trained weights provided with PhaseNet. Whereas in many cases, this simplistic approach already resulted in reasonable picks, there were also many instances where the no arrival was identified, *P* and *S* waves were not identified correctly, or the wrong part of the waveform was picked (Q. Shi *et al.* 2025). Therefore, to enhance the reliability of these labels, we developed an interactive software tool that allows manual correction of labelling results. We manually intervened every 5–10 channels, depending on the channel spacing for each DAS cable, correcting errors that might have arisen due to excessive noise on single-component DAS data (Fig. 5a). The arrival times of the *P* and *S* waves for the channels in between were determined through linear interpolation. We also conducted manual checks to ensure the accuracy of these interpolated times, and if any discrepancies were found, we intervened and corrected them manually (Fig. 5b). This hybrid method enabled us to efficiently and accurately label *P*- and *S*-wave seismic records, ensuring more accurate labelled data for subsequent analysis (Fig. 5c).

Due to the low SNR ratio in some DAS seismic data, there are instances where even manual labelling cannot accurately determine the *P*- and *S*-wave arrival times (SNR < 3). In such cases, we choose to discard this data to maintain the overall quality and accuracy of the data set.

2.3 Training data set preparation

Only the highest quality seismic records within the frequency range of 1 to 20 Hz, exhibiting clear waveforms and minimal noise interference, were selected for further analysis. All training data were carefully bandpass-filtered within this range to ensure consistency and optimal model performance. This frequency band effectively suppresses ocean wave noise, which predominantly occurs below 0.2 Hz (A. Sladen *et al.* 2019; N.J. Lindsey *et al.* 2019; H. Xiao *et al.* 2024), and mitigates interference from Scholte wave seismic noise, typically observed between 0.1 and 2 Hz (H. Xiao *et al.* 2025). Users are therefore advised to apply the same 1 to 20 Hz filtering when preparing their own data to ensure compatibility with the trained model. Filtering within this range preserves essential seismic information while discarding irrelevant background noise.

To standardize the data for analysis and model training, we resampled all waveforms to a uniform rate of 100 Hz, aligning with typical requirements for seismic analysis (e.g. default sampling rate in SeisBench models) while maintaining sufficient temporal resolution for accurate signal interpretation. However, we preserved the original channel spacing of each DAS data set to retain the different inherent spatial resolutions and specific characteristics of the seismic signals. This balance between standardization and preserving unique data set properties is critical for ensuring the generalization of the model.

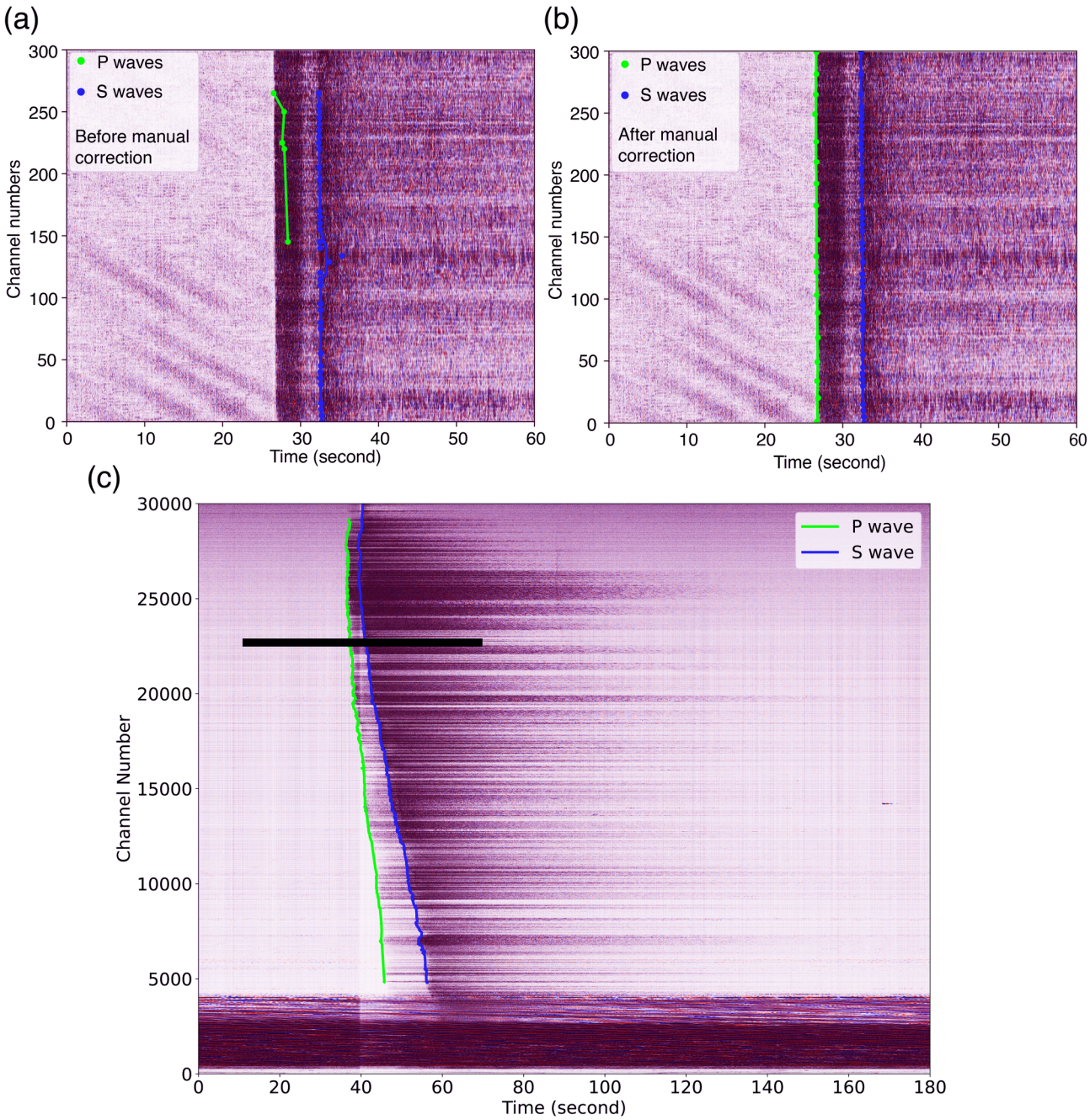

Figure 5. An example of semi-automatic labelling of *P* and *S* waves from the Chile data set. (a) We initially labelled the *P* and *S* waves using PhaseNet from the SeisBench platform, (b) then manually corrected any mislabelled data through our interactive software tool (c) The final label result from this semi-automatic process. The channel range in (a) and (b), corresponds to channels 22 500–22 800 in (c). We show the results in panel (c) using black lines only for clarity.

Subsequently, the *P*-wave arrival time in the middle channel of each patch was used as a reference. From these data, we extracted 40 s waveform segments, with start times randomly chosen between 2 and 10 s before the manually picked *P*-wave arrival in the middle channel. This method aligns with established data preparation techniques and ensures that the extracted segments capture a mix of seismic signals and background noise. We also randomly added real DAS seismic noise chosen before to the earthquake data. Incorporating this noise is essential for training models capable of robust detection across a variety of SNR conditions. To ensure balanced representation in the training data, we exclusively selected earthquake records containing both *P* and *S* waves, allowing the model to effectively learn the characteristics of both phases. This selection was necessary because, in the Alaska data set (Fig. 2), *S*-wave samples are far more numerous than *P*-wave samples; using all available data would have resulted in a heavily biased training set.

The processed data were then organized into manageable patches consisting of 200 to 500 DAS channels, with the specific number determined by the spacing between the channels to

ensure optimal segmentation for the data set. For training purposes, we exclusively used labelled earthquake data within these patches to guarantee the model learns from well-defined examples, avoiding any inclusion of unlabelled segments.

2.4 Training data augmentation

To further augment the data set, we applied slight temporal stretching and compression (less than 10 per cent) along the time axis. This approach is inspired by techniques used in image segmentation, where geometric transformations are applied to introduce variability in the training data without altering its core features and has been successful for training phase pickers across a wide range of frequencies Q. Shi & M.A. Denolle (2023). This technique simulates potential interchannel timing variations in real seismic recordings on the seabed due to the sediments and improves the model's ability to generalize across different ocean environments.

2.5 Test data set

To evaluate the performance of our model, we conducted a comprehensive comparative analysis using distinct test data sets. Specifically, we extracted 10 per cent of the data as test data and another 10 per cent as validation data. These subsets were randomly selected from the different DAS data sets to ensure a diverse representation, capturing a wide range of seismic event characteristics and environmental conditions. By including data from multiple regions and varying signal qualities, this approach enables a robust and generalized evaluation of the model's ability to detect seismic events under different circumstances. The random selection helps mitigate any biases that might arise from overfitting to specific subsets of the data, thus ensuring that the model's performance is a reliable reflection of its real-world applicability.

2.6 Model

We utilized the DeepLab model for our machine learning approach. DeepLab is a deep learning model designed specifically for semantic image segmentation, where the goal is to classify each pixel in an image into pre-defined categories (L.C. Chen *et al.* 2017). It leverages several advanced techniques to achieve high accuracy in segmenting complex images. One of the core innovations in DeepLab is the use of atrous convolutions, which allows the model to capture multiscale contextual information without losing resolution (Fig. 6). This technique effectively increases the receptive field of the network without adding extra parameters, making it particularly powerful for handling varied and complex input data.

We believe that atrous convolutions are especially well-suited for processing different DAS data, as each DAS setup can have different channel spacings. By leveraging atrous convolutions, DeepLab can better accommodate these variations in channel spacing, leading to improved performance and adaptability in analysing diverse DAS seismic data sets. Another key feature of DeepLab is its use of the fully connected conditional random field (CRF) as a post-processing step, which refines the boundaries of segmented objects, making them more precise and better aligned with actual object contours. The model also benefits from employing backbone networks like ResNet, which enhance its ability to learn deep and rich feature representations. Due to these innovations, DeepLab has demonstrated exceptional performance in image segmentation tasks across various challenging data sets.

3 TRAINING

We pre-trained our model using terrestrial DAS data from Ridgecrest, California (761 earthquakes and 6 316 300 records), capitalizing on the similarities in seismic signal characteristics (AI4EPS 2023; W. Zhu *et al.* 2023). Labelling this training data was also accomplished using our semi-automated labelling tool, ensuring the accuracy and consistency of the data set. Although seafloor DAS signals exhibit notable differences from land DAS data due to oceanic and environmental influences, there are also fundamental similarities in their underlying seismic wave characteristics. Therefore, we utilized land DAS data to pre-train our model before fine-tuning it with seafloor DAS data. This approach allows the model to learn general seismic feature representations from the accessible land data sets, which can then be effectively adapted to the marine environment. Such transfer learning strategies are widely adopted in other domains—for example, in medical image segmentation, where models pre-trained on natural image data sets (e.g. ImageNet) are fine-tuned for medical imaging tasks (H.E. Kim *et al.* 2022), demonstrating that leveraging shared features across different data domains can significantly improve model performance. This pre-training step is crucial, as it enables the model to rapidly learn the fundamental aspects of earthquake detection, such as identifying and labelling *P* and *S* waves, providing a strong foundation that can be effectively transferred to the more challenging task of submarine seismic detection.

To adapt the DeepLab framework for DAS data, we modified its architecture to accommodate single-channel 2-D inputs, representing the spatiotemporal features of DAS seismic data (instead of three channels in image analysis). The output layer was reconfigured to classify each input segment into one of three categories: *P* waves, *S* waves and noise. We tried different methods and found that this approach achieved the fastest training speed and the best convergence. This adjustment ensures that the model is specifically optimized for the characteristics of DAS data sets, which differ significantly from traditional three-component seismic data but also pose different structures than required in classic image analysis. For labelling *P* and *S* wave arrivals, we represent them using Gaussian functions, transforming each arrival into a probability distribution. This Gaussian representation allows for more flexible and probabilistic handling of arrival times compared to a single-point label. We can adjust the values of σ independently in both the x- and y-directions to control the width of the Gaussian distribution based on their specific needs. A smaller σ will produce a sharper, more precise peak, while a larger σ provides a broader and more tolerant representation of the arrival time. We use a larger $\sigma = 60$ in the early stages of training to help better convergence, and a smaller $\sigma = 20$ in the later stages of training to achieve higher precision.

We implemented a cross-entropy loss function, which is well-suited for classification tasks and effectively measures the divergence between the predicted probabilities and the true labels. To optimize the learning process, we employed the Adam optimizer

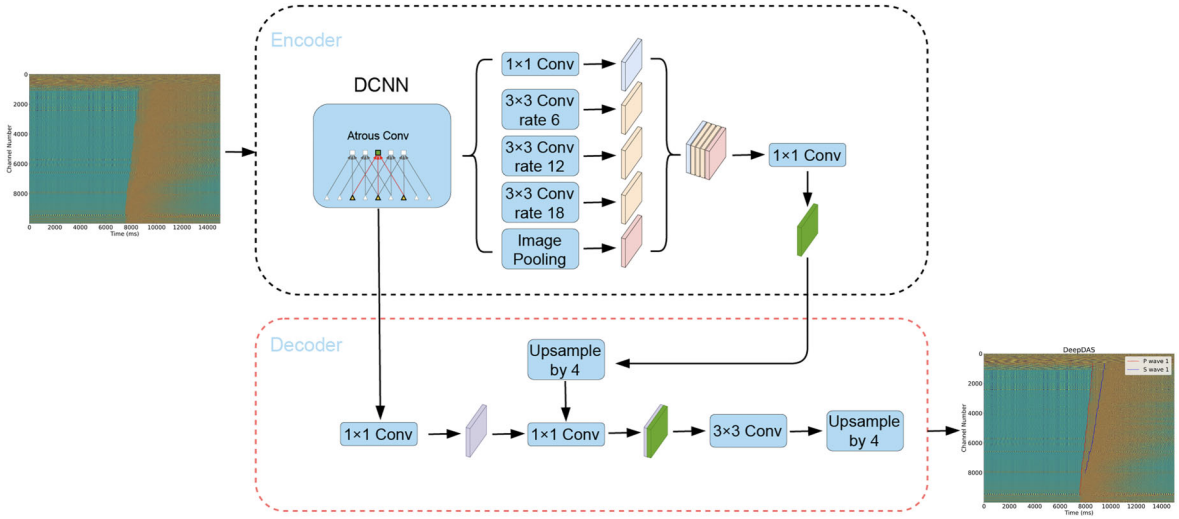

Figure 6. Architecture diagram of DeepLab v3. The model utilizes a Deep Convolutional Neural Network (DCNN) backbone (ResNet) for feature extraction, followed by an Atrous Spatial Pyramid Pooling (ASPP) module that captures contextual information at multiple scales. 'Conv' denotes convolutional layers. The ASPP module combines parallel atrous (dilated) convolutions with different rates to effectively enlarge the receptive field without increasing the number of parameters. A final decoder module refines the output predictions, enabling accurate phase localization in noisy DAS data.

(K. Diederik 2014), known for its adaptive learning rate capabilities, which helps to accelerate convergence while minimizing the risk of overshooting the optimal solution.

To further enhance model robustness and generalization, we applied data augmentation techniques mentioned before during training. These included adding real seismic noise to the DAS signals, slight temporal stretching and compression, and introducing temporal shifts. Specifically, the temporal shifts were implemented by randomly selecting the starting point of the input time window within the 10-s interval preceding the *P* wave arrival. The shifts followed a uniform distribution and were generated once per patch based on the first channel, then applied consistently across all channels within that patch to preserve spatial coherence. These augmentations ensured that the model was exposed to a wide range of conditions, preparing it to handle diverse and noisy submarine seismic data effectively.

Initially, we set the learning rate to 1×10^{-3} for the first five epochs to facilitate rapid learning during the early training phase. After this initial period, we reduced the learning rate to 1×10^{-5} for the subsequent 20 epochs, allowing for more fine-tuned adjustments as the model converged.

Additionally, we implemented a spatial dropout rate of 0.2 to prevent overfitting by randomly deactivating a portion of the units during training. This technique improves the robustness of the model by encouraging it to learn more general features. To further mitigate the risk of overfitting, we also implemented an early stopping mechanism that monitors the model's performance on the validation set (L. Prechelt 2002). This mechanism halts training when the validation loss begins to increase, ensuring that the model retains its ability to generalize well to unseen data.

For the main training phase, we employed four NVIDIA A100 GPUs, each with 80 GB of memory, providing good computational power and memory capacity to efficiently handle large-scale data sets. Prior to the main training, the model was pre-trained for 1 hr using terrestrial DAS data to establish a stable initialization, followed by training on underwater DAS data. It is important to note that during inference, the model does not require the same level of hardware resources; a single A100 GPU—or even one with lower memory—can adequately support model execution. The full training process lasted for 24 hr, during which we continuously monitored key performance metrics and made necessary adjustments to ensure stable convergence and robust generalization.

4 PERFORMANCE EVALUATION

We evaluated our model's performance using recall, precision and F1 score. The formulae for these metrics are as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$

where TP is the number of true positives and FP is the number of false positives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

where FN is the number of false negatives.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

These metrics provide a comprehensive evaluation of the model's ability to correctly identify positive instances.

The test results for Chile, Alaska and the other data sets are summarized in Table 2 when the threshold is set to 0.8. For *P*-wave detection in the Chilean data set, the model achieved a precision of 0.95 and a recall of 0.93, significantly outperforming its performance on the Alaska data (approximately 0.8 for both metrics). This improved performance may be attributed to the relatively larger earthquake magnitudes in the Chilean region and the noisier data in the Alaska region, which is more affected by the local storms. These factors result in more distinct and easily identifiable *P*-wave onsets. For *S*-wave detection, the model exhibited consistent performance across different regions, with precision, recall and F1 scores all around 0.88. This indicates that the model is relatively robust in identifying *S* waves, regardless of regional variations in earthquake characteristics. We also

Table 2. Performance comparison on Chilean and Alaskan coastal data. Outliers are defined as picks with absolute error > 1 s.

Region	Phase	Precision	Recall	F1	MAE (s)	Outlier (per cent)
Chile	P	0.95	0.93	0.94	0.22	3.2
	S	0.96	0.82	0.88	0.15	2.1
Alaska	P	0.88	0.79	0.83	0.21	4.2
	S	0.91	0.89	0.90	0.18	3.7
Other data sets	P	0.87	0.76	0.81	0.20	6.2
	S	0.88	0.85	0.86	0.17	7.1

evaluated the model on data sets from other data sets including Svalbard, Madeira and Valencia. The performance on these data sets is slightly lower than that on the main training regions, indicating that the model may be influenced by the dominant characteristics of the primary training data (Chile and Alaska). This result suggests that while the model generalizes reasonably well, further improvement could be achieved by incorporating more diverse data sets from different regions.

In addition, we compared the performance of DeepSubDAS and (original) PhaseNet DAS in picking seismic *P* and *S* waves from submarine DAS data (see Fig. 7). Our findings suggest that DeepSubDAS outperforms PhaseNet DAS in those examples. This result is understandable, as PhaseNet DAS was originally trained on DAS seismic data from land-based sources. PhaseNet DAS tends to exhibit a higher rate of false picks, which could be attributed to the differences in noise characteristics between marine DAS seismic data and those from terrestrial environments. It is worth noting that for the correctly picked *P* and *S* waves by PhaseNet DAS (Fig. 8, Table S1), the picking error is slightly higher than that of DeepSubDAS (Fig. 9), with most errors within 0.5 s. This indicates that, although PhaseNet DAS tends to produce many more false positive detections, the time error of its true positive picks remains within a reasonable range; however, there are still many outliers.

In the case of the Alaska data set, PhaseNet DAS produced erroneous picks for both *P* and *S* waves (Fig. S1 and Table S1). The inaccuracies in picking *P* and *S* waves for the Alaska data set are likely due to the short duration of the pre-event noise data. A longer time-series of background seismic noise prior to the event could provide more contextual information, potentially benefiting the performance of both models. This difference in performance highlights the importance of training models specifically for the offshore DAS and land-based data to which they will be applied, as well as the need for more diverse ocean-based training data sets to improve robustness.

Moreover, we observed that DeepSubDAS is more sensitive to local time variations caused by seafloor sediment layers, whereas PhaseNet DAS tends to produce smoother picks (Fig. 7). PhaseNet DAS's tendency toward smooth picks could be beneficial in more stable environments but becomes a disadvantage in complex, sediment-heavy seafloor regions where abrupt time delays or scattering may occur and should be visible in the picking time.

In addition, we found that DeepSubDAS is capable of detecting some signals with low SNRs that humans might miss (Fig. S2). Additionally, for some low SNR seismic records, DeepSubDAS outperforms semi-automated manual labelling (Figs S3a and S3b). This advantage likely arises from the interpolation methods used in semi-automated labelling, which can result in significant onset deviations in low SNR conditions, while human analysts struggle to correct these inaccuracies effectively. Furthermore, human fatigue during the labelling of large data sets can introduce errors (Figs S3c and S3d), which further highlights the model's stability and reliability, especially in challenging data conditions.

Fig. S5 illustrates the performance of our model when multiple earthquakes occur simultaneously. As shown, the model is still able to clearly identify both *P* and *S* waves. While the model inherently distinguishes *P* and *S* phases, we note that, given the single-component nature of DAS data, the relative order of the seismic phases may play an important role. Nevertheless, the model likely also learns additional features that aid in phase identification.

5 DISCUSSION

While the overall performance of DeepSubDAS demonstrates clear advantages over PhaseNet DAS on submarine DAS data, several limitations and edge-case behaviours merit further discussion. These insights not only help clarify the current boundaries of the model but also provide directions for future development.

1. Trade-off between context window and memory constraints. When running inference on systems with limited memory, processing long time-series data across thousands of channels (e.g. 5000 channels for 30 s) may exceed available memory. To mitigate this, we allow users to specify a smaller number of channels or shorter time windows for input. While effective in avoiding memory issues, this approach introduces a drawback: the model's limited global context. In DAS data, SNR often degrades at the cable's distal ends. In such scenarios, *S* wave arrivals are occasionally misclassified as *P* waves due to the weak *P* wave signals. This type of misclassification is more prevalent in single-component DAS recordings compared to traditional three-component seismometers, where polarization information aids in phase identification (see Fig. 10a).

2. Challenges in *S*-wave onset detection. In some cases, the onset of *S* wave is difficult to pick accurately. This typically occurs when the coda energy from the *P* wave masks the beginning of the *S* wave, leading to a decreased SNR for *S*-wave identification (see Fig. 10b). These edge cases illustrate the need for improved phase boundary discrimination, especially under overlapping waveforms.

3. Performance under extremely low SNR conditions. Finally, in cases where the *P*-wave signal is almost imperceptible to human analysts across the entire DAS array, the model may also fail to reliably identify the *P* wave. Interestingly, under such conditions, the model tends to label the only clearly visible waveform as a *P* wave, even when it is more likely to be an *S* wave due to its higher amplitude. This behaviour contrasts with human analysts, who can draw on domain-specific knowledge—such as typical amplitude differences between seismic phases to

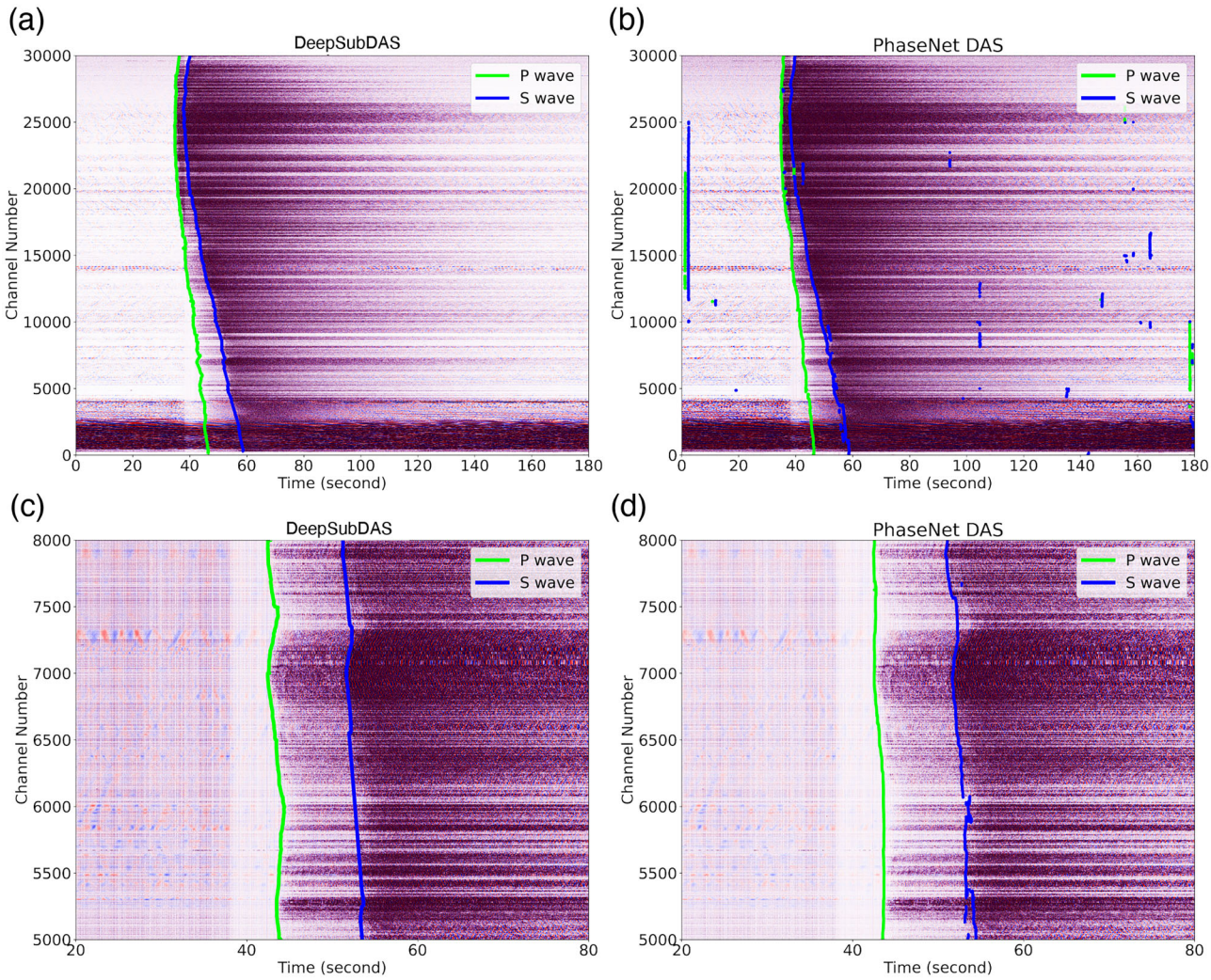

Figure 7. This figure compares the performance of DeepSubDAS (a) and PhaseNet DAS (b) in picking *P* and *S* waves for an ML 3.6 earthquake in Chile. Panels (c) and (d) provide a zoomed-in view of the results for channels 5000 to 8000.

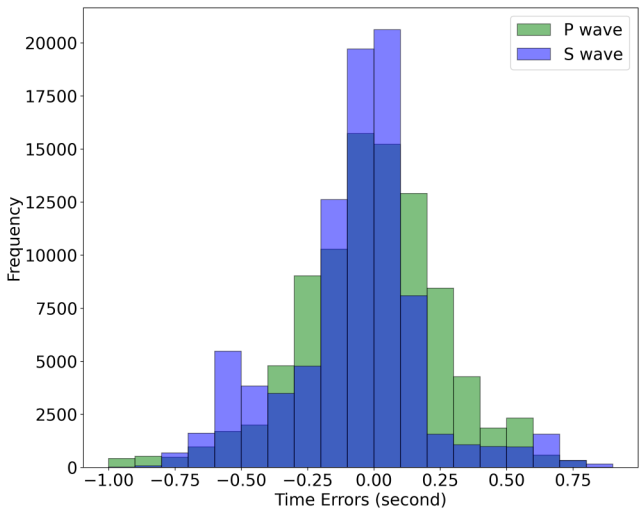

Figure 8. Histogram of *P* and *S* wave residuals of PhaseNet DAS on Alaska and ChileChilly data picks compared to semi-automatic picks.

make more accurate judgments (see Fig. 10c). This highlights both the model's sensitivity to phase order features and its current limitations in reliably distinguishing phases under extreme noise.

4. Limitations of the current training data set: Our current model is trained exclusively on manually labelled *P* and *S* phases, without inclusion of converted phases such as P–S or S–P. Consequently, the model is unable to detect or classify these converted phases (see Fig. 10d). This limitation arises not only from the training data itself but also from the model's output design, which currently considers only three classes: *P*, *S* and noise. Properly addressing converted phases would require either extending the classification scheme to include P–S and S–P phases or training separate models specialized for converted phases. Given the prevalence and often large amplitude of converted phases—particularly P–S conversions—this limitation is important to acknowledge. A more diverse and fully annotated data set that includes these phases would likely enhance the model's versatility and interpretability.

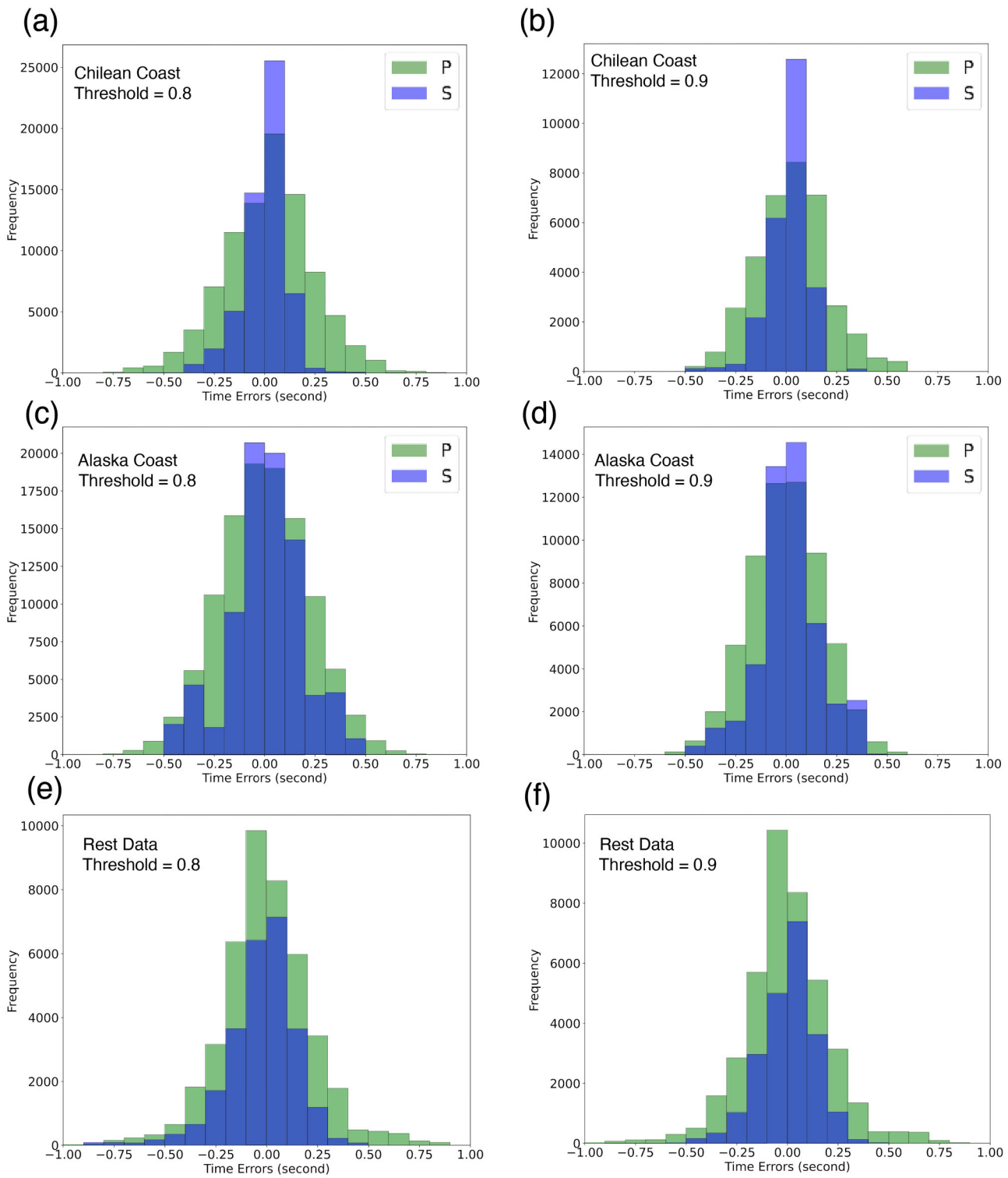

Figure 9. Histogram of *P* and *S* residuals of DeepSubDAS picks compared to semi-automatic picks. (a) and (b) show data from the Chilean coast, while (c) and (d) display data from Alaska, (e) and (f) are rest of the data. In (a), (c) and (e), the threshold we selected is 0.8, whereas in (b), (d) and (f) it is 0.9. It can be observed that the picking error follows a Gaussian distribution and is mostly within 0.2 s. See Table 2 for statistics of the distributions.

These observations emphasize that while DeepSubDAS performs robustly in many settings, its behaviour is still influenced by architectural constraints, input pre-processing strategies, and the composition of the training data set. Particularly, model generalization could benefit significantly from enriched data sets that include converted phases and broader signal conditions.

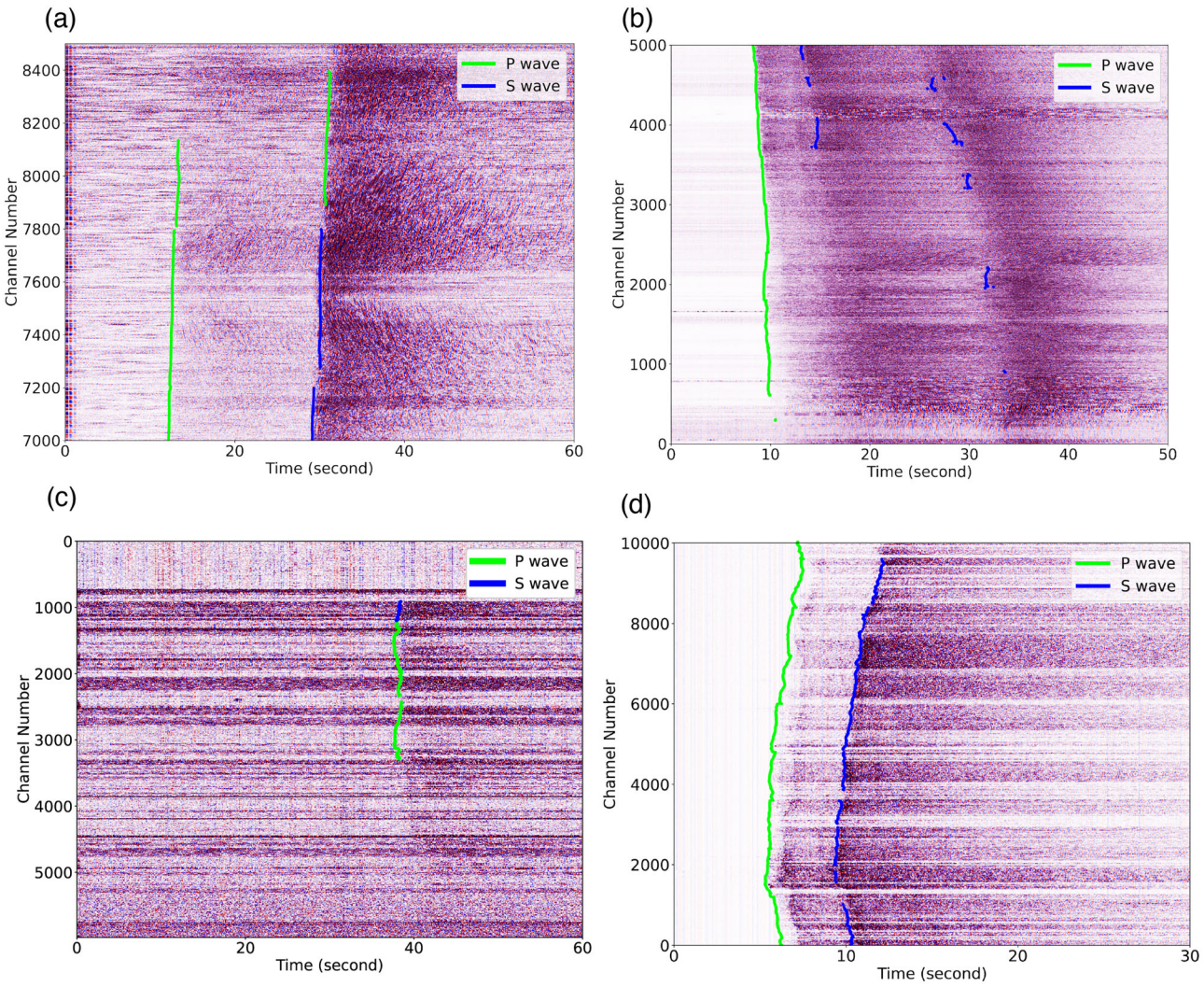

Figure 10. Representative examples of typical failure cases in DeepSubDAS seismic phase picking on submarine DAS data. (a) Limited context misclassification: In distal regions of the DAS array, where SNR is low for *P* waves, the model incorrectly classifies *S* waves as *P* waves due to the absence of global context when inference is limited to smaller channel subsets. (b) Obscured *S* wave onsets: The trailing energy of strong *P* wave arrivals can mask the onset of *S* waves, leading to incomplete or inaccurate *S* wave picks. (c) Severe low-SNR scenario: When the *P*-wave signal is below the visibility threshold for both human analysts and the model, DeepSubDAS tends to label the dominant visible wave—likely the *S* wave—as a *P* wave, highlighting the model's sensitivity to amplitude in the absence of clear *P* phase cues. (d) Undetected converted phases: Converted phases such as PS are not identified by the model, as they are not labelled in the training data set, which only includes *P* and *S* phases.

6 CONCLUSIONS

In this study, we developed and trained a deep learning model specifically for submarine DAS data using a framework based on DeepLab. To support model training, we collected a data set of 92 million seismic records from DAS data across 13 submarine fibre optic cables worldwide. Recognizing the challenge of creating a high-quality labelled data set, we developed a semi-automated labelling tool to accurately mark *P* and *S* waves within this extensive data set. Our model demonstrates superior performance compared to existing models, achieving higher accuracy in picking seismic *P* and *S* waves. For seismic records with low SNR ratios, it can detect arrivals, when human identification becomes difficult at best. To assert the quality of picks in such very low SNR ratios requires estimates of their performance in downstream location tasks, and is beyond the scope of this study.

This advancement is particularly valuable for the future application of submarine DAS data in seismic early warning systems, as it enhances both the reliability and speed of seismic event detection in underwater environments.

In this study, we did not perform earthquake relocation for two primary reasons: first, the precise locations of DAS channels along the cables were unavailable; secondly, relocation is not within the scope of this paper. Oceanic sediment layers, in particular, can heavily influence seismic velocities, adding complexity to precise relocation efforts.

Our model is designed with high flexibility, allowing it to process input data of varying sizes, making it well-suited for deployment across DAS data sets from different regions. This adaptability is critical for the scalability of submarine seismic monitoring, enabling more comprehensive and tailored data analysis

for diverse oceanic environments. Its ability to accommodate regional variability without extensive reconfiguration underscores the model's potential as a robust tool for global submarine seismic monitoring and earthquake early warning systems.

However, our model still has several limitations. For example, it currently cannot identify converted waves, and when the preceding *P* wave signal is too weak, the model may mistakenly classify the subsequent *S* wave as a *P* wave. Addressing these challenges will be an important direction for improvement and optimization.

Looking to the future, several strategies can be employed to improve the generalization and accuracy of our model. One key approach is the adoption of advanced deep learning architectures, such as combining convolution networks with recurrent or transformer models, which can enhance the model's ability to capture both spatial and temporal patterns in seismic data. Furthermore, data augmentation techniques and the generation of synthetic data could help expand the training data set, allowing the model to generalize over a wider range of seismic events and noise conditions. To further improve the adaptability of the model, we can explore multisource data fusion, integrating submarine DAS data with complementary sensors such as ocean bottom seismometers or satellite-based measurements. Furthermore, transfer learning and domain adaptation techniques would allow the model to be more easily adapted to new regions with fewer available data. Finally, expanding the data set to include a wider variety of seismic events from different geographic and geological conditions will ensure the model's robustness, allowing it to perform reliably in diverse underwater environments. Through these advances, our goal is to create a more accurate, flexible and scalable tool for submarine seismic monitoring and rapid earthquake detection systems.

ACKNOWLEDGMENTS

We gratefully acknowledge the support of EU-INFRATECH, Grant agreement ID: 101095055 (SUBMERSE project). We thank Zack Spica and the co-authors of PubDAS for sharing their data. Special thanks to Zhongwen Zhan for sharing the Ridgecrest earthquake data and to Alister Trabattoni for sharing the Chilean coastline earthquake data. Additionally, we appreciate the SeisBench and DeepLab projects for providing their code. This work utilized high-performance computing resources made possible by funding from the Ministry of Science, Research and Culture of the State of Brandenburg (MWFK) and is operated by the IT Services and Operations unit of the Helmholtz Centre Potsdam. We thank GCI for providing access to submarine fibres in Cook Inlet, Alaska. The data acquisition and storage are supported by the UW FiberLab, supported by a Murdock Charitable Trust fund, and The Jerome M. Paros Endowed Chair in Sensor Networks to Dr William Wilcock at University of Washington.

SUPPORTING INFORMATION

Supplementary data are available at *GJIRAS* online.

suppl_data

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

DATA AVAILABILITY

The data and trained model used in this study are available on SeisBench (J. Woollam *et al.* 2022) and Zenodo (H. Xiao 2025).

REFERENCES

AI4EPS, 2023. quakeflow das (revision 91b72d3), hugging face. Available at: https://huggingface.co/datasets/AI4EPS/quakeflow_das

Bornstein, T., Lange, D., Münchmeyer, J., Woollam, J., Rietbrock, A., Barcheck, G., Grevemeyer, I. & Tilmann, F., 2024. PickBlue: Seismic phase picking for ocean bottom seismometers with deep learning, *Earth Space Sci.,* **11**(1), e2023EA003332.

Bouffaut, L. *et al.*, 2022. Eavesdropping at the speed of light: Distributed acoustic sensing of baleen whales in the arctic, *Front. Mar. Sci.,* **9**. doi:

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A.L., 2017. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs, *IEEE Trans. Pattern Anal. Mach. Intell.,* **40**(4), 834–848.

Diederik, K., 2014. Adam: A method for stochastic optimization, preprint(arXiv:).

Jousset, P. *et al.*, 2018. Dynamic strain determination using fibre-optic cables allows imaging of seismological and structural features, *Nat. Commun.,* **9**(1), 2509. doi:

Kim, H.E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M.E. & Ganslandt, T., 2022. Transfer learning for medical image classification: a literature review, *BMC Med. Imaging,* **22**(1), 69.

Latorre, H., Ventosa, S., Ugalde, A., Villaseñor, A., Bartolomé, R. & Ranero, C.R., 2025. Kvp: a multiscale kurtosis approach for seismic phase picking, *Geophys. J. Int.,* **241**(3), 1923–1935.

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**(23), 11 792–11 799.

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.

Lior, I., Rivet, D., Ampuero, J.P., Sladen, A., Barrientos, S., Sanchez-Olavarria, R., Villarroel Opazo, G.A. & Bustamante Prado, J.A., 2023. Magnitude estimation and ground motion prediction to harness fiber optic distributed acoustic sensing for earthquake early warning, *Sci. Rep.,* **13**(1), 424.

Loureiro, A., Schlaphorst, D., Matias, L., Pereira, A., Corela, C., Gonçalves, S. & Caldeira, R., 2025. First DAS observations from the GeoLab fibre in Madeira, Portugal, *Seismica* **4** (2). McGill University Library and Archives.

Mata Flores, D., Mercerat, E.D., Ampuero, J.P., Rivet, D. & Sladen, A., 2023. Identification of two vibration regimes of underwater fibre optic cables by distributed acoustic sensing, *Geophys. J. Int.,* **234**(2), 1389–1400.

Mousavi, S.M., Zhu, W., Sheng, Y. & Beroza, G.C., 2019. CRED: a deep residual network of convolutional and recurrent units for earthquake signal detection, *Sci. Rep.,* **9**(1), 10267.

Mousavi, S.M., Ellsworth, W.L., Zhu, W., Chuang, L.Y. & Beroza, G.C., 2020. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking, *Nat. Commun.,* **11**(1), 3952. doi:

Münchmeyer, J. *et al.*, 2022. Which picker fits my data? a quantitative evaluation of deep learning based seismic pickers, *J. geophys. Res.: Solid Earth,* **127**(1), e2021JB023499.

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, *J. geophys. Res.: Machine Learning and Computation,* **1**(3), e2024JH000247.

Niksejel, A. & Zhang, M., 2024. OBSTransformer: a deep-learning seismic phase picker for OBS data using automated labelling and transfer learning, *Geophys. J. Int.,* **237**(1), 485–505.
Posey, R., Johnson, G. & Vohra, S., 2000. Strain sensing based on coherent Rayleigh scattering in an optical fibre, *Electron. Lett.,* **36,** 1688–1689.
Prechelt, L., 2002. Early stopping-but when?, in *Neural Networks: Tricks of the trade,* pp. 55–69, eds Montavon, G., Orr, G.B. & Müller, K.-R., Springer.
Romanowicz, B. *et al.*, 2023. SeaFOAM: A Year-Long DAS Deployment in Monterey Bay, California, *Seismol. Res. Lett.,* **94**(5), 2348–2359.
Ruppert, N.A., Barcheck, G. & Abers, G.A., 2022. Enhanced regional earthquake catalog with Alaska Amphibious Community Seismic Experiment Data, *Seismol. Res. Lett.,* **94**(1), 522–530.
Shi, Q. & Denolle, M.A., 2023. Improved observations of deep earthquake ruptures using machine learning, *J. geophys. Res.: Solid Earth,* **128**(12), e2023JB027334.
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, *J. geophys. Res.: Solid Earth,* **130**(2), e2024JB029728.
Sladen, A., Rivet, D., Ampuero, J.P., De Barros, L., Hello, Y., Calbris, G. & Lamare, P., 2019. Distributed sensing of earthquakes and ocean-solid earth interactions on seafloor telecom cables, *Nat. Commun.,* **10**(1), 5777. doi:
Spica, Z.J., Nishida, K., Akuhara, T., Pétrélis, F., Shinohara, M. & Yamada, T., 2020. Marine sediment characterized by ocean-bottom fiber-optic seismology, *Geophys. Res. Lett.,* **47**(16), e2020GL088360.
Spica, Z.J. *et al.*, 2023. PubDAS: A PUBlic distributed acoustic sensing datasets repository for geosciences, *Seismol. Res. Lett.,* **94**(2A), 983–998.
Strumia, C., Trabattoni, A., Supino, M., Baillet, M., Rivet, D. & Festa, G., 2024. Sensing optical fibers for earthquake source characterization using raw DAS records, *J. geophys. Res.: Solid Earth,* **129**(1), e2023JB027860.
Trabattoni, A., Vernet, C., van den Ende, M., Baillet, M., Potin, B. & Rivet, D., 2024. Sediment corrections for distributed acoustic sensing, *J. geophys. Res.: Solid Earth,* **129,** e2024JB029054. doi:
Tsuji, T., Ikeda, T., Matsuura, R., Mukumoto, K., Hutapea, F.L., Kimura, T., Yamaoka, K. & Shinohara, M., 2021. Continuous monitoring system for safe managements of CO_2 storage and geothermal reservoirs, *Sci. Rep.,* **11**(1), 19120.
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, *Seismol. Res. Lett.,* **93**(1), 351–363.
van den Ende, M., Trabattoni, A., Baillet, M. & Rivet, D., 2025. An analysis of the dynamic range of distributed acoustic sensing for earthquake early warning, *Seismica,* **4**(1). doi:
Wilcock, W.S.D., Abadi, S. & Lipovsky, B.P., 2023. Distributed acoustic sensing recordings of low-frequency whale calls and ship noise offshore Central Oregon, *JASA Express Lett.,* **3**(2), 026002.
Williams, E.F. *et al.*, 2023. Fiber-optic observations of internal waves and tides, *J. geophys. Res.: Oceans,* **128**(9), e2023JC019980.
Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S. & Trujillo, J., 1998. A comparison of select trigger algorithms for automated global seismic phase and event detection, *Bull. seism. Soc. Am.,* **88**(1), 95–106.
Woollam, J. *et al.*, 2022. Seisbench—a toolbox for machine learning in seismology, *Seismol. Soc. Am.,* **93**(3), 1695–1709.
Xiao, H., 2025. *DeepSubDAS: An earthquake phase picker from submarine distributed acoustic sensing data*[Dataset], Zenodo. doi:
Xiao, H., Tanimoto, T., Spica, Z.J., Gaite, B., Ruiz-Barajas, S., Pan, M. & Viens, L., 2022. Locating the precise sources of high-frequency microseisms using distributed acoustic sensing, *Geophys. Res. Lett.,* **49**(17), e2022GL099292.
Xiao, H., Spica, Z.J., Li, J. & Zhan, Z., 2024. Detection of earthquake infragravity and tsunami waves with underwater distributed acoustic sensing, *Geophys. Res. Lett.,* **51**(2), e2023GL106767.
Xiao, H., Tanimoto, T., Spica, Z.J. & Tilmann, F., 2025. Frequency dependent microseisms sources: a case study in Oregon, *Geophys. Res. Lett.,* **52**(19), e2025GL118297.
Yin, J., Soto, M.A., Ramírez, J., Kamalov, V., Zhu, W., Husker, A. & Zhan, Z., 2023. Real-data testing of distributed acoustic sensing for offshore earthquake early warning, *The Seismic Record,* **3**(4), 269–277.
Yoon, C.E., O'Reilly, O., Bergen, K.J. & Beroza, G.C., 2015. Earthquake detection through computationally efficient similarity search, *Sci. Adv.,* **1**(11), e1501057. doi:
Zhu, W. & Beroza, G.C., 2018. PhaseNet: a deep-neural-network-based seismic arrival-time picking method, *Geophys. J. Int.,* **216**(1), 261–273.
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, *Nat. Commun.,* **14**(1), 8192. doi: