Deliverable 1.4: DAS Preprocessing Workflow

DigiMon

Digital monitoring of CO_2 storage projects

Prepared by

Antony Butcher (UoB)

Tom Hudson (UoO)

Alan Baird (UoB)

Rob Mellors (LLNL)

Anna Stork (Silixa)

DigiMon Deliverable D.1.4 Version 2,

September 2020

Revision

Version	Date	Change	Page
1.0	23.09.2020	First version	All
2.0	03.11.2020	Noise analysis and data conditioning updated.	

Document distribution

ACT Coordinator

- Research Council of Norway

ACT national funding agencies

- Forschungszentrum Jülich GmbH, Projektträger Jülich, (FZJ/PtJ), Germany.
- Geniki Grammatia Erevnas kai Technologias/The General Secretariat for Research and Technology (GSRT), Greece.
- Ministry of Economic Affairs and Climate/Rijksdienst voor Ondernemend Nederland (RVO), the Netherlands.
- The Research Council of Norway (RCN), Norway.
- Gassnova, Norway.
- Development and Executive Agency for Higher Education, Research, Development and Innovation Funding (UEFISCDI), Romania.
- Department for Business, Energy and Industrial Strategy (BEIS), UK.
- Department of Energy (DoE), USA.

DigiMon partners

- NORCE Norwegian Research Centre AS
- OCTIO Environmental Monitoring AS
- NTNU Norwegian University of Science and Technology
- University of Bristol
- University of Oxford
- CRES Centre for Renewable Energy Sources and Saving
- Helmholtz–Centre for Environmental Research
- Sedona Development SRL
- TNO Nederlandse Organisatie voor toegepast -natuurwetenschappelijk Onderzoek
- Geotomographie GmbH
- LLC Lawrence Livermore National Security
- SILIXA LTD
- EQUINOR ASA
- REPSOL –NORGE AS

Table of contents

1 Introduction

This report addresses deliverable D1.4 of the DigiMon project, which covers the preprocessing workflow for datasets acquired by Distributed Acoustic Systems (DAS). The workflow seeks to capture the key stages required to prepare the raw seismic data for the main processing stages and demonstrates their application using both synthetic and real-world data. A description of the synthetic datasets can be found in DigiMon deliverable D1.3 report, while details of the real-world datasets are included in DigiMon reports D1.1 and D1.2.

The objective of the DigiMon project is to develop an early-warning system for Carbon Capture and Storage (CCS) which utilises a broad range of sensor technologies including DAS. While the system is primarily focused on the CCS projects located in shallow offshore environment of the North Sea, it is also intended to be adaptable to onshore settings. Some of the key areas that the systems will monitor include the movement of the plume within the reservoir, well integrity and CO2 leakage into the overburden. A combination of both active and passive seismic methods will be deployed to track the movement of CO2, for example seismic reflection to image seismic velocity changes or microseismics to capture fault activation. Acquiring seismic surveys using DAS is highly novel and offers cost-effective approach which can significantly increase the spatial resolution of the survey data. However, DAS generally suffer from poorer signal-to-noise ratios (SNR) than conventional sensors (e.g. geophones), and exploiting the spatial resolution of DAS datasets to improve SNR is one of the main challenges of this workflow.

We broadly divided preprocessing into four separate stages (Figure 1), which converts raw data into a dataset ready for the main processing workflows. For the purposes of this project, raw data represents the seismic data provided by the instrument supplier, which in this case is Silixa's TDMS file format recorded by the iDAS™ system. The final output may either be waveforms files in a standard industry format, such as SEGY, or a numerical array passed to the main processing algorithm. The workflow has is designed to be relatively generic in order to make applicable to both active and passive seismic surveys. Both the workflow and report will continue to be updated as the project progresses, due to rapidly evolving DAS processing methods and to support the future developments of the main processing workflows.

Figure 1: Preprocessing workflow comprising of four separate stages preparing the raw seismic data for the main processing stages.

2 Definition of Metadata and Geometry

2.1 Purpose

Establish survey parameters and geometry and define header information

2.2 Details

DAS data can come in a variety of different formats including some standard seismology formats like SEG-Y, or mini-SEED, or in less well known or proprietary formats like Silixa's TDMS. If a proprietary format is used the vendor should supply a reader. For example, Silixa provide a TDMS reader for both Python and Matlab on their website. From this the data can be saved in a different format if required.

Once the data is read in, it is important to establish if all the relevant metadata required for processing the data correctly is available. Much of this information is likely available in the data headers (e.g. time stamp, sampling rate etc.), however, depending on the data format used, some metadata specific to DAS may need to be manually edited, such as gauge length and channel spacing. You should also make sure that you know what units your data are given in. Most DAS systems provide raw units that are proportional to either strain or strain-rate but require a scalar factor to be applied. For example, the raw data for a Silixa iDAS system provides units of optical phase / time sample but can be converted to nanostrain / second (nm/m/s) by multiplying the raw signal by 116*sampling frequency*gauge length.

Next you need to establish the geometry of the array and work out a methodology to map channels to geographic coordinates. Ideally the geometry of the array was correctly surveyed to allow the distance along the cable to be mapped to a coordinate. From there, knowledge of the effective channel spacing will allow you to distribute the channels along the array geometry. Note that channel spacing is the spacing along the fibre rather than the spacing along the cable. These should be approximately equivalent for straight fibre cables, but for helically wound cables (HWCs) there is more fibre per unit length, meaning the effective channel spacing along the cable axis will be smaller. The effective channel spacing along a HWC is (channel_spacing)*sinϕ, where ϕ is the wrapping angle of the fibre.

Finally, you need to determine the channels of interest and calibrate the channel mapping. Often when you read data the first channel in the file does not correspond to the first channel in your array. This is because there is excess fibre between the interrogator and the start of the array that is not coupled with the ground. To establish your channel range of interest it is necessary to identify key locations where you can link a channel to a fixed point. These can be established in the field using tap tests or calibration shots to determine channel locations. For borehole deployments co-located geophones or depth of the borehole can be useful.

Figure 2 shows an example of establishing the channels of interest using data from the FORGE DAS array. This is a vertical downhole array and tap tests were used establish the distance between the well head and the base of the well as 996.51 (corresponding to 977 channels with spacing of 1.02m on a linear fibre). By plotting the data with the interval between the key locations marked, we can use the noise characteristics of the data to estimate what channel corresponds to the well head. Figure 3 shows an example from the

surface deployed Antarctic dataset where sharp changes in the apparent velocity of a recorded icequake allows us to confidently locate the channels at the corners of a triangular shaped array.

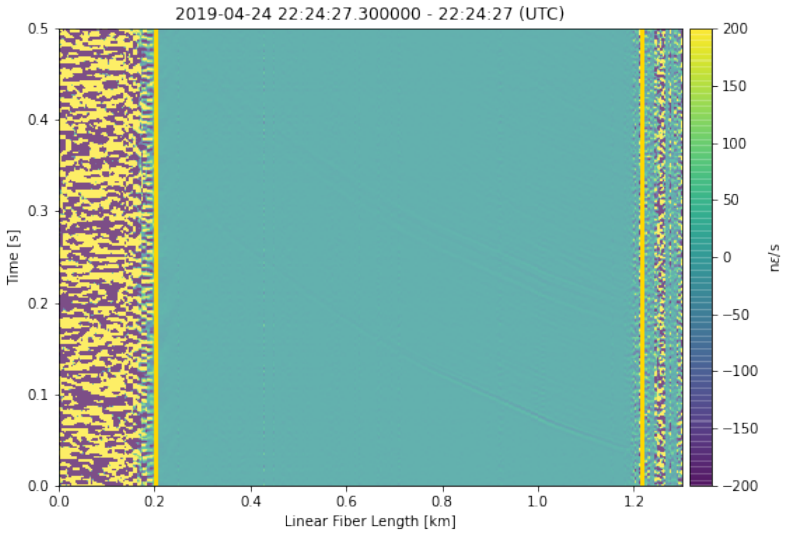

Figure 2: Example establishing the geometry and channels of interest from the FORGE dataset (from the 2020 IRIS DAS Virtual Workshop and Tutorial).

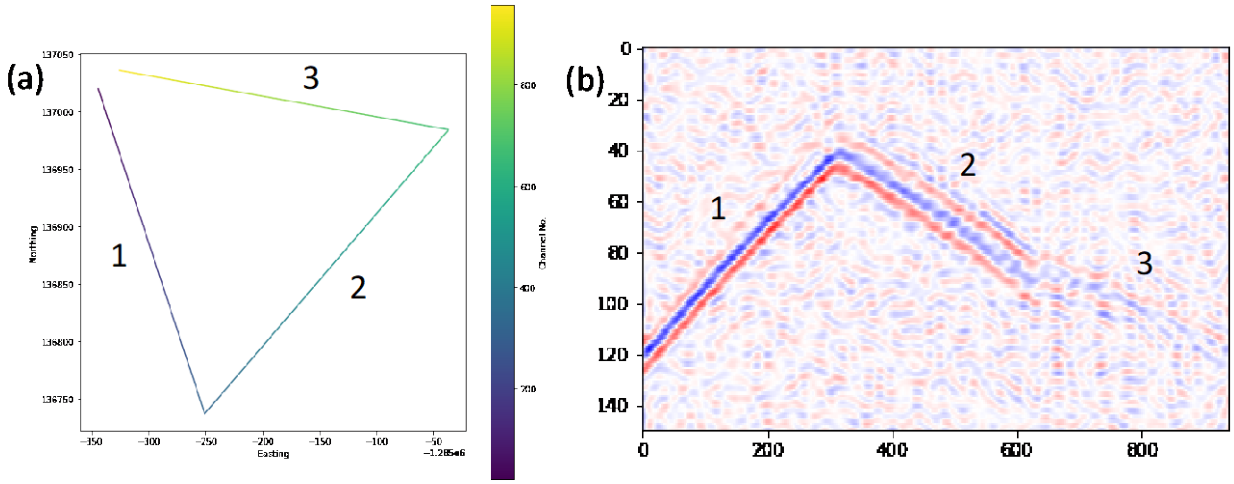

Figure 3: Example establishing the geometry of a triangular shaped array from the Antarctic dataset. (a) Map showing the array geometry, with the three sides labelled. (b) An example of a recorded S-wave arrival from an icequake. The sharp change in apparent velocity observed in the data allow us to locate the channels at the corners of the array.

3 Noise analysis

3.1 Purpose

Analysis of noise characteristic to determine filtering parameters required to increase the signal to noise ratio.

3.2 Nature of Seismic Noise

3.2.1 Instrument Noise

The types of noise observed in DAS data can be classified into three different types: common mode noise, random infinite-velocity spikes and coherent fading noise (Lindsey el al., 2020).

Common mode noise is characterised by an infinite velocity signal caused by vibrations of the interrogator unit, which projects an instantaneous signal over the entire fibre (Figure 6). This can be removed through applying a median filter in time, though this may be problematic if body waves arrive at a normal incident. Another approach is to create a pilot trace from a quiet, signal free section of the data and subtract this from the entire recording.

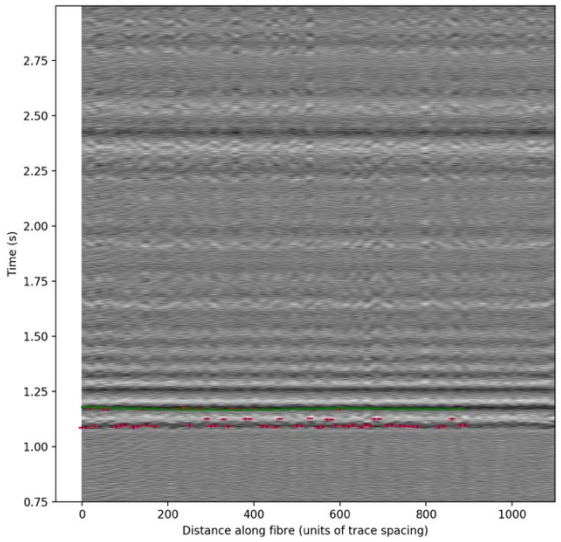

Figure 6: An example of common-mode noise from the Antarctic dataset, caused by the vibration of the interrogator unit.

Random infinite-velocity spikes are comparable to common-mode noise and result of laser frequency drift or laser noise (Zhirnov et al., 2016). This can create temporal jumps across the entire fibre length, and this can be removed with a median filter or the creation of a pilot trace.

Coherent fading noise is a reduced amplitude pattern, which is quasi-random in space but time invariant. This can result due to photonic fading or sensor coupling and distinguishing which is the cause is difficult. Photonic fading is the result of destructive interference of backscattered lightwaves (Shimizu et al., 1992) while sensor coupling relates to the nature of the fibre installation. Both can be removed through a median filter or a highpass filter.

3.2.2 Natural and Anthropogenic Noise

Seismic noise in the shallow marine environment can be generated by a range of both natural and anthropogenic sources, which span broad frequency and amplitude ranges. Anthropogenic sources include shipping traffic, wind turbines, marine construction and unrelated seismic surveys. Natural sources include marine life, tidal wave interaction and ocean current driven harmonic tremors. Most of the sound energy

emitted in the ocean originates from anthropogenic sources, which is dominated by commercial shipping (Figure 5) followed by airguns and explosions (Sertlek et al., 2019). The North Sea is one of the most heavily trafficked regions is the world, with the majority of the noise clustered along the main shipping channels and offshore infrastructure (Farcas et al., 2020). These anthropogenic sources are responsible of the largest amount of acoustic energy in the North Sea, contributing 100 times more energy than wind (Sertlek et al., 2019). These noise sources are likely to the most impact on CCS seismic monitoring arrays, as depleted reservoirs located within these 'noisy' areas of the North Sea are a key storage media for CO_2.

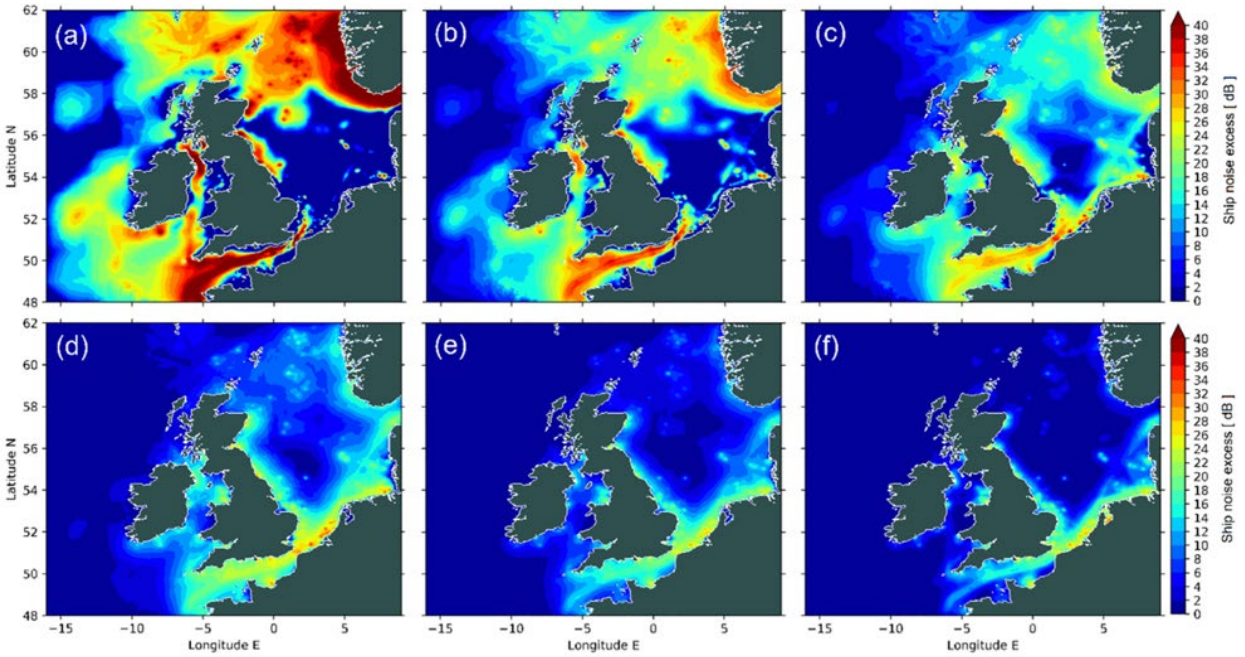

Figure 5: Annual median ship noise excess (ship noise levels above wind) modelled at various frequencies: (a) 63 Hz (b) 125 Hz (c) 250 Hz (d) 500 Hz (e) 1 kHz (f) 4 kHz. From Farcas et al. (2020)

Sound energy propagates within the ocean as compressional waves, travelling with very little attenuation at a velocity of ~1.5km/s. The apparent velocity and incident angle with depend on the location of the noise source relative to the seismic network. When these signals interact with the seafloor they can be transmitted directly into seabed where they travel as both body waves and dispersive waves along the fluid-solid interface as. Two types of dispersive waves can occur: Scholte waves, which travel mainly in the liquid, and Leaky Rayleigh waves which are predominately located within the solid seafloor. While the velocity of both Scholte and Leaky Rayleigh waves is dependent on frequency, in general they will propagate at speeds slightly lower than the S-wave velocity (Johansen et al., 2019). Filtering methods that therefore exploit both the frequency and the velocity of these wavefield generated by these noise sources improve the SNR.

3.3 Seismic Noise Assessments

As described in the Section 3.2, seismic noise can originate from range of different sources and affect both broad and narrow frequency ranges of the signal. A useful approach to assess the impact of this noise is through the use of power spectral density (PSD) plots, which provide an opportunity to examining the noise characteristics of seismic datasets. These can be used assess the quality of the dataset and compare the performance of individual traces within the array. Results are displayed as frequency dependent power distributions. This can be demonstrated through using the FORGE dataset, where PSD plots from individual channels can be compared as individual (Figure 6) or colour contoured (Figure 7) plots.

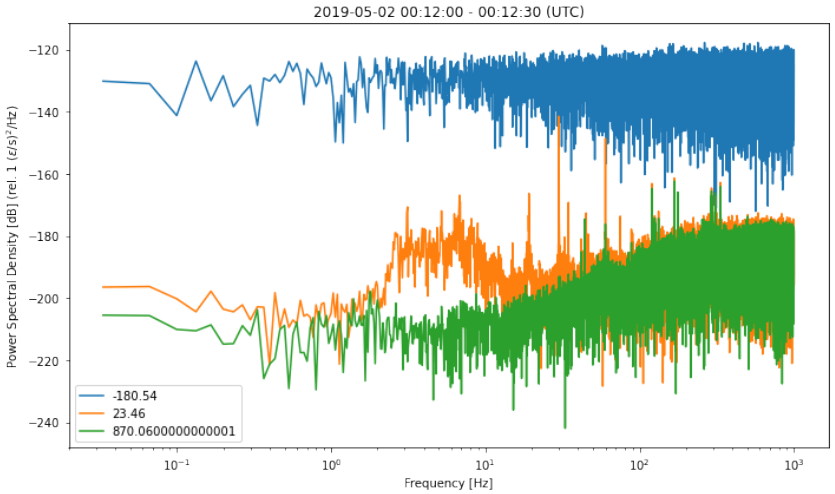

Figure 6: PSD plot of three individual channels from the FORGE dataset. Noisy channels are easily identified through a static shift (e.g. blue trace). (Data from the 2020 IRIS DAS Virtual Workshop and Tutorial).

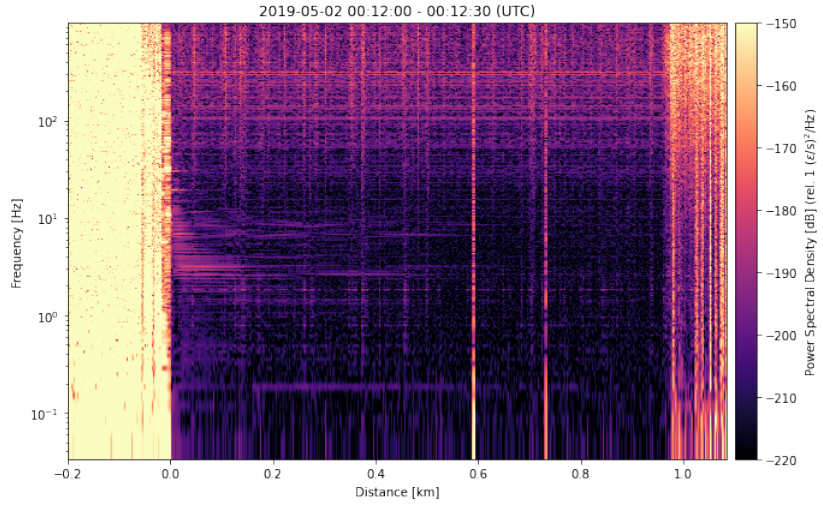

Figure 7: Presenting data using colour contour plots provides an simple overview of the data quality, which allows noise channels and frequency bands to be identified. (Data from the 2020 IRIS DAS Virtual Workshop and Tutorial).

4 Data conditioning

4.1 Purpose

After determining the nature of the noise and defining the filtering parameters, this stage seeks to increase both the signal-to-noise ratio (SNR) and reduce the data size ahead of the main processing stages.

4.2 Preprocessing filters

4.2.1 1D Bandpass Filter

High frequency noise can often dominate DAS datasets and often static amplitude shifts between channel can be observed. Both can be removed by applying an appropriate bandpass filter to individual channels (Figure 7). The parameters used for this filter will be dependent of the characteristic of the target signal and should be based on the noise analysis. Applying this filter at this stage allows for a clearer identification of arrivals during the f-k analysis.

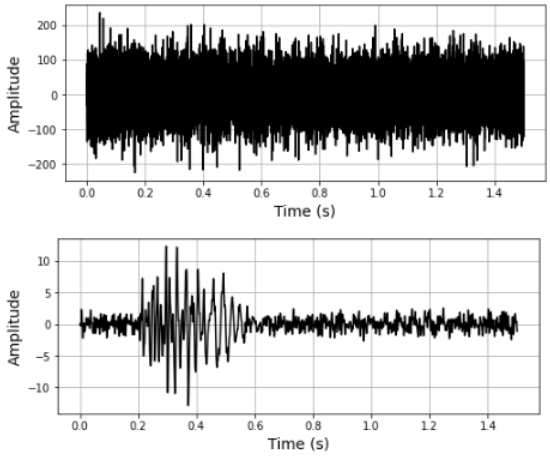

Figure 7: DAS signal before (top) and after (bottom) applying a bandpass filter to an individual channel.

4.2.2 Wiener Filtering

A Wiener filter is a type of de-blurring filter routine applied in image processing. To apply a Wiener filter, one assumes that the signal and noise processes are approximately linearly time-invariant. This assumption is appropriate for a Wiener filter applied to a 2D DAS time-distance along fibre image. An example for the effectiveness of a Wiener filter is shown by it applied to an icequake from Rutford Ice Stream, Antarctica, in Figure 9. In this case, the Wiener filter is effective at removing noise from the entire image, resulting in significantly lower noise levels of both the icequake phase arrival and the preceding noise field.

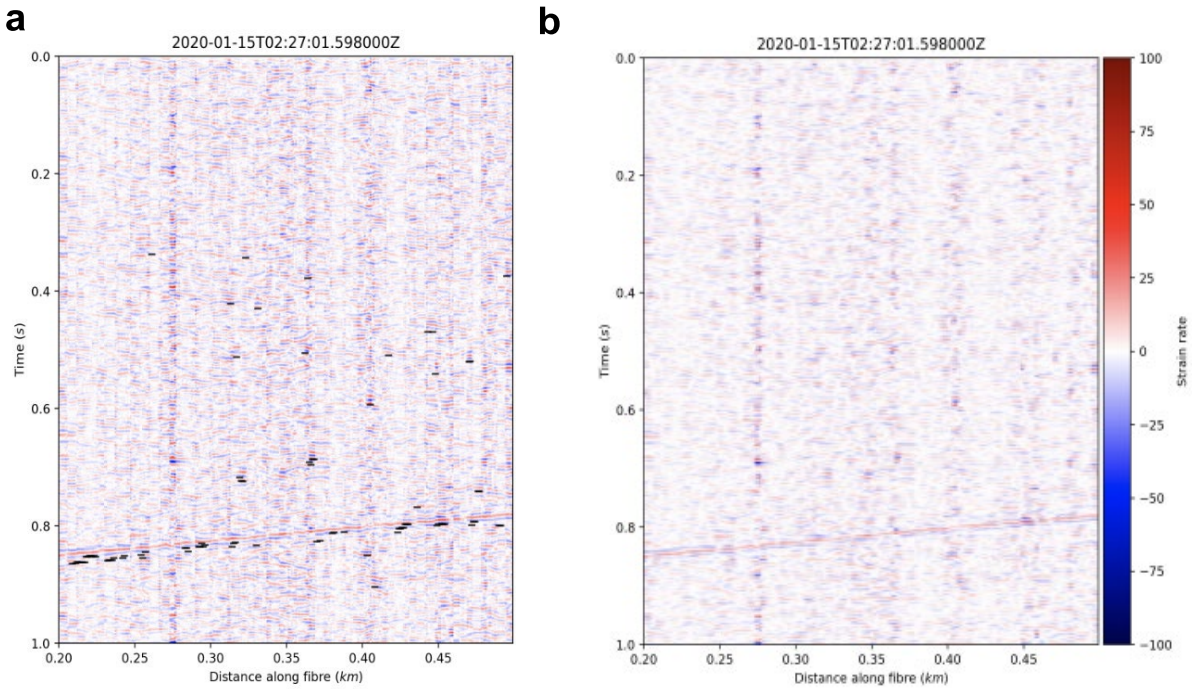

Figure 9 – Data with and without a Wiener filter applied. a) without a Wiener filter applied. b) With a Wiener filter applied. Note that the data in both plots have had a bandpass filter of 5-120 Hz applied.

4.2.3 F-K Filter

DAS datasets benefit from a very high spatial resolution due to the close and regular channel spacing along the fibre, which makes signal analysis and filtering in the wavenumber/ frequency domain an effective method of removing unwanted seismic noise. To perform the f-k, the time/space domain signal is transformed to the frequency/wavenumber domain using a 2D fast Fourier transform. Noise with different frequencies and wavenumbers than the signal of interest can then be masked before the signal is transformed back into the time/space domain. This provides a useful method or removing coherent seimic noise, which may result from instrument noise or local sources. Figure 10 provides an example from the Antarctica datasets of several arrivals with different velocities.

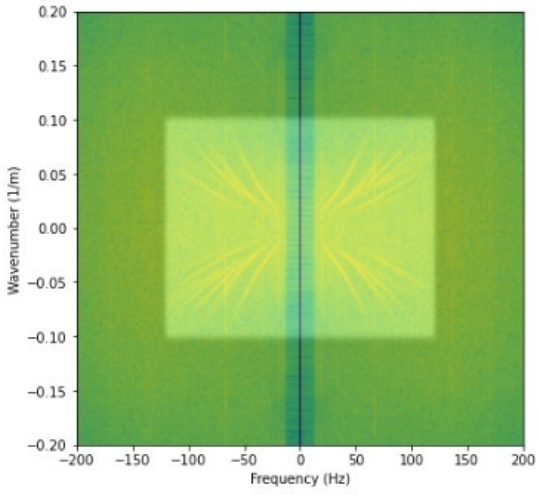

Figure 10: FK filter applied to active survey data from the Antarctic dataset. Different velocity arrivals are clearly identifiable in the plot which can be enhanced through removing signals outside of their frequency and wavenumber range (shaded area).

4.2.4 Notch filter

Surface waves from generators and other near-fibre sources can be detrimental for analysis of DAS data. One example of this is from an Antarctic dataset from Rutford Ice Stream, where a generator was situated at one end of the fibre (0m along fibre, Figure 11). This produced a surface wave source with a fundamental mode at 33Hz. One cannot necessarily remove such noise with a bandpass or other filter, since the surface waves are coherent and might be within the frequency range of interest. Figure 10 provides an example for an icequake with generator surface wave noise, with and without notch filters applied at 33Hz and 66Hz. These results clearly show the effectiveness of applying a notch filter.

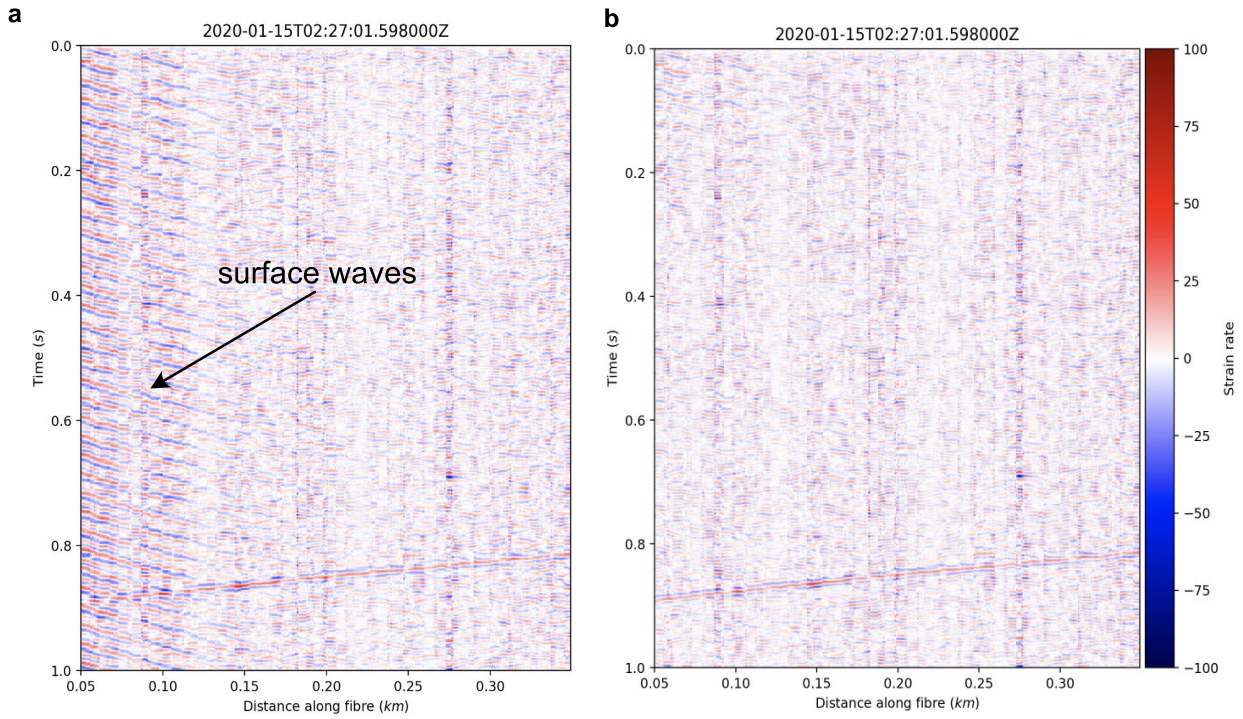

Figure 11 – Example of a notch filter applied to data from Rutford Ice Stream. a) Without a notch filter applied. b) With notch filters applied at 33Hz and 66Hz, the first and second order modes of the surface wave noise, respectively.

4.2.5 Data Resampling

The data size of DAS recordings is typically very large due to the high temporal and spatial sampling of the system, and often downsampling is essential to increase the speed of future processing stages. The typical sampling rates of the Silixa iDAS system is 1000Hz. For the PoroTomo dataset this was resampled to 100Hz for the purposes of earthquake detection (Li & Zang, 2018) while Ajo-Franklin et al. (2019) used a 125Hz sampling rate for ambient noise analysis. The choice of sampling rate will depend on the purpose of the dataset and the frequency content of the recorded signal.

4.2.6 Quantifying the performance of filters: SNR and processing time

Quantifying the performance of the various pre-processing filters is important for deciding what filter/s to use for a particular dataset. One method of quantifying the performance of a filter to remove noise while preserving the signal is the Signal-to-Noise Ratio (SNR). We provide an estimate of SNR using the method of Stork et al (2020). We take a window of 100 samples centred around a phase arrival, and a noise window of the same length, before the signal window, and define the SNR as the ratio of the root mean square (rms) amplitudes of the signal window divided by the noise window. This method requires phase arrival picks. We use a simple short-time-average to long-time-average (STA/LTA) detection algorithm with a Median Average Deviation (MAD) multiplier of 5 to obtain estimates for these phase picks. We then use an icequake from the Rutford Ice Stream dataset to assess the performance of the various filters.

The S phase picks used for the filter SNR analysis are shown in Figure 12. We apply a number of the aforementioned filters to the data in Figure 12 to find the SNR improvement, as well as the time taken to process the filter. Note that the timings are only for relative comparison, since actual processing times are dependent on quantity of data processed and the processing power of the machine they are processed on. The results are summarised in Figure 13.

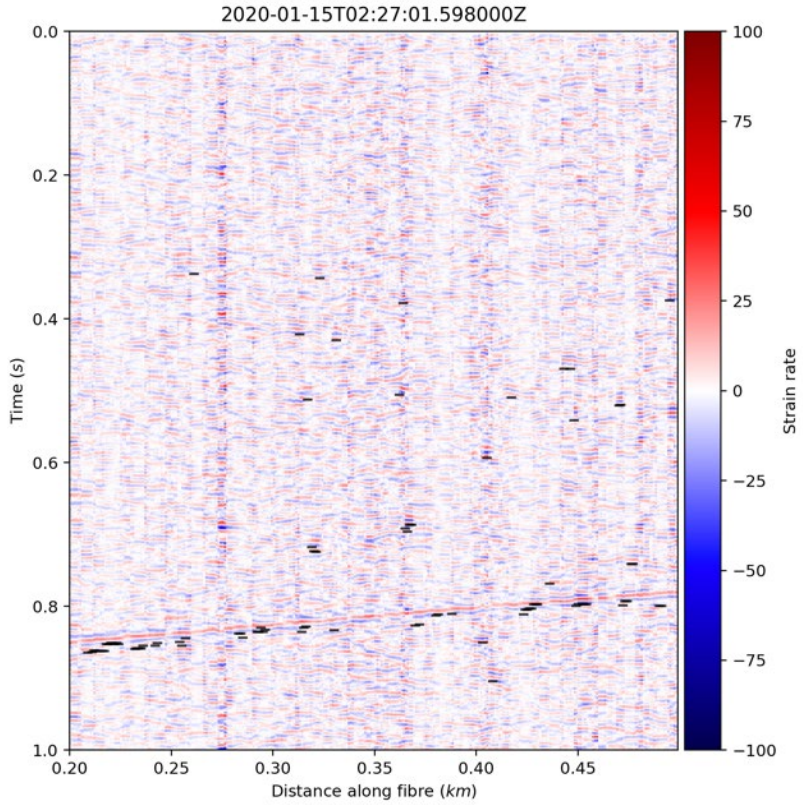

Figure 12 – Plot of time vs distance along the fibre for an icequake arrival at Rutford Ice Stream, Antarctica. The black lines indicate the STA/LTA detection phase picks. The amplitudes represent strain rate, but are uncalibrated in units of counts. Data used here for the phase detection is band-pass filtered.

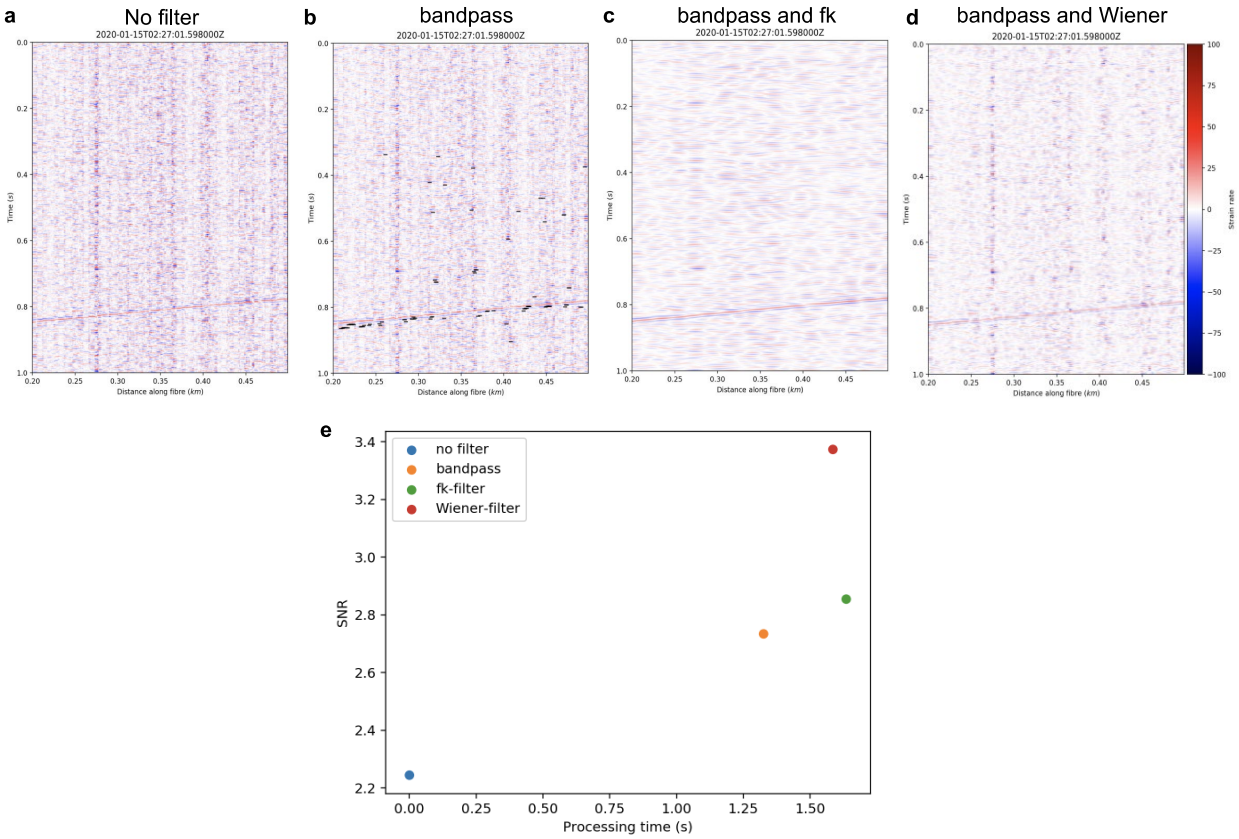

Figure 13 – Quantification of filter performance. a) to d) Data for the icequake in Figure 12, with no filter, a bandpass filter, a bandpass and fk filter, and a bandpass and Wiener filter, respectively. e) The SNR for a) to d) compared to the processing time for each filter. Note that the fk and Wiener filter processing times include the time taken to apply the bandpass filter.

5 Data Conversion

5.1 Purpose

Although DAS systems natively record either strain or strain-rate there may be situations where you would want to convert this to velocity or a geophone equivalent. For example, to take advantage of pre-existing processing algorithms that were developed for use on geophone data, or to make direct comparisons with a co-located seismic station. Additionally, if the instrument response of a co-located seismic station is known, this provides an approach to estimate an empirical transfer function. There is currently no standard method to perform this conversion.

5.2 Details

One approach to converting DAS data to velocity is by recognizing that strain-rate is just the spatial gradient of velocity, thus it can be estimated simply by integrating strain-rate along the cable. Unfortunately, this is difficult to achieve in practice due to channel specific noise and fading, which introduces errors into the spatial integration. An alternative approach is to take advantage of the array nature of DAS data to use the apparent velocity of coherent arrivals as a means to transform the spatial integral into a time integral (to convert from strain-rate to strain). For a cable oriented in the x direction the longitudinal strain ε_{xx} is related to particle velocity v_x and apparent slowness of a coherent arrival p_x through the following relation:

$$\varepsilon_{xx} = \frac{\partial u_x}{\partial x} = \frac{\partial u_x}{\partial t}\frac{\partial t}{\partial x} = v_x p_x.$$

Thus strain can be converted to particle velocity simply by scaling the arrival amplitude by the apparent velocity ($c_x = 1/p_x$). This can be done in the space-time domain by windowing coherent arrivals and scaling the amplitudes, however, this approach may be difficult for complicated recordings which have overlapping arrivals with different apparent velocities. A simpler approach is to first transform the data into the frequency wavenumber domain and then recognize that the apparent slowness is simply the ratio of wavenumber to frequency ($p_x = k/\omega$). Thus the conversion can be done in the FK domain using (e.g. Wang et al 2018, Lindsey et al, 2020),

$$V_x(\omega) = (\omega/k_x)E_{xx}(k_x, \omega).$$

This can then be transformed back to the space-time domain by applying the inverse FK transform.

Figure 1 shows an example of applying this technique to synthetic strain-rate DAS data. The technique provides good fit to the velocity data, but does introduce some artefacts, including low apparent slowness banding along the array. This is likely due to the instability of the scaling relationship for wavenumbers approaching zero. Lindsey et al. (2020) suggest an approach to minimize this effect by introducing a water-level parameter into the equation to avoid division by small numbers.

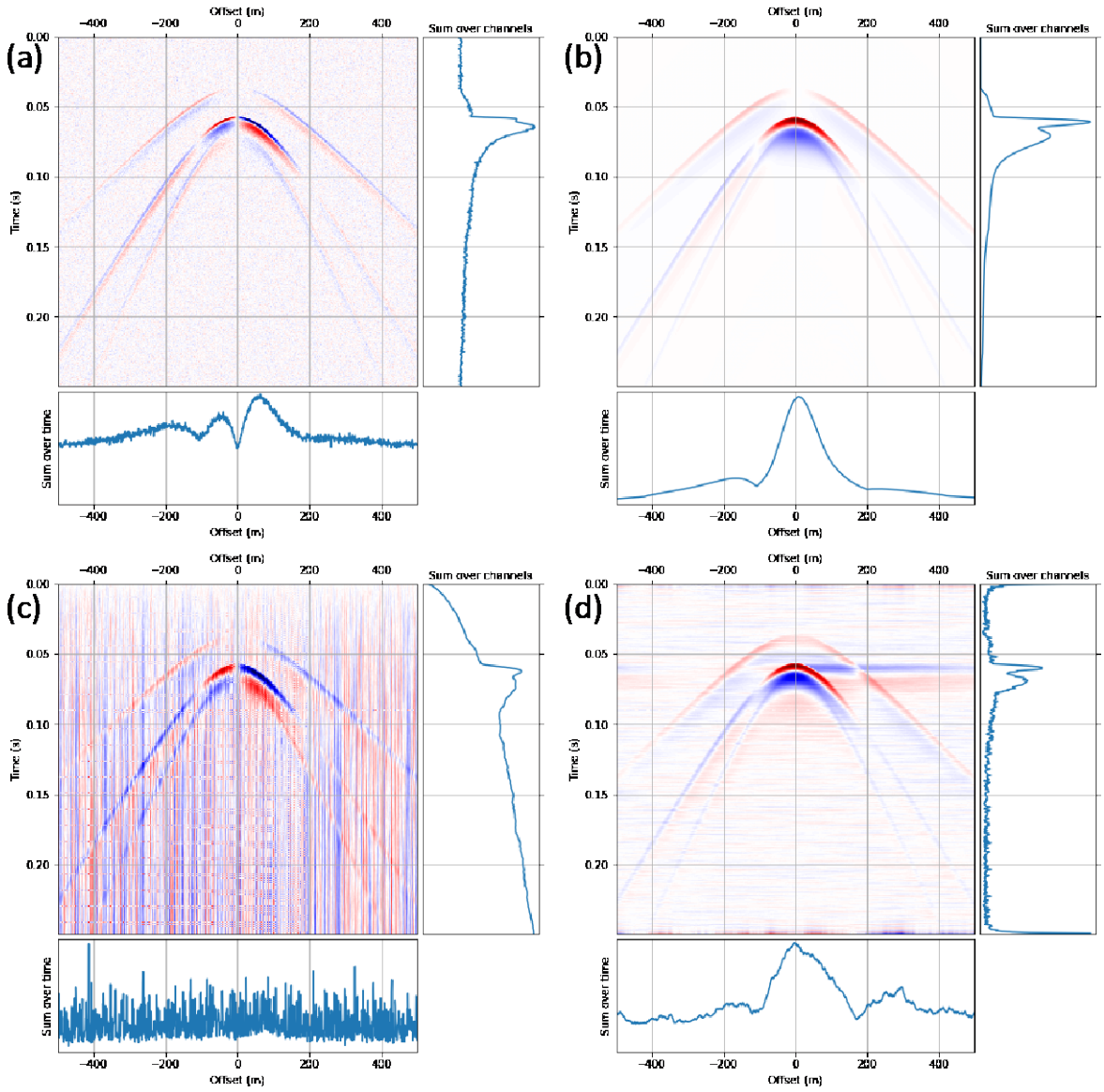

Figure 14 (a) Synthetic DAS strain-rate data with 10 m gauge length and added gaussian noise. (b) Noise free synthetic particle velocity data. (c) Strain data found by time integration of DAS data. (d) Particle velocity data estimated by scaling strain data in FK domain.

6 References

Farcas, A., Powell, C. F., Brookes, K. L., & Merchant, N. D. (2020). Validated shipping noise maps of the Northeast Atlantic. *Science of the Total Environment*, *735*, 139509. https://doi.org/10.1016/j.scitotenv.2020.139509

IRIS DAS workshop, 2020, https://www.iris.edu/hq/event/2020_DAS_Workshop

Johansen, T. A., Ruud, B. O., & Hope, G. (2019). Seismic on floating ice on shallow water: Observations and modeling of guided wave modes. *Geophysics*, *84*(2), P1–P13. https://doi.org/10.1190/geo2018-0211.1

Lellouch, A., N. J. Lindsey, W. L. Ellsworth, B. L. Biondi, 2020. Comparison between Distributed Acoustic Sensing and Geophones: Downhole Microseismic Monitoring of the FORGE Geothermal Experiment. Seismological Research Letters doi: https://doi.org/10.1785/0220200149

Lindsey, N. J., H. Rademacher, and J. Ajo-Franklin, 2020, On the Broadband Instrument Response of Fiber-Optic DAS Arrays, Jour. Geophys. Res., https://doi.org/10.1029/2019JB018145

SEAFOM Measuring Sensor Performance Document – 02 (SEAFOM MSP-02): DAS Parameter Definitions and Tests available at https://seafom.com/published-documents/

Sertlek, H. Ö., Slabbekoorn, H., ten Cate, C., & Ainslie, M. A. (2019). Source specific sound mapping: Spatial, temporal and spectral distribution of sound in the Dutch North Sea. *Environmental Pollution*, *247*, 1143–1157. https://doi.org/10.1016/j.envpol.2019.01.119

Stork, A. L., Baird, A. F., Horne, S. A., Naldrett, G., Lapins, S., Kendall, J.-M., Wookey, J., Verdon, J. P., Clarke, A., & Williams, A. 2020, Application of Machine Learning To Microseismic Event Detection in Distributed Acoustic Sensing (Das) Data. Geophysics, 1–53. https://doi.org/10.1190/geo2019-0774.1

Utah FORGE: Phase 2C Topical Report, available at https://gdr.openei.org/submissions/1187 Wang, H. F., Zeng, X., Miller, D. E., Fratta, D., Feigl, K. L., Thurber, C. H., & Mellors, R. J. (2018). Ground motion response to an Ml 4.3 earthquake using co-located distributed acoustic sensing and seismometer arrays. Geophysical Journal International, 213(3), 2020–2036.

Wenz, G. M. (1962). Acoustic Ambient Noise in the Ocean: Spectra and Sources. *The Journal of the Acoustical Society of America*, *34*(12), 1936–1956. https://doi.org/10.1121/1.1909155

Zhirnov, A., Fedorov, A., Stepanov, K., Nesterov, E., Karasik, V., Svelto, C., & Pnev, A. (2016). Effects of laser frequency drift in phase-sensitive optical time-domain reflectometry fiber sensors. arXiv preprint arXiv:1604.08854.