SANDIA REPORT
SAND2023-08140
August 2023

An Analysis of Published DAS Studies for Application to SPE Phase III

Robert W. Porritt and A. Christian Stanciu

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

Distributed Acoustic Sensing (DAS) is an emerging technology capable of recording the acoustic wavefield at unprecedented spatial resolution. However, this new tool requires significant refinements before it becomes operational for explosion monitoring objectives. Recent studies have shown significant development of array processing with DAS data. In this report we explore three such array processing methods including DAS strain-rate data versus geophone measured ground motion, beamforming for event parameters, and machine learning based denoising. Prior to applying these algorithms to the Source Physics Experiment Phase II and Phase III data, we validate these methods through replication analysis.

ACKNOWLEDGEMENTS

This Source Physics Experiment (SPE) research was funded by the National Nuclear Security Administration, Defense Nuclear Nonproliferation Research and Development (NNSA DNN R&D). The authors acknowledge important interdisciplinary collaboration with scientists and engineers from LANL, LLNL, NNSS, and SNL.

We thank Martjin van den Ende and Herb Wang for providing open access to software and datasets used in this study.

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Security Administration under contract DE-NA-0003525.

CONTENTS

Abstract....3
Acknowledgements....4
Executive Summary....7
Acronyms and Terms....8
1. Introduction....9
2. Data and Methods....10
  2.1. Data....10
  2.2. Methods....12
  2.2.1 Conversion between strain-rate and ground motion....12
  2.2.2 Beamforming....13
  2.2.3 Blind Deep Learning Denoising....13
3 Results....14
  3.1 Conversion between strain-rate and ground motion....14
    3.1.1 Euler's Method....14
    3.1.2 F-k integration....19
  3.2 Beamforming....23
  3.3 Blind Deep Learning Denoising....24
4 Strengths and weaknesses....30
  4.1 Conversion to Ground Velocity....30
  4.2 Array Beamforming....30
  4.3 Array Denoising....30
5 Conclusion....32
References....33
Distribution....34

LIST OF FIGURES

Figure 2-1: Location of the data arrays and sources....11
Figure 3-1: Example of Euler's Method modified from van den Ende and Ampuero (2021)....15
Figure 3-2: Determination of σ fit parameters for Euler's Method with DAG-2 data....18
Figure 3-3: Example waveforms after applying Euler's Method to the DAG-2 data....19
Figure 3-4: Frequency-Wavenumber (f-k) spectra of Geophones and DAS....21
Figure 3-5: Ground velocity for DAS and geophones for PoroTomo....22
Figure 3-6: DAG-2 record sections comparing f-k integration scaling of DAS data....23
Figure 3-7: Application of beamforming via the MUSIC algorithm using PoroTomo data and DAG-2 data....24
Figure 3-8: Denoising process of van den Ende et al. (2021)....25
Figure 3-9: Process to generate realistic Earth noise....26
Figure 3-10: Result of adding and removing synthetic noise to DAG-2 and DAG-3 data....27
Figure 3-11: Variance reduction and correlation coefficient for the DAG-2 and DAG-3 denoising test....28
Figure 3-12: Histograms of the denoising fit metrics....29

This page left blank

EXECUTIVE SUMMARY

Distributed Acoustic Sensing (DAS) is an emerging technology capable of recording the seismic wavefield at unprecedented spatial resolution. Recent studies have shown capabilities at sub-meter inter-station spacing for array apertures on the order of 10s of km. However, this new tool requires significant refinements before it becomes operational for explosion monitoring objectives. First, the data is in units of strain-rate whereas typical seismological analyses are defined for ground motion. Second, the noise levels, installation effects, and non-ideal wave incidence effect all create challenges to quantifying the amplitude response. Third, the data tend to have higher noise levels relative to conventional seismic monitoring instruments. Recent studies have shown significant development of array processing techniques using DAS data. In this report we explore three such array processing methods including a direct comparison between DAS strain-rate data and geophone measured ground motion, beamforming for event parameters, and machine learning based denoising. Prior to applying these algorithms to the Source Physics Experiment Phase II and Phase III data, we validate these methods through replication of published analyses. From this we can better understand the strengths and weaknesses of each approach. We conclude two major points. First, the proper approach to converting DAS strain-rate data to ground motion is situationally dependent. Second, the array nature of the recording method lends itself well to beamforming and array-based denoising methods.

ACRONYMS AND TERMS

Acronym/Term	Definition
DAS	Distributed Acoustic Sensing
SPE	Source Physics Experiment
RV/DC	Rock Valley Direct Comparison
NV	Nevada
DAG	Dry Alluvium Geology
DAG-2	Second chemical explosion in the DAG sequence
DAG-3	Third chemical explosion in the DAG sequence
TNT	Trinitrotoluene
2D	Two dimensional (spatial)
2.5D	Two and a half dimensional (i.e. 2 spatial dimensions and time)
f-k	Frequency-wavenumber
MUSIC	Multiple Signal Classification
VR	Variance Reduction
$\dot{u}$	Ground velocity
u	Ground displacement
x	Location (bold for vector notation)
L	Gauge Length
$\dot{\varepsilon}$	Strain-rate
ε	Strain
$\boldsymbol{k}$	Wavenumber vector
$\boldsymbol{\omega}$	Angular frequency vector
Δx	Spatial step
σ	Correction Factor
p, g, $\overline{p}$	Arbitrary series of values with bar indicating mean over series.
km	Kilometer
m	meter
Hz	Hertz
L1	Absolute value norm
L2	Root-mean squared norm
FFT	Fast Fourier Transform
UTM	Universal Transverse Mercator
CA	California
SNR	Signal-to-Noise Ratio

1. INTRODUCTION

Emerging distributed sensing capabilities with fiber optic cables are producing observables with sub-meter spatial resolution over array lengths of 10s of kilometers. This report focuses on Distributed Acoustic Sensing (DAS) and its capability to measure deformation in a fiber optic cable that can be related to ground motion, which is traditionally measured by geophones and broadband seismometers. This technology relies on Rayleigh scattering which is a type of elastic (no energy loss) optical scattering, sensitive to changes in strain within the fiber. Complementary technologies, such as Distributed Temperature Sensing and Distributed Strain Sensing use Raman and Brillouin (i.e., inelastic) scattering, which are sensitive to changes in temperature and strain respectively. Monitoring networks for the Source Physics Experiment (SPE) Phase III, Rock Valley Direct Comparison (RV/DC) will include both downhole and surface deployed fiber optic cables for distributed acoustic and temperature sensing. Our current goal is to develop new and improved computational capabilities for this technology to rapidly characterize the planned chemical explosions and background seismicity.

The strain-rate timeseries measured by DAS can be related to ground motion observations recorded by traditional seismic monitoring instruments. Quantifying the response of a DAS sensing unit (e.g., station) presents challenges, as the amplitude response is dependent on a large set of parameters; however, the phase response has been shown to be consistent with ground motion instruments (Lindsey et al., 2020). This allows a range of phase dependent seismological analyses to be performed on DAS data, such as arrival time determination, with little to no additional calibration (e.g., Porritt et al., 2022). Constraining the amplitude response and the broadside incidence effect (i.e., how the amplitude response decreases by a function of cosine squared of the angle between the cable and incident plane wave [e.g., Martin et al., 2021]) are additional challenges towards quantitatively determining the amplitude of ground motion, necessary for estimating earthquake magnitude or explosion yield.

The main goal of this report is to document progress of applying array analysis methods to DAS data. The methods tested here are derived from published literature and are evaluated using publicly available datasets. The replication analysis allows us to verify algorithm functionality and to investigate the strengths and weaknesses of these methodologies. Algorithm selection was driven by data and software availability in the published literature, and suitability for RV/DC monitoring and source characterization goals.

2. DATA AND METHODS

The seismic data processing methods discussed here are separated into three main tasks. First, we apply two algorithms to convert between ground motion (displacement, velocity, and acceleration) and strain-rate and vice-versa. The first algorithm, Euler's method, uses a finite differences method to relate the strain-rate and ground motion for the case of a passing plane wave. As this method operates by solving a partial differential equation, it requires data for a boundary condition and then the appropriate 1D gradients can be applied in the time domain. The other method uses integration in the frequency-wavenumber domain after scaling by the phase velocity. Second, we apply beamforming to an array of receivers to estimate the slowness and the back-azimuth of incoming waves (e.g., Rost and Thomas, 2002). Third, we apply an array based denoising algorithm based on "blind" machine learning to identify inconsistent DAS stations. In the following, we present a brief summary of the data and methodologies with a reference to the published literature for additional details.

2.1. Data

An important consideration when working with DAS data is the gauge length. This value is the length of a laser pulse in the cable as supplied by the interrogator unit (e.g., Dean et al., 2017; Lindsey et al., 2020). This length is then subset to infer phase shifts at individual DAS stations. Effectively, this results in a data stream that has been subject to a low-pass spatial filter equal to the gauge length. For the cases here, the gauge length in each case is a small fraction of the wavelengths of interest and therefore no adjustment was necessary.

The first replication study uses data from the PoroTomo experiment at Brady Hot Springs (Feigl & PoroTomo Team, 2017). This geothermal field was densely instrumented both at the surface and in observatory boreholes; it includes 238 Fairfield Nodal ZLand three component seismometers (for generality, this type of instrument is termed "geophone" throughout) and an irregular "zig-zag" geometry of fiber optic cable with a gauge length of 10 m. Waveform data for both the geophones and the DAS were provided by Wang et al. (2018) and van den Ende and Ampuero (2021) via public repositories. Datasets were pre-cut by the authors to focus on a regional M4.1 earthquake which occurred near Hawthorne, NV. See Figure 2-1 and Figure 2-1b for a regional scale map and the location of the geophone array. We applied the algorithms in Wang et al. (2018) and van den Ende and Ampuero (2021) to reproduce the results comparing DAS measured strain-rate and geophone measured ground velocity.

The second replication study uses data recorded on dark fibers (fiber optic cables deployed for telecommunications, but at times unused) deployed offshore southwest Greece (Figure 2-1c; van den Ende et al., 2021). One cable is operated by the Hellenic Centre of Marine Research and the other is part of the Neutrino Extended Submarine Telescope with Oceanographic Research project. The data was recorded with a gauge length of 19.2 m and made available at 50 Hz sampling rate. The data window is 41 seconds long and it is centered around the first arrival for each of the 21 earthquakes they tested the denoising method on. For this replication study, we used the dark fiber data as per the original publication to confirm algorithm functionality, and we refer the reader to the original publication for more information on the effectiveness of the denoising algorithm with passive source data.

After confirming successful application of the algorithms to the above publicly available datasets, we apply them to test datasets from the SPE Phase II, Dry Alluvium Geology (DAG) experiment (Snelson et al., 2013) The waveform data for chemical explosions DAG-2 and DAG-3 were recorded on ~2 km of fiber optic cable with a Silixa iDAS interrogator with a gauge length of 2 m and ~500 nodal geophones, of which 30 are nearly co-located with the DAS fiber optic cable. The cable was deployed on the surface with the nearest to the source end deployed at 26 m offset to the south-southeast and the furthest end of the line at 2248 m. The geophones were deployed in a regular grid with the source located in the northern quadrant of the array (Figure 2-1d). The geophones to source offset ranged from 325 m to 1774 m. The 3-component geophones were rotated into the radial component to align with the nearly co-located DAS. The DAG-2 chemical explosion had a yield of 50.997 metric tons (TNT Equivalent) and was detonated at 299.8 m depth. The DAG-3 chemical explosion was significantly smaller at 0.908 metric tons and shallower at 149.9 m depth. Detailed information about the DAG experiment can be found at: http://ds.iris.edu/data/reports/2021/21-022/DAG_DataReleaseReport_FINAL.pdf (accessed April 12th, 2022).

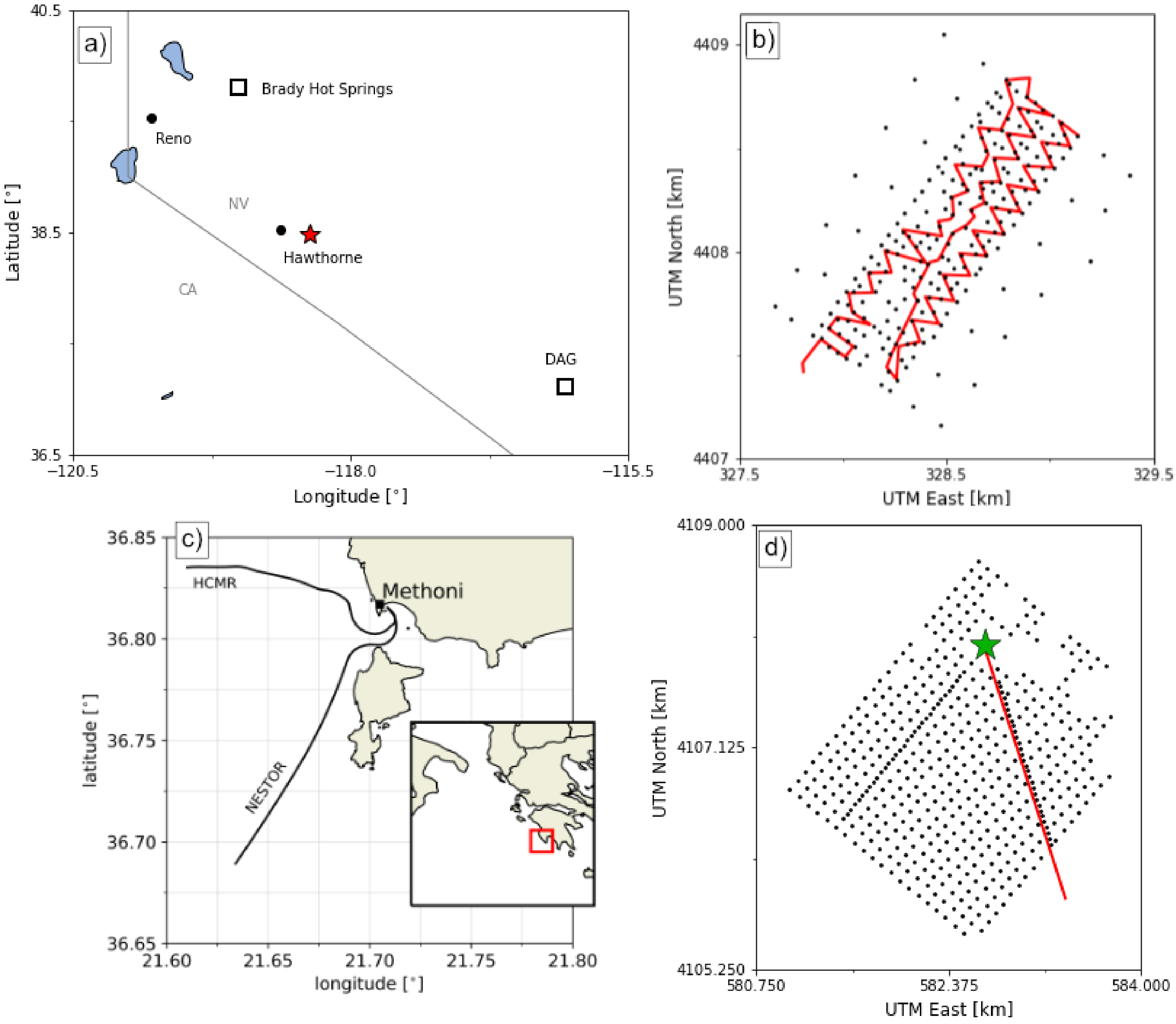

Figure 2-1: Location of the data arrays and sources.

(a) Regional view of the deployments and the M4.1 Hawthorne, NV earthquake (red star) and the Brady Hot Springs/PoroTomo experiment. (b) PoroTomo experiment geometry with nodal geophones shown as black dots and the DAS as a red line. (c) The HCMR and NESTOR cables offshore southwestern Greece. (d) DAG experiment with nodal geophones in black, the DAS stations as a red line, and the source ground-zero in green.

2.2. Methods

2.2.1 Conversion between strain-rate and ground motion

We present two methods of converting ground motion and strain-rate discussed in Wang et al. (2018) with equations derived in Bakku (2015) and Daley et al. (2016).

A) *Euler's method* (Bakku, 2015) uses first-order finite differences to solve a partial differential equation (Eq. 1). It relates the observed strain-rate and the spatial derivative of ground velocity. It can be applied when geophones are co-located with DAS stations:

$$\dot{\varepsilon}_{DAS}(x,t) = \frac{\dot{u}\left(x+\frac{L}{2},t\right)-\dot{u}\left(x-\frac{L}{2},t\right)}{L} \#(1)$$

where $\dot{\varepsilon}_{DAS}(x,t)$ is the strain-rate observed on the DAS at location, x, at time t, $\dot{u}(x,t)$ is the observed ground velocity on a geophone at location x and time t, and L is the gauge length used for the spatial derivative. The dot indicates a time derivative relative to strain (ε) or displacement (u).

This method is used to infer ground motion from DAS strain-rate at a given location relative to a reference measurement. Implementation of equation 1 to estimate ground velocity at location x can be expressed as (van den Ende and Ampuero, 2021, their equation 7):

$$\dot{u}(x,t) = \dot{u}(x_{ref},t) + \dot{\varepsilon}(x;x_{ref},t)\Delta x \#(2)$$

where $x_{ref}$is the reference location of a ground velocity ($\dot{u}$) measurement.

The second term on the right-hand side can be expanded as a finite sum for location x = x_{ref}+nL, where L is the gauge length and n is an integer (modified from van den Ende and Ampuero (2021), their equation 8):

$$\dot{\varepsilon}(x;x_{ref},t)\Delta x = L\sum_{i=1}^{n}\dot{\varepsilon}\left(x_{ref}+iL;x_{ref}+(i-1)L,t\right) \#(3)$$

B) *The frequency-wavenumber (f-k) method* (Bakku, 2015; Daley et al., 2016) scales the 2D wavefield spectrum by the reciprocal of the phase velocity – defined as the wavenumber divided by the angular frequency (Eq. 4). This is particularly effective for linear arrays parallel to the propagation direction of a plane wave, but it could be generalized for 2.5D f-k spectral analysis. Because it uses the apparent phase velocity, it provides an opportunity to suppress or highlight portions of the timeseries considered noise or signal, as relevant for the analysis.

$$\dot{u}(\boldsymbol{x},\boldsymbol{k},\boldsymbol{\omega}) = \pm\left(\boldsymbol{k}/\boldsymbol{\omega}\right)\varepsilon(\boldsymbol{x},\boldsymbol{k},\boldsymbol{\omega}) \#(4)$$

where **x** is a vector array of locations, **k** is the spatial wavenumber vector, $\boldsymbol{\omega}$ is the angular frequency, and the ratio $\boldsymbol{k}/\boldsymbol{\omega}$ is inversely proportional to the apparent phase velocity across the array. Note that this operates in the frequency-wavenumber space, which is the Fourier Transform of the time-space domain.

2.2.2 Beamforming

Beamforming is a method of shifting and stacking waveforms to determine the vector slownesses of waves propagating across an array. Due to its value to estimating source parameters, beamforming is a widely used array processing method (e.g. Rost and Thomas, 2002). Several variations have been developed and here we apply a method termed multiple signal classification (MUSIC, e.g., Meng et al., 2011). The highly linear DAS geometry and low broadside incidence sensitivity provide important complications to applying beamforming directly to DAS data, but techniques have been developed to address these limitations. The most notable steps are to consider the overlapping results from linear sub-arrays and to convert the strain-rate data to ground motion before beamforming.

2.2.3 Blind Deep Learning Denoising

A major limitation to the broad usage of DAS is the higher noise level relative to traditional instruments. This is a common trade-off in array seismology where many low-cost sensors can be used to provide a denser characterization of the spatial wavefield at lower signal-to-noise ratio (SNR) compared to broadband seismometers. Van den Ende et al. (2021) use the array characteristics of DAS to detect outlier waveforms and replace them with predictions based on neighboring channels. They formulate a J-Invariant deep neural network that can be used to identify portions of record sections where data does not fit the expectation and replace that with predicted data. This function can operate “blind” in that there is no target signal that the algorithm is seeking to replicate. This has an advantage over standard frequency filtering because it will not remove signal when the signal and noise overlap in the frequency domain, and it is able to reconstruct missing data. However, it requires dense array data to determine the reference space.

3 RESULTS

3.1 Conversion between strain-rate and ground motion

As discussed above, there are two separate ways to convert between the ground motion system and strain-rate system. While not necessarily required for seismic analysis, because these systems can be measured independently, the transfer function between the two can be used as a quality control to verify observations on separate instruments. Reproduceable versions of the two conversion algorithms are provided in Wang et al. (2018) and van den Ende and Ampuero (2021).

3.1.1 Euler's Method

Examples of implementing Euler's Method are shown in Figure 3-1. The reference ground velocity at location x_{ref} is plotted in blue, the target ground velocity in red, and the intermediary integrated waveforms at non-overlapping gauge lengths according to equations 2 and 3 in black (Figure 3-1a). Figure 3-1b shows the individual DAS stations at the intermediary locations in purple. All waveforms are filtered between 0.5 and 25 Hz. In comparison, Wang et al. (2018) and van den Ende and Ampuero (2021) used a high pass filter corner of 0.5 Hz and a low pass filter corner of either 2 or 5 Hz.

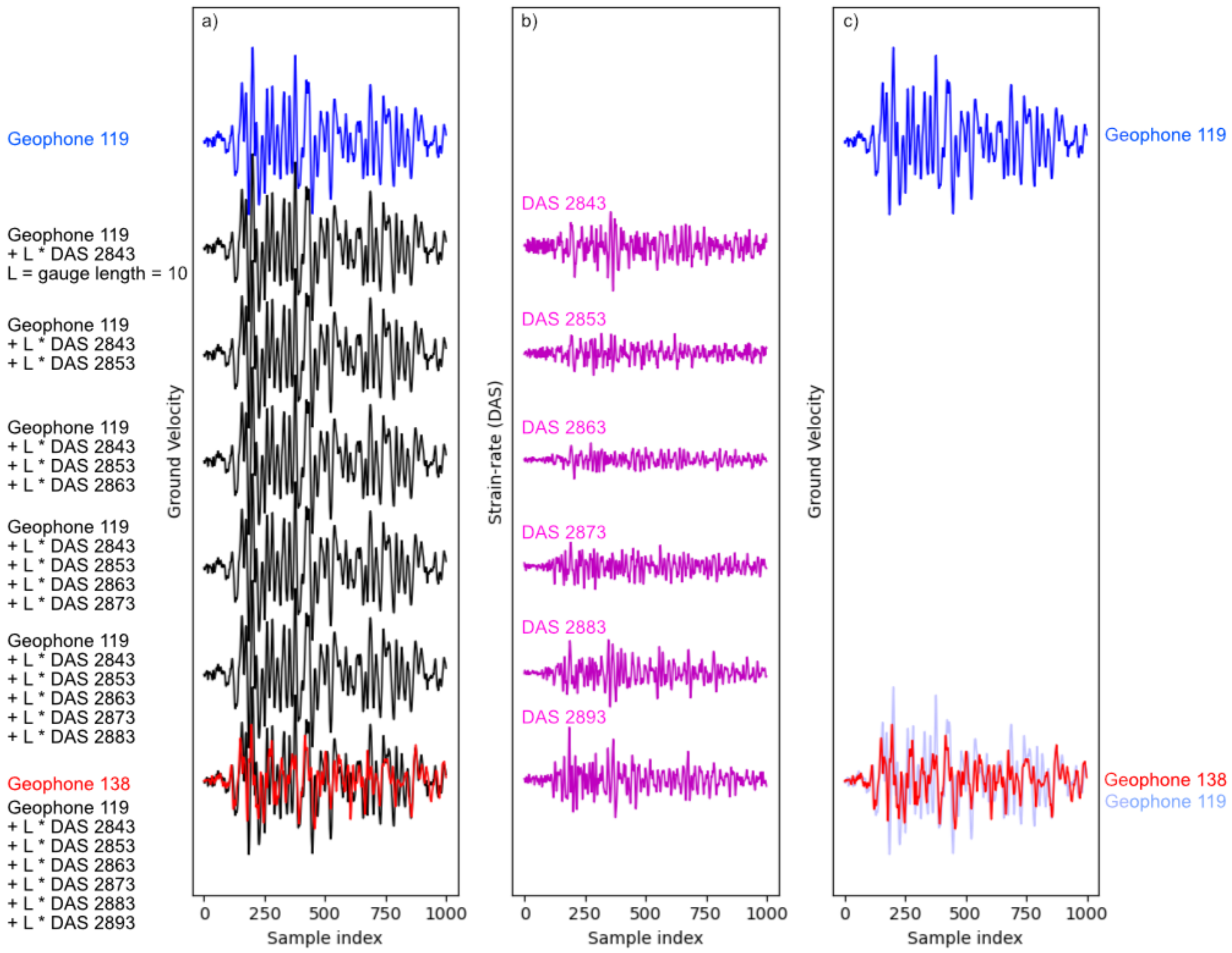

Figure 3-1: Example of Euler's Method modified from van den Ende and Ampuero (2021)

Observed ground motion at the beginning of a segment (blue) and end of a segment (red) for a regional P wave arrival at the PoroTomo experiment. Panel (a) shows black waveforms for the integrated DAS channels at non-overlapping gauge lengths. Panel (b) purple waveforms are the individual waveforms for DAS station IDs at non-overlapping gauge lengths. Panel (c) is the reference geophone 119 near the top and an overlay of geophones 119 and 138 at the bottom. Waveforms filtered between 0.5 Hz and 25 Hz. The offset between terminal geophones is 60 m and the source is ~157 km from the array.

From this, we can make several observations. First, there is a high qualitative similarity of target and integrated waveforms at high frequencies. We can quantify this with the variance reduction and cross-correlation:

$$VR = 100 * \left(1 - \frac{\sum (p-g)^2}{\sum p^2}\right) \#(5)$$

$$r = \frac{\sum(p-\overline{p})(g-\overline{g})}{\sqrt{\sum (p-\overline{p})^2 \sum (g-\overline{g})^2}} \#(6)$$

Where VR is the variance reduction, p and g are two waveforms being compared, r is the cross-correlation coefficient, and the mean of the series is designated with the bar (i.e. $\overline{p}$). These are both measures of fit with a maximum VR of 100% and maximum cross-correlation coefficient of 1.0. The results are provided in Table 3-1 for a low frequency band (0.5-5 Hz) and a high frequency band (0.5- 25 Hz). First and foremost, this indicates that the Euler integration results in far more similar waveforms compared to the two waveforms at the ends of the segment: Euler integration similarity ~72% VR, 0.91 r compared with ~30% VR and 0.60 r for low frequency and ~30% VR, 0.80 r compared with ~6-12% VR and ~0.45 r for high frequency. Secondarily, the cross-correlation coefficients between the two geophones of ~0.6 for low frequency and ~0.45 for high frequency are relatively high indicating significant similarity between the geophones. This is due to the signal being effectively a plane wave at the relevant spatial wavelengths and temporal frequencies. As a consequence of this observation, DAS has the potential to significantly densify observations of the spatial gradients of plane waves as required for methods such as gradiometry (e.g., Poppeliers, 2010).

Table 3-1: Measured variance reduction and cross-correlation coefficient for sample waveforms from the beginning of the test segment (geophone 119) to the end of the test segment (geophone 138). DAS integrated is the waveform following Euler integration of DAS data from geophone 119 to the location of geophone 138. Filter is applied before integration.

Waveform Pair	Variance Reduction [%]	Cross-correlation
Geophone 119 – Geophone 138 (0.5 – 5 Hz)	32.4	0.57
Geophone 119 – DAS integrated (0.5 - 5 Hz)	30.6	0.60
Geophone 138 – DAS integrated (0.5 - 5 Hz)	72.5	0.91
Geophone 119 – Geophone 138 (0.5 – 25 Hz)	11.8	0.41
Geophone 119 – DAS integrated (0.5 - 25 Hz)	6.0	0.48
Geophone 138 – DAS integrated (0.5 - 25 Hz)	29.2	0.80

We explore the generality of Euler's Method by applying it to the DAG-2 chemical explosion data. Initial tests found little improvement to the similarity metrics between the geophones and integrated DAS. To explore this, we reformulated the relationship as:

$$\dot{u}(x_{ref} + nL, t) = \dot{u}(x_{ref}, t) + \sum_{i=1}^{n} (\dot{\varepsilon}(x_{ref} + iL, t) * L * \sigma) \quad \#(7)$$

Where σ is a factor used to scale the strain-rate waveforms to compensate for unquantified factors such as coupling or broad-side incidence of scattered energy. This follows from the analysis by Lindsey et al., 2020 which shows that the phase response of a DAS is consistent with a geophone, but that the amplitude response is poorly constrained. We determine this amplitude response factor empirically by applying the algorithm to segments of DAS stations defined as the set of DAS sensing points between pairs of geophones. The fit between the geophone at the end of the segment and the integrated waveform was estimated with the variance reduction and cross-correlation coefficient as defined above. We also computed L1 (absolute value norm) and L2 (root mean squared norm) misfit functions, but found the results were consistent with the variance reduction metric. We find the optimum σ values vary strongly as a function of frequency by estimating over a series of narrow band filters. We also find that the spread of optimum values is significantly larger for the correlation coefficient method. This follows from observations of fit at constant frequency and segment which typically show a single peak for the variance reduction and often show an arctangent function for the correlation coefficient (Figure 3-2a, b). The variance reduction (and L1 and L2 misfits which showed the same relationship) is likely indicating a need for scaling to fit the amplitude while the correlation coefficient indicates that a time shift is needed to fit the phase. Furthermore, Figure 3-2 shows largely positive corrections needed for frequencies less than ~5-7 Hz and a negative correction at higher frequencies. As this change occurs near the peak response of the geophone, this may be an effect of the specific instrumentation.

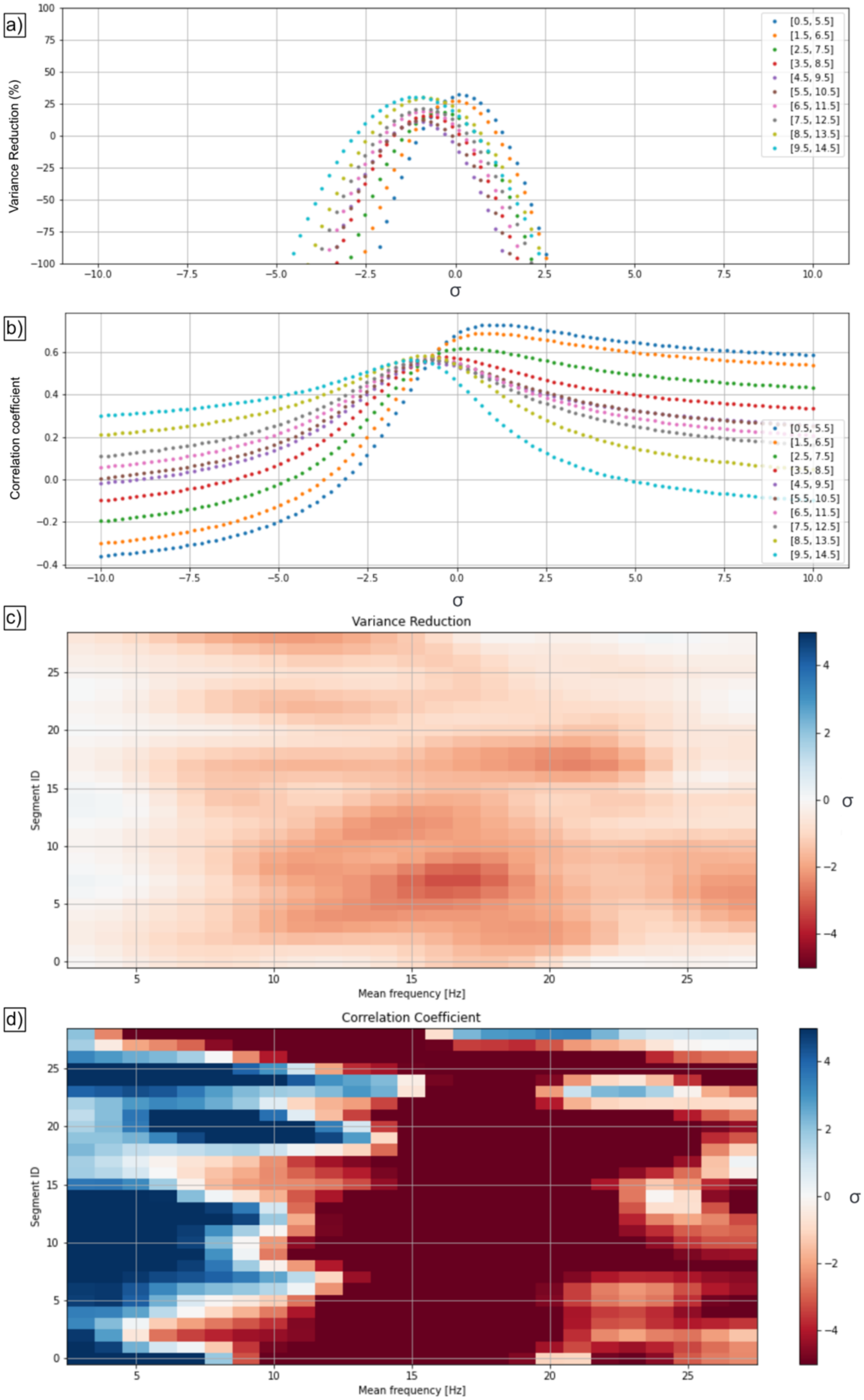

Figure 3-2: Determination of σ fit parameters for Euler's Method with DAG-2 data.
Measured variance reduction (a) and correlation coefficient (b) between integrated waveform and nodal geophone at the end of segment 16 (1124 m to 1174 m from the source ground-zero) for a set of test factors, σ. Colors indicate frequency band as enumerated in the legend. Best fit test factors for each segment and frequency band as measured by (c) variance reduction and (d) correlation coefficient.

The test results indicate that Euler's method can, in some cases, achieve quantifiable improvements of similarity between geophone data and DAS data. However, the observation that the final result of the integrated DAS data is highly similar to the original geophone data remains a significant shortcoming of this method. This is highlighted in Figure 3-3 where we compare the nodal waveforms at the start of a segment (Geophone 20), end of a segment (Geophone 21), and integrated waveforms at the end of the segment using the fit terms based on maximizing the variance reduction (i.e., matching amplitude). The integration has little effect relative to the effect of the ground velocity of the starting geophone for most cases.

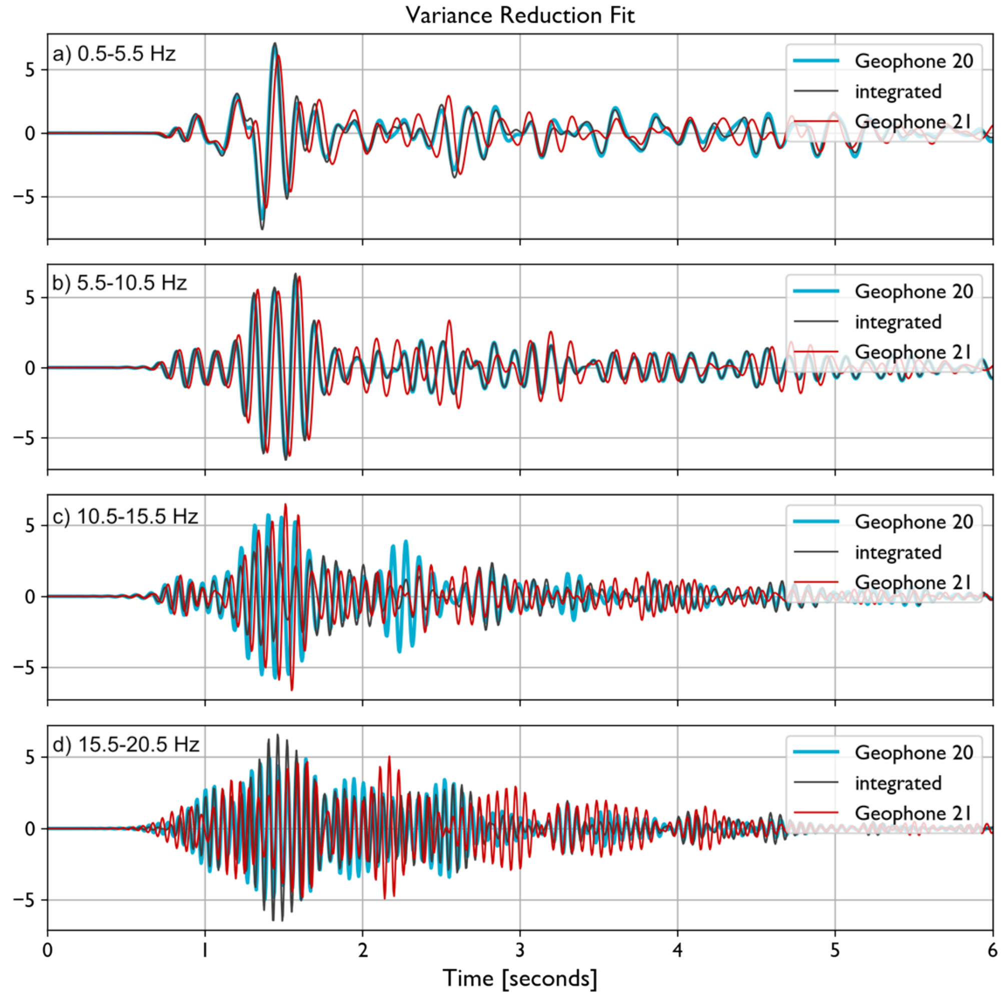

Figure 3-3: Example waveforms after applying Euler's Method to the DAG-2 data. Waveforms for the start of segment 20 (1325 m to 1374 m from source ground-zero) as Geophone 20 (blue), end of segment 20 as Geophone 21 (red), and the DAS integrated waveform (black) nearly co-located with Geophone 21 at 4 frequency bands as denoted in each panel.

3.1.2 F-k integration

In this section, we implement the f-k integration algorithm from Wang et al. (2018):

a) Integrate the DAS data with respect to time (from strain-rate to strain),

b) Apply a 2D Fast Fourier Transform (FFT) to the DAS record section data (i.e., time-space domain into frequency-wavenumber domain),

c) For each wavenumber and angular frequency,

    i) Multiply by the ratio of wavenumber to angular frequency ($\boldsymbol{k}/\boldsymbol{\omega}$),

d) Apply an inverse FFT to the 2D data array.

While this can appear straightforward, there are some important caveats that make direct replication challenging. First, this can only be applied for an evenly sampled linear array segment for accurate representations of the spatial vector wavenumber. Second, our testing found that the comparison to geophone-based data works best if we only consider wavenumber values that are recorded by both the geophone array and the DAS array. Because the geophones are more coarsely spaced than the DAS recording locations, this entails setting the strain spectrum values to 0 if their wavenumber in the f-k space is outside the range of wavenumbers in the geophone f-k space. In Figure 3-4 we illustrate this consideration. The geophone data only resolves a small subset of the DAS data wavenumber domain as indicated by the red lines Figure 3-4c, d. For the PoroTomo Experiment, this is poorly resolved due to only including three geophones in the spectrum, but the DAS data shows a more complex pattern. For DAG-2, because the geophone array is dense and linear, the spectrum is relatively well resolved and compares well with the DAS spectrum. The primary difference is that the DAS data contains increased low frequency energy relative to the 4-Hz geophones.

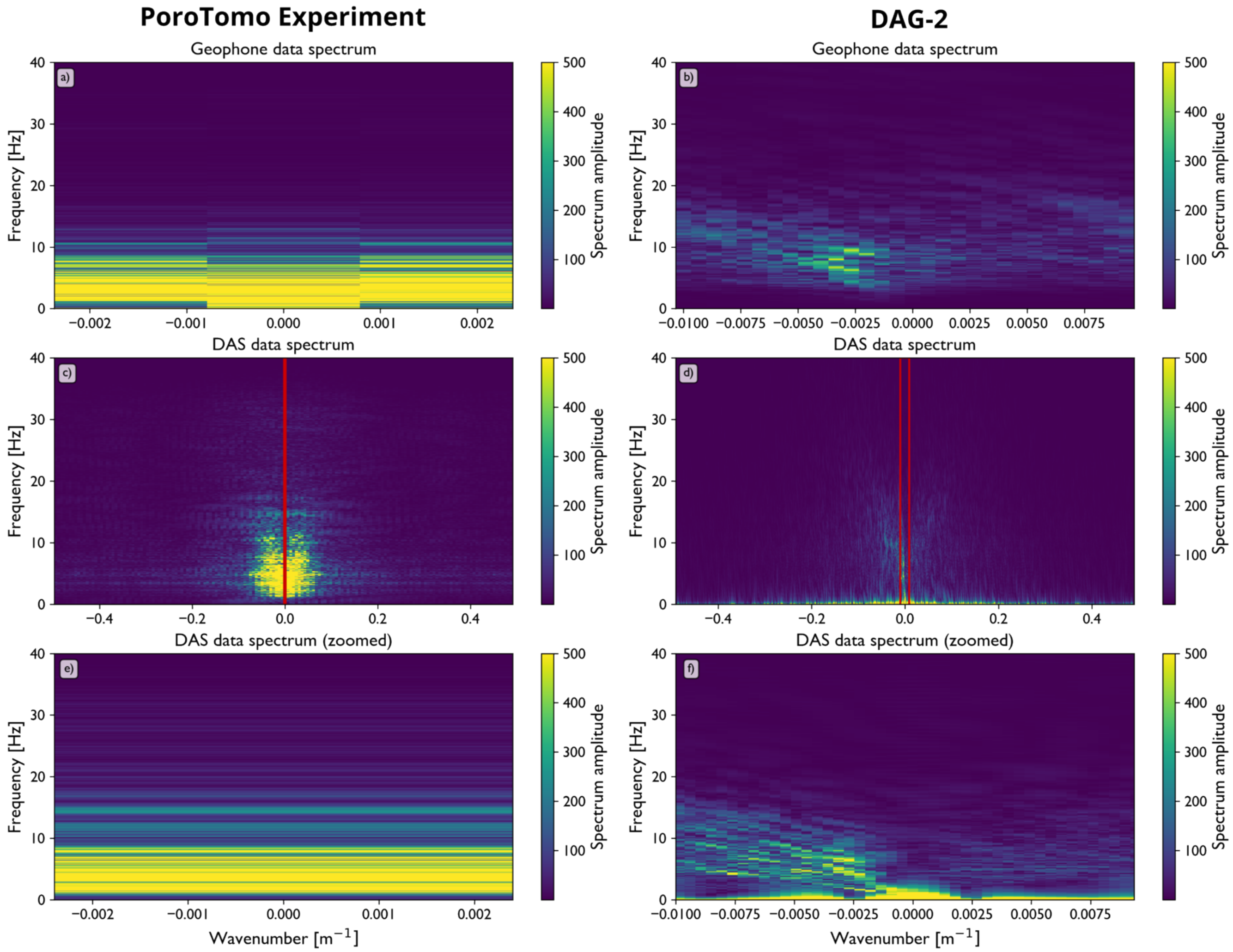

Figure 3-4: Frequency-Wavenumber (f-k) spectra of Geophones and DAS.

Panel (a) f-k spectrum computed for the 200-m spaced set of three geophones near DAS 7107 for the PoroTomo experiment. Panel (b) f-k spectrum computed for DAG-2 on the 30-m spaced geophones. Panels (c) and (d) are the corresponding DAS f-k spectra near DAS7107 (PoroTomo) and for DAG-2. Red lines indicate wavenumber space in (e) and (f) corresponding resolved limits in (a, b).

After applying the f-k integration and filtering and then returning the DAS spectra to the time-space domain, we can observe ground velocity from the DAS that is qualitatively similar to the geophones (Figure 3-5). The access to frequency dependence in the f-k conversion method may provide an unexpected solution to the significant challenge of dealing with unprecedented data volumes that DAS is able to generate. Much of the spectra in Figure 3-4c, d is effectively 0 and therefore little information is lost if we compress this space. Moreover, if we wish to replicate a nominally dense geophone array, we only need to extract a small subset of the f-k spectra. This represents a significantly smaller data volume than the raw data while still presenting sufficiently dense waveform data for standard seismological analyses.

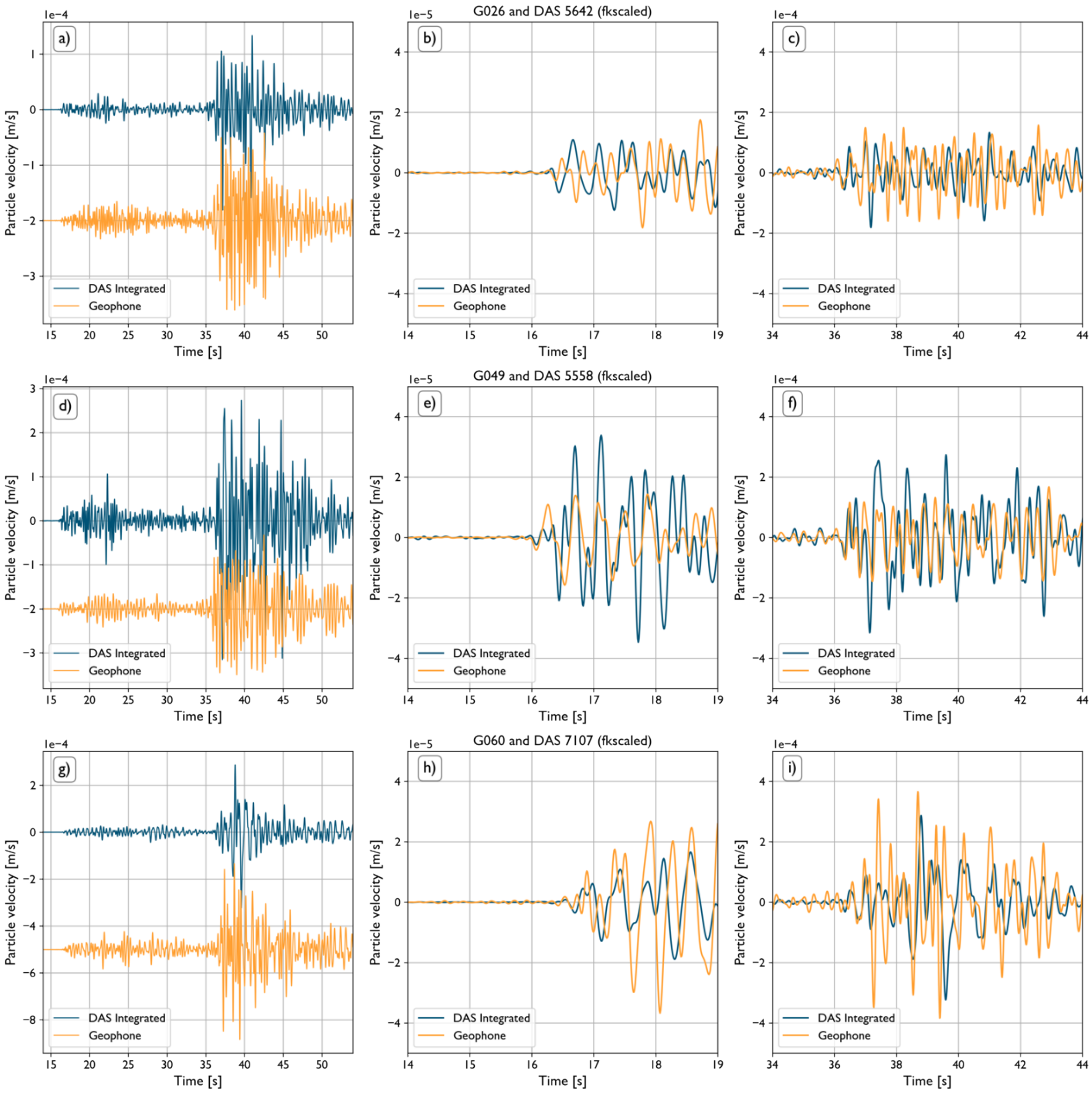

Figure 3-5: Ground velocity for DAS and geophones for PoroTomo.
Frequency-wavenumber integrated DAS and geophone ground velocity replicating Wang et al. (2018) Figure 18. Blue are the f-k integrated DAS and amber are the geophones scaled to ground motion. First arrivals have been visually aligned. Panels (a-c) compare geophone 26 and DAS station 5642. Panel (a) is the full waveform, (b) is focused on the P wave arrival and (c) is focused on the S wave arrival. Panels (d-f) similarly display the waveforms for geophone 49 and DAS 5558. Panels (g-i) are for geophone 60 and DAS 7107.

Application of f-k integration to DAG-2 DAS data is shown in Figure 3-6. The unscaled DAS data in Figure 3-6a shows qualitatively more high frequency energy than the geophone data in Figure 3-6c. The f-k scaled DAS data in Figure 3-6b is closer to the geophone data, with the significant first arrival between 1 and 2 seconds. Spatially decimating the f-k scaled DAS data to the nearest set of geophones in Figure 3-6d shows a significantly closer comparison with the geophone data. This confirms that the major features of the wavefield are observable in both datasets, but the higher spatial sampling of the DAS is capable of greater spatial resolution than the geophones.

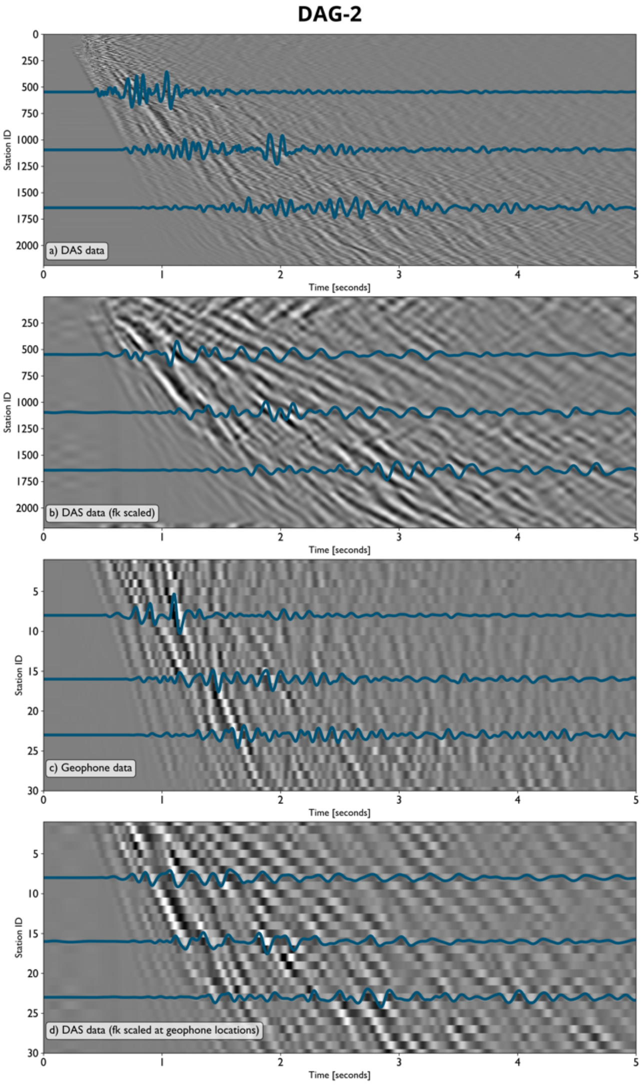

Figure 3-6: DAG-2 record sections comparing f-k integration scaling of DAS data.

Gray-scale image of full wavefield and sample waveforms from along the record section. Panel (a) is the raw DAS data. Panel (b) is the full f-k scaled DAS data. Panel (c) is the raw geophone data. Panel (d) is the f-k scaled DAS data spatially downsampled to the geophone locations.

3.2 Beamforming

Van den Ende and Ampuero (2021) show that beamforming results with DAS can be substantially improved by first integrating the strain-rate data into ground velocity (Figure 3-7, a-c). In an ideal case, beamforming returns an isolated peak at the vector slowness consistent with a plane wave propagating across the array. In the case of the Hawthorne earthquake recorded at the PoroTomo experiment, amplitudes are diffuse throughout the slowness space for the uncorrected DAS strain-rate data. In van den Ende and Ampuero (2021), following Euler integration of the DAS data, beamforming recovers a nearly identical result to that computed with the geophone data. This result follows the high waveform similarity achieved through Euler integration with their data and the application of stacking along multiple segments to simulate an array with non-colinear segments.

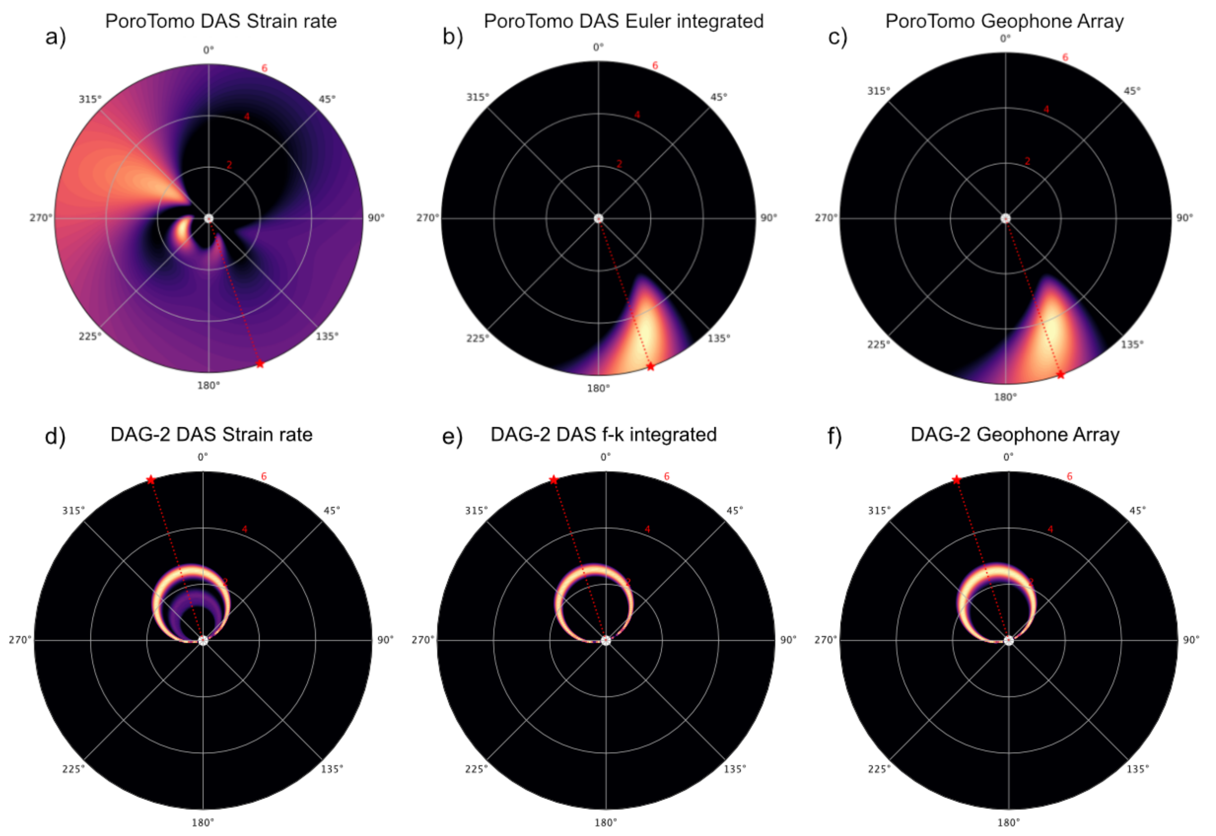

Figure 3-7: Application of beamforming via the MUSIC algorithm using PoroTomo data and DAG-2 data.

Slowness spaces (phase velocity increases radially outward) for the Hawthorne earthquake P wave arrival recorded on the (a) DAS, (b) Euler's method integrated DAS, and (c) geophones. Multiple non-parallel segments are used to improve the azimuthal resolution. Recomputed using data and software from van den Ende and Ampuero (2021). Panels (d-f) are as (a-c) but for the DAG-2 data using f-k integration for conversion from strain-rate to ground velocity.

Applying beamforming to the DAG-2 data, we find improvements after integrating the DAS data to ground velocity, but the difference is less significant when comparing to the PoroTomo experiment. The linear nature of the array generates a ring-shaped peak, rather than an isolated maximum. Figure 3-7, d-f, shows this relationship as well as the effect of using f-k integration on the DAG-2 DAS data. Beamforming on the strain-rate data alone can recover the same peak ring as the nodal array data, but there is a secondary maxima ring at lower phase velocity (closer to the origin). In contrast, the DAS data after f-k integration is visually identical to the beamforming from the linear nodal array.

3.3 Blind Deep Learning Denoising

We apply the array based denoising methodology presented in van den Ende et al. (2021) to their data from two dark fiber cables and data from the DAG-2 and DAG-3 chemical explosions. The process is illustrated in **Error! Reference source not found.**. Record sections of the original data show incidence of a plane wave near the center of the time window, but the arrival is obscured by significant noise unrelated to the incident wave. The denoising algorithm reconstructs the plane wave signal and this is further enhanced with a standard frequency filter.

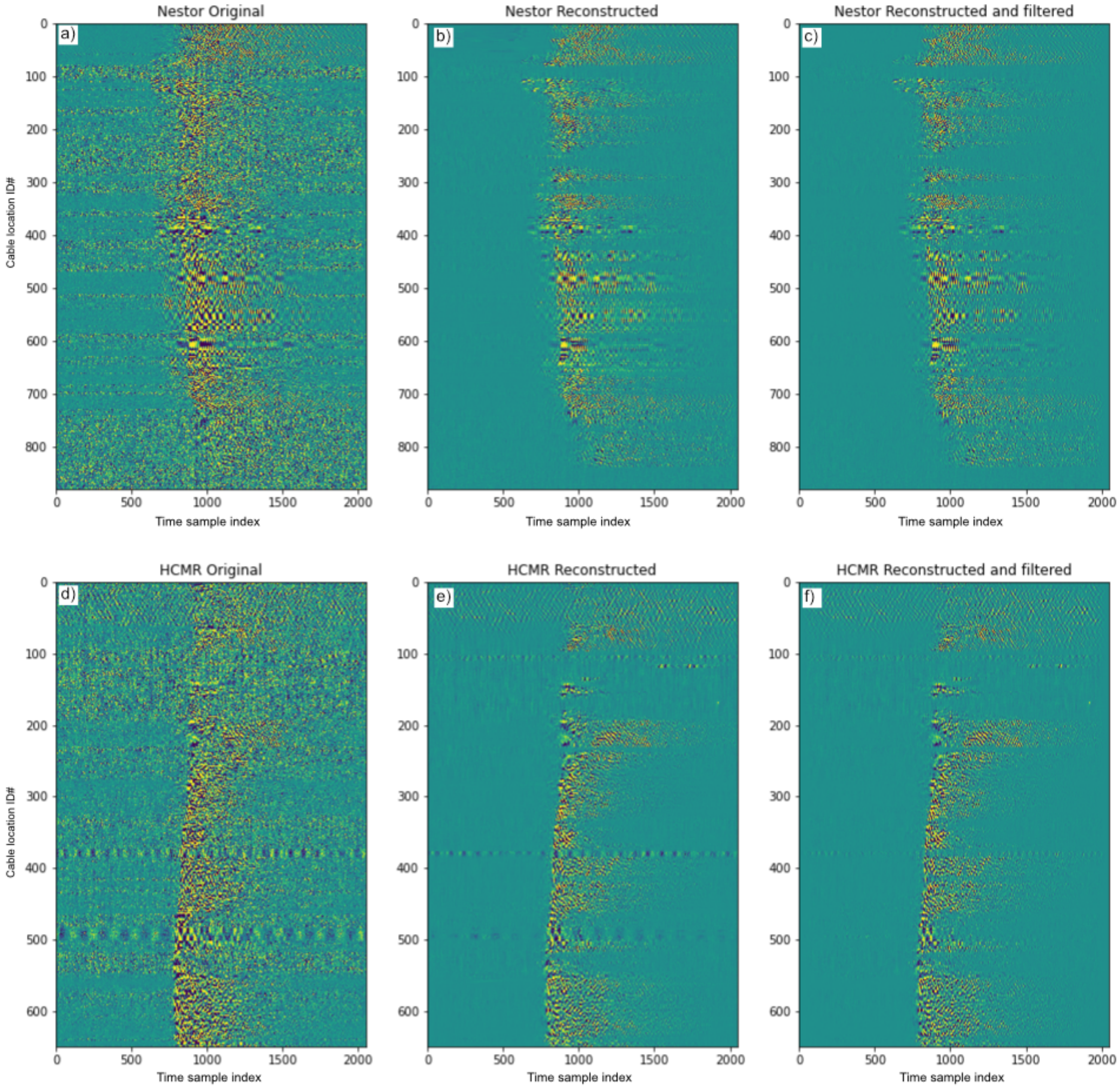

Figure 3-8: Denoising process of van den Ende et al. (2021).

Replication of denoising on the Nestor and HCMR cables from offshore Greece. Panel (a) is original data from the Nestor cable shown as a record section. Panel (b) reconstructed (aka denoised) record section of Nestor data. Panel (c) reconstructed and filtered record section of Nestor data. Panels (d-f) as (a-c), but for the HCMR cable.

As DAG-2 and DAG-3 are active source data at short offsets, the SNR is high in the raw data. This provides an opportunity to evaluate the effectiveness of the denoiser on constructed noise. We approach this by creating a noisy version of the dataset containing three primary noise factors, illustrated in Figure 3-9. First, we apply a constant level of background noise to every waveform. This constant noise amplitude is determined by the power spectrum of the Peterson, 1993 New Low Noise Model (NLNM), scaled to the input scale and set to a constant phase with small random perturbations (Figure 3-9a, c). Second, we apply a high amplitude, random phase spectrum to 25% of the DAS stations, with amplitude again following the NLNM, but scaled higher (Figure 3-9a, d). Third, because this method can also reconstruct missing stations, we set another 25% of waveforms to 0. These three additional noise sources represent the constant Earth background, spurious interfering signal-generated noise, and instrumentation failure, respectively. The resulting waveforms still show the primary arrivals, but also a much higher noise level compared to the original (Figure 3-10).

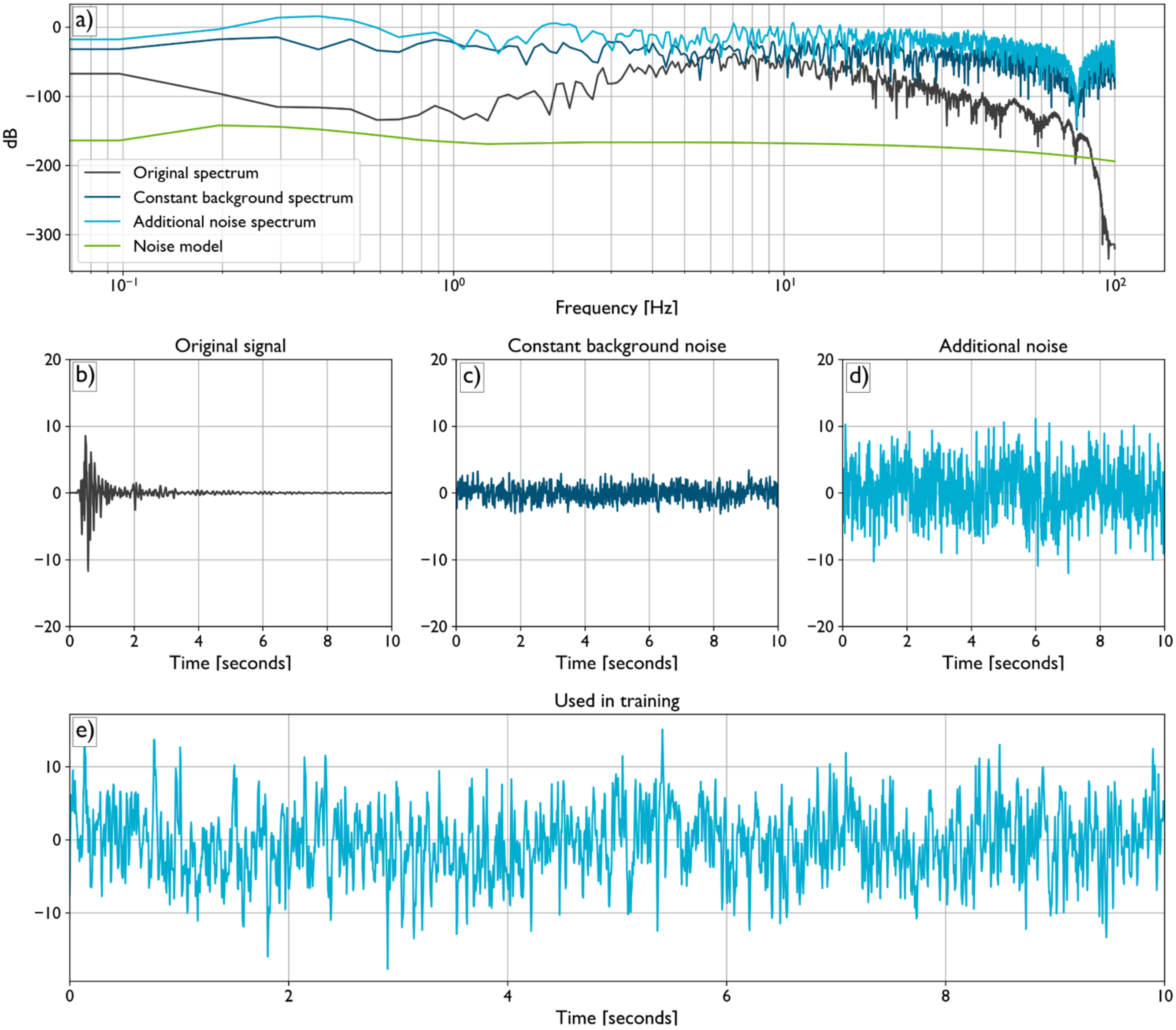

Figure 3-9: Process to generate realistic Earth noise

(a) Power spectra of the original data (black), added constant phase background noise (medium blue), random high amplitude noise (light blue), and reference model (NLNM: green). Panels (b-d) are the time domain representations of each. Panel (e) is the summation of the original signal, constant noise, and additional noise as was used to train the blind denoiser.

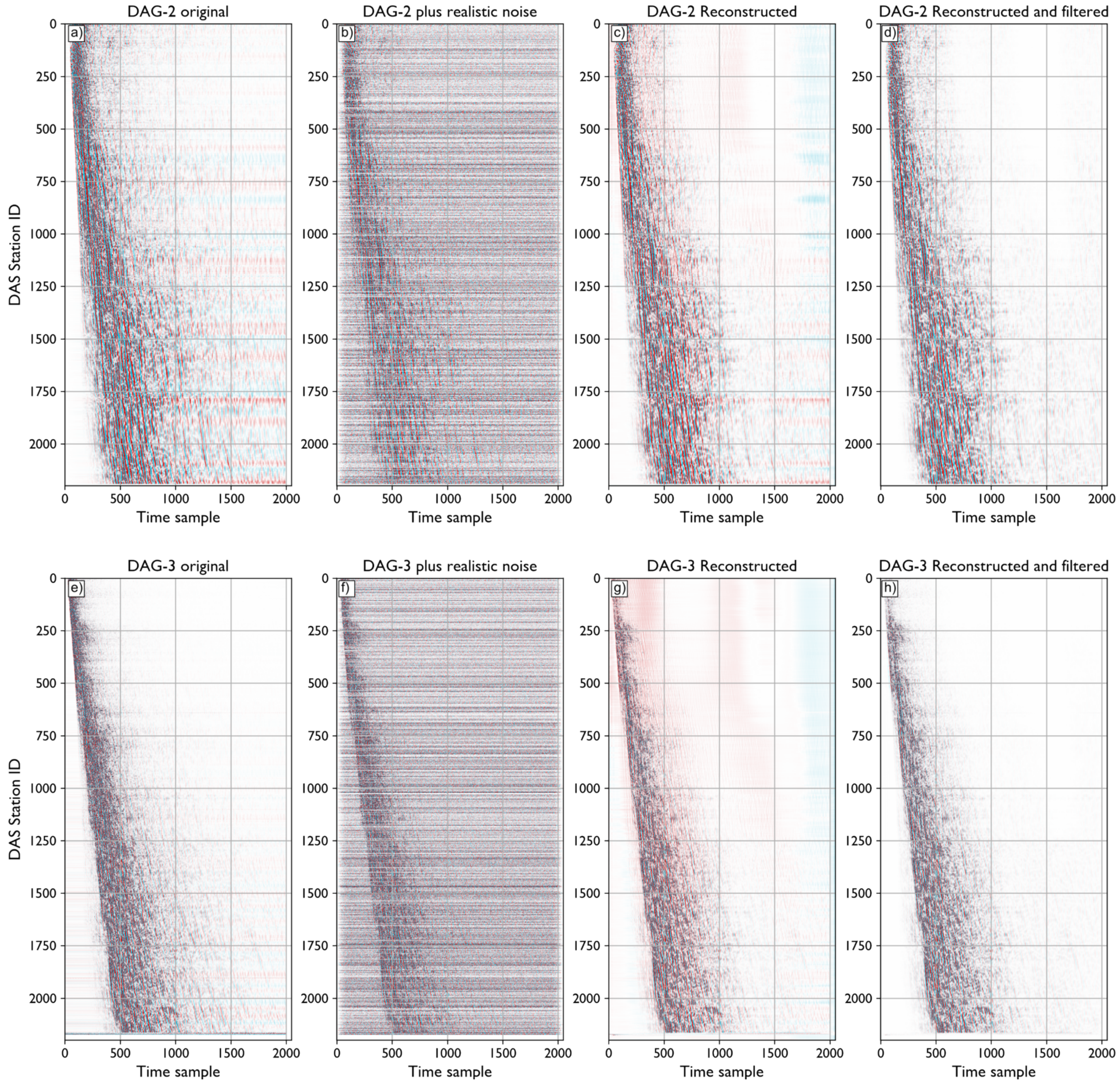

Figure 3-10: Result of adding and removing synthetic noise to DAG-2 and DAG-3 data. Panel (a) is the original data for DAG-2. Panel (b) is the DAG-2 after addition of synthetic noise and zeroing out random data. Panel (c) is the denoising reconstruction result. Panel (d) is the reconstructed data filtered at 3-20 Hz. Panels (e-h) are as (a-d), but for DAG-3 and panel (h) uses a low pass corner of 50 Hz.

Quantifying the similarity between the reconstructed and original data confirms the effectiveness of the denoising method. We use the variance reduction and correlation coefficient to assess waveform similarity, as defined above. Figure 3-11 shows the metrics for each DAS station for the cases of original vs. noise (Figure 3-11a), original vs. a bandpass filtered version of the noise-added data (Figure 3-11b), and original vs. reconstructed and filtered data (Figure 3-11c). For the first two cases, the trends are similar with correlation coefficients around 0.75 and variance reductions around 40-50% for most of the stations. There is another band of correlation coefficients around 0 corresponding to the zeroed-out stations. As these two images show highly similar results, we infer the bandpass filter is insufficient to reduce the constructed noise. However, the reconstructed data show far greater similarity to the original data than the other filtered data. These fits are around 85% variance reduction and 0.95 correlation coefficient, suggesting a good recovery of the original data. Moreover, the zeroed-out traces have been recovered and are consistent with the original waveforms.

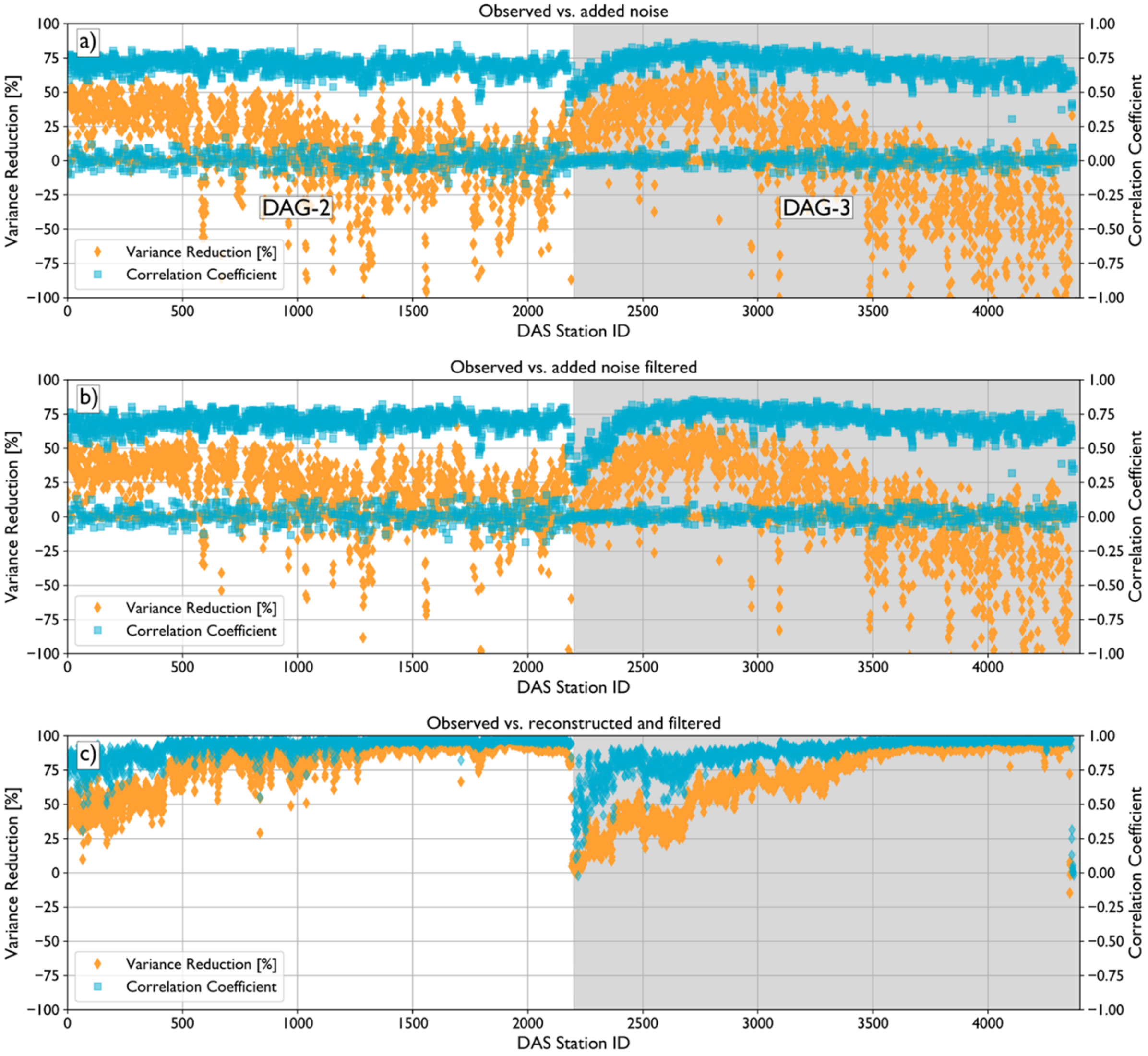

Figure 3-11: Variance reduction and correlation coefficient for the DAG-2 and DAG-3 denoising test.

Panel (a) is the variance reduction (amber) and correlation coefficient (blue) for the DAG-2 data (left) and DAG-3 data (right) comparing the original data with the data following addition of noise. Panel (b) is the same, but for the original data with the noise added version filtered with a bandpass filter (3-20 Hz for DAG-2 and 3-50 Hz for DAG-3). Panel (c) is for the case of comparing the original data with the reconstructed and then filtered data.

The distributions of the fit quality of the reconstructions are shown in Figure 3-12. As discussed above, the bandpass filter shows only a negligible improvement in the fit quality and there are distinct peaks at “0” in both the variance reduction and correlation coefficient. However, the fit metrics for the denoising reconstruction are substantially improved. This indicates the denoising algorithm can remove realistic noise applied to our dataset. As we prepare for DAS deployments aimed at long-term seismic monitoring, this may become an important tool for the common task of removing background noise. We will note, however, that this construction of noise does contain several parameter choices. By adjusting these parameters, we were able to trick the denoiser into maintaining only noise and removing the desired signal. Therefore, this tool will need to be carefully used in operation to only apply on cases where a significant signal can be observed amongst the noise.

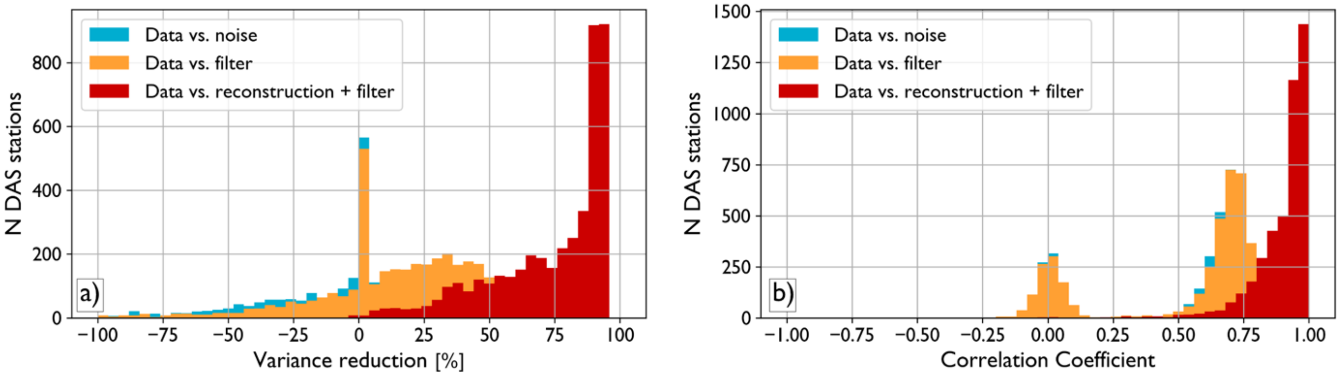

Figure 3-12: Histograms of the denoising fit metrics.

Summary statistics of the fit metrics in Figure 3-11. Panel (a) is for the variance reduction and panel (b) is the correlation coefficient.

4 STRENGTHS AND WEAKNESSES

The three tasks explored here cover some of the basic processing that will need to be utilized to maximize the amount of information returned from future DAS arrays. The conversion to ground velocity is necessary for tasks such as beamforming to derive plane wave propagation or estimating magnitude. Beamforming provides measurements that indicate the relative direction to the location of the source. Denoising helps enhance signals of interest from data. The versions of these algorithms we explored here are far from exhaustive but provide a baseline for additional future comparisons.

4.1 Conversion to Ground Velocity

The conversion of strain-rate data to ground velocity can take two approaches based on what extra data is available and the signal of interest. Euler's method works well for relatively low frequency (0.5-5 Hz) plane waves propagating across an array with co-located geophones. As a method based on solving a partial differential equation, it is sensitive to both the initial conditions (i.e., reference geophone) and order of the solution (typically first order). These limitations can provide an advantage, however, for arrays such as the PoroTomo experiment where the DAS array is better thought of as a set of sub-arrays, each anchored by pairs of geophones. However, this reduces the generalizability of the measurement and requires additional sources of data. Frequency-wavenumber integration, on the other hand, can be used without co-located geophones because it does not need similar boundary conditions. However, it relies on a 2D time-space to frequency-wavenumber transformation which is only defined for linear arrays. More complex fiber geometries will require either linear subsets, considerations of off-axis propagation, or numerical approximations to 2.5D transforms.

4.2 Array Beamforming

Beamforming is a primary tool for seismic monitoring because it derives information about the passage of seismic wavefields across an array which can provide more accurate information about the location of the source than single sensors. Tests with the raw data from an interrogator found minimally worse (DAG-2) to unusable (PoroTomo) beamforming results. However, in both cases, integration of the data to ground velocity recovered information identical to co-located geophones. This indicates that DAS is able to accurately recover wavefield parameters from beamforming after processing to ground motion. As DAS samples at much higher spatial and temporal resolution, this technique has the potential to uncover signals not observed by traditional arrays, but at the cost of a substantially higher data volume.

4.3 Array Denoising

The separation of signals of interest and those considered noise is a longstanding challenge. While innovative single-channel methods are being developed (e.g., Langston and Mousavi, 2019; Tibi et al., 2022), the array nature of DAS and recent developments in machine learning provide an opportunity to explore new approaches. The J-Invariant method of van den Ende et al., (2021) leverages these to enhance signals from passing earthquakes. Their work focused on real data with no ground truth resulting in a qualitative improvement in signal to noise. By fabricating synthetic noise and adding to a low-noise dataset, we were able to quantify that this approach can significantly reduce noise. However, it does need to be approached carefully as it could enhance coherent waveforms that would be considered noise rather than the user desired signal. As an emerging technology, we expect future work will show significant improvements, possibly by integrating single channel methods such as the continuous wavelet transform used in Langston and Mousavi (2019) or a convolutional neural network used in Tibi et al., (2022).

5 CONCLUSION

DAS is an emerging technology with significant potential for ground motion monitoring, however, significant limitations require further analysis to increase its efficiency and applicability to monitoring problems. In this report, we've replicated published studies and applied the methodologies to explosion data from SPE Phase II to document strengths and weaknesses of several processing methods. This exercise has revealed several key considerations when working with DAS data. First, conversion of strain-rate data to ground velocity data works best for the case where the passing wavefield approximates a plane wave for the array geometry. The Hawthorne earthquake evaluated with the PoroTomo experiment data was located at a sufficient distance so there was little phase delay across the array. This allowed the strain-rate records to effectively measure the gradient of the wavefield. However, this was not the case for DAG-2 data where the first order partial derivative solution was not applicable. The f-k integration method did prove effective when only considering the low wavenumber (i.e., long wavelength) components in the f-k spectrum. This is most readily applied for the case of a linear array, but algorithms using the 2D wavefield may provide increased generalization. The second observation is that beamforming with DAS strain-rate data tends to show artifacts which can be removed by integrating to ground velocity before beamforming. For the case of a high amplitude or small-offset source, this may not be a necessary step as the primary signal is significantly higher than the artifacts. However, we were unable to fully explore the 2D effects of a non-linear array and broadside incidence. The third observation is that the high spatial density of DAS data is effective at denoising by using the array characteristic of the dataset. While we have only explored one methodology, the observation that this denoising technique takes advantage of spatially local waveform similarity suggests there may be other algorithms that can be developed using this same concept in addition to denoising developments on single waveforms. Based on these observations, DAS has significant potential to augment existing technologies and operate as its own primary monitoring technology. Outstanding challenges include managing the large volumes of data and optimizing array geometry for monitoring applications.

REFERENCES

[1] Dean, T., Cuny, T., and Hartog, A. H. (2017). The effect of gauge length on axially incident P-waves measured using fibre optic distributed vibration sensing. *Geophysical Prospecting*, 65 (1): 184-193, doi: 10.1111/1365-2478.12419

[2] van den Ende, M. P. A. and Ampuero, J.-P. (2021). Evaluating seismic beamforming capabilities of distributed acoustic sensing arrays. *Solid Earth*, 12, 915-934, doi:10.5194/se-12-915-2021

[3] van den Ende, M. P. A., Lior, I., Ampuero, J.-P., Sladen, A., Ferrari, A., and Richard, C. (2021). A Self-Supervised Deep Learning Approach for Blind Denoising and Waveform Coherence Enhancement in Distributed Acoustic Sensing Data. *IEEE Transactions on Neural Networks and Learning Systems*, doi:10.1109/TNNLS.2021.3132832

[4] Feigl, K. L., and PoroTomo Team (2017). Overview and preliminary results from the PoroTomo project at Brady Hot Springs, Nevada: poroelastic tomography by adjoint inverse modeling of data from seismology, geodesy, and hydrology, in *42nd Stanford Workshop on Geothermal Reservoir Engineering*, Stanfrod University, Stanford, CA, SGP-TR-212, 15pp.

[5] Langston, C. A., and Mousavi, S. M. (2019). Separating signal from noise and from other signal using nonlinear thresholding and scale-time windowing of continuous wavelet transforms. *Bulletin of the Seismological Society of America*, *109* (5), 1691-1700. doi:10.1785/0120190073

[6] Lindsey, N. J., Rademacher, H., and Ajo-Franklin, J. B. (2020). On the Broadband Instrument Response of Fiber-Optic DAS Arrays. *Journal of Geophysical Research – Solid Earth*, 125 (2), doi:10.1029/2019JB018145

[7] Martin, E. R., Lindsey, N. J., Ajo-Franklin, J. B., and Biondi, B. L. (2021). Introduction to Interferometry of Fiber-Optic Strain Measurements. *Distributed Acoustic Sensing in Geophysics: Methods and Applications*, 111-129, doi:10.1002/9781119521808.ch9

[8] Poppeliers, C. (2010). Seismic Wave Gradiometry Using the Wavelet Transform: Application to the Analysis of Complex Surface Waves Recorded at the Glendora Array, Sullivan, Indiana, USA. *Bulletin of the Seismological Society of America*, 100 (3): 1211-1224, https://doi.org/10.1785/0120090304

[9] Porritt, R. W., Abbot, R. E., and Poppeliers, C. (2022). Quantitative assessment of Distributed Acoustic Sensing at the Source Physics Experiment, Phase II. Sandia National Laboratories, Technical Report, SAND2022-0715 702945. Doi:10.2172/1855336

[10] Rost, S. and Thomas, C. (2002). Array Seismology: Methods and Applications. *Reviews of Geophysics*, 40 (3) 1008, doi:10.1029/2000RG000100

[11] Snelson, C. M., Abbott, R. E., Broome, S. T., Mellors, R. J., Patton, H. J., Sussman, A. J., et al. (2013). Chemical explosion experiments to improve nuclear test monitoring. *Eos Transactions of AGU*, 94(27), 237– 239. https://doi.org/10.1002/2013EO270002

[12] Tibi, R., Young, C. J., and Porritt, R. W. (2022). Comparative Study of the Performance of Seismic Waveform Denoising Methods Using Local and Near-Regional Data. *Bulletin of the Seismological Society of America* 2022; doi: 10.1785/0120220105

[13] Wang, H. F., Zeng, X., Miller, D. E., Fratta, D., Feigl, K. L., Thurber, C. H., and Mellors, R. J. (2018). Ground motion response to an M_L 4.3 earthquake using co-located distributed acoustic sensing and seismometer arrays. *Geophys. J. Int.,* 213, 2020-2036, doi:10.1093/gji/ggy102

DISTRIBUTION

Name	Org.	Sandia Email Address
Christopher J. Young	6371	cjyoung@sandia.gov
Christian Poppeliers	8911	cpoppel@sandia.gov
Technical Library	1911	sanddocs@sandia.gov

This page left blank

Sandia
National
Laboratories