8-2024

Emerging Technologies and Advanced Analyses for Non-Invasive Near-Surface Site Characterization

Aser Abbas
Utah State University

EMERGING TECHNOLOGIES AND ADVANCED ANALYSES FOR NON-INVASIVE NEAR-SURFACE SITE CHARACTERIZATION

by

Aser Abbas

A dissertation proposal submitted in partial fulfillment
of the requirements for the degree
of

DOCTOR OF PHILOSOPHY

in

Civil and Environmental Engineering

Approved:

_____	_____
Brady Cox, Ph.D. Major Professor	Marv Halling, Ph.D. Committee Member
_____	_____
James Bay, Ph.D. Committee Member	Tony Lowry, Ph.D. Committee Member
_____	_____
John Rice, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2024

ABSTRACT

Emerging Technologies and Advanced Analyses for Non-Invasive Near-Surface Site Characterization

by

Aser Abbas, Doctor of Philosophy.

Utah State University, 2024

Major Professor: Dr. Brady R. Cox
Department: Civil and Environmental Engineering

Recent years have witnessed a growing interest in non-invasive subsurface imaging techniques for determining the small-strain shear modulus (G_{max}) and damping ratio (D) through seismic wave propagation. This dissertation explores two primary domains: analysis techniques for estimating G_{max} and D, and advancements in data acquisition systems. For estimating G_{max}, a novel frequency-velocity domain convolutional neural network (CNN) is developed to rapidly and non-invasively image the near-surface geo-materials' shear wave velocity (Vs). Operating in the frequency-velocity domain allows for significant flexibility in the linear-array, active-source experimental testing configurations used for generating the CNN input, which are normalized dispersion images. While normalized dispersion images retain the most important aspects of near-surface wavefields, they are relatively insensitive to the exact experimental testing configuration used to generate and record the wavefields, accommodating various source types, source offsets, numbers of receivers, and receiver

spacings. The CNN was validated through synthetic simulations and blind testing using field data collected at a site in Austin, Texas, USA.

To develop a comprehensive profile of the soil small-strain parameters, a methodology is presented for estimating frequency-dependent attenuation coefficients. These coefficients, alongside phase velocity dispersion data, can be utilized in a joint inversion to determine the *Vs* and *D* profiles of the subsurface. The methodology relies on recording ambient vibrations using two-dimensional (2D) arrays of surface seismic sensors and employs an attenuation-specific wavefield conversion and frequency-domain beamforming. Comparisons with synthetic and field data demonstrate the efficacy of the developed approach. The dissertation also introduces an open-access comprehensive subsurface imaging experiment conducted in Newberry, Florida, where distributed acoustic sensing (DAS) and three-component nodal stations were employed to record seismic waves generated from both active sources and ambient vibrations. The active-source data were generated using a vibroseis shaker truck and impact sources. The dissertation provides detailed information on the experiment layout, acquisition parameters, processing steps, and data availability.

Finally, the dissertation concludes with a review of recent advancements in 1D and 2D imaging techniques, including machine learning, full waveform inversion (FWI), and 2D multichannel analysis of surface waves (MASW). These methods were applied at the same site, enabling direct comparison between them.

(255 pages)

PUBLIC ABSTRACT

Emerging Technologies and Advanced Analyses for Non-Invasive Near-Surface Site Characterization

Aser Abbas

This dissertation introduces novel techniques for estimating the soil small-strain shear modulus (G_{max}) and damping ratio (D), crucial for modeling soil behavior in various geotechnical engineering problems. For G_{max} estimation, a machine learning approach is proposed, capable of generating two-dimensional (2D) images of the subsurface shear wave velocity, which is directly related to G_{max}. The dissertation also presents a method for estimating frequency dependent attenuation coefficients from ambient vibrations collected using 2D arrays of seismic sensors deployed across the ground surface. These attenuation coefficients can then be used in an inversion process to estimate D. The developed techniques for G_{max} and D estimation have undergone rigorous validation and testing through synthetic simulations and field experiments, demonstrating their effectiveness. Furthermore, the dissertation presents a comprehensive dataset collected using cutting-edge seismic sensing technologies, including distributed acoustic sensing, three-component seismometers, and a large mobile shaker truck. This dataset has been archived and made publicly available, aiding researchers worldwide in developing and testing new non-invasive imaging techniques. Finally, the dissertation concludes with a review and comparison of recent advancements in non-invasive subsurface imaging techniques and their application at the same site.

DEDICATION

To my beloved Parents and Siblings,

To my cherished Wife,

To the inspiring Teachers, Mentors, and Fellow Students,

And to the gracious Donors whose support made my scholarships and awards possible

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my PhD advisor, Professor Brady Cox, for his unwavering support, invaluable guidance, and profound insights throughout my doctoral journey. His mentorship has been instrumental in shaping both my research and personal growth.

I am also profoundly grateful to my PhD committee members, Professors Marvin Halling, James Bay, John Rice, and Anthony Lowry, for their expertise, constructive feedback, and encouragement at every stage of my dissertation development. Their collective wisdom has significantly enriched the quality of my work.

I extend my sincere appreciation to my collaborators, Sebastiano Foti, Mauro Aimar, Joseph Vantassel, Krishna Kumar, Khiem Tran, Isabella Corey, Nishkarsha Dawadi, Michael Yust, and Jodie Crocker for their collaboration, support, and contributions to various aspects of my research. Their diverse perspectives and expertise have greatly enhanced the depth and scope of my work.

Furthermore, I am indebted to the generous funding provided by the U.S. National Science Foundation (NSF) through Grant Numbers CMMI-2120155, CMMI-1930697, and CMMI-2037900. This support has been crucial in enabling me to pursue my research goals and achieve significant milestones in my academic journey.

Lastly, I would like to express my heartfelt thanks to my family, friends, and loved ones for their encouragement throughout this endeavor. Their unconditional support has been my source of strength and motivation.

This work would not have been possible without the collective contributions and support of all those mentioned above. Thank you.

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

1.1 Background

Non-invasive geophysical imaging techniques utilizing seismic wave propagation have garnered escalating attention in recent decades owing to their remarkable cost-effectiveness compared to conventional invasive site characterization methods and their potential to cover large areas. These methods capitalize on the fact that the characteristics of propagating seismic waves in the subsurface are associated with the medium's mechanical parameters through simple relationships. Geophysical testing often involves subsurface strain levels that are very small. For minimal strains, soils typically display a linear stress–strain relationship, which can be modeled using the theory of linear viscoelasticity. Linear viscoelasticity is the simplest constitutive theory capable of capturing the mechanical response of geomaterials undergoing low-amplitude dynamic oscillations, showcasing their ability to simultaneously store and dissipate strain energy over a finite period of time (Lai and Özcebe, 2016). Wave propagation in a linear viscoelastic media can be completely described using the real-valued primary or compressional wave (P-wave) velocities (*Vp*) and the secondary or shear wave (S-wave) velocities (*Vs*), along with their associated attenuation coefficients, α_P and α_S (Lai and Rix, 2002). In fully saturated, near-surface soil deposits, *Vp* is mainly controlled by the compressibility of water in the soil's pore structure and as a result is much faster than the *Vp* of dry soils (Foti et al., 2014). Hence, *Vp* is less revealing of the soil's structure when

the materials are relatively soft and saturated. Conversely, *Vs* is mainly a function of the soil skeleton parameters, with the influence of pore fluid on S-wave propagation being negligible, as pore fluid lacks shearing resistance and their influence is linked to changes in mass density (Biot, 1956a, 1956b; Foti et al., 2014). Therefore, for geotechnical engineering purposes, geophysical imaging techniques primarily focus on capturing two crucial soil parameters: *Vs* and the small-strain damping ratio (*D*). *Vs* is directly related to the small-strain shear modulus (G_{max}), representing the stiffness of the soil, while *D* quantifies the soil internal energy dissipation at low strains. Understanding these properties is vital for predicting ground shaking during seismic events, modeling ground-borne vibrations' impact on structures, and predicting deformation under static loads.

1.2 Research Objectives

The overall objective of the dissertation is to present new innovations in seismic waves data acquisition (DAQ) and introduce novel analysis techniques for estimating G_{max} and *D*.

To estimate G_{max}, a novel frequency-velocity convolutional neural network (CNN) is presented for rapid, non-invasive, 2D-*Vs* imaging of near-surface geo-materials. Operating in the frequency-velocity domain offers significant flexibility in the linear-array, active-source experimental testing configurations used to generate the CNN input, which are normalized dispersion images. These images retain crucial aspects of near-surface wavefields while accommodating various source types, offsets, receiver numbers, and spacings in the testing configuration. The ultimate vision of this imaging approach is to provide a cost-effective and accurate alternative to current pseudo-2D multichannel analysis of surface waves (MASW) imaging, or to develop 2D starting models for more

rigorous imaging methods such as full waveform inversion (FWI) (Tarantola, 1984; Mora, 1987), especially when fully developed for a wide range of geological conditions.

In regard to estimating *D*, the dissertation introduces a methodology for estimating frequency-dependent attenuation coefficients through the analysis of ambient noise wavefield data recorded by 2D arrays of surface seismic sensors. The approach relies on the application of an attenuation-specific wavefield conversion and frequency-domain beamforming (FDBF). Using an FDBF approach enables the direction of ambient noise propagation to be determined for each noise window and frequency, and does not require an equally-partitioned ambient noise wavefield such as in the case of noise cross-correlation methods (Sánchez-Sesma and Campillo, 2006; Snieder et al., 2007). The experimental dispersion and attenuation data can then be inverted to determine not only the *Vs* profile but also the *D* profile of the subsurface to greater depths.

The dissertation also presents a comprehensive and open-access dataset of seismic wave recordings gathered using some of the most advanced technologies available in geophysical-noninvasive subsurface imaging. A test site in Newberry, Florida, USA, was selected for this extensive subsurface imaging experiment due to its complex geology, which includes many known and unknown karstic voids of variable size and depth. A 2D layout of distributed acoustic sensing (DAS) fiber optic cable and a 2D array of three-component (3C) geophone nodal stations covering an area approximately 155 m x 75 m were used at the site to record both active-source and passive-wavefield seismic waves. The active sources used to initiate seismic wave propagation comprised both a broadband, three-dimensional, vibroseis shaker truck named T-Rex from the NHERI@UTexas experimental facility (Stokoe et al., 2020), and more-variable impact

sources. In total, approximately 2 km of DAS fiber optic cable and 144, 3C nodal stations were used to record wavefields from more than 367 shot locations. This unique and publicly accessible dataset is available on DesignSafe (Rathje et al., 2017; https://www.designsafe-ci.org/) under project PRJ-3521, “Active-source and Passive-wavefield DAS and Nodal Station Measurements at the Newberry Florida Site”. This dataset will serve as a valuable resource for researchers seeking to explore novel approaches for anomaly detection and subsurface imaging.

Lastly, the dissertation delves into a detailed examination of some of the most recent advancements in imaging techniques, encompassing both 1D and 2D methodologies. Within the realm of 1D imaging, the dissertation showcases the use of DAS as the DAQ for 1D MASW, as well as the joint estimation of phase velocity and phase attenuation data within a 1D MASW test setup, achievable through either geophones or DAS as the DAQ. Furthermore, it presents the application of DAS for simultaneously characterizing the stiffness and dissipative parameters of a test site. Moving into the domain of 2D imaging, the dissertation explores a range of techniques, including machine learning, 2D MASW utilizing DAS data, and FWI with DAS data. These state-of-the-art methodologies were successfully applied and tested at the Hornsby Bend test site at Austin, Texas, USA, a thoroughly documented case study location, providing an invaluable opportunity for comparative analysis and discussion of their findings.

1.3 Dissertation Organization

The dissertation comprises six chapters, four of which are presented in paper formats. Chapter 1 serves as a general introduction, delineating the significance of two

critical small strain soil parameters in geotechnical engineering: G_{max} and D. Additionally, it outlines the methodologies proposed within this dissertation to measure and process seismic wave data for their evaluation. In Chapter 2, a frequency-velocity CNN is introduced for the development of near-surface 2D-*Vs* images using linear-array, active-source wavefield measurements. Chapter 3 introduces a frequency-domain beamforming procedure for extracting Rayleigh wave attenuation coefficients from ambient noise recordings of 2D microtremor array measurements (MAM) arrays. Chapter 4 describes an open-access dataset comprising active-source and passive-wavefield DAS and nodal station measurements conducted at a site in Newberry, Florida, USA. Chapter 5 provides a comprehensive overview of emerging technologies and advanced analyses pertinent to non-invasive near-surface site characterization. Finally, Chapter 6 encapsulates a summary and conclusions derived from the research presented throughout the dissertation.

1.4 References

Biot, M. A., 1956a. Theory of Propagation of Elastic Waves in a Fluid-Saturated Porous Solid. I. Low-Frequency Range. The Journal of the Acoustical Society of America, 28(2).

Biot, M. A., 1956. Theory of Propagation of Elastic Waves in a Fluid-Saturated Porous Solid. II. Higher Frequency Range. The Journal of the Acoustical Society of America, 28(2).

Foti, S., Lai, C. G., Rix, G. J., Strobbia, C., 2014. Surface Wave Methods for Near-Surface Site Characterization. London: CRC Press.

Lai, C. G., & Özcebe, A. G. (2016). Causal Damping Ratio Spectra and Dispersion Functions in Geomaterials from the Exact Solution of Kramers-Kronig Equations of Viscoelasticity. Continuous Media with Microstructure 2. doi:DOI 10.1007/978-3-319-28241-1_24

Lai, C. G., & Rix, G. J. (2002). Media, Solution of the Rayleigh Eigenproblem in Viscoelastic. Bulletin of the Seismological Society of America, 92(6), 297–2309.

Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data. Geophysics, 52(9).

Rathje, E. M. et al., 2017. DesignSafe: New Cyberinfrastructure for Natural Hazards Engineering. Natural Hazards Review, 18(3).

Sanchez-Sesma, F. J., & Campillo, Mi. (2006). Retrieval of the Green's Function from Cross Correlation: The Canonical Elastic Problem. Bulletin of the Seismological Society of America, 96(3), 1182–1191. https://doi.org/10.1785/0120050181

Snieder, R., Wapenaar, K., & Wegler, U. (2007). Unified Green's function retrieval by cross-correlation; connection with energy principles. Physical Review E, 75(3), 036103. https://doi.org/10.1103/PhysRevE.75.036103

Stokoe, K., Cox, B., Clayton, P., Menq, F., 2020. NHERI@UTexas Experimental Facility With Large-Scale Mobile Shakers for Field Studies. Frontiers in Built Environment.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation. Geophysics, 49(8), pp. 1140-1395.

CHAPTER 2

A FREQUENCY-VELOCITY CNN FOR DEVELOPING NEAR-SURFACE 2D VS IMAGES FROM LINEAR-ARRAY, ACTIVE-SOURCE WAVEFIELD MEASUREMENTS

Abstract

This paper presents a frequency-velocity convolutional neural network (CNN) for rapid, non-invasive 2D shear wave velocity (V_S) imaging of near-surface geo-materials. Operating in the frequency-velocity domain allows for significant flexibility in the linear-array, active-source experimental testing configurations used for generating the CNN input, which are normalized dispersion images. Unlike wavefield images, normalized dispersion images are relatively insensitive to the experimental testing configuration, accommodating various source types, source offsets, numbers of receivers, and receiver spacings. We demonstrate the effectiveness of the frequency-velocity CNN by applying it to a classic near-surface geophysics problem, namely, imaging a two-layer, undulating, soil-over-bedrock interface. This problem was recently investigated in our group by developing a time-distance CNN, which showed great promise but lacked flexibility in utilizing different field-testing configurations. Herein, the new frequency-velocity CNN is shown to have comparable accuracy to the time-distance CNN while providing greater flexibility to handle varied field applications. The frequency-velocity CNN was trained, validated, and tested using 100,000 synthetic near-surface models. The ability of the proposed frequency-velocity CNN to generalize across various acquisition configurations is first tested using synthetic near-surface models with different acquisition configurations

from that of the training set, and then applied to experimental field data collected at the Hornsby Bend site in Austin, Texas, USA. When fully developed for a wider range of geological conditions, the proposed CNN may ultimately be used as a rapid, end-to-end alternative for current pseudo-2D surface wave imaging techniques or to develop starting models for full waveform inversion.

Keywords: machine learning; CNN; subsurface imaging; surface waves

2.1 Introduction

Non-invasive subsurface imaging techniques based on stress wave propagation have gained increased interest over the past few decades due to their significant cost savings over traditional invasive site characterization methods and their potential to cover large areas. The current study proposes a frequency-velocity domain, deep-learning technique for rapid, non-invasive 2D shear wave velocity (V_S) imaging of near-surface geo-materials. The ultimate vision of this type of imaging approach is that when fully developed for a wide range of geological conditions, it may be used as a cost-effective and accurate alternative to current pseudo-2D multichannel analysis of surface waves (MASW) (e.g., Park 2005) imaging, or in developing 2D starting models for more rigorous imaging methods like full waveform inversion (FWI) (Tarantola, 1984; Mora,1987). The information provided below illustrates the need for new approaches to near-surface site characterization, presents background information on previous work related to deep-learning subsurface imaging, and explains why a frequency-velocity domain approach is desirable for field application flexibility.

At depths of greatest interest to geotechnical engineering (less than ~ 30 m), surface waves dominate the energy of the elastic wavefield (Miller and Pursey, 1955). As a result,

surface wave methods are the most common techniques for developing 1D Vs profiles for near-surface application (e.g., Stokoe et al., 1994; Park et al., 1999; Foti, 2000; Louie, 2001; Okada, 2003; Tokimatsu et al., 1992). Surface wave methods work by exploiting the dispersive properties of surface waves in vertically heterogeneous media to develop 1D Vs profiles through the solution of an inverse problem. Solving the inverse problem involves assuming a 1D model with elastic properties (Foti et al., 2014; 2018) and iteratively solving a theoretical wave propagation problem (i.e., the forward problem) until the theoretical dispersion curves from the assumed model match the dispersion data extracted from experimental measurements of surface waves phase velocity. The surface wave inverse problem, commonly referred to as surface wave inversion, has been explored extensively in the literature and is known to be particularly challenging due to it being ill-posed and without a unique solution (Vantassel & Cox 2021a, b; Cox & Teague 2016; Foti et al., 2014; 2018). Despite these challenges, surface wave methods have been applied widely in practice and are commonly used to develop 1D, and even pseudo-2D, subsurface models, for example with methods such as 2D MASW (Park 2005; Ivanov et al., 2006). It is important to note, however, that these are not true 2D models due to underlying 1D assumptions in the numerical solution of the dispersion data forward problem used during inversion. Thus, the process of spatially interpolating between numerous 1D Vs profiles collected along a linear array produces a pseudo-2D subsurface image rather than a true 2D subsurface image. Presently, the only linear-array, active-source subsurface imaging method capable of producing a true 2D subsurface image is FWI.

While FWI can produce a true 2D model, it is less commonly used than surface wave methods for near-surface imaging due to its more complex and time-consuming field

acquisition and data processing requirements. However, it is a more promising approach for recovering true 2D and 3D subsurface images, as it utilizes the entirety of the seismic wavefield (rather than only the surface wave dispersion in surface wave methods). FWI can be described as a data-fitting procedure that seeks to minimize the misfit between the experimentally acquired seismic waveforms and the synthetic wavefield obtained by solving a wave propagation simulation through a candidate model. The FWI optimization process can be performed using either a global or a local search algorithm. Even though numerical methods for modeling the propagation of elastic waves through 2D and 3D earth models exist (e.g., finite-difference, spectral-element), they are computationally expensive, making global search methods, which are already computation demanding, uncommon for FWI (Virieux and Operto 2009). Local search methods (e.g., Pratt et al., 1998; Pratt 1999; Nocedal and Wright 2006) are less computationally expensive, as they begin with a predefined starting model and iteratively refine that model until the misfit between the recorded seismic waveforms and the calculated wavefield becomes sufficiently small. As a result, they require solving fewer forward problems than their global counterparts, making them computationally less expensive. However, if the starting model is not sufficiently similar to the true subsurface model, these methods are likely to be trapped in a local minimum, or saddle point, that prevents them from converging to the true solution (Monteiller et al., 2015; Smith et al., 2019; Feng et al., 2021; Vantassel and Cox 2022). Given the sensitivity of FWI results to the starting model (Shah et al., 2012; Vantassel et al., 2022a), rapid and accurate ways of generating 2D and 3D starting models are needed to more fully take advantage of FWI in engineering practice.

There has been growing interest in the past few years in using deep-learning methods to either enhance or completely replace FWI. An extensive review on integrating these deep-learning methods in various parts of the FWI can be found in Alder et al., (2021). Several end-to-end techniques, which aim to retrieve subsurface models directly from seismic wavefield data are also available in the literature (e.g., Araya-polo et al., 2018; Mosser et al., 2018a, b; Mao et al., 2019; Yang and Ma, 2019; Li et al., 2020). However, most previous works have either targeted recovering crustal-scale subsurface velocity models, as indicated by Vantassel et al. (2022a), or/and suffered from a weak generalization ability, preventing them from being used for a wide variety of field applications, as noted by Feng et al. (2021) and Liu et al. (2020). Additionally, a significant portion of the deep learning seismic imaging literature has focused on developing 2D velocity models using the acoustic approximation (Araya-polo et al., 2018; Mosser et al., 2018a, b; Mao et al., 2019; Yang and Ma, 2019; Li et al., 2020), which only models the propagation of compression wave velocities (Vp) and makes them poorly suited for near-surface applications. In fully saturated near-surface soil deposits, Vp is mainly controlled by the compressibility of the water and is much faster than the Vp of dry materials (Foti et al., 2014). Hence, Vp is less revealing of actual subsurface soil properties when the materials are relatively soft and saturated. On the other hand, the small strain shear modulus calculated from V_S represents that of the soil skeleton only and is independent of the ground saturation (Aziman et al., 2016). Due to the interest in retrieving Vs for engineering site characterization and the predominance of surface waves in actively-generated wavefields, deep learning approaches based only on acoustic wave propagation are not applicable for near-surface site characterization. To the authors' knowledge, the sole deep-learning, end-

to-end, 2D imaging technique for near-surface geotechnical engineering purposes was proposed by Vantassel et al. (2022a).

Vantassel et al. (2022a) demonstrated the ability of deep-learning methods in utilizing complicated wavefields comprised of surface and body waves to image the near-surface. They designed a CNN that could be used to generate 2D Vs images for subsurface profiles consisting of soil over undulating rock. Their CNN could predict a 24-m deep and 60-m wide V_S image directly from waveforms recorded by 24 receivers at 2-m spacing, which is a common configuration for active-source, linear-array imaging techniques for FWI. They used 100,000 synthetic soil-over-rock models in training the CNN and tested it on an additional 20,000 synthetic models. Their CNN showed great promise for developing starting models for near-surface FWI and, in some cases, yielded 2D subsurface models that could not be improved upon by local search FWI. However, the authors acknowledged that their approach could not generalize beyond the data acquisition configurations selected during CNN training (e.g., source type, source location, number of receivers, and receiver spacing).

The present work aims to show that developing a CNN with a frequency-velocity domain input image can yield comparable accuracy to the time-distance domain input approach proposed by Vantassel et al. (2022a), while providing the flexibility necessary to generalize for a broader range of field applications. We demonstrate that once trained for an appropriate set of geological conditions, the proposed frequency-velocity CNN approach can be used to instantly generate a 2D subsurface Vs image directly from a normalized dispersion image obtained from linear-array, active-source wavefield measurements. Normalized dispersion images are shown to be relatively insensitive to the

experimental testing configuration and can be easily generated due to their wide use in surface wave testing. The effectiveness of our frequency-velocity CNN is demonstrated by applying it to a classic near-surface geophysics problem; namely, imaging a two-layer, undulating, soil-over-bedrock interface. A total of 100,000 models were developed to train, validate, and test the frequency-velocity CNN. The ability of the proposed frequency-velocity CNN to generalize across various acquisition configurations is first tested using synthetic near-surface models using different acquisition configurations from that of the training set, and then applied to experimental data collected at the Hornsby Bend site in Austin, Texas, USA. We also compare the performance of the frequency-velocity CNN with the time-distance CNN for different near-surface models.

2.2 Overview of the frequency-velocity CNN

The 2D Vs imaging approach proposed herein builds on the work of Vantassel et al. (2022a) by developing a frequency-velocity CNN that accommodates different linear-array, active-source experimental testing configurations. A schematic illustrating the similarities and differences between the time-distance CNN (left) and the frequency-velocity CNN (right) required for subsurface Vs imaging is illustrated in Figure 2-1. The time-distance CNN proposed by Vantassel et al. (2022a) receives an input seismic wavefield recorded at specific receiver locations relative to a single source type and location and predicts a 2D V_S image. The proposed frequency-velocity CNN generalizes beyond specific receiver locations, source types and source locations by using a frequency-dependent normalized dispersion image as its input for predicting a 2D sub-surface V_S-image. Unlike the time-distance CNN, which would require additional training and tuning to handle multiple source types and receiver spacings, the frequency-velocity-based CNN

is relatively insensitive to the experimental testing configuration and, therefore, saves the network from needing to learn that additional complexity. In both cases, time-distance and frequency-velocity, developing and training the CNN takes a significant amount of time and effort, but after it has been trained it can be used to instantaneously produce a V_S image from the input.

To demonstrate the key difference between training a CNN using a time-distance input (i.e., seismic wavefield) versus a frequency-velocity input (i.e., dispersion image), several synthetic seismic wavefields acquired with different testing configurations on the same subsurface model are shown in Figure 2-2 along with their associated dispersion images. Figure 2-2a depicts a synthetic soil-over-rock V_S image. At its surface, 48 receivers with 1-m spacing and two source locations at 5 m and 20 m to the left of the first receiver are shown. Figures 2-2b through 2-2i show the seismic wavefields resulting from several different source and receiver configurations and their associated dispersion images. For example, Figures 2-2b and 2-2c show the wavefield sampled by 48 receivers at a 1-m receiver spacing and its corresponding dispersion image, respectively, due to a 30-Hz Ricker wavelet source (Figure 2-3a) at 5-m distance from the first receiver. This will be referred to as the base configuration. The experimental configuration used to obtain the seismic wavefield and dispersion image illustrated in Figures 2-2d and 2-2e, respectively, differs from the base configuration in that the source is excited at 20 m from the first receiver. While the wavefield image in Figure 2-2d is clearly different from the wavefield image of the base configuration (Figure 2-2b) due to the increased travel time associated with a greater source offset, the dispersion images from both configurations (Figure 2-2c and 2-2e) are similar. The experimental configuration used to obtain the wavefield and

dispersion image illustrated in Figures 2-2f and 2-2g, respectively, differ from the base configuration in that the 30-Hz Ricker wavelet excited 5 m from the first receiver is now recorded by only 24 receivers at a 2-m spacing (i.e., using half the number of receivers at two-times the spacing). Once again, the wavefield image for the alternate testing configuration (Figure 2-2f) is clearly different from the base configuration (Figure 2-2b) due to wavefield sampling at one-half the spatial resolution, but the dispersion images (Figures 2-2c and 2-2g) are very similar. Figure 2-2h shows a wavefield sampled by 48 receivers at a 1-m receiver spacing from a source located 5 m from the first receiver (similar to the base configuration), however, the source function is now a 12-second-long linear chirp from 3-Hz to 80-Hz (Figure 2-3c). Once again, while this wavefield is drastically different from the base configuration wavefield (Figure 2-2b), its dispersion image is visually identical to the others.

To quantitatively illustrate the good agreement between the dispersion images across different experimental testing configurations, the mean structural similarity index (MSSIM) proposed by Wang et al. (2004) is used. In this case, the MSSIM is used to compare the similarity between the base configuration dispersion image (Figure 2-2c) and the dispersion images obtained using the varied testing configurations (refer to Figures 2-2e, 2-2g, and 2-2i). The value of the MSSIM index between two images can range between 0 and 1, where a value of 0 indicates no structural similarity, while a value of 1 means perfect structural similarity. More information about MSSIM is provided in section 5.1. In the meantime, the high values of MSSIM for the different testing configurations (0.81 – 1.0) echo the qualitative observations made previously that the four dispersion images are very similar to one another. This insensitivity of dispersion images to acquisition

configurations of sources and receivers grants the frequency-velocity CNN approach the flexibility needed to predict on diverse testing configurations independent of the training dataset on which the CNN was trained, which is imperative for field applications.

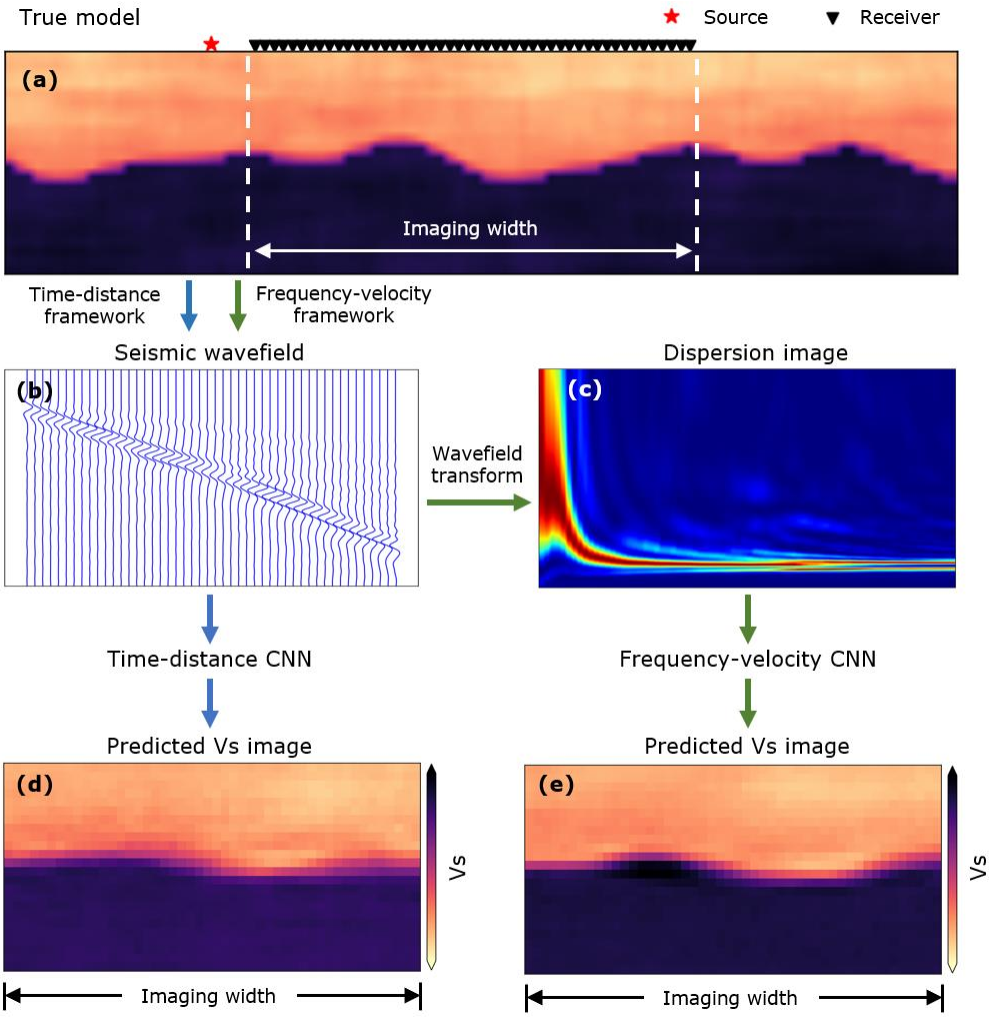

Figure 2-1. Time-distance CNN framework proposed by Vantassel et al. (2022a), which follows the blue arrow's path (left), and the new frequency-velocity CNN framework, which follows the green arrow's path (right).

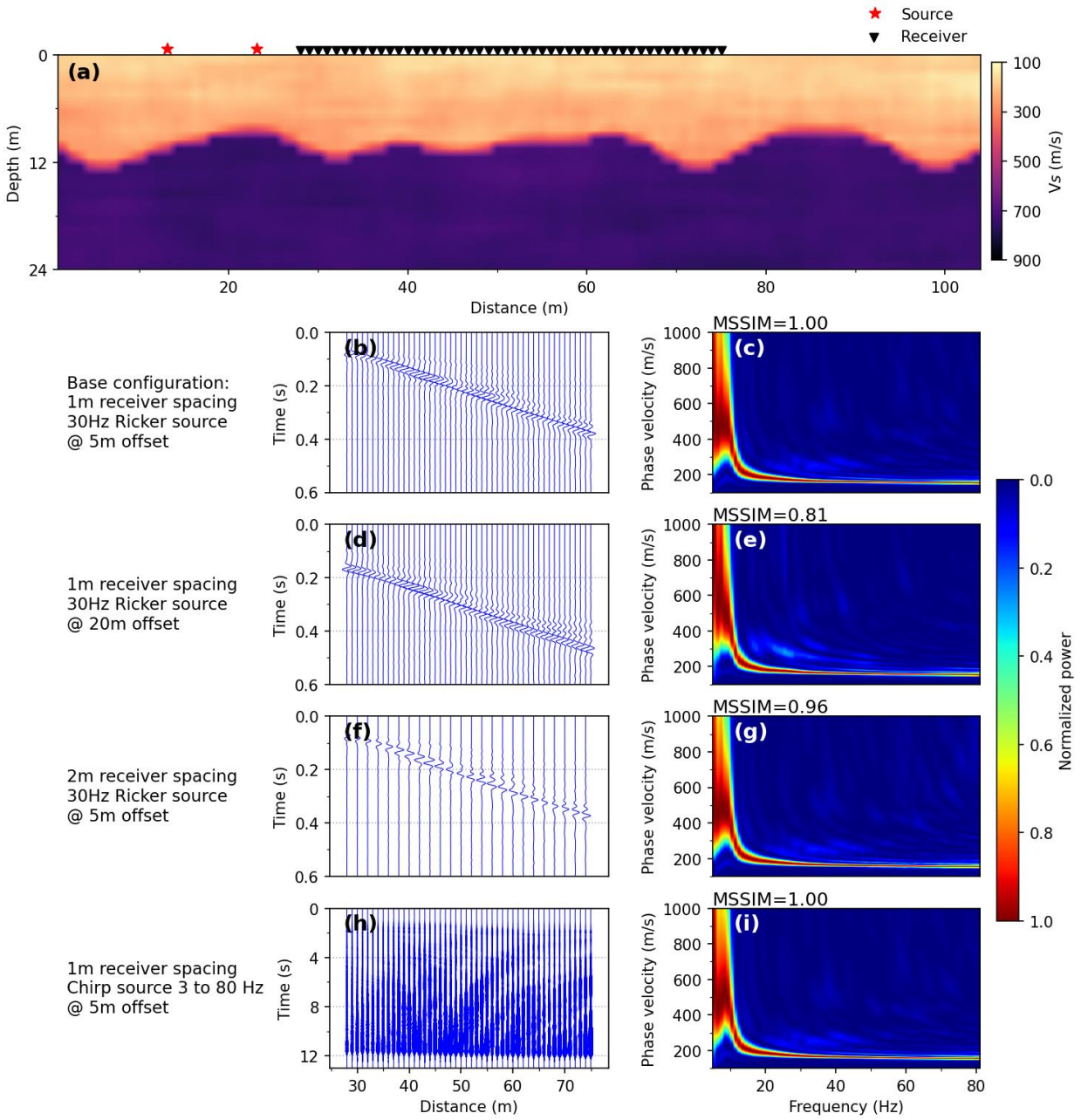

Figure 2-2 (a) a 104-m wide and 24-m deep soil-over-rock 2D model with a 47-m array of receivers at 1-m spacing and two source locations at 5 and 20 m off the end of the array. (b) the seismic wavefield recorded by 48 receivers from a 30-Hz Ricker source at the 5 m source location and (c) its associated dispersion image. (d) the seismic wavefield recorded by 48 receivers from a 30-Hz Ricker source at the 20 m source location and (e) its associated dispersion image. (f) the seismic wavefield recorded by 24 receivers from a 30-Hz Ricker source at the 5 m source location and (g) its associated dispersion image. (h) the seismic wavefield recorded by 48 receivers from a 3-Hz to 80-Hz chirp/sweep over 12-seconds at the 5 m source location and (i) its associated dispersion image. The mean structural similarity index (MSSIM) of each dispersion image relative to the base case (i.e., panel c) is presented above each panel.

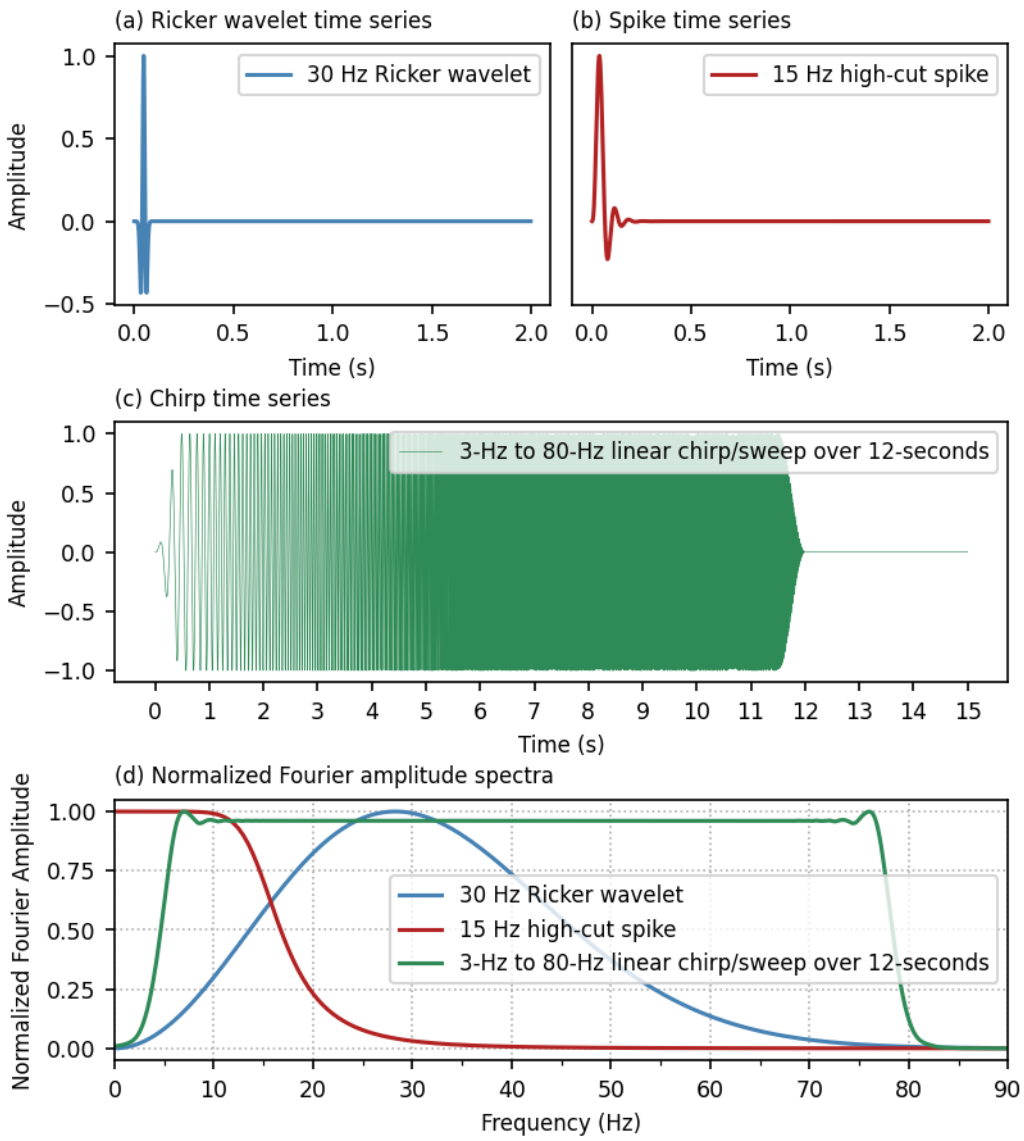

Figure 2-3 Source functions used in this study and their associated normalized Fourier amplitude spectra. (a) the time series of a 30-Hz Ricker wavelet. (b) the time series of a 15-Hz high-cut filtered spike wavelet. (c) the time series of a 3-Hz to 80-Hz linear chirp/sweep over 12-seconds. (d) the normalized Fourier amplitude spectra of each source function.

The present study is structured as follows. We start by describing the synthetic soil models developed for this study. Next, the details of the wave propagation simulations and the post-processing of the results used in constructing the dataset for this study are presented. We then outline the architectures of the time-distance and frequency-velocity

CNNs and evaluate their relative performances on the developed dataset. Subsequently, we demonstrate the advantages of operating in the frequency-velocity domain in terms of flexibility and generalization across different testing configurations that are not present in the training set relative to the time-distance CNN. Lastly, we validate the potential of the frequency-velocity CNN for handling experimental data based on comparisons with field observations at the Hornsby Bend site in Austin, Texas, USA.

2.3 Development of synthetic near-surface models

The synthetic near-surface models developed for this study used a slightly modified version of the framework developed by Vantassel et al. (2022a) for generating realistic soil-over-undulating bedrock subsurface profiles. A brief description of the models with changes relative to those implemented by Vantassel et al. (2022a) is highlighted in this section. A total of 100,000 synthetic near-surface models representative of soil overlying irregular bedrock were developed to train, validate, and test the CNN. A 104-m wide and 24-m deep domain was utilized in the present work. This domain is larger than the 60-m by 24-m domain used by Vantassel et al. (2022a) to accommodate the different receiver and source configurations required for testing the abilities of the proposed frequency-velocity approach to generalize across different field acquisition setups.

Similar to Vantassel et al. (2022a), the base Vs models were constructed by first assuming the vertical variation of Vs in the overlying soil layer (i.e., the upper part of the model) followed the approximate relationship between Vs and mean effective confining pressure for dense granular soils proposed by Menq (2003). To avoid unrealistically low velocities, the Vs relationship was truncated near the ground surface (i.e., at low mean effective stresses) to ensure no Vs less than 200 m/s. To model a broader range of realistic

dense soil-velocities, the relationship was scaled up and down by a random variable, the soil velocity factor, between 0.9 and 1.1. The average interface boundary between the stiff soil layer and bedrock ranged between 5 m and 20 m in depth. Several different soil-to-bedrock interface conditions were simulated; namely, 30% of the models had highly undulating interfaces (e.g., Figure 2-4a through 2-4f), 60% had slightly undulating interfaces (e.g., Figure 2-4g through 2-4r), and 10% of the models had planar soil-rock interfaces (e.g., Figure 2-4s and 2-4t). These percentages of models with different bedrock undulation conditions were preserved throughout the CNN training, validation, testing, and generalization evaluation stages. Three overlapping spatial undulation frequencies were used to control the interface undulation intensity. The range of spatial frequencies used for the highly undulating interfaces was between 1/5 m^{-1} and 1/60 m^{-1}, while the range of frequencies used for the slightly undulating interfaces was between 1/10 m^{-1} and 1/60 m^{-1}. These values are smaller than those used by Vantassel et al. (2022a), making the soil-bedrock interface less variable in the present study. The bedrock V_S was randomly varied following a uniform distribution between a lower bound of 360 m/s and an upper bound of 760 m/s. Lateral and vertical perturbations were imposed on both the soil and rock portions of the V_S model to simulate inhomogeneities present in natural materials. The small-scale irregularities introduced by the perturbations were 1 m to 2 m in the vertical direction and 4 m to 6 m in the horizontal direction. Further details on the computations used to introduce these small-scale irregularities, including the assumed correlation structure, can be found in Vantassel et al. (2022a).

The 100,000 synthetic models were developed by randomly changing the stiff soil V_S multiplier values, the weathered rock V_S, the V_S lateral and vertical perturbations, the

interface depth, and the interface undulation frequencies within the upper and lower bound for each variable. Following the development of each Vs image, a Vp image was generated by using a Poisson's ratio of 0.33 for soil and 0.2 for rock. The mass density image was constructed by assigning a value of 2000 kg/m^3 for soil and 2100 kg/m^3 for rock. Twenty randomly selected Vs images generated using the procedures described above are shown in Figure 2-4. Note that while the Vs images in Figure 2-4 have been selected at random, the number of images for each of the three model types (i.e., highly undulating, lightly undulating, and planar) have been selected to follow the distribution of model types in the training set (i.e., 30%, 60%, and 10%, respectively).

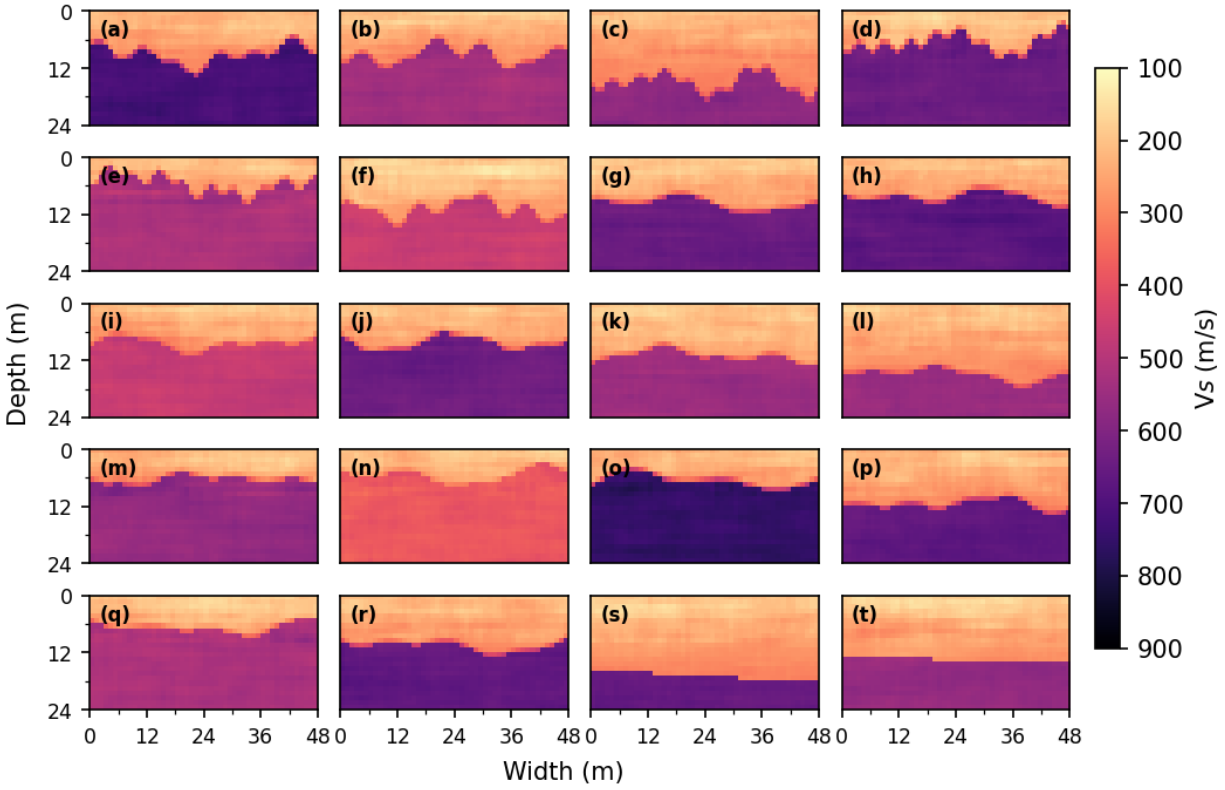

Figure 2-4 Twenty randomly selected synthetic, near-surface 2D V_S images from the 20,000 models used to test the convolutional neural networks (CNNs). The first six models (from a to f) have a highly undulating soil-rock interface, the following 12 models (from g to r) have a slightly undulating soil-rock interface, and the last two models (s and t) have a planar soil-rock interface.

2.4 Development of synthetic seismic wavefields

The 2D finite-difference software DENISE (Köhn, 2011; Köhn et al., 2012) was used to simulate elastic wave propagation through the synthetic models. Forty-eight receivers were placed at 1-m spacing across the center of the 104-m wide models to sample the wavefield generated by a source located 5 m to the left of the first receiver. This contrasts with the 24 receivers at a 2-m spacing and source located at the center of the array used by Vantassel et al. (2022a). The receivers occupied the distance between 28 m to 75 m, so they were as far away from the model boundaries as possible (refer to Figure 2-2a). A 30-Hz Ricker wavelet (Figure 2-3a) was used as a forcing function during the development of the training, validation, and testing datasets. This is a higher-frequency wavelet than that used by Vantassel et al. (2022a), allowing, in theory, high-resolution predictions to be developed. The finite-difference simulations were used to model two seconds of wave propagation to allow for recording of all wave types by the receivers. A sixth-order finite-difference operator in space and a second-order finite difference operator in time were utilized during the simulations. Perfectly matched layer absorbing boundaries (Komatitsch and Martin, 2007) were placed at the sides and the bottom of the domain, while the ground surface was modeled using a free boundary condition (Levander, 1988). Prior to wave propagation simulations, the soil models were interpolated from 1-m pixels to 0.2-m pixels to ensure that the wavefield was spatially sampled with a small enough grid to avoid numerical artifacts and instabilities, as recommended by Köhn et al. (2011). A 5E-5 second time step was used to satisfy the Courant-Friedrichs-Lewy criterion (Courant et al., 1967). The waveforms at each receiver location were recorded at a 400-Hz sampling rate. The simulations were performed on the Texas Advanced Computing Center's

(TACCs) high-performance cluster Stampede2 using a single Skylake (SKX) compute node. Each simulation took approximately 5 seconds to complete.

1.1.1 Post-processing to obtain the wavefield and dispersion input images

To evaluate the relative performance between the time-distance and frequency-velocity CNNs, two identical datasets for training and testing of the two CNNs were developed. A total of 100,000 near-surface V_S images and their corresponding seismic wavefields formed the image pairs in the dataset developed for the time-distance CNN. The same soil models and the dispersion images derived from their respective wavefields composed the dataset produced for the frequency-velocity CNN. Even though the synthetic models were 104-m wide and 24-m deep, only the 48 m at the center of the models immediately below the receivers were considered in training and testing the CNN, making the utilized soil models 48-m wide and 24-m deep, as illustrated in Figure 2-4. The near-surface V_S images were then normalized by the maximum V_S value in the training set to facilitate the training process. The waveforms recorded by the 48 receivers used for the time-distance CNN were normalized by the maximum amplitude across all receivers to preserve the rate of amplitude decay. The input shape for the time-distance CNN was 48x800x1, which represents the 48-receiver wavefield sampled for two seconds at 400 Hz.

The frequency-velocity CNN utilizes a normalized dispersion image as input. The dispersion image is computed using a wavefield transformation of the time-distance wavefield recorded by a linear array of receivers. Several wavefield transformation techniques can be used to generate a dispersion image from a recorded time-distance wavefield. For example, the frequency-wavenumber (f-k) (Gabriels et al., 1987; Nolet and Panza, 1976), slant-stack (McMechan and Yedlin, 1981), phase-shift (Park et al., 1998),

and frequency-domain beamformer (FDBF) (Zywicki, 1999) transformations are all commonly used in various commercial and open-source surface wave processing software. The FDBF approach with a plane wave steering vector and no amplitude weighting (Zywicki, 1999) was used in the present study to develop the dispersion images used in training and testing the frequency-velocity CNN. We chose the FDBF over the more common frequency-wavenumber (f-k) approach because the FDBF technique allows the transformation of time-distance wavefields collected using non-uniformly spaced arrays of receivers and the calculation of dispersion power above the aliasing wavenumber (Vantassel and Cox, 2022a), thereby making the inputs to the CNN easier to acquire in the field. Furthermore, the FDBF approach has been judged to be superior to other wavefield transformation methods as assessed by Rahimi et al. (2021). Along with the FDBF, we use the plane wave steering vector to be consistent with the wavefield simulations. In particular, the seismic wavefield data were produced under the 2D, plane-strain assumption, and as a result are representative of waveforms generated by a line source (i.e., a plane wave source). Importantly, the FDBF method can be adjusted when applied to field data that is truly 3D in nature by swapping out the plane-wave steering vector and no amplitude weighting for cylindrical-wave steering and square-root-distance weighting to appropriately compensate for the effects of radiation damping in 3D data (Zywicki and Rix, 2005). This substitution removes the need to “correct” 3D waveforms collected in the field to 2D equivalents before dispersion processing and use in the frequency-velocity CNN.

The dispersion images in this study were generated using a frequency range of 5 Hz to 80 Hz with a 1-Hz frequency step and a phase velocity range of 100 m/s to 1000 m/s

with 2.25 m/s velocity step. The dispersion image, therefore, had an input shape of 75x400x1 (i.e., 75 frequencies x 400 phase velocities). After computing the MASW dispersion images we utilized frequency-dependent normalization to further simplify the learning task of the CNN. With frequency-dependent normalization, sources with different frequency contents and offsets appear more similar than with other forms of normalization, such as absolute maximum normalization proposed by Park et al. (1998). By performing frequency-dependent normalization, we remove the need for the CNN to learn this aspect of the underlying physics, allowing for better generalization across various source types and locations. All dispersion images in this study were generated programmatically using the open-source Python package *swprocess* (Vantassel, 2021).

2.5 Time-distance and frequency-velocity CNN architectures

CNNs are an excellent tool for computer vision tasks and have shown great potential for use in seismic imaging (Wu and Lin, 2019; Yang and Ma, 2019; Liu et al., 2020; Vantassel et al., 2022a). The time-distance and frequency-velocity CNNs follow the architecture proposed by Vantassel et al. (2022a) of five convolutional layers interspersed with max-pooling layers. The convolutional layers employ a set of kernels, also called feature detectors or stencils, to capture the relevant patterns (i.e., feature maps) in the dataset images. Once the relevant features of the images are detected, subsampling or pooling layers are utilized to decrease the feature maps' spatial resolution, which in turn reduces the reliance on precise positioning within feature maps produced by the convolutional layers. Disregarding the exact position of features within a feature map while maintaining the relative position of features with respect to each other allows for a better CNN performance on inputs that differ from the training data. Max-pooling layers were

used in the current study, as they were shown to be superior in capturing invariances in image-like data compared to subsampling layers (Scherer et al., 2010). The final convolutional layer is flattened and connected to a fully connected layer to perform the regression task. Table 2-1 shows the architectures used for the two CNNs. Google Collab and the open-source machine learning library Keras (Chollet et al., 2015) were used in training and testing the CNNs. 70%, 10%, and 20% of the developed 100,000 image pairs were used in the two CNNs training, validation, and testing stages, respectively. The two CNNs' architectures were adjusted from that originally proposed by Vantassel et al. (2022a) to accommodate their respective input sizes.

In addition to the networks' architectures, the model's hyperparameters need to be rigorously tuned to provide optimal performance. In the present study, we tuned the following hyperparameters by varying them between the upper and lower bounds listed below. We then selected the set that produced the best performance on the validation set. The hyperparameters considered include: the learning rate (0.1 to 0.0001), batch size (8 to 64), number of training epochs (10 to 100), optimizer (RMSprop and Adam), and loss function (mean squared error and mean absolute error). Ultimately, a learning rate of 0.0005, batch size of 16, training epoch of 40, Adam optimizer (Kingma and Ba, 2014), and mean absolute error (MAE) were selected for both the time-distance and frequency-velocity CNNs. We note these hyperparameters are similar to the ones selected by Vantassel et al. (2022a) despite being chosen after independent hyperparameter tuning exercises. The validation dataset MAEs for the time-distance and frequency-velocity CNNs using the selected hyperparameters are 0.022 and 0.025, respectively.

Table 2-1 Architectures for the time-distance and frequency-velocity convolutional neural networks (CNNs) developed in this study.

Network layer type	Time-distance CNN		Frequency-velocity CNN	
	Filter size	Size of output layer	Filter size	Size of output
2D Convolution	1x3	48x798x32	3x1	398x76x32
2D Max Pooling	1x3	48x266x32	3x1	132x76x32
2D Convolution	1x3	48x264x32	3x1	130x76x32
2D Max Pooling	1x3	48x88x32	3x1	43x76x32
2D Convolution	1x3	48x86x64	3x1	41x76x64
2D Max Pooling	2x3	24x28x64	1x3	41x25x64
2D Convolution	3x3	22x26x128	3x3	39x23x128
2D Max Pooling	2x2	11x13x128	3x3	13x7x128
2D Convolution	3x3	9x11x128	3x3	11x5x128
Flatten		12672		7040
Dense		1152		1152
Reshape		24x48		24x48

2.5.1 CNNs accuracy evaluation

The accuracy of the time-distance and frequency-velocity CNNs were evaluated using their respective 20,000 testing image pairs, which the networks were not trained on. The mean absolute percent error (MAPE) and MSSIM were used to provide a quantitative assessment of the CNN's performance. MAPE is the mean of the absolute value of the pixel-by-pixel percent error of each predicted V_S image in physical units. The MSSIM index assess the quality of one image relative to another that is deemed to be of perfect quality based on three key features: luminance, contrast, and structure (Wang et al., 2004). MSSIM was calculated using a Gaussian windowing approach to match the implementation of Wang et al. (2004) and using a dynamic range equal to the difference

between the absolute maximum and minimum V_S values of the true models in the testing set.

The average values of MAPE and MSSIM for the time-distance CNN are 5.3% and 0.80, respectively, while the MAPE and MSSIM for the frequency-velocity CNN are 6.0% and 0.78, respectively (refer to Table 2-2). While the time-distance CNN provides slightly better accuracy than the frequency-velocity CNN, their performances are quite similar. Furthermore, we will show that the advantages permitted by operating in the frequency-velocity domain in terms of flexibility and generalization across data acquisition configurations outweigh the minor loss of accuracy. Figures 2-5 and 2-6 show the Vs image predictions for the time-distance and the frequency-velocity CNNs, respectively, for comparison with the 20 true images depicted in Figure 2-4. As noted above, Figure 2-4 depicts six highly undulating, 12 slightly undulating, and two linear soil-rock interface models. The MAPE and MSSIM values for each CNN prediction relative to the true image are also provided in Figures 2-5 and 2-6. While, on average, the time-distance CNN slightly outperforms the frequency-velocity CNN in terms of overall MAPE and MSSIM, there are some individual models for which the frequency-velocity CNN is slightly more effective (e.g., model S in Figures 2-5s and 2-6s). Furthermore, by-eye it would be difficult to distinguish the time-distance and frequency-velocity predictions from one another.

Table 2-2 The mean absolute percent error (MAPE) and mean structural similarity index (MSSIM) between the true V_S images and the V_S images predicted using the time-distance and frequency-velocity convolutional neural networks for the testing set and the different acquisition configurations. The values in the table represent the average MAPE and MSSIM over the number of images used for each acquisition configuration. The base acquisition configuration comprised 48 receivers with a 1-m spacing, a source offset of 5 m relative to the first receiver in the linear array, and a 30-Hz Ricker wavelet source forcing function.

Acquisition variation	CNN used	Deviation from base configuration	Number of images	MAPE (%)	MSSIM
Base case	Time-distance	-	20,000	5.3	0.80
	Frequency-velocity	-	20,000	6.0	0.78
Varying receiver spacing	Frequency-velocity	24 receivers at 2-m spacing	5,000	6.0	0.77
		16 receivers at 3-m spacing	5,000	9.9	0.65
		12 receivers at 4-m spacing	5,000	24.0	0.32
Varying source location	Frequency-velocity	Source at 6 m from first receiver	5,000	6.8	0.77
		Source at 10 m from first receiver	5,000	11.0	0.69
		Source at 20 m from first receiver	5,000	12.7	0.64
		Average between 5-m source and 20-m source offsets	5,000	8.1	0.73
		Average between 10-m source and 20-m source offsets	5,000	11.2	0.67
Varying source forcing function	Frequency-velocity	15-Hz high-cut filtered spike forcing function	5,000	5.9	0.78
		3-Hz to 80-Hz linear sweep over 12-seconds forcing function	4,887	7.3	0.75

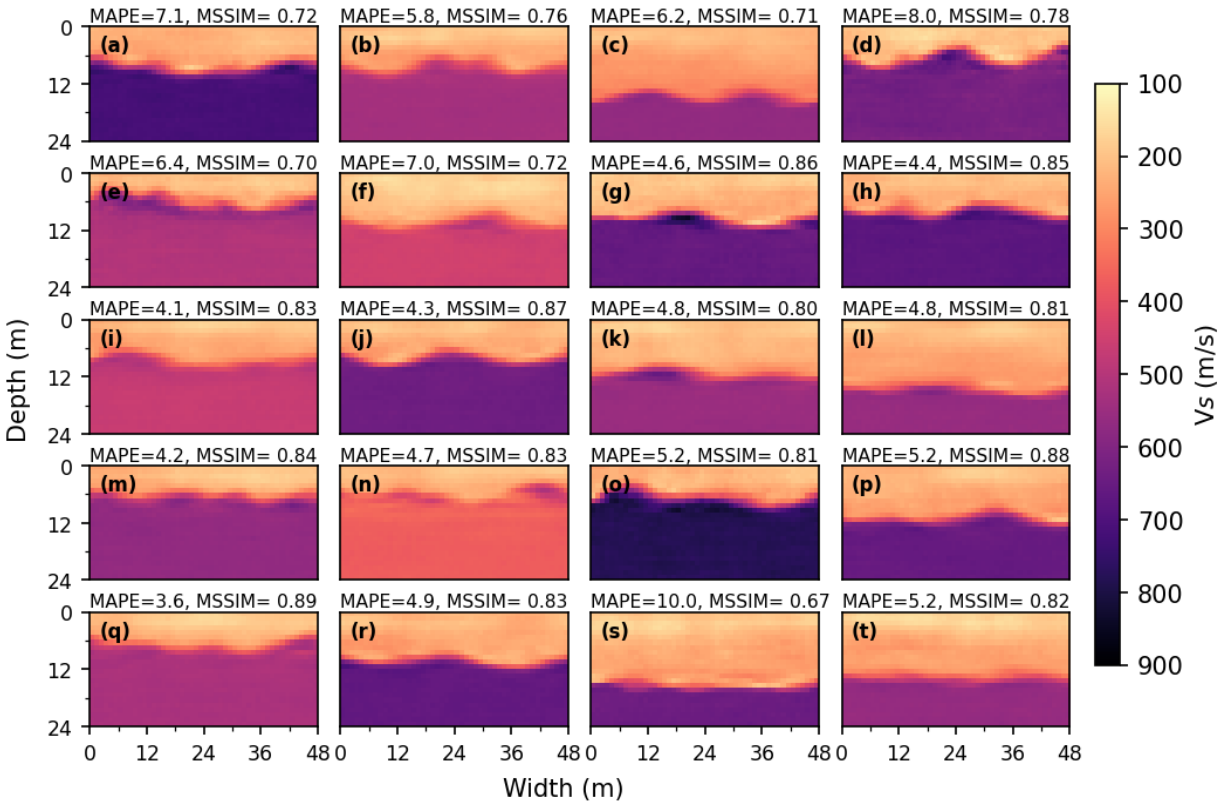

Figure 2-5 The time-distance CNN's predictions of the true synthetic 2D Vs images presented in Figure 2-4. The inputs used to obtain these predictions are the wavefields recorded by 48 receivers at 1-m spacings, which were excited by a 30-Hz Ricker source wavelet at 5 m from the first receiver. The mean absolute percent error (MAPE) and mean structural similarity index (MSSIM) of each predicted image are presented above each panel.

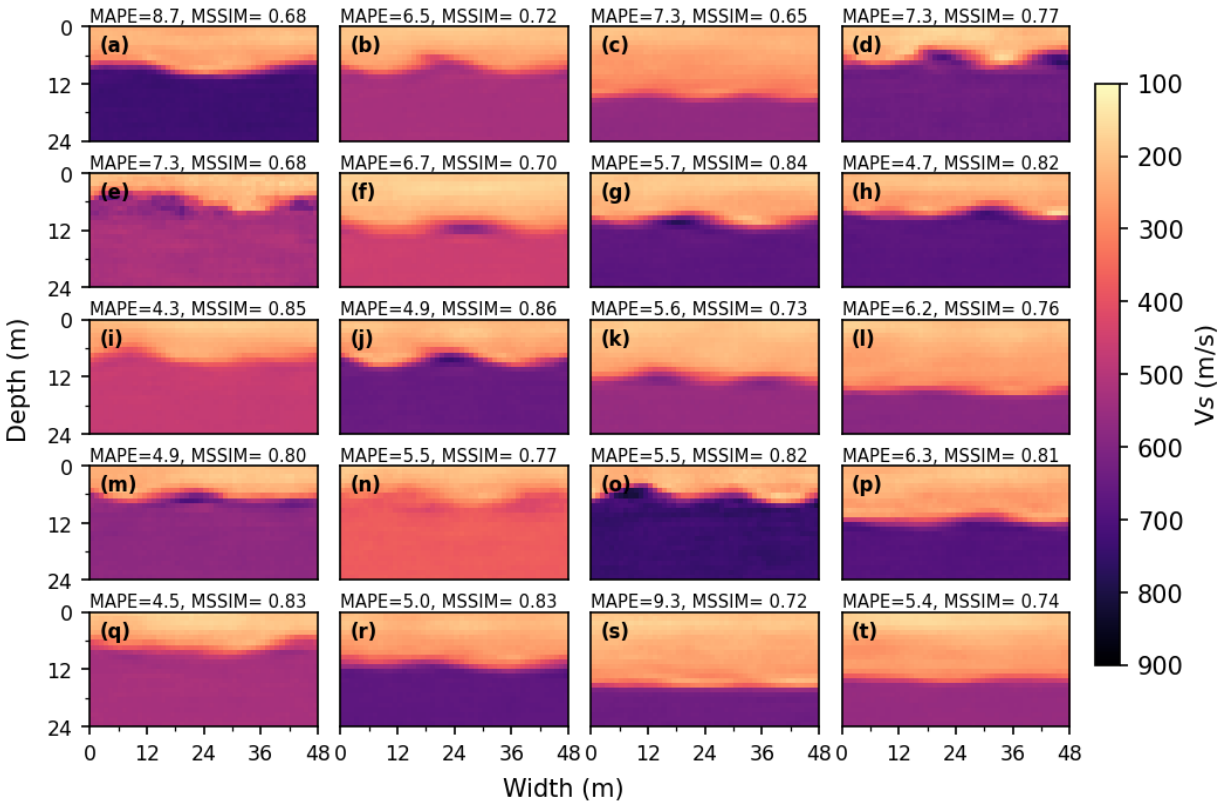

Figure 2-6 The frequency-velocity CNN's predictions of the true synthetic 2D Vs images presented in Figure 2-4. The inputs used to obtain these predictions are the normalized dispersion images obtained by post-processing the wavefields recorded by 48 receivers at 1-m spacings, which were excited by a 30-Hz Ricker source wavelet at 5 m from the first receiver. The mean absolute percent error (MAPE) and mean structural similarity index (MSSIM) of each predicted image are presented above each panel.

Figure 2-7 shows the residuals between the predicted frequency-velocity V_S images (Figure 2-6) and the true images (Figure 2-4). It can be seen that large portions of the residual images have neutral colors, indicating relatively small differences in Vs between the true and predicted images. Specifically, the Vs of the soil and rock layers are generally well predicted, whereas most of the error is concentrated at the undulating soil-rock interfaces. While only shown herein for the frequency-velocity CNN, similar, localized interface errors are present in the time-distance predicted images and were also reported by Vantassel et al. (2022a) for their time-distance CNN. Nonetheless, the interface locations and major undulations are still fairly well preserved in the predictions.

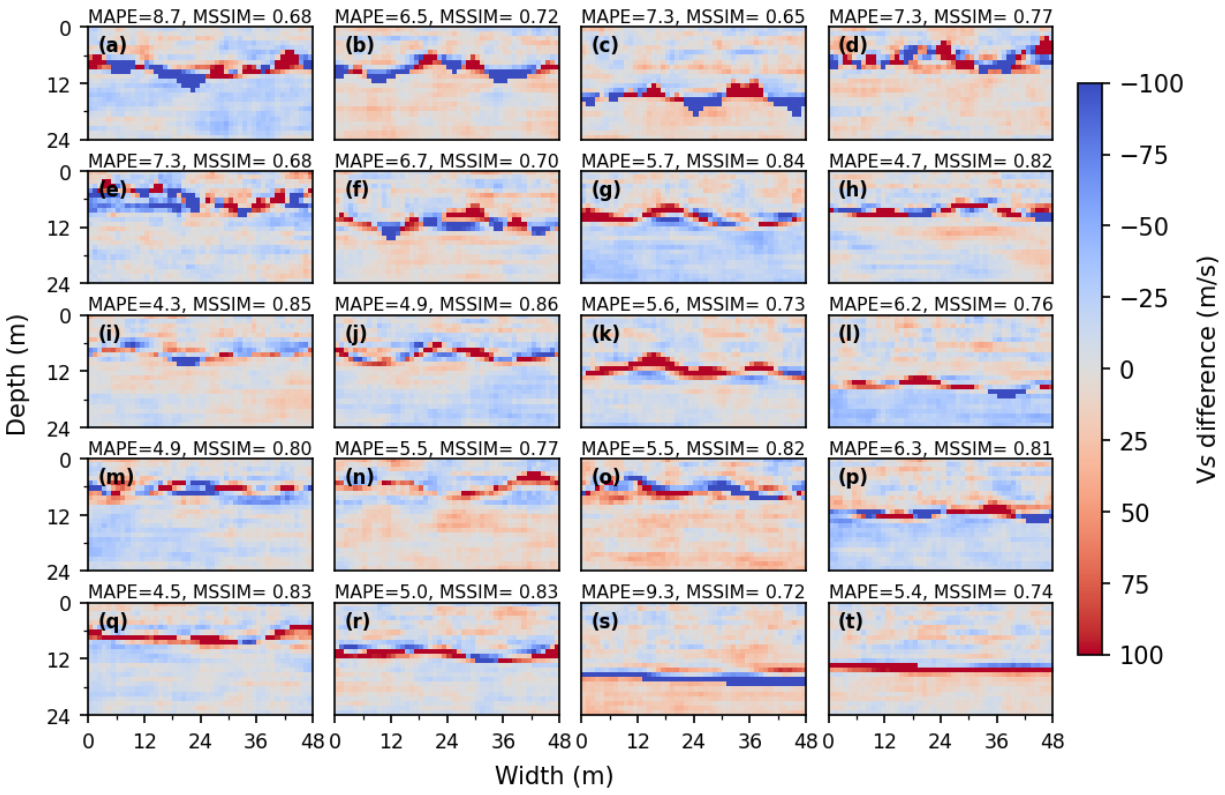

Figure 2-7 The pixel-by-pixel difference between the twenty true 2D V_S near-surface images presented in Figure 2-4 and the corresponding frequency-velocity CNN image predictions shown in Figure 2-6. The mean absolute percent error (MAPE) and mean structural similarity index (MSSIM) of each predicted image are presented above each panel.

2.6 Generalizing the frequency-velocity CNN across acquisition configurations

The acquisition generalization capabilities of the frequency-velocity CNN are evaluated by simulating different testing configurations than the base case configuration used during model training. The testing setup variations investigated herein include modifications to the number of receivers and receiver spacings, the source offset, and the source forcing function. The effects of each variation in the testing configuration are assessed separately, while the remainder of the setup is maintained identical to the base case configuration used in training the CNN. As a reminder, this base case training configuration was: 48 receivers with a 1-m spacing, a source offset of 5 m relative to the

first receiver in the linear array, and a 30-Hz Ricker wavelet source forcing function (refer to Figure 2-2b). Due to the high computational costs required to run numerous wave propagation simulations with various acquisition configurations, only 5,000 of the 20,000 testing models were evaluated during this stage. These 5,000 models were randomly selected, but with the stipulation that this smaller population of models maintained the same ratios of 30% highly undulating interfaces, 60% slightly undulating interfaces, and 10% linear soil-rock interfaces as the original training and testing sets. We evaluate the performance of the frequency-velocity CNN image predictions for these 5,000 testing models with wavefields recorded using a wide range of simulated acquisition testing configurations using the same MAPE and MSSIM statistics used to evaluate the full 20,000 testing model set. The following sections discuss the performance of the frequency-velocity CNN for different testing configurations.

2.6.1 Generalizing to the number of receivers and receiver spacings

In the field, the number of receivers used to image the subsurface is dependent on equipment availability, testing space, and the objective of the experiment (e.g., better near-surface resolution versus greater imaging depth). Therefore, a CNN that can provide accurate 2D images from wavefields collected with different numbers of receivers is desirable. To test the frequency-velocity CNN for such acquisition generalization ability, three sets of 5,000 input dispersion images were obtained from wave propagation simulations on the 5,000 testing models. The first set of dispersion images was generated using the wavefields from 24 receivers at 2-m spacing, the second set from 16 receivers at 3-m spacing, and the third set from 12 receivers at 4-m spacing. The accuracy of the CNN's predictions for the three sets of inputs is presented in Table 2-2 in terms of MAPE and

MSSIM. From Table 2-2, it is clear that the frequency-velocity CNN is capable of generalizing across various numbers of receivers and receiver spacings, but only within reasonable adjustments to the base configuration. For example, the MAPE and MSSIM values for 24 receivers at 2-m spacing are equal to 6.0 and 0.77, respectively, essentially equivalent to those from the base configuration. The MAPE and MSSIM values for 16 receivers at 3-m spacing are equal to 9.9 and 0.65, respectively, indicating a slight degradation in performance. However, the MAPE and MSSIM values for 12 receivers at 4-m spacing are equal to 24.0 and 0.32, respectively, which show a significant reduction in predictive capabilities as fewer receivers with larger receiver spacings are used to record the wavefield. Reasons for these observations are investigated further by considering plots presented in Figure 2-8.

Figure 2-8a shows the true V_S image depicted in Figure 2-4j. Figures 2-8b and 2-8c show the input dispersion image and CNN output Vs image, respectively, for the 48-receiver base configuration. The input dispersion images and output Vs images for modified acquisition configurations are shown in Figures 2-8d and 2-8e, respectively, for 24 receivers and in Figures 2-8f and 2-8g, respectively, for 12 receivers. Also shown in all dispersion images are the high frequency spatial array resolution limits (sometimes called the f-k aliasing limits) for the respective receiver configurations, which are plotted as dashed white lines. The f-k aliasing limits represent the largest wavenumber (k), or equivalently the smallest wavelength (λ), that can be measured without concern for spatial aliasing. The spatial aliasing limit is a constant wavelength that is equal to two-times the receiver spacing (i.e., at least two measurements per spatial wavelength). Note that proper spatial sampling is analogous to proper time-domain sampling following the Nyquist

sampling theorem. However, unlike in time-domain sampling the presence of contaminating short wavelength (i.e., high frequency) waves are less common due to material damping, and as a result clear dispersion data above the spatial aliasing limit can be used, albeit cautiously (Foti et al., 2018).

As can be seen in Figures 2-8b, 2-8d, and 2-8f, the portion of the input dispersion image that exists below the spatial aliasing line increases as the number of receivers decreases and the receiver spacing increases. This means that when larger receiver spacings are used the higher frequency data may not be resolved accurately due to spatial aliasing. For example, the high-power trend at frequencies greater than 60-Hz in Figure 2-8f is not a true higher mode, but rather an artefact of spatial aliasing. While the frequency-velocity CNN can generalize across different receiver spacings to a certain extent, as evident by comparing the Vs-images MAPE and MSSIM values for the 48-receiver and 24-receiver configurations (refer to Figs. 8c and 8e), it cannot accurately generalize across receiver spacings that are drastically different from the base configuration, as the input dispersion images are affected by spatial aliasing. Even though the MAPE increases significantly when 12 receivers at 4-m increments are used (refer to Figure 2-8g), the CNN was still able to qualitatively predict the location of the soil rock interface quite well. Nonetheless, the actual Vs values for the soil and rock are not well resolved, a clear result of the limitations imposed by the CNN's training data (i.e., no spatial aliasing was present in the training set).

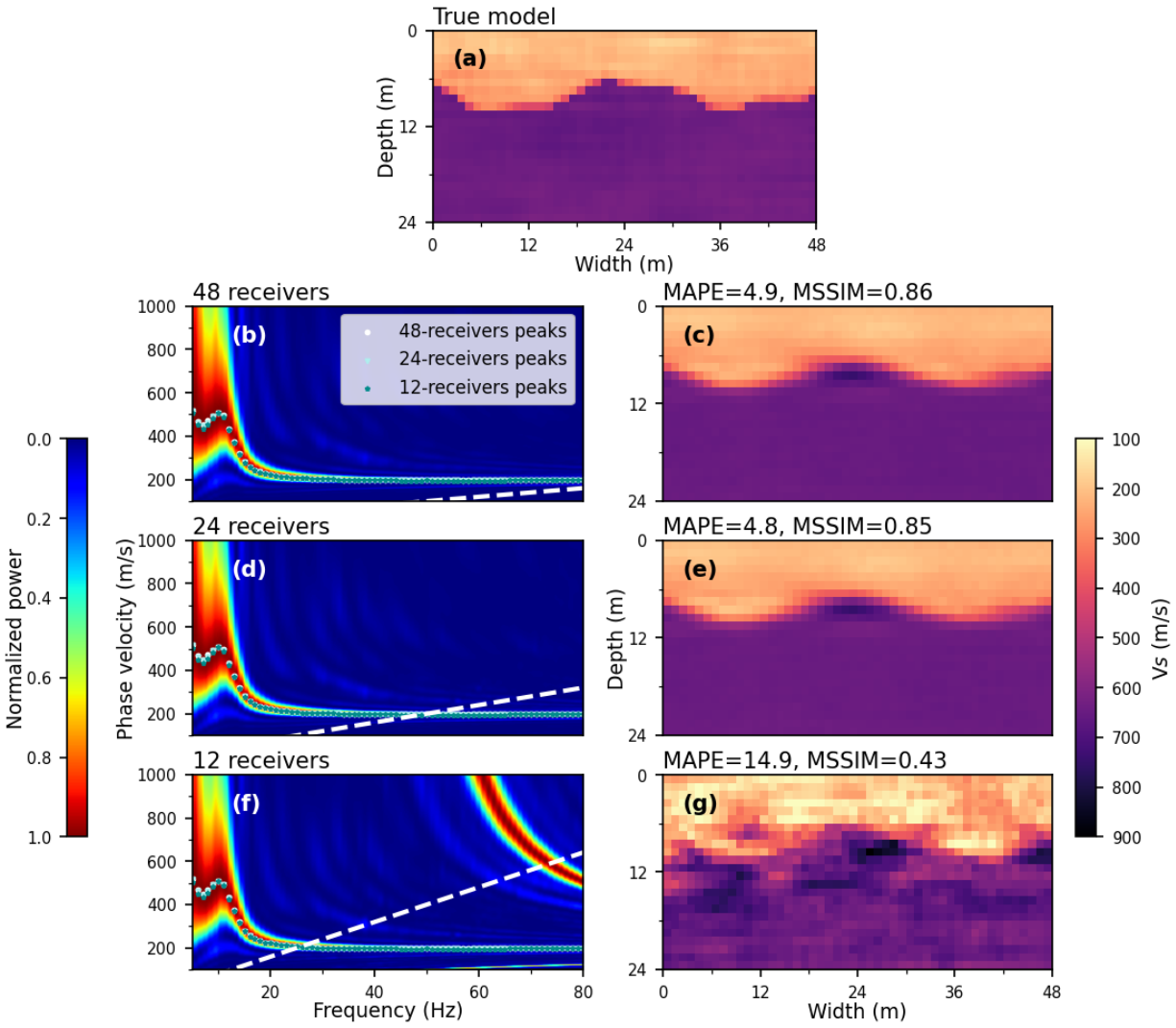

Figure 2-8 (a) the true synthetic image from Figure 2-4j. (b & c) the input dispersion image and output frequency-velocity CNN Vs image prediction, respectively, from the base receiver configuration (i.e., 48 receivers at 1-m spacing). (d & e) the input dispersion image and output frequency-velocity CNN Vs image prediction, respectively, from 24 receivers at 2-m spacing. (f & g) the input dispersion image and output frequency-velocity CNN Vs image prediction, respectively, from 12 receivers at 4-m spacing. The white dashed lines in (b), (d), and (f) represent the spatial array resolution limit for each array configuration. Peak power points obtained from the apparent fundamental Rayleigh wave mode of each dispersion image are also shown. The mean absolute percent error (MAPE) and mean structural similarity index (MSSIM) of each predicted image are presented above each panel.

2.6.2 Generalizing to source offset

When performing linear-array, active-source surface wave testing, it is good practice to record the waveforms generated at several different source locations with increasing offset distance to help identify dispersion data contaminated by near-field effects and quantify dispersion uncertainty (Cox and Wood, 2011; Vantassel and Cox, 2022b). Near-field effects, which result from measuring low frequency (i.e., long wavelength) waves too close to the source location, are known to bias phase velocity estimates to lower velocities (Rosenblad and Li, 2011; Li and Rosenblad, 2011; Yoon and Rix 2009). As such, one needs to be cautious about placing an active source too close to a linear array. To complicate matters, near-field affects are site-dependent and, as a result, it is difficult to know *a-priori* what source offset distances are appropriate at a site. In addition to near-field effects, other wave propagation phenomena like body wave reflections and refractions influence the recorded seismic wavefield when varying source offset distances are used. This is particularly true when the subsurface conditions are neither 1D nor homogeneous. Therefore, a CNN that is capable of generalizing to different, or possibly multiple, source offsets is desirable.

To test the frequency-velocity CNN's ability to generalize in terms of source offset distance, five sets of 5,000 input dispersion images were obtained from wave propagation simulations on 5,000 testing models using different source offset distances relative to 48 receivers with 1-m spacing. Specifically, the first set of dispersion images was generated using a 6-m source offset, the second set using a 10-m offset, and the third set using at 20-m offset. The fourth and fifth sets of dispersion images were generated by combining dispersion images from two different source offsets in the frequency-phase velocity domain

as follows: the fourth set resulting from stacking dispersion images from a 5-m and 20-m offset, and the fifth set from stacking dispersion images from a 10-m and 20-m offset. The accuracy of the CNN's predictions for the five sets of varied source offset inputs is presented in Table 2-2 in terms of MAPE and MSSIM. From Table 2-2, it is clear that, as with the receiver number and spacing discussed previously, the frequency-velocity CNN is capable of generalizing across various source offset distances, but only within reasonable adjustments to the base configuration. For example, the MAPE and MSSIM values for the 6-m source offset are only slightly worse than the base configuration, while the values for the 10-m offset and 20-m offset show increasing error as the source offset distance increases. Nevertheless, the average MAPE for the 20-m offset only increase by approximately 7% relative to the base configuration (i.e., 12.7% compared to 6.0%).

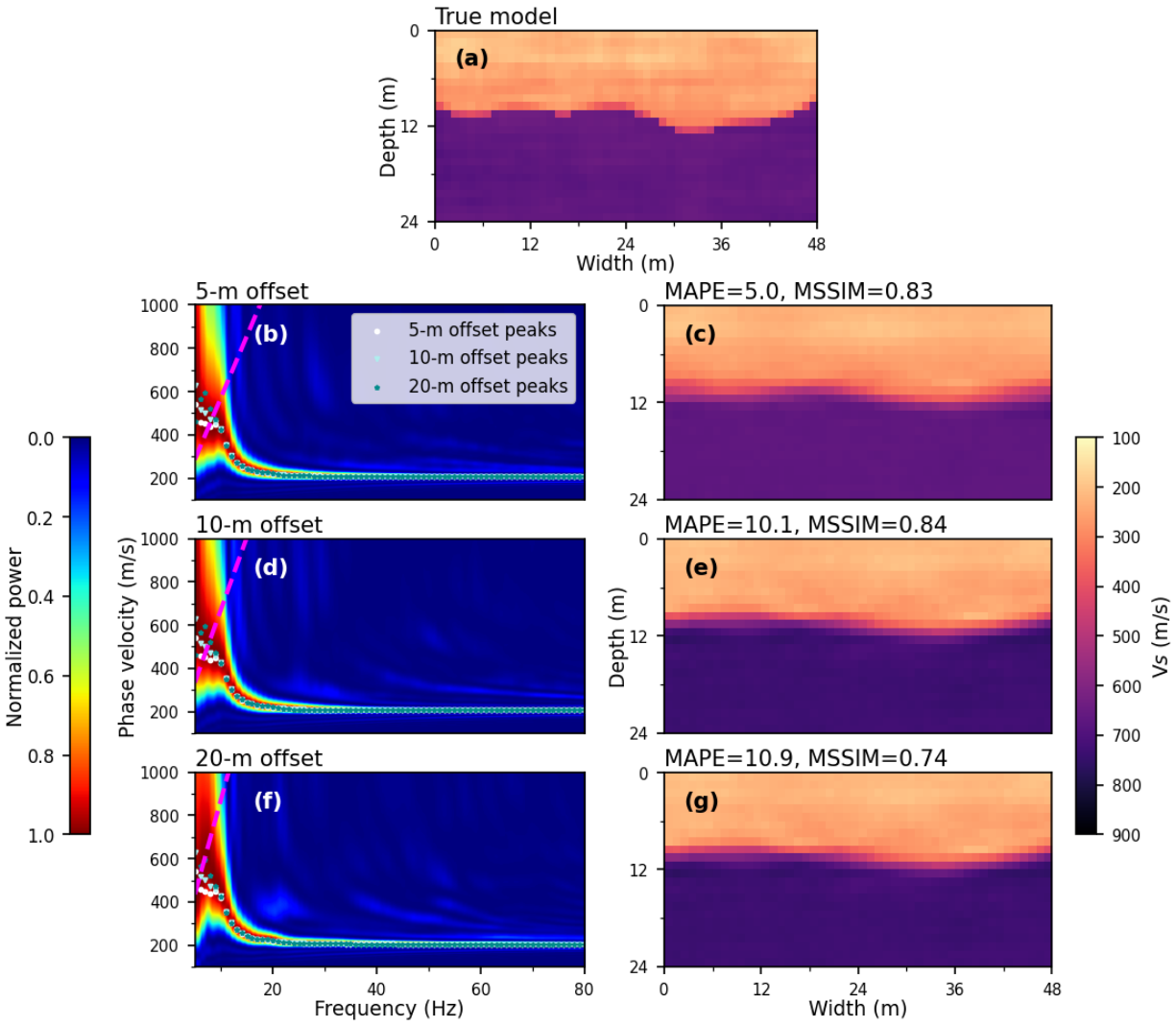

Figure 2-9 (a) the true synthetic model from Figure 2-4r. (b & c) the input dispersion image and output frequency-velocity CNN Vs image prediction, respectively, from the base receiver configuration (i.e., 5-m source offset). (d & e) the input dispersion image and output frequency-velocity CNN Vs image prediction, respectively, from a 10-m source offset. (f & g) the input dispersion image and output frequency-velocity CNN Vs image prediction, respectively, from a 20-m source offset. The magenta dashed lines represent constant wavelengths set equal to two-times the array-center distance that correspond to phase velocity errors less than 5% as presented by Yoon and Rix (2009). Peak power points obtained from the apparent fundamental Rayleigh wave mode of each dispersion image are also shown. The mean absolute percent error (MAPE) and mean structural similarity index (MSSIM) of each predicted image are presented above each panel.

Figure 2-9a shows the true V_S image depicted in Figure 2-4r. Figures 2-9b and 2-9c show the input dispersion image and CNN output Vs image, respectively, for the 5-m source offset base configuration. The input dispersion images and output Vs images for modified acquisition configurations are shown in Figures 2-9d and 2-9e, respectively, for a source offset of 10-m and in Figures 2-9f and 2-9g, respectively, for a source offset of 20-m. Also shown in all dispersion images are magenta dashed lines that delineate zones where the low frequency (i.e., long wavelength) dispersion data from the various source offset distances may be influenced by near-field effects. These lines represent constant wavelengths set equal to two-times the array-center distances that correspond to phase velocity errors less than 5%, as presented by Yoon and Rix (2009). The array-center distance is equal to the distance from the source to the center of the linear array. Hence, the larger the source offset distance, the greater the array-center distance, and the longer the maximum wavelength that can be extracted from the dispersion data without contamination from near-field effects. As noted above, near-field effects manifest in dispersion data as phase velocity estimates that are lower than actual conditions. Near-field effects can be observed at low frequencies in the dispersion images shown in Figure 2-9 by comparing the peak power points from the 5-m, 10-m, and 20-m source offset dispersion images. The peak power points from the 20-m offset have the highest phase velocity values at low frequencies, while those from the 5-m offset have the lowest phase velocity values. These differences are only visible for frequency-phase velocity pairs near the magenta dashed lines that delineate zones where low frequency data may be influenced by near-field effects. Other than these low frequency zones, the dispersion images from the various source offsets appear very similar to one another. As such, the increasing errors resulting

from increasing source offset are simply due to the fact that the CNN was trained on dispersion images generated from only a single source offset (i.e., 5 m) and this single source offset had dispersion data at lower frequencies that were slightly biased to lower velocities due to near-field effects. This does not imply that the dispersion images used for training the CNN are incorrect, but rather that the network learned to associate the near-field dispersion data with its corresponding true Vs image. We observe that the dispersion images obtained from larger source offsets have lesser near-field effects and, therefore, result in Vs image predictions that are stiffer at depth than the ones obtained from a 5-m offset. This is the greatest source of increasing error when larger source offset distances are used.

The dispersion images generated from several different offsets can be stacked together to balance out the impact of near-field effects and other differences in the wavefields caused by source location. Table 2-2 demonstrates that the increased error caused by using larger source offset distances can be combatted by stacking dispersion images from both near and far source offsets. This stacking of dispersion images is facilitated by normalizing each image by its absolute maximum power to counteract the varying dispersion image powers caused by the same source type being excited at different offset distances (i.e., closer sources having higher absolute dispersion power than distant sources). Once all images are normalized by their absolute maximum power, the images are summed and re-normalized by the maximum power at each frequency (i.e., frequency-dependent normalization). As discussed above regarding Table 2-2, this results in reduced MAPE values relative to using only a single, larger source offset. Vs images obtained from the frequency-velocity CNN after stacking the 5-m and 20-m source offset dispersion

images are shown in Figure 2-10. These 20 predicted images can be compared to their ground truth images shown in Figure 2-4, time-distance CNN predictions in Figure 2-5, and the base case frequency-velocity CNN predictions in Figure 2-6. While the MAPE values associated with the predicted Vs images in Figure 2-10 are slightly greater than those in Figure 2-5 and Figure 2-6, the images, except for a few high and low velocity artifacts, do not look significantly different. In fact, as detailed in Table 2-2, the average MAPE values across all 5,000 testing models for the combined 5-m and 20-m source offsets are only about 2% higher than the average MAPE values for the base case configuration. This demonstrates the ability of the frequency-velocity CNN to generalize to a number of different source offset distances, enabling more flexible field data acquisition.

2.6.3 Generalizing to source forcing function

Different source types are commonly used to excite the ground surface during linear-array, active-source wavefield testing. These active sources can range from large shaker trucks, to accelerated weight drops, to sledgehammer impacts, depending on the desired depth of profiling and the relative importance of the experiment. Therefore, a CNN that is capable of generalization in terms of providing accurate Vs images from wavefields collected with different source forcing functions is imperative for handling field applications.

To test the frequency-velocity CNN for acquisition generalization ability in terms of source forcing function, two sets of 5,000 input dispersion images were obtained from wave propagation simulations on 5,000 testing models using two different source forcing functions. Specifically, the first set of dispersion images was generated using a 15-Hz high-

cut filtered spike wavelet, while the second set of dispersion images was generated using a 12-second-long linear chirp over frequencies from 3-Hz to 80-Hz. Recall that these two forcing functions, as well as the base case forcing function (i.e., a 30-Hz Ricker wavelet), are shown in both the time and frequency domains in Figure 2-3. It is clear from Figure 2-3 that the time and frequency domains of the source forcing functions are drastically different. However, we will demonstrate that the exact frequency-dependent amplitude of the source is not critical when using normalized dispersion images, provided the source induces broadband energy across the frequencies of interest.

The accuracy of the CNN's predictions for the two sets of varied source forcing functions is presented in Table 2-2 in terms of MAPE and MSSIM. From Table 2-2, it is clear that the frequency-velocity CNN is capable of generalizing across source forcing functions that are drastically different from one another. For example, the MAPE and MSSIM values for the 15-Hz high-cut filtered spike are virtually identical to those of the base configuration, which used a 30-Hz Ricker wavelet. Furthermore, the MAPE and MSSIM values for the 12-second chirp are only slightly worse than those for the base configuration. Note that the finite difference wave propagation simulations had to be extended to 13 seconds to capture the reflected and refracted waves from the extended chirp propagating through the models, and a total of 113 models out of the 5,000 testing models had to be discarded due to significant numerical artifacts caused by the length and complexity of the simulations. Despite these challenges, the average MAPE based on the 12-second chirp was less than 1.5% greater than the base configuration.

The effects of using a 12-second chirp as a forcing function are more clearly visualized in Figure 2-11 by observing the Vs image predictions obtained from the

frequency-velocity CNN. These 20 predicted images are for the same 20 true synthetic models shown in Figure 2-4, and the results can be compared directly to those shown for the time-distance CNN predictions in Figure 2-5 and the base case frequency-velocity CNN predictions in Figure 2-6. On average, the MAPE values associated with the predicted Vs images in Figure 2-11 are higher than those in Figures 2-5 and 2-6, but the differences are minimal, and the images appear to be quite similar. This exhibits notable generalization capabilities for the frequency-velocity CNN in terms of using different source forcing functions, provided they contain energy across the CNN input frequency band.

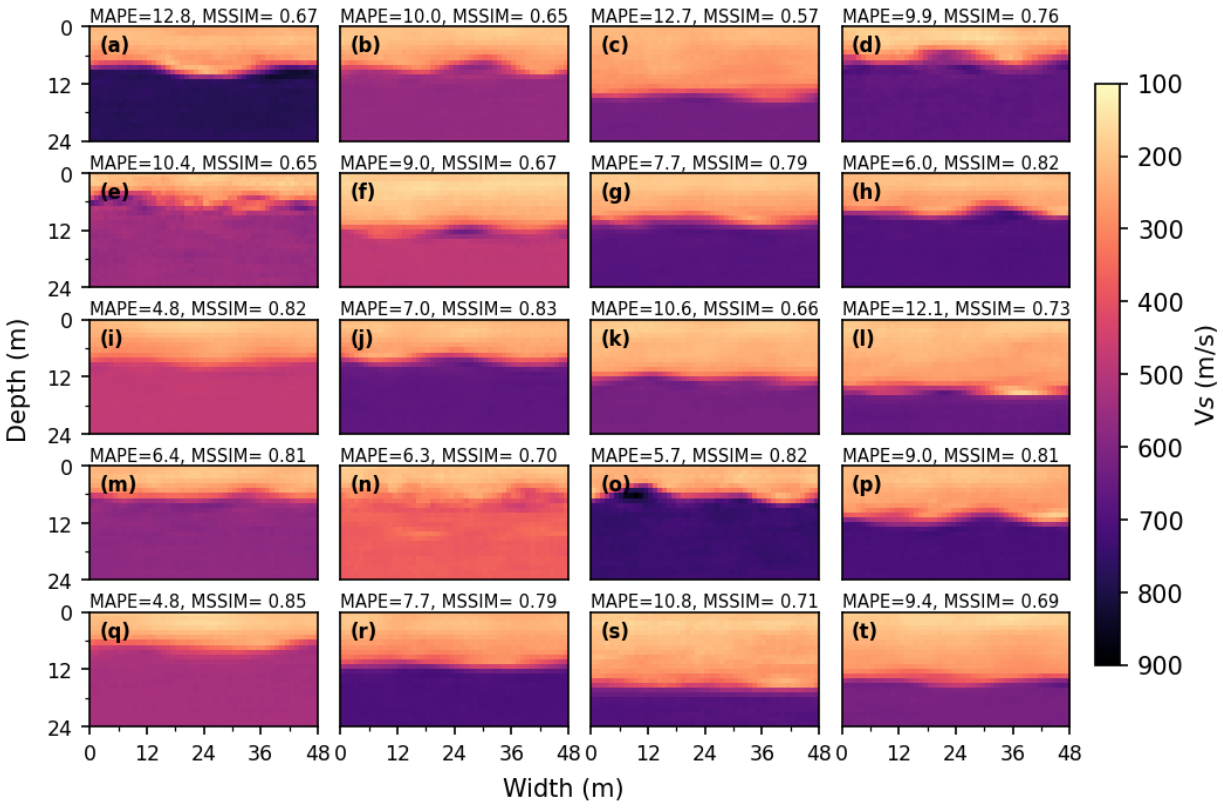

Figure 2-10 The frequency-velocity CNN's predictions of the true synthetic 2D Vs images presented in Figure 2-4. The inputs used to obtain these predictions are the normalized dispersion images obtained by averaging the 5-m and 20-m source offset dispersion images. The mean absolute percent error (MAPE) and mean structural similarity index (MSSIM) of each predicted image are presented above each panel.

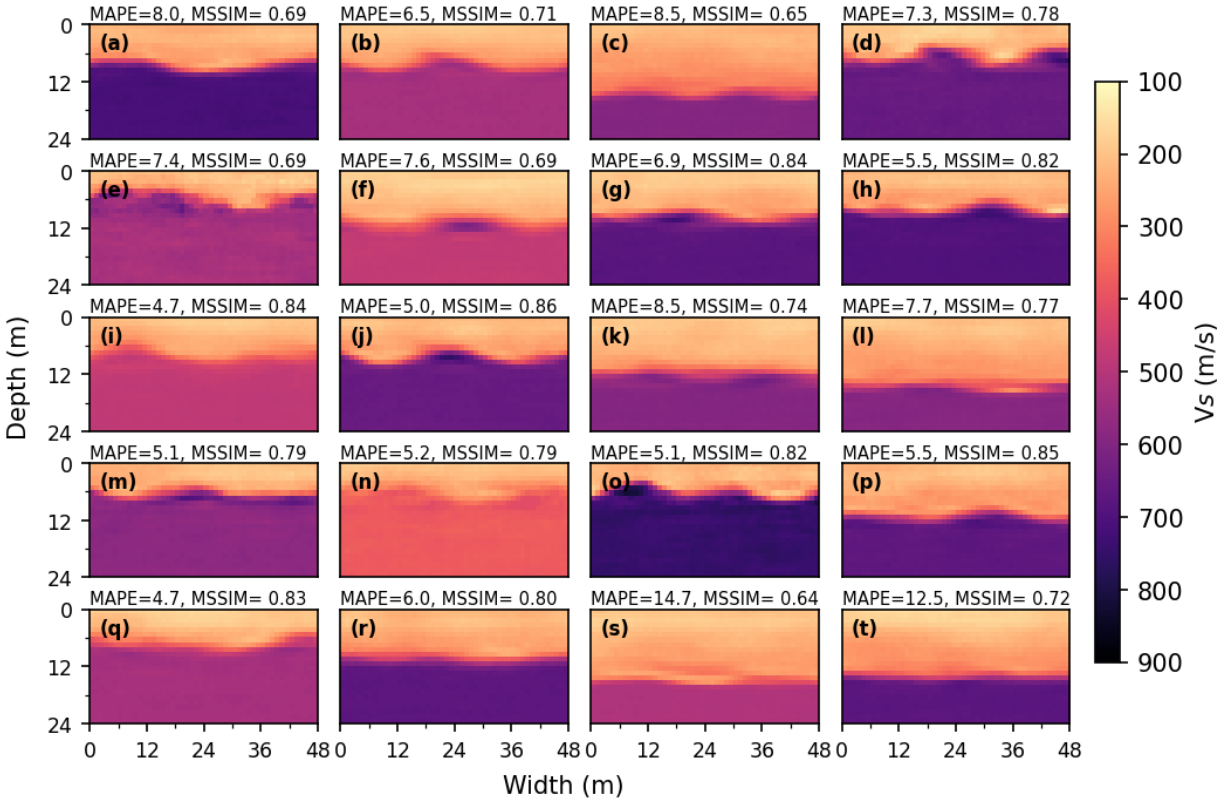

Figure 2-11 The frequency-velocity CNN's predictions of the true synthetic 2D Vs images presented in Figure 2-4. The inputs used to obtain these predictions are the dispersion images obtained by post-processing the waveforms recorded by 48 receivers at 1-m spacings, which were excited by a 12-seconds long 3-Hz to-80 Hz sweep/chirp at 5 m from the first receiver. The mean absolute percent error (MAPE) and mean structural similarity index (MSSIM) of each predicted image are presented above each panel.

2.7 Field application and validation

The proposed frequency-velocity CNN was used to predict the near-surface 2D V_S image at the NHERI@UTexas Hornsby Bend site in Austin, Texas, USA, where extensive site characterization studies have been conducted in recent years (e.g., Stokoe et al., 2020; Vantassel et al., 2022b). The wavefields used for testing the frequency-velocity CNN were actively generated using a sledgehammer to strike vertically on a square aluminum strike-plate. The wavefields were recorded by 24, 4.5-Hz vertical geophones. Five distinct

sledgehammer blows were recorded at a distance of 5 m relative to the first geophone for subsequent stacking in the time domain to increase the signal-to-noise ratio. The input to the CNN was the dispersion image obtained from the FDBF with cylindrical-steering vector and square-root-distance weighting. Figure 2-12 shows the stacked waveforms, input dispersion image, and the predicted frequency-velocity CNN 2D V_S image. The predicted Vs image indicates stiff soil (Vs ~ 200 to 300 m/s) overlying gently dipping rock, with an interface at ~ 12 to 14 m.

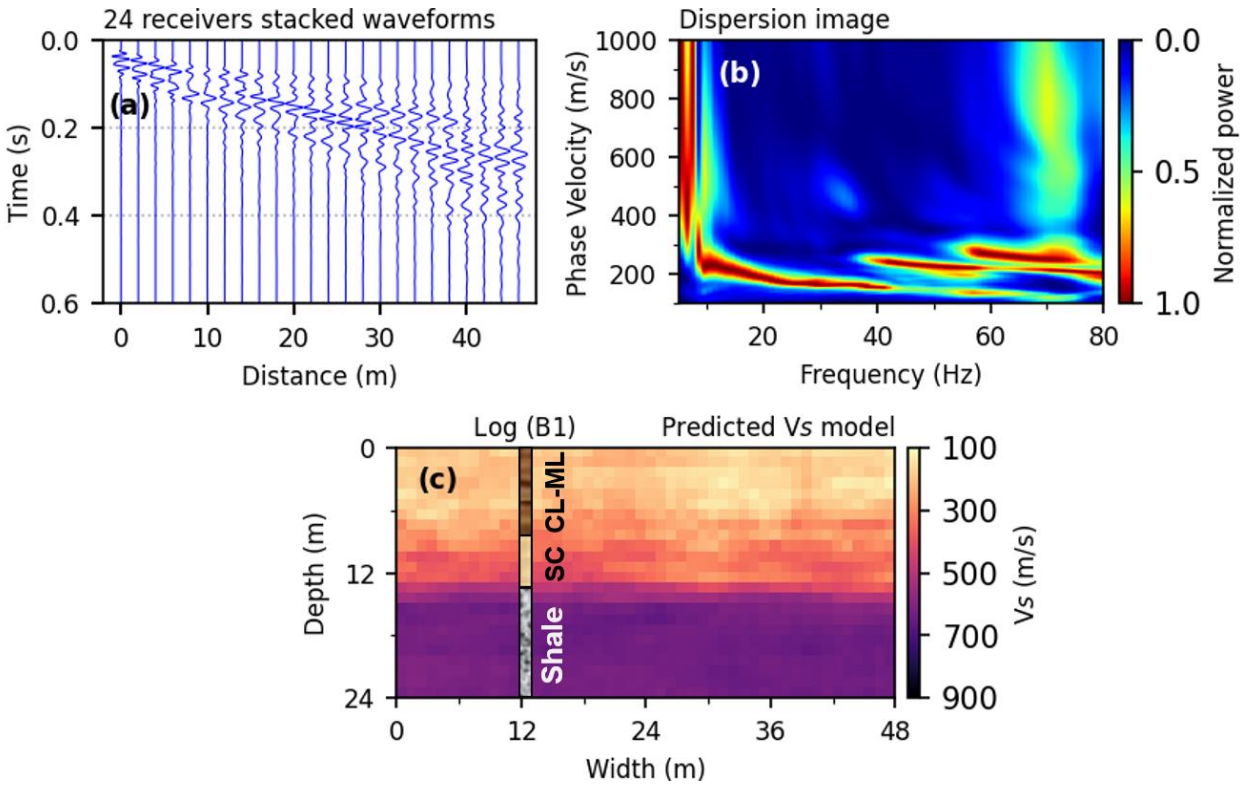

Figure 2-12 Application of the frequency-velocity CNN to linear-array, active-source wavefield measurements collected at the Hornsby Bend site. (a) the staked waveforms from five sledgehammer impacts at 5-m source offset relative to the first geophone in a linear array of 24 receivers at 2-m spacing. (b) the dispersion image associated with the wavefield presented in (a), which was used as input for the frequency-velocity CNN. (c) the frequency-velocity CNN output 2D V_S image for the Hornsby Bend site. For comparison with actual field conditions, a borehole log (i.e., B1) is superimposed on the predicted V_S image at 12.5 m, which is the location where the boring was conducted.

Two boreholes were recently drilled at the Hornsby Bend site to investigate subsurface layering down to rock; the first (i.e., B1) was located 12.5 m from the start of the geophone array, while the second (i.e., B2) was located 137.5 m away. Both borehole logs indicated a shale layer at approximately 13.5 m below the ground surface. Figure 2-12c shows a schematic of the lithology and 1D layer boundaries obtained from borehole log B1 superimposed at its correct location on the CNN-predicted 2D V_S image. Based on Figure 2-12c, the CNN was not only capable of precisely determining the depth of the shale layer, but it was also capable of characterizing the increase in stiffness from the near-surface sandy silty clay (CL-ML) soils to the underlying clayey sand (SC) soils. This is an interesting finding since the frequency-velocity CNN was only trained on two-layer synthetic models (i.e., variable soil overlying undulating rock). Nonetheless, the 2D Vs image predicted by the frequency-velocity CNN appears to properly capture the expected trends in Vs for this three-layer field site. This is a particularly notable finding given that the boreholes at the site were drilled after the CNN predictions were developed, such that the 2D Vs image was produced in a truly blind manner without any *a priori* constraints from boring logs. It is worth mentioning that the CNN input dispersion image in Figure 2-12b was generated in less than one minute, while the 2D Vs prediction in Figure 2-12c took less than two seconds to obtain.

While the borehole lithology log from B1 provides great insights into the 1D subsurface layering, it has not yet been used to perform downhole Vs profiling and can therefore not be used to judge the accuracy of the Vs predictions. So, to compare the CNN predicted 2D V_S image with other V_S estimates at the Hornsby Bend site, the extensive 1D surface wave inversions performed by Vantassel et al. (2022c) were used. Vantassel et al.

(2022c) processed MASW data collected at the Hornsby Bend site and inverted it using four layering parameterizations of 3, 5, 7, and 9 layers, based on the layering by number (LN) approach, to investigate subsurface layering uncertainty (Vantassel and Cox, 2021). They reported a suite of 1D Vs profiles representing the uncertainty in Vs across the MASW array. The median 1D Vs profiles obtained from each of these four layering parameterizations, along with the lognormal discretized median V_S profile from all the inversion realizations investigated by Vantassel et al. (2022c), are provided in Figure 2-13. The entire MASW inversion process performed by Vantassel et al. (2022c) took approximately six hours to complete on four SKX nodes on the Stampede2 supercomputer cluster. Vantassel et al. (2022c) also reported three V_S profiles along the array that were obtained from correlations to cone penetration testing (CPT) data, which are also shown in Figure 2-13. Each CPT-based 1D Vs profile was obtained from a CPT sounding by averaging the correlated V_S values from three CPT-Vs relationships developed by Hegazy and Mayne (2006), Andrus et al. (2007), and Robertson (2009). To facilitate comparison between the CNN-predicted 2D V_S image and the 1D Vs profiles reported by Vantassel et al. (2022c), the predicted 48-m wide 2D V_S image was discretized into 48 1D V_S profiles by slicing vertically through the 2D Vs image at 1-m increments, as shown in Figure 2-13. As can be seen in Figure 2-13, the median trend from the 1D MASW inversions reported by Vantassel et al. (2022c) is slightly stiffer (higher Vs) than the median trend from the 1D slices through the 2D Vs image obtained from the CNN prediction over the top 13 m, and slightly softer at depths greater than 13 m. Nevertheless, the agreement between the CNN and MASW Vs predictions is quite good. Figure 2-13 also shows good agreement between the 1D Vs profiles from the CNN prediction and the 1D V_S profiles obtained from the CPT-

Vs correlations reported by Vantassel et al. (2022c). Log B1 is also repeated in Figure 2-13 so that the expected soil-type lithology can be visualized relative to the 1D Vs profiles. The increasing stiffness from the CL-ML soils to the underlying SC soils is further validated by observing the trends in the standard penetration test (SPT) raw blow count (N) values, which are presented next to the B1 lithology log.

The strong agreement between the CNN predictions and the 1D Vs profiles derived from both MASW and CPT-Vs correlations, as well as the precision in locating the rock depth relative to the ground truth (i.e., log B1), demonstrate that the frequency-velocity CNN is capable of rapidly generating accurate Vs images for geologic conditions similar to those on which it was trained (i.e., stiff soil overlying rock). Furthermore, even though the network was trained using a 30 Hz Ricker wavelet source recorded by 48 receivers with a 1-m spacing, it was capable of utilizing field data collected with sledgehammer impacts recorded by 24 receivers with a 2-m spacing. This illustrates that the proposed frequency-velocity CNN can generalize across different field data acquisition configurations.

To further assess the 2D Vs image predicted by the frequency-velocity CNN, Figure 2-14 compares the measured and predicted MASW dispersion images. In particular, Figure 2-14a shows the measured MASW dispersion image that was computed from the stacked experimental waveforms from the sledgehammer source at the Hornsby Bend site. However, the predicted MASW dispersion image shown in Figure 2-14b required some extra work to obtain. First, the frequency-velocity CNN's predicted 2D Vs image (recall Figure 2-12c) was used to obtain a full 2D predicted subsurface model by applying the simple Vp and mass density rules discussed above in regards to synthetic model development. Then, finite difference wave propagation simulations were performed using

a 30 Hz Ricker wavelet located 5 m from the array. The waveforms were measured on a 24-receiver array at a 2-m spacing to be consistent with the actual field data. Then, the dispersion image was obtained using the FDBF with plane-wave steering vector and no amplitude weighting. For the purpose of comparison, the peak dispersion data amplitudes in Figure 2-14a are repeated in Figure 2-14b. Overall, we observe consistency between the modal trends in the measured and predicted MASW dispersion image, with the R0 mode between 10 and 40 Hz showing particularly strong agreement. However, the higher modes observed in Figure 2-14a are not as pronounced in Figure 2-14b. Nonetheless, on careful inspection of Figure 2-14b it is possible to observe some higher mode trends that show some consistency with the locations of the higher mode dispersion data present in the experimental wavefield. This comparison serves to emphasize that, while not perfect, the 2D Vs image predicted by the frequency-velocity CNN (i.e., Figure 2-12b) is relatively consistent with the experimental wavefield data and measured dispersion image (i.e., Figure 2-12a) from the field experiment.

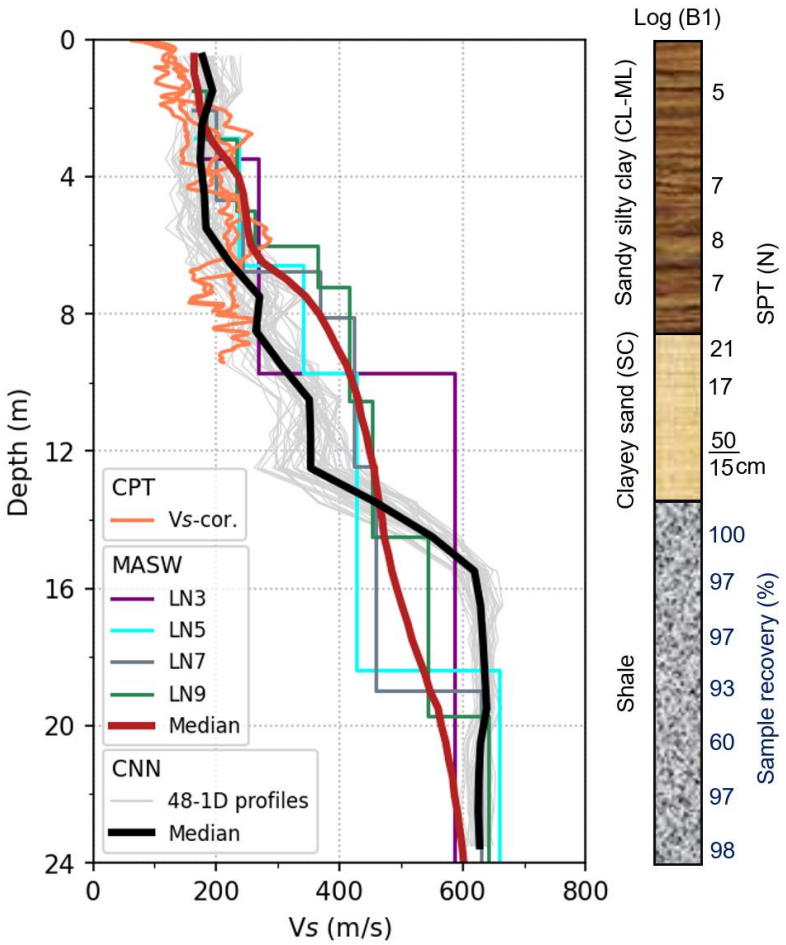

Figure 2-13 Comparison of 1D Vs profiles for the area imaged by the frequency-velocity CNN at the Hornsby Bend site. The figure shows four layer-by-layer median 1D V_S profiles from four layering by number (LN) MASW inversion parameterizations, as well as the overall lognormal discretized median 1D V_S profile from the 1D MASW inversions performed by Vantassel et al. (2022c). The figure also shows three 1D V_S profiles obtained from correlations with three CPT soundings, as reported by Vantassel et al. (2022). These 1D V_S profiles are plotted relative to the 48 1D Vs profiles extracted from the 2D Vs image obtained from the frequency-velocity CNN and their lognormal median. To compare the CNN predictions with ground truth, the lithology log from borehole B1 is also provided.

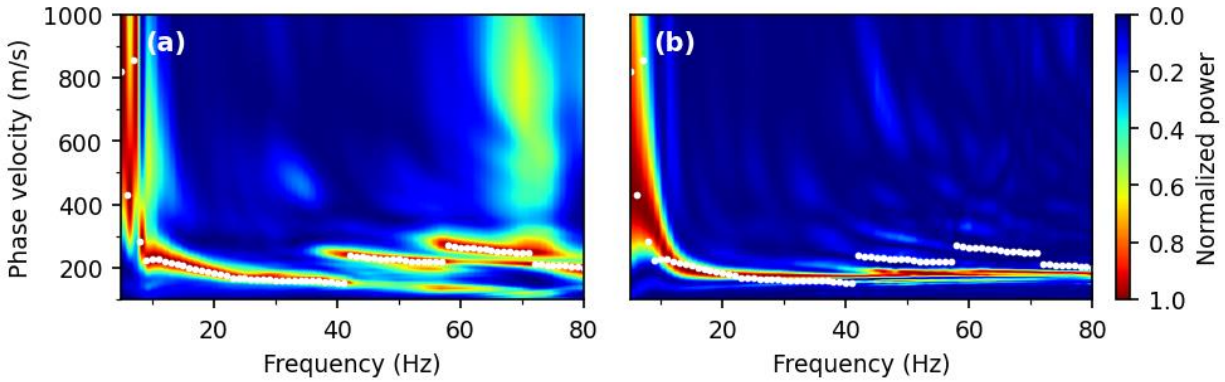

Figure 2-14 (a) dispersion image developed using field measurements, and (b) the dispersion image obtained from the predicted 2D Vs model. To allow comparisons between the dispersion images, the peak power at each frequency in panel (a) is shown in panel (b) using white circles.

2.8 Conclusions

A frequency-velocity CNN has been developed for rapid, non-invasive 2D Vs near-surface imaging using stress waves. The CNN uses a normalized dispersion image as an input and outputs a 2D Vs image. The proposed framework provides significant flexibility in the linear-array, active-source experimental testing configuration used in generating the CNN input at a given site, accommodating various source types, source offsets, numbers of receivers, and receiver spacings. Such acquisition flexibility permits the use of the developed CNN as an end-to-end imaging technique, or as a means for generating rapid starting models for FWI. A total of 100,000 soil-over-rock synthetic models were used to train, validate, and test the CNN. The testing metrics of the developed frequency-velocity CNN revealed similar prediction accuracy to the time-distance CNN recently developed within our research group, which showed great promise but lacked flexibility important for field applications. The acquisition generalization ability of the proposed frequency-velocity CNN was first demonstrated using sets of 5,000 synthetic near-surface models. For each set of 5,000 models, the inputs to the CNN were dispersion images obtained using

different testing configurations than the ones used during training the CNN. The CNN showed remarkable acquisition generalization ability with regards to the number of receivers, receiver spacings, source offset distances, and source forcing functions, as long as the testing configuration was not drastically different relative to the base case configuration on which the CNN was trained. Finally, the ability of the proposed CNN to handle field data was demonstrated using the experimental tests conducted at the Hornsby-Bend site in Austin, Texas, USA. The good agreement between the CNN's predicted 2D Vs image and the actual subsurface structure determined through 1D surface wave inversions, CPT-Vs correlations, and boring logs reinforce the capabilities of the proposed CNN for accurately retrieving 2DVs images using field data from testing configurations different from the one used during training.

2.9 Acknowledgments

The open-source software DENISE (Köhn 2011; Köhn et al., 2012) was used for all the wave propagation simulations conducted in this study. The Texas Advanced Computing Center's (TACC's) cluster Stampede2 was used in the construction of seismic wavefield-image pairs, with an allocation provided by DesignSafe-CI (Rathje et al., 2017). Google Colaboratory along with the open-source machine learning library Keras (Chollet et al., 2015) were used in training and testing the CNNs presented herein. The wavefield transformations were performed using the open-source Python package swprocess (Vantassel 2021). Matplotlib 3.1.2 (Hunter, 2007) was used to create the figures in this study. This work was supported primarily by the U.S. National Science Foundation (NSF) grant CMMI- 2120155 with equipment resources for field testing at the Hornsby Bend site associated with grant CMMI-2037900. However, any opinions, findings, and conclusions

or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

2.10 References

Adler, A., Araya-Polo, M., Poggio, T., 2021. Deep Learning for Seismic Inverse Problems. IEEE signal processing magazine, 38(2), pp. 89-119.

Andrus, R. D. et al., 2007. Predicting shear-wave velocity from cone penetration resistance. s.l., s.n.

Araya-Polo, M., Jennings, J., Adler, A., Dahlke, T., 2018. Deep-learning tomography. The Leading Edge, 37(1), pp. 58-66.

Aziman, M., Hazreek, Z. A. M., Azhar, A. T. S., Haimi, D. S., 2016. Compressive and Shear Wave Velocity Profiles using Seismic Refraction Technique. Journal of Physics: Conference Series 710 012011.

Chollet, F. others, 2015. Keras. [Online] Available at: https://github.com/fchollet/keras

Clayton, C. R. I., 2011. Stiffness at small strain: research and practice. Géotechnique, 61(1), pp. 5-37.

Courant, R., Friedrichs, K., Lewy, H., 1967. On the Partial Difference Equations of Mathematical Physics. IBM Journal of Research and Development, 11(2), pp. 215-234.

Cox, B. R., Teague, D. P., 2016. Layering ratios: a systematic approach to the inversion of surface wave data in the absence of a priori information. Geophysical Journal International, 207(1), p. 422–438.

Cox, B. R., Wood, C. M., 2011. Surface Wave Benchmarking Exercise: Methodologies, Results, and Uncertainties. s.l., ASCE.

Fam, M., Santamarina, J. C., 1997. A study of consolidation using mechanical and electromagnetic waves. Géotechnique, 47(2), pp. 203-219.

Feng, S., Lin, Y., Wohlberg, B., 2021. Multiscale Data-Driven Seismic Full-Waveform Inversion With Field Data Study. IEEE transactions on geoscience and remote sensing, Volume 60, pp. 1-14.

Forbriger, T., Groos, L., Schäfer, M., 2014. Line-source simulation for shallow-seismic data. Part 1: theoretical background. Geophysical Journal International, 198(3), pp. 1387-1404.

Foti, S., 2000. Multi-station methods for geotechnical characterisation using surface waves, s.l.: PhD Dissertation, Politecnico di Torino.

Foti, S. et al., 2018. Guidelines for the good practice of surface wave analysis: a product of the InterPACIFIC project. Bulletin of Earthquake Engineering.

Foti, S., Lai, C. G., Rix, G. J., Strobbia, C., 2014. Surface Wave Methods for Near-Surface Site Characterization. London: CRC Press.

Gabriels, P., Snieder, R., Nolet, G., 1987. In situ measurements of shear-wave velocity in sediments with higher-mode Rayleigh waves. s.l., s.n., pp. 187-196.

Hegazy, Y. A., Mayne, P. W. M., 2006. A Global Statistical Correlation between Shear Wave Velocity and Cone Penetration Data. Shanghai, China, ASCE.

Hunter, J. D., 2007. Matplotlib: A 2D Graphics Environment. Computing in Science & Engineering, 9(3), pp. 90 - 95.

Ivanov, J. et al., 2006. Delineating a shallow fault zone and dipping bedrock strata using multichannal analysis of surface waves with a land streamer. Geophysics, Volume 71.

Kingma, D. P., Ba, J. L., 2015. Adam: a method for stochastic optimization. San Diego, s.n.

Köhn, D., 2011. Time domain 2D elastic full waveform tomography, Kiel, Germany: s.n.

Köhn, D. et al., 2012. On the influence of model parametrization in elastic full waveform tomography. Geophysical Journal International, 191(1), pp. 325-345.

Komatitsch, D., Martin, R., 2007. An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation. GEOPHYSICS, 72(5), pp. 1SO-Z83.

Levander, A. R., 1988. Fourth-order finite-difference P-SV seismograms. GEOPHYSICS, 53(11), pp. 1379-1492.

Li, J., Rosenblad, B., 2011. Experimental study of near-field effects in multichannel array-based surface wave velocity measurements. Near Surface Geophysics, 9(4), pp. 357-366.

Li, S. et al., 2020. Deep-Learning Inversion of Seismic Data. IEEE Transactions on Geoscience and Remote Sensing, 58(3), pp. 2135-2149.

Liu, B. et al., 2020. Deep-Learning Inversion of Seismic Data. IEEE Transactions on Geoscience and Remote Sensing, 58(3), pp. 2135-2149.

Louie, J. N., 2001. Faster, Better: Shear-Wave Velocity to 100 Meters Depth From Refraction Microtremor Arrays. Bulletin of the Seismological Society of America.

Mao, B., Han, L.-G., Feng, Q., Yin, Y.-C., 2019. Subsurface velocity inversion from deep learning-based data assimilation. Journal of Applied Geophysics, Volume 167, pp. 172-179.

McMechan, G. A., Yedlin, M. J., 1981. Analysis of dispersive waves by wave field transformation. Geophysics, 46(6), pp. 832-958.

Menq, F.-Y., 2003. Dynamic properties of sandy and gravelly soils, s.l.: The University of Texas at Austin.

Miller, G. F., Pursey, H., 1955. On the partition of energy between elastic waves in a semi-infinite solid. s.l., Royal Society, p. 55–69..

Monteiller, V., Chevrot, S., Komatitsch, D., Wang, Y., 2015. Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the SEM–DSM hybrid method. Geophysical Journal International, 202(2), pp. 811-827.

Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data. Geophysics, 52(9).

Mosser, L., Dubrule, O., Blunt, M. J., 2018b. Stochastic seismic waveform inversion using generative adversarial networks as a geological prior. arXiv:1806.03720 [physics.geo-ph].

Mosser, L. et al., 2018a. Rapid seismic domain transfer: Seismic velocity inversion and modeling using deep generative neural networks. arXiv:1805.08826v1 [physics.geo-ph].

Nocedal, J., Wright, S. J., 2006. Numerical Optimization. s.l.:Springer Series in Operations Research and Financial Engineering.

Nolet, G., Panza, G. F., 1976. Array analysis of seismic surface waves: Limits and possibilities. pure and applied geophysics, 114(5), pp. 775-790.

Okada, H., 2003. The microtremor survey method: Geophysical Monograph Series, No. 12. s.l.:society of exploration geophysicists.

Park, C. B., 2005. MASW - Horizontal Resolution in 2D Shear-Velocity (Vs) Mapping, Kansas: Kansas Geological Survey.

Park, C. B., Miller, R. D., Xia, J., 1998. Imaging dispersion curves of surface waves on multi-channel record. s.l., s.n., pp. 1377-1380.

Park, C. B., Miller, R. D., Xia, J., 1999. Multichannel analysis of surface waves. GEOPHYSICS, 64(3), pp. 800-808.

Pratt, R. G., 1999. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. Geophysics, 64(3), pp. 659-992.

Pratt, R. G., Shin, C., Hick, G. J., 1998. Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. Geophysical Journal International, 133(2), p. 341–362.

Rahimi, S., Wood, C. M., Teague, D. P., 2021. Performance of Different Transformation Techniques for MASW Data Processing Considering Various Site Conditions, Near-Field Effects, and Modal Separation. Surveys in Geophysics, Volume 42.

Rathje, E. M. et al., 2017. DesignSafe: New Cyberinfrastructure for Natural Hazards Engineering. Natural Hazards Review, 18(3).

Robertson, P. K., 2009. Interpretation of cone penetration tests — a unified approach. Canadian Geotechnical Journal, 46(11), pp. 1337-1355.

Rosenblad, B. L., Li, C.-H., 2011. Influence of Poisson's Ratio on Surface Wave Near-Field Effects. Dallas, Texas, United States, ASCE.

Schäfer, M., Groos, L., Forbriger, T., Bohlen, T., 2014. Line-source simulation for shallow-seismic data. Part 2: full-waveform inversion—a synthetic 2-D case study. Geophysical Journal International, 198(3), pp. 1405-1418.

Scherer, D., Müller, A., Behnke, S., 2010. Evaluation of Pooling Operations in Convolutional Architectures for Object Recognition. s.l., Springer, Berlin, Heidelberg, pp. 92 - 101.

Shah, N. et al., 2012. Quality assured full-waveform inversion: Ensuring starting model adequacy. s.l., Society of Exploration Geophysicists.

Smith, J. A. et al., 2019. Tunnel detection at Yuma Proving Ground, Arizona, USA — Part 2: 3D full-waveform inversion experiments. GEOPHYSICS, 84(1), pp. 1JF-Z5.

Stokoe, K., Cox, B., Clayton, P., Menq, F., 2020. NHERI@UTexas Experimental Facility With Large-Scale Mobile Shakers for Field Studies. Frontiers in Built Environment.

Stokoe, K. I., Wright, S., Bay, J. A., Roesset, J. M., 1994. Characterization of geotechnical sites by SASW method. s.l., R.D. Woods, Oxford, IBH Publication, pp. 15-25.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation. Geophysics, 49(8), pp. 1140-1395.

Tokimatsu, K., Shinzawa, K., Kuwayama, S., 1992. Use of Short-Period Microtremors for Vs Profiling. Journal of Geotechnical Engineering, 118(10).

Vantassel, J., 2021. jpvantassel/swprocess:v0.1.0b0. Zenodo., s.l.: s.n.

Vantassel, J. et al., 2022b. Active-Source, Near-Surface, Surface-Wave Measurements using Distributed Acoustic Sensing (DAS) and Traditional Geophones, s.l.: in Characterization of the NHERI@UTexas Hornsby Bend Test Site. DesignSafe-CI.

Vantassel, J., Cox, B. R., 2021b. A procedure for developing uncertainty-consistent Vs profiles from inversion of surface wave dispersion data. Soil Dynamics and Earthquake Engineering, Volume 145.

Vantassel, J. P., Cox, B. R., 2021a. SWinvert: a workflow for performing rigorous 1-D surface wave inversions. Geophysical Journal International, 224(2), p. 1141–1156.

Vantassel, J. P., Cox, B. R., 2022. SWprocess: a workflow for developing robust estimates of surface wave dispersion uncertainty. Journal of Seismology.

Vantassel, J. P., Cox, B. R., Hubbard, P. G., Yust, M., 2022c. Extracting High-Resolution, Multi-Mode Surface Wave Dispersion Data from Distributed Acoustic Sensing Measurements using the Multichannel Analysis of Surface Waves. arXiv:2202.04779 [physics.geo-ph].

Vantassel, J. P., Kumar, K., Cox, B. R., 2022a. Using convolutional neural networks to develop starting models for near-surface 2-D full waveform inversion. Geophysical Journal International, 231(1), pp. 72-90.

Virieux, J., Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics. Geophysics, 74(6), pp. 1ND-Z107.

Virta, K., 2016. Numerics of Elastic and Acoustic Wave Motion, s.l.: Uppsala University Publications.

Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E., 2004. Image quality assessment: from error visibility to structural similarity. s.l., IEEE, pp. 600-612.

Wu, Y., Lin, Y., 2019. InversionNet: An Efficient and Accurate Data-Driven Full Waveform Inversion. IEEE Transactions on Computational Imaging, Volume 6, pp. 419-433.

Yang, F., Ma, J., 2019. Deep-learning inversion: A next-generation seismic velocity model building method. Geophysics, 84(4), pp. 1JA-Z21.

Yoon, S., Rix, G. J., 2009. Near-Field Effects on Array-Based Surface Wave Methods with Active Sources. Journal of Geotechnical and Geoenvironmental Engineering, 135(3).

Zywicki, D. J., 1999. Advanced signal processing methods applied to engineering analysis of seismic surface waves, Atlanta, Georgia, United States: Georgia Institute of Technology.

Zywicki, D. J. , Rix, G. J., 2005. Mitigation of Near-Field Effects for Seismic Surface Wave Velocity Estimation with Cylindrical Beamformers. Journal of Geotechnical and Geoenvironmental Engineering, 131(8).

CHAPTER 3

A FREQUENCY-DOMAIN BEAMFORMING PROCEDURE FOR EXTRACTING RAYLEIGH WAVE ATTENUATION COEFFICIENTS AND SMALL-STRAIN DAMPING RATIO FROM 2D AMBIENT NOISE ARRAY MEASUREMENTS

Abstract

The small-strain damping ratio plays a crucial role in assessing the response of soil deposits to earthquake-induced ground motions and general dynamic loading. The damping ratio can theoretically be inverted for after extracting frequency-dependent Rayleigh wave attenuation coefficients from wavefields collected during surface wave testing. However, determining reliable estimates of in-situ attenuation coefficients is much more challenging than achieving robust phase velocity dispersion data, which are commonly measured using both active-source and ambient-wavefield surface wave methods. This paper introduces a new methodology for estimating frequency-dependent attenuation coefficients through the analysis of ambient noise wavefield data recorded by two-dimensional (2D) arrays of surface seismic sensors for the subsequent evaluation of the small-strain damping ratio. The approach relies on the application of an attenuation-specific wavefield conversion and frequency-domain beamforming. Numerical simulations are employed to verify the proposed approach and inform best practices for its application. Finally, the practical efficacy of the proposed approach is showcased through its application to field data collected at a deep, soft soil site in Logan, Utah, USA, where phase velocity and attenuation coefficients are extracted from surface wave data and then simultaneously inverted to develop deep shear wave velocity and damping ratio profiles.

Keywords: attenuation; damping; surface wave testing; inversion; in situ; noninvasive; ambient noise; vibrations

3.1 Introduction

The small-strain shear modulus (G_{max}) and small-strain damping ratio (D) form the starting point for many soil constitutive models and play a crucial role in assessing the response of soil deposits to earthquake-induced ground motions and general dynamic loading. G_{max} is directly related to the in-situ shear wave velocity (V_s), and it represents the soil stiffness and its resistance to deformation under applied shear stress. D characterizes the energy dissipation properties of the material. The influence of D on the amplitude and frequency content of seismic waves has been recognized since at least 1940 (Ricker, 1940), with subsequent research establishing it as a pivotal parameter for seismic site response studies and for modeling ground-borne vibrations (e.g., Anderson et al., 1996; Tao and Rathje, 2019; Papadopoulos et al., 2019; Foti et al., 2021). Despite its significance, the in-situ estimation of D has received far less attention when compared to measurements of V_s (Parolai, 2014). D can theoretically be inverted for after extracting frequency-dependent Rayleigh wave phase velocity and attenuation coefficients (α) from wavefields collected during surface wave testing (Lai, 1998; Foti, 2004). However, in-situ α values are generally much more difficult to reliably measure than phase velocities (Haendel et al., 2016; Parolai et al., 2022), which are commonly measured using both active-source and ambient-wavefield surface wave methods.

This paper introduces a new noninvasive method to estimate frequency-dependent Rayleigh wave α using ambient noise wavefield data collected with two-dimensional (2D) arrays of surface seismic sensors for the subsequent evaluation of D. The approach relies

on frequency-domain beamforming (FDBF) and applies an attenuation-specific wavefield conversion, known as the FDBFa approach. While Aimar et al. (2024a) previously used this approach for active-source surface wave testing, it has not been applied to ambient noise surface wave testing. In this paper, we introduce a new method called the noise FDBFa (NFDBFa) approach and document its development and application.

The subsequent sections of this paper are organized as follows: first, we cover important background information on attenuation and damping. Second, we present a concise overview of the FDBF technique introduced by Lacoss et al. (1969) and the FDBFa wavefield conversion methodology proposed by Aimar et al. (2024a), along with the integration of these methods within our proposed NFDBFa approach. Then, synthetic studies are presented to showcase the capabilities of the proposed NFDBFa approach and inform best practices for its application. The synthetic studies offer valuable insights into the influence of 2D array size and proximity to noise sources on attenuation estimates. For example, it is demonstrated that the optimal 2D ambient noise array design principles for attenuation estimation differ from the principles governing 2D array design for phase velocity estimation. Finally, we demonstrate the practical utility of our proposed NFDBFa technique through a field application at a deep, soft soil site in Logan, Utah, USA. In this field application, phase velocity and attenuation coefficients are extracted from surface wave data and then simultaneously inverted to develop deep V_s and D profiles. The good agreement observed between the attenuation estimates derived from our new NFDBFa technique and those obtained through the standard FDBFa analysis of active-source data collected using the multichannel analysis of surface waves (MASW) provides compelling evidence of the effectiveness of our new ambient noise approach.

3.2 Background

The attenuation of seismic waves in a continuum is related to the damping ratios of both compression waves (D_p) and shear waves (D_s). Surprisingly, little is known about the relative relationship between D_p and D_s, and one can find instances in the literature where researchers have assumed $D_p = D_s$ (Badsar et al., 2010; Verachtert et al., 2017; Aimar et al., 2024b), $D_p > D_s$ (Bergamo et al., 2023), and $D_p < D_s$ (Xia et al., 2002). In this paper, when 'D' is used without a subscript, it implies that the statement or equation is valid for both D_p and D_s. The damping ratio (D) is commonly used in engineering, while its inverse, the quality factor (Q), where $Q^{-1} = 2D$, is more prevalent in seismological and geophysical literature (Foti, 2004). Consequently, Q, being the inverse of D, also varies for compressional waves (Q_p) and shear waves (Q_s).

Seismic wave attenuation is commonly attributed to three mechanisms: material damping, geometric spreading, and apparent attenuation (Zywicki, 1999). Material damping, or anelastic attenuation, arises from the collective interaction of diverse factors (Johnston et al., 1979). These factors encompass frictional losses among solid particles and fluid flow losses due to the relative motion between solid and fluid phases, a phenomenon particularly notable in coarse-grained soils (Biot, 1956; Walsh, 1966 and 1968; Stoll, 1974). Fine-grained soils, however, showcase more intricate phenomena influenced by electromagnetic interactions between water dipoles and microscopic solid particles (Lai, 1998). This intrinsic material damping is typically approximated as frequency-independent, especially within the seismic frequency band, primarily spanning 0.1 to 10 Hz (Aki and Richards, 1980; Shibuya et al., 1995). Material damping gives rise to a cyclic stress-strain curve exhibiting a hysteretic loop and is commonly referred to as hysteretic damping (Rix et al., 2000; Parolai et al., 2022).

Geometric or radiation damping involves the spread of a fixed amount of energy over a broader area or volume as the wavefront moves away from the source. Take, for instance, a harmonic unit point load applied along the normal direction to the surface of a homogeneous and isotropic half-space; this perturbation generates both body waves and Rayleigh waves. The body waves propagate radially from the source, forming a hemispherical wave front, while Rayleigh waves travel outward along a cylindrical wave front. As these waves travel, they traverse an expanding volume of material, leading to a decrease in energy density as the distance from the source increases. The amplitude of the body waves attenuates in proportion to the ratio of r^{-1} (where r is the radial distance from the source), except when along the surface of the half-space. In that case, the amplitude attenuates proportionally to r^{-2}. Conversely, the amplitude of the Rayleigh waves attenuates as $r^{-0.5}$ (Lamb, 1904; Ewing et al., 1957; Richart et al., 1970). Consequently, at substantial distances from the surface source, the dominant influence on overall particle motion stems from the surface wavefield (Lai, 1998). It is worth mentioning that these geometric spreading rules do not hold with transient waveforms (Keilis-Borok, 1989) or non-homogeneous media (Lai, 1998).

Apparent attenuation includes wave scattering, which arises from the interaction of waves with heterogeneities along the seismic path (O'Doherty and Anstey, 1971; Spencer et al., 1977), as well as the reflection and transmission of seismic waves at interfaces and mode conversions (Rix et al., 2000). Therefore, apparent attenuation is highly site-specific and difficult to generalize.

Both laboratory tests and in-situ methods have been proposed to estimate *D*. Laboratory tests, such as the resonant column (ASTM D4015-21), are valuable for

parametrically studying the material/intrinsic damping ratio, but they cannot capture the other two mechanisms contributing to the attenuation of seismic waves in situ. Conversely, the damping ratio estimates obtained using in-situ methods are influenced by all the seismic wave damping mechanisms mentioned above (Parolai et al., 2022). In-situ methods also have the advantage of assessing soil characteristics in their natural and undisturbed state (Rix et al., 2000). Additionally, in-situ tests encompass a greater soil volume, effectively reducing biases that might arise from localized variations in soil properties (Badsar et al., 2010). Furthermore, they provide parameter estimates on a spatial scale relevant to common engineering applications (e.g., Comina et al. 2011). In the scope of estimating *D*, in-situ methods can be dissected into two categories: invasive and noninvasive methods. Invasive methods encompass techniques such as cross-hole testing (Jongmans, 1989; Hall and Bodare, 2000) and downhole testing (Michaels, 1998; Crow et al., 2011). Noninvasive methods, particularly surface wave techniques, offer numerous advantages. By situating sensors at the ground surface, surface wave methods accelerate data acquisition, minimize costs, streamline validation of soil-receiver coupling, and encompass a frequency range closely aligned with those pertinent to earthquake engineering applications (Rix et al., 2000; Verachtert et al., 2017; and Parolai et al., 2022).

Surface wave testing became popular in the 1980's as an effective way to non-invasively develop 1D layering and V_s profiles for both soil deposits and pavement systems (e.g., Nazarian et al., 1983; Stokoe et al., 1989). Typically, the use of surface wave methods involves acquiring experimental phase velocity dispersion data through active-source methods, ambient-noise methods, or a combination of both (Tokimatsu, 1995). These dispersion data are then inverted to obtain layered subsurface models, with the primary

goal of resolving changes in V_S. The combined use of active-source and ambient-noise methods facilitates the generation of dispersion data across a wide frequency range, which enables resolution of both near-surface and deeper layers. Active sources predominantly produce energy concentrated at higher frequencies, typically ranging from several Hertz to perhaps 100 Hertz, with limited energy generation below 5-10 Hz for small sources like sledgehammers and drop weights. Consequently, the effective profiling depth using active-source methods is often constrained to approximately 15 to 40 m, contingent on the subsurface velocity and source mass (Foti et al., 2018). The primary hindrance to achieving increased penetration depths lies in generating lower frequency (i.e., longer wavelength) waves with affordable and highly-portable sources. This difficulty is circumvented by ambient-noise methods, which do not involve the active generation of wave energy. Instead, they rely on ground motions induced by cultural noise and microtremors (i.e., ambient noise), encompassing an abundance of low-frequency components (Lai, 1998). Consequently, ambient-noise surveys offer valuable insights for deep characterization, extending to depths of hundreds of meters or more (Foti et al., 2014; Teague et al., 2018a). Nevertheless, the spectral power of microtremors is generally low at higher frequencies (Peterson, 1993), which limits their ability to resolve changes in stiffness near the ground surface (Tokimatsu, 1995; Foti et al., 2014). Combining both active and ambient-noise measurements offers a solution to overcome this limitation.

Ambient-noise surveys typically employ 2D arrays of surface seismic sensors due to the a-priori unknown location of the ambient noise sources. Unlike linear arrays, 2D arrays allow for the determination of wave propagation direction, which is necessary for resolving the true phase velocity (Cox and Beekman, 2011). While 2D ambient noise array

measurements have been referred to using several names, in this paper we will refer to them as microtremor array measurements (MAM; Ohrnberger et al., 2004; Teague et al., 2018b). A schematic representation of a typical survey utilizing both active and ambient-noise arrays is presented in Figure 3-1a. The active-source array in Figure 3-1a is in accordance with the MASW method (Park et al., 1999), utilizing a linear array of receivers to capture the wavefield generated by active sources off each end of the array. Example waveforms recorded by 24 receivers placed in-line with one of the active sources to the left of the array are depicted in Figure 3-1b. The ambient wavefield array depicted in Figure 3-1a is in accordance with MAM testing, where surface sensors are deployed in a 2D circular pattern (note that other 2D geometries are also permissible). Example ambient noise waveforms recorded by nine sensors in the circular array are depicted in Figure 3-1d. Figure 3-1c schematically illustrates phase velocity dispersion data that are commonly extracted from active-source MASW waveforms and ambient noise MAM waveforms using various well-known wavefield transformation techniques (Vantassel and Cox, 2022). Examples of these techniques include frequency-domain beamforming (FDBF; Lacoss et al., 1969), high-resolution frequency-wavenumber (f-k) spectrum analysis (Capon 1969), cylindrical FDBF (Zywicki 1999; Zywicki and Rix, 2005), and Rayleigh three-component beamforming (Wathelet et al., 2018). The combined dispersion data from MASW and MAM spans a wide frequency range, encompassing both low frequencies obtained from the MAM testing and high frequencies obtained from the MASW testing, with some overlap in between. The phase velocity dispersion data are then typically used to solve the parameter identification problem (i.e., inversion) and obtain 1D V_s profiles of the

subsurface (Foti et al., 2018; Vantassel and Cox, 2021). Note that the inversion step and resulting V_s profiles are not illustrated schematically in Figure 3-1.

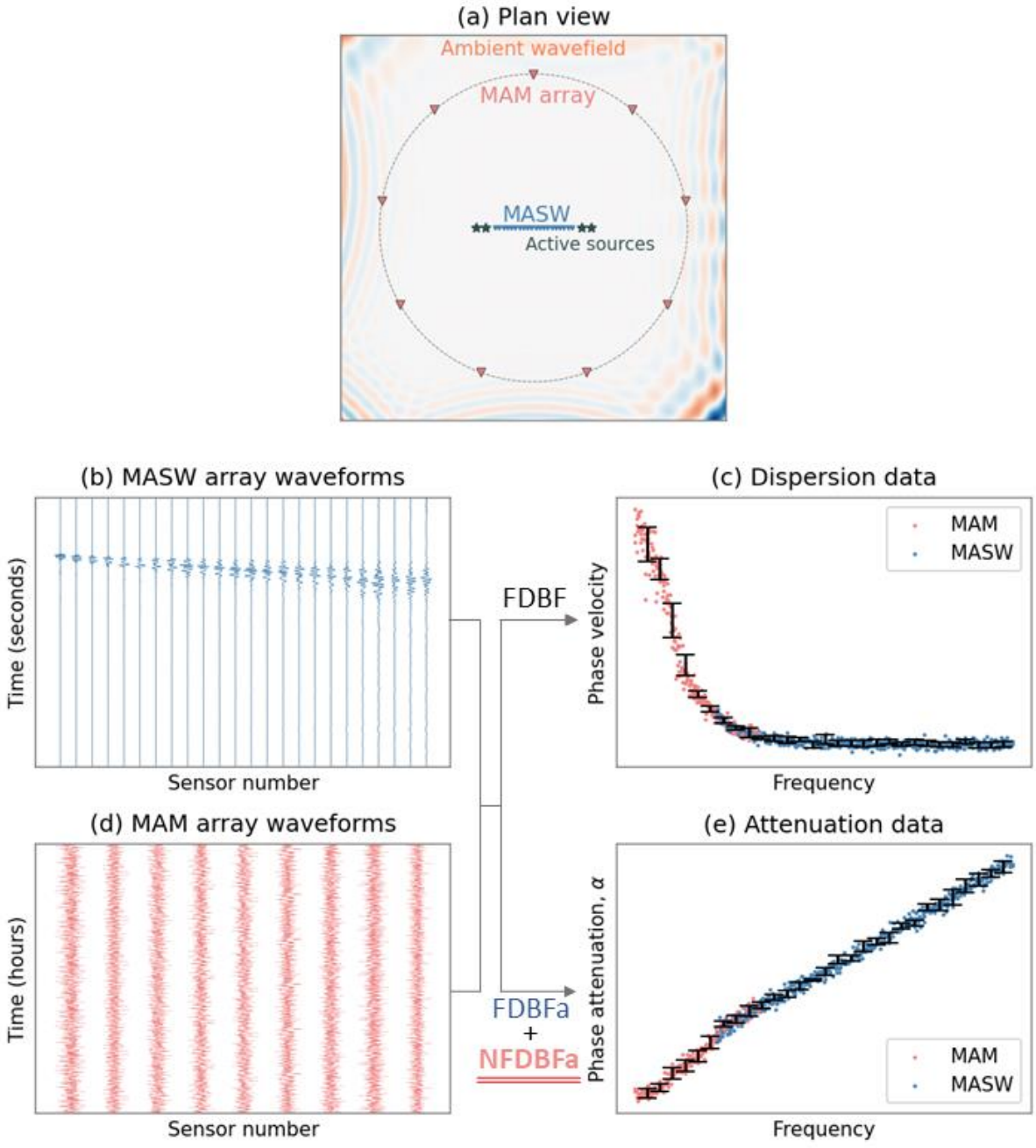

Figure 3-1 Schematic illustrating the data acquisition and processing stages of active-source and ambient-wavefield surface wave testing used to extract phase velocity and phase attenuation data. Panel (a) presents a typical acquisition setup consisting of concentric MASW and MAM arrays, featuring active sources for the MASW array and an ambient wavefield for the MAM array. Panel (b) shows waveforms from a single active-

source location collected using the MASW array, while Panel (c) presents the combined phase velocity dispersion data resulting from MASW and MAM Frequency Domain Beamforming (FDBF) processing. Panel (d) depicts the ambient noise waveforms collected from the MAM array. In Panel (e), phase attenuation data processed through active-source FDBFa and ambient-wavefield NFDBFa techniques are illustrated.

As noted above, much more effort has been devoted to extracting phase velocity information from surface wave approaches than to extracting attenuation information. Nonetheless, multiple active-source methods have been developed to estimate the attenuation of surface waves. The methods introduced by Lai (1998), Lai et al. (2002), Rix et al. (2000), Xia et al. (2002), and Foti (2004) are founded on assessing the spatial decay of Rayleigh waves, a phenomenon that is influenced by both D_p and D_s, as described by Aki and Richards (1980). These approaches assume the dominance of a single Rayleigh wave mode of propagation. Consequently, they might yield inaccurate results in soil profiles where multiple surface wave modes significantly contribute to the wavefield propagation (Rix et al. 2001). Badsar et al. (2010) introduced the half-power bandwidth method, originally developed in the field of mechanical and structural dynamics to determine the modal damping ratio of a structure, to assess Rayleigh modal attenuation by analyzing the width of the Rayleigh peaks in the f-k domain. Verachtert et al. (2017) employed the circle fit method, originally developed to determine eigenfrequencies and modal damping ratios in structural dynamics (Ewins 1984), to estimate multimodal Rayleigh dispersion and attenuation curves. Both the half-power bandwidth and circle fit methods facilitated the determination of modal attenuation curves from multimode wavefields (Verachtert et al., 2017). Recently, Aimar et al. (2024a) introduced an innovative technique that combines a novel wavefield conversion approach coupled with FDBF (Lacoss et al., 1969) for processing active-source data collected using MASW to

estimate the frequency-dependent α values. They called this the FDBF attenuation (FDBFa) method. Notably, the wavefield conversion proposed by Aimar et al. (2024a) to extract α differs from the conventional wavefield transformations commonly used to go from the time-distance domain to the f-k domain, as detailed in the following section. To avoid confusion, we will refer to the wavefield transformation proposed by Aimar et al. (2024a) as 'wavefield conversion,' while reserving the term 'wavefield transformation' specifically for the more common f-k domain transformations used to extract phase velocity data.

While important research on extracting phase attenuation coefficients using active-source methods is ongoing, similar to phase velocity data, combining active-source and ambient noise methods is desirable for resolving attenuation data over a broader frequency band. The majority of ambient noise techniques aimed at estimating the attenuation of surface waves were developed for regional-scale estimation (Haendel et al., 2016; Parolai et al., 2022). Only a limited number of approaches have considered local scales that hold relevance for engineering purposes, like site-specific seismic ground response analyses or dynamic vibration studies. These local-scale approaches are predominantly based on retrieving attenuation properties from the cross-correlation of seismic noise (e.g., Albarello and Baliva, 2009; Parolai, 2014; Haendel et al., 2016). Albarello and Baliva (2009) proposed a methodology that reconstructs the Green's function based on the temporal derivative of averaged cross-correlations from noise recordings obtained by pairs of geophones, thereby incorporating attenuation effects into the process. They further validated this approach by demonstrating its potential in estimating attenuation coefficients at two distinct sites. Parolai (2014) estimated the Rayleigh phase velocity and attenuation

coefficients by fitting a damped zero-order Bessel function, introduced by Prieto et al. (2009), using data generated from the space correlation function introduced by Aki (1957). To mitigate the impact of uneven source distribution on cross-correlations, Haendel et al. (2016) employed a higher-order noise cross-correlation technique to extract the phase velocity and attenuation coefficient of Love waves. They illustrated that their approach yields correlation functions with higher signal-to-noise ratios compared to simple noise cross-correlations.

The importance of seismic noise cross-correlation methods cannot be underestimated. Nonetheless, in theory, the reconstruction of the full Green's function requires the noise wavefield energy to be equally partitioned in all directions (Sánchez-Sesma and Campillo, 2006; Snieder et al., 2007). This is a highly specific condition that rarely met rigorously by ambient noise on Earth (Cupillard and Capdeville, 2010; Tsai, 2011; Haendel et al., 2016). Furthermore, while travel time measurements from cross-correlation of ambient noise are theoretically understood, amplitude measurements lack a corresponding theoretical background, except when the noise is equipartitioned (Snieder et al., 2007; Tsai, 2011). Studies by Cupillard and Capdeville (2010) and Tsai (2011) have shown that attenuation estimates using cross-correlations are significantly influenced by the distribution of the noise sources. In light of the challenges posed by the equipartitioning condition for the reconstruction of the full Green's function in ambient noise studies (Sánchez-Sesma and Campillo, 2006; Snieder et al., 2007), and considering the limitations highlighted by Cupillard and Capdeville (2010) and Tsai (2011) regarding the influence of noise source distribution on attenuation estimates, we introduce a paradigm-shifting approach herein for calculating attenuation coefficients from ambient noise. This novel

method not only eliminates the need for an equipartitioned noise wavefield, but also remains robust in the face of uneven noise source distribution, marking a departure from existing methodologies.

This paper builds upon Aimar et al.'s (2024a) work on developing an FDBFa technique for estimating α from active-source MASW testing and expands the FDBFa approach to ambient noise data recorded using MAM. Importantly, using an FDBF approach enables the actual direction of ambient noise propagation to be determined for each noise window and frequency, and does not require equipartitioning of ambient noise energy. Furthermore, using an FDBF approach enables the phase attenuation data generated from MASW and that from MAM to be combined in order to generate phase attenuation data spanning a broader frequency range, as illustrated schematically in Figure 3-1e. The experimental dispersion and attenuation data can then be combined and inverted to determine not only the V_s profile but also the D profile of the subsurface to greater depths. This inversion of dispersion and attenuation data to obtain V_s and D profiles can be carried out either sequentially, as demonstrated in the work of Rix et al. (2000), or simultaneously, as shown by both Lai (1998) and Aimar et al. (2024b).

3.3 Wavefield conversion proposed by Aimar et al., (2024a)

The method introduced by Aimar et al. (2024a) to estimate Rayleigh wave attenuation (α) assumes that the recorded wavefield is dominated by planar surface waves, specifically Rayleigh waves observed in the far field, with a dominant propagation mode. Several techniques have been developed to estimate the wavenumber (k) and therefore the phase velocity from such wavefields (e.g., Lacoss et al., 1969; Capon, 1969; Zywicki and Rix, 2005; Wathelet et al., 2018). Aimar et al. (2024a) harnessed this concept and

introduced a novel wavefield conversion approach that provides a pathway for calculating α by utilizing methods from existing literature originally developed for estimating *k*. The methodology involves converting the recorded wavefield into a function interpreted as a pseudo-wave. This pseudo-wave exhibits dispersion characteristics reflecting the phase attenuation of the original wave. The determination of α then becomes straightforward through the application of existing techniques for estimating *k*.

Consider the harmonic, exponentially decaying displacement wavefield, *U(r)*, depicted in Figure 3-2a and expressed by Equation 1. This wavefield is observed at many discrete distances at a specific moment in time and is induced by the passage of a monochromatic plane wave. Within this wavefield, α governs the amplitude decay resulting from material damping in accordance with a viscoelastic constitutive model. When the wavefield is plotted as log amplitude versus radial distance (*r*) from the source, the slope of the amplitude decay is α, as illustrated in Figure 3-2b. When the wavefield is plotted as phase angle versus *r*, *k* denotes the slope of the unwrapped phase (i.e., the linear phase shift), as shown in Figure 3-2c. Aimar et al. (2024a) proposed raising the recorded wavefield, *U(r)*, to the power of the imaginary number, i (see Equation 2). Consequently, a pseudo displacement wavefield, $v(r)$, is generated, wherein the wavenumber is modulated by α, signifying that when the unwrapped phase of the converted wavefield is graphed against radial distance, the slope of that phase corresponds to the value of α (refer to Figure 3-2d). Conversely, when the log amplitude of the converted wavefield is plotted against distance, the slope manifests as *k*, with an inverted sign (refer to Figure 3-2e). This wavefield conversion allows for estimating α using any of the already established and

common wavefield transformation techniques for calculating k (e.g., f-k or FDBF methods).

$$U(r) = e^{-\alpha r} e^{-ikr} \quad (1)$$

$$v(r) = [U(r)]^i = e^{-i\alpha r} e^{kr} \quad (2)$$

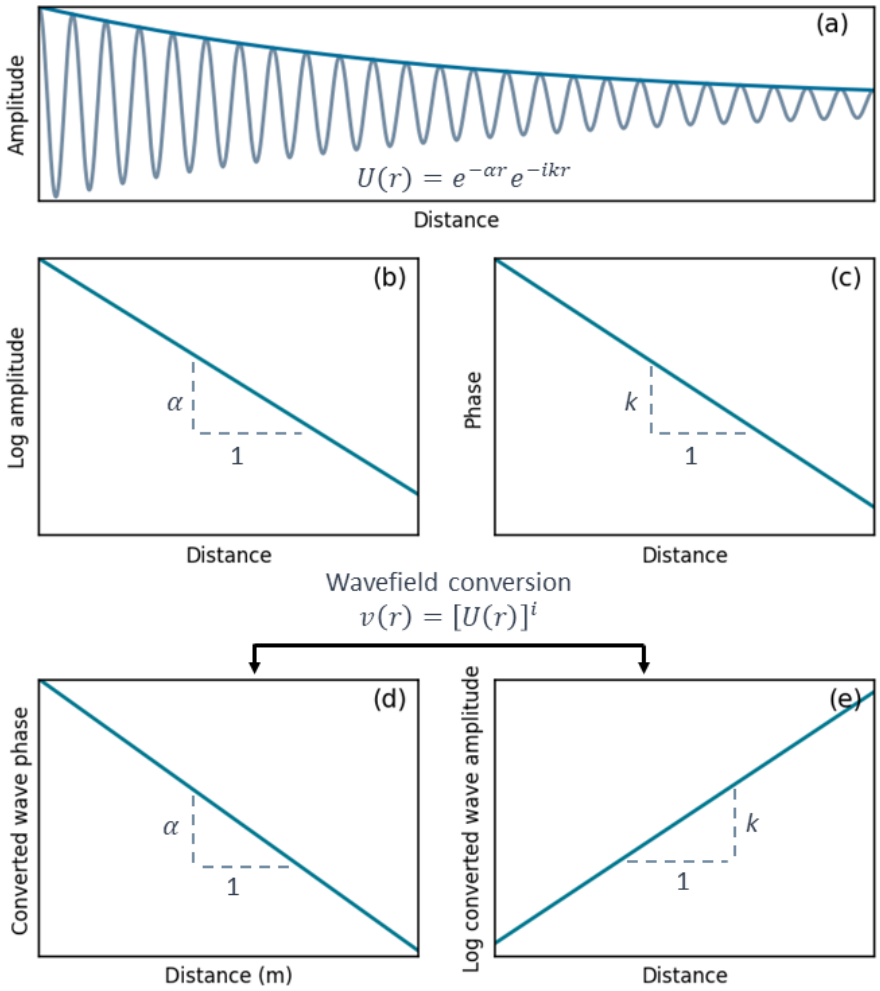

Figure 3-2 Schematic illustrating the wavefield conversion approach proposed by Aimar et al. (2024a) to extract attenuation coefficients (α). Panel (a) displays the particle displacement of a monochromatic wave experiencing exponential amplitude decay with distance, indicative of material damping in a viscoelastic constitutive model. Panel (b) depicts linear amplitude decay in log amplitude versus linear distance space, where the slope represents the phase attenuation coefficient. In Panel (c), the modulation of the unwrapped phase slope with distance by the wavenumber (k) is demonstrated. Panel (d) illustrates the modulation of the unwrapped phase slope by the phase attenuation coefficient in the converted wavefield. Panel (e) showcases the control of the slope of the log amplitude decay with linear distance by the wavenumber, albeit with an inverted sign.

This wavefield conversion can also be extended to a broadband wavefield, comprised of a superposition of monochromatic plane waves by exponentiating the wavefield in the frequency domain with the power of the imaginary number. To address numerical artifacts introduced by the wrapped phase on the pseudo wavefield, Aimar et al. (2024a) recommended normalizing *v(r)* by its amplitude on a frequency-by-frequency basis. Aimar et al. (2024a) showed that this wavefield conversion can be successfully applied to active-source wavefields recorded using MASW as a means to estimate α. In this paper, we extend this approach to estimate α from ambient noise wavefields recorded using MAM arrays, employing the FDBF technique introduced by Lacoss et al. (1969).

3.4 Noise frequency domain beam forming - attenuation (NFDBFa)

The inherent challenge in ambient noise measurements stems from the lack of *a priori* information about the source location or the direction of wave propagation, necessitating the use of spatial 2D arrays to determine the noise propagation directions during post-processing (Zywicki, 1999). As ambient noise wavefields operate in two spatial dimensions (e.g., x and y), it is necessary to represent the wavenumber using 2D vectors (Johnson and Dudgeon, 1993; Zywicki, 1999), where $\vec{k} = k_x\hat{\imath} + k_y\hat{\jmath}$, and $\hat{\imath}$ and $\hat{\jmath}$ are unit vectors in the x and y directions, respectively. Similarly, $\vec{\alpha}$ is also expressed as a 2D vector (i.e., $\vec{\alpha} = \alpha_x\hat{\imath} + \alpha_y\hat{\jmath}$) in this paper. Beamforming refers to a diverse set of array processing algorithms that concentrate the signal-capturing capabilities of the array in a specific direction. The fundamental concept behind beamforming is straightforward: when a propagating signal exists within an array's aperture, the outputs of the sensors, delayed by appropriate amounts and added together, enhance the coherent signal while mitigating the incoherent signal from waves propagating in different directions. The delays that

enhance the signal are directly linked to the time it takes for the signal to travel between sensors (Johnson and Dudgeon, 1993). Delays in the time domain correspond to linear phase shifts in the frequency domain, providing information about the wavenumber. FDBF calculations are exclusively performed within the frequency domain. Applying FDBF to the original wavefield, $U(\vec{r})$, provides information about $\vec{k}$, which informs the estimation of the phase velocity. This paper aims to demonstrate that applying FDBF to the converted, normalized pseudo wavefield, $v(\vec{r})$, informs the estimation of $\vec{\alpha}$. Henceforth, in this paper, we will denote FDBF applied to the converted noise wavefield as NFDBFa, emphasizing its role in estimating the phase attenuation from ambient noise.

In the NFDBFa approach, the first step is to partition the noise data collected by a 2D array of *m* sensors into *B* time windows. The *m* sensors are located at the ground surface at coordinates (x_i, y_i) denoted by the vector $\vec{r_i}$, where *i* varies from 1 to *m*. For each time window, Fourier spectra are calculated. Following this, the complex number at each frequency in the spectra is exponentiated to the imaginary power. Then, each exponentiated complex number is normalized by dividing it by its absolute amplitude. This process is conducted to obtain the normalized spectra of the pseudo wavefield (Aimar et al., 2024a). These spectra are then used to compute the Hermitian symmetric spatio-spectral correlation matrix, R_{ij}, with i and j representing indices of the *m* sensors in the 2D array, using Equation 3:

$$R_{ij}(\omega) = \frac{1}{B}\sum_{n=1}^{B} v_{i,n}(\omega)\, v_{j,n}^{*}(\omega) \tag{3}$$

where $R_{ij}(\omega)$ is the averaged pseudo cross-power spectrum between the i^{th} and j^{th} sensors in the array across all windows, $v_{i,n}(\omega)$ is the normalized pseudo spectra of the i^{th} sensor's data in the n^{th} window, * indicates complex conjugation, and ω is the angular

frequency. Despite being frequency-dependent, the spatio-spectral correlation matrix conveys spatial wavefield properties. Power within specific frequency-phase attenuation (f-$\vec{\alpha}$) pairs is determined by steering the array towards various directions and potential phase attenuation values. Array steering involves exponential phase shift vectors determined by trial $\vec{\alpha}$ values in pseudo space, as given by Equation 4:

$$e(\vec{\alpha}) = [\exp(-i\vec{\alpha}.\overrightarrow{r_1}), \dots, \exp(-i\vec{\alpha}.\overrightarrow{r_m})]^T \quad (4)$$

where $e(\vec{\alpha})$ is a steering vector associated with a trial $\vec{\alpha}$ and T denotes the transpose of the vector. The power in a particular f-$\vec{\alpha}$ pair, $P_{NFDBFa}(\vec{\alpha}, \omega)$, is estimated by multiplying $R_{ij}(\omega)$ by $e(\vec{\alpha})$ and summing the total power over all sensors, as given by Equation 5:

$$P_{NFDBFa}(\vec{\alpha}, \omega) = e^H(\vec{\alpha}) R_{ij}(\omega) e(\vec{\alpha}) \quad (5)$$

where H indicates the Hermitian transpose. The steering vectors aim to align the array with plane waves propagating from a specified direction and phase attenuation for each frequency. The successful alignment results in a peak within the $P_{NFDBFa}(\vec{\alpha}, \omega)$ pseudo-spectrum estimate. Thus, the NFDBFa technique presented herein allows for estimating α from ambient noise data without requiring an equipartitioned wavefield.

Even though there are similarities between the FDFBa method proposed by Aimar et al. (2024a) for estimating α using an MASW test setup and the NFDBFa method introduced in this study, there are notable differences between the two. Part of the difference is a consequence of the inherent dissimilarities between MASW and MAM. In the FDBFa method, the source location is predetermined and the array is aligned with the source, simplifying the problem and enabling the use of wavefield transformations like cylindrical frequency domain beamforming (Zywicki and Rix, 2005). Moreover, the signal-to-noise ratio can be readily enhanced by time-domain or frequency-domain

stacking, as advocated by Vantassel and Cox (2022) and Foti et al., (2018). Additionally, dispersion and attenuation uncertainties can be quantified using the multiple source offset approach proposed by Cox and Wood (2011). In contrast, the NFDBFa approach developed in this study encounters distinct challenges, primarily arising from the *a priori* unknown location of the source(s). This necessitates the utilization of 2D arrays and involves azimuthally scanning the 2D space to ascertain the direction of the most coherent source of energy at each frequency for each window. Furthermore, the enhancement of the coherent noise-to-incoherent noise ratio involves averaging multiple time windows, while uncertainty quantification involves analyzing various time blocks, each composed of different windows. Thus, in the NFDBFa approach, data are recorded for significantly longer durations (i.e., hours) compared to FDBFa (i.e., seconds). Additionally, the NFDBFa approach relies on measurements of ambient noise, which is typically assumed to be generated by sources located in the far field. If this assumption holds true, it helps to mitigate the impact of geometric spreading, which plays a significant role on attenuation estimates near an active source (Badsar, 2012). Nearfield noise sources lead to complications in extracting accurate attenuation estimates, as discussed in greater detail below.

Figure 3-3 presents examples of the FDBF and NFDBFa responses obtained from a synthetic wavefield recorded by a ten-receiver circular MAM array for a single frequency and single time window. The array comprises nine sensors equally spaced on the perimeter of the circle and one sensor in the middle. The FDBF method is utilized to estimate $\vec{k}$ and the NFDBFa method is utilized to estimate $\vec{\alpha}$. In Figures 3-3a and 3-3b, the results of applying the FDBF technique to the original noise wavefield recorded by the array are

depicted. Figure 3-3a illustrates the f-$\vec{k}$ spectrum at the considered frequency in a 2D wave number space (k_x-k_y). Stronger powers are represented by a darker purple color. This spectrum offers insights into the power and vector velocities of propagating waves. In this example, a wave propagates along the x-axis with a velocity represented by a vector wave number $\vec{k}$ at the chosen frequency. Consequently, a spectrum peak emerges on the positive k_x axis at a distance of $|\vec{k}|$ from the origin. The associated phase velocity can be calculated as $V_r = 2\pi f/|\vec{k}|$, and the wavelength, λ, can be determined as $\lambda = 2\pi/|\vec{k}|$. Figure 3-3b illustrates the cross-section a-a from Figure 3-3a, revealing the main and side lobes. Generally, the narrower the main lobe and the shorter the side lobes the better the array and processing algorithm are at accurately identifying the correct $\vec{k}$ values for a given frequency.

Figures 3-3c and 3-3d display the f-$\vec{\alpha}$ spectrum obtained from applying the NFDBFa method to the converted noise wavefield for the same time window used to develop Figure 3-3a. In this case, instead of presenting the beamforming peak powers in the k_x-k_y space, as seen in Figure 3-3a, they are now depicted in the α_x-α_y space. This transition occurs because the phase in the pseudo wavefield is modulated by α (refer to Figure 3-2), rather than *k*. Figure 3-3c employs a different color scheme, where stronger powers are represented by darker blue colors. The f-$\vec{\alpha}$ spectrum shown in Figure 3-3c illustrates wave propagation for a single frequency along the x-axis with a phase attenuation represented by the vector $\vec{\alpha}$. Figure 3-3d illustrates the cross-section x-x from Figure 3-3c, revealing the main and side lobes along the positive x-axis (i.e., direction of wave propagation). Similar to estimating $\vec{k}$, the narrower the main lobe and the shorter the side lobes the better the array and processing algorithm are at accurately identifying the

correct $\vec{\alpha}$ values for a given frequency. The ability of the NFDBFa approach to develop phase attenuation estimates from ambient-noise recorded using MAM arrays is investigated in the following section using synthetic data.

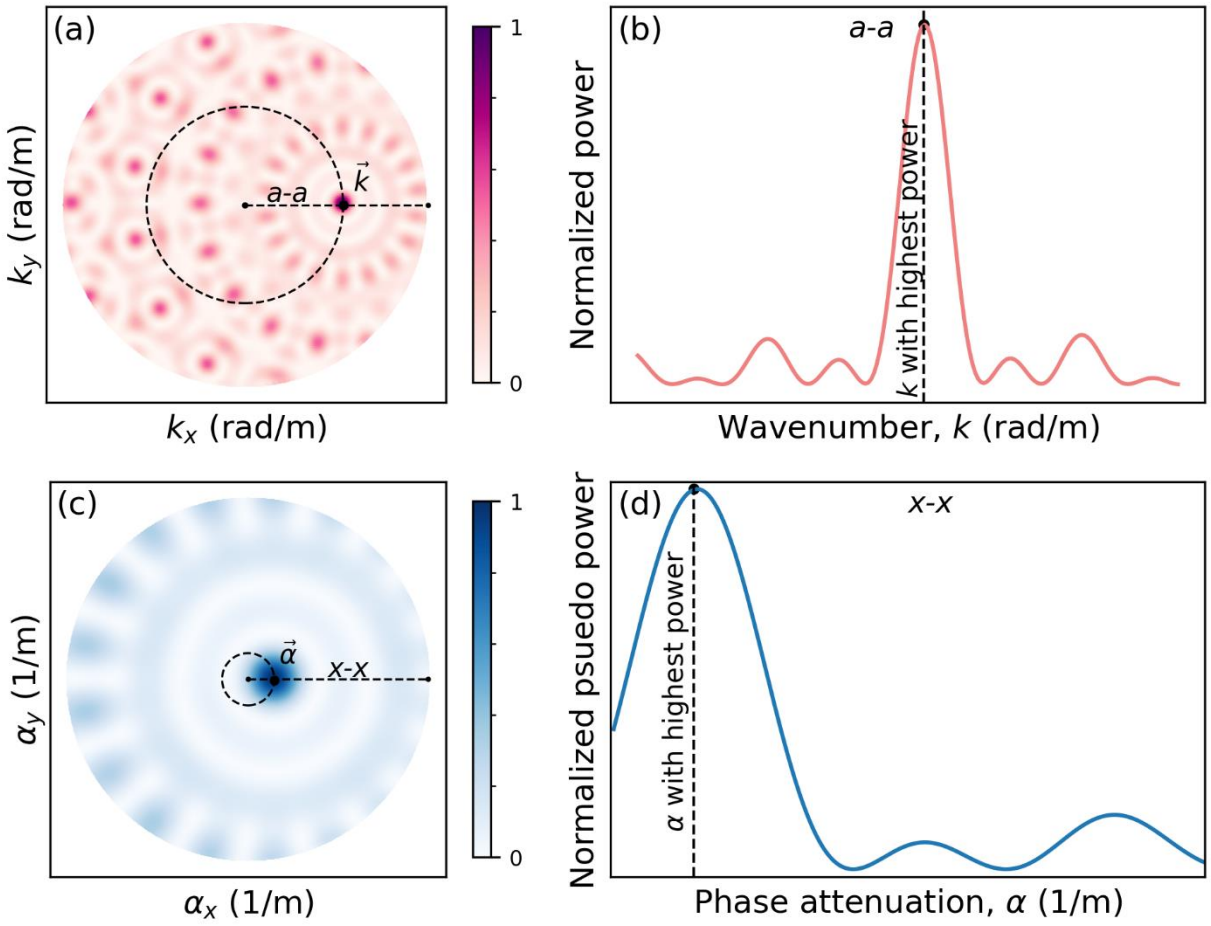

Figure 3-3 Schematic illustrating the FDBF and NFDBFa responses obtained from an ambient noise wavefield recorded by a ten-receiver circular MAM array for a single frequency and single time window. Panel (a) presents the f-$\vec{k}$ spectrum resulting from applying the FDBF method to the original wavefield, displaying the beamforming peak powers in k_x-k_y space. Panel (b) shows the cross-section a-a from Figure 3-3a, revealing the main and side lobes. Panel (c) presents the f-$\vec{\alpha}$ spectrum resulting from applying the NFDBFa technique to the pseudo wavefield, presenting the beamforming peak powers in α_x-α_y space. Panel (d) illustrates the cross-section x-x from Figure 3-3c, showing the main and side lobes along the direction of wave propagation.

3.5 NFDBFa evaluation with synthetic wavefields

This section uses synthetic data to validate the effectiveness of the NFDBFa approach in estimating phase attenuation from ambient noise recorded using MAM arrays. Specifically, the approach is tested on two soil models: a half-space model and a single layer above a half-space model. All numerical simulations discussed in this section were executed using Salvus (Afanasiev et al., 2019), a comprehensive 2D and 3D full-waveform modeling software suite based on the spectral element method. The simulations were performed on the Texas Advanced Computing Center's (TACCs) high-performance cluster Lonestar6 using two compute nodes.

3.5.1 Half-space model

This subsection presents a simple wave propagation simulation consisting of a single surface source generating body and surface waves propagating through a half-space soil model. Despite the simplicity of the model, the outcomes obtained from this simulation offer key insights into the attenuation of a wavefield generated by a surface source and elucidate the capabilities of the NFDBFa approach. Figure 3-4 depicts a schematic plan view illustrating the source location and MAM array configurations employed in the half-space simulation. The wavefield was generated by a point source acting in the vertical direction at coordinates (0, 0, 0) in an x, y, z cartesian coordinate system. The source was a single Ricker wavelet with a center frequency of 5 Hz. This source function produces broadband energy over a frequency range of approximately 1 to 10 Hz. The wavefield emanating from the source was recorded using five circular MAM arrays, each comprising 10 sensors, with one sensor at the center and nine sensors evenly spaced around the perimeter. In this paper, the arrays are named using the convention 'C' followed by the

diameter of the array, where 'C' denotes that the array is circular. Therefore, the first array, located two kilometers away from the source and with a diameter of one kilometer, is denoted as C1000 at 2 km. The remaining four arrays, concentrically-centered five kilometers from the source, have diameters of 60 m (C60), 300 m (C300), 1000 m (C1000 at 5 km), and 2000 m (C2000). It is noteworthy that, although currently only the vertical component of the displacement wavefield is utilized in NFDBFa, each sensor recorded both horizontal and vertical displacement components, and plans for utilizing all components from noise recordings are ongoing. Additionally, the NFDBFa processing operated independently of any knowledge about the source location, mirroring the conditions of an ambient noise MAM survey, ensuring an unbiased analysis.

The half-space constitutive soil parameters are presented in Figure 3-5a, where V_p, and υ are the compression wave velocity, and Poisson's ratio, respectively. Due to the large spatial extent of the model and the substantial computational expense associated with running a simulation over such a vast domain, a 2D simulation was conducted rather than a 3D simulation. In the 2D simulations, the sensor locations were projected onto a 2D plane, as illustrated in Figure 3-5a. This entailed setting the y-coordinate to zero for each surface sensor location shown in Figure 3-4, resulting in their positions being determined exclusively by their x-axis coordinates. For example, the sensor initially situated at coordinates (2321.4, 383, 0) in an x, y, z system (as depicted in Figure 3-4), transformed to (2321.4, 0) in the 2D x, z system presented in Figure 3-5. However, it is important to note that, during NFDBFa processing, the coordinates assigned to each sensor were derived from those shown in Figure 3-4; consequently, the aforementioned sensor retained coordinates of (2321.4, 383, 0) during processing. This approach not only substantially

reduced the computational cost of the simulations, but also ensured that the arrays were measuring plane waves. The simulation required 4 hours and 20 minutes of computation utilizing 256 threads on the high-performance cluster Lonestar6.

Before describing the application of the NFDBFa method, some preliminary features of the amplitude decay versus distance are discussed, as they directly influence attenuation estimates. To better observe this decay pattern, the wavefield emanating from the source was recorded every ten meters along the surface. Those time histories were then filtered at discrete frequencies so the amplitude decay at each frequency could be observed. The decay of Fourier amplitudes with distance from the vertical Ricker wavelet source for frequencies 1, 2, 3, 4, and 5 Hz are shown in Figures 3-5b and 3-5c. In Figure 3-5b, the amplitudes for each frequency are normalized by their respective maximum values at the source and plotted on a log scale, while the distances are not normalized and plotted on a linear scale. In contrast, in Figure 3-5c, the distances from the source are normalized by the wavelength (λ) corresponding to each plane wave frequency and plotted on a linear scale. The figures depict a sharp amplitude decrease near the source due to nearfield effects. Following this, amplitude oscillations with diminishing power are superimposed over a linear decay trend. Note that a linear decay trend in log amplitude scale corresponds to an exponential decay in linear amplitude scale. These amplitude oscillations tend to flatten greatly after propagating approximately 10 λ away from the source. It is noteworthy that these oscillations, although verified using other software packages, such as the ElastoDynamics Toolbox (EDT; Schevenels et al., 2009), challenge conventional intuition regarding wave attenuation in a half-space. Neither the geometric spreading of Rayleigh waves nor the attenuation due to material damping should exhibit such oscillations in a

half-space, as detailed by Lai (1998). The oscillating amplitude decay pattern in a half-space model is a result of body wave amplitude decay oscillations, as shown by Tokimatsu (1995) and Holzlohner (1980). Hence, when estimating phase attenuation using ambient noise, it is essential for the MAM arrays to be at a sufficient distance (more than approximately 10 λ) away from any potential surface sources, such that wave amplitude oscillations do not contaminate the expected trend of amplitude decay with distance.

It is worth noting that in layered media, oscillating amplitude decay of Rayleigh waves due to geometric spreading has been reported and accounted for in attenuation studies, as observed in the work of Lai (1998). Thus, in layered media, wave amplitude oscillations can be more pronounced and may extend beyond 10 λ from the surface source, as demonstrated by Tokimatsu (1995). This may be thought of as a type of near-field effect specific to attenuation studies, wherein the wavefield amplitude decay patterns are significantly more complicated at distances less than approximately 10 λ from source. This is distinct from, and more severe than, the typical range of near-field effects for phase velocity estimations, which generally deteriorate between 0.5 λ and 2 λ from the source, depending on the subsurface velocity structure (Tokimatsu 1995; Rix et al., 2001).

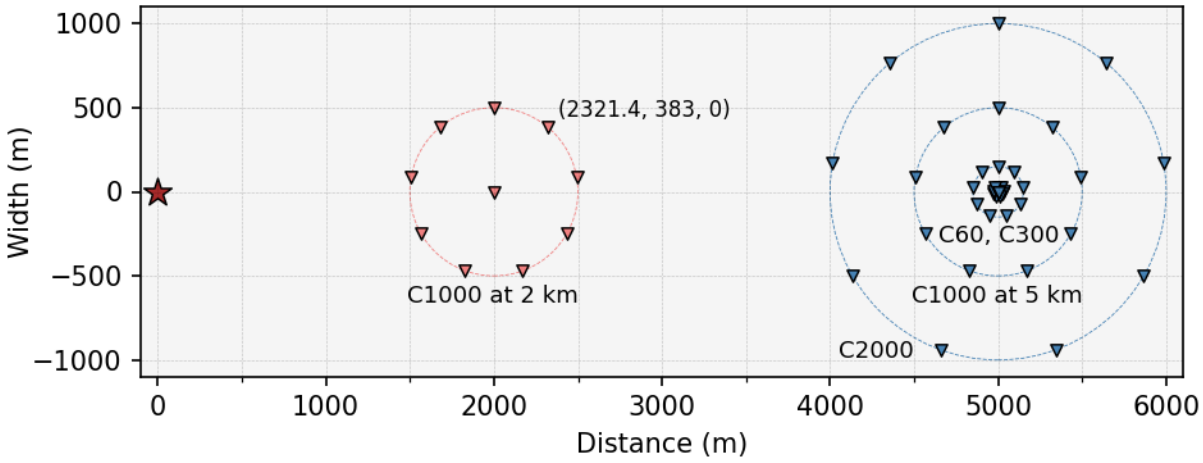

Figure 3-4 Plan view of the source (star symbol) and receiver (inverted triangle symbols) configurations used for synthetic wavefield simulations. The source was a single Ricker wavelet with a center frequency of 5 Hz. The wavefield was recorded using five MAM arrays. The first array (C1000 at 2 km) has a diameter of 1 km and is positioned 2 km from the source. The remaining four arrays are concentrically-centered 5 km away from the source and have diameters of 60 m (C60), 300 m (C300), 1 km (C1000 at 5 km), and 2 km (C2000), respectively.

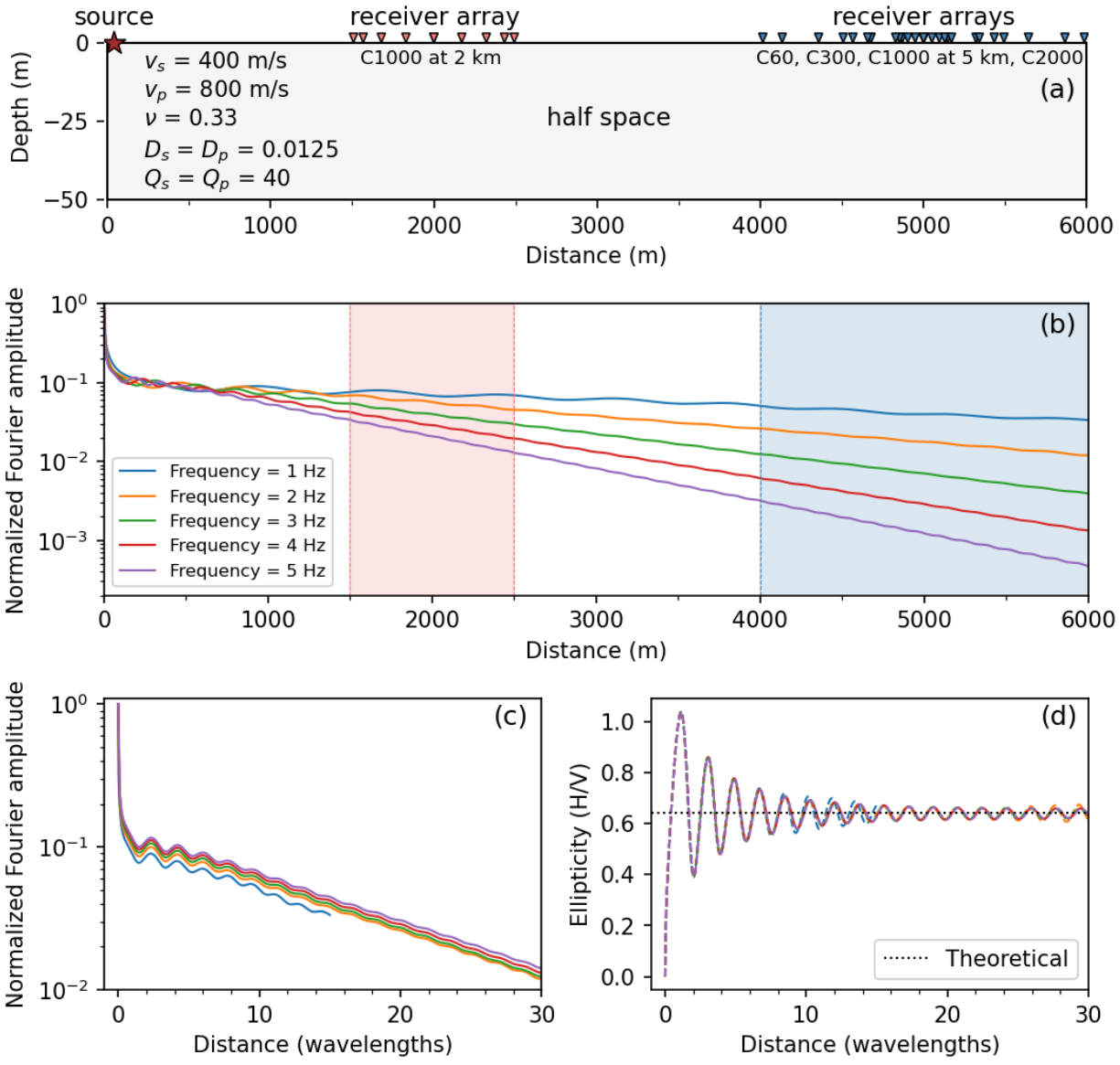

Figure 3-5 Half-space wavefield simulation: Panel (a) presents a cross-section view of the configuration of the source and receivers shown in Figure 3-4, along with the half-space soil properties. Panel (b) shows the decay of particle vertical displacement as a function of distance from the source for five distinct frequencies, each normalized by its maximum amplitude at the source. Panel (c) presents the particle displacement decay patterns from Panel b, with distance now normalized by the wavelength for each frequency. Panel (d)

shows the particle ellipticities for each frequency, expressed as the horizontal particle displacement divided by the vertical particle displacement, with the dotted horizontal line indicating the theoretical ellipticity calculated based on the Poisson's ratio of the half-space soil model.

To further demonstrate the more severe near-field effects associated with amplitude decay, Figure 3-5d presents the simulated wavefield ellipticity, expressed through the horizontal-to-vertical (H/V) ratio of particle displacement, measured with distance in wavelengths for the same frequencies outlined in Figure 3-5b. The ellipticity also displays oscillations that decrease and stabilize at normalized distances greater than about 10 λ from the source. This observation underscores that the near-field amplitude decay oscillations stem from body waves, as Rayleigh wave ellipticity in a half-space is determined solely by Poisson's ratio (Tokimatsu 1995) and should not oscillate. In Figure 3-5d, we observe that the calculated ellipticities oscillate around the theoretical value anticipated for Rayleigh wave ellipticity in a half-space with Poisson's ratio equal to 0.33, depicted by the dotted horizontal line in Figure 3-5d.

The synthetic time histories recorded by the C1000 at 2 km and the C1000 at 5 km MAM arrays (refer to Figures 3-4 and 3-5) were processed using the FDBF and NFDBFa methods to estimate phase velocity and attenuation, respectively, as illustrated in Figure 3-6. Figure 3-6 aims to highlight the impact of wave amplitude decay patterns on the attenuation estimates. In terms of abilities to resolve phase velocity, both the C1000 arrays seem to perform approximately the same, whether 2 km away from the source (Figure 3-6a) or 5 km away from the source (Figure 3-6b). However, upon inspecting Figures 3-6c and 3-6d, it becomes evident that the array located 5 km from the source (i.e., Figure 3-6d) provides more reliable attenuation estimates at lower frequencies compared to the array closer to the source. This observation can be explained by referring to Figure 3-5b, where

the amplitude decay patterns measured by the array positioned 2 km from the source are shaded in pink. It is apparent that in close proximity to the source, the low-frequency waves have not traveled a sufficient number of wavelengths, resulting in amplitude decay that does not conform to pure exponentials (i.e., linear decay in log scale). However, by the time these waves reach the array positioned 5 km from the source (blue shading in Figure 3-5b), the oscillations in amplitude decay have diminished significantly, approaching a pure exponential decay. Therefore, it is noteworthy that in an ambient-noise survey, even though the source location is unknown, if the noise source is close to the array in terms of wavelengths traveled by the desired frequency, it may lead to unreliable and scattered attenuation results. Nonetheless, Figures 3-6c and 6d clearly demonstrate the reliability of the new NFDBFa approach in retrieving phase attenuation estimates over a broad range of frequencies.

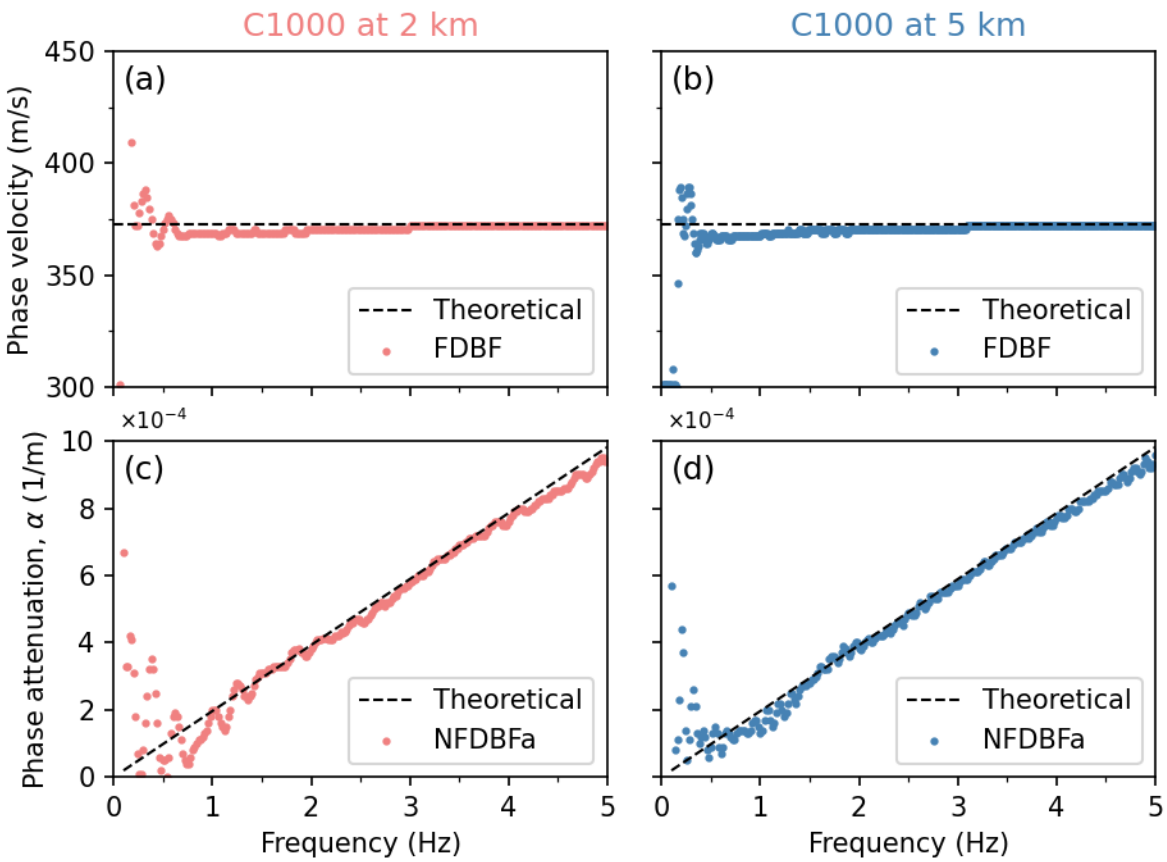

Figure 3-6 Half-space wavefield simulation: phase velocity (top) and phase attenuation (bottom) dispersion data estimated with FDBF and NFDBFa, respectively, from 1 km arrays positioned at two distinct distances from the ambient noise source: (left) at two kilometers (C1000 at 2 km), and (right) at five kilometers (C1000 at 5 km).

Finally, the performance of the NFDBFa in the presence of incoherent noise is investigated. For this purpose, Figure 3-7 illustrates the influence of incoherent noise and array size on phase attenuation estimates using the same half-space simulation results. The analysis focuses on the four arrays of different sizes concentrically-centered 5 km from the source (refer to Figures 3-4 and 3-5a). Incoherent noise was introduced to the signal, with a target signal-to-noise ratio (SNR) at 20 dB, which resulted in the frequency-dependent amplitude decay patterns depicted in Figure 3-7a (compared to Figure 3-5b). Figures 3-7b, 3-7c, 3-7d, and 3-7e display the attenuation estimates obtained using the C60, C300,

C1000, and C2000 MAM arrays, respectively. It becomes evident that larger arrays yield more accurate attenuation estimates in the presence of incoherent noise. Figure 3-7a elucidates the rationale behind this enhanced performance for larger arrays across all frequencies. The C2000 MAM array samples a significantly larger area, enabling it to discern the exponential amplitude decay even in the presence of noise. The C60 MAM array samples a significantly smaller area, and thus is considerably more sensitive to amplitude fluctuations caused by incoherent noise, resulting in the significant scatter observed in the attenuation estimates shown in Figure 3-7b.

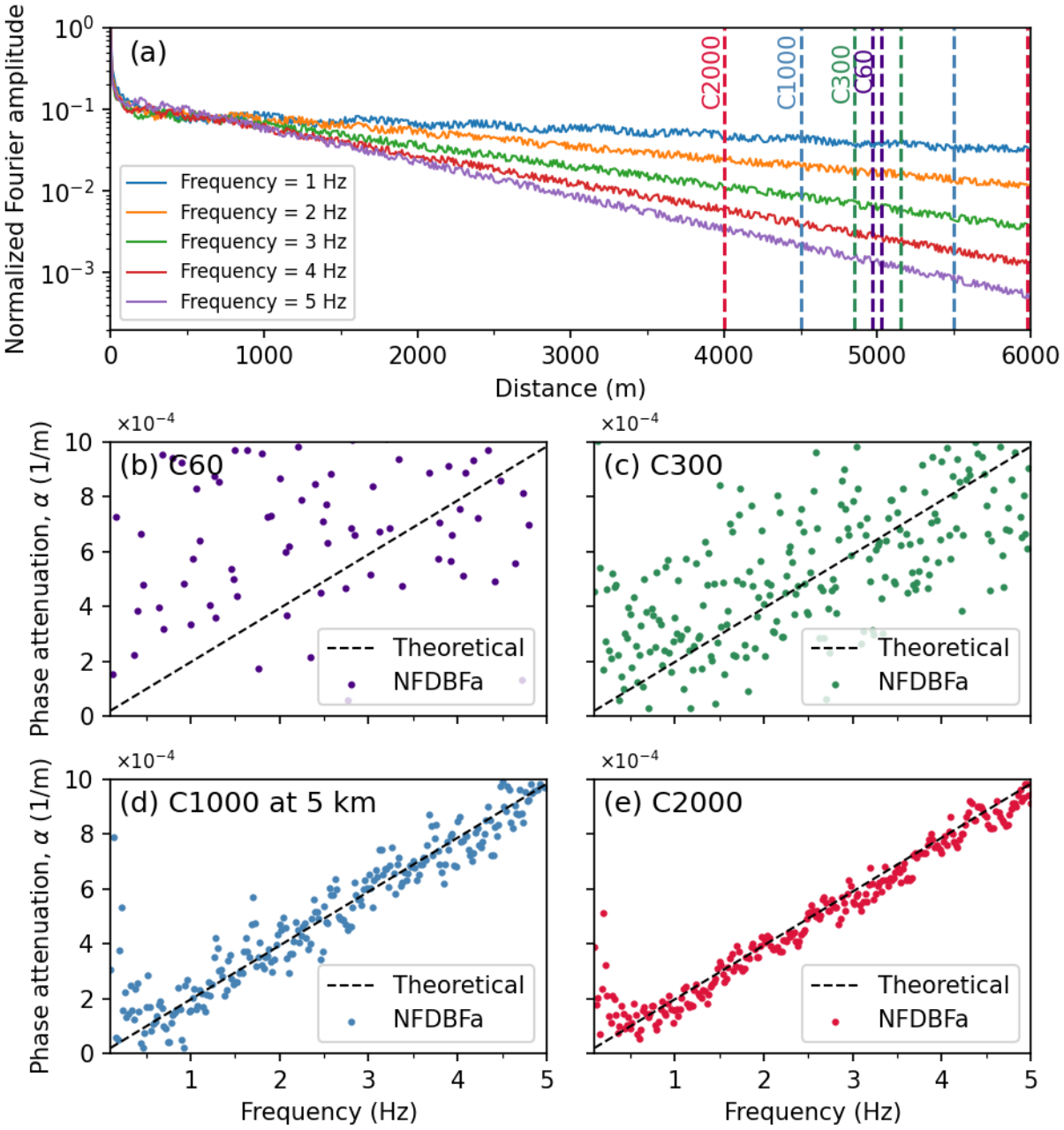

Figure 3-7 Half-space wavefield simulation with noise: Panel (a) shows the amplitude decay of the same five frequencies depicted in Figure 3-5 but now with added incoherent noise to the signal, setting the signal-to-noise ratio (SNR) at 20 dB. Panels (b) to (e) present the predicted phase attenuation data from the NFDBFa analysis for four arrays concentrically-centered at five kilometers from the source, with diameters of 60 m (C60), 300 m (C300), 1 km (C1000 at 5 km), and 2 km (C2000), respectively.

Figure 3-8 further illustrates the impact of array size on resolving attenuation coefficients by showcasing the f-$\vec{\alpha}$ spectra for a frequency of 3 Hz that were calculated from the wavefield recorded by the four concentrically-centered arrays located 5 km from the source. Notably, the mainlobe (dark blue shaded area) is considerably narrower for larger arrays, resulting in more reliable estimates of phase attenuation. Two key points warrant attention here. First, the feasibility of employing larger arrays might be restricted due to limitations in access at a given site, or to help maintain approximately a one-dimensional (1D) subsurface condition beneath the array, which is an implicit assumption in the analysis technique (i.e., no lateral spatial variability). Meeting this assumption becomes more challenging as the array size expands. Second, it is essential to highlight that the method used to determine the optimal MAM array size for attenuation estimates differs from the one employed in obtaining phase velocity estimates. In dispersion estimation, smaller MAM arrays are more effective at capturing high frequency phase velocities, whereas larger arrays are better suited for resolving lower frequency phase velocities (Foti et al., 2018; Vantassel and Cox, 2022). However, according to the results depicted in Figure 3-7, the larger arrays demonstrated superior ability in resolving phase attenuation across the entire considered frequency range compared to the smaller arrays.

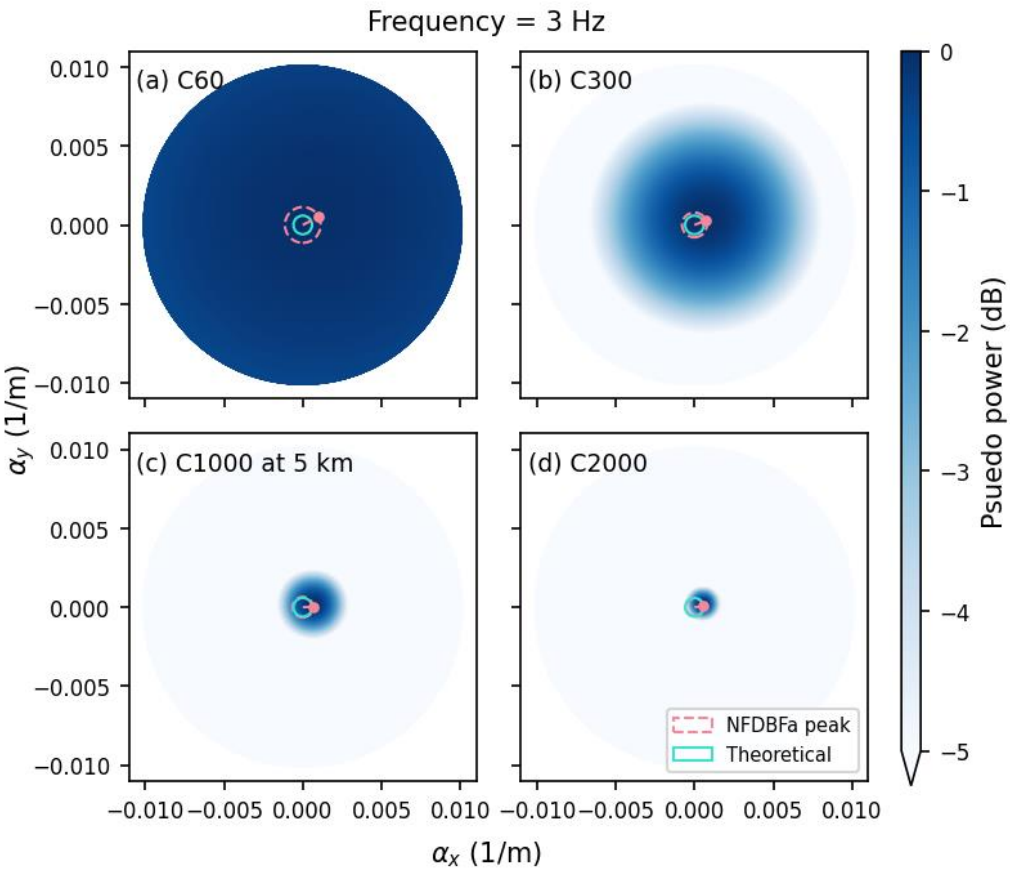

Figure 3-8 Half-space wavefield simulation with noise: Panels (a) through (d) present the f-$\vec{\alpha}$ spectra obtained through NFDBFa analysis for a frequency of 3 Hz. The spectra are derived from the wavefield recorded by the four arrays concentrically-centered five kilometers from the source with diameters of 60 m (C60), 300 m (C300), 1 km (C1000), and 2 km (C2000), respectively, as depicted in Figure 3-7.

3.5.2 Layer above a half-space model

The performance of the NFDBFa approach on a synthetic model consisting of a single layer above a half-space is illustrated in this subsection. The model's constitutive small-strain parameters and the source and receiver configurations are provided in Figure 3-9a. For this synthetic study, 150 vertical point sources with varying forcing functions and trigger times were activated. The sources were triggered five kilometers away from the center of a one-kilometer diameter circular array consisting of 10 sensors: one in the center and nine equally distributed around its perimeter (just like the C1000 at 5 km MAM array

depicted in Figure 3-4). The waveforms recorded by the array are depicted in Figure 3-9b. These waveforms were subsequently processed using FDBF and NFDBFa to derive the Rayleigh wave phase velocity dispersion data shown in Figure 3-9c and the phase-attenuation data shown in Figure 3-9d, respectively. The theoretical Rayleigh-wave phase velocity dispersion and attenuation curves for the model are also presented in Figures 3-9c and 9d, respectively. In these figures, the fundamental theoretical mode is denoted as Mode 1, while the 1st-higher mode is denoted as Mode 2.

The FDBF method is able to extract experimental phase velocity dispersion data from the synthetic wavefield that well-matches the theoretical dispersion curves and captures the transition from Mode 1 to Mode 2 at approximately 7 Hz. A strong agreement is also observed between the theoretical attenuation curves and the experimental attenuation data extracted from the synthetic wavefield using the NFDBFa method, particularly for Mode 1. Interestingly, the attenuation data shifts to Mode 2 at the same frequency where the phase velocity dispersion data transitions to Mode 2. A similar observation about possible links between the frequencies where phase velocity and attenuation mode transitions occur was also reported by Aimar et al., (2024a) using MASW data. While further studies are needed to validate the observations that phase velocity and phase attenuation data tend to jump modes at identical frequencies, this is a potentially important point, as patterns in attenuation modes are much more complex than phase velocity modes.

The effectiveness of the proposed NFDFBa approach has been successfully demonstrated through the analyses conducted on synthetic datasets, as discussed above.

Now, we shift our focus to applying this approach to real field data, offering a thorough demonstration of its effectiveness in a practical, real-world situation.

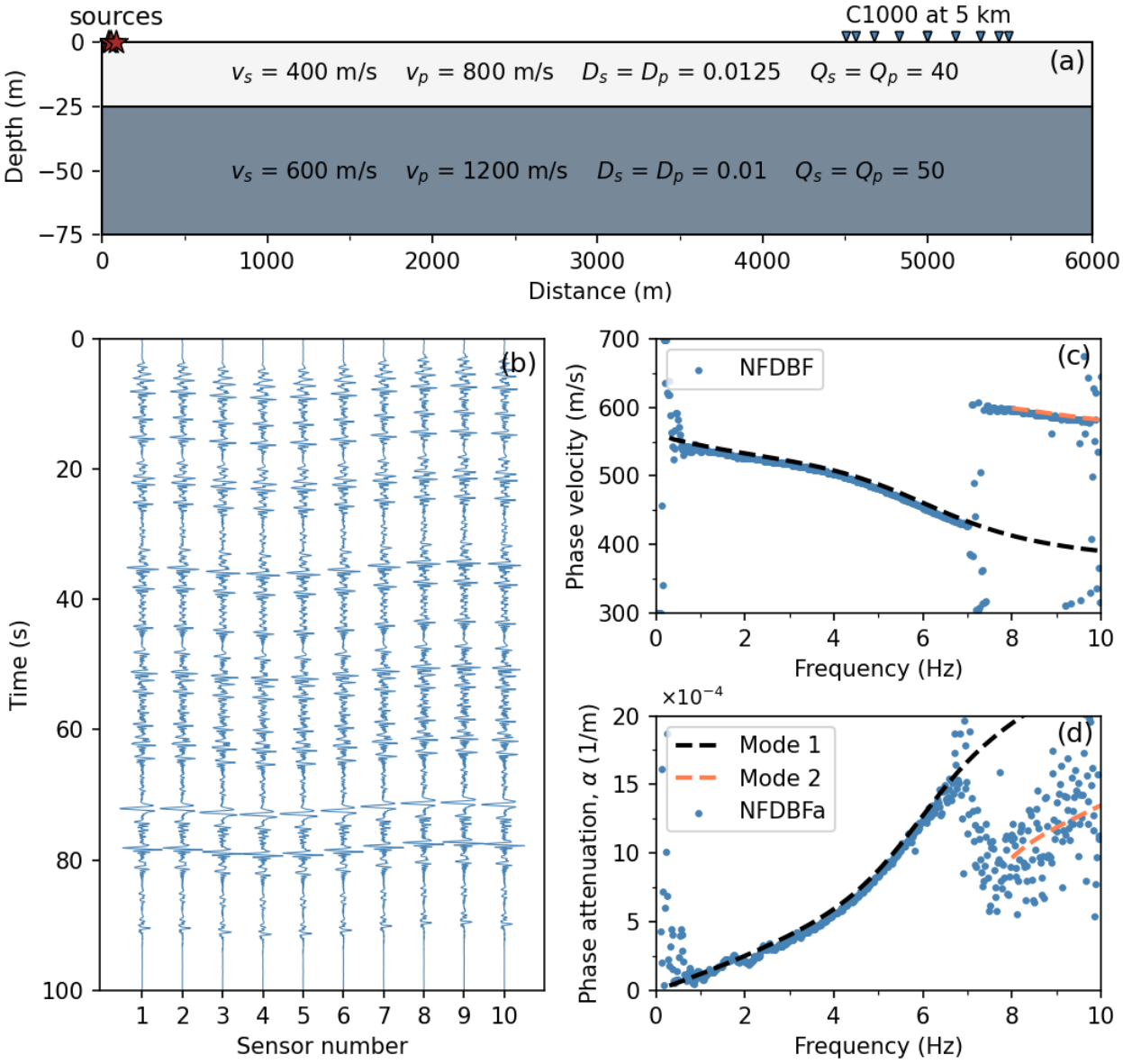

Figure 3-9 Layered model simulation: Panel (a) presents the soil properties utilized in the simulation for the soil layer and the half-space, along with the surface sources and 1-km receiver array located 5 km from the source (C1000 at 5 km). Panel (b) displays the waveforms collected from the C1000 array. In Panel (c), the good agreement between the theoretical Rayleigh-wave phase velocity curves (Mode 1 and Mode 2) and the experimental phase velocity data obtained through the FDBF approach on the original wavefield is demonstrated. Finally, Panel (d) showcases the good agreement between the theoretical phase attenuation curves and the experimental phase attenuation data extracted from the converted wavefield using the proposed NFDBFa approach.

3.6 Field application and validation

A surface wave field-testing campaign was conducted at the Drainage Farm Site in Logan, Utah, USA (refer to Figure 3-10), a property owned by Utah State University (USU). Structural geology indicates that Southern Cache Valley, encompassing the Drainage Farm Site and located in the northeastern part of the Basin and Range province, is a graben bounded by high-angle normal faults (Williams, 1962). The site is underlain by Paleozoic rocks, which are overlain by Tertiary formations such as the Wasatch and Salt Lake formations, composed of conglomerate, siltstone, and tuffaceous sandstone. In certain areas of Cache Valley, these formations reach thicknesses of up to 2,440 m (Evans et al., 1996). The near-surface geology of the Drainage Farm Site is characterized by sediments from ancient Lake Bonneville, which receded to form the Provo shoreline. These sediments include alluvial, lacustrine, and deltaic deposits (Williams, 1962; Evans et al., 1996). Well logs presented by Williams (1962) reveal alternating layers of silt and clay, sand, and gravel above the Salt Lake formation. Moreover, limited deep well logs from the vicinity of the Drainage Farm Site indicate that rock can be encountered at depths ranging from 176 m to more than 350 m (Perez, 1969).

The goal of the testing was to collect a high-quality surface wave dataset that could be used for attenuation studies to validate the proposed NFDBFa technique. The field testing involved both active-source MASW testing and ambient noise MAM testing. The sensor array configurations utilized for MASW and MAM at the Drainage Farm Site are illustrated in Figure 3-10. MASW testing was performed using 24, 4.5-Hz vertical geophones placed with a spacing of two meters between successive geophones, resulting

in an array length of 46 m. Wavefields with strong Rayleigh wave content were actively generated by striking vertically on a strike-plate with a sledgehammer. The sledgehammer was used at eight distinct "shot" locations that were offset by 5, 10, 15, and 20 m relative to the first/last geophone off each end of the array. Five distinct sledgehammer blows were recorded at each location for subsequent stacking to increase the signal-to-noise ratio (Foti et al. 2018). MAM testing utilized three concentric circular arrays that were aligned with the middle of the MASW array, as depicted in Figure 3-10. The three arrays were 700-m, 300-m, and 60-m in diameter, and will be referred to as C700, C300, and C60, respectively. Each array consisted of nine evenly distributed three-component broadband seismometers (Nanometrics Inc. Trillium Compact 120s seismometers) along its circumference to capture ambient vibrations. The three arrays did not record data simultaneously; instead, the nine sensors were used to collect noise data for each of the MAM arrays one array at a time. First, the sensors recorded seismic noise for 13 hours and 30 minutes for the C700 array. Subsequently, the sensors were relocated to their designated locations for the C60 and C300 arrays, recording ambient noise for an hour and a half and three hours, respectively.

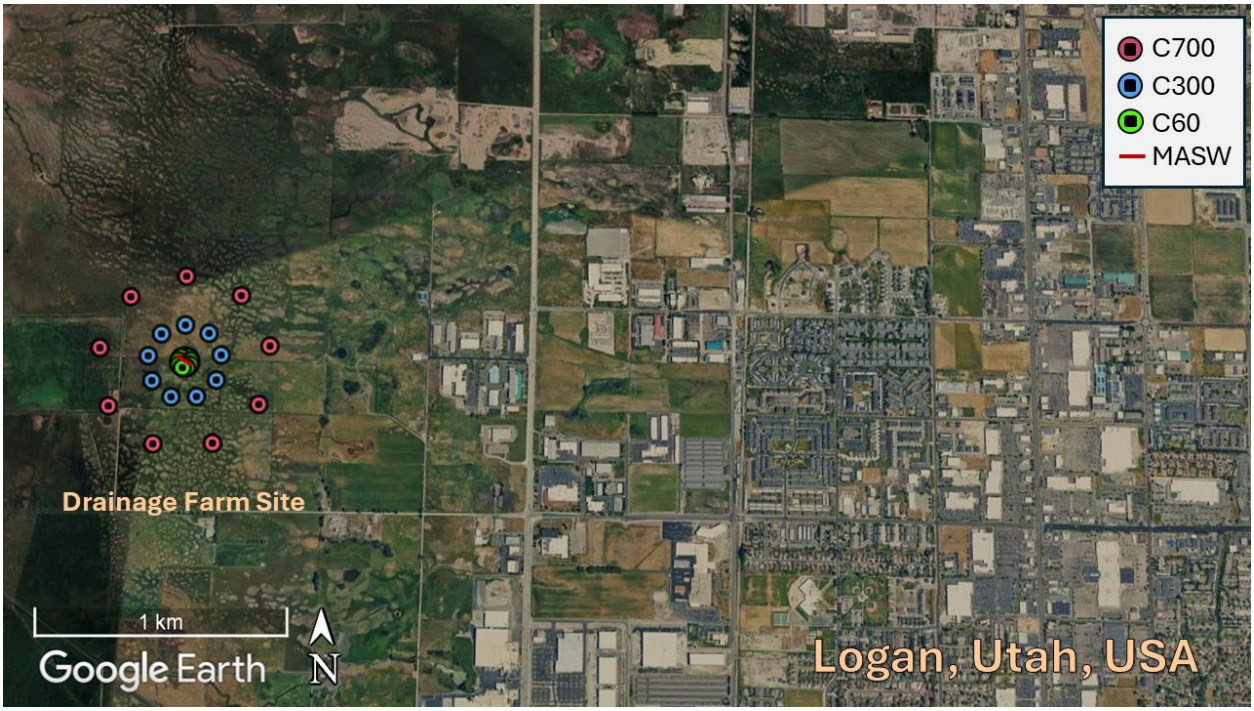

Figure 3-10 Plan view of the MASW and MAM arrays employed for testing at the Drainage Farm Site in Logan, Utah, USA. The concentric MAM arrays featured diameters of 60 m (C60), 300 m (C300), and 700 m (C700), while the MASW array comprised 24, 4.5-Hz vertical geophones, spanning 46 m.

For Rayleigh-wave phase velocity dispersion analysis, MASW data were analyzed using the FDBF method with cylindrical-wave steering (Zywicki and Rix, 2005), as coded in the open-source surface wave processing package swprocess (Vantassel, 2021). This processing was coupled with the multiple source-offset technique for identifying near-field contamination and quantifying dispersion uncertainty (Cox and Wood, 2011; Vantassel and Cox, 2022). As a result, eight phase velocity estimates were obtained for each frequency, corresponding to one phase velocity estimate from each of the eight shot locations. MASW Rayleigh wave dispersion data influenced by near-field effects or significant offline noise were trimmed before calculating phase velocity dispersion statistics.

The three-component beamforming approach (Wathelet et al., 2018) coded in the open-source software package Geopsy (Wathelet et al., 2020) was used to generate Rayleigh-wave phase velocity dispersion data for each of the MAM arrays. The recorded time for each array was discretized into blocks, with each block further divided into at least 30-time windows. The window lengths were selected to contain at least 30 cycles (periods) at the lowest processing frequency that could be extracted from each MAM array (Vantassel and Cox, 2022). For each MAM array, eight phase velocity estimates were extracted at each analyzed frequency using the three-component beamforming (Wathelet et al., 2018) approach to ensure consistency with the eight phase velocity estimates obtained from the MASW processing. Spurious dispersion data stemming from high-amplitude noise in the near-field (e.g., traffic noise close to the sensors) and incoherent noise were manually eliminated before calculating dispersion statistics. Ambient noise phase velocity dispersion data from all MAM arrays were combined with the active phase velocity dispersion data obtained from MASW processing, as shown in Figure 3-11a. The combined data, used to compute mean and ± one standard deviation dispersion estimates (Vantassel and Cox, 2022), are displayed in Figure 3-11b relative to the individual MASW and MAM dispersion data points for the Drainage Farm Site.

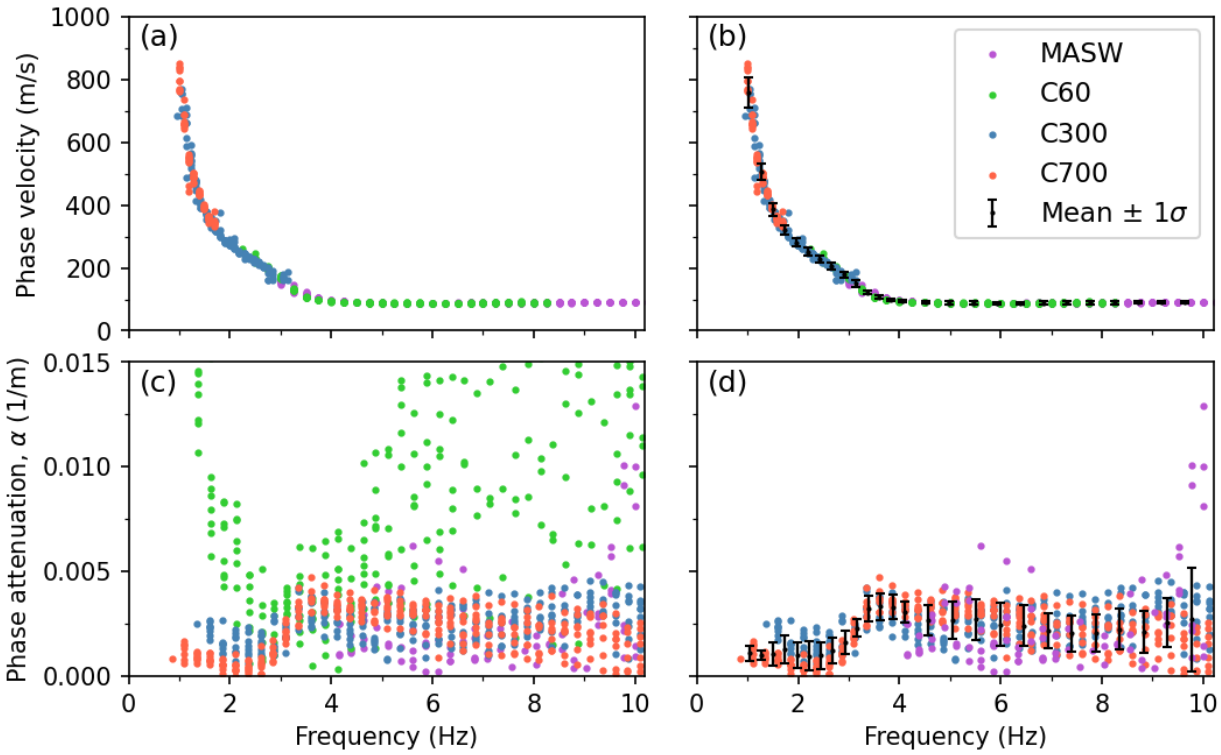

Figure 3-11 Experimental phase velocity and attenuation data extracted from MASW and MAM testing at the Drainage Farm Site in Logan, UT, USA. Panel (a) displays the experimental phase velocity dispersion data of Rayleigh waves processed from an MASW array and three circular MAM arrays, with diameters of 60 m, 300 m, and 700 m. Panel (b) showcases the mean and ± one standard deviation of the experimental Rayleigh wave phase velocity dispersion data derived from the combined MASW and MAM datasets. Panel (c) displays the experimental phase attenuation data from MASW and three circular MAM arrays. Panel (d) illustrates the mean ± one standard deviation of the experimental phase attenuation data calculated from the combined MASW, C300, and C700 MAM arrays.

The cylindrical FDBFa (CFDBFa) approach, as proposed by Aimar et al. (2024a), was employed to derive attenuation estimates from the MASW data. Mirroring the MASW phase velocity dispersion analysis, the multiple source-offset technique was utilized for quantifying attenuation uncertainty. Thus, eight attenuation estimates were extracted from the MASW data at each analyzed frequency using CFDBFa. For the MAM attenuation estimates, the new NFDBFa approach introduced in this study was employed. The recorded time for each array was discretized into eight blocks, with each block further divided into

30 windows. Consequently, the window length employed for each MAM array can be determined by dividing the total recording time of the array by the product of 8 blocks and 30 windows (i.e., 240). Similar to MAM phase velocity dispersion analysis, the window lengths were selected to contain at least 30 periods at the lowest processing frequency that could be extracted from each MAM array (Vantassel and Cox, 2022). Averaging the estimates from all windows within each block yielded a single data point per block, thus providing eight unique attenuation estimates per frequency. This processing approach ensured that an equal number of attenuation data points were obtained at each frequency for all of the MASW and MAM arrays. The combined ambient-noise attenuation data from all MAM arrays and the active attenuation data from the MASW array are plotted together in Figure 3-11c. A good agreement is observed between the attenuation estimates derived from the MASW array and those obtained from the C300 and C700 arrays for frequencies ranging from 4 to 10 Hz. The MASW testing did not generate coherent attenuation data at frequencies less than 4 Hz, due to the limitations of the active sledgehammer source. However, the MAM testing was able to extract coherent attenuation data at frequencies below 1 Hz. The agreement observed between the active-source and ambient noise attenuation estimates serves as compelling evidence for the efficacy of the proposed NFDBFa approach. However, it is notable that there is significant scatter in the attenuation estimates obtained using the C60 array. This variability is likely attributed to the challenges previously discussed in regards to using smaller MAM arrays for attenuation studies, as the phase velocity data extracted from the C60 array was very good (refer to Figure 3-11a). Hence, the attenuation data from the C60 array was removed prior to calculating attenuation statistics. The combined attenuation estimates from the MASW, C300, and

C700 arrays, and the mean and ± one standard deviation attenuation estimates obtained from those three arrays, are depicted in Figure 3-11d. While a noticeable agreement exists among the three arrays, there is significantly greater scatter in the attenuation estimates (Figure 3-11d) compared to the phase velocity dispersion estimates (Figure 3-11c). This observation aligns with the findings reported by Aimar (2022). The application of the new NFDBFa approach in this field test showcases its effectiveness in estimating attenuation coefficients from ambient noise wavefield data.

Finally, the statistical experimental Rayleigh-wave phase velocity and attenuation parameters derived from both the MASW and MAM testing (refer to Figures 3-11b and 3-11d) were used to invert for V_s and D_s profiles at the Drainage Farm Site. This was achieved through the Monte Carlo-based joint inversion of phase velocity and phase attenuation data developed by Aimar et al. (2024b). Although the effectiveness of the joint inversion procedure has been proven for active surface wave data (Lai and Rix, 1998; Aimar et al., 2024b), its application to combined dispersion data from MASW and MAM testing, covering a broad frequency range, is novel. This is because past studies on inverting MAM-based attenuation data to retrieve damping properties at large depths typically adopted an uncoupled inversion approach, based on a separate inversion of Rayleigh-wave phase velocity and α (e.g., Prieto et al., 2009; Parolai, 2014).

The inversions performed herein involved 50,000 five-layer trial Earth models with progressively increasing thicknesses, covering a comprehensive range of layer thicknesses, V_s, and D_s values. The layering was informed by a preliminary inversion study based solely on phase velocity dispersion data, which is omitted here for simplicity. Realistic values were fixed for the Poisson's ratio and mass densities. A constant D_p/D_s ratio of 1.4 was

used during the inversion, similar to the approach taken by Bergamo et al. (2023). Forward dispersion and attenuation modeling were conducted using the Computer Programs for Seismology software (Herrmann, 2013). The fit to the experimental data was quantitatively assessed using a normalized root mean square (RMS) error that accounts for estimation uncertainty, similar to the metric proposed by Wathelet et al. (2004).

The ten best inversion results are shown in Figure 3-12. The theoretical phase velocity and attenuation curves are shown relative to the experimental data in Figures 3-12a and 3-12b, respectively. The V_s and D_s profiles are shown in Figures 3-12c and 3-12d, respectively, down to a depth of 400 m, which is approximately 1/2 of the maximum resolved phase velocity wavelength. The V_s profiles in Figure 3-12c collectively feature a shallow layer about 25 m thick with velocities ranging from approximately 90 to 185 m/s, including a low-velocity zone, which is consistent with known near-surface layering. Below this, there is generally a thicker layer extending down to approximately 180 m with velocities varying around 500 m/s. At depths of 150-200 m, a stiff layer with velocities around 1500 m/s is commonly identified across the profiles. These depths, while variable, are consistent with the location of Salt Lake Formation rock surface, as discussed above. The D_s profiles in Figure 3-12d indicate that damping in the top 25 m is less than approximately 1%. Below this depth, there is a noticeable variability in the estimated D_s values between the ten best profiles. Nonetheless, D_s can be observed to increase to approximately 2% to 4% in the deeper soil deposits, which consist of alternating clay, sand, and gravel layers. At the top of the Salt Lake Formation rock surface, D_s collectively decreases again to less than 2% for all of the ten best profiles. The large variability in D_s can be attributed to the complex geology of the site, the significant standard deviation in

the experimental attenuation data, and the moderately low sensitivity of theoretical attenuation curves to D_s at greater depths (e.g., Badsar et al., 2012; Aimar et al., 2024b).

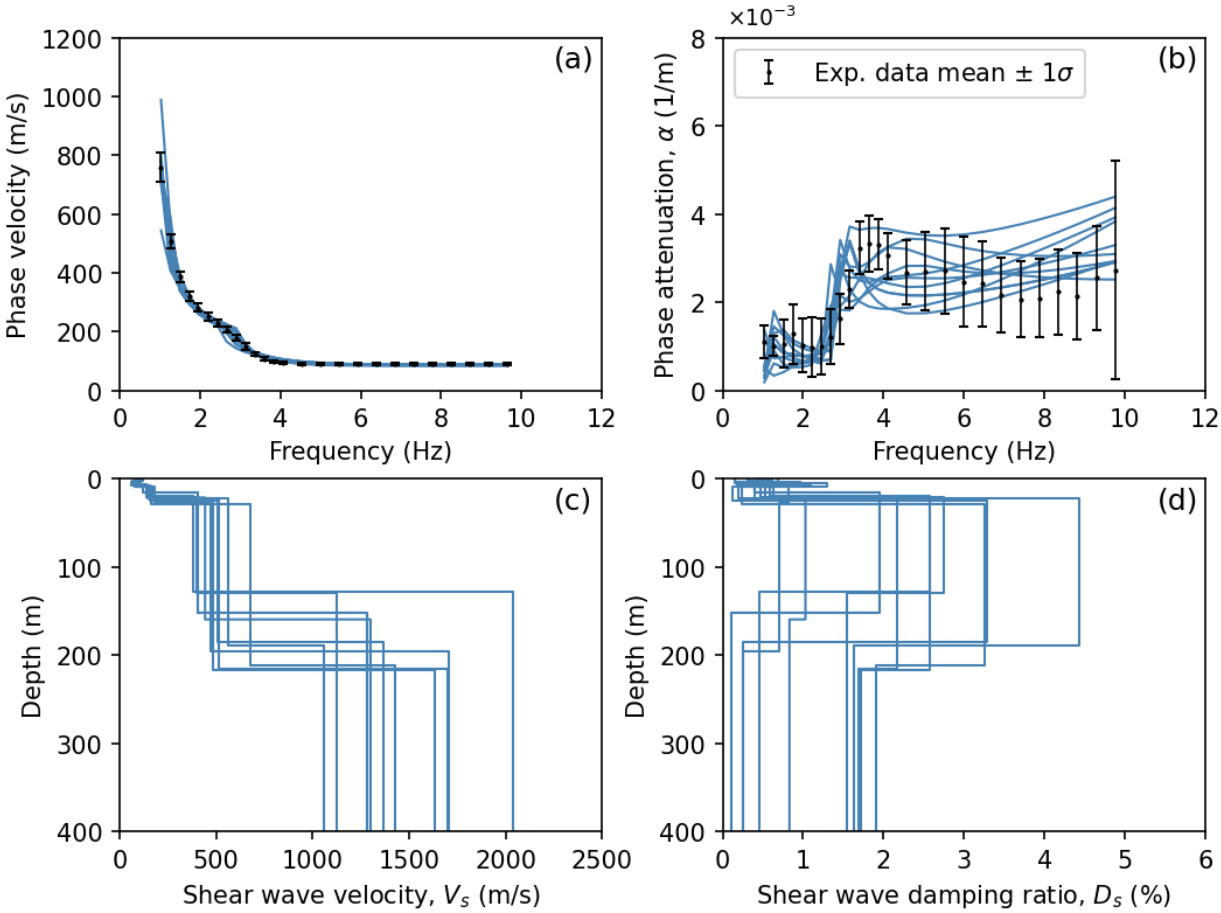

Figure 3-12 Inversion results for the experimental Rayleigh-wave phase velocity and attenuation data collected at the Drainage Farm Site in Logan, UT, USA. The figure highlights the ten best-fitting models, with Panels (a) and (b) comparing the theoretical curves for phase velocity and attenuation, respectively, against the experimental data represented by mean values with ± one standard deviation error bars. Panels (c) and (d) display the V_s and D_s profiles, respectively, for the ten best theoretical models.

Despite these limitations, the joint inversion procedure successfully provided in-situ estimates of D_s at depths not reached by conventional site characterization techniques. This confirms the advantages of combining MASW and MAM data for the combined estimation of stiffness and dissipation parameters of soil deposits.

3.7 Conclusions

A new methodology for estimating frequency-dependent attenuation coefficients through the analysis of ambient noise wavefield data recorded by 2D arrays of surface seismic sensors has been presented. The approach relies on the application of an attenuation-specific wavefield conversion and frequency-domain beamforming (FDBF). It has been termed the noise FDBF attenuation (NFDBFa) method. Importantly, using an FDBF approach, as opposed to a noise cross-correlation approach, enables the direction of ambient noise propagation to be determined for each noise window and frequency, and does not require an equipartitioned ambient noise wavefield. Furthermore, using an FDBF approach enables the phase velocity and attenuation data generated from active-source testing like MASW to be combined with phase velocity and attenuation data generated from ambient noise testing like MAM in order to span a broader frequency range. This enables the joint inversion of phase velocity and attenuation to be performed as a means to extract shear wave velocity and small-strain damping ratio profiles to significantly greater depths than previously possible using only active-source data.

Numerical simulations were conducted to deepen our understanding of the proposed NFDBFa method. These simulations aimed to evaluate how the proximity of the MAM array to the noise source, the presence of incoherent noise, and the size of the array affect the estimates of phase attenuation. The results demonstrated that near-field effects are more pronounced and extend over greater distances for phase attenuation estimates in comparison to those considered for phase velocity estimation. Furthermore, it was discovered that larger array sizes consistently provided more accurate phase attenuation estimates across all considered frequencies, contrary to the conventional MAM design criteria used for phase velocity dispersion estimation, where larger arrays are typically

preferred for resolving lower frequencies while smaller arrays excel at resolving higher frequencies. This distinction emphasizes the need for unique design criteria when planning a MAM array for attenuation estimation.

The proposed NFDBFa approach underwent validation through numerical wave propagation simulations, comparing predicted frequency-dependent phase attenuation values against theoretical phase attenuation curves for two synthetic models. Furthermore, validation of the developed technique was reinforced using MASW and MAM field data collected at the Drainage Farm Site in Logan, Utah, USA. The phase velocity and attenuation data extracted from the MASW and MAM recordings agreed well over a common bandwidth, while the ambient noise MAM data allowed the phase velocity and attenuation estimates to be extracted at significantly lower frequencies. The joint inversion of the experimental Rayleigh-wave phase velocity and phase attenuation data obtained from both MASW and MAM testing facilitated the estimation of shear wave velocity and small-strain damping ratio profiles to significant depths (400 m) at the Drainage Farm Site.

As noted herein and in other studies like Aimar et al. (2024a), attenuation data are significantly more variable and more complex to understand (e.g., modal curves that repeatedly cross one another) than phase velocity data. As such, there is a need for future studies to better understand attenuation data and how to invert them to retrieve reliable in-situ profiles of the small-strain damping ratio. Future efforts should involve additional numerical and experimental testing of diverse subsurface conditions, coupled with comparisons to damping estimates obtained from invasive tests. With the validity of this approach demonstrated on the vertical component, future research will also explore the

utilization of the three components of the noise wavefield to enhance attenuation estimates beyond the current method's capabilities.

3.8 Acknowledgements

The numerical simulations were run on the Texas Advanced Computing Center's (TACC's) cluster Lonestar6, with an allocation provided by DesignSafe-CI (Rathje et al., 2017). This work was supported by the U.S. National Science Foundation (NSF) Grant Number CMMI-2120155. 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.

3.9 Research data and code availability

The field test data used to validate the NFDBFa approach presented in this paper are available in the dataset by Abbas et al. (2024).

3.10 References

Abbas A, Cox BR, Dawadi N, Jackson N, and Cannon K (2024) Geotechnical site characterization at the Drainage Farm Site. DesignSafe-CI. https://doi.org/10.17603/ds2-sx2h-8s20 v1

Aimar M, Foti S and Cox BR (2024a accepted) Novel Techniques for In-situ Estimation of Shear-wave Velocity and Damping Ratio through MASW testing Part I: A Beamforming Procedure for Extracting Rayleigh-wave Phase Velocity and Phase Attenuation. (accepted to Geophysical Journal International).

Aimar M, Foti S and Cox BR (2024b accepted) Novel Techniques for In-situ Estimation of Shear-wave Velocity and Damping Ratio through MASW testing Part II: A Monte Carlo Algorithm for the Joint Inversion of Phase Velocity and Attenuation. (accepted to Geophysical Journal International).

Afanasiev M, Boehm C, van Driel M, Krischer L, Rietmann M, May DA, Knepley MG and Fichtner A (2019) Modular and flexible spectral-element waveform modelling in two and three dimensions. Geophysical Journal International, 216(3), 1675–1692. https://doi.org/10.1093/gji/ggy469.

Albarello D and Baliva F (2009) In-Situ Estimates of Material Damping from Environmental Noise Measurements. In Increasing Seismic Safety by Combining Engineering Technologies and Seismological Data (pp. 73–84). https://doi.org/10.1007/978-1-4020-9196-4_6

Anderson JG, Lee Y, Zeng Y and Day S (1996) Control of strong motion by the upper 30 meters. Bulletin of the Seismological Society of America, 86(6), 1749–1759. https://doi.org/10.1785/BSSA0860061749.

Aki K (1957). Space and Time Spectra of Stationary Stochastic Waves, with Special Reference to Microtremors. Bulletin of the Earthquake Research Institute, 35, 415–459.

Aki K and Richards PG (1980) Quantitative Seismology, Theory and Methods. Volume I: 557 pp., 169 illustrations. Volume II: 373 pp., 116 illustrations. San Francisco: Freeman. Price: Volume I, U.S. $35.00; Volume II, U.S. $35.00. ISBN 0 7167 1058 7 (Vol. I), 0 7167 1059 5 (Vol. II). Geological Magazine, 118(2), 208–208. https://doi.org/10.1017/S0016756800034439.

Badsar SA, Schevenels M, Haegeman W and Degrande G (2010) Determination of the material damping ratio in the soil from SASW tests using the half-power bandwidth method. Geophysical Journal International, 182(3), 1493–1508. https://doi.org/10.1111/j.1365-246X.2010.04690.x.

Badsar S (2012) In situ determination of material damping in the soil at small deformation ratios. Katholieke Universiteit Leuven.

Bergamo P, Marano S and Fah D (2023) Joint estimation of S-wave velocity and damping ratio of the near-surface from active Rayleigh wave surveys processed with a wavefield decomposition approach. Geophysical Journal International 233 (3), 1560–1579. https://doi.org/10.1093/gji/ggad010.

Biot MA (1956) Theory of Propagation of Elastic Waves in a Fluid-Saturated Porous Solid. I. Low-Frequency Range. The Journal of the Acoustical Society of America, 28(2), 168–178. https://doi.org/10.1121/1.1908239.

Capon J (1969) High-resolution frequency-wavenumber spectrum analysis. Proceedings of the IEEE, 57(8), 1408–1418. https://doi.org/10.1109/PROC.1969.7278.

Lai CG (1998) Simultaneous inversion of Rayleigh phase velocity and attenuation for near-surface site characterization. Georgia Institute of Technology.

Cupillard P and Capdeville Y (2010) On the amplitude of surface waves obtained by noise correlation and the capability to recover the attenuation: a numerical approach. Geophysical Journal International. https://doi.org/10.1111/j.1365-246X.2010.04586.x.

Comina C, Foti S, Boiero D and Socco LV (2011) Reliability of V_S, 30 evaluation from surface-wave tests. Journal of Geotechnical and Geoenvironmental engineering, 137(6), 579-586.

Cox BR and Beekman AN (2011) Intra-Method Variability in ReMi Dispersion and Vs Estimates at Shallow Bedrock Sites. Journal of Geotechnical and Geoenvironmental Engineering, 137(4), pp. 354-362.

Cox BR and Wood CM (2011) Surface Wave Benchmarking Exercise: Methodologies, Results, and Uncertainties. GeoRisk 2011, 845–852. https://doi.org/10.1061/41183(418)89.

Crow H, Hunter JA and Motazedian D (2011) Monofrequency in situ damping measurements in Ottawa area soft soils. Soil Dynamics and Earthquake Engineering, 31(12), 1669–1677. https://doi.org/10.1016/j.soildyn.2011.07.002.

Zywicki D (1999) Advanced signal processing methods applied to engineering analysis of seismic surface waves. Georgia Institute of Technology.

Johnson D and Dudgeon D (1993) Array signal processing: Concepts and techniques. Englewood Cliffs, NJ: P T R Prentice Hall.

Evans JP, McCalpin JP and Holmes DC (1996) Geologic Map of the Logan 7.5' Quadrangle, Cache County, Utah. Miscellaneous Publication 96-1, Utah Geological Survey, a division of Utah Department of Natural Resources, Salt Lake City, Utah. ISBN 1-55791-375-7.

Ewing WM, Jardetzky WS, Press F and Beiser A (1957) Elastic Waves in Layered Media. Physics Today, 10(12), 27–28. https://doi.org/10.1063/1.3060203.

Foti S (2004) Using transfer function for estimating dissipative properties of soils from surface-wave data. Near Surface Geophysics, 2(4), 231–240. https://doi.org/10.3997/1873-0604.2004020.

Foti S, Aimar M and Ciancimino A (2021) Uncertainties in Small-Strain Damping Ratio Evaluation and Their Influence on Seismic Ground Response Analyses (pp. 175–213). https://doi.org/10.1007/978-981-16-1468-2_9.

Foti S, Hollender F, Garofalo F, Albarello D, Asten M, Bard PY, Comina C, Cornou C, Cox B, Di Giulio G, Forbriger T, Hayashi K, Lunedei E, Martin A, Mercerat D, Ohrnberger M, Poggi V, Renalier F, Sicilia D and Socco V (2018) Guidelines for the good practice of surface wave analysis: a product of the InterPACIFIC project. Bulletin of Earthquake Engineering, 16(6), 2367–2420. https://doi.org/10.1007/s10518-017-0206-7.

Foti S, Lai C, Rix GJ and Strobbia C (2014) Surface Wave Methods for Near-Surface Site Characterization. CRC Press. https://doi.org/10.1201/b17268.

Haendel A, Ohrnberger M and Krüger F (2016) Extracting near-surface Q L between 1–4 Hz from higher-order noise correlations in the Euroseistest area, Greece. Geophysical Journal International, 207(2), 655–666. https://doi.org/10.1093/gji/ggw295.

Hall L and Bodare A (2000) Analyses of the cross-hole method for determining shear wave velocities and damping ratios. Soil Dynamics and Earthquake Engineering, 20(1–4), 167–175. https://doi.org/10.1016/S0267-7261(00)00048-8.

Herrmann RB (2013) Computer programs in seismology: an evolving tool for instruction and research, Seismol. Res. Lett., 84, 1081–1088. 10.1785/0220110096.

Holzlöhner U (1980) Vibrations of the elastic half-space due to vertical surface loads. Earthquake Engineering & Structural Dynamics, 8(5), 405–414. https://doi.org/10.1002/eqe.4290080504.

Johnston DH, Toksöz M N and Timur A (1979) Attenuation of seismic waves in dry and saturated rocks: II. Mechanisms. Geophysics, 44(4), 691–711. https://doi.org/10.1190/1.1440970

Jongmans D (1990) In-situ attenuation measurements in soils. Engineering Geology, 29(2), 99–118. https://doi.org/10.1016/0013-7952(90)90001-H.

Keilis-Borok V (1989) Seismic Surface Waves in a Laterally Inhomogeneous Earth (V. I. Keilis-Borok, Ed.; Vol. 9). Springer Netherlands. https://doi.org/10.1007/978-94-009-0883-3.

Lacoss RT, Kelly EJ and Toksöz MN (1969) Estimation of seismic noise structure using arrays. Geophysics, 34(1), 21–38. https://doi.org/10.1190/1.1439995

Lai CG, Rix GJ, Foti S and Roma V (2002) Simultaneous measurement and inversion of surface wave dispersion and attenuation curves. Soil Dynamics and Earthquake Engineering, 22(9–12), 923–930. https://doi.org/10.1016/S0267-7261(02)00116-1.

Lamb H (1904) On the propagation of tremors over the surface of an elastic solid. Proceedings of the Royal Society of London, 72(477–486), 128–130. https://doi.org/10.1098/rspl.1903.0029.

Aimar A (2022) Uncertainties in the estimation of the shear-wave velocity and the small-strain damping ratio from surface wave analysis. Politecnico di Torino.

Michaels P (1998) In Situ Determination of Soil Stiffness and Damping. Journal of Geotechnical and Geoenvironmental Engineering, 124(8), 709–719. https://doi.org/10.1061/(ASCE)1090-0241(1998)124:8(709).

Nazarian S, Stokoe KH and Hudson WR (1983) Use of spectral analysis of surface waves method for determination of moduli and thicknesses of pavement systems. Transportation Research Record, 38–45. https://api.semanticscholar.org/CorpusID:58935998.

O'doherty RF and Anstey NA (1971) Reflections on Amplitudes*. Geophysical Prospecting, 19(3), 430–458. https://doi.org/10.1111/j.1365-2478.1971.tb00610.x.

Ohrnberger M, Schissele E, Cornou C, Wathelet M, Savvaidis A, Scherbaum F, . . . Kind F (2004) Microtremor array measurements for site effect investigations: Comparison of analysis methods for field data crosschecked by simulated wavefields. 13th World Conference on Earthquake Engineering. Vancouver, B.C., Canada.

Papadopoulos M, François S, Degrande G and Lombaert G (2018) The influence of uncertain local subsoil conditions on the response of buildings to ground vibration. Journal of Sound and Vibration, 418, 200–220. https://doi.org/10.1016/j.jsv.2017.12.021.

Park CB, Miller RD and Xia J (1999) Multichannel analysis of surface waves. Geophysics, 64(3), 800–808. https://doi.org/10.1190/1.1444590.

Parolai S (2014) Shear wave quality factor Qs profiling using seismic noise data from microarrays. Journal of Seismology, 18(3), 695–704. https://doi.org/10.1007/s10950-014-9440-5.

Parolai S, Lai CG, Dreossi I, Ktenidou OJ and Yong A (2022) A review of near-surface QS estimation methods using active and passive sources. Journal of Seismology, 26(4), 823–862. https://doi.org/10.1007/s10950-021-10066-5.

Peterson JR (1993) Observations and modeling of seismic background noise. In. Albuquerque, New Mexico: United States Geological Survey.

Prieto GA, Lawrence JF and Beroza GC (2009) Anelastic Earth structure from the coherency of the ambient seismic field. Journal of Geophysical Research: Solid Earth, 114(B7). https://doi.org/10.1029/2008JB006067.

Perez M (1969) The Capability of the Utah State University Drainage Farm as an Irrigation and Drainage Demonstration Project. All Graduate Plan B and other Reports, Spring 1920 to Spring 2023. 623. https://doi.org/10.26076/6e25-d8ed.

Ricker N (1940) The form and nature of seismic waves and the structure of seismograms. Geophysics, 5(4), 348–366. https://doi.org/10.1190/1.1441816.

Rix GJ, Lai CG and Spang AW (2000) In Situ Measurement of Damping Ratio Using Surface Waves. Journal of Geotechnical and Geoenvironmental Engineering, 126(5), 472–480. https://doi.org/10.1061/(ASCE)1090-0241(2000)126:5(472).

Rix G, Lai C and Foti S (2001) Simultaneous Measurement of Surface Wave Dispersion and Attenuation Curves. Geotechnical Testing Journal, 24(4), 350. https://doi.org/10.1520/GTJ11132J.

Richart FE, Hall JR and Woods RD (1970) Vibrations of Soils and Foundations. Prentice-Hall.

Sanchez-Sesma FJ and Campillo MI (2006) Retrieval of the Green's Function from Cross Correlation: The Canonical Elastic Problem. Bulletin of the Seismological Society of America, 96(3), 1182–1191. https://doi.org/10.1785/0120050181.

Schevenels M, François S and Degrande G (2009) EDT: An ElastoDynamics Toolbox for MATLAB. Computers & Geosciences, 35(8), 1752–1754. https://doi.org/10.1016/j.cageo.2008.10.012.

Shibuya S, Mitachi T, Fukuda F and Degoshi T (1995) Strain Rate Effects on Shear Modulus and Damping of Normally Consolidated Clay. Geotechnical Testing Journal, 18(3), 365. https://doi.org/10.1520/GTJ11005J.

Snieder R, Wapenaar K and Wegler U (2007) Unified Green's function retrieval by cross-correlation; connection with energy principles. Physical Review E, 75(3), 036103. https://doi.org/10.1103/PhysRevE.75.036103.

Spencer TW, Edwards CM and Sonnad JR (1977) Seismic wave attenuation in nonresolvable cyclic stratification. Geophysics, 42(5), 939–949. https://doi.org/10.1190/1.1440773.

Stoll RD (1974) Acoustic Waves in Saturated Sediments. In Physics of Sound in Marine Sediments (pp. 19–39). Springer US. https://doi.org/10.1007/978-1-4684-0838-6_2.

Stokoe K, Rix G, Nazarian S (1989) In situ seismic testing with surface waves. 12th International Conference on Soil Mechanics and Foundation Engineering.

Tao Y and Rathje E (2019) Insights into Modeling Small-Strain Site Response Derived from Downhole Array Data. Journal of Geotechnical and Geoenvironmental Engineering, 145(7). https://doi.org/10.1061/(ASCE)GT.1943-5606.0002048.

Teague DP, Cox BR, Bradley B and Wotherspoon L (2018) Development of Deep Shear Wave Velocity Profiles with Estimates of Uncertainty in the Complex Inter-Bedded Geology of Christchurch, New Zealand. Earthquake Spectra, 34(2), 639-672. (https://doi.org/10.1193/041117EQS069M).

Teague DP, Cox BR, and Rathje ER (2018) Measured vs. Predicted Site Response at the Garner Valley Downhole Array Considering Shear Wave Velocity Uncertainty from Borehole and Surface Wave Methods. Soil Dynamics and Earthquake Engineering, 113(10), 339-355. https://doi.org/10.1016/j.soildyn.2018.05.031.

Tsai VC (2011) Understanding the amplitudes of noise correlation measurements. Journal of Geophysical Research, 116(B9), B09311. https://doi.org/10.1029/2011JB008483.

Tokimatsu K (1998) Geotechnical site characterization using surface waves. International Conference on Earthquake Geotechnical Engineering, ISBN 90-5410-581-X, 1333–1368.

Vantassel JP and Cox BR (2020) SWinvert: a workflow for performing rigorous 1-D surface wave inversions. Geophysical Journal International, 224(2), 1141–1156. https://doi.org/10.1093/gji/ggaa426.

Vantassel JP and Cox BR (2022) SWprocess: A Workflow for Developing Robust Estimates of Surface Wave Dispersion Uncertainty, Journal of Seismology, 26 (1): 731-756 (https://doi.org/10.1007/s10950-021-10035-y).

Vantassel JP (2021) jpvantassel/swprocess: v0.1.0b0. Zenodo https://doi.org/10.5281/zenodo.4584129.

Verachtert R, Lombaert G and Degrande G (2018) Multimodal determination of Rayleigh dispersion and attenuation curves using the circle fit method. Geophysical Journal International, 212(3), 2143–2158. https://doi.org/10.1093/gji/ggx523.

Walsh JB (1966) Seismic wave attenuation in rock due to friction. Journal of Geophysical Research, 71(10), 2591–2599. https://doi.org/10.1029/JZ071i010p02591.

Walsh JB (1968) Attenuation in partially melted material. Journal of Geophysical Research, 73(6), 2209–2216. https://doi.org/10.1029/JB073i006p02209.

Wathelet M, Jongmans D and Ohrnberger M (2004) Surface-wave inversion using a direct search algorithm and its application to ambient vibration measurements. Near surface geophysics, 2(4), 211-221.

Wathelet M, Guillier B, Roux P, Cornou C and Ohrnberger M (2018) Rayleigh wave three-component beamforming: signed ellipticity assessment from high-resolution frequency-

wavenumber processing of ambient vibration arrays. Geophysical Journal International, 215(1), 507–523. https://doi.org/10.1093/gji/ggy286.

Wathelet M, Chatelain JL, Cornou C, Di Giulio G, Guillier B, Ohrnberger M and Savvaidis A (2020) Geopsy: A User-Friendly Open-Source Tool Set for Ambient Vibration Processing. Seismological Research Letters, 91(3), 1878--1889, doi: 10.1785/0220190360.

Williams JS (1962) Lake Bonneville: Geology of Southern Cache Valley, Utah. Geological Survey Professional Paper 257-C. United States Government Printing Office, Washington, D.C.

Xia J, Miller RD, Park CB and Tian G (2002) Determining Q of near-surface materials from Rayleigh waves. Journal of Applied Geophysics, 51(2–4), 121–129. https://doi.org/10.1016/S0926-9851(02)00228-8.

Zywicki DJ and Rix GJ (2005) Mitigation of Near-Field Effects for Seismic Surface Wave Velocity Estimation with Cylindrical Beamformers. Journal of Geotechnical and Geoenvironmental Engineering, 131(8), 970–977. https://doi.org/10.1061/(ASCE)1090-0241(2005)131:8(970).

Zywicki D (1999) Advanced signal processing methods applied to engineering analysis of seismic surface waves. Georgia Institute of Technology.

CHAPTER 4

AN OPEN-ACCESS DATA SET OF ACTIVE-SOURCE AND PASSIVE-WAVEFIELD DAS AND NODAL STATION MEASUREMENTS AT THE NEWBERRY FLORIDA SITE

Abstract

This paper documents a comprehensive subsurface imaging experiment using stress waves in Newberry, Florida, at a site known for significant spatial variability, karstic voids, and underground anomalies. The experiment utilized advanced sensing technologies, including approximately two kilometers of distributed acoustic sensing (DAS) fiber optic cable, forming a dense 2D array of 1920 channels, and a 2D array of 144 three-component nodal stations, to sense active-source and passive-wavefield stress waves. The active-source data were generated using a vibroseis shaker truck and impact sources, and it was simultaneously sensed by both the DAS and the nodal stations. The vibroseis truck was used to excite the ground in the three directions at 260 locations inside and outside the instrumented array, while the impact sources were used at 268 locations within the instrumented array. The passive-wavefield data recorded using the nodal stations comprised 48 hours of ambient noise collected over a period of four days in four twelve-hour time blocks. Meanwhile, the passive wavefield data collected using DAS consisted of four hours of ambient noise recordings. This paper aims to provide a comprehensive overview of the testing site, experiment layout, the DAS and nodal station acquisition parameters, implemented processing steps, and potential use cases of the dataset. While 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 paper is on documenting this unique dataset rather than on processing the data for detecting anomalies or generating subsurface 2D/3D imaging results. The raw and processed data, along with detailed documentation of the experiment and Python tools to aid in visualizing the DAS dataset have been archived and made publicly available on DesignSafe under project PRJ-3521.

Keywords: DAS; SmartSolo; Nodal Stations; Vibroseis Shaker Truck; T-Rex; FWI; Imaging; Surface Wave; Inversion; Ambient Noise; Dataset

4.1 Introduction

Non-invasive imaging techniques are increasingly being used for geotechnical site characterization due to their advantages in time, cost, and spatial coverage when compared to traditional invasive methods. Geophysical imaging based on stress 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 damping estimation. High-quality field data are essential for developing and testing these emerging techniques. This paper presents a comprehensive and open-access dataset of stress wave recordings gathered using some of the most advanced technologies available in geophysical-non-invasive subsurface imaging. A test site in Newberry, Florida was selected for this extensive subsurface imaging experiment due to its complex geology, which includes many known and unknown karstic voids of variable size and depth. A 2D layout of distributed acoustic sensing (DAS) fiber optic cable and a 2D array of three-component (3C) geophone nodal stations covering an area approximately 155 m x 75 m were used at the site to record both active-source and passive-wavefield stress waves (refer

to Figure 4-1). The active sources used to initiate seismic wave propagation comprised both a broadband, three-dimensional, vibroseis shaker truck named T-Rex from the NHERI@UTexas experimental facility (Stokoe et al., 2020), and more-variable, narrow-band, impact sources. In total, approximately 2 km of DAS fiber optic cable and 144, 3C nodal stations were used to record wavefields from more than 367 shot locations. This unique and publicly accessible dataset is available on DesignSafe (Rathje et al., 2017; https://www.designsafe-ci.org/) under project PRJ-3521, "Active-source and Passive-wavefield DAS and Nodal Station Measurements at the Newberry Florida Site". The ensuing paragraphs offer a concise overview of the sensing technologies employed in this experiment and the potential value of the dataset documented herein.

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 sensors (Cox et al., 2012; Yu et al., 2019). It is increasingly being used to sense active and passive stress waves for geophysical imaging and seismic monitoring of the near surface (e.g., Dou et al., 2017; Hubbard et al., 2022; Vantassel et al., 2022). DAS measures dynamic strain by using an interrogator unit (IU) to fire a series of laser pulses (probe pulses) through a fiber optic 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 et al. (2022) have demonstrated that DAS

measurements can yield very 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 length scales (on the order of tens of kilometers), surpassing traditional sensing technologies (Soga & Luo, 2018). For instance, DAS measurements with the ~2-km long fiber optic cable used in the present study resulted in 1920 channels of vibration data sensed at a spatial resolution of 1.02 m, which is equivalent to deploying 1920 unidirectional geophones. In recent years, several studies have been conducted to evaluate the potential of using DAS for non-invasive near-surface imaging. Most of these studies utilized either a 1D (i.e., line) DAS cable (e.g., Hubbard et al., 2022; Vantassel et al., 2022), 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 (US). 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 stations and two DAS fiber optic cables, along with a vibroseis shaker. Their research focused on imaging the top four kilometers of the crust, and thus, their nodal stations were spaced out over several kilometers in each direction, with interstation distances varying from tens to hundreds of meters. Furthermore, their DAS cables had a nearly linear configuration, and the vibroseis source only generated vertical vibrations. In contrast, the experiment documented herein focuses on the near surface depths relevant to geotechnical engineering (less than ~ 50 m). The study stands out for its use of a densely

spaced 2D configuration of DAS cable to capture both passive- and active-wavefields generated by vibroseis shaking in three directions. These wavefields were simultaneously recorded by a dense 2D array of 3C geophone nodal stations, allowing for a detailed comparison of DAS and nodal station measurements, further setting this dataset apart from prior work.

The second sensing technology deployed at the test site was 3C nodal stations, which 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, albeit at approximately 1/5 the spatial resolution of the DAS measurements across the site. 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, while the horizontal crossline component is better suited to detect Love waves (Foti et al., 2018; Vantassel, Cox, et al., 2022; Vantassel & Cox, 2022). Since different wave types carry information about different mechanical properties of the subsurface (Sheriff & Geldart, 1995), analyzing the data collected by 3C nodal stations can provide valuable insights into the 3D mechanical properties of the site. The versatility of the 3C nodal station measurements open avenues for researchers to use the dataset who are working on techniques that leverage either the vertical, any of the horizontal, or any combination of the three directions of ground shaking measurements (e.g., Cheng et al., 2020, 2021; Cox et al., 2020; Fathi et al., 2016; Kristekova et al., 2020; Nakamura, 1989; Pan et al., 2016; Smith et al., 2019; Wang et al., 2019; Wathelet et al., 2018).

The karstic voids at the testing site add an intriguing dimension to this experiment. Several non-invasive seismic techniques have been developed to detect voids and other underground anomalies (e.g., Belfer et al., 1998; Branham & Steeples, 1988; Cook, 1965; Kolesnikov & Fedin, 2018; Kristekova et al., 2020; Pernod P. and Piwakowski, 1989; Sloan et al., 2012; Smith et al., 2019; Wang et al., 2019). However, detecting anomalies using real field data remains challenging (Grandjean & Leparoux, 2004; Sloan et al., 2010; Smith et al., 2019). For example, Smith et al. (2019) and Wang et al. (2019) conducted an experiment to image a known 0.9 m x 1.5 m x 96 m tunnel situated ten meters below the surface at Yuma Proving Ground, Arizona, US. The experiment utilized a dense 2D array comprising 720 vertical and 720 horizontal geophones and an accelerated weight-drop source. They used both 2D and 3D full waveform inversion (FWI) to image the tunnel. Smith et al. (2019) highlighted the advantages of using various source orientations with multicomponent seismic sensors when imaging for voids. Their research revealed that different combinations of source orientations with receiver components produced varying resolutions of the tunnel. The Smith et al. (2019) and Wang et al. (2019) studies also showed that 2D FWI was effective in imaging the known location of the tunnel, because the experiment was designed specifically for that purpose. However, for imaging unknown void locations with complex shapes, 3D FWI would be more suitable. Smith et al. (2019) noted that the resolution of the tunnel was limited by the lack of higher-frequency data used in the inversion. We anticipate that the present dataset, 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 void imaging.

4.2 Site Overview

The test site is a dry retention pond located in Newberry, Florida (29°39'0.39" N, 82°35'51.20" W) along State Road 26 (refer to Figure 4-1). Sinkholes are common in this area and generally an immense problem in parts of Florida due to karstic geology. The Newberry retention pond site has undergone thorough investigations over the years, utilizing both invasive and non-invasive methods. Findings from these previous studies indicate that the subsurface is comprised of medium-dense fine sand and silt that range in depth from two to ten meters overlying highly variable limestone (Tran et al., 2013; Tran & Hiltunen, 2011). The site also contains sinkholes of different sizes and depths, some of which have been repaired. 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 non-invasive subsurface imaging research. Tran and Hiltunen (2011) conducted ten cone penetration tests (CPT), eight geotechnical borings with standard penetration tests (SPT), and 12 consecutive 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 (Vp) 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 ten 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 three-dimensional 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 one-dimensional, small vibroseis shaker truck named Thumper from the 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 x 12 configuration. The source and receiver 2D grids were uniform, with 3-m spacing, covering a total area of 12 m x 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.

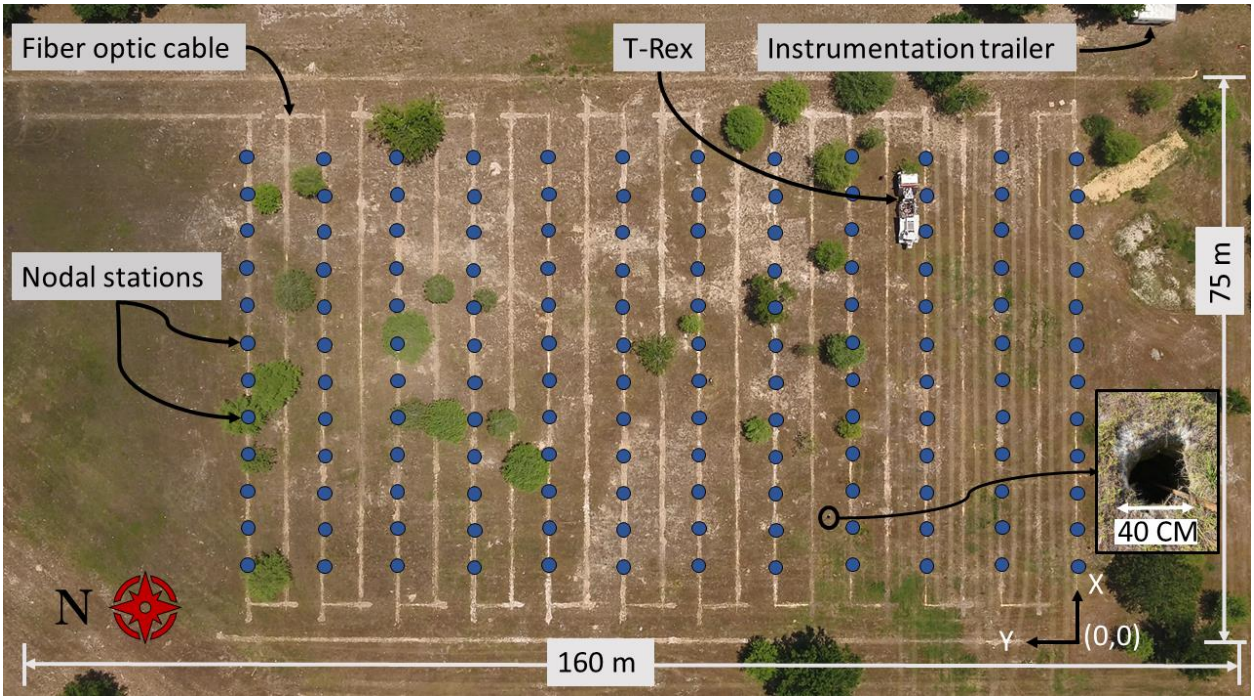

Figure 4-1 Overhead-view of the Newberry test site captured from a drone, showing the trenches used for fiber optic cable installation, as indicated by white, linear markings on the ground surface, and the 3C geophone nodal station locations, indicated by blue circle symbols. Additionally, the figure 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.

Despite being used for a number of previous subsurface imaging studies, the authors felt there was an opportunity to collect a unique subsurface imaging dataset at the Newberry site that would improve upon previous studies by: (a) covering a larger spatial area, (b) using DAS to enable much more dense spatial sampling, (c) using 3C nodal stations to enable multi-component processing, (d) using a dense grid of more numerous shot locations, and (e) incorporating 3C shaking capabilities from a broadband and powerful vibroseis shaker truck. It is hoped that this will enable more advanced processing techniques to be applied, both in the present and in the future, with the goal of achieving deeper and higher resolution imaging to resolve subsurface anomalies.

4.3 Overview of the Dataset

The subsurface imaging dataset documented in this paper was collected over the course of eight days, beginning Monday, May 9th, 2022, and ending Monday, May 16th, 2022. The dataset consists of both active-source and passive-wavefield stress wave recordings that were sensed using 2D arrays of DAS fiber optic cable and 3C nodal stations. Approximately two kilometers of continuous DAS fiber optic cable was laid out in the zigzag pattern shown schematically by a black line in Figure 4-2, which is also visible through the white markings from the fiber optic cable trenches shown in the drone image presented in Figure 4-1. The fiber optic cable was interrogated using an OptaSense ODH 4+ interrogator unit (IU) configured to have a 1.02-m channel separation, resulting in a total of 1920 channels. A 2D array of 144, 3C geophone nodal stations was also deployed on site, as indicated by the blue-solid circles in Figures 4-1 and 4-2. These stations were arranged in a 12 x 12 grid, evenly spaced every five and ten meters in the X (west-east) and Y (south-north) directions, respectively. The active-source data were generated from 260 shot locations inside and outside of the instrumented area where T-Rex was used to shake the ground in all three directions, and 286 shot locations inside the instrumented area where impact sources were used to strike the ground vertically (refer to Figure 4-2). The passive wavefield data consisted of approximately four hours of ambient noise recordings using DAS, and approximately 48 hours of ambient noise recordings using the nodal stations. The dataset is permanently archived and publicly available on DesignSafe (Rathje et al., 2017; https://www.designsafe-ci.org/) under a project titled “Active-source and Passive-wavefield DAS and Nodal Station Measurements at the Newberry Florida Site”. The project houses three parent folders: a "Raw data" folder, which contains both the DAS

and the nodal station data in their original, as-collected form; a "Processed data" folder, which organizes the data in a user-friendly format after undergoing preprocessing; and a "Supporting documents" folder, which includes complete and thorough documentation of the experiment. The experimental setup and dataset organization are explained in detail in the following sections.

4.4 Experiment layout

The experiment was laid out on a 2D survey grid, as illustrated in Figure 4-2. 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 five meters and were labeled with the letters A through W in order from south to north, respectively. The lowermost line, Z, was 15 m south of line A, while the uppermost line, ZZ, was 30 m north of line W. The vertical lines were spaced five meters apart and named from west to east, 101, 102, 1, 2, 3, … through 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 dataset, this paper 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.

The site is relatively flat and the grass was generally quite short at the time of testing. The lines and points of the survey grid, which mark the fiber optic cable and nodal

station locations, as detailed in subsequent sections, were surveyed into position using two total stations and several 100-m tape measures. The first total station was deployed at point A102 (coordinate 5,0) and was oriented in the positive direction of the Y axis (refer to Figure 4-2). With the bearing set, the intersection points between all the horizontal lines (e.g., B, C, D, etc.) with line 102 were established. The total station was then rotated 90° towards the east to align with the positive direction of the X axis. With this bearing set, the intersection points between all the vertical lines (e.g., 1, 2, 3, etc.) with line A were established. At that point, the second total station was deployed at point A103 (coordinate 70,0) to survey the intersection points between all the horizontal lines (e.g., B, C, D, etc.) with line 103. This process was repeated for all lines of the grid, with forward- and back-sights established whenever a total station was relocated to maintain orthogonality. Projecting the experiment layout onto the test site was completed in the first day, although minor surveying was performed throughout the experiment to ensure the proper placement of the fiber optic cable, nodal stations, shots, etc.

4.5 Distributed Acoustic Sensing (DAS)

Selecting the right fiber optic cable is crucial for good DAS measurements, since the cable functions as both the strain sensing element and the means of transmitting optical signals (Hartog, 2018). In this experiment, a 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 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/over it. The length of cable installed at the site was approximately 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 Figure 4-1), and the other end appropriately terminated at the northwest corner of the site to reduce end reflections. The cable was buried in a shallow trench to ensure optimal coupling between the ground and the cable. To facilitate precise trenching along the grid lines, a total station was positioned at the endpoint of each cable route and a tape measure was pulled tight to mark the trenching path. A second tape measure was then horizontally offset from the first by an amount equal to the distance between the trencher blade and its tire, as illustrated in Figure 4-3a. The trencher was guided along the surveyed line to create a trench that was approximately 20-cm wide and 10- to 15-cm deep, ensuring its straightness. The cable was then manually placed at the bottom of the trench by rolling the cable spool over it (Figure 4-3b), with slight tension applied to minimize slack. The corners of the trench were rounded to a radius of approximately 20 cm (Figure 4-3c), 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 (Figure 4-3d and 3e). The backfilled soil was then manually compacted over the cable to ensure good coupling with the native ground (Figure 4-3f). 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 noting the DAS channels that responded with 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 shown in Figure 4-2. 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. As noted below with respect to DAS data acquisition, the gauge length used throughout the study was 2.04 m, so the absolute spatial location of each channel is accurate to within about 1 m. However, the relative location/spacing between channels is exactly equal to one-half of the gauge length, or 1.02 m. The corners were backfilled and compacted to ensure proper coupling between the cable and the surrounding soil. Trenching started on Monday, May 9th, from point Z104 and ended at point J102. On Tuesday, May 10th, trenching was completed, and the cable installation began from Z104, reaching J07 by the end of the day. The cable installation was completed on Wednesday, May 11th.

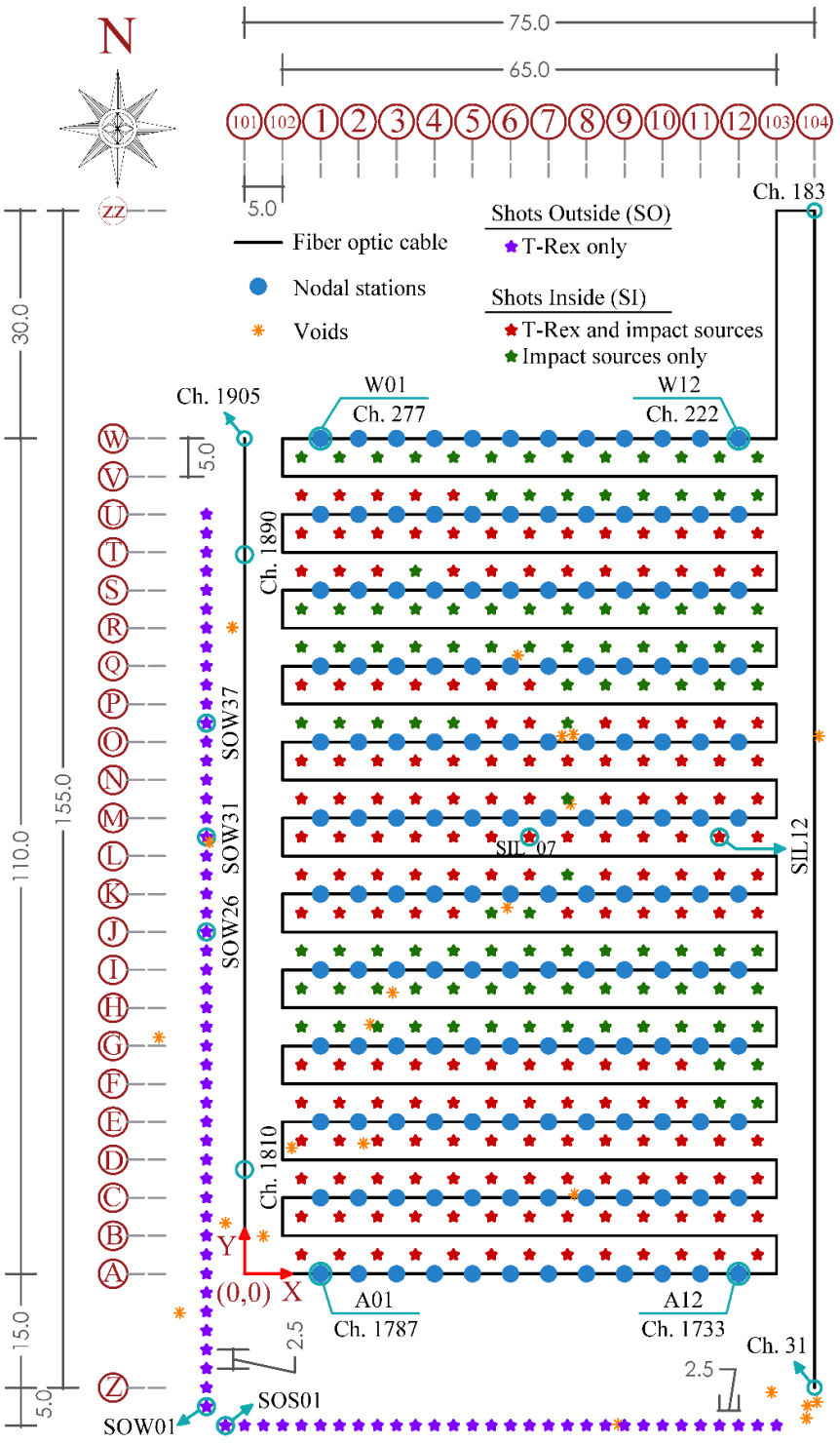

Figure 4-2 Schematic layout of the test site showing locations of the: 3C geophone nodal stations, fiber optic cable, T-Rex and impact shots, and voids that are visible from the ground surface. The layout is comprehensive, including all of the line numbers/letters and dimensions used to arrange the equipment.

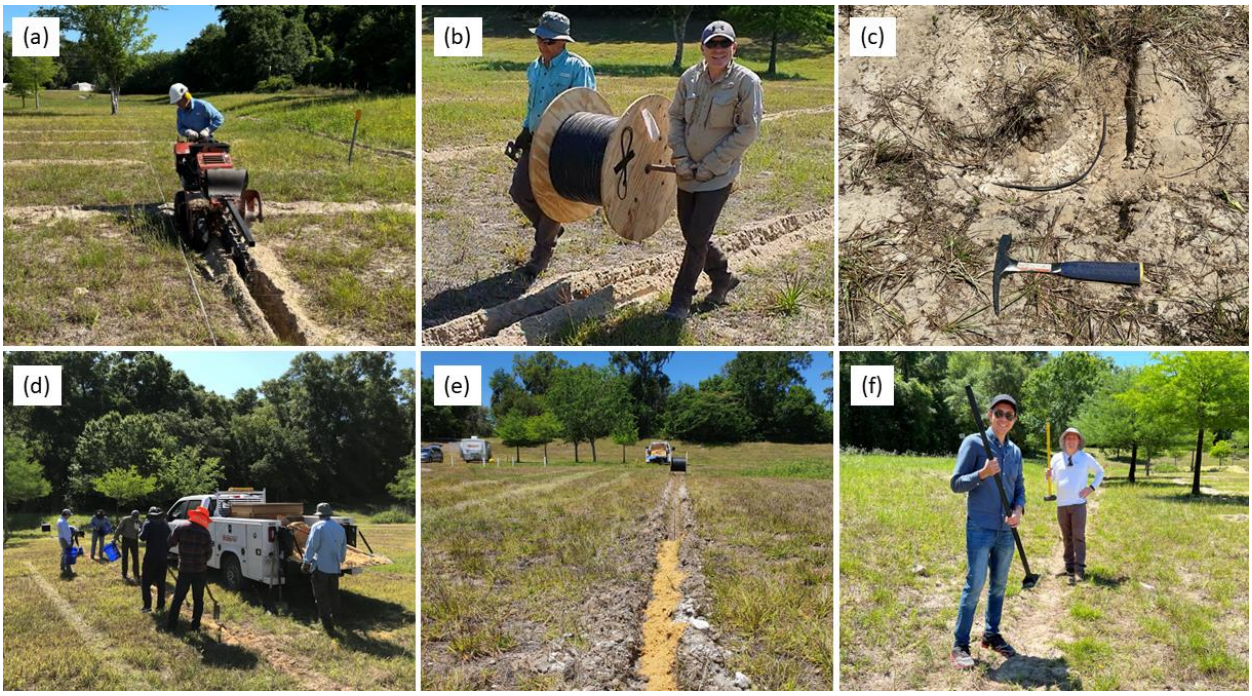

Figure 4-3 Pictures illustrating the fiber optic cable installation procedure, starting with: (a) trenching the cable route, (b) laying the cable by rolling it off the cable spool along the trench, (c) rounding the cable at the corners, (d) and (e) filling the trench with native soil or clean sand as required, and finally (f) compacting the backfilled soil to ensure proper coupling between the cable and the ground.

4.6 Nodal stations

4.6.1 Instrumentation

The nodal stations used in this experiment were SmartSolo IGU-16HR 3C. They have a compact, all-in-one modular design, with a GPS-synced, 32-bit digitizer (accurate to ±10 microseconds), a maximum input signal of ±2.5 Volts at 0 dB gain, and a storage capacity of 64 GB. Each station is equipped with three, orthogonal, 5-Hz geophones 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 an upward voltage departure in the geophone oriented along that axis (refer to Figure 4-4d). Four conical spikes were mounted on each nodal station to ensure good coupling with the ground. These IGU-16HR 3C stations have

a small footprint of 95 mm x 103 mm, a height of 30 cm, and weigh around 2.4 kg with spikes attached. The 144 nodal stations used in this experiment were sourced from two locations: 88 stations from the Earthquake Engineering and Subsurface Imaging Lab at Utah State University (labeled USU01 through USU90, excluding stations USU07 and USU42), and 56 stations from the NHERI@UTexas experimental facility (labeled UT01 through UT56). The stations from both sources were independently labeled in ascending order based on their serial numbers.

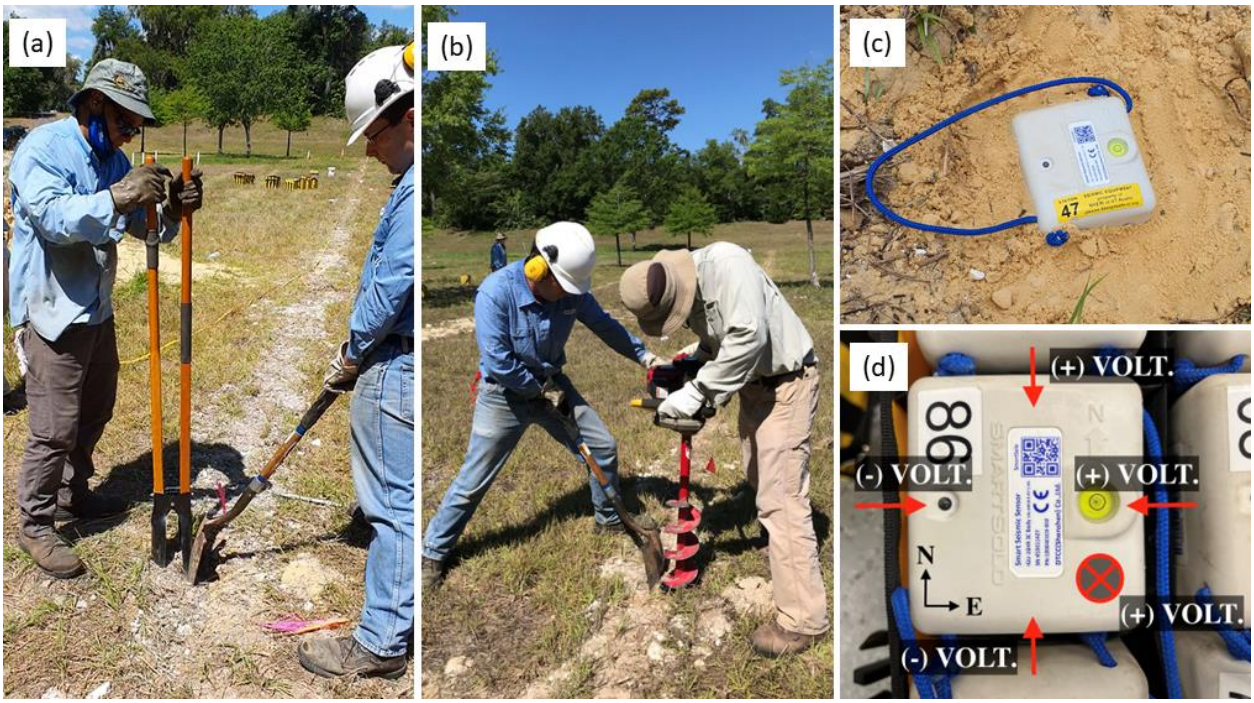

Figure 4-4 Picture illustrating the installation of nodal stations next to the fiber optic calbe. In panels (a) and (b), excavation is performed with either a post-hole digger or a gas-powered auger, respectively, while the fiber optic cable is protected with a shovel. Panel (c) depicts a completed installation, with the station securely in place and its top exposed. Panel (d) describes the voltage polarity of the three geophones in each nodal station.

4.6.2 Installation

The 144, 3C geophone nodal stations were arranged in a 12 x 12 grid configuration with ten- and five-meter spacings in the south-north and west-east directions, respectively, as shown in Figures 4-1 and 4-2. Starting with the USU stations, the stations were placed in order from south to north along line 1 (USU01 to USU13), then along line 2 (USU14 to USU25), and subsequently all the way to line 12. Remember, stations USU07 and USU42 were not used in this experiment. Thus, the UT stations started at I08 and followed the same pattern until the last station was installed at W12. The stations were deployed right next to the fiber optic cable in their respective locations. Depending on the stiffness of the ground, either a gas-powered earth auger or a manual post-hole digger was used to excavate a hole approximately 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 Figures 4-4a and 4-4b. The orientation of the stations was such that the north arrow pointed towards the +Y direction, which was orthogonal to the fiber optic cable, conveniently aligning with true north (within 1° tolerance). This ensured that the east-west geophone/channel in the three-component 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 4-4c. Following deployment of all stations, the stations were activated for continuous recording. This process involved booting up each station using a magnet switch, confirming successful activation by observing a flashing green LED state indicator, conducting a quality scan using a designated handheld device, and surveying its final location. Deploying and activating the stations started on Wednesday, May 11th, and ended on Thursday, May 12th.

4.7 Active-source wavefield generation

On Friday, May 13th, following installation of the fiber optic cable and all SmartSolo nodal stations, active-source wavefield recording commenced. Two source types were used for active wavefield generation: a highly controlled, powerful, broadband vibroseis source, and more variable, narrow-band, impact sources. The vibroseis source used in this experiment was the NHERI@UTexas, large mobile shaker truck named T-Rex (Figure 4-5a). T-Rex is a 29-ton, tri-axial vibroseis truck capable of shaking its baseplate in the vertical, longitudinal, and transverse directions (Stokoe et al., 2020). It has a maximum force output of about 267 kN in the vertical direction and 134 kN in each horizontal direction. In this experiment, T-Rex was used to generate a 12-second-long chirp with a linear frequency sweep 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 where 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 4-2. 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. Note that the three SON locations are not shown in Figure 4-2 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 furthest to the south and inside the instrumented area (refer to Figure 4-2) is named 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 labelled in Figure 4-2. While the inside shot locations are organized on a regular 5-m x 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 weight of the truck. Thus, the SI locations illustrated in Figure 4-2 are clearly denoted as those where 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., in-line 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. Due to 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).

Figure 4-5 Pictures illustrating the various sources utilized in the experiment. Panel (a) shows the large, tri-axial vibroseis shaker truck, T-Rex, from the NHERI@UTexas experimental facility, panel (b) shows the PEG-40kg impact source mounted on a Ford F-350 pickup truck, and panel (c) shows an eight-pound sledgehammer.

The impact sources consisted of a 40-kg propelled energy generator (PEG-40kg), also known as an accelerated weight drop, manufactured by R.T. Clark Companies (Figure 4-5b), and an eight-pound sledgehammer (Figure 4-5c). The PEG-40kg 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-40kg: single-cycle and continuous-cycle. For this experiment, the single-cycle mode was used. To ease mobility, the PEG-40kg was mounted on a Ford F-350 pickup truck, as shown in Figure 4-5b. The impact sources were only used to excite the ground inside the instrumented area, as shown in Figure 4-2. Since 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-40kg, ensuring that all inside shot locations in the 5-m x 5-m grid were covered by one of the impact sources.

4.8 Passive-wavefield monitoring

The DAS passive-wavefield data consisted of approximately four hours of ambient noise recordings on May 15th between 16:58 to 21:04 Universal Time Coordinated (UTC). During this period, there were two instances of rainfall which took place from 17:20 to 17:35 UTC and 20:54 to 20:56 UTC. Additionally, there was notable interference (i.e., noise) from an electric power generator on channels 30 to 66 and 1720 to 1742 during this time.

Every day from May 12th to May 16th the nodal stations were left to record ambient noise between 23:00 UTC, when work at the site was completed, to 11:00 UTC, when work resumed the next day. This resulted in a total of 48 hours of ambient noise data distributed over four, 12-hour time blocks gathered over a period of four days.

4.9 Data acquisition

4.9.1 DAS

An OptaSense ODH4+ IU was used in this experiment. The IU was borrowed from the NHERI@UTexas experimental facility (Stokoe et al., 2020) 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. The IU configuration and the cable length resulted in a total of 1920 DAS channels. 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 demodulation errors. Phase demodulation errors arise when the strain change 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 time resolution/sampling rate, and the signal can be subsequently decimated to lower sampling rates relevant to the study being conducted in order to save memory space. On Friday, May 13th, during the first day of active wavefield data acquisition, the OptaSense 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 data collected on Friday, May 13th are of slightly higher quality because a 50 kHz ping rate was used. This should not be interpreted to mean that the data collected on other days are of poor quality due to it being acquired with a 20 kHz ping rate. Indeed, a 20 kHz ping rate is still more than adequate for interrogating a 2-km long cable. It's noteworthy that, on Friday, May 13th, T-Rex executed all shots along the SOS and SOW lines.

The OptaSense ODH4+ IU outputs the raw data in consecutive one-minute H5 files that encompass all the DAS channels. Each channel contains the digitized output from the

light phase difference over a gauge length, presented in radians relative to the average wavelength of the source light. Specifically, radians scaled by a factor of $2\pi/2^{16}$. H5 files utilize the hierarchical data format (HDF) to store vast amounts of data in the form of multidimensional arrays. At the end of the experiment, 538 GB of active and passive wavefield raw DAS data were collected and organized into three parent folders according to the source/excitation type: T-Rex, Impact sources, and Ambient noise.

4.9.2 Nodal stations

For this experiment, the SmartSolo nodal stations were configured with a sampling rate of 250 Hz. Every SmartSolo nodal station generated a folder, titled after its serial number, that housed three miniSEED files with amplitude units of counts, each file being associated with one of the station's three channels. These files contained the recording of the full duration of the experiment, from activation to deactivation of the stations. At the end of the experiment, the SmartSolo nodal stations had collected 48.6 GB of active-source and passive-wavefield data. The MiniSEED format, a commonly used subset of the Standard for the Exchange of Earthquake Data (SEED) format, was selected to facilitate data sharing and analysis. This format is widely recognized in the geophysical community, and there are open-source libraries like obspy and libmseed that provide support for reading and converting MiniSEED files to other formats. Thus, the raw data provided in the dataset comprises 144 folders, each corresponding to one of the 144 deployed nodal stations in the experiment.

4.10 Supporting metadata documents

The field handwritten notes were carefully reviewed, converted to digital format, and synthesized with other electronic notes and files to describe the dataset. All of this metadata have been organized in the “Supporting documents” folder. The information on the SmartSolo nodal stations, including serial numbers, station names, location codes, and both local and global coordinates, can be found in the "Nodal stations information" Excel file provided with the archived electronic dataset. The frequency instrument response curves for the vertical and horizontal geophones in the nodal stations are provided in the "DT-Solo5Hz(HP305) Frequency Response" Excel file. The "Cable information" Excel file contains the DAS channel indices, local and global coordinates for each corner along the fiber optic cable, as well as the cable's start and end points. The "T-Rex and impact sources information" Excel file provides information on all shots, whether from T-Rex or impact sources, including local and global coordinates for each shot, the shot location code, and the shot start time. Local and global coordinates for all the observed surface voids and trees in the vicinity of the instrumented area are provided in the “Voids and trees locations” Excel file. While these documents are best viewed using Excel due to the presence of illustrative images and multiple tabs for organizing information, alternative open-source CSV files are also provided, albeit without the images and advanced organization of the Excel files. A comprehensive and interactive PDF layout schematic (similar to what is shown in Figure 4-2, but with more information) has also been added to the supporting documents to enhance visual understanding of the experimental configuration. This multi-layered layout includes a representation of the fiber optic cable, nodal stations, and shot locations, with the assigned DAS channel for each cable corner and the experiment dimensions, voids, and trees locations all clearly marked.

4.11 Processed data

The electronic dataset publicly accessible through DesignSafe and documented herein contains both the raw and processed data from the DAS and SmartSolo sensing systems. While the raw data may be of interest to some readers, parsing and interpreting it in preparation for subsequent analyses can be a time-consuming task. Thus, we anticipate that the processed data will be of greater interest to most readers. Nonetheless, this paper and the accompanying electronic dataset provide all of the necessary information to completely re-interpret the raw data if desired. Figure 4-6 depicts the folder structure of the "Processed data" folder, showcasing two parent folders, "DAS" and "SmartSolo." Within these parent folders, the data are organized, based on the specific type of wavefield source, into three folders, namely: "T-Rex," "Impact sources," and "Ambient noise." Figure 4-6 also indicates the number of subfolders and files in each folder, along with the naming convention used for the files. The following subsections outline the processing steps used for the DAS and the SmartSolo nodal stations data and provide further details about the files stored in each folder and their naming convention.

4.11.1 DAS

As mentioned above, the DAS parent folder contains folders that correspond to data collected from T-Rex, Impact sources, and Ambient noise. For each T-Rex shot, the following processing steps were followed: (1) 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 pre-trigger delay. As noted above, these times are also provided in the "T-Rex and impact sources information" Excel file. (2) A 15-second window encompassing the shot was extracted from the one-minute-long DAS

H5 files, which included a one-second pre-trigger delay, 12 seconds of T-Rex shaking, and two seconds of listen time post-T-Rex shaking to capture the waves reaching the array extremities. The 15-second windows were checked for dropped samples due to IU digitization and data storage errors in any of the 1920 DAS channels. Two of the 780 T-Rex shots had dropped samples on one or more channels. Specifically, Shots Z_SID11 and Z_SIF07 failed this check and were excluded from the processed dataset. (3) The raw data were scaled by $2\pi/2^{16}$ to convert it into phase change measurements in radians. (4) 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). (5) The DAS waveforms from the SOS and SOW shots were decimated from 10 kHz to 1 kHz to align with the rest of the collected data. (6) Phase data were converted to strain using Equation 1 (Hubbard et al. 2022):

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

where 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, approximately 1.47; ξ is the photoelastic scaling factor for longitudinal strain in an isotropic medium, equal to 0.78; g is the gauge length, approximately 2.04 m; and ε is the normal strain per single gauge length. (7) 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. The event object was created using a Python class,

designed to efficiently organize the data collected during this experiment. To ensure that researchers with minimal programming proficiency can utilize the dataset, we have included Python tools in the Supporting documents folder for effortless querying and visualization of the H5 files (i.e., event objects). Each H5 file was given a unique label composed of T-Rex shaking direction, shot location, and shot timestamp. For example, file Z_SOS01_20220513154814.50 contains the data obtained when T-Rex was shaking vertically (i.e., in the P-mode Z-direction) at shot location SOS01 on 13 May 2022 at 15:48:14.5 UTC. The H5 files were organized in the T-Rex folder by shot location into two folders, “Inside shots” and “Outside shots”, as illustrated in Figure 4-6. The Inside shots folder contains 535 H5 files, with a file for each T-Rex shaking direction for each of the 179-T-Rex inside shot locations (two shots were discarded as mentioned above). The Outside shots folder contains 243 H5 files, with a file for each T-Rex shaking direction for each of the 81 shot locations outside the array (refer to Figure 4-2). (8) All the DAS channels for each shot were cross-correlated with the T-Rex pilot signal and stored in H5 files, categorized into two folders: "Inside shots" and "Outside shots" based on the shot location. These folders contain the same number of files as the uncorrelated shots folders noted above. Additionally, the naming convention for the cross-correlated H5 files follow the same format as the uncorrelated shots, but with "CC" added to indicate "cross-correlated" (refer to Figure 4-6). For example, for the file described above, the cross-correlated data are stored in a file named Z_SOS01_CC_20220513154814.50, where CC distinguishes it from the uncorrelated records.

The same processing steps were followed for the impact sources, with a few notable exceptions. The shot start times were extracted from the field datasheet and verified

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. These times are provided in the “T-Rex and impact sources information” Excel file. The impact time was used to trim a three-second time window from the raw, one-minute H5 files, which included a one-second pre-impact portion and a two-second post-impact portion. The header of each processed H5 file for an impact source indicates if the shot was generated by an accelerated weight drop (i.e., the PEG-40kg) or the 8-lb hammer. The processed impact shot files were named using the same convention as noted above, but started with an "I" to denote "Impact" instead of an "X," "Y," or "Z" used to describe the T-Rex shaking orientation. These files are in a folder named “Inside shots” within the Impact sources folder to indicate that all the impact sources were excited inside the instrumented area. The Inside shots folder contains 286 H5 files corresponding to the 286 impact-source shot locations (refer to Figure 4-2).

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 one-minute-long segments and labeled with the prefix "N" to distinguish them as noise recordings. These files are stored in the Ambient noise folder, which contains 246, one-minute-long H5 files that correspond to approximately four hours of ambient noise DAS recordings.

4.11.2 Nodal stations

The SmartSolo nodal station data processing was limited to extracting the shot time windows (15-second windows for T-Rex shots and 3-second 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-second time window from all three components of the nodal stations for each shot. This included a one-second pre-trigger delay, a 12-second T-Rex shaking duration, and a 2-second listen segment post-T-Rex shaking. The T-Rex shot data collected from the nodal stations follows a comparable folder structure as that of the DAS, featuring two principal folders for "Inside shots" and "Outside shots". Nonetheless, each of these two folders is additionally subdivided into three folders based on the T-Rex shaking direction, namely "X Shaking", "Y Shaking", and "Z Shaking," corresponding to the local coordinate system. In each shaking direction folder for "Inside shots," individual folders are present for the 179-T-Rex inside shot locations. These folders encompass 145 files, out of which 144 correspond to the 144 nodal stations, while the remaining file comprises the T-Rex pilot signal, base plate and mass accelerations, and ground force in engineering units. Likewise, for the "Outside shots," the shaking directions folder comprise individual folders for the 81 outside shot locations, each of which contains 145 files. The miniSEED files for each nodal station include three components: DHN, DHE, and DHZ, following the naming convention recommended by FDSN (2012). The "D" in the name refers to the use of a 250 Hz sampling rate, "H" indicates the use of a high gain seismometer, and "N", "E", and "Z" indicate the geophone orientation (north, east, or vertical). 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 related to T-Rex shaking were named according to the format presented in Figure 4-6, including the

T-Rex shaking direction, shot location, UTC date and time, station location, and the designation "3c" to indicate that the file includes the station's three perpendicular geophone records. For example, the file "X_SIA01_20220514143604.85_A01_3C.miniseed" is a 15-second miniSEED recording of T-Rex shaking in the X direction at shot location SIA01 on 14 May 2022 at 14:36:04.85 UTC captured by a nodal station located at A01. The miniSEED files containing the T-Rex source information follow the same naming convention, except they end with "Source.miniseed" instead of the station location name and "3c.miniseed".

The miniSEED files corresponding to the impact sources have a 3-second duration, with 1 second before the impact and 2 seconds following it. The SmartSolo data from the impact sources are organized in a folder named "Inside shots" within the Impact sources folder, following a structure similar to that of the DAS folder. Nevertheless, in contrast to the DAS folder structure, each of the 286 impact-source shots is stored in a separate folder. The impact source signature was not recorded. As a result, each folder for a given impact shot location contains 144 miniSEED files (one for each nodal station). The file naming convention is similar to that previously described for T-Rex shots, with the only difference being that the files begin with the letter "I" instead of the T-Rex shaking direction, as shown in Figure 4-6.

As noted above, the nodal stations were used to record 48 hours of ambient noise over a period of four days, in 12-hour increments from 23:00 to 11:00 UTC. The recorded data are stored in the Ambient noise folder, which is subdivided into four folders, one for each day during which the ambient noise was recorded. Each of these four folders

comprises 1728 miniSEED files, consisting of 12 one-hour miniSEED files for each of the 144 nodal stations. These files are distinguished by the prefix "N" at the beginning, as depicted in Figure 4-6.

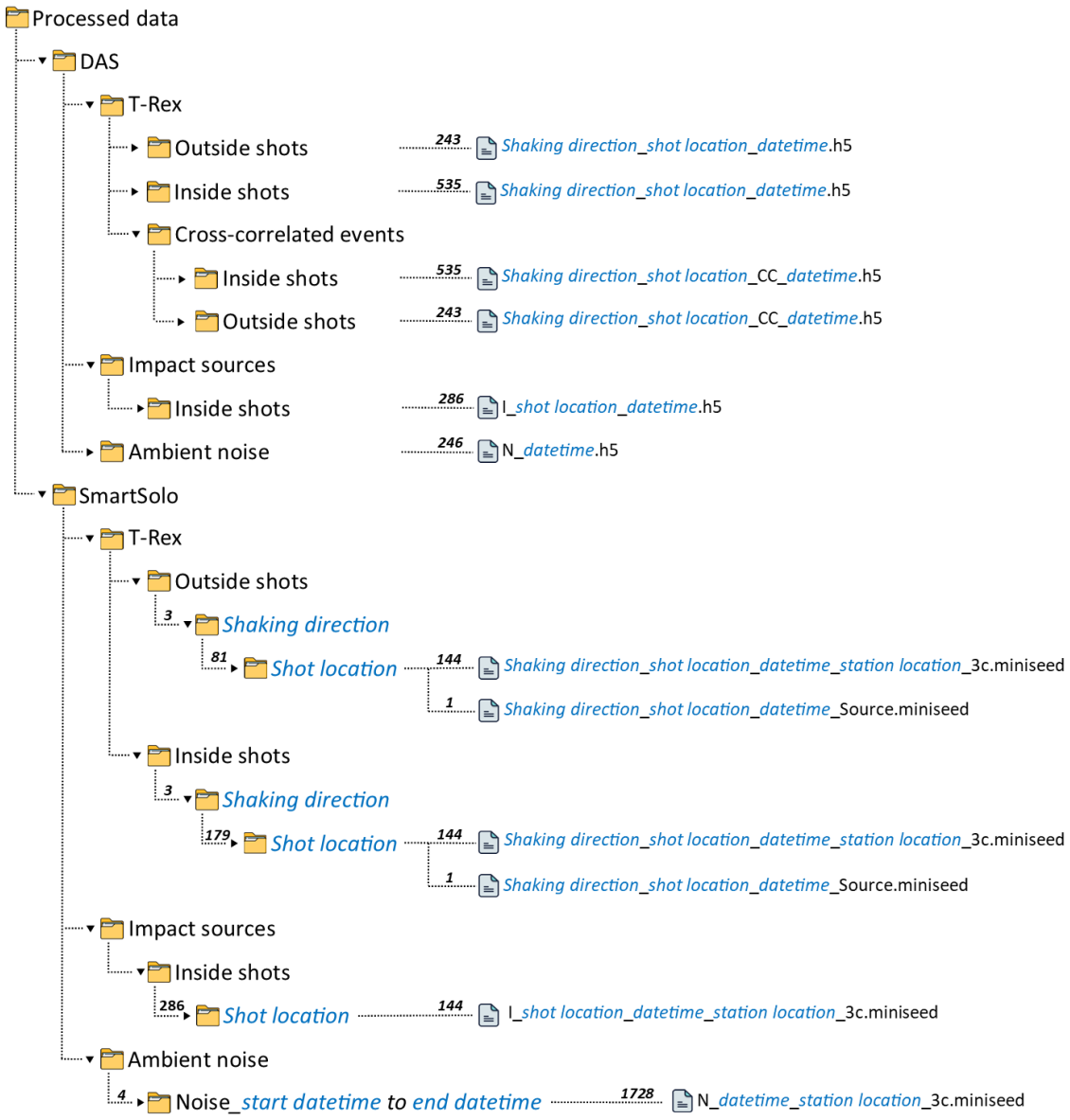

Figure 4-6 Schematic of the hierarchical folder structure for the processed data, indicating the number of subfolders and files within each folder. The blue, italicized text represents dynamic content, such as the T-Rex "shaking direction", which can be "X Shaking", "Y Shaking", or "Z Shaking", or the "shot location", which can be any of the numerous outside or inside shot locations, such as "SOS01". All dates and times are in Universal Time Coordinated (UTC).

4.12 Potential Dataset Use Cases

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

4.12.1 One-, two- and three-dimensional imaging

Previous research conducted at this site by Tran and Hiltunen (2011), Tran et al. (2013), and (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, despite being more sparsely spaced in comparison to the DAS. Successful attempts have been reported in the literature in imaging the subsurface using active source 2D FWI (e.g., Wang et al., 2019; Tran et al., 2013) 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/results, we instead focus on illustrating the quality of the collected waveforms and some potential ideas for taking advantage of the multi-direction sensing and multi-directional shaking on such a dense grid.

Figure 4-7 shows the waveforms generated by T-Rex shot X_SIL07 (refer to Figure 4-2) and recorded by the DAS channels and the DHE component of the nodal stations at the four furthest corners of the instrumented area (i.e., DAS channels 1787, 1733, 277, and 222, and nodal stations A01, A12, W01, and W12), which are labelled in Figure 4-2. These plots 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 (i.e., line 104) for T-Rex shot Y_SIL12, as shown in Figure 4-8. This waterfall plot presents the cross-correlated waveforms captured by DAS channels 31 through 183 (refer to Figure 4-2), with each trace normalized by its absolute maximum amplitude. Figure 4-8 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 4-7 and 4-8 are typical of the quality contained in this extensive dataset. As such, the waveforms should be more than adequate for performing various 1D, 2D and/or 3D imaging studies.

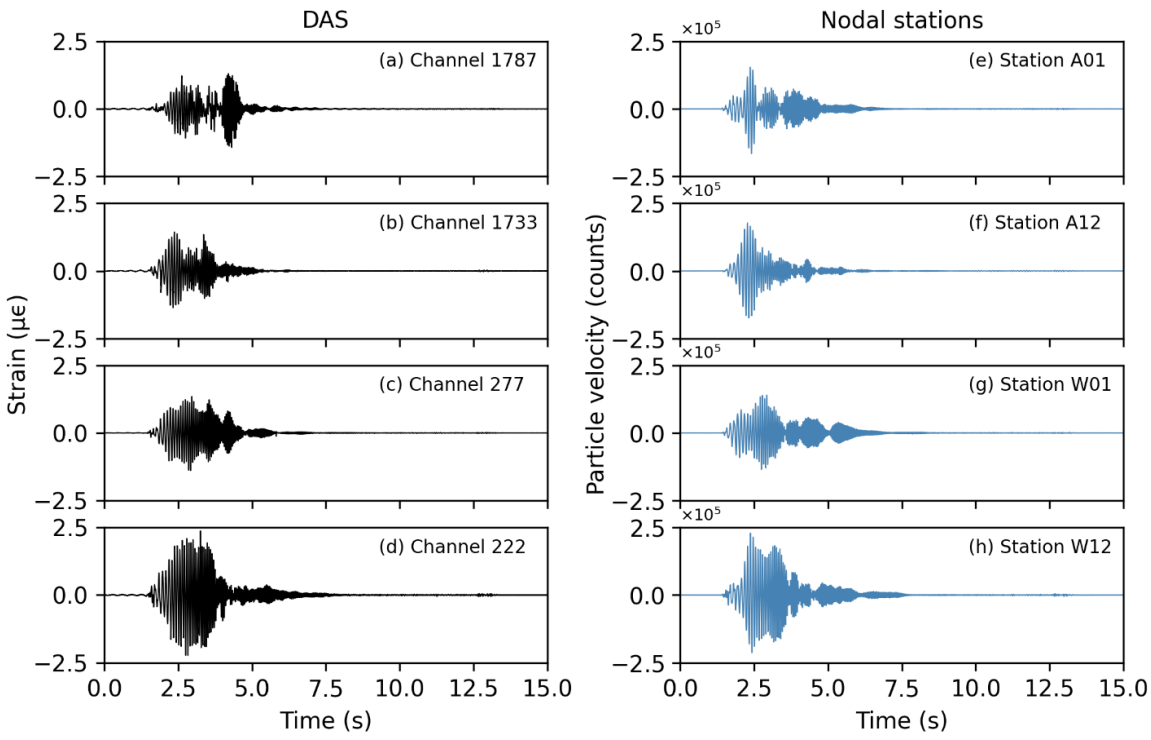

Figure 4-7 The waveforms generated by T-Rex shot X_SIL07 (refer to Figure 2), as captured by both the DAS channels and the DHE component of nodal stations positioned at the four furthest corners of the instrumented area (i.e., channels 1787, 1733, 277, and 222, and nodal stations A01, A12, W01, and W12).

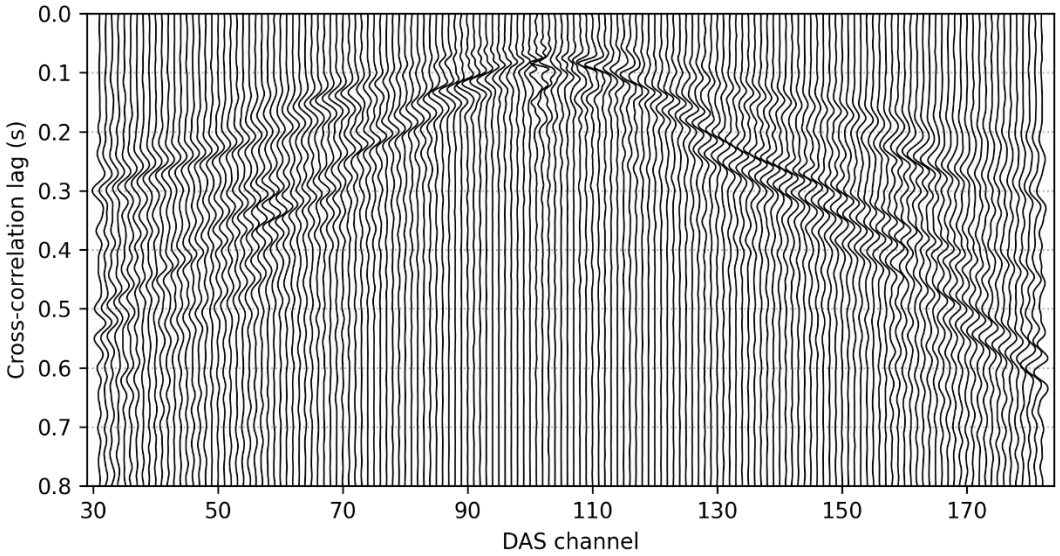

Figure 4-8 Waveforms recorded by DAS channels 31-183 along Line 104 (refer to Figure 4-2) after cross correlating with the T-Rex pilot signal for T-Rex shot Y_SIL12, and normalizing each waveform by its absolute maximum value.

The numerous shot locations and spatially-dense DAS channels and nodal stations provide ample opportunities to visualize wave propagation patterns in an attempt to locate anomalies prior to more rigorous processing, such as that used for psedu-2D MASW or 2D/3D FWI. One way to visualize the data is by observing the wave propagation as a function of time across an entire sensing array (i.e., either the DAS or the nodal stations) initiated by a T-Rex shot at one location. Alternatively, the wavefield recorded by a section of the sensing array, like a DAS line, generated by T-Rex shaking at multiple shot locations can be used to create wavefield animations. Figure 4-9 illustrates an example of the latter, where the cross-correlated waveforms generated by shots Z_SOW26, Z_SOW31, and Z_SOW37 and recorded by DAS channels 1810 through 1890 (refer to Figure 4-2) are shown, along with the location of a known surface void. In these waveforms, the peaks and troughs are filled with red and blue shading, respectively, to help better visualize disruptions to the wave polarities and wave propagation directions. Some of the backscatter events at the location of a known surface void can be observed in Figures 4-9a and 4-9c for shots Z_SOW26 and Z_SOW37, respectively, as highlighted by the dashed blue and red lines. Additionally, there is a noticeable increase in the amplitude of the late-arriving waves at the void location, as indicated by the intensity of the color and circumscribed by the dashed ellipses in Figures 4-9a and 4-9c. These observations are consistent with the outcomes of a synthetic study conducted by (Rector et al., 2015) that investigated a wavefield produced by an active source and measured by geophones situated on the surface directly above underground voids. Moreover, significant backscatter is evident between channels 1830 and 1840 for shots Z_SOW26 and Z_SOW31 (Figures 4-9a and 4-9b), which may indicate the presence of an underground anomaly, although no surface voids

were observed at that location. The low signal strength in this zone could be caused by the presence of an anomaly and/or poor cable coupling. The waveforms from this zone need to be examined from other shot locations in order to conclusively determine the cause(s).

Regarding the potential to use the dataset for 1D and/or pseudo-2D surface wave imaging; Figure 4-10 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 4-10a. The dispersion images displayed in Figure 4-10 were generated using the frequency-domain beamformer (FDBF) 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). Figures 4-10b and 4-10c display dispersion images derived from channels DHN (crossline) and DHZ (vertical) in nodal stations Q01 through Q12 for shots Y_SOW40 and Z_SOW40, respectively, conforming to the customary multi-channel analysis of surface waves (MASW) configuration used for processing Love and Rayleigh waves with nodal stations, 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 Figures 4-10b and 4-10c, respectively, along with some potentially 1st-higher mode Rayleigh wave (R1?) trends.

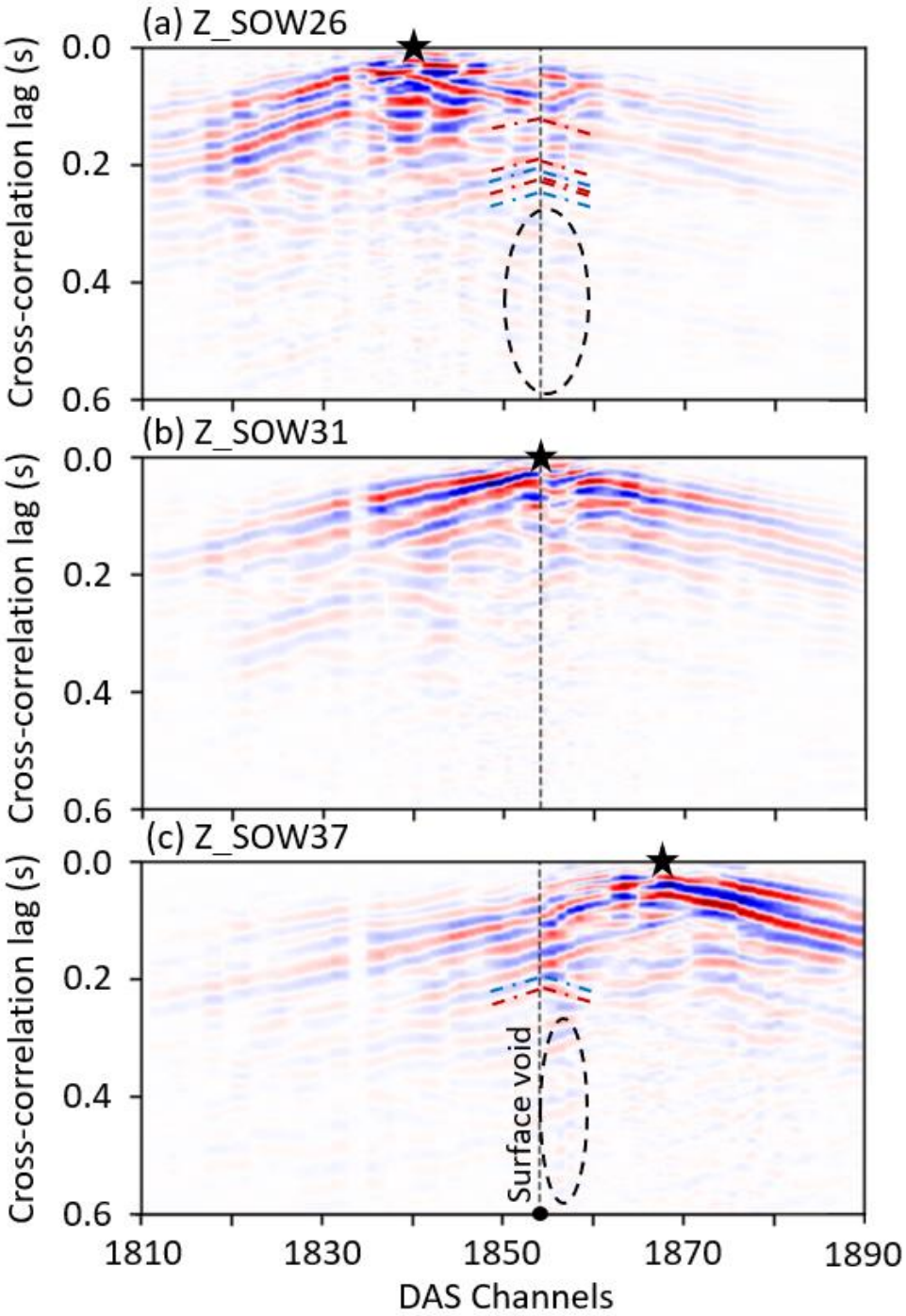

Figure 4-9 Waveforms recorded by DAS channels 1810 through 1880 for shots Z_SOW26, Z_SOW31, and Z_SWO37 (refer to Figure 4-2) in Panels (a), (b), and (c), respectively. A surface void location is indicated by a black dashed line in all panels, with backscatter evident at its location, highlighted by dashed blue and red lines. The dashed ellipses in panels (a) and (c) circumscribe the relatively higher amplitude of the late-arriving waves at the void location compared to other locations in the wavefield at similar time lags, as indicated by the intensity of the color.

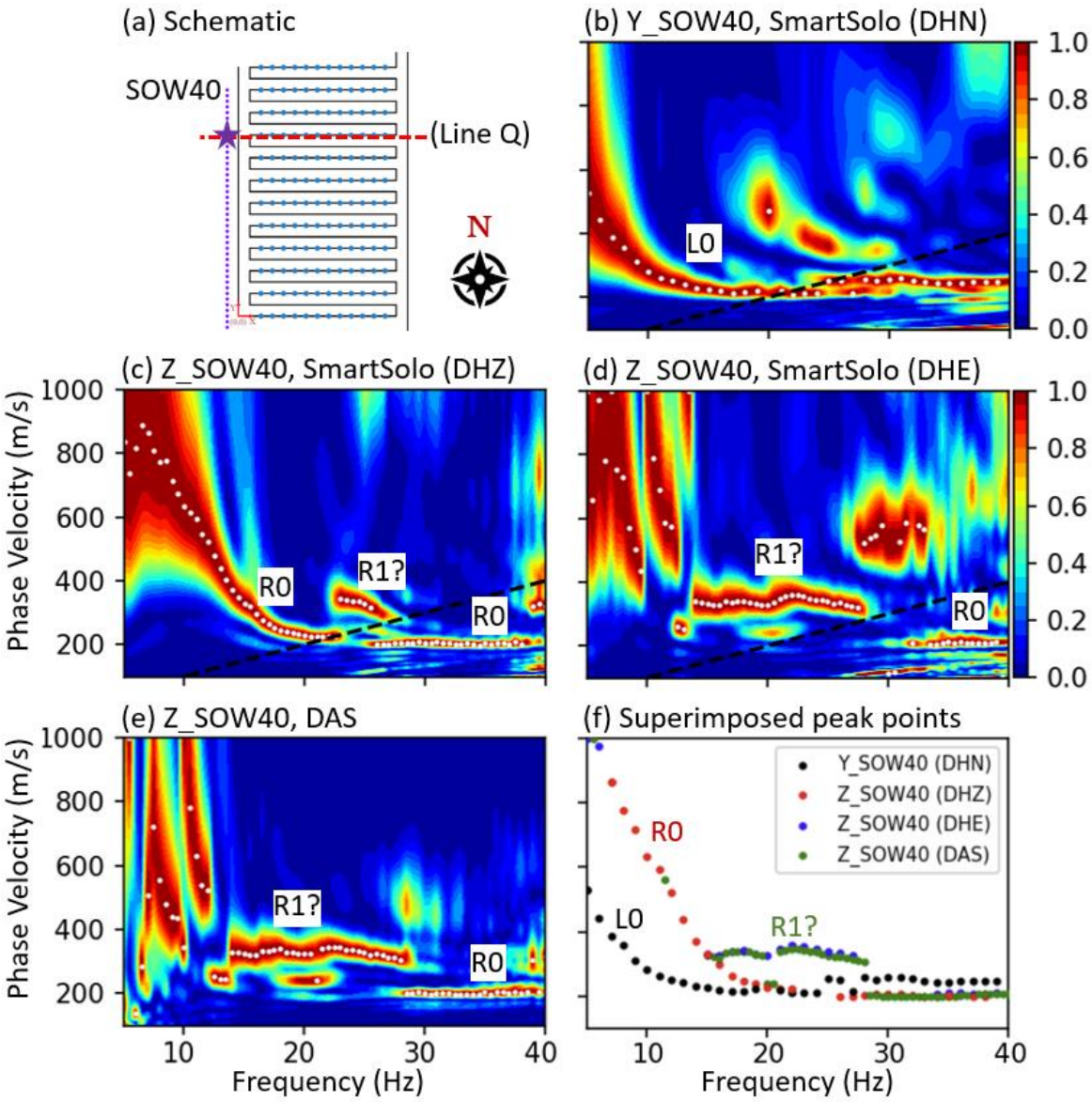

Figure 4-10 Dispersion images obtained from DAS and nodal stations along line Q due to T-Rex shaking at location SOW40. Panel (a) presents a schematic map highlighting the locations of shot location SOW40 and line Q. Panels (b), (c), and (d) showcase dispersion images from nodal stations Q01 through 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 to 693 located at line Q, for shot Z_SOW40. Panel (f) shows the superimposed peak power points from panels (b), (c), (d) and (e).

While Figure 4-10c 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 particle motion (Vantassel et al., 2022). Figures 4-10d and 4-10e present dispersion images obtained from the DHE (inline) channels in nodal stations Q01 through Q12 and DAS channels 629 through 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 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 panels 10d and 10e, respectively, the resemblance between the resulting dispersion images is evident, which agrees with the observations of Vantassel et al. (2022). Nonetheless, the DAS line offers a significant advantage in such a spatially variable site, as it enables 2D MASW-type processing (M. 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 4-10c to the ones in Figures 4-10d and 4-10e, 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 4-10f, which compares the dispersion data obtained from combining the peak power trends from all four

dispersion images in Figure 4-10b – 4-10e. This combined approach enables a more clear 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 (Soomro et al., 2016). In summary, the multi-direction shaking and multi-component sensing allows for more robust surface wave dispersion processing.

The dataset 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., Abbas et al., 2023; Crocker et al., 2023; Vantassel et al., 2022). Vantassel et al. (2022) trained a CNN to take a wavefields inputs and generate 2D V_S images of the near surface, while Abbas 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 4-7 through 4-10 demonstrate the potential of using the dataset 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 3-component (3C) noise data could be utilized in 2D microtremor array measurements (e.g., Wathelet et al., 2018) and 2D/3D ambient noise tomography (Wang et al., 2021, 2023). Additionally, the horizontal-to-vertical spectral ratio (HVSR) measurement technique has been demonstrated to provide valuable information about the subsurface, as discussed in the following section.

4.12.2 Horizontal-to-Vertical Spectral Ratio (HVSR)

HVSR can be used to infer the spatial variability of fundamental site period (T_0) at each 3C nodal station. The fundamental site period can reveal important information about the location/depth of strong impedance contrasts beneath each nodal station (Bard & SESAME Team, 2004). For example, stations with higher T_0 values (or lower fundamental frequency, f_0) are expected to have deeper impedance contrasts, while stations with lower T_0 values (or higher f_0) are expected to have more shallow 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 the 14 May at 4:00 UTC (i.e., files N_20220514040000_*station location*_3C). The data were processed with the open-source Python package hvsrpy (Vantassel, 2021). The one-hour long recording for each station was divided into 30, 120-second-long time windows and the horizontal components were combined using the geometric-mean, as recommended by Cox et al., (2020) Further, smoothing was performed using the filter proposed by (Konno & Ohmachi, 1998) with b=40. A color-mapped representation of the 144 stations' fundamental frequency from the HVSR median curve ($f_{0,mc}$) can be found in Figure 4-11, indicating a notable fluctuation in $f_{0,mc}$ throughout the site, with a range of values spanning from 4.19 to 7.03 Hz. HVSR amplitude with frequency plots for stations A01, M01, and W01 are also shown in Figure 4-11. Higher $f_{0,mc}$ values are indicative of shallower depths to limestone, while lower $f_{0,mc}$ 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) and (2020). The HVSR data

could be processed in more rigorous ways to extract more qualitative estimates for the depth to bedrock (e.g., Bignardi et al., 2016; Hobiger et al., 2009; Scherbaum et al., 2003).

4.12.3 DAS reception patterns

Using one-dimensional strain measurements to detect stress waves can greatly impact the waves measured phase and magnitude, owing to their directional sensitivity. The sensitivity of a DAS array to stress 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 stress 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 4-12a 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. The ε_{xx} values shown in Figure 4-12a resemble those that a horizontally placed DAS cable oriented in the X direction would measure. In this representation, red indicates tension while blue represents compression. Notably, a distinct change in wavefield polarity is evident between the left and right sides of the shot location in Figure 4-12a, with a zone of zero sensitivity directly above and below the shot location (i.e., at 90 and 270 degrees from the zero X axis). The dataset 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 DAS array.

Figure 4-12b displays a snapshot of the waveforms recorded on DAS channels 218 through 1791 at 1.9 seconds into shot X_SIL07 (refer to Figure 4-2). In this figure, a clear polarity flip can be observed between the right and left 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 right of the shot location by multiplying their values by -1. Figure 4-12c shows the results from this reversal of the DAS polarity on the channels to the right of the source. After flipping the polarity of the channels to the right of the source location, the two dotted circles in Figure 4-12c mainly encompass a tension wave propagating away from the source, as inferred by the red cable color. By reproducing Hubbard et al.'s (2022) numerical simulations with real-field data, Figure 4-12b underscores the potential of the dataset for investigating DAS reception patterns and any potential effects of underground anomalies on them.

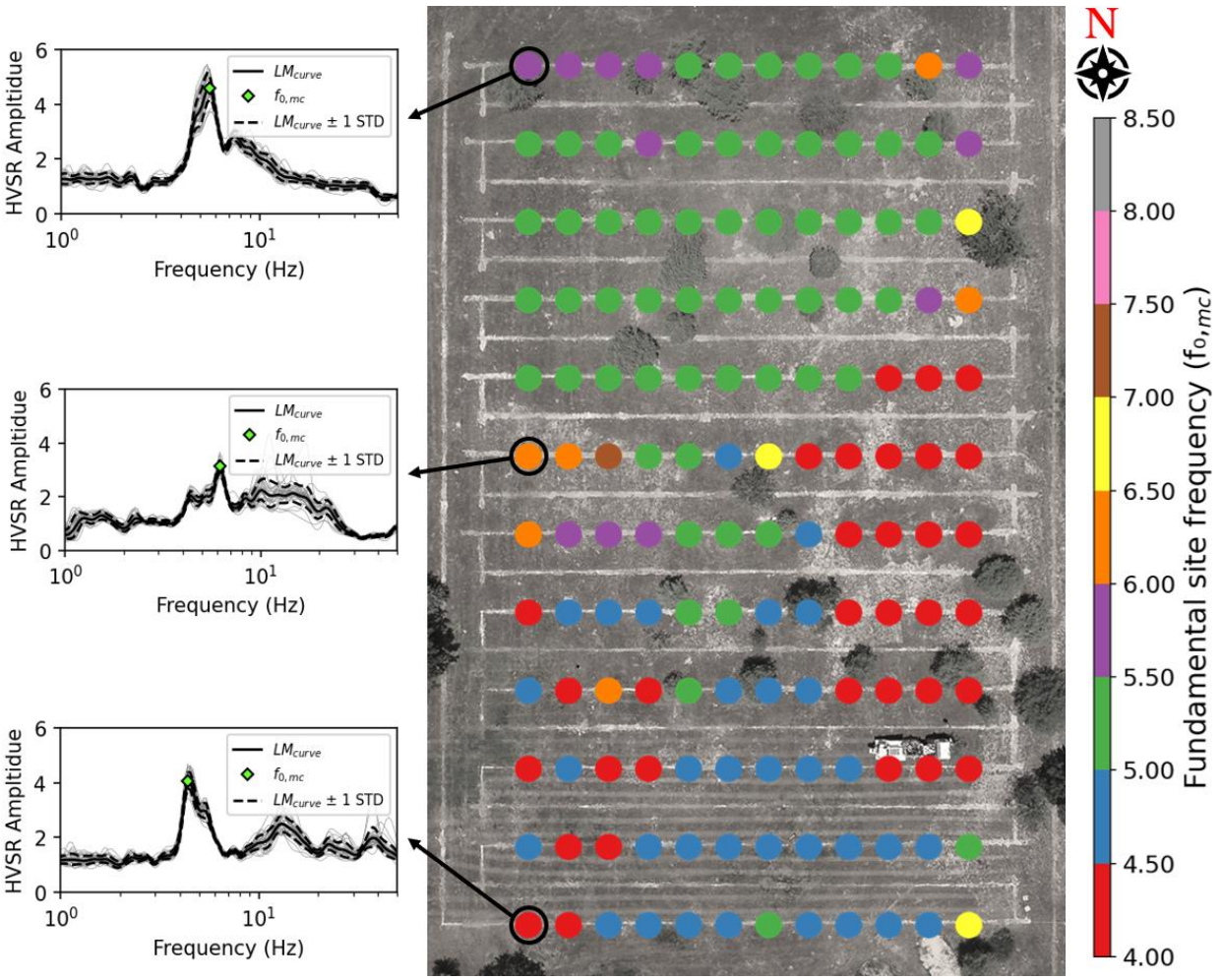

Figure 4-11 Spatial distribution of the fundamental site frequency (f_0), as determined by the peak of the lognormal median curve ($f_{0,mc}$), obtained from the Horizontal-to-vertical spectral ratio (HVSR) analysis of one hour of ambient noise data collected at each nodal station. Detailed HVSR plots 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 ($f_{0,mc}$).

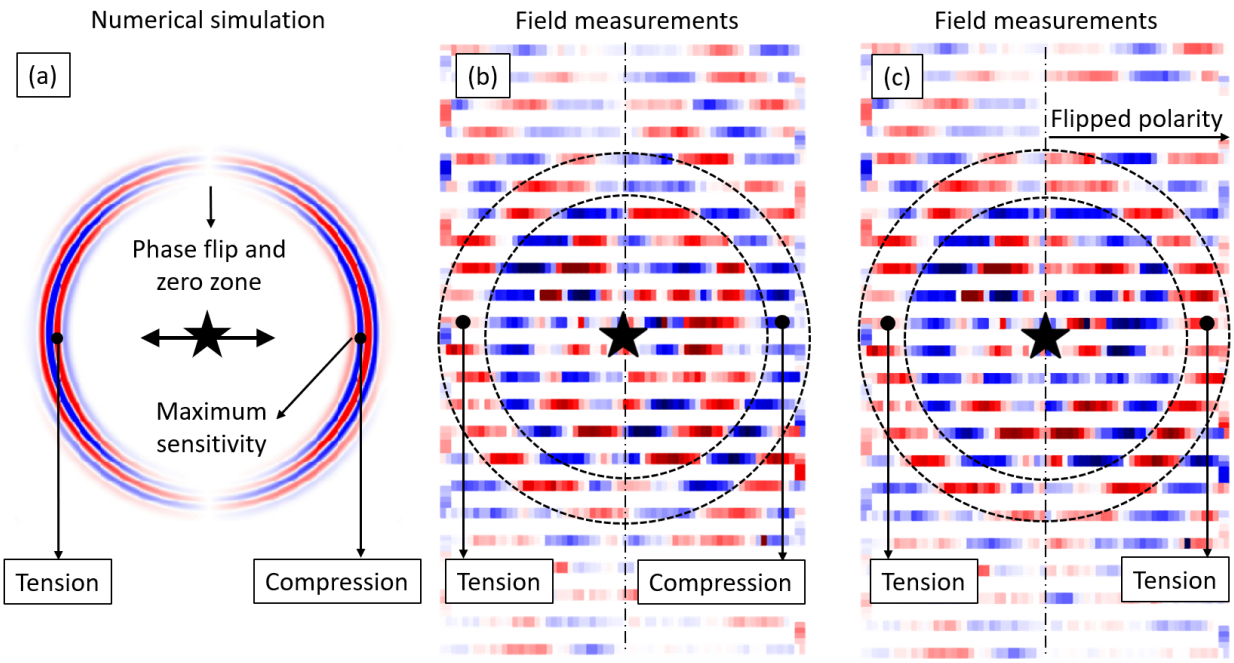

Figure 4-12 Reception patterns of DAS due to a horizontal excitation at the ground surface, as obtained from a numerical simulation in Panel (a) and field measurements in Panels (b) and (c). Panel (a) displays the numerical x-direction surface strain (ε_{xx}) results, with compression shown in blue and 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 through 1791, 1.9 seconds 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 left and right 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 right of the shot location have been flipped.

4.13 Conclusions

This research paper outlines a comprehensive subsurface imaging experiment in Newberry, Florida using stress waves. The site is spatially variable and contains karstic surface and underground voids and anomalies, which have been documented in the literature and indicated through preliminary processing of the collected data. The sensing technologies used comprised a dense 2D array of 1920 DAS channels and a 12 x 12 grid of 144 SmartSolo 3C nodal stations, which covered an area of 155 m x 75 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, namely: a powerful three-dimensional vibroseis shaker truck, a 40-kg propelled energy generator (PEG-40kg), and an 8-lb sledgehammer. The vibroseis shaker truck was used to vibrate the ground in the three directions at 260 locations inside and outside the instrumented area, while 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 the DAS, while the nodal stations recorded 48 hours of ambient noise in four 12-hour increments over a period of four days. The waveforms obtained from the 1920 DAS channels for every active-source shot or passive-wavefield time block were extracted, processed, and stored in H5 files. These files can be easily visualized using a Python script incorporated with the open-access dataset. Additionally, the three-component data gathered from each SmartSolo nodal station were consolidated into a single miniSEED file, and the data from all 144 nodal stations obtained during each active-source shot or passive-wavefield time block were extracted and saved into a separate folder. To enable efficient retrieval of all necessary information, the dataset was systematically organized into three parent folders - raw data, processed data, and supporting documents - with a consistent naming convention employed for all files and folders. The raw and processed dataset, along with complete and detailed documentation of the experiment, have been archived and made publicly available on DesignSafe. We anticipate that this dataset will be a valuable resource for researchers developing techniques for void and anomaly detection using non-invasive, stress wave-based subsurface imaging techniques. We have provided examples of the data and potential use cases as a means to

inspire present and future researchers who need a high-quality experimental dataset with known and unknown anomaly locations for testing imaging methods.

4.14 Acknowledgements

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.

4.15 References

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

Bard, & SESAME Team. (2004). *Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations measurements, processing and interpretation*.

Belfer, I., Bruner, I., Keydar, S., Kravtsov, A., & Landa, E. (1998). Detection of shallow objects using refracted and diffracted seismic waves. In *Journal of Applied Geophysics* (Vol. 38).

Bignardi, S., Mantovani, A., & Abu Zeid, N. (2016). OpenHVSR: imaging the subsurface 2D/3D elastic properties through multiple HVSR modeling and inversion. *Computers & Geosciences*, *93*, 103–113. https://doi.org/10.1016/j.cageo.2016.05.009

Branham, K., & Steeples, D. (1988). Cavity detection using high-resolution seismic reflection methods. *Min. Eng*, *40*(2). https://www.osti.gov/biblio/5243315

Castongia, E., Wang, H. F., Lord, N., Fratta, D., Mondanos, M., & Chalari, A. (2017). An Experimental Investigation of Distributed Acoustic Sensing (DAS) on Lake Ice. *Journal of Environmental and Engineering Geophysics*, *22*(2), 167–176. https://doi.org/10.2113/JEEG22.2.167

Cheng, T., Cox, B. R., Vantassel, J. P., & Manuel, L. (2020). A statistical approach to account for azimuthal variability in single-station HVSR measurements. *Geophysical Journal International*, *223*(2), 1040–1053. https://doi.org/10.1093/gji/ggaa342

Cheng, T., Hallal, M. M., Vantassel, J. P., & Cox, B. R. (2021). Estimating Unbiased Statistics for Fundamental Site Frequency Using Spatially Distributed HVSR Measurements and Voronoi Tessellation. *Journal of Geotechnical and Geoenvironmental Engineering*, *147*(8). https://doi.org/10.1061/(ASCE)GT.1943-5606.0002551

Cook, J. C. (1965). SEISMIC MAPPING OF UNDERGROUND CAVITIES USING REFLECTION AMPLITUDES. *GEOPHYSICS*, *30*(4), 527–538. https://doi.org/10.1190/1.1439618

Cox, B. R., Cheng, T., Vantassel, J. P., & Manuel, L. (2020). A statistical representation and frequency-domain window-rejection algorithm for single-station HVSR measurements. *Geophysical Journal International*, *221*(3), 2170–2183. https://doi.org/10.1093/gji/ggaa119

Cox, B., Wills, P., Kiyashchenko, D., Mestayer, J., Lopez, J., Bourne, S., Lupton, R., Solano, G., Henderson, N., Hill, D., & Roy, J. (2012). *Distributed Acoustic Sensing for Geophysical Measurement, Monitoring and Verification*.

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

Daley, T. M., Miller, D. E., Dodds, K., Cook, P., & Freifeld, B. M. (2016). Field testing of modular borehole monitoring with simultaneous distributed acoustic sensing and geophone vertical seismic profiles at Citronelle, Alabama. *Geophysical Prospecting*, *64*(5), 1318–1334. https://doi.org/10.1111/1365-2478.12324

Dou, S., Lindsey, N., Wagner, A. M., Daley, T. M., Freifeld, B., Robertson, M., Peterson, J., Ulrich, C., Martin, E. R., & Ajo-Franklin, J. B. (2017). Distributed Acoustic Sensing for Seismic Monitoring of the Near Surface: A Traffic-Noise Interferometry Case Study. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-11986-4

Fathi, A., Poursartip, B., Stokoe II, K. H., & Kallivokas, L. F. (2016). Three-dimensional P- and S-wave velocity profiling of geotechnical sites using full-waveform inversion driven by field data. *Soil Dynamics and Earthquake Engineering*, *87*, 63–81. https://doi.org/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*. https://www.fdsn.org/pdf/SEEDManual_V2.4.pdf

Foti, S., Hollender, F., Garofalo, F., Albarello, D., Asten, M., Bard, P.-Y., Comina, C., Cornou, C., Cox, B., Di Giulio, G., Forbriger, T., Hayashi, K., Lunedei, E., Martin, A., Mercerat, D., Ohrnberger, M., Poggi, V., Renalier, F., Sicilia, D., & Socco, V. (2018). Guidelines for the good practice of surface wave analysis: a product of the InterPACIFIC project. *Bulletin of Earthquake Engineering*, *16*(6), 2367–2420. https://doi.org/10.1007/s10518-017-0206-7

Grandjean, G., & Leparoux, D. (2004). The potential of seismic methods for detecting cavities and buried objects: experimentation at a test site. *Journal of Applied Geophysics*, *56*(2), 93–106. https://doi.org/10.1016/j.jappgeo.2004.04.004

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

Hobiger, M., Bard, P.-Y., Cornou, C., & Le Bihan, N. (2009). Single station determination of Rayleigh wave ellipticity by using the random decrement technique (RayDec). *Geophysical Research Letters*, *36*(14), L14303. https://doi.org/10.1029/2009GL038863

Hubbard, P. G., Vantassel, J. P., Cox, B. R., Rector, J. W., Yust, M. B. S., & Soga, K. (2022). Quantifying the Surface Strain Field Induced by Active Sources with Distributed Acoustic Sensing: Theory and Practice. *Sensors*, *22*(12). https://doi.org/10.3390/s22124589

Kolesnikov, Yu. I., & Fedin, K. V. (2018). Detecting underground cavities using microtremor data: physical modelling and field experiment. *Geophysical Prospecting*, *66*(2), 342–353. https://doi.org/10.1111/1365-2478.12540

Konno, K., & Ohmachi, T. (1998). Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor. *Bulletin of the Seismological Society of America*, *88*(1), 228–241. https://doi.org/10.1785/BSSA0880010228

Kristekova, M., Kristek, J., Moczo, P., & Labak, P. (2020). The Finite-interval Spectral Power method for detecting underground cavities using seismic ambient noise. *Geophysical Journal International*, *224*(2), 945–960. https://doi.org/10.1093/gji/ggaa494

Lancelle, C., Lord, N., Wang, H., Fratta, D., Nigbor, R., Chalari, A., Karaulanov, R., Baldwin, J., & Castongia, E. (2014). *Sample Data from a Distributed Acoustic Sensing Experiment at Garner Valley, California*. AGU Fall Meeting. https://dx.doi.org/10.15121/1177104

Martin, E. R., Lindsey, N. J., Ajo-Franklin, J. B., & Biondi, B. L. (2021). *Introduction to Interferometry of Fiber-Optic Strain Measurements* (pp. 111–129). https://doi.org/10.1002/9781119521808.ch9

Nakamura, Y. (1989). A method for dynamic characteristics estimation of subsurface using microtremor on the ground surface. *Quarterly Report of RTRI*, *30*(1).

Obermann, A., Sánchez-Pastor, P., Wu, S.-M., Wollin, C., Baird, A. F., Isken, M. P., Clinton, J., Goertz-Allmann, B. P., Dahm, T., Wuestefeld, A., Shi, P., Lanza, F., Gyger, L., Wetter, S., Hjörleifsdóttir, V., Langet, N., Brynjarsson, B., Jousset, P., & Wiemer, S. (2022). Combined Large- *N* Seismic Arrays and DAS Fiber Optic Cables across the Hengill Geothermal Field, Iceland. *Seismological Research Letters*, *93*(5), 2498–2514. https://doi.org/10.1785/0220220073

Pan, Y., Xia, J., Xu, Y., Gao, L., & Xu, Z. (2016). Love-wave waveform inversion in time domain for shallow shear-wave velocity. *GEOPHYSICS*, *81*(1), R1–R14. https://doi.org/10.1190/geo2014-0225.1

Pernod P. and Piwakowski, B. and D. B. and T. J. C. (1989). Detection of Shallow Underground Cavities by Seismic Methods: Physical Modelling Approach. In N. and K. J. Shimizu Hiroshi and Chubachi (Ed.), *Acoustical Imaging* (pp. 705–713). Springer US. https://doi.org/10.1007/978-1-4613-0791-4_74

Rathje, E. M., Dawson, C., Padgett, J. E., Pinelli, J.-P., Stanzione, D., Adair, A., Arduino, P., Brandenberg, S. J., Cockerill, T., Dey, C., Esteva, M., Haan, F. L., Hanlon, M., Kareem, A., Lowes, L., Mock, S., & Mosqueda, G. (2017). DesignSafe: New Cyberinfrastructure for Natural Hazards Engineering. *Natural Hazards Review*, *18*(3). https://doi.org/10.1061/(asce)nh.1527-6996.0000246

Rector, J. W., Pfeiffe, J., Hodges, S., Kingman, J., & Sprott, E. (2015). Tomographic imaging of surface waves: A case study from the Phoenix Mine, Battle Mountain, Nevada. *The Leading Edge*, *34*(11), 1360–1364. https://doi.org/10.1190/tle34111360.1

Scherbaum, F., Hinzen, K.-G., & Ohrnberger, M. (2003). Determination of shallow shear wave velocity profiles in the Cologne, Germany area using ambient vibrations. *Geophysical Journal International*, *152*(3), 597–612. https://doi.org/10.1046/j.1365-246X.2003.01856.x

Sheriff, R. E., & Geldart, L. P. (1995). *Exploration Seismology*. Cambridge University Press. https://doi.org/10.1017/CBO9781139168359

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

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

Smith, J. A., Borisov, D., Cudney, H., Miller, R. D., Modrak, R., Moran, M., Peterie, S. L., Sloan, S. D., Tromp, J., & Wang, Y. (2019). Tunnel detection at Yuma Proving Ground, Arizona, USA — Part 2: 3D full-waveform inversion experiments. *GEOPHYSICS*, *84*(1), B107–B120. https://doi.org/10.1190/geo2018-0599.1

Soga, K., & Luo, L. (2018). Distributed fiber optics sensors for civil engineering infrastructure sensing. *Journal of Structural Integrity and Maintenance*, *3*(1), 1–21. https://doi.org/10.1080/24705314.2018.1426138

Soomro, R. A., Weidle, C., Cristiano, L., Lebedev, S., & Meier, T. (2016). Phase velocities of Rayleigh and Love waves in central and northern Europe from automated, broad-band, interstation measurements. *Geophysical Journal International*, *204*(1), 517–534. https://doi.org/10.1093/gji/ggv462

Stokoe, K. H., Cox, B. R., Clayton, P. M., & Menq, F. (2020). NHERI@UTexas Experimental Facility With Large-Scale Mobile Shakers for Field Studies. *Frontiers in Built Environment*, *6*. https://doi.org/10.3389/fbuil.2020.575973

Tran, K. T., & Hiltunen, D. R. (2011). Inversion of First-arrival Time Using Simulated Annealing. *Journal of Environmental and Engineering Geophysics*, *16*(1), 25–35. https://doi.org/10.2113/JEEG16.1.25

Tran, K. T., McVay, M., Faraone, M., & Horhota, D. (2013). Sinkhole detection using 2D full seismic waveform tomography. *GEOPHYSICS*, *78*(5), R175–R183. https://doi.org/10.1190/geo2013-0063.1

Tran, K. T., Nguyen, T. D., Hiltunen, D. R., Stokoe, K., & Menq, F. (2020). 3D full-waveform inversion in time-frequency domain: Field data application. *Journal of Applied Geophysics*, *178*, 104078. https://doi.org/10.1016/j.jappgeo.2020.104078

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

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

Vantassel, J. P., & Cox, B. R. (2022). SWprocess: a workflow for developing robust estimates of surface wave dispersion uncertainty. *Journal of Seismology*, *26*(4), 731–756. https://doi.org/10.1007/s10950-021-10035-y

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

Vantassel, J. P., Kumar, K., & Cox, B. R. (2022). Using convolutional neural networks to develop starting models for near-surface 2-D full waveform inversion. *Geophysical Journal International*, *231*(1), 72–90. https://doi.org/10.1093/gji/ggac179

Wang, Y., Khorrami, M., Tran, K. T., & Horhota, D. (2023). Application of ambient noise tomography for deep void detection. *Journal of Applied Geophysics*, *209*, 104922. https://doi.org/10.1016/j.jappgeo.2022.104922

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

Wang, Y., Tran, K. T., & Horhota, D. (2021). Road sinkhole detection with 2D ambient noise tomography. *GEOPHYSICS*, *86*(6), KS123–KS135. https://doi.org/10.1190/geo2020-0739.1

Wathelet, M., Guillier, B., Roux, P., Cornou, C., & Ohrnberger, M. (2018). Rayleigh wave three-component beamforming: signed ellipticity assessment from high-resolution frequency-wavenumber processing of ambient vibration arrays. *Geophysical Journal International*, *215*(1), 507–523. https://doi.org/10.1093/gji/ggy286

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. *Geophysical Research Letters*, *46*(3), 1320–1328. https://doi.org/10.1029/2018GL081195

Yust, M. B. S., Cox, B. R., Vantassel, J. P., Hubbard, P. G., Boehm, C., & Krischer, L. (2023). Near-Surface 2D Imaging via FWI of DAS Data: An Examination on the Impacts of FWI Starting Model. *Geosciences*, *13*(3), 63. https://doi.org/10.3390/geosciences13030063

Yust, M., Cox, B. R., Vantassel, J. P., & Hubbard, P. G. (2022). *DAS for 2D MASW Imaging: A Case Study on the Benefits of Flexible Sub-Array Processing*. https://doi.org/10.48550/arXiv.2210.14261

CHAPTER 5

EMERGING TECHNOLOGIES AND ADVANCED ANALYSES FOR NON-INVASIVE NEAR-SURFACE SITE CHARACTERIZATION

Abstract

The in-situ small-strain shear modulus of soil and rock materials is a parameter of paramount importance in geotechnical modeling. It can be derived from non-invasive geophysical surveys, which provide the possibility of testing the subsurface in its natural and undisturbed condition by inferring the velocity of propagation of shear waves. In addition, for soil dynamics and earthquake engineering applications, the small-strain damping ratio plays a relevant role, yet its estimation is still challenging, lacking consolidated approaches for its in-situ evaluation. Recent advancements in instrumentation, such as distributed acoustic sensing (DAS), combined with advanced analysis methodologies for the interpretation of stress wave propagation (e.g., machine learning and full waveform inversion), open new frontiers in site characterization. This paper presents and compares some advanced applications of measuring 1D and 2D variations in shear wave velocity and attenuation in-situ with reference to a specific case history.

Keywords: Shear wave velocity; Rayleigh waves; DAS; Damping; FWI; Machine learning

5.1 Introduction

Non-invasive subsurface imaging techniques utilizing stress wave propagation have garnered escalating attention in recent decades owing to their remarkable cost-effectiveness compared to conventional invasive site characterization methods and their

potential to cover large areas. These imaging techniques primarily focus on capturing two crucial soil parameters of particular interest in geotechnical engineering: the small-strain shear wave velocity (Vs) and the small-strain damping ratio in shear (DS). Vs is directly related to the small-strain shear modulus (G_0 or G_{max}), representing the stiffness of the soil, while DS quantifies the soil internal energy dissipation at low strains. This paper provides a review of some of the latest advancements in non-invasive subsurface imaging techniques for the estimation of Vs and DS and their practical application at a well-characterized case history site called the Hornsby Bend site in Austin, Texas, USA. The subsequent paragraphs highlight the importance of Vs and DS in geotechnical engineering applications and discuss the challenges and advancements in the non-invasive techniques developed for their estimation.

Vs and Ds play a key role in evaluating the response of soil deposits to both general dynamic loading and ground motion amplification caused by earthquakes. In regard to seismic loads, Vs and Ds are especially important parameters to quantify when the soil is subjected to low-intensity shaking (e.g., Tao & Rathje, 2019; Rodriguez-Marek et al. 2021; Fernandes et al. 2023). For instance, Rodriguez-Marek et al. (2021) observed that Ds is the most influential parameter at high frequencies, with an impact even more relevant than Vs, whereas the low-frequency soil response is mainly affected by Vs of shallow layers. According to Foti et al. (2021), Ds has a substantial influence on the seismic amplification in deformable soil deposits. However, this influence is less pronounced under conditions of strong shaking that strain the soil sufficiently to induce nonlinear soil behavior. Additionally, in a site-specific study, Foti et al. (2021) compared the amplification resulting from the epistemic uncertainty in Ds with that caused by Vs and the nonlinear soil behavior

modeled using modulus reduction and damping (MRD) curves. They found that a change in Ds leads to a significant variation in amplification compared to the overall variability of the results. This effect is particularly relevant at high frequencies and near the resonance peak, even under higher seismicity conditions.

Furthermore, studies such as Kouroussis et al. (2011), Papadopoulos et al. (2019), and Dos Santos et al. (2016) highlight the importance of Vs and Ds in assessing vibrational impact. The small-strain and anelastic properties of the soil significantly influence the energy transmission from the source, its propagation, and the resulting motion at the receiver. These properties directly impact the amplitude and frequency content of the vibrations. For instance, Lombaert & Degrande (2003) and Lombaert et al. (2006) observed that when dealing with rail traffic as the noise source, uncertainties in defining the spatial variation of dynamic soil characteristics lead to poor agreement between simulated and experimental data. Rail vehicles mainly generate high-frequency signals, reaching up to 200 Hz (Pyl, 2004). Due to their short wavelengths, these signals are highly sensitive to local heterogeneities in the soil deposit. Moreover, Schevenels (2007) demonstrated the impact of uncertainties in Ds and Vs on free-field wave propagation, whose variability exponentially increases with the frequency, especially at large distances from the source.

The small-strain dynamic soil properties also play an important role in soil-structure interaction problems, where the deformability of the supporting soil impacts both the fundamental period and the energy dissipation of the system (Veletsos & Meek, 1974). Among these properties, Vs holds particular significance as it directly governs soil deformability, which is a key factor in this phenomenon (Veletsos & Meek, 1974). Energy dissipation, on the other hand, arises from various sources, including inelastic phenomena

within the structure, hysteretic dissipation in the soil deposit, and radiation damping. Radiation damping refers to the geometric effect of waves radiating from the foundation-soil interface, carrying energy away from the foundation system as they propagate outward. It is commonly assumed that radiation damping represents the primary mechanism for energy dissipation, especially at small strains. However, Martakis et al. (2017) observed through centrifuge tests that significant dissipation also occurs due to hysteretic effects linked to the intrinsic dissipation of the soil, even at small strains.

The preceding paragraphs underscore the importance of accurately estimating Vs and Ds for dynamic modeling of seismic and general vibration problems. Traditionally, these small-strain soil properties have been estimated through laboratory testing or empirical relationships (e.g., Darendeli, 2001; Menq, 2003; Ciancimino et al., 2020). However, their in-situ estimated values often deviate from those obtained in the laboratory. This deviation can be attributed to the disturbances that inevitably occur during the acquisition of soil samples for laboratory testing, leading mostly to a reduction in soil stiffness (e.g., Stokoe & Santamarina, 2000). Additionally, at the site scale, complex wave propagation phenomena (e.g., wave scattering) result in additional energy dissipation beyond material dissipation, which cannot be captured accurately through laboratory tests (e.g., Stewart et al., 2014; Tao & Rathje, 2019). Geophysical field measurements offer the advantage of estimating the ground response in its natural state thus mitigating the uncertainties associated with sample disturbance and scale effects often encountered when working with rock-like materials. Overall, non-invasive techniques investigate a large volume of the medium, whose size depends on the array geometry (Comina et al., 2011; Passeri, 2019), providing parameter estimates at a scale compatible with those of geotechnical systems.

Furthermore, some design criteria in geoengineering directly rely on these testing procedures. For instance, rail infrastructure design requires the train speed to be smaller than a “critical” speed, corresponding to the Rayleigh phase velocity, VR, of the underlying medium (e.g., Connolly et al., 2015). Indeed, at higher speeds, the amplitude of track vertical deflection dramatically increases (Timoshenko, 1927; Krylov, 1995; Madshus & Kaynia, 2000; Madshus et al., 2004). For this reason, the rail operational speed is often determined through a dispersion diagram, which involves the experimental VR at various frequencies as an input parameter, determined through surface wave-based geophysical techniques (Thompson, 2009). Finally, the field-based small-strain estimates of Vs and Ds can be used in conjunction with laboratory tests to map and un-normalize the nonlinear mechanical response of soil, which is most easily characterized in the lab at strains ranging from moderate to large. This approach enables the development of advanced numerical models or simplified procedures for evaluating the behavior of geotechnical systems subjected to either static or dynamic loading (e.g., settlement of shallow foundations and seismic site response) across a broad range of induced strains.

Given the importance of accurately estimating Vs and Ds in situ, the field of geophysical imaging based on stress wave propagation is continuously advancing, introducing new innovations aimed at increasing imaging resolution and reducing uncertainty. These innovations encompass improvements in both data acquisition systems (DAQ) and imaging methodologies. A notable recent development in data acquisition is the utilization of distributed acoustic sensing (DAS) for stress wave measurements. DAS offers unprecedented spatial resolutions (in the order of meters) and length scales (on the order of tens of kilometers), surpassing conventional sensing technologies (Soga & Luo, 2018).

Further details on DAS technology, which is employed as the DAQ for most of the 1D and 2D imaging techniques discussed in this paper, are provided in a separate, dedicated section later in this paper. This paper also highlights some of the significant advancements in imaging techniques, encompassing both 1D and 2D approaches. In the field of 1D imaging, two notable developments have emerged. Firstly, there is the utilization of DAS as the DAQ for 1D multichannel analysis of surface waves (MASW). Secondly, the joint estimation of phase velocity and phase attenuation data within a 1D MASW test setup has become achievable using either geophones or DAS as the DAQ. This paper demonstrates the pioneering use of DAS for jointly characterizing the stiffness and dissipative parameters of a soil deposit. In the realm of 2D imaging, the presented techniques comprise the application of 2D MASW using DAS data, the utilization of machine learning for 2D imaging, and the use of full waveform inversion (FWI) with DAS data. All of these 1D and 2D imaging techniques were successfully applied at a well-characterized case history site called the Hornsby Bend test site, providing a valuable opportunity to compare and discuss their results.

The subsequent sections of the paper are organized as follows. First, a brief overview of DAS technology is presented, highlighting its key features and capabilities. Following that, the testing conducted at the Hornsby Bend site is discussed. The paper then delves into the advancements in 1D and 2D imaging techniques in regard to measuring Vs and Ds at the Hornsby Bend site. Then, a discussion about the advantages and disadvantages of each technique is presented. Lastly, a comprehensive subsurface imaging experiment conducted at the Newberry site in Florida, USA utilizing some of the latest sensing technologies is showcased.

5.2 The DAS technology

DAS is an innovative technique that transforms fiber-optic cables into a distributed array of ground motion sensors (Cox et al., 2012; Yu et al., 2019). This rapidly evolving technology allows for the simultaneous collection of high-resolution data with small channel separations (e.g., 1-m) over long arrays spanning kilometers, making it highly suitable for near-surface imaging applications. Unlike traditional geophones that measure particle velocity ($\partial u/\partial t$) at discrete points along the acquisition array (Figure 5-1a), DAS records the spatially-averaged axial strain e(r,t) induced on the fiber-optic cable by the passing wavefield. Here, "r" represents the location coordinate and "t" represents the time instant. When properly coupled with the ground, the passage of mechanical waves generates an axial strain in the fiber-optic cable that coincides with the horizontal, in-line strain ε(r,t) in the ground. An interrogator unit (IU) reads the consequent shift in phase lag of a laser pulse traveling in the cable, induced by the variation in the length of the cable. Thus DAS measurements represent the variation in phase difference over a reference length 2g, called gauge length, around the investigated location, from which the average strain e(r,t) is derived (Figure 5-1b; Grattan & Sun, 2000). The resulting average strain at each measurement point can be linked with the displacement u(r,t), as it equals the difference of the radial displacement at two points separated by a distance equal to the gauge length (Mateeva et al. 2014; Bakku, 2015; Jousset et al. 2018; Vantassel et al. 2022a):

$$e(r,t) = \frac{1}{2g}\left[u(r+g,t) - u(r-g,t)\right] \tag{1}$$

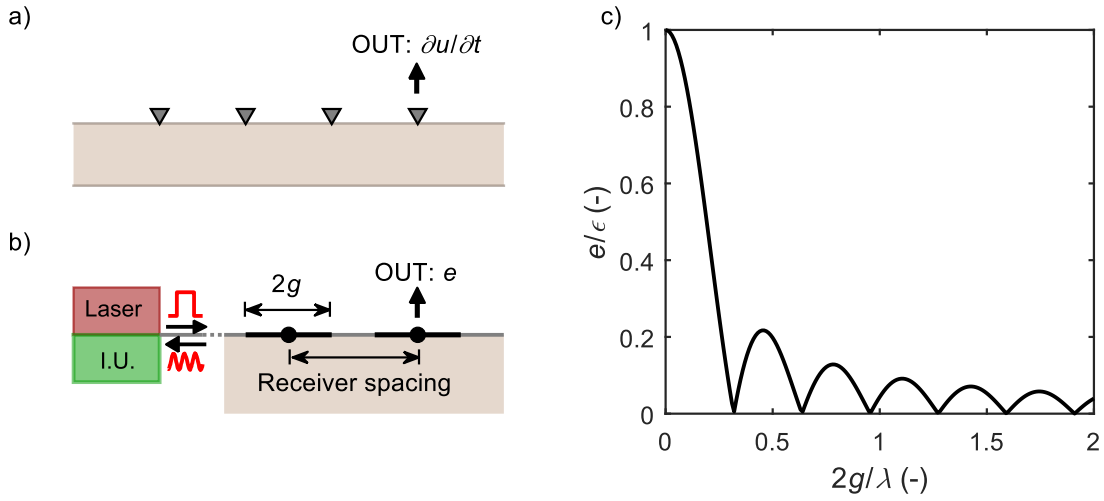

Figure 5-1 a) Schematic model of an acquisition system based on geophones, wherein the output (labeled as "OUT") is the particle velocity ∂u/∂t; b) Schematic model of the DAS system, where a source generates a laser pulse which is then interpreted by an interrogator unit (labeled as I.U.) and the output (labeled as "OUT") is the average strain e (modified from Bakku, 2015); c) Amplitude response in terms of e/ε ratio, as a function of the wavelength-normalized gauge length 2g/λ.

It is noteworthy that the gauge length is not necessarily linked with the channel separation (i.e., the distance between two subsequent measurement points). The gauge length plays a crucial role in the spatial sampling quality, as it limits the range of investigable wavelengths. In the simple scenario of a spatially harmonic radial displacement field, it is demonstrated that the averaging procedure (i.e., the mapping from ε(r,t) to e(r,t)) is equivalent to applying a lowpass filter (such as a sinc filter) in the wavenumber domain (e.g., Bakku, 2015). This filtering tends to attenuate wave components with shorter wavelengths λ, particularly affecting the high-frequency components of the Rayleigh wavefield (Figure 5-1c). Increasing the gauge length results in a greater loss of information for short-wavelength data, although it improves the overall signal quality and signal-to-noise ratio (e.g., Bakulin et al., 2020). Therefore, selecting an optimal gauge length is a critical task that should consider various factors such as the acquisition setup, source

quality, magnitude of incoherent noise, and the desired range of investigated wavelengths. A possible strategy to overcome the resolution issues induced by spatial averaging is to conduct multiple measurements where the gauge length is modified at each step (Bakku, 2015). It is worth noting that interpreting DAS-recorded data, which provides a spatially averaged measure of the strain field, is slightly more complex compared to conventional acquisition devices. Nonetheless, DAS is increasingly being used for invasive geophysical tests (e.g., Mateeva et al., 2014; Kuvshinov, 2016), ambient noise vibrations (e.g., Hornman et al., 2013; Freifeld et al., 2016; Yavuz et al., 2016; Ajo-Franklin et al., 2017) and MASW testing (Galan-Comas, 2015; Lancelle, 2016; Costley et al., 2018; Song et al., 2018 Vantassel et al. 2022).

5.3 Innovative processing techniques and DAS

5.3.1 Reference dataset: Hornsby Bend

The Hornsby Bend site (HB, 30°13.918′N, 97°38.631′W – in the WGS84 Datum) is located on the outskirts of Austin, Texas, USA. This site has been the subject of extensive invasive and non-invasive site characterization studies in recent years. The non-invasive testing conducted at the site and utilized in this paper involved the deployment of two parallel 200-meter-long fiber-optic cables (refer to Figure 5-2a), one manufactured by NanZee and the other by AFL. These cables were interconnected at the far end of the array by splicing the NanZee and AFL cables together, enabling simultaneous recording on both cables. On the near-side of the array, the NanZee cable was connected to an OptaSense ODH4 IU, while the AFL fiber was properly terminated to minimize end-reflections. The ODH4 IU was configured with the gauge length and channel separation set to 2.04 m and 1.02 m, respectively. These values represent the shortest gauge length and channel separation

allowed by the ODH4. Consequently, the DAS recorded waveforms represent an average response over the 2.04-m gauge length surrounding each channel location (i.e., every 1.02 m). The IU sampling frequency, or ping rate, was set at 100 kHz. After acquiring the data, the raw measurements underwent down sampling to 1 kHz and high-pass filtering above 3 Hz to remove low-frequency artifacts linked with laser drift and static strains.

Two geophone arrays were also deployed in conjunction with the fiber-optic cables: a vertical geophone array and a horizontal geophone array oriented in line with the DAS fiber optic cables. Each array comprised 48 geophones, uniformly spaced at 2-m intervals, resulting in a total array length of 94 m as shown in Figure 5-2a. To capture the geophone signals, four interconnected 24-channel Geometrics Geode seismographs were utilized, enabling simultaneous recording from both the vertical and horizontal geophone arrays. All signals were acquired using a sampling rate of 1 kHz. The geophone array and DAS fiber-optic cables were employed to simultaneously record actively-generated surface waves from various sources. These sources encompassed highly-controlled vibroseis shaker trucks and more variable impact sources. The vibroseis sources comprised the three-dimensional shaker, T-Rex, and the highly-mobile one-dimensional shaker, Thumper, both from the Natural Hazards Engineering Research Infrastructure at the University of Texas at Austin (NHERI@UTexas) experimental facility (Stokoe et al., 2020). Additionally, an instrumented 5.4 kg sledgehammer from PCB Piezotronics was used as an impact source. T-Rex was utilized for shaking in all three directions: vertically, horizontally in-line, and horizontally cross-line. It generated a 12-second chirp signal with frequencies linearly swept from 3 to 80 Hz, providing a maximum force output of approximately 270 kN in the vertical direction and 130 kN in the horizontal directions. Thumper was used to produce a

12-second chirp signal in the vertical direction with frequencies linearly swept from 5 to 200 Hz, offering a maximum force output of approximately 270 kN in the vertical direction. For the vibroseis sources, three sweeps were conducted at each source location, whereas five impacts were performed using the sledgehammer. These sources were used at various locations around the site, however, for the purposes of this paper, only the source locations along the linear array alignment will be discussed. A significant portion of the non-invasive dataset from testing conducted at the Hornsby Bend site is accessible to the public through DesignSafe-CI (Vantassel et al., 2022c).

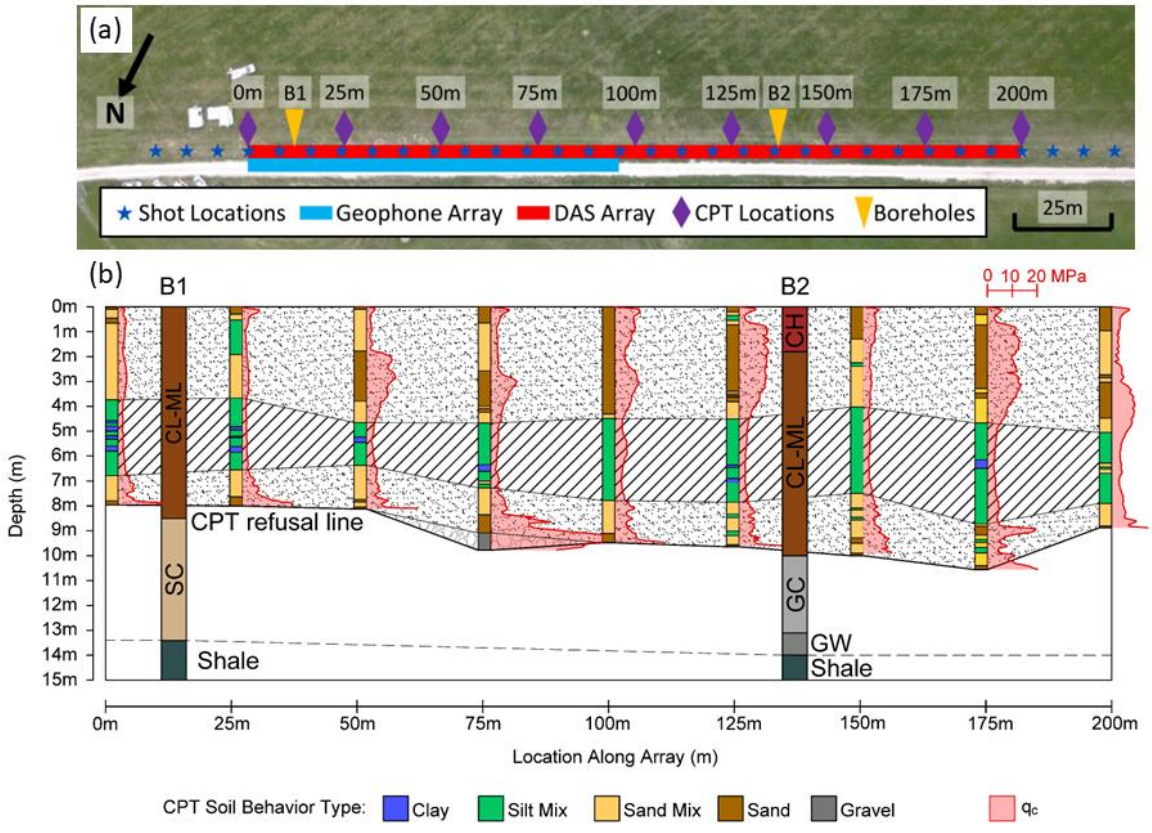

Figure 5-2 a) Aerial view of the Hornsby Bend test site showing the locations of CPT tests and boreholes as well as the DAS fiber optic cable, the geophone array, and the vibroseis shot locations; b) Geological cross section.

In addition to the non-invasive testing, nine cone penetration tests (CPT) soundings were carried out at 25-m intervals along the fiber-optic cable, covering the range from 0 to 200 m, as indicated in Figure 5-2a. These CPT soundings were performed on three different dates, spaced out over a span of approximately eight months. Yust et al. (2022) utilized the soil behavior type index value (Ic) method developed by Robertson (2009) to analyze the collected CPT data. Based on their analysis, they developed the subsurface cross section shown in Figure 5-2b down to the depths of CPT refusal. According to Yust et al. (2022), the site comprises three distinct layers above the depth of CPT refusal. These layers include a shallow granular layer (between depths of approximately 0 to 4 m) consisting of sand and sand mix, an intermediate cohesive layer (between depths of approximately 4 to 7 m) comprising clay and silt, and a deeper granular layer (between depths of approximately 7 to 10 m) composed of sand and sand mix. The depth of CPT refusal along the cable varied between 7.96 and 10.56 m, with an average depth of 9.15 m. In this study, the original cross-section developed by Yust et al. (2022) has been further extended from its original depth of approximately ten meters down to 15 m, as shown in Figure 5-3b. This extension enables the depiction of the shale layer depth, which was determined using data acquired from two recently drilled boreholes at the Hornsby Bend site. Both the first borehole (B1) positioned 12.5 m from the starting point of the geophone array, and the second borehole (B2), located 137.5 m away (refer to Figure 5-2b), confirmed the existence of a shale layer at an approximate depth of 13.5 m beneath the ground surface. Furthermore, seismic downhole testing was conducted in borehole B1, reaching a depth of 24 m with a receiver interval of 1 m, which led to the identification of four distinct velocity layers in the subsurface (discussed later in the paper).

5.3.1.1 1D MASW processing for the joint estimation of S-wave velocity and damping ratio

A promising technique for obtaining in-situ estimates of Vs and Ds relies on MASW (Nolet & Panza, 1976; McMechan & Yedlin, 1981; Gabriels et al., 1987; Park et al., 1999; Foti, 2000). This technique relies on the measurement of propagation characteristics of surface waves (typically, Rayleigh waves) and the testing procedure can be divided into three main steps:

- Data acquisition: a waveform generated by an artificial source is recorded along a linear array of sensors (typically, geophones) on the ground surface.
- Data processing: based on variations of phase lag and amplitude of surface waves along the array, the corresponding propagation speed (i.e., the phase velocity VR) and spatial attenuation of the amplitude (i.e., the phase attenuation αR) are derived, as a function of the frequency. The frequency-dependence of VR and αR is a combined effect of geometric dispersion, which results from the variation of mechanical properties with depth, and intrinsic dispersion, due to the constitutive behavior of linear viscoelastic media.
- Inversion: the Vs and the Ds profile with depth are obtained through an inversion scheme, where a theoretical soil model is calibrated to match the experimental VR and αR.

In MASW testing, the main advantage of the DAS technology with respect to conventional acquisition devices is the enhanced spatial resolution using low-cost instrumentation. Indeed, DAS allows for dense spatial sampling of the wavefield, potentially along a broad array extent. Conversely, achieving the same spatial resolution with ordinary receiver arrays would require a large number of sensors, entailing severe economic and logistic issues. Furthermore, high quality measurements can be obtained from conventional fiber-optic cables, that are not specifically designed for seismic investigation and already

deployed in the ground (e.g., the telecommunication infrastructure; Jousset et al., 2018). Therefore, the per-channel cost is moderately low. Applications of this technology to MASW surveys demonstrated that the dispersion estimates well match those obtained from geophone measurements (Galan-Comas, 2015; Vantassel et al., 2022a). Furthermore, the DAS acquisition tends to better identify higher propagation modes (Galan-Comas, 2015). However, fiber-optic systems are uniaxial devices, recording only perturbations acting in the longitudinal direction, and the correct location of measurement points may be uncertain in some cases (e.g., in the case of fiber overstuffing; Bakku, 2015). Also, the signal-to-noise ratio of measured data is lower compared to geophones. The lower quality in recorded traces limits the repeatability of the survey (Costley et al., 2018) and the reliability of the estimated wave parameters in the presence of weak signals (Mestayer et al., 2012). Finally, as explained above, the measurement technique involved in this technology partially limits the minimum investigable wavelengths at greater values than the one defined by the Nyquist-Shannon theorem (Lancelle, 2016; Bakulin et al., 2020). Therefore, the characterization of high-frequency R-wave data might be challenging.

Different acquisition layouts were investigated at the Hornsby Bend site to assess the influence of the DAQ type on the estimated dispersion and attenuation data, through the canonical 1D processing procedure. The first 94-m section of the fiber-optic cable, which is adjacent to the geophone array, is selected in this study to ensure a consistent comparison between results, as they sample a comparable volume of the soil deposit. This study refers to waveforms generated at shot points located at 5 m, 10 m, 20 m, and 40 m offset from the closest measurement point (i.e., 0m; refer to Figure 5-3a). Additionally, it only utilizes waveforms created by the Thumper truck, which generated a 12-s long chirp signal, with

frequency shifting from 5 Hz to 200 Hz. As an example, Figure 5-3 a-b reports the time histories of recorded data in the Hornsby Bend site using the geophone and the DAS arrays, with the active source located at an offset equal to 10 m.

As noted above, Vantassel et al. (2022a) showed that it is possible to extract equivalent surface wave dispersion data from seismic measurements made using a traditional geophone array and DAS. Moreover, they demonstrated that frequency-dependent normalization of the dispersion image removes the effect of scaling, integration, and differentiation on the acquired waveforms, thereby mitigating the need to convert the measurements into consistent engineering units prior to comparing dispersion data. Thus, Vantassel et al. (2022a) rigorously demonstrated the potential for extracting high-resolution, multi-mode surface wave dispersion data using DAS measurements and MASW-type processing.

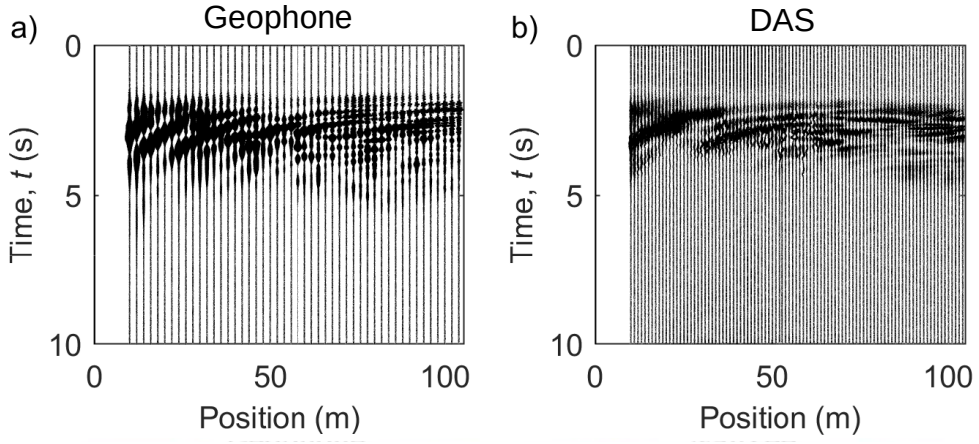

Figure 5-3 Recorded data at the Hornsby Bend site: a) Time histories of particle velocity recorded by the geophone array; b) Time histories of average radial strain recorded by DAS. Data refer to the wavefield generated from the active source located at an offset equal to 10 m.

Aimar et al. (2023) extended the assessment of the potential of exploiting DAS data for the joint estimation of the phase velocity and phase attenuation data. These quantities were

obtained from the vertical geophone waveforms through application of the recently developed Cylindrical Frequency-Domain BeamForming – Attenuation algorithm with Modal Filtering (CFDBFaMF; Aimar, 2022). The extraction of the R-wave parameters from DAS data adopts a modified version of the CFDBFaMF, which implements an average strain-based beamforming (further details are available in Aimar, 2022). In this way, the procedure properly models the spatial variation of the amplitude and phase of the average radial strain. Figure 5-4 compares the estimated modal dispersion and attenuation data for the first two propagation modes (labeled as R0 and R1, respectively), obtained from the interpretation of the DAS and the geophone array data. In this case, the data distribution is represented by the interval around the median value, the width of which equals one logarithmic standard deviation. Data statistics are obtained by combining results from different source offsets, consist with the multi-offset approach (Cox and Wood, 2011).

In general, dispersion and attenuation data well match with each other, particularly for the R0 mode. For the R1 mode, the DAS data does not allow phase velocity and attenuation estimates to be made over as broad a frequency range, being more limited at lower frequencies. This partially limits the capability of the DAS system in characterizing deeper layers. However, the corresponding degree of variability in the phase velocity and attenuation data derived from DAS is generally less than or equal to the variability affecting the geophone-based parameters. This result is quite surprising, as the signal-to-noise ratio of DAS records has been reported in other studies to be slightly lower than geophone records, hence, higher variability in the DAS-derived attenuation was expected. A possible reason behind the low data scatter can be the remarkably larger number of measurement

points that the DAS system includes, that provides a more exhaustive dataset of wavefield values to better constrain the velocity and the attenuation estimates.

Finally, experimental Rayleigh-wave data were mapped into profiles of Vs and Ds versus depth, by means of an inversion procedure. This operation was carried out through an improved Monte Carlo scheme, which implements a smart sampling technique of the model parameter space, by exploiting the scaling properties of the Rayleigh wave parameters in linear viscoelastic media. These properties allow a significant saving in computation time, preserving the quality of the resulting ground models at the same time. For simplicity, the model identification adopts a three-layer ground model, consistently with the stratigraphy inferred by cone penetration soundings carried out close to the DAS array (refer to Figure 5-2b). For each layer, an adequate range of layer thicknesses, S-wave velocities and damping ratios were investigated, whereas the mass density and Poisson's ratios were fixed at realistic values. The inversion was run using 10,000 trial earth models. Forward dispersion and attenuation modeling was carried out through the EDT toolbox (Schevenels et al., 2009). Model selection was based on a proper misfit function, wherein fitting errors between theoretical curves and experimental data were weighted as a function of the uncertainties affecting $VR(\omega)$ and $\alpha R(\omega)$.

Figure 5-5 shows results for the best fitting 30 models. Inverted S-wave velocity and damping ratio profiles are relatively well constrained, and the velocity and dissipation structures are clearly identifiable in the near-surface layers. Specifically, the resulting Vs model exhibits a gradual increase in stiffness with depth. The depths of the identified layers interfaces are about 4 m and 12-13 m. This result is consistent with the main geological interfaces inferred at the site and with information from past geophysical surveys. The

estimated DS profiles are affected by greater variability, which increases with depth. Specifically, DS is about 5% in the near-surface layer, and increases to around 8% in the layer below. As for the half-space, the variability in both Vs and Ds dramatically increases with respect to shallow layers. Indeed, the small amount of experimental data at long wavelengths does not allow effective constraint of estimated profiles at greater depths. Thus, the Vs oscillates between 330 m/s and 400 m/s, whereas Ds spans over a much broader range, mostly between 0.5% and 5% (that is, the variation is about one order of magnitude). This is the combined effect of the large variability in low-frequency attenuation data, the relevant influence of Vs on phase velocity and attenuation data, and the moderately low sensitivity of theoretical attenuation curves to Ds at great depths (e.g., Verachtert, 2018), that does not allow a constraint on Ds as effective as in the stiffness modeling. On the other hand, it should be noted that, particularly in the near-surface layers, both the velocity and dissipation structures show well defined trends.

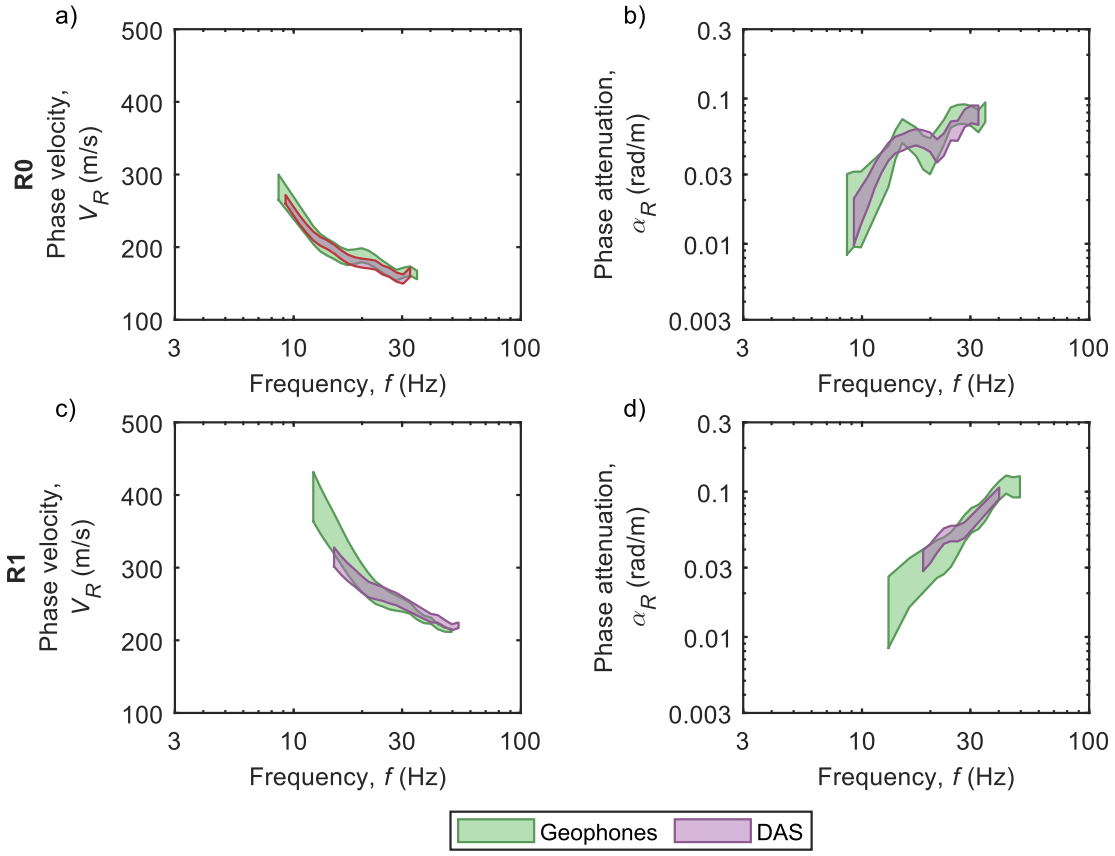

Figure 5-4 Comparison between the estimated dispersion and attenuation curves from the DAS and the geophone data at the Hornsby Bend site: a-b) Resulting dispersion (a) and attenuation (b) curves for the fundamental mode, R0; c-d) Resulting dispersion (c) and attenuation (d) curves for the first higher mode, R1. Estimated data are represented in terms of intervals given by one logarithmic standard deviation around the median value; after Aimar et al. (2023).

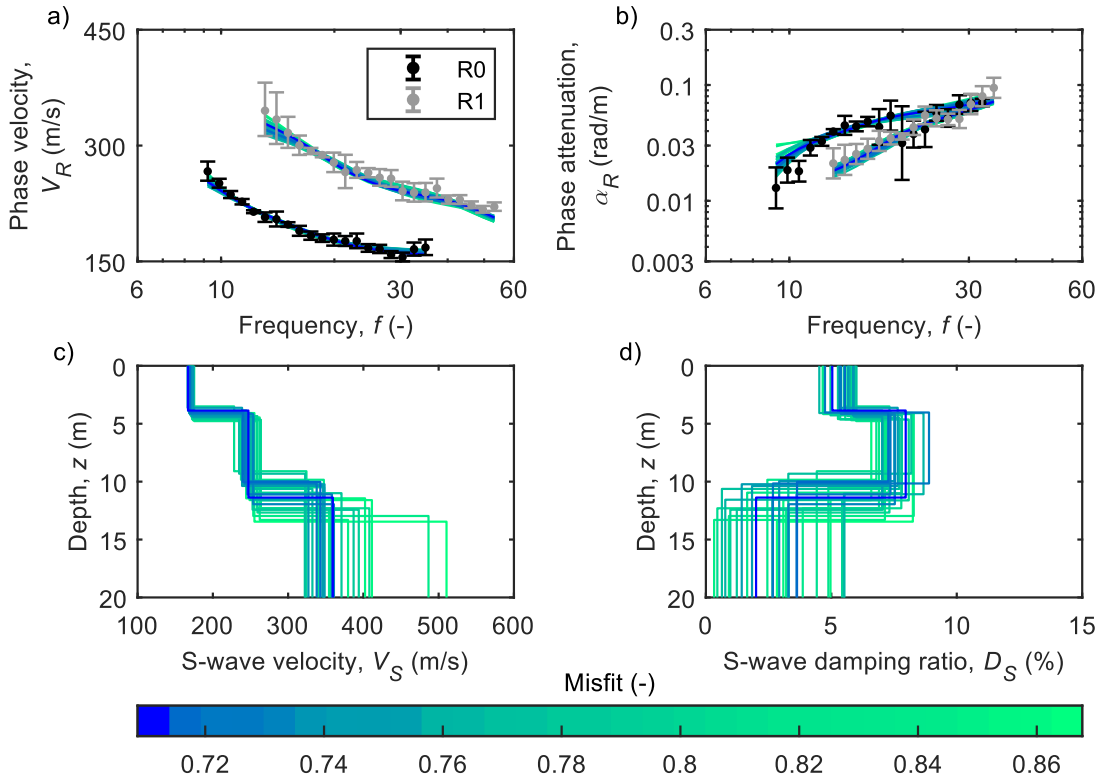

Figure 5-5 Best fitting inverted ground models to DAS experimental data from the Hornsby Bend site: a-b) Theoretical and experimental data for the phase velocity (a) and phase attenuation (b); c-d) Resulting S-wave velocity (c) and damping ratio (d) profiles; after Aimar et al. (2023).

In summary, the DAS technology can be successfully used to jointly estimate the phase dispersion and attenuation data, obtaining the same level of reliability of the canonical geophone array. Furthermore, the potentially stronger influence of incoherent noise on DAS data is balanced by the significant increase in the number of measurement points, thus resulting in a reduction in data variability, entailing an improvement in the accuracy of this system.

5.3.1.2 2D MASW using DAS

Two-dimensional (2D) MASW (e.g., Park 2005) is a technique used to produce a pseudo-2D Vs cross-section of the subsurface by expanding upon the 1D MASW approach. This

technique relies on spatially interpolating between numerous 1D MASW Vs profiles obtained from overlapping sub-arrays along a linear testing alignment. One of the main challenges when performing 2D MASW using conventional equipment, such as geophones and 24-channel seismographs, is that the geophone spacing, the length of the sub-arrays, the spatial interval between sub-arrays, and the positions of shots relative to the sub-arrays must be determined prior to/during data acquisition, making it difficult to adjust them during data processing (i.e., after data acquisition). This poses a challenge because the geophone spacing and sub-array length can have a significant impact on the lateral resolution, maximum characterization depth, and anomaly detection capabilities (Yust et al., 2022). Therefore, the ability to modify parameters such as the sub-array length after initial processing results have been investigated can be highly advantageous. According to a study by Yust et al. (2022), the use of DAS rather than traditional 2D MASW equipment can be highly advantageous in addressing these challenges.

In their recent study, Yust et al. (2022) investigated the effects of 2D MASW sub-array length using the DAS data collected at the Hornsby-Bend site. Three sets of sub-arrays with varying lengths were used to develop pseudo-2D Vs cross-sections along a 200-m long DAS line (refer to Figure 5-2a). The sub-arrays investigated by Yust et al. (2022) consisted of: (a) 12-channel sub-arrays approximately 11-m long, (b) 24-channel sub-arrays approximately 23-m long, and (c) 48-channel sub-arrays approximately 47-m long. They used an equivalent sub-array spatial interval of four channels (approximately 4 m) for all sub-arrays and performed 129 individual MASW analyses in total. The pseudo-2D Vs cross-sections obtained using the 12-, 24-, and 48-channel sub-arrays had lateral extents of 187.68 m, 175.44 m, and 150.96 m, respectively, while maintaining a consistent depth

of 15 m, as shown in Figures 5-6a, 5-6b, and 5-6c, respectively. These cross-sections are quite similar over the top 7-8 m, but show noticeable differences at greater depths, highlighting the sensitivity of 2D MASW results to the choice of sub-array length. Nonetheless, each cross-section obtained from the different sub-array lengths was found to correlate better with a different key feature of the subsurface, as verified by comparisons with invasive data collected along the array alignment. For instance, Yust et al. (2022) found that, for the Hornsby-Bend site, the most prominent impedance contrast in the Vs cross-section obtained using 12-channel sub-arrays corresponded to the depth of CPT refusal (see Figure 5-6a), while the Vs cross-section obtained using longer, 48-channel sub-arrays corresponded to the deeper shale impedance contrast, as indicated by the boring lithology log superimposed on the Vs cross-section (see Figure 5-6c). It is worth noting that Yust et al. (2022) were able to investigate these different sub-array lengths post data acquisition, owing to the flexibility allowed by DAS technology. Unlike traditional seismic equipment, DAS records the wavefield generated at each of the considered shot locations simultaneously along the entire length of the array at a constant channel separation, eliminating the need to pre-determine sub-array length and spatial sampling interval during data acquisition. This feature enables the investigation of multiple sub-array geometries during the processing stage of the analysis, thereby providing greater flexibility and control over acquisition parameters.

The study by Yust et al. (2022) shed light on the advantages of utilizing DAS in 2D MASW, and emphasized the significance of incorporating a priori information, such as invasive testing data, to fine-tune the 2D MASW analysis and achieve project-specific objectives, whenever feasible. In cases where there is insufficient conclusive information

to constrain subsurface layering, Yust et al. (2022) recommended examining multiple 2D MASW sub-array configurations to gain a comprehensive understanding of the subsurface conditions and accurately assess the uncertainty of the results.

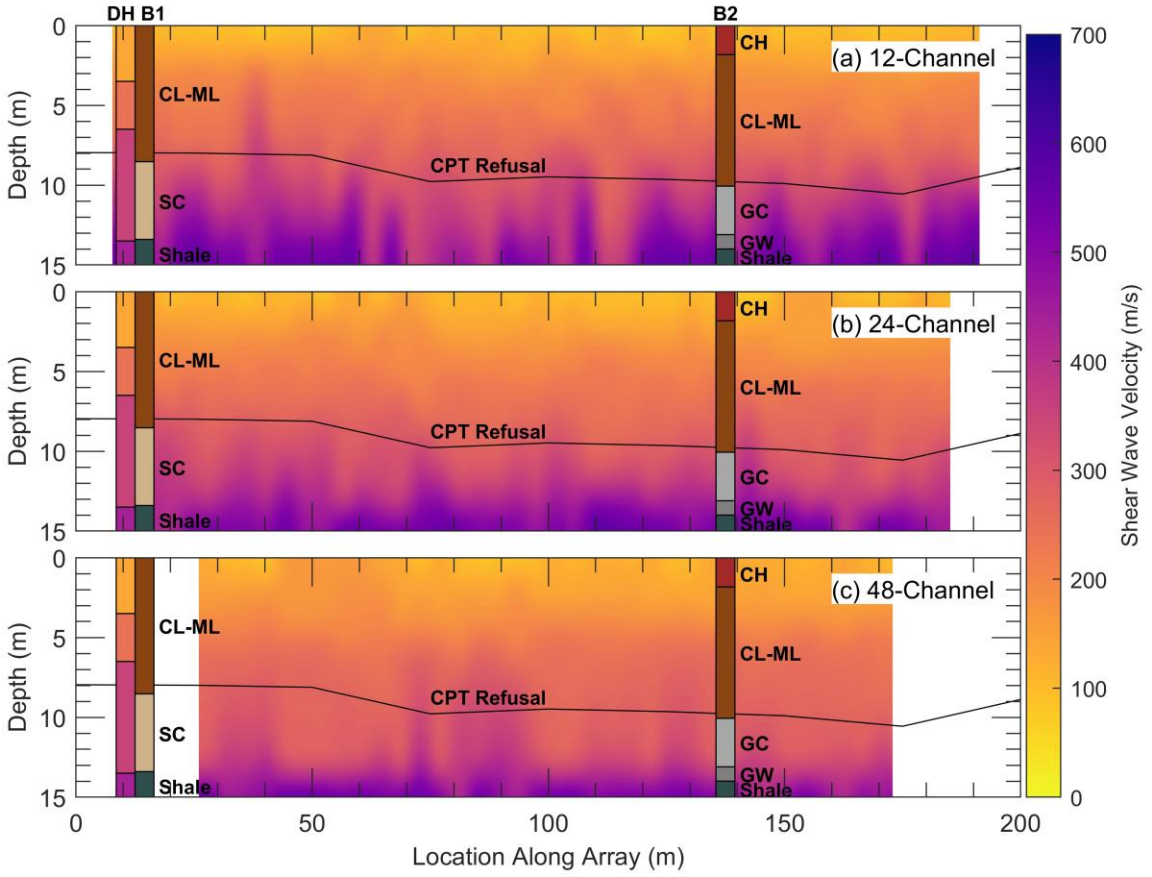

Figure 5-6 Pseudo-2D Vs cross-sections after Yust et al. (2022) from the: (a) 47, 12-channel MASW sub-arrays, (b) 44, 24-channel MASW sub-arrays, and (c) 38, 48-channel MASW sub-arrays inverted using a 15-layer inversion parameterization. The depths of refusal for 9 CPT soundings along the array are shown on all plots with a solid black line.

5.3.1.3 Machine learning

In recent years, there has been a growing interest in utilizing deep learning/machine learning/artificial intelligence (DL/ML/AI) techniques for non-invasive subsurface imaging (Adler et al., 2021). For example, with a fully trained and adaptable neural network it would theoretically be possible to swiftly generate subsurface images directly

from wavefield measurements without the need to perform costly and complicated inversions. Furthermore, this type of imaging could be performed by an analyst without any significant understanding of machine learning. This would enable the reuse of fully trained neural networks to rapidly produce site-specific results, eliminating the need for specialized expertise, which is often required by conventional inversion methods. To train a neural network for subsurface imaging, a large dataset consisting of numerous input-output image pairs is required. The input can be a representation of the raw data acquired from the field, or some post-processed version of it, while the output can be a subsurface image of the parameter of interest, such as a Vs cross-section. Due to the considerable number of image pairs needed to train a neural network, all research studies thus far have relied on numerically-developed image pairs, as in the case of the convolutional neural network (CNN) developed by Vantassel et al. (2022b), which takes a seismic wavefield input image and outputs a 2D Vs image. A significant challenge that has impeded the use of machine learning as an end-to-end imaging technique for real-field applications is the lack of generalizability (Li et al., 2020; Feng et al., 2022). In other words, these neural networks often struggle when presented with real-field data, particularly if the data was acquired using a different acquisition configuration than the one used during network training (e.g., Vantassel et al., 2022b).

Abbas et al. (2023a) have recently introduced a CNN that shows promise for rapidly generating 2D Vs images of near-surface soil-over-bedrock geology using real-field data. To train and test their CNN, they utilized 100,000 synthetic near-surface models with varying soil-over-bedrock conditions. Their CNN takes a frequency-dependent normalized dispersion image as input (rather than a seismic wavefield image) and produces a 2D Vs

image as output, as illustrated schematically in Figure 5-7. Abbas et al. (2023a) demonstrated that while using different testing configurations in terms of source type, source offset, number of receivers, and receiver spacings leads to significantly different measured wavefields for the same subsurface structure, the normalized dispersion images processed from these different wavefields are quite similar, although not identical. Abbas et al. (2023a) leveraged this feature and demonstrated that a CNN trained on normalized dispersion images processed from wavefields acquired using a specific testing configuration can still perform well when presented with dispersion images processed from wavefields acquired using different testing configurations. This acquisition flexibility significantly improves the CNN's generalization capability, enabling it to be utilized as an end-to-end imaging method or as a tool to create rapid starting models for full-waveform inversion (FWI).

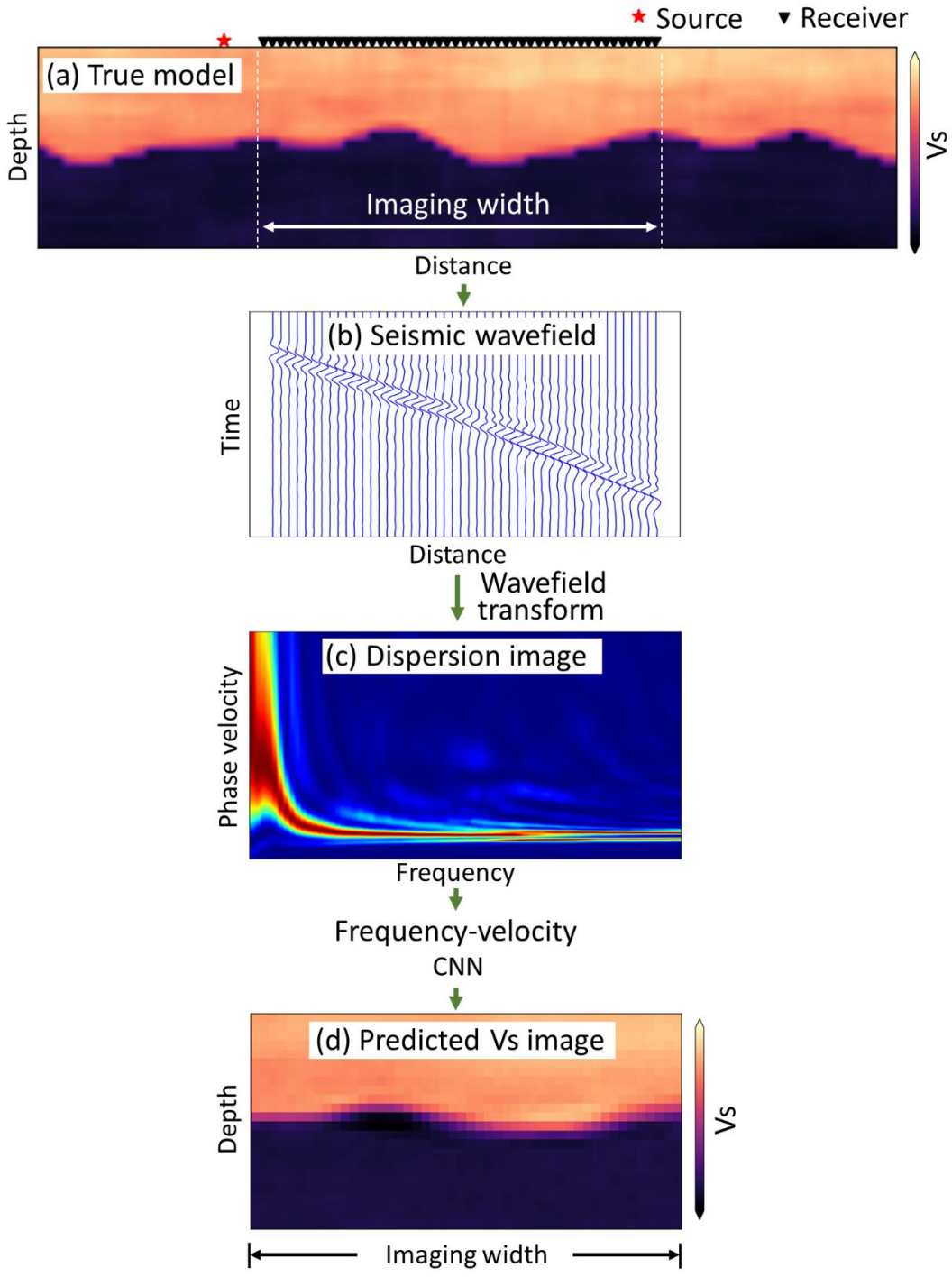

Figure 5-7 Frequency-velocity CNN framework proposed by Abbas et al. (2023a) for 2D Vs imaging of near-surface soil-over-bedrock geology.

Abbas et al. (2023a) demonstrated the practical capability of their CNN by applying it to experimental field data collected at the Hornsby Bend site. Their CNN generated a high-resolution 48-m wide by 24-m deep Vs subsurface image, which agrees well with the actual subsurface structure determined through invasive tests conducted at the site, as shown in Figure 5-8, thereby establishing the CNN's promise in handling real-field data. While Abbas et al. (2023a) used geophone-derived dispersion images in their study, their approach is equally applicable to DAS-derived dispersion images, as geophone and DAS-derived dispersion images are shown to be equivalent by Vantassel et al. (2022a).

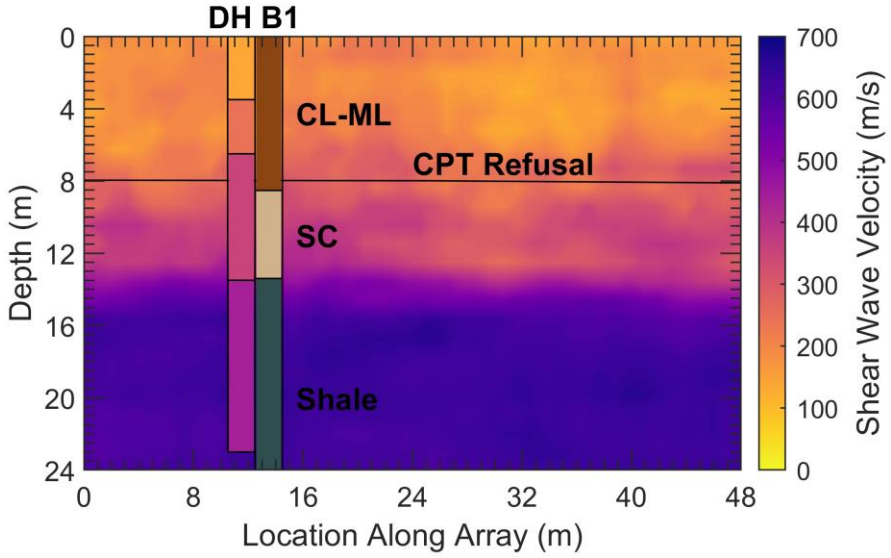

Figure 5-8 The frequency-velocity CNN output 2D Vs image for the Hornsby Bend site after Abbas et al. (2023a). For comparison with actual field conditions, a borehole log (i.e., B1) is superimposed on the predicted Vs image at 12.5 m, which is the location where the boring was conducted.

5.3.1.4 Full wave form inversion using DAS

Full-waveform inversion (FWI) is a robust imaging technique that produces 2D/3D images of the subsurface by matching a synthetic seismic wavefield, generated by numerically solving the wave equation, to an experimental seismic wavefield acquired in the field. The

process involves an iterative data-fitting procedure that requires modifying an initial model through which the synthetic waveforms propagate until the synthetic and experimental wavefields match, based on a selected inversion objective function. This is achieved by adjusting the material properties of the synthetic subsurface model, thereby reducing the misfit between the synthetic and recorded data. Unlike 2D MASW, which can only produce pseudo-2D images, FWI generates true 2D and 3D images by leveraging all available information in the seismic wavefield, including phase and amplitude. This sets FWI apart from other approaches, such as seismic refraction that relies only on wavefield first arrivals or surface wave testing that uses only Rayleigh dispersion. In addition to providing insights into material parameters like Vs and compression wave velocity (Vp), which are of high interest to engineers, FWI can also be used to assess any other material properties that impact seismic wave propagation, such as density and damping ratio.

Despite its numerous advantages, FWI also encounters significant challenges. The computational costs associated with FWI are considerable, and the accuracy of its results can be heavily influenced by the initial model (Vantassel et al., 2022b; Yust et al., 2023), particularly in near-surface applications such as imaging the top 30 m. In the near surface, material properties tend to exhibit rapid variations over short distances, leading to a complex mixture of different wavefield components. This amalgamation includes compression, shear, and surface waves, which have not yet propagated far enough to separate from each other.

Traditionally, FWI analysis has utilized data acquired through geophones, which capture particle velocity wavefields. However, the high spatial resolution provided by DAS has garnered interest for use in FWI studies. Nonetheless, there is a dearth of literature on the

use of FWI with DAS data for near-surface characterization, particularly in regards to directly inverting DAS strain data without first converting it to particle displacement or velocity.

In a recent study conducted by Yust et al. (2023), a direct inversion of strain measurements obtained from DAS was performed at the Hornsby Bend site. The study aimed to image a cross-section measuring 200-m wide and 30-m deep. Four distinct 1D and 2D starting models specifically tailored to the site were utilized. The first model was based on a 1D Vs profile obtained through traditional MASW testing. The second starting model utilized a 1D Vs profile derived from seismic down-hole invasive testing. The third starting model relied on the CNN 2D Vs subsurface model developed by Abbas et al. (2023a) and depicted in Figure 5-8. Lastly, the fourth model was constructed using the 2D MASW analysis with the 48-channel subarrays developed by Yust et al. (2022) and discussed in the 2D MASW section (refer to Figure 5-6). The four FWI starting models used by Yust et al. (2023) are visually depicted in Figure 5-9. Yust et al. (2023) inverted for Vs, Vp, and density, ρ, while assuming constant quality factors to characterize the attenuation of compression and shear waves in the model. Nonetheless, they noted that the attenuation parameters significantly impact the FWI results and noted that additional studies are required to help constrain these values. Yust et al. (2023) observed enhancements in each of the initial models through the implementation of FWI. This improvement was evident when comparing the misfit between synthetic and recorded waveforms for both the initial and updated models. For example, Figure 5-10 illustrates the misfit between the synthetic and experimental waveforms of the initial models, while Figure 5-11 displays the misfit for the FWI updated models following the first stage of a four-stage FWI. By comparing the waveform misfit

values in Figure 5-10 and Figure 5-11, which are based on a graph space optimal transport distance (GSOTD) algorithm, one can clearly see that the FWI iterations in the first stage resulted in better waveform matches (i.e., lower misfit values) for all four starting models. This reduction in misfit values continued through each stage of the FWI process, ultimately resulting GSOTD misfit values for the four final models that were very similar to one another and only varied between 1.91 to 1.46. Despite the similarities in the final waveform misfit values, the final subsurface models did not vary significantly from their respective starting models and exhibited noticeable visual differences from one another, as depicted in Figure 5-12. Within the upper 7 m, the final Vs images are quite similar. Yet, below this depth there are evident disparities in the magnitudes and rates at which Vs increase. This finding highlights the intrinsic non-uniqueness associated with the FWI process, underscoring the importance of incorporating multiple starting models. Assessing the sensitivity of results to the choice of starting model and attaining consistent outcomes across different models engenders confidence in the subsurface regions where the models yield congruent results.

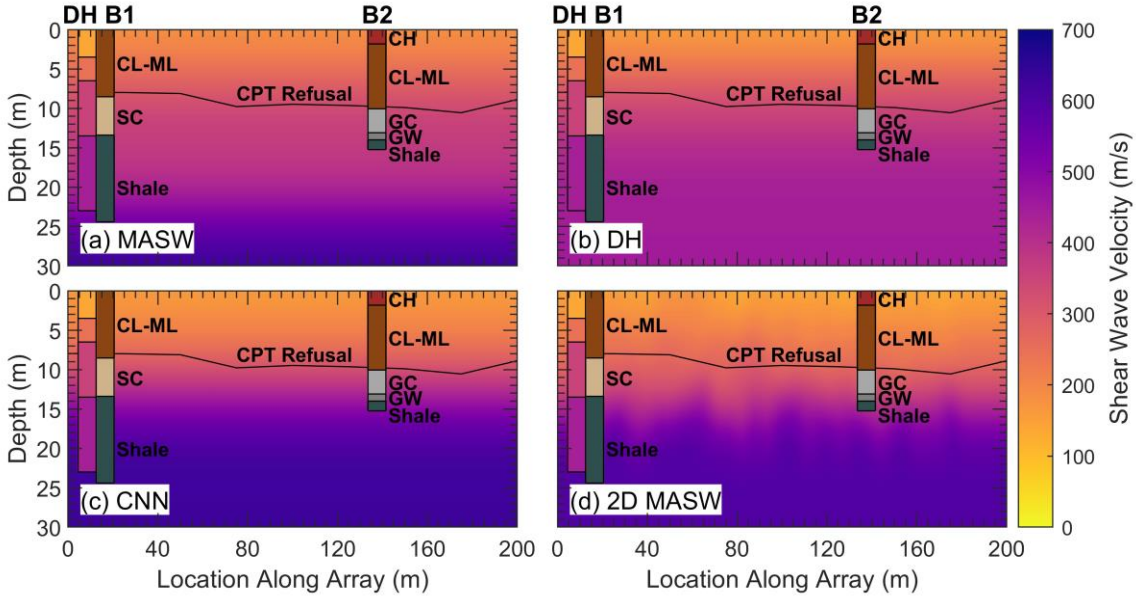

Figure 5-9 The four smoothed 2D Vs starting models used by Yust et al. (2023) for FWI based on: (a) 1D MASW, (b) downhole testing, (c) CNN machine learning, and (d) pseudo-2D MASW.

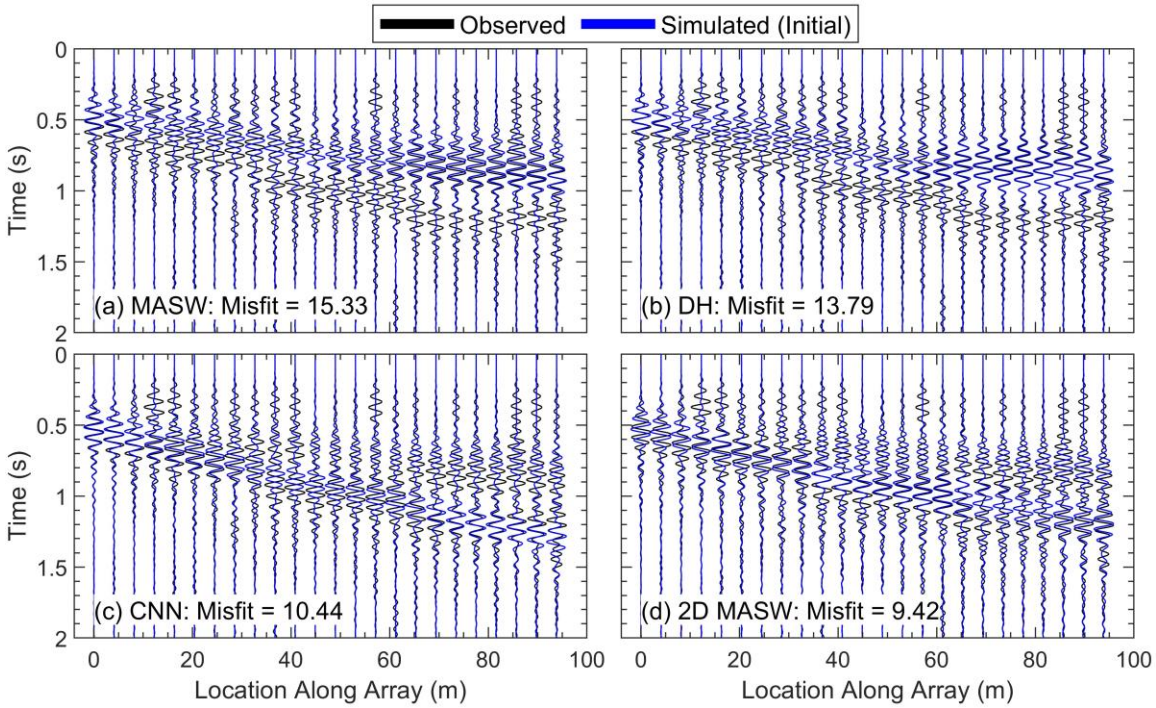

Figure 5-10 Normalized observed and simulated waveforms by Yust et al. (2023) from Shot 1 (−24 m) of Stage 1 (10 to 15 Hz) for the: (a) MASW, (b) DH, (c) CNN, and (d) 2D MASW starting models. GSOTD misfit values for each set of simulated waveforms are shown in the bottom left of each plot. Note that for clarity purposes, the waveforms are only shown for every fourth channel used for FWI.

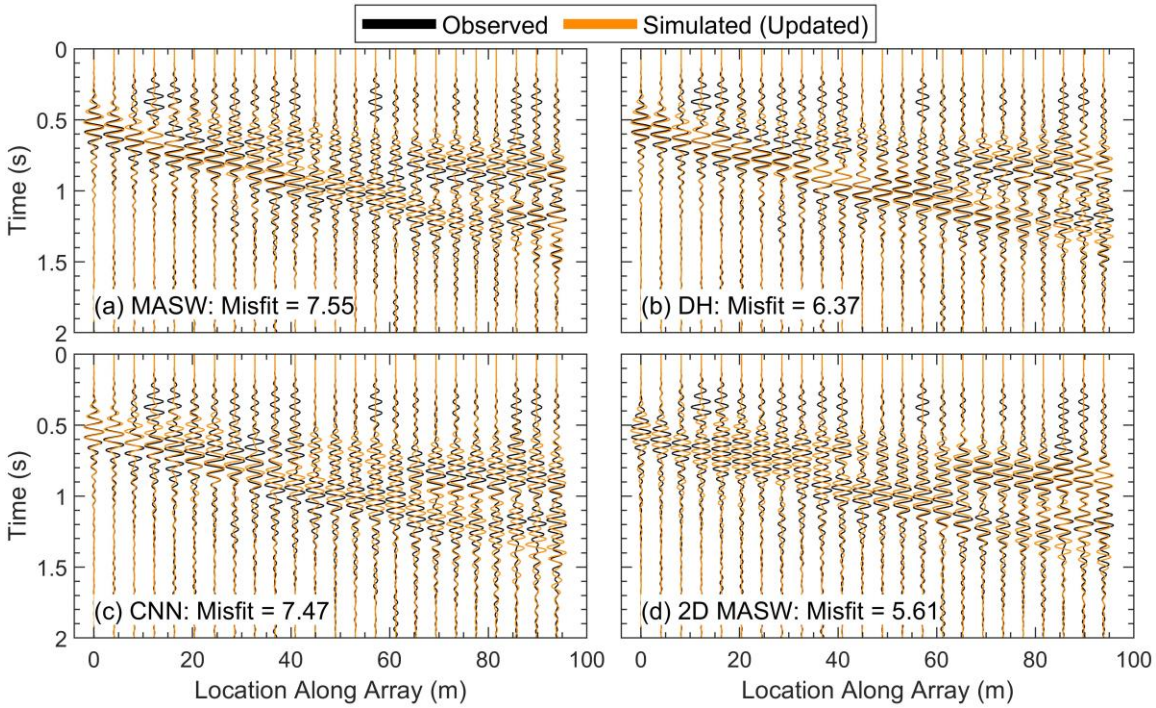

Figure 5-11 Normalized observed and simulated waveforms by Yust et al. (2023) from Shot 1 (−24 m) for the updated models at the end of FWI Stage 1 (10 to 15 Hz) based on the: (a) MASW, (b) DH, (c) CNN, and (d) 2D MASW starting models. GSOTD misfit values for each set of simulated waveforms are shown in the bottom left of each plot. Note that for clarity purposes, the waveforms are only shown for every fourth channel used for FWI.

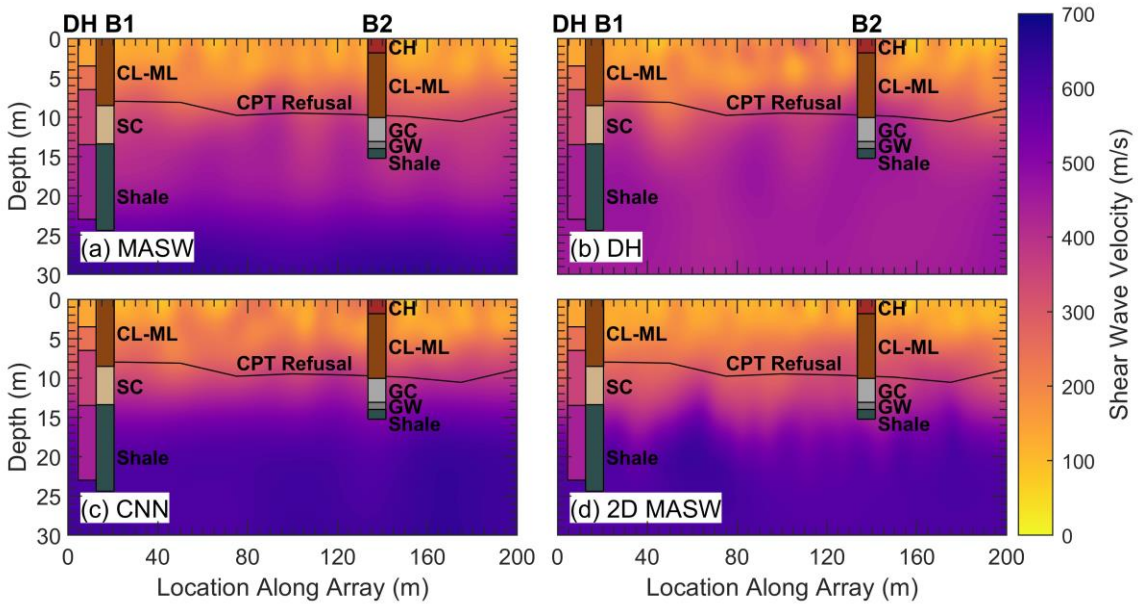

Figure 5-12 Borehole logs, downhole Vs results, and the depth to CPT refusal overlaid on the final, updated 2D Vs images at the end of FWI Stage 4 (10 to 30 Hz) for the: (a) MASW, (b) DH, (c) CNN, and (d) 2D MASW starting models after Yust et al. (2023).

5.3.2 Discussion

In the preceding sections, various subsurface imaging techniques have been reviewed. This section aims to provide an analysis of the advantages and disadvantages associated with each technique. The first method reviewed was 1D MASW, which is a well-established technique for Vs imaging using geophones. Extensive research supports this method, offering best practices and recommended workflows (Foti et al., 2018; Vantassel & Cox, 2022). However, this study presents two advancements to the conventional 1D MASW workflow. The first involves using DAS instead of geophones for data acquisition, while the second is the joint inversion for damping and Vs. The utilization of DAS significantly enhances measurement scales and spatial resolution beyond the capabilities of traditional measurement technologies. Moreover, when appropriate precautions are taken, the

measurements obtained through DAS exhibit good agreement with those acquired using traditional methods like geophones (Daley et al., 2016; Hubbard et al., 2022; Vantassel et al., 2022a). It should be noted, though, that DAS provides deformation measurements only along the fiber optic cable direction, while geophones can capture particle motion point measurements in all three directions individually or simultaneously.

The second advancement focuses on the joint inversion for damping and Vs using data acquired through a 1D MASW setup, which represents a significant breakthrough, particularly due to the crucial role of damping in various domains, including site response. Aimar et al. (2023) showed that their approach is applicable on both geophone and DAS acquired data. Furthermore, this paper illustrates the innovative application of DAS in jointly characterizing the stiffness and dissipative parameters of a soil deposit based on a fiber-optic array. Figure 5-4 illustrates the favorable agreement between attenuation estimates obtained using both DAS and geophones for the Hornsby Bend site, with DAS-based measurements exhibiting lower uncertainty compared to geophone-based measurements. However, the uncertainty in attenuation estimates generally tends to be higher than that observed for dispersion estimates, as detailed by Aimar (2022). In addition, the results presented were obtained from a portion of a large experimental dataset. Future studies will thoroughly investigate these data to obtain a more comprehensive view of DAS performance compared to geophones and to seek for improved estimates of the phase attenuation, especially at low frequencies. It is worth noting that the 1D MASW technique generates a 1D subsurface profile by averaging the soil properties beneath the instrumentation array. To obtain a 2D representation of the subsurface, the 2D MASW technique was utilized. However, 2D MASW generates a pseudo 2D image rather than a

true one, as it involves interpolating between multiple 1D MASW-derived soil profiles. Yust et al. (2022) demonstrated that this method produced Vs images that aligned with significant features at the Hornsby-bend site. Nevertheless, it was also observed that the resulting Vs image is reliant on the specific parameters chosen for data acquisition.

Another approach presented in this paper for 2D Vs imaging is the machine learning CNN proposed by Abbas et al. (2023a). This approach holds great promise as it can generate true 2D images of the subsurface, in contrast to the pseudo 2D images obtained through 2D MASW. Additionally, the machine learning approach is notable for its remarkable speed, as a trained neural network can deliver imaging results within seconds. However, this technique is still in its early stages of development and requires extensive research and testing before it can be reliably employed for subsurface imaging purposes.

The final imaging approach explored in this paper is 2D FWI. What sets FWI apart from the previously discussed methods is its utilization of the entire measured wavefield to generate true 2D and 3D subsurface images. Moreover, FWI differs from the machine learning approach by not operating as a black box method. As FWI continues to undergo further development, it is anticipated to become the preferred imaging technique. However, FWI currently faces several challenges, including notable computational costs and time-consuming complexity of the analysis process. Furthermore, studies by Yust et al. (2023) and Vantassel et al. (2022b) have demonstrated that the imaging results obtained through FWI are significantly influenced by the starting model used.

5.4 Reference dataset: Newberry site, Florida

The preceding paragraphs have showcased the latest developments in stress wave-based non-invasive subsurface imaging, with a practical demonstration of their potential at the

Hornsby Bend site. However, it should be noted that the Hornsby Bend site is a relatively simple site with no documented underground anomalies. Furthermore, the seismic data from the Hornsby Bend site was collected using relatively short and linear DAS and geophone arrays. To provide researchers with a more comprehensive dataset at a more challenging site, Abbas et al. (2023b) conducted a field test in Newberry, Florida, at a site known for its spatial variability, karstic voids, and underground anomalies. The experiment utilized cutting-edge sensing technologies, including a two-kilometer DAS fiber optic cable, forming a dense 2D array of 1920 channels, and a 2D array of 144 SmartSolo three-component nodal stations to sense active-source and passive-wavefields, as illustrated schematically in Figure 5-13. The active-source data was generated using a powerful three-dimensional vibroseis shaker truck and impact sources, and it was simultaneously sensed by both the DAS and nodal stations. The vibroseis truck was used to vibrate the ground in the three directions (two horizontal and one vertical) at 260 locations inside and outside the instrumented array, while the impact sources were used at 268 locations within the array (refer to Figure 5-13). The passive wavefield data, recorded using the nodal stations, consisted of 48 hours of ambient noise collected over four days in four twelve-hour time blocks. The active-source and passive-wavefield DAS and nodal station data have been preprocessed and organized in an easy-to-navigate folder structure. The raw and processed data, along with detailed documentation of the experiment and Python tools to aid in visualizing the DAS dataset have been archived and made publicly available on DesignSafe (Abbas et al., 2023c). The Newberry dataset (Abbas et al., 2023b; Abbas et al., 2023c), featuring a powerful, triaxial vibroseis shaker and 3C sensors, in conjunction with the dense

DAS array, offers a valuable resource for researchers exploring novel non-invasive subsurface imaging approaches that utilize stress waves.

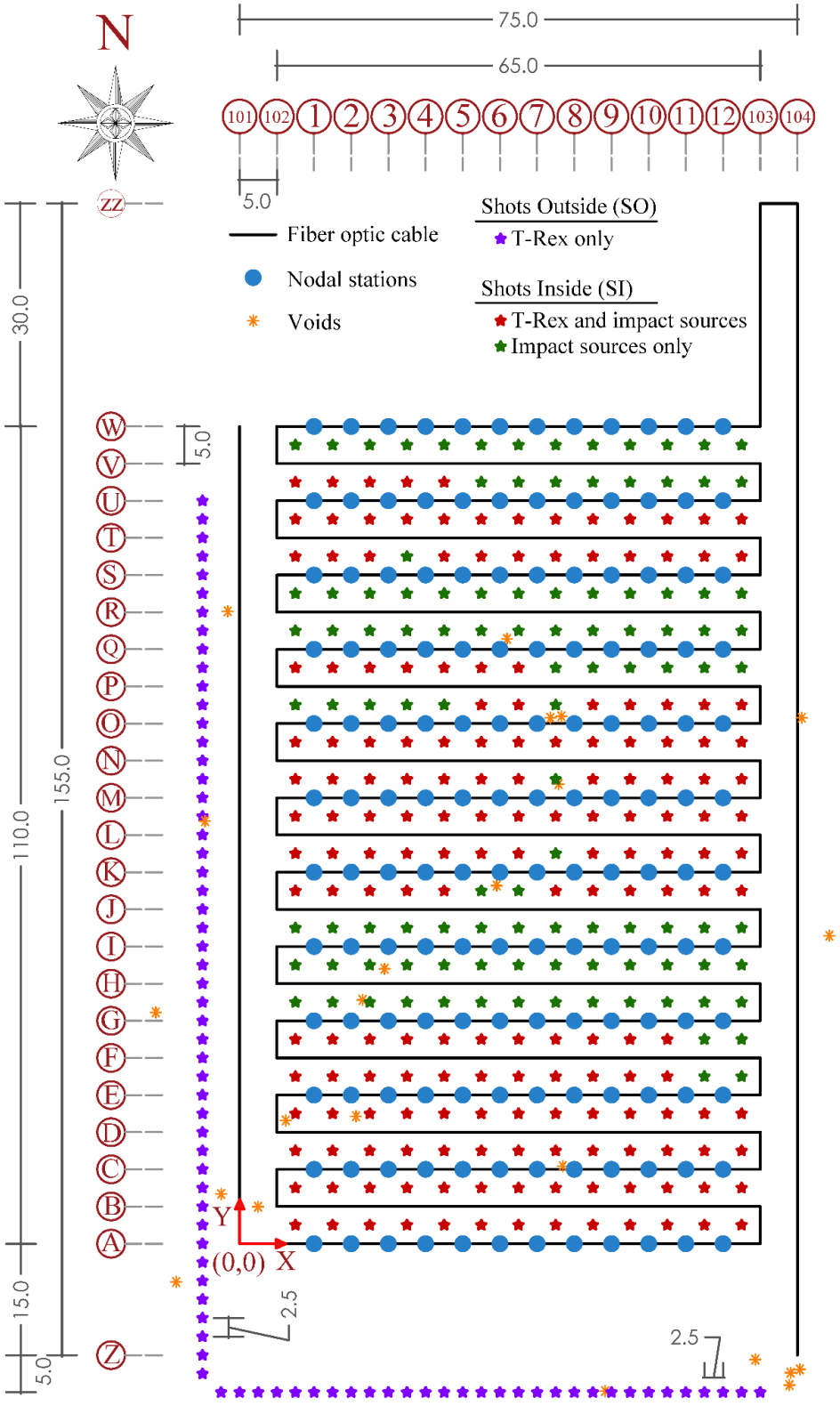

Figure 5-13 Newberry site testing configuration after Abbas et al. (2023b).

5.5 Conclusion

The paper covers advancements in non-invasive subsurface imaging technologies that utilize stress waves for site characterization. These advancements span both innovative sensing technologies and advanced 1D and 2D imaging techniques for retrieving the small-strain shear modulus and damping ratio. Regarding sensing innovations, a particular focus is placed on the use of DAS as the data acquisition system for stress wave sensing in near-surface imaging applications. In terms of imaging techniques, the joint inversion of attenuation and dispersion data is presented, utilizing 1D MASW in conjunction with either DAS or traditional sensing techniques such as geophones. With this approach, 1D shear wave velocity and damping profiles can be obtained. The presented 2D imaging techniques comprise 2D MASW using DAS, machine learning for 2D subsurface imaging, and 2D FWI using DAS data. All of these advancements, whether in sensing technologies or analysis methods, were implemented at the well-characterized Hornsby Bend site in Austin, Texas, enabling a discussion of the advantages and disadvantages associated with each method. Finally, a comprehensive and open-access subsurface imaging experiment conducted in Newberry, Florida, has been presented, where state-of-the-art technologies for sensing stress waves and generating wavefields have been implemented.

5.6 References

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

Abbas, A., B. R. Cox, K. T. Tran, I. Corey, & N. Dawadi (2023b). An Open-Access Data Set of Active-Source and Passive-Wavefield DAS and Nodal Seismometer Measurements at the Newberry Florida Site, Seismol. Res. Lett., 1–17, doi: 10.1785/0220230216.

Abbas, A., Cox, B. R., Tran, K. T., Corey, I., Dawadi, N., & Menq, F.-Y. (2023c). Active-source and Passive-wavefield DAS and Nodal Station Measurements at the Newberry Florida Site. https://doi.org/10.17603/ds2-50eh-7v93

Adler, A., Araya-Polo, M., & Poggio, T. (2021). Deep learning for seismic inverse problems: towards the acceleration of geophysical analysis workflows. IEEE Signal Processing Magazine, 38(2), 89-119.

Aimar, M. (2022). Uncertainties in the estimation of the shear-wave velocity and the small-strain damping ratio from surface wave analysis. (PhD thesis), Politecnico di Torino.

Aimar, M., Cox, B. R., & Foti, S. (2023). Surface wave testing with distributed acoustic sensing measurements to estimate the shear-wave velocity and the small-strain damping ratio. Paper presented at the National Conference of Researchers in Geotechnical Engineering (CNRIG), Palermo, Italy.

Ajo-Franklin, J., Dou, S., Daley, T., Freifeld, B., Robertson, M., Ulrich, C., Wood, T., Eckblaw, I., Lindsey, N., Martin, E., Wagner, A. (2017). Time-lapse surface wave monitoring of permafrost thaw using distributed acoustic sensing and a permanent automated seismic source. Paper presented at the SEG Technical Program Expanded Abstracts 2017.

Bakku, S. K. (2015). Fracture characterization from seismic measurements in a borehole. (PhD Thesis), Massachusetts Institute of Technology.

Bakulin, A., Silvestrov, I., & Pevzner, R. (2020). Surface seismics with DAS: An emerging alternative to modern point-sensor acquisition. The Leading Edge, 39(11), 808-818.

Ciancimino, Lanzo, Alleanza, Amoroso, Bardotti, Biondi, Cascone, Castelli, Di iulio, d' Onofrio, Foti, Lentini, Madiai, and Vessia. (2020). Dynamic characterization of fine-grained soils in Central Italy by laboratory testing, Bulletin of Earthquake Engineering: 29.

Cox, B., Wills, P., Kiyashchenko, D., Mestayer, J., Lopez, J., Bourne, S., Lupton, R., et al., (2012). Distributed Acoustic Sensing for Geophysical Measurement, Monitoring and Verification.

Cox, B. R., and Wood, C. M. (2011). "Surface Wave Benchmarking Exercise: Methodologies, Results, and Uncertainties." GeoRisk 2011, American Society of Civil Engineers, Atlanta, Georgia, United States, 845–852.

Comina, C., Foti, S., Boiero, D., & Socco, L. V. (2011). Reliability of VS,30 Evaluation from Surface-Wave Tests. Journal of Geotechnical and Geoenvironmental Engineering, 137(6), 579-586. doi:10.1061/(ASCE)GT.1943-5606.0000452.

Connolly, D. P., Kouroussis, G., Laghrouche, O., Ho, C. L., & Forde, M. C. (2015). Benchmarking railway vibrations – Track, vehicle, ground and building effects. Construction and Building Materials, 92, 64-81. https://doi.org/10.1016/j.conbuildmat.2014.07.042

Costley, D. R., Galan-Comas, G., Kirkendall, C. K., Simms, J. E., Hathaway, K. K., Parker, M. W., Ketcham, S., Smith, E., Folks, W., Milburn, T., Wadman, H. M. (2018). Spectral Analysis of Surface Waves with Simultaneous Fiber Optic Distributed Acoustic Sensing and Vertical Geophones. Journal of Environmental and Engineering Geophysics, 23(2), 183-195.

Daley, T. M., Miller, D. E., Dodds, K., Cook, P., & Freifeld, B. M. (2016). Field testing of modular borehole monitoring with simultaneous distributed acoustic sensing and geophone vertical seismic profiles at Citronelle, Alabama. Geophysical Prospecting, 64(5), 1318-1334.

Darendeli. (2001). Development of a new family of normalized modulus reduction and material damping curves, Doctoral Dissertation, University of Texas at Austin.

Dos Santos, Colaço, Costa, and Calçada. (2016). Experimental analysis of track-ground vibrations on a stretch of the Portuguese railway network, Soil Dynamics and Earthquake Engineering, 90, 358-380.

Feng, S., Lin, Y., & Wohlberg, B. (2022). Multiscale data-driven seismic full-waveform inversion with field data study. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-14.

Fernandes, F. C., Riz de Moura, B. L., Soares de Almeida, M. S., Souza Junior, L. O., Tarazona, S. F. M., Ferreira de Almeida, M. C., de Camargo Barros, J. M. (2023). Determination of the dynamic parameters of Speswhite kaolin with resonant column and centrifuge tests, Soil and Rocks, 46(2), DOI: 10.28927/SR.2023.013422, https://www.soilsandrocks.com/sr-2023-013422

Foti, S. (2000). Multistation methods for geotechnical characterization using surface waves. (PhD thesis), Politecnico di Torino.

Foti, Aimar, and Ciancimino. (2021). Uncertainties in small-strain damping ratio evaluation and their influence on seismic ground response analyses. in, Latest Developments in Geotechnical Earthquake Engineering and Soil Dynamics (Springer: Singapore).

Foti, S., Hollender, F., Garofalo, F., Albarello, D., Asten, M., Bard, P.-Y., Comina, C., Cornou, C., Cox, B., Di Giulio, G., Forbriger, T., Hayashi, K., Lunedei, E., Martin, A., Mercerat, D., Ohrnberger, M., Poggi, V., Renalier, F., Sicilia, D., Socco, V. (2018). Guidelines for the good practice of surface wave analysis: a product of the InterPACIFIC project. Bulletin of Earthquake Engineering, 16(6), 2367-2420.

Freifeld, B.M., Pevzner, R., Dou, S., Correa, J., Daley, T.M., Robertson, M., Tertyshnikov, K., Wood, T., Ajo-Franklin, J., Urosevic, M., Gurevich, B. (2016). The CO2CRC Otway Project deployment of a distributed acoustic sensing network coupled with permanent rotary sources. Paper presented at the 78th EAGE Conference and Exhibition 2016.

Gabriels, P., Snieder, R., & Nolet, G. (1987). In situ measurements of shear-wave velocity in sediments with higher-mode Rayleigh waves. Geophysical Prospecting, 35(2), 187-196.

Galan-Comas, G. (2015). Multichannel analysis of surface waves using distributed fiber optic sensors. (PhD Thesis), Mississippi State University.

Grattan, and Sun. 2000. Fiber optic sensor technology: an overview, Sensors and Actuators A: Physical, 82, 40-61.

Hornman, K., Kuvshinov, B., Zwartjes, P., & Franzen, A. (2013). Field trial of a broadside-sensitive distributed acoustic sensing cable for surface seismic. Paper presented at the 75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013. https://doi.org/10.3997/2214-4609.20130383

Hubbard, P. G., Vantassel, J. P., Cox, B. R., Rector, J. W., Yust, M. B. S., & Soga, K. (2022). Quantifying the surface strain field induced by active sources with distributed acoustic sensing: theory and practice. Sensors, 22(12)

Jousset, P., Reinsch, T., Ryberg, T., Blanck, H., Clarke, A., Aghayev, R., Hersir, G. P., Henninges, J., Weber, M., Krawczyk, C. M. (2018). Dynamic strain determination using fibre-optic cables allows imaging of seismological and structural features. Nature communications, 9(1), 1-11.

Kouroussis, Verlinden, and Conti. (2011). Free field vibrations caused by high-speed lines: measurement and time domain simulation, Soil Dynamics and Earthquake Engineering, 31, 692-707.

Krylov, V. V. (1995). Generation of ground vibrations by superfast trains. Applied Acoustics, 44, 149-164. https://doi.org/10.1016/0003-682X(95)91370-I

Kuvshinov, B. N. (2016). Interaction of helically wound fibre-optic cables with plane seismic waves. Geophysical Prospecting, 64(3), 671-688. https://doi.org/10.1111/1365-2478.12303

Lancelle, C. (2016). Distributed acoustic sensing for imaging near-surface geology and monitoring traffic at Garner Valley, California. (PhD Thesis), The University of Wisconsin-Madison.

Lombaert, and Degrande. (2003). The experimental validation of a numerical model for the prediction of the vibrations in the free field produced by road traffic, Journal of Sound and Vibration, 262, 309-331.

Lombaert, Degrande, Kogut, and François. (2006). The experimental validation of a numerical model for the prediction of railway induced vibrations, Journal of Sound and Vibration, 297, 512-535.

Li, S., Liu, B., Ren, Y., Chen, Y., Yang, S., Wang, Y., & Jiang, P. (2020). Deep learning inversion of seismic data. IEEE Transactions on Geoscience and Remote Sensing, 58(3), 2135-2149.

Martakis, Taeseri, Chatzi, and Laue. (2017). A centrifuge-based experimental verification of Soil-Structure Interaction effects, Soil Dynamics and Earthquake Engineering, 103, 1-14.

Madshus, C., & Kaynia, A. M. (2000). High-speed railway lines on soft ground: dynamic behaviour at critical train speed. Journal of Sound and Vibration, 231(3), 689-701.

Madshus, C., Lacasse, S., Kaynia, A. M., & Harvik, L. (2004). Geodynamic challenges in high-speed railway projects. Paper presented at the Geotechnical Engineering for Transportation Projects, Los Angeles, CA.

Mateeva, A., Lopez, J., Potters, H., Mestayer, J., Cox, B., Kiyashchenko, D., Wills, P., Grandi, S., Hornman, K., Kuvshinov, B., Berlang, W., Yang, Z., Detomo, R. (2014). Distributed acoustic sensing for reservoir monitoring with vertical seismic profiling. Geophysical Prospecting, 62, 679-692.

McMechan, G. A., & Yedlin, M. J. (1981). Analysis of dispersive waves by wave field transformation. Geophysics, 46(6), 869-874.

Mestayer, J., Grandi Karam, S., Cox, B., Wills, P., Mateeva, A., Lopez, J., Hill, D., Lewis, A. (2012). Distributed acoustic sensing for geophysical monitoring. Paper presented at the 74th EAGE Conference and Exhibition incorporating EUROPEC 2012.

Menq. (2003). Dynamic properties of sandy and gravelly soils, The University of Texas at Austin.

Nolet, G., & Panza, G. F. (1976). Array analysis of seismic surface waves: limits and possibilities. Pure and Applied geophysics, 114(5), 775-790.

Papadopoulos, Kuo, Germonpré, Verachtert, Zhang, Maes, Lombaert, and Degrande. (2019). Numerical prediction and experimental validation of railway induced vibration

in a multi-storey office building. In Proceedings of the 13th International Workshop on Railway Noise, 16-20 September 2019. Ghent: Springer.

Park, C. B., Miller, R. D., & Xia, J. (1999). Multichannel analysis of surface waves. Geophysics, 64(3), 800-808.

Park, C.B., (2005). MASW - Horizontal Resolution in 2D Shear-Velocity (Vs) Mapping. Kansas Geological Survey, Kansas.

Passeri, F. (2019). Development of advanced geostatistical models of shear wave velocity profiles to manage uncertainties and variabilities in Ground Response Analyses. (PhD thesis), Politecnico di Torino, Torino.

Pyl. (2004). Development and experimental validation of a numerical model for traffic induced vibrations in buildings, K.U. Leuven.

Rodriguez-Marek, Bommer, Youngs, Crespo, Stafford, & Bahrampouri. (2021). Capturing epistemic uncertainty in site response. Earthquake Spectra, 37(4), 921-936.

Robertson, P. K. (2009). Interpretation of cone penetration tests - a unified approach. Canadian Geotechnical Journal, 46(11), 1337-1355.

Schevenels. (2007). The impact of uncertain dynamic soil characteristics on the prediction of ground vibrations. KU Leuven.

Schevenels, M., Degrande, G., & François, S. (2009). EDT: an elastodynamics toolbox for MATLAB. Computers & Geosciences. https://bwk.kuleuven.be/bwm/edt

SESAME. (2004). Guidelines for the Implementation of the H/V Spectral Ratio Technique on Ambient Vibrations Measurements, Processing, and Interpretation. In (Vol. WP12-Deliverable D23.12): SESAME European research project.

Song, Z., Zeng, X., Thurber, C. H., Wang, H. F., & Fratta, D. (2018). Imaging shallow structure with active-source surface wave signal recorded by distributed acoustic sensing arrays. Earthquake Science, 31. DOI: 10.29382/eqs-2018-0208-4

Soga, K., & Luo, L., (2018). Distributed fiber optics sensors for civil engineering infrastructure sensing. J. Struct. Integr. Mainten., 3, 1–21. https://doi.org/10.1080/24705314.2018.1426138.

Stokoe, & Santamarina. (2000). Seismic-wave-based testing in geotechnical engineering. In ISRM International Symposium. International Society for Rock Mechanics.

Stokoe II, Cox, Clayton, & Menq. (2020). NHERI@UTexas Experimental Facility With Large-Scale Mobile Shakers for Field Studies. Frontiers in Built Environment, 6, 575973.

Stewart, Afshari, & Hashash. (2014). Guidelines for performing hazard-consistent one-dimensional ground response analysis for ground motion prediction. In PEER Report 2014.

Tao, & Rathje. (2019). Insights into Modeling Small-Strain Site Response Derived from Downhole Array Data. Journal of Geotechnical and Geoenvironmental Engineering, 145, 04019023.

Thompson, D. (2009). Railway Noise and Vibration: Mechanisms, Modelling and Means of Control. Elsevier Ltd.

Timoshenko, S. (1927). Methods of analysis of statical and dynamical stresses in rail. Paper presented at the Proceedings of the Second International Congress on Applied Mechanics, Zurich.

Vantassel, J. P., & Cox, B. R. (2022). SWprocess: A Workflow for Developing Robust Estimates of Surface Wave Dispersion Uncertainty. Journal of Seismology, 26(4), 731-756. https://doi.org/10.1007/s10950-021-10035-y.

Vantassel, J. P., Cox, B. R., Hubbard, P. G., & Yust, M. (2022a). Extracting High-Resolution, Multi-Mode Surface Wave Dispersion Data from Distributed Acoustic Sensing Measurements using the Multichannel Analysis of Surface Waves. Journal of Applied Geophysics, Volume 205, https://doi.org/10.1016/j.jappgeo.2022.104776

Vantassel, J. P., Kumar, K., & Cox, B. R. (2022b). Using convolutional neural networks to develop starting models for near-surface 2-D full waveform inversion. Geophysical Journal International, 231(1), 72-90. https://doi.org/10.1093/gji/ggac179

Vantassel, J., Cox, B., Hubbard, P., Yust, M., Menq, F. (2022c) "Active-Source, Near-Surface, Surface-Wave Measurements using Distributed Acoustic Sensing (DAS) and Traditional Geophones", in Characterization of the NHERI@UTexas Hornsby Bend Test Site. DesignSafe-CI. https://doi.org/10.17603/ds2-bz52-ep82 v1.

Verachtert, R. (2018). Deterministic and probabilistic determination of dynamic soil characteristics. (PhD thesis), KU Leuven.

Veletsos, & Meek. (1974). Dynamic behaviour of building-foundation systems. Earthquake Engineering & Structural Dynamics, 3, 121-138.

Yavuz, S., Freifeld, B. M., Pevzner, R., Tertyshnikov, K., Dzunic, A., Ziramov, S., Shulakova, V., Robertson, M., Daley, T. M., Kepic, A., Urosevic, M., Gurevich, B. (2016). Subsurface imaging using buried DAS and geophone arrays - preliminary results from CO2CRC Otway Project. Paper presented at the 78th EAGE Conference and Exhibition 2016.

Yust, M. B. S., Cox, B. R., Vantassel, J. P., & Hubbard, P. G. (2022). DAS for 2D MASW Imaging: A Case Study on the Benefits of Flexible Sub-Array Processing. Retrieved from arXiv preprint arXiv:2210.14261.

Yust, M. B. S., Cox, B. R., Vantassel, J. P., Hubbard, P. G., Boehm, C., & Krischer, L. (2023). Near-Surface 2D Imaging via FWI of DAS Data: An Examination of the Impacts of FWI Starting Model. Geosciences, 13(3), 63. https://doi.org/10.3390/geosciences13030063.

Yu, C., Zhan, Z., Lindsey, N.J., Ajo-Franklin, J.B. and Robertson, M., (2019). The Potential of DAS in Teleseismic Studies: Insights From the Goldstone Experiment. Geophys Res Lett, 46, 1320–1328, Blackwell Publishing Ltd. doi:10.1029/2018GL081195.

CHAPTER 6

SUMMARY AND CONCLUSIONS

This dissertation presents multiple advancements in non-invasive subsurface imaging techniques using seismic waves. The dissertation encompasses four key studies: the first and second studies introduce novel analysis techniques for estimating the small-strain shear modulus (G_{max}) and the small-strain damping ratio (D), respectively. The third study presents a unique dataset collected using state-of-the-art testing technologies. The fourth study provides an overview of current progress in non-invasive subsurface imaging using seismic waves. These studies are summarized in more detail below.

The first study introduces a frequency-velocity Convolutional Neural Network (CNN) developed for rapid, non-invasive 2D shear wave velocity (Vs) near-surface imaging using seismic waves. This CNN uses a normalized dispersion image as input to output a 2D-Vs image. The proposed framework offers significant flexibility in the linear-array, active-source experimental testing configuration used for generating the CNN input at a given site. The framework accommodates various source types, source offsets, numbers of receivers, and receiver spacings. Such acquisition flexibility allows for the use of the developed CNN as a comprehensive imaging technique or as a tool for rapidly generating starting models for full waveform inversion (FWI). The CNN was trained, validated, and tested using a total of 100,000 synthetic soil-over-rock models. The acquisition generalization ability of the proposed frequency-velocity CNN was initially demonstrated using sets of 5,000 synthetic near-surface models. For each set, the inputs to the CNN were dispersion images obtained using different testing configurations and near-surface models than those used for the CNN's training. The CNN exhibited remarkable

acquisition generalization ability with respect to the number of receivers, receiver spacings, source offset distances, and source forcing functions, as long as the testing configuration was not drastically different from the base case configuration on which the CNN was trained. Finally, the capability of the proposed CNN to handle field data was demonstrated using a field test conducted at the Hornsby Bend site in Austin, Texas, USA. The good agreement between the CNN's predicted 2D-*Vs* image and the actual subsurface structure determined through 1D surface wave inversions, Cone Penetration Test (CPT)-*Vs* correlations, and boring logs reinforces the capabilities of the proposed CNN for accurately retrieving 2D-*Vs* images using field data from testing configurations different from the one used during training.

For the second study, a new methodology for estimating frequency-dependent attenuation coefficients through the analysis of ambient vibration wavefield data recorded by 2D arrays of surface seismic sensors is presented. This approach, termed the noise Frequency-Domain Beamforming attenuation (NFDBFa) method, relies on an attenuation-specific wavefield conversion and frequency-domain beamforming (FDBF). Unlike noise cross-correlation methods, NFDBFa enables the determination of the ambient vibration propagation direction for each noise window and frequency without requiring an equally partitioned ambient noise wavefield. Furthermore, using an FDBF approach enables the phase velocity and attenuation data generated from active-source testing like the Multichannel Analysis of Surface Waves (MASW) to be combined with phase velocity and attenuation data generated from ambient noise testing like the Microtremor Array Measurements (MAM) in order to span a broader frequency range. This integration allows for joint inversions of phase velocity and attenuation to extract *Vs*

and *D* profiles to significantly greater depths than previously possible using only active-source data. Numerical simulations were conducted to deepen our understanding of the proposed NFDBFa method. These simulations aimed to evaluate how the proximity of the MAM array to the noise source, the presence of incoherent noise, and the size of the array affect the estimates of phase attenuation. The results demonstrated that near-field effects are more pronounced and extend over greater distances for phase attenuation estimates compared to those considered for phase velocity estimation. Furthermore, it was discovered that larger array sizes consistently provided more accurate phase attenuation estimates across all considered frequencies, contrary to the conventional MAM design criteria used for phase velocity dispersion estimation, where larger arrays are typically preferred for resolving lower frequencies while smaller arrays excel at resolving higher frequencies. This distinction emphasizes the need for unique design criteria when planning a MAM array for attenuation estimation. The proposed NFDBFa approach underwent validation through numerical wave propagation simulations, comparing predicted frequency-dependent phase attenuation values against theoretical phase attenuation curves for two synthetic models. Furthermore, the validation of the developed technique was reinforced using real MASW and MAM field data collected at the Drainage Farm site in Logan, Utah, USA. The phase velocity and attenuation data extracted from the MASW and MAM recordings agreed well over a common bandwidth, while the ambient noise MAM data allowed the phase velocity and attenuation estimates to be extracted at significantly lower frequencies. The joint inversion of the experimental Rayleigh-wave phase velocity and phase attenuation data obtained from both MASW and MAM testing facilitated the estimation of *Vs* and *D* profiles to significant depths (400 m)

at the Drainage Farm site. As noted herein and in other sources like Aimar et al. (2024a), attenuation data are significantly more variable and complex to understand (e.g., model curves that repeatedly cross one another) than phase velocity data. As such, there is a need for future studies to better understand attenuation data and how to invert it to retrieve reliable in-situ profiles of *D*. Future efforts should involve additional numerical and experimental testing under diverse subsurface conditions, coupled with comparisons to damping estimates from invasive tests. With the validity of this approach demonstrated on the vertical component, future research will also explore the utilization of the three components of the wavefield to enhance attenuation estimates beyond the current method's capabilities.

In the third study, a comprehensive subsurface imaging experiment conducted in Newberry, Florida, USA, using seismic waves is presented. The site exhibits spatial variability and features karstic surface and underground voids and anomalies, as documented in the literature. The sensing technologies employed included a dense 2D array of 1920 horizontal-component distributed acoustic sensing (DAS) channels and a 12 × 12 grid of 144 SmartSolo 3C nodal stations, covering an area of 75 m × 155 m. These technologies were used to record both active-source and passive-wavefield data. The active-source data were generated using various vibrational and impact sources, including a powerful triaxial vibroseis shaker truck, a 40-kg PEG-40 kg, and an eight-pound sledgehammer. The vibroseis shaker truck was utilized to vibrate the ground in three directions at 260 locations both inside and outside the instrumented area, while the impact sources were employed at 268 locations within the instrumented area.

Additionally, four hours of ambient noise were recorded using DAS. The nodal stations captured 48 hours of ambient noise in four 12-hour increments over four days.

Both the raw and processed datasets, along with comprehensive and detailed documentation of the experiment, have been archived and made publicly accessible. Examples of the data quality and potential use cases are provided to inspire current and future researchers who require a high-quality experimental dataset with known and potentially unknown anomaly locations for testing non-invasive imaging methods.

The fourth and final study included in this dissertation provides a review of the recent advancements in non-invasive subsurface imaging technologies that utilize seismic waves for site characterization. These advancements encompass both innovative sensing technologies and advanced 1D and 2D imaging techniques for retrieving G_{max} and D. In terms of sensing innovations, a special emphasis is given to the use of DAS as the data acquisition system for seismic wave sensing in near-surface imaging applications. As for imaging techniques, the joint inversion of attenuation and dispersion data is highlighted, employing 1D MASW in combination with either DAS or traditional sensing techniques such as geophones. This approach allows for the development of 1D *Vs* and damping profiles. The 2D imaging techniques presented include 2D MASW using DAS, machine learning for 2D subsurface imaging, and 2D full waveform inversion (FWI) using DAS data. All these advancements, whether in sensing technologies or analysis methods, were implemented at the well-characterized Hornsby Bend site in Austin, Texas, USA, facilitating a discussion on the advantages and disadvantages associated with each method.

APPENDICES

Appendix C: Authorship Permission

Paper I

Dear All,

I am in the process of preparing my dissertation in the Civil and Environmental Engineering Department at Utah State University. I am requesting your permission to include the journal paper "A frequency-velocity CNN for developing near-surface 2D Vs images from linear-array, active-source wavefield measurements", of which you are a coauthor, as a chapter in my dissertation. Please indicate your approval of this request.

Thank you,
Aser

Dr. Joseph Vantassel

Hi Aser,
I approve.

Dr. Krishna Kumar

Yes I approve

Dr. Jodie Crocker

Hi Aser,

I approve.

Paper II

Dear All,

I am in the process of preparing my dissertation in the Civil and Environmental Engineering Department at Utah State University. I am requesting your permission to include the journal paper “A frequency-domain beamforming procedure for extracting Rayleigh wave attenuation coefficients and small-strain damping ratio from 2d ambient noise array measurements”, of which you are a coauthor, as a chapter in my dissertation. Please indicate your approval of this request.

Thank you,

Aser

Dr. Mauro Aimar

Dear Aser,

On my part, I approve your request.

Mauro

Mauro Aimar, Ph.D., Eng.
Research Fellow

Dept. of Structural, Building and Geotechnical Engineering
Politecnico di Torino
Corso Duca degli Abruzzi 24,
10124 Torino (Italy)

Dr. Sebastiano Foti

Fine with me

Paper III

Dear All,

I am in the process of preparing my dissertation in the Civil and Environmental Engineering Department at Utah State University. I am requesting your permission to include the journal paper "An open-access data set of active-source and passive-wavefield das and nodal station measurements at the Newberry Florida site", of which you are a coauthor, as a chapter in my dissertation. Please indicate your approval of this request.

Thank you,

Aser

Dr. Khiem Tran

Hi Aser,

Thanks for asking. I approve your request.

Regards,
KT

Khiem T. Tran, Ph.D.
Associate Professor
Department of Civil & Coastal Engineering
University of Florida, Gainesville, FL 32611
Office: Weil Hall 265N, Phone: (352) 294-3369
Email: khiem.tran@essie.ufl.edu
https://faculty.eng.ufl.edu/khiem-tran/

Isabella Corey

Hi Aser,

I hope you are doing well. You have my permission.

Thanks,

Isabella
Nishkarsha Dawadi

Hi Aser,

Please go ahead. Thanks for asking.

Best,
Nishkarsha

Nishkarsha Dawadi
Ph.D. Student | Graduate Research Assistant
President, EERI Student Chapter
Department of Civil and Environmental Engineering
Utah State University
Phone: (435) 754-9721
Email: nishkarsha.dawadi@usu.edu

Paper IV

Dear All,

I am in the process of preparing my dissertation in the Civil and Environmental Engineering Department at Utah State University. I am requesting your permission to include the journal paper “Emerging technologies and advanced analyses for non-invasive near-surface site characterization”, of which you are a coauthor, as a chapter in my dissertation. Please indicate your approval of this request.
Thank you,
Aser

Dr. Mauro Aimar

Dear Aser,

On my part, I approve your request.

Mauro

Mauro Aimar, Ph.D., Eng.
Research Fellow

Dept. of Structural, Building and Geotechnical Engineering
Politecnico di Torino
Corso Duca degli Abruzzi 24,
10124 Torino (Italy)

Dr. Sebastiano Foti

Absolutely fine with me
Good luck !
Sebastiano

Dr. Michael Yust

Hi Aser,

Congratulations on finishing up you PhD. I am sorry I was not able to watch you defense but I hope it went well. You definitely have my permission to include the paper we coauthored in your dissertation. Please let me know if there is anything else you need from me and I will do so as soon as possible.

Best, Michael

Aser Abbas

Website: www.aserabbas.com
Phone: +1 (256) 468-2493
Email: aser.abbas@usu.edu

EDUCATION

Ph.D. in Civil and Environmental Engineering (Geotechnical) Aug 2024
Logan, Utah, USA
Utah State University (USU)
Dissertation: *Emerging Technologies and Advanced Analyses for Non-invasive Near-surface Site Characterization*
Advisor: Brady Cox | GPA: 4.0/4.0

M.S. in Systems and Materiel Engineering Aug 2020
Alabama A&M University Huntsville, Alabama, USA
Research: *Serviceability of piles under compression and downdrag forces*
Advisor: Mohamed Ashour | GPA: 4.0/4.0

M.S. in Structural Engineering Aug 2017
Mansoura University Mansoura, Egypt
Thesis: *Pile response under axial tension force in sandy soil*
Advisor: Mohamed Ashour

B.S. in Civil Engineering May 2012
Mansoura University Mansoura, Egypt
Graduation Project: *Highway and Airport Engineering*

RESEARCH EXPERTISE

Research interests:

My research interests can be grouped into two broad areas. The first area involves in-situ dynamic site characterization of soils and rocks, primarily for seismic site response analysis. The second area focuses on soil-pile interaction. In these domains, I have contributed to 12 peer-reviewed journal articles, 7 technical reports, 1 open-access dataset, 1 conference paper, and several software packages. My research has been applied in several high-profile projects, such as the Champlain Towers South collapse forensic investigation, underscoring its significant real-world impact.

Specializations:

- In-situ subsurface imaging using noninvasive and invasive seismic wave techniques
- Advanced sensing technologies (e.g., distributed acoustic sensing (DAS), three-component seismometers, direct-push crosshole sensors, etc.)
- Numerical and analytical modeling, machine learning, high performance computing
- Anomaly detection using Full Waveform Inversion (FWI)
- Fundamental and applied research in geotechnical earthquake engineering
- Structural response to seismic forces
- Soil-pile interaction (e.g., mobilized axial and/or lateral responses, downdrag, pile-group-cap systems, etc.)
- Programing using: Python, C#, Visual Basic, and Fortran

EXPERIENCE

Graduate Research and Teaching Assistant Aug 2020 – Aug 2024
Utah State University Logan, Utah

- Deep (+400 m) shear wave velocity profiling for the Palisades nuclear power plant, Michigan, USA
- Forensic soil investigation of the Champlain Towers South collapse, Florida
- Deep (+1 km) shear wave velocity profiling for the Duynefontein nuclear power plant, South Africa
- 3D imaging and seismic site response of Salt Lake City, Utah
- Dynamic soil characterization for a large bridge interchange, South Carolina
- Deep (+200 m) shear wave velocity profiling with surface waves for Temple Square buildings, Salt Lake City, Utah
- Subsurface anomaly detection for Texas DOT, Texas
- Establish the Earthquake Engineering and Subsurface Imaging Lab at USU with Professor Brady Cox
- Develop machine learning techniques for subsurface imaging
- Develop a noninvasive testing approach for characterizing in-situ damping
- Serve as a frequent guest lecturer for the Deep Foundations and the Site Characterization courses
- Assist Professor Brady Cox in developing teaching materials for the Deep Foundations course
- Mentor graduate and undergraduate students

Graduate Research and Teaching Assistant Jan 2019 – Jul 2020
Alabama A&M University Huntsville, Alabama

- Fulfilled the role of a teaching assistant in the following courses at Alabama A&M University: Foundation Design, Soil Mechanics, Soil Lab, and Hydraulic Design
- Researched a new methodology for characterizing the permanent load-displacement relationship of piles subjected to downdrag

- Developed a methodology for characterizing the settlement and strength of inundated collapsible soils
- Developed a semi-analytical approach for estimating the settlement and rotation of pile group cap systems, when the piles have different stiffnesses due to variations in pile lengths and/or diameters
- Developed the proprietary software utilized by the Alabama DOT for the design of axially loaded piles and pile groups
- Updated the Strain Wedge Model (SWM) technique, which is used for lateral pile analysis by various companies and DOTs in the US and worldwide
- Developed software for evaluating the downdrag loads on bridge pile foundations in inundated collapsible soils, completed for the National Center for Transportation Infrastructure & Life-Extension (TriDurLE)

Structural Engineer - Owner Representative Dec 2012 – Dec 2018
Egyptian Maintenance Company (EMC) as Technical Support for PetroBel Egypt

Duties:

- Led the structural engineering design team at PetroBel's South Sinai division
- Ensured that projects' specifications were met during the construction phase by conducting regular inspections
- Ensured compliance with health, safety, and legal regulations throughout all project stages
- Led a team of junior engineers and technicians in supervising all civil engineering-related projects in the Belayim area, South Sinai, ensuring their successful execution
- Prepared and presented detailed status reports, closely monitoring budgets, contractors, and schedules

Selected projects:

- Designed and supervised the construction of the foundations for over fifty oil rigs and manifolds
- Inspected the use of conductor pipes as pile foundations for an offshore platform (Group action)
- Designed and supervised the construction of several concrete residential buildings
- Oversaw the repair of Patrico facilities during Phase two
- Reviewed designs and supervised the construction of Belayim’s seawater treatment plant
- Reviewed designs and supervised the construction of Belayim’s wastewater treatment plant
- Reviewed designs and supervised the construction of Belayim's sewage network, including the collection sump Achievements:
- Increased the percentage of in-house-made designs by over 100%, resulting in a promotion to lead design engineer in the South Sinai division
- Through meticulous revision of consultancy office projects' designs, successfully reduced construction costs by an average of 30%

HONORS AND AWARDS

- Awarded the Doctoral Researcher of the Year by USU's Civil and Environmental Engineering Department (2024)
- Awarded the Doctoral Researcher of the Year by USU's College of Engineering (2024)
- Selected as one of the top six best Doctoral Researchers at USU (2024)
- Received the W.O. Carter Scholarship for graduate students in Structural and Geotechnical Engineering (2021)
- Awarded the Outstanding Research Achievement Certificate by the Chairman of EMC Company (2017)

PEER-REVIEWED JOURNAL ARTICLES

In Preparation:

[J13] **Abbas, A.**, Aimar, M., Cox, B.R., Foti, S., (2023) *A Frequency-Domain Beamforming Procedure for Extracting Rayleigh Wave Attenuation Coefficients and Small-Strain Damping Ratio from 2D Ambient Noise Array Measurements*" submitted to Earthquake Spectra.

Published Articles:

[J12] **Abbas, A.**, Aimar, M., Yust, M., Cox, B.R., Foti, S., (2023) "*Emerging Technologies and Advanced Analyses for Non-Invasive Near-Surface Site Characterization*" Soils and Rocks. DOI: 10.28927/SR.2024.006923. link

[J11] **Abbas, A.**, Cox, B. R., Tran, K. T., Corey, I., Dawadi, N. (2023) "*An Open-Access Dataset of Active-source and Passive-wavefield DAS and Nodal Station Measurements at the Newberry Florida Site*." Seismological Research Letters (SSA). link

[J10] **Abbas, A.**, Vantassel, J., Cox, B.R., Kumar, K., Crocker, J., (2023) "*A Frequency-Velocity CNN for Developing Near-Surface 2D Vs Images from Linear-Array, Active-Source Wavefield Measurements*" Computers and Geotechnics (Elsevier). DOI: 10.1016/j.compgeo.2023.105305. link

[J9] Xu, X., **Abbas, A.**, Lee, J., (2022) "*Characterization of Fracture Process Zones in Scaled Centre-Notched Quasi-Isotropic Carbon/Epoxy Laminates Using a Convolutional Neural Network*" Engineering Fracture Mechanics (Elsevier). DOI: 10.1016/j.engfracmech.2022.108768. link

[J8] Ashour, M., **Abbas, A.** (2021) "*Mobilized Response of Piles Subjected to Downdrag*" International Journal of Geomechanics (ASCE). DOI: 10.1061/(ASCE)GM.1943-5622.0002087. link

[J7] Ashour, M., **Abbas, A.** (2020) "*Response of Piles in Multilayers of Soil Under Uplift Forces*" International Journal of Geomechanics (ASCE). 0.1061/(ASCE)GM.1943-5622.0001676. link

[J6] Ashour, M., **Abbas, A.,** (2020) "*Axially Loaded Piles in Inundated Collapsible Soils under Compression and Tension Forces*" Canadian Journal of Civil Engineering (NCR). https://doi.org/10.1139/cjce-2019-0506. link

[J5] Ashour, M., **Abbas, A.**, Altahrany, A., Alaaeldin, A. (2020) "*Modeling the Behavior of Inundated Collapsible Soils*" Engineering Reports (Wiley). https://doi.org/10.1002/eng2.12156. link

[J4] Ashour, M., **Abbas, A.**, Boskovic, S. (2019) "*Pile Cap Interaction With Bridge Pile Foundations Under Lateral Loads*" Journal of Bridge Engineering (ASCE). 10.1061/(ASCE)BE.1943-5592.0001408. link

[J3] Ashour, M., Altahrany, **Abbas, A.** (2019) "*Development Of T–Z Curve for Piles in Sands Under Uplift Force*" Innovative Infrastructure Solutions (Springer). 10.1007%2Fs41062-019-0210-7. link

[J2] **Abbas, A.**, Ashour, M., Altahrany, A. (2017) "*Behaviour of Piles Subjected to Uplift Loads in Sand*" New York Science Journal, Marsland press. 10.7537/marsnys100517.02. link

[J1] **Abbas, A.**, Ashour, M., Altahrany, A. (2017) "*Pile Response Under Axial Tension Forces in Sandy Soils*" Journal of Bridge Engineering (ASCE). 10.1061/(ASCE)BE.1943-5592.0001142. link

DATASETS

[D2] **Abbas, A.**, Cox, B., Dawadi, N., Jackson, N., Cannon, K., (2024) "*Geotechnical site characterization at the Drainage Farm Site.*" DesignSafe-CI. https://doi.org/10.17603/ds2-sx2h-8s20 v1

[D1] **Abbas, A.**, B. Cox, K. Tran, I. Corey, N. Dawadi, F. Menq. (2023) "*Active-source and Passive-wavefield DAS and Nodal Station Measurements at the Newberry Florida Site.*" DesignSafe-CI. link

CONFERENCE PAPERS

[C1] Ashour, M., Singh, J., **Abbas, A.** (2018) "*Load Transfer Curve of Piles in Sands under Uplift Forces*" Proceedings of the 43rd Annual Conference on Deep Foundations, 2018, Anaheim, CA, USA, (DFI).

REPORTS

[R8] Cox, B., Dawadi, N., **Abbas, A.,** (2024) " *Deep Shear Wave Velocity Profiling Using MASW and MAM Surface Wave Methods: Palisades NPP Site* " Geotechnical Engineering Report COX24-01, Utah State University, Logan, Utah.

[R7] Ashour, M., Bhattacharjee, S., **Abbas, A.** (2024) "*Assessment and Evaluation of Post-Liquefaction Lateral Spread Impact on Bridge Deep Foundations*" National Center for Transportation Infrastructure Durability & Life-Extension (TriDurLE). - In Progress

[R6] Cox, B., **Abbas, A.,** (2023) "*Deep Shear Wave Velocity Profiling Using MASW and MAM Surface Wave Methods: Duynefontein Project, South Africa*" Geotechnical Engineering Report COX23-01, Utah State University, Logan, Utah.

[R5] Cox, B., **Abbas, A.,** Aimar, M., Foti, S., (2022) "*Champlain Towers South Soil Attenuation/Damping Measurements*" Geotechnical Engineering Report COX22-02, Utah State University, Logan, Utah.

[R4] Cox, B., Cannon, K., Jackson, T., **Abbas, A.,** (2022) "*Deep Shear Wave Velocity Profiling Using MASW and MAM Surface Wave Methods: Assembly Hall on Temple Square, Salt Lake City, Utah*" Geotechnical Engineering Report COX22-01, Utah State University, Logan, Utah.

[R3] Ashour, M., **Abbas, A.,** (2022) "*Evaluation of downdrag loads on bridge pile foundations in inundated collapsible soils*" National Center for Transportation Infrastructure Durability & Life-Extension (TriDurLE).

[R2] Cox, B., **Abbas, A.,** (2021) "*Deep Shear Wave Velocity Profiling Using MASW and MAM Surface Wave Methods: North Myrtle Beach, South Carolina*" Geotechnical Engineering Report, Utah State University, Logan, Utah.

[R1] Ashour, M., **Abbas, A.,** Bhattacharjee, S. (2021) " *Serviceability of Piles under Compression and Downdrag Forces* " Alabama Department of Transportation (ALDOT) Contract No. 930-978.

SOFTWARE DEVELOPMENT PROJECTS

CNN for 2D subsurface imaging:

A frequency-velocity convolutional neural network (CNN) for rapid, non-invasive 2D shear wave velocity imaging of near-surface geo-materials. The CNN uses a normalized dispersion image from MASW testing as an input and outputs a 2D subsurface Vs image. This framework provides significant flexibility in the linear-array, active-source experimental testing configuration used in generating the CNN input at a given site, accommodating various source types, source offsets, numbers of receivers, and receiver spacings

DASPy:

A Python module for processing DAS collected waveforms (e.g., converting DAS raw data to engineering units, filtering, trimming, etc.)

attPy:

A Python module for estimating in-situ attenuation using frequency domain beam forming (FDBF) and microtremor array measurements (MAM)

WBuzzPile:

A proprietary software utilized exclusively by the Alabama DOT for the design of axially loaded piles and pile groups. WBuzzPile is developed using the Visual Basic programming language

SWM (CGI-DFSAP):

The SWM technique is referenced by the FHWA (2006, 2010, 2016) and employed by various DOTs, including the Washington DOT and Nevada DOT, for modeling the lateral response of piles and pile groups. I Incorporated the capability of modeling the lateral response of the pile-group-cap system into the strain wedge model technique (SWM). The software engine is written in FORTRAN, while the interface is developed using a combination of C# Windows Forms application and Visual Basic

AMR:

The Axial Mobilized Response (AMR) is my personal pile modeling software, currently utilized in teaching the Deep Foundations course at USU. The software is developed using C#, and the interface is implemented using Windows Presentation Foundation (WPF)

In-situ site characterization:

A Python package that includes multiple modules for interpreting SPT, CPT, and DMT data. It also incorporates statistical analysis for generating soil cross-sections and facilitates the generation of plots to aid in subsurface interpretations

TRAINING CERTIFICATES

- Completed the DesignSafe Academy training and NHERI Hackathon on High-Performance Computing (HPC) in Austin, Texas, USA (2022)

- Competent scaffolding inspector (Scaffold Training Institute. Texas, USA) (2018)

ACADEMIC AND COMMUNITY SERVICE

- Co-founded and currently serve as the inaugural president of the Earthquake Engineering Research Institute (EERI) student chapter at USU (2022 - 2024)
- Reviewed journal articles for the Soil Dynamics and Earthquake Engineering journal and Computers and Geotechnics Journal
- Selected by the Civil and Environmental Engineering Department to administer the Engineering State (E-State) earthquake engineering challenge at USU in the summer of 2023
- Mentored indigenous students in the summer of 2023 at USU
- Chosen by the Natural Hazards Engineering Research Infrastructure (NHERI) to mentor students in their Research Experience for Undergraduates (REU) internship program during the summer of 2022
- Served as a foreign associate with the National Institute of Standards and Technology (NIST) on the forensic investigation of the Champlain Towers South collapse project (2021-2022)

SELECTED COURSES TAKEN

- In-situ site characterization
- Geotechnical earthquake engineering
- Shear strength of soils
- Consolidation settlement
- Ground reinforcement, improvement, and treatment
- Deep foundations
- Structural dynamics and seismic design
- Inverse theory
- Theory of plasticity
- Theory of elastic stability
- Finite element analysis
- Statistical methods for engineers
- Engineering project management
- Advanced concrete design