

DAS-N2N: machine learning distributed acoustic sensing (DAS) signal denoising without clean data

S. Lapins¹,¹ A. Butcher,¹ J.-M. Kendall,² T.S. Hudson,² A.L. Stork,³ M.J. Werner¹,¹ J. Gunning¹ and A.M. Brisbane⁴

¹School of Earth Sciences, University of Bristol, Bristol BS8 1RL, UK. E-mail: sacha.lapins@bristol.ac.uk

²Department of Earth Sciences, University of Oxford, Oxford OX1 3AN, UK

³Silixa Ltd, Elstree, Bristol WD6 3SN, UK

⁴British Antarctic Survey, Cambridge CB3 0ET, UK

Accepted 2023 November 15. Received 2023 October 31; in original form 2023 April 18

SUMMARY

This paper presents a weakly supervised machine learning method, which we call DAS-N2N, for suppressing strong random noise in distributed acoustic sensing (DAS) recordings. DAS-N2N requires no manually produced labels (i.e. pre-determined examples of clean event signals or sections of noise) for training and aims to map random noise processes to a chosen summary statistic, such as the distribution mean, median or mode, whilst retaining the true underlying signal. This is achieved by splicing (joining together) two fibres hosted within a single optical cable, recording two noisy copies of the same underlying signal corrupted by different independent realizations of random observational noise. A deep learning model can then be trained using only these two noisy copies of the data to produce a near fully denoised copy. Once the model is trained, only noisy data from a single fibre is required. Using a data set from a DAS array deployed on the surface of the Rutford Ice Stream in Antarctica, we demonstrate that DAS-N2N greatly suppresses incoherent noise and enhances the signal-to-noise ratios (SNR) of natural microseismic icequake events. We further show that this approach is inherently more efficient and effective than standard stop/pass band and white noise (e.g. Wiener) filtering routines, as well as a comparable self-supervised learning method based on masking individual DAS channels. Our preferred model for this task is lightweight, processing 30 s of data recorded at a sampling frequency of 1000 Hz over 985 channels (approximately 1 km of fibre) in <1 s. Due to the high noise levels in DAS recordings, efficient data-driven denoising methods, such as DAS-N2N, will prove essential to time-critical DAS earthquake detection, particularly in the case of microseismic monitoring.

Key words: Antarctica; Machine learning; Instrumental noise; Distributed acoustic sensing; Earthquake monitoring and test-ban treaty verification; Denoising.

1 INTRODUCTION

Distributed acoustic sensing (DAS) is a novel form of seismic monitoring, measuring changes in strain acting along a buried or encased fibre-optic cable through reflectometry. In recent years, DAS has seen a growing range of applications, including passive and active experiments to detect seismic events, monitor urban and anthropogenic activity, image the subsurface and monitor changes in material and ambient properties (e.g. Dou *et al.* 2017; Ajo-Franklin *et al.* 2019; Lindsey *et al.* 2019; Hudson *et al.* 2021b; Nayak *et al.* 2021; Jousset *et al.* 2022; Kennett 2022; Zhou *et al.* 2022; van den Ende *et al.* 2023). A DAS interrogator unit sends short, finite-

duration light pulses along an optical fibre and measures the phase of Rayleigh backscattering caused by small density variations and defects in the fibre (Parker *et al.* 2014; Hartog 2017; Lindsey *et al.* 2020). The backscattered light from a given section of fibre returns to the interrogator with a predictable two-way traveltime: this allows any changes in the light's phase or intensity between successive pulses (e.g. from disturbances as a result of incoming seismic waves) to be mapped to specific sections along the fibre within some known precision, known as the 'gauge length' (e.g. Dean *et al.* 2017; Hartog 2017; Lindsey *et al.* 2020). In this manner, the entire fibre-optic cable acts as a series of distributed seismic sensors, sampling changes to the strain field acting on the fibre at regularly spaced

intervals along its length (typically shorter than the gauge length; Parker *et al.* 2014; Hartog 2017), the locations of which are often referred to as ‘channels’.

Optical fibres have many attractive properties for seismic monitoring. They are flexible, durable, highly sensitive to vibrations and changes in strain field (Hartog 2017) and can extend many kilometres from the DAS interrogator unit and power source (e.g. Parker *et al.* 2014; Ajo-Franklin *et al.* 2019; Lindsey *et al.* 2019; Shinohara *et al.* 2022). As such, they are well-suited to seismic monitoring in harsh or remote environments, such as volcanoes (Jousset *et al.* 2022), regions with extreme climate (e.g. glacial settings; Walter *et al.* 2020; Hudson *et al.* 2021b; Zhou *et al.* 2022) and beneath oceans (Lindsey *et al.* 2019; Shinohara *et al.* 2022). Furthermore, the vast networks of existing telecommunications fibre present the opportunity to heavily augment the coverage of existing seismic networks, both on local and global scales (Ajo-Franklin *et al.* 2019; Nayak *et al.* 2021; Kennett 2022; Shinohara *et al.* 2022).

However, despite these advantageous properties, optical fibres are also highly sensitive to temperature (Hartog 2017), local disturbances from the interrogator (Lindsey *et al.* 2020), ground/coupling conditions (Hartog *et al.* 2014; Ajo-Franklin *et al.* 2019) and properties of the fibre/instrument components used (Isken *et al.* 2022), most of which are heterogeneously distributed along the optical cable, leading to greater levels of seemingly random observational noise in DAS recordings when compared to conventional seismometers (Hudson *et al.* 2021b). DAS fibres are also only sensitive to along-cable strain, which leads to challenges in recording the full seismic wavefield and relating measurement units to actual ground motion. Finally, the large data volumes acquired by sampling along long extents of fibre (sometimes on the order of TBs per day) require highly efficient and optimized signal processing methods, especially for real-time monitoring operations and earthquake early warning systems.

Optical cables are regularly manufactured with multiple fibres for added capacity (e.g. for telecommunication providers). For DAS applications, additional fibres are typically left unused as DAS interrogators often process measurements from a single light pulse and fibre at a time to avoid interference (Parker *et al.* 2014). However, the availability of multiple fibres can provide highly useful redundancy for enhancing the signal-to-noise ratio (SNR) of any external signal through application of so-called ‘weakly supervised’ machine learning. By splicing (joining together) two fibres at one end of the cable, the light pulse from the DAS interrogator effectively travels ‘there-and-back’ along the length of the cable, recording two copies of the same underlying seismic signal but with different random measurement noise due to differences in scatterers and photon behaviour between the two fibres. A deep learning model can then be trained using only these two noisy copies of the underlying signal to produce a denoised copy of the data through a method known as ‘Noise2Noise’ (N2N; Lehtinen *et al.* 2018), a form of weakly supervised machine learning that exploits the point estimation properties of certain loss functions and does not require clean (i.e. noise-free) target data or manual curation/labelling for training.

In this paper, we present the first known application of N2N for suppressing strong random (i.e. incoherent) noise processes in DAS data, which we refer to as DAS-N2N. This approach has previously been used to suppress synthetically generated noise in individual photographic, MRI scan and microscopy images (Lehtinen *et al.* 2018; Calvarons 2021) but never previously (to our knowledge) to suppress real noise in continuously acquired noisy DAS or seismic

data. In Section 2, we provide an overview of both ‘conventional’ (i.e. non-machine learning) and machine learning approaches for seismic signal noise suppression, including N2N. In Section 3, we provide details of our example data set, acquired by a DAS deployment on the surface of the Rutford Ice Stream in Antarctica. In Section 4, we describe the theory behind N2N, and the procedure for training and implementing a DAS-N2N model. In Section 5, we compare icequake data denoised by DAS-N2N against three benchmark methods: conventional Butterworth bandpass filtering, Wiener filtering, and an existing self-supervised deep learning method for denoising DAS data, known as jDAS (van den Ende *et al.* 2021), that also requires no clean target data or manual curation during model training. The article ends with a discussion of DAS-N2N model performance, generalization to other data sets and some concluding remarks in Sections 6 and 7, respectively. Example code and data for implementing DAS-N2N have been archived and made available by Lapins *et al.* (2023) (see Data Availability section).

2 BACKGROUND

2.1 Conventional seismic signal filtering

Pass and stop band filters, designed to remove certain frequencies from a recorded signal, are a ubiquitous processing step for suppressing unwanted noise in seismic signals. The general aim is to identify a frequency range that contains as much of the desired signal and as a little of the undesired background noise as possible, with all other frequencies removed or suppressed by a chosen filter (e.g. Butterworth, Chebyshev or Gaussian filters). These filters are typically applied by convolution of the recorded signal with a polynomial approximating an idealized filter response (i.e. approximating a uniform and complete response in the pass band with full attenuation in the stop band, which cannot be expressed by a finite order polynomial; Proakis & Manolakis 1996). Although simple, interpretable and relatively fast for individual seismic traces, such methods have several drawbacks, both for DAS applications and for seismic signals more generally.

For noise suppression, the greatest drawback of conventional pass and stop band filters is the inability, by design, to suppress noise that lies in the same frequency range as the desired signal. For well-deployed geophones and broadband seismometers, random measurement noise is considered to be low and signals of interest are usually in distinct frequency bands from other external coherent noise sources (e.g. ocean microseisms in the 0.1–0.5 Hz range; Bromirski *et al.* 2013; Koper & Burlacu 2015; Lapins *et al.* 2020). However, environmental, financial, cultural and political factors mean that deploying large numbers of high-cost seismometers in quiet or well-insulated environments is rarely feasible. DAS offers a relatively low-cost, straightforward and densely sampled alternative; however, random measurement noise along the fibre is often observed to be much stronger than that of geophones (Hudson *et al.* 2021b; du Toit *et al.* 2022; Isken *et al.* 2022) and occurs across the entire observed frequency spectrum (see Section 5). As such, the frequency range of interest is much more contaminated by unwanted noise. Furthermore, for passive monitoring applications, this frequency range must be assumed *a priori*, which typically leads to more conservative (i.e. wider pass band) filtering and greater noise contamination. Choice of filter family and polynomial order is also subject to certain trade-offs, including the degree of amplitude ‘ripples’ in the pass and stop bands, the abruptness of the

transition between bands, and susceptibility to detrimental artefacts such as ringing, signal polarity changes and non-linear phase shifts (e.g. Proakis & Manolakis 1996; Scherbaum 2001; Havskov & Ottemöller 2010). Finally, and importantly, repeated application of a chosen filter over hundreds or thousands of individual DAS channels is computationally costly when required for (near-)real-time processing and monitoring.

Some of the drawbacks outlined above can be mitigated: for example through use of adaptive algorithms that adjust filter coefficients or parameters (e.g. Duncan & Beresford 1994; Jeng *et al.* 2009; Isken *et al.* 2022); filtering in both time and space frequency (f - k) domains (e.g. Duncan & Beresford 1994; Bacon *et al.* 2003; Mousa 2019; Hudson *et al.* 2021b; Isken *et al.* 2022); applying a statistical estimation method for identifying additive or incoherent noise (e.g. Wiener filters; Williams *et al.* 2020) or combining multiple methods (Chen *et al.* 2023). However, these approaches are still limited when noise and signal overlap within a given frequency range or by their computational demands, model/method assumptions, or the requirement for manual parametrization (see Section 5).

2.2 Deep learning denoising

Across the broader fields of science and engineering, noise suppression (or ‘denoising’) is being increasingly addressed through deep learning methods, with the greatest advancements occurring in the field of image processing. Unlike linear, f - k (time and space frequency) or statistical estimation filters, deep learning models are not restricted by explicit statistical assumptions, response trade-offs (e.g. choice of filter family or order), or manual parametrization (e.g. choosing a pass band or statistical model). Desirably, they have the capacity to ‘learn’ empirical, abstract and non-linear hierarchical data representations directly from sample data, allowing them to perform effective signal filtering and feature extraction without manual input or prior assumptions on the distribution of the signal or noise. Model implementation is also heavily optimizable through use of GPUs and compression/pruning strategies (Zhu & Gupta 2017), allowing for rapid signal processing.

2.2.1 Supervised learning

Initial success in this area was driven by the ‘standard’ fully supervised paradigm, using a large number of noisy/clean signal pairs for model training; that is both noisy and noise-free copies of each training sample are available and the model is trained to directly map noisy signals to their noise-free counterparts. This approach is sometimes referred to as ‘Noise2Clean’ (N2C) in the denoising literature and has been previously applied to seismic signals with apparent success (Zhu *et al.* 2019; Klochikhina *et al.* 2020; Li & Ma 2021; Tibi *et al.* 2021; Yang *et al.* 2022). However, in many applications, it can be difficult or even impossible to acquire sufficient quantities of noise-free or high signal-to-noise recorded signals for robust model training, and thus this approach is limited in its ‘real-world’ applicability. This situation is particularly true in the case of DAS recordings, where the observed data are heavily contaminated by strong random noise processes and simulating seismic wave propagation to generate realistic noise-free signals across a long extent of fibre is computationally intensive and challenging to model. This restricted applicability has led to the wider development of denoising methods that do not require noise-free ground-truth signals, such as weakly supervised (Zhou 2018; van

Engelen & Hoos 2020) or self-supervised (Ericsson *et al.* 2022) learning methods.

2.2.2 Weakly supervised learning

Weakly supervised learning relaxes the requirement for noise-free ‘ground-truth’ target data during training. One pioneering method for weakly supervised denoising, which we base our proposed DAS-N2N methodology on, is known as ‘Noise2Noise’ (N2N; Lehtinen *et al.* 2018), where the aim is for a model to learn to transform noisy images into clean images using only noisy copies of the same image as both input and target training data. A N2N model suppresses random noise by exploiting the point estimation properties of certain loss functions during model training (Lehtinen *et al.* 2018; Pang *et al.* 2021); for example, mean squared error (MSE) and mean absolute error (MAE) loss functions are minimized by the mean and median of a set of observations, respectively. Intuitively, as long as the noise in the input and target data are independently and randomly drawn from some (known or unknown) noise distribution, it is impossible for a model to predict the random noise values in the target data from the random noise values in the input data. As such, to minimize its expected loss, the model learns to map noise in the input data to the value of smallest average deviation from the noise in the target data (e.g. the mean, median or mode of the noise distribution; Lehtinen *et al.* 2018), according to the chosen loss function. Simultaneously, as long as the underlying clean signal in the input and target data are identical, the model’s expected loss is minimized by learning a direct 1-to-1 mapping between the two (see Section 4.1).

It has been demonstrated both theoretically and empirically that models trained using only noisy signals in this manner can perform as well as, or even better than, those trained in a fully supervised manner using noisy/clean signal pairs (Lehtinen *et al.* 2018; Pang *et al.* 2021). For example, it can be shown that the loss minimization problem is effectively the same for fully and weakly supervised learning and a MSE loss function (Section 4.1).

For DAS, the applicability of N2N is motivated by the fact that optical fibres can be spliced so that they effectively double-back on themselves within their cable sleeve (Fig. 1), recording two (near-) identical copies of any external seismic source but with different independent realizations of any random noise processes (Fig. 2a). Furthermore, when continuous recordings are available, vast training sets of independent noisy signal pairs are readily available for training without the need for any manual labelling, providing a fully automatable approach that can be applied to any DAS deployment.

The one main drawback of N2N is that, in some situations, recording multiple noisy copies of the same underlying signal is not possible; for example when analysing previously recorded unspliced DAS data or using so-called ‘dark’ fibres (existing unused telecommunication fibre networks) that may not always be feasibly spliced. In this case, N2N is not directly applicable, but an extension of this method, based on ‘recorrupting’ the recorded signal with additional noise (known as ‘Recorrupted-to-Recorrupted’, or R2R; Pang *et al.* 2021), can be applied. With R2R, the additional noisy copies of the signal required for model training are generated (as opposed to recorded) using a secondary noise distribution so that the noise in each new noisy copy is now independently drawn from a new noise distribution. These new noisy copies can then be used to train a model in the same manner as N2N, with similar performance (Pang *et al.* 2021). However, sufficiently corrupting the original observed

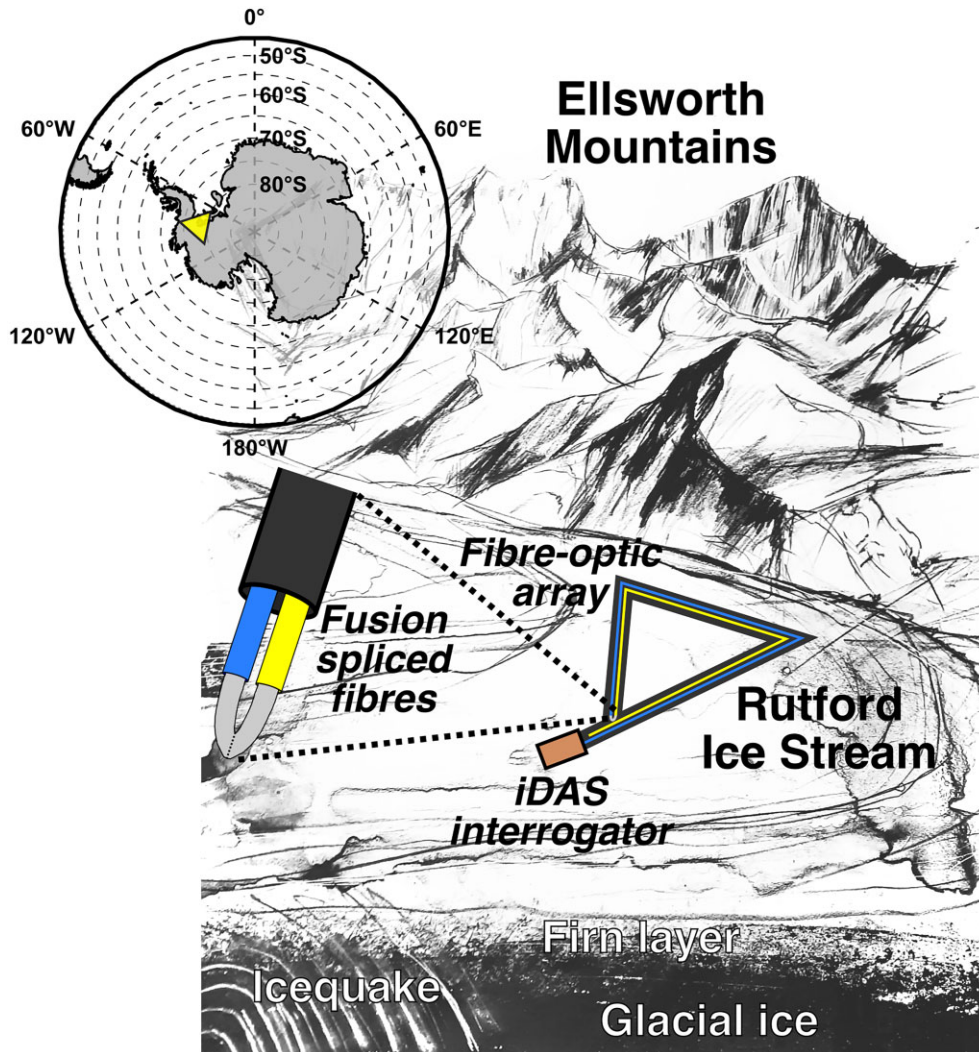


Figure 1. Map and schematic illustration showing the DAS experiment setup. Top left-hand side: map showing geographic location of DAS array (gold triangle) on Rutford Ice Stream, Antarctica. Main: Schematic illustration of the DAS experiment. DAS fibre array was deployed in triangular configuration on the surface of Rutford Ice Stream, with two single-mode fibres hosted within a single cable jacket spliced at cable end. See Hudson *et al.* (2021a) and Supporting Information in Hudson *et al.* (2021b) for further details.

noise to produce independent realizations from a new noise distribution means one must generally have some prior knowledge of the original noise distribution, which is not always known or may be challenging to model. Due to this added non-trivial requirement, we do not explore this method further in this paper and restrict our focus solely on the N2N approach with recorded noisy signal pairs.

2.2.3 Self-supervised learning

An alternative approach that requires no additional noisy or clean target data for training (i.e. an unspliced fibre can be used) is self-supervised learning. Self-supervised learning is often formulated as learning from ‘fill-in-the-gap’ problems (Ericsson *et al.* 2022), where some section of input data is hidden or masked and the model is tasked with predicting the values of the missing data. When applied to the task of denoising (sometimes known as ‘Noise2Self’ or ‘Noise2Void’; Krull *et al.* 2018; Batson & Royer 2019), the intuition behind such an approach is that, through training, the model will learn to interpolate or predict missing coherent or broad-scale signal features, based on the surrounding data and exposure to many

training samples, but will be unable to predict random, incoherent or fine-scale signal features. As with weakly supervised learning, the model minimizes its expected loss by learning to map the latter to the value (or point estimate) of smallest average deviation, according to some loss function (e.g. MSE).

One such self-supervised method, known as jDAS, after the concept of j-invariance (Batson & Royer 2019), has been previously applied to DAS (van den Ende *et al.* 2021). With jDAS, individual DAS channels are dropped during training and the model learns to predict these missing data using data from neighbouring channels; that is by effectively learning to interpolate any coherent signal across missing channels. This step of masking and predicting missing data is then repeated for each DAS channel at run-time (van den Ende *et al.* 2021). Similarly, Birnie *et al.* (2021) apply the same concept by treating dense-array post-stack seismic data as a 2-D image and masking rectangular/square sections of the image during training. Alternatively, Liu *et al.* (2022) divide the data into odd (input data) and even (target data) channel numbers and train a model to map one to the other, effectively amounting to the same task as masking every other channel and predicting the missing data.

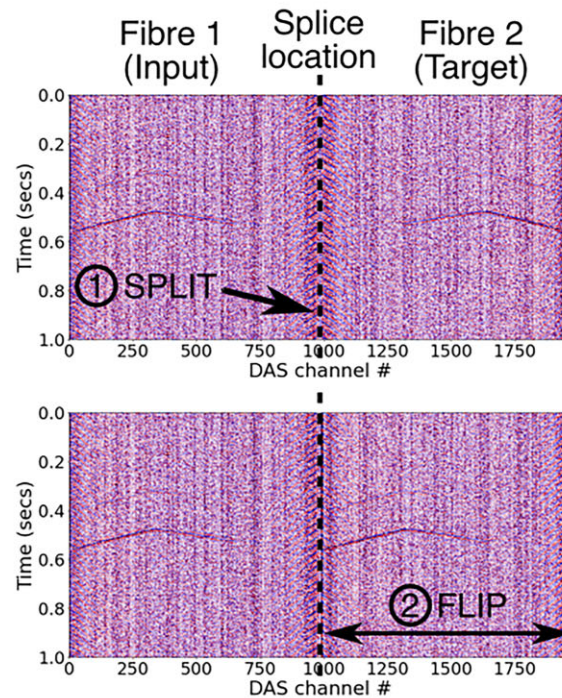
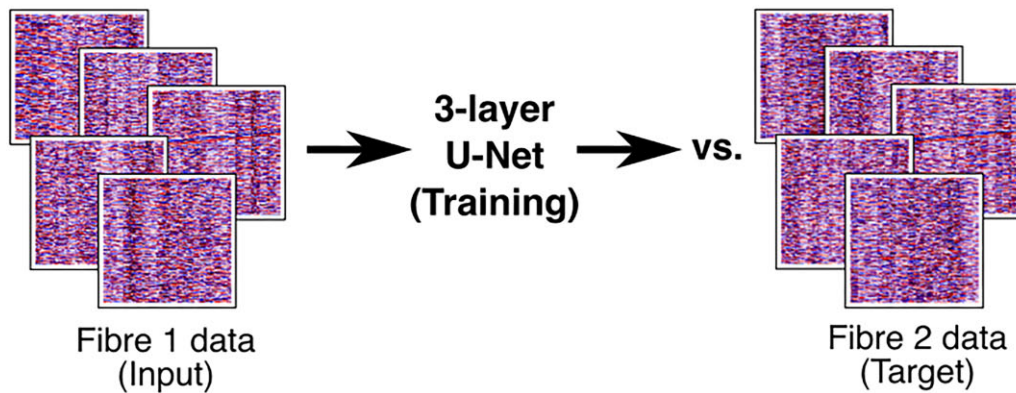
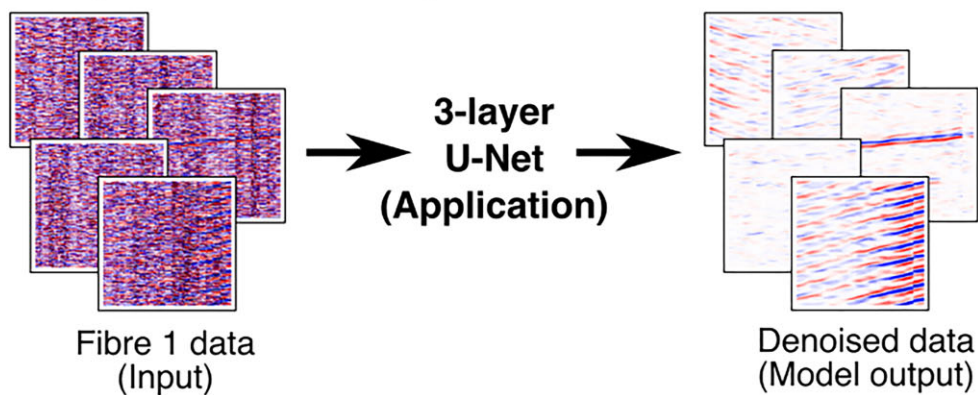
(a) Identify input / target data**(b) Chop up data for training (w/ augmentation)****(c) Only Fibre 1 data required at run time**

Figure 2. Implementing DAS-N2N. (a) Raw data is split into input (Fibre 1) and target (Fibre 2) training data. (b) Data are divided into smaller sections (128 samples \times 96 channels) for model training, with augmentation (vertical/horizontal flipping) randomly applied to each training sample pair. (c) Once the model is trained, only the input data (Fibre 1) is required for denoising.

Although self-supervised learning has the highly desirable trait of not requiring any additional clean or noisy copies of the data, its effectiveness becomes increasingly limited when noise levels are high (van den Ende *et al.* 2021) and the number of DAS channels to process is large. The desired signal features of interest must be interpolated by the model, rather than ‘retained’ through a 1-to-1 mapping as in the case of weakly or fully supervised learning, and thus signal quality can suffer as a result of the same point estimation properties (e.g. mapping to average of all possible outcomes) being used to suppress incoherent and fine-scale noise. Furthermore, by masking and predicting only one channel or a small number of datapoints at a time, self-supervised methods, when formulated as fill-in-the-gap problems, are an order of magnitude slower than standard or weakly supervised methods and can become prohibitively slow when the signal sample frequency and number of DAS channels to process are high.

3 DATA

In this paper, we demonstrate DAS-N2N using data acquired by a DAS array deployed on the surface of the Rutford Ice Stream in Antarctica (Fig. 1). Despite being a low anthropogenic noise environment, strong random noise processes (e.g. optical noise caused by random scattering/coupling of photons and environmental factors) in the raw recorded data from this deployment dominate the signal from microseismic icequake events. The deployment consists of a Silixa iDASv2 interrogator (Parker *et al.* 2014) and a 1 km cable, with a sample frequency of 1000 Hz, channel spacing of 1 m and gauge length of 10 m (see Hudson *et al.* 2021b for further details). Two single-mode optical fibres hosted within the cable jacket were spliced at one end to form a ‘there-and-back’ loop (Fig. 1). These data were collected to investigate the suitability of DAS for studying natural microseismicity (Hudson *et al.* 2021b) and imaging the near-surface ice structure (Zhou *et al.* 2022).

Over the course of 14 d (2020-01-11–2020-01-24), the DAS fibre-optic cable was repeatedly deployed in different horizontal arrangements on the surface of the ice stream, comprising a linear, triangular and ‘hockey stick’ array. The cable was coupled to the ground by placing it in a skidoo track and back-covering with snow. The data presented here were chosen from the time period during which the fibre was deployed in a triangular configuration (2020-01-17 0100–0500 UTC), with each linear section of the triangle approximately 330 m in length. Data recorded between 0100 and 0300 UTC were used for model training, with the remaining two hours data used as test data. As the DAS cables were deployed horizontally and P waves arrive at the surface with near-vertical incidence due to the presence of a low-velocity firm layer, only S -wave phase arrivals are observed by the fibre during the deployment (Hudson *et al.* 2021b). The example icequake events presented in Section 5 were detected using a localized Radon-transform-based detection method (Butcher *et al.* 2021).

4 METHODS

4.1 DAS-N2N theory

Seismic signals are contaminated by both coherent (i.e. seismic waves generated by some undesired external source) and incoherent (i.e. random) noise. A noisy signal, \mathbf{y} , can be expressed as a sum of independent signal components, such that

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \quad (1)$$

where \mathbf{x} is a single- or multidimensional array representing the underlying ‘clean’ signal from any recordable external seismic source (including external coherent noise sources) and \mathbf{n} are samples randomly drawn from some noise distribution, with one sample of \mathbf{n} drawn for each element in \mathbf{x} . The random noise distribution is often assumed to be Gaussian, as a result of the Central Limit Theorem, but this is not a requirement for N2N.

When DAS fibres are spliced, a second copy of the underlying signal is near-simultaneously recorded, with

$$\tilde{\mathbf{y}} = \mathbf{x} + \tilde{\mathbf{n}}, \quad (2)$$

where $\tilde{\mathbf{y}}$ is a second noisy copy of clean signal \mathbf{x} , corrupted by random noise samples $\tilde{\mathbf{n}}$ (drawn independently of noise samples \mathbf{n}). We observe that samples drawn from \mathbf{n} and $\tilde{\mathbf{n}}$ need not be locally identically and independently distributed (i.i.d.; see Section 5).

With DAS-N2N, these two noisy signals, \mathbf{y} and $\tilde{\mathbf{y}}$, serve as input and target data, respectively, for training a neural network, f_θ , parametrized by model weights, θ . This neural network is trained to minimise the expected loss between $f_\theta(\mathbf{y})$ and $\tilde{\mathbf{y}}$ according to some loss function, L . For an MSE loss function (i.e. $L(x, y) = \frac{1}{N} \sum (x - y)^2$), this expected loss can be expressed as

$$\mathbb{E} \{L[f_\theta(\mathbf{y}_i), \tilde{\mathbf{y}}_i]\} = \mathbb{E} \left\{ \frac{1}{M} \sum_{i=0}^M [(\mathbf{x}_i + \tilde{\mathbf{n}}_i) - f_\theta(\mathbf{y}_i)]^2 \right\}, \quad (3)$$

where i is training sample index and M is the number of training samples in a training batch. Eq. (3) can be trivially expanded (Pang *et al.* 2021), such that

$$\begin{aligned} & \mathbb{E} \left\{ \frac{1}{M} \sum_{i=0}^M [(\mathbf{x}_i + \tilde{\mathbf{n}}_i) - f_\theta(\mathbf{y}_i)]^2 \right\} \\ &= \mathbb{E} \left\{ \frac{1}{M} \sum_{i=0}^M [\mathbf{x}_i - f_\theta(\mathbf{y}_i)]^2 \right\} + \mathbb{E} \left\{ \frac{2}{M} \sum_{i=0}^M \tilde{\mathbf{n}}_i \mathbf{x}_i \right\} \\ & - \mathbb{E} \left\{ \frac{2}{M} \sum_{i=0}^M \tilde{\mathbf{n}}_i f_\theta(\mathbf{y}_i) \right\} + \mathbb{E} \left\{ \frac{1}{M} \sum_{i=0}^M \tilde{\mathbf{n}}_i^2 \right\}, \end{aligned} \quad (4)$$

where the first term, $\mathbb{E} \left\{ \frac{1}{M} \sum_{i=0}^M [\mathbf{x}_i - f_\theta(\mathbf{y}_i)]^2 \right\}$, is equivalent to the expected MSE loss when training using noisy/clean training pairs (i.e. the standard supervised case). As long as \mathbf{n} and $\tilde{\mathbf{n}}$ are independent, the remaining expectation terms are constant (Pang *et al.* 2021): the two intermediate terms are equal to zero if signal, noise and model output are all zero-mean (enforced by simple subtraction of recorded signal mean, a near-ubiquitous seismic pre-processing step) and summed over sufficiently large M , with the final expectation term equal to the variance of the noise distribution in the target data. As such, the loss minimization task when training a model with DAS-N2N can be expressed as

$$\mathbb{E} \{L[f_\theta(\mathbf{y}_i), \tilde{\mathbf{y}}_i]\} = \mathbb{E} \left\{ \frac{1}{M} \sum_{i=0}^M [\mathbf{x}_i - f_\theta(\mathbf{y}_i)]^2 \right\} + c, \quad (5)$$

which is equivalent to the standard noisy/clean supervised case, up to a constant, c , relating to the variance of the noise. It is for this reason that DAS-N2N can perform as well as a model trained with noisy/clean signal data, with the advantage that all recorded data can be used for model training without any manual curation or the need to ‘generate’ noisy/clean signal pairs.

4.2 Implementing DAS-N2N

As mentioned, a DAS-N2N model is trained by using data recorded by one of the spliced fibres as input data, with data recorded by the other spliced fibre as target data (Fig. 2a). The only pre-processing steps applied in this work are to remove the signal mean (across all channels) and normalize the data (i.e. divide through by the standard deviation). When training a model using data from longer fibres that have spatially changing or highly non-linear noise processes (e.g. from hanging sections or light decay) channel-wise normalisation will likely be required to ensure the data remain centred around zero and consistently normalised. The raw data were originally stored as 30 s TDMS files (the standard file type for data acquired using a Silixa iDAS interrogator; Parker *et al.* 2014), and thus these pre-processing steps are applied to 30 s sections of data at a time.

The input and target data are then split into corresponding 128×96 size arrays (no. of time samples \times no. of DAS channels, respectively), with a batch size of 24 used for model training (Fig. 2b). Training data are augmented by randomly flipping both the input and target data along their vertical (time) and horizontal (channel) axes. The loss between the model-processed input data and the noisy target data is calculated for each batch using an MSE loss function, with model weights updated using the Adam optimization algorithm (Kingma & Ba 2014). The model was trained for 30 epochs, with learning rate decreasing between epochs from 10^{-3} to 10^{-5} over the course of model training.

N2N is based on exploiting the point estimation properties of L2 and L1 loss functions (Lehtinen *et al.* 2018), and therefore its performance is relatively agnostic to choice of model architecture (i.e. any model with sufficient capacity can be trained to perform N2N denoising). However, certain model architectures and components will have advantageous qualities for denoising, such as hierarchical feature representation (e.g. from convolutional layers) and use of dense/residual connections (e.g. to retain underlying signal as it passes from layer to layer). With this in mind, we choose to implement DAS-N2N using a shallow, 3-layer U-Net (Ronneberger *et al.* 2015, see the Appendix). By limiting the number of model layers and using skip connections, the underlying signal can be easily retained from layer-to-layer and computational processing time is kept low. The final 3-layer DAS-N2N model has just 47 065 model parameters, processing 30 seconds of recorded data across 985 channels ($30\,000 \times 985$ data points) in <1 s (average processing time over 10 runs using Python 3.7.12, TensorFlow version 2.3.0 (Abadi *et al.* 2015) and a single NVIDIA GeForce RTX 2080 Ti GPU).

Following training, only the input data (i.e. data from a single fibre) are required for data processing (Fig. 2c). The normalization step applied during training is also reversed at this stage (i.e. the model-processed data are multiplied by the original data standard deviation) to recover absolute signal amplitude.

4.3 jDAS implementation

For comparison with our proposed DAS-N2N methodology, we implement the self-supervised jDAS approach described by van den Ende *et al.* (2021), using the same model architecture as our DAS-N2N model and applying the same data normalization/augmentation steps (Section 4.2), to serve as a benchmark for comparable ‘noisy data only’ machine learning approaches. The data are split into 2048×11 data blocks for model training, as proposed by van den Ende *et al.* (2021), with a mask randomly applied to a single DAS channel for each training sample.

Both the DAS-N2N and jDAS models are trained using the same 3-layer U-Net architecture, MSE loss function, Adam optimizer, learning rate schedule and number of epochs for direct comparison of method effectiveness.

4.4 Conventional bandpass filtering

For comparison with standard seismic filtering steps, we bandpass filter the raw DAS data between 10 and 100 Hz, based on icequake signal characteristics from Hudson *et al.* (2021b), using a 4th order Butterworth infinite impulse response (IIR) filter. We apply a two-pass filter to remove any nonlinear phase shift, allowing for more direct comparison between methods (Section 5). Filtering is performed using the open-source ObsPy Python library (Beyreuther *et al.* 2010; Megies *et al.* 2011; Krischer *et al.* 2015), which uses optimized low-level C programming language routines from the popular and widely used SciPy library (Virtanen *et al.* 2020). Butterworth filters have a near-uniform response in the pass band and are thus a popular choice for seismic signal processing as they adequately retain underlying signal amplitude information used for further seismic signal analysis (e.g. earthquake magnitude estimation). This near-uniform response also provides a benchmark for comparing absolute signal amplitudes against DAS-N2N and jDAS processing methods.

4.5 Wiener filtering

Wiener filtering is a classical technique for removing additive white noise and is commonly used to suppress unwanted incoherent noise in seismic data. These filters estimate the power of the underlying signal and additive noise by calculating the mean and variance over localised regions of the data, and optimise the separation of these processes through minimising an MSE loss function. These filters work best when the noise is constant-power (‘white’) additive noise, such as Gaussian noise, and provide a useful comparison for benchmarking the performance of DAS-N2N for suppressing incoherent noise in raw DAS data.

We apply a Wiener filter to our data with a window size of 7×7 , which is the size of the receptive field for the 3-layer U-Net used to implement DAS-N2N and jDAS (i.e. the area of input data that a deep learning model can ‘see’, given its depth, filter kernel size, etc.). Smaller window sizes are less effective at suppressing incoherent noise, and larger window sizes more aggressively suppress underlying signal. Filtering is performed using the widely used, open-source SciPy library (Virtanen *et al.* 2020).

5 RESULTS

5.1 Denoising example #1 (in-sample data)

Fig. 3 shows the *S*-wave arrivals from two icequakes recorded by the DAS deployment. These two events occur during the two hours of continuous data (2020-01-17 0100–0300 UTC) used for model training and are thus regarded as ‘in-sample’ data (i.e. data the model has ‘seen’ during training).

From Fig. 3, it is clear that the raw recorded DAS data are corrupted by very strong random noise (Fig. 3a), with the two *S*-wave arrivals (arriving at approximately 0.4 and 0.55 s on DAS channel 0, respectively) almost completely indistinguishable from the random background noise. The intensity of this noise varies spatially along the fibre but appears to show some uniformity over

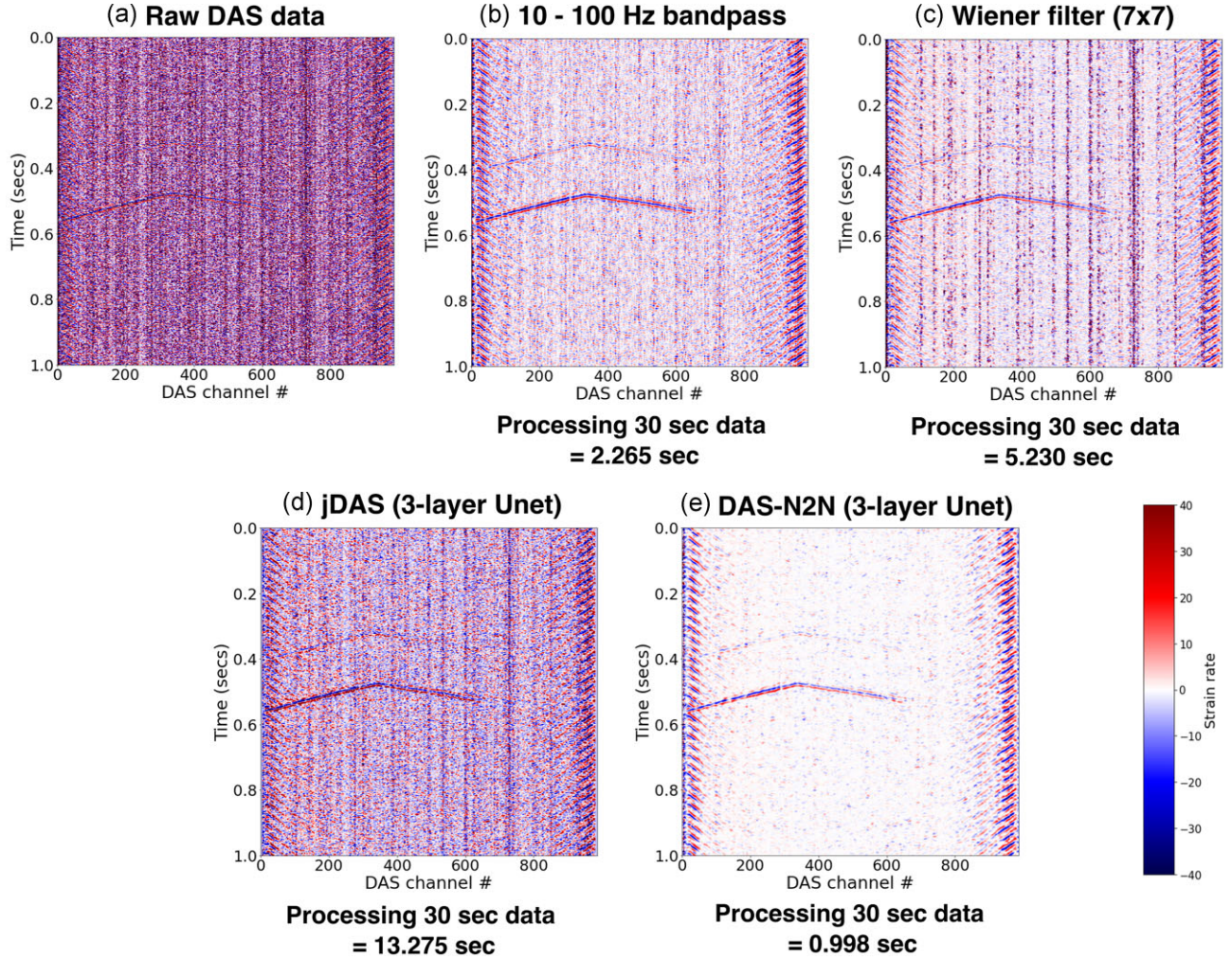


Figure 3. In-sample example of two icequakes (S -wave arrivals only) recorded by DAS deployment (time in seconds after 2020-01-17 01:30:19.232 UTC). (a) Raw DAS data. (b) Butterworth (2-pass, 4th order) 10–100 Hz bandpass filtered DAS data. (c) Wiener filtered (7×7 window size) DAS data. (d) jDAS filtered DAS data. (e) DAS-N2N filtered DAS data. Icequake S waves arrive at DAS channel 0 at time 0.4 and 0.55 s, respectively. Strain rate is recorded in units of strain/s (counts).

small sections (i.e. vertical streaks of high intensity noise are visible over multiple contiguous channels). This suggests that the noise in these data may not be independent across neighbouring channels (a required assumption for jDAS individual channel masking procedure).

Application of bandpass or Wiener filtering (Figs 3b and c) clearly improves the signal-to-noise ratio (SNR) of these arrivals, along with that of the surface waves produced by the DAS power generator visible at each end of the fibre (channels 0–100 and 850–986). However, the higher intensity vertical noise streaks present in the raw data are still present, particularly in the Wiener filtered data. SNR also appears to be improved over the raw data when using either the jDAS (Fig. 3d) or DAS-N2N (Fig. 3e) models, although the degree of noise suppression clearly differs between methods. Data processed by the jDAS model appears to still be strongly contaminated by random noise, including the same higher intensity noise streaks present in the raw and bandpass/Wiener filtered data, whereas, of all the methods presented, the DAS-N2N model appears to perform the greatest degree of background noise suppression (Fig. 3d), without any discernible noise streaks (noisy channels), and is therefore likely to yield the greatest improvement in SNR.

To confirm these observations, we examine estimates of local SNR determined using semblance (a measure of signal similarity across DAS channels; Neidell & Taner 1971). A moving window of size 19 time samples \times 13 DAS channels is applied to the data, with channel-wise cross-correlation and a minimum correlation coefficient of 0.7 used to correct for any local moveout within a window. Semblance is then calculated for each moveout-corrected window using the formula

$$S = \frac{\sum_{i=1}^N \left(\sum_{j=1}^M x_{ij} \right)^2}{M \sum_{i=1}^N \sum_{j=1}^M x_{ij}^2}, \quad (6)$$

where x_{ij} is the moveout-corrected DAS data with time index i and DAS channel j . Eq. (6) effectively represents the ratio of signal coherency to total signal energy. This value can then be used to estimate local SNR (Bakulin *et al.* 2022) by

$$\text{SNR}_{\text{local}} = S/(1 - S). \quad (7)$$

Intuitively, when random noise levels are low, coherent phase arrival signals will be very similar across neighbouring DAS channels, resulting in a high semblance score, S , and thus a high estimate of SNR, according to eq. (7). On the other hand, signals that are

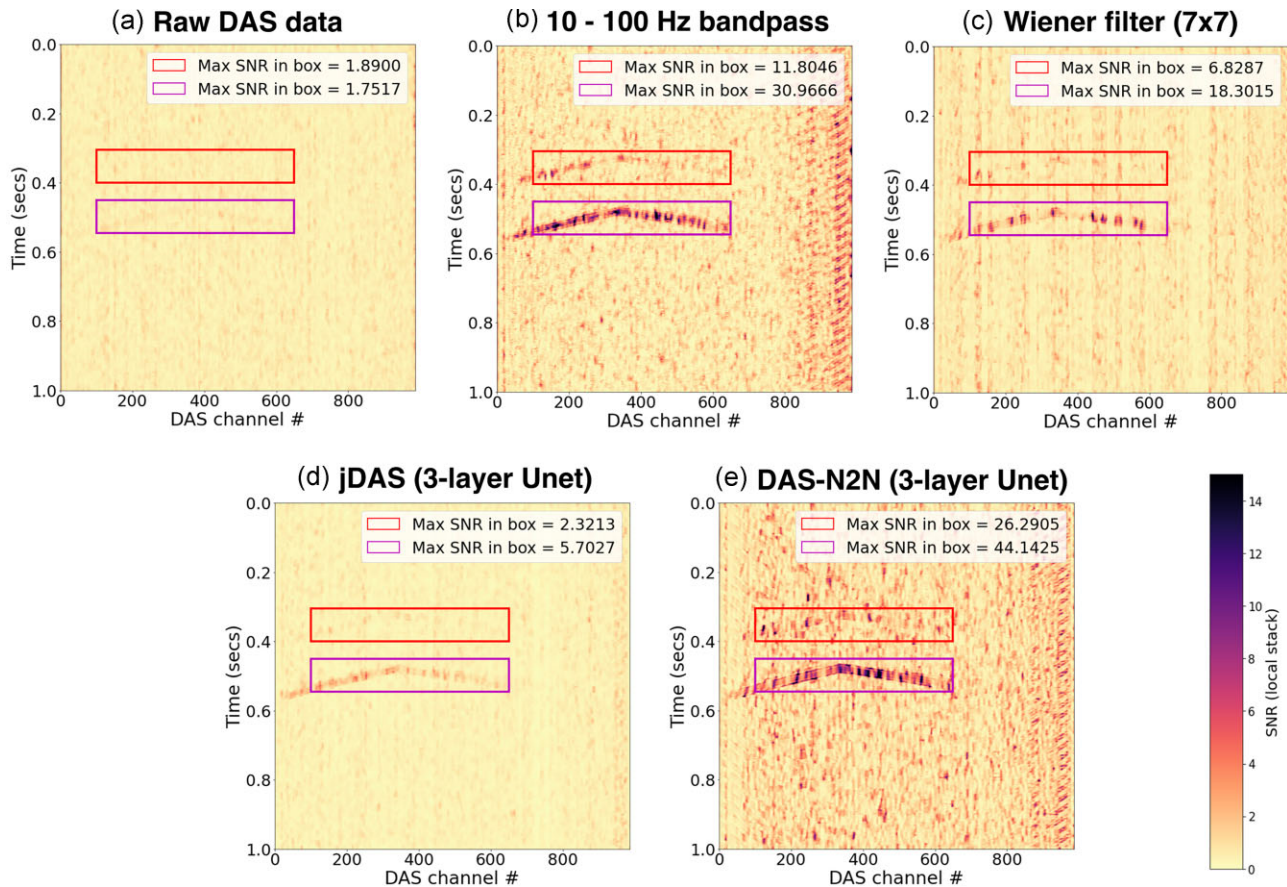


Figure 4. Local signal-to-noise ratio (SNR) estimates for each example in Fig. 3. SNR is calculated using semblance (eq. 7) and a $13\text{-channel} \times 19\text{-sample}$ 2-D moving window.

corrupted by strong random noise will have lower similarity across neighbouring channels and therefore yield a lower semblance score, S , resulting in a lower estimate of SNR. Of all the methods used, DAS-N2N results in the highest SNR for these two S -wave arrivals (Fig. 4e). This is true regardless of window size or chosen summary statistic (e.g. maximum, mean, median or quantile) used to compare local SNR for an arrival, and is also true for all events examined across this DAS deployment. The jDAS model (Fig. 4d) yields a higher SNR than the raw data (Fig. 4a) but fails to suppress background noise as well as conventional bandpass filtering (Fig. 4b), Wiener filtering (Fig. 4c) and DAS-N2N (Fig. 4e).

Fig. 5 shows a single DAS channel trace for each noise suppression method (top of each panel), along with their corresponding time–frequency spectrograms (bottom of each panel). From the spectrogram of the raw data (Fig. 5a), the random measurement noise appears to follow a ‘blue noise’ process, with the power or intensity of the noise increasing with frequency and remaining (locally) time-invariant. This observation could be useful for other machine learning DAS denoising methods, such as R2R or generating noisy/clean signal pairs for supervised learning, where the recorded signal must be corrupted to generate new, independent noise samples for model training.

From these individual traces and spectrograms, it is evident that the DAS-N2N model yields the greatest degree of noise suppression, with both S -wave arrivals clearly visible against background noise in both time (Fig. 5e, top) and time–frequency (Fig. 5e, bottom) domains. Furthermore, the DAS-N2N spectrogram (Fig. 5e, bottom) demonstrates that this method goes beyond simple spectral

filtering: noise within the 10–100 Hz range, which encompasses the dominant frequencies of the two recorded phase arrivals, is also greatly suppressed when compared with bandpass filtering (Fig. 5b), and low-amplitude high-frequency signal components (>100 Hz) are also retained. It is in this manner that DAS-N2N and other machine learning methods can exceed the performance of conventional stop/pass band filtering.

Although, in relative terms, DAS-N2N signals are stronger (with respect to background noise), absolute signal amplitudes after DAS-N2N processing are weaker than their corresponding bandpass filter benchmark (by a factor of approximately $4/5$; see vertical-axis labels on traces in Fig. 5). This $4/5$ scaling appears to be consistent across all events examined from this deployment (Butcher *et al.* 2021). This amplitude difference likely relates to signal leakage, where some of the desired underlying signal is suppressed with the noise, and is a common issue with denoising methods based on MAE and MSE loss functions (e.g. Birnie & Alkhalifah 2022), and where the raw data are very noisy or large regions of the underlying data are ‘empty’ (i.e. vast majority of the data contain no events; see Section 6). A similar degree of signal leakage occurs with the Wiener filtered data (Fig. 5c), which is also optimised using an MSE loss function.

5.2 Denoising example #2 (out-of-sample data)

In Fig. 6, we present another short section of recorded DAS data from a time period outside of our training set (2020-01-17 04:42:07.903 UTC). This section was chosen to demonstrate the performance of our model on so-called ‘out-of-sample’ data, with

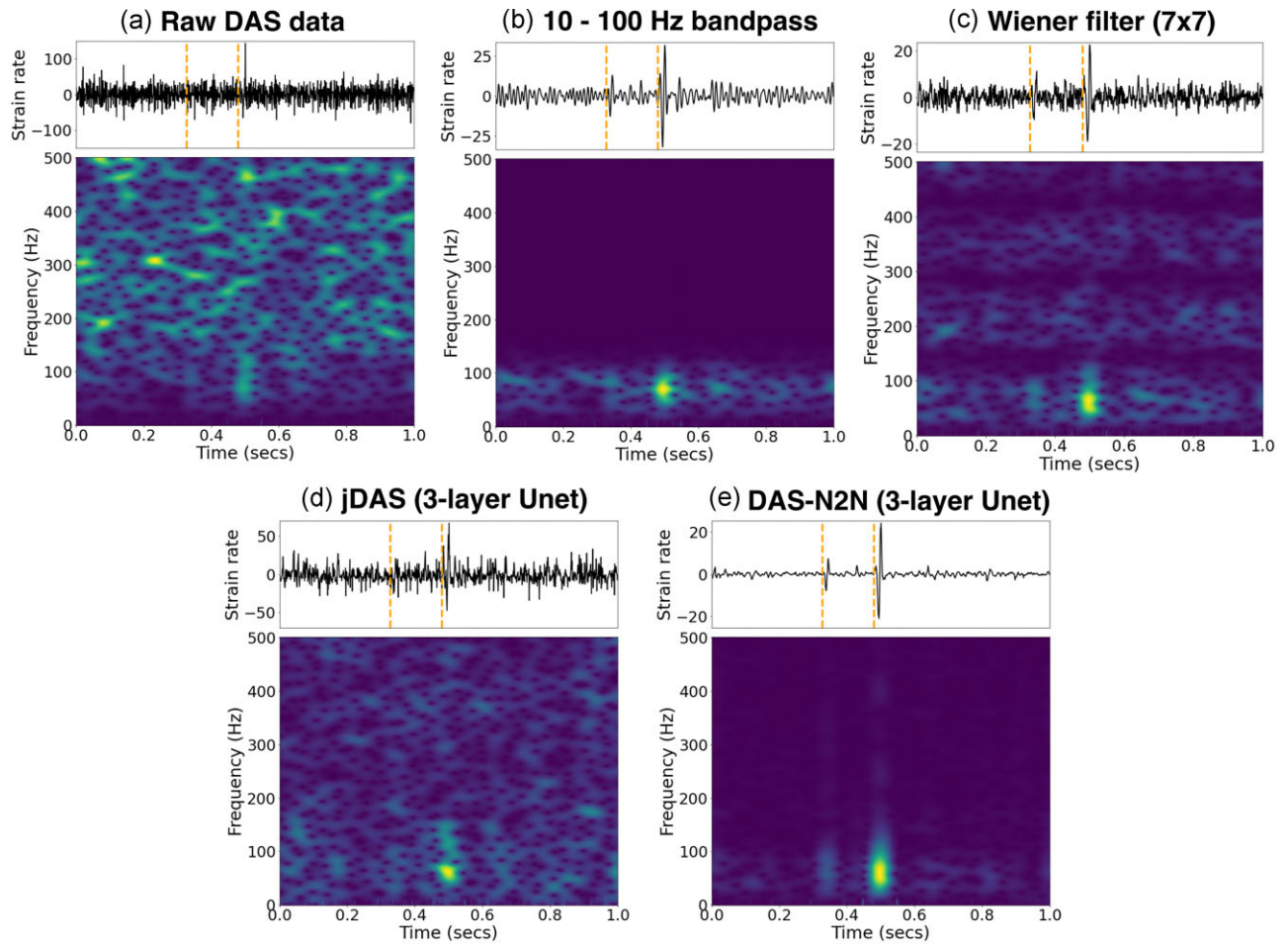


Figure 5. Individual DAS trace (top panel) and corresponding spectrogram (bottom panel) for DAS channel 255 in each example in Fig. 3. Strain rate is recorded in units of strain/s (counts).

three *S*-wave arrivals from discrete icequake events (arrival times on DAS channel 0 at approximately 0.4, 0.63 and 0.95 s, respectively) observed in the bandpass/Wiener filtered, jDAS denoised and DAS-N2N denoised data (Figs 6b–e).

From Fig. 6, it is clear that performance of all methods on out-of-sample data is similar to that on in-sample data (Fig. 3), with DAS-N2N unequivocally performing the greatest degree of noise suppression (Fig. 6e). This suggests that the DAS-N2N model has been adequately trained to generalize to sections of data outside of the training set and can be used for continual monitoring for this specific deployment. The DAS-N2N model also yields the highest local SNR for all three *S*-wave arrivals (again, regardless of window size or chosen summary statistic; Fig. 7e), with bandpass filtering also performing better than Wiener filtering and the jDAS model (Figs 7b–d).

As with the in-sample data, Fig. 8 shows the trace and spectrogram for an individual DAS channel processed by each method for this out-of-sample section of data. The three *S*-wave arrivals are difficult to discern in any of the traces or spectrograms except for in the DAS-N2N case (Fig. 8e), where all three arrivals appear as distinct features in both the time (top panel) and time–frequency (bottom panel) domains. Again, the raw observational noise appears to broadly follow a blue noise process, albeit with an apparent high frequency ‘ridge’ at approximately 285 Hz (Fig. 8a). Unlike the other methods, our DAS-N2N model adequately suppresses noise

across the full spectrum (including the frequency band encompassing the phase arrivals) and retains weaker high-frequency signal components (Fig. 8e), a feat beyond the capability of standard spectral filtering methods. The approximately 4/5 scaling of absolute signal amplitudes (i.e. signal leakage) for DAS-N2N and Wiener filtering when compared with bandpass filtering is also present in this example.

6 DISCUSSION

In terms of random (incoherent) noise suppression, DAS-N2N unequivocally performs better than conventional Butterworth bandpass/Wiener filtering and a comparable self-supervised machine learning approach (jDAS) for the data presented here. This improved performance is immediately apparent in plots of the processed data (Figs 3 and 6), where vertical bands of higher intensity noise over contiguous channels are suppressed only by DAS-N2N (as their locations differ between the two spliced fibres), and when estimates of SNR are determined through semblance (Figs 4 and 7). Spectrograms from individual DAS channels (Figs 5 and 8) show that part of this improved performance relates to the ability of machine learning models to suppress noise that lies in the same frequency band as the desired underlying signal. Such a feat will never be fully achievable for filters that rely on isolating or suppressing certain frequency bands, even when such techniques are enhanced

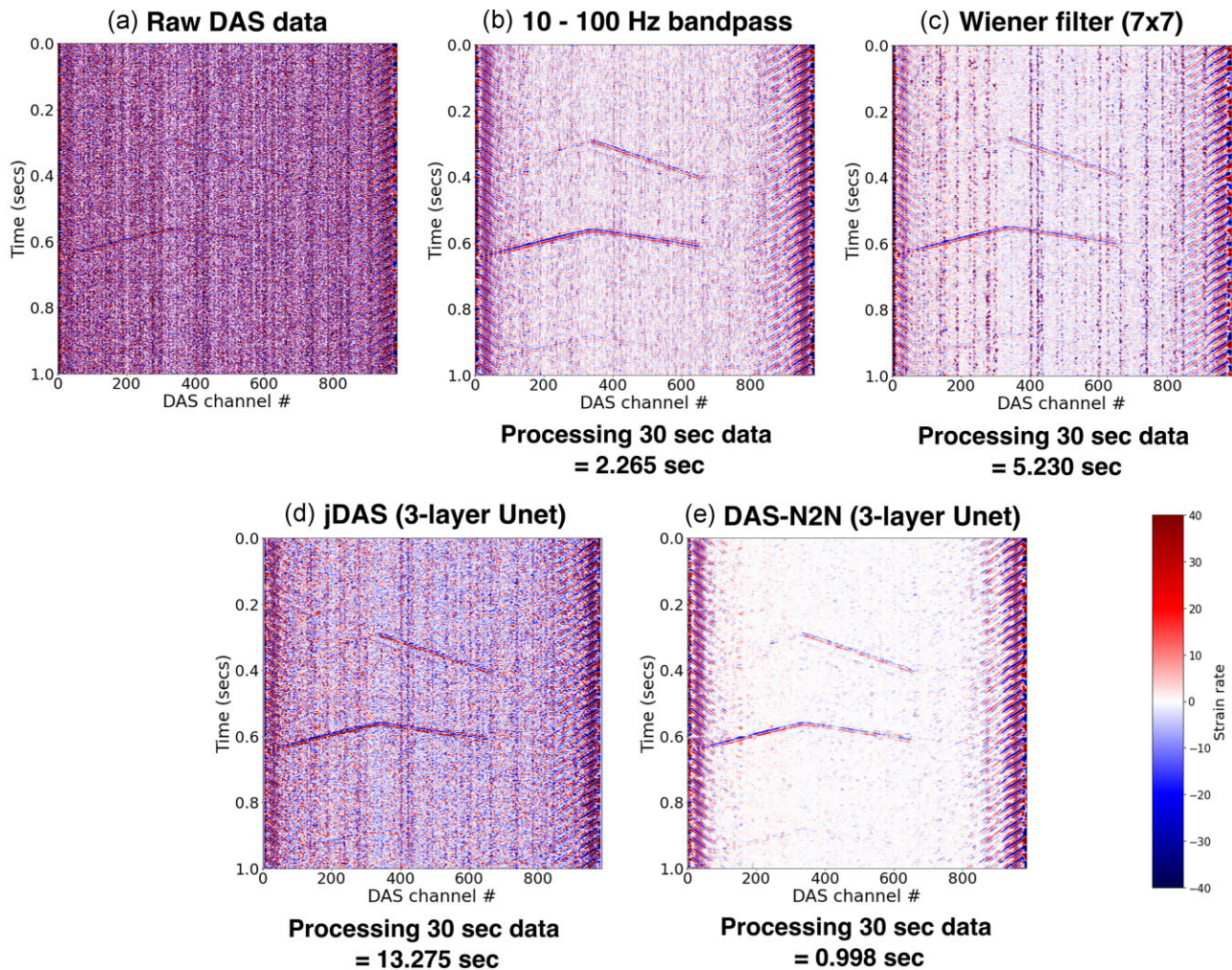


Figure 6. Out-of-sample example of three icequakes (*S*-wave arrivals only) recorded by DAS deployment (time in seconds after 2020-01-17 04:42:07.903 UTC). (a) Raw DAS data. (b) Butterworth (2-pass, 4th order) 10–100 Hz bandpass filtered DAS data. (c) Wiener filtered (7×7 window size) DAS data. (d) jDAS filtered DAS data. (e) DAS-N2N filtered DAS data. Icequake *S* waves arrive at DAS channel 0 at time 0.4, 0.63 and 0.95 s, respectively. Strain rate is recorded in units of strain/s (counts).

through adaptive parametrization algorithms or the use of both temporal and spatial frequencies. Furthermore, common yet undesired causal filtering artefacts, such as precursory ringing before phase arrivals and signal polarity changes, will not be present in DAS-N2N processed data as the model processes the raw data directly and such features would only serve to increase model loss between the processed data and the target data. We note, however, that our DAS-N2N model does exhibit a degree of signal leakage, consistently reducing the absolute amplitude of the underlying signal by a factor of 1/5. Extensive experimentation with model depth, kernel size, choice of loss function (e.g. MAE, Huber), pre-processing steps (e.g. median removal and quantile normalization to reduce the impact of outliers), and architecture style (e.g. ResNet) did not yield any consistent improvement in this regard. As such, the issue of signal leakage is one that cannot be trivially solved here, and we leave this for future areas of research. It is worth mentioning that, regardless of this observed signal leakage, data processed by DAS-N2N exhibits higher signal-to-noise levels than any of the other methods presented, and its GPU-optimised implementation is also much more efficient (two of the primary factors controlling the effectiveness of subsequent imaging/event detection techniques and the viability of the method for processing large DAS data sets).

Furthermore, once trained, our DAS-N2N model also shows an impressive degree of generalisation to other iDAS data sets, without the need for any retraining or fine-tuning (Fig. 9).

Fig. 9 shows application of our pre-trained Antarctica model on data collected during a 4-d DAS experiment conducted on two submarine cables extending off the U.S. west coast from Pacific City, Oregon (Wilcock & OOI 2023). The south-most cable, which we examine here (Fig. 9), was interrogated by an iDASv3 DAS interrogator and extends over 80 km offshore. Large amplitude, long period ocean microseisms are clearly visible over background noise in both the unfiltered raw (Fig. 9a) and DAS-N2N processed (Fig. 9d) data. This is most apparent in the individual channel traces, where DAS-N2N filters strong high-frequency noise contaminating these long period signals (Figs 9a and d, bottom). Application of a 10 Hz high-pass filter (Fig. 9b) reveals the presence of a lower amplitude blue whale ‘A’ call (vertical pulse-like signal observed approximately 40 km along fibre; Wilcock *et al.* 2023) and a much higher degree of incoherent noise as you go further along the fibre (due to decay of the interrogator light source). Subsequent application of DAS-N2N (Fig. 9e) greatly suppresses incoherent noise along the full extent of the fibre, revealing the individual pulses of the blue whale ‘A’ call (as well as other fin whale calls) in incredible

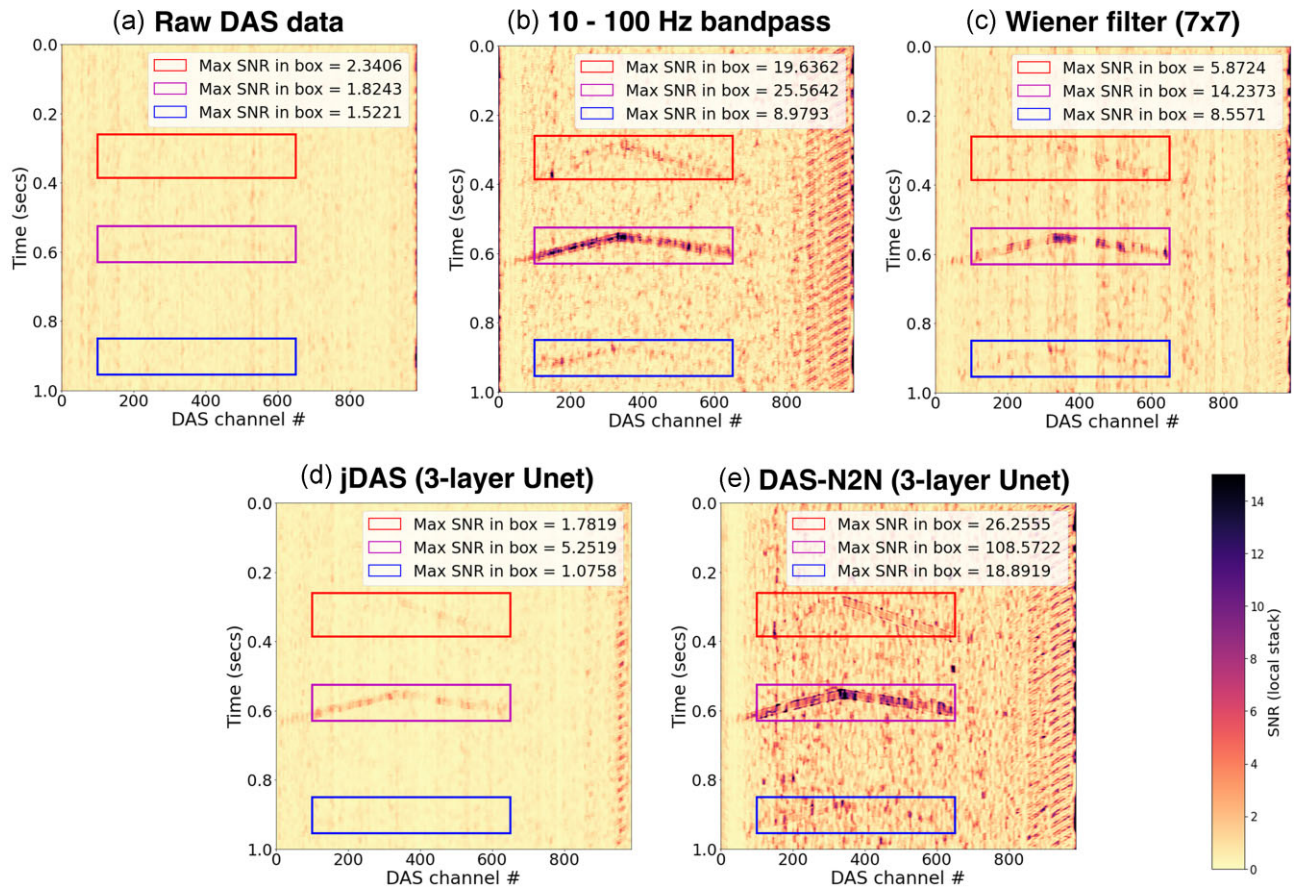


Figure 7. Local SNR estimates for each example in Fig. 6. SNR is calculated using semblance and a $13\text{-channel} \times 19\text{-sample}$ 2-D moving window.

detail (Fig. 9f). It is likely that DAS-N2N will also generalise well to other iDAS data sets (as this was the interrogator model used to acquire its training data), but will almost certainly need retraining to perform well on data collected by other interrogator models (due to differences in light source power, components used, measurement standards, etc.).

By learning to map random noise to the distribution mean, DAS-N2N learns to perform the equivalent of a large- N stack (sum or average) over many noisy copies of the signal, analogous to the averaging of many short, independent, noisy exposures acquired during long-exposure low-light photography (Lehtinen *et al.* 2018). The advantage of DAS-N2N over simple stacking, however, is that it only requires the acquisition of two noisy copies of the data for training, and only a single noisy copy of the data once trained. Furthermore, the noise in DAS-N2N processed data will be mapped to its distribution mean, whereas the noise in stacked data will only be mapped to its (statistically weaker) point-wise sample mean.

The DAS-N2N approach, in general, is an order of magnitude faster than self-supervised ‘fill-in-the-gap’ approaches, such as jDAS (Figs 3 and 6), as the latter’s masking procedure means it must process N times more data (where $1/N$ is the fraction of input data masked). When compared with a jDAS model trained with the same model architecture, training hyperparameters and data pre-processing steps, DAS-N2N also performs better at the task of noise suppression on the microseismic icequake data presented here. In general, self-supervised learning methods will likely struggle to match or exceed the performance of weakly supervised learning methods, particularly on data with very high noise levels, as they

are tasked with interpolating missing sections of data, which will always suffer from a degree of averaging over all possible values. On the other hand, weakly and fully supervised learning methods have the complete unmasked signal present in both the input and target data, meaning a direct 1-to-1 mapping can, theoretically, be learned.

In terms of computational efficiency, our 3-layer DAS-N2N model processes 30 s of recorded data in less than 1 s (Figs 3e and 6e) using the TensorFlow (version 2.3.0) Python framework and a single NVIDIA GeForce RTX 2080 Ti GPU. This is more than twice as fast as conventional channel-wise bandpass filtering using optimized low-level C programming language routines (Figs 3b and 6b). Any further algorithmic or filtering steps that yield improvements over bandpass filtering (Isken *et al.* 2022; Chen *et al.* 2023) will obviously have further computational demands, making them increasingly less feasible for real-time passive monitoring purposes.

Arguably, the largest observed drawback of DAS-N2N against the other noise suppression methods presented is the degree of apparent signal leakage observed after data processing. This signal leakage is most likely a consequence of using an MSE loss function during training, but could also be due to unforeseen issues with our data pre- and post-processing steps (e.g. dividing and rescaling by standard deviation of raw data) or an engineering aspect of the two fibres (e.g. channels on two fibres not lining up exactly). In any case, the degree of signal leakage appears to be consistent across observed signals in the data presented here and therefore, once a consistent scaling between DAS-N2N and bandpass filtered event signal amplitudes has been determined, one can apply a simple correction (e.g. for

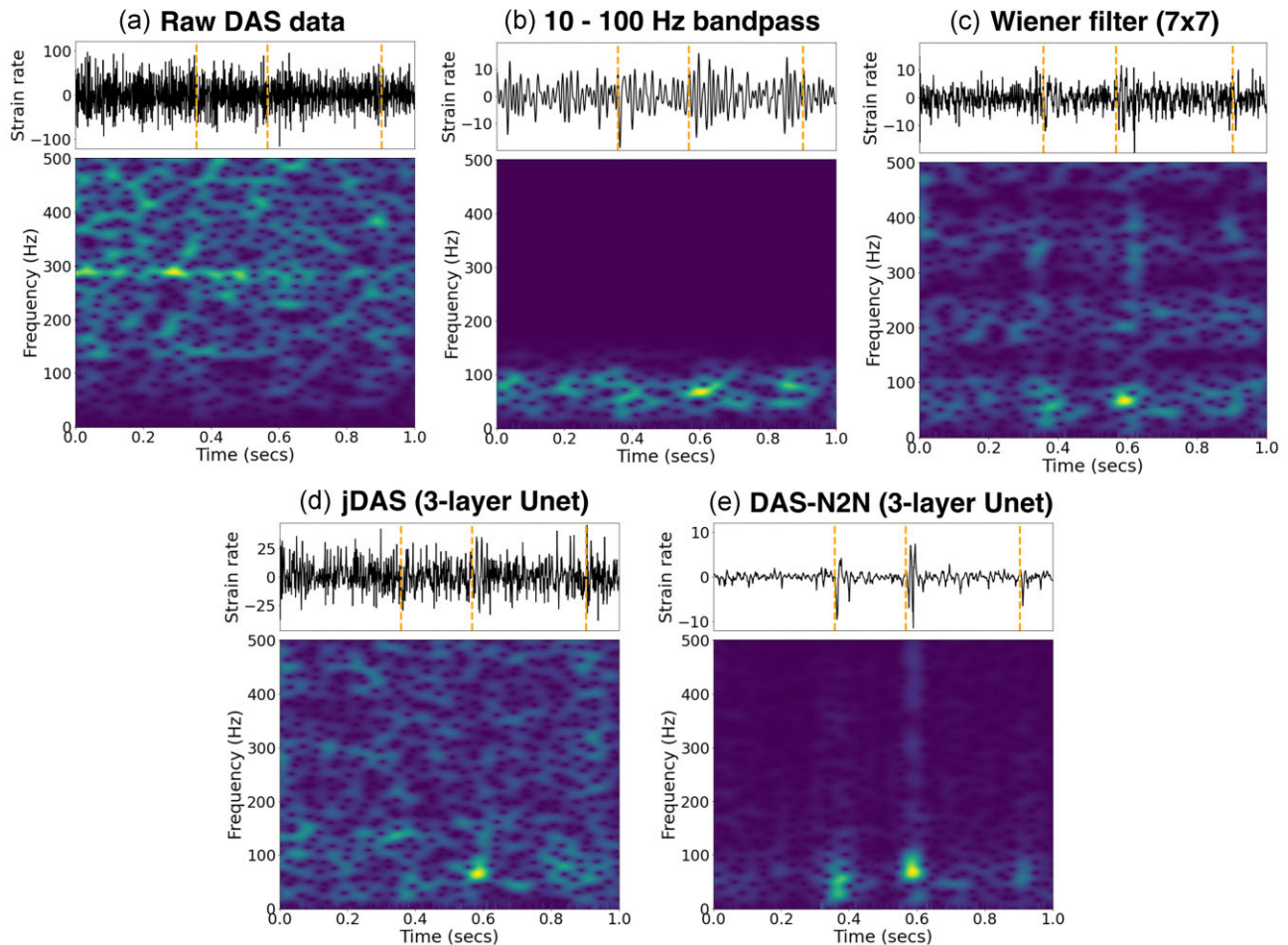


Figure 8. Individual DAS trace (top panel) and corresponding spectrogram (bottom panel) for DAS channel 587 in each example in Fig. 6. Strain rate is recorded in units of strain/s (counts).

earthquake magnitude and source parameter estimation). However, we do not perform any correction here in order to keep processing methods as transparent, comparable and simple as possible.

Another apparent drawback of all the methods presented is the inability to suppress unwanted coherent noise (e.g. the surface waves produced by the power generator for this DAS deployment). At present, this is likely still best performed by standard frequency filtering techniques (e.g. stop band/notch, filters), as such processes tend to produce signals with predictable and narrow-band frequency content (e.g. 33 and 66 Hz for the power generator surface waves in Figs 3 and 6; Hudson *et al.* 2021b).

Finally, in terms of model architecture, we follow Lehtinen *et al.* (2018) and van den Ende *et al.* (2021) in using a simple U-Net architecture (Ronneberger *et al.* 2015). However, there are likely to be more effective model design choices for DAS-N2N and jDAS denoising than the ones chosen in these studies. Identifying optimal model architectures and training hyperparameters is often a challenging and sizeable task, involving either extensive manual trial-and-error or computationally expensive iterative search strategies (e.g. Elsken *et al.* 2019; Hutter *et al.* 2019; White *et al.* 2023). We therefore focus the scope of this paper on the general applicability of N2N as a simple, effective strategy for denoising spliced-fibre DAS data without any clean training data or manual data curation. Furthermore, by demonstrating the effectiveness of DAS-N2N using a very small model (by deep learning standards), we provide evidence

that DAS-N2N processing can be applied rapidly (well within ‘real-time’ constraints) and could be suitable for low-powered devices and edge networks.

7 CONCLUSIONS

In this paper, we demonstrate the use of a weakly supervised machine learning method for fully automated random noise suppression in DAS data (which we call DAS-N2N after the corresponding N2N technique in image processing; Lehtinen *et al.* 2018). The method is ideally suited to DAS and other distributed optical fibre measurements (e.g. distributed temperature sensing; DTS) due to the ability to simultaneously record data across two spliced fibres within a single cable jacket. Advantageously, a DAS-N2N model can be trained end-to-end without any manual curation or labelling: simply, a section of data recorded on one of the spliced fibres serves as input data, with the corresponding section of data recorded on the other spliced fibre serving as target data (Figs 2a and b). Once trained, the model only requires input data from a single unspliced fibre (Fig. 2c), meaning there is no increase in data volumes to be stored after model training. Given the model’s ability to generalise to other DAS settings, or if fibres can be temporarily (i.e. mechanically) or more permanently (i.e. fusion) spliced at some later point in time to facilitate model retraining, this approach can be applied retroactively to existing deployments with unspliced fibres.

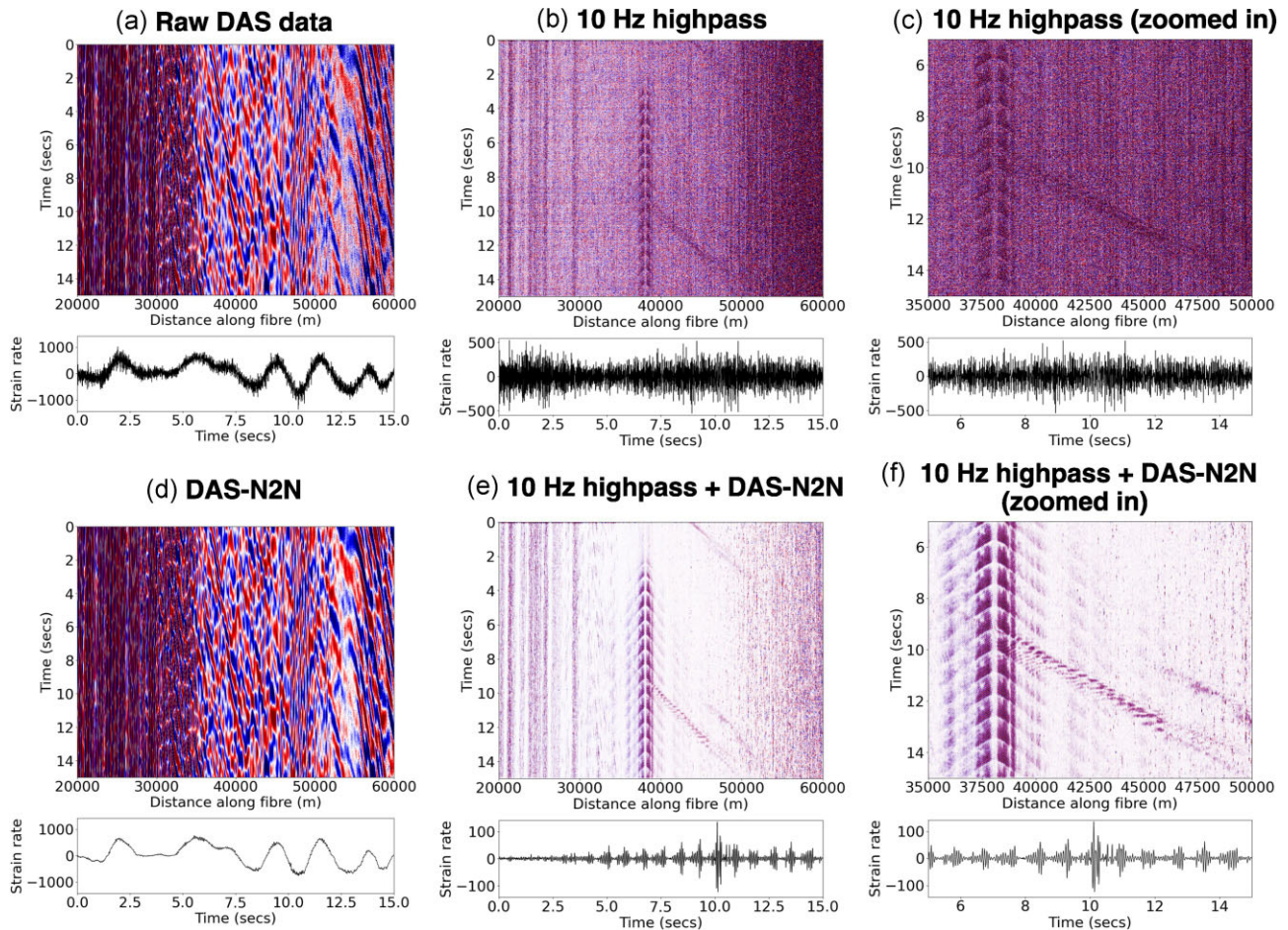


Figure 9. DAS data recorded by Wilcock & OOI (2023) on submarine cable extending off the Oregon coast, USA (time given in seconds after 2021-11-02 10:36:09.839 UTC). Top panel: data for all DAS channels between 20 and 60 km along south cable. Bottom panel: data for individual DAS channel (40.06 km along fibre). (a) Raw DAS data. (b) Butterworth (2-pass, 4th order) 10 Hz highpass filtered DAS data. (c) Zoomed in version of highpass filtered data in (b). (d) DAS-N2N filtered DAS data. (e) Application of Butterworth 10 Hz highpass filter, followed by DAS-N2N. (f) Zoomed in version of highpass filtered + DAS-N2N processed data in (e). Strain rate is recorded in units of strain/s (counts).

We demonstrate that DAS-N2N is inherently more effective and efficient than conventional bandpass filtering, Wiener filtering and self-supervised learning approaches. In particular, DAS-N2N is able to suppress noise lying within the same frequency range as the signal of interest (which is not possible for frequency-based filtering) and is an order of magnitude faster than self-supervised learning, due to the latter's masking procedure. Furthermore, the presence of the complete unmasked underlying signal in both the input and target data when training a DAS-N2N model means that the signal can be retained through a 1-to-1 mapping, whereas self-supervised learning effectively performs a form of interpolation to predict the masked signal, which becomes more challenging as noise levels increase. Lastly, we demonstrate that a DAS-N2N model can be extremely lightweight (e.g. three model layers) and efficient, processing data in a fraction of the acquisition time (1/30 in the examples presented here) when optimized with a single GPU, and faster than standard frequency filtering routines optimized using compiled low-level programming languages, such as C. This offers the possibility of such models being further optimized, compiled and compressed for processing on low-powered devices and edge networks, which will be crucial for offshore or remote early warning monitoring settings.

ACKNOWLEDGMENTS

We thank NERC British Antarctic Survey for logistics and field support, with particular thanks to Sofia Kufner for her part in deploying the DAS in the field. We also thank Silixa for the loan of an iDAS interrogator. This work was funded by a NERC Collaborative Antarctic Science Scheme grant (grant number CASS-166), the BEAMISH project (NERC grants: NE/G014159/1, NE/G013187/1) and the Digital Monitoring of CO₂ storage project (DigiMon; project no. 299622), which is part of the Accelerating CCS Technologies (ACT2) program (<https://www.act-ccs.eu/>). Author Sacha Lapins was funded by the above DigiMon project and a Leverhulme Trust Early Career Fellowship (<https://www.leverhulme.ac.uk/>). Authors Antony Butcher, Michael Kendall and Thomas Hudson were also funded by the DigiMon project. Author Maximilian Werner was funded by Natural Environment Research Council (NERC) grant NE/R017956/1 ('EQUIPT4RISK'). Author Jemma Gunning was supported by an EarthArt Fellowship provided by the School of Earth Sciences, University of Bristol. We are very grateful for comments and feedback provided by Feng Cheng and an anonymous reviewer. This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol - <http://www.bris.ac.uk/acrc/>.

DATA AVAILABILITY

The seismic data will be made available through the UK NERC Polar Data Centre. At the time of submission, the models, example data and code to reproduce the results in this paper were made available on GitHub (<https://github.com/sachalapins/DAS-N2N>), with the version associated with this paper archived through Zenodo (Lapins et al. 2023, doi:10.5281/zenodo.7825683).

REFERENCES

- Abadi, M. et al., 2015. *TensorFlow: large-scale machine learning on heterogeneous systems*, Software available from [tensorflow.org](https://www.tensorflow.org).
- Ajo-Franklin, J.B. et al., 2019. Distributed acoustic sensing using dark fiber for near-surface characterization and broadband seismic event detection, *Sci. Rep.*, **9**(1), 1–14.
- Bacon, M., Simm, R. & Redshaw, T., 2003. *3-D Seismic Data Acquisition and Processing*, pp. 17–56, Cambridge Univ. Press.
- Bakulin, A., Silvestrov, I. & Protasov, M., 2022. Research note: signal-to-noise ratio computation for challenging land single-sensor seismic data, *Geophys. Prospect.*, **70**(3), 629–638.
- Batson, J. & Royer, L., 2019. Noise2self: blind denoising by self-supervision, in *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, California, PMLR 97, pp. 524–533.
- Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y. & Wassermann, J., 2010. Obspy: a python toolbox for seismology, *Seismol. Res. Lett.*, **81**(3), 530–533.
- Birnie, C. & Alkhalifah, T., 2022. Transfer learning for self-supervised, blind-spot seismic denoising, *Front. Earth Sci.*, **10**, doi:10.3389/feart.2022.1053279.
- Birnie, C., Ravasi, M., Liu, S. & Alkhalifah, T., 2021. The potential of self-supervised networks for random noise suppression in seismic data, *Artif. Intell. Geosci.*, **2**, 47–59.
- Bromirski, P.D., Stephen, R.A. & Gerstoft, P., 2013. Are deep-ocean-generated surface-wave microseisms observed on land?, *J. geophys. Res.*, **118**(7), 3610–3629.
- Butcher, A., Hudson, T., Kendall, J., Kufner, S., Brisbourne, A. & Stork, A., 2021. Radon transform-based detection of microseismicity on das networks: a case study from Antarctica, in *EAGE GeoTech 2021 Second EAGE Workshop on Distributed Fibre Optic Sensing*, Online, 1–5 March 2021, Vol. 2021, pp. 1–4.
- Calvarons, A.F., 2021. Improved Noise2Noise denoising with limited data, in *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Nashville, TN, USA, 19–25 June 2021, pp. 796–805, IEEE, doi:10.1109/CVPRW53098.2021.00089.
- Chen, Y. et al., 2023. Denoising of distributed acoustic sensing seismic data using an integrated framework, *Seismol. Res. Lett.*, **94**(1), 457–472.
- Dean, T., Cuny, T. & Hartog, A.H., 2017. The effect of gauge length on axially incident P-waves measured using fibre optic distributed vibration sensing, *Geophys. Prospect.*, **65**(1), 184–193.
- Dou, S. et al., 2017. Distributed acoustic sensing for seismic monitoring of the near surface: a traffic-noise interferometry case study, *Sci. Rep.*, **7**(1), doi:10.1038/s41598-017-11986-4.
- du Toit, H.J., Goldswain, G. & Olivier, G., 2022. Can das be used to monitor mining induced seismicity?, *Int. J. Rock Mech. Min. Sci.*, **155**(5), doi:10.1016/j.ijrmms.2022.105127.
- Duncan, G. & Beresford, G., 1994. Slowness adaptive f-k filtering of prestack seismic data, *Geophysics*, **59**(1), 140–147.
- Elsken, T., Metzen, J.H. & Hutter, F., 2019. Neural architecture search: a survey, *J. Mach. Learn. Res.*, **20**(55), 1–21.
- Ericsson, L., Gouk, H., Loy, C.C. & Hospedales, T.M., 2022. Self-supervised representation learning: introduction, advances, and challenges, *IEEE Sig. Proc. Mag.*, **39**(3), 42–62.
- Glorot, X. & Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, Vol. 9 of Proceedings of Machine Learning Research*, pp. 249–256, PMLR, Chia Laguna Resort, Sardinia, Italy.
- Hartog, A.H., 2017. *An Introduction to Distributed Optical Fibre Sensors*, CRC Press.
- Hartog, A.H., Frignet, B., Mackie, D. & Clark, M., 2014. Vertical seismic optical profiling on wireline logging cable, *Geophys. Prospect.*, **62**(4), 693–701.
- Havskov, J. & Ottemöller, L., 2010. *Routine Data Processing in Earthquake Seismology*, Springer.
- Hudson, T., Baird, A., Kendall, J., Kufner, S., Brisbourne, A., Smith, A. & Butcher, A., 2021a. *Rutford ice stream distributed acoustic sensing dataset associated with the publication: distributed acoustic sensing (DAS) for natural microseismicity studies: a case study from Antarctica*, Zenodo, doi:10.5281/zenodo.4778368.
- Hudson, T.S. et al., 2021b. Distributed acoustic sensing (DAS) for natural microseismicity studies: a case study from antarctica, *J. geophys. Res.*, **126**(7), 1–19.
- Hutter, F., Kothhoff, L. & Vanschoren, J., 2019. *Automated Machine Learning, the Springer Series on Challenges in Machine Learning*, Springer International Publishing.
- Isken, M.P., Vasyura-Bathke, H., Dahm, T. & Heimann, S., 2022. Denoising distributed acoustic sensing data using an adaptive frequency-wavenumber filter, *J. geophys. Int.*, **231**(2), 944–949.
- Jeng, Y., Li, Y.-W., Chen, C.-S. & Chien, H.-Y., 2009. Adaptive filtering of random noise in near-surface seismic and ground-penetrating radar data, *J. Appl. Geophys.*, **68**(1), 36–46.
- Jouset, P. et al., 2022. Fibre optic distributed acoustic sensing of volcanic events, *Nat. Commun.*, **13**(1), doi:10.1038/s41467-022-29184-w.
- Kennett, B.L., 2022. The seismic wavefield as seen by distributed acoustic sensing arrays: local, regional and teleseismic sources, *Proc. R. Soc., A*, **478**(2258), doi:10.1098/rspa.2021.0812.
- Kingma, D.P. & Ba, J., 2014. Adam: a method for stochastic optimization, in *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, San Diego, pp. 1–15.
- Klochikhina, E., Crawley, S., Frolov, S., Chemingui, N. & Martin, T., 2020. Leveraging deep learning for seismic image denoising, *First Break*, **38**(7), 41–48.
- Koper, K.D. & Burlacu, R., 2015. The fine structure of souble-frequency microseisms recorded by seismometers in north america, *J. geophys. Res.*, **120**, 1677–1691.
- Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C. & Wassermann, J., 2015. Obspy: a bridge for seismology into the scientific python ecosystem, *Comput. Sci. Discov.*, **8**, doi:10.1088/1749-4699/8/1/014003.
- Krull, A., Buchholz, T.-O. & Jug, F., 2018. Noise2Void - learning denoising from single noisy images, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 15–20 June 2019, IEEE, doi:10.48550/arXiv.1811.10980.
- Lapins, S., Roman, D.C., Rougier, J., De Angelis, S., Cashman, K.V. & Kendall, J.-M., 2020. An examination of the continuous wavelet transform for volcano-seismic spectral analysis, *J. Volc. Geotherm. Res.*, **389**, doi:10.1016/j.jvolgeores.2019.106728.
- Lapins, S., Butcher, A., Kendall, J.-M., Hudson, T.S., Stork, A.L., Werner, M.J., Gunning, J. & Brisbourne, A.M., 2023. *Data, model and code to accompany paper “DAS-N2N: Machine learning distributed acoustic sensing (DAS) signal denoising without clean data”*, Zenodo, doi:10.5281/zenodo.7825683.
- Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M. & Aila, T., 2018. Noise2Noise: learning image restoration without clean data, in *Proceedings of the 35th International Conference on Machine Learning, ICML 2018*, Stockholm, Sweden, PMLR 80, pp. 4620–4631.
- Li, Y. & Ma, Z., 2021. Deep learning-based noise reduction for seismic data, *J. Phys.: Conf. Ser.*, **1861**(1), doi:10.1088/1742-6596/1861/1/012011.
- 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.

- Lindsey, N.J., Rademacher, H. & Ajo-Franklin, J.B., 2020. On the broadband instrument response of fiber-optic das arrays, *J. geophys. Res.*, **125**(2), 1–16.
- Liu, B., Yue, J., Zuo, Z., Xu, X., Fu, C., Yang, S. & Jiang, P., 2022. Unsupervised deep learning for random noise attenuation of seismic data, *IEEE Geosci. Remote Sens. Lett.*, **19**, 1–5.
- Megies, T., Beyreuther, M., Barsch, R., Krischer, L. & Wassermann, J., 2011. Obspy - what can it do for data centers and observatories?, *Ann. Geophys.*, **54**(1), 47–58.
- Mousa, W.A., 2019. *Advanced Digital Signal Processing of Seismic Data*, Cambridge Univ. Press.
- Nayak, A. et al., 2021. Distributed acoustic sensing using dark fiber for array detection of regional earthquakes, *Seismol. Res. Lett.*, **92**(4), 2441–2452.
- Neidell, N.S. & Taner, M.T., 1971. Semblance and other coherency measures for multichannel data, *Geophysics*, **36**(3), 482–497.
- Pang, T., Zheng, H., Quan, Y. & Ji, H., 2021. Recorruped-to-recorruped: unsupervised deep learning for image denoising, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2043–2052.
- Parker, T., Shatalin, S. & Farhadiroushan, M., 2014. Distributed acoustic sensing - a new tool for seismic applications, *First Break*, **32**(2), 61–69.
- Proakis, J.G. & Manolakis, D., 1996. *Digital Signal Processing: Principles, Algorithms and Applications*, 3rd edn, Prentice-Hall, Inc.,
- Ronneberger, O., Fischer, P. & Brox, T., 2015. *U-Net: Convolutional Networks for Biomedical Image Segmentation*, Vol. 9351 of *Lecture Notes in Computer Science*, pp. 234–241, Springer International Publishing.
- Scherbaum, F., 2001. *Of Poles and Zeros*, 2nd edn, Springer Netherlands.
- Shinohara, M., Yamada, T., Akuhara, T., Mochizuki, K. & Sakai, S., 2022. Performance of seismic observation by distributed acoustic sensing technology using a seafloor cable off Sanriku, Japan, *Front. Mar. Sci.*, **9**(April), 1–13.
- Tibi, R., Hammond, P., Brogan, R., Young, C.J. & Koper, K., 2021. Deep learning denoising applied to regional distance seismic data in utah, *Bull. seism. Soc. Am.*, **111**(2), 775–790.
- van den Ende, M., Lior, I., Ampuero, J.P., Sladen, A., Ferrari, A. & Richard, C., 2021. A self-supervised deep learning approach for blind denoising and waveform coherence enhancement in distributed acoustic sensing data, *IEEE Trans. Neural Networks Learn. Syst.*, **34**(7), 1–14.
- van den Ende, M., Ferrari, A., Sladen, A. & Richard, C., 2023. Deep deconvolution for traffic analysis with distributed acoustic sensing data, *IEEE Trans. Intell. Transport. Syst.*, **24**(3), 2947–2962.
- van Engelen, J.E. & Hoos, H.H., 2020. A survey on semi-supervised learning, *Mach. Learn.*, **109**(2), 373–440.
- Virtanen, P. et al., 2020. Scipy 1.0: fundamental algorithms for scientific computing in python, *Nat. Methods*, **17**(3), 261–272.
- Walter, F., Gräff, D., Lindner, F., Paitz, P., Köpfl, M., Chmiel, M. & Fichtner, A., 2020. Distributed acoustic sensing of microseismic sources and wave propagation in glaciated terrain, *Nat. Commun.*, **11**(1), doi:10.1038/s41467-020-15824-6.
- White, C., Safari, M., Sukthanker, R., Ru, B., Elksen, T., Zela, A., Dey, D. & Hutter, F., 2023. *Neural architecture search: Insights from 1000 papers*, arXiv, doi:10.48550/arXiv.2301.08727.
- Wilcock & OOI, 2023. *Rapid: A community test of distributed acoustic sensing on the ocean observatories initiative regional cabled array [data set]*, Ocean Observatories Initiative, doi:10.58046/5J60-FJ89.
- Wilcock, W. S.D., 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, doi:10.1121/10.0017104.
- Williams, A., Kendall, J., Verdon, J., Clarke, A. & Stork, A., 2020. Applying conventional filtering and picking approaches to das microseismic data, *First EAGE Workshop on Fibre Optic Sensing*, **2020**, 1–5.
- Yang, L., Liu, X., Zhu, W., Zhao, L. & Beroza, G.C., 2022. Toward improved urban earthquake monitoring through deep-learning-based noise suppression, *Sci. Adv.*, **8**(15), 3564, doi:10.1126/sciadv.abl3564.
- Zhou, W., Butcher, A., Brisbourne, A.M., Kufner, S.K., Kendall, J.M. & Stork, A.L., 2022. Seismic noise interferometry and distributed acoustic sensing (das): inverting for the firm layer S-velocity structure on ruford ice stream, antarctica, *J. geophys. Res.*, **127**(12), doi:10.1029/2022JF006917.
- Zhou, Z.-H., 2018. A brief introduction to weakly supervised learning, *Natl. Sci. Rev.*, **5**(1), 44–53.
- Zhu, M. & Gupta, S., 2017. *To prune, or not to prune: exploring the efficacy of pruning for model compression*, arXiv, doi:10.48550/arXiv.1710.01878.
- Zhu, W., Mousavi, S.M. & Beroza, G.C., 2019. Seismic signal denoising and decomposition using deep neural networks, *IEEE Trans. Geosci. Remote Sens.*, **57**(11), 9476–9488.

APPENDIX: MODEL ARCHITECTURE

Table A1 gives a summary of the U-Net model architecture (Ronneberger et al. 2015) used to implement DAS-N2N in this study. Prior to model training, model weights were initialized following Glorot & Bengio (2010). No batch normalization, dropout or other regularization techniques were used.

Table A1. Model architecture used to implement DAS-N2N in this study. Output shape given in rows \times columns \times feature maps. Param # is number of trainable parameters in model layer. All convolutions use padding mode ‘same’, and except for the final layer are followed by leaky ReLU activation function with $\alpha = 0.1$. Final layer has linear activation. Upsample 2×2 repeats data in each row and column.

Layer name (type)	Output shape	Param #	Function
input (InputLayer)	(128, 96, 1)	0	
conv00 (Conv2D)	(128, 96, 24)	240	Conv 3×3 then LeakyReLU
down10 (MaxPooling2D)	(64, 48, 24)	0	Max Pool 2×2
conv10 (Conv2D)	(64, 48, 24)	5208	Conv 3×3 then LeakyReLU
up01 (UpSampling2D)	(128, 96, 24)	0	Upsample 2×2
concat01 (Concatenate)	(128, 96, 48)	0	Concatenate with output of conv00
conv01a (Conv2D)	(128, 96, 48)	20 784	Conv 3×3 then LeakyReLU
conv01b (Conv2D)	(128, 96, 48)	20 784	Conv 3×3 then LeakyReLU
out01 (Conv2D)	(128, 96, 1)	49	Conv 1×1

Total trainable parameters: 47 065.