

Wavefield-based evaluation of DAS instrument response and array design

Jack B. Muir¹ and Zhongwen Zhan

Seismological Laboratory, Division of Geological and Planetary Sciences, California Institute of Technology, CA, USA. E-mail: jmuir@caltech.edu

Accepted 2021 October 18. Received 2021 September 27; in original form 2021 July 15

SUMMARY

Distributed acoustic sensing (DAS) networks promise to revolutionize observational seismology by providing cost-effective, highly dense spatial sampling of the seismic wavefield, especially by utilizing pre-deployed telecom fibre in urban settings for which dense seismic network deployments are difficult to construct. However, each DAS channel is sensitive only to one projection of the horizontal strain tensor and therefore gives an incomplete picture of the horizontal seismic wavefield, limiting our ability to make a holistic analysis of instrument response. This analysis has therefore been largely restricted to pointwise comparisons where a fortuitous coincidence of reference three-component seismometers and colocated DAS cable allows. We evaluate DAS instrument response by comparing DAS measurements from the PoroTomo experiment with strain-rate wavefield reconstructed from the nodal seismic array deployed in the same experiment, allowing us to treat the entire DAS array in a systematic fashion irrespective of cable geometry relative to the location of nodes. We found that, while the phase differences are in general small, the amplitude differences between predicted and observed DAS strain rates average a factor of 2 across the array and correlate with near-surface geology, suggesting that careful assessment of DAS deployments is essential for applications that require reliable assessments of amplitude. We further discuss strategies for empirical gain corrections and optimal placement of point sensor deployments to generate the best combined sensitivity with an already deployed DAS cable, from a wavefield reconstruction perspective.

Key words: Inverse theory; Joint inversion; Wavelet transform; Earthquake ground motions; Site effects.

1 INTRODUCTION

Distributed acoustic sensing (DAS) networks utilize time-of-flight interferometry of Rayleigh backscatter from natural imperfections within fibre-optic cables to obtain spatially resolved strain measurements. Current technological standards allow a single DAS interrogator unit to produce metres spaced channels along cables of length ~ 20 –100 km; in other words, creating a distributed strain network of many thousands of sensors. While most original application of DAS was in industry, where the spatial resolution and high environmental tolerance of fibre-optic cable makes DAS eminently suitable for borehole deployment, the falling costs of the interrogator units and the increasing utilization of ‘dark fibre’ (already laid inactive telecom fibre) have made surface deployments an attractive proposition for fundamental research purposes. Recent studies have shown that onshore DAS can detect both teleseismic (Lindsey *et al.* 2017; Yu *et al.* 2019) and local (Wang *et al.* 2018; Karrenbach *et al.* 2020; Zhan 2020) earthquakes with waveforms that match colocated seismometers. Several studies utilizing offshore cables have also reported success in observations of both earthquakes and

ambient environmental seismology (Lindsey *et al.* 2019; Sladen *et al.* 2019; Williams *et al.* 2019; Spica *et al.* 2020; Ide *et al.* 2021; Matsumoto *et al.* 2021). DAS also promises to be one of the key data sources in the nascent field of social seismology, with recent success in monitoring parade traffic in an urban setting (Wang *et al.* 2020) and real time monitoring of rail traffic (e.g. Ferguson *et al.* 2020; Wiesmeyr *et al.* 2020). Mixed networks of DAS and point sensors have also proven to be a successful combination for fine scale monitoring of seismic properties associated with hydrothermal fields (Feigl 2017; Feigl & Parker 2019), and hold promise in integration into traditional ambient-noise tomography workflows (Nayak *et al.* 2021).

While DAS presents great opportunities for very dense deployments, it also comes with several notable challenges. The signal-to-noise ratio of DAS data is typically poorer than that of conventional seismometers, and the instrument response in realistic deployments is also not fully understood, leading to a generally less informative single-channel observations. DAS measures the integrated strain rate or strain in a particular direction over a finite gauge length, meaning that individual stations or straight lengths of cable do not

have access to the full horizontal particle motion as would be obtained from a three-component point sensor. Finally, the huge volumes of data produced by DAS are highly redundant except at very high frequency, and storing and analysing them poses a great computational challenge. Many of these challenges can be overcome, or better understood, by treating the DAS sensor array as a means of accessing a single underlying wavefield, rather than as a collection of individual channels—that is, seeking a representation of the data as $u(x, y, t)$ rather than $u_i(t)$ where x and y are spatial components and i is a channel number. Such a representation allows us to study in detail the spatial response of the array, including gradient terms, which are essential for strain based measurements like DAS. The wavefield representation offers a simple and coherent mechanism for converting between strains and displacements, which makes it especially well suited for studying the lateral variations of DAS amplitude response. Characterization of this response will be essential for realizing the promise of DAS as a spatially dense sensing modality for strong ground motions and earthquake early warning (Karrenbach *et al.* 2020).

In the preceding paper (Muir & Zhan 2021), we developed a framework for combining an irregular network of spatially distributed sensors into a single unified data product by using wavefield reconstruction. In this study, we apply our compressive framework to the combined DAS and nodal array deployed during the PoroTomo Experiment at Brady, NV (Feigl 2016a, 2017; Feigl & Parker 2019). Nodal arrays are constructed from self-contained cable-free instruments that sense the conventional wavefield (i.e. time derivatives of the particle displacement). Nodal arrays have proven highly successful in temporary dense deployments (i.e. Lin *et al.* 2013; Jia & Clayton 2021). Nodes strongly complement DAS arrays in the burgeoning earthquake rapid-response space by providing extra spatial coverage to supplement dark-fibre DAS deployments. We perform a comparison between the strain-rates recorded on the DAS array and those predicted by reconstructing the velocity wavefield recorded on the nodal array. While this analysis has been performed for individual DAS segments using colocated seismometers (e.g. Wang *et al.* 2018; Lindsey *et al.* 2020; van den Ende & Ampuero 2021), our framework allows us to evaluate the DAS array using the entire recorded nodal data volume simultaneously, allowing for non-optimally aligned segments, without favourable reference seismometers, to be studied. The wavefield approach also sidesteps the need to apply f/k rescaling (or similar methods that rely on estimating a reference phase velocity which become unstable for small wavenumbers k (Lindsey *et al.* 2020)) when converting from strain rate to velocity. This aids in the simplicity of application and the robustness of the results. We further develop methods for simultaneously reconstructing DAS and nodal data, correcting DAS data using the observed nodal field, and optimal design strategies for mixed arrays.

2 INVERSION OF DAS RECORDS FOR PARTICLE VELOCITY

DAS measurements are performed by observing the change in backscattering characteristics from laser pulses sent into a fibre-optic cable. Measurements ideally correspond to the strain or strain rate averaged along a gauge-length at some point in the fibre. The strain ϵ_c along a cable is given by the projection of the horizontal

strain tensor along the cable azimuth α

$$\begin{aligned}\epsilon_c &= [\sin \alpha \cos \alpha] \begin{bmatrix} \epsilon_{xx} & \epsilon_{xy} \\ \epsilon_{xy} & \epsilon_{yy} \end{bmatrix} \begin{bmatrix} \sin \alpha \\ \cos \alpha \end{bmatrix} \\ &= \sin^2 \alpha \epsilon_{xx} + \sin 2\alpha \epsilon_{xy} + \cos^2 \alpha \epsilon_{yy}.\end{aligned}\quad (1)$$

with the strain components given as usual by $\epsilon_{ij} = \frac{u_{i,j} + u_{j,i}}{2}$. The theoretical DAS strain-rate response ϵ_{DAS} is in turn given by averaging the strain or strain rate (Bakku 2015; Wang *et al.* 2018) over a gauge length L

$$\epsilon_{\text{DAS}} = \frac{1}{L} \int_{-L/2}^{L/2} \epsilon_c(l) dl, \quad (2)$$

where l indexes the length along the gauge length. Typical gauge lengths are on the order of ~ 10 m. Work to characterize the phase and amplitude response of DAS cables is rapidly progressing. If a DAS array is deployed such that it has sensitivity to both components of the horizontal wavefield, the general framework of Muir & Zhan (2021) may be employed to cast the recovery of the horizontal wavefield from DAS records as an inverse problem—we develop the necessary inverse problem machinery in Appendix A. Qualitatively, the framework of Muir & Zhan (2021) creates a sparse representation of the wavefield using a reweighted curvelet basis (Candes & Demanet 2005), scaled to promote continuity of the wavefield Laplacian. This rescaling, coupled with the use of a curvelet representation which is natively well suited towards representing wave propagation, promotes interpolations that satisfy the seismic wave equation and result in smooth derivatives that can be used to calculate the predicted DAS strain rate based on the observed velocity field.

While most extant DAS deployments are typically linear or loop-shaped, the PoroTomo experiment deployed at the Brady, Nevada thermal field in 2016 has a suitable space-filling zig-zag deployment geometry for wavefield reconstruction using DAS (Feigl 2017). This suitability has also separately lead to analysis of the beamforming capabilities of the PoroTomo DAS array (van den Ende & Ampuero 2021), where the authors discovered issues with coherence within the array inhibited beamforming performance. The lack of coherence was interpreted to be variously due to the cable geometry, coupling and the effects of small-scale scattering. By integrating DAS strain records, the beamforming issues highlighted by van den Ende & Ampuero (2021) were largely ameliorated, with these results highlighting the potential utility of wavefield reconstruction for improving the spatial continuity of the seismic waveform, thereby making it more suitable for spatial techniques such as beamforming.

The seismic component of the PoroTomo experiment ran for 15 d in 2016 March, with the instrumentation consisting of 8700 m of fibre-optic cable filling a rectangle with approximately 4:1 aspect ratio, and additionally instrumented with 238 three-component seismic nodes. The distribution of nodes and the geometry of the DAS cable are shown in Fig. 1. An additional 400 m of cable was deployed in a borehole, which we do not utilize in our wavefield reconstruction. Wang *et al.* (2018) showed that by appropriately differencing co-deployed nodal seismic instruments and comparing them to the average strain rate along cable segments, quantitatively similar strain-rate waveforms could be observed for an M_L 4.3 that occurred 2016 March 21 approximately 150 km SSE of the PoroTomo experiment. The methodology proposed in this study allows us to perform similar quantitative analysis on the entire DAS array, using a robust wavefield recorded on the entire nodal array, as well as the reverse experiment—to invert for the velocity field perturbations given the strain rates.

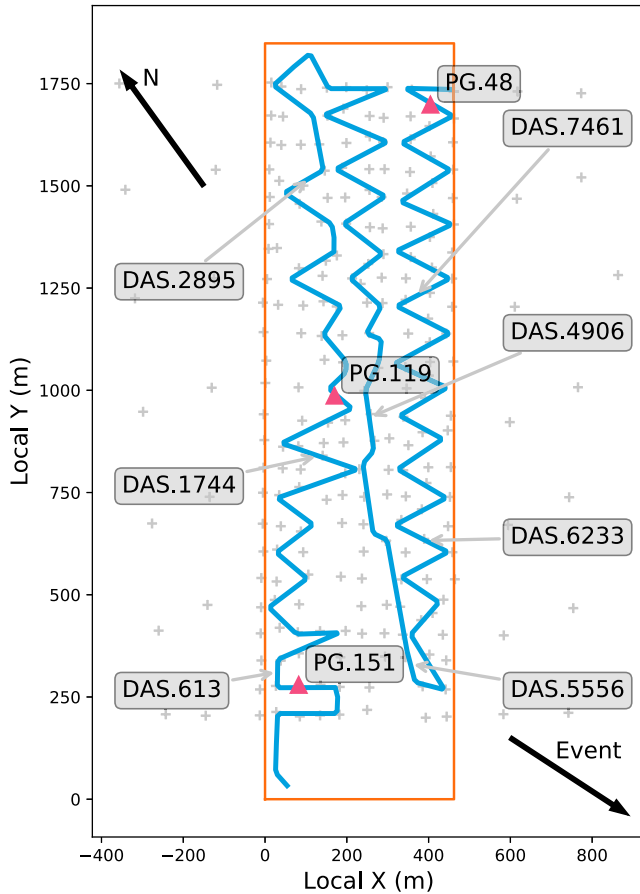


Figure 1. Deployment of DAS cable (blue lines) and nodal seismic stations (grey crosses) during the PoroTomo experiment at Brady, Nevada. The orange box shows the reconstruction domain, and pink stations are used for comparison in Fig. B1.

The PoroTomo team has provided data for the Fairfield ZL and 5 Hz three-component nodal seismometers, corrected for instrument gain to give units of coil-case velocity in $\mu\text{m s}^{-1}$ (Feigl 2017). Ringler *et al.* (2018) found that a collection of similar instruments can be well characterized by assuming a damped oscillator instrument response, with a mean free oscillation frequency of 4.87 Hz and a damping factor of 0.98 (averaged across all three channels for a collection of three instruments). We use these parameters to correct the nodal data to give ground velocity. DAS records were cleaned by removing channels for which the DAS gauge length included a corner (10 m for the PoroTomo deployment), and by removing traces below the 1st or above the 99th percentile of maximum log amplitude.

To begin with, we first emulate Wang *et al.*'s (2018) experiment using the compressive sensing framework developed in this study. Utilizing all nodes within the orange study area shown in Fig. 1, we inverted for a 32×128 pixel curvelet decomposition of the discrete wavelet transform (DWT) of the nodal ground velocity records. We then projected these results onto the DAS cable, and inverse transformed to give the predicted strain rate. Both nodes and DAS were bandpass decimated to a 10 Hz sampling rate, and a 1–2.5 Hz bandpass filter was applied. 2.5 Hz was chosen as an upper limit based on experimentation, which found it to be the highest frequency able to be reasonably reconstructed using DAS only data as seen in Appendix B. The Daubechies db12 wavelet with five levels was used for the time-domain transform (the optimal wavelet was

determined by experimentation), and the compressive sensing optimization used pure L1 regularization. The PoroTomo deployment is known to suffer from clock timing mismatches between the nodes and DAS (Wang *et al.* 2018). We estimated a best-fitting uniform clock correction by minimizing the least-squares misfit between the reconstructed DAS strain-rate data and the true DAS strain rate records as a function of lag time. This resulted in a best relative time-shift of 17 samples, or 1.7 s, with the DAS clock being faster. We plot the results (including time-shift) for seven cable locations (labelled in Fig. 1) in Fig. 2. We see that the recovered strain rate typically match the phase and relative amplitudes of the DAS records well. Absolute amplitudes are normally well recovered for the P wave. The machinery of Muir & Zhan (2021) can also be used to perform the reverse operation of predicting nodal data using the observed DAS strain rate without recourse to a plane-wave assumption and the consequent difficulties in obtaining a representative phase velocity, and also to jointly invert DAS and nodal data into a unified data product. The results of these experiments are shown in Appendices B and C respectively—notably the framework of wavefield reconstruction allows for a conversion of DAS strain rates to particle velocity without relying on an inferred horizontal phase velocity that may depend on local constitutive properties and that becomes difficult to stabilize for near-vertical incidence ground motions or where cable segments are insufficiently short for adequate k -space (wavenumber) resolution.

In Appendix C, we perform a fivefold cross-validation experiment to compare the reconstruction performance of a mixed network of DAS and nodes to the node only network. The low prediction error in the nodes-only cross-validation experiment shown in Fig. C1 illustrates that the nodal sensors at Brady are essentially able to fully capture the major details of M_L 4.3 Hawthorn NV wavefield within the 1–2.5 Hz frequency band, an argument that has also been made on the basis of beamforming coherence by van den Ende & Ampuero (2021). The average root-mean-square error (RMSE) when using 80 per cent of the nodes as a training set to predict the remaining 20 per cent left-out test data is 0.13. This suggests that we should be able to predict the DAS strain-rate data from the reconstructed velocity wavefield derived from the nodal sensors with a high degree of accuracy. While the results of Fig. 2 indicate that we can capture many features of the DAS data, there still remain substantial discrepancies to be explained. Sources of these discrepancies must relate either to incorrect metadata (imprecisions in the locations of the DAS channels, for example), unexplained errors in the ability of the DAS data to accurately record the true wavefield (i.e. instrument-ground coupling), or theory errors in the prediction of strain from the observed velocity. Given that strain measurements are acutely sensitive to very small-scale structure, the distinction between true instrument response, coupling between the instrument and the ground, and very local path effects are less distinct for DAS (and other strain-sensing modalities) than they are for point seismometers acting at typical 1 Hz frequencies (e.g. King *et al.* 1976; Ringler *et al.* 2019; Capdeville *et al.* 2020). Attempting to ascribe the apparent response of the cable to any one of these factors using only the predicted ground motions is difficult, however by studying the characteristics of response across the whole array it may be possible to build a hypothesis as to the predominant factors by searching for a correlation (or lack of correlation) with other data sets, such as tomographic models of the subsurface.

We now look at the characteristics of errors in the predicted wavefield in more detail. While the PoroTomo experiment unfortunately recorded only one earthquake source suitable for wavefield reconstruction, which precludes analysis of the DAS transfer

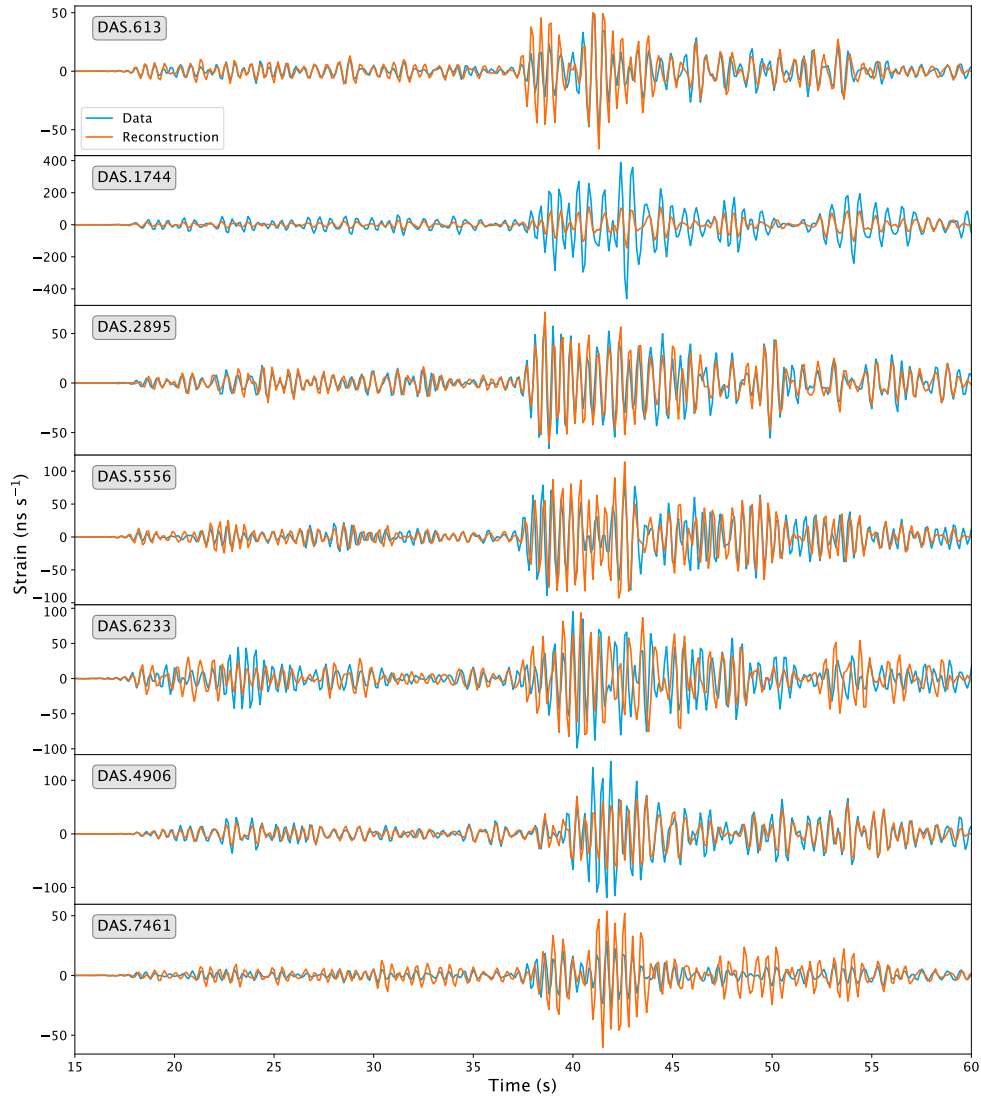


Figure 2. Comparisons between DAS channels (blue) and reconstructed strain rate from nodes (orange) at the seven stations highlighted in Fig. 1. Waveforms are bandpass filtered at 1–2.5 Hz, and nodal data have been corrected using a nominal Fairfield ZLand 3C instrument response.

function in detail as was performed for the FOSSA experiment by Lindsey *et al.* (2020), the areal coverage characteristics of the PoroTomo experiment offer a unique opportunity to investigate the spatial behaviour of the gross-scale metrics of instrument response. We computed zero-lag normalized cross-correlations between observed DAS strain rates and those predicted from the reconstructed nodal field as a metric of waveform shape matching or phase fidelity. We used the default settings for normalized cross-correlation in the ObsPy library (Beyreuther *et al.* 2010, i.e. demeaning and normalization of the signals in running windows). Averaged over the entire DAS array, we saw a median normalized zero-lag cross-correlation of 0.69, with a histogram of cross-correlation values plotted in Fig. 3, showing that the majority of the array is clustered at high cross-correlation values, indicating a good overall phase fit, with a long tail of poor cross-correlations, including some with reversed polarity. In order to assess the match in amplitudes, we took the log of the waveform envelopes, obtained using the Hilbert transform method, and compared their normalized RMSE (normalized by division by the interquartile range of the observed log-envelope

to provide a robust assessment of the scale of improvement irrespective of the original scale of amplitudes). The log-envelope metric gives a holistic view of the amplitude fit throughout the waveform, including low and high amplitude sections, with a RMSE ~ 0 under this metric indicating that the amplitude scale is matched throughout the waveform and an RMSE ~ 1 indicating amplitude errors on the scale of the interquartile range of the observations—the median RMSE for this amplitude metric was 0.61, and the histogram of RMSE values across the array is also plotted in Fig. 3.

To look at the amplitude fits in greater detail, we performed a time–frequency analysis by calculating the continuous wavelet transform (CWT) of the observed and predicted DAS strain rate signals using a Morelet wavelet, and then used these to calculate the RMS amplitude of both during the *P* wave (16–21 s) and *S* wave (36–41 s), for three frequency bands (1.0–1.5, 1.5–2.0 and 2.0–2.5 Hz, and also the full 1.0–2.5 Hz), with the results plotted in terms of log–amplitude ratios in Fig. 4. We observe spatially coherent patterns of mispredicted amplitudes across both *P* and *S* phases and across

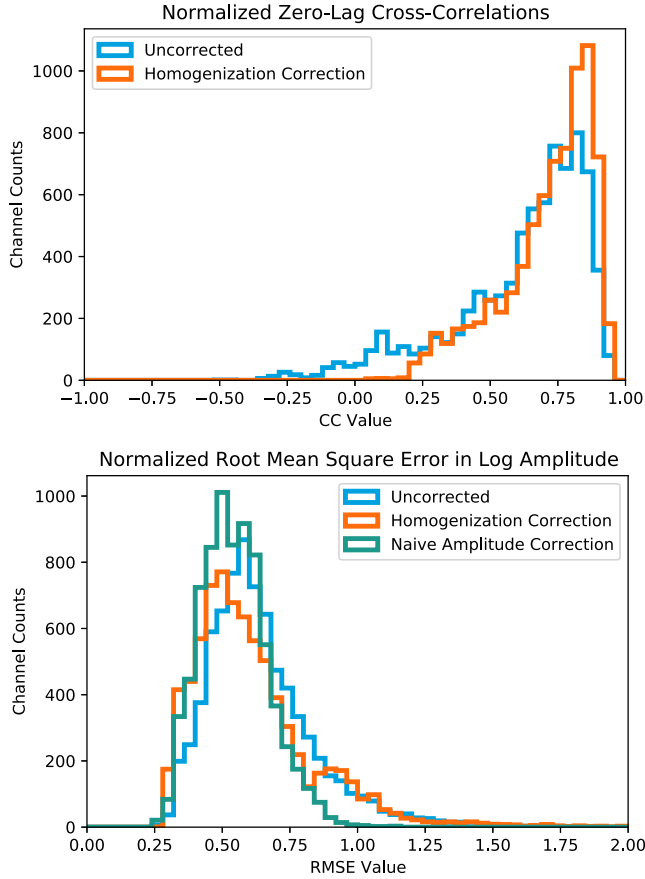


Figure 3. Histograms of the normalized zero-lag cross-correlation (after clock-correction) and normalized RMSE in log amplitudes between the observed DAS strain rates and strain rates predicted from the reconstructed nodal-seismic data, both before and after empirical correction based on two-scale homogenization theory.

frequency bands. The most obvious features are the two red areas of overpredicted amplitude in the lower left and upper right of the array. These areas are coincident with quaternary alluvial surface geology and particularly low near-surface V_P , as reported by Parker *et al.* (2018), which together with the large spatial scale of the amplitude response features and the fact that the PoroTomo cable was buried directly in a fresh trench (Feigl 2017) suggests that the geological conditions encountered by the DAS cable are substantively responsible for the differences between observed and predicted waveforms. In particular there are significant areas in which the observed amplitude of the DAS waveforms is substantially less than what one would expect based on the data recorded on the nodes. This implies that future studies that utilize DAS for amplitude-critical applications, such as earthquake early-warning, must carefully evaluate the fidelity of DAS amplitudes to expected values corresponding to the true displacement ground motions. Observing that the features in Fig. 4 are relatively independent of frequency band and wave type, a naive correction method for amplitude response is to simply take the median amplitude ratio from the six rightmost panels and multiply the predicted waveforms; doing this improves the median RMSE log-amplitude fit from 0.61 to 0.54, with the histogram of corrected values also shown in Fig. 3 showing significant improvement. We note that the two definitions of amplitude fit (RMS ratio for the P - and S -wave first arrivals vs log-amplitude RMSE for the entire waveform) are quite different so that this naive correction is not

just simply curve-fitting. Of course, the naive correction does not improve the normalized cross-correlation values whatsoever, as it does not affect the shape of the waveforms, and as such we are motivated to develop more detailed correction methods that may result in further improvements. Given our previous observation that the spatial distribution of amplitude responses suggest that the strains observed by the DAS for the PoroTomo experiment are affected by near-surface geology, we will develop an approach that models the DAS waveforms as a perturbation of the predicted long-wavelength strain wavefield using homogenization theory.

3 TOWARDS CORRECTING DAS AMPLITUDE RESPONSE USING TWO-SCALE HOMOGENIZATION THEORY

Homogenization theory (Capdeville *et al.* 2010a, b) gives a framework for understanding the differences between observed strain rates and those calculated from the reconstructed velocity field as observed by the nodes. The tutorial study by Capdeville *et al.* (2020) shows the potential issue starkly—material property distributions which produce smooth displacement waveforms and stress fields may produce counterintuitively complicated strain fields, which will complicate the study of the resulting DAS measurements which sense the integrated strain along a gauge length. Singh *et al.* (2019) studied means to correct for this effect, using the theory of two-scale homogenization, and applied it to rotation measurements recorded using the G-ring laser at Wetzell, Germany. They found substantial improvements could be made to the predicted rotation measurements calculated using PREM by making an additional first-order correction, with coefficients trained using observed rotation seismograms. A similar procedure may be used to obtain corrections for DAS seismograms, which we will outline here. For a full description of the theory, we refer the reader to the extensive development presented in Capdeville *et al.* (2020). We begin with the first-order expansion of the true displacement gradients $\nabla \mathbf{u}(x, y, t)$ in terms of reference displacements $\mathbf{u}^0(x, t)$ and strains $\epsilon^0(x, t)$, where x is the ‘large scale’ space variable and y is the ‘microscale’ space variable:

$$\nabla \mathbf{u}(x, y, t) = \nabla_x \mathbf{u}^0(x, t) + (\nabla_y \chi(y)) : \epsilon^0(x, t) \quad (3)$$

The reference displacements and strains are calculated using suitably averaged properties, while the term $\chi(y)$ is the first-order correction operator, which is a degree-three tensor. Equivalently, in Einstein-summation notation, and for notational convenience dropping the dependence of the fields on x, y and t we have

$$u_{i,j} = u_{i,j}^0 + \chi_{inm,jy} \epsilon_{nm}^0. \quad (4)$$

Inserting this relationship into eq. (1) and assuming that we have no coupling between the vertical and horizontal components of the wavefield, we have

$$\begin{aligned} \epsilon_c = & \sin^2 \alpha (u_{1,1x}^0 + \chi_{1nm,1y} \epsilon_{nm}^0) + \frac{\sin 2\alpha}{2} (u_{1,2x}^0 + u_{2,1x}^0 \\ & + (\chi_{1nm,2y} + \chi_{2nm,1y}) \epsilon_{nm}^0) \\ & + \cos^2 \alpha (u_{2,2x}^0 + \chi_{2nm,2y} \epsilon_{nm}^0), \end{aligned} \quad (5)$$

where the numerical indices are not summed over and represent the two orthogonal horizontal components and α is the clockwise azimuth relative to component 2. Averaging this equation along the

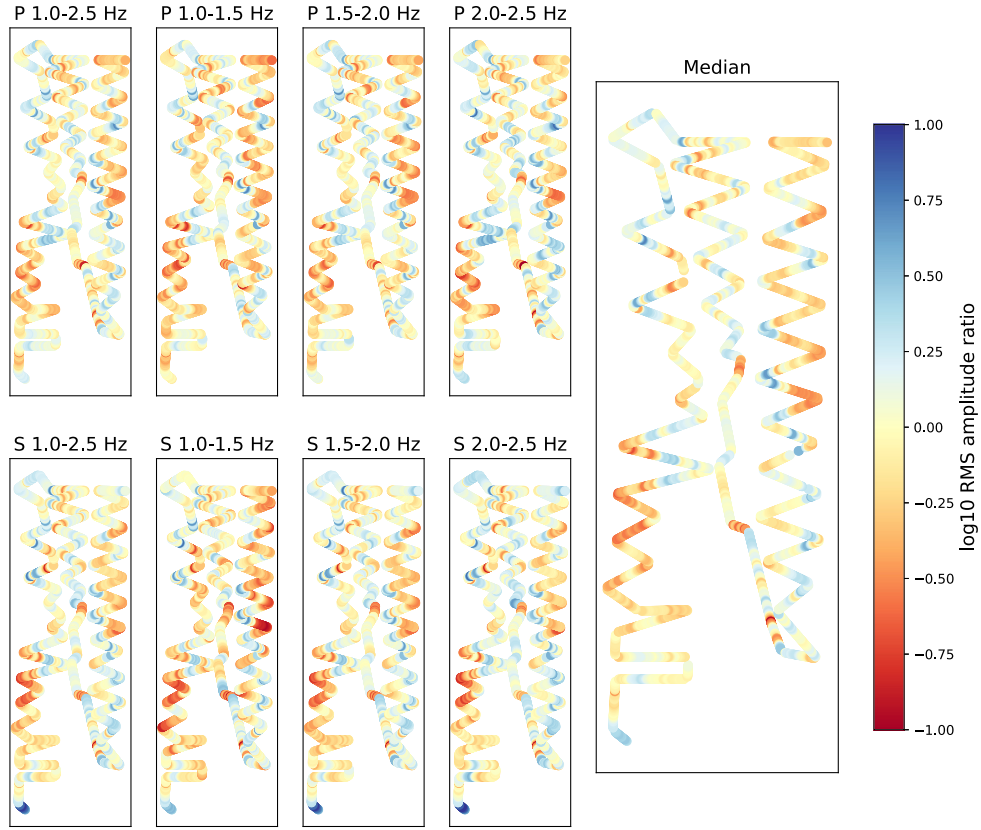


Figure 4. RMS amplitude ratios between the observed DAS data and the DAS strain rates predicted from the nodal data reconstruction, for the *P* (upper row) and *S* (lower row) energy packets of the *ML* 4.3 Hawthorn NV earthquake, both for the full 1–2.5 Hz range of the reconstruction and in 1.0–1.5, 1.5–2.0 and 2.0–2.5 Hz bands. The pattern of ratios shows clear spatial patterns that cut across frequency bands, with the most coherent features being the two overpredicted red patches in the lower left and upper right of the array, which is apparent in the median amplification across the six time-frequency bands.

DAS cable over a gauge length L gives us the DAS strain

$$\epsilon_{\text{DAS}} = \epsilon_{\text{DAS}}^0 + \frac{1}{L} \int_{-L/2}^{L/2} \sin^2 \alpha \chi_{1nm,1y}(l) \epsilon_{nm}^0(l) + \frac{\sin 2\alpha}{2} (\chi_{1nm,2y}(l) + \chi_{2nm,1y}(l)) \epsilon_{nm}^0(l) dl, \quad (6)$$

where we have assumed a straight cable about the reference location at $l = 0$. If we further assume that we are in a regime such that the reference strains ϵ_{nm}^0 are approximately constant over a gauge length, which we would expect to be true in the regime where displacements are well modelled and which allows the scale separation required by homogenization theory to be valid, then we can write

$$\epsilon_{\text{DAS}} \approx \epsilon_{\text{DAS}}^0 + J_{11} \epsilon_{11}^0 + J_{12} \epsilon_{12}^0 + J_{22} \epsilon_{22}^0, \quad (7)$$

where

$$J_{11} = \frac{1}{L} \int_{-L/2}^{L/2} \sin^2 \alpha \chi_{111,1y}(l) + \frac{\sin 2\alpha}{2} (\chi_{211,1y}(l) + \chi_{111,2y}(l)) + \cos^2 \alpha \chi_{211,2y}(l) dl, \quad (8)$$

etc. The coupling coefficients J are time independent and thus constant both within a particular recorded strain-rate time-series, and between records. This analysis depends on assumptions that seem well justified (scale separation allows us to remove the reference strains from the gauge integrand), as well as those that are potentially less so (lack of coupling between vertical and horizontal strains), but does serve as a starting point for the correction of DAS records to local heterogeneity using the theory of two-scale homogenization. The coefficients J can be obtained for each DAS channel

by computing

$$\mathbf{J} = (\mathbf{E}^{0T} \mathbf{E}^0)^{-1} \mathbf{E}^{0T} \delta \quad (9)$$

where for signal sample times $t = t_1, t_2, \dots, t_n$

$$\mathbf{J} = [J_{11} \quad J_{12} \quad J_{22}]^T, \quad (10)$$

$$\mathbf{E}^0 = \begin{bmatrix} \epsilon_{11}^0(t_1) & \epsilon_{12}^0(t_1) & \epsilon_{22}^0(t_1) \\ \epsilon_{11}^0(t_2) & \epsilon_{12}^0(t_2) & \epsilon_{22}^0(t_2) \\ \vdots & \vdots & \vdots \\ \epsilon_{11}^0(t_n) & \epsilon_{12}^0(t_n) & \epsilon_{22}^0(t_n) \end{bmatrix}, \quad (11)$$

$$\delta = \begin{bmatrix} \epsilon_{\text{DAS}}(t_1) - \epsilon_{\text{DAS}}^0(t_1) \\ \epsilon_{\text{DAS}}(t_2) - \epsilon_{\text{DAS}}^0(t_2) \\ \vdots \\ \epsilon_{\text{DAS}}(t_n) - \epsilon_{\text{DAS}}^0(t_n) \end{bmatrix}. \quad (12)$$

When multiple records are used for the inversion, the different t_i need not be from the same record as J is time independent, assuming that the local heterogeneity is not changing. While the PoroTomo experiment only recorded one significant event, which is insufficient to perform a robust analysis similar to Singh *et al.* (2019), we can solve for J by using the nodal reconstructed wavefield of the Hawthorn NV event to compute the reference strain rate and see what sort of improvements can be made using the assumptions developed in this section. It is worthwhile to note that the DAS records used here have significantly more high-frequency detail than the long-period rotation measurements in Singh *et al.* (2019), so the danger of overfitting is lessened. In Fig. 5, we show the corrected

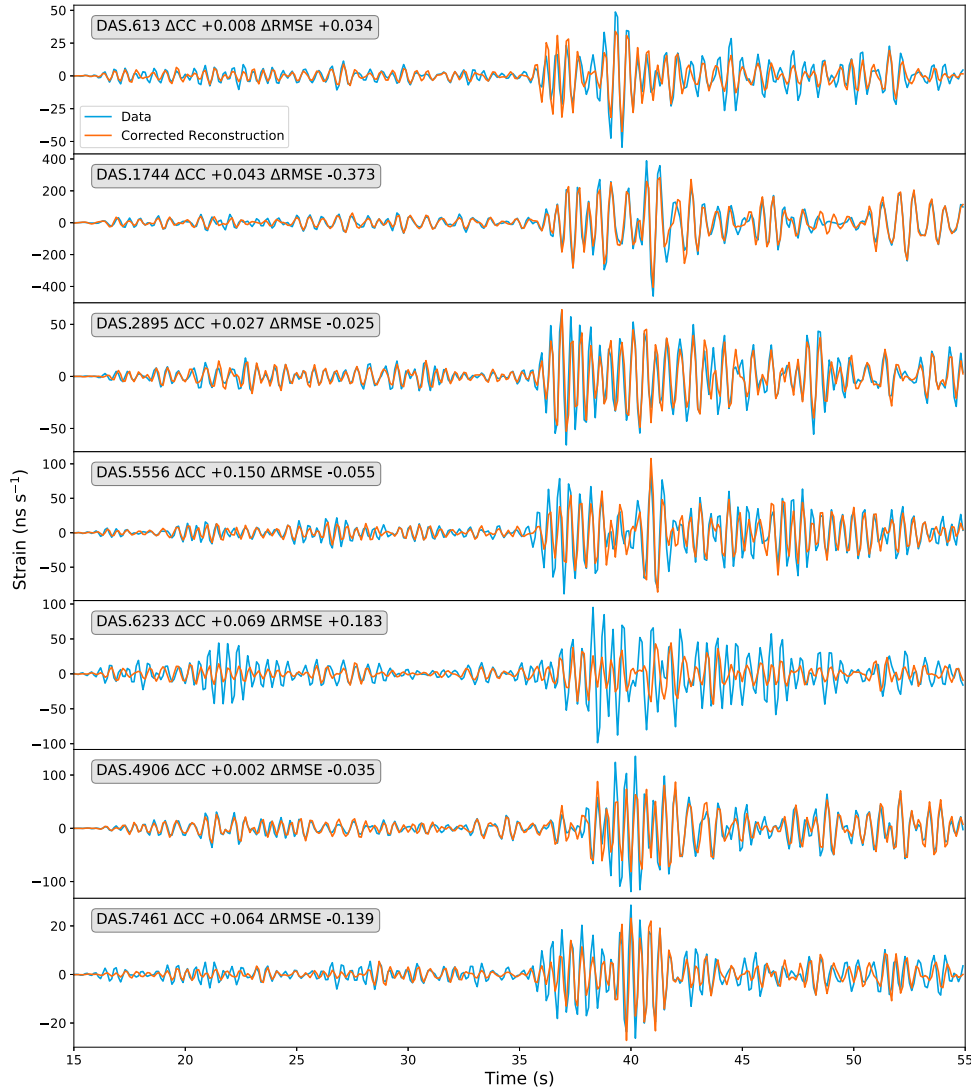


Figure 5. Comparisons between DAS channels (blue) and reconstructed corrected strain rate from nodes (orange) at the seven stations highlighted in Fig. 1. Corrections are obtained from applying a first-order expansion of the observed DAS signal in terms of the predicted DAS signal and predicted strain rates, as described in the text. Changes to the observed zero-lag normalized cross-correlation ΔCC and log-amplitude error $\Delta RMSE$ show substantially improved fits compared to Fig. 2 in almost all cases.

strain rates calculated from the nodal reconstructed wavefield. In five out of seven of the example waveforms, we see a significant improvement in this amplitude metric. Because of the requirement for clock-corrections, these error metrics are calculated using the time range of 15–55 s. Averaged over the entire DAS array, we saw the median normalized zero-lag cross-correlation increase from 0.69 to 0.76, and the median normalized RMSE in log amplitude decrease from 0.61 to 0.56—the average gain in waveform shape accuracy is significant, with the tail of negative zero-lag cross-correlations eliminated completely. However, the improvement in RMSE amplitude fit is not as good as the naive amplitude correction. Histograms showing the distributions of these values across channels are shown in Fig. 3, with the histogram of homogenized RMSE showing a secondary hump that may be due to the prioritization of the L2 metric of fitting phase over amplitude, which can result in the ‘best-fit’ corrected amplitude (in the L2 sense) having very small amplitude. Further advancements of homogenization-theory based corrections may allow for more accurate matching of recorded DAS strains to predicted waveforms, which may substantially improve the efficacy

of DAS arrays for studying source processes. Additionally, there is potentially scope for further inversion of the correction term J for small-scale structure, although the requisite theory has yet to be developed.

4 OPTIMAL DESIGN OF MIXED NETWORKS

In this study, we have so far explored the potential of wavefield reconstruction to characterize a combined DAS and point-seismometer network, with the view that such deployments will become increasingly common in the future. Given that new arrays offer new opportunities for deployment design, a natural extension of our study of wavefield reconstruction is how to best design such mixed networks to maximize their reconstruction performance and ability to correct for DAS gain, and in particular how we might use our mathematical framework to formalize the array design. Within the general framework of Muir & Zhan (2021), the final, spatially resolved step of the wavefield analysis is a linear inverse problem,

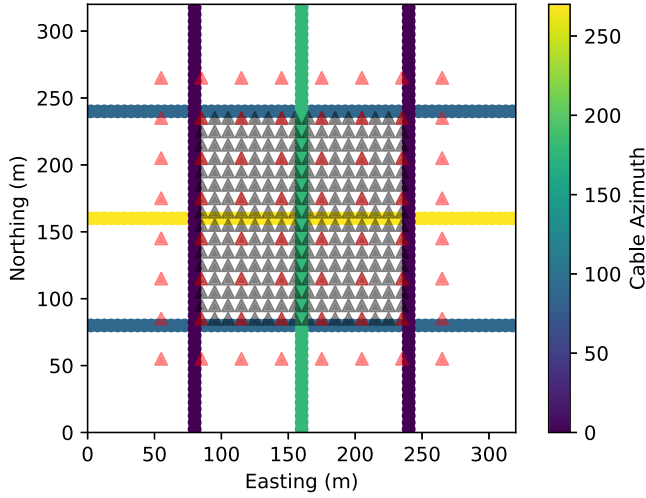


Figure 6. Network geometry used for synthetic optimal network design. DAS cable channel locations are shown as circles coloured by cable azimuth. Red triangles show the locations of candidate stations, and grey triangles show the locations of the validation sensors placed within the central region where we should expect to be able to achieve good recovery of the wavefield.

utilizing an L1 regularization to promote sparse solutions. Procedures for how to make optimal measurements for linear inverse problems have been widely studied for over a century, and have been collectively termed methods for Optimal Experimental Design (OED). While OED is a well-established concept, methods for OED within the context of compressive sensing have only been recently developed. In the following, we use an algorithm proposed in Ravi *et al.* (2017) to determine optimal mixed network designs and evaluate their performance for improving wavefield reconstruction relative to unoptimized networks.

Following Ravi *et al.* (2017), the information matrix $\mathbf{M}'$ is defined by the product $\mathbf{G}^T \mathbf{G}$. For compressive sensing problems, $\mathbf{M}'$ is not full rank, so is regularized by $\mathbf{M} = \mathbf{M}' + \epsilon \mathbf{I}$. Using the D-optimal design criterion, which minimizes the confidence ellipsoid of the estimate for the inverted curvelet coefficients, good designs will minimize $\log \det \mathbf{M}^{-1}$. We introduce a selection vector $\mathbf{s}$, such that each element $s_i \in \{0, 1\}$. We can write the matrix $\mathbf{M}^* = \sum_i s_i \mathbf{g}_i^T \mathbf{g}_i + \epsilon \mathbf{I}$ where $\mathbf{g}_i$ are the row vectors of $\mathbf{G}$ —each row vector corresponds to a candidate measurement location. The optimal design problem asks, given a budget of B stations, what the best $\mathbf{s}$ with $\text{sum}(\mathbf{s}) \leq B$ is, such that the solution of the compressive sensing problem $\|\text{Diag}(\mathbf{s})(\mathbf{G}\mathbf{m} - \mathbf{d})\|_2 + \|\mathbf{m}\|_1$ is similar to that of the unrestricted problem $\|\mathbf{G}\mathbf{m} - \mathbf{d}\|_2 + \|\mathbf{m}\|_1$.

Ravi *et al.* (2017) proposed a design criterion for compressive-sensing problems based on combining D-optimal design with an additional term designed to minimize the coherence of the selected stations. The coherence is defined as the maximum diagonal term of the hat matrix $\mathbf{H}^* = \text{Diag}(\mathbf{s})\mathbf{G}\mathbf{M}^{*-1}\mathbf{G}^T\text{Diag}(\mathbf{s})$ (this criterion is also known in the literature as G-optimal design). The full objective function to be minimized is therefore given by

$$f(\mathbf{s}) = -\log \det(\mathbf{M}^*) + \lambda \max [\text{Diag}(\mathbf{H}^*)], \quad (13)$$

for some regularization parameter $\lambda > 0$, and where the $\max$ is the maximum element of the diagonal of the hat matrix $\mathbf{H}^*$. Minimizing coherence tends to produce better solutions for compressive sensing, so combining these two criteria promotes a balance of solutions that allows for good sparse signal recovery (from the incoherence condition) while also maintaining overall sensitivity (from the D-optimal condition). Solving this problem for binary $\mathbf{s}$

is an NP-hard nonlinear mixed-integer problem and is thus generally intractable as a polynomial time solution is not known, leading to two solution strategies. The first is to obtain an approximate solution via sequential optimization by finding the minimum of f for an initial station, then fixing that station and subsequently optimizing for the second station, etc. This sequential approach is computationally tractable and will often give a good solution but is not guaranteed to find the optima of $f(\mathbf{s})$ —this approach has been commonly used for geophysical optimal design problems (e.g. see Bloem *et al.* 2020, for a recent overview including fully nonlinear design principles for small networks) and has the additional advantage that the same scheme can be also used to study the variant in which the candidate locations are allowed to vary, although we will not consider that possibility in this study. The second approach is to work with a relaxation $0 \leq s \leq 1$, and then minimize $f(\mathbf{s})$ simultaneously for all stations; this approach was advocated by Ravi *et al.* (2017). The resulting relaxed weights give some indication of favourable designs; for instance they could be used as probabilistic weights for random network designs, or further schemes can be used to round the relaxed weights into a binary solution. We implemented the relaxed scheme using the JuMP interface (Dunning *et al.* 2017) to *ipopt* (Wächter & Biegler 2006), which utilizes a log-barrier term to relax the constraints into the objective function.

For our mixed network problem, we have $\mathbf{s} = [\mathbf{s}_{\text{DAS}}; \mathbf{s}_x; \mathbf{s}_y]$. Typically DAS networks will either use existing dark-fibre instrumentation, or if using a greenfield deployment, will require complex constraints on the geometry due the required continuity of the cable, attempts to minimize corners, obstacles etc.—consequently, we will restrict our discussion to the case that the DAS cable geometry is known and we wish to optimize the locations of deployed point seismometers to supplement the cable. In that case, we have the additional restriction that $\mathbf{s}_x = \mathbf{s}_y = \mathbf{s}_{\text{Seismometers}}$, $\mathbf{s}_{\text{DAS}} = \mathbf{1}$, and the restriction on the number of sensors is then $\text{sum}(\mathbf{s}_{\text{Seismometers}}) \leq B_{\text{Seismometers}}$. With these changes, the problem statement conforms to that in Ravi *et al.* (2017), and can be solved as described above.

To test both the sequential and relaxed methods of network design, we utilize a square synthetic array consisting of 384 DAS channels arranged in a cross-hatch formation, with 5-m channel spacing and 10-m gauge length, along with 64 candidate two-component horizontal seismometer locations evenly spaced over a central square. The geometry of the synthetic setup is shown in Fig. 6. To determine reconstruction performance, we generated a synthetic Rayleigh-wave propagating across the array by simulating the acoustic wave equation with a point source for a weakly varying checkerboard velocity model (average velocity $\sim 1810 \pm 90 \text{ m s}^{-1}$), and then taking the gradient to determine horizontal motions. We calculate the synthetic traces at each DAS channel and candidate seismometer location, including 10 per cent added Gaussian noise, and also at 256 evenly spaced validation locations within the central square of the DAS array where we expect to be able to achieve adequate reconstruction results. All performance metrics reported are given by the per-trace normalized RMS misfit of the reconstructed wavefield at the validation stations.

For the relaxed analysis, we drew 10 lots of 1000 designs by sampling without replacement using the weights obtained by optimizing eq. (13). We set they hyperparameters to $\lambda = 1$ and $\epsilon = 10^{-2}$, with the latter motivated by the cutoff of the eigenvalue spectrum of the design matrix for a DAS array of totally random sensor locations and azimuths, which we would expect to have the best potential reconstruction performance—although that design is certainly not

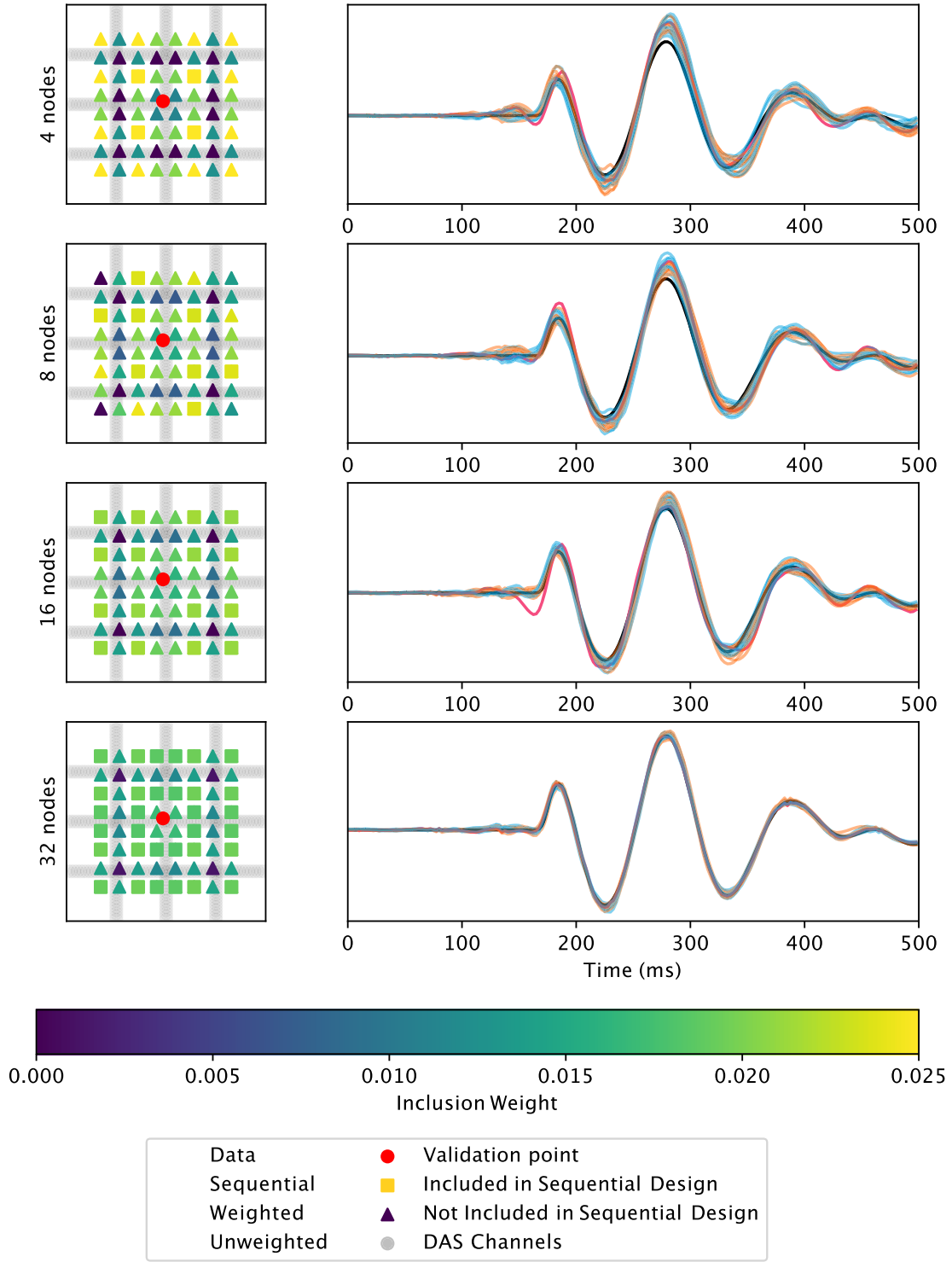


Figure 7. Array designs and reconstruction performance for different numbers of included nodes. The colours of the left-hand column shows the inclusion weight for the relaxed OED problem. The shapes show the sequential insertion design, with square symbols showing included stations and triangular symbols showing non-included stations. The right-hand column shows the evaluated reconstruction for the x component of the validation station at location (155,165), near the centre of the array, with the black line showing the true data, the pink line the reconstruction for the sequential insertion design, blue lines showing unweighted random designs and orange lines showing weighted random designs.

achievable in practice. The choice of λ was then made to give the incoherence term a similar scale to the typical range of perturbations to the D-optimal design term in eq. (13). For each of the 10 lots, we then chose the design with the lowest value of eq. (13) amongst the

1000 candidates, and performed a wavefield reconstruction on the data. We also performed a wavefield reconstruction on 10 designs chosen by unweighted random selection without replacement. We tested designs with 4, 8, 16, 32 and 64 stations (with 64 being the

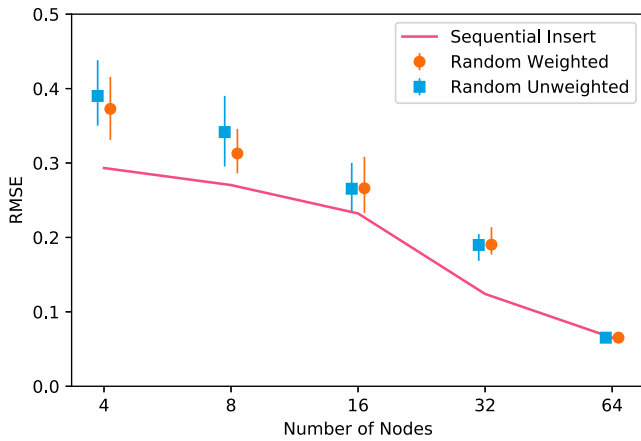


Figure 8. Per-trace normalized reconstruction RMSE calculated using the 256 validation stations shown in Fig. 6 as a function of number of utilized nodes in the design. The pink line shows the sequentially optimized design, while blue squares and orange circles show results from unweighted and weighted randomized designs respectively. The error bars show the maximum and minimum RMSE across the 10 samples used for each randomized design.

complete set of potential locations, so that only 1 design was in fact tested). We also performed wavefield reconstructions for the sequential optima for those numbers of stations. Fig. 7 shows the relaxed design weights, sequentially optimized design and example resultant fits to the true synthetic waveform for the tested designs up to 32 included nodes. Overall performance at the example station is quite good, and unsurprisingly gets better as the number of included stations increases. The weights generally prefer stations that are further away from the DAS cable, particularly the outer crossing, which is somewhat intuitive as those areas already have provided information; it is however interesting to note that the center of the array is always weighted at least moderately despite the presence of a DAS crossing, presumably as the strong improvement to the D-optimal design term from occupying the center location outweighs the coherence penalty there. Fig. 8 shows the per-trace-normalized RMSE for this synthetic experiment as a function of the number of included nodes. In this case, despite there being no guarantee that the sequential design is particularly performant, it consistently has significantly lower RMSE than both the weighted and unweighted random designs, potentially as the regular, dense DAS array design makes choosing the locations of the next element of the sequence simple. For low node density, the weighted design slightly outperforms the unweighted design on average, while it confers no advantage once one quarter of the nodes are included.

5 CONCLUSIONS

DAS networks promise a paradigm shift within observational seismology by making large- N , highly spatially dense networks financially and logistically feasible for the research community. However, the strain based measurement procedure is highly susceptible to the effects of local heterogeneity and is furthermore sensitive to only one projection of the horizontal strain tensor. Using a wavelet-curvelet compressive sensing based wavefield reconstruction, we have developed a method for simultaneous spatial assessment of DAS array response using a theory that does not require potentially unstable f_k rescaling, when a colocated but potentially off-cable point sensor array is available. We have shown that the DAS phase

response in the 1–2.5 Hz range is relatively accurate to true ground motion for the ML 4.3 Hawthorne, NV earthquake recorded on the Porotomo DAS array, but that there are however significant amplitude errors that correlate with near-surface geology. These amplitude errors can be well characterized by a single empirical gain coefficient within the studied frequency band, and further detailed studies using the framework of homogenization theory may yield corrections that improve both phase and amplitude response. Furthermore, we have shown that wavefield reconstruction permits a cohesive framework for combining DAS with point sensors such as 3-D nodes, with the combined DAS and nodes network outperforming the nodal network by itself for low node density using the PoroTomo array geometry. With a view to generalizing the applicability of these results beyond the PoroTomo deployment, we have investigated an optimal design strategy for improving the efficiency of mixed DAS and point sensor deployments for wavefield reconstruction. While at this stage we have focused on the relatively simpler task of deciding the best locations for point sensors given a fixed DAS deployment, as this is computationally more tractable and also corresponds to the common case of utilizing pre-existing dark fibre deployments, the general concept of optimizing network sensitivity while also promoting the incoherence of measurements, will allow for optimal design of general DAS networks in the future. As DAS becomes increasingly prevalent and further integrated with traditional seismic networks, wavefield reconstruction therefore represents a flexible framework for overcoming the weaknesses of DAS as a single component large- N array, instead optimizing its strength as a single unified areal sensor.

ACKNOWLEDGEMENTS

The authors would like to thank the editor, Andrew Valentine, and the reviewers, Eileen Martin and Karen Lythgoe, for their comments, which have greatly improved the manuscript. The authors would like to thank Ettore Biondi for useful discussions during the preparation of this manuscript. The authors would like to acknowledge the University of Wisconsin-Madison PoroTomo team for early access and usage instructions for the PoroTomo DAS data set. JBM acknowledges the financial support of the Origin Energy Foundation and the General Sir John Monash Foundation during his PhD studies. ZZ thanks the support from NSF CAREER award EAR 1848166 and the NSF/IUCRC GMG Center. Seismic data were processed using Obspy (Beyreuther *et al.* 2010), and figures were created using Matplotlib (Hunter 2007).

DATA AVAILABILITY

Data for the Hawthorn NV earthquake can be found at Feigl (2016a, b). The code used to produce the analysis for this study is stored at <https://github.com/jbmuir/DAS-Reconstruction>.

REFERENCES

- Bakku, S.K., 2015. *Fracture Characterization from Seismic Measurements in a Borehole*, PhD thesis, Massachusetts Institute of Technology.
- 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.
- Bloem, H., Curtis, A. & Maurer, H., 2020. Experimental design for fully nonlinear source location problems: which method should I choose?, *Geophys. J. Int.*, **223**(2), 944–958.

- Burnham, K.P. & Anderson, D.R., 2004. Multimodel inference: understanding AIC and BIC in model selection, *Sociol. Methods Res.*, **33**(2), 261–304.
- Candes, E.J. & Demanet, L., 2005. The curvelet representation of wave propagators is optimally sparse, *Commun. Pure Appl. Math.*, **58**(11), 1472–1528.
- Capdeville, Y., Guillot, L. & Marigo, J.-J., 2010a. 1-D non-periodic homogenization for the seismic wave equation, *Geophys. J. Int.*, **181**(2), 897–910.
- Capdeville, Y., Guillot, L. & Marigo, J.-J., 2010b. 2-D non-periodic homogenization to upscale elastic media for P-SV waves: 2-D non-periodic homogenization, PSV case, *Geophys. J. Int.*, **182**(2), 903–922.
- Capdeville, Y., Cupillard, P. & Singh, S., 2020. An introduction to the two-scale homogenization method for seismology, in *Advances in Geophysics*, Vol. **61**, pp. 217–306, Elsevier.
- Dunning, I., Huchette, J. & Lubin, M., 2017. JuMP: a modeling language for mathematical optimization, *SIAM Rev.*, **59**(2), 295–320.
- Feigl, K., 2016a. Brady's Geothermal Field DAS Earthquake Data.,
- Feigl, K., 2016b. Brady's Geothermal Field Nodal Seismometer Earthquake Data.,
- Feigl, K.L., 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 Workshop on Geothermal Reservoir Engineering*, Stanford University.
- Feigl, K.L. & Parker, L.M., 2019. *PoroTomo Final Technical Report: Poroelastic Tomography by Adjoint Inverse Modeling of Data from Seismology, Geodesy, and Hydrology*, Tech. Rep. 3.1, pp. 1499141.
- Ferguson, R.J., McDonald, M.A. & Basto, D.J., 2020. Take the Eh? train: Distributed Acoustic Sensing (DAS) of commuter trains in a Canadian City, *J. Appl. Geophys.*, **183**, 104201.
- Friedman, J., Hastie, T. & Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent, *J. Stat. Softw.*, **33**(1), 1–22.
- Hunter, J.D., 2007. Matplotlib: a 2D graphics environment, *Comput. Sci. Eng.*, **9**(3), 90–95.
- Ide, S., Araki, E. & Matsumoto, H., 2021. Very broadband strain-rate measurements along a submarine fiber-optic cable off Cape Muroto, Nankai subduction zone, Japan, *Earth Planets Space*, **73**(1), 1–10.
- Jia, Z. & Clayton, R.W., 2021. Determination of near surface shear-wave velocities in the Central Los Angeles Basin with dense arrays, *J. geophys. Res.: Solid Earth*, **126**(5), 1–14.
- Karrenbach, M. *et al.*, 2020. Rapid deployment of distributed acoustic sensing systems to track earthquake activity, in *SEG Technical Program Expanded Abstracts 2020*, SEG Technical Program Expanded Abstracts, pp. 490–494, Society of Exploration Geophysicists.
- King, G., Zurn, W., Evans, R. & Emter, D., 1976. Site correction for long period seismometers, tiltmeters and strainmeters, *Geophys. J. Int.*, **44**(2), 405–411.
- Lin, F.-C., Li, D., Clayton, R.W. & Hollis, D., 2013. High-resolution 3D shallow crustal structure in Long Beach, California: application of ambient noise tomography on a dense seismic array, *Geophysics*, **78**(4), Q45–Q56.
- Lindsey, N.J., Martin, E.R., Dreger, D.S., Freifeld, B., Cole, S., James, S.R., Biondi, B.L. & Ajo-Franklin, J.B., 2017. Fiber-optic network observations of earthquake wavefields, *Geophys. Res. Lett.*, **44**(23), 11792–11799.
- 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.: Solid Earth*, **125**(2), 1–16.
- Martin, E.R., Lindsey, N., Ajo-Franklin, J. & Biondi, B., 2018. *Introduction to Interferometry of Fiber Optic Strain Measurements*, preprint, Earth-ArXiv.
- Matsumoto, H. *et al.*, 2021. Detection of hydroacoustic signals on a fiber-optic submarine cable, *Sci. Rep.*, **11**(1), 1–12.
- Muir, J.B. & Zhan, Z., 2021. Seismic wavefield reconstruction using a pre-conditioned wavelet – curvelet compressive sensing approach, *Geophys. J. Int.*, **227**(1), 303–315.
- Nayak, A. & Ajo-Franklin, J., The Imperial Valley Dark Fiber Team 2021. Measurement of surface-wave phase-velocity dispersion on mixed inertial seismometer – distributed acoustic sensing seismic noise cross-correlations, *Bull. seism. Soc. Am.*.
- Parker, L.M. *et al.*, 2018. Active-source seismic tomography at the brady geothermal field, Nevada, with dense nodal and fiber-optic seismic arrays, *Seismol. Res. Lett.*, **89**(5), 1629–1640.
- Ravi, S.N., Ithapu, V.K., Johnson, S.C. & Singh, V., 2017. Experimental design on a budget for sparse linear models and applications, in *Proceedings of the 33rd International Conference on Machine Learning*, Vol. **48**, pp. 583–592, New York.
- Ringler, A.T., Anthony, R.E., Karplus, M.S., Holland, A.A. & Wilson, D.C., 2018. Laboratory tests of three Z-land fairfield nodal 5-Hz, three-component sensors, *Seismol. Res. Lett.*, **89**(5), 1601–1608.
- Ringler, A.T., Wilson, D.C., Zürn, W. & Anthony, R.E., 2019. Rayleigh wave ellipticity measurement uncertainty across the IRIS/USGS and New China Digital Seismograph Networks, *Geophys. J. Int.*, **217**(1), 219–237.
- Singh, S., Capdeville, Y. & Igel, H., 2020. Correcting wavefield gradients for the effects of local small-scale heterogeneities, *Geophys. J. Int.*, 996–1011.
- Sladen, A., Rivet, D., Ampuero, J.P., De Barros, L., Hello, Y., Calbris, G. & Lamare, P., 2019. Distributed sensing of earthquakes and ocean-solid Earth interactions on seafloor telecom cables, *Nat. Commun.*, **10**(1), 1–8.
- Spica, Z.J., Nishida, K., Akuhara, T., Pétrélis, F., Shinohara, M. & Yamada, T., 2020. Marine sediment characterized by ocean-bottom fiber-optic seismology, *Geophys. Res. Lett.*, **47**(16), 1–10.
- van den Ende, M. P.A. & Ampuero, J.-P., 2021. Evaluating seismic beam-forming capabilities of distributed acoustic sensing arrays, *Solid Earth*, **12**(4), 915–934.
- Wächter, A. & Biegler, L.T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, *Math. Program.*, **106**(1), 25–57.
- 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, *Geophys. J. Int.*, **213**(3), 2020–2036.
- Wang, X., Williams, E.F., Karrenbach, M., Herráez, M.G., Martins, H.F. & Zhan, Z., 2020. Rose parade seismology: signatures of floats and bands on optical fiber, *Seismol. Res. Lett.*, **91**(4), 2395–2398.
- Wiesmeyer, C., Litzemberger, M., Waser, M., Papp, A., Garn, H., Neunteufel, G. & Döller, H., 2020. Real-time train tracking from distributed acoustic sensing data, *Appl. Sci.*, **10**(2), 1–12.
- Williams, E.F., Fernández-Ruiz, M.R., Magalhaes, R., Vanthillo, R., Zhan, Z., González-Herráez, M. & Martins, H.F., 2019. Distributed sensing of microseisms and teleseisms with submarine dark fibers, *Nat. Commun.*, **10**(1), 1–11.
- Yu, C., Zhan, Z., Lindsey, N.J., Ajo-Franklin, J.B. & Robertson, M., 2019. The potential of DAS in teleseismic studies: insights from the goldstone experiment, *Geophys. Res. Lett.*, **46**(3), 1320–1328.
- Zhan, Z., 2020. Distributed acoustic sensing turns fiber-optic cables into sensitive seismic antennas, *Seismol. Res. Lett.*, **91**(1), 1–15.

APPENDIX A: WAVEFIELD RECONSTRUCTION THEORY

In Muir & Zhan (2021), we proposed a wavelet decomposition in the time domain, coupled with a compressive-sensing based pre-conditioned curvelet decomposition in the spatial domain. Mathematically this involves a transform $\mathbf{w}(j_w, s) = WT(\mathbf{u}(t))$, operating trace-by-trace, from the collection of time-domain signals $\mathbf{u}$ to the wavelet domain indexed by scale j_w and time–position variable s , followed by the solution of an $L1$ regularized linear inverse problem $\mathbf{G}\hat{\mathbf{c}}(j_w, s) = \mathbf{w}(j_w, s)$ expressing the spatial behaviour of the wavelet coefficients $\mathbf{w}(j_w, s)$ in terms of the curvelet coefficients $\hat{\mathbf{c}}(j_w, s)$. Individual curvelet coefficients are indexed by a scale j_c , as well as rotation and translation indices. The design matrix

$\mathbf{G} = \mathbf{\Psi} \mathbf{\Phi} \mathbf{P}$ in our formulation includes both the spatial behaviour of the curvelet basis functions, $\mathbf{\Phi}$, the spatial sampling matrix $\mathbf{\Psi}$ and a variable preconditioning matrix $\mathbf{P} = \text{Diag}(2^{-j_c^n})$ that allows us to promote smooth wavefields by penalizing the scale factor j_c of the curvelets.

We now develop the theory further to include DAS measurements of the strain field. As indicated in eq.(2), DAS strain metrics are calculated by averaging the true strain or strain rate over a gauge length L . As such, we can invert for the horizontal wavefield curvelet coefficients by solving

$$\left[\frac{1}{L} \int_{-L/2}^{L/2} \mathbf{A}_1 \frac{\partial \mathbf{G}}{\partial x} + \mathbf{A}_2 \frac{\partial \mathbf{G}}{\partial y} dl \quad \frac{1}{L} \int_{-L/2}^{L/2} \mathbf{A}_2 \frac{\partial \mathbf{G}}{\partial x} + \mathbf{A}_3 \frac{\partial \mathbf{G}}{\partial y} dl \right] \begin{bmatrix} \mathbf{c}_x(j_w, s) \\ \mathbf{c}_y(j_w, s) \end{bmatrix} = \mathbf{w}_{\text{DAS}}(j_w, s) \quad (\text{A1})$$

$$\begin{bmatrix} \mathbf{G}_{\text{DAS},x} & \mathbf{G}_{\text{DAS},y} \end{bmatrix} \begin{bmatrix} \mathbf{c}_x(j_w, s) \\ \mathbf{c}_y(j_w, s) \end{bmatrix} = \mathbf{w}_{\text{DAS}}(j_w, s), \quad (\text{A2})$$

where $\mathbf{A}_1 = \text{Diag}(\sin^2 \alpha)$, $\mathbf{A}_2 = \text{Diag}(\frac{\sin 2\alpha}{2})$ and $\mathbf{A}_3 = \text{Diag}(\cos^2 \alpha)$. The integrals are applied to each row, each of which corresponds to a single measurement for a cable segment with gauge length L and cable azimuth α . This inversion process is unable to recover static components of the velocity field due to the derivatives involved, but given the normally oscillatory characteristics of solid earth ground motion, this is unlikely to be an issue in practice. Once the curvelet components are found, the ground velocity wavelet coefficients may be recovered by applying the undifferentiated curvelet basis matrix giving $\mathbf{w}_{u_x}(j_w, s) = \mathbf{G} \mathbf{c}_x(j_w, s)$ and $\mathbf{w}_{u_y}(j_w, s) = \mathbf{G} \mathbf{c}_y(j_w, s)$. The ground velocities in time domain can then be recovered by the inverse wavelet transform. In the reverse process, the horizontal components $\mathbf{u}_x$ and $\mathbf{u}_y$ of 3C seismic deployments may be inverted using an appropriate design matrix $\mathbf{G}$, and the above equations used to obtain the predicted DAS signal. Finally, we may simultaneously invert both 3C seismometers and DAS deployments by forming the block matrix equation

$$\begin{bmatrix} \mathbf{G}_{\text{DAS},x} & \mathbf{G}_{\text{DAS},y} \\ \mathbf{G}_{\text{Seismometers}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{c}_x(j_w, s) \\ \mathbf{c}_y(j_w, s) \end{bmatrix} = \begin{bmatrix} \mathbf{w}_{\text{DAS}}(j_w, s) \\ \mathbf{w}_{u_x}(j_w, s) \\ \mathbf{w}_{u_y}(j_w, s) \end{bmatrix}, \quad (\text{A3})$$

which in the compressive sensing framework is given by L1 regularization

$$\begin{bmatrix} \hat{\mathbf{c}}_x(j_w, s) \\ \hat{\mathbf{c}}_y(j_w, s) \end{bmatrix} = \arg \min \left[\left\| \begin{bmatrix} \mathbf{G}_{\text{DAS},x} & \mathbf{G}_{\text{DAS},y} \\ \mathbf{G}_{\text{Nodes}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{c}_x(j_w, s) \\ \mathbf{c}_y(j_w, s) \end{bmatrix} - \begin{bmatrix} \mathbf{w}_{\text{DAS}}(j_w, s) \\ \mathbf{w}_{u_x}(j_w, s) \\ \mathbf{w}_{u_y}(j_w, s) \end{bmatrix} \right\|_2 + \lambda \left\| \begin{bmatrix} \mathbf{c}_x(j_w, s) \\ \mathbf{c}_y(j_w, s) \end{bmatrix} \right\|_1 \right], \quad (\text{A4})$$

solved using the Lasso.jl Julia module (a fast reimplementation

of the R code glmnet, Friedman *et al.* 2010), with the regularization parameter λ determined automatically using the corrected Akaike information criterion or AICc (Burnham & Anderson 2004), or via fivefold cross-validation where computationally feasible.

APPENDIX B: RECONSTRUCTION OF NODAL DATA USING DAS

Having shown that the dense nodal deployment is able to well recover the DAS strain rate, we then performed the opposite procedure—recovery of both horizontal components of the wavefield using only the DAS cable. We note that this is a significantly more challenging task than the previous nodes-to-DAS experiment, as we can only access the horizontal wavefield through its derivative. In order to stabilize the selection of curvelets when only using DAS, we use mixed L1 and L2 regularization (i.e. Elastic Net Regression) of the form $\alpha \|\mathbf{c}\|_1 + (1 - \alpha) \|\mathbf{c}\|_2$ with $\alpha = 0.95$. We have found that this small additional amount of L2 regularization is not necessary when any nodal data is included. Fig. B1 shows the resulting horizontal components at three node locations, using timing lags optimized for the P wave on each instrument—individually calibrated time corrections were used for this figure as the overall DAS to nodal reconstruction is not of sufficient accuracy to calculate a global correction. In this case, the fit to the phases and amplitudes of the initial P wave is generally good, and performance declines throughout the P wave coda. Waveform fits improve on the East components during the initial S wave, but fits to the North component are generally poor. Sensitivity analysis following Martin *et al.* (2018) is shown in Fig. B2 assuming a near surface V_P of 1300 ms^{-1} and a Poisson solid relationship for V_S indicates that relative sensitivities for P , SV and SH motions are almost independent of frequency across the frequency band studied here. The cable geometry, near surface velocity structure and earthquake source location combine to give highest sensitivity to P , followed by SH , with little sensitivity to SV . The overall better fit of the P -wave data is therefore unsurprising, as the effective wavelength of this low-slowness phase is large across the array, and the relative sensitivity to P motions is large, allowing the CS inversion scheme to better integrate the P wavefield. In the case of the S wave, the higher amplitude of the East versus North nodal channels suggests that the SH component of the S wavepacket is dominant (given that the earthquake is south-southeast of the array). The higher sensitivity to SH , combined with higher SH component, explains the better fit to the East components of the motion at the nodes. In general, however, it is apparent that the DAS cable is by itself insufficient to recover the wavefield to a desirable accuracy within the frequency range of this study.

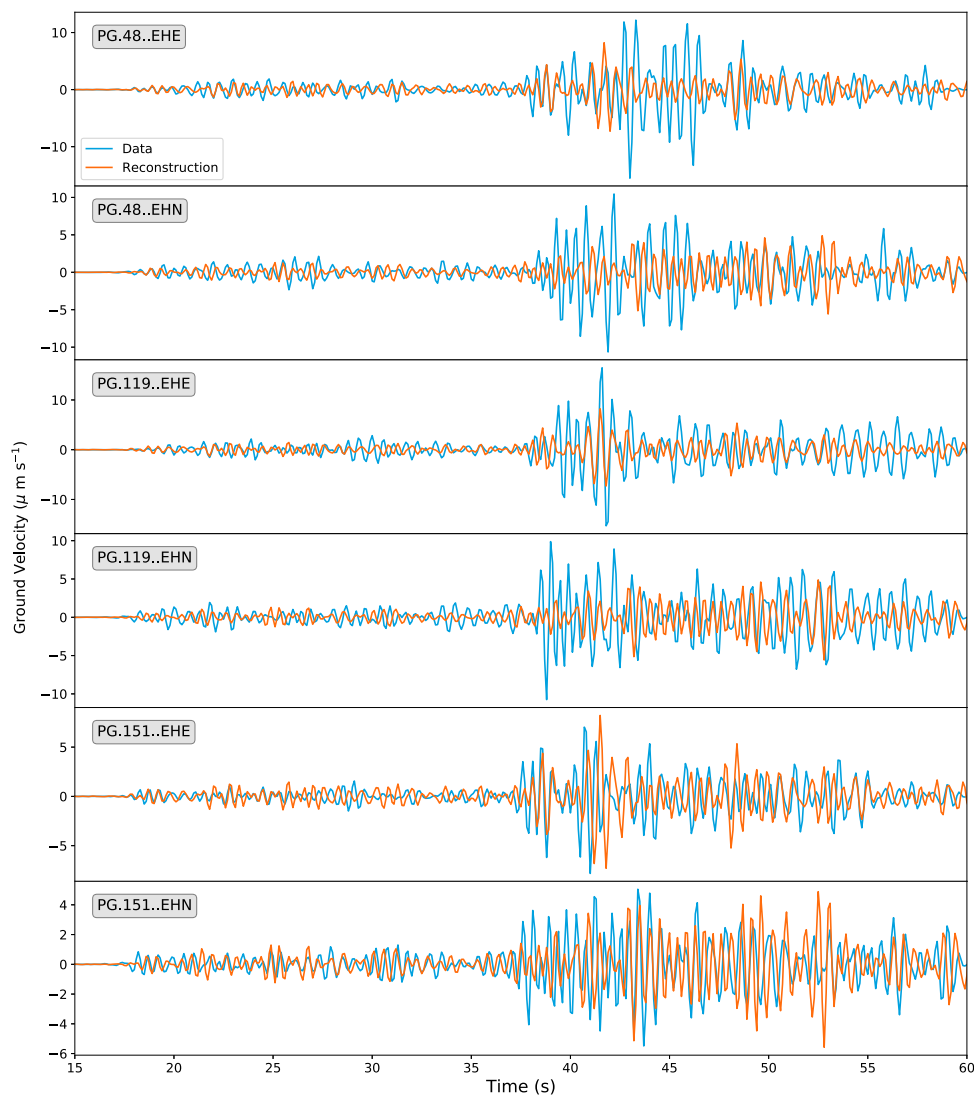


Figure B1. Comparisons between nodal seismic channels (blue) and reconstructed ground velocity from DAS (orange) at the three stations highlighted in Fig. 1. Waveforms are bandpass filtered at 1–2.5 Hz, and nodal data have been corrected using a nominal Fairfield ZLand 3C instrument response.

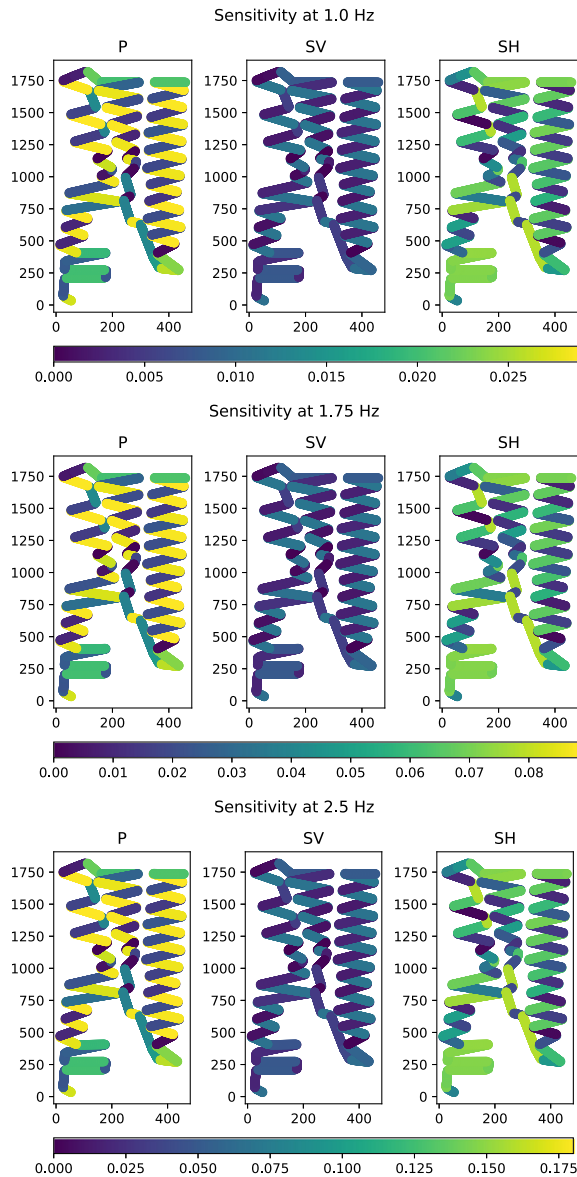


Figure B2. DAS cable sensitivity of the PoroTomo experiment at Brady, NV, at a frequencies of 1.0, 1.75 and 2.5 Hz to motions originating from the M_L 4.3, 2016 March 21 studied in Wang et al. (2018). Notably, the relative sensitivity of the DAS cable is independent of cable azimuth across this period band.

APPENDIX C: JOINT RECONSTRUCTION OF DAS AND NODES

Finally, we study the performance of a mixed deployment of horizontally sensitive point sensors (nodes in this case) and DAS cable.

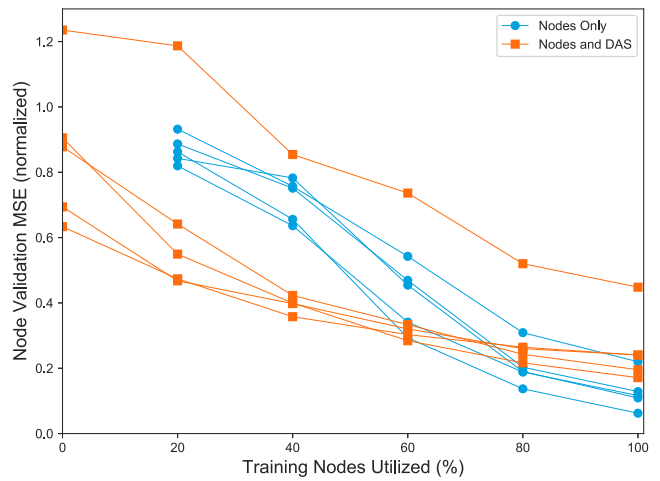


Figure C1. MSE paths for a fivefold cross-validation experiment. Blue lines show paths trained using only the remaining four-fifths of nodal data. Orange lines show paths with the addition of the full processed and quality-controlled DAS data set.

We performed a fivefold cross-validation experiment to obtain standardized MSE paths for nodal horizontal component reconstruction as a function of nodal instrument density. Specifically, we split the nodal data set into five folds. With each of these folds acting as a validation data set, we used between 20 per cent and 100 per cent of the remaining four folds as a training data set. We then calculated the MSE of the reconstruction both with and without including the DAS data in the reconstruction. We also calculated the MSE of the DAS for each validation fold with no nodal data included. The results are shown in Fig. C1. With the exception of one poorly performing fold, we see that the inclusion of DAS data substantially improves the MSE performance of reconstruction for low station densities. As the station density increases, the performance of the mixed network saturates whilst the performance of the nodal only network continues to improve. We interpret this saturation effect to be due to the inconsistencies between the nodal and DAS data due to clock errors and uncalibrated DAS-ground coupling—with a mixed network designed from the outset for joint wavefield reconstruction, improved performance at all station densities seems likely. This result highlights the potential of a mixed deployment to act as a unified areal sensor via wavefield reconstruction, even while utilizing a small number of three component seismometers. A network designed and calibrated with wavefield reconstruction in mind from the outset would likely have strongly improved performance when compared to the Porotomo experiment. This style of mixed network would be especially powerful for long-term or permanent deployment, where the costs of installing and maintaining dense three component arrays are substantial.