Data Mine

An Open-Access Data Set of Active-Source and Passive-Wavefield DAS and Nodal Seismometer Measurements at the Newberry Florida Site

Aser Abbas*[1], Brady R. Cox[1], Khiem T. Tran[2], Isabella Corey[1], and Nishkarsha Dawadi[1]

Abstract

This article documents a comprehensive subsurface imaging experiment using seismic waves in a well-studied outdoor laboratory at Newberry, Florida, which is known for significant spatial variability, karstic voids, and underground anomalies. The experiment used approximately two kilometers of distributed acoustic sensing (DAS) fiber-optic cable, forming a dense 2D array of 1920 horizontal-component channels, and a 2D array of 144 SmartSolo three-component nodal seismometers, to sense active-source and passive-wavefield seismic waves. The active-source data were generated using a powerful, triaxial vibroseis shaker truck (T-Rex) and impact sources (accelerated weight drop and an eight-pound sledgehammer) that were simultaneously recorded by both the DAS and nodal seismometers. The vibroseis truck was used to excite the ground in three directions (two horizontal and one vertical) at 260 locations inside and outside the instrumented array, whereas the impact sources were used at 268 locations within the instrumented array. The passive-wavefield data recorded using the nodal seismometers comprised 48 hr of ambient noise collected over a period of four days in four 12-hour time blocks, whereas the passive wavefield data collected using DAS consisted of four hours of ambient noise recordings. This article aims to provide a comprehensive overview of the testing site, experiment layout, the DAS and nodal seismometer acquisition parameters, and implemented raw data processing steps. Although potential use cases, such as surface-wave testing, full-waveform inversion, and ambient noise tomography, are discussed relative to example data, the focus of this article is on documenting this unique data set and presenting its initial data quality rather than on generating subsurface imaging results. The raw and processed data, along with detailed documentation of the experiment and Python tools to aid in visualizing the DAS data set, have been made publicly available.

Cite this article as Abbas, A., B. R. Cox, K. T. Tran, I. Corey, and N. Dawadi (2024). An Open-Access Data Set of Active-Source and Passive-Wavefield DAS and Nodal Seismometer Measurements at the Newberry Florida Site, *Seismol. Res. Lett.* **95**, 1082–1098, doi: 10.1785/0220230216.

Supplemental Material

Introduction

Noninvasive imaging techniques such as surface-wave testing and full-waveform inversion (FWI; e.g., Tarantola, 1984; Foti *et al.*, 2018) are increasingly being used for geotechnical site characterization due to their advantages in time, cost, and spatial coverage when compared to traditional invasive methods such as standard penetration test (SPT) borings and cone penetration test (CPT) soundings (refer to Loehr *et al.*, 2016). Geophysical imaging based on seismic wave propagation continues to evolve, with new innovations emerging to meet increasingly complex demands, such as higher imaging resolution for elastic moduli, anomaly detection, and attenuation estimation. High-quality field data are essential for developing and testing these emerging techniques.

This article documents an open-access data set from a comprehensive subsurface imaging experiment using seismic waves in a well-studied outdoor laboratory at Newberry, Florida, which is known for significant spatial variability, karstic voids, and

1. Department of Civil and Environmental Engineering, Utah State University, Logan, Utah, U.S.A., https://orcid.org/0000-0003-1462-8539 (AA); 2. Department of Civil and Coastal Engineering, University of Florida, Gainesville, Florida, U.S.A.

*Corresponding author: aser.abbas@usu.edu

underground anomalies (Tran and Hiltunen, 2011). The experiment utilized ~2 km of distributed acoustic sensing (DAS) single-mode fiber-optic cable, creating a dense 2D grid array comprising 1920 horizontal-component channels, covering an area measuring 75 m × 155 m. In addition, a 2D grid array of 144 SmartSolo three-component (3C) nodal seismometers was deployed at the site. Both sensing systems were used to simultaneously record active-source and passive-wavefield seismic waves. The active-source data were generated using a powerful, triaxial vibroseis shaker truck (T-Rex) and two types of impact sources (accelerated weight drop and an eight-pound sledgehammer). The experiment focuses on the near-surface depths relevant to geotechnical engineering (less than ~ 50 m). The known and possibly unknown karstic voids of variable size and depth at the testing site add an intriguing dimension to this experiment.

Several noninvasive seismic techniques have been developed to detect voids and other underground anomalies (e.g., Cook, 1965; Branham and Steeples, 1988; Pernod *et al.*, 1989; Belfer *et al.*, 1998; Kolesnikov and Fedin, 2018; Kristekova *et al.*, 2020; Sloan *et al.*, 2012; Smith *et al.*, 2019; Wang *et al.*, 2019). However, detecting anomalies using real-field data remains challenging (Grandjean and Leparoux, 2004; Sloan *et al.*, 2010; Smith *et al.*, 2019). We anticipate that this open-access data set, featuring a powerful, triaxial vibroseis shaker, and 3C sensors, in conjunction with the dense DAS array, will serve as a valuable resource for researchers seeking to explore novel approaches for anomaly detection and subsurface imaging.

This article aims to provide a comprehensive overview of the testing site, experiment layout, the DAS, and nodal seismometer acquisition parameters, and implemented raw data processing steps. Although potential use cases, such as surface-wave testing, FWI, and ambient noise tomography, are discussed relative to example data, the focus of this article is on documenting this unique data set and presenting its initial data quality rather than on generating subsurface imaging results. The ensuing paragraphs offer a concise overview of the sensing technologies employed in this experiment and the potential value of the data set documented herein.

Overview of Sensing Technologies

The first sensing technology used in this experiment was DAS, which is a rapidly evolving technique for transforming fiber-optic cables into a distributed array of ground motion (i.e., strain or strain rate) sensors (Cox *et al.*, 2012; Yu *et al.*, 2019). It is increasingly being used to sense active and passive seismic waves for geophysical imaging and seismic monitoring of the near surface (e.g., Dou *et al.*, 2017; Hubbard *et al.*, 2022; Vantassel, Cox, *et al.*, 2022; Yust *et al.*, 2023). DAS measures dynamic strain along a fiber-optic cable by utilizing an interrogator unit (IU) that emits a series of laser pulses (probe pulses) through the cable. The interaction between a probe pulse and the fiber's inhomogeneities returns a backscattered signal, mainly composed of Rayleigh backscatter, to the launching end. The relative phase of the Rayleigh backscattered light is used to determine changes in length between scattering regions along the cable. This determination is repeated for each resolvable point along the fiber, resulting in a measure of the dynamic strain as a function of time and location (Hartog 2018). Studies by Daley *et al.* (2016), Hubbard *et al.* (2022), and Vantassel, Cox, *et al.* (2022) have demonstrated that DAS measurements can yield similar waveforms and processed data (e.g., surface-wave dispersion data) as geophones when proper care is taken. Furthermore, DAS can provide unprecedented spatial resolutions (on the order of meters) and coverage (on the order of tens of kilometers), surpassing the economically feasible simultaneous data acquisition capabilities of traditional sensing technologies.

In recent years, several studies have been conducted to evaluate the potential of using DAS for noninvasive near-surface imaging. Most of these studies utilized either a 1D (i.e., line) DAS cable and active sources (e.g., Hubbard *et al.*, 2022; Vantassel, Cox, *et al.*, 2022; Yust *et al.*, 2023), a 2D DAS cable configuration and an impact source (e.g., Castongia *et al.*, 2017), or a 2D DAS cable configuration and ambient noise (e.g., Dou *et al.*, 2017). Lancelle *et al.* (2014) employed a 2D DAS fiber-optic cable configuration along with shear and vertical vibrational sources at the Garner Valley testing site in California, United States. However, the cable runs were sparsely spaced and the sources were utilized at a limited number of locations. Obermann *et al.* (2022) conducted a seismic study in the Hengill geothermal area in southwest Iceland using a network of 3C nodal seismometers and two DAS fiber-optic cables, along with a uniaxial vibroseis shaker. Their research focused on imaging the top four kilometers of the crust, and thus, their nodal seismometers were spaced out over several kilometers in each direction, with interstation distances varying from tens to hundreds of meters. Parker *et al.* (2018) utilized a triaxial vibroseis shaker and deployed 238 3C nodal seismometers along with 8700 m of DAS fiber-optic cable at the natural laboratory of Brady Hot Springs to gather active-source and passive-wavefield data. Notably, the nodal seismometer array had 60 m spacing in its most densely instrumented region.

As mentioned earlier, the second sensing technology deployed at the test site was 3C nodal seismometers, which are often referred to as "nodal stations" in the scientific literature. These stations allow for the concurrent measurement of ground shaking in all three directions. This complements the unidirectional sensitivity of the DAS system, thereby providing a more comprehensive view of the seismic wavefield. The three perpendicular geophones (two horizontal and one vertical) exhibit sensitivity to different types of waves. For instance, Rayleigh and compression waves are best identified on the vertical and horizontal inline components, whereas the horizontal crossline component is better suited to detect Love waves

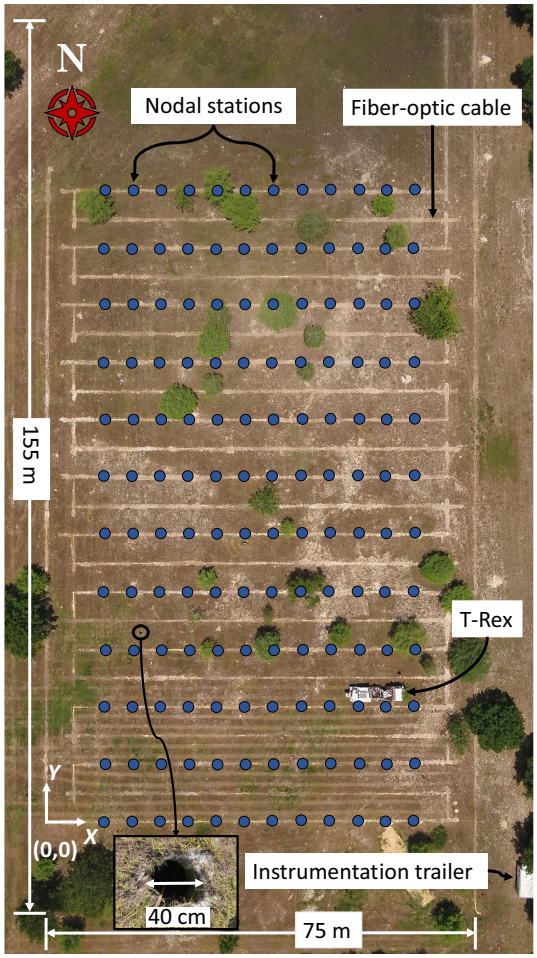


Figure 1. Overhead drone photograph of the 75 m × 155 m instrumented area at the Newberry test site, showing ~2 km trenches used for fiber-optic cable installation, as indicated by white, linear markings on the ground surface, and the 144-3C geophone nodal seismometer locations, indicated by blue circle symbols. In addition, the figure highlights one of the T-Rex shot locations, one of the voids present inside of the array, and the instrumentation trailer outside of the array. The color version of this figure is available only in the electronic edition.

(Foti *et al.*, 2018; Vantassel and Cox, 2022; Vantassel, Cox, *et al.*, 2022). Because different wave types carry information about different mechanical properties of the subsurface (Sheriff and Geldart 1995), analyzing the data collected by 3C nodal seismometers can provide valuable insights into the 3D mechanical properties of the site. The versatility of 3C nodal seismometer measurements open avenues for researchers who are working on techniques that leverage only the vertical, either of the horizontal, or any simultaneous combination of the three components of ground-shaking measurements (e.g., Nakamura, 1989; Fathi *et al.*, 2016; Pan *et al.*, 2016; Wathelet *et al.*, 2018; Smith *et al.*, 2019; Wang *et al.*, 2019; Cheng *et al.*, 2020, 2021; Cox *et al.*, 2020; Kristekova *et al.*, 2020).

Site Overview

The test site is a dry retention pond located in Newberry, Florida, at latitude 29.650109° N and longitude −82.597557° W, along State Road 26 (refer to Fig. 1). Sinkholes are common in this area and generally an immense problem in parts of Florida due to the dissolution of carbonate rocks like limestone by water (Upchurch *et al.*, 2019). The Newberry retention pond site has undergone thorough investigations over the years, utilizing both invasive and noninvasive methods. Findings from these previous studies indicate that the subsurface is composed of medium-density fine sand and silt that range in depth from 2 to 10 m overlying highly variable limestone (Tran and Hiltunen, 2011; Tran *et al.*, 2013). The site also contains sinkholes of different sizes and depths, some of which have been repaired by dumping uncontrolled fill. Notably, the largest chimney observed at the site had a diameter of ~40 cm (refer to Fig. 1). Noninvasive imaging studies conducted over smaller portions of the site by Tran *et al.* (2013) and Mirzanejad *et al.* (2020) have detected voids as large as several meters tall and wide at depths ranging from 3 to 10 m below the ground surface. The locations and vertical dimensions of some of these voids have been confirmed with geotechnical borings, as described subsequently.

The varied stiffness and depth of the limestone layer, along with the presence of surficial and underground anomalies, make this site a prime location for noninvasive subsurface imaging research. Tran and Hiltunen (2011) conducted 10 CPTs, eight geotechnical borings with SPTs, and 12 consecutive 36.6-m-long seismic refraction tests using a linear array of 31, 4.5 Hz vertical geophones and a sledgehammer source. The first arrival times from the refraction tests were inverted using simulated annealing to develop 2D compression wave velocity (V_P) profiles of the site. However, no voids were identified via seismic refraction testing by Tran and Hiltunen (2011). Tran *et al.* (2013) used a linear array of 24, 4.5 Hz vertical geophones and a sledgehammer source to collect seismic data at 10 different locations at the site. The seismic data were then inverted using 2D FWI, which identified an underground anomaly that was later verified to be a void through an SPT sounding. Nonetheless, they observed that the predicted depth of the void was greater than its actual depth. They attributed this discrepancy to the difference between the measured wavefield, which is affected by the 3D variations in the subsurface, and the assumed plane strain condition used in their 2D FWI. Tran *et al.* (2020) expanded on the FWI studies at the site by utilizing the uniaxial, small vibroseis shaker truck named Thumper from the Natural Hazards Engineering Research Infrastructure at the University of Texas at Austin (NHERI@UTexas) experimental facility (Stokoe *et al.*, 2020) to excite the ground at 65 locations within and around a 2D grid of 48, 4.5 Hz vertical geophones arranged in a 4 × 12 configuration. The source and receiver 2D grids were uniform, with 3-m spacing, covering a total area of 12 m × 36 m. Using the collected data and 3D FWI analyses, Tran *et al.* (2020) created a 3D subsurface model

below the sensor grid, identifying a low-velocity anomaly and a void that was confirmed through an SPT sounding. Mirzanejad *et al.* (2020) employed 3D Gauss–Newton FWI to image the subsurface and identify voids at the Newberry site. Their seismic experiment featured 72 vertical 4.5 Hz geophones and 91 impact shot locations using a 40 kg dropweight. Their geophones and shot locations were arranged in 6 × 12 and 7 × 13 uniform grids, respectively, with 3 m spacing. Their FWI identified three voids, but during verification with SPT soundings, they confirmed the presence of only two voids, whereas no voids were found at the third location.

The experiment documented herein improves upon previous noninvasive studies at the Newberry site by (1) covering a larger spatial area, (2) using DAS to enable much denser spatial sampling, (3) using 3C nodal stations to enable multicomponent processing, (4) using a dense grid of more numerous shot locations, and (5) incorporating 3C shaking capabilities from a broadband and powerful vibroseis shaker truck named T-Rex from NHERI@UTexas experimental facility.

Experimental Data Set

The subsurface imaging data set documented in this article was collected over the course of eight days, beginning Monday, 9 May 2022, and ending Monday, 16 May 2022. The data set consists of both active-source and passive-wavefield seismic-wave recordings that were sensed using 2D surface arrays of DAS fiber-optic cable and 3C nodal stations. The approximately 2 km of DAS fiber-optic cable were laid out in the zigzag pattern shown schematically by a black line in Figure 2, which is also visible through the white markings from the drone image presented in Figure 1. The 144 deployed 3C geophone nodal stations were arranged in a 12 × 12 grid with a 5 m spacing in the *X* (west–east) direction and a 10 m spacing in the *Y* (south–north) direction, as indicated by the blue solid circles in Figures 1 and 2. T-Rex shook the ground in all three directions at 260 shot locations both inside and outside the instrumented area, whereas impact sources struck the ground vertically at 286 shot locations within the instrumented area (refer to Fig. 2). Between T-Rex and the impact sources, a total of 367 shot locations were utilized to generate the active-source data. The passive-wavefield data consisted of ∼4 hr of ambient noise recordings using DAS, and ∼48 hr of ambient noise recordings using the nodal stations.

Experiment layout

The experiment was laid out on a 2D survey grid, as illustrated in Figure 2. The grid consisted of 25 horizontal lines pointing approximately east (bearing 89°) and 16 vertical lines pointing approximately north (bearing 359°). The horizontal lines, except for the lowermost and uppermost, were uniformly spaced at 5 m and were labeled with the letters A–W, in order from south to north, respectively. The lowermost line Z was 15 m south of line A, whereas the uppermost line ZZ was 30 m north of line W. The vertical lines were spaced 5 m apart and named from west to east, 101, 102, 1, 2, 3, … to 12, 103, and 104, respectively. The grid intersection points will be referenced first by the letter and then by the number representing the intersecting horizontal and vertical lines, respectively (e.g., A101). Although the global latitude and longitude coordinates of the grid points were surveyed and are included in the electronic data set, this article will utilize a local coordinate system for ease of reference, with the origin at point A101 (local coordinate 0,0), the positive *X* direction pointing eastward, and the positive *Y* direction pointing northward.

DAS

An OptaSense ODH4+ IU was used in this experiment, which was borrowed from the NHERI@UTexas experimental facility (Stokoe *et al.*, 2020). The ODH4+ IU can provide Global Positioning System (GPS) time with microsecond accuracy and was configured with a 2.04 m gauge length and 1.02 m channel separation, the minimum allowed by the OptaSense ODH4+. The gauge length refers to the average straight-line distance between the consecutive origins of the Rayleigh backscatter. Hence, the measurements of vibrations at each sampling location (channel separation) represent the average over the 2.04 m gauge length. Selecting the right fiber-optic cable is crucial for good DAS measurements because the cable functions as both the strain sensing element and the means of transmitting optical signals (Hartog, 2018). In this experiment, a single mode fiber-optic tactical cable (AFL X3004955180H-RD) consisting of four tight buffered fibers coated in a layer of aramid yarn and enclosed by a polyurethane jacket was used. Studies by Hubbard *et al.* (2022) and Vantassel, Cox, *et al.* (2022) confirm that this cable offers good deformation coupling between the internal optical fiber and the ground when buried with soil compacted around and over it. The length of cable installed at the site was ∼2 km (exactly 1958.4 m), with one end connected to the IU at the instrumentation trailer located in the southeast corner of the testing site (refer to Fig. 1), and the other end appropriately terminated at the northwest corner of the site to reduce end reflections. The DAS IU configuration resulted in 1920 horizontal-component channels along the fiber-optic cable. The cable was buried in a 20 cm wide and 10–15-cm-deep trench to ensure optimal coupling between the ground and the cable (Fig. 3a,b). The corners of the trench were rounded to a radius of ∼20 cm (Fig. 3c), which is greater than the AFL cable's allowable bend radius. Following cable installation, the trench was backfilled with native soil, or with clean sand when the native soil was too hard and clotted to allow for good coupling. The backfilled soil was then manually compacted over the cable to ensure good coupling with the native ground (Fig. 3d).

All cable corners were left exposed until tap tests could be performed to index the DAS cable (i.e., map the DAS channel numbers to their physical locations). The tap tests involved lightly tapping on the fiber-optic cable at all corners and other important locations (such as the start and end of the cable) and

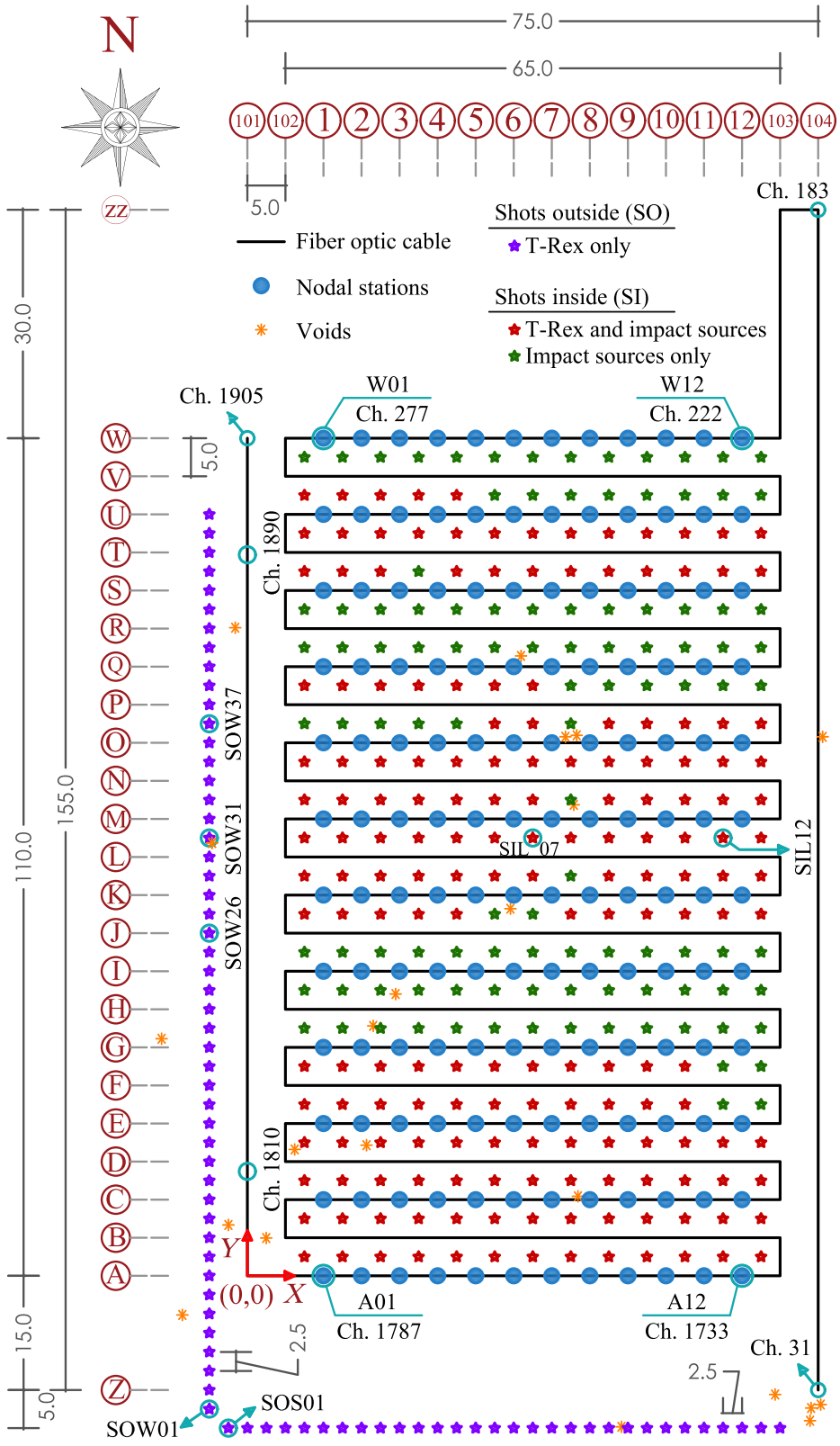


Figure 2. Schematic layout of the test site showing locations of the: 3C geophone nodal stations, fiber-optic cable, T-Rex and impact shots, and voids that are visible from the ground surface. The layout is comprehensive, including all of the line numbers, letters, and dimensions in meters used to arrange the equipment. The color version of this figure is available only in the electronic edition.

noting the DAS channels that responded with the most significant energy. Based on the tap tests, the first and last channels on the cable with usable data (i.e., the first and last buried channels) are channels 31 and 1905, respectively, as indicated in Figure 2. A tap test can only locate the measurement point with an accuracy of the gauge length of the DAS system, meaning it has a ±one-half gauge length margin of error. Consequently, the absolute spatial location of each channel is accurate to within about 1 m. However, the spacing between channels is exactly equal to one-half of the gauge length, or 1.02 m. After indexing, the corners were backfilled and compacted to ensure proper coupling between the cable and the surrounding soil.

In DAS, it is desirable to set the laser pulse repetition (ping rate) as rapid as possible for a given cable length to increase signal-to-noise ratio and to mitigate phase determination errors. Phase determination errors arise when the strain rate exceeds the DAS system's capability to detect, similar to how amplitudes that are too high for traditional seismographs result in clipping (Hubbard *et al.*, 2022). This sometimes happens when a large source is activated near the fiber-optic cable. To mitigate such errors, the ping rate can be set much higher than the desired sampling rate, and the signal can be subsequently decimated to lower sampling rates relevant to the study being conducted to save memory space. On Friday, 13 May, during the first day of active wavefield data acquisition, the

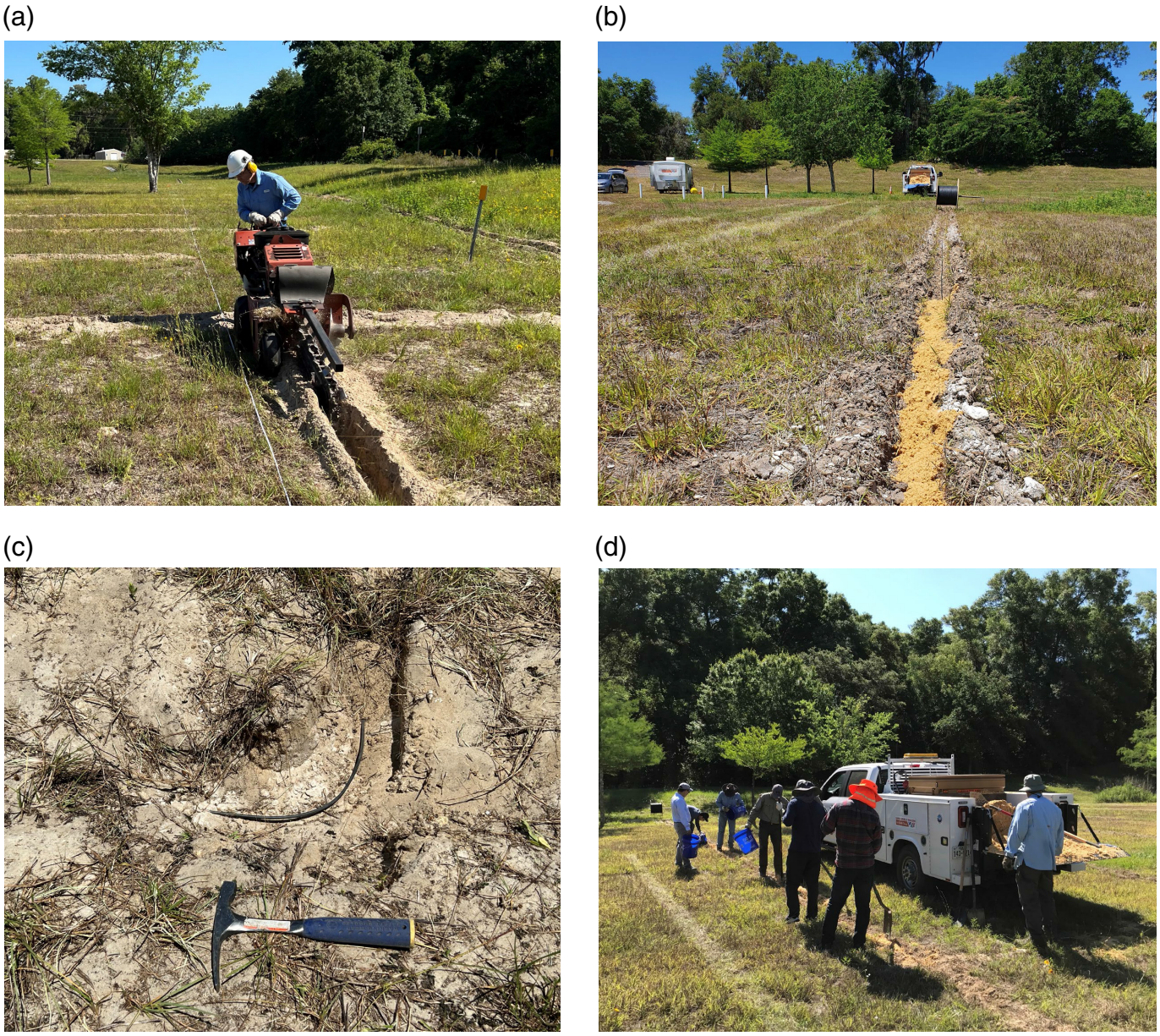


Figure 3. Pictures illustrating the fiber-optic cable installation procedure, starting with (a) trenching the cable route, (b) laying the cable by rolling it off the cable spool along the trench, (c) rounding the cable at the corners, and (d) filling the trench with native soil or clean sand, followed by compacting the backfilled soil to ensure proper coupling between the cable and the ground. The color version of this figure is available only in the electronic edition.

ODH4+ IU was configured with a ping rate of 50 kHz, and the DAS data were decimated to 10 kHz in real time. On subsequent days, the ping rate was reduced to 20 kHz, and the data were decimated to 1 kHz. This lowering of the ping rate was necessitated by unanticipated problems with the laser pulse balancing on the ODH4+. Thus, the raw data collected on Friday, 13 May were of slightly higher quality because a 50 kHz ping rate was used. For more details regarding the influence of ping rate on the quality of data collected using an ODH4 + IU and the same fiber-optic cable type employed in this experiment, specifically on phase determination error, please refer to Hubbard *et al.* (2022).

Nodal stations

The 144 SmartSolo IGU-16HR 3C nodal stations used in this experiment (refer to Fig. 4) were sourced from two locations: 88 stations from the Earthquake Engineering and Subsurface Imaging Lab at Utah State University (USU) and 56 stations from the NHERI@UTexas experimental facility. Each station has three, orthogonal, 5 Hz geophones, a GPS clocking unit, and a self-contained power supply with a 30-day battery life. The geophones are wired such that a tap from the north, east, or top causes a positive voltage departure in the geophone oriented along that axis (refer to Fig. 4c). The stations come

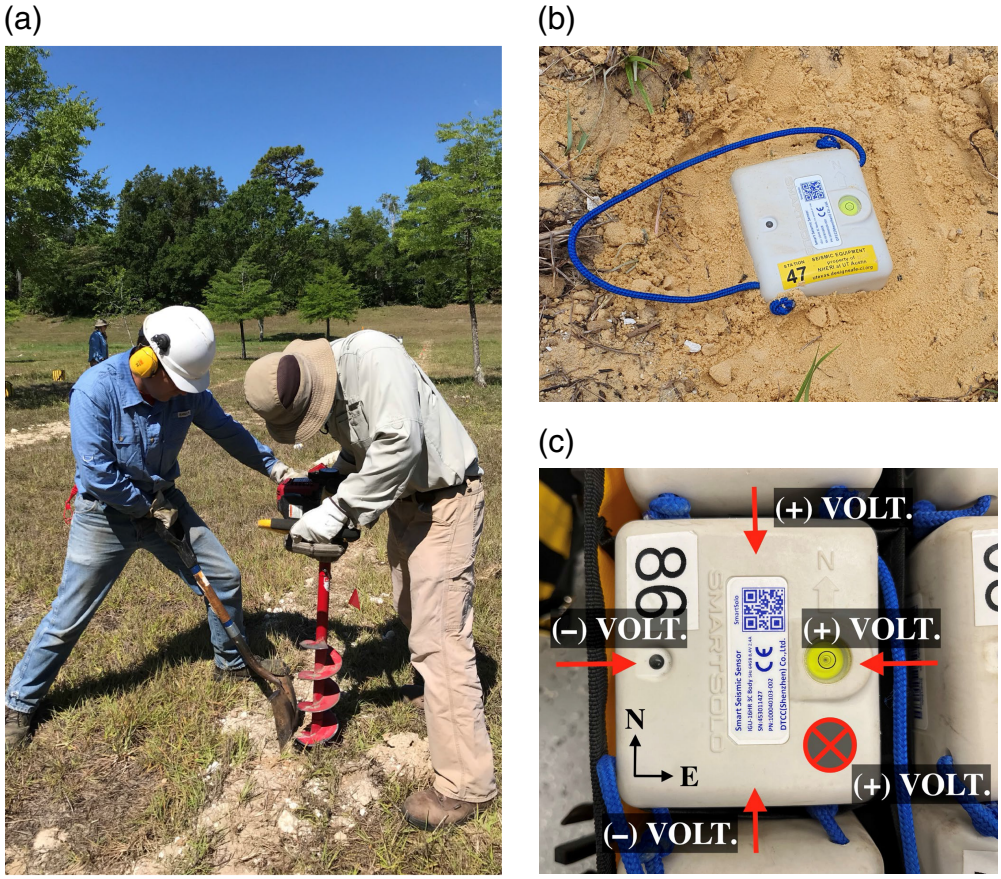


Figure 4. Picture illustrating the installation of nodal stations next to the fiber-optic cable. In panel (a) excavation is performed with a gas-powered auger whereas the fiber-optic cable is protected with a shovel. Panel (b) depicts a completed installation, with the station securely in place and its top exposed. Panel (c) describes the voltage polarity of the three geophones in each nodal station. The color version of this figure is available only in the electronic edition.

equipped with four conical spikes to ensure good coupling with the ground. For this experiment, the SmartSolo nodal stations were configured with a sampling rate of 250 Hz. The stations were deployed right next to the fiber-optic cable in their respective locations, as shown in Figures 1 and 2. Depending on the stiffness of the ground, either a gas-powered earth auger or a manual posthole digger was used to excavate a hole ~20 cm in diameter and 25 cm deep for the stations. The fiber-optic cable was protected from the hole digging operations using a shovel, as shown in Figure 4a. The orientation of the stations was such that the north arrow pointed toward the $+Y$ direction, which was orthogonal to the fiber-optic cable, conveniently aligning with true north. This ensured that the east–west geophone in the 3C nodal stations was properly aligned with the fiber-optic cable. The stations were then leveled and the holes backfilled using either excavated soil or clean sand when needed, leaving only the station's top exposed, as shown in Figure 4b.

Active-source wavefield generation

Two source types were used for active wavefield generation: a highly controlled, powerful, broadband vibroseis source (T-Rex), and more variable, impact sources. T-Rex, which is shown in Figure 5a, is a 29 ton, triaxial vibroseis truck capable of shaking its baseplate in the vertical, longitudinal, and transverse directions (Stokoe *et al.*, 2020). The experiment harnessed T-Rex's maximum force output of ~267 kN vertically and 134 kN horizontally in each direction to generate 12 s linear frequency sweeps from 5 to 80 Hz. In addition to GPS time and coordinates for each shot location, the T-Rex electronics recorded the baseplate and mass accelerations and the ground force with 1 kHz sampling rate for each shot. Herein, a "shot" refers to an instance for which a source was used to excite the ground at a given location. In total, T-Rex vibrated at 260 distinct locations: 81 outside and 179 inside the instrumented area, as illustrated in Figure 2. Shots outside (SO) of the instrumented area were distributed among 30 locations to the south (S), 48 locations to the west (W), and three locations to the north (N). These locations will be referred to as SOS, SOW, and SON, respectively. The three SON locations are not shown in Figure 2 due to their significant offsets from the main instrumentation grid.

The location numbers for each shot location increase in order with the positive X direction for the SOS and SON locations and with the positive Y direction for the SOW locations. Shots inside (SI) of the instrumented area are referred to by SI and the horizontal line directly south of them. For instance, the first line of shot locations farthest to the south and inside the instrumented area (refer to Fig. 2) is named shot inside to the north of line A (SIA). Similar to the outside shot locations, the inside shot location numbers along any given line increase with the positive X direction. For example, to aid the reader with orientation, shot locations SOW01, SOS01, SIL07, SIL12, and others are labeled in Figure 2. Although the inside shot locations are organized on a regular 5 m × 5 m grid, T-Rex could not be used at all shot locations inside the instrumented area, either because of restricted maneuverability around trees or the presence of voids nearby that might collapse due to the

Figure 5. Pictures illustrating the various sources utilized in the experiment. Panel (a) shows the large, triaxial vibroseis shaker truck, T-Rex, from the NHERI@UTexas experimental facility, panel (b) shows an eight-pound sledgehammer, and panel (c) shows the PEG-40 kg impact source mounted on a Ford F-350 pickup truck. The color version of this figure is available only in the electronic edition.

weight of the truck. Thus, the SI locations illustrated in Figure 2 are clearly denoted as those for which both T-Rex and impact sources were used and those where only impact sources were used (i.e., those where T-Rex could not be used). At each shot location where T-Rex was used, all three shaking modes were utilized: P mode for vertical shaking, SL mode for shear longitudinal shaking (i.e., inline with the truck), and ST mode for shear transverse shaking (i.e., cross-line to the truck). However, these three modes of shaking were not excited consecutively at each location. Instead, for each line of shots (e.g., SOS), a vibration mode was set (e.g., P mode) and shots were performed along the entire line. Then, the shaking mode was switched (e.g., to SL mode) and shot locations along the line were revisited using the updated mode. This was found to be more efficient than switching shaking modes at each shot location. Because of the significant number of shot locations and shaking modes used, and given the powerful nature of the source, no shot stacking was performed (i.e., only a single shot was collected for a given shaking mode at each shot location).

The impact sources consisted of a 40 kg propelled energy generator (PEG-40 kg), also known as an accelerated weight drop, manufactured by R. T. Clark Companies (Fig. 5b), and an eight-pound sledgehammer (Fig. 5c). The PEG-40 kg is a portable source that generates seismic energy when a hammer mass is propelled downward by an elastomer band onto an impact plate. The hammer drop mass and height are 40 kg and ~40 cm, respectively, generally producing an impact frequency range of ~10–250 Hz. There are two modes of operation for the PEG-40 kg: single cycle and continuous cycle. For this experiment, the single-cycle mode was used. To ease mobility, the PEG-40 kg was mounted on a Ford F-350 pickup truck, as shown in Figure 5b. The impact sources were only used to excite the ground inside the instrumented area, as shown in Figure 2. Because the F-350 pickup truck was significantly lighter and smaller than T-Rex, it was able to reach more shot locations. In the rare cases when the F-350 was unable to reach a shot location, an eight-pound sledgehammer was utilized instead of the PEG-40 kg, ensuring that all inside shot locations in the 5 m × 5 m grid were covered by one of the impact sources. Unlike T-Rex shots, the impact source signature and GPS times were not automatically recorded. Therefore, the impact times were manually noted by the researchers using GPS time from a smartphone and recorded in the field data sheet. These approximate times were later verified and refined through inspection of the closest nodal station waveforms. The impact arrival time at the nearest station was manually selected as

the first instance of energy surpassing the noise floor. For further details on the amplitude and frequency content of the wavefields generated by T-Rex and PEG-40 kg, please refer to the supplemental material, available to this article.

Passive-wavefield monitoring

The DAS passive-wavefield data consisted of approximately four hours of ambient noise recordings on 15 May between 16:58 and 21:04 Universal Time Coordinated (UTC). During this period, there were two instances of rainfall that took place from 17:20 to 17:35 UTC and 20:54 to 20:56 UTC. The rainfall times were recorded using smartphones by the researchers conducting the experiment. Because there may be a difference between the actual start time of the rainfall and when the researchers perceived it, these times are not highly accurate. There was also notable interference (i.e., noise) from an electric power generator on DAS channels 30–66 and 1720–1742 during this time.

The nodal stations continuously recorded data from deployment to termination, which are archived in the "Raw data" folder of the electronic data set (please refer to the supplemental material for more details). However, for the passive-wavefield data, we specifically focused on the nonworking hours when all researchers were offsite. It is important to note that during working hours, we were actively using T-Rex and impact sources. By considering only the nonworking hours, the passive-wavefield data comprises a total of 48 hr of ambient noise, distributed over four 12-hr time blocks spanning a four-day period. Further details are provided in the Processed data section.

Electronic data set organization

The open-access electronic data set published on DesignSafe (Rathje *et al.*, 2017) includes both raw and processed data from the DAS and SmartSolo sensing systems, along with all the necessary supporting documents. The data set is organized into three main folders. First, the "Raw Data" folder contains the original, as-collected data in its raw format from both the DAS and the nodal stations. Second, the "Supporting Documents" folder is a comprehensive repository of essential information for potential data reprocessing, ensuring that users have all the resources needed to revisit and reanalyze the raw data. Finally, the "Processed Data" folder is developed for user convenience, offering preprocessed data for easy access and utilization. For a more detailed understanding of the electronic data set structure and its components, please refer to the supplemental material, for which all these details are thoroughly covered.

Processed data

DAS. The DAS data include waveforms collected from T-Rex shots, impact sources, and ambient noise. Subsequent processing steps were applied to each individual T-Rex shot as part of the preprocessing and the development of the processed data folder. The shot start time was retrieved from the T-Rex electronics trigger file, where the GPS times for all shots were recorded. The times for the T-Rex shots automatically included a one-second pretrigger delay. A 15 s window encompassing the shot was extracted from the 1-min-long DAS raw H5 files produced by the ODH4+ IU, which included a one-second pretrigger delay, 12 s of T-Rex shaking, and 2 s of listen time post-T-Rex shaking to capture the waves reaching the array extremities. The 15 s windows were examined for dropped samples or errors in any of the 1920 DAS channels. Two of the 780 T-Rex shots had dropped samples or errors on one or more channels, namely, Shots Z_SID11 and Z_SIF07. These particular shots correspond to T-Rex shaking in the Z direction at locations SID11 and SIF07, respectively, and as a result, they were excluded from the processed data set. The raw data were multiplied by $2\pi/2^{16}$ to convert it into phase change measurements in radians. A 3-Hz high-pass filter was applied to remove low-frequency artifacts from laser drift and static strains caused by shaking close to the fiber-optic cable, as recommended by Hubbard *et al.* (2022). The DAS waveforms from the SOS and SOW shots were decimated from 10 to 1 kHz to align with the rest of the collected data. Phase data were then converted to strain using the following equation (Hubbard *et al.*, 2022):

$$\varepsilon_{xx} = \frac{\lambda d\phi}{4\pi n g \xi}, \tag{1}$$

in which λ is the average laser wavelength of the DAS system in a vacuum, equal to 1550 nm; $d\phi$ is the phase change measured by the DAS in units of radians; n is the group refractive index of the sensing fiber, ~1.47; ξ is the photoelastic scaling factor for longitudinal strain in an isotropic medium, equal to 0.78; g is the gauge length, ~2.04 m; and ε is the normal strain per single gauge length. For each shot, the processed DAS data, along with the T-Rex pilot signal, base plate and mass accelerations, ground force in engineering units, and other shot-related information, such as local and global coordinates, shot time, and sampling rate, were organized into an "event" object and saved in an H5 file. All the DAS channels for each shot were also cross correlated with the T-Rex pilot signal and stored in another H5 file.

The same processing steps were followed for the impact sources, with a few notable exceptions. For each impact shot, the impact time was manually selected as the first instance of energy surpassing the noise floor of the closest nodal station waveforms to the shot location. This impact time was then used to extract a 3 s time window from the raw 1 min H5 files. This window included a 1 s preimpact portion and a 2 s postimpact portion. In addition, the header of each processed H5 file for an impact source indicates whether the shot was generated by an accelerated weight drop (i.e., the PEG-40 kg) or the eight-pound sledgehammer.

The same processing steps followed for the active-source data were also followed for the four hours of ambient noise

recorded by the DAS. However, the files were maintained as 1-min-long segments.

Nodal stations. The SmartSolo nodal station data processing was limited to extracting the shot time windows (15 s windows for T-Rex shots and 3 s windows for impact shots), merging the three individual components into a single miniSEED file, editing the header file information, and arranging them in a user-friendly format. For T-Rex shots, the T-Rex trigger file was utilized to obtain the shot times, which were then used to trim a 15 s time window from all three components of the nodal stations for each shot. This included a 1 s pretrigger delay, a 12 s T-Rex shaking duration, and a 2 s listen segment post-T-Rex shaking. The miniSEED files for each nodal station include three components: DHN, DHE, and DHZ, following the naming convention recommended by Federation of Digital Seismographic Networks (FDSN) (2012). The miniSEED header file holds important information, such as the record's sampling rate, start and end times, and the station location. The miniSEED files corresponding to the impact sources have a 3 s duration, with 1 s before the impact and 2 s following it. It should be noted that the impact source signature was not recorded. The processing of passive data involved extracting data from 12 to 16 May, spanning the period from 23:00 UTC (when work at the site concluded) to 11:00 UTC (when work resumed the following day). This process resulted in a total of 48 hr of ambient noise data, organized into four 12 hr time blocks, collected over a span of four days.

Potential Data Set Use Cases

This section presents raw and preprocessed data examples from the archived, open-access data set, with the aim of inspiring potential use cases for those interested in the data set. The high-quality active-source data can be used for seismic reflection, refraction tomography, surface-wave inversion, FWI, and other imaging techniques. Similarly, the passive-wavefield data can be employed in techniques such as horizontal-to-vertical spectral ratio (HVSR), microtremor array method (MAM), ambient noise tomography, and so on. Although no imaging results are presented in this section, we showcase that the waveforms and dispersion data extracted from the data set are of high quality and can be used in such active-source and passive-wavefield imaging techniques. In addition, we provide one or two example papers for each potential use case to aid readers in exploring further details.

1D, 2D, and 3D imaging

Previous research conducted at this site by Tran and Hiltunen (2011), Tran *et al.* (2013, 2020), and Mirzanejad *et al.* (2020) revealed a high degree of spatial variability in its subsurface. This variability, coupled with the existence of karstic voids, suggests that employing 2D and 3D imaging techniques would be more suitable than 1D methods in effectively characterizing the subsurface. These imaging techniques can capitalize on the high spatial sensing resolution provided by the DAS, as well as the 3D sensitivity of the nodal stations. Successful attempts have been reported in the literature in imaging the subsurface using active source 2D FWI (e.g., Tran *et al.*, 2013; Wang *et al.*, 2019) and 3D FWI (e.g., Fathi *et al.*, 2016; Smith *et al.*, 2019; Tran *et al.*, 2020) using geophone data, and recently 2D FWI using DAS data (e.g., Yust *et al.*, 2023). The efficacy of active-source imaging techniques critically depends on the wavefield generated by the source being sufficiently strong throughout the entire spatial extent of the instrumented area. Thus, rather than demonstrating specific imaging strategies and results, we instead focus on illustrating the quality of the collected waveforms and some potential ideas for taking advantage of the multidirectional sensing and multidirectional shaking on such a dense grid.

Figure 6 displays the waveforms generated by T-Rex shaking in the *X* direction at location SIL07, referred to as shot X_SIL07 (see Fig. 2). The different panels in Figure 6 illustrate the waveforms recorded by the DAS channels and the DHE component of the nodal stations at the four farthest corners of the area instrumented by nodal stations, specifically, DAS channels 1787, 1733, 277, and 222, along with nodal stations A01, A12, W01, and W12, as labeled in Figure 2. These panels demonstrate that the wavefields generated by T-Rex were clearly sensed by both the DAS and nodal stations. Further evidence of this can be seen by examining a waterfall plot of the waveforms recorded by the longest DAS line, line 104, for T-Rex shot Y_SIL12 (i.e., shaking in the *Y* direction at location SIL12), as shown in Figure 7. This waterfall plot presents the cross-correlated waveforms captured by DAS channels 31–183 (refer to Fig. 2), with each trace normalized by its absolute maximum amplitude. Figure 7 reveals that clear waveforms were sensed by the entire DAS line. Furthermore, disturbances to the classical linear arrival-time moveout patterns can be observed in the spatially dense DAS waveforms, potentially resulting from local heterogeneities and spatial variability at the site. The waveforms shown in Figures 6 and 7 are typical of the quality contained in this extensive data set. As such, the waveforms should be of sufficient quality for performing various 1D, 2D, and 3D imaging studies.

Regarding the potential to use the data set for 1D and pseudo-2D surface-wave imaging; Figure 8 shows a selection of surface wave dispersion images that can be derived from various permutations of T-Rex shaking directions at just one-shot location, namely SOW40, and utilizing a single line, line Q, of DAS channels and nodal stations. The locations of shot SOW40 and line Q are shown relative to the entire experiment in the schematic map presented in Figure 8a. The dispersion images displayed in Figure 8 were generated using the frequency-domain beamformer technique with cylindrical-wave steering, square-root-distance weighting (Zywicki and Rix, 2005), and frequency-dependent normalization, as implemented in the open-source Python package swprocess (Vantassel, 2022).

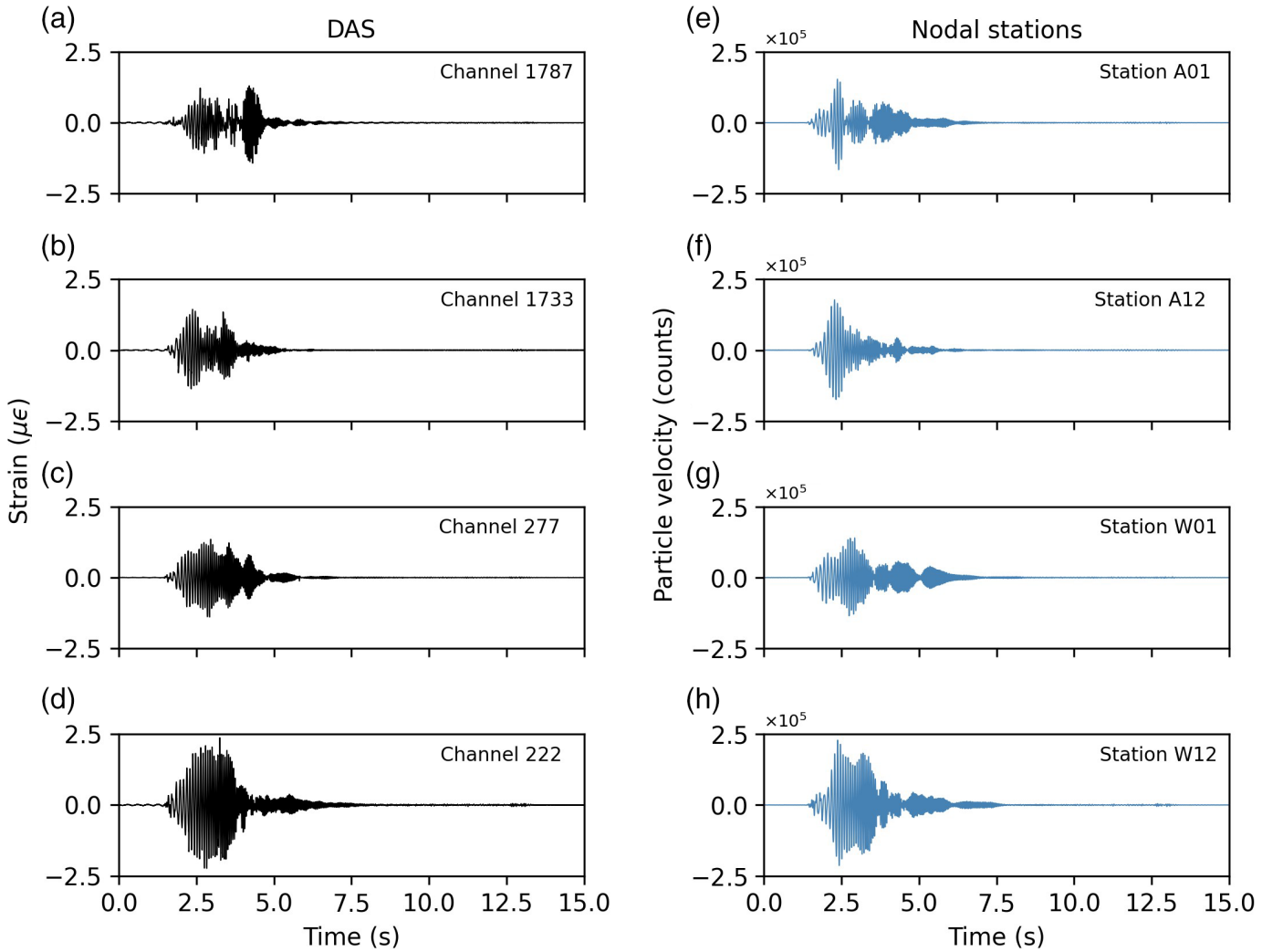


Figure 6. The waveforms generated by T-Rex shot X_SIL07 (refer to Fig. 2), as captured by both the distributed acoustic sensing (DAS) channels and the DHE component of nodal stations positioned at the four farthest corners of the instrumented area. Panels (a–d) showcase DAS channels 1787, 1733, 277, and 222, while panels (e–h) present the DHE component of nodal stations A01, A12, W01, and W12, respectively. The color version of this figure is available only in the electronic edition.

Figure 8b,c displays dispersion images derived from channels DHN (crossline or tangential) and DHZ (vertical) in nodal stations Q01–Q12 for shots Y_SOW40 and Z_SOW40, respectively, conforming to the customary multichannel analysis of surface waves (MASW) configuration used for processing Love and Rayleigh waves with nodal stations, respectively. Shots Y_SOW40 and Z_SOW40 correspond to T-Rex shaking in the *Y* and *Z* directions at location SOW40, respectively. The peak power points at each frequency are indicated by white dots. Clear fundamental mode Love- (L_0) and Rayleigh- (R_0) wave trends are visible in the dispersion images in Figure 8b and 8c, respectively, along with some potentially first-higher mode Rayleigh wave ($R1$?) trends. Although Figure 8c illustrates the more common approach to calculating Rayleigh-wave dispersion images (i.e., using the vertical components to record the wavefield from a vertical source), theoretically, Rayleigh-wave dispersion data could be obtained from a vertical excitation using either the vertical particle motion or the horizontal inline (radial) particle motion (Vantassel, Cox, *et al.*, 2022). Figure 8d,e presents dispersion images obtained from the DHE (inline) channels in nodal stations Q01–Q12 and DAS channels 629–693, respectively, for shot Z_SOW40. This configuration of vertical shaking and horizontal inline DAS channels is typically employed as a standard MASW arrangement for processing Rayleigh waves with DAS; however, the use of inline geophones for Rayleigh waves is not common practice. Nonetheless, for comparative purposes with the DAS Rayleigh-wave dispersion image, we include them in this study, as previously done by Vantassel, Cox, *et al.* (2022). Despite the disparity in the number of line Q DAS channels (65) and nodal stations (12) utilized in developing the dispersion images in Figure 8d,e, respectively, the resemblance between the resulting dispersion images is evident, which agrees with the observations of Vantassel, Cox, *et al.* (2022). Nonetheless, the DAS line offers a significant

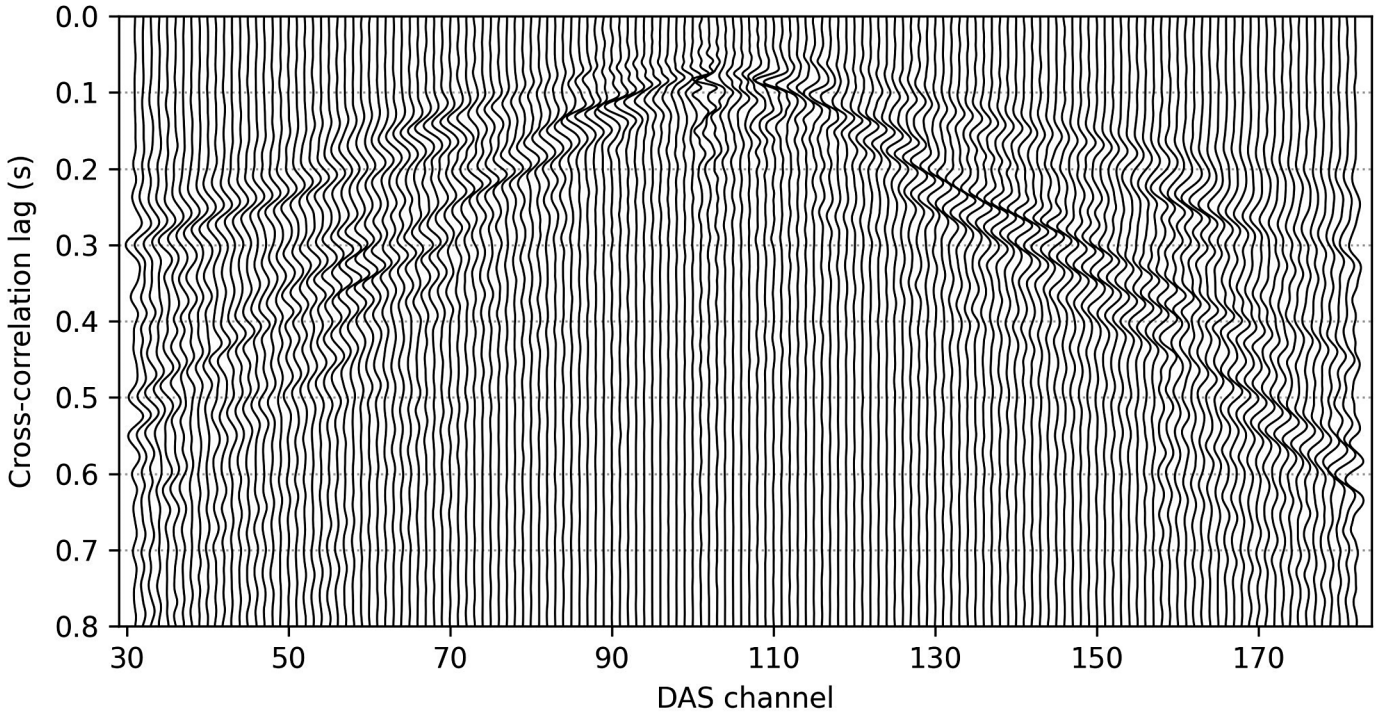


Figure 7. Waveforms recorded by DAS channels 31-183 along line 104 (refer to Fig. 2) after cross correlating with the T-Rex pilot signal for T-Rex shot Y_SIL12, and normalizing each waveform by its absolute maximum value.

advantage in such a spatially variable site because it enables 2D MASW-type processing (Yust *et al.*, 2022), which is not feasible using the nodal stations due to the limited number of stations deployed at each line. Comparing the dispersion image in Figure 8c to the ones in Figure 8d and 8e, it seems that the vertical particle motion recorded by the SmartSolo DHZ components resolves more of the apparent R_0 trend. However, there is a benefit to using both the vertical and horizontal particle motions to obtain a better understanding of the Rayleigh-wave propagation. For example, the benefits of integrating Rayleigh-wave dispersion data from both vertical and horizontal particle motions becomes apparent when examining Figure 8f, which compares the dispersion data obtained from combining the peak power trends from all four dispersion images in Figure 8b–e. This combined approach enables a clearer identification of fundamental and higher mode trends and clearly highlights the anticipated higher Love-wave phase velocities at higher frequencies, as compared to Rayleigh waves (James, 1989; Soomro *et al.*, 2016). In summary, the multidirectional shaking, multicomponent sensing, and dense configuration of horizontal DAS channels allows for more robust surface-wave dispersion processing.

The data set can also be utilized in machine learning imaging studies, which have been gaining significant interest in the last few years. For instance, recent studies have showcased the potential of applying convolutional neural networks (CNNs) to image the near surface (e.g., Vantassel, Kumar, and Cox, 2022; Abbas, Vantassel, *et al.*, 2023; Crocker *et al.*, 2023). Vantassel, Kumar, and Cox (2022) trained a CNN to take a wavefield input and generate a 2D V_S image of the near surface, whereas Abbas, Vantassel, *et al.* (2023) developed a CNN that employs dispersion images as inputs to produce 2D V_S near-surface images that was validated on field data. The high-quality waveforms and their derived dispersion images shown in Figures 6–8 demonstrate the potential of using the data set in such machine learning studies.

In the preceding paragraphs, potential use cases for the application of active-wavefield data in imaging have been presented. However, it should be noted that successful imaging of the subsurface using passive-wavefield data has also been documented in the literature. For instance, the 3C noise data could be utilized in MAM (e.g., Wathelet *et al.*, 2018) and both 2D and 3D ambient noise tomography (Wang *et al.*, 2021, 2023). In addition, the HVSR measurement technique has been demonstrated to provide valuable information about the subsurface, as discussed in the following section.

HVSR

HVSR can be used to infer the spatial variability of the fundamental site period (T_0) at each 3C nodal station. The fundamental site period can reveal important information about the depth of strong impedance contrasts beneath each nodal station (Bard and SESAME Team, 2004). For example, for a site with small dimensions, for which the overlying sediments are expected to have relatively similar velocities, stations with higher T_0 values (or lower fundamental frequency f_0) are expected to have deeper

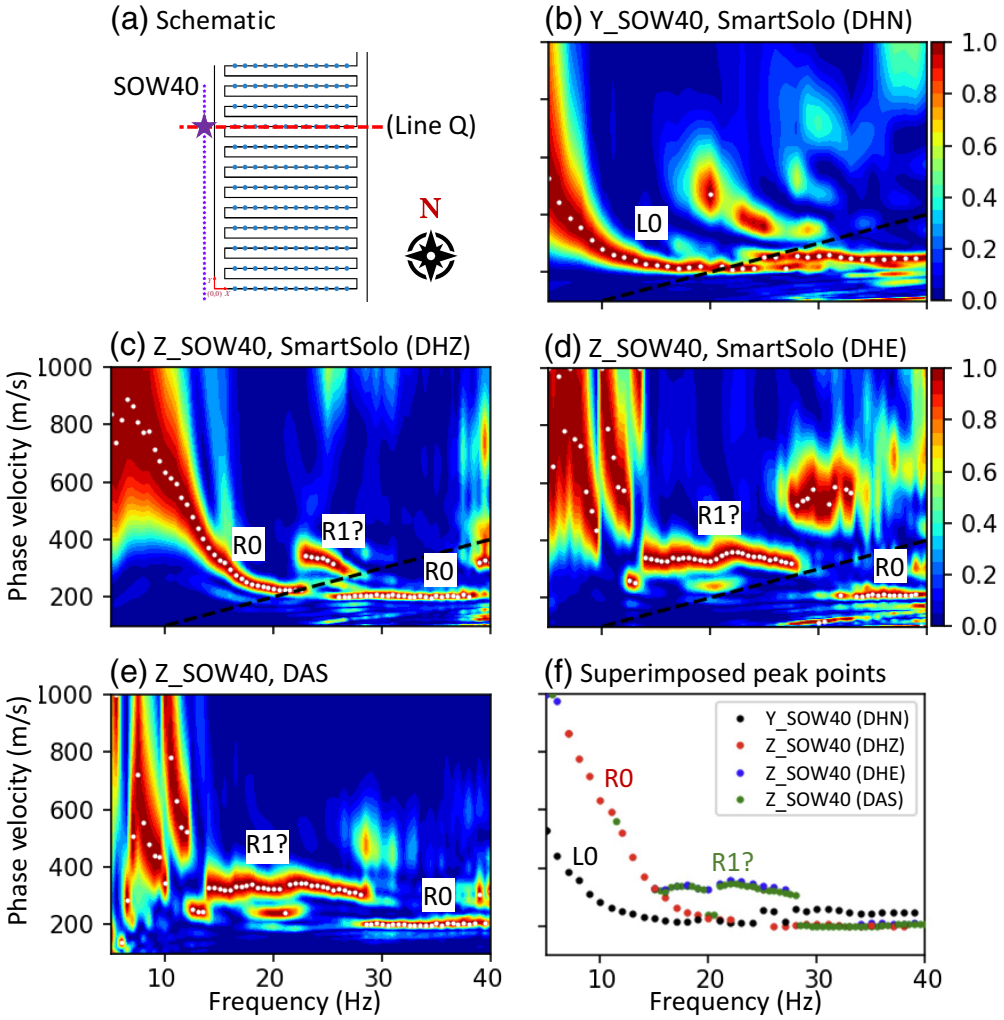


Figure 8. Dispersion images obtained from DAS and nodal stations along line Q due to T-Rex shaking at location SOW40. Panel (a) presents a schematic map highlighting the locations of shot location SOW40 and line Q. Panels (b–d) showcase dispersion images from nodal stations Q01–Q12 for shots Y_SOW40, Z_SOW40, and Z_SOW40, respectively, derived from DHN, DHZ, and DHE components, respectively. Panel (e) displays the dispersion image derived from DAS waveforms, spanning channels 629–693 located at line Q, for shot Z_SOW40. Panel (f) shows the superimposed peak power points from panels (b,c,d,e). The color version of this figure is available only in the electronic edition.

impedance contrasts, whereas stations with lower T_0 values (or higher f_0) are expected to have shallower impedance contrasts. The calculation of HVSR can be accomplished using the passive-wavefield data collected by the nodal stations. The HVSR was computed using only one hour of ambient noise recordings captured on 14 May at 4:00 UTC. The data were processed with the open-source Python package hvsrpy (Vantassel, 2021). The 1-hr-long recording for each station was divided into 30, 120-s-long time windows, and the horizontal spectra were combined using the geometric-mean, as recommended by Bard and SESAME Team (2004). Further, smoothing was performed using the filter proposed by Konno and Ohmachi (1998) with b equals 40. A color-mapped representation of the 144 stations' fundamental frequency from the HVSR lognormal median curve ($f_{0,\mathrm{mc}}$) can be found in Figure 9, indicating a notable fluctuation in $f_{0,\mathrm{mc}}$ throughout the site, with a range of values spanning from 4.19 to 7.03 Hz. For more information on the advantages of using a lognormal distribution of HVSR, please refer to Cox *et al.* (2020). HVSR amplitude with frequency plots for stations A01, M01, and W01 is also shown in Figure 9. Higher $f_{0,\mathrm{mc}}$ values are indicative of shallower depths to limestone, whereas lower $f_{0,\mathrm{mc}}$ values correspond to deeper depths. This lends additional support to the site's significant spatial variability, as concluded by Tran and Hiltunen (2011) and Tran *et al.* (2013, 2020). The HVSR data could be processed in more rigorous ways to extract more qualitative estimates for the depth to bedrock (e.g., Scherbaum *et al.*, 2003; Hobiger *et al.*, 2009; Bignardi *et al.*, 2016).

DAS reception patterns

Using 1D strain measurements to detect seismic waves can greatly impact the waves' measured phase and magnitude, owing to their directional sensitivity. The sensitivity of a DAS array to seismic waves depends on the angle at which the waves impinge on the cable and the ratio between the wavelength and the gauge length (Martin *et al.*, 2021). It is, therefore, essential to study and consider this phenomenon when using DAS for active-source seismic-wave measurements. Martin *et al.* (2021) developed a comprehensive analytical full waveform representation of pointwise and distributed strain-rate measurements for all kinds of planar surface and body waves. Similarly, Hubbard *et al.* (2022) developed numerical representations of DAS reception patterns for different source orientations and wavelength-to-gauge-length ratios. Figure 10a demonstrates their work for a wavelength-to-gauge-length ratio of five, depicting the horizontal strain measurements (ε_{xx}) resulting from X-direction Ricker wavelet excitation caused by a point force at the surface.

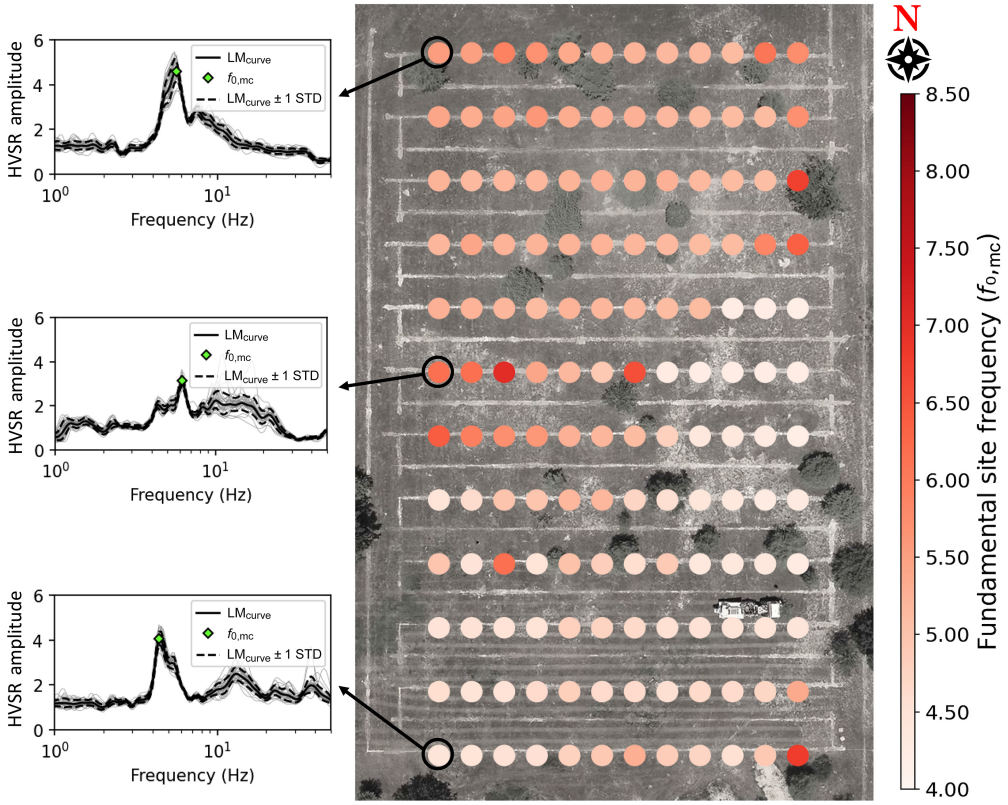
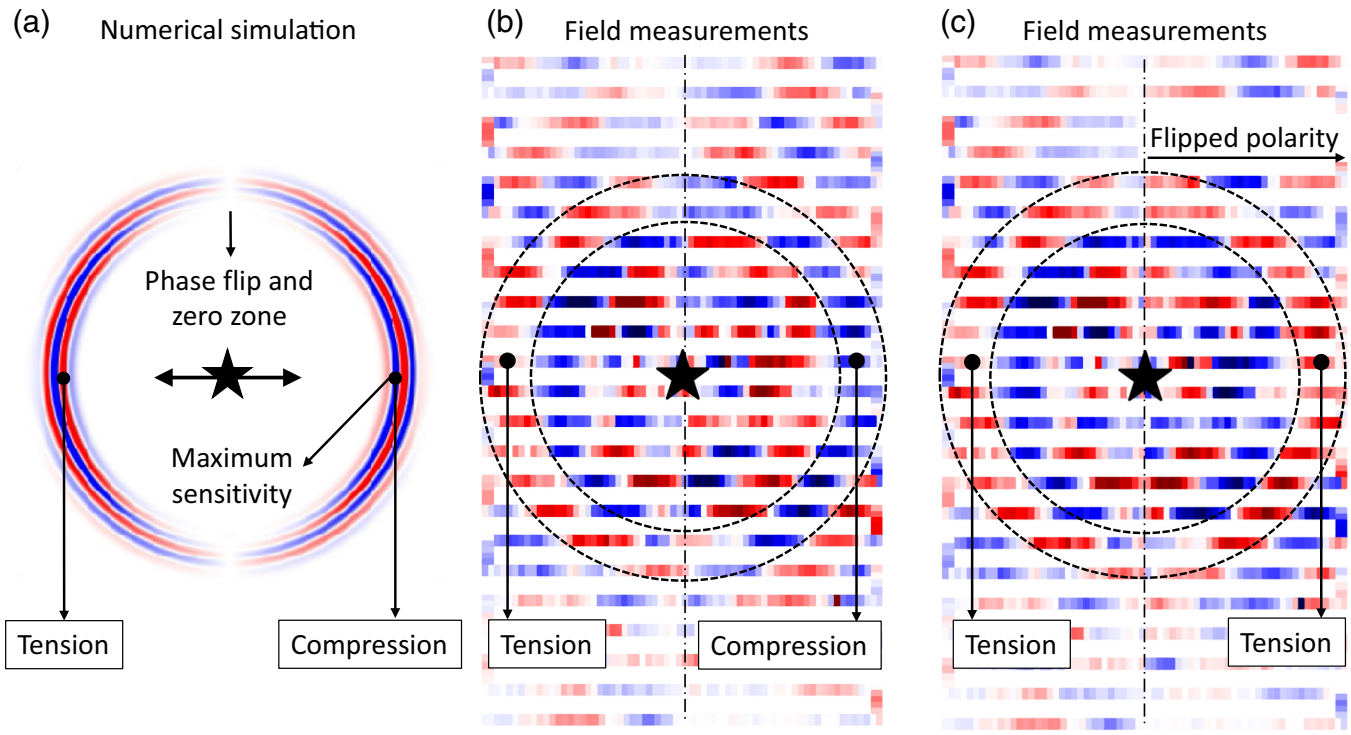


Figure 9. Spatial distribution of the fundamental site frequency (f_0), as determined by the peak of the lognormal median curve ($f_{0,mc}$), obtained from the horizontal-to-vertical spectral ratio (HVSR) analysis of one hour of ambient noise data collected at each nodal station. Detailed HVSR plots are shown for selected nodal stations (i.e., A01, M01, and W01), depicting the HVSR calculations for each time window, the lognormal median curve (LM_{curve}), the ± 1 lognormal standard deviation (STD) curves, and the fundamental site frequency from the median curve ($f_{0,mc}$). The color version of this figure is available only in the electronic edition.

The ε_{xx} values shown in Figure 10a resemble those that horizontally placed DAS cables oriented in the X direction would measure. In this representation, red indicates tension whereas blue represents compression. Notably, a distinct change in wavefield polarity is evident between the west and east sides of the shot location in Figure 10a, with a zone of zero sensitivity directly above and below the shot location (i.e., at 90° and 270° from the zero X axis). The data set documented herein offers a valuable resource for analyzing DAS reception patterns using real-field data, thanks to its abundance of shot locations and shaking directions, as well as the utilization of a dense 2D grid DAS array. Figure 10b displays a snapshot of the uncorrelated waveforms recorded on DAS channels 218–1791 at 1.9 s into the 12-s-long sweep at shot X_SIL07 (refer to Fig. 2). In Figure 10b, a clear polarity flip can be observed between the right (east) and left (west) of the shot location, particularly between the two wavefronts marked by the dotted black circles. To highlight this contrast, we inverted the polarity of all channels to the east of the shot location by multiplying their values by −1. Figure 10c shows the results from this reversal of the DAS polarity on the channels to the east of the source. After flipping the polarity of the channels to the east of the source location, the two dotted circles in Figure 10c mainly encompass a tension wave propagating away from the source, as inferred by the red cable color. By reproducing the numerical simulations of Hubbard *et al* (2022) with real-field data for only a single shot location and direction of shaking, Figure 10b underscores the potential of the data set for investigating DAS reception patterns and any potential effects of underground anomalies on them.

Conclusions

This research article outlines a comprehensive subsurface imaging experiment in Newberry, Florida, using seismic waves. The site is spatially variable and contains karstic surface and underground voids and anomalies, which have been documented in the literature. The sensing technologies used comprised a dense 2D array of 1920 horizontal-component DAS channels and a 12 × 12 grid of 144 SmartSolo 3C nodal stations, which covered an area 75 m × 155 m and were used to record both active-source and passive-wavefield data. The active-source data were generated by a variety of vibrational and impact sources: a powerful triaxial vibroseis shaker truck, a 40-kg PEG-40 kg, and an eight-pound sledgehammer. The vibroseis shaker truck was used to vibrate the ground in the three directions at 260 locations inside and outside the instrumented area, whereas the impact sources were used at 268 locations inside the instrumented area. In addition to active-source data, four hours of ambient noise were recorded using DAS, whereas the nodal stations recorded 48 hr of ambient noise in four 12-hr increments over a period of four days. The raw and processed data set, along with complete and detailed documentation of the experiment, have been archived and made publicly available. We have provided examples of the data quality and potential use cases as a means to inspire present and future researchers who need a high-quality experimental data set with known and potentially unknown anomaly locations for testing imaging methods.



Data and Resources

The data underlying this article can be found on DesignSafe (Rathje *et al.* 2017), in the work by Abbas, Cox, *et al.* (2023). The data set is publicly accessible through the following doi: 10.17603/ds2-50eh-7v93. The supplemental material to this article contains the experiment's timeline and the sequence of activities undertaken. In addition, it provides a hierarchical folder structure of the data set, accompanied by explanations of the naming conventions used during the archiving process. Furthermore, it offers insights into the force outputs and frequency contents of both T-Rex and PEG-40 kg. Finally, the supplement includes a practical data visualization example derived from the data set.

Declaration of Competing Interests

The authors acknowledge that there are no conflicts of interest recorded.

Acknowledgments

The authors would like to acknowledge FarnYuh Menq, Robert Kent, Andrew Valentine, Ruoyu Chen, and Scott Wasman for assisting in the field data acquisition. This work was supported by the U.S. National Science Foundation (NSF) Grant Numbers CMMI-2120155, CMMI-1930697, and CMMI-2037900. However, any opinions, findings, conclusions, or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the NSF.

References

Abbas, A., B. Cox, K. Tran, I. Corey, N. Dawadi, and F. Menq (2023). Active-source and passive-wavefield DAS and nodal station measurements at the Newberry Florida Site, doi: 10.17603/ds2-50eh-7v93.

Abbas, A., J. P. Vantassel, B. R. Cox, K. Kumar, and J. Crocker (2023). A frequency-velocity CNN for developing near-surface 2D vs images from linear-array, active-source wavefield measurements, *Comput. Geotech.* **156,** 105305, doi: 10.1016/j.compgeo.2023.105305.

Bard and SESAME Team (2004). Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations measurements, processing and interpretation, available at http://sesame.geopsy.org/Delivrables/Del-D23-HV_User_Guidelines.pdf (last accessed March 2023).

Belfer, I., I. Bruner, S. Keydar, A. Kravtsov, and E. Landa (1998). Detection of shallow objects using refracted and diffracted seismic waves, *J. Appl. Geophys.* **38,** 155–168.

Bignardi, S., A. Mantovani, and N. ZeidAbu (2016). OpenHVSR: Imaging the subsurface 2D/3D elastic properties through multiple

Figure 10. Reception patterns of DAS due to a horizontal excitation at the ground surface, as obtained from a numerical simulation in panel (a) and field strain measurements in panels (b,c). Panel (a) displays the numerical x direction surface strain (ε_{xx}) results, with compression shown in blue and tension in red, obtained from an elastic half-space excited by a Ricker wavelet in the X direction with wavelength-to-gauge length ratio (λ/g) of five, after Hubbard *et al.* (2022). Panel (b) displays the DAS field measurements, as detected by channels 218–1791, 1.9 s after the initiation of T-Rex-induced shaking in the X direction at location SIL07. Panel (b) also highlights a clear reversal of polarity between the west and east sides of the shot, which is best observed by examining the colors between the wavefronts indicated by dotted lines. To accentuate the distinction, Panel (c) replicates panel (b) after the polarity of channels to the east of the shot location have been flipped. The color version of this figure is available only in the electronic edition.

HVSR modeling and inversion, *Comput. Geosci.* **93,** 103–113, doi: 10.1016/j.cageo.2016.05.009.

Branham, K., and D. Steeples (1988). Cavity detection using high-resolution seismic reflection methods, *Min. Eng.* **40,** available at https://www.osti.gov/biblio/5243315 (last accessed February 2023).

Castongia, E., H. F. Wang, N. Lord, D. Fratta, M. Mondanos, and A. Chalari (2017). An experimental investigation of distributed acoustic sensing (DAS) on Lake Ice, *J. Environ. Eng. Geophys.* **22,** 167–176, doi: 10.2113/JEEG22.2.167.

Cheng, T., B. R. Cox, J. P. Vantassel, and L. Manuel (2020). A statistical approach to account for azimuthal variability in single-station HVSR measurements, *Geophys. J. Int.* **223,** 1040–1053, doi: 10.1093/gji/ggaa342.

Cheng, T., M. M. Hallal, J. P. Vantassel, and B. R. Cox (2021). Estimating unbiased statistics for fundamental site frequency using spatially distributed HVSR measurements and voronoi tessellation, *J. Geotech. Geoenviron. Eng.* **147,** doi: 10.1061/(ASCE)GT.1943-5606.0002551.

Cook, J. C. (1965). Seismic mapping of underground cavities using reflection amplitudes, *Geophysics* **30,** 527–538, doi: 10.1190/1.1439618.

Cox, B., P. Wills, D. Kiyashchenko, J. Mestayer, J. Lopez, S. Bourne, R. Lupton, G. Solano, N. Henderson, D. Hill, *et al.* (2012). Distributed acoustic sensing for geophysical measurement, monitoring and verification, *CSEG Recorder* **37,** no. 2, 7–13.

Cox, B. R., T. Cheng, J. P. Vantassel, and L. Manuel (2020). A statistical representation and frequency-domain window-rejection algorithm for single-station HVSR measurements, *Geophys. J. Int.* **221,** 2170–2183, doi: 10.1093/gji/ggaa119.

Crocker, J., K. Kumar, and B. Cox (2023). Using explainability to design physics-aware CNNs for solving subsurface inverse problems, *Comput. Geotech.* **159,** 105452, doi: 10.1016/j.compgeo.2023.105452.

Daley, T. M., D. E. Miller, K. Dodds, P. Cook, and B. M. Freifeld (2016). Field testing of modular borehole monitoring with simultaneous distributed acoustic sensing and geophone vertical seismic profiles at Citronelle, Alabama, *Geophys. Prospect.* **64,** 1318–1334, doi: 10.1111/1365-2478.12324.

Dou, S., N. Lindsey, A. M. Wagner, T. M. Daley, B. Freifeld, M. Robertson, J. Peterson, C. Ulrich, E. R. Martin, and J. B. Ajo-Franklin (2017). Distributed acoustic sensing for seismic monitoring of the near surface: A traffic-noise interferometry case study, *Sci. Rep.* **7,** doi: 10.1038/s41598-017-11986-4.

Fathi, A., B. Poursartip, K. H. Stokoe, and L. F. Kallivokas (2016). Three-dimensional P- and S-wave velocity profiling of geotechnical sites using full-waveform inversion driven by field data, *Soil Dynam. Earthq. Eng.* **87,** 63–81, doi: 10.1016/j.soildyn.2016.04.010.

Federation of Digital Seismographic Networks (FDSN) (2012). SEED reference manual standard for the exchange of earthquake data SEED Format Version 2.4, available at https://www.fdsn.org/pdf/SEEDManual_V2.4.pdf (last accessed February 2023).

Foti, S., F. Hollender, F. Garofalo, D. Albarello, M. Asten, P.-Y. Bard, C. Comina, C. Cornou, B. Cox, G. Di Giulio, *et al.* (2018). Guidelines for the good practice of surface wave analysis: a product of the InterPACIFIC project, *Bull. Earthq. Eng.* **16,** 2367–2420, doi: 10.1007/s10518-017-0206-7.

Grandjean, G., and D. Leparoux (2004). The potential of seismic methods for detecting cavities and buried objects: Experimentation at a test site, *J. Appl. Geophys.* **56,** 93–106, doi: 10.1016/j.jappgeo.2004.04.004.

Hartog, A. H. (2018). *An Introduction to Distributed Optical Fibre Sensors*, Boca Raton, Florida.

Hobiger, M., P.-Y. Bard, C. Cornou, and N. BihanLe (2009). Single station determination of Rayleigh wave ellipticity by using the random decrement technique (RayDec), *Geophys. Res. Lett.* **36,** L14303, doi: 10.1029/2009GL038863.

Hubbard, P. G., J. P. Vantassel, B. R. Cox, J. W. Rector, M. B. S. Yust, and K. Soga (2022). Quantifying the surface strain field induced by active sources with distributed acoustic sensing: Theory and practice, *Sensors* **22,** doi: 10.3390/s22124589.

James, D. E. (1989). Surface waves, in *Encyclopedia of Earth Science Geophysics*, Springer, Boston, Massachusetts, 1255–1266, doi: 10.1007/0-387-30752-4_150.

Kolesnikov, Y. I., and K. V. Fedin (2018). Detecting underground cavities using microtremor data: Physical modelling and field experiment, *Geophys. Prospect.* **66,** 342–353, doi: 10.1111/1365-2478.12540.

Konno, K., and T. Ohmachi (1998). Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor, *Bull. Seismol. Soc. Am.* **88,** 228–241, doi: 10.1785/BSSA0880010228.

Kristekova, M., J. Kristek, P. Moczo, and P. Labak (2020). The finite-interval spectral power method for detecting underground cavities using seismic ambient noise, *Geophys. J. Int.* **224,** 945–960, doi: 10.1093/gji/ggaa494.

Lancelle, C., N. Lord, H. Wang, D. Fratta, R. Nigbor, A. Chalari, R. Karaulanov, J. Baldwin, and E. Castongia (2014). Sample data from a distributed acoustic sensing experiment at garner valley, California, *AGU Fall Meeting*, doi: 10.15121/1177104.

Loehr, E., L. Alan, R. Brent, and B. Andrew (2016). Geotechnical site characterization. Geotechnical engineering circular NO. 5, FHWA NHI-16-072, Federal Highway Administration, Washington, D.C., available at https://www.fhwa.dot.gov/engineering/geotech/pubs/nhi16072.pdf (last accessed April 2023).

Martin, E. R., N. J. Lindsey, J. B. Ajo-Franklin, and B. L. Biondi (2021). Introduction to interferometry of fiber-optic strain measurements, in *Distributed Acoustic Sensing in Geophysics: Methods and Applications*, Y. Li, M. Karrenbach, and J. B. Ajo-Franklin (Editors), American Geophysical Union and John Wiley & Sons, Inc., 111–129, doi: 10.1002/9781119521808.ch9.

Mirzanejad, M., K. T. Tran, M. McVay, D. Horhota, and S. Wasman (2020). Sinkhole detection with 3D full seismic waveform tomography, *Geophysics* **85,** no. 5, B147–B157.

Nakamura, Y. (1989). A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface, *Quarterly Rept. RTRI*, Vol. 30, Railway Technical Research Institute/Tetsudo Gijutsu Kenkyujo, available at http://www.rtri.or.jp/eng/ (last accessed February 2023).

Obermann, A., P. Sánchez-Pastor, S.-M. Wu, C. Wollin, A. F. Baird, M. P. Isken, J. Clinton, B. P. Goertz-Allmann, T. Dahm, A. Wuestefeld, *et al.* (2022). Combined Large- *N* seismic arrays and DAS fiber optic cables across the hengill geothermal field, Iceland, *Seismol. Res. Lett.* **93,** 2498–2514, doi: 10.1785/0220220073.

Pan, Y., J. Xia, Y. Xu, L. Gao, and Z. Xu (2016). Love-wave waveform inversion in time domain for shallow shear-wave velocity, *Geophysics* **81,** R1–R14, doi: 10.1190/geo2014-0225.1.

Parker, M., H. Thurber, X. Zeng, P. Li, E. Lord, D. Fratta, F. Wang, C. Robertson, M. Thomas, S. Karplus, *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,** no. 5, 1629–1640, doi: 10.1785/0220180085.

Pernod, P., B. Piwakowski, B. Delannoy, and J. C. Tricot (1989). Detection of shallow underground cavities by seismic methods: physical modelling approach, in *Acoustical Imaging*, H. Shimizu, N. Chubachi, and K. Ji (Editors), Springer US, Boston, Massachusetts, 705–713, doi: 10.1007/978-1-4613-0791-4_74.

Rathje, E. M., C. Dawson, J. E. Padgett, J.-P. Pinelli, D. Stanzione, A. Adair, P. Arduino, S. J. Brandenberg, T. Cockerill, C. Dey, *et al.* (2017). DesignSafe: New cyberinfrastructure for natural hazards engineering, *Nat. Hazards Rev.* **18,** doi: 10.1061/(asce)nh.1527-6996.0000246.

Scherbaum, F., K.-G. Hinzen, and M. Ohrnberger (2003). Determination of shallow shear wave velocity profiles in the Cologne, Germany area using ambient vibrations, *Geophys. J. Int.* **152,** 597–612, doi: 10.1046/j.1365-246X.2003.01856.x.

Sheriff, R. E., and L. P. Geldart (1995). *Exploration Seismology*, Cambridge University Press, New York, doi: 10.1017/CBO9781139168359.

Sloan, S. D., S. L. Peterie, J. Ivanov, R. D. Miller, and J. R. McKenna (2010). 12. Void Detection Using Near-Surface Seismic Methods, in *Advances in Near-surface Seismology and Ground-penetrating Radar*, R. D. Miller, J. H. Bradford, and K. Holliger (Editors), Society of Exploration Geophysicists, American Geophysical Union, Environmental and Engineering Geophysical Society, 201–218, doi: 10.1190/1.9781560802259.ch12.

Sloan, S. D., S. L. Peterie, R. D. Miller, J. Ivanov, J. R. McKenna, S. W. Broadfoot, and O. M. Metheny (2012). Tunnel detection using near-surface seismic methods, *SEG Technical Program Expanded Abstracts 2012*, Society of Exploration Geophysicists, 1–5, doi: 10.1190/segam2012-1442.1.

Smith, J. A., D. Borisov, H. Cudney, R. D. Miller, R. Modrak, M. Moran, S. L. Peterie, S. D. Sloan, J. Tromp, and Y. Wang (2019). Tunnel detection at Yuma Proving Ground, Arizona, USA — Part 2: 3D full-waveform inversion experiments, *Geophysics* **84,** B107–B120, doi: 10.1190/geo2018-0599.1.

Soomro, R. A., C. Weidle, L. Cristiano, S. Lebedev, and T. Meier (2016). Phase velocities of Rayleigh and Love waves in central and northern Europe from automated, broad-band, interstation measurements, *Geophys. J. Int.* **204,** 517–534, doi: 10.1093/gji/ggv462.

Stokoe, K. H., B. R. Cox, P. M. Clayton, and F. Menq (2020). NHERI@UTexas experimental facility with large-scale mobile shakers for field studies, *Front. Built Environ.* **6,** doi: 10.3389/fbuil.2020.575973.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation, *Geophysics* **49,** no. 8, 1140–1395.

Tran, K. T., and D. R. Hiltunen (2011). Inversion of first-arrival time using simulated annealing, *J. Environ. Eng. Geophys.* **16,** 25–35, doi: 10.2113/JEEG16.1.25.

Tran, K. T., M. McVay, M. Faraone, and D. Horhota (2013). Sinkhole detection using 2D full seismic waveform tomography, *Geophysics* **78,** R175–R183, doi: 10.1190/geo2013-0063.1.

Tran, K. T., T. D. Nguyen, D. R. Hiltunen, K. Stokoe, and F. Menq (2020). 3D full-waveform inversion in time-frequency domain: Field data application, *J. Appl. Geophys.* **178,** 104078, doi: 10.1016/j.jappgeo.2020.104078.

Upchurch, S., T. Scott, M. Alfieri, B. Fratesi, and T. Dobecki (2019). *The Karst Systems of Florida: Understanding Karst in a Geologically Young Terrain*, Springer Nature, New York, New York.

Vantassel, J. (2021). jpvantassel/hvsrpy: v1.0.0, *Zenodo*, doi: 10.5281/zenodo.5563211.

Vantassel, J. (2022). jpvantassel/swprocess: v0.1.1, doi: 10.5281/zenodo.6481915.

Vantassel, J. P., and B. R. Cox (2022). SWprocess: A workflow for developing robust estimates of surface wave dispersion uncertainty, *J. Seismol.* **26,** 731–756, doi: 10.1007/s10950-021-10035-y.

Vantassel, J. P., B. R. Cox, P. G. Hubbard, and M. Yust (2022). Extracting high-resolution, multi-mode surface wave dispersion data from distributed acoustic sensing measurements using the multichannel analysis of surface waves, *J. Appl. Geophys.* **205,** doi: 10.1016/j.jappgeo.2022.104776.

Vantassel, J. P., K. Kumar, and B. R. Cox (2022). Using convolutional neural networks to develop starting models for near-surface 2-D full waveform inversion, *Geophys. J. Int.* **231,** 72–90, doi: 10.1093/gji/ggac179.

Wang, Y., M. Khorrami, K. T. Tran, and D. Horhota (2023). Application of ambient noise tomography for deep void detection, *J. Appl. Geophys.* **209,** 104922, doi: 10.1016/j.jappgeo.2022.104922.

Wang, Y., R. D. Miller, S. L. Peterie, S. D. Sloan, M. L. Moran, H. H. Cudney, J. A. Smith, D. Borisov, R. Modrak, and J. Tromp (2019). Tunnel detection at Yuma Proving Ground, Arizona, USA — Part 1: 2D full-waveform inversion experiment, *Geophysics* **84,** B95–B105, doi: 10.1190/geo2018-0598.1.

Wang, Y., K. T. Tran, and D. Horhota (2021). Road sinkhole detection with 2D ambient noise tomography, *Geophysics* **86,** KS123–KS135, doi: 10.1190/geo2020-0739.1.

Wathelet, M., B. Guillier, P. Roux, C. Cornou, and M. Ohrnberger (2018). Rayleigh wave three-component beamforming: Signed ellipticity assessment from high-resolution frequency-wavenumber processing of ambient vibration arrays, *Geophys. J. Int.* **215,** 507–523, doi: 10.1093/gji/ggy286.

Yu, C., Z. Zhan, N. J. Lindsey, J. B. Ajo-Franklin, and M. Robertson (2019). The potential of DAS in teleseismic studies: Insights from the goldstone experiment, *Geophys. Res. Lett.* **46,** 1320–1328, doi: 10.1029/2018GL081195.

Yust, M., B. R. Cox, J. P. Vantassel, and P. G. Hubbard (2022). DAS for 2D MASW Imaging: A case study on the benefits of flexible subarray processing, available at http://arXiv.org/abs/2210.14261 (last accessed February 2023).

Yust, M. B. S., B. R. Cox, J. P. Vantassel, P. G. Hubbard, C. Boehm, and L. Krischer (2023). Near-surface 2D imaging via FWI of DAS Data: An examination on the impacts of FWI starting model, *Geosciences* **13,** 63, doi: 10.3390/geosciences13030063.

Zywicki, J., and J. Rix (2005). Mitigation of near-field effects for seismic surface wave velocity estimation with cylindrical beamformers, *J. Geotech. Geoenviron. Eng.* **131,** no. 8, 970–977.

Manuscript received 5 July 2023
Published online 2 January 2024