AUTOMATING NEAR-SURFACE CHARACTERIZATION USING AMBIENT NOISE DISTRIBUTED ACOUSTIC SENSING DATA

by

Nikhil Punithan

A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Master of Science (Geophysical Engineering).

Golden, Colorado

Date ________________________

Signed: ________________________

Nikhil Punithan

Signed: ________________________

Dr. Eileen R. Martin
Thesis Advisor

Signed: ________________________

Dr. Jeffrey Shragge
Thesis Advisor

Golden, Colorado

Date ________________________

Signed: ________________________

Dr. Paul C. Sava
Department Head
Department of Geophysics

ABSTRACT

As urban populations grow, geohazards that affect the shallow subsurface increasingly impact more people and infrastructure. These hazards can impact large regions, and occur on a wide variety of time scales. This motivates a large-scale, repeatable time-lapse methods to monitor the subsurface. Typical near-surface seismic exploration methods fall short as land access for geophone deployment may be impossible in developed areas, and active seismic data acquisition can be both logistically challenging and expensive to repeat.

Distributed Acoustic Sensing (DAS) using existing telecommunications infrastructure mitigates the data acquisition problem, as buried fiber optic cables are abundant in dense urban environments. Since these installations are semi-permanent, ambient seismic data can be collected over long time scales, enabling the use of seismic interferometry to retrieve surface wave Green's function approximations. Large-scale near-surface characterization can be conducted using Multichannel Analysis of Surface Waves (MASW) at many locations, followed by stitching together neighboring 1-D shear wave velocity profiles. Since processing and analyzing large amounts of seismic data may be infeasible for small teams of geotechnical engineers and infrastructure managers, an automated workflow can be employed to facilitate large-scale near-surface characterization with limited resources.

In this thesis, I explore an automated approach to near-surface characterization using passive DAS data. I automatically pre-process passive DAS data to remove unwanted coherent signals. I then use seismic interferometry between successive channel pairs to create noise correlation functions along the fiber array. Surface wave dispersion analysis is conducted automatically through a combination of threshold filtering and density-based spatial clustering. Lastly, I invert each dispersion curve for a 1-D shear wave velocity profile to create a pseudo 2-D shear-wave velocity section. By applying this workflow to two dark fiber arrays, I show that large scale shear wave velocity interfaces can be resolved, however clusters of 1-D profiles show unrealistic lateral heterogeneities that can be attributed to challenges in mode picking. I also compare results between two subsets of data from each fiber to show the time-lapse repeatability of the automated processing workflow.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Density-Based Cluster of Applications with Noise	DBSCAN
Distributed Acoustic Sensing	DAS
Fiber A, first time period	FA_1
Fiber A, second time period	FA_2
Fiber B, first time period	FB_1
Fiber B, second time period	FB_2
Gauge Length	GL
Interrogator Unit	IU
Multichannel Analysis of Surface Waves	MASW
Noise Correlation Function	NCF
Root-mean-square	RMS
Root-mean-square deviation	RMSD
Shear wave velocity	V_s
Spectral Analysis of Surface Waves	SASW

ACKNOWLEDGMENTS

Firstly, I would like to thank my advisor, Dr. Eileen Martin for the abundance of knowledge and time that she has shared with me over the past two years. Her support and advice was essential to inspiring my dedication to this work and in shaping the beginnings of my research career. I also want to thank my co-advisor, Dr. Jeffrey Shragge. His attention to detail and positive outlook towards research served as a driving force for me to try new things within my studies. I have to thank both of my advisors for giving me the opportunity to learn about urban fiber optic sensing and helping me find out the kind of researcher that I want to become.

I'd like to thank my committee members Dr. Rich Krahenbuhl and Dr. David Wald for their perspective and assistance on this project. I'd also like to thank Ivan Lim Chen Ning who worked with me and helped me improve as both a grad student and researcher. I owe a big thanks to the members of Martin Group: Ahmad, Tomas, Rachel, Shihao, Hafiz, Yida, Reynaldo, Georgia, Badr, Jimmy. You all made weekly group meetings a load of fun and something I actively looked forward to. I'm grateful to have such good friends in Arsya, Cullen, Nick, Brett, Hannah, Zach, Samara, and Reinaldo. You all played an important role in showing me what it means to work hard and conduct good, interesting research while also making room for some fun.

Grad school would not have been possible for me without the support of my parents. Their endless support made this work possible. They were the ones who first pushed me to apply for Mines and I am eternally grateful for all they have done for to get me to this point. Above all I'd like to thank my Brother, Nithin. You have been there by my side all this time and for that I must say that this achievement is yours as much as it is mine.

For my Brother.

CHAPTER 1

INTRODUCTION AND BACKGROUND

Transportation networks and civil infrastructure can face damage from flooding, liquefaction, slope failure, earthquakes, and many other natural phenomena. These geohazards can affect the subsurface on a wide range of time scales and spatial scales. Surface-wave vibrations during earthquakes can cause damage on city-wide scales in a matter of seconds to minutes. In contrast, rock slope failure can occur on the scale of minutes to months [1]. Furthermore, earthquakes can trigger secondary geohazards long after the initial event has ended, such as landslides [2].

Over large distances and long time scales, active near-surface seismic exploration can be inefficient. Additionally, survey crews may find it difficult to install and acquire data from traditional geophones. Paved surfaces can make proper ground coupling untenable and urban foot and vehicle traffic can generate noise that interferes with active shot records.

In some frequency ranges, the ambient seismic wavefield can replace active source data acquisition by retrieving approximate Green's functions via seismic interferometry [3, 4]. Distributed Acoustic Sensing (DAS) can be used to replace geophones or seismometers to acquire data. Existing telecommunications infrastructure, also known as dark fiber, can be leveraged to sample DAS data in urban environments [5]. The combination of these two can be used for near-surface characterization in urban environments [6, 7].

1.1 Distributed acoustic sensing

DAS functions by sending pulses of light through a fiber optic cable using an Interrogator Unit (IU). Impurities within the fiber made during manufacturing can backscatter the light pulses which are then received by the IU. When seismic waves interact with the fiber optic cable, the fiber compresses or extends, causing the locations of these impurities to shift with the cable, resulting in a phase change of the backscattered light pulse. This phase change is then related to strain or strain rate which allows for the estimation of the particle velocity of the transiting wave. DAS enables seismic sensing at high spatial and temporal rates. Fiber optic cables can cover 10s of kilometers that can be converted into a large-scale distributed network of seismic sensors. Additionally, cables can be installed semi-permanently which facilitates time-lapse (4-D) investigation. By using dark fiber, large-scale, 4-D surveying can be accomplished in urban environments without the costs associated with repeatedly deploying new fiber optic cables. With this, the ambient seismic wavefield can be sampled over long time periods and over large distances. When using DAS, the fiber optic cable is effectively partitioned into short segments, which are called channels. These channels record strain or strain rate data that are averaged over a distance called a Gauge Length (GL). Shorter GLs improve spatial resolution but reduce the signal-to-noise ratio (SNR). Longer GLs improve SNR at the cost of spatial resolution. GLs between 8 m and 10 m are typically used to optimize the trade-off between SNR and spatial resolution when sampling seismic waves [8].

1.2 Seismic interferometry

The 1-D case of seismic interferometry involves a plane wave propagating through two geophones separated by some distance. The traces from each geophone show the arrival of this wave, separated in time by a time lag. The cross-correlation of these traces results in a new trace that has an arrival at the time lag between the arrivals on the original traces. In 2-D, a single plane-wave source is replaced by many sources randomly distributed around the receivers. The expected travel time of a wave traveling between these two receivers can be retrieved assuming random, uncorrelated noise sources evenly distributed around the receivers [4]. As such, ambient seismic data can be used to retrieve approximate Green's function via cross-correlation of pairs of receivers. By cross-correlating a single channel with numerous adjacent channels, a Noise Correlation Function (NCF) is constructed with a stationary virtual source. These data are similar to active-source data that would be acquired if a source were applied at the virtual source location. In a near-surface setting, the ambient wavefield is dominated by surface waves that decay more slowly than body waves, which leads the signals reconstructed by seismic interferometry to be dominated by surface waves. This process can also be applied to DAS data to retrieve Rayleigh wave travel times assuming the use of co-linear fiber segments. As DAS is a one-component strain sensor, a mixture of modes can occur when using segments with varying geometries. Assuming an ideal ambient energy source distribution, the NCFs of an array of co-linear fiber channels yields primarily Rayleigh waves. The resulting signals can then be analyzed by surface-wave processing tools such as Multichannel Analysis of Surface Waves (MASW) [9].

1.3 Near-surface characterization using surface waves

Surface waves are elastic waves that propagate along the Earth's surface. Rayleigh waves are a subset of surface waves that exhibit elliptical particle motion. A property of Rayleigh waves is that their lower-frequency components probe deeper into the subsurface while higher frequency components "see" shallower. Since the near-subsurface velocities typically increase with depth, the different wave components travel at different velocities which is known as dispersion. This property has been used to infer subsurface shear-wave velocity (V_s) assuming horizontal layering.

Spectral Analysis of Surface Waves (SASW) is a method that involves recording Rayleigh waves across a pair of geophones to construct a dispersion curve which maps the surface-wave phase velocities as a function of frequency [10]. However, this method did not initially account for higher-order modes. This method has since been used to identify an "apparent" dispersion curve [11].

MASW was developed to overcome the limitations of SASW. The multichannel approach enables field crews to acquire data in the form of active-source shot gathers or through the use of ambient energy through NCFs [6, 12]. The dispersive energy in the data can be analyzed through slant stacking or a phase-shift technique which transforms the multichannel recording from a time-distance to a phase velocity-frequency representation [9]. From this output, mode picking and labeling typically occur manually or with the aid of commercial or open-source software [13, 14]. The picked mode(s) are then inverted for a 1-D V_s subsurface profile, again assuming horizontal layering. These methods can be expanded along multiple subarrays to construct pseudo-2-D V_s profiles [15, 16]. MASW has been useful in non-invasive estimation of near-surface velocities for use in earthquake hazard site classification [12] and for geotechnical studies [17].

A $V_{s_{30}}$ estimate is the average V_s of the top 30 m of the subsurface. It was first used as an empirical metric for ground stability in geotechnical site assessment and has since found use in site classification for earthquake hazards as well as in ground motion amplification studies [18, 19].

1.4 Thesis contributions

Through the combined use of DAS, seismic interferometry, and MASW, near-surface characterization can be achieved over long distances. However, in practical applications, infrastructure managers and geotechnical engineers may not have the specific skills or workforce needed to calculate and analyze hundreds of NCFs. As such, an automated workflow can be employed to improve the accessibility of using these technologies for near-surface characterization.

In this thesis, I introduce a semi-automated workflow that leverages ambient DAS data and MASW to infer pseudo-2-D near-surface V_s maps. I then implement this workflow on data acquired by two different dark-fiber arrays. Time-lapse repeatability is assessed by comparing the models generated from ambient wavefield energy recorded during two different time periods over which no subsurface changes were expected. Additionally, convergence of NCFs is assessed to determine the optimal stack duration. Furthermore, I show the strengths and drawbacks of this semi-automated workflow and describe a path forward to improving its reliability. Previous studies using similar methods have shown promising near-surface characterization [16, 20] and time-lapse repeatability [6] results. However, these cases present fibers with primarily straight geometries, locations where the distribution of seismic sources and local geology are known, as well as with fiber lengths of under 1 km. In this thesis, I show the challenges associated with semi-autonomous near-surface characterization in a non-ideal scenario where fiber geometry is not easily segmented into straight sections, local geology is largely unknown, and with fiber lengths greater than 5 km.

1.5 Thesis outline

This thesis is separated into three chapters. The first chapter is the introduction and background where an overview of DAS, seismic interferometry, and ground characterization through surface wave methods are described. The case for a semi-automated workflow using these methods is also stated in this chapter. The second chapter details the semi-autonomous workflow and its application to two data subsets from two different fiber arrays. This includes analyzing the results of individual pseudo 2-D V_s maps as well as comparison between results on the same fiber. This chapter is to be submitted in *Geophysics*. The third chapter presents the conclusions and discusses potential areas of future research. These include improvements to preprocessing steps, multi-modal dispersion curve inversions and near real-time near-surface characterization.

CHAPTER 2

AUTOMATED INTERFEROMETRIC NEAR-SURFACE CHARACTERIZATION USING DISTRIBUTED ACOUSTIC SENSING ON DARK FIBER ARRAYS

From a paper to be submitted to *Geophysics*

Nikhil Punithan[1,2], Eileen R. Martin[3], Jeffrey Shragge[4], Ahmad Tourei[5], Ivan Lim Chen Ning[6]

2.1 Abstract

Seismic imaging in near-surface environments is a useful tool in monitoring groundwater, infrastructure stability, and geohazard studies over long distances – especially when using preinstalled telecommunications fiber to acquire ambient seismic wavefield data. Processing these data, though, continues to include many manual steps, hindering the ability to rapidly characterize the subsurface over large spatial extents when addressing pressing environmental and infrastructure challenges. We present an automated workflow for generating 2-D shear-wave velocity (V_s) profiles using ambient wavefield energy acquired through distributed acoustic sensing (DAS). Seismic interferometry is used to generate virtual sources along the fiber. The resulting noise correlation functions are analyzed using an automated multichannel analysis of surface waves approach to generate pseudo 2-D V_s models. We apply this automated workflow on DAS data acquired during different time windows from one fiber to show repeatability. We also apply this workflow on two different fibers to highlight the successes and ongoing challenges associated with workflow generalization for multiple fiber geometries. We show that large-scale subsurface feature can be resolved using this method; however, individual 1-D profiles can show potentially unrealistic lateral heterogeneities.

2.2 Introduction

The near-surface environment is susceptible to geohazards that may adversely affect urban populations and infrastructure. Liquefaction, ground subsidence, and sinkholes can cause substantial damage over a wide range of time scales and may not be initially visible at the surface. This issue motivates the development of methods capable of estimating subsurface properties in areas with elevated geohazard risk that can cover expansive urban locations while facilitating repeatability on time scale ranging from weeks to months.

---

[1] Primary researcher and author

[2] Author for correspondence

[3] Associate Professor, Colorado School of Mines

[4] Professor, Colorado School of Mines

[5] Graduate Student, Colorado School of Mines

[6] Research Geophysicist, Chevron U.S.A. Inc.

Conventional near-surface seismic investigations typically fall short in both of these deployment challenges as cabled geophones or nodal systems may not be suitable for urban settings and seismic land streamers lack significant range for long-distance monitoring. Additionally, near-surface active-source seismic data acquisition can be prohibitively expensive to repeat for long-term monitoring due to the costs associated with survey personnel and hardware.

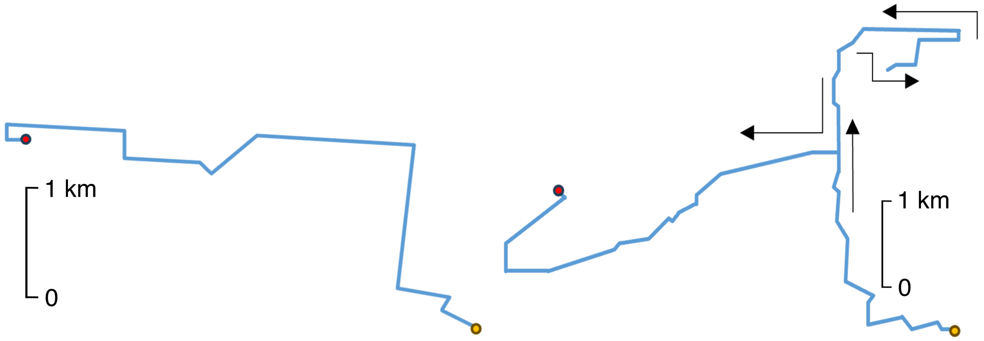
Distributed acoustic sensing (DAS) in the form can be a substitute for geophone- or nodal-based seismic acquisition. DAS enables fine-scale spatial sampling of broadband seismic waves across kilometer-scale distances and over long periods of time [7]. Furthermore, exploiting existing semipermanent telecommunications infrastructure in so-called "dark fiber" arrays can reduce seismic sensor deployment and maintenance costs [16]. One can also replace active seismic energy sources with ambient seismic energy, which have proven useful for retrieving empirical Green's functions by comparing signals recorded by pairs of sensors [3, 4, 21]. Numerous studies have investigated ambient DAS seismic methods for near-surface characterization in urban settings using multichannel analysis of surface waves (MASW) [6, 22–24]. These methods have proven useful in estimating $V_{s_{30}}$ values in site classification for earthquake hazards and ground motion amplification studies [18, 19]. These estimates are also used within the geotechnical field to non-invasively assess site conditions [25].

Logistical challenges, such as long and complex fiber geometries as well as large data volumes, may prevent the generation and analysis of large numbers of dispersion curves resulting from ambient wavefield analysis. This motivates the development of automated analysis methods that can be employed to improve the feasibility of urban near-surface characterization using DAS data. To address this, we automate passive seismic data preprocessing and surface wave analysis into a data processing workflow that requires minimal human input. Aside from user determined inputs and information on the geometry of individual fibers, we automate a raw DAS data to shear wave velocity profile processing scheme.

This paper presents a case study using ambient DAS data, seismic interferometry, and MASW for (semi-)automated near-surface characterization. We demonstrate the effectiveness of this framework on ambient seismic data from two dark fiber arrays. We compare the shear wave velocity profiles generated from two different data subsets on the same fiber to assess time-lapse repeatability of our automated workflow. We test our workflow on a second fiber to understand the generalizability and limitations that exist with using automated methods in variable conditions. We assess potential failure points within the semi-automated workflow and identify avenues to remedy these shortcomings.

2.3 Data Acquisition

Seismic data was collected for this experiment using DAS. Passive seismic data was acquired on two separate fiber locations: fibers A and B with a relative deployment geometry shown in Figure 2.1. Fibers A and B have measured lengths of 8,352 m and 21,725 m, respectively. Both survey lines have buried fiber sections as well as portions of aerial fiber secured on poles. The fiber A and B data sets were recorded with the same acquisition parameters of 2-m channel spacing, 10-m gauge length, and a 500 Hz sampling rate. DAS data were recorded on fibers A and B for 14 and 7 days, respectively. The recorded data were partitioned into 30-s files that serve as the time-window duration for the subsequent ambient wavefield processing.



(a) (b)

Figure 2.1 Deployment geometry schematics for fibers (a) A and (b) B. The yellow and red circles respectively indicate the start and end locations on each fiber.

For this study, two 36-hour data subsets were chosen from each fiber for analysis (Section 2.9 - Appendix A analyzes shorter duration windows of the ambient data subsets). For fiber A, the first and second data analysis periods were 11 days apart while for fiber B, we chose data acquired 7 days apart. These time periods were selected to investigate the time-lapse repeatability of the data processing workflow as no notable subsurface changes are expected within this time span. These data subsets are referred to below as FA_1 and FA_2 for fiber A and FB_1 and FB_2 for fiber B.

2.4 Methods

This section details the automated processing steps taken to create pseudo 2-D V_s models from raw ambient DAS data. The automatic data processing workflow is detailed in Figure 2.2.

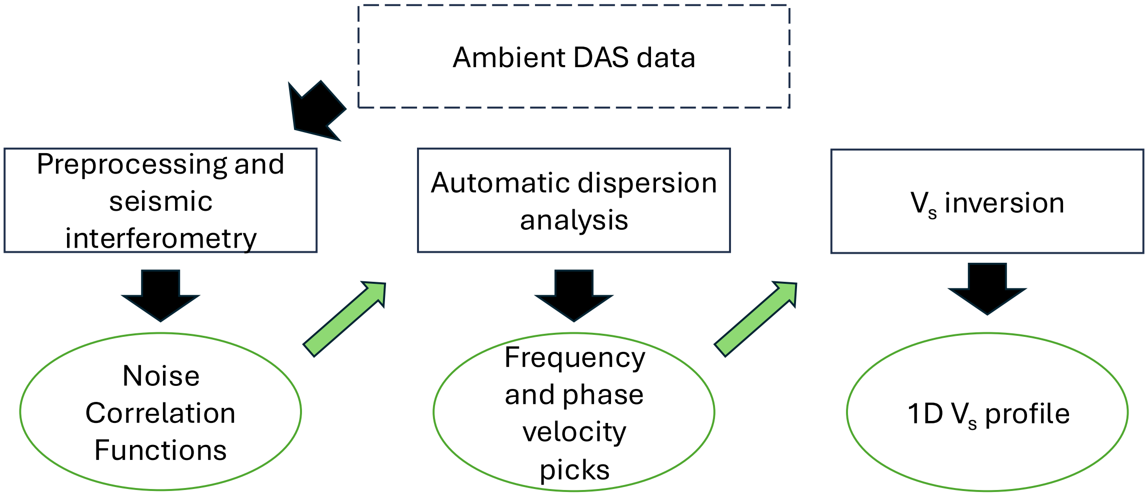


Figure 2.2 Schematic of the automated data processing workflow to generate 1D V_s profiles from ambient DAS data. The dashed frame box indicates raw data. The solid frame boxes indicate data processing steps, and the green ovals indicate data products.

2.4.1 Data selection

The first data processing step is to remove high root mean square (RMS) energy time windows [26]. This is done to remove spurious high-amplitude events that may interfere with the long-duration cross-correlation plus stack results. An energy threshold is determined for consecutive 10-minute data chunks by calculating short-term RMS values using 30-s windows. Windows with the most energetic 40% of RMS values are removed from the dataset before further processing.

2.4.2 Ambient data cross-correlation

Next, we preprocess the ambient data and perform interferometric cross-correlation analysis. We follow several standard preprocessing steps before cross-correlating ambient data [27]. We first linearly detrend the data on both the temporal and spatial axes to remove DAS common-mode noise. We then apply spectral whitening to mitigate the effect of any remaining events with highly imbalanced magnitude spectra as well as correlated noise sources. A noise correlation function (NCF) is then computed at selected virtual source locations by cross-correlating the virtual source trace with traces nearby. NCF virtual source locations are selected along each fiber at 20 m spacing with the subarray spanning 400 m to either side of each virtual source. These parameters were chosen to ensure sufficient lateral aperture given the investigation depth of 70 m and lateral resolution requirement of 20 m (roughly one-third the maximum investigation depth). Each NCF contains both positive and negative along-fiber offsets of 400 m as well as causal and acausal correlation-lag components within the range of ±5 s. We then stack the causal and acausal components and choose an along-fiber offset range based on fiber geometry. This step addresses the issue that non-colinear fiber segments can have signals that affect both apparent travel times and the Rayleigh-to-Love wave ratio deviating from the empirical Green's function calculated from seismic interferometry [21, 28, 29]. Lastly, we apply an $f - k$ filter to remove waves that appear to be propagating toward the virtual source, indicative of local noise sources that do not fit the theoretical ambient interferometry assumptions or backscattered energy that is irrelevant for future processing steps. We determined that 352 NCF gathers from fiber A and 462 NCF gathers from fiber B were sufficient for further analysis. This was done by manually checking for changes in fiber geometry that would be present in 10-minute NCFs. Changes in fiber geometry were identified by a break in signal that exist across all time lags at a single location within the NCFs.

2.4.3 Dispersion analysis and surface-wave inversion

We now detail the approached developed to automatically analyze NCF surface-wave dispersion and invert for individual 1-D V_S profiles. Dispersion images are calculated using the open-source software DASCore [30] which uses a phase-shift technique to transform the ambient NCF data from a space-time to phase velocity-frequency domain representation [9]. We use a modified version of the automated mode picking algorithm described in [31] to identify the fundamental-mode Rayleigh-wave energy in each dispersion image. Starting with a raw normalized dispersion image, we first introduce an amplitude threshold of 0.3 to isolate the primary surface-wave energy from noise. We then use density-based clustering of applications with noise (DBSCAN) [32] to distinguish between the fundamental and higher-order modes. After isolating the fundamental mode, a 200 m wavelength restriction is imposed to avoid picking unreasonably fast velocities that are not supported by the maximum $L = \pm 400$ m array length [33]. Final picks are made by automatically choosing maximum amplitude values at $\Delta f = 0.1$ Hz intervals in the dispersion image. Lastly, we invert the fundamental mode picks using MASWavesPy [34] to estimate a 1-D shear-wave velocity (V_s) profile that is spatially associated with the virtual source location . This process is repeated over the selected fiber length to build a pseudo 2-D V_s profile. Lastly, we estimate $V_{s_{30}}$ values by calculating the mean value of the V_s of the top 30 m of each profile.

2.5 Results

We apply the automated data processing workflow on FA_1, FA_2, FB_1, and FB_2 to generate 2-D V_s models and $V_{s_{30}}$ values along each profile. Fiber B is over twice the length of fiber A and features a more complex deployment geometry. Thus, to better understand the results of applying this method and the extent to which it generalizes, we first apply the workflow to fiber A and then repeat the analysis for fiber B data sets.

2.5.1 Application to FA_1

Figure 2.3 shows the inversion results from the first selected 36-hour data window FA_1. On a large scale, the image highlights a consistent velocity interface between 45 m to 60 m and a shallow velocity interface between 15 m and 25 m. However, there are numerous locations where the depth to and the estimated velocity of the lower layer seem anomalously shallow and fast. Profiles such as those located at 4,500 m, 5,000 m, and 6,400 m show considerable lateral heterogeneity between neighboring profiles. This is unexpected given the 20 m profile spacing and the 400 m subarrays used to generate the NCFs, which should result in lateral smoothing, suggesting that some profiles may contain spurious results. Gaps between individual V_s profiles indicate locations where coherent NCFs could not be analyzed due to low SNRs.

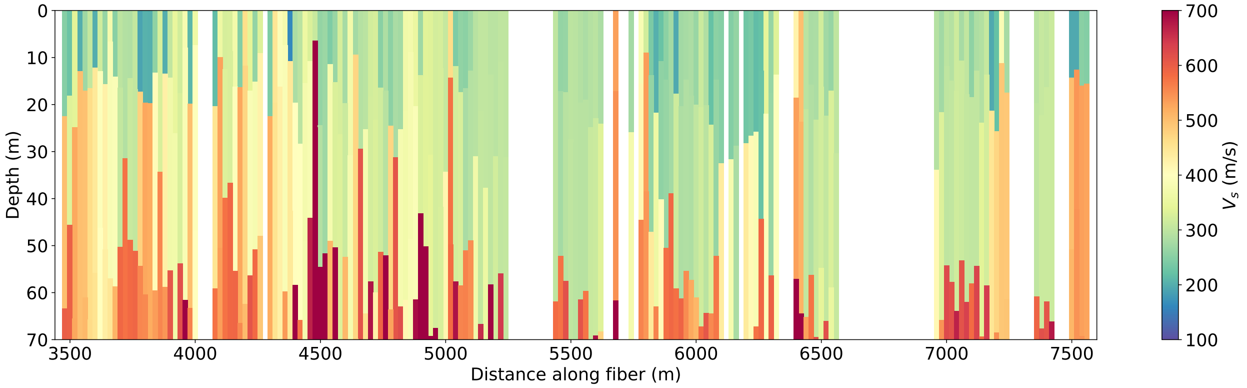


Figure 2.3 Pseudo 2-D V_s profile underlying fiber A from FA_1.

Figure 2.4 presents the $V_{s_{30}}$ estimates derived from the results presented in Figure 2.3. Figure 2.4 shows that the median $V_{s_{30}}$ estimate along fiber A is between 300 m/s and 380 m/s. The box and whisker plots show the distribution of $V_{s_{30}}$ estimates grouped into 200 m sections. We observe increased lateral heterogeneity between 3,500 m and 5,000 m compared to the rest of the fiber. The interquartile ranges (IQRs) corroborate this as we observe a high velocity range toward the left side, whereas the right side shows tightly clustered $V_{s_{30}}$ estimate values. It is important to note the presence of outliers throughout the section, which implies that fiber geometry may not be the only factor causing anomalously high apparent velocity estimates.

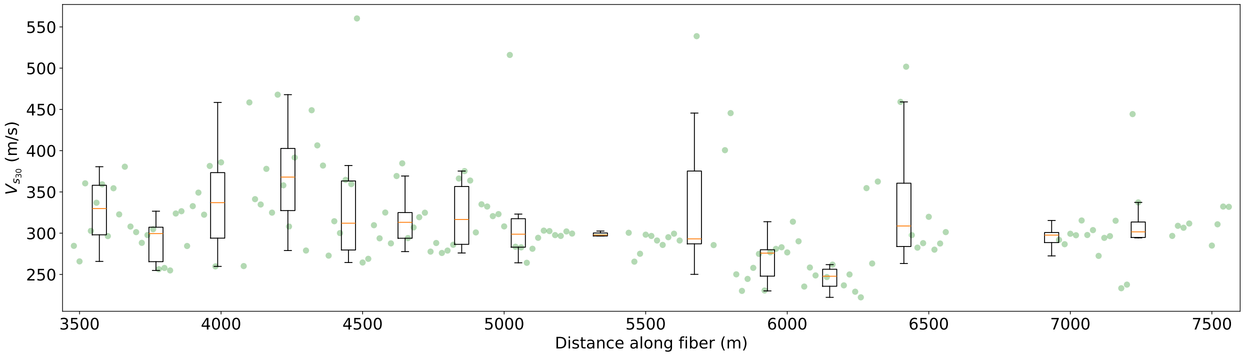


Figure 2.4 $V_{s_{30}}$ estimates derived from the FA_1 surface-wave inversions.

2.5.2 Application to FA_2

The second fiber dataset FA_2 was acquired 11 days after FA_1. No major environmental events (e.g., substantial rainfall or landslides) were reported during this period near the survey area, so significant 4-D velocity variations are not expected between the two inverted models. We applied the same automatic workflow as in the FA_1 analysis to calculate the FA_2 pseudo 2-D V_s profile and extract $V_{s_{30}}$ estimates, which are shown in Figure 2.5 and Figure 2.6, respectively. In this example, we can resolve a velocity interface between 50 m and 65 m; however, stronger lateral heterogeneity appears present in the left half of FA_2 compared to the right half. Between 3,500 m and 5,800 m along fiber distance, we observe consistently shallow and fast velocities that are absent from the FA_1 analysis. This is also noted in the corresponding $V_{s_{30}}$ estimates as the left section shows greater IQRs ranging from 270 m/s to 470 m/s. However, the right-side results are more consistent with the FA_1 observations with IQRs ranging from 250 m/s to 350 m/s.

Without a ground truth for comparison, we are unable to assess which of the FA_1 or FA_2 sections is more realistic, but a drastic change in shear wave velocity at the same depth within a two-week period would be geologically unrealistic. However, we note the similarity of the results on certain sections, which highlights favorable repeatability of the automated processing workflow for some fiber segments.

2.5.3 Application to FB_1

Figure 2.7 shows the inversion results from FB_1 constructed between 5,550 m and 14,180 m along the fiber. The fiber section doubles back on itself around 9,600 m, so we expect the inverted model to exhibit symmetry about this location. This is apparent to some extent as we can see repeated V_s profiles occurring

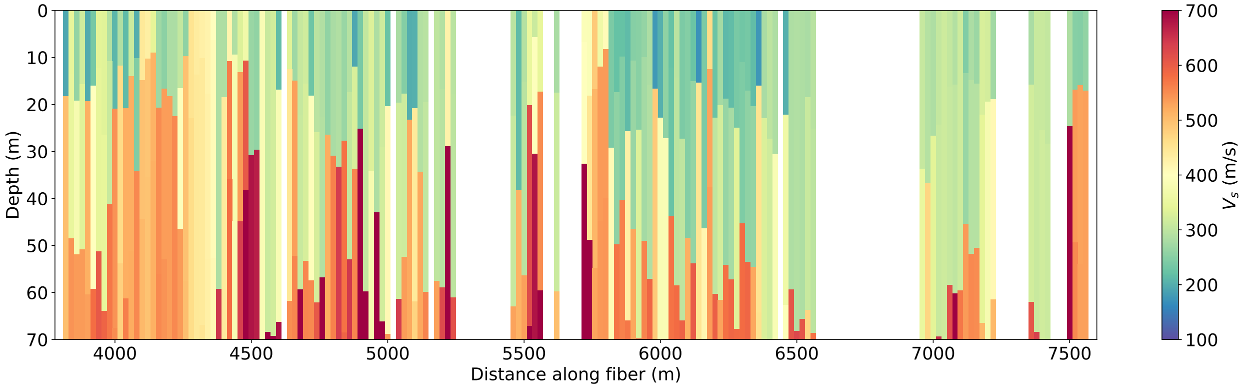


Figure 2.5 Pseudo 2-D V_s profile for fiber dataset FA$_2$.

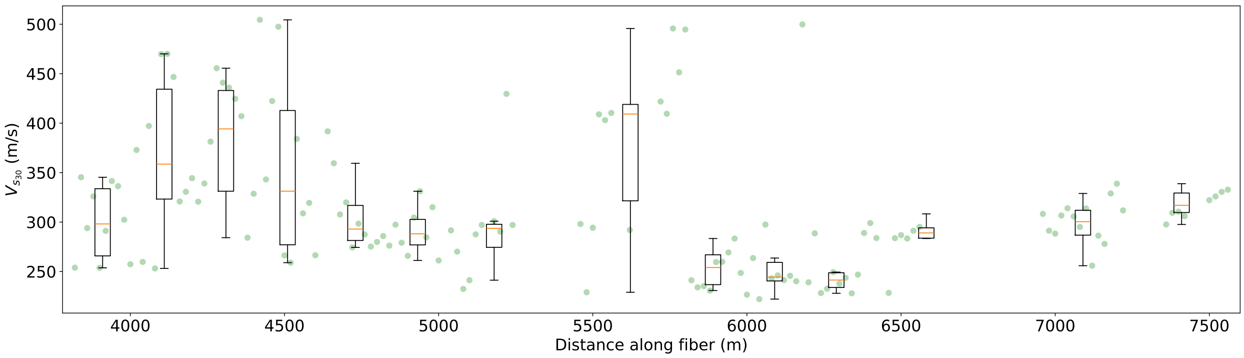


Figure 2.6 $V_{s_{30}}$ estimates derived from the FA$_1$ surface-wave inversions.

at roughly 1,500 m and 3,000 m away from the turn-back in either direction. However, profiles are not identical in the forward and reverse fiber directions.

A consistent lower depth velocity interface cannot be identified across the full 9,000 m profile (Section 2.10 Appendix B presents an analysis of these features over shorter array segments). We do observe a laterally continuous shallow velocity interface at roughly 20 m depth. We again note the presence of potentially anomalous lateral heterogeneity. Shallow, high-velocity anomalies occur between 6,200-7,000 m, 7,700-9,000 m, and 10,400-11,100 m. Deep, low-velocity anomalies also appear between 9,600-9,800 m and 10,200-10,400 m. Figure 2.8 shows the $V_{s_{30}}$ estimates from FB$_1$. Sections of the profile, such as between 6,400-8,000 m and 12,000-13,000 m, have relatively small IQRs of approximately 50 m/s. However, the general trend suggests the presence of notable lateral heterogeneity in the $V_{s_{30}}$ estimates. The median values of the $V_{s_{30}}$ estimates when grouped into 200 m subsets tend to fall between 300-350 m/s; however, the IQRs span up to 100 m/s.

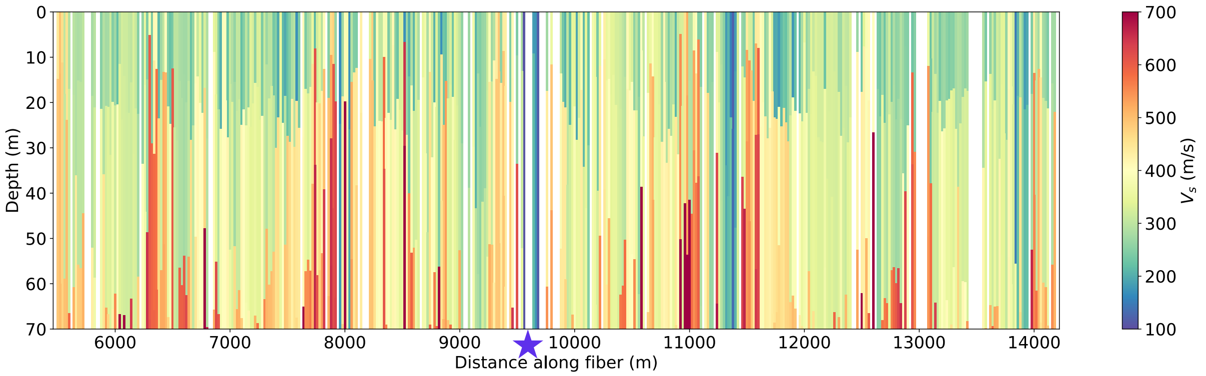


Figure 2.7 Psuedo-2D shear wave velocity profile of subsurface under fiber B from FB_1. The purple star indicates the location of the turnback on the fiber.

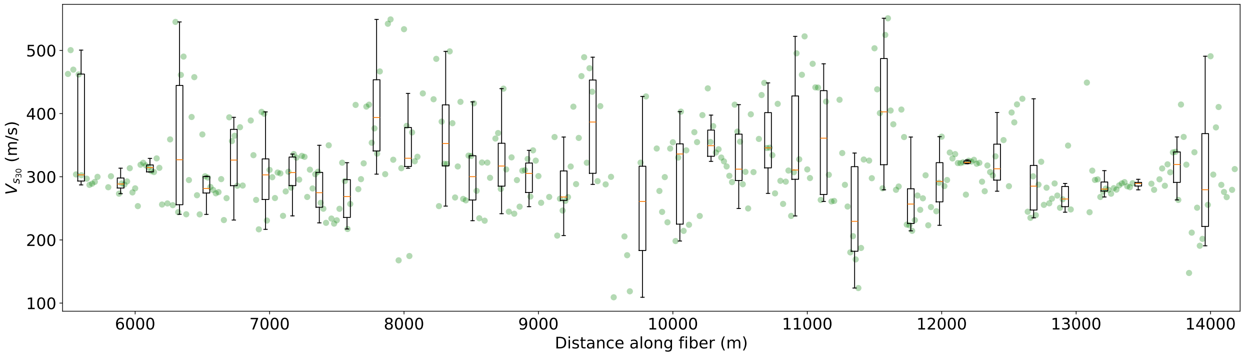


Figure 2.8 $V_{s_{30}}$ estimates derived from the FB_1 surface-wave inversions.

2.5.4 Application to FB_2

Figure 2.9 shows the inverted V_s profile for FB_2. Similar to FB_1, a shallow velocity interface appears around 20 m depth. We again observe shallow, high velocity anomalies between 6,200-7,000 m, 7,700-9,000 m, and 10,400-11,100 m and deep low-velocity anomalies between 9,600-9,800 m and 10,200-10,400 m that align with the FB_1 findings. It is important to note that the high velocity anomalies appear faster than those from the FB_1 analysis. Figure 2.10 presents the FB_2 $V_{s_{30}}$ estimates, which are consistent with those from the FB_1 analysis, with the caveat that some upper quartile velocity values are greater than 600 m/s.

2.6 Discussion

While previous studies note the consistency of inverted results with local geology and leverage primarily straight fiber geometries, dark fiber often features complex deployment geometries in areas of limited geologic knowledge [6, 16, 20, 23]. We present two examples of applying a single automated processing

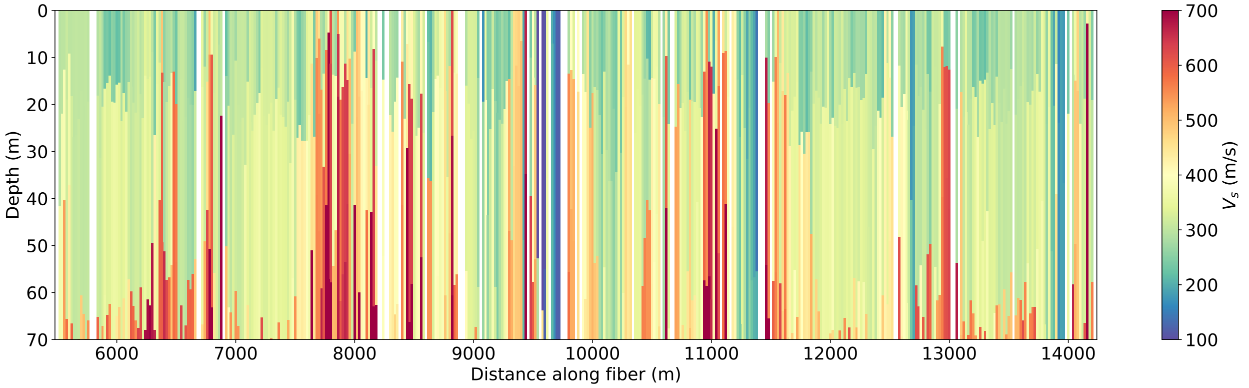


Figure 2.9 Estimated FB_2 pseudo 2-D V_s profile underlying fiber B.

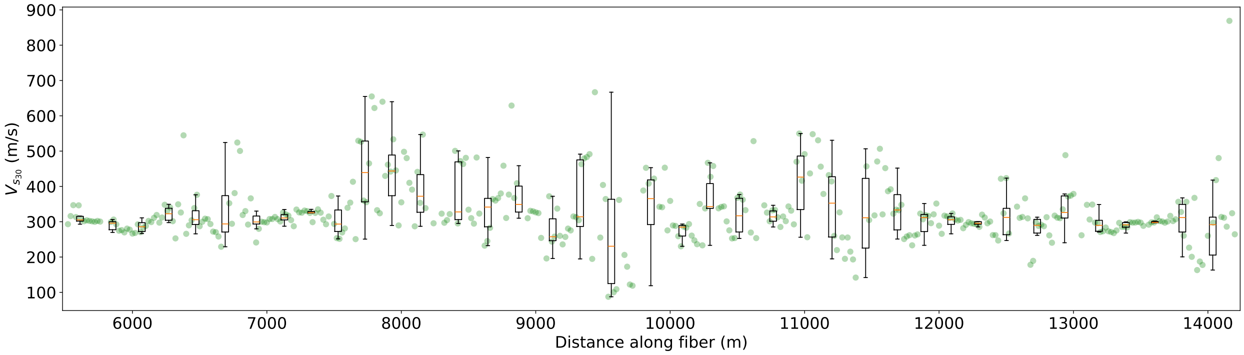
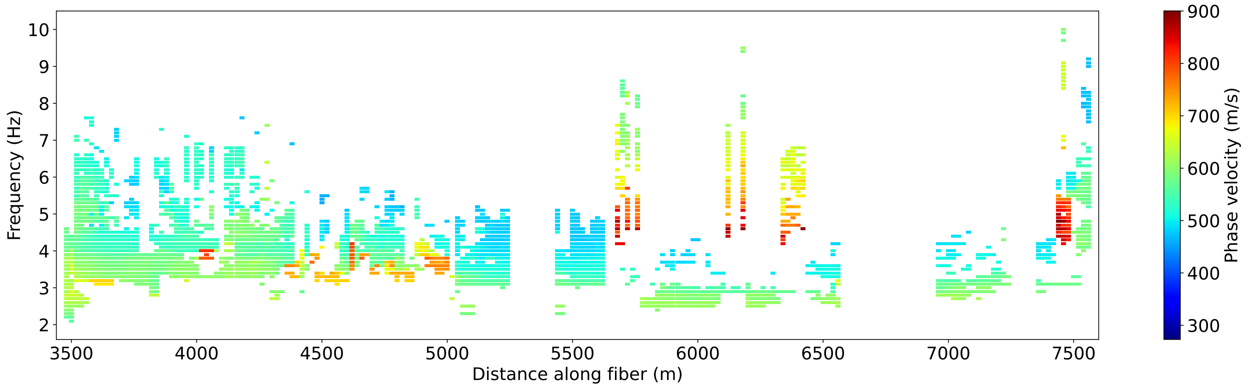


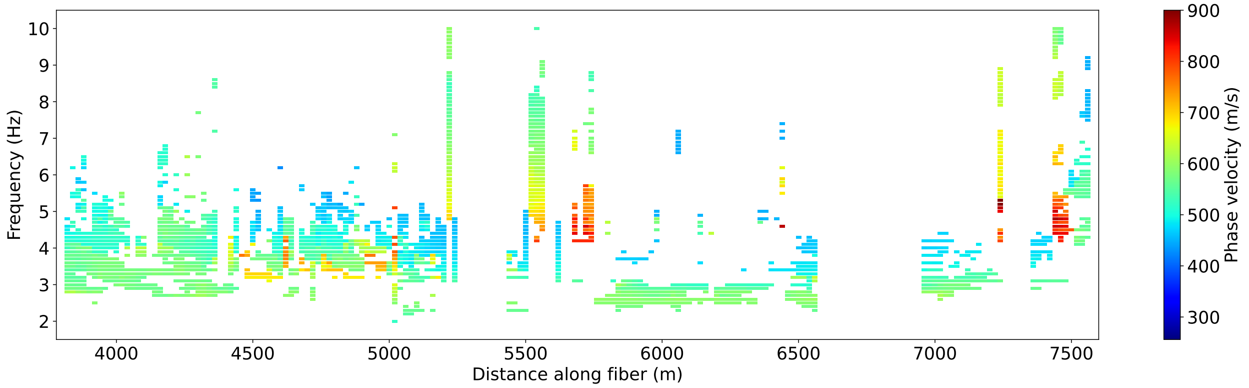
Figure 2.10 $V_{s_{30}}$ estimates derived from the FB_2 surface-wave inversions.

scheme to fiber with non-ideal geometries and limited background geology. As the results show, the combination of these challenging factors and processing workflow deficiencies could result in erroneous V_s estimates in some locations and when using different data subsets on the same fiber.

Because surface-wave inversion primarily depends on the veracity of the input dispersion curves, we can analyze the variation in automated picks to determine where outliers may be present. From Figure 2.11(a) we note lateral continuity in dispersion picks along 100 m sections. Outliers tend to be grouped together and show relatively high phase velocities for the given frequency ranges. This is most exemplified by the picks between 5,500 m and 5,750 m in FA_1. Within the same data subset, we observe laterally consistent dispersion picks between 5,500-5,640 m. However, less than 100 m along the fiber, we see picks increase in both frequency range (from 3-5 Hz to 4-9 Hz), as well as in phase velocity (from 450-550 m/s to 550-900 m/s). This is also noted in Figure 2.11(b) suggesting that certain locations may exhibit strong higher-order surface-wave modes that are being picked instead of the fundamental mode. Differences in the left half of Figure 2.3 and Figure 2.5 can be attributed to variations within lower-frequency picks that are biased toward higher phase-velocity estimates.



(a) FA_1



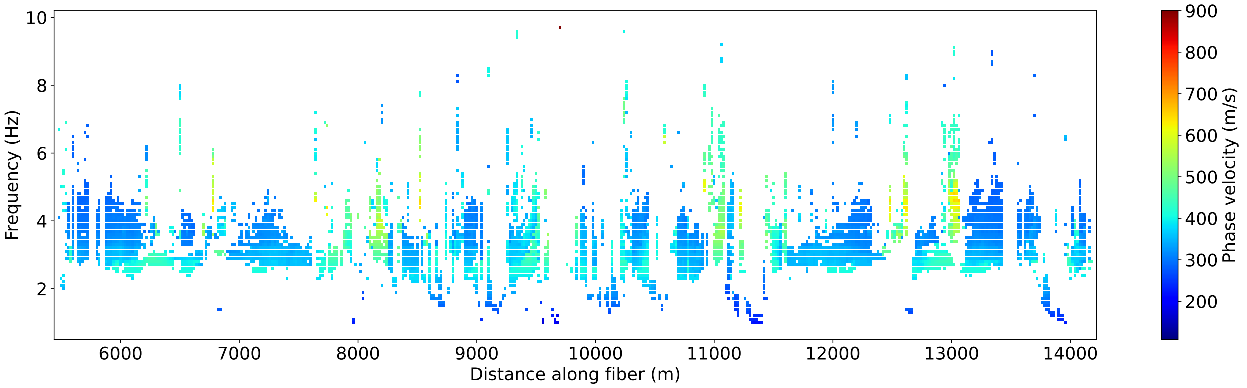
(b) FA_2

Figure 2.11 Fiber A dispersion curve picks.

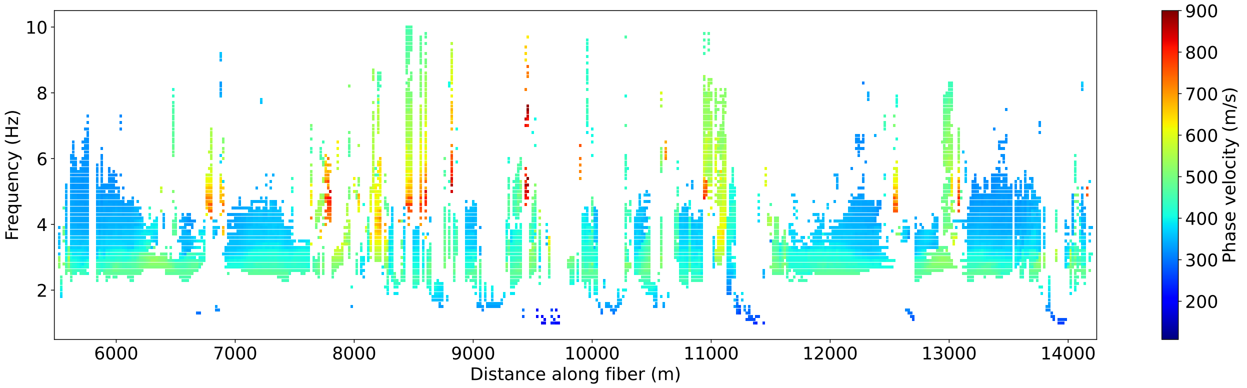
The FB_1 dispersion curve picks similarly show sections of higher velocity pick for given frequencies, which result in anomalously fast V_s profiles. We also note that lower velocity picks are present at a wide range of frequencies. This may be possible with unconsolidated sediments. However, we also observe inconsistent sampling of picks at high frequencies as well as picks with inconsistent sampling at broad frequency ranges. This could result in error-prone forward modeling of dispersion curves that would result in unrealistic V_s profile estimates.

The FB_2 picks, presented in Figure 2.12(b), show a contrast from Figure 2.12(a) between 2.0-3.5 Hz, with a consistent phase-velocity increase of around 100 m/s. We also note more anomalously high picks and higher picked phase velocity values compared to the FB_1 results.

An automated multimodal dispersion curve picking algorithm may be necessary to ensure spatial and temporal consistency of phase velocity picks. Additionally, differences between V_s profiles generated from different data subsections on the same fiber indicate the effect of noise distribution variations and their



(a) FB_1



(b) FB_2

Figure 2.12 Fiber B dispersion curve picks.

impact on estimated surface-wave travel times between fiber segments from seismic interferometry. Furthermore, generalizing automated processing strategies for dark fiber arrays faces a challenge from the variety of fiber geometries and coupling conditions. All of these factors can influence the validity and repeatability for near-surface characterization using the methods detailed in this paper.

2.6.1 Application in earthquake hazards assessment

Earthquake engineers use $V_{s_{30}}$ estimates as an input to classify site amplifications [35, 36]. Table 2.1 shows the site classes associated with specific $V_{s_{30}}$ values. Figure 2.13 shows the NEHRP classification applied to the $V_{s_{30}}$ estimates from FA_1 and FA_2. From FA_1, 129 points fall in Class D while 26 fall in Class C, where most estimates are anomalously high. FA_2 shows similar results with 114 and 26 points falling in Classes D and C, respectively. FA_2 appears to have more anomalously high $V_{s_{30}}$ estimates; however, FA_1 has 71 points higher than FA_2 at the same locations compared to 58 for FA_2. 83% of FA_1 estimates and 81% of FA_2 estimates fall in Class D. This shows the repeatability of these estimates for site classification in an earthquake hazards context for fiber A.

Table 2.1 NEHRP site classes based on $V_{s_{30}}$, adapted from [36].

Site class	Soil Profile name	$V_{s_{30}}$ (m/s)
A	Hard rock	> 1500 m/s
B	Rock	$760 - 1500$ m/s
C	Very dense soil and soft rock	$360 - 760$ m/s
D	Stiff soil	$180 - 360$ m/s
E	Soft soil	< 180 m/s

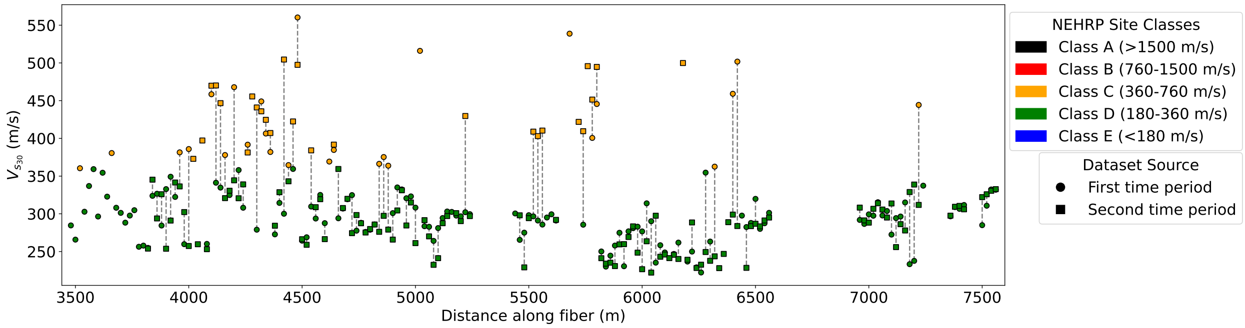


Figure 2.13 $V_{s_{30}}$ estimates from FA$_1$ (circles) and FA$_2$ (squares) plotted together. Dashed lines show locations where both FA$_1$ and FA$_2$ have a $V_{s_{30}}$ value. Green indicates site class D (stiff soil) and yellow indicates site class C (very dense soil and soft rock).

Figure 2.14 shows the NEHRP classification applied to $V_{s_{30}}$ estimates from FB$_1$ and FB$_2$. For FA$_1$, there are 8 locations in Class E, 282 locations in Class D, and 82 locations in Class C. For FA$_2$, there are 10 locations in Class E, 295 locations in Class D, 87 locations in Class C, and 1 location in Class B. Both FA$_1$ and FA$_2$ show low-velocity estimates that fall in Category E; however, these account for just 2% of all $V_{s_{30}}$ estimates across the entire array. Similarly, only 1 point out of 765 falls in Class B. Additionally, we note that the majority of these extreme Classes B and E locations are in the second half of the fiber and are not observed in the first half of the fiber at the same locations. Given higher sensor noise levels at longer DAS fiber observation distances, this suggests that many of these likely are unreliable. These show cases of anomalous $V_{s_{30}}$ estimates that may be removed in a practical application of this method for automated site classification. However, fiber B does show a more even split between locations in Classes D and C which may require a more individualized classification approach.

Despite the challenges with the current automated workflow, augmentations can be applied that can improve the quality of the generated profiles. Multimodal strategies for dispersion-curve picking can be implemented to improve 1-D inversions [37, 38]. Raw data processing can be enhanced by implementing

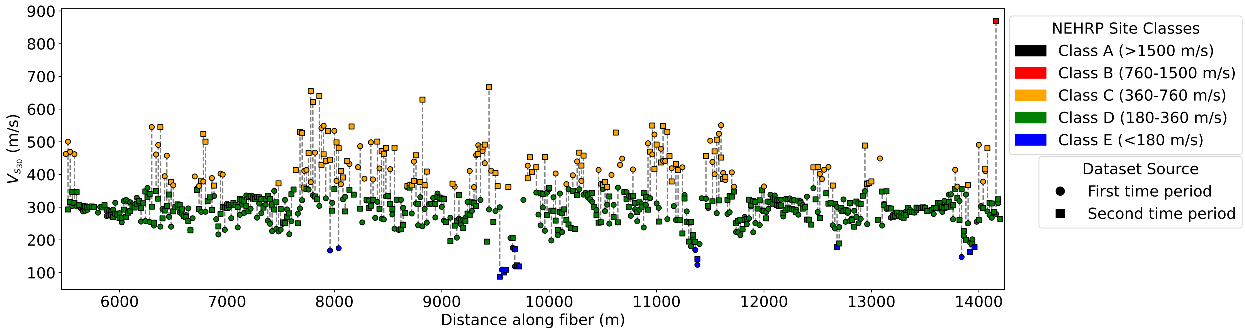


Figure 2.14 $V_{s_{30}}$ estimates from FB$_1$ (circles) and FB$_2$ (squares) plotted together. Dashed lines show locations where both FB$_1$ and FB$_2$ have a $V_{s_{30}}$ value. Blue indicates site class E, green indicates site class D, yellow indicates site class C, and red indicates site class B.

both time and frequency debursting strategies to improve the quality of NCFs [26]. Furthermore, applications in earthquake hazards monitoring using $V_{s_{30}}$ may only require tolerances within specific site classification bounds.

2.7 Conclusions

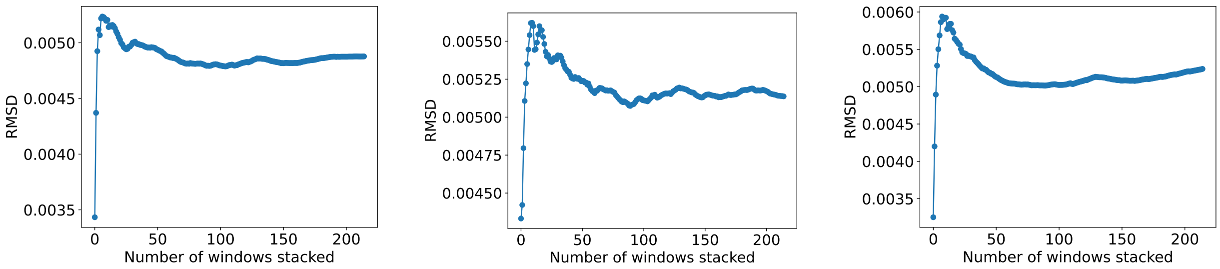
This study demonstrates that automated data processing schemes can be developed and applied to ambient DAS datasets acquired on dark fiber arrays for near-surface characterization purposes. By automating this analysis, time-lapse near-surface investigations can be conducted to monitor urban areas and potentially detect subsurface changes that may pose a danger to urban infrastructure and the people who rely on it. We applied our automated workflow on two separate dark fiber arrays and found that while near-surface characterization can be achieved, augmentation of the data processing strategy may be necessary to improve the final result. Our analysis shows that automated surface-wave mode picking imposes a critical issue for the reliability of V_s inversions. The lack of mode labeling can lead to the inconsistent picking of the fundamental and higher-order modes resulting in inverted profiles that exhibit high V_s values at shallow depths. Additionally, time-lapse repeatability is dependent on optimizing the ambient seismic wavefield for retrieving accurate Green's function estimates. We showed that on a large scale, shallow interfaces could be resolved by the presented workflow. While unrealistic V_s profiles were present in the pseudo 2-D V_s models, we determined that these tended to occur in close proximity to each other. A survey crew could use this information for follow-on site-specific investigations.

2.8 Acknowledgments

We thank the Colorado School of Mines Research Computing team for providing HPC resources. We also thank the sponsors of the Colorado School of Mines Center for Wave Phenomena industrial consortium for funding this research. We also thank Ranjan Dash and Guangchi Xing for their contributions in reviewing this paper.

2.9 Appendix A: Analysis of stacking duration on automated workflow

Assessing the quality of the NCFs at multiple locations can be useful in identifying where errors in the automated processing workflow may begin. We can use the root-mean-square deviation (RMSD) of a baseline NCF with each successive stacked NCF to visualize convergence as a function of number of stacks. Three examples from FA_1 are shown in Figure 2.15. Figure 2.15(a) and Figure 2.15(c) show NCFs that converge by around 75 stacks. This represents around 12.5 hours of stacked ambient noise data. Both of these examples also show a slight bump in RMSD around 125 sub-stacks. Further stacking cause the RMSD to gradually increase, indicating that the optimal amount of stacking may occur prior to 125 sub-stacks at these locations. Figure 2.15(c) shows a location that both appears to converge after more than 125 sub-stacks, and has more noise. In this case, convergence seems to be met around 130 sub-stacks or 21.67 hours. This shows that convergence rates can vary considerably along the same fiber.



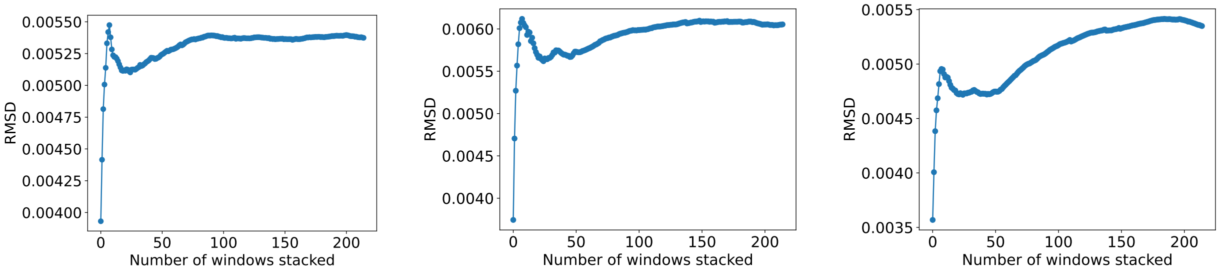
(a) Channel 2,000

(b) Channel 3,000

(c) Channel 3,800

Figure 2.15 The convergence of three NCFs along fiber A from FA_1 using the root-mean-square deviation of a baseline NCF with each stack. Each sub-stack accounts for 10 minutes of ambient noise data. Plots show the RMSD of NCFs centered at (a) channel 2,000, (b) channel 3,000, (c) channel 3,800.

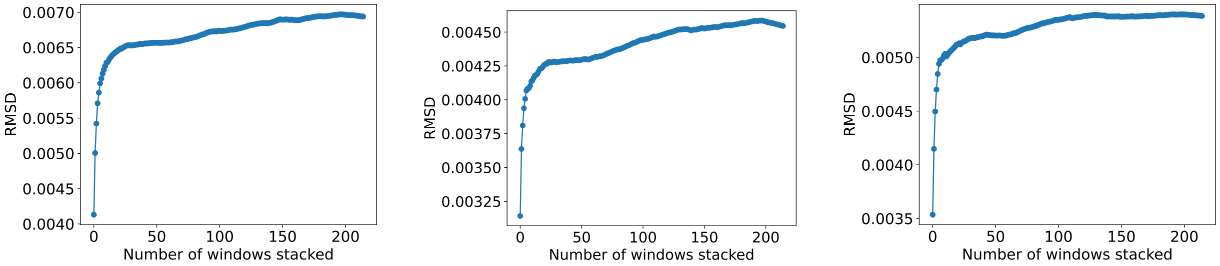
We also compare the RMSD between FA_1 and FA_2 (Figure 2.16). Figure 2.15(a) shows convergence after 100 sub-stacks and does not appear to change substantially with further stacking. Compared to the same location in FA_1 we see a difference of around 4 hours in stacking. Channel 3000 shows a marked change between FA_1 and FA_2 as the second time period appears less noisy and converges around 125 sub-stacks (Figure 2.15(b)). Figure 2.15(c) shows another deviation from FA_1 as the RMSD increases after 60 sub-stacks until around 200 sub-stacks. This indicates that convergence was met with a lower number of stacks, with additional stacks increasing noise in the NCF. By comparing FA_1 and FA_2, we also note that variances in convergence can occur from data collected at different times along the same fiber, which could affect the repeatability of automated processing schemes.



(a) Channel 2,000 (b) Channel 3,000 (c) Channel 3,800

Figure 2.16 The convergence of three NCFs along fiber A from FA_2. Plots show the RMSD of NCFs centered at channels (a) 2000, (b) 3000, and (c) 3800.

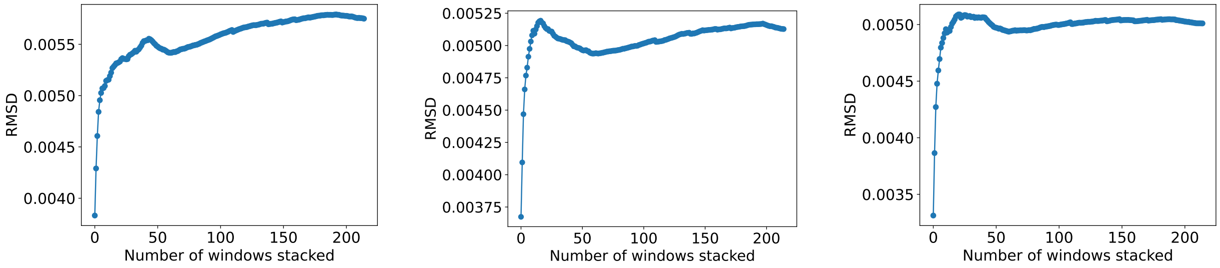
Figure 2.17 and Figure 2.18 show the RMSD plots for FB_1 and FB_2 respectively. FB_1 appears to reach convergence at all three locations after 125 stacks. Channel 3700 has a decreasing RMSD around 200 stacks which is also seen in the same channel on FB_2. The RMSD plots of FB_1 and FB_2 have similar trends across like channels, however, they do not retain this similarity across channels along the same fiber.



(a) Channel 3,000 (b) Channel 3,700 (c) Channel 4,500

Figure 2.17 The convergence of three NCFs along fiber B from FB_1. Plots show the RMSD of NCFs centered at (a) channel 3,000, (b) channel 3,700, (c) channel 4,500.

The convergence of NCFs also impacts dispersion analysis, which we can assess by viewing the dispersion picks of smaller stacking windows. Figure 2.19 shows the changes in dispersion curve picks from FA_1 over 12, 24, 36 hours of stacking. With the increase in stacking, we can see a general increase in the number of anomalously fast and high frequency picks. In particular, these types of picks occur near 7,500 m in Figure 2.19(b) and Figure 2.19(c). Furthermore, Figure 2.19(c) shows high velocity picks between 6,000 m and 6,500 m that do not appear in Figure 2.19(a) and Figure 2.19(b). That same section also has fewer coherent picks as the length of stacking increases. This is contrasted with the section of fiber from 3,500 m to 5,500 m as those picks stay largely consistent, or improve in spatial coverage and bandwidth.



(a) Channel 3,000 (b) Channel 3,700 (c) Channel 4,500

Figure 2.18 The convergence of three NCFs along fiber B from FB_2. Plots show the RMSD of NCFs centered at (a) channel 3,000, (b) channel 3,700, (c) channel 4,500.

We replicate this analysis on FB_1 to further compare the application of the automated workflow on another dataset (Figure 2.20). In this case, we see that the 24 hour dispersion picks are on average 100 m/s faster across most frequency ranges compared to both the 12 and 36 hour picks. This could indicate a scenario in which apparently fast noise sources or higher modes dominate the ambient wavefield for a period in time.

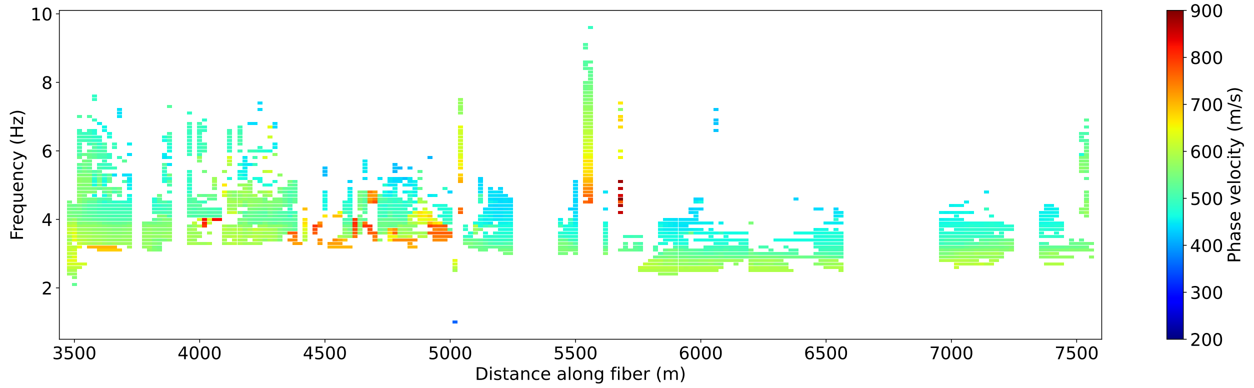
Lastly, we view the effects of stacking time on the psuedo-2D V_s profiles from FA_1 (Figure 2.21), and FB_1 (Figure 2.22). All of the computed profiles from FA_1 appear to have a consistent shallow velocity interface between 10 m and 20 m in depth. We also see a basement layer between 50 m and 60 m in depth that gets clearer with larger amounts of stacking. With increasing stacking, the high velocity anomaly between 4,000 m and 5,000 m along the fiber appears to diminish in vertical extent. However, Figure 2.21(c) does contain a few anomalously fast profiles that have very fast layers as shallow as 15 m depth.

The profiles from FB_1 show similarity. All three have a consistent shallow velocity interface at roughly 20 m in depth. Additionally, the locations and depths of the high velocity anomalies largely remain the same. This indicates an issue in mode picking, rather than convergence of NCFs. Lateral heterogeneity that may not be physically realistic does appear between 11,000 m and 11,500 m in the 24 hour and 36 hour plots.

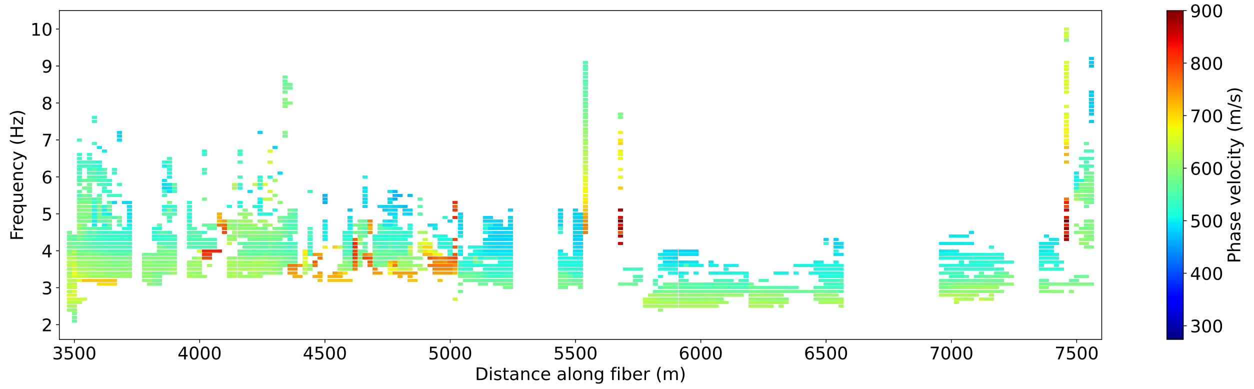
2.10 Appendix B: Spatial variation of inverted results

2.10.1 Fiber A, first 36 hours

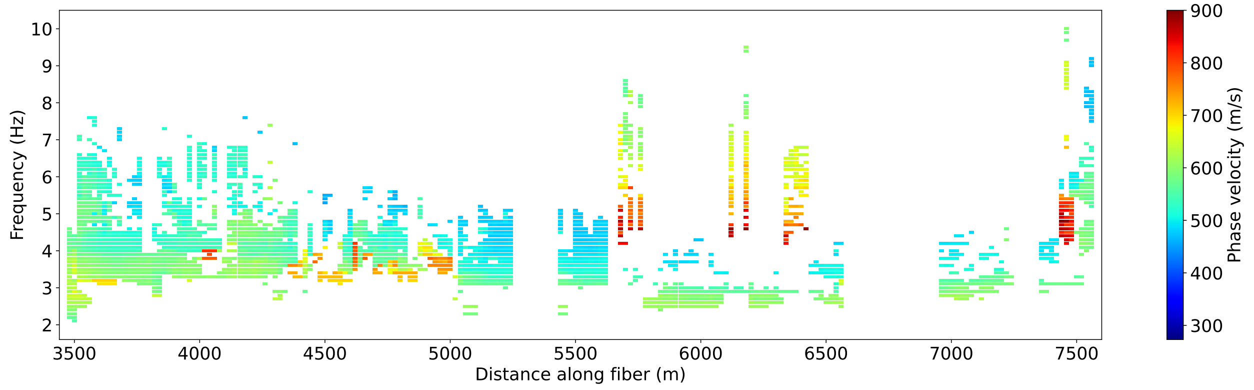
By windowing into specific sections of the inverted V_s profiles and the dispersion curve picks, we can more closely analyze locations that seem anomalous. Figure 2.23 shows a section of the V_s profile and its dispersion curve picks that appears to contain considerable lateral heterogeneities. We see in this section,



(a) 12 hours

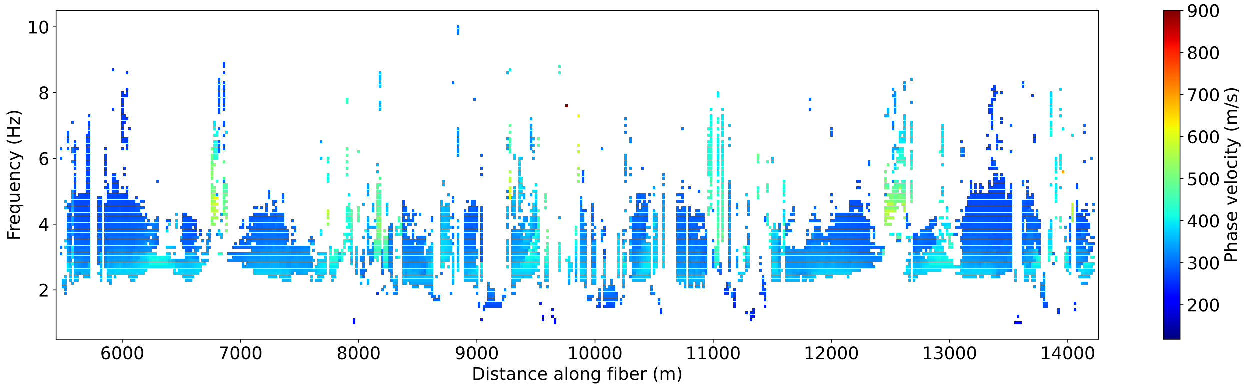


(b) 24 hours

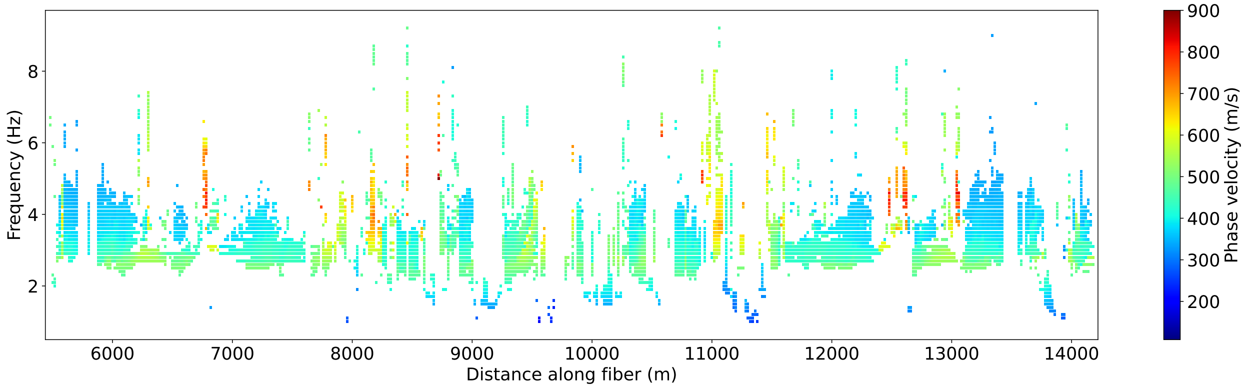


(c) 36 hours

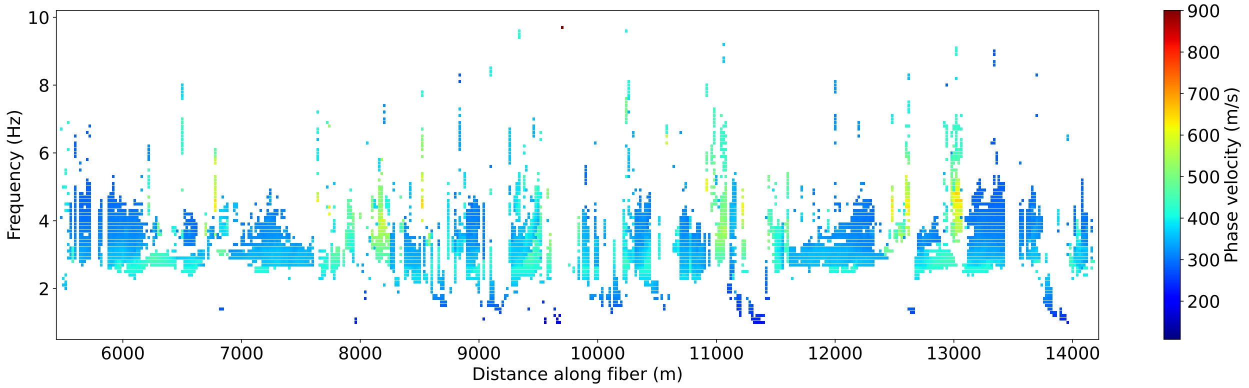
Figure 2.19 Dispersion curve picks along fiber A from (a) 12 hours, (b) 24 hours, and (c) 36 hours of stacked NCFs from FA_1.



(a) 12 hours

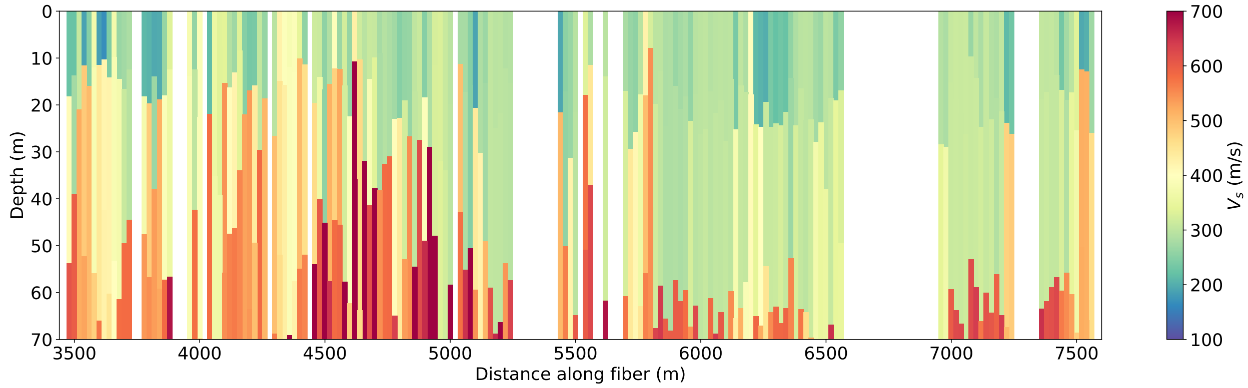


(b) 24 hours

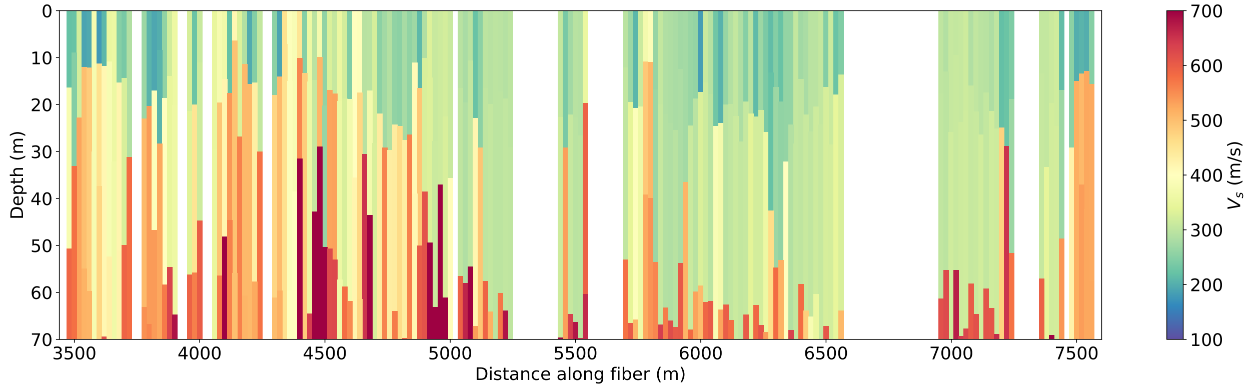


(c) 36 hours

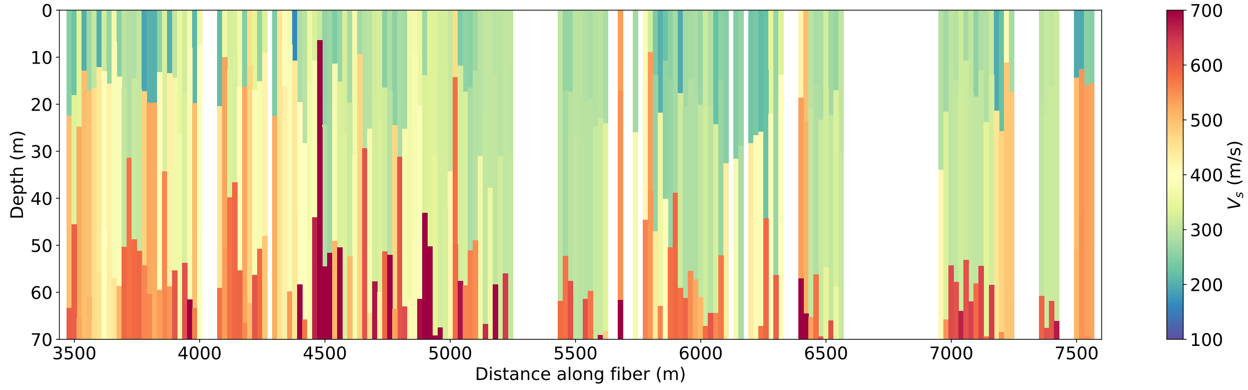
Figure 2.20 Dispersion curve picks along fiber B from (a) 12 hours, (b) 24 hours, and (c) 36 hours of stacked NCFs from FB_1.



(a) 12 hours

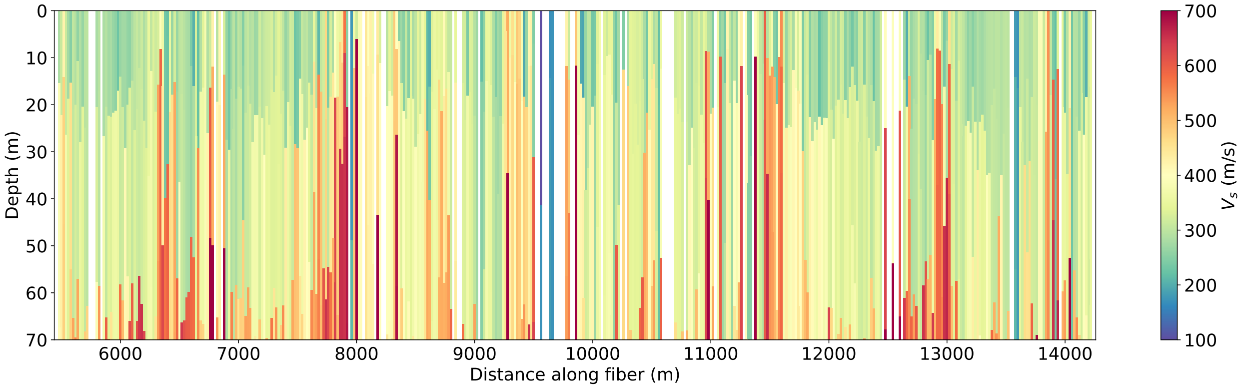


(b) 24 hours

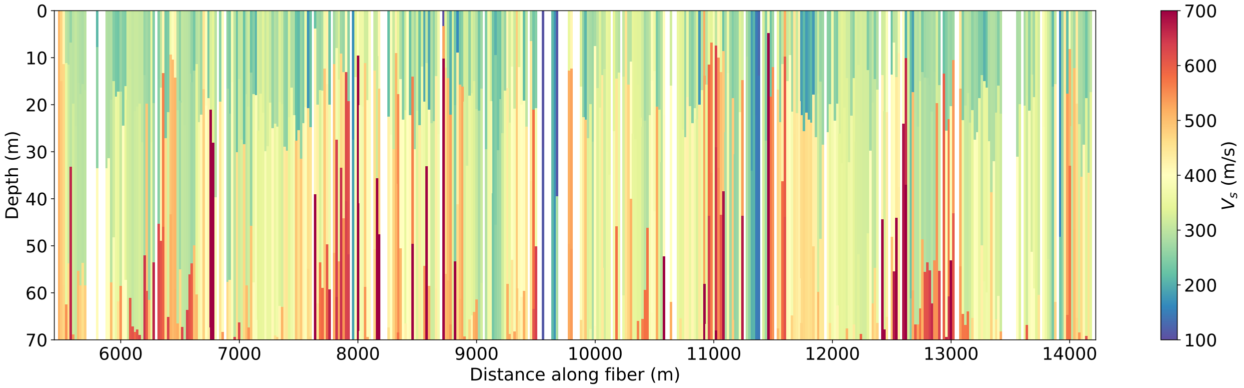


(c) 36 hours

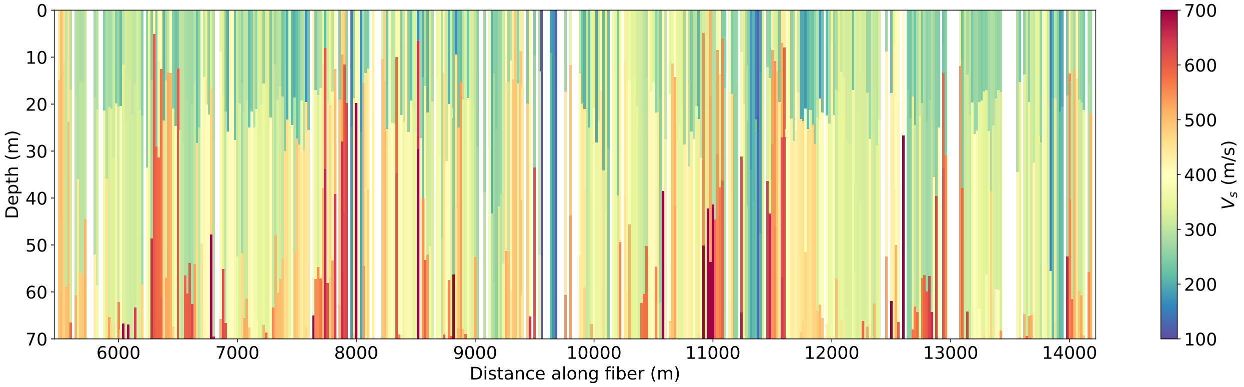
Figure 2.21 Psuedo 2D shear wave velocity profiles along fiber A using (a) 12 hours, (b) 24 hours, and (c) 36 hours of stacked NCFs from FA_1.



(a) 12 hours



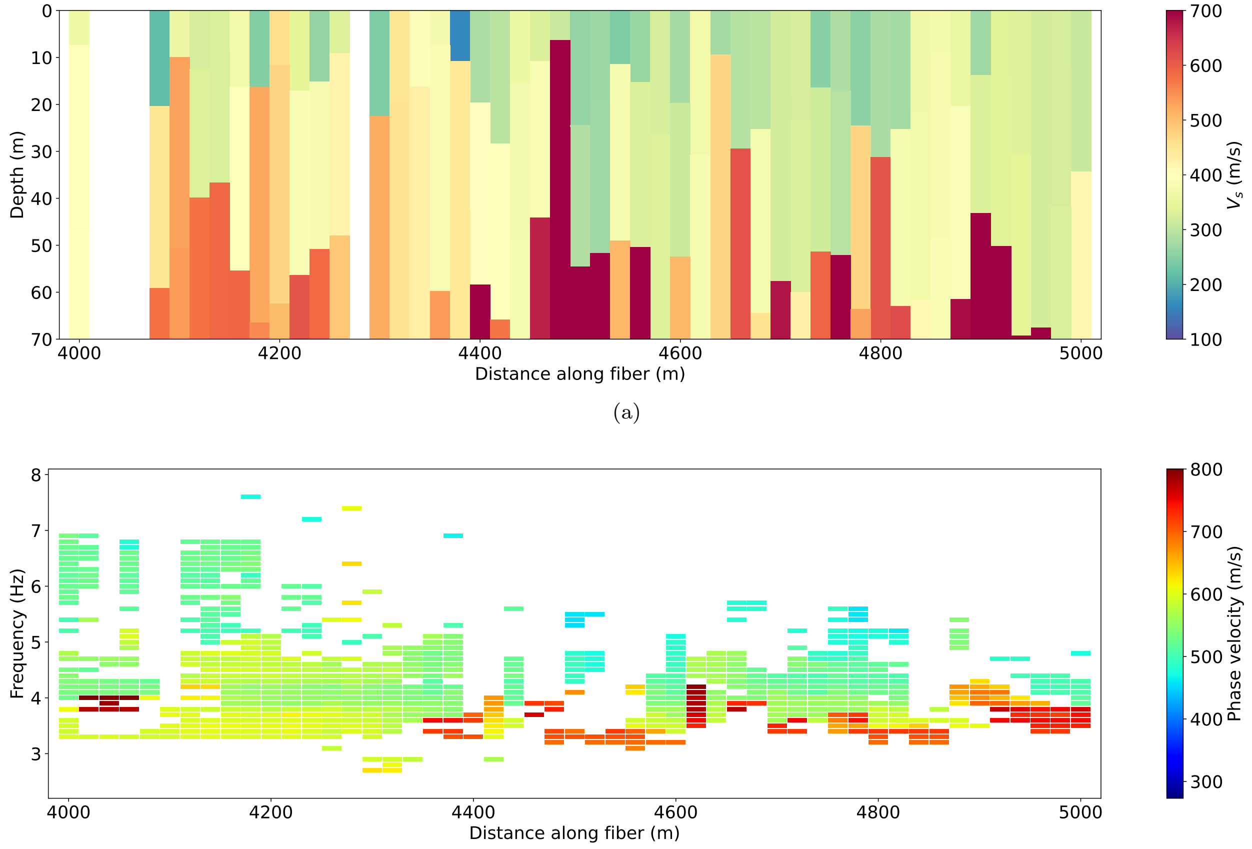
(b) 24 hours



(c) 36 hours

Figure 2.22 Psuedo 2D shear wave velocity profiles along fiber B using (a) 12 hours, (b) 24 hours, and (c) 36 hours of stacked NCFs from FB_1.

there is a high contrast in the low-frequency picks between 3 Hz to 4 Hz. In cases such as at 4,020 m, 4,040 m, and 4,060 m this results in a failed inversion where no profile can be inverted for. Between 4,100 m and 4,300 m we see picks that appear to initially increase in phase velocity as frequency increases before decreasing. Additionally, some of these picks appear to be contaminated with higher-order information. In these cases, phase velocity initially decay as frequencies increase which is the expected signature of a single mode. However, at some higher frequencies we see a sudden increase in picked phase velocities, indicating the picking of points from higher modes. Between 4,400 m and 5,000 m, a strong basement layer contrast appears. Without prior geologic information we cannot assess the validity of this section. However, we can attribute the high velocities to high phase velocity contrasts in the low frequencies. It is interesting to note that at 4,480 m, there is a clearly anomalous high velocity contrast in the inverted result, however the corresponding dispersion curve consist of five low frequency, high velocity picks.

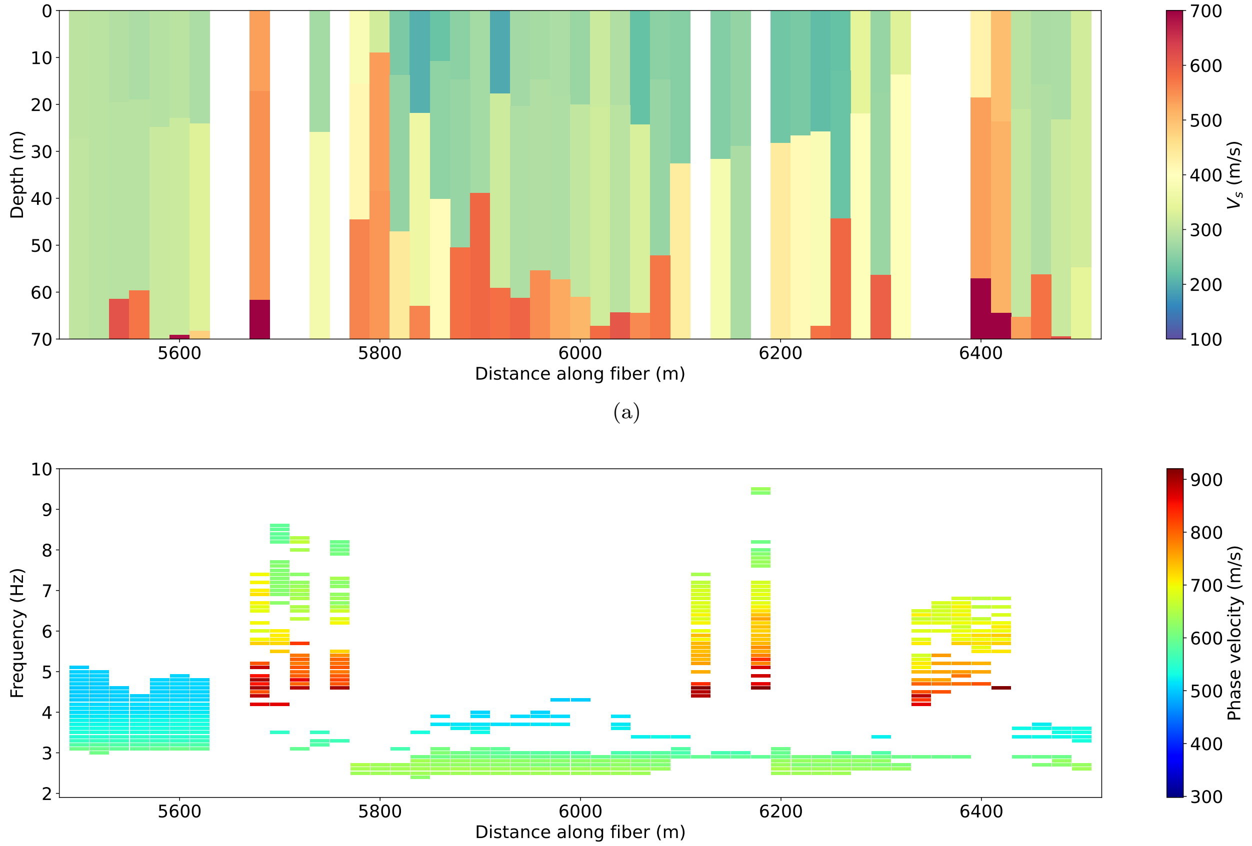


(a)

(b)

Figure 2.23 (a) Psuedo 2D shear wave velocity profile along fiber A from FA_1 and (b) its corresponding dispersion curve picks. Windowed between 4,000 m and 5,000 m.

The section between 5,500 m and 6,500 m (Figure 2.24) shows clusters of lateral heterogeneity around locations with consistent results. Between 5,600 m and 5,800 m there is one clearly anomalous V_s profile that is surrounded by locations where the inversion process failed, resulting in white spaces on the psuedo-2D V_s plot. The corresponding dispersion curve picks show high phase velocity gradients that do not appear similar to the picks at 5,600 m and 5,800 m. This is similarly seen between 6,340 m and 6,420 m. Furthermore, the dispersion curve picks at 6,120 m and 6,180 m also appear anomalous, however these locations were not inverted. It appears that locations that can not be inverted often have low velocity picks at low frequencies followed by what appears to be a higher order mode at higher frequencies.



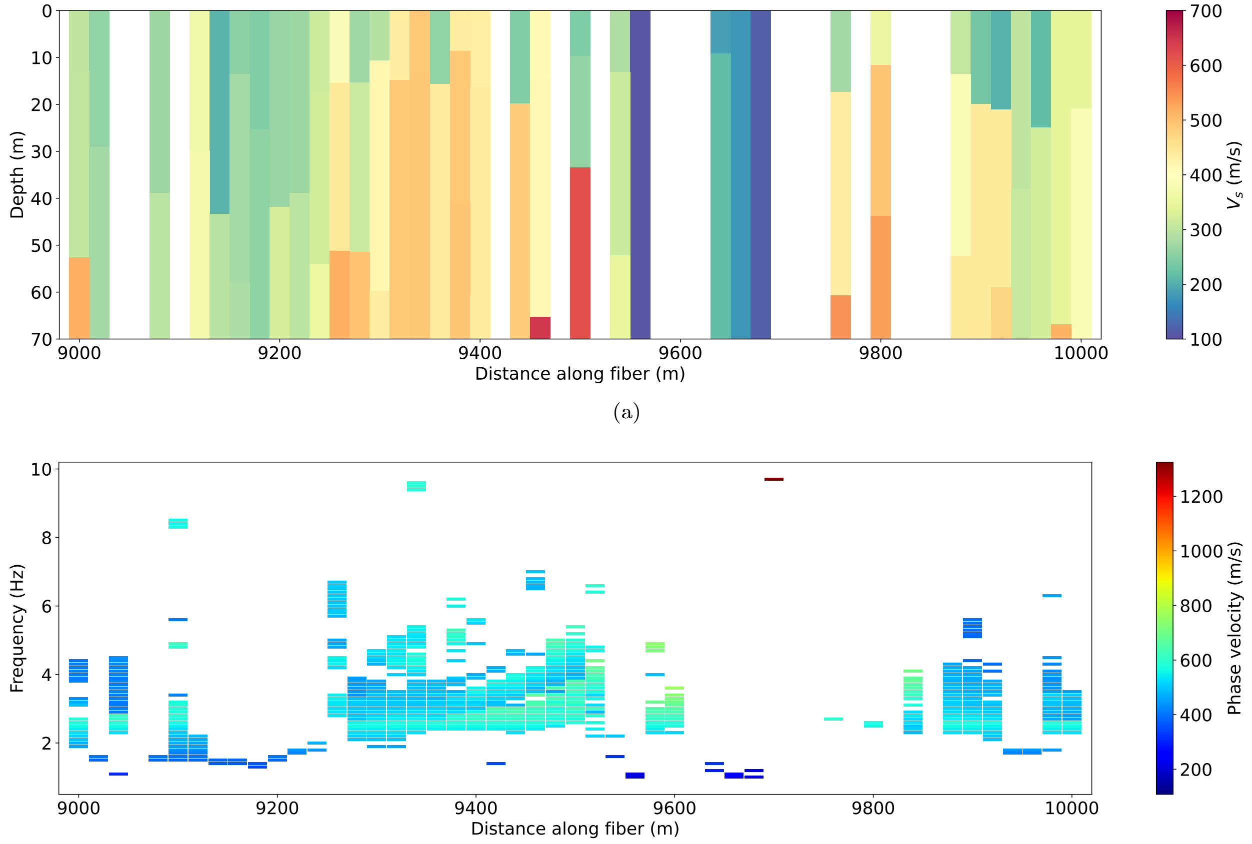
(a)

(b)

Figure 2.24 (a) Psuedo 2D shear wave velocity profile along fiber A from FA_1 and (b) its corresponding dispersion curve picks. Windowed between 5,500 m and 6,500 m.

2.10.2 Fiber B, first 36 hours

Figure 2.25 shows a windowed section of FB_1 between 9,000 m and 10,000 m. In this case, relatively high velocity sections appear adjacent to a relatively low velocity section. For the high velocity sections (between 9,260 m and 9,500 m) we can see that the dispersion curve picks show a pattern of increase, suddenly decreasing, and then increasing again in phase velocity as frequency increases. The low velocity profiles (at 9,560 m, 9,640 m, 9,660 m, and 9,680 m) appear to be inverted from only a couple of low velocity and low frequency picks. We see similar low velocity picks between 9,140 m and 9,240 m, however the phase velocities are slightly higher so the inverted profiles don't stand out as explicitly anomalous. This indicates two problems faced by the automated mode picking scheme. First, mode labeling is a challenging problem that can result in anomalously fast inverted profiles. Second, inverted results from a small number of picks can result in poor inversions.

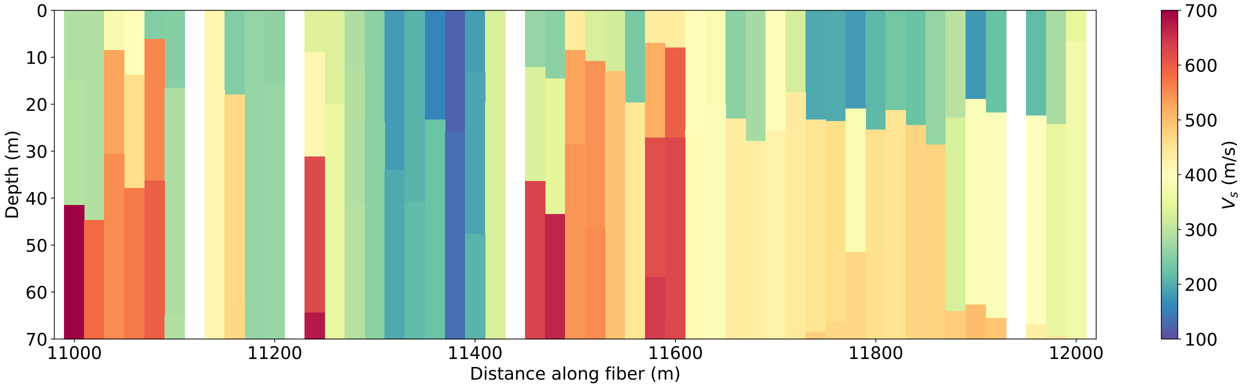


(a)

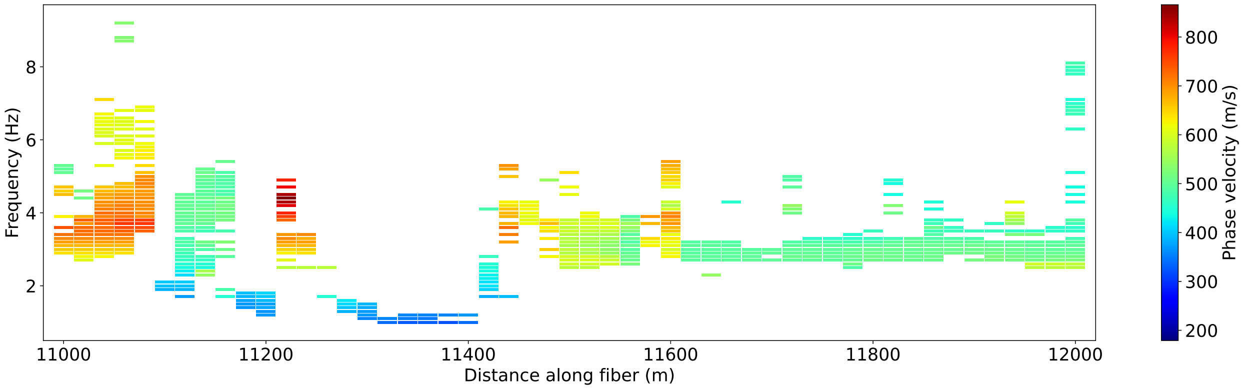
(b)

Figure 2.25 (a) Psuedo 2D shear wave velocity profile along fiber B from FB_1 and (b) its corresponding dispersion curve picks. Windowed between 9,000 m and 10,000 m.

Figure 2.26 presents another example from FB_1, windowed between 11,000 m and 12,000 m. In this example, there are again substantial lateral heterogeneities that appear close together spatially. High velocity anomalies, such as between 11,000 m and 11,080 m and between 11,440 m and 11,600 m, appear to have a mixture of first picking increasing phase velocity and then following the expected decrease, or picking between multiple modes. The low velocity profiles between 11,180 m and 11,400 m again show cases where only a few number of low frequency picks are made, resulting in unrealistic inversions. These examples are contrasted with the section between 11,620 m and 12,000 m. This section exhibits some homogeneity as well as a clear example of an interface at 20 m in depth. Furthermore, the corresponding dispersion curves also appear mostly homogeneous with slight variations at low frequencies near 12,000 m.



(a)

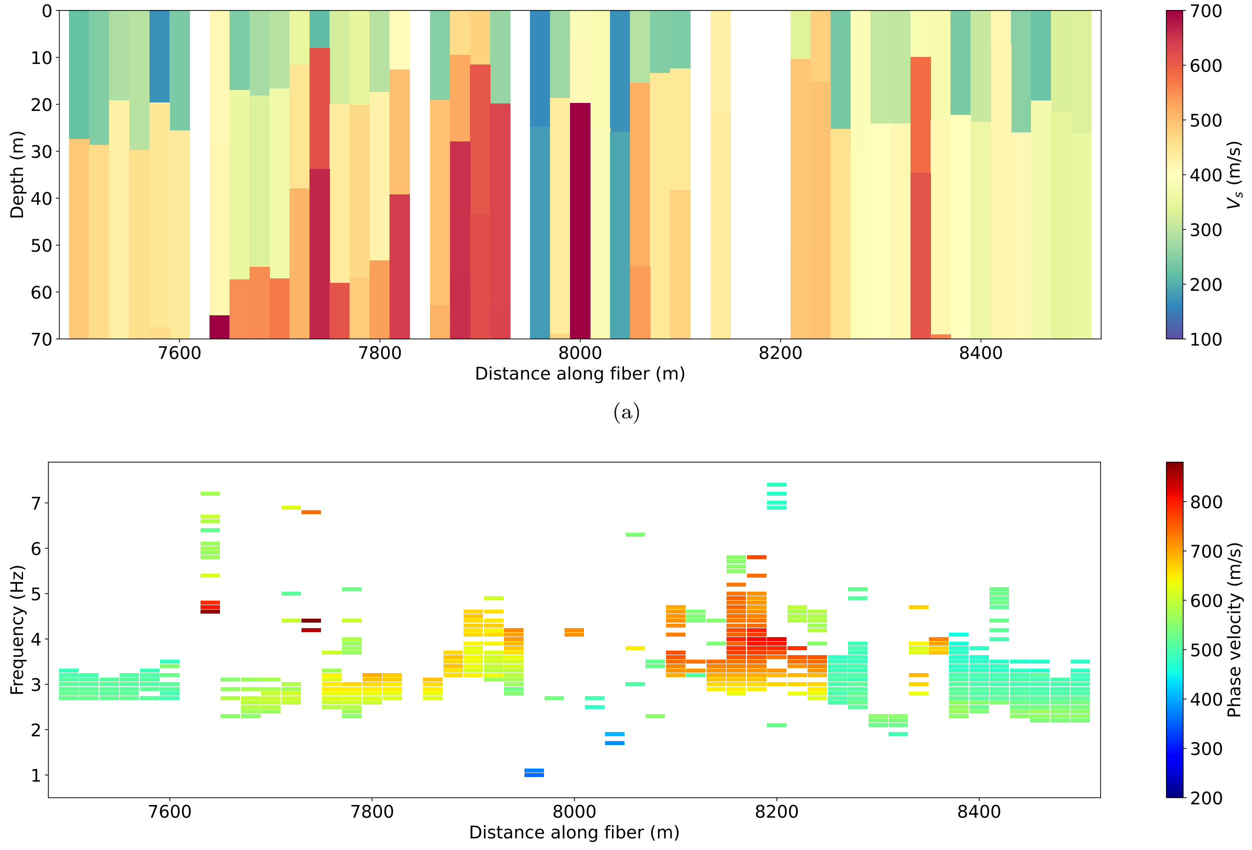


(b)

Figure 2.26 (a) Psuedo 2D shear wave velocity profile along fiber B from FB_1 and (b) its corresponding dispersion curve picks. Windowed between 11,000 m and 12,000 m.

Figure 2.27(b) shows a section of FB_1 that approximately mirrors the section from Figure 2.26(b). Both cases show low velocity anomalies that are bounded by high velocity anomalies. We expect the later section of fiber to have worse results since the light pulses that propagate through the fiber attenuates as a factor of fiber distance. We see that the low velocity section in the earlier section consists of only 2 profiles, whereas the later portion of fiber contains 9 anomalously slow profiles. The low velocity anomaly between 11,100 m and 11,200 m does not appear within the earlier section of fiber. It is possible that the lack of sensitivity to ground motion in the later portions of fiber, reflect as apparently slow surface waves. The high velocity anomalies between 7,740 m and 7,920 m appears similar to the anomaly between 11,460 m and 11,600 m with the exception that the anomaly in the earlier portion of fiber spans a greater distance.

It is important to note that there is no consistent underestimation of shear wave velocities or phase velocities in the later section of fiber. In this case, it is not clear whether fiber distance in relation to the interrogator has a noticeable effect on the computed V_s models. Further study may be necessary within this area to assess the feasibility of the automated workflow on larger fiber arrays.



(a)

(b)

Figure 2.27 (a) Psuedo 2D shear wave velocity profile along fiber B from FB_1 and (b) its corresponding dispersion curve picks. Windowed between 7,500 m and 8,500 m.

CHAPTER 3

CONCLUSIONS AND FUTURE WORK

This thesis explores a semi-automated data processing workflow to build pseudo 2-D, near-surface V_s profiles using ambient DAS data. Virtual sources are created across a fiber array by the cross-correlation of a primary trace with adjacent traces. Due to the volume of NCFs generated along a fiber, I employ an automated fundamental mode picking strategy to extract dispersion curves. These estimates are then inverted for individual V_s profiles that are then concatenated to construct a pseudo 2-D V_s section.

I applied this workflow to two dark fiber arrays to create V_s sections using 12, 24, and 36 hours of ambient DAS data. These analyses was repeated on data subsets to assess the time-lapse repeatability of the processing and analysis workflow. From this, I calculated $V_{s_{30}}$ estimates from each inverted profile to evaluate the efficacy of this method for earthquake hazard site classification. When analyzing these results, I found that large-scale subsurface structures were resolvable through this methodology; however, smaller-scale and geological unrealistic heterogeneities were also present in the inversion results.

To analyze potential causes of anomalous heterogeneities, I reviewed potential shortcomings within the automated workflow. I computed root-mean-square deviation curves to determine the NCF convergence rates at various locations along both fibers. I also assessed the automatically picked dispersion curves and found that inconsistent mode picking contributed significantly to the unreasonable lateral heterogeneity observed in the V_s sections.

3.1 Future work

The semi-autonomous workflow is shown to be repeatable with errors primarily associated with automated mode picking. However, full automation of the method is not yet complete and further work is required to improve the robustness of the resulting velocity models.

First, an autonomous strategy for identifying and handling directional changes in acquisition geometry is necessary to fully automate the ambient model-building workflow. The mixture of surface-wave modes resulting from fiber directionality changes can greatly affect dispersion analysis and the resulting inversion [29]. Improvements in this regard can also increase the coverage of virtual sources along DAS arrays. This problem has an increased importance in urban environments as ambient wavefields can change significantly in urban areas, and mapping of fiber geometry may not be accurate or possible.

Next, the removal of unwanted strong, coherent, or near-field energy sources is necessary to ensure that seismic interferometry accurately reconstructs surface-wave Green's functions between receiver pairs. Fully implementing the automated ambient processing workflow shown in [26] could improve the quality of NCFs and reduce the volume of ambient data needed to achieve convergence. Further analysis of the ambient energy distribution could also be included in the workflow in the form of beamforming to identify optimal fiber subarrays from which to compute NCFs and thereby account for the changing ambient wavefield.

A multimodal dispersion inversion approach [16, 37–39] could be useful in improving the quality of individual inverted V_s profiles and forestall picking issues associated with mode classification. This could lead to V_s profiles generated with greater confidence and repeatability.

Lastly, implementing this methodology using either cloud or edge computing could enable near-real-time, near-surface characterization during ambient data acquisition. Future work could be done to assess processing workflow scalability as well as parallelizing and optimizing for GPU applications. Further studies could identify pathways to continuous updating of near-surface models within urban areas and could lead to detection of subsurface hazards before they present on the surface and/or cause potential damage to urban populations and infrastructure.

REFERENCES

[1] J. Leinauer, S. Weber, A. Cicoira, J. Beutel, and M. Krautblatter. An approach for prospective forecasting of rock slope failure time. *Communications Earth & Environment*, 4(1):1–10, 2023. doi: 10.1038/s43247-023-00909-z.

[2] X. Fan, G. Scaringi, O. Korup, A. J. West, C. J. van Westen, H. Tanyas, N. Hovius, T. C. Hales, R. W. Jibson, K. E. Allstadt, L. Zhang, S. G. Evans, C. Xu, G. Li, X. Pei, Q. Xu, and R. Huang. Earthquake-induced chains of geologic hazards: Patterns, mechanisms, and impacts. *Reviews of Geophysics*, 57(2):421–503, 2019. doi: 10.1029/2018RG000626.

[3] N. M. Shapiro and M. Campillo. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, 31(7), 2004. doi: 10.1029/2004GL019491.

[4] K. Wapenaar, D. Draganov, R. Snieder, X. Campman, and A. Verdel. Tutorial on seismic interferometry: Part 1 — basic principles and applications. *Geophysics*, 75(5):75A195–75A209, 2010. doi: 10.1190/1.3457445.

[5] B. Biondi, R. G. Clapp, S. Yuan, and F. Huot. Scaling up to city-wide dark-fiber seismic arrays: Lessons from five years of the Stanford DAS array project. In *First International Meeting for Applied Geoscience & Energy Expanded Abstracts*, SEG Technical Program Expanded Abstracts, pages 3225–3229. 2021. doi: 10.1190/segam2021-3594539.1.

[6] S. Dou, N. Lindsey, A. M. Wagner, T. M. Daley, B. Freifeld, M. Robertson, J. Peterson, C. Ulrich, E. R. Martin, and J. B. Ajo-Franklin. Distributed acoustic sensing for seismic monitoring of the near surface: A traffic-noise interferometry case study. *Scientific Reports*, 7(1):11620, 2017. doi: 10.1038/s41598-017-11986-4.

[7] N. J. Lindsey and E. R. Martin. Fiber-optic seismology. *Annual Review of Earth and Planetary Sciences*, 49(1):309–336, 2021. doi: 10.1146/annurev-earth-072420-065213.

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

[9] C. B. Park, R. D. Miller, and J. Xia. Imaging dispersion curves of surface waves on multi-channel record. In *SEG Technical Program Expanded Abstracts 1998*, pages 1377–1380, 1998. doi: 10.1190/1.1820161.

[10] S. Nazarian, K. Stokoe, and W. Hudson. Use of spectral analysis of surface waves method for determination of moduli and thicknesses of pavement systems. *Transportation Research Record*, (930): 38–45, 1983.

[11] N. Gucunski and R. D. Woods. Numerical simulation of the SASW test. *Soil Dynamics and Earthquake Engineering*, 11(4):213–227, 1992. doi: 10.1016/0267-7261(92)90036-D.

[12] C. Park. MASW for geotechnical site investigation. *The Leading Edge*, 32(6):656–662, 2013. doi: 10.1190/tle32060656.1.

[13] E. A. Olafsdottir, S. Erlingsson, and B. Bessason. Tool for analysis of multichannel analysis of surface waves (MASW) field data and evaluation of shear wave velocity profiles of soils. *Canadian Geotechnical Journal*, 55(2):217–233, 2018. doi: 10.1139/cgj-2016-0302.

[14] E. A. Olafsdottir, B. Bessason, and S. Erlingsson. Combination of dispersion curves from MASW measurements. *Soil Dynamics and Earthquake Engineering*, 113:473–487, 2018. doi: 10.1016/j.soildyn.2018.05.025.

[15] J. Xia, R. D. Miller, C. B. Park, and J. Ivanov. Construction of 2-D vertical shear-wave velocity field by the multichannel analysis of surface wave technique. *Proceedings of the Symposium on the Application of Geophysics to Engineering and Environmental Problems*, pages 1197–1206, 2000. doi: 10.4133/1.2922726.

[16] J. B. Ajo-Franklin, S. Dou, N. J. Lindsey, I. Monga, C. Tracy, M. Robertson, V. Rodriguez Tribaldos, C. Ulrich, B. Freifeld, T. Daley, and X. Li. Distributed acoustic sensing using dark fiber for near-surface characterization and broadband seismic event detection. *Scientific Reports*, 9(1):1328, 2019. doi: 10.1038/s41598-018-36675-8.

[17] A. Tourei, X. Ji, G. Rocha dos Santos, R. Czarny, S. Rybakov, Z. Wang, M. Hallissey, E. R. Martin, M. Xiao, T. Zhu, D. Nicolsky, and A. Jensen. Mapping permafrost variability and degradation using seismic surface waves, electrical resistivity, and temperature sensing: A case study in Arctic Alaska. *Journal of Geophysical Research: Earth Surface*, 129(3):e2023JF007352, 2024. doi: 10.1029/2023JF007352.

[18] D. J. Wald and T. I. Allen. Topographic slope as a proxy for seismic site conditions and amplification. *Bulletin of the Seismological Society of America*, 97(5):1379–1395, 2007. doi: 10.1785/0120060267.

[19] Y. Choi and J. P. Stewart. Nonlinear site amplification as function of 30m shear wave velocity. *Earthquake Spectra*, 21(1):1–30, 2005. doi: 10.1193/1.1856535.

[20] K. T. Smolinski, D. C. Bowden, P. Paitz, F. Kugler, and A. Fichtner. Shallow subsurface imaging using challenging urban DAS data. *Seismological Research Letters*, 96(1):168–181, 2025. doi: 10.1785/0220240052.

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

[22] C. B. Park, R. D. Miller, and J. Xia. Multichannel analysis of surface waves. *Geophysics*, 64(3): 800–808, 1999. doi: 10.1190/1.1444590.

[23] Z. J. Spica, M. Perton, E. R. Martin, G. C. Beroza, and B. Biondi. Urban seismic site characterization by fiber-optic seismology. *Journal of Geophysical Research: Solid Earth*, 125(3):e2019JB018656, 2020. doi: 10.1029/2019JB018656.

[24] J. Shragge, J. Yang, N. Issa, M. Roelens, M. Dentith, and S. Schediwy. Low-frequency ambient distributed acoustic sensing (DAS): case study from Perth, Australia. *Geophysical Journal International*, 226(1):564–581, 2021. doi: 10.1093/gji/ggab111.

[25] C. Comina, S. Foti, D. Boiero, and L. V. Socco. Reliability of VS,30 evaluation from surface-wave tests. *Journal of Geotechnical and Geoenvironmental Engineering*, 137(6):579–586, 2011. doi: 10.1061/(ASCE)GT.1943-5606.0000452.

[26] A. J. Girard and J. Shragge. Automated processing strategies for ambient seismic data. *Geophysical Prospecting*, 68(1):293–312, 2020. doi: 10.1111/1365-2478.12794.

[27] G. D. Bensen, M. H. Ritzwoller, M. P. Barmin, A. L. Levshin, F. Lin, M. P. Moschetti, N. M. Shapiro, and Y. Yang. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophysical Journal International*, 169(3):1239–1260, 2007. doi: 10.1111/j.1365-246X.2007.03374.x.

[28] Y. Zhao, Y. E. Li, and B. Li. On beamforming of DAS ambient noise recorded in an urban environment and Rayleigh-to-Love wave ratio estimation. *Journal of Geophysical Research: Solid Earth*, 128(8):e2022JB026339, 2023. doi: 10.1029/2022JB026339.

[29] Q. Ji, B. Luo, and B. Biondi. Exploiting the potential of urban DAS grids: Ambient-noise subsurface imaging using joint Rayleigh and Love waves. *Seismological Research Letters*, 95(3):1794–1811, 2024. doi: 10.1785/0220230104.

[30] D. Chambers, G. Jin, A. Tourei, A. H. S. Issah, A. Lellouch, E. Martin, D. Zhu, A. Girard, S. Yuan, T. Cullison, T. Snyder, S. Kim, N. Danes, N. Punithan, M. S. Boltz, and M. M. Mendoza. DASCore: A Python library for distributed fiber optic sensing. *Seismica*, 3(2), 2024. doi: 10.26443/seismica.v3i2.1184.

[31] Z. Wang, C. Sun, and D. Wu. Automatic picking of multi-mode surface-wave dispersion curves based on machine learning clustering methods. *Computers & Geosciences*, 153:104809, 2021. doi: 10.1016/j.cageo.2021.104809.

[32] M. Ester, H. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD'96)*, pages 226–231, 1996. doi: 10.5555/3001460.3001507.

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

[34] E. A. Olafsdottir, B. Bessason, S. Erlingsson, and A. M. Kaynia. A tool for processing and inversion of MASW data and a study of inter-session variability of MASW. *Geotechnical Testing Journal*, 47(5): 1006–1025, 2024. doi: 10.1520/GTJ20230380.

[35] R. Dobry, R. D. Borcherdt, C. B. Crouse, I. M. Idriss, W. B. Joyner, G. R. Martin, M. S. Power, E. E. Rinne, and R. B. Seed. New site coefficients and site classification system used in recent building seismic code provisions. *Earthquake Spectra*, 16(1):41–67, 2000. doi: 10.1193/1.1586082.

[36] T. L. Holzer, A. C. Padovani, M. J. Bennett, T. E. Noce, and J. C. Tinsley. Mapping NEHRP $\mathbf{V_{S30}}$ site classes. *Earthquake Spectra*, 21(2):353–370, 2005. doi: 10.1193/1.1895726.

[37] M. Maraschini and S. Foti. A Monte Carlo multimodal inversion of surface waves. *Geophysical Journal International*, 182(3):1557–1566, 2010. doi: 10.1111/j.1365-246X.2010.04703.x.

[38] M. Maraschini, F. Ernst, S. Foti, and L. V. Socco. A new misfit function for multimodal inversion of surface waves. *Geophysics*, 75(4):31–43, 2010. doi: 10.1190/1.3436539.

[39] Y. Yang, Z. Zhan, M. Karrenbach, A. Reid-McLaughlin, E. Biondi, D. A. Wiens, and R. C. Aster. Characterizing South Pole firn structure with fiber optic sensing. *Geophysical Research Letters*, 51 (13):e2024GL109183, 2024. doi: 10.1029/2024GL109183.