Ambient Acoustics as Indicator of Environmental Change in the Beaufort Sea: Experiments & Methods for Analysis

by

Rui Chen

Submitted to the Department of Mechanical Engineering

and

Joint Program in Applied Ocean Science and Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

and the

WOODS HOLE OCEANOGRAPHIC INSTITUTION

June 2021

Author .....
Joint Program in Applied Ocean Science and Engineering
May 01, 2021

Certified by .....
Henrik Schmidt
Professor of Mechanical and Ocean Engineering
Thesis Supervisor

Accepted by .....
David Ralston
Chair, Joint Committee for Applied Ocean Science and Engineering

Accepted by .....
Nicolas G. Hadjiconstantinou
Chair, Committee on Graduate Students for Mechanical Engineering

Ambient Acoustics as Indicator of Environmental Change in the Beaufort Sea: Experiments & Methods for Analysis

by

Rui Chen

Submitted to the Department of Mechanical Engineering
and
Joint Program in Applied Ocean Science and Engineering
on May 01, 2021,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Abstract

The Arctic Ocean is a vital component of Earth's climate system experiencing dramatic environmental changes. The changes are reflected in its underwater ambient soundscape as its origin and propagation are primarily dependent on properties of the ice cover and water column.

The first component of this work examines the effect on ambient noise characteristics due to changes to the Beaufort Sea sound speed profile (SSP) and ice cover. Specifically, the emergence of a warm water intrusion near 70 m depth has altered the historical Arctic SSP while the ice cover has become thinner and younger due to the rise in average global temperature. Hypothesized shifts to the ambient soundscape and surface noise generation due to these changes are verified by comparing the measured noise data during two experiments to modeled results. These changes include a broadside notch in noise vertical directionality as well as a shift from uniform surface noise generation to discrete generation at specific ranges.

Motivated by our data analyses, the second component presents several tools to facilitate ambient noise characterization and generation monitoring. One is a convolutional neural network (CNN) approach to noise range estimation. Its robustness to SSP and bottom depth mismatch is compared with conventional matched field processing. We further explore how the CNN approach achieves its performance by examining its intermediate outputs. Another tool is a frequency domain, transient event detection algorithm that leverages image processing and hierarchical clustering to identify and categorize noise transients in data spectrograms. The spectral content retained by this method enables insight into the generation mechanism of the detected events by the ice cover. Lastly, we present the deployment of a seismo-acoustic system to localize transient events. Two forward approaches that utilize time-difference-of-arrival are described and compared with a more conventional, inverse technique. The

examination of this system's performance prompts recommendations for future deployments.

With our ambient noise analysis and algorithm development, we hope these contributions provide a stronger foundation for continued study of the Arctic ambient soundscape as the region continues to grow in significance.

Thesis Supervisor: Henrik Schmidt
Title: Professor of Mechanical and Ocean Engineering

Acknowledgments

This thesis is the culmination of five transformative years at MIT and WHOI. I would like to express gratitude to the numerous mentors, colleagues, friends, and family members who have supported me and helped me grow as a scientist and as a person.

First and foremost, thank you to my advisor, Professor Henrik Schmidt. I am inspired by your knowledge and optimism. Your encouragement and guidance have made me a more skilled and independent researcher. Further gratitude goes to Dr. Erin Fischell for her constant support and kindness. Thank you for the confidence that you instill in me. To Professor Arthur Baggeroer and Dr. Julien Bonnel, thank you for your time and valuable insights. Your comments and suggestions have challenged me to think more critically and strengthened this work. A big shoutout goes to Andrew Poulsen, Toby Schneider, Scott Carper, and Thomas Howe, without whom, the collection of the ICEX-16 data would not have been possible. Kevin Maganini, Brendan O'Neill, and Caileigh Fitzgerald were instrumental in carrying out the SIDEx experiment; I am sincerely grateful for their efforts. On the administrative side, thank you to Geoff Fox, Kris Kipp, and Lea Fraser for their assistance over the past few years.

Thank you to my colleagues and friends for the many great memories during this journey. To EeShan, Oscar, Kristen, Nick, Greg, Bradli, Dan, Michael, Blake, Supun, Craig, and everyone in LAMSS, I will always cherish our many discussions, field experiments, and social hours. Thank you for all of your help and encouragement. I am honored to have worked alongside you. To Jeffrey and Tom, time spent with you has brought me much joy. Thank you for our many moments of levity and your friendship.

To my family, I would like to thank my parents for their unwavering love and support. Your sacrifices have made this endeavor much easier. Thank you to my fiancé, Christine, for her love and encouragement. Your confidence in me keeps me motivated everyday; I could not have done this without you. Thank you as well to

the entire Schlaug family for their hospitality and support of Christine and I as we begin the next chapter of our lives.

Last but not least, this research was funded and supported by a number of sponsors; they are gratefully acknowledged below.

- Office of Naval Research (ONR) via the University of California - San Diego (UCSD) under award number N00014-16-1-2129.
- Defense Advanced Research Projects Agency (DARPA) via Applied Physical Sciences Corp. (APS) under award number HR0011-18-C-0008.
- Office of Naval Research (ONR) under award number N00014-17-1-2474.
- Office of Naval Research (ONR) under award number N00014-19-1-2741.
- National Science Foundation (NSF) under grant number 2389237.

Contents

List of Figures

List of Tables

List of Acronyms

AIDJEX Arctic Ice Dynamic Joint Experiment

AUV autonomous underwater vehicles

CE cross-entropy

CNN convolutional neural network

CNN-c classification convolutional neural network

CNN-r regression convolutional neural network

CTD conductivity-temperature-depth

DFT discrete Fourier transform

DOA direction of arrival

FC fully connected

FFT fast Fourier transform

FN false negative

FPR false positive rate

GPS global positioning system

HICEAS Hawaiian Islands Cetacean & Ecosystem Assessment Survey

h-clustering hierarchical clustering

ICEX Ice Exercise

LAMSS Laboratory for Autonomous Marine Sensing Systems

ltbb long-time-broadband

LTI linear time invariant

ltnb long-time-narrowband

ITP ice-tethered profiler

MAE mean absolute error

MCM multiple constraints method

MFP matched field processing

MIT Massachusetts Institute of Technology

ML machine learning

MPD motion product detector

MSE mean squared error

NATO North Atlantic Treaty Organization

NN neural network

ONR Office of Naval Research

PSD power spectral density

RMS root-mean-squared

SCM sample covariance matrix

SELU scaled exponential linear unit

SH horizontally polarized transverse

SIDEx Sea Ice Dynamics Experiment

SIMI Sea Ice Mechanics Initiative

SNR signal-to-noise ratio

SSP sound speed profile

stbb short-time-broadband

stnb short-time-narrowband

TDoA time difference of arrival

TP true positive

TPR true positive rate

UTC Coordinated Universal Time

VLA vertical line array

WHOI Woods Hole Oceanographic Institution

Chapter 1

Introduction

Although the Arctic Ocean is the smallest of the world's five main bodies of water, it holds no less climatological, economical, and geopolitical significance. The region is the most affected by the rise in average global temperature [1, 2], contains vast amounts of natural resources, and is an emerging space for economic and military activity. For much of the year, the Arctic is also unique in that it is covered with a layer of ice, making naval operations both on and below the surface more difficult. This combination of strategic importance and inaccessible environment has prompted growing interest in Arctic research in recent decades, both to study how the region is being affected by the changing climate as well as how to safely and effectively conduct naval missions in its waters. As with other ocean environments, the application of underwater acoustics is an essential tool for studying the Arctic as sound propagation depends on various water properties such as temperature, pressure, and salinity [3]. Acoustics is also the primary medium for underwater communication and navigation, which enable under-ice naval operations with both manned and autonomous underwater vehicles (AUV). A major component of the acoustic environment of any region is the ambient soundscape; it is the foundational noise above which other signals propagate. Thus, an accurate characterization of the passive ambient soundscape leads to more effective transmission of active signals, which, in turn, contributes to more reliable signal processing for undersea missions. In the Arctic, the ambient soundscape is also

particularly important because the primary origin of ambient noise is the ice cover [4]. As a result, changes in ice cover properties also affect the ambient soundscape in addition to water properties. Inversely, the ambient noise environment can be studied to better understand shifts in the Arctic ice cover and water column that result from broader changes in climate.

The earliest studies of Arctic ambient noise were conducted by Milne and Ganton during the ICE PACK experiments in 1961 and 1963 [5]. Underwater noise measurements were collected in the Canadian Archipelago below 20 kHz and linked to thermal stress and relative floe motions in the ice cover. Following this, the Arctic Ice Dynamic Joint Experiment (AIDJEX) in 1975-1976 enabled examination of the seasonal spatial and temporal scales of ambient noise in the Beaufort Sea below 1000 Hz [6] as well as noise generation mechanisms in the ice cover [6, 7, 8]. Then, during the four year span between 1979-1982, the Fram experiments collected ambient noise in the Fram strait below 10 kHz; this dataset led to analyses of the temporal distribution of transient noise events [9] and environmental correlates to low frequency ambient noise between 10-20 Hz [10]. The next major Arctic expedition occurred in 1994; as part of the Sea Ice Mechanics Initiative (SIMI), seismo-acoustic emissions of ice cover cracking [11] and the physics behind ice-induced acoustic events [12] were studied using a combination of hydrophone and geophone sensors. More recently, the U.S. Navy has continued to support acoustics research in the Arctic region through the Ice Exercise (ICEX) experiments (most recently in 2020). These exercises allow the Navy to test operational strategies in the region while enabling scientists to document the effect of environmental changes on underwater ambient noise and develop new frameworks for under-ice acoustic communication and navigation [13, 14].

This work is a further contribution to the study of Arctic underwater noise and, in particular, focuses on the Beaufort Sea region (Fig. 1.1), which has been experiencing dramatic shifts to both its ice cover properties as well as its underwater environment. This thesis consists of two main components. The first presents a hypothesis for how environmental changes in the Beaufort Sea have impacted the underwater ambient soundscape; we further demonstrate using acoustic modeling that our hypothesis is

consistent with properties of measured ambient noise. The second component describes various tools that aid in our data analysis and noise processing and showcases their benefits over more conventional techniques. More broadly, we hope to promote more accurate characterizations of the Arctic ambient soundscape, which would enable more successful future underwater operations in this region.

1.1 Significance of the Arctic Region

In this section, we further elaborate on the importance of the Arctic Ocean to global climate, economy, and geopolitics that motivates its continued study and monitoring.

1.1.1 Climatological Importance

In addition to being more sensitive to changes to the global climate than other regions, the Arctic Ocean also has critical influence over many components of the climate system, particularly in the northern hemisphere. For example, the ice cover is a significant regulator of material flux between the Arctic Ocean and the atmosphere as it reduces the exchange of heat, momentum, and water vapor [16]. With decreasing ice cover extent, increased heat flux from the ocean has been linked to a rise in annual mean air temperature (measured at 2 m) north of 60^oN, with the largest increase in coastal areas surrounding the Arctic [17, 18]. There is also evidence that Arctic sea ice decline contributes to snowier winters and wetter summers in northern Europe [16, 19]. Circulation patterns within the Arctic Ocean also cause it to be an exporter of water into the north Atlantic. In recent decades, the movement of water out of the Arctic basin has become increasingly fresh as a result of greater sea ice melting and runoff from river tributaries [20, 21, 22]. Consequently, the influx of fresh water has reduced the north Atlantic deep water formation and the Atlantic thermohaline circulation [23]. There are also observed effects to the spatial distribution of underwater biota in both the Arctic and Atlantic Oceans. Specifically, the biogeographic ranges of boreal plankton and fish species have retreated while those of subtropical and temperate species in the Northeast Atlantic have expanded northward [24].

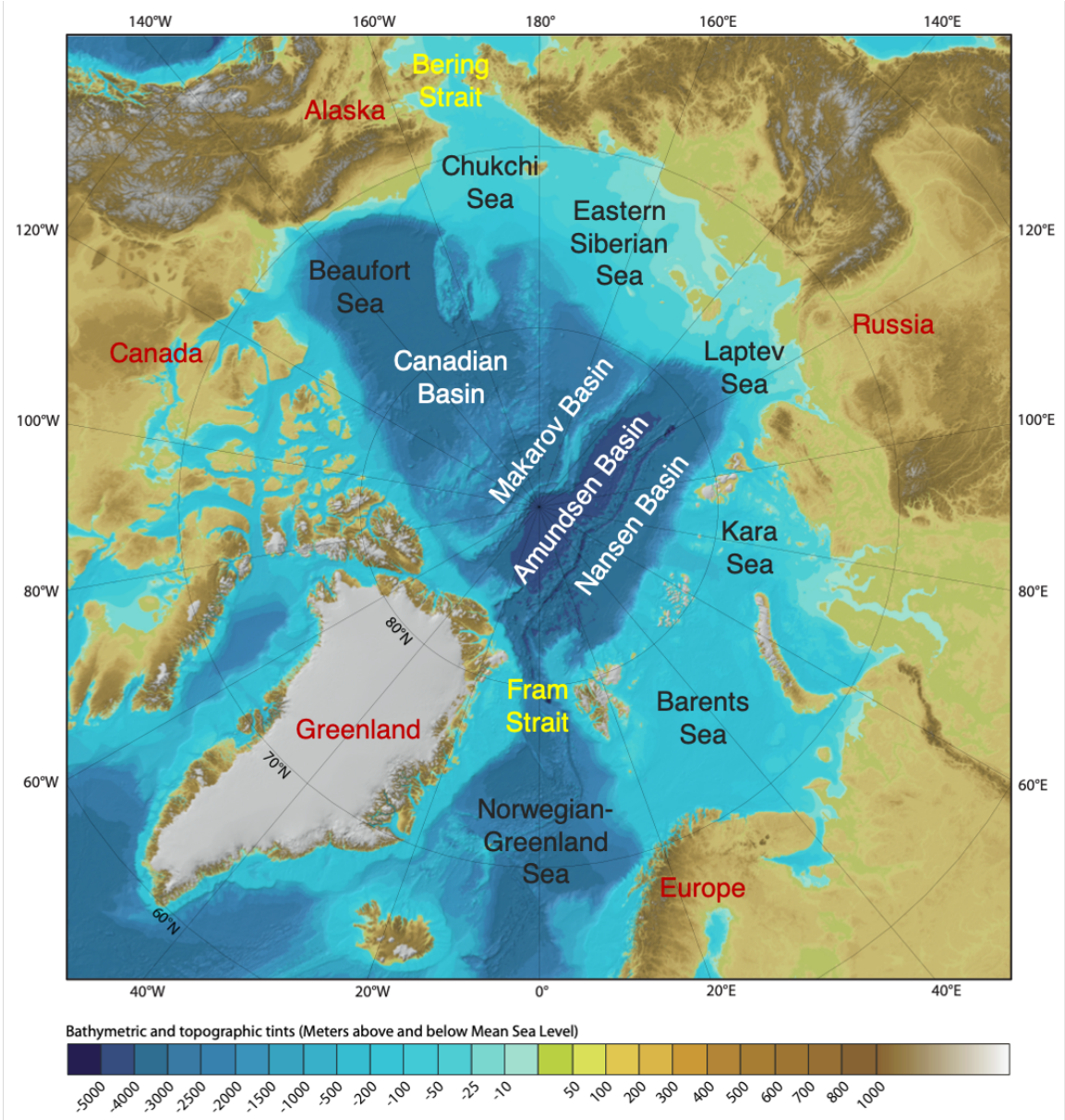


Figure 1.1: Arctic Ocean geography and bathymetry [4, 15].

1.1.2 Economical Importance

Despite its relative inaccessibility, the Arctic Ocean hosts a non-trivial portion of the world's economic activity in the form of aquaculture, resource extraction, and more recently, shipping. The region contains the world's largest cod stock in the form of the Northeast Arctic Cod (*Gadus morhua*)[25]. With rising ocean temperatures, the recruitment of Arctic cod is expected to increase as well [26], further contributing to the region's economic potential. However, this relationship is not linear and the correlation turns negative for water temperatures higher than 9^oC [27]. In regards to natural resources, the Arctic region accounts for 1/10 of the world's total oil production and 1/4 of the world's total natural gas production [28]. It also contains an estimated 30% of the world's undiscovered natural gas and 13% of the undiscovered oil, mostly offshore under less than 500 m of water [29]. With diminishing ice coverage, these resources are likely to become increasingly accessable, further bolstering the activity of extraction industries in the region. Also as a result of diminished ice coverage, another economic perspective in the Arctic is the opening of new shipping lanes between Asia and Europe that are 2/3 of the average distance of existing routes [30, 31]. Such shipping lanes would undoubtedly further stimulate international commercial development and investment in the Arctic.

1.1.3 Geopolitical Importance

The Arctic Ocean is bordered by many of the world's most influential sovereignties, namely, the United States, Canada, the European Union, and Russia. As a result, the region holds immense geopolitical significance. While the nations that surround the Arctic share many known common interests such as increasing political stability, scientific cooperation, and environmental protection and assessment [32], there are, nonetheless, also efforts of military strength projection. In particular, Russia has increased its investment in military infrastructure in the Arctic by constructing new bases, manufacturing new icebreakers, as well as establishing two special "Arctic brigades" [33]. Similarly, the NATO allied nations have established joint military

training exercises and the Arctic Joint Security Forces round-table meetings [34]. With continued environmental changes and further accessibility of the Arctic Ocean, the region is bound to attract even more scientific and military activity in the future.

1.2 Common Tools of Signal Processing

As evident by the significance of the Arctic Ocean presented in the previous section, there will likely be increased scientific, economic, and military activity in the region in the future. Much of these activities would likely require the use of underwater acoustics, whether for environmental characterization, sensing, or underwater communication and navigation. In this section, we provide a broad overview of ocean acoustics and introduce some common tools for acoustic signal processing and modeling. This overview also serves as background for later chapters as the presented tools are used in ambient noise data analysis.

1.2.1 Acoustic Wave Equation

Underwater acoustic propagation is governed by the wave equation, which is derived following the principles of conservation of mass, Newton's second law, and the adiabatic relation between pressure and density [3]. In an ideal fluid, these three principles manifest, respectively, in the equations below.

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \rho \mathbf{v}, \tag{1.1}$$

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = -\frac{1}{\rho}\nabla p(\rho), \tag{1.2}$$

$$p = p_0 + \rho' \left[\frac{\partial p}{\partial \rho}\right]_s + \frac{1}{2}(\rho')^2 \left[\frac{\partial^2 p}{\partial \rho^2}\right]_s + \cdots, \tag{1.3}$$

where ρ is the density, p is the pressure, and $\mathbf{v}$ is the particle velocity. Small perturbations to the ambient pressure p_0 and density ρ_0 are denoted by p' and ρ', respectively. The subscript s denotes that thermodynamic derivatives are taken at constant en-

tropy. We further define the sound speed, c, in the medium as

$$c^2 \equiv \left[\frac{\partial p}{\partial \rho}\right]_s . \tag{1.4}$$

By only keeping the first order terms in the equations above and combining together, the linear acoustic wave equation is derived as

$$\rho \nabla \cdot (\frac{1}{\rho} \nabla p) - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0. \tag{1.5}$$

Furthermore, if the density is constant in space, the equation above simplifies to

$$\nabla^2 p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = 0. \tag{1.6}$$

Instead of pressure, Eq. 1.6 is commonly re-formulated in terms of the displacement potential, defined with regards to the particle displacement, $\mathbf{u}$, as

$$\mathbf{u} = \nabla \psi. \tag{1.7}$$

The displacement potential relates to the pressure following

$$p = -\rho \frac{\partial^2 \psi}{\partial t^2}, \tag{1.8}$$

which means that it is also governed by the wave equation

$$\nabla^2 \psi - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = 0. \tag{1.9}$$

With the right-hand side of Eq. 1.9 equal to 0, this equation represents the scenario where there is no acoustic source in the medium. However, a source can easily be implemented by including a forcing term on the right-hand side, leading to the inhomogeneous wave equation

$$\nabla^2 \psi - \frac{1}{c^2} \frac{\partial^2 \psi}{\partial t^2} = f(\mathbf{r}, t), \tag{1.10}$$

where $f(\mathbf{r}, t)$ represents a volume injection as a function of space and time. This equation can be solved analytically for some simple forcing terms in propagation environments with straight-forward boundary conditions. However, most often, it is solved numerically. To simplify computation, numerical approaches typically solve the wave equation in the frequency domain instead of the time domain. This is because by taking the Fourier transform of Eq. 1.10, the time dependence of the equation can be removed, forming what is known as the Helmholtz equation

$$\left[\nabla^2 + k^2(\mathbf{r})\right] \psi(\mathbf{r}, \omega) = f(\mathbf{r}, \omega), \tag{1.11}$$

where $k(\mathbf{r})$ is the medium wavenumber at the radian frequency ω,

$$k(\mathbf{r}) = \frac{\omega}{c(\mathbf{r})}. \tag{1.12}$$

As we will present in Section 1.2.5, one approach to numerically solve the Helmholtz equation is through wavenumber integration, which is the basis of the modeling software OASES [35].

1.2.2 Power Spectral Density and Spectrogram

The power spectral density (PSD) and the spectrogram are both tools that enable visualization of an acoustic time series' spectral content. In particular, the PSD describes a signal's distribution of power into frequencies. Given a time domain signal snapshot of the form $x(t)$, its PSD is computed as

$$S(\omega) = \frac{1}{T}|X(\omega)|^2, \tag{1.13}$$

where T is the length of $x(t)$, $X(\omega)$ is the Fourier transform of $x(t)$, and ω is the radian frequency. In practice, for a discrete, wide band signal $x\,[t]$, the time series would be segmented into snapshots of length T to calculate the PSD of each snapshot.

For example, for a snapshot centered at $t = t_1$, the PSD is calculated as

$$S_{t_1}[\omega] = \frac{1}{T}|X_{t_1}[\omega]|^2, \tag{1.14}$$

where $X_{t_1}[\omega]$ is the discrete Fourier transform (DFT) of the snapshot $x\left[t_1 - \frac{T}{2}, \ldots, t_1 + \frac{T}{2}\right]$. The choice of T, along with the sampling frequency, F_s, of $x[t]$, determine the frequency domain resolution of $S_{t_1}[\omega]$. This relationship is described by

$$df = \frac{F_s}{T}. \tag{1.15}$$

The spectrogram is simply an extension of the PSD that shows the power distribution of a time series in frequency as well as in time. To generate a spectrogram, the PSD of consecutive snapshots of $x[t]$ are calculated and formed into a matrix

$$[S_{t_1}[\omega], S_{t_2}[\omega], \ldots]. \tag{1.16}$$

Often, $t_1, t_2, \ldots$ are selected so that there is some overlap between one snapshot and the next (50% overlap is a common choice). This step ensures no information is lost at the boundaries of the snapshots when computing the spectrogram.

1.2.3 Conventional Beamforming

Beamforming is a technique for estimating a signal's direction of arrival (DOA). From a set of time series collected with an array of recievers, it generates a distribution of measurement values over all spatial angles that reflect the most likely arrival direction of the recorded acoustics data. The conventional approach to beamforming is known as the delay-and-sum or Bartlett beamformer. Conceptually, it is quite straightforward. Given an array of receivers, each arbitrarily positioned in space and a plane wave incident upon the center of the array, the wave's DOA can be estimated by leveraging the time at which the wave reaches each receiver [36, 37]. For any incident acoustic signal, the plane wave assumption is satisfied if the array is located in the far field of source.

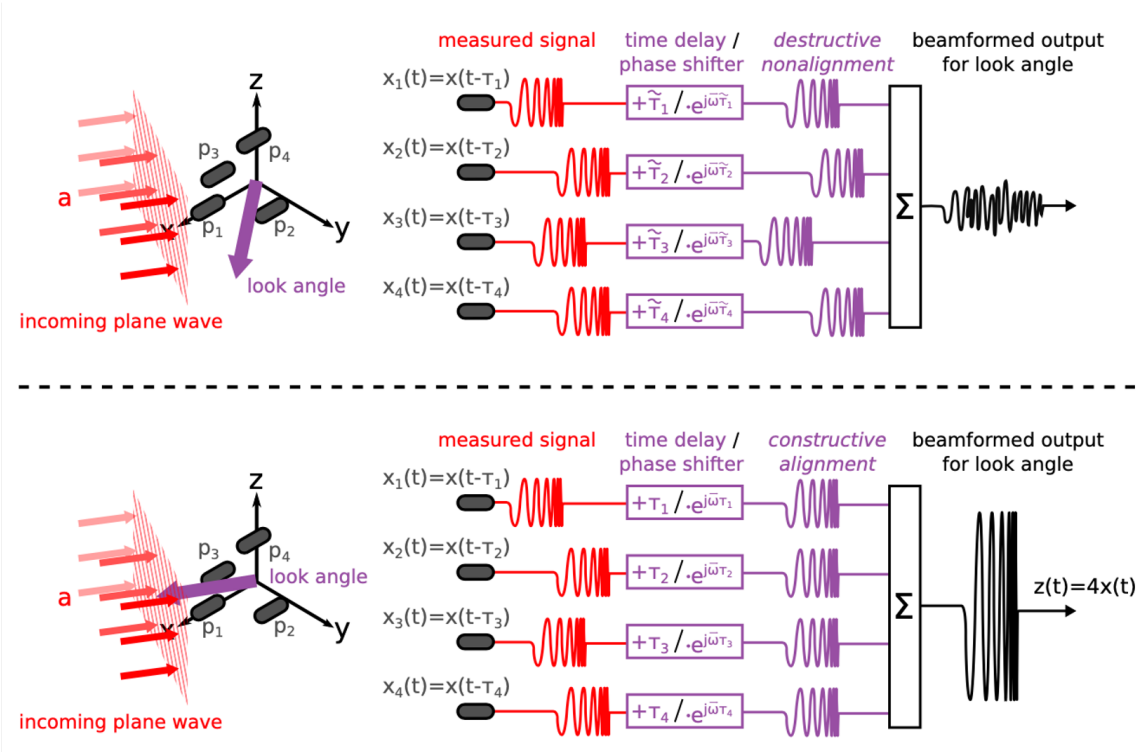


Figure 1.2: (Top) Beamformer look-angle is not pointing in the opposite direction of the incoming plane wave. Thus, the filtered signals are not aligned, resulting in destructive interference and a low output power. (Bottom) Beamformer look-angle *is* pointing in the opposite direction of the incoming plane wave. Thus, the filtered signals *are* aligned, resulting in *constructive* interference and a maximum output power. [37].

Consider an array of N omni-directional receivers at positions $\mathbf{p}_n$ subject to an incoming plane wave (Fig. 1.2). The sensors spatially sample the signal field at their respective locations $\mathbf{p}_n : n = 1, \ldots, N$, yielding a set of measurements

$$\mathbf{x}(t, \mathbf{p}) = \begin{pmatrix} x_1(t, \mathbf{p}_1) \\ \vdots \\ x_N(t, \mathbf{p}_N) \end{pmatrix}. \tag{1.17}$$

The measurements at each sensor are then processed using a linear, time invariant (LTI) filter with impulse response $h_n(\tau)$. The processed outputs are then summed to form an array output $y(t)$ following

$$y(t) = \sum_{n=1}^{N} \int_{-\infty}^{\infty} x_n(t, \mathbf{p}_n) h_n(t - \tau) d\tau = \int_{-\infty}^{\infty} \mathbf{h}^T(t - \tau) \mathbf{x}(t, \mathbf{p}) d\tau, \tag{1.18}$$

where

$$\mathbf{h}(\tau) = \begin{pmatrix} h_1(\tau) \\ \vdots \\ h_N(\tau) \end{pmatrix}. \tag{1.19}$$

Applying the Fourier transform to both sides of the equality, Eq. 1.18 can be written in the frequency domain as

$$Y(\omega) = \mathbf{H}^T(\omega) \mathbf{X}(\omega). \tag{1.20}$$

Now, consider a plane wave with radian frequency ω and propagation direction $\mathbf{a}$ (Fig. 1.2). Working in spherical coordinates, $\mathbf{a}$ is a unit vector described by

$$\mathbf{a} = \begin{pmatrix} -sin(\theta)cos(\phi) \\ -sin(\theta)sin(\phi) \\ -cos(\theta) \end{pmatrix}, \tag{1.21}$$

where θ is the inclination ($\theta = 0$ points along the positive z-axis) and ϕ is the azimuth ($\phi = 0$ points along the positive x-axis and positive rotation is around the z-axis following the right-hand rule). The time series measured at each receiver position,

$x(t, \mathbf{p})$, relative to measurements at some arbitrary origin of the array, $x(t)$, can be written as

$$\mathbf{x}(t, \mathbf{p}) = \begin{pmatrix} x(t - \tau_1) \\ \vdots \\ x(t - \tau_N) \end{pmatrix}, \tag{1.22}$$

where

$$\tau_n = \frac{\mathbf{a}^T \mathbf{p}_n}{c} \tag{1.23}$$

and c is the propagation speed of the wave. Again, taking the Fourier transform of every component of Eq. 1.22, $\mathbf{x}(t, \mathbf{p})$ is expressed in the frequency domain as

$$\mathbf{X}(\omega, \mathbf{p}) = \begin{pmatrix} X(\omega)e^{-j\omega\tau_1} \\ \vdots \\ X(\omega)e^{-j\omega\tau_N} \end{pmatrix}. \tag{1.24}$$

Using the relation between the spatial wavenumber of a wave and its radian frequency,

$$\mathbf{k} = \frac{\omega}{c}\mathbf{a}, \tag{1.25}$$

we can see that

$$\omega\tau_n = \mathbf{k}^T \mathbf{p}_n. \tag{1.26}$$

As a result, we can define the equivalent vectors

$$\mathbf{v}(\mathbf{k}) = \begin{pmatrix} e^{-j\mathbf{k}^T\mathbf{p}_1} \\ \vdots \\ e^{-j\mathbf{k}^T\mathbf{p}_N} \end{pmatrix} \quad \Leftrightarrow \quad \mathbf{v}(\omega) = \begin{pmatrix} e^{-j\omega\tau_1} \\ \vdots \\ e^{-j\omega\tau_N} \end{pmatrix}. \tag{1.27}$$

These vectors are known as the array manifold vectors. They contain all relevant characteristics of the array such as the sensor positions as well as how a plane wave of a certain frequency ω and incident direction (described by $\mathbf{k}$) is received by each

element in the array. Substituting Eq. 1.27 into Eq. 1.24, we can re-write $\mathbf{X}(\omega, \mathbf{p})$ as

$$\mathbf{X}(\omega, \mathbf{p}) = X(\omega)\mathbf{v}(\mathbf{k}) = X(\omega)\mathbf{v}(\omega). \tag{1.28}$$

Now, comparing the form of Eq. 1.28 to the form of $Y(\omega) = \mathbf{H}^T\mathbf{X}(\omega)$ (Eq. 1.20), it becomes apparent that the received field on the array due to its geometry is consistent with processing the array sensors with a set of LTI filters with transfer functions $H_n(\omega) = e^{-j\omega\tau_n}$. In other words, the array geometry shifts the signal received on receiver n by τ_n in time, which corresponds to the time difference in signal reception between receiver n and the array origin. This collection of LTI filters is what is referred to as the conventional beamformer. Consequently, the output of the beamformer for any look-angle, $\mathbf{u}$, is computed as

$$\mathbf{B}(\omega, \mathbf{u}) = \frac{1}{N}\mathbf{H}^T\mathbf{v}(\mathbf{u}), \tag{1.29}$$

where $\frac{1}{N}$ is included as a normalization factor.

As nicely illustrated by Rypkema [37] using Fig. 1.2, applying the set of LTI filters defined above at the look angle exactly opposite the wave propagation direction (*i.e.* $\mathbf{u} = -\mathbf{a}$) results in constructive alignment between the received time series on all array elements, creating a maximum beamformer output. Thus, if we simply loop through all possible look angles, we can make an estimate of the wave incident angle by computing

$$\underset{\mathbf{u}}{\text{argmax}}(\mathbf{B}(\omega, \mathbf{u})). \tag{1.30}$$

1.2.4 Raytracing

Raytracing is one of the earliest methods for modeling underwater acoustic propagation. Much of the understanding for the behavior of ray propagation was acquired long before its application to underwater sound, stemming from optics experiments by figures such as Euclid and Snell [3]. While newer underwater acoustics modeling tools have gained favor over raytracing over the years due to the latter's compara-

tively lower accuracy, particularly at low frequencies, raytracing is still widely used due to its fast computational speed [3, 38]. Other advantages of raytracing are its capability to model broadband propagation efficiently and its suitability for range-dependent environments [38]. In addition, the insight derived from raytracing outputs is important in corroborating and interpreting the results of other models [3].

The foundational strategy of raytracing as a modeling tool is to simulate the propagation paths and amplitudes of a fan of rays from an acoustic source out into the environment. The propagation paths of the rays are governed by the ray equations, which are derived from the Helmholtz equation (Eq. 1.11) assuming a point source forcing term and a high frequency approximation [3]:

$$\begin{aligned} \frac{dr}{ds} &= c\xi(s), \quad \frac{d\xi}{ds} = -\frac{1}{c^2}\frac{\partial c}{\partial r}, \\ \frac{dz}{ds} &= c\zeta(s), \quad \frac{d\zeta}{ds} = -\frac{1}{c^2}\frac{\partial c}{\partial z}, \end{aligned} \tag{1.31}$$

where s is arc-length along a ray, $c(s)$ is the propagation speed, $[r(s), z(s)]$ is a ray's trajectory in cylindrical coordinates, and $[\xi(s), \zeta(s)]$ describes the tangent vector to the ray ($\mathbf{t}_{ray} = c\,[\xi(s), \zeta(s)]$). The initial conditions of the ray specify the source position (r_0, z_0) and the starting slope of the ray given a specified take-off angle θ_0 (Fig. 1.3):

$$\begin{aligned} r &= r_0, \quad \xi = \frac{cos\theta_0}{c(0)}, \\ z &= z_0, \quad \zeta = \frac{sin\theta_0}{c(0)}. \end{aligned} \tag{1.32}$$

With these initial conditions, the propagation path of a ray can be obtained by solving the ray equations (Eq. 1.31) iteratively. After each step ds along the ray path, the new propagation angle of the ray is calculated by applying Snell's Law:

$$c_1 cos\theta_1 = c_2 cos\theta_2, \tag{1.33}$$

where c_1, c_2 are the sound speeds and θ_1, θ_2 are the ray angles before and after taking the step along the ray path, respectively.

Along with computing the ray paths, we also need to calculate the acoustic in-

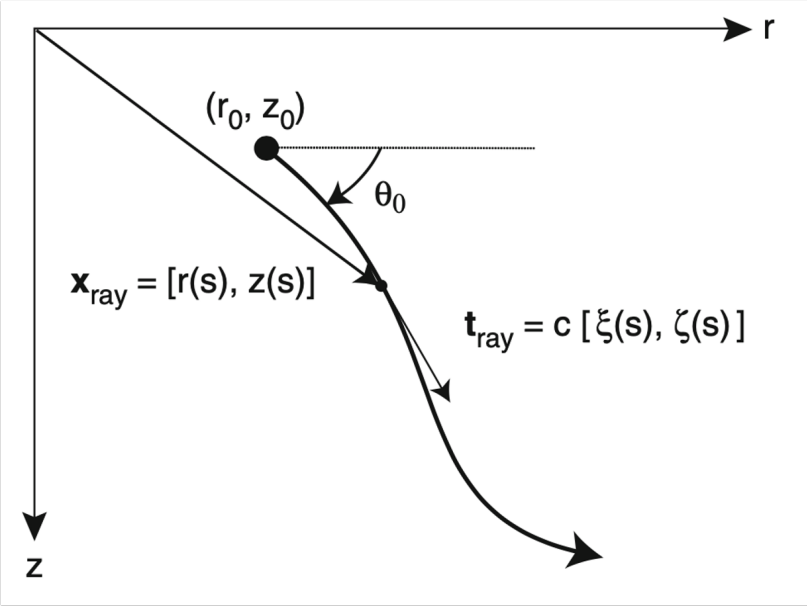


Figure 1.3: A schematic of 2-D ray tracing geometry [3].

tensity along each ray. The differential equations describing the intensity are the dynamic ray equations [39]:

$$\frac{dq}{ds} = cp(s), \quad \frac{dp}{ds} = -\frac{c_{nn}}{c^2(s)}q(s), \tag{1.34}$$

where c_{nn} is the curvature of sound speed in the direction normal to the ray path and can be expressed as

$$c_{nn} = c^2(\frac{\partial^c}{\partial r^2}\zeta^2 - 2\frac{\partial^2 c}{\partial r \partial z}\zeta\xi + \frac{\partial^2 c}{\partial z^2}\xi^2). \tag{1.35}$$

These equations characterize the change in ray path due to general changes to the ray's initial condition. For the purpose of tracing a web of rays, we need to perturb the ray with respect to take-off angle. The appropriate initial conditions for the dynamic ray equations in this case are:

$$q(0) = 0, \quad p(0) = \frac{1}{c(0)}. \tag{1.36}$$

Consequently, the ray amplitude along its path is then given as

$$A(s) = \frac{1}{4\pi}\left|\frac{c(s)cos\theta_0}{rc(0)q(s)}\right|^{1/2}. \tag{1.37}$$

Similar to the computation of the ray trajectory before, the intensity along the ray path can be determined through iterative computation of the dynamic wave equations (Eq. 1.34) followed by Eq. 1.37. In practice, to calculated the full acoustic field, each simulated ray is given some "width" through the application of a shape function (common shapes include triangle or Gaussian). The pressure field at any point in the environment is then determined by summing the intensity of all rays (eigenrays) that pass through that point.

To implement raytracing, a popular and publicly available modeling software is BELLHOP [40]. This tool is used in this thesis to generate raytracing outputs.

1.2.5 Wavenumber Integration

Wavenumber integration is a numerical approach for solving the Helmholtz equation Eq. 1.11 for a source distribution in horizontally stratified media (environment varies in depth only) [3]. As a result, it is applied for range-independent propagation problems. This method divides the water column into many depth layers and then reduces the Helmholtz equation to a set of one-dimensional differential equations, one within each layer of the water column. These depth-separated equations are then solved analytically and the full field can be computed by matching the boundary conditions between the layers. There are no simplifying assumptions in the derivation of the depth-separated equations; in this sense, the pressure field calculated through wavenumber integration is exact, in contrast to techniques such as raytracing.

Consider a source distribution along a vertical axis in a horizontally stratified environment, the acoustic field in layer m containing the source is expressed in terms of the displacement potential following the Helmholtz equation,

$$\left[\nabla^2 + k_m^2(z)\right]\psi_m(r,z) = f_s(z,\omega)\frac{\delta(r)}{2\pi r}, \tag{1.38}$$

where $k_m(z) = \omega/c(z)$ is the medium wavenumber for layer m. Similarly, in layers without sources, the acoustics field satisfies the homogenous Helmholtz equation with $f_s(z, \omega) = 0$. Now, we apply to Eq. 1.38 the forward Hankel transform, defined as

$$f(k_r, z) = \int_0^\infty f(r, z) J_0(k_r r) r dr, \tag{1.39}$$

where $J_0(\cdot)$ is zeroth order Bessel function of the first kind. This operation results in the depth-separated wave equation,

$$\left[\frac{d^2}{dz^2} - \left[k_r^2 - k_m^2(z)\right]\right] \psi_m(k_r, z) = \frac{f_s(z)}{2\pi}, \tag{1.40}$$

which is an ordinary differential equation in depth. The solution to this type of differential equation is the sum of a particular solution, $\hat{\psi}_m(k_r, z)$, to Eq. 1.40 and a linear combination of the two independent solutions, $\psi_m^+(k_r, z)$ and $\psi_m^-(k_r, z)$, to the homogeneous equation [3]. Therefore, the total solution of the depth-dependent field is the, so-called, *depth-dependent Green's function*:

$$\psi_m(k_r, z) = \hat{\psi}_m(k_r, z) + A_m^+(k_r)\psi_m^+(k_r, z) + A_m^-(k_r)\psi_m^-(k_r, z), \tag{1.41}$$

where A_m^+ and A_m^- are coefficients to be determined by satisfying the boundary conditions at the interfaces between the layers. For the particular solution, the most convenient solution is simply the field produced by the source distribution in an environment absent of any boundary conditions. After the unknown coefficients are solved for, the last step to compute the full acoustic field is to apply the inverse Hankel transform, which is defined as

$$f(r, z) = \int_0^\infty f(k_r, z) J_0(k_r r) k_r dk_r. \tag{1.42}$$

As written, this inverse transform is an integration over the wavenumber domain, which is the namesake for the wavenumber integration approach. To solve for the coefficients in Eq. 1.41, one technique developed by Schmidt [41, 42] is the global

matrix approach. In this method, the local boundary conditions at each interface are assembled into a global system of equations expressing the boundary conditions at all interfaces. The numerical solution of the global system then yields the field in all layers simultaneously. The global matrix approach and the subsequent wavenumber integration is implemented by the publicly available software OASES [35]. This software is used in this thesis to carry out range independent acoustic modeling.

1.3 Objectives & Contributions

The primary objective of this thesis is to examine changes to the Beaufort Sea ambient soundscape and demonstrate that they are indicative of environmental shifts occurring in the region. This goal is motivated by the significant benefits that a better description of ambient noise provides in tracking environmental changes as well as enabling successful underwater naval operations in the region. To this end, the contributions of this work include: (i) a comparison of ambient noise characteristics near the same location at identical times of the year more than 20 years apart. The earlier ambient noise data was collected in 1994 while the latter in 2016. Differences in noise characteristics between these two times include spectral content, vertical directionality, and statistical distribution of transient events. (ii) A demonstration of how the changed noise features are consistent with modeled effects due to a shift in the sound speed profile (SSP) and surface noise generation. Specifically, Beaufort Lens SSP is responsible for a notch near horizontal arrivals in noise vertical directionality while the peak elevation angles in the noise directionality profile result from discrete surface noise generation, as expected from a thinner and younger ice cover.

Another objective of this work is to develop noise detection and localization approaches that aid in our analysis of ambient noise data and compare their performances to more conventional methods. To this end, this work makes the following contributions: (i) a model-based convolutional neural network (CNN) approach to source range estimation. This approach demonstrates more robustness to SSP and ocean bottom depth mismatch than conventional matched field processing (MFP).

(ii) A frequency domain event detection algorithm that identifies transients in the data spectrograms and categorizes them based on duration and bandwidth. This approach is an improvement over time domain, amplitude-based detection methods because the spectral shapes of the detected events provide insight into their generation mechanisms. (iii) Two forward approaches to localize ice-generated transient events recorded by a geophone array using time-difference-of-arrival (TDoA) that show similar performance to an inverse TDoA matching technique. Together, these methods demonstrate the importance of array coverage area to localization success. As secondary contributions in the area of transient localization using geophones, we further showcase the capability of remote geophone nodes for ice cover transient event monitoring and make suggestions for the deployment of future seismo-acoustic systems that combine both geophones and hydrophones.

1.4 Thesis Organization

The major components of this thesis are organized into the following chapters:

- **Chapter 2 - The Historical Arctic Environment**: This chapter describes the historical propagation environment in the Arctic Ocean pertaining to its underwater SSP and surface ambient noise generation. As a demonstration of the ambient soundscape of such an environment, noise data collected during the SIMI-94 experiment are analyzed and compared with modeled outputs. The temporal distribution of transient noise events within the collected data is also examined using a time domain, amplitude-based event detection algorithm.

- **Chapter 3 - Environmental effects on Beaufort Sea Ambient Noise**: This chapter introduces on-going environmental changes in the Beaufort Sea region of the Arctic Ocean and their hypothesized effects on ambient noise directionality and generation. Data from the ICEX-16 experiment are analyzed and compared with modeling outputs to showcase how the measured ambient noise characteristics are indicative of the observed Beaufort Sea environment.

The analysis results are further compared with those of the SIMI-94 data to identify differences that may be attributed to the changed environment.

- **Chapter 4 - Noise Generation Range Estimation**: This chapter presents a model-based CNN approach to noise range estimation as a more robust alternative to conventional MFP. The CNN method is trained using simulated data and tested with both simulated and real data against both SSP (using data from ICEX-16) and ocean bottom depth mismatch (using data from SWellEx-96). The reason for the CNN approach's improved performance is explored by examining the trained networks' intermediate outputs and comparing against the template replica vectors of MFP.

- **Chapter 5 - Transient Ice Noise Detection & Characterization**: This chapter describes a frequency domain, transient event detection algorithm as a more useful alternative to time domain, amplitude-based methods. By combining image processing techniques with hierarchical clustering (h-clustering), this method detects and groups transient features in the data spectrogram into events and then categorizes the events based on their bandwidth and duration. The spectral properties of the detected events can then aid in hypothesizing their generation mechanism by the ice cover. This frequency domain method is applied to ambient noise collected during ICEX-16 to characterize the temporal distribution of transient noise events during this experiment.

- **Chapter 6 - Ice Cover Transient Noise Monitoring**: This chapter presents the deployment of a seismo-acoustic system of geophones and hydrophones to monitor ice cover dynamics and detect transient cracking events during the SIDEx experiment. As part of the data processing, two forward event localization approaches are described that utilize the motion product detector (MPD) and TDoA between sensor pairs. These methods show similar performance to an inverse TDoA matching technique. The deployment of the seismo-acoustic system also demonstrates the utility of remote geophones for ice cover monitoring. Due to their cable-less nature, these remote units can be easily deployed to

form an array that covers a much larger area than what is feasible with cabled geophones.

Chapter 2

The Historical Arctic Environment

Before we highlight the environmental changes occurring in the Arctic Ocean, we first introduce the historical Arctic propagation environment in this chapter. As we describe below, the defining signature of the historical Arctic environment consists of a monotonically increasing sound speed profile (SSP) that promotes upward refraction of sound with a single surface acoustic duct. While this profile is not an exact description for the entire Arctic Ocean, it is exemplary of much of the central Arctic. In order to study the environment-induced effects on the underwater ambient soundscape, we need to first establish a control description of the noise in the historical Arctic environment. To do so, underwater ambient sound characteristics during the Sea Ice Mechanics Initiative (SIMI-94) experiment are analyzed and presented below. We further compare the measured data with modeling output using traditional assumptions for noise generation to examine their agreement. We begin with a brief background on the historical Arctic environment and noise modeling before introducing the SIMI-94 experiment and presenting noise analysis results.

2.1 Historical Arctic Propagation Environment

The underwater ambient noise soundscape in the ocean is governed by two main components - the propagation environment and the noise generation mechanism. The

propagation environment is largely defined by the SSP in the water and any boundaries such as the sea surface or bottom. In the deep ocean, the SSP has the most significant effect on acoustic propagation. Analogous to light refracting in different media with varying indices of refraction, sound waves refract towards lower sound speed in accordance to Snell's law (Eq. 1.33). The sound speed within a medium varies with the medium's density and compressibility. In water, these properties are dependent on temperature, salinity, and static pressure, the latter being a function of water depth. Thus, sound speed in water can be expressed as an empirical function of temperature (T) in degrees Celsius, salinity (S) in parts per thousand, and depth (z) in meters [3]. Eq. 2.1 demonstrates that, in general, sound speed increases with rise in temperature and salinity and with deepening water depth.

$$\begin{aligned} c(T,S,z) = 1449.2 + 4.6T - 0.055T^2 + 0.00029T^3 \\ +(1.34 - 0.01T)(S-35) + 0.016z. \end{aligned} \tag{2.1}$$

In the Arctic region, the existence of ice cover reflects much of the sun's energy and prevents it from penetrating into the water column. Consequently, the under-ice water column is relatively isothermal and the SSP can be typically described as monotonically increasing with depth (Fig. 2.1). Near the surface (upper $\sim$200 m), the increase in sound speed with depth is generally more dramatic because of a steeper salinity gradient - freshwater from ice cover melt results in lower salinity closer to the surface. Below this surface layer, the water column is close to isosaline and the increase in sound speed with depth becomes more gradual and constant. This SSP's effect on acoustic propagation is demonstrated in Fig. 2.1. For a near surface source, acoustic rays emitted at shallow departure angles are trapped within the surface duct due to the steeper sound speed gradient while steeper rays follow deeper refracted paths.

Another governing factor of underwater ambient noise is the noise generation mechanism. In the open ocean, primary contributors of ambient noise include wind-driven waves and surface vessel traffic [43]. In the Arctic, however, ice cover of

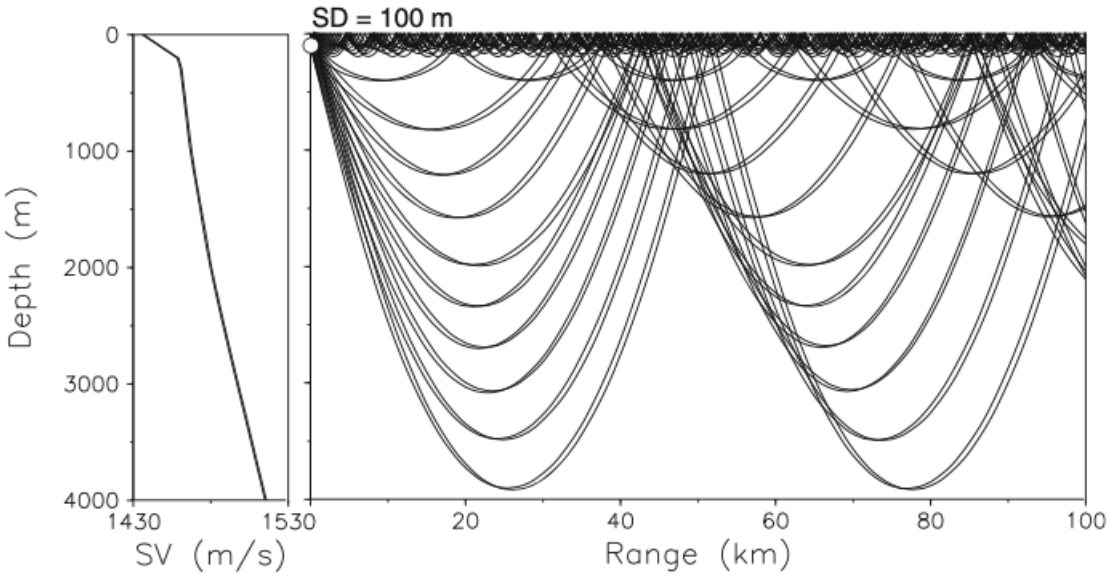


Figure 2.1: (Left) A historical, monotonically increasing Arctic ocean SSP. (Right) Propagation paths of acoustic rays emitted by a near surface source at 100 m. Rays with shallow departure angles are trapped within the surface duct while steeper rays follow deeper refracted paths [3].

variable thickness and age exists for most of the year [44, 45, 46, 47] and largely precludes these sources. Instead, the ice cover itself is responsible for much of the ambient noise generation in this region [4]. Environmental forcings such as wind, temperature, ocean current, and air pressure can all prompt noise generation through mechanical deformation in the ice cover [7, 48]. Under a packed or continuous ice cover, thermal cracking due to atmospheric cooling [48, 49] is a main contributor of noise between 100-900 Hz [50]. The spatial distribution of these cracking events is uniform [51] and the overall noise amplitude forms a qausi-Gaussian distribution [48, 52]. The magnitude of ice-generated noise driven by wind also follows a Gaussian distribution between 1-10 kHz [48, 49, 50]. Noise level has been demonstrated to be proportional to average wind speed with an increase of $\sim$0.5 dB/m/s at 250 Hz [53] and an increase of 16 dB per doubling of wind speed within the 3.2-6.4 kHz octave band [49]. At frequencies below 100 Hz, Makris and Dyer [10] found that ambient noise variations over periods of $>$1 hour correlate highly with composite measures

of stress applied to the ice by wind and current, whereas temperature plays a less important role.

Near lead openings and pressure ridges in the ice cover, as well as in marginal ice zones, ice characteristics such as thickness, concentration, and drift also affect ambient noise. Noise level has been shown to correlate positively with ice concentration [54] and ice drift [55]. Spectral properties of noise are also tied to different ice motion mechanisms such as cracking and shearing. For example, large scale lead openings can lead to broadband transients [56] while rubbing and shearing between different ice floes can result in pure and varying tonals [56, 57, 58, 59]. Taken all together, the cacophony of different cryo-acoustic emissions by the ice cover over large areas results in a mostly stationary Gaussian ambient noise soundscape [60].

2.1.1 Modeling of Historical Arctic Ambient Noise

In the historical Arctic, noise generation is generally uniformly distributed throughout the ice cover [51, 56]. Thus, to model Arctic ambient noise, one approach is to assume a uniform distribution of sources near the surface. Such a noise model was first proposed by Kuperman and Ingenito [61] and its mathematical formulation is briefly described here. Fig. 2.2 shows the geometry for this model, where an infinite plane of monopole sources is placed below the surface at depth z'. Each monopole has random source strength $s(\mathbf{r}, t)$, where $\mathbf{r}$ is the radial vector in the source plane and t is the time variable. The source distribution is assumed to be radially homogeneous and isotropic with correlation $N(s)$, where s is the horizontal distance between any two sources. Following the derivation in [3, 62], the acoustic pressure intensity due to this source distribution at any frequency ω is calculated as

$$|p_\omega^2(z)| = 4\pi^2 q^2 \int [F(k_r)|g(k_r, z, z')|^2] k_r dk_r, \tag{2.2}$$

where k_r is the horizontal component of the wavenumber k, $g(k_r, z, z')$ is the depth dependent Green's function satisfying Helmholtz equation for a single monopole source, q is the spectral source level of the monopoles, and $F(k_r)$ is the spatial wave vector

spectrum of the distributed noise source, which is defined as the Fourier transform of the correlation function $N(s)$. Assuming the monopole sources are uncorrelated and their far-field radiation pattern follows $cos(\theta)$, with θ defined from the vertical axis, the correlation function is derived by [63] as

$$N(s) = \frac{2\delta(s)}{k^2(z')s}. \tag{2.3}$$

Taking the Fourier transform of this expression and substituting into Eq. 2.2 yields the noise intensity

$$|p_\omega^2(z)| = \frac{8\pi^2 q^2}{k^2(z')} \int_0^\infty |g(k_r, z, z')|^2 k_r dk_r. \tag{2.4}$$

This expression depends on the unknown source depth z', which is taken to be much smaller compared to the vertical wavelength. To eliminate this dependence, the monopole source strength q is normalized to yield a pressure level Q in an infinitely deep ocean [62] by assigning it the value

$$q^2(z') = \frac{Q^2}{16\pi(z')^2}. \tag{2.5}$$

Given this normalization and a user defined source level Q, Eq. 2.4 can be solved numerically using wavenumber integration techniques. In practice, OASES [35] is used to model noise generation by a uniform distribution of sources in this thesis.

2.2 The SIMI-94 Experiment

As part of a larger Arctic initiative by the Office of Naval Research (ONR), the SIMI-94 experiment was a collective effort between the Massachusetts Institute of Technology (MIT) and the Woods Hole Oceanographic Institution (WHOI) to study acoustic emissions from ice fracturing processes during the spring of 1994. Both a horizontal cross array and a vertical line array (VLA) of hydrophones were deployed in the Beaufort Sea (approximately 400 km north of Prudhoe Bay, Alaska) below packed ice conditions to collect ambient noise data. The exact location of the experiment

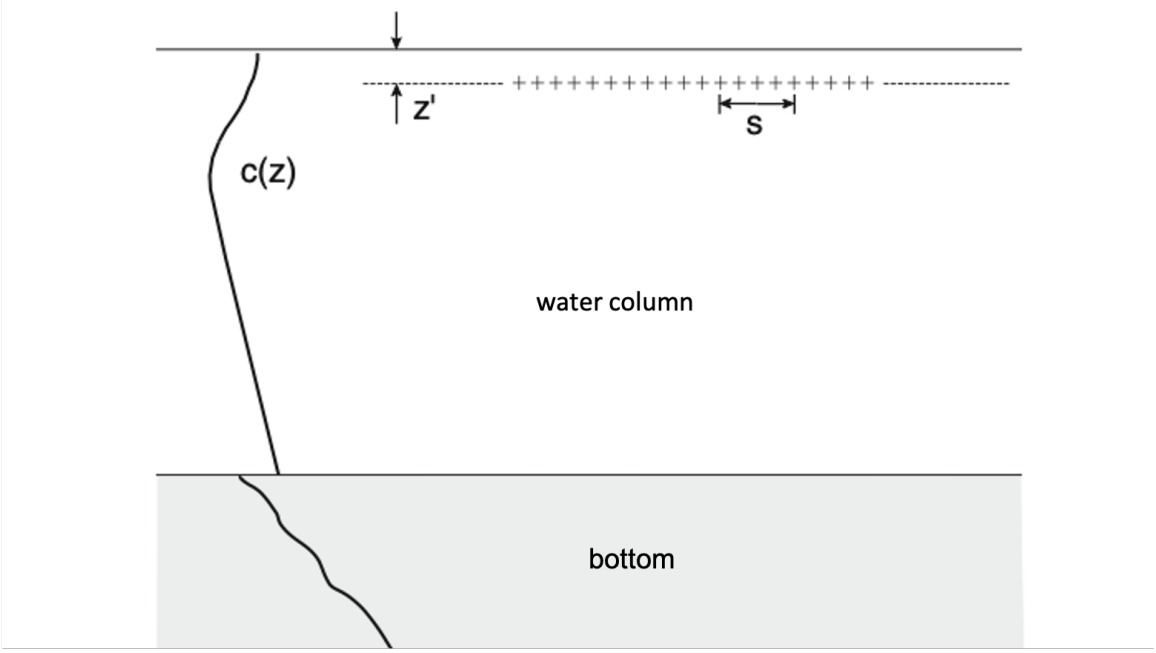


Figure 2.2: Source model for uniformly distributed noise generation [3].

camp site on April 22, 1994 was $73^o00'56''N$ and $149^o53'55''W$ (Fig. 2.3) [11], which is off of the Alaskan continental shelf and in deep ocean of ∼3000 m depth. Results presented in this chapter focus specifically on data recorded on the VLA. This array consist of 32 omni-directional channels with linear 7 m spacing recording continuously at a sampling rate of 1000 Hz. It was deployed in the water from a depth of 62 to 279 m with the center element positioned at 167 m (Fig. 2.4). The sensitivity of the hydrophones was -175 dB re 1V/μPa. The data collected on the array were recorded with a 16-bit amplitude quantization and bandpass filtered between 1 and 350 Hz (24 dB/octave roll-off at low end, 48 dB/octave roll-off at high end) by the acquisition system before being stored on tapes. In total, 35 tapes of data were recorded during the entire experiment; however, the VLA was only active during tapes 23–32 (from April 18–22 UTC). All tapes had a continuous interference that is consistent at 60 Hz, caused by the camp generators. Many tapes also included strong noise interference at 70 Hz, 80 Hz and higher frequencies. For our analysis, 6 hours of data on tape 23 were analyzed because this segment contained lower levels of noise interference compared to the rest of the data. While this shorter dataset does not allow us to correlate

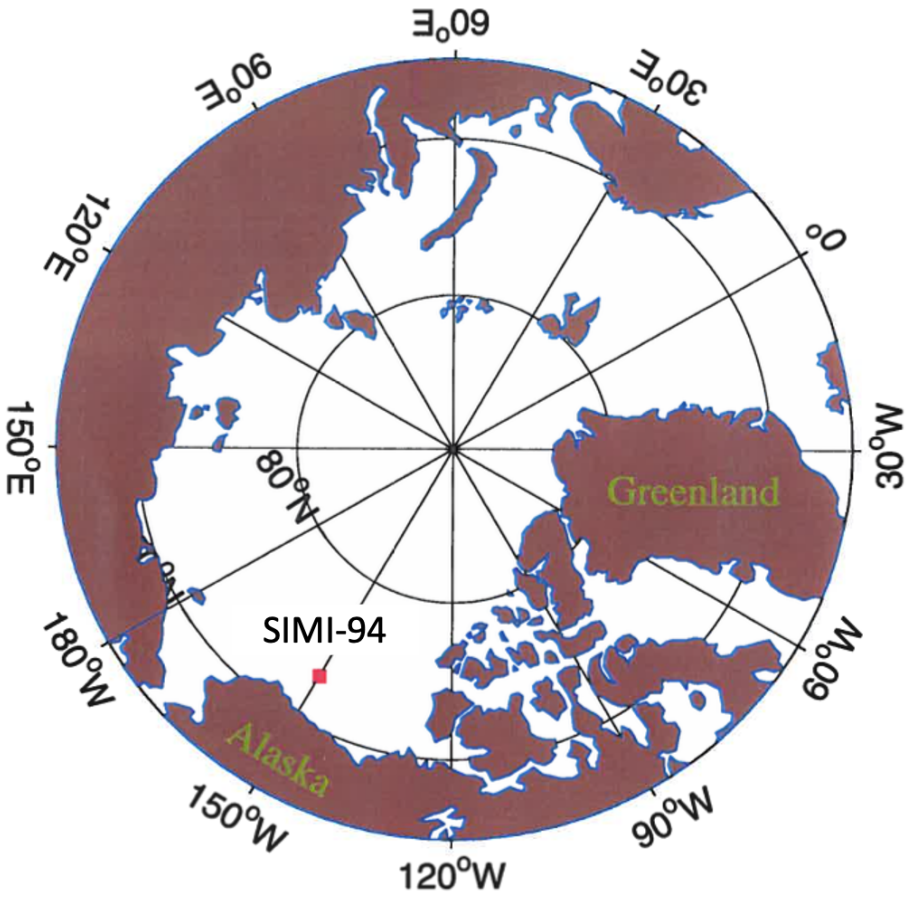


Figure 2.3: Location of the SIMI-94 experiment camp site on April 22, 1994 [11].

the ambient noise to longer time-scale environmental parameters such as daily or seasonal temperature or wind measurements, it is, nonetheless, a useful snapshot of the ambient soundscape that would allow for a valid study of how features of the ambient noise can be connected to environmental characteristics during SIMI-94.

The SSP measured at the experiment site during data collection is shown in Fig. 2.5. The upper 800 m of the profile is plotted based on recorded water temperature data while the values at deeper depths are extrapolated by continuing the profile's slope at 800 m to the depth of the ocean bottom. The overall shape of this SSP matches the description of the historical Arctic profile. The sound speed monotonically increases with depth. The gradient of the increase is steeper above ∼350 m due to the salinity effect mentioned previously, creating a knee in the profile at that

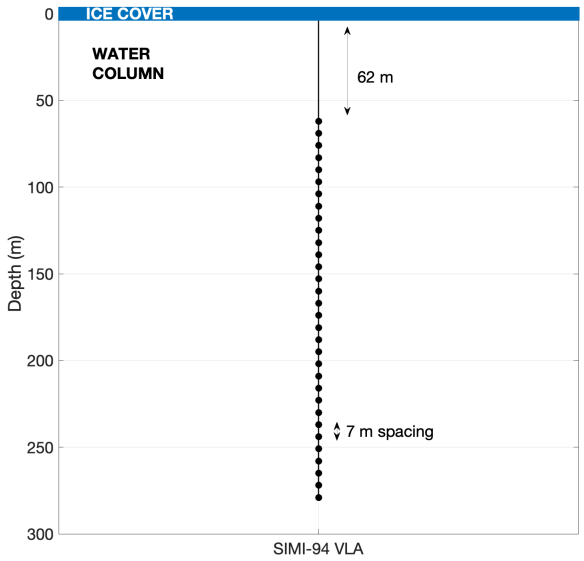


Figure 2.4: Deployment configuration of the SIMI-94 VLA.

depth.

2.3 Overall Ambient Noise Soundscape

In this section, we present the spectral and spatial characteristics of ambient noise collected during SIMI-94. We further compare our measured result with modeling output computed using the historical, uniform source distribution to model ice cover noise generation.

2.3.1 Spectral Analysis

The spectrogram and PSD estimate of the ambient noise data are presented in Fig. 2.6. To generate these plots, the time series data from the top hydrophone in the VLA (62 m depth) are segmented using a 512-point Hanning window with 50% overlap before carrying out a fast Fourier transform (FFT). This produces a frequency resolution of ∼2 Hz. For the PSD estimate, the median of calculated power values at each frequency bin is plotted. The noise spectrogram (Fig. 2.6(a)) shows many

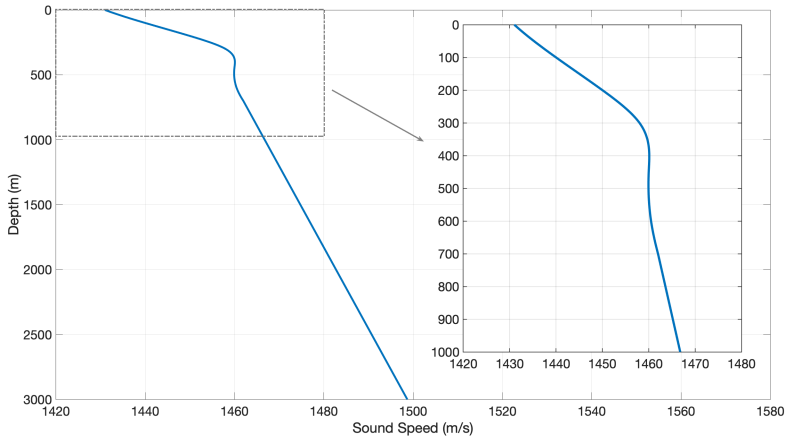
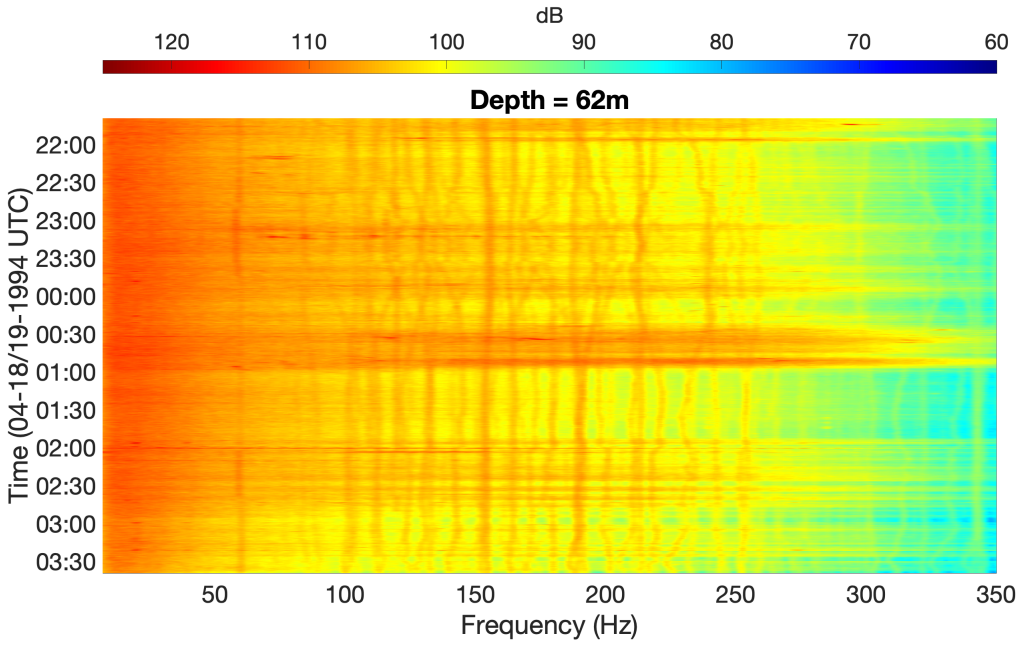


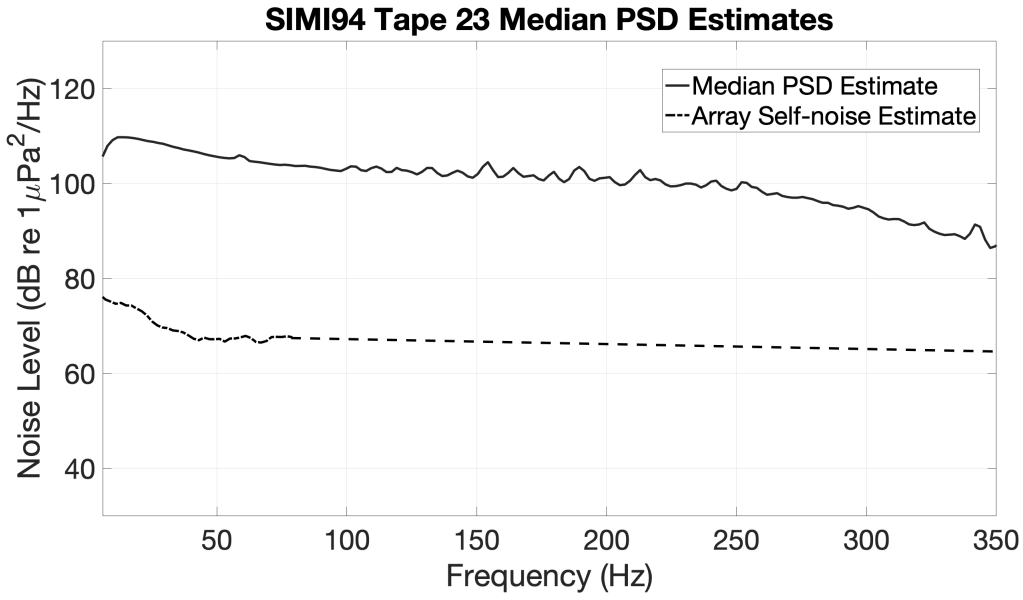
Figure 2.5: SSP during SIMI-94. Its monotonically increasing nature matches the description of the historical Arctic profile.

persistent, high level bands at distinct frequencies that create spikes in the median PSD estimate (Fig. 2.6(b)). The peak at 60 Hz is caused by interference from the camp generators; the cause for the higher frequency peaks is unclear but is likely noise from other man-made sources because of their narrow bandwidth. The absolute noise level remains between 85-110 dB for all frequencies examined. These values are high compared to previously published ambient noise measurements from the Arctic region within this frequency band (Fig. 2.7 [57]) but are similar to the results shown in a prior study of the same SIMI-94 data by Stamoulis [12]. The general shape of the calculated noise profile more closely matches the results from previous works shown in Fig. 2.7. The profile has a peak in noise level near 20 Hz and follows the overall trend that as frequency increases, the noise level decreases.

To further investigate the high noise level, we estimated the self-noise of the SIMI-94 VLA. To do so, conventional beamforming is performed using all VLA elements to produce the wavenumber-frequency (k-f) [64] plot shown in Fig. 2.8. The frequency limit (100 Hz) of this plot corresponds to the spatial aliasing frequency limit constrained by the VLA dimension. The solid lines in the plot denote an apparent propagation speed of 1435 m/s, as calculated through division of the frequency and wavenumber indices along the lines. This speed is equal to the approximate sound



(a)



(b)

Figure 2.6: (a) SIMI-94 tape 23 spectrogram. Persistent bands are visible at distinct frequencies, which are likely caused by anthroprogenic interfering sources due to their narrow bandwidth. For instances, the consistent spectral band at 60 Hz and its harmonics are due to the camp generator. (b) SIMI-94 tape 23 median PSD estimates. Dot-dashed line shows estimated array self noise derived from k–f beamforming; values for self-noise above 80 Hz are extrapolated.

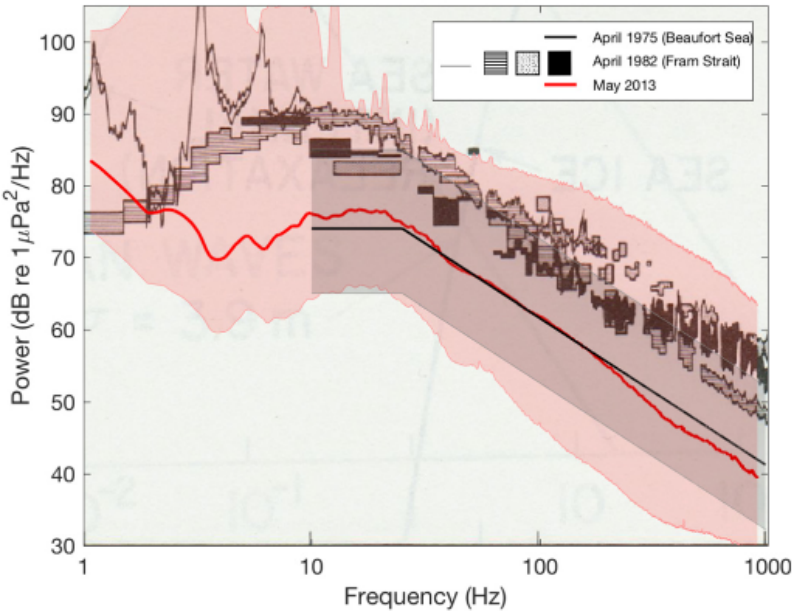


Figure 2.7: Median PSD estimate published in previous Arctic ambient noise studies. The 10% to 90% spectral level interval for May 2013 is shaded. The 5% to 95% spectral levels interval for April 1975 is also shaded [57].

speed in the water column at the depth of the VLA (Fig. 2.5). All signal arrivals within the cone outlined by these lines represent realistic arrivals with apparent propagation speeds greater than 1435 m/s. For example, along the zero wavenumber index, the apparent propagation speed is infinite because this index represents signals that arrived at the VLA horizontally, thus being received by all elements within the array at the exact same time. Similarly, the solid lines represent signals that arrived at the VLA vertically, thus the apparent propagation speed is the same as the actual sound speed. Signal arrivals outside the cone denote apparent propagation speeds that are less than the water sound speed, which is unphysical. Thus, they are likely caused by array self-nose. Using Fig. 2.8, the array self-noise is estimated for frequencies below 80 Hz by calculating the mean noise level outside the cone at each frequency bin. Above 80 Hz, since there is limited to no data outside the cone, self-noise is extrapolated by setting it equal to the value calculated at 80 Hz. The resultant estimate is plotted as the dot-dashed line on Fig. 2.6(b). Unsurprisingly, the array self-noise level is quite high, with values near 75 dB. Thus, this elevated noise floor may explain why

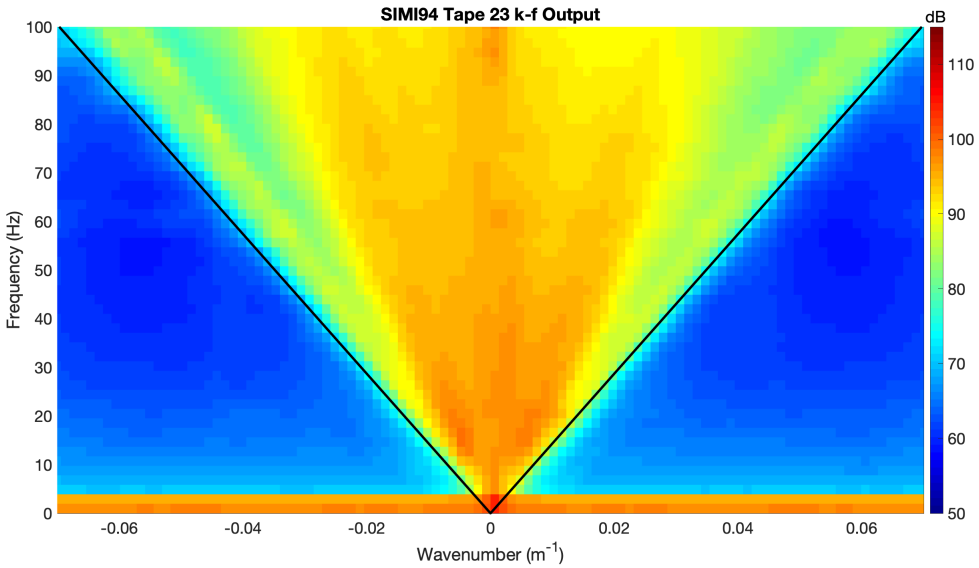


Figure 2.8: SIMI-94 data k–f beamforming output showing unaliased frequencies below 100 Hz. The solid lines denote apparent propagation speed of 1435 m/s.

the measured noise level during SIMI-94 was much louder than results from previous studies.

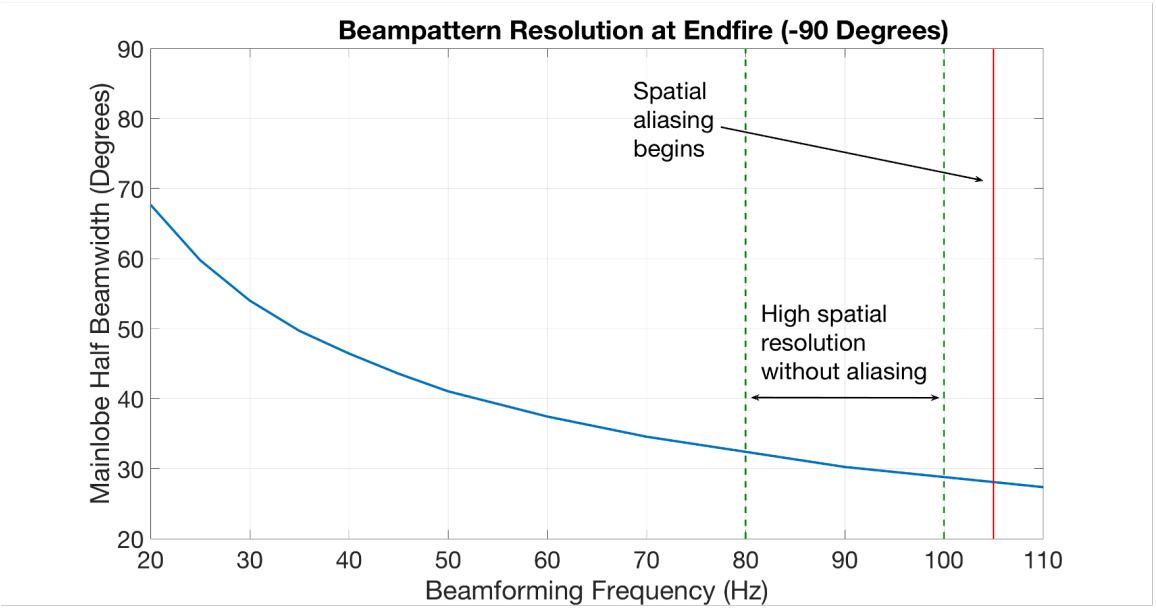
2.3.2 Vertical Directionality Analysis

As a benefit of using an array to collect ambient noise data, conventional beamforming can be applied to analyze the spatial distribution of the received data. Specifically, the VLA allows for insight into the noise vertical directionality. For the SIMI-94 tape 23 data, the noise vertical directionality over time is calculated by beamforming the data between 20-100 Hz. The frequency averaged output in the 80-100 Hz frequency interval (center frequency = 90 Hz) is then computed to generate a plot of noise level at each steering elevation and moment in time (Fig. 2.10). This specific frequency bin is selected to ensure that the beamformed output has good spatial resolution without encountering spatial aliasing based on the VLA geometry. Fig. 2.9(a) demonstrates, for the SIMI-94 VLA, spatial aliasing occurs near 105 Hz when the array is steered towards endfire and the frequencies leading up to the aliasing frequency have the best spatial resolution. At lower frequencies, there is still no spatial aliasing but the

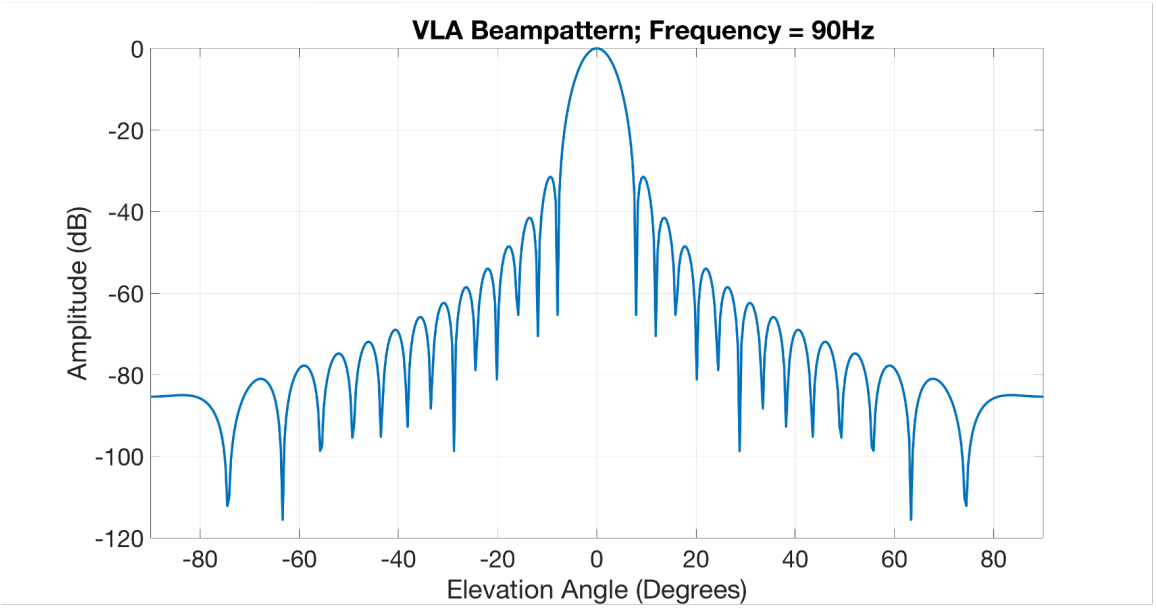
spatial resolution of the output is lower. When steered towards broadside, the array beampattern at 90 Hz has a 3-dB-down beamwidth of ∼3 degrees (Fig. 2.9(b)). Fig. 2.10 shows that during the entire recording period, noise level peaks at 0 degrees elevation angle. This means ambient noise arrival was highest at array broadside (top of the array is +90 degrees, bottom is -90 degrees). To examine noise arrived from broadside more closely, the beamform output at a specific steering elevation angle as a function of frequency and time is determined. Fig. 2.11 shows the noise from broadside contains numerous tonal bands that vary in frequency with time. Because of the meandering pattern of these tonal bands in frequency, they are likely not caused by the camp generator or other anthropogenic sources like the peaks observed in the PSD estimates (Fig. 2.6(a)). Rather, their source is likely the ice cover. Xie and Farmer [58] have noted that intermittent pure tones are generated by shearing and rubbing of adjacent ice masses and that ice of different thicknesses resonate at different frequencies. Since the tones observed here are persistent with time, one possible explanation for them is the ice cover during SIMI-94 experienced constant shear stress as ice masses of variable thicknesses persistently rubbed against each other; furthermore, the ice was thick (up to 8 m [12]) and strong enough during SIMI-94 to maintain this pressure to generate these tonal bands instead of ridging or breaking apart.

2.3.3 Surface Noise Generation Modeling

To better understand why ambient noise during SIMI-94 exhibits its vertical directionality, the time averaged profile at the 80-100 Hz interval (Fig. 2.10 bottom) is compared with modeled noise vertical directionality profiles. The model assumed for ice cover noise generation is the historical, distributed source approach presented in Section 2.1.1. The modeling is carried out using OASES [35]. The modeled environment consists of a 2 m thick surface ice layer on top of a 3000 m deep water column. Below that is a solid bottom halfspace. Physical parameters used to model these layers are presented in Table 2.1. As noted in Section 2.1.1, the only restriction on the placement of the uniform noise distribution below the surface is that its depth



(a)



(b)

Figure 2.9: (a) Spatial aliasing for the SIMI-94 VLA occurs at endfire near 105 Hz. Spatial resolution becomes higher as frequency increases towards the aliasing frequency. (b) The VLA beampattern steered towards broadside at 90 Hz shows 3-dB-down mainlobe width of $\sim$3 degrees.

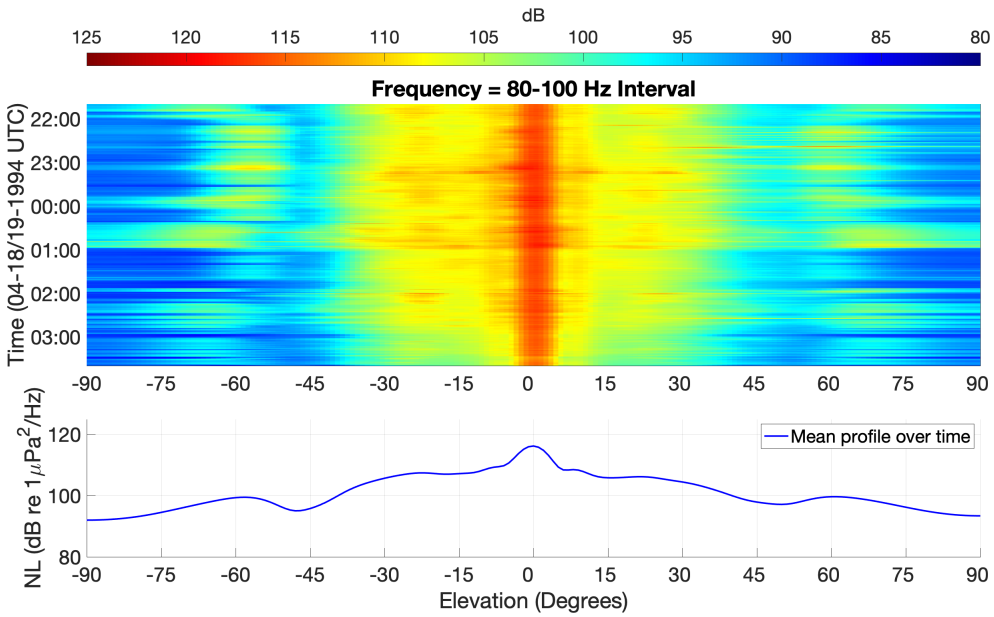


Figure 2.10: Beamform output over time averaged over the 80-100 Hz frequency interval. Noise level peaks near 0 degrees elevation during the entire recording period. Bottom plot shows mean vertical directionality profile generated by averaging top plot in time.

must be small compared to the expected acoustic wavelength. OASES automatically sets this depth as $\lambda_{top}/30$, where λ_{top} is the wavelength calculated at the surface of the water column. We set the modeled source frequency to be 90 Hz, matching that used to generate the measured vertical directionality profile. Given that the surface sound speed in the SIMI-94 SSP is 1435 m/s, the depth of uniform source distribution is calculated to be 0.5 m below the surface. With these settings, the comparison between the measured and modeled noise vertical directionality profiles is shown in Fig. 2.12. The noise level of the profiles have been normalized since we are not concerned about absolute levels, only the location of their peaks. Based on this metric, the two profiles agree well in that both peak near the horizontal at 0 degrees elevation. Besides the location of the peaks, the two profiles do differ in their noise levels at certain elevation angles. In particular, the modeled profile has a broader peak near 0 degrees elevation and a steeper decrease in noise level between $\pm$15-30 degrees compared to the measured profile. These differences may possibly be

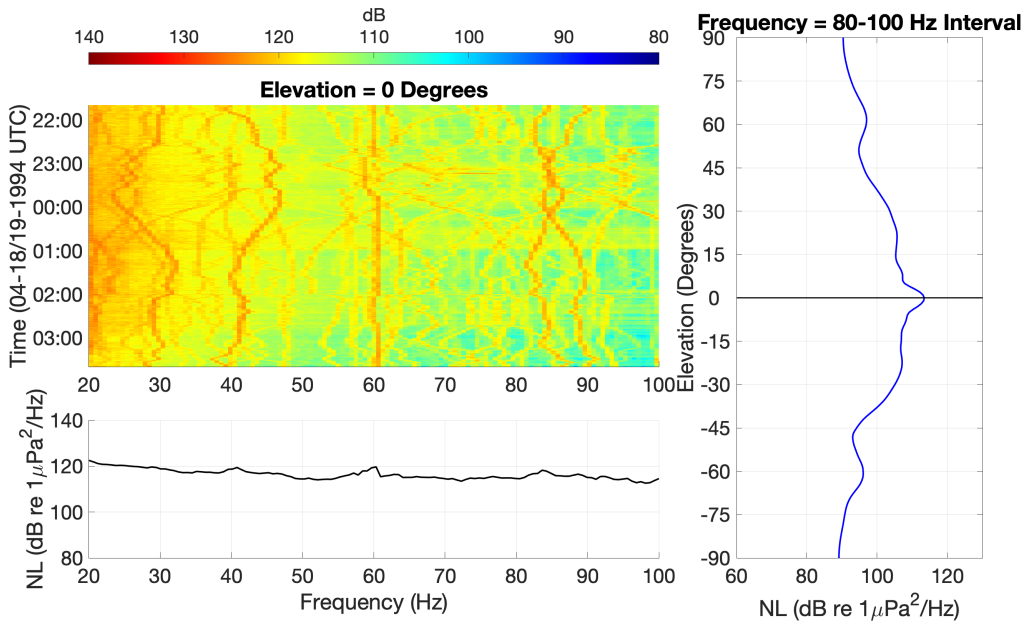


Figure 2.11: Beamform output at 0 degree elevation angle as a function of frequency and time. Meandering tones that vary with frequency and persist with time are observed. The non-frequency varying, persistent tone at 60 Hz is due to camp generator noise. Bottom plot shows mean noise level profile generated by averaging top plot in time. Right plot shows mean vertical directionality profile over time in the 80-100 Hz frequency interval.

Table 2.1: Parameters used to model the SIMI-94 environment. C_p denotes compressional speed, C_s denotes shear speed, ρ denotes density, λ denotes spatial wavelength, h_{RMS} denotes root-mean-square roughness, η denotes roughness correlation length, D denotes layer thickness. Parameter values are selected based on discussions in [3, 65].

Layer	**Parameters**
Ice	$C_p = 3600$ m/s, $C_s = 1800$ m/s $\rho = 0.9\ g/cm^3$, $D = 2$ m $h_{RMS} = 0.2$ m, $\eta = 20$ m
Water column	$\rho = 1.0\ g/cm^3$, $D = 3000$ m C_p = SIMI-94 SSP (Fig. 2.5)
Bottom halfspace	$C_p = 2200$ m/s, $C_s = 1500$ m/s $\rho = 2.9\ g/cm^3$

attributed to slight mismatches between the real and modeled environments relating to the SSP, the ice cover, the VLA, and even the the surface noise distribution - it is unrealistic to expect that our model is a perfect replica of the actual environment during SIMI-94. Nonetheless, the modeled profile appears to capture the essence the measured profile in regards to its peak location and general shape. This agreement suggests that our modeling parameters, which includes the historical, uniform surface source distribution, is a close, albeit not perfect, description of the environment during the SIMI-94 experiment. Thus, ice cover noise generation during this experiment can still be suitably described by the traditional, uniform distribution model.

2.4 Transient Event Characteristics

In addition to spectral and spatial noise analysis, detection and characterization of transient ambient noise events in the data time series can provide useful information on the temporal distribution of ice cover activity. One of the earliest studies of transient ambient noise events is conducted by Manning, which provided insight into the temporal distribution of noise events below 80 Hz in the central Arctic [9]. Greening *et al.* [51] and Zakarauskas *et al.* [66] conducted similar studies in shallow Arctic

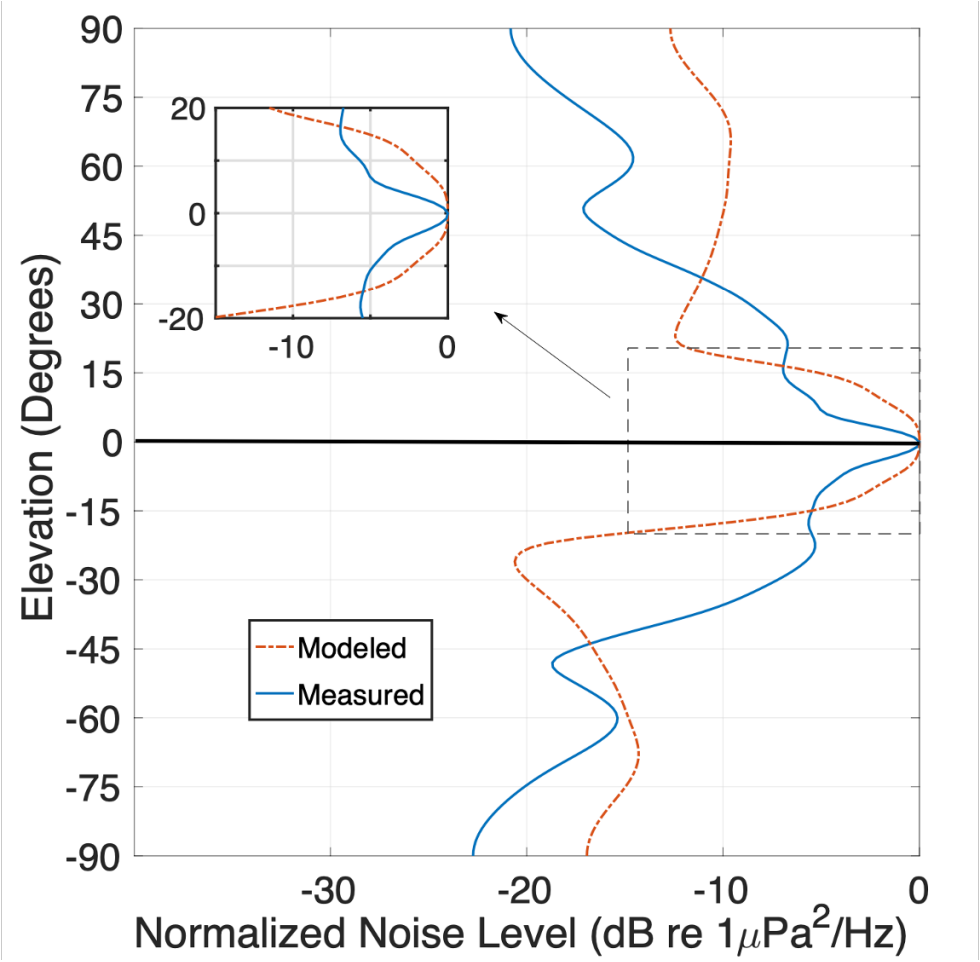


Figure 2.12: Comparison of measured SIMI-94 noise vertical directionality profile with modeling result assuming uniform surface noise generation. The two profiles agree well in that both have peaks near 0 degrees elevation angle, as marked by the solid black line.

environments. They extended the analysis to higher frequencies and focused more exclusively on the spatial distribution and directionality of the detected events. Stamoulis [12] also studied acoustic transients recorded during the SIMI-94 experiment by characterizing the motion of their generation mechanism in the ice cover. Relating to our study, this work noted the temporal grouping of transient events where multiple arrivals may exist in a single event and multiple events may be grouped to belong within a combined cluster. In this section, we provide an overview of an amplitude-based event detection method developed based on an approach by Zakarauskas *et al* [66, 67]. We then present results relating to temporal statistical characteristics of transients detected in the SIMI-94 tape 23 ambient noise data.

2.4.1 Amplitude-based Event Detection Method

An amplitude-based transient event detection algorithm is used to find transient acoustic emissions in the ambient noise data. This method requires three parameters (w, r, p_{thres}) and searches for improbable clusters of high amplitude peaks within the data time series. Briefly, it works as follows:

1. Given the input time series $\phi(t)$, form a histogram of its absolute peak amplitude $\phi_{peaks}(t)$, where $\phi_{peaks}(t)$ is calculated by applying the MATLAB function *findpeaks* to $|\phi(t)|$.

2. Set the amplitude value exceeded by $r\%$ of the histogram as a threshold value, m.

3. Partition the time series into adjoining segments of w peaks; within each segment, count the number of peaks, N, whose amplitude exceed m.

4. Calculate the probability of having at least N peaks from the top $r\%$ of the histogram within a window of w peaks as a sum of binomial distributions, i.e. $P = \sum_{i=N}^{w} \binom{w}{i} r^i (1-r)^{(w-i)}$.

5. If P is less than the chosen threshold probability p_{thres}, the window contains an event and the peaks within the window that have amplitude greater than m

are identified.

6. Consecutive peaks that are less than 0.05 seconds apart are grouped into a single composite event. This threshold is set based on the description of transient event types in previous studies by Chen [68] and Stamoulis [12], where different arrivals within the same event are typically less than 0.05 seconds apart.

The algorithm parameters used to analyze the SIMI-94 data are $w = 10$, $r = 1\%$, and $p_{thres} = 0.00001$. $w = 10$ is chosen because an impulsive transient event contains, typically, less than 10-20 peaks [12, 68]. If w is too large, shorter events may be missed because the likelihood of having a few large peaks within a long window may still be fairly high. Thus, to be conservative, a small value of 10 is chosen. However, it is also unlikely for a transient event to have only 1 or 2 peaks and ideally, p_{thres} should be small so that the number of false alarms is kept to a minimum; $p_{thres} = 0.00001$ seems to be a reasonable threshold through some empirical testing. r is chosen in complement to the other two parameters so that, as shown in Table 2.2, as N increases to more than 2 peaks within a window of 10, the window would be selected to contain a transient. As another check for false alarms, this method is applied to data on all 32 channels and detected transients are only confirmed to be events if they appear on at least 3/4 of the channels. After event selection, the beginning and end times of each event are documented.

Table 2.2: Probability of N or more peaks with amplitude greater than the top $r\%$ of the PDF in a sample of 10 peaks.

	r = 10%	r = 5%	r = 1%	r = 0.5%	r = 0.1%
N = 1	0.264	0.086	0.004	0.001	4.476×10^{-5}
N = 2	0.070	0.012	1.138×10^{-4}	1.461×10^{-5}	1.194×10^{-7}
N = 3	0.013	0.001	2.001×10^{-6}	1.281×10^{-7}	2.089×10^{-10}

Since ambient noise level generally increases towards lower frequencies, the data must be bandpass filtered into separate frequency bands before being processed by

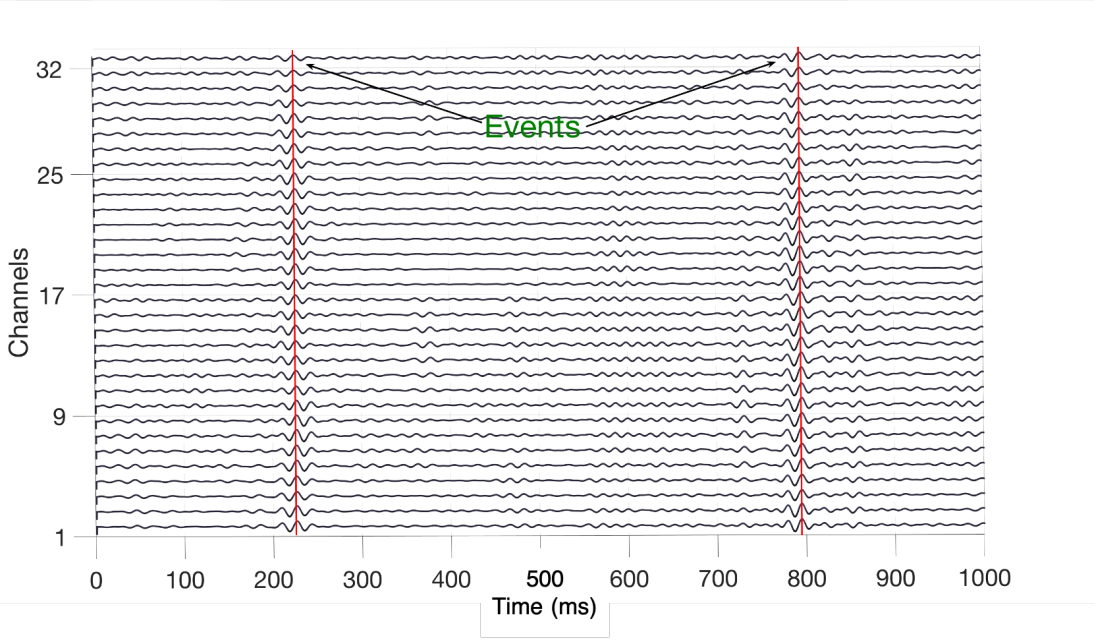


Figure 2.13: Example of two transients detected in the data time series in the 40-80 Hz interval.

the event detector or higher frequency transients may be missed [66, 67]. Accordingly, the time series are filtered into 3 octave bands between 40-320 Hz (40-80 Hz, 80-160 Hz, 160-320 Hz). The narrow-band interferences observed in Fig. 2.6(a) should not affect the output of the event detector because they are persistent with time and the event selector is focused on large amplitude transient events. An example of transient events detected in the 40-80 Hz interval is shown in Fig. 2.13.

Before we present the results from applying the detection algorithm, it is important to discuss a limitation of this approach. As with all datasets without ground truth information regarding the presence of transients events, it is extremely difficult to quantify the performance of this algorithm applied to the SIMI-94 data. It should be expected that, with any automated detection system, there will be missed events as well as false identifications. We specifically chose this algorithm because it has been applied previously in published literature [66, 67] and its probabilistic approach provided more confidence that it would detect true transients containing a series of high amplitude peaks while ignore random noise that may only contain a single peak. We further selected the parameters to be more conservative with detection, further

reducing the possibility of false identifications. There are certainly faint events that are missed by the algorithm, however, we are fairly confident that the algorithm would detect most of the stronger transient events that are inherently a more significant component of the ambient soundscape. This first limitation highlights the difficulty of transient characterization for ambient noise in general . Even if detection is done by trained human expert (which is incredibly tedious for any substantial dataset), there would still be some unknown uncertainty regarding the accuracy of the detection process given that there is no ground truth for ambient noise transient events. Thus, it is perhaps more apt to view our temporal analysis of the SIMI-94 data in this case (and in any future applications) as not a means to identify all true transient events within the ambient noise, but instead, as a way to tabulate high amplitude features in the noise that are representative of transient events and provide a description for their temporal distribution. This is still a useful and informative result as, regardless of their origin, these transients are, nonetheless, part of the ambient soundscape.

2.4.2 Temporal Distribution Analysis

Table 2.3 tallies the number of transient events detected in the dataset in each octave band. The number of transients increases with frequency. This result suggests as frequency increases, higher amplitude transient noise events contribute more and more to the total ambient noise soundscape. This finding is in agreement with the result of a previous study by Zakarauskas in a shallower Arctic environment [66].

Table 2.3: Number of detected transient events in each frequency band.

40-80 Hz	80-160 Hz	160-320 Hz
173	451	1043

Post event detection, the start and end times of the detected transients are recorded to gain a better understanding of their temporal density and distribution. These times are calculated by subtracting the end time of an event from the start

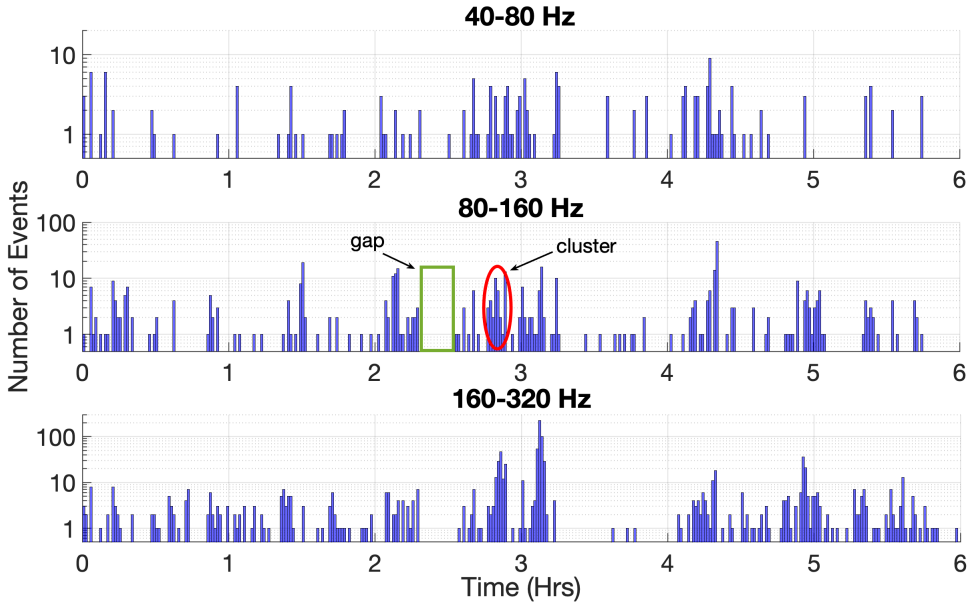


Figure 2.14: Temporal distribution of transient events during the SIMI-94 experiment at three octave bands. Examples of a clustering of events and a time gap that separate the clusters are shown.

time of the next. The only known previous work to have characterized inter-arrival times between transient events is by Manning [9]. In that study, a J-shaped gamma distribution is found to best describe the statistical spread of inter-arrival times if very large outliers (>280 s) are excluded. However, it may be natural to have very large inter-arrival times if the weather conditions are calm and the ice cover experiences low environmental forcing. A more natural way to characterize inter-arrival times may be to view the occurrence of transient events as a clustering process. Within a cluster, the inter-arrival times are short but large gaps may separate consecutive clusters. To visualize and verify this clustering process, the data time series are segmented into 1 minute bins and the number of events detected within each bin is counted. The results are presented in Fig. 2.14. This plot appears to confirm this hypothesis. Transient events do seem to occur in clusters that are separated by gaps, with a cluster defined as any period of time during which at least 1 event occurs in each 1 minute bin and a gap defined as any period of time during which no events occur in each 1 minute bin.

Table 2.4: Time gap between successive event clusters (min)

	40-80 Hz	80-160 Hz	160-320 Hz
75th Percentile	9	4.5	3
Median	3	2	2
25th Percentile	2	1	1

Table 2.5: Time length of event clusters (min)

	40-80 Hz	80-160 Hz	160-320 Hz
75th Percentile	2	2	3
Median	1	1	2
25th Percentile	1	1	1

Table 2.6: Number of events in each event cluster

	40-80 Hz	80-160 Hz	160-320 Hz
75th Percentile	5	7	10
Median	3	2	3
25th Percentile	1	1	2

To better quantify the clustering characteristics of transient activity during SIMI-94 within each frequency octave, the statistical median and quartile values of the event gap and cluster lengths, as well as the number of events within each cluster are presented in Tables 2.4, 2.5, and 2.6, respectively. The time gap between successive event clusters decreases with frequency. Table 2.4 shows that the 75th percentile value for the 40-80 Hz interval is 9 min and this value decreases to 3 min for the 160-320 Hz interval. This result is in agreement with the trend that there is more transient activity at higher frequencies. The lower gap length means that more event clusters are occurring at higher frequencies. However, while the number of clusters increases with frequencies, Table 2.5 shows the duration of clusters are fairly consistent over all frequency intervals - most event clusters last less than $\sim$3 minutes regardless of frequency. Lastly, Table 2.6 shows while the median number of events within any cluster is 2-3 for all frequency intervals, higher frequency clusters have a higher ceiling for the number of events within any single cluster. This is demonstrated by the increase in the 75th percentile value with frequency interval.

2.5 Summary

In this chapter, we examined noise data collected during the SIMI-94 experiment as an example of the ambient soundscape in the historical Arctic environment. Such an environment has a monotonically increasing SSP and noise generation from its ice cover can be modeled by a uniform distribution of sources near the surface. The spectral shape of ambient noise collected during SIMI-94 is similar to results from previous studies. Noise level generally decreases with increasing frequency with a peak near 20 Hz. The absolute noise level during SIMI-94, however, was higher than previous Arctic recordings. This discrepancy may be caused by higher array self-noise of the SIMI-94 VLA. The vertical directionality of the recorded noise shows a peak in noise level arriving broadside to the VLA (0 degrees elevation angle). This result matches well with modeling output assuming uniform noise generation by the ice cover. Lastly, an amplitude-based transient event detection algorithm was applied

to study the temporal distribution of transient noise events during SIMI-94. This analysis demonstrated that transient occurrence during SIMI-94 may be described as a clustering process, where events occur in quick succession followed by a gap during which there are no transients. Furthermore, more events were detected at higher frequency octave bands. This result indicates that transient noise accounts for a larger portion of the total acoustic soundscape at higher frequencies. A drawback of this event detection method is the lack of visualization of the detected events in the frequency domain. Performing detections in separate frequency octaves also fragment broadband transients and does not provide a complete picture for these events. Furthermore, the spectral shapes of the detected transients would provide insight into their generation mechanisms in the ice cover and can also be used as a means to further eliminate false identifications (*e.g.*, a very narrowband, short duration detection is likely not a true transient event). Motivated by these shortcomings, we propose an alternative, frequency domain event detection algorithm in Chapter 5 of this thesis. Further work regarding this algorithm and temporal analysis of transients in general also include comparing the occurrence of the detected transient against statistical models such as one that assumes a Poisson distribution (as an example) for event occurrence. Such comparisons would provide insight into whether the transients' temporal characteristics follow any known statistical distributions.

Meanwhile, in the following chapter, we present noise analysis results from a more recent Arctic experiment in comparison to SIMI-94 observations to demonstrate the effects on the ambient soundscape due to environmental changes. These changes relate to the two main components that dictate acoustics propagation introduced in this chapters, which are the SSP and ice cover noise generation.

Chapter 3

Environmental effects on Beaufort Sea Ambient Noise

The historical Arctic environment outlined in Chapter 2 has been an apt description for much of the region in the past. However, the Arctic ocean is dynamic and has been undergoing significant changes in recent years. In this chapter, we introduce observed shifts to the underwater SSP and the surface ice cover and discuss their effects on the ambient soundscape. We focus our presentation on the Beaufort Sea, where the SIMI-94 experiment was conducted and where a more recent ambient noise dataset was collected during the ICEX-16 experiment. We again leverage noise modeling to explain how features observed in the measured noise data connect to environmental changes in the region.

3.1 Beaufort Sea Environmental Changes

3.1.1 Beaufort Lens SSP

In the Beaufort Sea, an influx of warm Pacific water entering the region from the Bering Strait, combined with subduction of solar-heated water from the Chukchi Sea during the summer months, has dramatically altered the underwater SSP by increasing water temperature between ∼40-80 m depth [69, 70, 71]. While this circulation

phenomenon, named the "Beaufort Lens", has been noted since the 1970s [72], more recent measurements by the WHOI Ice-Tethered Profiler (ITP) program[1] [73, 74] found the Lens' intensity has increased and its geographical extent has spread. The ITP program consists of a sparse network of oceanographic profilers located throughout the Arctic ocean. Each ITP is deployed by securing it within the surface ice cover; it then travels with the ice motion and collects data in the water column below along a wire rope suspended through the ice to a depth of up to 800 m. A conductivity, temperature, depth (CTD) sensor travels along the wire to measure the water salinity, temperature, and pressure (depth). These values can then be used to calculate the SSP. Each ITP performs 2-6 traverses along its wire per day and the data collected is transmitted to a shore-based data server using an Iridium transmitter in the above-ice portion of the unit (Fig. 3.1). From data collected using the ITP program, the presence of the Beaufort Lens has been observed in much of the Canadian Basin and water temperature at the Lens' depth can exceed that just outside its depth interval by $> 1^{o}$C (Fig. 3.2 [69]).

As the result of the Beaufort Lens, the historical, monotonically increasing Arctic SSP is disrupted. Due to the increase in temperature, a sound speed maximum now occurs at the Lens' depth (Fig. 3.3(b)). This change affects acoustic propagation in the region by creating a double duct environment - one at the surface as usual and another just below the sound speed local maximum. In the surface channel, acoustic waves encounter frequent interactions with the ice-water interface, resulting in severe attenuation. In contrast, the lower duct has been shown to promote long range propagation (Fig. 3.3) by effectively trapping sound above 300 Hz [13, 14, 75, 76, 77]. The effect of this ducted environment on acoustic propagation will likely increase in significance in the future. For instance, with the modeled rise in Arctic Ocean pH, sound is predicted to propagate even further within the lower duct due to decreased attenuation [78]. Thus, continued study of this phenomenon and its effects are of critical importance to improving our ability to characterize the Arctic Ocean

[1]The ITP program is deployed throughout the Arctic Ocean and the Beaufort Sea is just part of its coverage area.

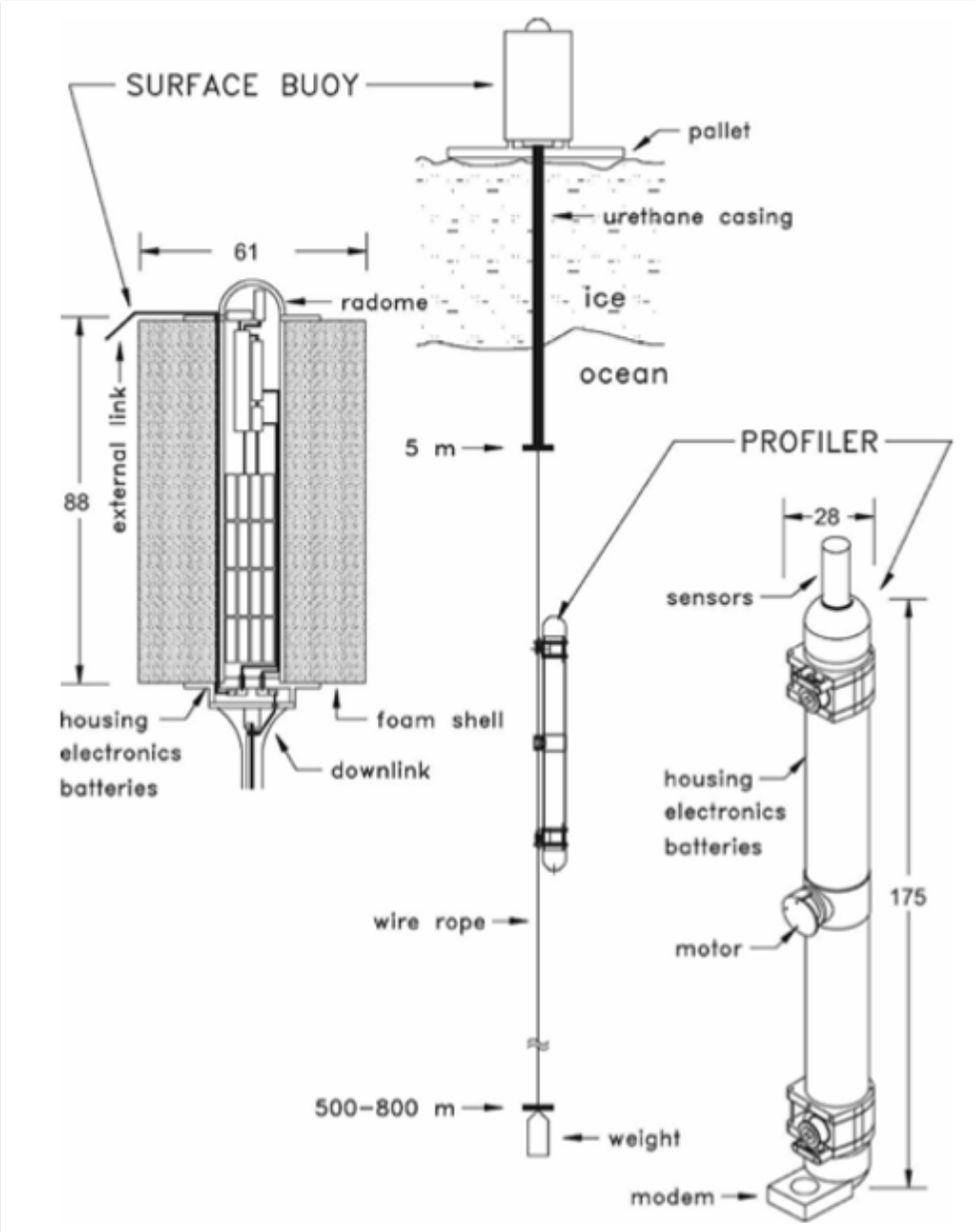


Figure 3.1: Schematic of an Ice-Tethered Profiler. Dimensions are presented in cm, except where specified. [74]

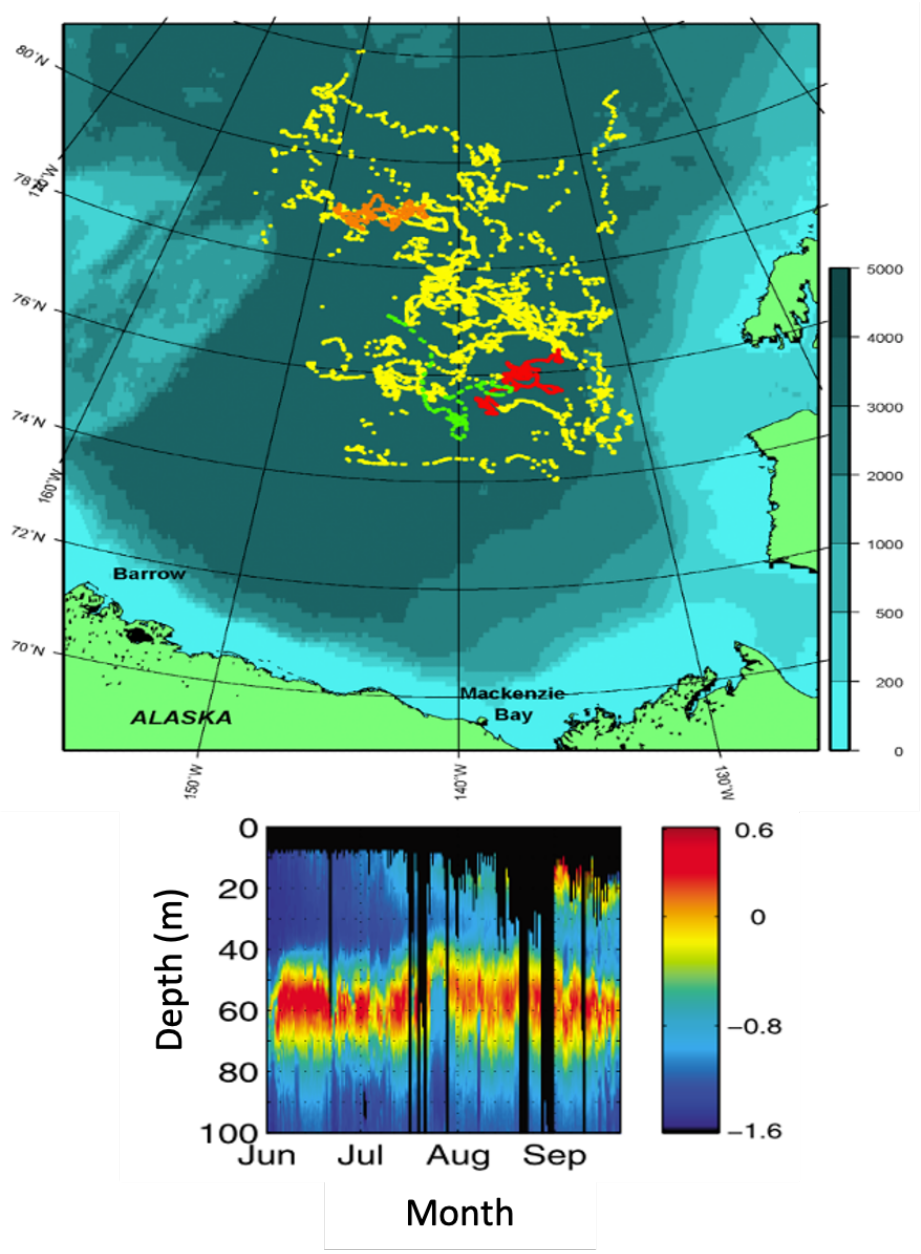


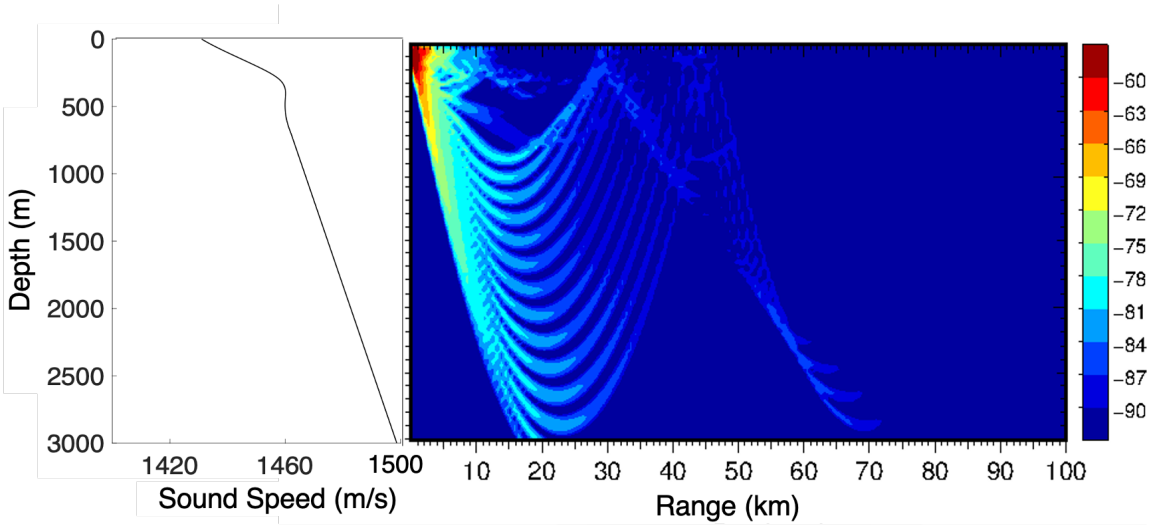
Figure 3.2: (Top) Spatial distribution of temperature and salinity profiles acquired by ITPs from August 2004 to September 2009 analyzed by Toole *et al.*. Color shading indicates the depth in meters. Special attention is given to profiles from ITP 6 acquired in summer 2007 (green), ITP 3 from winter 2005–2006 (red), and ITP 4 from winter 2006–2007 (orange). (Bottom) Depth-time contour plot of the observed upper ocean temperature (^{o}C) for the summer of 2007 as recorded on ITP 6. [69]

soundscape and successfully conduct underwater acoustic operations in the region.

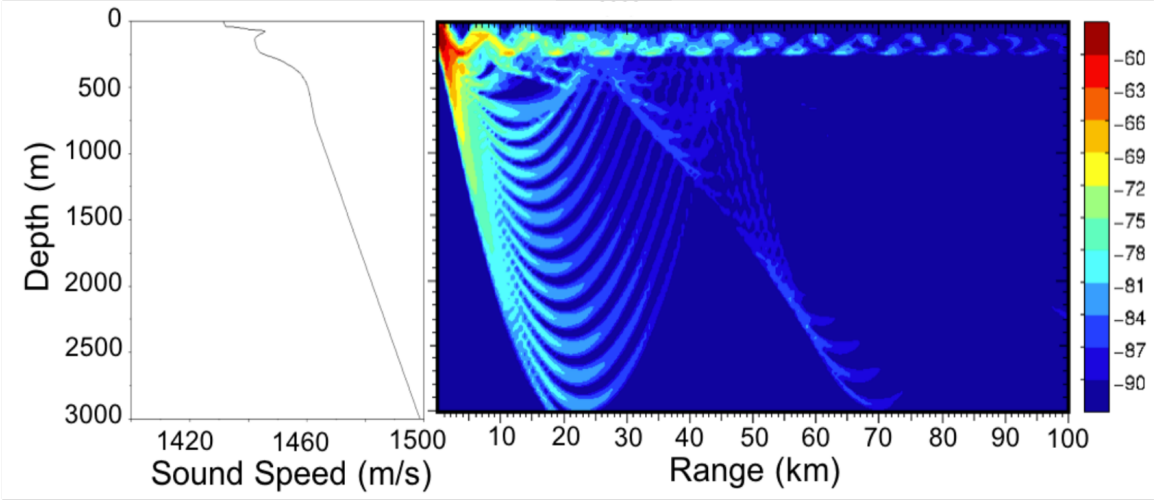
In addition to the ITP data presented by Toole *et al.* [69], we also conduct our own analysis of ITP data collected in the Beaufort Sea in more recent years from January 2014 to January 2020. The positions of 2331 SSP measurements we examine are shown in Fig. 3.4(a). They are all located within the Beaufort Sea region of the Arctic and their collection year is denoted by different markers. Most measurements were collected in 2014 (1756) while 2018 and 2020 have the least with 20 and 26 measurements, respectively. Fig. 3.4(b) shows the yearly mean of the measured profiles, as well as the sound speed standard deviation with depth. Similar to Toole *et al.* [69], we observe a local sound speed maximum near $\sim$70 m depth caused by the Beaufort Lens. Furthermore, in most years, the largest temporal variability in the SSP occurs around the depth of the local sound speed maximum and just below the lower duct. This is evident by the slight increase in sound speed standard deviation near those depths. While the SSP typically has a single, sharp local maximum, it is also possible for the profile to exhibit a broader peak or even multiple peaks. These features can be seen in the more recent profile averages from 2018 to 2020 (Fig. 3.4(b)). In fact, the mean of the 26 measurements collected in January 2020 show two distinct peaks instead of just one. The observation of these new SSP features demonstrates the dynamic nature of the Beaufort Sea. The shifting underwater SSP invites continued monitoring to ensure we can accurately describe the region's undersea environment in the future.

3.1.2 Changes to Surface Ice Cover

Over the past 100 years, the global average surface temperature has been steadily increasing [2] with the most significant increase occurring since 1976 [1, 2]. More significantly, the Arctic region has experienced more dramatic warming than any other region on Earth [2]. In the historical Arctic, ice cover exists in most parts throughout the year [79]. However, with the region's continued warming, climate models project that the ocean could be completely ice-free during the summer within the century [47, 80]. Indeed, a general trend in the Arctic ocean has been the decline in

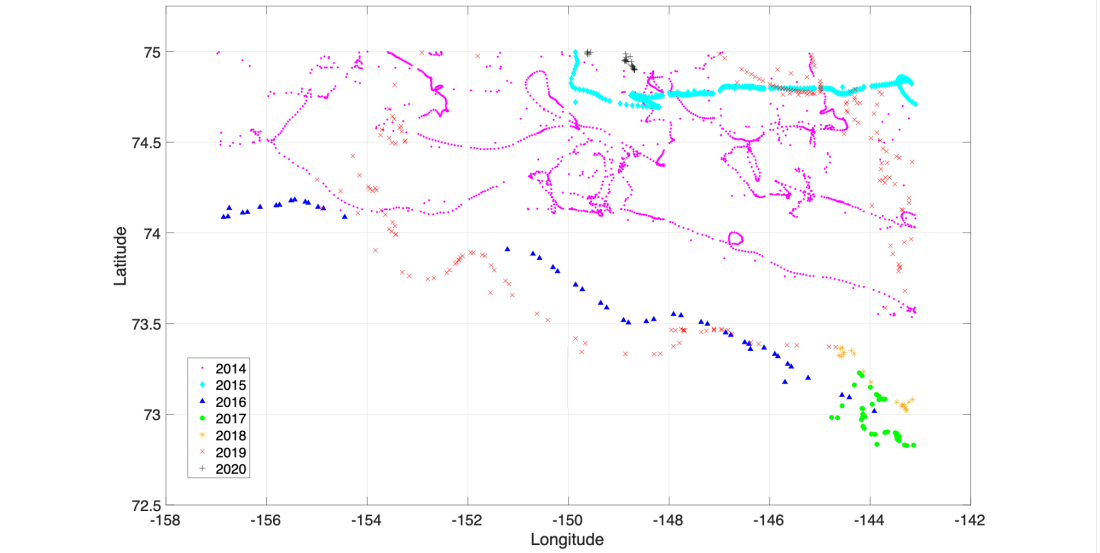


(a)

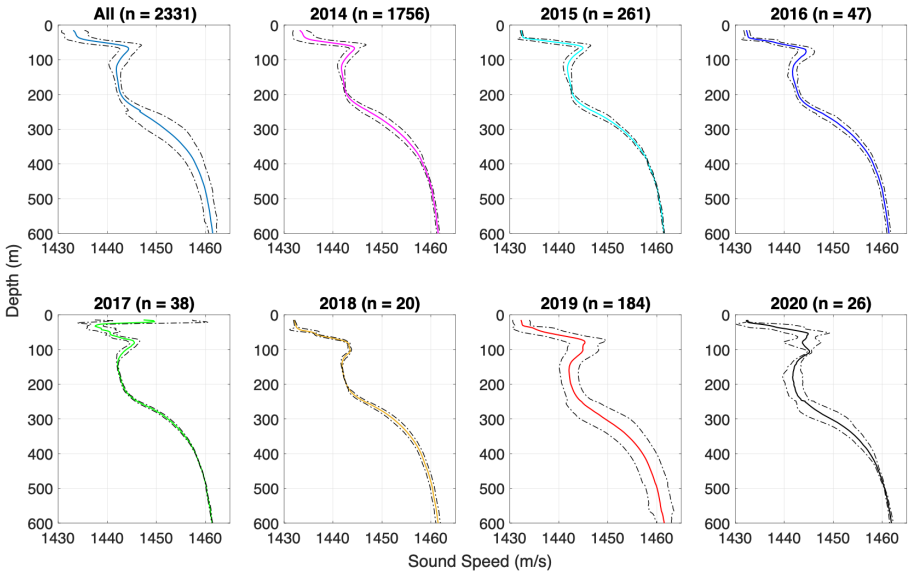


(b)

Figure 3.3: (a) Representative SSP of the historic Arctic profile. Corresponding transmission loss (dB) plot with a 900 Hz source at 100 m depth. (b) Representative SSP demonstrating the effect of the Beaufort Lens. The increase in temperature at the Lens' depth (∼40-80 m) creates a local sound speed maximum. Corresponding transmission loss (dB) plot with a 900 Hz source at 100 m depth showing the ability of the lower sound duct to trap sound and allow propagation to long ranges. [13]



(a)



(b)

Figure 3.4: (a) Locations of measured SSPs as part of the ITP program from 2014 to early 2020 in the Beaufort Sea. (b) Mean (solid lines) and ±2 standard deviations (dashed-dotted lines) of the measured profiles. The first panel shows all profiles while the rest show profiles collected in a single year. The n value in each panel title denotes the number of profiles that were collected during that year.

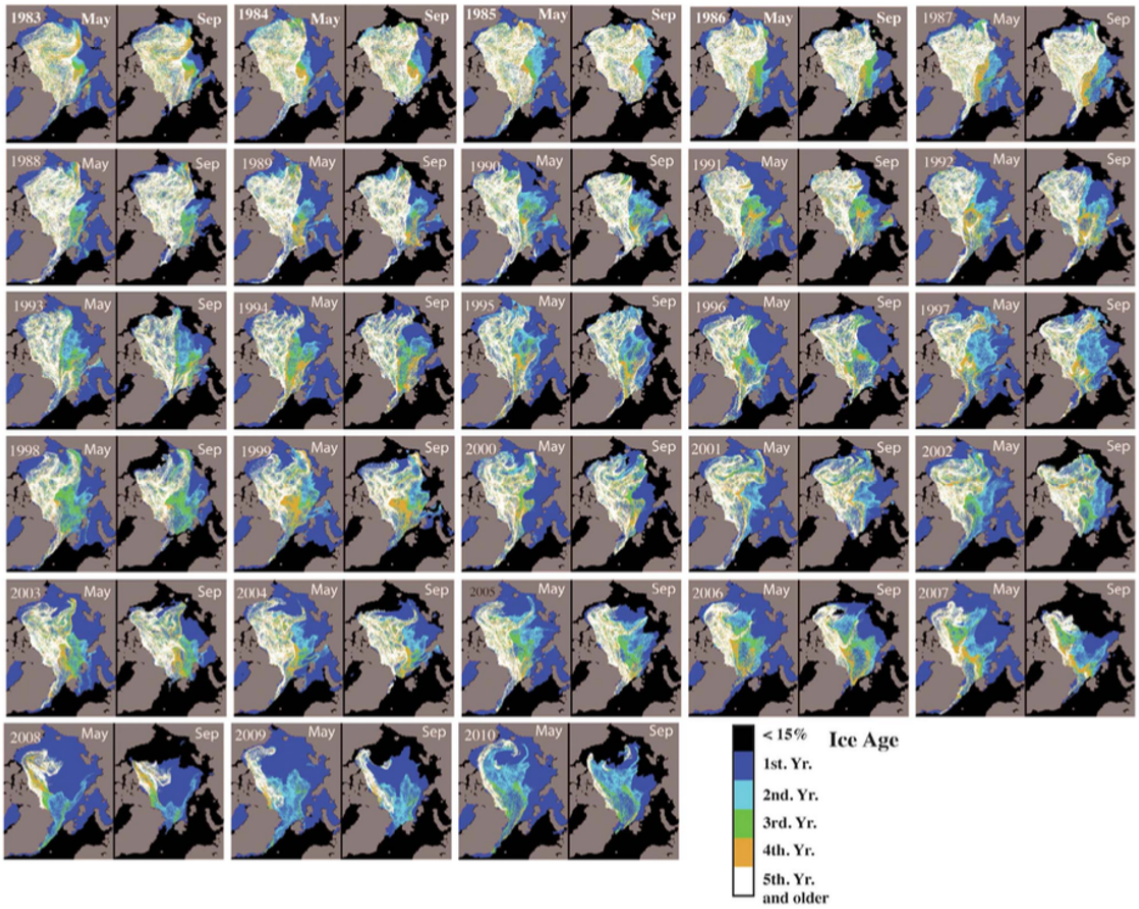


Figure 3.5: Ice cover age during May and September of 1983-2010 showing the decline in multi-year ice over time. [45]

ice cover thickness and spatial extent. The mean winter thickness of central Arctic ice cover declined at more than 10 cm/year between 1993-2007 [81] and Arctic ice extent has also steadily declined since 1979 [82, 83, 84]. More specifically, the percentage of the ice cover that consists of multi-year ice, or ice that has existed for longer than 1 year, has dramatically reduced (Fig. 3.5) [45, 85]. Instead, more fragile first-year ice is more prevalent.

The effect that thinner and younger ice cover has on ambient noise generation has not been extensively studied. However, a hypothesis is that noise may be produced at more localized positions in the ice cover such as along ridge formations. Such a result would be a change from the uniform surface noise generation assumption of the

historical Arctic. As part of our analysis in this chapter, we examine whether this hypothesis fits with observed data.

3.2 The ICEX-16 Experiment

In March of 2016, the MIT Laboratory for Autonomous Marine Sensing Systems (LAMSS), Applied Physical Sciences Corporation, Bluefin Robotics, and GobySoft jointly participated in the U.S. Navy's 2016 Ice Exercise (ICEX-16) in the Beaufort Sea region of the Arctic ocean. Coincidentally, the location of the ICEX-16 camp site was nearly identical to that of the SIMI-94 experiment more than two decades earlier (Fig. 3.6). One scientific goal of ICEX-16 was to characterize the effect that environmental changes, such as that occurring in the SSP and ice cover, have on the region's underwater acoustic environment. Continuous ambient noise data were recorded with a 32 VLA tethered below a suspended Bluefin-21 autonomous underwater vehicle (AUV) under packed ice conditions. The data sampling rate was 12000 Hz and a 16-bit amplitude quantization was used by the recording system. The VLA had omni-directional hydrophones with nested spacings of 1.5 and 0.75 m and its center was 38 m below the AUV during deployment (Fig. 3.7). The sensitivity of the hydrophones was -176 dB re 1v/μPa [86]. Two datasets were gathered during the experiment. First, approximately 8 hours of ambient noise were recorded with the AUV at the ocean surface (center of the VLA at 38 m depth) in a man-made ice hole on March 13. Then, on March 14, more ambient noise data were collected at different water depths by traversing the AUV from 50 m down to 200 m below the ice hole and back up to 25 m depth at $\sim$25 m increments (the center of the VLA varied from 63 to 238 m). At each depth, roughly 15 minutes of data were collected on each leg of the AUV deployment. Similar to the SIMI-94 dataset, while the data recordings during ICEX-16 are not long enough for long-term environmental correlation, they are representative samples of the ambient noise that allows us to study how the soundscape may be shaped by the specific environmental characteristics during ICEX-16.

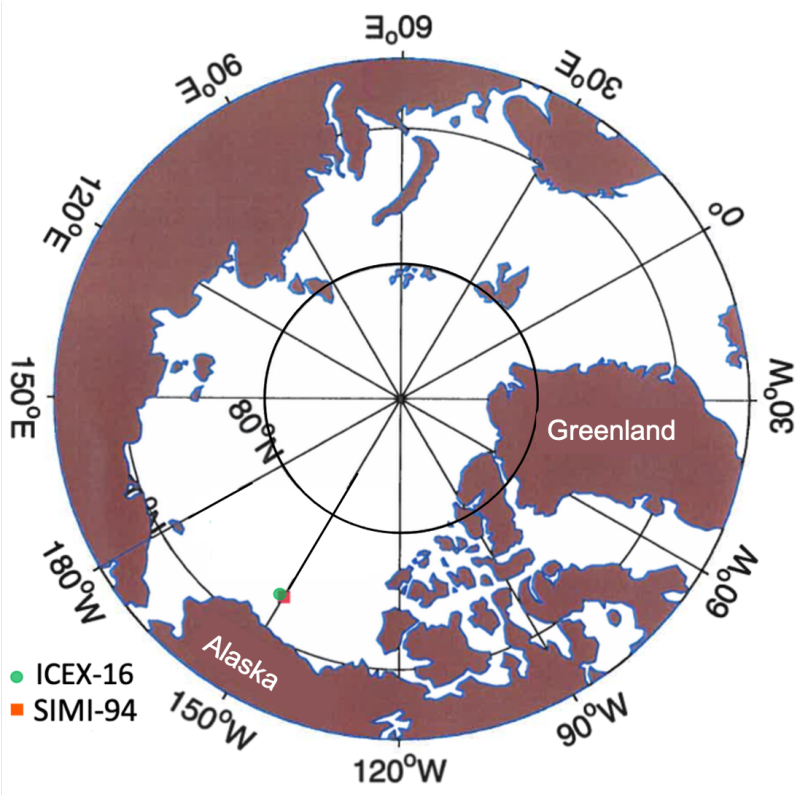


Figure 3.6: Location of ICEX-16 experiment camp site compared to SIMI-94. The coordinates of the ICEX-16 camp on March 13, 2016 were $73^o02'60"N$ and $149^o35'60"W$, which is only ∼6 nautical miles from the SIMI-94 camp location.

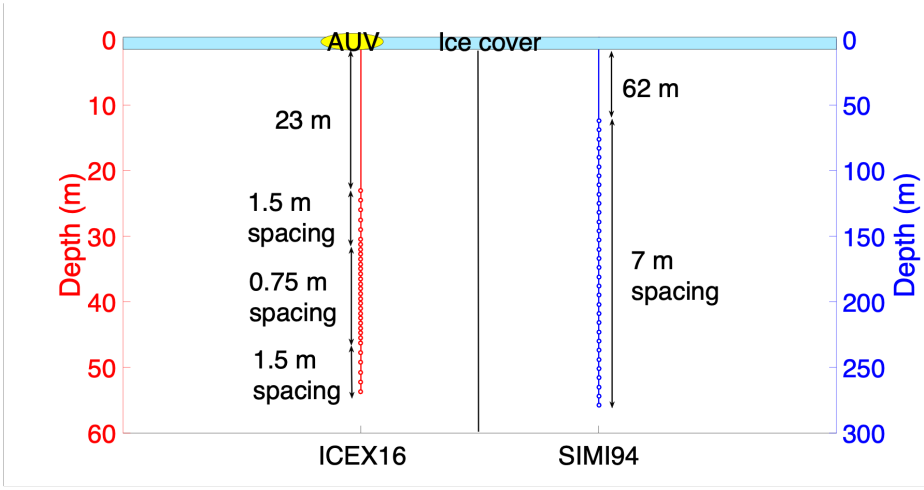


Figure 3.7: The ICEX-16 VLA, deployed with AUV at the surface, compared with the SIMI-94 VLA. Note the depth scale difference between left and right.

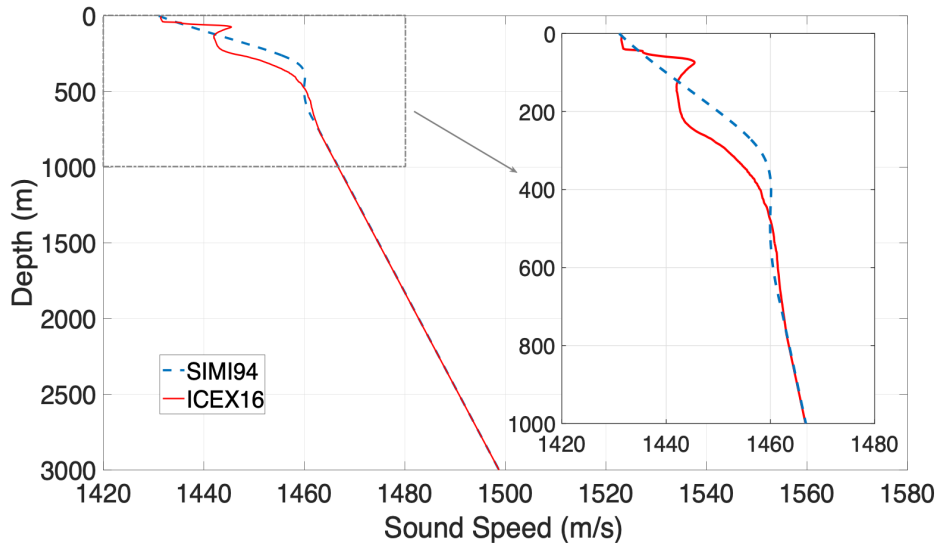
The underwater SSP measured during the ICEX-16 experiment is shown in Fig. 3.8(a). In comparison to the monotonically increasing profile from SIMI-94, the ICEX-16 SSP exhibits the effect of the Beaufort Lens, with a local sound speed maximum near 75 m depth. In addition to the SSP, the ice cover during ICEX-16 also reflects the broad environmental changes occurring in the region. The camp was positioned on top of first year ice with thickness of ∼1-2 m. A ridge in the ice cover was present ∼30-50 km to the north-northeast of the camp location; It can be observed on the ice-temperature satellite imagery on March 13 [87] as a line of higher ice temperature (Fig. 3.8(b)). The presence of this ridge further motivates the hypothesis that ambient noise generation in the new Arctic, such as the environment during ICEX-16, may be more localized in the ice cover along such features rather than uniformly distributed in space. In our later analysis, we test this hypothesis by comparing the measured data during ICEX-16 with model output assuming discrete surface noise generation in addition to the conventional, uniform noise generation approach.

3.3 Impacts of Environmental Changes to Overall Ambient Noise

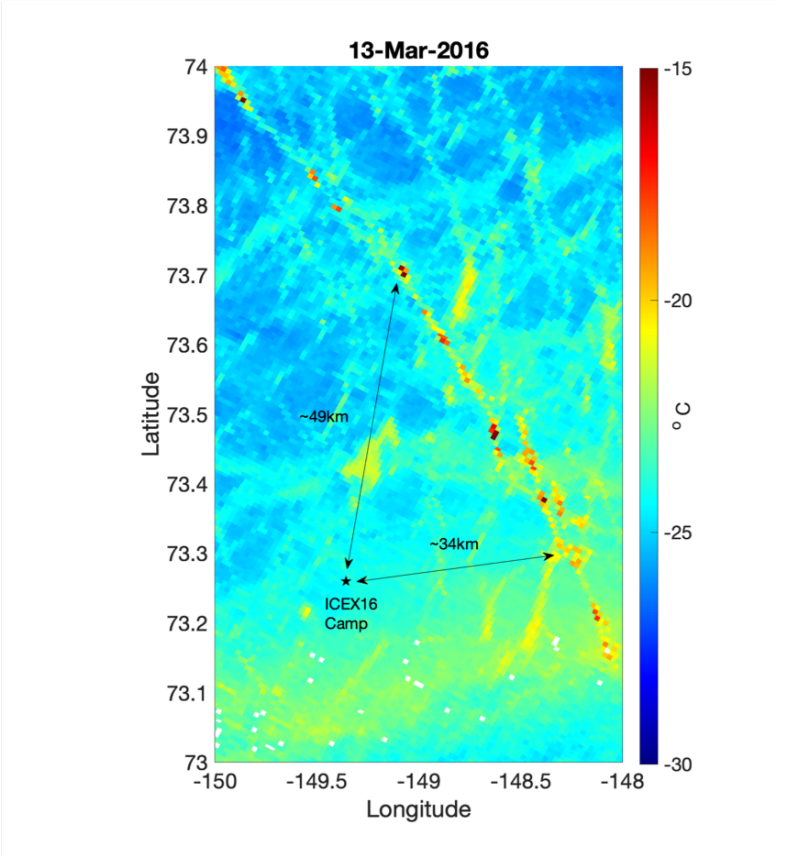
In this section, we analyze the spectral and spatial properties of ambient noise collected during ICEX-16 to explore characteristics that may be attributed to changes in the SSP and ice cover. We start with spectral analysis of the noise data, followed by comparisons of the noise profile and directionality with modeled outputs.

3.3.1 Spectral Analysis

To generate spectrograms of the ICEX-16 data, the recorded time series are segmented using a 2048-point Hanning window with 50% overlap before applying the FFT. For the long recording at 38 m depth, spectral information from the lowest hydrophone in the VLA, at 54 m, is presented in Fig. 3.9(a). This hydrophone is selected because it was the closest in depth to the hydrophone used to showcase the SIMI-94 spectral



(a)



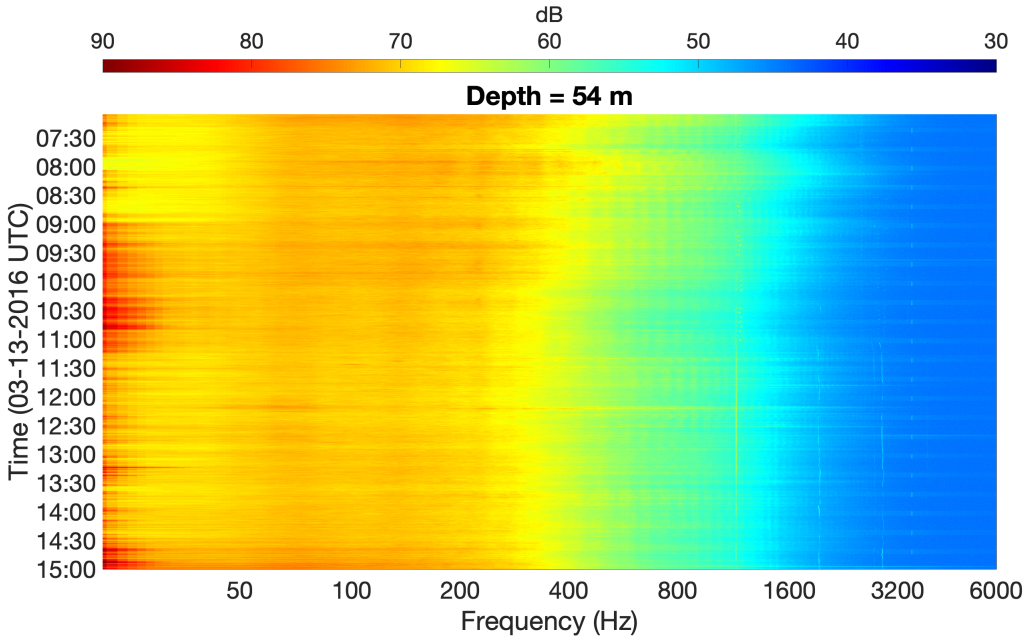
(b)

Figure 3.8: (a) SSP during ICEX-16, in comparison with the SIMI-94 profile, exhibits the effect of the Beaufort Lens with a local sound speed maximum near 75 m. (b) Ice temperature satellite imagery of the ICEX-16 camp site on March 13, 2016. A line of higher ice temperature to the north- northeast of the camp indicates the presence of a ridge formation.

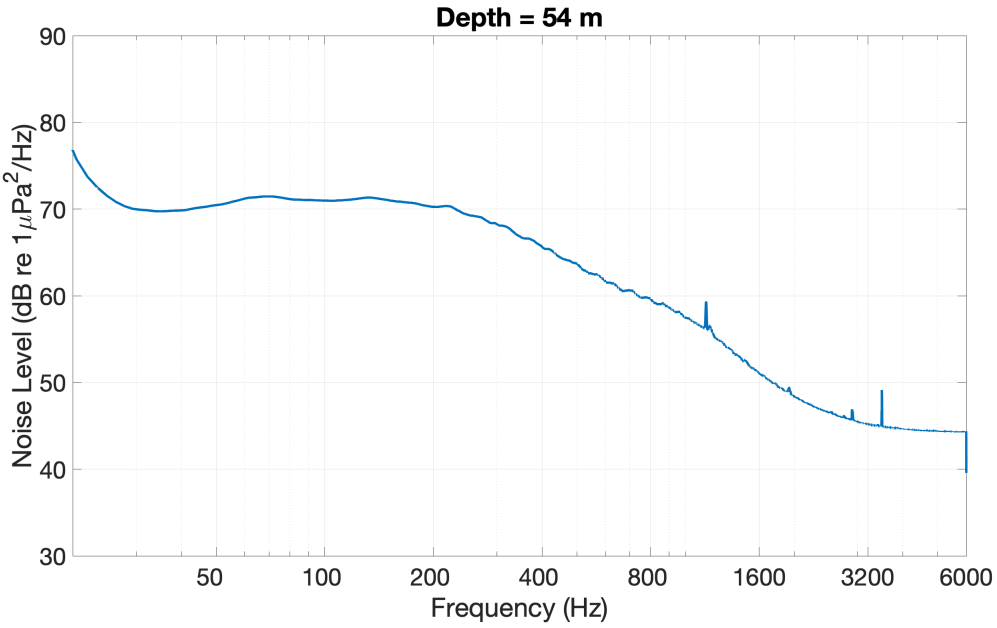
results in Chapter 2. Its spectrogram demonstrates that noise level decreases with increasing frequency and this trend is consistent during the entire recording period. The noise level in the 100-1000 Hz interval varies between ∼75-55 dB. These values are much lower than the levels measured during SIMI-94 and are more in line with results from previous studies (Fig. 2.8). Near 3.5 kHz, a periodic, narrowband interference can be observed. This feature is caused by an acoustic modem used to communicate with the AUV during deployment. Other faint, narrowband interferences are also present near 1, 2, and 3 kHz. These are likely man-made noise from the camp site or AUV noise. Fig. 3.9(c) compares the median PSD estimate of this long recording with the PSD of the SIMI-94 data in the frequency interval that the two datasets overlap. It shows noise level measured during SIMI-94 was ∼30 dB higher than the level during ICEX-16. However, this large increase is likely attributed to the higher array self-noise of the SIMI-94 VLA and may not be indicative of the actual noise level difference between the two experiments.

3.3.2 Noise profile with Depth

On March 14, ambient noise data were collected at various depths in the water column by moving the deployed AUV between 25-200 m depth (the VLA center positioned between 38-238 m). As a result, this dataset is used to construct a profile of ambient noise level vs. depth. Fig. 3.10(a) shows the VLA center depth with time during this recording session as well as the corresponding noise spectrogram at each depth calculated from the data on the center hydrophone. On the spectrogram, broadband interferences, which are more prominent at lower frequencies, are observed whenever the array center depth changes. Thus, these features are likely associated with AUV or array movement noise as the vehicle changes depth in the water column. Periodic interferences are also observed near 3.5 kHz. These are again due to the acoustic modem. Overall, the spectral shape of the noise at each depth is similar to the data recorded on March 13 - as frequency increases, the noise level diminishes. Fig. 3.10(c) presents the noise level as a function of depth. This profile is estimated from the spectrograms of all 32 channels in the array but excluding data that contain artificial

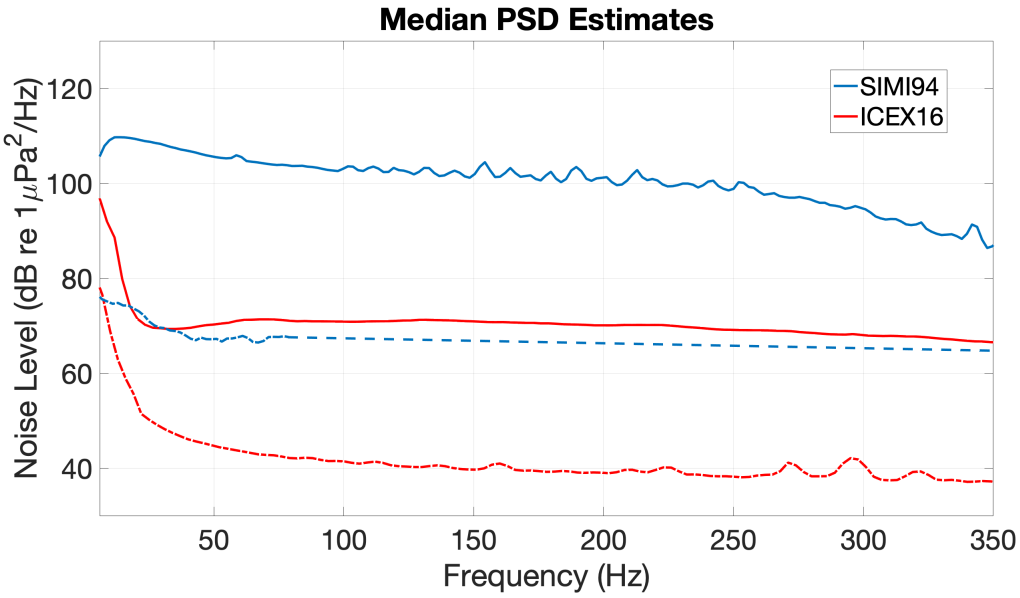


(a)



(b)

Figure 3.9: (a) Spectrogram of an ∼8 hour recording during ICEX-16 on March 13. Periodic interference at 3.5 kHz is caused by an acoustic modem used to communicate with the AUV. Note the frequency axis scale is logarithmic. (b) Median PSD estimate of this recording up to 6 kHz. The noise level roll-off between the 300-3000 Hz decade is ∼23 dB. (c) Median PSD estimate of this recording compared with SIMI-94 Tape 23 PSD over the frequency interval that the datasets overlap. The array self-noise estimates are presented as dot-dashed lines. The ICEX-16 VLA self-noise is estimated using its k-f beamforming result.



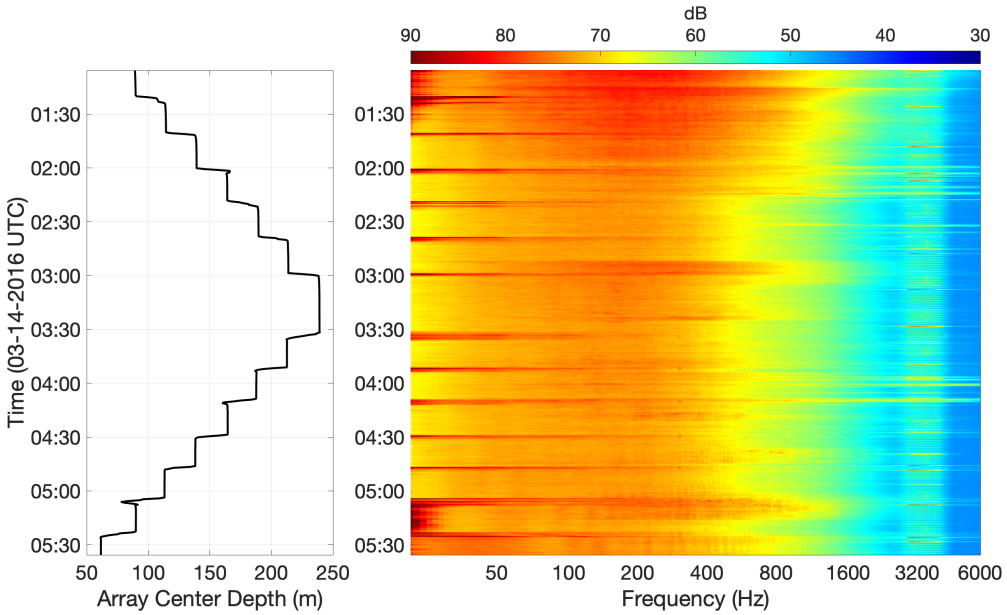
(c)

Figure 3.9: (a) Spectrogram of an ∼8 hour recording during ICEX-16 on March 13. Periodic interference at 3.5 kHz is caused by an acoustic modem used to communicate with the AUV. Note the frequency axis scale is logarithmic. (b) Median PSD estimate of this recording up to 6 kHz. The noise level roll-off between the 300-3000 Hz decade is ∼23 dB. (c) Median PSD estimate of this recording compared with SIMI-94 Tape 23 PSD over the frequency interval that the datasets overlap. The array self-noise estimates are presented as dot-dashed lines. The ICEX-16 VLA self-noise is estimated using its k-f beamforming result.

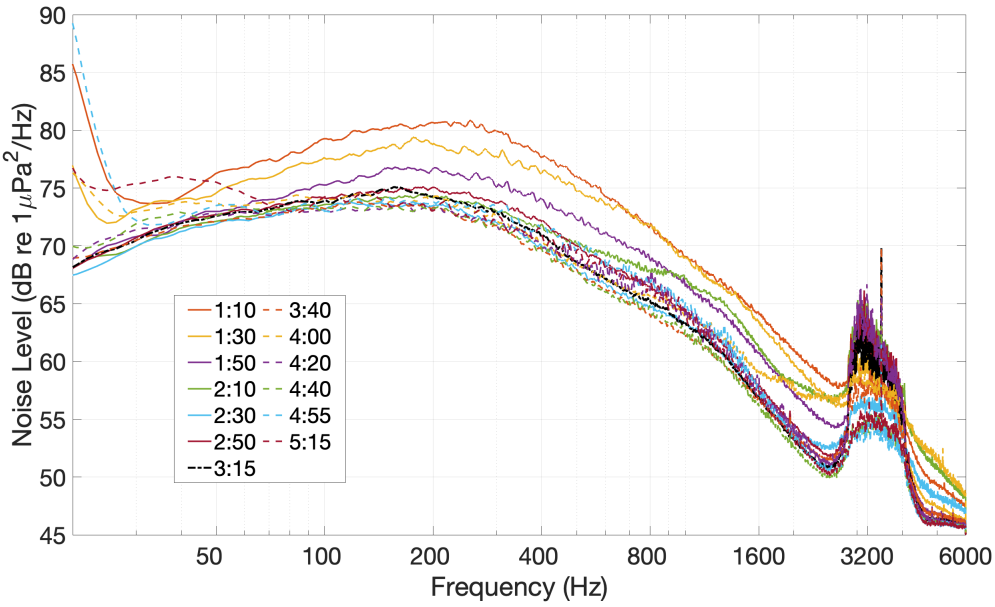
interferences. Each dot in this plot represents the averaged spectrogram value over a 15 s snapshot at 850 Hz while the solid line represents the moving average over every 750 snapshots. 850 Hz is selected as an example here because it aligns with the frequency interval used in our beamforming analysis later. However, the noise profiles at other frequencies are similar in shape to the result at 850 Hz. From this profile, we see that noise level is not uniform in depth. Instead, it peaks just above 75 m, then decreases with depth to about 225 m. The cause of the noise level variation is likely the Beaufort Lens. This theory is highlighted by plotting the SSP during ICEX-16 next to the noise profile. The noise level peak aligns with the local sound speed maximum in the SSP, which indicates much of the noise is trapped within the surface duct by the positive sound speed gradient. Consequently, less noise propagates into the lower duct, creating a quieter ambient environment below SSP maximum.

To further study the effect of the Beaufort Lens on ambient noise level vs. depth, as well as possible effects due to a change in the spatial distribution of noise generation in the ice cover, we compute modeled noise level under various scenarios to compare with the measured data. The modeling parameters selected for the ice cover, water column, and bottom halfspace are similar to those used in Chapter 2. They are presented in Table 3.1. As before, OASES [35] is used to carry out the modeling.

First, to isolate the effect of the Beaufort Lens, we examine the modeled noise level vs. depth in two cases. The first uses the historical, monotonically increasing, Arctic SSP, while the second uses a generic Beaufort Lens SSP, which is taken as the average of all measured profiles shown in Fig. 3.4(b). For surface noise generation, both cases use a uniform noise distribution with the source frequency at 850 Hz (the modeled noise profiles at other frequencies show similar shapes). Here, our goal is not to model the exact value of noise level at different depths; instead, we are interested in how the noise level changes, relatively, with depth. Thus, the particular choice of source strength is not important and we select an arbitrary value. Fig. 3.11(a) shows the modeled noise level in the top 250 m for these two cases. With the historical Arctic SSP, noise level decreases consistently with depth (left side of figure). A slight knee in the SSP at 100 m may trap some acoustic rays that propagate at very shallow

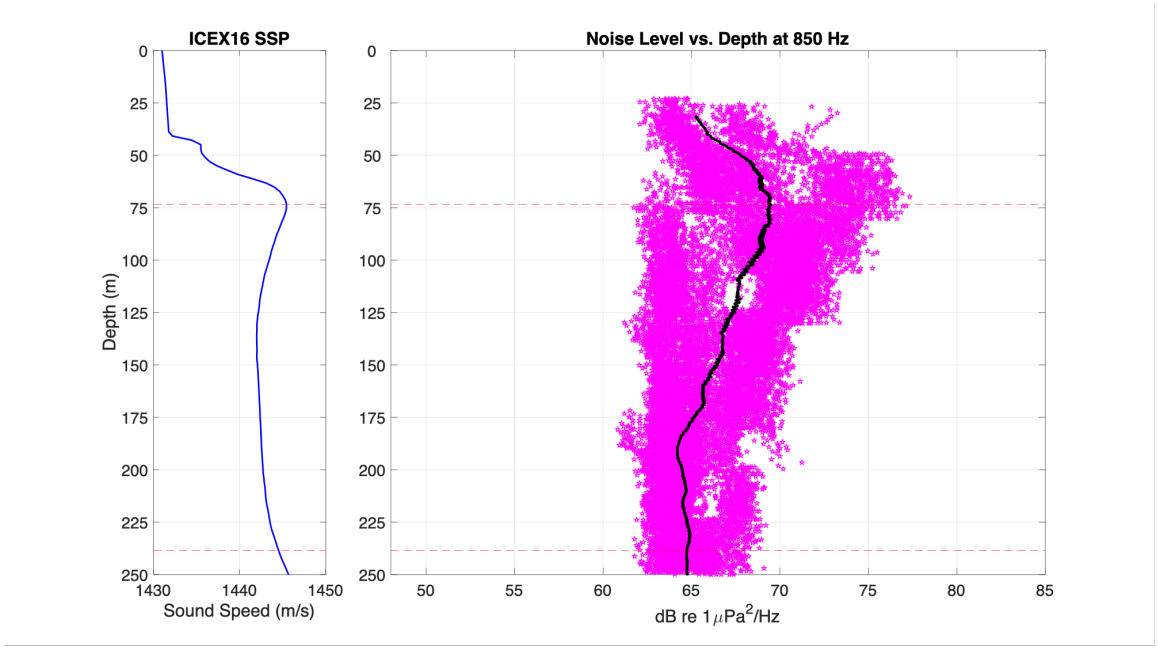


(a)



(b)

Figure 3.10: (a) Spectrogram of data recorded during ICEX-16 on March 14. Left plot shows the center depth of VLA with time during data acquisition. Broadband interferences in the spectrogram are associated with AUV or array movement as they changed depth. Periodic interference at 3.5 kHz is caused by the acoustic modem. Note the frequency axis scale is logarithmic. (b) Median PSD estimate during the time periods that the VLA was not changing depth. The legend denotes the approximate center times of each of these periods. The noise level roll-off between the 260-2600 Hz decade is between ∼17-22 dB for these PSDs. Again, the increase in noise near 3.5 kHz is caused by the acoustic modem.



(c)

Figure 3.10: (c) Noise level vs. depth of the collected data. Dots represent calculated noise level for individual snapshot windows; solid line represents moving average over every 750 windows. Plot on the left shows the SSP during ICEX-16 with the red dashed lines denoting the extent of the lower duct created by the Beaufort Lens.

angles near the surface; however, in general, the water column becomes quieter with depth in a mostly linear trend. In contrast, the right side of Fig. 3.11(a) shows that with a generic Beaufort Lens SSP, the noise profile generated with a uniform surface source distribution (black line) has higher noise level above the sound speed local maximum and the level dips quickly with depth near the top of the lower duct (outlined by dashed red lines). This result makes sense as the sharp local maximum in the Beaufort Lens SSP more effectively traps surface generated noise near the top of the water column than the historical profile. Consequently, less noise is able to propagate into the lower duct, creating a more distinct drop in noise level in comparison.

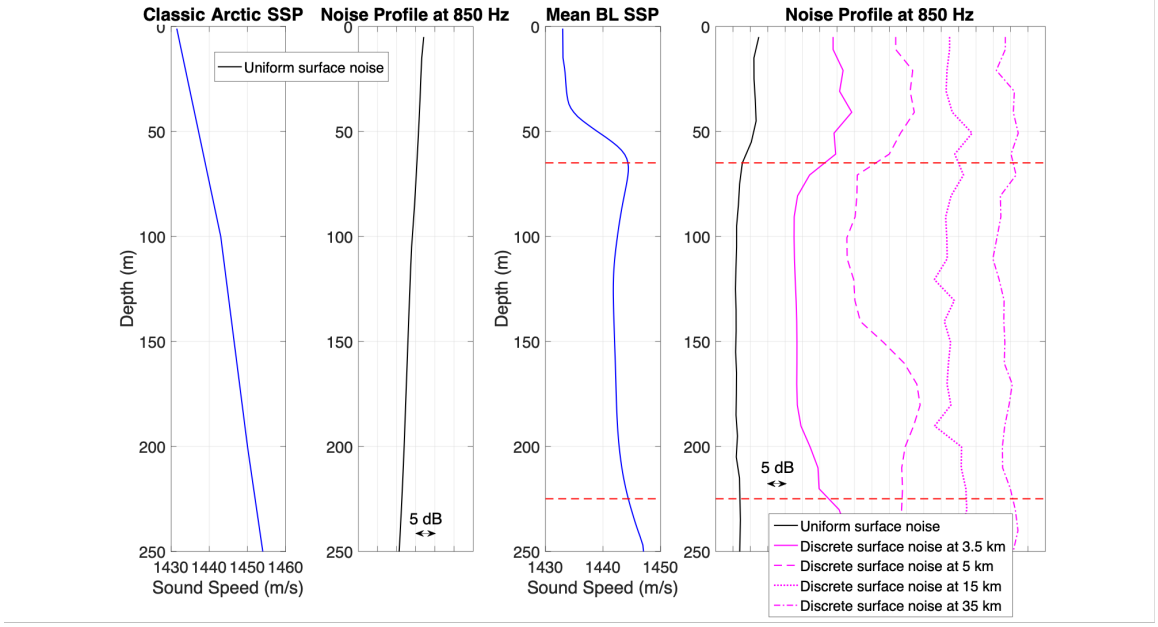
Table 3.1: Parameters used to model the ICEX-16 environment. C_p denotes compressional speed, C_s denotes shear speed, ρ denotes density, λ denotes spatial wavelength, h_{RMS} denotes root-mean-square roughness, η denotes roughness correlation length, D denotes layer thickness. Parameter values are selected based on discussions in [3, 65].

Layer	**Parameters**
Ice	$C_p = 3600$ m/s, $C_s = 1800$ m/s $\rho = 0.9$ g/cm^3, $D = 1$ m $h_{RMS} = 0.2$ m, $\eta = 20$ m
Water column	$\rho = 1.0$ g/cm^3, $D = 3000$ m $C_p =$ Varies (see text)
Bottom halfspace	$C_p = 2200$ m/s, $C_s = 1500$ m/s $\rho = 2.9$ g/cm^3

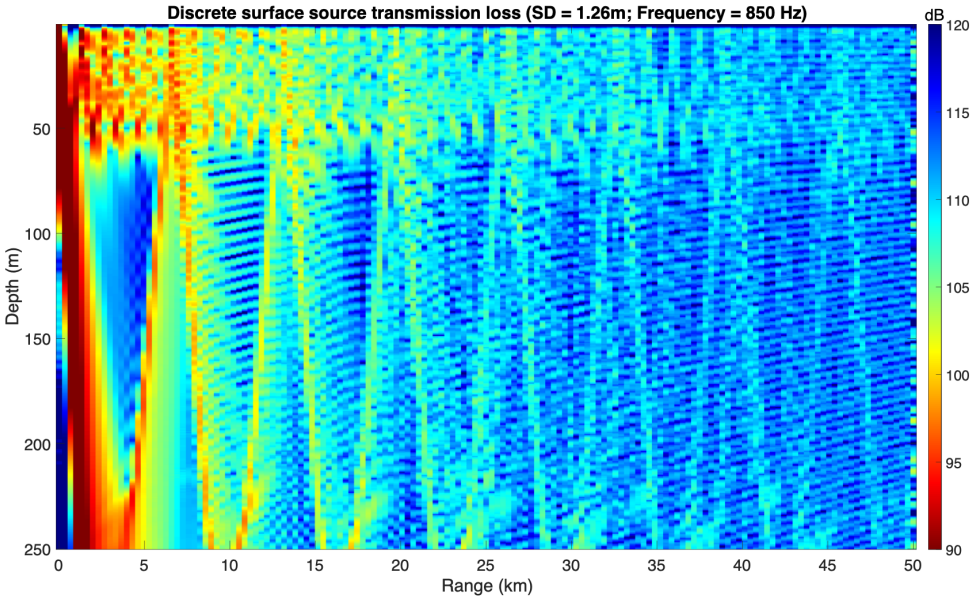
In addition to the underwater SSP, we also test the effect of surface noise generation on noise level vs. depth. As hypothesized, noise generation in the Arctic may be shifting to occur in more localized areas in the ice cover such as along ridge formations. Thus, we model a case with discrete surface noise generation to compare with the uniform noise generation result. For both, the Beaufort Lens SSP is adopted in the water column. To model discrete surface noise, we follow a similar approach as the uniform case described in Chapter 2 - we place a single, near surface ($\lambda_{top}/30$

below the ice), monopole source at the range of expected noise generation. By using a discrete noise source, the modeled noise level becomes inherently range dependent. As a result, we plot outputs for various source ranges, shown by the magenta lines on the right side of Fig. 3.11(a). From this plot, we see that the shapes of the discrete source noise level profiles are similar to the uniform noise case. Due to the Beaufort Lens SSP, the noise level is again higher in the surface duct and dips in the lower duct. However, one key difference between the two is that for the discrete noise case, depending on the source range, the noise level within the lower duct changes. For example, at 3.5 km range (solid magenta line), the noise level within the lower duct shows a larger decrease from the surface noise level than the profiles for the source at 15 and 35 km (dotted and dot-dashed lines). Furthermore, with the discrete source at 5 km, the noise profile shows an increase in noise level about halfway down the lower duct near 150 m. These features can be explained by examining the transmission loss plot of a near surface source in an environment with the Beaufort Lens SSP, shown in Fig. 3.11(b). At 3.5 km, there is a distinct shadow zone in the lower duct. Below the duct, the noise level increases again near 225 m depth due to shallow convergence zone propagation. Similarly, at 5 km, the increase in noise level within the lower duct is due to convergence zone propagation refracting back to the surface. Further out in range, at 15 and 35 km, the shadow zone in the lower duct becomes less prominent and the convergence zone propagation is weaker as well due to attenuation, which is why the decrease in noise level compared to the surface is less dramatic. Consequently, the noise level profiles at these further ranges become more similar to the profile generated with uniform surface noise.

Given our modeling results, we compare the measured noise profile (Fig. 3.10(b)) to the modeled outputs. As discussed, the measured profile clearly demonstrates the effect of the Beaufort Lens, with noise level peaking near the depth of the sound speed maximum. However, it is less evident whether the measured profile more closely resembles the modeled profiles generated with uniform surface noise or discrete surface noise (Fig. 3.11(a), right). The reason for this ambiguity is partly because the modeled profiles in both cases are very similar. Furthermore, if we assume surface



(a)



(b)

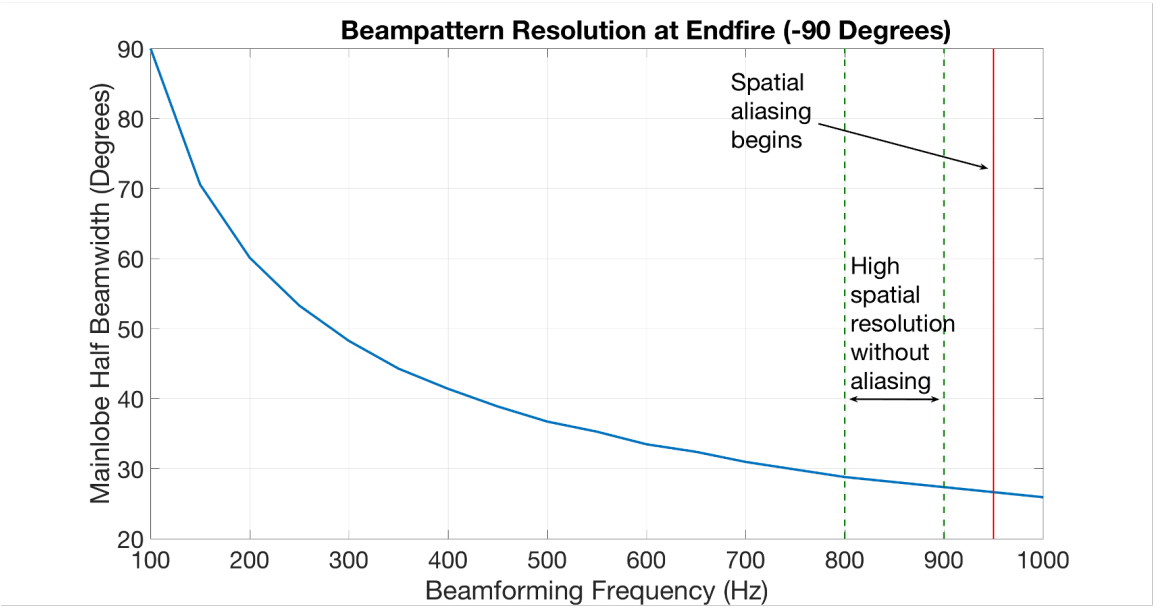
Figure 3.11: (a) Modeled ambient noise level vs. depth with historical Arctic SSP and uniform surface noise (left); Beaufort Lens SSP and uniform surface noise (right, left-most profile); Beaufort Lens SSP and discrete surface noise (right, all other profiles). The distance between the x-axis tick marks on the profile plots denotes a 5 dB change in noise level. (b) Transmission loss of a discrete near surface source in an environment with the Beaufort Lens SSP.

noise generation during ICEX-16 occurred ∼30-50 km away from the VLA along the ridge formation shown in Fig. 3.8(b), at these distances, the range dependency of the discrete source noise profiles is much weaker compared to closer ranges. As a result, the profiles more closely resemble the shape of the uniform source distribution noise profile. Thus, in regards to underwater ambient noise level during ICEX-16, the Beaufort Lens SSP appears to be a much more influential factor than the spatial distribution of noise generation in the ice cover.

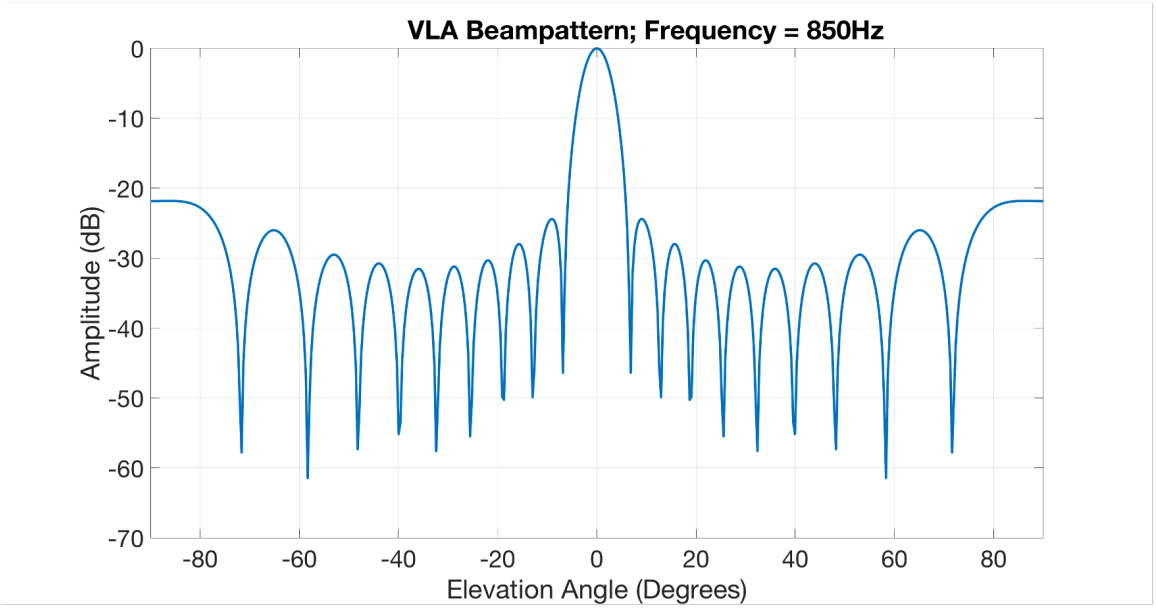
3.3.3 Noise Vertical Directionality

Similar to the analysis of the SIMI-94 data, we again apply conventional beamforming to examine the vertical directionality of noise recorded during ICEX-16. The geometry of the ICEX-16 VLA constricts the maximum beamforming frequency at 950 Hz. Below this frequency, there is no spatial aliasing even when the array is steered towards endfire (Fig. 3.12(a)). The best spatial resolution occurs between 800-900 Hz. Thus, the beamform output within this frequency interval is averaged over time to create noise level vs. elevation plots. When steered towards broadside, the ICEX-16 VLA beampattern at 850 Hz has a 3-dB-down beamwidth of ∼2.5 degrees (Fig. 3.12(b)).

Fig. 3.13(a) shows that throughout most of the recording period on March 13, noise level peaks near -10-15 degrees in elevation angle. When averaged over time, the mean profile peaks near -10 degrees. This result differs from the noise directionality during SIMI-94, which peaks at 0 degrees elevation. Fig. 3.13(b) takes a closer examination of noise arriving from the peak elevation angle. Again, unlike noise recorded during SIMI-94, no persistent tonal bands exist in the ICEX-16 data. The beamform output for ambient noise recorded on March 14 shows a similar result. The noise vertical directionality peaks near ±15 degrees, creating a noise notch near the horizontal. This feature is consistent with depth, as demonstrated by the mean profiles at 138 and 238 m (Fig. 3.14(a)). Again, no persistent tonal bands are observed in the noise that arrives at the peak elevation angle (Fig. 3.14(b)). The lack of persistent tones in the ICEX-16 data may be indicative of the shift in Arctic



(a)



(b)

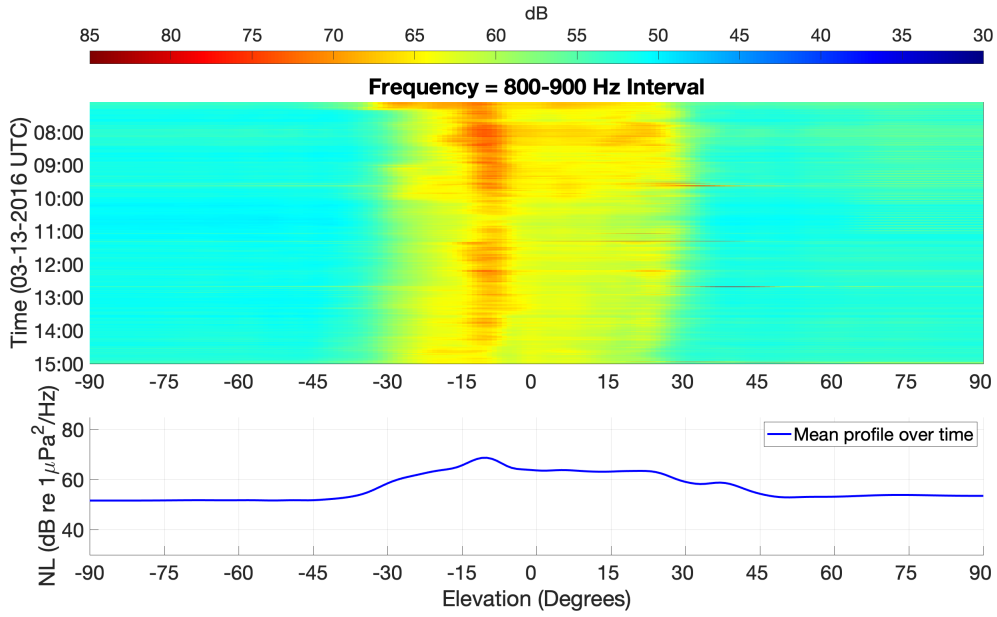
Figure 3.12: (a) Spatial aliasing for the ICEX-16 VLA occurs at endfire near 950 Hz. Spatial resolution becomes higher as frequency increases towards the aliasing frequency. (b) The VLA beampattern steered towards broadside at 850 Hz shows 3-dB-down mainlobe width of $\sim$2.5 degrees.

ice cover characteristics. It is possible the thinner and younger ice cover during ICEX-16 could not sustain the constant shear from rubbing past adjacent floes to produce the persistent tones observed in the SIMI-94 data. Instead, the ice more likely fractured or formed ridges, leading to more discrete surface noise generation. To further investigate this hypothesis and better explain the observed features in the ICEX-16 noise vertical directionality profiles, we again turn to modeling.

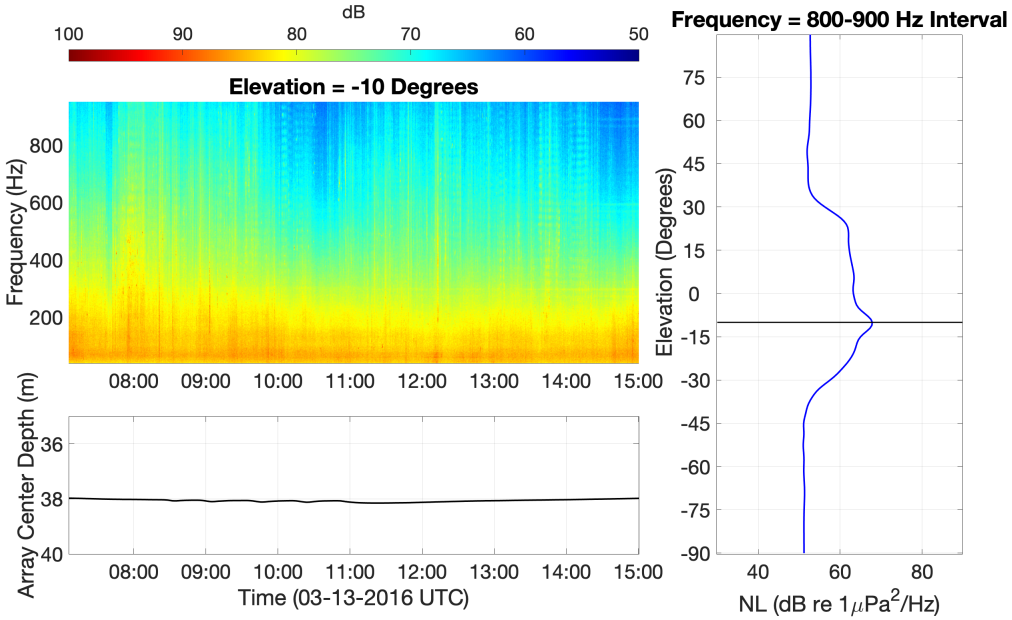
There are two distinct features in the ICEX-16 noise vertical directionality profiles. One is the existence of a noise notch near horizontal elevation angles. The notch is not as evident at 38 m depth but is pronounced at deeper depths as seen in the 138 and 238 m profiles. The second feature is the peak in the profiles, which occurs near -10 to -15 degrees. Here, we take a closer look at how the Beaufort Lens SSP and discrete surface noise generation by the ice cover can explain these features.

Using raytracing carried out by Bellhop [40], we explain why horizontal noise notches are observed at 138 and 238 m depths but not so much at 38 m depth. For our modeling environment, we use the SSP measured during ICEX-16 (Fig. 3.8(a)) and assume discrete noise generation by the ice cover. First, we place a monopole source at 38 m depth. The raytraces with departure angles between ± 7 degrees are shown in Fig. 3.15(a). Green lines represent rays that interact with the surface, while red lines represent rays that do not interact with the surface. With the source at 38 m, all shallow angle rays interact with the surface. By reciprocity, this means that rays generated by a discrete surface noise source at any range can arrive at a receiver at 38 m depth at very shallow arrival angles. Hence, this is why we do not observe a strong horizontal noise notch in the data collected at 38 m. With a source placed at 138 or 238 m depth, however, there are rays with very shallow departure angles that will never interact with the surface. Again, by reciprocity, this means that fewer rays generated by a discrete surface noise source would arrive at a receiver at those depths at shallow angles. In the real environment, this would translate to a reduction in noise arriving from the horizontal, which is why we observe the noise notches in the directionality profiles for 138 and 238 m depths.

The second feature in the ICEX-16 directionality profiles is the peak in noise

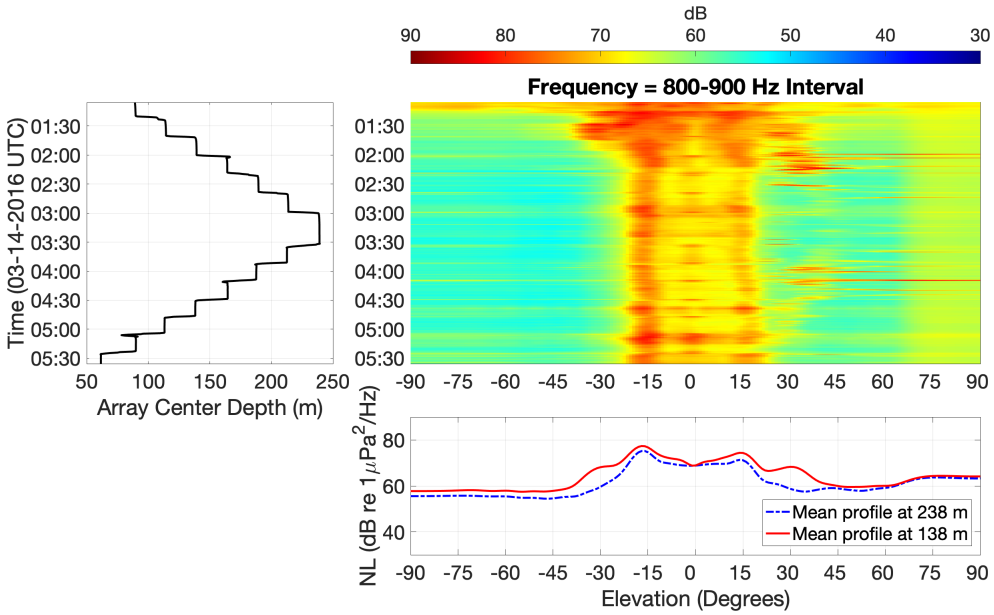


(a)

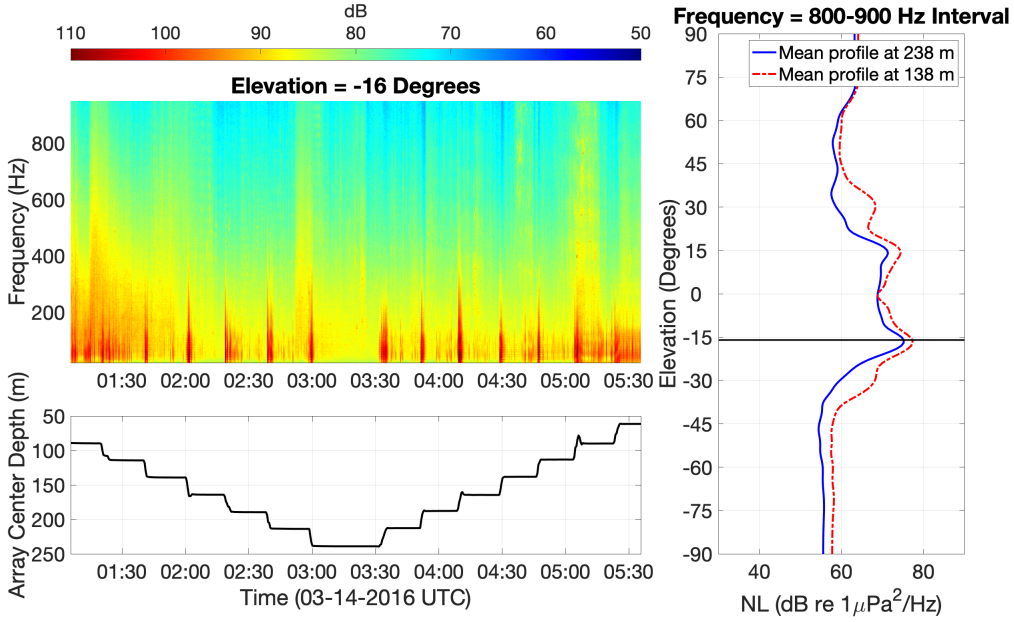


(b)

Figure 3.13: (a) Beamform output over time averaged over the 800-900 Hz frequency interval for data collected on March 13. Noise level peaks near -10 degrees elevation. Bottom plot shows mean vertical directionality profile generated by averaging top plot in time. (b) Beamform output at -10 degree elevation angle as a function of frequency and time. No persistent meandering tones are observed. Bottom plot shows VLA center depth during the recording session. Right plot shows mean vertical directionality profile over time in the 800-900 Hz frequency interval.



(a)



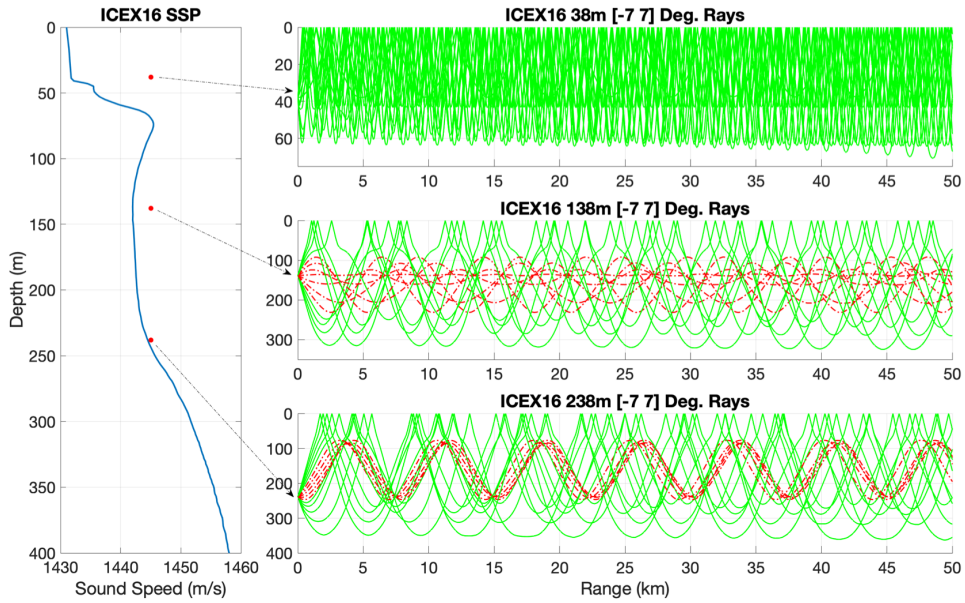
(b)

Figure 3.14: (a) Beamform output over time averaged over the 800-900 Hz frequency interval for data collected on March 14. Noise level peaks near ±16 degrees elevation. Bottom plot shows mean vertical directionality profiles generated by averaging the results for 138 and 238 m depth, respectively. (b) Beamform output at -16 degree elevation angle as a function of frequency and time. No persistent meandering tones are observed. Broadband interferences are present when the array center depth changes, indicating contamination from AUV or array movement noise. Bottom plot shows VLA center depth during the recording session. Right plot shows mean vertical directionality profiles generated by averaging the results for 138 and 238 m depth, respectively.

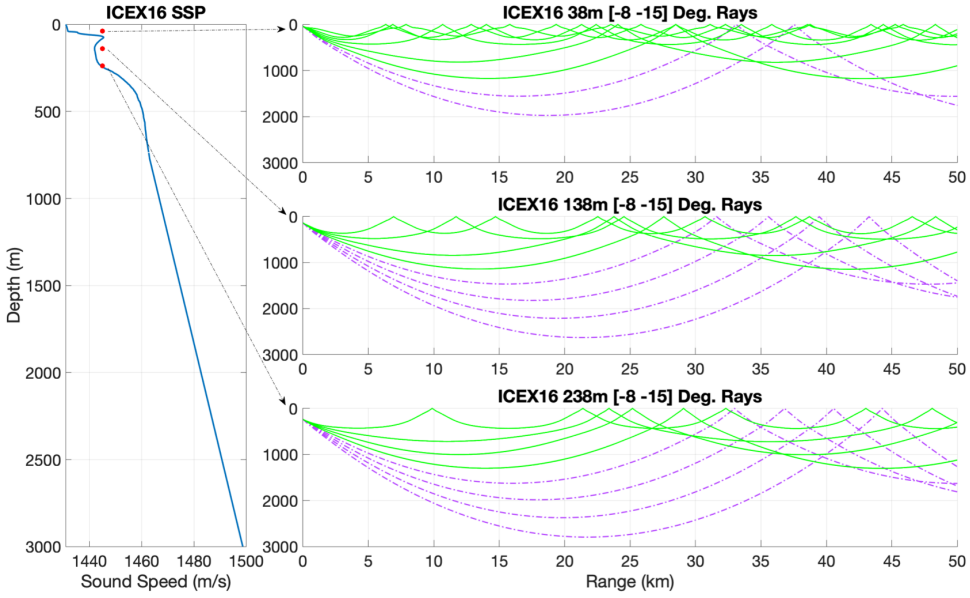
level from near -10 to -15 degrees. This can again be explained using raytracing and by invoking reciprocity. In Fig. 3.15(b), we show the raytraces for a monopole source placed at 38, 138, and 238 m depth, respectively, for departure angles between -8 and -15 degrees. Based on the location of the ridge formation with respect to the ICEX-16 camp shown in Fig. 3.8(b), we assume that surface noise is generated between 30-50 km away from the VLA. With this in mind, the rays highlighted by the dot-dashed purple lines in Fig. 3.15(b) are the ones that arrive at the surface between 30-50 km away from the VLA via a direct path. The departure angles of these highlighted rays range from -12 to -15 degrees. Through reciprocity, this result means that for a discrete surface noise source placed between 30-50 km away from the origin, the noise level received at 0 km range at 38, 138, and 238 m depths should be the highest between -12 to -15 degrees because these angles correspond to the arrival angles of direct paths. For steeper or shallower arrival angles, the noise level would be weaker because the ray paths would experience at least one surface or bottom bounce. Thus, these raytracing results suggest that the noise directionality profiles should peak near -12 to -15 degrees, which is very similar to what we observe in the ICEX-16 directionality profiles. Thus, our raytracing results are consistent with the hypothesis that surface noise generation during ICEX-16 was more localized in space, specifically, noise was predominately generated $\sim$30-50 km away from the VLA.

Further evidence of surface noise generation during ICEX-16 being more discrete in space can be seen by modeling the noise vertical directionality directly. OASES [35] is used to carry out the modeling here. First, we present the result with uniform surface noise generation. Fig. 3.16 shows the modeled profiles in this case differ significantly from the measured ICEX-16 profiles at all 3 examined depths. Peak elevation angles in the modeled profiles are positioned closer to the horizontal plane. Furthermore, although the modeled profiles do show a notch at 0 degrees elevation, the extent of this notch is more confined compared to the measured profiles. Thus, these differences suggest that the assumed uniform surface noise distribution is not a valid description of noise generation during ICEX-16.

For a second comparison, we model the noise directionality profiles assuming dis-



(a)



(b)

Figure 3.15: Raytraces for monopole sources placed at 38, 138, and 238 m depths in the ICEX-16 environment. Solid lines denote rays that interact with the surface. (a) For 138 and 238 m depths, some rays with shallow departure angles would never interact with the surface (dot-dashed lines). (b) Rays that reach the surface 30-50 km away from the source via direct paths (dot-dashed lines) have departure angles between -12 to -15 degrees.

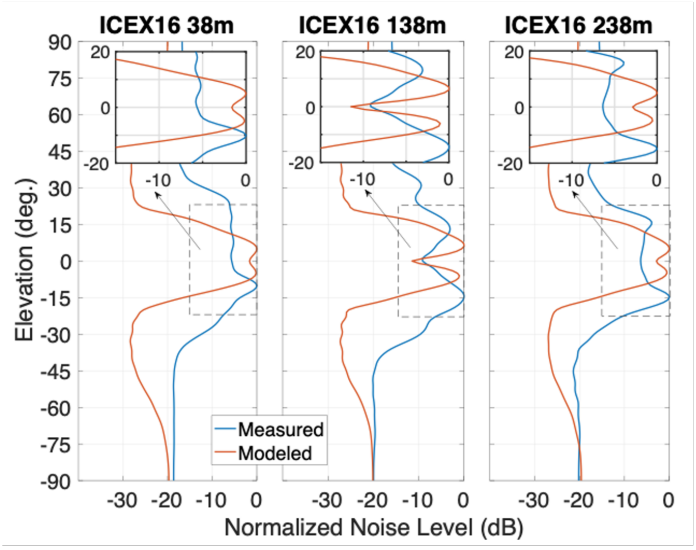


Figure 3.16: Modeled noise vertical directionality profiles with uniform surface noise generation in the ICEX-16 environment with the center of the array at 38 m, 138 m, and 238 m depths. Peak elevation angles of the modeled profiles do not matches those of the measured profiles.

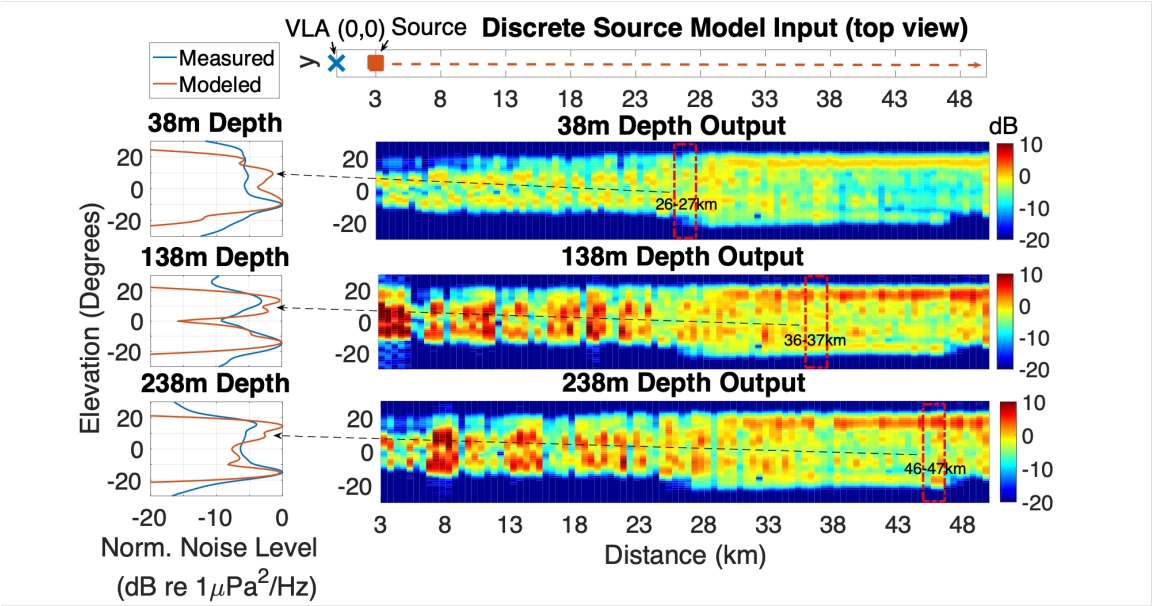


Figure 3.17: Modeled noise vertical directionality profiles with discrete surface noise generation in the ICEX-16 environment with the center of the VLA at 38 m, 138 m, and 238 m depths. Top plot shows the range placement of the surface monopole source. On the right, modeled vertical directionality is shown with respect to source distance as a contour plot. On the left, average model output over the indicated ranges on the contour plot are compared with measured profiles at the respective depths.

crete surface noise generation at different ranges. This is accomplished by placing a single monopole source near the surface (λ_{top}/30 m depth) at various distances (3-50 km at 0.5 km increments) from the VLA. The set-up and result for this case is presented in Fig. 3.17. At each source position, the resultant vertical directionality profile is shown in the form of a contour plot below. At all three examined depths, as the source is positioned further away from the VLA, the peak elevation angle of the modeled output shifts further away from the horizontal and the noise notch becomes more apparent. Thus, the output of this discrete source model is more similar to the measured profiles than the uniform noise generation model. Particularly, as shown on the left side of Fig. 3.17, for 38 m depth, the averaged output between 26-27 km matches the measured output most closely, suggesting that ambient noise recorded during this time is dominated by ridging activity at that distance. For 138 and 238 m, the averaged model output between 36-37 km and 46-47 km closely match the measured outputs, respectively. These results are again consistent with the hypothesis that noise generation during ICEX-16 occurred between ∼30-50 km away from the VLA.

3.4 Transient Event Characteristics

As noted in Chapter 2, analysis of the temporal distribution of transient ambient noise events allows for a better understanding of ice activity as a function of time and the mechanisms by which transient events are generated. In this section, we provide a complementary analysis of transient event characteristics during ICEX-16 to the SIMI-94 results. Specifically, we compare the temporal characteristics of transient occurrences during the two experiments and discuss whether any differences may be attributed to the shift in the Beaufort Sea environment.

3.4.1 Event Detection Approach

Ambient noise recorded on March 13 is analyzed to detect transient events. This dataset is selected because, similar to the SIMI-94 Tape 23 data, it is a continuous

period of recording at a constant depth. The length of the March 13 dataset is about 8 hours. However, as noted, periodic interference exists in the data due to the acoustic modem. While the frequency of this interference is primarily centered at 3.5 kHz, its signature is more faintly present at other frequencies as well. Thus, periods of data that contain the modem noise is excluded from the dataset. This results in a total of 5.65 hours of clean ambient noise data, which is comparable to the 6 hours of data analyzed for the SIMI-94 experiment.

The algorithm used to detect transient events within the ICEX-16 dataset is the same, amplitude-based method presented in Chapter 2. The three algorithm parameters are again set as $w = 10$, $r = 1\%$, and $p_{thres} = 0.00001$. Consistent with the procedure for the SIMI-94 analysis, the time series data are filtered into three separate octave bands between 40 and 320 Hz. The detection method is then applied to data on all 32 channels and detected transients are only confirmed to be events if they appear on at least 3/4 of the channels. After event selection, the beginning and end times of each event are documented.

3.4.2 Comparison with SIMI-94 Results

Similar to the results for the SIMI-94 dataset, Table 3.2 shows the number of detected transient events in the ICEX-16 data increases with frequency. Comparison of the count within each frequency octave reveals the number of transients is much lower during ICEX-16 than SIMI-94. This decrease suggests the ice cover during ICEX-16 was less active at producing transients than the ice cover during SIMI-94. However, this conclusion is incomplete. Fig. 3.18 shows the detected events during ICEX-

Table 3.2: Number of detected transient events in each frequency band.

	40-80 Hz	80-160 Hz	160-320 Hz
SIMI-94	173	451	1043
ICEX-16	43	86	260

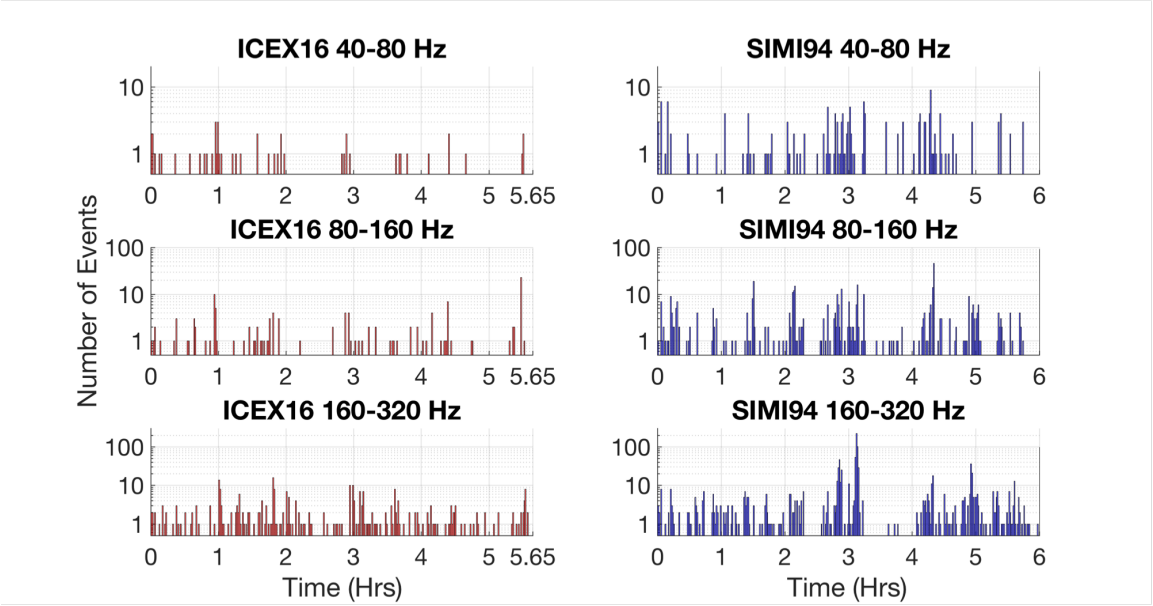


Figure 3.18: Temporal distribution of transient events during ICEX-16 and SIMI-94 at three octave bands. Plots show the clustering of events and time gaps that separate the clusters.

16, as well as during SIMI-94, grouped into clusters, where the y-axis shows the number of events within each 1-minute time window. The statistical properties of these clusters, such as the length of the time gap between successive clusters, the length of each cluster, and the number of events within each cluster, are presented in Fig. 3.19. The comparison between ICEX-16 and SIMI-94 results in this figure shows the gap length distribution during the two experiments are similar at all three frequency octaves. This means event clusters occurred at similar rates during both experiments. However, event clusters are likely to be longer during SIMI-94 than ICEX-16 as shown by the distribution of event cluster lengths. The 75th and 99th percentile values for this metric are higher during SIMI-94 than during ICEX-16 at all three frequency octaves. The bottom plot of Fig. 3.19 further shows the number of transient events within each cluster are also likely to be greater during SIMI-94 than ICEX-16. Combined, these results suggest that if an event cluster can be assumed as a period of high ice activity, then SIMI-94 and ICEX-16 have comparable statistics regarding the occurrence of such high activity periods. However, within such a period,

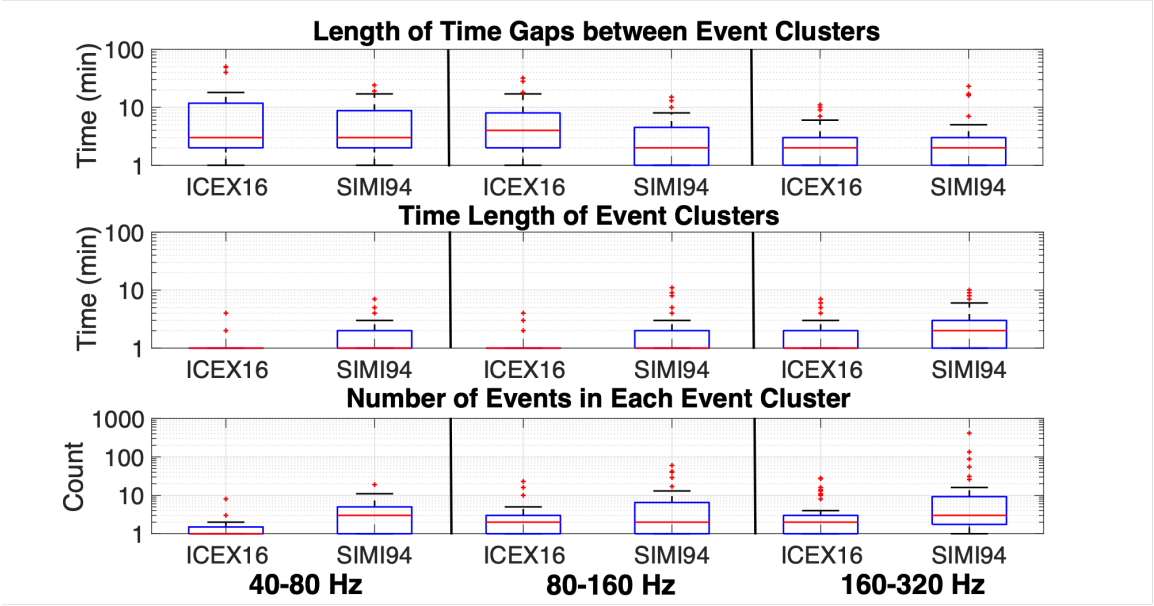


Figure 3.19: (Top) Statistical distribution of time gap lengths, cluster lengths, and number of events in each cluster during ICEX-16 and SIMI-94 at three octave bands. Red lines represent the median, box edges represent the 25th and 75th percentiles, whiskers extend to the 1st and 99th percentiles, and red crosses signify extreme values that are much larger than the rest of the distribution. Gap lengths of the two datasets have comparable distributions. (Middle) Although the median values are similar, event clusters are more likely to be longer during SIMI-94 than ICEX-16. (Bottom) The number of transient events within each cluster are also likely to be greater during SIMI-94 at all three frequency bands.

the ice cover during SIMI-94 was likely more acoustically active and produced more transient events than the ice cover during ICEX-16. The reason for this difference may simply be that environmental forces such as wind or temperature were more prominent during SIMI-94, thus causing the ice cover to produce transient events. However, another plausible explanation may result from the proposed change in the surface noise generation between the two experiments. More transient events would be expected to be generated from a uniform distribution of sources, as is suspected to be the case during SIMI-94, than from a few distinct sources located along a pressure ridge, as is suspected to be the case during ICEX-16. Thus, the difference in the transient noise temporal statistics between the two experiments may reflect a change in surface noise generation by the ice cover.

3.5 Summary

In this chapter, we discussed two major areas of change occurring in the Beaufort Sea region of the Arctic Ocean. One is the underwater SSP, which exhibits a local sound speed maximum near 70 m depth due to intrusion of warm Pacific water into the region. This "Beaufort Lens" SSP creates a double ducted propagation environment in which the lower duct has diminished ambient noise levels and is capable of promoting long-range acoustic propagation. The overall shape of this SSP is still dynamic as more recent measurements demonstrate the possibility of two local maxima within the water column instead of just one. Thus, this component of the Beaufort Sea environment invites continued observation and study.

Another change ongoing in the Beaufort Sea, as well as much of the Arctic Ocean, relates to the surface ice cover. The Arctic ice cover is thinning and becoming younger. The amount of multi-year ice has diminished, replaced by fragile first-year ice that is more susceptible to ridging. Consequently, ambient noise generation by the ice cover may no longer be adequately described by a uniform distribution of sources. Instead, noise generation has become more spatially discrete and better modeled with a single source positioned at a specific range. Modeling results with discrete surface noise generation more consistently match measured data from the ICEX-16 experiment. Specifically, the elevation angle of the peak in noise vertical directionality profiles at various depths suggests surface noise generation during ICEX-16 may have occurred ∼30-50 km away from the deployed VLA. This range interval matches the distance of a ridge formation observed during the experiment to the north-northeast of the camp site.

Lastly, we compared the temporal distribution of transient noise events during ICEX-16 to results from SIMI-94. Our analysis shows the median time gap length between consecutive event clusters is comparable during the two experiments; however, cluster lengths and the number of events in each cluster are greater during SIMI-94 than ICEX-16. These results demonstrate that when the ice cover was active, more transient events occurred during SIMI-94 than ICEX-16. This conclusion

is consistent with the proposed change in surface noise generation in the ice cover from uniform to more spatially discrete.

Chapter 4

Noise Generation Range Estimation

So far in this thesis, we have analyzed ambient noise data from two prior experiments to highlight acoustic characteristics that are indicators of environmental change in the Beaufort Sea. Through our study, we have come to identify two areas where improved analysis tools would further aid in ambient noise characterization. In this and the following chapter, we present these needs for improvement and propose alternative methods to better address them.

One component of the ongoing change in the Arctic Ocean is the shift in ice cover properties. As we have detailed in Chapter 3, the thinner and younger ice cover challenges the assumption that surface noise generation is uniformly distributed in space. Instead, during the ICEX-16 experiment, ambient noise generation was more spatially localized, likely along a ridge formation ∼30-50 km away from the experiment site. We further demonstrated that this shift has a profound and observable impact on the ambient noise soundscape. Thus, in order to more comprehensively describe the noise field in this new Arctic, there is a need for reliable estimation of surface noise generation range; in other words, how far from the location of interest is ambient noise generated?

To address this problem, we introduce a model-based convolution neural network (CNN) approach to surface source range estimation, which maps received noise on

an array to noise generation range through training on simulation data. We compare this method's performance against conventional matched field processing (MFP) and explore how the approach achieves its effectiveness. We begin with an overview of conventional MFP before presenting the architecture and training process of the CNN method. We then test this approach with two real datasets from past experiments and discuss the results.

4.1 Conventional Matched Field Processing

MFP refers to signal processing algorithms which exploit the full field structure of acoustic propagation to infer parameters of the acoustic source or the propagation environment [88, 89, 90, 91]. These parameters may be the source range, depth, or environmental properties such as sub-bottom information [92] or water column SSP [93]. The conceptual basis behind MFP is quite simple. The measured (true) acoustics data in an environment is "matched" with simulated data derived from acoustic modeling of the environment. The parameter of interest is varied during modeling to generate many iterations of simulated data for matching. The parameter value that results in the best match is then outputted as the estimate. The literature on MFP is immense. Bucker was the first to formulate what is now considered conventional MFP [94]. Since then, numerous variants have emerged such as ones that are adaptive [95, 96, 97], constrained [98, 99, 100], take a statistical approach [101, 102, 103, 104], reduce computation load [105], for uncertain [106, 107, 108] or range-dependent environments [109], and for broadband [110, 111, 112, 113] or moving sources [114].

In this chapter, we focus solely on the conventional approach. As described in [89], there are three components of an MFP algorithm. The first is replica generation, where modeling is used to compute the expected acoustic arrival on the receiving array in the environment, otherwise known as the replica field. At a specific frequency f and for a particular set of parameters used for modeling, $\mathbf{A}$, the replica field on an array with L sensors is denoted by $\mathbf{R}(f;\mathbf{A}) = [R_1(f;\mathbf{A})), ..., R_L(f;\mathbf{A}))]^T$. This expression

is complex valued as it involves both the amplitude and phase information of the modeled field. Typically, $\mathbf{R}(f;\mathbf{A})$ is normalized so that it represents the relative pressure values along the array rather than the absolute levels. This normalization is done according to

$$\mathbf{R_n}(f;\mathbf{A})) = \frac{\mathbf{R}(f;\mathbf{A})}{\sqrt{\sum_{l=1}^{L}|R_l(f;\mathbf{A}))|^2}} = \frac{\mathbf{R}(f;\mathbf{A}))}{\|\mathbf{R}(f;\mathbf{A}))\|}. \tag{4.1}$$

The second component of MFP is the estimation of the sample covariance matrix (SCM). Measured data from the real environment are used to form an estimate of the covariance matrix of the true acoustic field. This estimate is obtained at frequency f by first taking the FFT of pressure time series snapshots recorded on the array, denoted as $\mathbf{p}_s(f) = [p_1(f), ..., p_L(f)]^T$, where s represents the sth snapshot. $\mathbf{p}(f)$ is then normalized through division by its norm in accordance with Eq. 4.1 to compute $\mathbf{p_n}(f)$. Following this, the SCM is estimated by averaging over N snapshots as

$$\mathbf{K_n}(f) = \frac{1}{N}\sum_{s=1}^{N}\mathbf{p}_s(f)\mathbf{p}_s^H(f), \tag{4.2}$$

where H denotes the conjugate transpose operator. Similar to the modeled replica vectors, $\mathbf{R_n}$, this matrix is complex valued - containing both the amplitude and phase information of the measured data. Its size is $L \times L$, where L is the number of sensors in the recording array. A few assumptions are made in computing this matrix. The statistics of the acoustics field are assumed to be Gaussian. This means $\mathbf{p}(f)$ is Gaussian and we further consider it to have a mean of zero. In addition, we assume that the expected value of $\mathbf{K_n}$ equals the true acoustic field covariance matrix. This is the case if the number of snapshots, N, that are averaged is large enough to adequately capture the temporal variance of the data.

The final component of MFP is the algorithm itself. For conventional MFP, the output is simply formed as the projection of the measured data onto the normalized

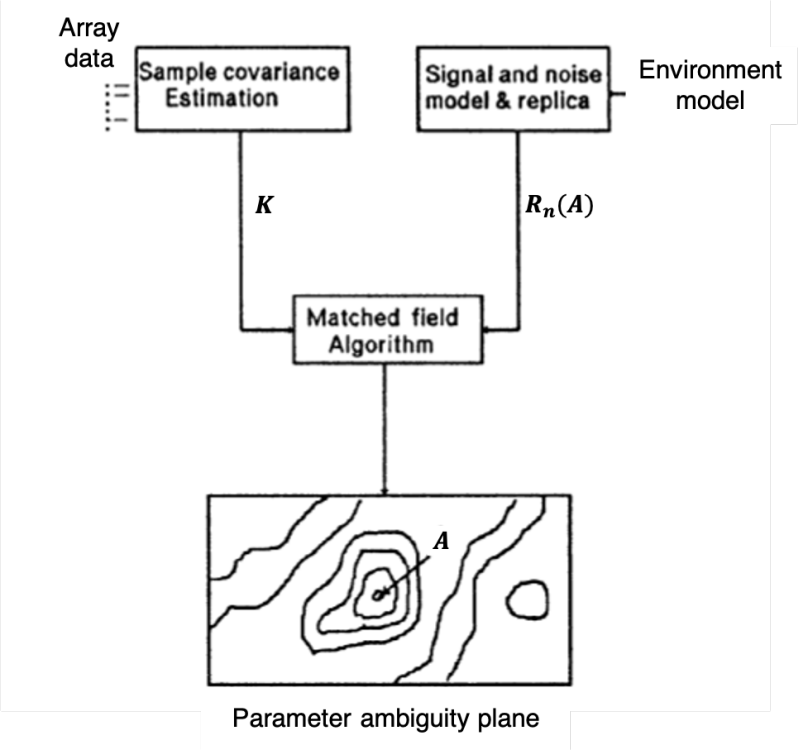


Figure 4.1: Components of a MFP algorithm [89].

replica vectors. This can be written in a quadratic form in terms of the SCM, $\mathbf{K_n}$, as

$$\mathbf{B}(f;\mathbf{A}) = \mathbf{R_n}^H(f;\mathbf{A})\mathbf{K_n}\mathbf{R_n}(f;\mathbf{A}). \tag{4.3}$$

$\mathbf{B}(f;\mathbf{A})$ represents an ambiguity plane over the parameters of interest showing the MFP output value vs. the location/environment prescribed by $\mathbf{A}$. The estimate for the parameters of interest, $\mathbf{A}$, is then made according to

$$\mathbf{A_{est}} = \underset{\mathbf{A}}{\operatorname{argmax}}(\mathbf{B}(f;\mathbf{A})). \tag{4.4}$$

Eq. 4.3 demonstrates how MFP essentially compares the measured acoustic data with a set of modeled replicas to invert for source or environmental parameters. Consequently, the accuracy of MFP is inherently sensitive to the exactness of the modeled environment and mismatch between the modeled and actual propagation

environments can degrade the algorithm's performance. With our CNN approach, we hope to devise a more robust method and improve upon MFP's sensitivity to environmental mismatch. In this chapter, conventional MFP is applied to estimate source range and it serves as a baseline for comparison with our proposed CNN approach.

4.2 Motivation for Model-based Convolutional Neural Network

4.2.1 Applications of Machine Learning

Broadly, machine learning (ML) refers to a class of statistical methods that automatically detect and utilize patterns within a dataset to make inferences on future or unseen data of the same type [115]. By same type, it is assumed that the existing (training) dataset and the unseen (testing) dataset are from similar statistical distributions. The applications of machine learning have enabled advances in numerous fields such as computer vision, image and signal processing, and physical sciences [116, 117, 118, 119, 120, 121, 122, 123].

Relevant to underwater acoustics, machine learning has been utilized to classify underwater targets [124, 125, 126] and identify marine mammals or other biologics [127, 128, 129]. It has also been applied extensively for the purpose of acoustic source localization as an alternative to MFP. Numerous works have proposed data-driven ML techniques that learn noise field features directly from collected data without the need for any environmental modeling, thus circumventing the MFP drawback with respect to modeling mismatch [130, 131, 132, 133]. Results from these studies demonstrate that data-driven ML methods are capable of performing on par or better than MFP when given adequate training data under a variety of environments. However, drawbacks exist as well for this approach; it is often limited by the impracticality of collecting enough acoustic data over a sampled space of source locations in order to build the required training dataset. This limitation makes data-driven ML potentially

costly to implement due to increased ship time and experiment logistics. Thus, in this study, we present a different ML approach that does not rely on training with real data.

4.2.2 Model-based Approach

Instead of data-driven ML, model-based ML for underwater source localization hopes to take advantage of the performance of ML but without the challenge of large scale data collection. Such an approach is similar to MFP in that it is trained using simulated data derived from acoustic modeling. This does mean that it, too, requires a certain degree of accuracy in modeling the propagation environment to perform well. However, many works that have examined model-based ML for source localization found that it can have improved performance compared to conventional MFP [134, 135, 136, 137, 138, 139, 140]. Thus, this approach shows promise as a compromise between performance and ease of data generation. Nonetheless, some research questions still remain:

1. How does the performance of model-based ML methods compare to MFP when tested on simulated environments outside the bounds of the originally modeled parameters? To what degree do they suffer the same environmental robustness issue as MFP?

2. Does the performance of model-based ML methods transfer to real data collected in the field? How does their performance compare to MFP in this case?

3. If model-based ML methods show improvement over MFP for questions 1 and 2, how may they be achieving their better performance?

In this chapter, we take an initial step to address these questions. We propose a model-based CNN approach to source-range estimation and test its performance against MFP in two separate environments with different types of environmental mismatch. All environmental modeling is done using OASES [35].

4.2.3 CNN Overview

Neural Networks

Neural networks (NN) are a type of machine learning algorithm that learns a non-linear mapping between input data and labeled outputs. They are *networks* because they have a graphical structure in which data is processed in a large number of simple equivalent components (nodes) arranged into successive layers [123, 141]. At the qth node within a layer of Q nodes , the input, typically a vector $\mathbf{x} \in \mathbb{R}^N$, is modified according to

$$z_q = g(\sum_{n=1}^{N} W_{nq} x_n + W_{q0}), \tag{4.5}$$

where W_{nq} and W_{q0} are the weights and biases applied to the input at this node, g is the activation function of the node (see following section) - which is typically applied element-wise, and z_q is the output of the node. Given that there are Q nodes, the output of this layer and the input to the next layer is a vector $\mathbf{z} \in \mathbb{R}^Q$. Following the last layer of the network, the output of the NN is denoted as $\mathbf{y} \in \mathbb{R}^P$. This output is compared with the true output of the pre-labeled training data using an objective function J (see following section). The weight and bias matrices within all of the layers are then updated through optimization of J using methods such as stochastic gradient descent [142]. Fig. 4.2 shows an example of a 3-layer, fully connected, NN [123]. The circles in the figure represent individual node units and the arrows signify information passage between nodes.

Activation and Objective Functions

There are many choices for the activation function within a network node depending on the task that the NN is designed to accomplish (*e.g.* classification, regression) and where a node is positioned in the NN. Relevant to our study, we present four different functions. They are the scale-exponential linear unit (SELU), sigmoid, linear, and softmax (Fig. 4.3).

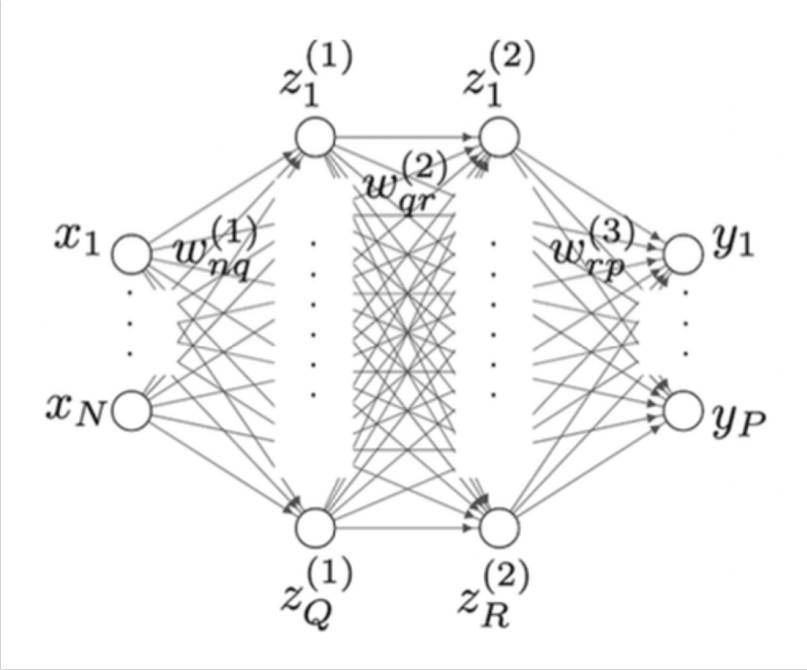


Figure 4.2: An example of a 3 layer neural network [123].

The SELU function is defined as:

$$y(x) = \begin{cases} \lambda x, & \text{if } x > 0, \\ \lambda\alpha(\exp(x) - 1), & \text{if } x \leq 0. \end{cases} \tag{4.6}$$

The parameters α and λ are pre-defined constants. Their values are $\alpha = 1.67326324$ and $\lambda = 1.05070098$ and are chosen to help standardize the input between consecutive neural network layers [143]. This is done to decrease the chance of vanishing or exploding gradients during gradient descent which are detrimental to network training. This function is typically used within nodes of the hidden (non-input or output) layers.

The sigmoid function is another popular activation function; it is a logistic function whose output ranges from 0 to 1. It is also often applied within hidden layers but can be used in the output layer for binary classification.

$$y(x) = \frac{\exp(x)}{\exp(x) + 1}. \tag{4.7}$$

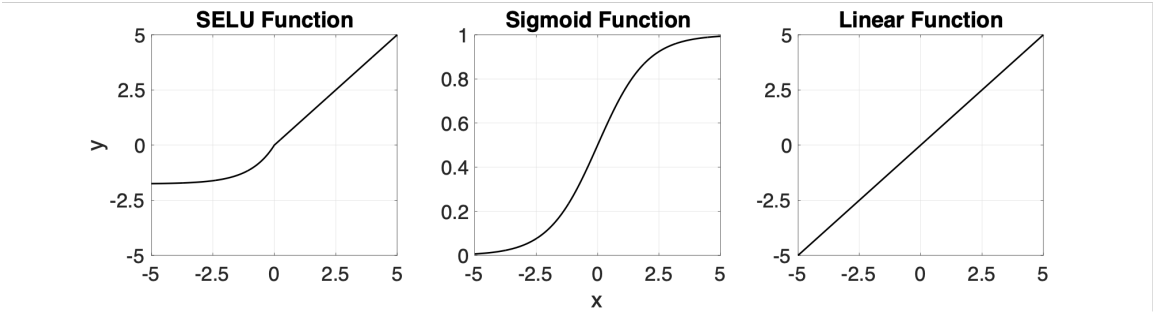


Figure 4.3: Activation functions used in this study for x from -5 to 5. The softmax function has a vector input and is not shown here.

The linear function is commonly used in the output layer of regression NNs. It is simply defined as:

$$y(x) = x. \tag{4.8}$$

The softmax function is typically used in the output layer of multi-class classification NNs. It takes in a vector input $\mathbf{x}$ of length K and outputs a probability distribution based on the exponential of each entry in the vector.

$$y_k(\mathbf{x}) = \frac{\exp(x_k)}{\sum_{j=1}^{K} \exp(x_j)}, \text{ where } k = 1, ..., K. \tag{4.9}$$

Similar to activation functions, the choice for the objective function, J, for a NN usually depends on the task at-hand. If the network is designed for classification, the typical objective function is the cross-entropy loss. This function is almost always used following a softmax activation in the output layer. It applies the probability distribution generated by the softmax to output a probability for each class in the classification problem. For example, for a problem with K classes, the cross entropy output for class k is

$$CE_k = -log(\frac{\exp(x_k)}{\sum_{j=1}^{K} \exp(x_j)}). \tag{4.10}$$

If the NN is designed to perform regression, the most common choice for J is likely

the mean-squared-error loss function, defined as

$$MSE = \frac{1}{N}\sum_{j=1}^{P}(y_{j,true} - y_j)^2, \tag{4.11}$$

where $\mathbf{y} \in \mathbb{R}^P$ is the NN output and $\mathbf{y}_{true} \in \mathbb{R}^P$ is the true output label of the training data.

CNN Specialization

CNNs [144, 145] are a special type of NN mainly used for problems involving image processing. Given the input to a CNN layer (say layer l) as a stack of Q images (feature maps), the qth feature map is denoted as $\mathbf{z}_q^{l-1}$. If the lth layer produces P feature maps as output, the relationship between $\mathbf{z}_q^{l-1}$ and the pth output feature map at layer l is described by

$$\mathbf{z}_p^l = g(\sum_{q=1}^{Q} \mathbf{W}_{pq}^l * \mathbf{z}_q^{l-1} + b_{pq}^l), \text{ for } p = 1, ..., P, \tag{4.12}$$

where $*$ denotes discrete convolution, $\mathbf{W}_{pq}^l$ are spatial filters of some defined size K (*e.g.* 3×3), b_{pq}^l are scalar biases, and g is an activation function [123]. The full output of layer l is then the stack of P feature maps, each denoted as a channel (Fig. 4.4).

Between successive convolutional layers, special layers are usually inserted into the network to help increase network robustness and prevent overfitting to training data. One type of such layer is a pooling layer, where input information is condensed by reducing the data resolution but preserving the number of channels. The most common variety of pooling is max pooling with a given filter size M, for which the values in a $M \times M$ grid of the feature map is replaced by their maximum. If the max pooling filter is applied without overlap, the input information is effectively reduced by a factor of M^2. The condensation of information helps the network train more efficiently while focusing on the most pronounced features in the input data.

Other kinds of special layers include dropout [146] and batch normalization [147]. The former randomly prohibits a proportion ($r = 0$ to 1) of the inputs from passing

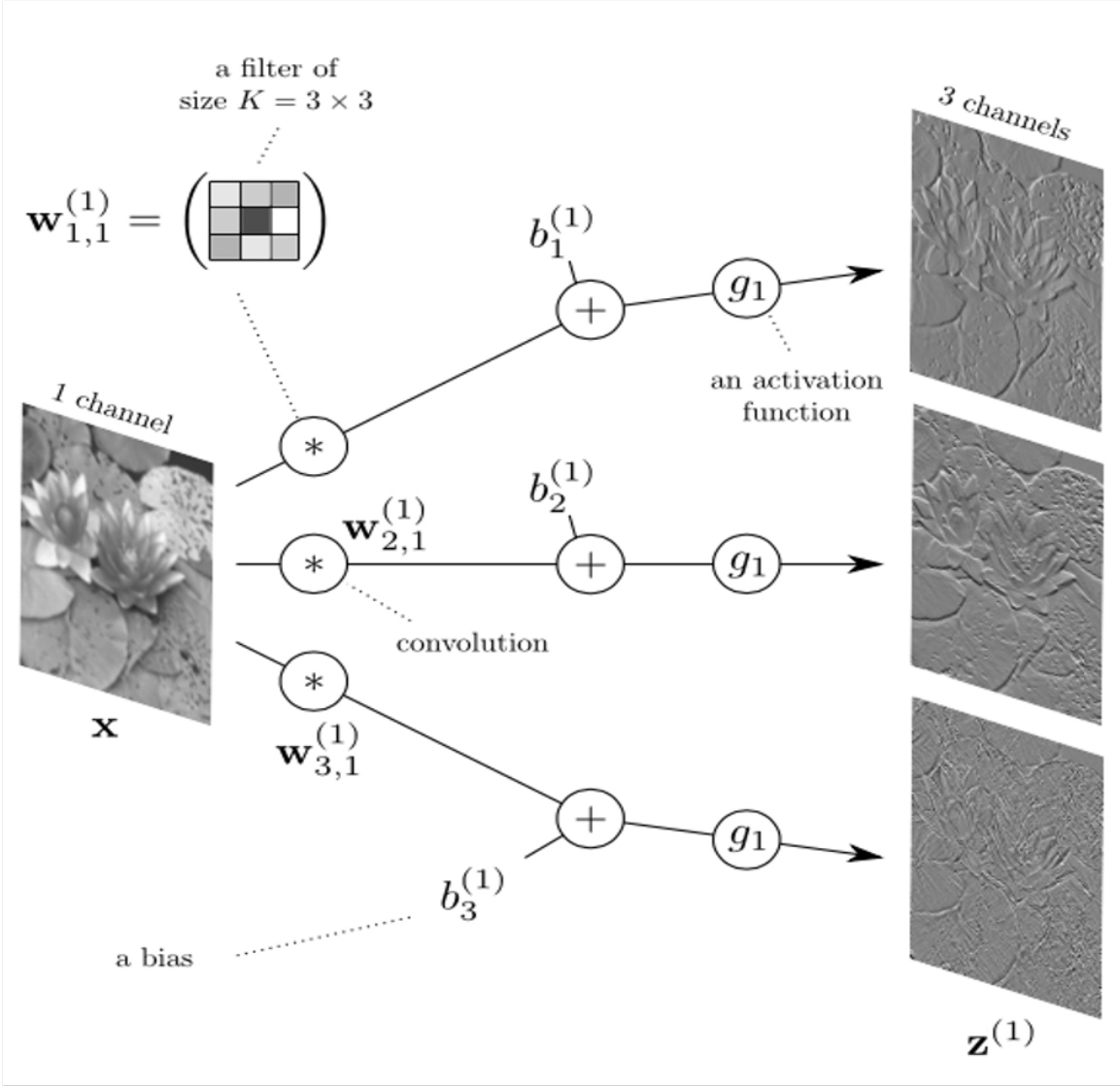


Figure 4.4: An example of a convolutional layer with 3 channels and a filter size of $K = 3 \times 3$ [123].

onto the next layer by setting them to 0. This essentially shuts down a random part of the network during each iteration (epoch) of training, forcing the network to not overly rely on any subset of nodes to make a prediction. The result is that dropout regularizes the network, helping it generalize better to data unseen in the training set. The batch normalization layer works by standardizing the input data to zero mean and unit variance to address the problem of covariate shift. Simply, the distribution of input values between successive network layers changes over time as the prior layer's filter weights change; this shift makes it extra difficult for the network to learn a mapping between input and output since the distribution of the input is constantly changing. Batch normalization alleviates this issue while also acting as another regularization tool. Each standardization procedure mildly perturbs the input data, which again helps to improve network generalization.

Following convolutional layers, CNNs will typically also include a few layers of regular neural network nodes before the final output layer. The connection between the convolutional and regular layers is made through a vectorization step, where the multi-dimensional output of the convolutional layer is reshaped into a vector in 1-D containing all of the values in the higher dimensional output before it is passed on to the regular layer.

For this study, CNN is used for source range estimation because the input data for this problem are SCMs of the measured field. These are represented as matrices of size $L \times L \times 2$, where L is the number of hydrophone sensors in the recording array and the 2 channels denote the real and imaginary parts of each SCM. Thus, source range estimation can be reframed into an image processing problem with the SCMs as input and the source ranges as output. As noted, CNN are specialized for this type of setup.

4.3 Network Architecture and Training

For both propagation environments that we examine in this chapter, we train two separate CNNs to perform source range estimation. One takes a classification approach

(CNN-c), while the other takes a regression approach (CNN-r). The general design of the two CNNs are the same except in their output layer activation function and training objective function. For classification, softmax activation is used in the output layer with the cross entropy objective function. For regression, linear activation is used in the output layer with MSE as the objective function.

The rest of the network architecture for both approaches consists of three convolutional layers with 16, 128, and 256 SELU activated filters in each layer, respectively. The sizes of the filters in each layer are presented in later sections as they differ between the two environments we studied. These layers are followed by a fully connected layer of 256 sigmoid activated nodes, and then the output layer. This results in five total layers in the networks. This configuration is selected after some initial empirical testing on the performance of various CNN architectures with greater and fewer layers of different types. Batch normalization is performed after each convolutional layer while dropout is performed after each convolutional ($r = 0.5$) and fully connected layer ($r = 0.25$). A schematic of the CNN architectures is shown in Fig. 4.5 (right).

The CNNs are implemented and trained in Python using the Keras and Tensorflow libraries [148]. Prior to input to the networks, the training dataset is first standardized to zero mean and unit variance. This pre-processing step helps to speed up network training [149]. The input data are then randomly segmented by an 80/20 split, where 80% is used for training and 20% is used for validation to see how well the networks perform on non-training data after each training epoch. To carry out stochastic gradient descent, the Adam optimizer [150] is used with a batch size of 128 and an initial learning rate of $\gamma = 0.0001$. γ subsequently decreases by 90% if the validation cost does not decrease for 75 epochs. Training stops if the validation cost does not decrease for 125 epochs to help prevent over-training.

To further optimize the CNN architectures, network pruning [151] is done to strip away under-activated filters. Such filters have small values for their weights, which means that the output after applying these filters is small compared to outputs from other filters. As such, these filters are under-activated because data that they are

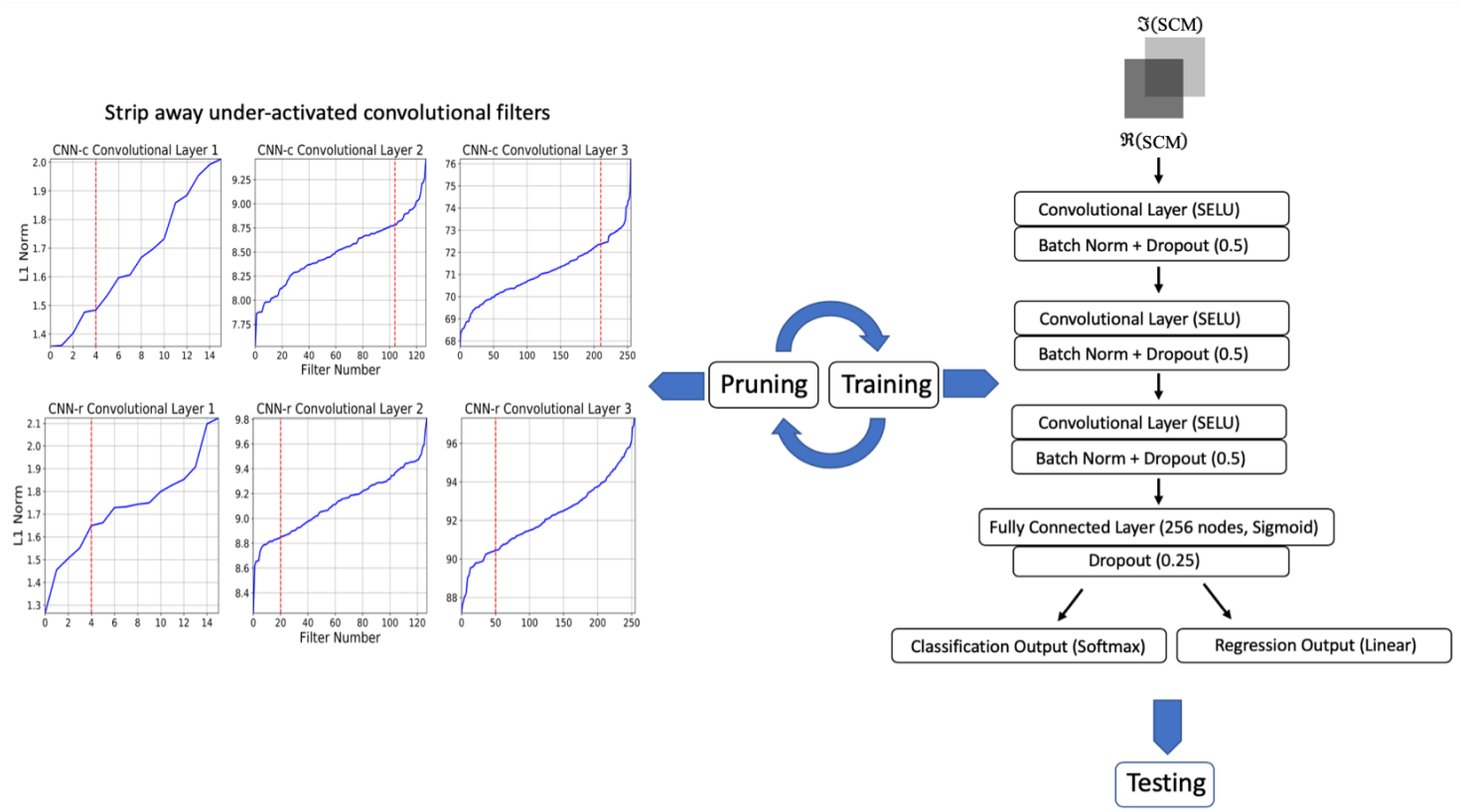


Figure 4.5: (Right) Architecture of CNNs trained for source range estimation. CNN-r and CNN-c differ only in the output layer. The number of filters in each convolutional layer depends on the pruning process. The size of convolutional filters depends on the environment case. (Left) $L1$ norms of CNN-c and CNN-r convolutional layer filters for the ICEX-16 environment (see Section 4.4). Filters to the left of the vertical line on each plot are deleted in the final reduced networks after successive rounds of pruning.

applied to are suppressed within the network and thus do not have a significant effect in the final network output. We identify and remove these filters by examining the sum of weight values (*i.e.* $L1$ norm) of all filters. This pruning process is as follows:

1. The original, full network is trained until stoppage.

2. The L1 norms of all filters in each convolutional layer of the trained network are plotted (example shown in Fig. 4.5, left); all filters whose $L1$ norm is small compared to the largest $L1$ norm value are deleted from the network. This can be done empirically during each training round or by setting a threshold such as removing filters with $L1$ norms in the bottom 10th percentile.

3. Training is continued on the updated, smaller network to re-adjust the weights of the kept filters, until stoppage again. The initial training rate is set as the same as when training last stopped.

4. Steps 2 and 3 are repeated until the validation accuracy of the reduced network decreases from that of the original, full size network.

These pruning steps reduce the trained networks' complexity by decreasing their number of parameters, making the final models more lightweight and better optimized. Following this procedure, the final network architectures for the two environments we examine are different. Thus, they are presented in later sections for each test case.

4.4 Test Case 1 - ICEX-16 Experiment

4.4.1 Simulated Data Tests

Training and Testing Data Generation

As presented in Chapter 3, the main dynamic factor affecting acoustic propagation in the region of the ICEX-16 experiment is the SSP, which is primarily dictated by the form of the Beaufort Lens. As a result, any viable model-based approach for surface

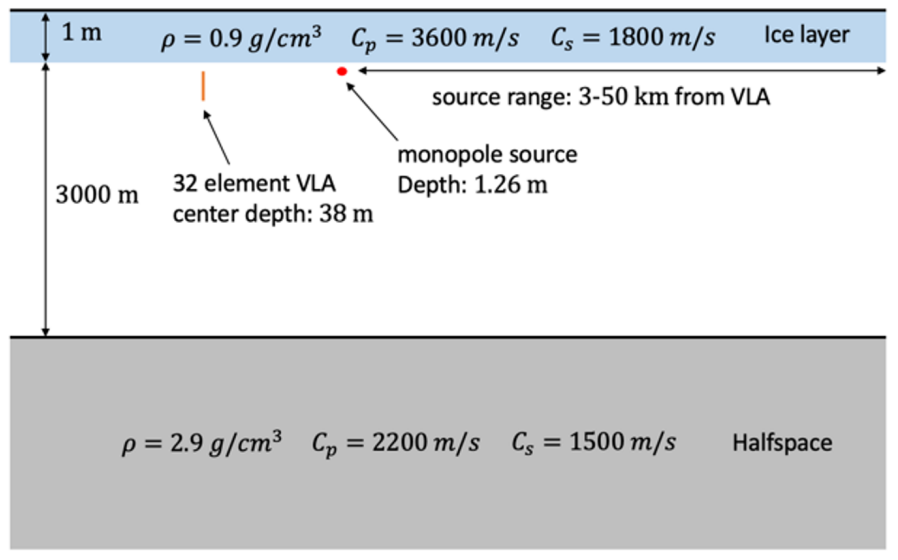
source range estimation in this region must show robustness to some degree of SSP mismatch. With this in mind, we begin with a comparison of the CNN approach performance to MFP under various amounts of simulated SSP mismatch. To model surface noise generation at various ranges in the ICEX-16 environment, we follow the setup presented in Table 3.1 and shown again in Fig. 4.6(a). The environment consists of a 1 m ice cover, 3000 m water column, and a solid bottom half space. To generate the training data for the CNN approach and the replica vectors for MFP[1], a single monopole source at 850 Hz is placed just below the surface (1.26 m) at 10 m increments between 3 to 50 km range from a recording VLA. The VLA is the same as the one deployed during ICEX-16, which consists of 32 omni-directional elements with nest 0.75 m and 1.5 m spacing. At each source range, the corresponding simulated SCM/replica vector is normalized and recorded. The SSP used within the water column is the profile measured during the ICEX-16 experiment, shown as the "original SSP" in Fig. 4.6(b).

For the CNN and MFP testing data, because we are comparing the two methods' robustness to SSP mismatch, we generate several testing datasets by applying various amounts of deviation to the original SSP in the simulated water column. These deviations are described by changes to the Beaufort Lens strength, which is defined as the sound speed difference between the local SSP maximum and the local minimum below. For each testing dataset, 1000 SCMs and replica vectors are simulated by placing the near surface monopole source at random ranges within the training interval of 3-50 km.

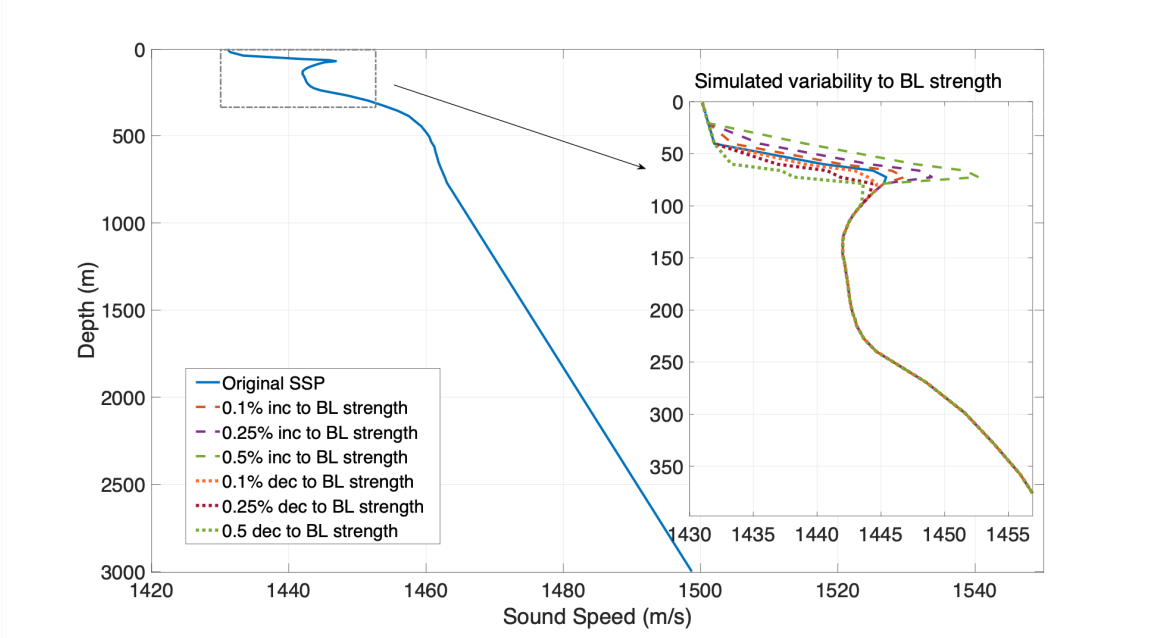
CNN Architecture

For the ICEX-16 environment, the filter sizes used in the three convolutional layers of the CNNs (CNN-c and CNN-r) are 3×3, 5×5, and 7×7, respectively. All filters in each layer are applied with a stride size of 2 to condense information from one layer to the next. We decided to increase filter size with network depth because some

[1]Recall the input to the CNNs are the data SCMs, which are equivalent to the replica vectors given to MFP because the vectors can be used to directly compute the SCMs. Thus, the information used to train the CNNs are exactly the same as the information given to MFP to perform matching.



(a)



(b)

Figure 4.6: (a) Simulation setup for generating training and testing datasets in the ICEX-16 environment. (b) Original, measured ICEX-16 SSP used to generate training dataset (solid line); SSPs with deviations to the Beaufort Lens strength used to generate testing datasets (dashed and dotted lines).

Table 4.1: ICEX-16 CNN architectures after iterative pruning. Table shows number of filters in each of the three convolutional layers and the total number of parameters in the trained networks.

	CNN-c	CNN-r
# of Conv. Filters	12; 24; 46	12; 108; 206
# of Parameters	275 009	1 968 687

initial testing demonstrated smaller filters in the first layer and larger filters in the deeper layers performed better than the reverse. We suspect this is because smaller filters capture more detail in the input SCMs that gets passed on to the later layers while the larger filters in the later layers help more with condensing information to pass onto the fully connected layer and generate an output. The rest of the CNN architectures and the training process follow the presentation in Section 4.3. After training and pruning, the final number of filters in each convolutional layer of the CNN-c model is 12, 24, and 46, respectively. These numbers for the CNN-r model are 12, 108, and 206, respectively, in each layer. The details of the pruned networks are shown in Table 4.1.

For the networks' outputs, CNN-c generates a prediction to 1 of 95 "classes" between 3-50 km range. These classes represent equal range intervals at every 500 m increment. For CNN-r, the predictions are simply a float number since the method takes a regression approach.

Performance Comparison

The performance of MFP and the two trained CNNs on the simulated testing datasets are compared using two metrics. One is the percentage of testing predictions that are within 1 km of the actual source range. This metric reflects how accurate each individual prediction is to the corresponding correct range value. The other metric is mean-absolute-error (MAE) and reflects the averaged error over all predictions for a

testing dataset. This metric is formally defined as

$$MAE = \frac{1}{M}\sum_{i=1}^{M}|Prediction_i - Actual_i|, \tag{4.13}$$

where M denotes the number of entries in a testing dataset and the expression within the absolute value is the difference between the predicted and actual source range for entry i.

Figs. 4.7 and 4.8 show the performances for MFP and the CNNs on testing data with varying degrees of SSP mismatch. As expected, as the magnitude of Beaufort Lens strength deviation increases, the performances of all three methods decrease by both metrics. However, CNN-c and CNN-r show improvement over MFP with SSP mismatch while performing similarly to MFP with no mismatch (Fig. 4.7). At the maximum amount of mismatch, the percentage of predictions within 1 km of the actual distance is ∼3-20 points higher for the CNNs than MFP; meanwhile, the MAE over all predictions is ∼2.5-5 km lower for the CNNs than MFP.

The panels within Fig. 4.8 further reveal that there is high variability in the MFP predictions. This means that close-to-correct predictions can often be extremely accurate while for incorrect predictions, the margins of the mistakes can be quite large. In contrast, the CNN methods, particularly CNN-r, show lower variability in their predictions. As a result, although the accuracy of any individual CNN prediction may not be as high as the corresponding MFP prediction, the overall predictions are more consistent with ground truth. Thus, the CNNs appear to gain robustness to mismatch over MFP by trading off individual data-point prediction accuracy for overall prediction consistency. Furthermore, the reason that CNN-r is the more prominent example of this trade-off over CNN-c is likely because of the difference in the objective function used during their training. For CNN-r, training focuses on minimizing the MSE loss, which inherently leads to more consistent predictions and lower overall error than predictions derived from categorical association, as is the case for CNN-c. Further discussion on how the CNNs achieve their robustness is presented in Section 4.5. Another point to note is that while the individual predictions of all methods

could have large errors, in practice, noise range is typically done through a moving average of predictions (see Fig. 4.9) or through a tracker system [152] that ensures a smooth, realistic source track.

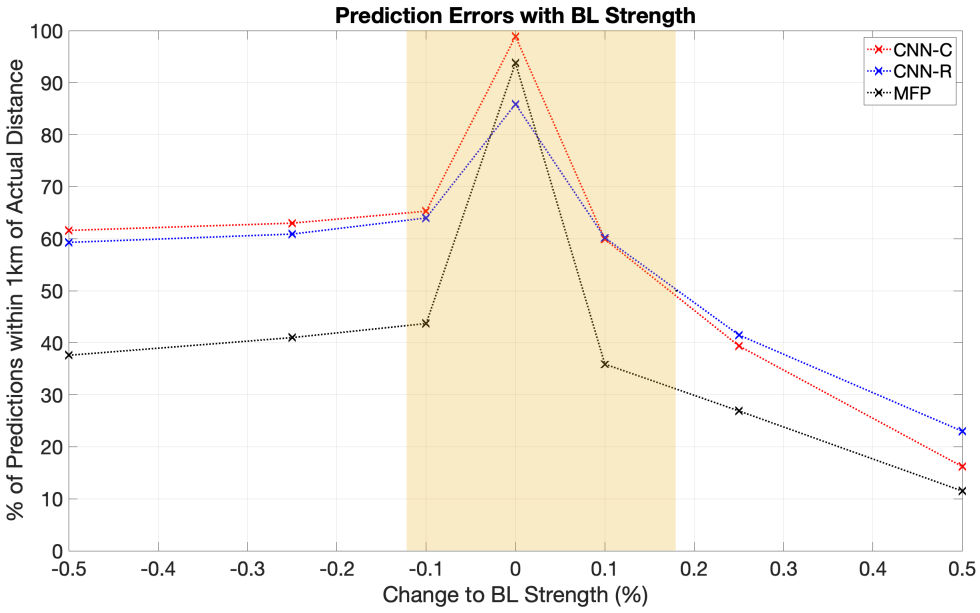
4.4.2 Real Data Test

Testing Data Generation

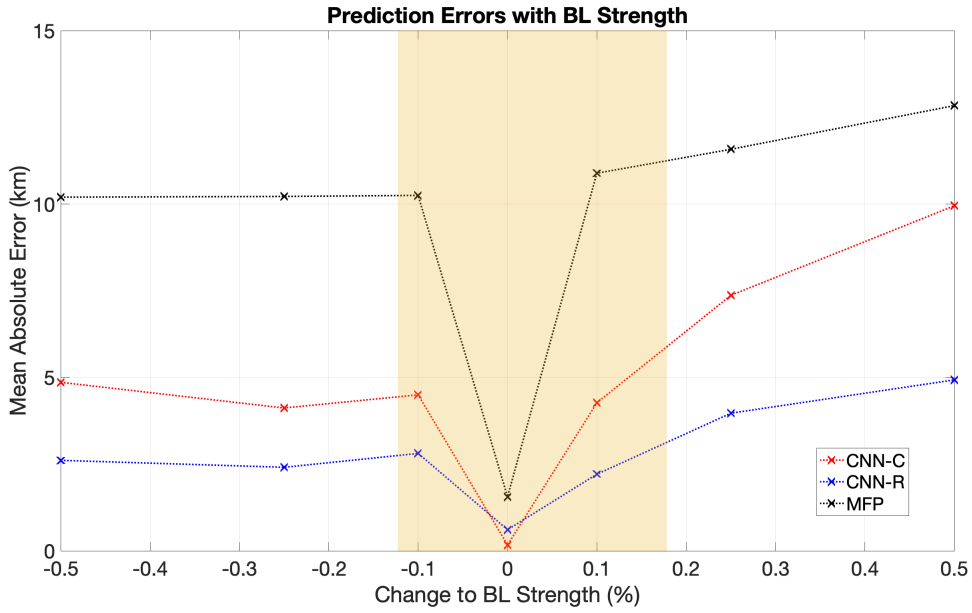
The trained CNNs are also applied to real data collected during ICEX-16 to see if their predictive capabilities transfer and how they perform compared to MFP. The CNNs are not re-trained with real data in this test; they are the same CNNs from before, trained using simulated data. As described in Chapter 3, ∼8 hours of ambient noise data were collected on March 13, 2016 using a 32 element VLA with a sampling frequency of 12000 Hz; the recorded noise was generated by the ice cover ∼30-50 km away from the experiment site. To prepare the testing dataset, the collected data is segmented into 10240-point Hanning windows with 50% overlap. The SCM averaged over 800-900 Hz is then calculated for each data snapshot. This frequency interval is selected because the simulated training data were generated for a 850 Hz source. The SCMs over every 32 snapshots are averaged to form the testing dataset. Although not shown in the results, longer snapshot averages were tried as well and the prediction outputs for all approaches were consistent across all averaging lengths.

Performance Comparison

Fig. 4.9 presents the source range estimations of MFP, CNN-c, and CNN-r on the ICEX-16 ambient noise dataset. The dots show the individual predictions for each snapshot window while the solid line represents the 10-minute moving average of the predictions. From the top plot, MFP predictions deviate up to ∼20 km from the expected range of noise generation (∼30-50 km). In contrast, the CNN-c and CNN-r predictions are more consistent with the expected range. This is especially true for CNN-r, whose predictions show very little variability and largely remain between ∼25-40 km. This result matches the observations from the simulated data tests where

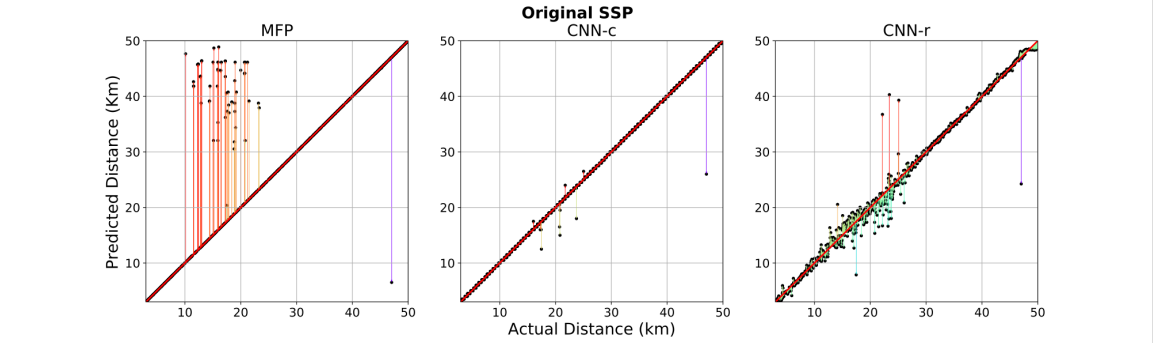


(a)

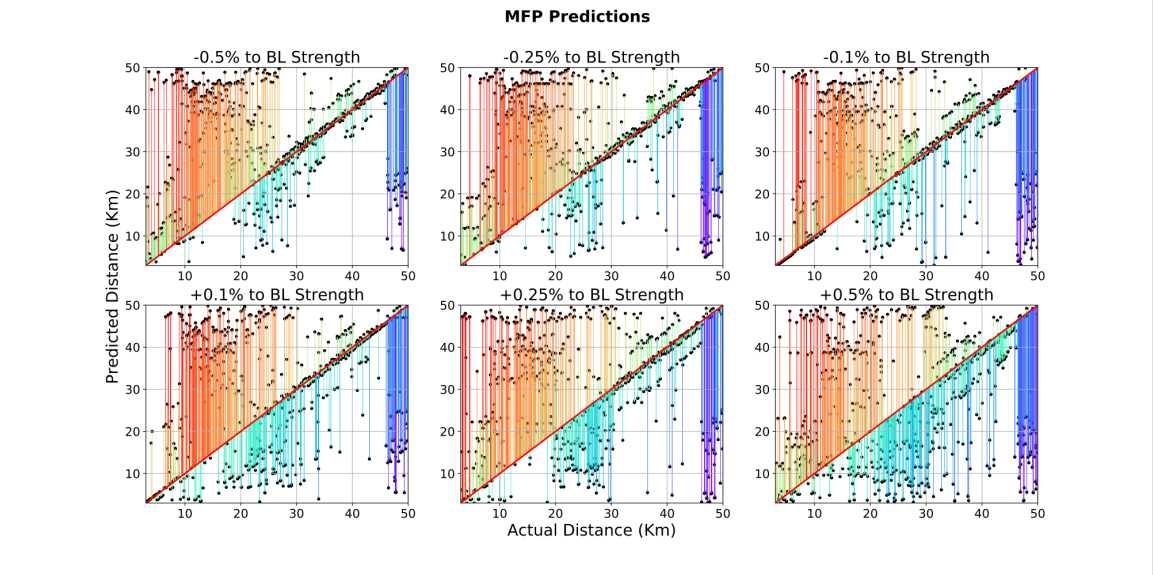


(b)

Figure 4.7: Performance metric comparison between MPF, CNN-c, and CNN-r with varying amounts of deviation to the original ICEX-16 SSP. The shaded interval represents the ± 2 standard deviation values of the amount of natural variability observed in all measured Beaufort Sea SSPs presented in Fig. 3.4(b) - this demonstrates that the maximum amount of SSP variability introduced to generate the testing datasets exceeds the amount observed in the measured data. By both metrics, CNNs show similar performance to MFP without any SSP mismatch but improved performance with SSP mismatch.

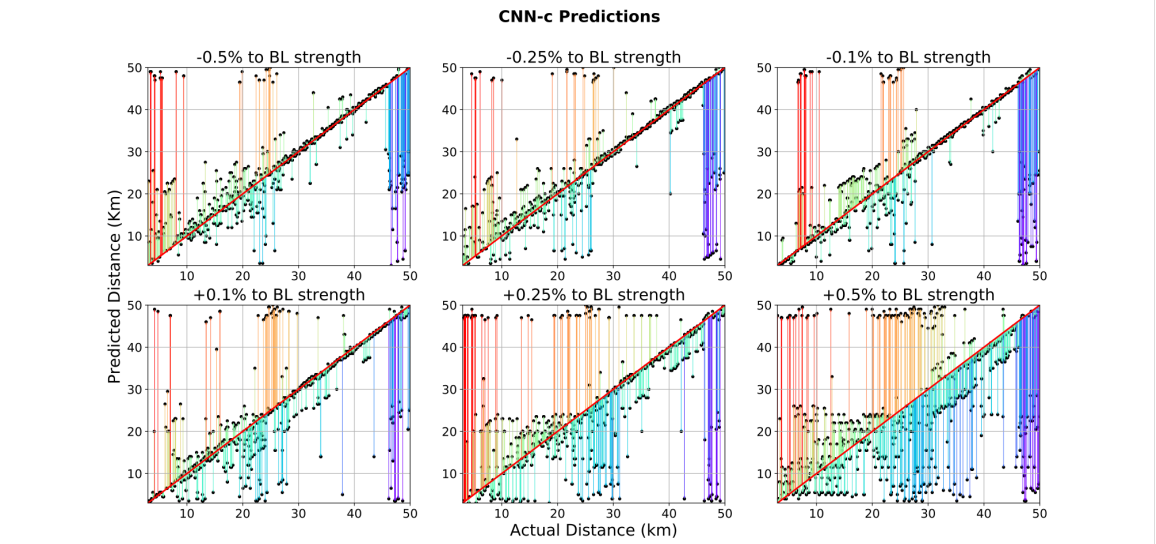


(a)

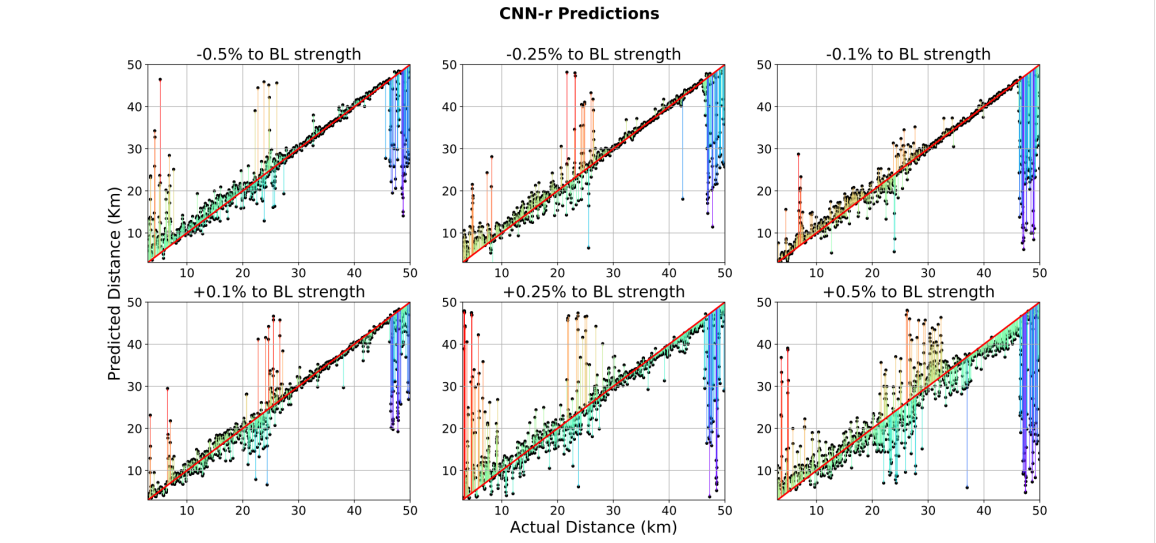


(b)

Figure 4.8: Diagonal solid line represents ground truth. Dots show prediction values. Quivers represent the difference between each prediction and the corresponding ground truth value. (a) Performance of MFP, CNN-c, and CNN-r on testing data generated using the original ICEX-16 SSP. (b)–(d) Performance of each approach on testing data generated with deviations to the original ICEX-16 SSP.



(c)



(d)

Figure 4.8: Diagonal solid line represents ground truth. Dots show prediction values. Quivers represent the difference between each prediction and the corresponding ground truth value. (a) Performance of MFP, CNN-c, and CNN-r on testing data generated using the original ICEX-16 SSP. (b)–(d) Performance of each approach on testing data generated with deviations to the original ICEX-16 SSP. *(Continued).*

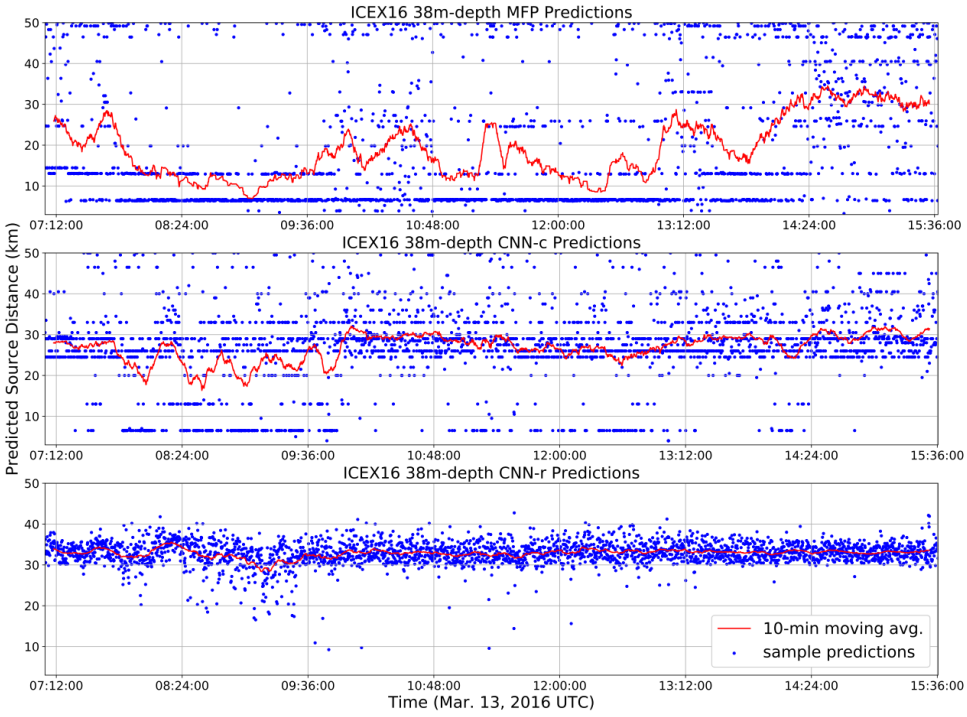


Figure 4.9: Range predictions from MFP, CNN-c, and CNN-r for ambient noise data collected on March 13, 2016. Dots represent individual predictions; solid lines denote 10-minute moving averages.The expected range of noise generation is ∼30-50 km.

the CNNs' outputs also show less variability and more robustness to mismatch than MFP. Consequently, the CNNs performance gain over MFP on simulated data does indeed transfer to real data. Furthermore, these model-based CNNs do have utility as an alternative to MFP and data-driven ML approaches. Another interesting observation from Fig. 4.9 is that for CNN-c, some of the predictions outside of the expected ground truth range match those made by MFP in that they also cluster around 5 and 15 km. This similarity suggests the CNNs may be arriving at their prediction outputs in a similar manner as MFP. However, the CNNs have fewer of these incorrect, outlier predictions. This means the CNNs are somehow better guarded against making incorrect predictions than MFP. More discussion on how this may be achieved is presented in section 4.5.

4.5 Insight into CNN Robustness

In this section, we explore in more detail how the CNNs achieve their more robust performance compared to MFP. To do this, it is helpful to examine an intermediate output of the networks. In MFP, source range predictions are made by essentially comparing the measured data replica vector (in the form of data SCM) with the template of modeled replica vectors. Analogous to the MFP template replica vectors for the CNN approach would be the pre-prediction output vectors of the fully connected (FC) layer[2] in the trained CNN-c and CNN-r networks. The MFP replica vectors consist of 32 complex-valued entries while the CNN FC-layer template vectors are 256-element, real-valued vectors[3]. Both sets of vectors are used in the last calculation step in their respective approach before a prediction is outputted. We use the ICEX-16 environment with the source at 33 km as a specific example and demonstrate how the MFP, CNN-c, and CNN-r vector sets are affected by different amounts of SSP mismatch. For this demonstration, we first need to define a quantitative measure to describe how different one vector is from another. We adopt the Euclidean distance between two vectors, $\vec{v}_i$ and $\vec{v}_j$, which is defined as

$$Distance = \sqrt{\sum_{n=1}^{N} [\overline{(\vec{v}_i[n] - \vec{v}_j[n])} * (\vec{v}_i[n] - \vec{v}_j[n])]} \tag{4.14}$$

where $^-$ denotes the complex conjugate operator and N is the length of the vectors.

Taking a look at MFP first, Fig. 4.10(a) shows the normalized Euclidean distance between the MFP template replica vectors with a data replica vector simulated with the source at 33 km under no SSP mismatch (0% change to Beaufort Lens strength). The normalization is accomplished by dividing by the maximum distance between the data vector and every vector in the template set. Unsurprisingly, because there is no SSP mismatch, the data replica vector is exactly the same as the MFP template replica vector with source at 33 km. Thus, the Euclidean distance between the two vectors is 0. Away from the correct source distance of 33 km, the Euclidean distance

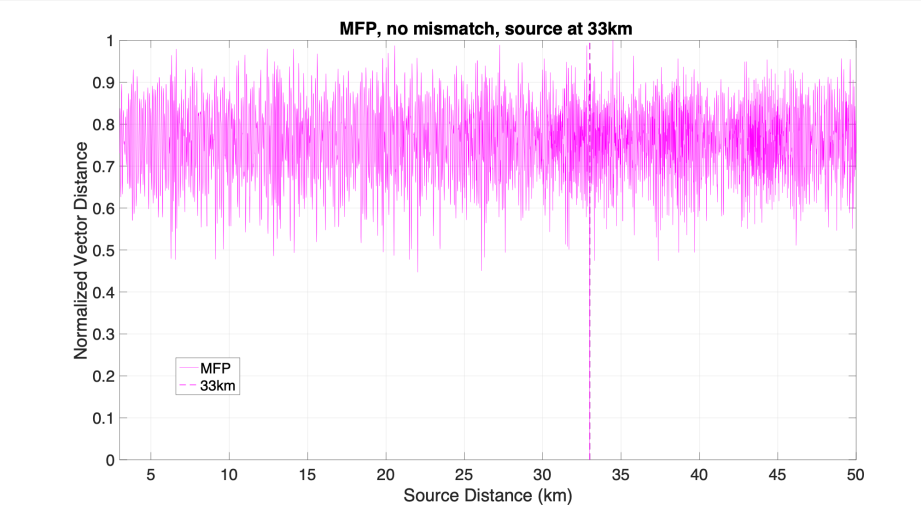
[2]The fully connected layer is the last layer before the output layer of the CNNs.

[3]The length of the vectors match the number of nodes in the FC-layer.

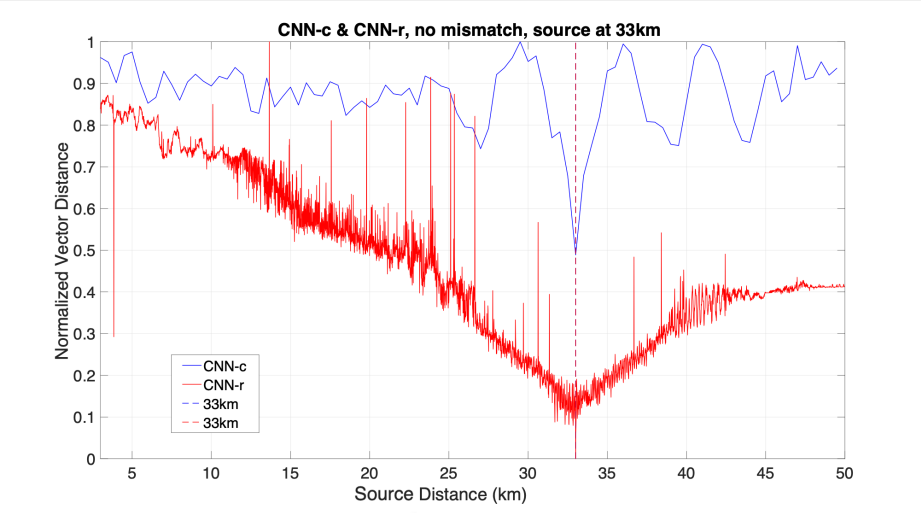
values increase very rapidly such that there is a sharp and narrow minimum at 33 km[4]. Because of this steep minimum, it is obvious from this plot that MFP should output 33 km as the correct prediction, which it does, as shown by the dotted vertical line in Fig. 4.10(a). However, when SSP mismatch is introduced, the correct output becomes much less obvious in the Euclidean distance plot. Figure 4.11(a) shows when the data replica vector is generated under SSP mismatch (again with source at 33 km), the normalized Euclidean distance between the data replica vector and the MFP template vector for source at 33 km grows to about the same value as that of any other vector in the template set - nearly all Euclidean distance values in Fig. 4.10(a) are between 0.75 and 1 and there is no longer a steep and obvious minimum. While MFP may still output a prediction close to the correct answer in this case (as shown by plots for 0.1% inc., 0.25% inc., and 0.25% dec. to Beaufort Lens strength), it is also more likely than the no mismatch case to output a very inaccurate prediction (as is the case for 0.5% inc., 0.1% dec., and 0.5% dec. to Beaufort Lens strength). Thus, we can view the Euclidean distance metric as a proxy for predictive confidence. When there is no SSP mismatch, MFP has much higher confidence that the correct source range is 33 km than any other range value. However, when mismatch is introduced, MFP's predictive confidence for the correct source range decreases significantly more compared to other range values. As a result, the MFP prediction may become very inaccurate.

Now we examine the Euclidean distance plots for CNN-c and CNN-r. Similar to the MFP case, we first plot the distance between the CNN FC-layer template vectors and the FC-layer output vector when the source is at 33 km under no SSP mismatch. Fig. 4.10(b) shows that for both CNN-c and CNN-r, the minimum Euclidean distance

[4]Besides the single minimum, there is an apparent lack of comparable "sidelobes" in this plot partly because the y-axis in not in dB scale - halving a value in linear scale leads to a decrease of just 3 dB in log scale. This was done to better showcase the steep minimum in this case. Furthermore, the metric of Euclidean distance is different than the conventional correlation metric used when presenting MFP ambiguity maps. Whereas the correlation metric varies between 0 to 1, the Euclidean distance can span a much larger interval prior to normalization. Because of this greater span, this distance metric is specifically chosen since it better demonstrates the disparity between the strong match at the correct source range compared to the much weaker match at the incorrect ranges under the no mismatch case for MFP.

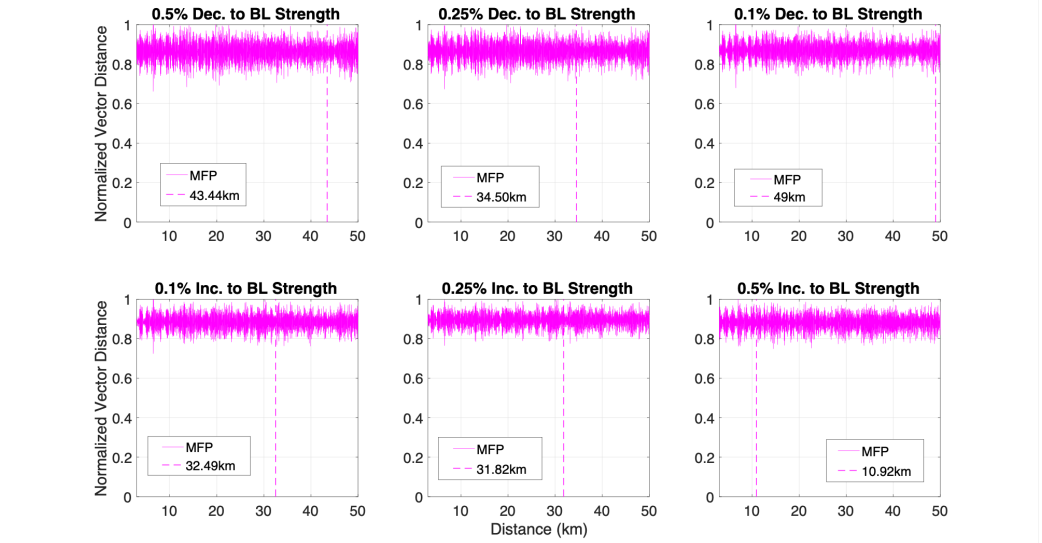


(a)

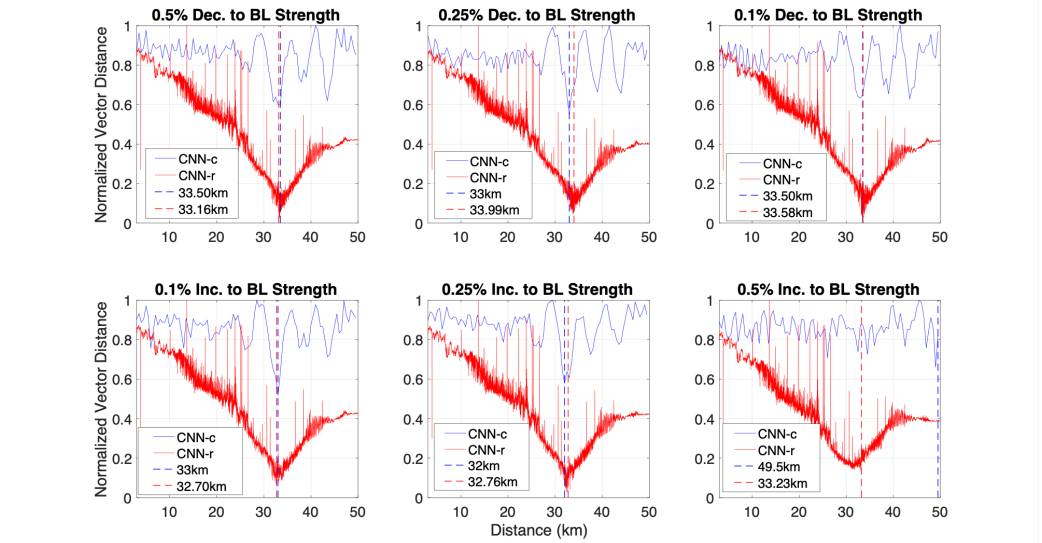


(b)

Figure 4.10: Plots are generated from simulated data in the ICEX-16 environment with no SSP mismatch. (a) Normalized Euclidean distances between replica vectors in the MFP template set and data replica vector with source at 33 km. By definition, the normalized distance of the vector with itself is 0 (at 33 km); away from 33 km, the distance values quickly increase from 0, forming a sharp and narrow minimum. Dotted line shows MFP prediction for this case. (b) Normalized Euclidean distances between CNN-c (blue) and CNN-r (red) FC-layer template vectors with their respective FC-layer output vector with source at 33 km. Note, the normalized distance of the vector with itself (at 33 km) is not 0 for CNN-c because there are multiple samples in each output class; the distance shown in the figures represents the averaged distance between an input with every sample in each output class. The CNN plots show more gradual increase away from the Euclidean distance minima than the plot for MPF. Dotted lines show CNN-c and CNN-r predictions for this case.



(a)



(b)

Figure 4.11: Plots are generated from simulated data in the ICEX-16 environment with varying amounts of SSP mismatch. (a) Normalized Euclidean distances between replica vectors in the MFP template set and data replica vector with source at 33 km. With SSP mismatch, the expected minimum at 33 km increases to around the same value as at any other range. As a result, the MFP predictions (dotted lines) may become very inaccurate. (b) Normalized Euclidean distances between CNN-c (blue) and CNN-r (red) FC-layer template vectors with their respective FC-layer output vector with source at 33 km. The CNN plots retain their minima near the correct range (33 km) much better than the MFP plots under mismatch. As a result, their predictions remain consistent and closer to the correct range value.

occurs at the correct source range of 33 km. However, different from the MFP plot (Fig. 4.10(a)), away from the correct range, the increase in Euclidean distance is more gradual. This is especially true for the CNN-r plot. The more gradual increase away from the minimum means, unlike MFP template replica vectors, CNN FC-layer template vectors for neighboring source range inputs are also closer to each other in Euclidean space. For example, the CNN FC-layer vector for source at 33 km is closer to the vector for source at 32.5 km in Euclidean space than it is to the vector for source at 10 km. This neighboring property of the CNN FC-layer template vectors may be what increases the CNNs' robustness compared to MFP. Given the same amount of environmental mismatch which causes a slight change in the data input, the CNNs are more constrained from outputting a prediction drastically different from the true value than MFP because their FC-layer template vectors closest to the correct vector also represent source ranges near the correct range value. Of course, if the mismatch causes a large enough change to the data input, the CNNs are not immune from making a very inaccurate prediction (as is the case for CNN-c under 0.5% inc. to Beaufort Lens strength). The CNNs' increased robustness can be seen in Fig. 4.11(b), which shows their Euclidean distance plots under varying degrees of SSP mismatch. These plots retain their minima near the correct range (33 km) much better than the MFP plots (Fig. 4.10(b)). Although the CNN predictions in these cases (dotted lines) are not exactly equal to 33 km, they are very close to the correct value and remain more consistent between different amounts of mismatch than MFP. However, the more gradual increase from the Euclidean distance minima also suggests that the CNNs are less certain of the exact correct source range compared to MFP when the environment is precisely modeled (MFP has a very sharp and narrow minimum in this case). Thus, this appears to be the trade-off for the CNNs' improved performance when mismatch is present.

4.6 Test Case 2 - SWellEx-96 Experiment

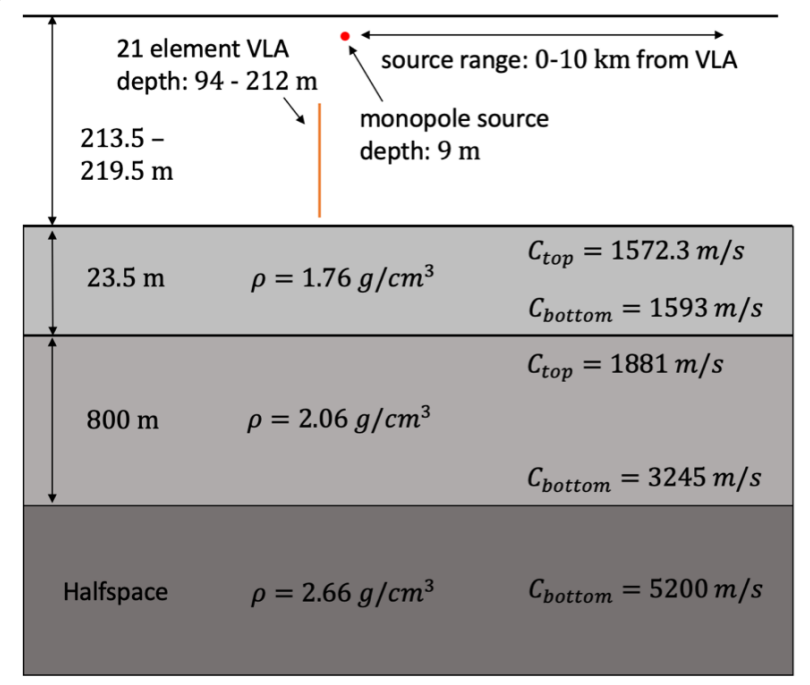
As a second test case for our model-based CNN approach and to further demonstrate its utility, we apply the method to perform source range estimation in another propagation environment. This environment is modeled after the site of the SWellEx-96 experiment [153] off the coast of Point Loma, CA between May 10-18, 1996. As part of this experiment, a 109 Hz source was towed below a vessel at 9 m depth while transmitting and the acoustic data were recorded on a VLA. The modeled environment of the experiment site consists of a 216.5 m water column on top of three solid bottom layers with increasing density (Fig. 4.12(a)). The SSP used in the simulated environment is the average of all profiles collected during the experiment. The simulated array matches the VLA deployed during SWellEx-96 and contains 21 evenly spaced, omnidirectional elements between $94.125 - 212.25$ m depth (Fig. 4.12(b)).

4.6.1 Simulated Data Tests

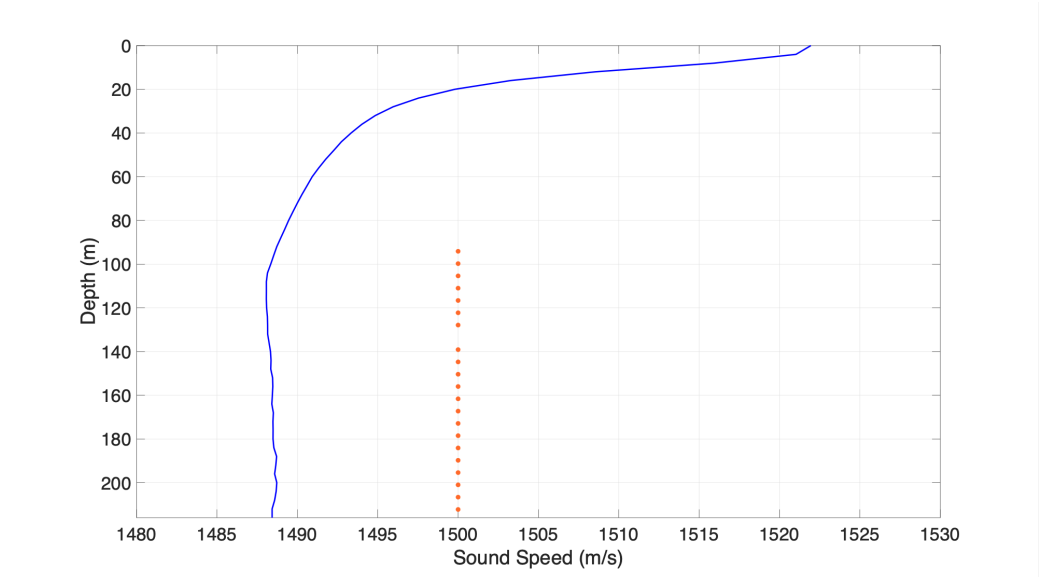
Training and Testing Data Generation

At the SWellEx-96 experiment site, the main environmental variability results from the depth of the ocean bottom as the bathymetry around the VLA varies from ∼150-270 m. Thus, we use this environment to test the robustness of the CNN approach to ocean bottom depth mismatch. To generate the training dataset, the ocean bottom is set at 216.5 m depth in a range-independent model. A 109 Hz monopole source is placed 9 m below the ocean surface and 0-10 km away from the VLA at 10 m increments (Fig. 4.12(a)). At each source location, the SCM recorded on the simulated VLA is normalized and added to the training dataset.

Four testing datasets are generated with ocean bottom depth set to 213.5, 215.5, 217.5, and 219.5 m, respectively. For each, the source is placed at 500 random ranges within the training interval.



(a)



(b)

Figure 4.12: (a) Simulation setup for generating training and testing datasets in the SWellEx-96 environment. To generate the training dataset, ocean bottom depth is set at 216.5 m. To generate the testing datasets, ocean bottom depth is varied between 213.5-219.5 m at 2 m increments. (b) SSP used to generate training and testing datasets and VLA element locations.

Table 4.2: SWellEx-96 CNN architectures after iterative pruning. Table shows number of filters in each of the three convolutional layers and the total number of parameters in the trained networks.

	CNN-c	CNN-r
# of Conv. Filters	6; 38; 40	16; 128; 256
# of Parameters	162 433	1 463 025

CNN Architecture

Compared to the data SCMs for the ICEX-16 test case, the SCMs for SWellEX-96 are smaller in size. The former are size 32×32×2 while the latter are 21×21×2. This reduction is caused by the difference in the number of hydrophone elements in the VLAs of the two experiments (32 vs. 21). Due to the smaller SCM input dimensions, we chose a slightly smaller filter size for the last convolutional layer in the CNNs for the SWellEx-96 case; the filter sizes for the layers are set to 3×3, 5×5, and 5×5, respectively, for the three convolutional layers. The rest of the CNN architectures and the training process follow the presentation in Section 4.3. After the training and pruning process, the final number of filters in each convolutional layer of the CNN-c model is 6, 38, and 40, respectively. These numbers for the CNN-r model are 16, 128, and 256, respectively, in each layer. Note, for CNN-r, the network architecture did not decrease after pruning as any reduction in the number of filters caused a decrease in performance. The details of the pruned networks are shown in Table 4.2.

For the networks' outputs, CNN-c generates a prediction to 1 of 101 "classes" between 0-10 km range. These classes represent equal range intervals at every 100 m increment. For CNN-r, the predictions are simply a float number since the method takes a regression approach.

Performance Comparison

Figs. 4.13 and 4.14 compare the performance of MFP to CNN-c and CNN-r based on the two metrics introduced in Section 4.4.1. These plots show a similar trend

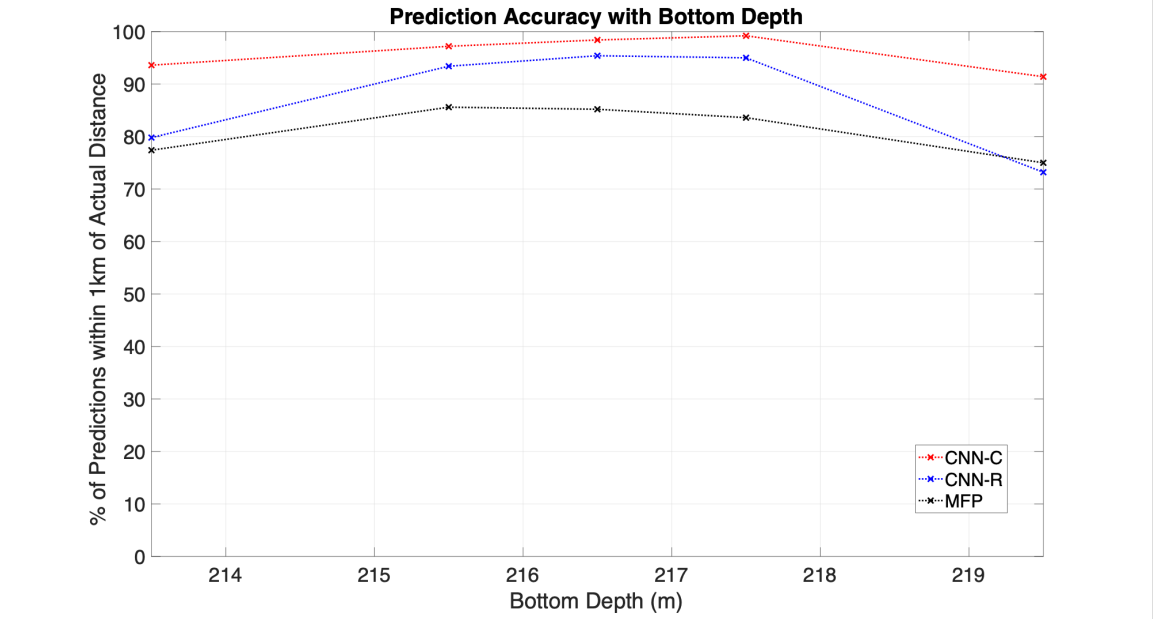
compared to the simulated data test results from the ICEX-16 environment. However, in this case, the CNN-c approach shows a clear improvement in performance compared to the other two methods by both metrics for all bottom depths (Fig. 4.13). At the maximum amount of mismatch, the percentage of predictions within 1 km of the actual distance is ∼15-20 points higher for CNN-c than MFP; meanwhile, the MAE over all predictions is ∼0.75-1 km lower for CNN-c than MFP.

Comparing CNN-r to MFP, the results again demonstrate CNN-r trades off accuracy of individual predictions for lower overall error. However, for this environment, this trade-off may be overdone. While the MAE of CNN-r is lower than MFP for all bottom depths (Fig. 4.13), Fig. 4.14 shows CNN-r typically has a larger error margin than MFP when comparing individual predictions, especially as the bottom depth mismatch increases from 216.5 m. Thus, for this environment, the CNN-r approach may not be a more robust alternative to MFP while CNN-c shows promise.

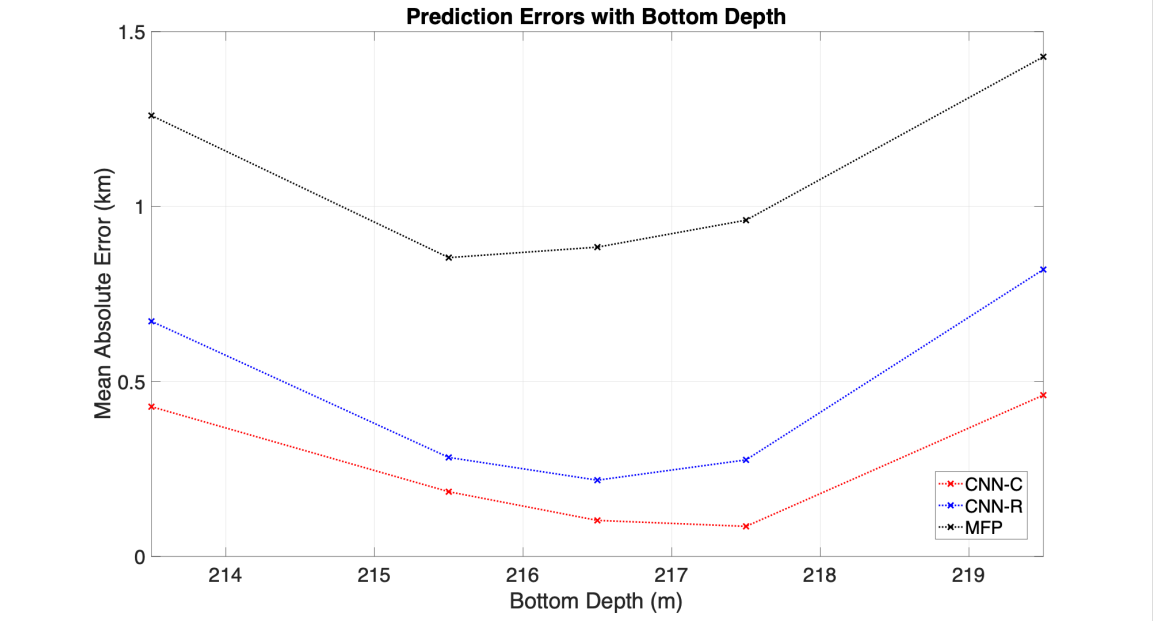
4.6.2 Real Data Test

Testing Data Generation

Acoustic recording by the 21 element VLA (sampling frequency = 1500 Hz) from the SWellEx-96 experiment is used to test the performance of our trained CNNs on real data from this environment. Again, the CNNs are trained with simulated data generated with ocean bottom depth equal to 216.5 m. As part of the SWellEx-96 experiment, a 9 m deep source emitting at 109Hz was towed by a vessel along the blue track shown in Fig. 4.15(a). While the recording VLA was deployed at a location with ocean bottom depth of 216.5 m, the bathymetry along the source track varied between ∼180-220m depth. Thus, there is mismatch between the simulated training environment and the real testing environment. Similar to the processing of the ICEX-16 data, the SWellEx-96 dataset is segmented into Hanning windows, in this case of size 512, with 50% overlap. The SCM averaged over 108.5-109.5Hz is then calculated for each data snapshot. Following this, the SCMs over every 25 snapshots are averaged to create the testing dataset.

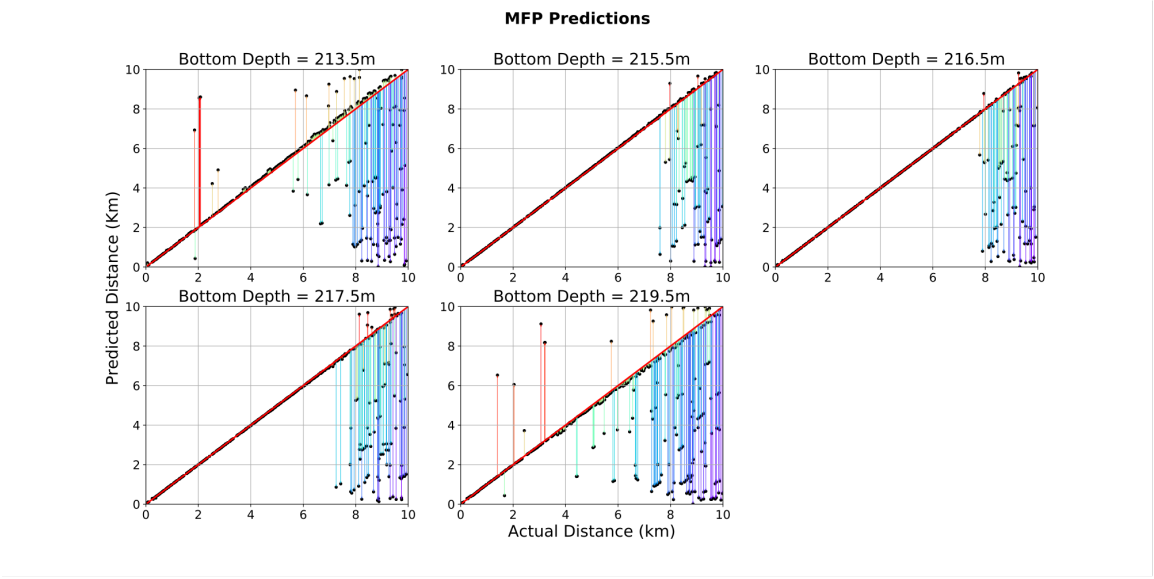


(a)



(b)

Figure 4.13: Performance metric comparison between MFP, CNN-c, and CNN-r with various bottom depths. CNN-c shows improved performance to MFP in all cases by both metrics. CNN-r shows similar or improved performance over MFP in all cases by both metrics; however, it still may not be preferable to MFP based on results shown in Fig. 4.14.



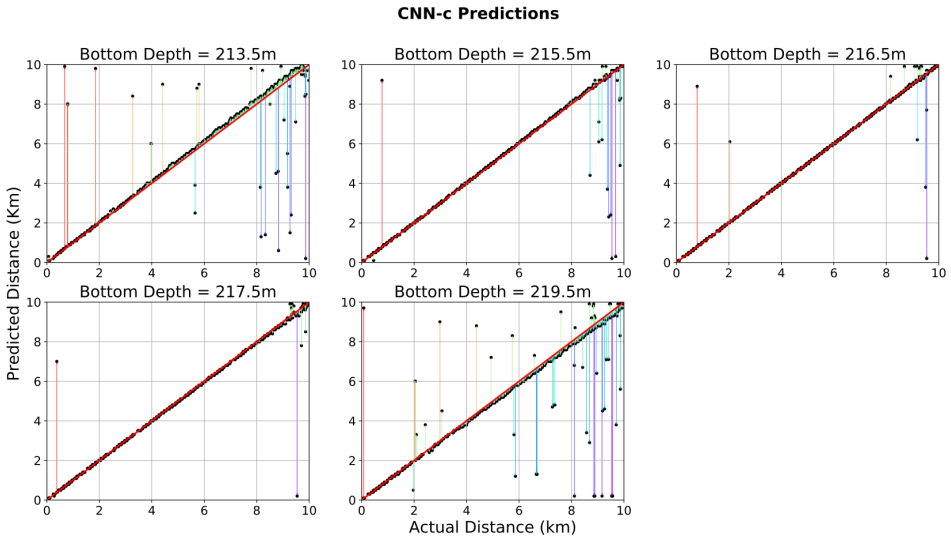
(a)

Figure 4.14: Diagonal solid line represents ground truth. Dots show prediction values. Quivers represent the difference between each prediction and the corresponding ground truth value. (a)–(c) Performance of each approach on testing data generated with different bottom depths.

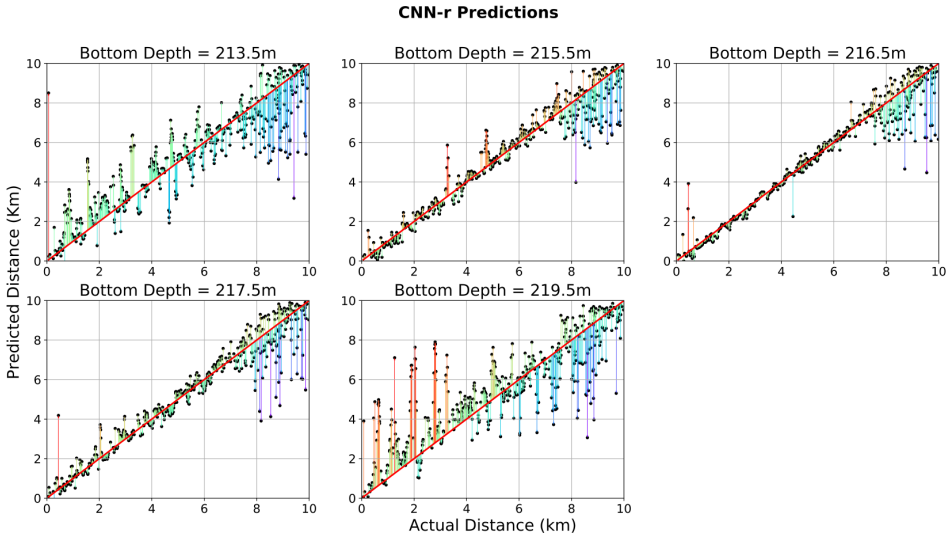
The ground truth range of the towed source with time is calculated from recorded GPS coordinates of the VLA and the source during the experiment; this is shown as the solid line in Fig. 4.15(b).

Performance Comparison

Fig. 4.15(b) shows the range predictions by MFP, CNN-c, and CNN-r. From this plot, it appears that MFP and CNN-c have similar performances, while CNN-r does worse in comparison. Table 4.3 shows that based on the two performance metrics introduced, CNN-c has the highest percentage of predictions within 1 km of ground truth with 70.7%, followed by MFP at 57.5%, and last, CNN-r with 37.8%. However, comparing their MAEs, CNN-r and CNN-c have similar performances by this metric, with 1.4 and 1.41 km, respectively. Both out-perform MFP, which has a MAE of 1.73 km. These results match the observations from the simulated testing cases, where CNN-c had the best performance of the three methods and CNN-r had lower MAE than MFP but lacked accuracy in individual predictions. This similarity



(b)



(c)

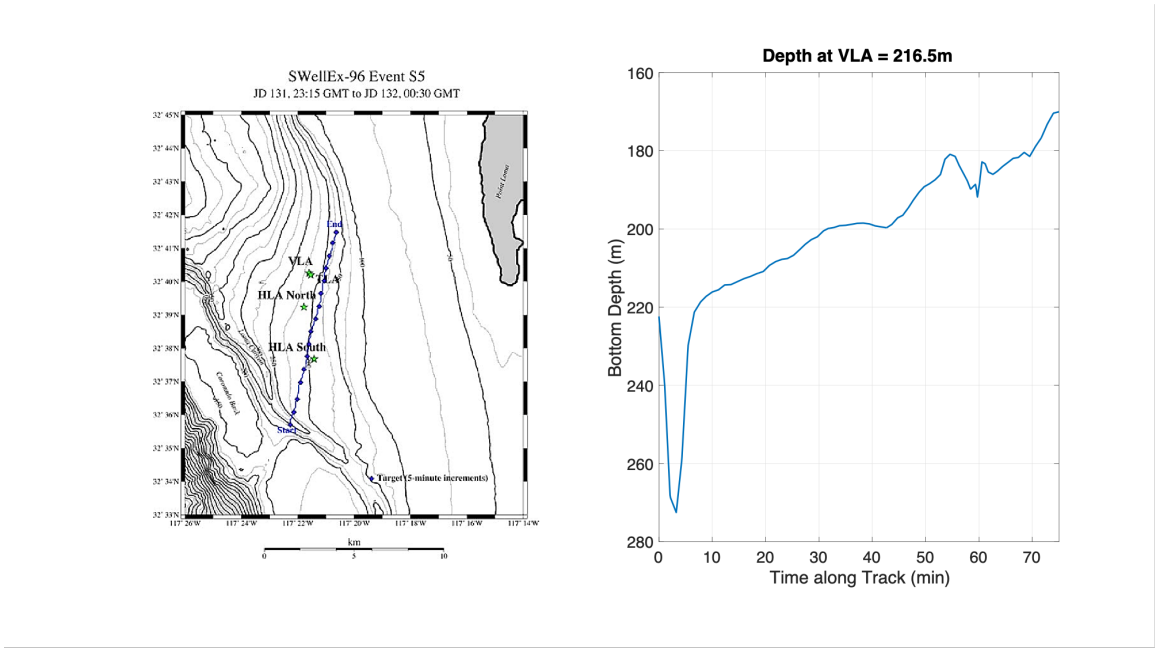
Figure 4.14: Diagonal solid line represents ground truth. Dots show prediction values. Quivers represent the difference between each prediction and the corresponding ground truth value. (a)–(c) Performance of each approach on testing data generated with different bottom depths. *(Continued).*

Table 4.3: Performance of MFP, CNN-c, and CNN-r on real data collected during SWellEx-96.

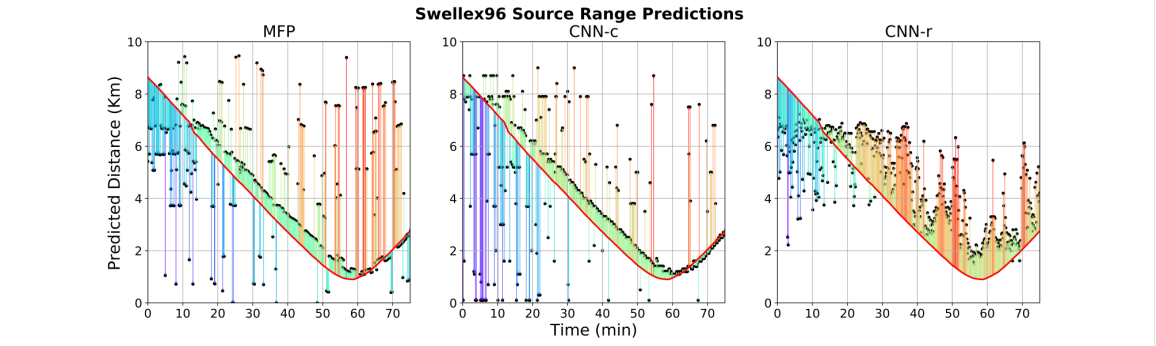
	MFP	CNN-c	CNN-r
% within 1 km of Actual	57.5	70.7	37.8
MAE (km)	1.73	1.41	1.40

again demonstrates the performance of our model-based CNN approach does indeed transfer to real data. Here, as with the simulated testing cases, CNN-r appears to over-compensate for lowering overall error at the expense of less accurate individual predictions.

In contrast, CNN-c shows better performance than MFP but is not immune to environmental mismatch, as demonstrated by its prediction errors at similar source ranges as MFP shown in Fig. 4.15(b). Furthermore this plot shows a consistent overestimation of $\sim$1 km or more in the MFP and CNN-c predictions from ground truth after the $\sim$10-minutes mark along the source track. The CNN-r predictions also greatly overestimate the source range during this part of the track. This overestimation may be attributed to the mirage effect in shallow water [154] - as a source moves over shallower bathymetry than what is modeled, the mismatch leads to the appearance of the source at a greater range than ground truth. Fig. 4.15(a) shows that, indeed, after around the 10 minutes mark, the bathymetry along the source track begins to become shallower than the modeled bottom depth of 216.5 m. Evidence of the mirage effect on the CNN predictions again demonstrates that although this approach may achieve more robust performance than MFP (CNN-c in this case), it is, nonetheless, subject to the same mismatch challenges that exist for all model-based methods.



(a)



(b)

Figure 4.15: (a) Left plot shows source track (solid line with dots) and the location of the VLA during the SWellEx-96 experiment [153]. Right plot shows the bottom depth along the source track; the bottom depth at the VLA is 216.5 m. (b) Range prediction outputs from MFP, CNN-c, and CNN-r. Solid lines represent ground truth of source range to VLA with time. Dots show individual predictions. Quivers show difference between individual predictions and corresponding ground truth values.

4.7 Summary

In this chapter, we explored the feasibility of using a model-based CNN approach for surface source range estimation. Specifically, we set out to answer three questions. The first is whether the CNN method would show more robustness to environmental mismatch than MFP on simulated testing data. To this end, we proposed two CNNs (CNN-c and CNN-r) and compared their performance to MFP under SSP and ocean bottom mismatch, respectively, in two separate simulated environments. CNN-c showed improvements over MFP in both cases. On the other hand, CNN-r performed better than MFP against SSP mismatch but less so against ocean bottom mismatch. The reason for CNN-r's inconsistent performance is likely due to its goal of lowering the overall MSE cost during training. This specification causes CNN-r to have less variability in its predictions, which lowered the overall error of the estimates but increased error on individual predictions compared to MFP.

Second, we used real field data collected in the two presented environments to test whether the performances of our model-based CNN-c and CNN-r transfer to real data. Our results confirm this to be the case. For the ICEX-16 dataset, both CNNs returned predictions consistent with the expected source range. For the SWellEx-96 dataset, CNN-c outperformed MFP by both of the metrics examined while CNN-r showed better overall MAE than MFP but was less accurate on individual predictions. These results are consistent with our simulated data test results.

Lastly, we discussed how our model-based CNNs may be achieving their performance by examining their intermediate outputs. Using the ICEX-16 environment as an example, we compared the Euclidean distance plots of MFP's template replica vectors to those of the CNNs' FC-layer (pre-prediction) output vectors (Figs. 4.10 and 4.11). For MFP, these plots show a sharp and narrow minimum when there is no SSP mismatch; the steep minimum disappears when mismatch is introduced. This result demonstrates that MFP can make very accurate predictions under no mismatch but can become very inaccurate with mismatch. In comparison, Euclidean distance plots for CNN-c and CNN-r show more gradual increases away from their respec-

tive minimum. This result means that CNN FC-layer output vectors for neighboring source ranges are also near each other in Euclidean space. Thus, any slight change to the FC-layer vectors as a result of environmental mismatch is less likely to cause the CNNs to output a prediction that is drastically different from the correct output, as would be the case with MFP.

Chapter 5

Transient Ice Noise Detection & Characterization

WITHIN the overall underwater ambient soundscape, there typically lies a cacophony of transient noise events. In the Arctic region, these transients usually result from various physical mechanisms in the ice cover such as fracturing and shearing [58]. Thus, detection and characterization of transient events in ambient noise can provide useful information on the temporal distribution of ice cover activity and mechanisms by which noise is generated. As we mentioned in Chapter 2, amplitude-based event detection algorithms in the time domain has a significant drawback in that the spectral content of the transients are fragmented. To improve upon such approaches, we propose a frequency domain transient detection algorithm in this chapter. We first discuss the benefit of this approach to previous time domain methods before dissecting into how various components of the algorithm work together to identify transients. We then apply it to data collected during ICEX-16 to demonstrate how it can better characterize transient distribution and their possible generation mechanisms during the experiment. We conclude by providing a quantification for the performance of our approach by applying it to two datasets with ground-truth transient labels.

5.1 Drawbacks of Amplitude-based Detection Approach

In Chapter 2, we presented a time domain transient detection algorithm by Zakarauskas *et al.* [66] that searches for improbable clusters of large amplitude peaks in the time series. Such an amplitude-based approach is commonly used for event detection not only in the field of underwater acoustics, but other areas such as seismology. For example, several studies [11, 155, 156, 157] have used the ratio of short and long-term averages over time series windows as a threshold to detect transient occurrences in geophone data. The benefit of time domain algorithms is that they are straightforward to implement and their detection sensitivity can often be easily tuned by adjusting the ratio threshold. However, the drawback is that they do not provide a complete description of the detected events' frequency characteristics. Oftentimes, the data time series are processed in separate frequency bands so that, in particular, weaker, narrowband transients at high frequencies are not obscured by the general background noise [67]. As a result of the lack of spectral information, time domain methods are less helpful in hypothesizing the mechanisms by which the detected transients may have been generated, as different ice cover mechanics can produce different spectral signatures [56, 58].

5.2 Frequency Domain Transient Detection Algorithm

As an alternative to time domain, amplitude-based methods, we present an algorithm for transient detection in the frequency domain. This method takes an image processing approach to identify transients within the noise spectrograms. It then easily allows the user to categorize the detected events based on their bandwidth and duration. We apply this method to ambient noise data collected during ICEX-16 on March 13. Because this dataset is only an $\sim$8 hour snapshot of the Beaufort Sea noise

environment, we design our analysis to detect shorter time scale transient events (on the order of seconds to tens of seconds) that result from ice cracking or floe interactions. These types of events commonly occur near pressure ridges where weaker ice readily fractures and breaks off into floes due to temperature, wind, or current stress.

5.2.1 Data Pre-processing

As noted, the input to our detection algorithm is the spectrogram of a data snapshot. For the ICEX-16 data, spectrograms of data on each hydrophone channel in the VLA are computed using a 4096-point Hanning window with 50% overlap. All channels are then averaged to form a composite spectrogram. We focus our analysis to below 2048 Hz and place more emphasis on lower frequencies by transforming the frequency axis to Mel scale with 128 bins. The conversion from Hertz to Mel scale is linear for frequencies <1 kHz and logarithmic above 1 kHz (Fig. 5.1). This step serves three purposes. First, noise generated by ice cover activity has most of its spectral energy below 2 kHz, with much of that focused below 1 kHz [158, 159]. The use of the Mel scale conveniently places more focus on lower frequencies during event detection. Second, while the ICEX-16 VLA has the best beamforming spatial resolution between 800-900 Hz, it would not encounter any aliasing between ±30 degrees for frequencies up to ∼2048 Hz[1]. Since much of the ambient noise arrives at the VLA between ±30 degrees (Fig. 3.13), we can safely perform beamforming on any transient event that we detect to determine its vertical directionality. The final purpose is that the decreased number of frequency bins as the result of using the Mel scale decreases the total number of pixels in the input spectrograms. This allows our algorithm to process the data more efficiently.

5.2.2 Image Processing Tools

Before presenting the transient detection algorithm, we first introduce a few operators that are commonly used in computer vision and image processing for the purpose of

[1]Spatial aliasing occurs first at VLA endfire and last at broadside.

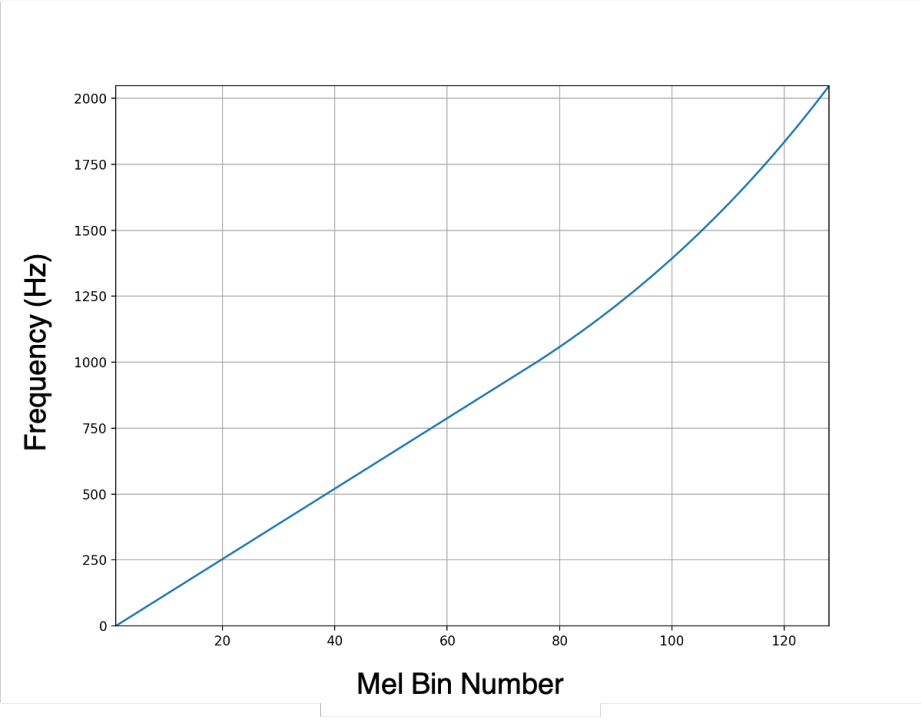


Figure 5.1: Conversion from linear frequency to Mel scale is linear below 1 kHz and logarithmic otherwise.

noise reduction and edge detection. These operators are all applied to the input image via discrete convolution ($*$) with a kernel matrix. For notation, we define the input to an operator as the source matrix, $S(x, y)$, and the output as the destination matrix, $D(x, y)$, where (x, y) denote indices of a 2-D matrix.

Gaussian Blurring

Gaussian blurring is an effective way of removing random, Gaussian noise from an input image. In this method, a 2-D discrete, Gaussian kernel, $G(x, y)$, is defined using two variance parameters, σ_x and σ_y. The size of this kernel is user defined, but is typically a positive, odd number. The output of the blurring process is calculated as

$$D(x, y) = G(x, y) * S(x, y). \tag{5.1}$$

Erosion and Dilation

In images where there are foreground objects against a background, erosion thins the boundaries of the foreground objects and gets rid of very small features, or "salt" noise. Contrarily, dilation expands the boundaries of the foreground objects and eliminates small gaps between objects, or "pepper" noise. When used together, they help diminish the amount of "salt and pepper" noise. The erosion operator is defined as

$$D(x, y) = \min_{(x', y')}(S(x + x', y + y')), \tag{5.2}$$

where (x', y') denote the indices around (x, y) that are included in the convolution with the operator kernel and thus depend on the kernel size. Similarly, the dilation operator is defined as

$$D(x, y) = \max_{(x', y')}(S(x + x', y + y')). \tag{5.3}$$

Edge Detection

Edge detectors identify the edges of objects in an image by computing the gradient of the image in both the x and y directions. One common approach for calculating the derivative is with the Sobel filter [160], which is a discrete differentiation operator. For a kernel size of 3×3, the horizontal derivative is calculated as

$$G_x(x, y) = \begin{bmatrix} -1 & 0 & +1 \\ -2 & 0 & +2 \\ -1 & 0 & +1 \end{bmatrix} * S(x, y), \tag{5.4}$$

and the vertical derivative as

$$G_y(x, y) = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ +1 & +2 & +1 \end{bmatrix} * S(x, y). \tag{5.5}$$

For our application, we combine the two directional derivatives to compute the total derivative of the input as

$$D(x, y) = max(G_x(x, y), G_y(x, y)). \tag{5.6}$$

5.2.3 Feature Detection

With the input spectrograms formed, the detection algorithm is applied to identify high SNR *features* that may be part of transient events. This algorithm is regulated by three threshold parameters - T_{mask}, T_{dist}, and T_{area}. T_{mask} is a threshold that controls which pixels in the spectrogram are kept based on how much higher their values are compared to the background noise estimate. As shown in Eq. 5.7, it is a unit-less measure of linear SNR.

$$A_{mask} = \frac{A_{pixel}}{\eta_b}, \tag{5.7}$$

where A_{mask} is the mask value for a pixel that is compared to the threshold T_{mask}, A_{pixel} is the spectrogram amplitude of that pixel, and η_b is the background noise estimate for the input spectrogram (more on this below). The decision for the value of T_{mask} depends on the expected loudness of the events. If the transients of interest are fairly faint compared to the general background noise, then T_{mask} should be set small, and *vise versa*. However, there is a positive correlation between true detection rate and false detection rate - a smaller T_{mask} value leads to increased detection of fainter transients but also more mis-classifications of random noise as events. Thus, some initial empirical testing is recommended when setting this parameter.

T_{dist} is a measure of distance between features based on their positions in the spectrogram. The distance between two features is computed using a custom metric shown in Eq. 5.9 (see Section 5.2.4). It is used to determine whether two or more features should be grouped together as a single transient event or be counted separately. Thus, it is a parameter that controls the amount of over- or under-counting that may occur. T_{area} is simply a measure of the number of pixels in a feature. Its

purpose is to eliminate features with very few pixels that are likely random noise. In this sense, it sets a minimum pixel count requirement for the definition of a feature. With these threshold values, the detection algorithm works as follows:

1. The background noise level at each frequency (η_b), defined as the noise level when no transient events are occurring, is estimated for the first input data snapshot by calculating the spectrogram of a data window much longer than the duration of the first snapshot and then averaging the spectrogram values over time. For the analysis of the ICEX-16 data, the individual input snapshot lengths are set to 30 seconds while the duration of data used to estimate the background noise is 30 minutes. This approach for estimating the background noise level works off of two assumptions. One is that for any sufficiently long window of data, there are no transient events during a large majority of the time. The other is that there is enough total data available for such a sufficiently long window. In our testing with the ICEX-16 dataset, a 30-minute long window was adequate for us to detect even very low SNR transients in the first input snapshot. An example of an input spectrogram is shown in Fig. 5.2 (top left).

2. The frequency-normalized spectrogram (Fig. 5.2, top middle) is calculated for the current data snapshot by dividing the original spectrogram values at each frequency by the estimated background noise level at the corresponding frequency. This step helps to mitigate the issue that noise level is typically higher at lower frequencies, which can mask transient events at higher frequencies if the spectrogram is not normalized.

3. Smoothing, dilation, and erosion are applied to the frequency-normalized spectrogram to filter out noise in the image that may be mistaken for transient events. Smoothing is done by applying a 3×3 Gaussian blurring filter over the spectrogram. Dilation and erosion are performed using the Python library OpenCV [161], also with a 3×3 kernel. After noise reduction, edge detection is done using the Sobel filter to construct a gradient map of the input (Fig. 5.2, top right).

4. A mask is created by thresholding the gradient map - all pixel values in the map that are less than T_{mask} are set to 0 while the rest are set to 1 (Fig. 5.2, bottom left). The mask is then applied to the filtered spectrogram through multiplication and all zero-value pixels are removed from the output. This results in a post-processed image with disconnected pixel groups that we call *features* (Fig. 5.2, bottom middle).

5. The features in the post-processed image are further grouped together using hierarchical clustering (h-clustering) with a custom distance metric that is designed to conjoin separate features that are likely to be part of a single transient event (see Section 5.2.4). Two features are conjoined if the distance between them is less than T_{dist}. At the end of the clustering process, very small features with less than T_{area} number of pixels are deleted to further eliminate noise. The detected transients are then labeled by type based on their bandwidth and duration (Fig. 5.2, bottom right, see below for type definitions).

6. After transient detection within the current snapshot, background noise is estimated for the next snapshot. Pixels in the detected transients are deleted from the current input spectrogram to create a "noise-only" image. An update to the background noise level ($\eta_{b,\delta}$) is calculated by averaging this "noise-only" image over time. The new background noise level is then calculated by averaging together the previous background noise level and this new update following

$$\eta_{b,new} = \frac{\eta_{b,old} + \eta_{b,\delta}}{2}. \tag{5.8}$$

7. Repeat steps 2-6 on the spectrogram of the next data snapshot window.

For analysis of the ICEX-16 data, the three threshold parameters are set as $T_{mask} = 17$, $T_{dist} = 5.5$, $T_{area} = 20$. These values are selected empirically after some testing on a small subset of the data to check the algorithm is detecting all transients that are observed by a human analyst.

Based on the bandwidth and duration of the detected transients, we define four

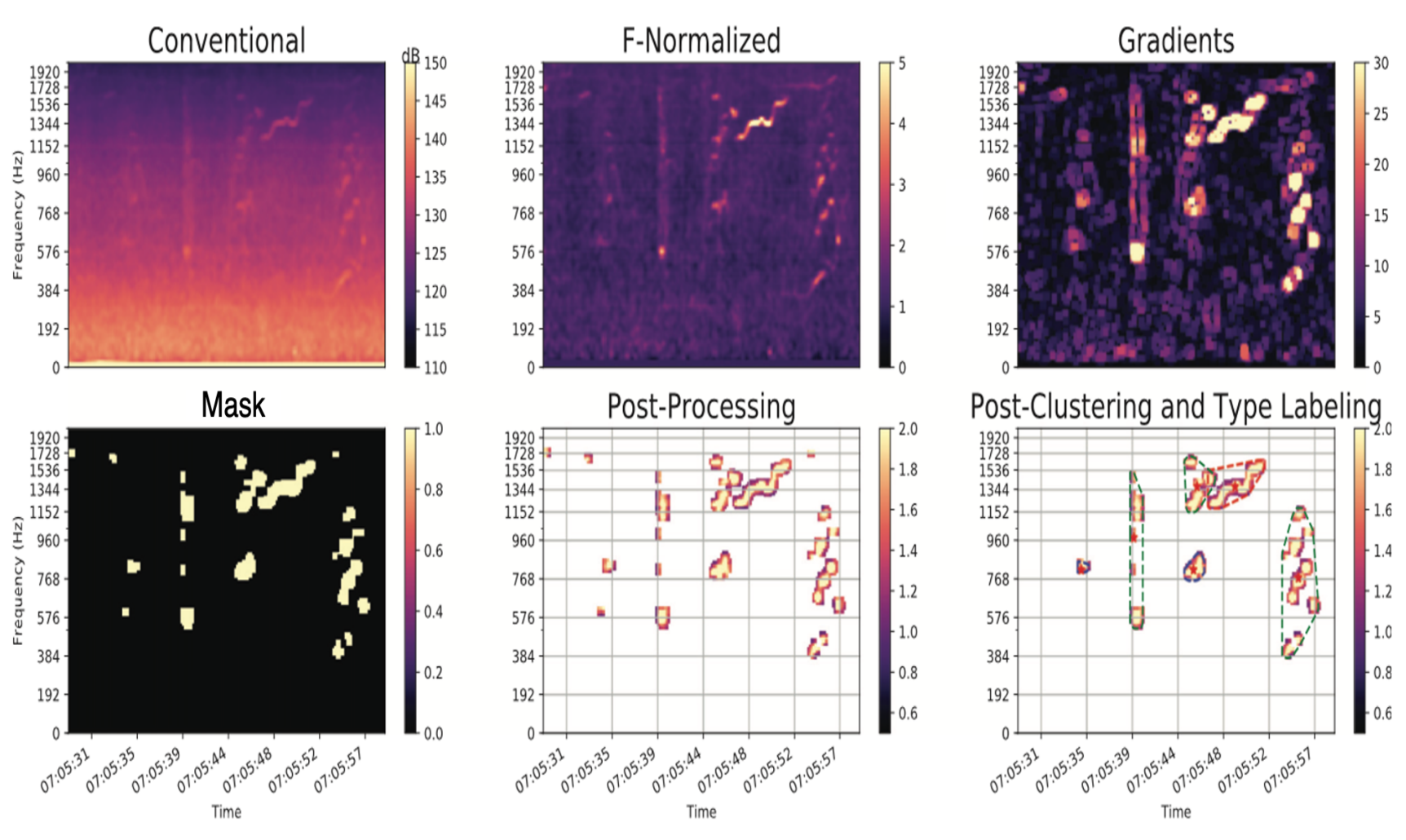


Figure 5.2: Example of an ICEX-16 data snapshot spectrogram that is the input into the detection algorithm (top left). Input spectrogram after frequency normalization (top middle). Gradient map created following image processing to reduce "salt and pepper" noise (top right). Mask created based on the gradient map (bottom left). Post-processed spectrogram after applying the mask; only pixels in detected transients are kept (bottom middle). Outline of detected transients after h-clustering and labeling. Blue, stnb; red, ltnb; green, stbb; black, ltbb. Red star denotes the center of the detected event (bottom right).

types of events: short-time-narrowband (*stnb*, duration <5 s, bandwidth <200 Hz), long-time-narrowband (*ltnb*, duration >5 s, bandwidth >200 Hz), short-time-broadband (*stbb*, duration <5 s, bandwidth >200 Hz), and long-time-broadband (*ltbb*, duration >5 s, bandwidth >200 Hz). Examples of each event type are shown in Fig. 5.3. We believe the *stnb* events mostly encompass high SNR background noise within the spectrogram that have been falsely detected as events or perhaps small features that may have been part of another event but were incorrectly clustered. Thus, the *stnb* category serves as another noise removal procedure. The representation of the other three categories is discussed in Section 5.3.

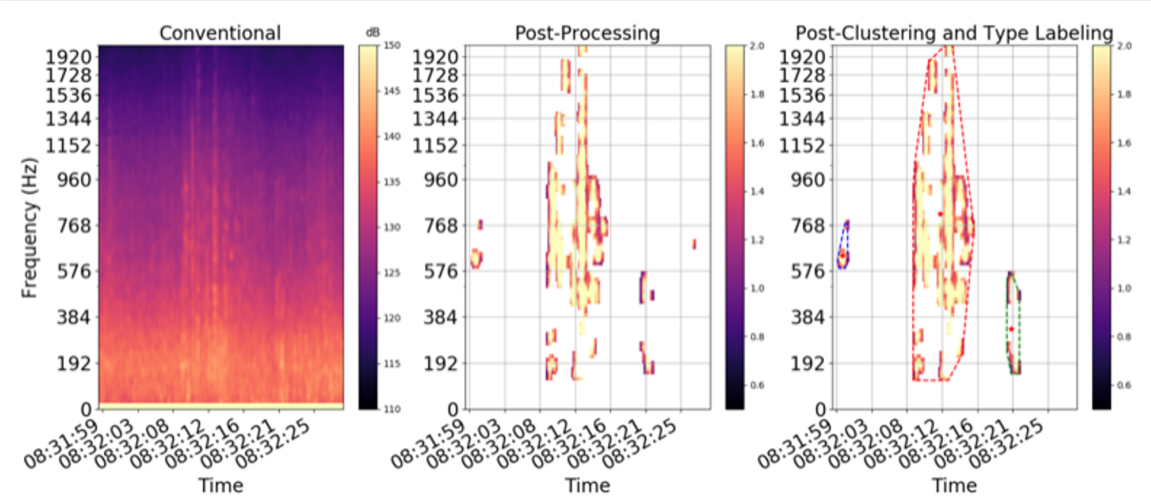
5.2.4 Hierarchical Clustering and Distance Metric Design

h-clustering

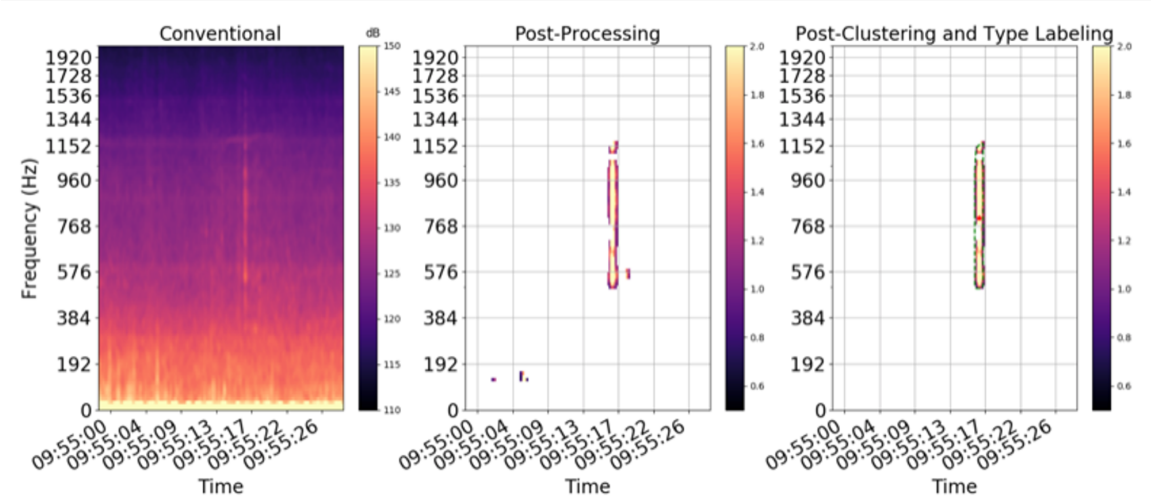
Hierarchical clustering (h-clustering) [162] is used in our event detection method to group together features that may be part of the same transient event. We take an agglomerative or "bottom-up" approach, which means that each feature in the post-processed image starts off as an individual cluster of one. Then, starting with the largest cluster by the number of pixels, we iterate through each cluster and combine that cluster with its closest neighbor as defined by a distance metric (Fig. 5.4). We stop the clustering process once all remaining clusters are at least T_{dist} apart from each other.

Distance Metric

The distance metric used in our h-clustering procedure is designed based on two principles. The first is that, intuitively, features that occur at or near the same time are more likely to be part of the same transient event. To implement this principle, we elect to use an average of the absolute pixel distance and the pixel distance along the time axis only as the distance metric between two features (Eq. 5.9). The absolute pixel distance between two features is further defined as the minimum distance between all pixels in one feature to all pixels in the other feature

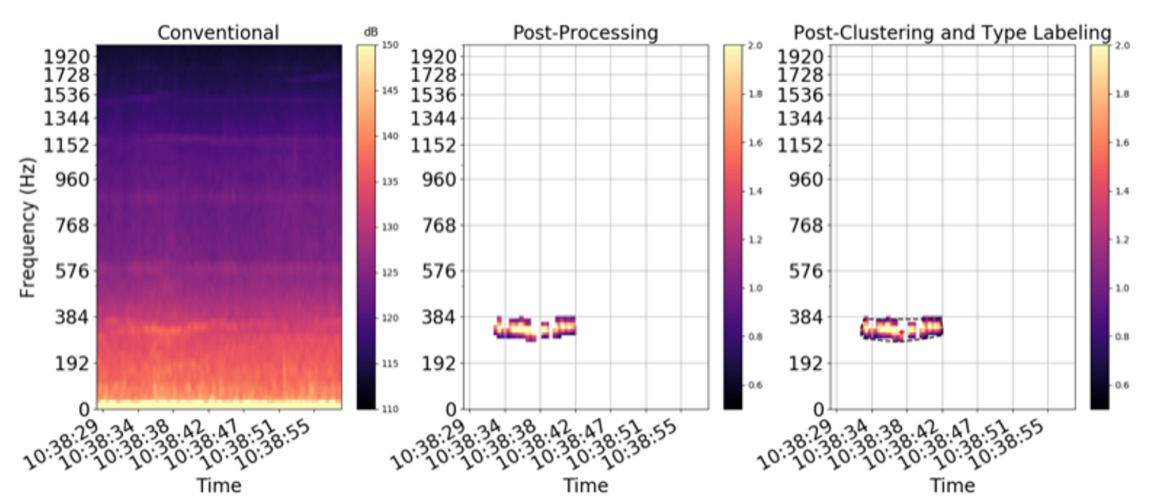


(a)

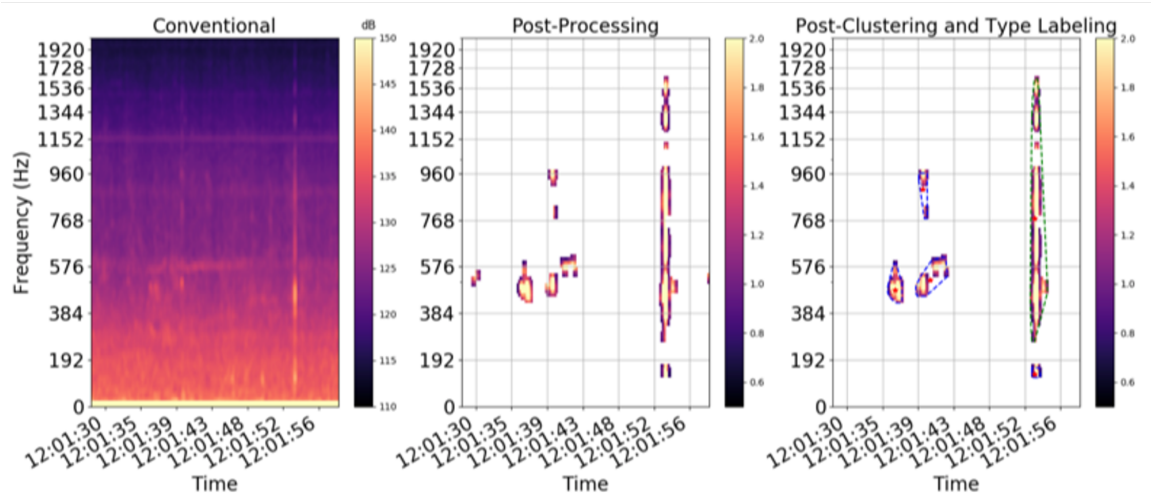


(b)

Figure 5.3: Examples of different transient event categories detected in ICEX-16 data from March 13. (a) A *ltbb* event followed by a *stbb* event. (b) A *stbb* event. (c) A *ltnb* event. (d) Some *stnb* noise followed by a *stbb* event.



(c)



(d)

Figure 5.3: Examples of different transient event categories detected in ICEX-16 data from March 13. (a) A *ltbb* event followed by a *stbb* event. (b) A *stbb* event. (c) A *ltnb* event. (d) Some *stnb* noise followed by a *stbb* event. *(Continued).*

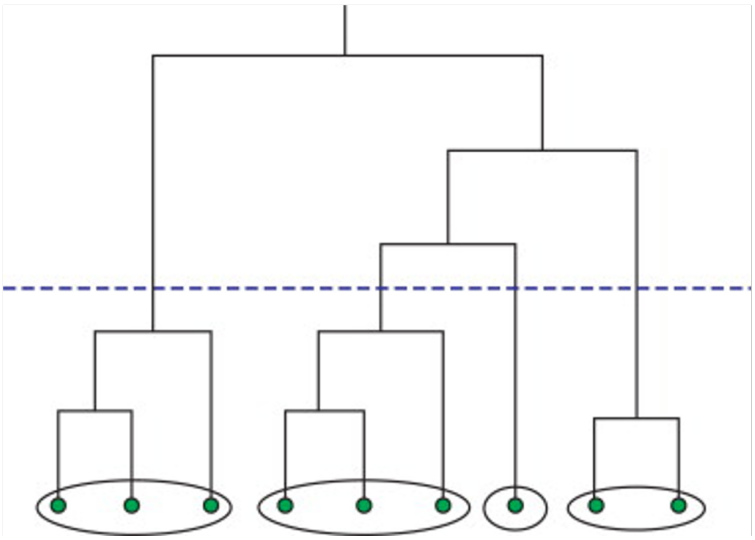


Figure 5.4: Demonstration of agglomerative h-clustering. Green dots represent individual starting clusters, larger ovals represent output clusters following grouping, dashed line represents the clustering hierarchy at which the stop condition is met.

(Eq. 5.10). Similarly, the pixel distance along the time axis between two features is defined as the minimum distance between the time indices of all pixels in one feature and the time indices of all pixels in the other feature (Eq. 5.11). This custom distance metric represents a balance between grouping together features that are close together in terms of relative proximity, while also combining more distant features that are close in time only - some more distant features may be part of a single broadband event and would not otherwise be grouped.

The second design principle places a constraint on the first and holds that it is unlikely for a horizontal and a vertical feature to be part of the same transient event. This is because a horizontal feature is more likely to be part of a narrowband tonal while a vertical feature is more likely to be part of a broadband impulse. Thus, as part of our distance metric calculation, we first determine the slopes of the features. If the slope is greater than 1.2, we classify that feature as vertical. If the slope is less than 0.8, we classify the feature as horizontal. Otherwise, we do not assign an orientation label. Note, a slope of 1 means a feature has an equal number of pixels in height and width. In calculating the distance between two features, if one has vertical orientation while the other has horizontal orientation, we manually assign

their distance to be infinite so that they are never clustered together. This ensures that tonals and broadband events are not grouped as a single event, even if they occur closely in time.

Formally, the distance metric between two features f_1 and f_2, is computed as

$$d(f_1, f_2) = \begin{cases} \infty, & \text{if } f_1 \perp f_2 \\ \frac{d_p + d_t}{2}, & \text{otherwise,} \end{cases} \tag{5.9}$$

where, $\perp$ signifies that f_1 and f_2 have opposing orientations. d_p and d_t denote the absolute pixel distance and the pixel distance along the time axis between f_1 and f_2, respectively. Given the pixels in f_1 and f_2 are indexed by $(\vec{x_1}, \vec{y_1})$ and $(\vec{x_2}, \vec{y_2})$, respectively, where x represents the time axis and y represents the frequency axis, the two distances are calculated as

$$d_p = min(\sqrt{(\vec{y_2} - \vec{y_1})^2 + (\vec{x_2} - \vec{x_1})^2}), \tag{5.10}$$

$$d_t = min(|\vec{x_2} - \vec{x_1}|). \tag{5.11}$$

5.3 Characteristics of Transients detected in ICEX-16 Dataset

A total of 3178 transient events are detected by applying our algorithm to the ∼8 hours of ICEX-16 data collected on March 13. Of these, 2081 are *stnb* events, which, as we noted, are likely background noise that are not ice-generated transients. Thus, we do not consider these in our discussion and instead focus on the other three event categories.

Details of the 1097 non-*stnb* events are shown in Table 5.1. *stbb* events are the most common event type, with 773 occurrences accounting for 70.5% of the detections. We believe this event type is indicative of small, impulsive ice fractures that may occur in the periphery of larger cracking events. As described by Xie and Farmer [58], ice

Table 5.1: Detected transient events by category.

	# of Detections	% of Total	Combined Duration (s)	% of Recording Time	Mean Peak Elevation (Deg.)
stbb	773	70.5	1800.5	6.3	-10.7
ltnb	82	7.5	691.9	2.4	-11.0
ltbb	242	22.0	2366.4	8.3	-10.5

breaking can be partitioned into three phases. First, environmental forcing initiates ice breaking near weakened locations such as ridge formations. This process emits short, impulsive, sound pulses, similar to the *stbb* events that we observe. Second, the initial fractures promote more failures and further cracking. In turn, a positive feedback loop develops and leads to a more extended period of broadband noise generation. We believe this second phase is descriptive of the *ltbb* events. They are the most prevalent event type with 2366.4 s of total combined duration over 242 detections. This accounts for ∼8.3% of the total data recording time. The third phase of ice breaking accounts for ice floe interactions. As small ice floes break off and rub past each other, pure tones are emitted into the water column. We believe the *ltnb* events are representative of these tonal features. They are the rarest event type from our analysis. Only 82 are detected and their combined duration accounts for less than 2.5% of the total data recording time. The distribution of event durations for the detected events is shown in Fig. 5.5. From this plot, we see that the majority of *stbb* events have durations less than 3 s. For *ltnb* and *ltbb* events, their durations are longer than 5 s by definition but the majority last no longer than 10 s. These relatively shorter event durations compared to findings from previous studies [56, 57] suggest that ice cracking during ICEX-16 was likely occurring on a small temporal and spatial scale, absent of any prolonged major activity in the ice cover (*e.g.*, no large lead openings).

After event detection, the peak vertical directionality arrival angle at the VLA is determined for each transient. This step is done by segmenting the portion of data

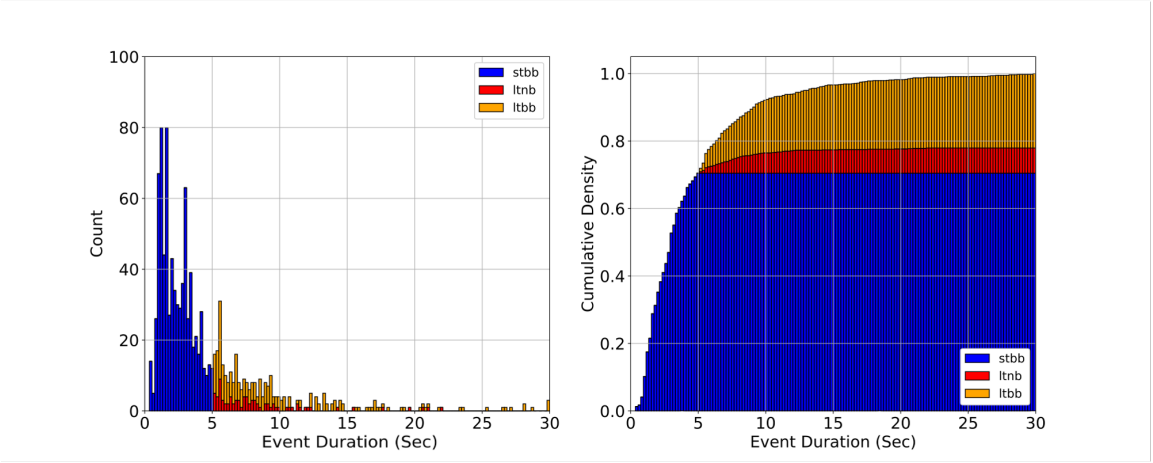


Figure 5.5: Distribution of transient event durations by type.

that contains the event using a 512-point Hanning window and then beamforming over the frequency range of the event. The beamformed output is then averaged over frequency and time to derive its directionality profile. From there, the peak elevation angle is determined. As shown in Fig. 5.6, these peak angles mostly cluster near -10 degrees and are fairly consistent over time for all event types. The mean peak elevation angles for different event types are all between -10 and -11 degrees (Table 5.1). These results agree with that of the overall ambient noise presented in Chapter 3. Thus, in accordance with our previous discussion in Chapter 3, the peak arrival angle values support the claim that the detected events are generated by the ridge formation ∼30-50 km away from the ICEX-16 camp.

Table 5.2 shows the frequency content of the detected events. Five frequency intervals are defined at octave ranges (<160, 160-320, 320-640, 640-1280, and 1280-2048 Hz). Of these intervals, 640-1280 Hz and 320-640 Hz are the most active with 79.8% and 70.9% of events having some frequency content within these two octaves, respectively. In addition, the distribution of the event bandwidths is shown in Fig. 5.7. For *ltnb* events, most have a bandwidth of less than 150 Hz. For broadband events of both types, most have bandwidths of less than 750 Hz. Of course, the event bandwidths are capped by our choice of the analysis frequency limit at 2048 Hz. However, the values in Fig. 5.7 are reasonable compared with transient events

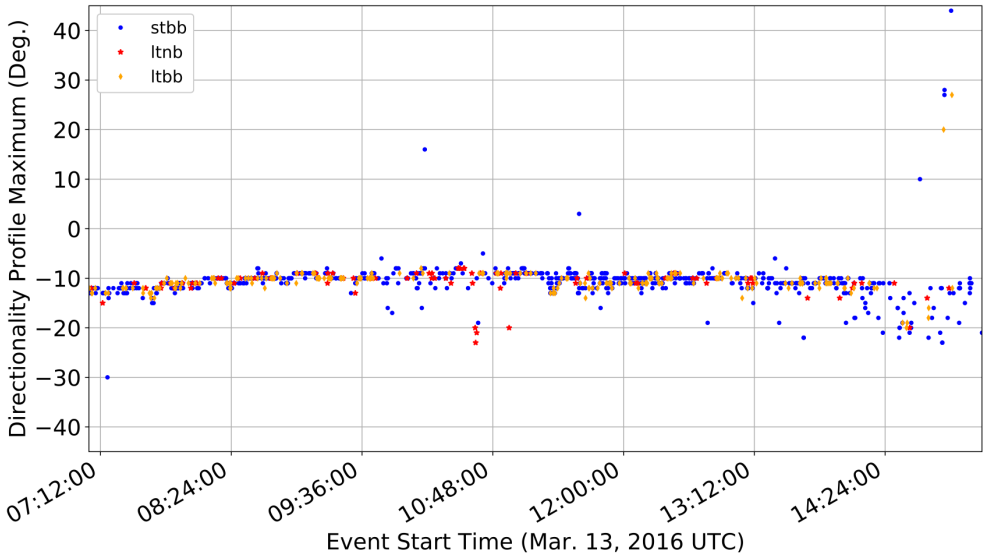


Figure 5.6: Peak arrival angles of detected events mostly cluster near -10 degrees, in agreement with that of the overall ambient noise. This plot also shows the temporal distribution of detected events by category.

Table 5.2: Percentage of events that are at least partly within a frequency interval (excludes *stnb* events).

Frequency Interval (Hz)	<160	160-320	320-640	640-1280	1280-2048
% of Events	8.8	22.3	70.9	79.8	19.3

presented in previous studies [56, 57, 58].

The final noise characteristic that we present is the time gap between the starts of consecutive transients. Fig. 5.8 (left) shows the majority of events occurred less than 50 s from the previous. This result suggests transient events were occurring quite consistently throughout the data recording session. If we examine the time between consecutive events of the same type, Fig. 5.8 (right) shows the time gaps can be much longer, especially for *ltnb* and *ltbb* events. This makes intuitive sense as these events are rarer than *stbb* events. However, their rarity does not mean that they can not occur in quick succession of one another. The majority of the *ltnb* and *ltbb* events

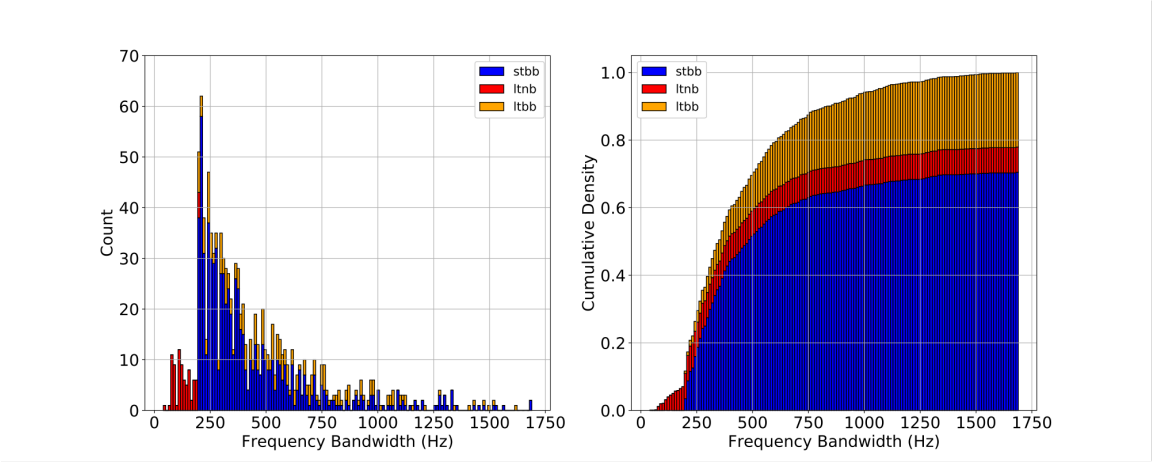


Figure 5.7: Distribution of event bandwidths by category.

nonetheless occurred within 100 s of the previous event of their respective type. Fig. 5.6 shows the temporal distribution of detected events by category. For *ltnb* and *ltbb* events, we observe times when they happen in clusters (e.g., near 08:24:00) and times when there is a longer gap in their occurrence (e.g., after 13:12:00). In contrast, the occurrence of *stbb* events appears to be more consistent over time. These observations suggest small fractures may have occurred fairly regularly during ICEX-16, whereas larger cracking events and the subsequent ice floe interactions were more sparse in time. However, when larger cracking events did occur, they tend to have happened in clusters, resulting in a sudden increase in transient noise generation. This clustering result is in agreement with our previous temporal analysis of the ICEX-16 data in Chapter 3.

5.4 Case Studies to Assess Algorithm Performance

In this section, we present more information on the performance of our frequency domain transient event detection algorithm by applying it to two datasets with ground truth transient event labels. We evaluate our algorithm with two metrics. The first is the true positive rate (TPR), which is the percentage of detected transients that are known to be true events based on the ground truth labels. It is formally defined

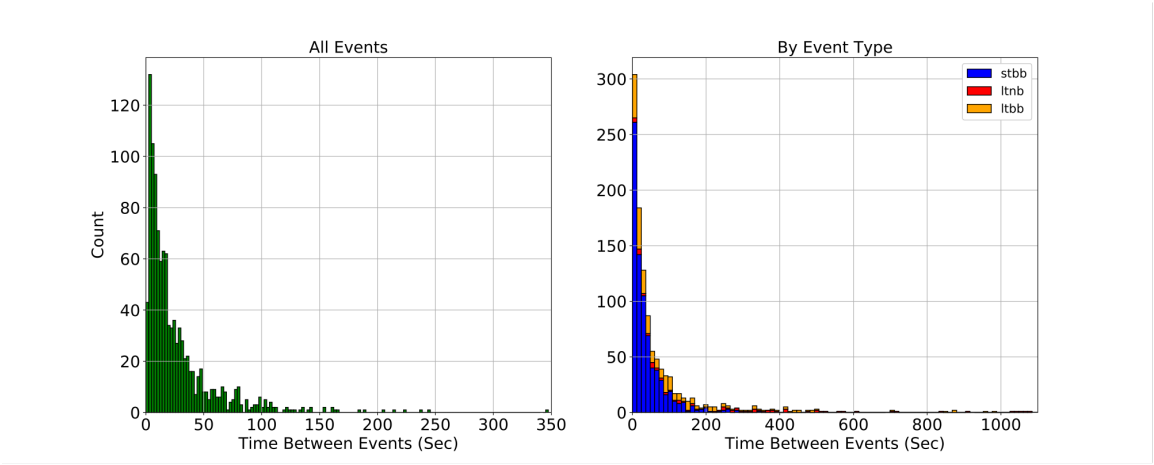


Figure 5.8: Distribution of time gaps between successive event start times (left) for all events and (right) for events of the same category.

as

$$TPR = \frac{TP}{TP + FN} \times 100, \tag{5.12}$$

where, TP is the number of true positives - transient events correctly identified as transients, and FN is the number of false negatives - transient events that are not detected.

A complementary metric to TPR is the false positive rate (FPR), which is the percentage of detected transients that are incorrectly identified as transients based on the ground truth labels. It is computed as

$$FPR = \frac{FP}{FP + TN} \times 100, \tag{5.13}$$

where, FP is the number of false positives - non-events incorrectly identified as transients, and TN is the number of true negatives - non-events that are correctly not identified as transients. Typically, for a detection problem, TPR and FPR are positively correlated - as the TPR increases, FPR increases as well. The goal of any detection algorithm is to maximize TPR while minimizing FPR. In practice, for the user, this means setting the parameters of the algorithm appropriately so that there is a balance between an adequately large TPR and acceptable FPR. For our exam-

ples below, we have selected the algorithm parameters such that there is diminishing returns on TPR, *i.e.*, changing the parameter values further to boost FPR does not lead to an equivalently large increase in TPR.

5.4.1 HICEAS Bioacoustics Dataset

The first dataset is a subset of the acoustic data collected during the Hawaiian Islands Cetacean and Ecosystem Assessment Survey (HICEAS) in 2017 [163]. It was collected using a multi-channel towed hydrophone array [164] with a sampling frequency of 500 kHz. For our analysis, we focus on detecting Minke whale "boing" vocalizations within a ∼3.5 hour segment of data collected on December 1, 2017. The ground truth labels for these calls were annotated using the software program PAMGuard between 1100 and 1800 Hz [163]. To generate the input spectrograms, a Hanning window of length 51200 is applied to 1-minute long data snapshots with 50% overlap. The initial background noise estimate is calculated using 10 minutes of data. The application of our algorithm is the same as described in Section 5.2 but with some parameters changed to suit this dataset. Most notably, we set $T_{mask} = 50$, $T_{dist} = 5.5$, and $T_{area} = 75$. These values are again empirically chosen after some testing on a subset of the dataset. The major cause of false predictions in this dataset is the presence of short, impulsive noise (Fig. 5.9). Thus, we devise two event categories: short-time (duration <1.25 s) and long-time (duration >1.25 s). Only events classified as long-time are considered whale call detections. With this set-up, our algorithm achieves a TPR of 70.9% and a FPR of 13.9% compared to ground truth annotations.

A brief note of our algorithm's performance on this dataset compared to other whale call detection methods presented in the Bioacoustics literature. We do not expect our algorithm to achieve the same level of accuracy as other methods that are dedicated to whale call detection [165, 166]. This is because such methods typically incorporate and leverage common features of the whale calls' spectral signatures to aid detection. In its current state, our algorithm is designed as a tool that would allow the user to detect all transient features in a dataset regardless of spectral signature and then categorize them to gain a better understanding of the entire soundscape. In this

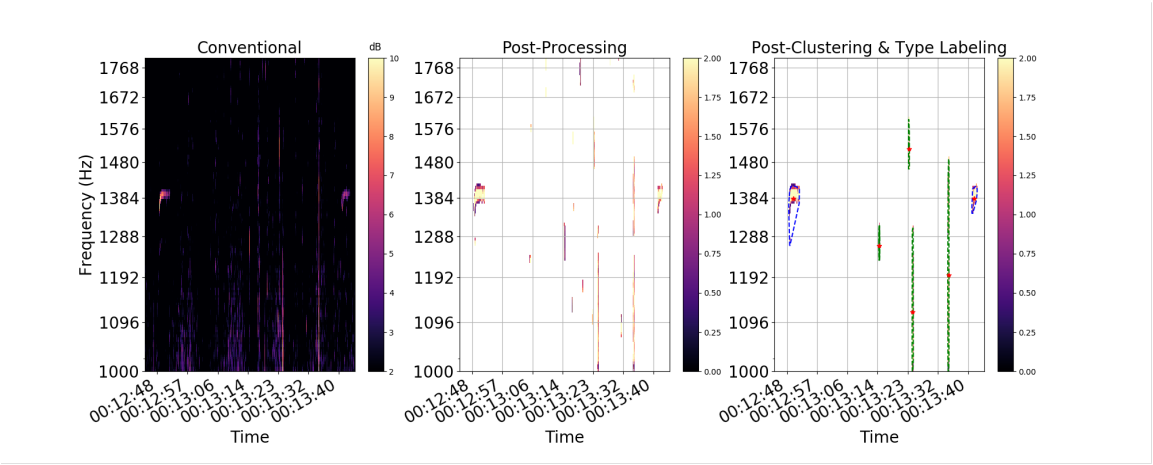


Figure 5.9: Examples of whale call detections (near 00:12:48 and 00:13:40) and impulsive noise (all others) in the HICEAS dataset. The x-axes denote UTC time on December 1, 2017.

sense, the algorithm provides more utility in situations where the spectral shape of the transients of interest are not well known or may vary (such as for characterizing ambient noise generated by ice cover). The algorithm's categorization metrics can certainly be designed based on the user's need to specialize detection of a certain type of transient. Thus, this is an area for improvement for this approach.

5.4.2 SIDEx Dataset

The second dataset is collected during the Sea Ice Dynamics Experiment (SIDEx), which deployed a geophone array to monitor transient cracking events in the surface ice cover just north of Utqiagvik, AK in January, 2020. As part of this experiment, a series of calibration events were generated by the field team by synchronously jumping and landing at various locations on the ice cover. At each location, four calibration events are generated in quick succession. These events were recorded by four, 3-axes, cabled geophones sampling at 1000 Hz and their generation times were noted for ground truth comparison. The calibration events appear as impulsive, broadband signals in the data spectrogram, with much of their energy below 100 Hz. For more information on the SIDEx experiment and data, please see Chapter 6.

For our analysis here, we focus on detecting these calibration events using data

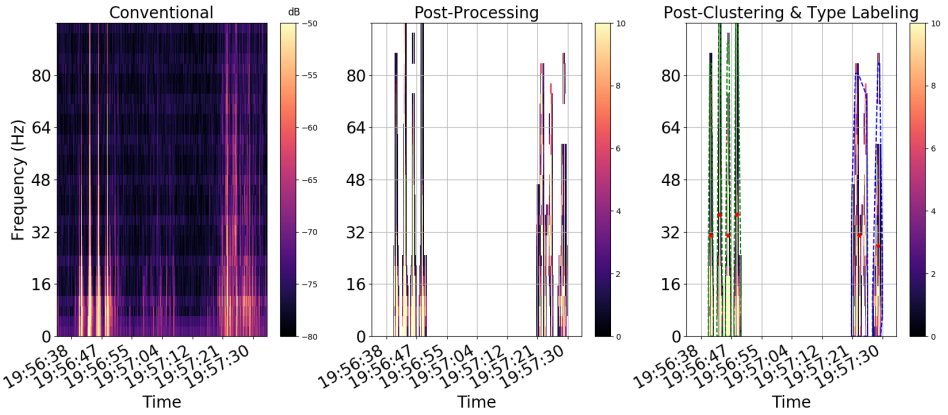


Figure 5.10: Examples of calibration event detections (between 19:46:07 and 19:46:16) and noise (all others) in the SIDEx dataset. The x-axes denote UTC time on January 25, 2020.

from the vertical component of the geophones. To generate the input spectrograms, a Hanning window of length 512 is applied to 1-minute long data snapshots with 50% overlap. The initial background noise estimate is calculated using 10 minutes of data. After some initial testing, we set the threshold parameters as $T_{mask} = 200$, $T_{dist} = 3$, and $T_{area} = 25$. The major cause of false predictions in this dataset is the presence of longer, impulsive noise greater than 2 s in duration (Fig. 5.10). Thus, we again devise two event categories: short-time (duration <2 s) and long-time (duration >2 s). Only events classified as short-time are considered calibration event detections. With this set-up, our algorithm achieves a TPR of 73.9% and a FPR of 21.7% compared to ground truth annotations.

Fig. 5.10 shows the spectral shape of the calibration events are very similar to the noise in this case. This undoubtedly diminishes the ability of our algorithm to accurately separate the two into different categories. Thus, this is a limitation of our current algorithm. As noted in the previous example, for future improvement, less empirical metrics for event categorization may be developed by leveraging known characteristics of the transients of interest. For example, in this case, the fact that calibration events occur in groups of four may have been incorporated to more accurately separate them from other noise transients.

5.5 Summary

In this chapter, we presented a frequency domain transient event detection algorithm which leverages image processing techniques and h-clustering to identify transients in data spectrograms. This approach is an improvement over amplitude-based methods in the time domain in that it retains the spectral information of the detected transients, which can then be used to further categorize the event and hypothesize their generation mechanisms. Using this method, we characterized transient events that occurred during the ICEX-16 experiment based on their duration and bandwidth. Our results suggest that ice cracking during this time likely happened on small temporal and spatial scales. Small fractures occurred regularly, whereas larger cracking events and ice floe interactions were more rare. When larger cracking events did occur, they tend to have happened in clusters, resulting in a sudden increase in transient noise generation.

We further tested our algorithm on two datasets with known ground truth labels. It achieved a TPR (FPR) of 70.9 (13.9)% and 73.9 (21.7)%, respectively. For future improvement, less empirical metrics for event categorization can be developed based on knowledge of the spectral signature of the transients of interest. This would allow for more specialized identification of events and further improvement of detection accuracy.

Chapter 6

Ice Cover Transient Noise Monitoring

In the previous chapters, we have analyzed experiment data and presented methods for characterizing underwater ambient noise that facilitate a broad description of noise generation range and mechanisms by which transient noise events are produced in the ice cover. Combined, these two components form an overall description of the far field ambient noise-scape and how changes to the soundscape are indicative of environmental changes. In addition to the work presented so far, another approach to monitor Arctic ambient noise activity, specifically its generation in the ice cover, is through the deployment of geophone arrays on the ice surface. Such a system can document transient events generated by sea ice dynamics as well as enable localization of the events to study their spatial distribution over an area of interest. Specific focus on the recorded waveforms of these transient events also means a closer examination of near field noise propagation that contributes to the overall soundscape. This aspect of ice generated noise was not a focus of the previous chapters but, nonetheless, an interesting part of the ambient soundscape worthy of study.

With these motivations in mind, in this chapter, we outline a seismo-acoustic system consisting of geophones and hydrophones to monitor ice cover dynamics and the resulting transient noise activity. This system was deployed on Elson Lagoon, north of

Utqiagvic, AK, as part of the Sea Ice Dynamics Experiment (SIDEx) during January to March of 2020. We first provide some background on wave propagation in the ice cover before presenting the deployed array and an example of a recorded event during the experiment. Then, continuing with our presentation of data analysis algorithms, we describe two forward approaches for event localization that primarily leverage time-difference-of-arrival (TDoA) of the event signal on different receiver elements. We test the performance of these approaches using a calibration dataset collected during the experiment and compare with a more conventional, inverse method. We further investigate the utility of remote node geophones for ice cover seismo-acoustic monitoring. Because of their cable-less nature, these units can be easily deployed to form an array that covers a much larger area than what is possible with cabled geophones. Using the remote geophones deployed, we provide a localization estimate for a natural event that occurred during the experiment. We conclude by suggesting improvements for the next iteration of this system based on lessons learned during this deployment.

6.1 Background

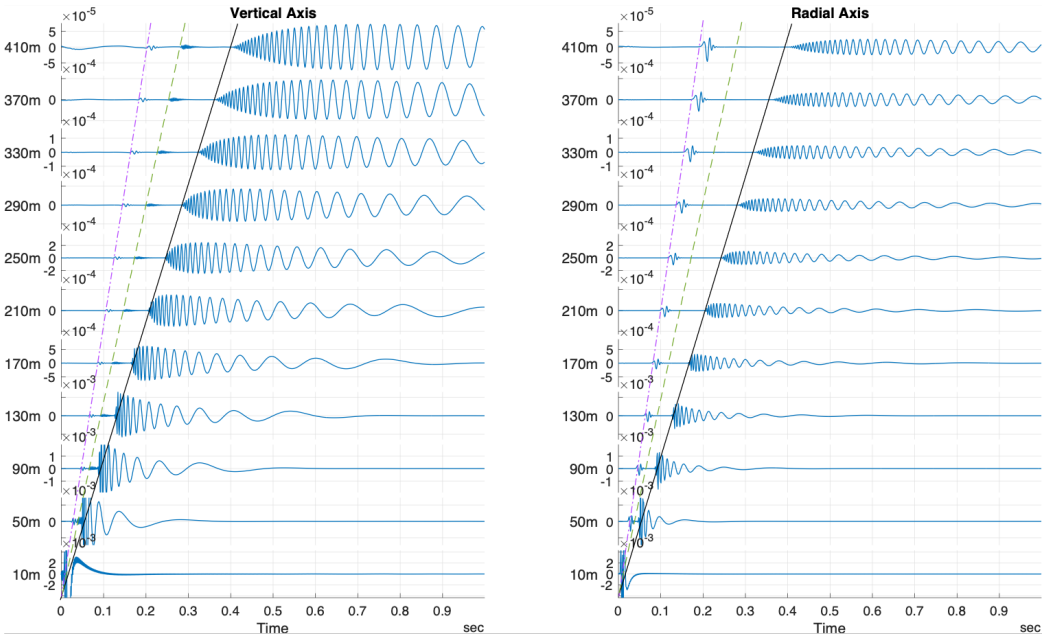
6.1.1 Sea Ice Dynamics Studies

The toolkit for studying sea ice dynamics is quite abundant. For example, stress and strain sensors [167, 168, 169], ice buoys [170, 171, 172, 173], and satellite-based observations or radar interferometry [174, 175, 176, 177, 178] are the standard. In addition to these, other works describe another tool - using passive acoustic and seismic measurements on and below the ice cover to characterize ice properties and localize transient events [179, 180, 181, 182, 183, 184, 185, 186]. The challenges associated with deploying and maintaining a seismo-acoustic system in the Arctic are significant. The extreme environment, processing and storing of large amounts of data, and the interdisciplinary knowledge required for data analysis all contribute to the difficulty of long-term monitoring of ice dynamics. One recent deployment of this

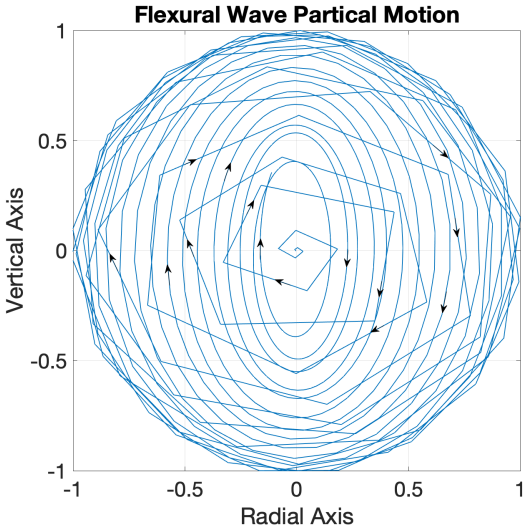
type was actually during the SIMI-94 experiment, in which an array of geophones and hydrophones was deployed on and under the Beaufort sea ice cover to localize transient noise events caused by ice cracking [11]. In the SIDEx experiment, the goal is to again deploy a system based on the SIMI-94 arrays and further explore the capability of such a system to localize transient noise events in the ice cover, which would construct a more complete picture of ice dynamics and noise generation.

6.1.2 Wave Propagation in the Ice Cover

Due to its elastic nature, wave propagation in the ice cover is much more complex than in water. Three fundamental wave types are observed to propagate in the ice cover - longitudinal (P), horizontally polarized transverse (SH), and flexural waves [187, 188, 189, 190, 191]. The longitudinal and transverse waves travel at the compressional and shear speeds in the ice cover, respectively. They are faster than the flexural wave, which propagates as the result of bending of the ice over. In terms of amplitude, the flexural wave is the most prominent in the time series and thus commonly the focus of analysis. One property of the flexural wave is frequency dispersion, meaning the wave's propagation speed depends on frequency. In a time series, the resultant waveform shows high frequencies arriving first, followed by lower frequencies. Another flexural wave property is elliptical particle motion - its vertical and radial components are $\pi/2$ out of phase and vertically polarized in the plane containing the source and receiver. Fig. 6.1(a) shows an example of wave arrivals at various ranges in both the vertical and radial directions for a simulated source at the surface of a 1 m thick ice cover. The first arrival is the P wave, followed by the SH wave, and then the dispersive flexural wave. Fig. 6.1(b) shows the particle motion of the flexural wave arrival, which exhibits an elliptical shape. Theses figures are generated using OASES [35] with the parameters shown in Table 6.1.



(a)



(b)

Figure 6.1: (a) Simulated time series move-out of a vertically forced source on the surface of a 1 m thick ice cover with parameters presented in Table 6.1. Dashed-dot line marks the move-out of the P wave, dashed line marks the SH wave, solid line marks the flexural wave. (b) Particle motion of the flexural wave component of the simulated event showing complex elliptical polarization (horizontal ellipses for early, higher frequency arrivals turning to vertical ellipses for later, lower frequency arrivals).

Table 6.1: Parameters used to model example wave arrivals in an ice cover. C_p denotes compressional speed, C_s denotes shear speed, ρ denotes density, η_{rms} denotes root-mean-square roughness, η_{cl} denotes roughness correlation length.

Medium	C_p	C_s	ρ	η_{rms}	η_{cl}
Vacuum halfspace	-	-	-	-	-
Ice layer (top)	3000 m/s	1200 m/s	0.9 g/cm^3	-	-
Ice layer (bottom)	2400 m/s	600 m/s	0.9 g/cm^3	0.2 m	19.1 m
Water halfspace	1435 m/s	-	1.0 g/cm^3	-	-

6.2 Experiment Setup

6.2.1 System Overview

The seismo-acoustic system deployed during SIDEx consists of two parts: a central cabled array with four 3-axis geophones and two hydrophones, plus an un-cabled array with four independent seismic recording units (Figure 6.2). The geophones (both cabled and stand-alone) were frozen onto the surface of the ice cover using a bit of water. For the two hydrophone units, one was deployed on the surface of the ice (in air and covered by snow) while the other was lowered into the water column ∼1 m below the ice cover. The purpose of this configuration was to test whether we can detect transient event arrivals at the air and water propagation speeds, respectively. Each of the 3-axis geophones measure the velocity of the ice motion in a Cartesian coordinate system to form recorded time series. These measurements can be represented as $\vec{v_i} = [\vec{v_{xi}}, \vec{v_{yi}}, \vec{v_{zi}}]'$ and $\vec{v_{xi}} = [v_{xi}[0], v_{xi}[1], ...]$, where i denotes the numbering for each element and x, y, z denote the three axes. The geophones were oriented with the longitudinal axis pointed approximately East-West and the transverse axis pointed approximately North-South.

At the Elson Lagoon deployment site, the ice cover was thin during the experiment, ∼1 m thick. The water depth beneath the ice is quite shallow, approximately 3-4 m deep. Below the water column, the sub-bottom consists of mostly muddy sediment. A crack in the ice cover was observed at the site, which ran diagonally through the

deployed system (Fig. 6.2).

Cabled Array

The cabled geophone array and hydrophones were deployed near the center of the overall system. They are denoted as G0-G3 and H0-H1, respectively, in Fig. 6.2. The geophone array was configured as a center unit (G3) with three peripheral units (G0-G2), each approximately 50 m away from the center unit. The hydrophone elements were co-located with the center geophone. All elements recorded at a sampling rate of 1000 Hz. The collected data were transmitted wirelessly via long-range wi-fi to a shore-side laptop, which automatically backed-up the data to a dedicated online drive. The entire cabled system was powered using eight batteries that were periodically recharged by a 600 W wind turbine. The power requirements for wireless communications and data transfer were significant (10 W); as a result, the wireless link time was limited to 1 hour per day. Thus, the recorded data was also stored locally on a SD card in the system.

Remote Array

The remote, un-cabled array also consisted of four 3-axis geophone units, labeled S0-S3 in Fig. 6.2. Because the units are not tethered by cables, they were able to be more easily deployed to form a larger array. Similar to the cabled array, the remote array was again configured as a central unit (S3) with three peripheral units (S0-S2), each approximately 1000 m away from the center. The center S3 unit was positioned immediately next to unit G3 from the cabled array. Each remote unit is a Sigicom INFRA C22 wireless vibration monitor, which recorded at 4096 Hz and was powered by an external battery. The data recorded at each unit was stored locally on a SD card. To our knowledge, this was the first time that these units have ever been used for ice cover seismo-acoustic measurements; they are generally used for monitoring building vibrations. Thus, testing the viability of these units for deployment on the ice was an important function of this experiment.

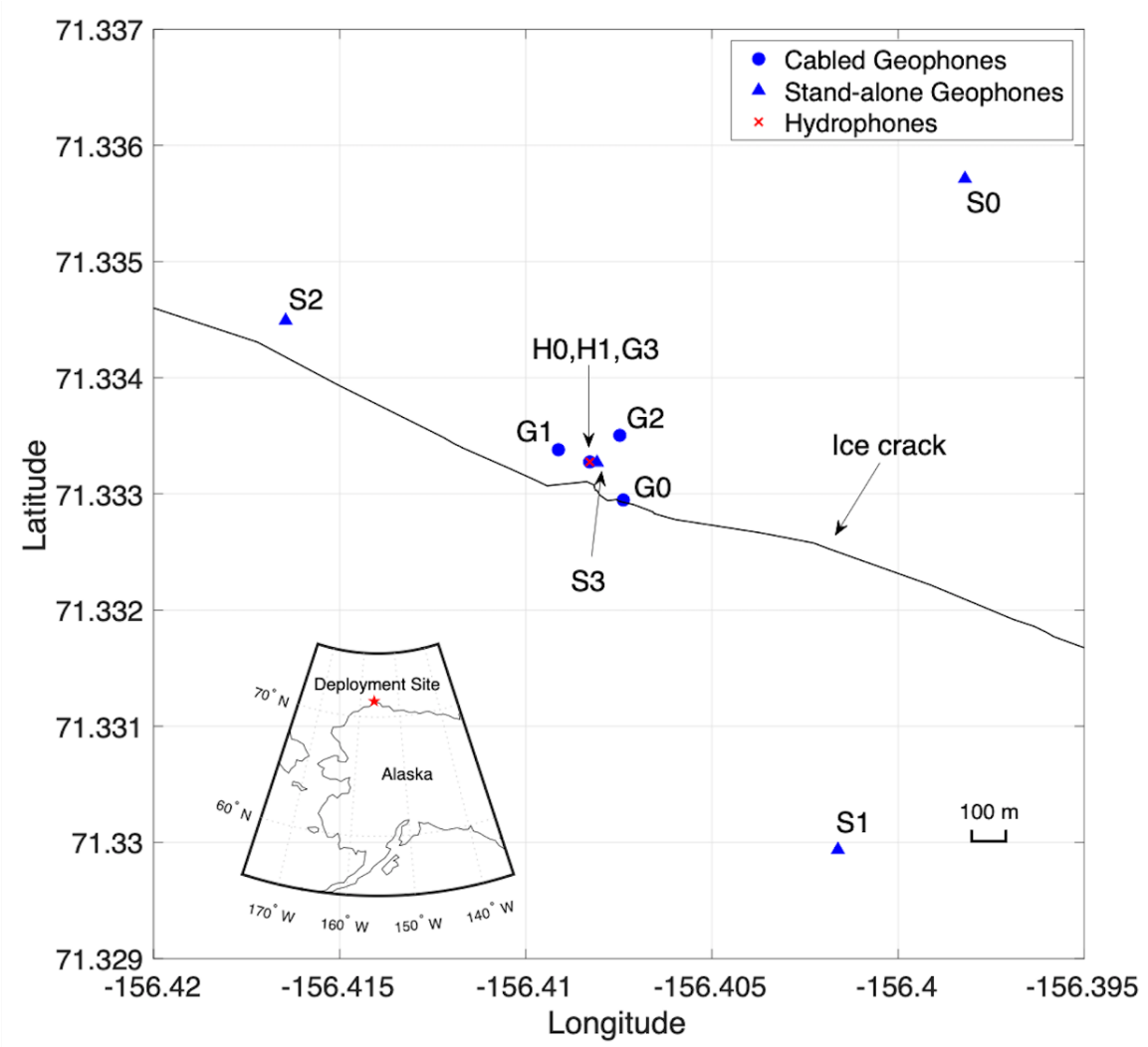


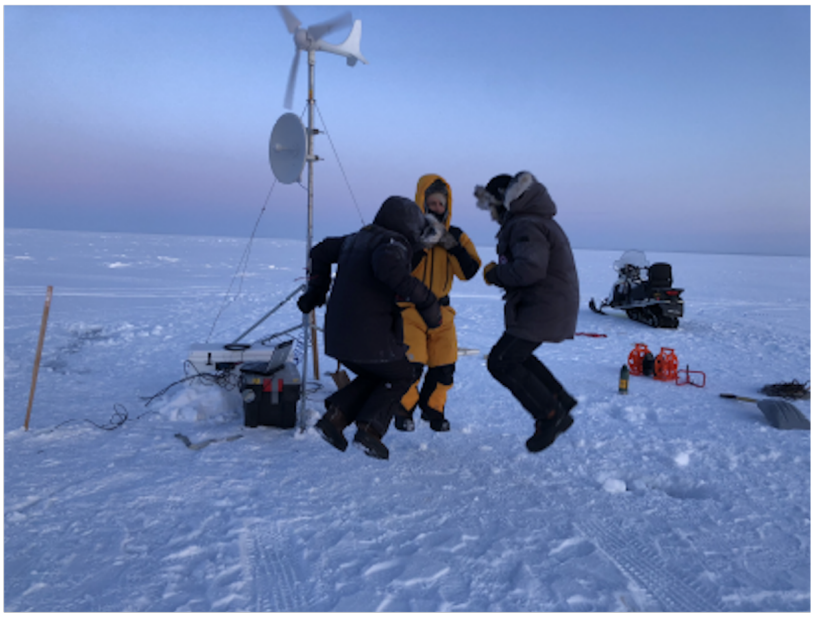
Figure 6.2: GPS locations of the cabled geophones, hydrophones, and stand-alone geophones in the deployed system. The location of the experiment site and an observed crack in the ice cover are also shown.

6.2.2 Calibration Dataset

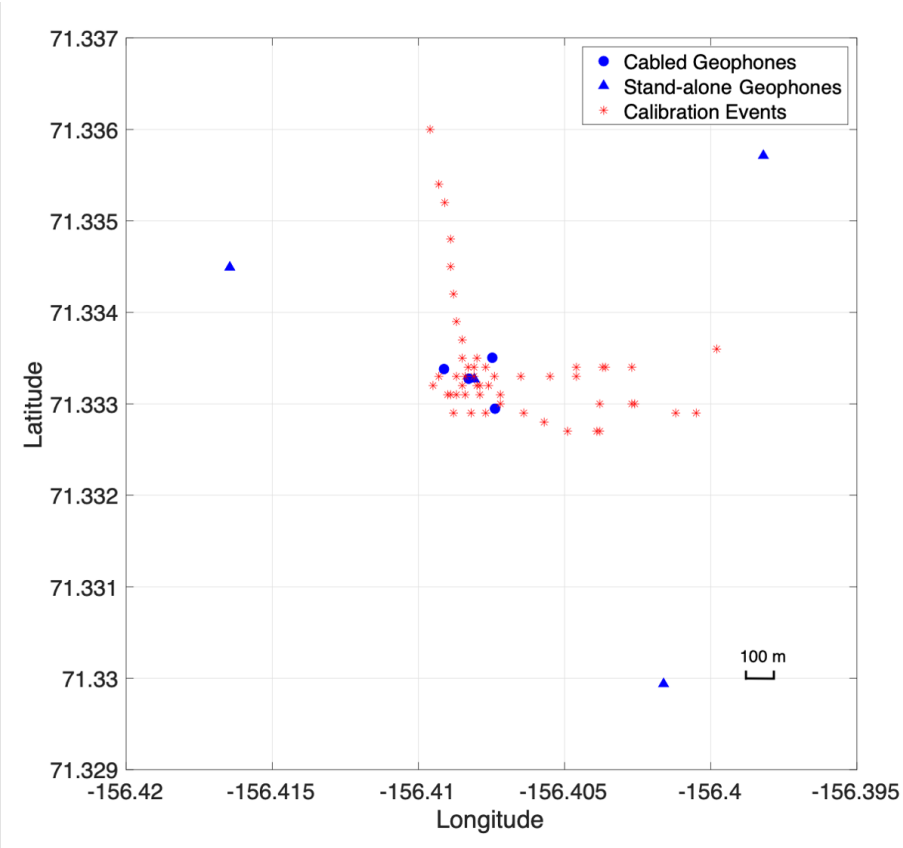
A calibration dataset was recorded after system deployment. This dataset serves three purposes: first, it allows us to estimate the flexural wave propagation speed in the ice cover by analyzing the TDoA of the calibration events at each geophone unit. Second, it provides ground truth for testing our localization algorithms presented in later sections. Finally, this dataset can also demonstrate any limitations of the deployed system itself and help inform improvements for future system design. To carry out the calibration, the field team created impulsive events on the ice by synchronously jumping and landing at various locations (Fig. 6.3(a)). An accelerometer and GPS combination was used to record the location and timing of the jumps. This calibration method was selected as it was the most feasible way to impart a large force into the ice cover to simulate a transient event without having to transport weights to the site or arrange for explosives permits.

Calibration events were generated at 50 locations around and within the cabled array perimeter, at ranges of up to 400 m from the array center (Fig. 6.3(b)). At each location, 4 impulses (jumps) were generated. Unfortunately, since most of the events were near the cabled array and far away from the remote units, many events are only present on the cabled array units. Thus, only the cabled array calibration data were analyzed. In addition, while the jump method of calibration may not fully replicate the form or intensity of the signals seen in natural cracking events, it does provide a clean, clear signal that can be used to estimate the propagation speed of flexural waves through the ice cover.

Of the 200 calibration events, 188 are clearly visible on the cabled array recordings. As an example, calibration event 1's location and time domain move-out on the cabled array geophones are shown in Fig. 6.4(a,b). On the time series plot, we see only the flexural wave component is clearly observable. As noted previously, this component is much more prominent than the P and SH waves. Furthermore, the signals on the vertical axes are much stronger than the arrivals on the radial axes. This is likely due to the fact that the calibration events impart much more vertical energy into



(a)



(b)

Figure 6.3: (a) Method for generating calibration events - 3 members of the field team jumped up and down 4 times at each of the 50 locations. (b) Locations of all calibration events.

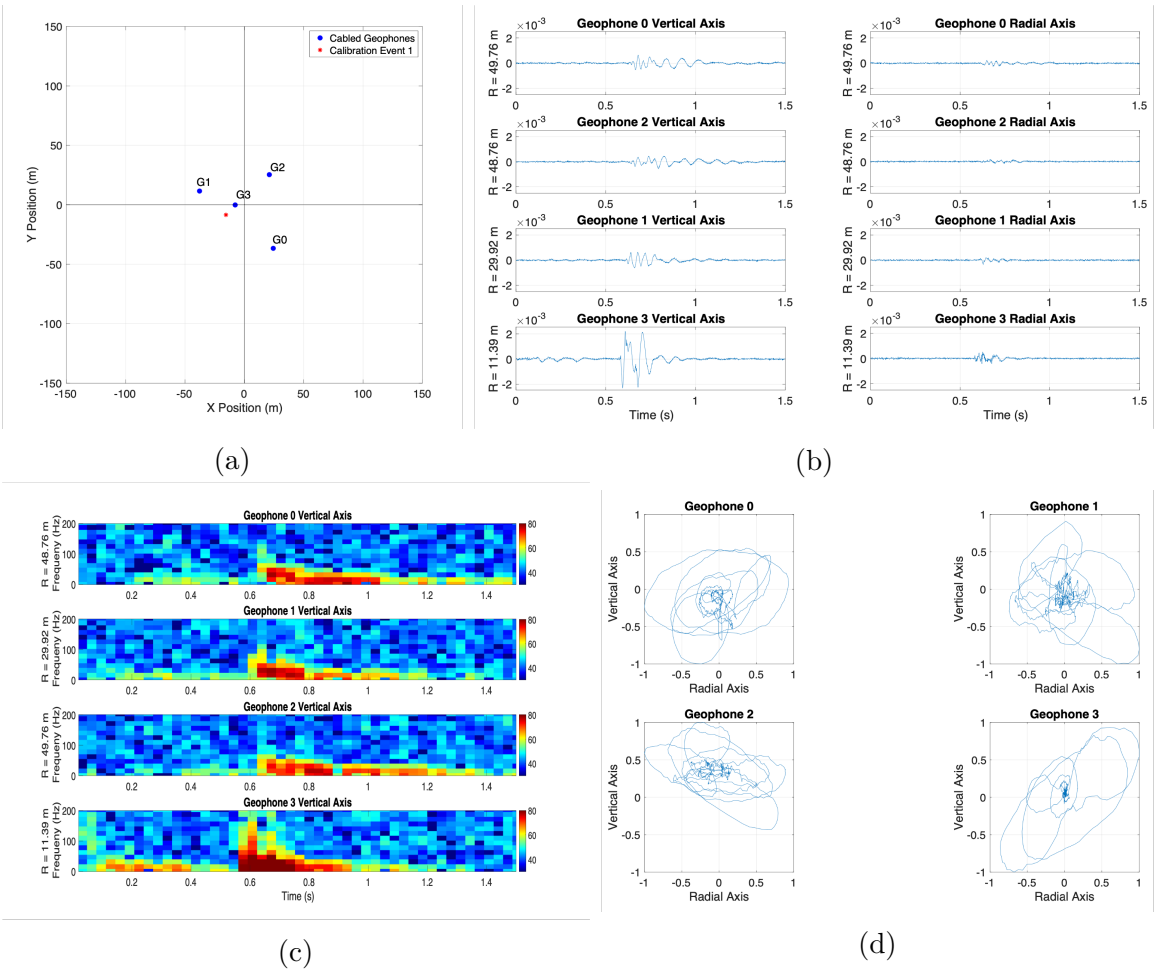


Figure 6.4: (a) Calibration event 1's location with respect to the cabled array center. (b) Time domain move-out on each cabled geophone element; the radial component is a combination of the x-y axes data. (c) Event spectrogram from the vertical axis of each element showing the dispersive nature of the flexural wave arrival. (d) Particle motion on each element showing close-to-elliptical polarization.

the ice cover than in the radial direction. We can confirm the waves observed in the time series are flexural by plotting their spectrograms (Fig. 6.4(c)), which highlight the signals' dispersive nature. Similarly, the particle motion of the arrivals on all channels are close to elliptical (Fig. 6.4(d)). Although, the ellipses are less well defined compared to the modeled event shown in Fig. 6.1(b).

6.3 Hydrophone Data

The motivation for deploying two hydrophones along with the geophone arrays was to test whether we would observe the airborne and waterborne arrivals of transient events in addition to waves propagating in the ice cover. Accordingly, one hydrophone was placed on the surface of the ice cover while the other was lowered into the water column, 1 m below the ice. Although we were skeptical about seeing airborne arrivals on the surface hydrophone, we did expect to observe waterborne arrivals, propagating at the speed of sound in the water ($\sim$1430 m/s), in the lowered hydrophone. However, upon inspection of the recorded calibration events on the hydrophones, we find that, similar to the time series recorded on the geophones, only the ice cover flexural wave is prominently visible. Fig. 6.5 demonstrates this by showing the time series of calibration event 1 recorded on the hydrophones and the co-located G3 geophone. As we mentioned, only the dispersive flexural wave is observed on all three units. The in-air, surface hydrophone time series contain significantly more high frequency noise than the in-water hydrophone data. This is expected as the hydrophones were not designed to be deployed in air and the surface unit was likely more affected by other environmental noise such as from wind.

Faced with this result, we suspect the reason only the ice cover flexural wave is observed on the hydrophones is because they were deployed too close to the ice cover. As a result, the flexural wave arrival is much stronger than the air or waterborne arrivals and effectively obscures them in the time series. Particularly, if we were able to deploy the in-water hydrophone lower in the water column, the increased distance from the ice cover would have allowed the flexural wave to dissipate and we would then be able to observe the waterborne pressure wave arrival. To test this theory, we modeled received time series caused by a surface source at 200 m range on hydrophones placed at various depths in an environment with a 1 m thick ice cover, 1000 m deep water column, and a muddy bottom halfspace (Table 6.2). The modeling results are shown in Fig. 6.6 and they confirm our hypothesis. At depths within $\sim$10 m of the ice cover, only the flexural wave is observed in the hydrophone time series;

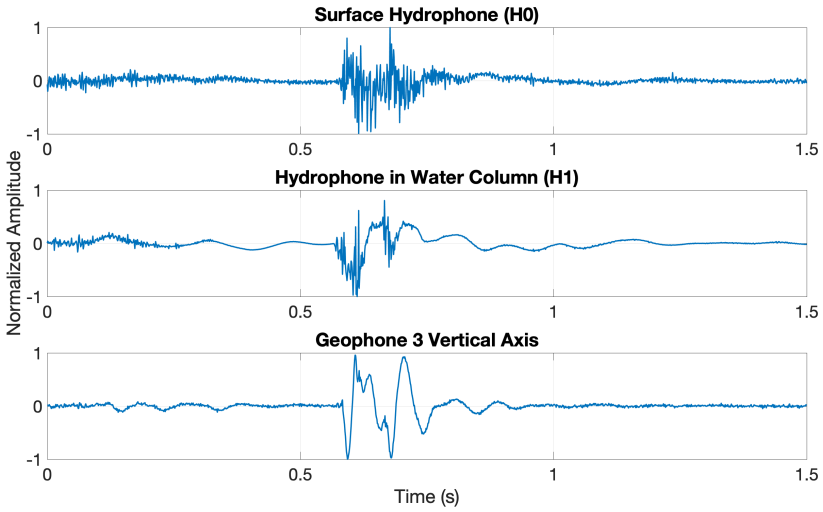


Figure 6.5: Calibration event 1 time series recorded on the hydrophones and the co-located G3 geophone. Only the flexural wave arrival is clearly observed on all three elements.

not until below 10 m depth do we start to observe the waterborne pressure wave arrival. Thus, in order to make use of waterborne propagation in our data analysis, we needed to deploy the in-water hydrophone deeper in the water column. This was not possible at the Elson Lagoon experiment site as the water depth is only ∼3-4 m. However, this finding will help inform the design of future system deployments.

Table 6.2: Parameters used to model example wave arrivals at hydrophones. C_p denotes compressional speed, C_s denotes shear speed, ρ denotes density, η_{rms} denotes root-mean-square roughness, η_{cl} denotes roughness correlation length.

Medium	C_p	C_s	ρ	η_{rms}	η_{cl}
Vacuum halfspace	-	-	-	-	-
Ice layer (top)	3000 m/s	1200 m/s	0.9 g/cm^3	-	-
Ice layer (bottom)	2400 m/s	600 m/s	0.9 g/cm^3	0.2 m	19.1 m
Water column	1435 m/s	-	1.0 g/cm^3	-	-
Bottom halfspace	1550 m/s	100 m/s	1.6 g/cm^3	-	-

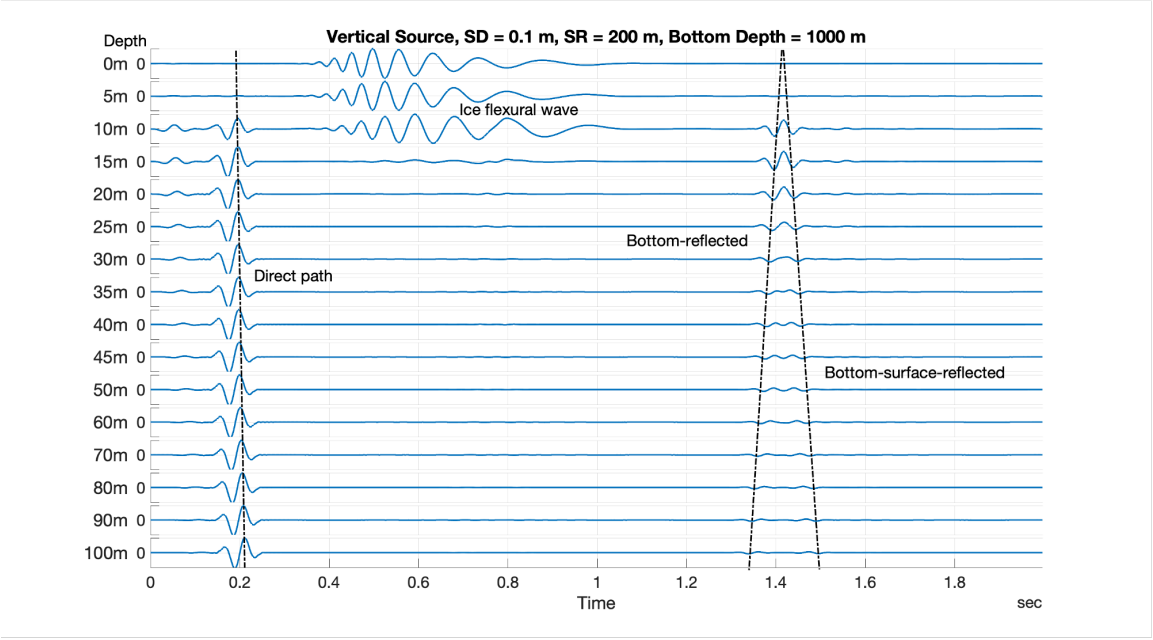


Figure 6.6: Modeled time series recorded on hydrophones at various depths for a surface source in the ice cover at 200 m range. For hydrophones near the surface, only flexural waves are clearly observed. The waterborne pressure wave is not visible until below 10 m depth. Note, the time series amplitudes are normalized.

6.4 Propagation Speed Estimation

As mentioned previously, one purpose of collecting the calibration dataset is to estimate the flexural wave propagation speed (*i.e.* group speed) in the ice cover. This information will then allow us to apply our localization algorithms later. Since flexural waves are dispersive, we need to make group speed estimates as a function of frequency. To do so, we first pass each calibration event through a bank of bandpass filters. We selected the center frequency of these filters at 2 Hz increments between 2-50 Hz and set the bandwidth of the filters equal to 8 Hz. This frequency interval was selected after computing the PSD of each calibration event arrival on each of the three axes of the cabled geophone elements. The PSDs are calculated by segmenting the data into snapshots of 128 samples with 50% overlap and then performing a FFT. Fig. 6.7(a) presents the results. The colored dots denote the PSDs for individual events while the solid lines show the mean PSD of all events on each of the three axes. From this plot, we observe the events have most of their energy below 50 Hz

with a peak near 16 Hz. As a result, we limit our bandpass filters to below 50 Hz.

After filtering, the Hilbert transform of each event is calculated, which provides the complex envelope of the filtered time series. The absolute maximum of this envelope is then used to denote the arrival time of an event at a geophone element. Following this, a group speed estimate at the filtered frequency band can be calculated for each geophone pair in the cabled array by dividing the distance between the geophones by the event's TDoA at the elements. For example, for two geophones i and j and a filtered calibration event on the vertical axis, $z_f(t)$, where f denotes frequency, the group speed estimate is computed as

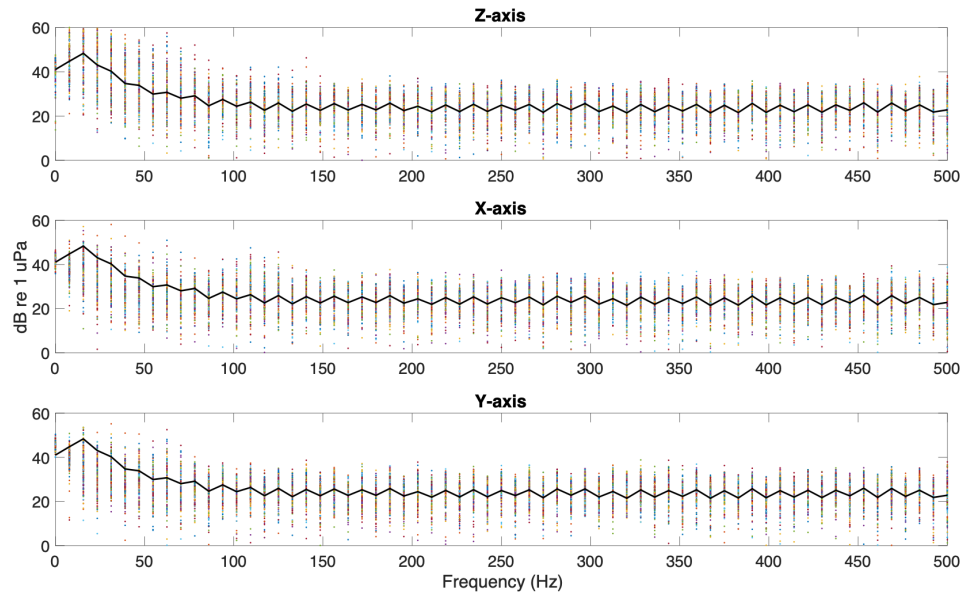
$$v(i,j,f) = \frac{D_{ij}}{\tau(i,j,f)}. \tag{6.1}$$

Here, D_{ij} is the distance in meters between geophones i and j. $\tau(i,j,f)$ is the TDoA of $z_f(t)$ at the two geophones in seconds. It, as noted, is calculated as

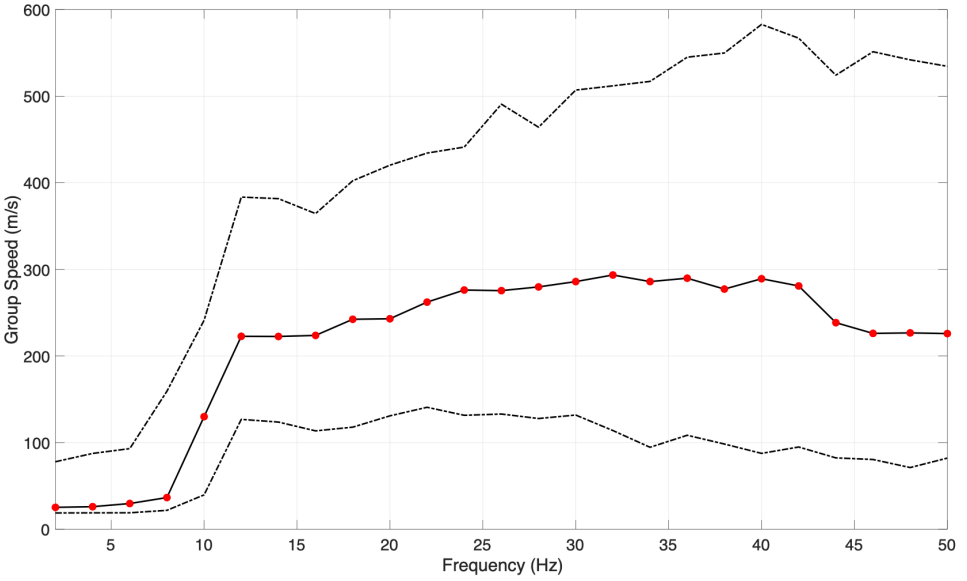
$$\tau(i,j,f) = |\max_t(|H(z_{i,f}(t)|) - \max_t(|H(z_{j,f}(t)|)|, \tag{6.2}$$

where H denotes the Hilbert transform and $z_{i,f}(t)$ and $z_{j,f}(t)$ represent the arrival waveform of $z_f(t)$ on geophones i and j, respectively.

Repeating this calculation for all calibration events on all three axes results in a distribution of group speed estimates for each frequency bin. Fig. 6.7(b) shows the median of this distribution plotted as a function of frequency; this plot also displays the 25th and 75th percentiles in dashed-dot lines. As expected, the group speed varies with frequency. Below 10 Hz, the speed is low (<100 m/s). Above 10 Hz, the median estimate increases to between 200-300 m/s. At the PSD peak frequency of 16 Hz, the middle 50% of group speed estimates vary between ∼100-400 m/s. With this result, we now have an estimated interval of group speeds to use for event localization.



(a)



(b)

Figure 6.7: (a) PSD estimates of all calibration event arrivals. Colored dots represent estimates for individual events, solid line denotes the mean estimate with a peak in power near 16 Hz. (b) Flexural wave group speed estimates vs. frequencies. Solid line denotes median and dashed-dot lines denote the 25th and 75th percentiles.

6.5 Localization Algorithm

In this section, we introduce the components of our transient event localization algorithm. We again use calibration event 1 as an example to demonstrate the output of each component. We then present how the components are combined to form a location estimate. First, a note regarding input data pre-processing. Because the flexural wave is dispersive, we need to analyze the data in a narrow frequency interval when using TDoA. Thus, we choose to bandpass filter the input data to between 12-20 Hz prior to applying our algorithm. The reason for this selection is because, as demonstrated in the previous section, the peak power in the signal arrivals occurs at 16 Hz, the center frequency of this interval.

6.5.1 Motion Product Detector

One component of our event localization approach is the application of the motion product detector (MPD). First described by White [192], MPD enhances waves with a specified particle motion while also indicating the bearing of arrival for that wave type. Recall that for transient events recorded by the geophone array, flexural waves are the predominant arrival and they have elliptical particle motion with $\pi/2$ phase difference between the vertical and radial axes. Thus, we can apply MPD to further enhance flexural wave arrivals within each event and estimate the event's direction of arrival. We follow the method described by Dudko [11] in their previous analysis of geophone data collected during SIMI-94. Briefly, it works as follows:

1. Given an event arrival, for each geophone element, shift the vertical component (Z) of the signal time series in phase by $\pi/2$ radians. This is done by taking the imaginary component of the signal's Hilbert transform:

$$Z^{\frac{\pi}{2}} = \Im(H(Z)). \tag{6.3}$$

2. Multiply each of the horizontal time series components (X, Y) by the shifted vertical component ($Z^{\frac{\pi}{2}}$). This step enhances the flexural wave arrival.

3. The product of each horizontal component with the shifted vertical component is then averaged through integration over a short time window. The window length was selected empirically as 0.2 s. The averaged result of the $XZ^{\frac{\pi}{2}}$ component product is denoted as X_{HiV} while that for the $YZ^{\frac{\pi}{2}}$ component product is denoted as Y_{HiV}. Formally, these two terms are expressed at time t as:

$$X_{HiV}(t) = \frac{1}{T_1} \int_{t-\frac{T_1}{2}}^{t+\frac{T_1}{2}} XZ^{\frac{\pi}{2}} dt, \tag{6.4}$$

$$Y_{HiV}(t) = \frac{1}{T_1} \int_{t-\frac{T_1}{2}}^{t+\frac{T_1}{2}} YZ^{\frac{\pi}{2}} dt, \tag{6.5}$$

where T_1 is the averaging interval.

4. To estimate the bearing from a geophone element to the source of the event, plot Y_{HiV} against X_{HiV} for unit. Since the flexural wave is vertically polarized in the plane containing the source and receiver, the line of best fit of this scatter plot naturally denotes the direction of arrival estimate for the event. This is done using the method of least squares [193].

By applying MPD, a series of bearing lines can be plotted for an event, one for each geophone element in the array. Examples of these lines plotted for calibration event 1 are shown in Fig 6.8. These MPD lines are incorporated into our localization algorithm to make a final localization estimate for each calibration event.

6.5.2 Time Difference of Arrival

The second component of our event localization algorithm makes use of the TDoA of an event between each unique geophone pair in the array to estimate the event's origin. As before, the TDoA of an event between any geophone pair is calculated following Eq. 6.2. For analysis of the calibration events, we use the z-axis time series to determine the TDoAs for each event since the signal arrivals are the most prominent on the vertical axis.

For each event, as before, we denote the TDoA calculated for each geophone pair

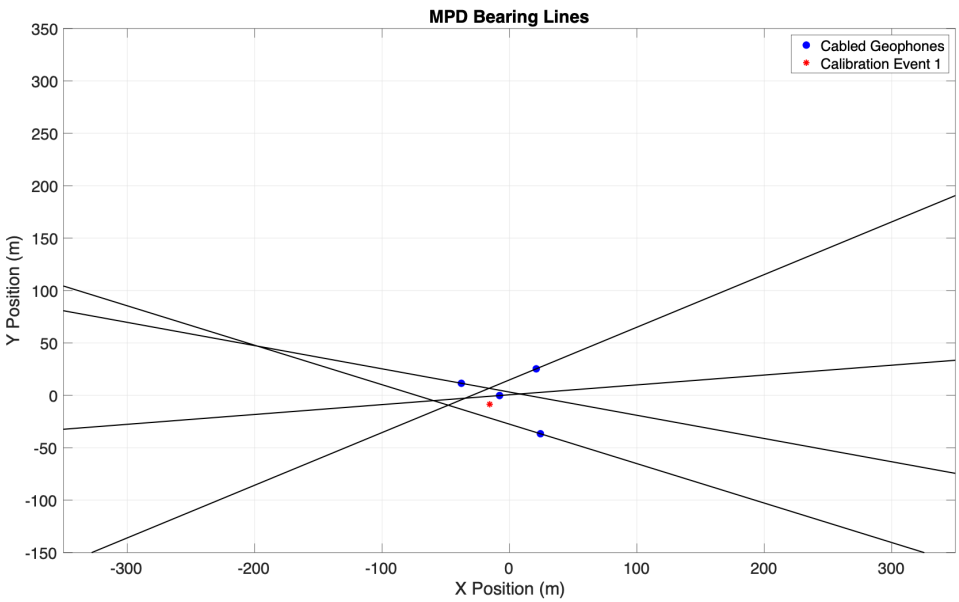


Figure 6.8: MPD bearing lines for all four geophone elements in the cabled array. Red dot denotes the true location of calibration event 1.

as τ. To localize an event, we need to find the locus of points, $\mathcal{P}$, in Euclidean space such that for any point P within the locus, the absolute difference in distance from P to the locations of the two geophones (N_1, N_2) is equal to $c_0\tau$, where c_0 is the assumed propagation speed. Formally, this is defined as

$$\mathcal{P} = \{P : |PN_1 - PN_2| = c_0\tau\}, \tag{6.6}$$

where PN_1 and PN_2 are the distances from P to N_1 and N_2, respectively. This locus of points is exactly the definition of a hyperbola in $\mathbb{R}^2$. Thus, for each geophone pair in the array, we can plot a hyperbola that represents the set of possible source locations.

For our purpose, the equation for a hyperbola in the geophone-centric coordinate frame (x', y'), where the midpoint between a pair of geophones is the origin and both geophones are positioned along the x'-axis can be written as

$$\frac{x'^2}{a^2} - \frac{y'^2}{c^2 - a^2} = 1. \tag{6.7}$$

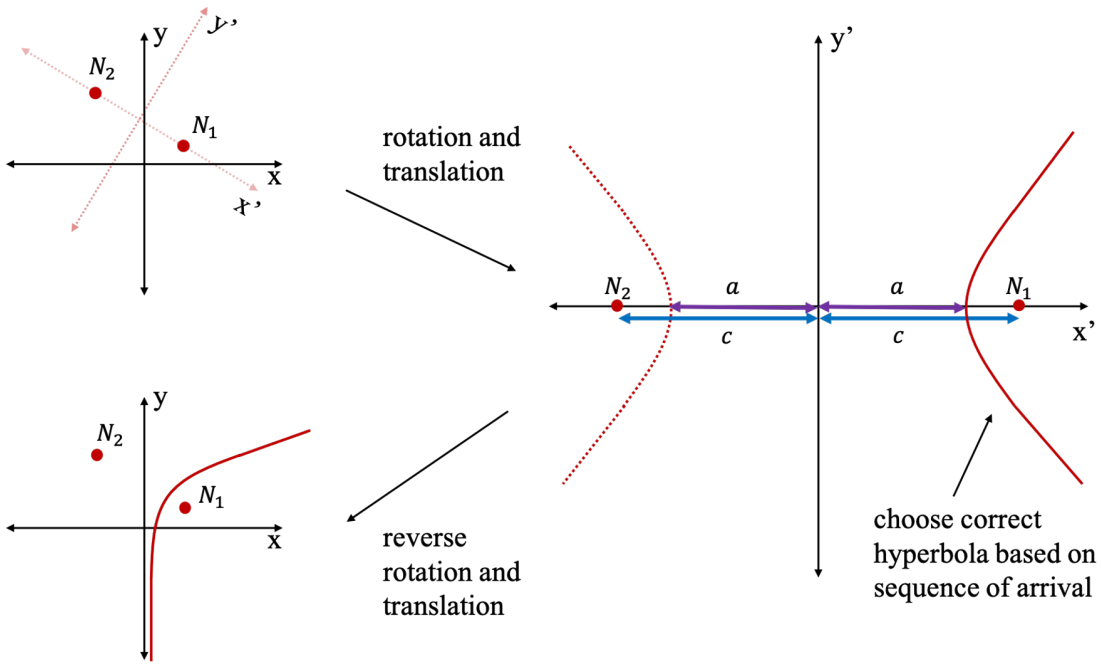


Figure 6.9: Process for plotting a hyperbola for a geophone pair in the geophone frame (x', y') and transformation to the array frame (x, y). Red dots denote the location of the geophones.

By definition, $a = \frac{1}{2}c_0\tau$ and c is the distance between each geophone in the pair to the midpoint along the x'-axis. Naturally, this equation results in two hyperbolas. However, the correct hyperbola can be determined by examining at which geophone the event first arrived. Furthermore, because the resultant hyperbola is in the frame of the geophone pair, a coordinate rotation and translation must also be performed to correctly plot the hyperbola in the array-centered frame. This transformation is demonstrated in Fig. 6.9.

Similar to the MPD bearing lines, the resultant hyperbola for each geophone pair in the array is incorporated in the localization algorithm to make an event location estimate. As an example, the plotted hyperbolas for calibration event 1 are shown in Fig. 6.10. The four geophones in the cabled array form 10 unique pairs; thus, there are 10 hyperbolas. The method for selecting the propagation speed, c_0 when forming the hyperbolas for an event is discussed in the following section.

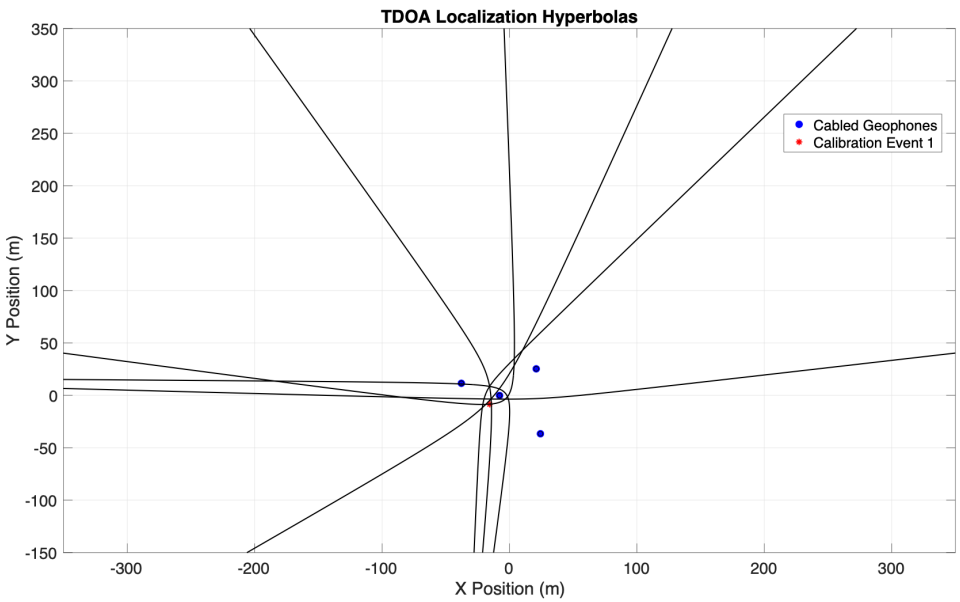


Figure 6.10: TDoA hyperbolas for all unique geophone pairs in the cabled array assuming a propagation speed of 242 m/s. Red dot denotes the true location of calibration event 1.

6.5.3 Localization using Intersections

Intersections Method

With the MPD lines and TDoA hyperbolas, a natural approach to making a localization estimate is to consider all of their intersections. To do so, we identify all intersections of the lines and hyperbolas following a method described by Schwarz [194]. Consider two curve segments C_1 and C_2 with endpoints $[(x_1(1), y_1(1)), (x_1(2), y_1(2))]$ and $[(x_2(1), y_2(1)), (x_2(2), y_2(2))]$, respectively. To find any intersections of the two curves, one can compose four equations with four unknowns. The four unknowns are t_1, t_2, x_{int}, and y_{int}, where (x_{int}, y_{int}) is the intersection of C_1 and C_2, t_1 is the distance from an end point of C_1 to the intersection point relative to the length of C_1, and t_2 is the distance from an end point of C_2 to the intersection point relative

to the length of C_2. Accordingly, the four equations are

$$\begin{aligned}(x_1(2) - x_1(1))t_1 &= x_{int} - x_1(1),\\ (x_2(2) - x_2(1))t_2 &= x_{int} - x_2(1),\\ (y_1(2) - y_1(1))t_1 &= y_{int} - y_1(1),\\ (y_2(2) - y_2(1))t_2 &= y_{int} - y_2(1).\end{aligned} \tag{6.8}$$

These equations can be written in matrix form, $AT = B$, as

$$\begin{bmatrix} x_1(2) - x_1(1) & 0 & -1 & 0 \\ 0 & x_2(2) - x_2(1) & -1 & 0 \\ y_1(2) - y_1(1) & 0 & 0 & -1 \\ 0 & y_2(2) - y_2(1) & 0 & -1 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ x_{int} \\ y_{int} \end{bmatrix} = \begin{bmatrix} -x_1(1) \\ -x_2(1) \\ -y_1(1) \\ -y_2(1). \end{bmatrix} \tag{6.9}$$

T can be solved as $A\backslash B$ and the values of t_1 and t_2 can be examined to determine whether C_1 and C_2 intersect. If $0 \leq t_1 < 1$ and $0 \leq t_2 < 1$, then their definitions are satisfied, which means (x_{int}, y_{int}) is indeed an intersection point. Following this approach, we can identify all intersections of two curves by incrementally marching along one of the curves and performing this test at each point along the curve.

We apply this technique to all unique pairings of MPD lines and TDoA hyperbolas to find all intersection points. To derive a localization estimate from these points, we first fit a non-parametric, kernel distribution to their X and Y position values, respectively. We denote these as $K_X(s)$ and $K_Y(s)$. The formulation of the kernel distribution that we use is given by

$$K(s) = \frac{1}{N} \sum_{i=1}^{N} \phi(s - s_i), \tag{6.10}$$

where $[s_1, ..., s_N]$ are the data samples being fit and $\phi(\cdot)$ is the standardized normal distribution. Following this, we compute the localization estimate as

$$[\hat{x}, \hat{y}] = [\max_s(K_X(s)), \max_s(K_Y(s))]. \tag{6.11}$$

Fig. 6.11 shows a demonstration of this approach for calibration event 1. Again, the TDoA hyperbolas are formed in this plot assuming a propagation speed of 242 m/s. The reason for this choice is discussed below.

Refinement of Propagation Speed Estimate

Recall in Section 6.4, we estimated the flexural wave group speed in the ice cover at the experiment site to be between 200-400 m/s at 16 Hz. In this section, we explain how we use our localization algorithm to further refine this estimate. Quite simply, we iterate through the group speed interval at 1 m/s increments and compare our localization estimate using the intersections method at each speed value with ground truth. We then select the propagation speed for each event that results in the lowest estimation error. Using this approach, the propagation speed estimate for calibration 1 is determined as 242 m/s. While performing this iteration does not explicitly help us localize natural events for which we do not have ground truth, it does help to build a better sense of the general propagation speed in the area that the calibration events took place. It is important to note that the propagation speed is likely not spatially uniform in the ice cover. Thus, the speed estimates for these events represent an overall *average* speed from the event location to the geophone units. Fig. 6.12 shows a histogram of propagation speeds estimates for the calibration events as well as a fitted kernel distribution. For the majority of the events, the propagation speed value is closer to the higher end of the 200-400 m/s interval. The peak of the kernel distribution occurs at 378 m/s while the mean of all speed estimates is 336 m/s.

6.5.4 Ambiguity Map Generation

While the intersections method is an effective approach for transient event localization, it does have some noticeable drawbacks. The first is that it is indiscriminate to the quality of data used to form the MPD lines and TDoA hyperbolas. High SNR signal arrivals, for which the MPD bearing and TDoA estimates are likely more accurate, are weighted the same as lower SNR signals. As a result, the algorithm may

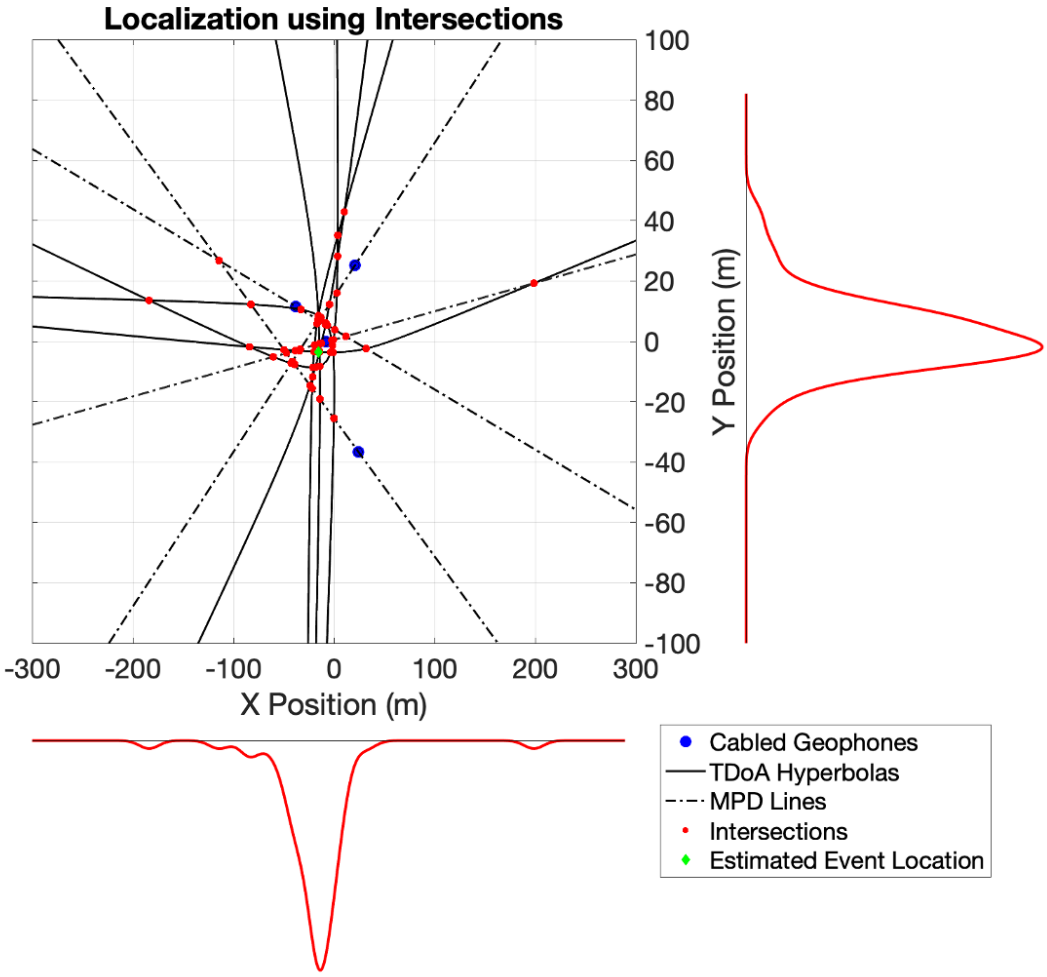


Figure 6.11: Calibration event 1 localization estimate generated using the intersections method. The TDoA hyperbolas are formed assuming a propagation speed of 242 m/s. The distributions along the x and y axes are fitted using the kernel approach.

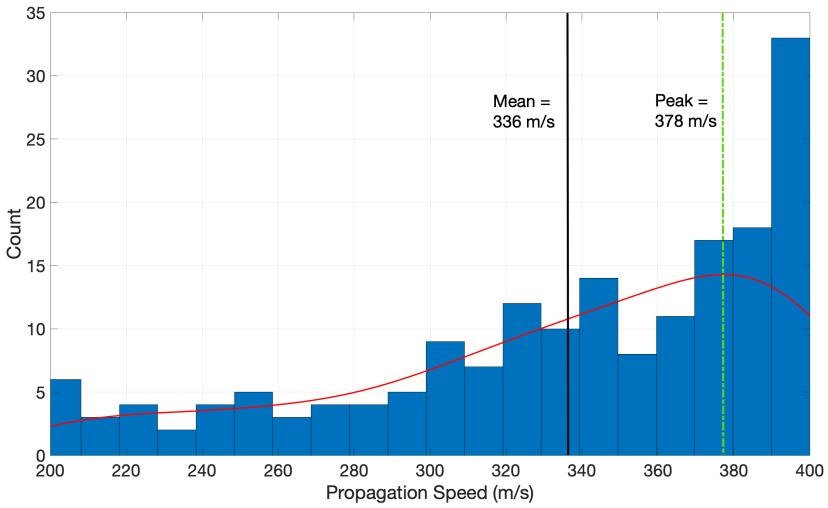


Figure 6.12: Histogram of estimated propagation speeds for all calibration events. Red line denotes a fitted kernel distribution, the solid line denotes the mean of the histogram, the dashed line denotes the peak of the fitted kernel distribution.

be biased by low quality data to make less accurate localization estimates. Another drawback is that the intersections method outputs an estimate without a measure of how good that estimate may be compared to other locations. Ideally, the algorithm's output would be an ambiguity map over an area that shows, by some metric, the likelihood of each position being the event location. With these drawbacks in mind, we present an another approach in addition to the intersections method that combines the MPD lines and TDoA hyperbola to generate an ambiguity map of the event location. This method further weighs the lines and hyperbolas by the SNR of the data used to compute them and is thus more selective on data quality.

To enact this ambiguity map approach, we convert each MPD line and TDoA hyperbola into a 2-D surface. The values of each surface are determined by summing over a Gaussian function formed around each point (x_0, y_0) in the line or hyperbola following

$$G(x,y) = \sum_{(x_0,y_0)} A_0 exp(-(\frac{(x-x_0)^2}{2\sigma} + \frac{(y-y_0)^2}{2\sigma})). \tag{6.12}$$

Two parameters control this Gaussian surface. The variance, σ, is controlled by the

SNR of the recorded event used to generate the line or hyperbola. It is calculated as

$$\text{SNR} = 10log_{10}(S/N), \tag{6.13}$$

where S is the maximum amplitude of the event arrival and N is the noise amplitude estimate computed by taking the average over the entire data time series recorded on a geophone. This definition of SNR is adequate for the bearing lines. However, for the hyperbolas, which are generated from data on a pair of geophones, the SNR is taken as the lesser of the two event SNRs calculated from both geophones. Fig. 6.13 (left) shows the relationship between SNR and variance; it is defined empirically using a logistic function with the design of having high variance at low SNR and vise versa. The parameters of the function are selected after some initial testing with localizing the calibration events.

$$\sigma = \frac{-50}{1 + 20exp(-0.16\text{SNR})} + 53. \tag{6.14}$$

The other parameter of the Gaussian surface is the amplitude, A_0, which is inversely related to variance as shown in Fig. 6.13. This relationship is again determined empirically to ensure event arrivals with high SNR have smaller variances and thus larger amplitudes in their Gaussian surfaces.

$$A_0 = \frac{-1}{1 + 10exp(-0.077\sigma)} + 1. \tag{6.15}$$

After forming a Gaussian surface for each MPD line and TDoA hyperbola, the final ambiguity map ($A(x, y)$) is generated by summing all Gaussian surfaces. A localization estimate is made by finding the maximum of the ambiguity map,

$$[\hat{x}, \hat{y}] = \max_{(x,y)}(A(x, y)). \tag{6.16}$$

Fig. 6.14 shows the ambiguity map for calibration event 1. Following the intersections

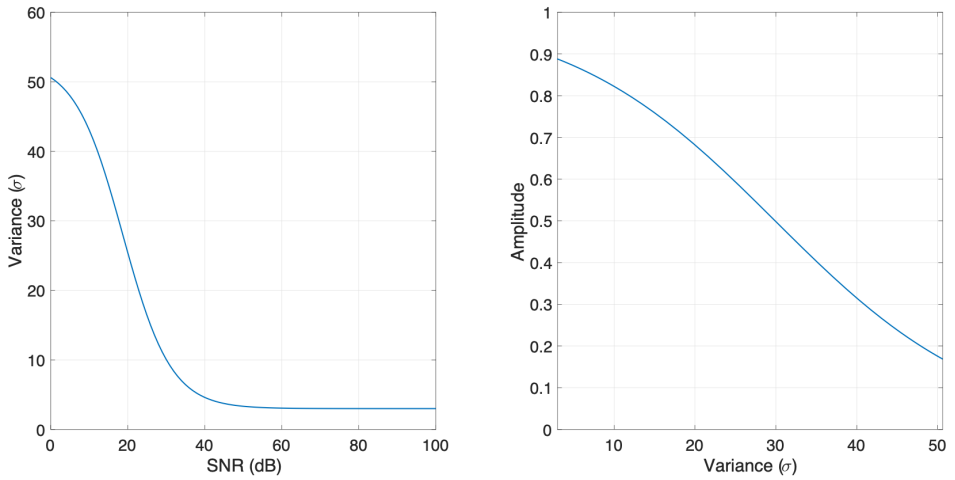


Figure 6.13: (Left) The variance of a Gaussian surface depends on the SNR of the event(s) used to compute the MPD line or TDoA hyperbola. (Right) The amplitude of a Gaussian surface depends on its variance.

approach, the propagation speed is set as 242 m/s. The amplitude values in the plot are normalized by the mean of the surface and then converted to dB scale. The resultant localization estimate in this case is similar to that of the intersections method because the two follow the same underlying approach of using the MPD lines and TDoA hyperbolas. However, the ambiguity map output is beneficial in that it returns a comparative measure between selecting any position on the map as the event location in addition to a single position estimate.

6.6 Calibration Event Localization Results

6.6.1 Performance Analysis

Applying both the intersections and ambiguity map methods to the SIDEx calibration dataset, Fig. 6.15 shows the localization outcome for the events. The propagation speed assumed for each event is determined through iterative application of the intersections method as described in Section 6.5.3. As expected, the localization estimates from both methods are comparable. A notable feature on both plots is that the algorithm's performance is good within the perimeter of the cabled array. The majority

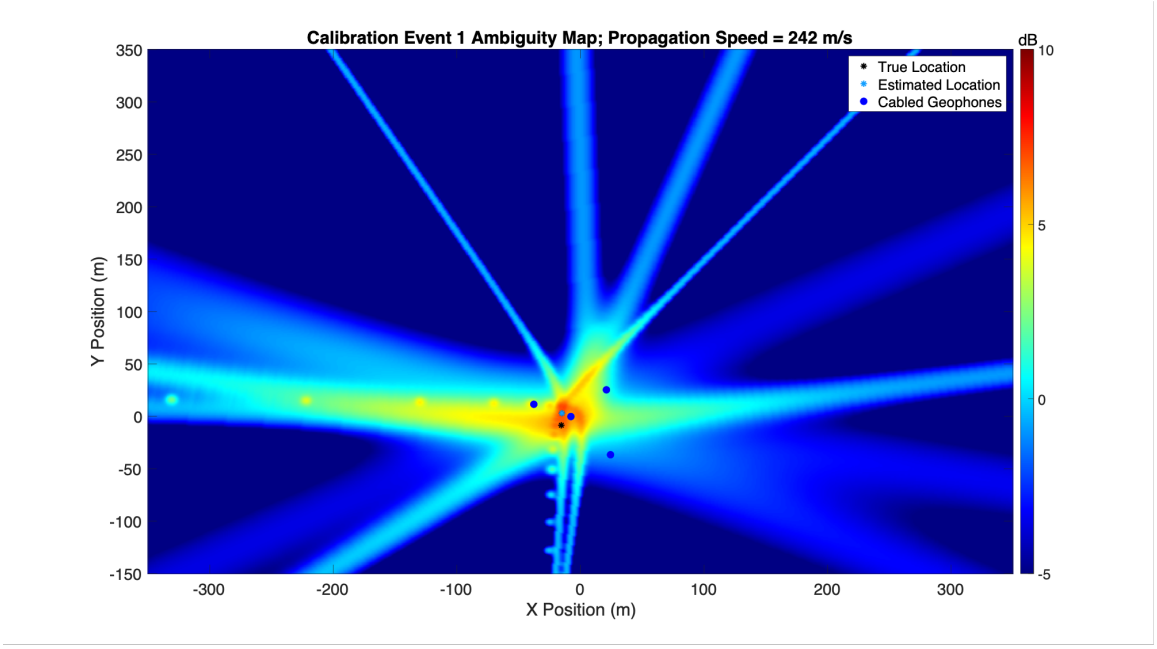
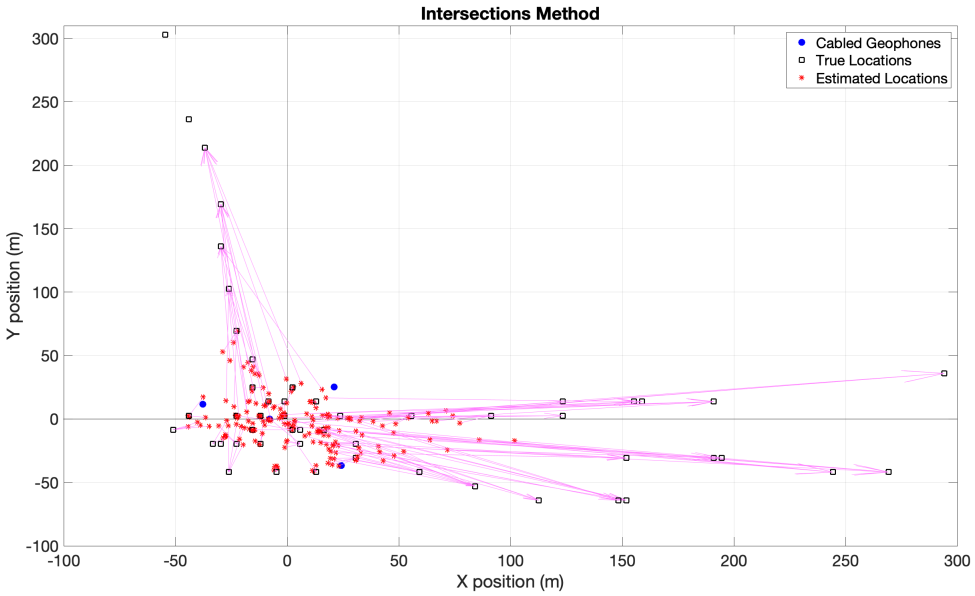


Figure 6.14: Ambiguity map and localization estimate for calibration event 1.

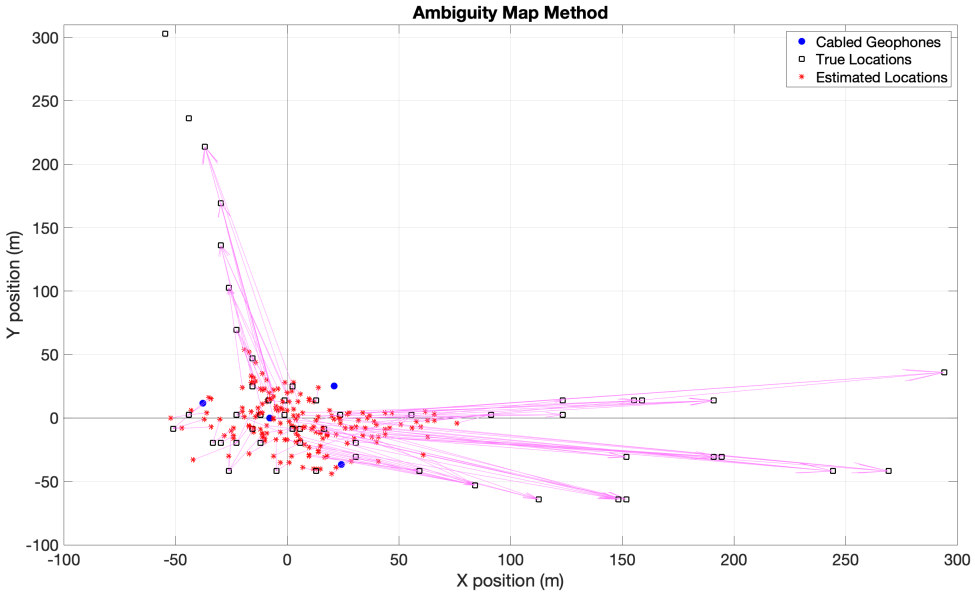
of estimates have less than 20 m error for events within 50 m from the array center (Fig. 6.17). However, localization accuracy degrades severely for events outside the cabled array perimeter. This result makes intuitive sense as the further away an event is from the array center, the more precise its bearing and TDoA estimate must be in order to have the MPD lines and TDoA hyperbolas intersect at the correct location. Thus, this result demonstrates that both algorithms are only effective within the perimeter of the deployed array and highlights the importance of array coverage area in localizing transient events.

6.6.2 Comparison with TDoA Matching Approach

As a comparison to our intersections and ambiguity map methods, we present a conventional, inverse approach to transient event localization that we call TDoA Matching. As the name suggests, this technique again utilizes TDoA for localization. However, it is an inverse approach because instead of using the measured TDoAs between all geophone pairs to directly infer event location, we generate a set of simulated TDoA vectors (each vector contains the TDoAs between all geophone pairs) for



(a)



(b)

Figure 6.15: Calibration event localization estimates by (a) the intersections method and (b) the ambiguity map method. The algorithms are most effective within the perimeter of the deployed array.

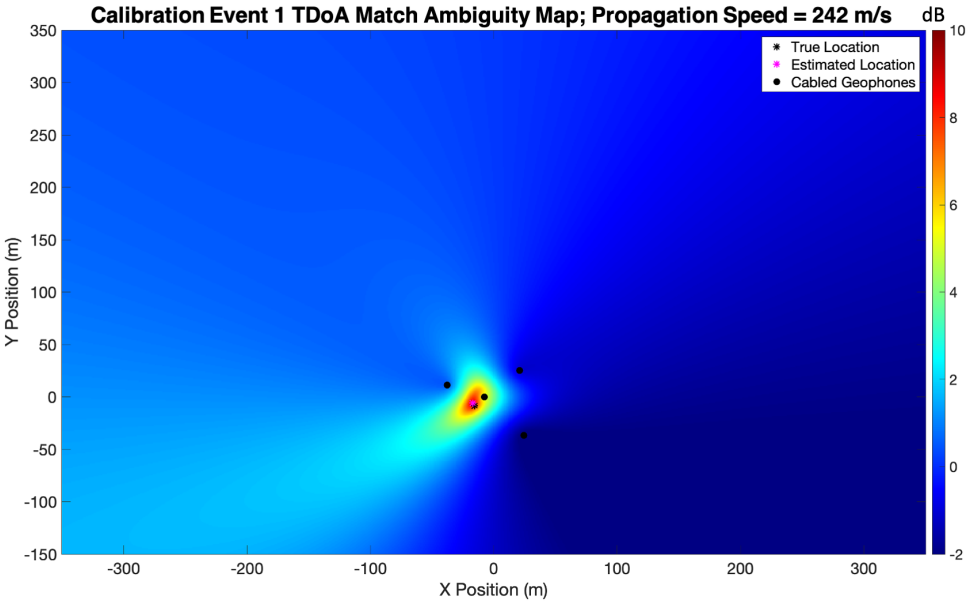
a grid of event locations and then *match* the simulated TDoA vectors to the measured values.

For the calibration events, we choose an area outlined by $x = [-350, 350]$ and $y = [-150, 350]$. These bounds are selected because all calibration events were generated within this region. We set the increment of our grid to 1×1 m; this also inherently defines the localization resolution of this approach. For each point, p, in this grid, we compute the expected TDoA between a pair of geophones, i and j, of an event originating from point p as

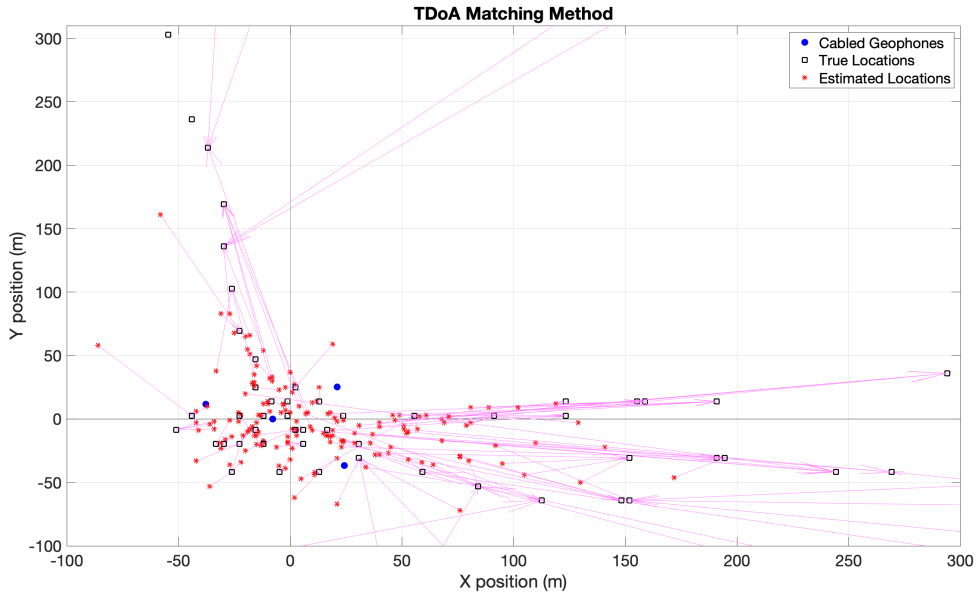
$$\tau_{ij} = \frac{|D_i - D_j|}{c_0}, \tag{6.17}$$

where D_i and D_j are the distances from point p to geophones i and j, respectively and c_0 is the assumed flexural wave propagation speed in the ice cover. Repeating this calculation for every unique geophone pair in the cabled array gives us a vector of simulated TDoAs for every point in the grid. For each calibration event, we assume the propagation speed as the refined estimate for that event using the intersections method in Section 6.5.3. We then calculate the mean squared error between the simulated TDoA vectors for each point and the measured TDoA vector computed from the recorded data. Finally, we select the point within the grid with the lowest error as the output localization estimate for the event.

The benefit of this method is that, similar to the previous ambiguity map approach, it also produces an ambiguity map. This result is shown for calibration event 1 in Fig. 6.16(a). The localization estimates for all calibration events are shown in Fig. 6.16(b). Here, as with before, localization accuracy is good within the perimeter of the array but degrades for further events. A comparison of localization error of the three methods is presented in Fig. 6.17. The performances are essentially identical with mean error values near 60 m for all calibration events. Thus, unfortunately, our forward approaches do not show an improvement over the inverse method in localizing the calibration events. While this result is disappointing, it does further affirm the importance of array coverage area to localization performance. For all three methods, the localization error is reasonable within the array perimeter. So, to apply the



(a)



(b)

Figure 6.16: (a) Ambiguity map generated for calibration event 1 using TDoA matching approach. (b) Localization estimates for all calibration events using TDoA matching method. Similar to the forward approach, the algorithm is most effective within the perimeter of the deployed array.

algorithms we have presented, the array deployed should ideally cover the entire area of interest to enable accurate event localization.

Given the result above, is there any advantage to using the forward approaches over the TDoA matching method? To this point, we do want to highlight that, in particular, the intersections method can have a much lower computational cost than the TDoA matching approach. Specifically, the slowest step in the intersections method is finding the intersections, which is essentially N 4×4 matrix divisions, where N is the number of points we use to segment the MPD lines and TDoA hyperbolas. We then repeat this calculation for all unique geophone pairs, which for m geophones, equal $(m-1)^2+1$. Thus, the total computation complexity for the intersections method is $\mathcal{O}((m-1)^2N)$. For the TDoA matching method, the slowest step is calculating the TDoA vector for each simulated event location. On its own, this step is only one division for every unique geophone pair for a total of $(m-1)^2+1$ divisions for all unique pairs. However, we must repeat this calculation for every point in our designated grid. Assuming a grid size of $L \times L$, the computation cost for this approach is $\mathcal{O}((m-1)^2L^2)$. In practice, N is typically much smaller than L^2. For example, in our analysis of the calibration dataset, N was set as 1000 while L^2 was equal to 351201. Thus, the computation cost of the inverse technique can be much higher, especially if we want finer resolution in the designated grid. In addition, while having lower computational cost is good in general, it is especially useful when the propagation speed in the ice cover is uncertain. As demonstrated in Section 6.5.3, with the intersections method, one can quickly try an interval of propagation speeds to see which gives the best localization result. Doing so with the TDoA matching approach would require much longer computation time.

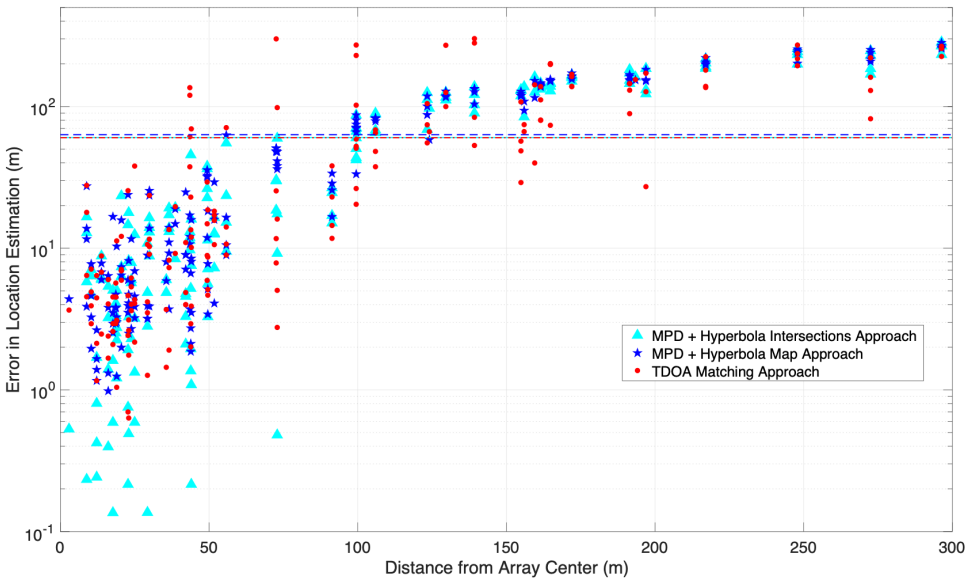


Figure 6.17: Localization error as a function of calibration event distance for the three methods presented. The dashed lines represent the mean error of all estimates for the three methods.

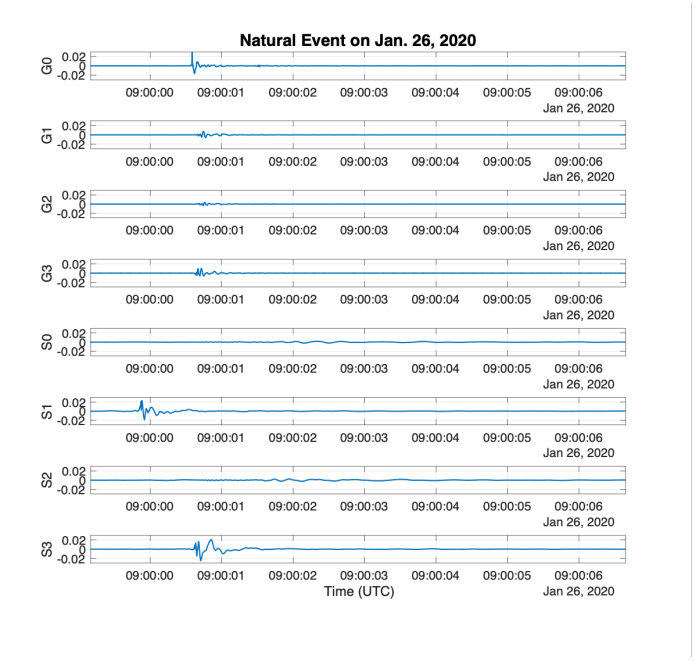
6.7 Example of a Natural Event on both the Cabled and Stand-alone Geophone Arrays

The main objective of deploying an array of remote geophone nodes in addition to the cabled geophones was to test whether the remote units can be utilized in the same manner as the cabled units to record events generated by the ice cover. As mentioned, this is the first instance that we are aware of that these remote geophones have been used for this purpose. Because the calibration events were not powerful enough to be observed on the further away remote nodes, we present an example of a stronger, natural event that was recorded by all cabled and remote geophones. We then demonstrate the capability of the stand-alone geophone array in localizing this event. Because we do not have ground truth information on the location of this event's origin, we instead show that the estimated location is at least consistent with modeling results.

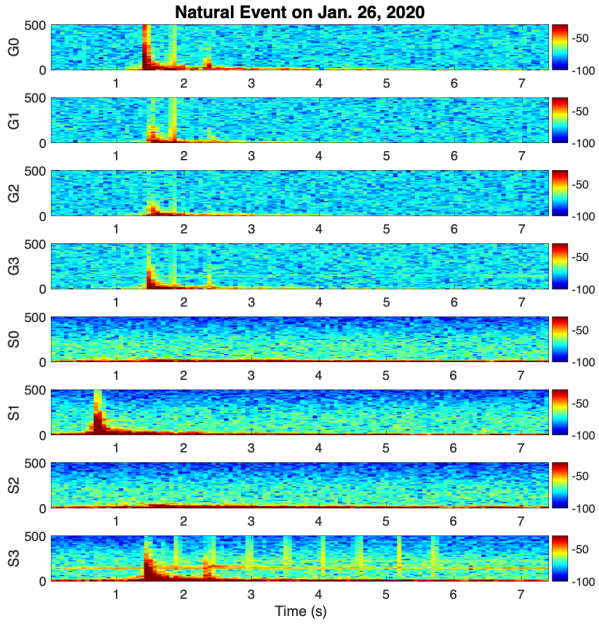
6.7.1 The January 26 Event

On January 26, 2020, a natural event occurred in the ice cover which was recorded by all cabled and remote geophones. The time series of this event is shown in Fig. 6.18(a). The first four panels in this plot show the recordings on the cabled geophones (G0-G3). The event arrives on each unit at around the same time, before 09:00:01. The bottom four panels show the recording on the stand-alone geophones (S0-S3). Since S3 is approximately co-located with G3, the event arrival on S3 aligns with the arrival time on G3. On S1, the event arrives much earlier, before 09:00:00. This indicates the event occurred at a location closest to S1. On the S0 and S2 time series, the event is barely observable. The only evidence of its existence is a faint, very low frequency waveform near 09:00:02 on both channels. We suspect the signal amplitude is much lower on these nodes because they were the furthest away from the event location. Consequently, the higher frequency portion of the event dissipated before reaching these geophones. Fig. 6.18(b) shows the spectrogram of the time series recorded on every geophone. On the top four panels, the arrivals on the cabled geophones show the expected flexural wave dispersion. We observe a similar pattern on the S1 and S3 spectrograms. For S0 and S2, since their time series consist of a very faint waveform, their spectrograms correspondingly lack a strong flexural wave feature. Nonetheless, the similar time and frequency domains features between the recorded data on G0-G3 and S1 & S3 demonstrate the remote geophones can indeed be utilized for seismo-acoustic monitoring of the ice cover, provided that they are deployed close enough to the locations of transient event sources.

Next, we demonstrate that, similar to the localization of the calibration events using the cabled array data, we can localize this natural event using the data collected by the remote geophones. We use the forward ambiguity map method here and assume a propagation speed of 336 m/s, which is the average of all calibration event speed estimates shown in Fig. 6.12. The resultant localization estimate is presented in Fig. 6.19. Consistent with our expectation from examining the time series, the estimated event location is closer to S1 and S3 and further from S0 and S2. Furthermore, this



(a)



(b)

Figure 6.18: Time series and spectrograms of the natural event on January 26. (a) Time series recorded on the cabled (G0-G3) and stand-alone (S0-S3) geophones vertical axes; the y-axes of the panels are unnormalized amplitude values. (b) Spectrograms of the event on every geophone unit (vertical axis); the y-axes of the panels are frequency in Hz.

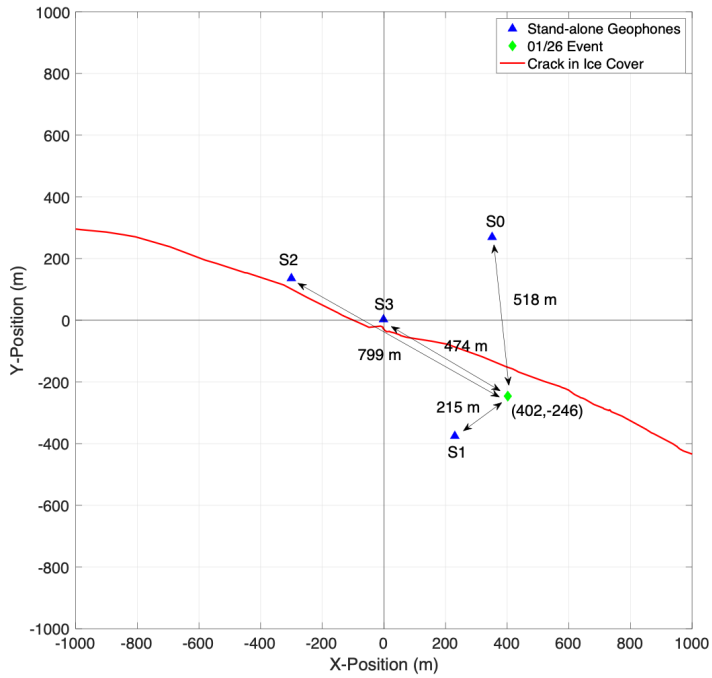


Figure 6.19: Localization estimate for the January 26 event using data from the remote geophones. Consistent with expectation, the event is closer to S1 and S3 and further from S0 and S2.

location is not far from the crack in the ice cover observed during system deployment, which may be the source of the event.

6.7.2 Comparison with Modeling

Although we do not have ground truth on where the January 26 event occurred, we can loosely verify the accuracy of its localization estimate by comparing the stand-alone geophone data with modeling results. To do so, we assume in our model that the source of the event is located at the estimated position. We then simulate the recorded time series on geophones at the four ranges indicated in Fig. 6.19. To model the SIDEx experiment site, we assumed an environment consisting of a 1 m thick ice layer, a 4 m deep water column, and a muddy bottom halfspace. The parameters used in these layers are presented in Table 6.3. For the event source, a vertically

Table 6.3: Parameters used to model the January 26 Event. C_p denotes compressional speed, C_s denotes shear speed, ρ denotes density, η_{rms} denotes root-mean-square roughness, η_{cl} denotes roughness correlation length.

Medium	C_p	C_s	ρ	η_{rms}	η_{cl}
Vacuum halfspace	-	-	-	-	-
Ice layer (top)	3000 m/s	1200 m/s	0.9 g/cm^3	-	-
Ice layer (bottom)	2400 m/s	600 m/s	0.9 g/cm^3	0.2 m	19.1 m
Water column	1435 m/s	-	1.0 g/cm^3	-	-
Bottom halfspace	1550 m/s	100 m/s	1.6 g/cm^3	-	-

forced impulse is used with a center frequency of 16 Hz[1] and placed 0.1 m within the ice cover. The simulation is carried out using OASES [35].

The modeled time series are shown in Fig. 6.20(a). Assuming there is minimal mismatch between the simulated and real environments and sources, if the modeled time series are similar to the measured time series in terms of their arrival times, then the localization estimate should be accurate because the simulated source was placed at the estimate location. While the no mismatch assumption is unlikely to hold given the complexity of ice cover properties and noise generation, comparison with the simulated time series is still helpful because it allows us to confirm whether our localization estimate is consistent with the physics of a simplified propagation environment. However, we do caution that a match between the simulated and measured time series is not conclusive verification of localization accuracy, it only suggests the estimate is consistent with our modeling of the event. Comparing Fig. 6.20(a) with Fig. 6.18(a), the arrival order at the four simulated geophones matches the measured data, which means the bearing of the localization estimate should be close to correct. The simulated time series appear to arrive slightly too early on S0 while the simulated S3 arrival is more attenuated at the higher frequencies (Fig. 6.20(b)). These differences suggest that, based on our modeling, the estimated source location should have been closer to S3 and further away from S0. Despite this, the simulated

[1]This frequency is chosen based on the calibration data PSD (Fig. 6.7(a)).

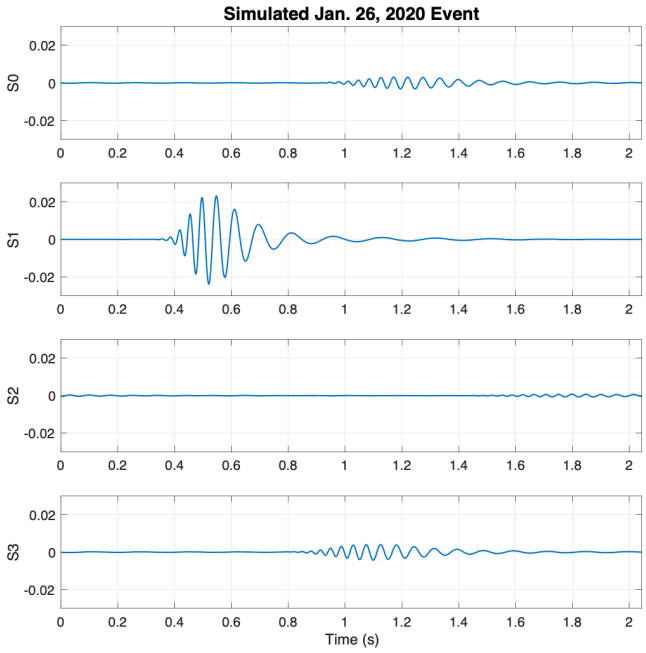
time series are, in general, consistent with the measured data.

As a final check that the application of our localization algorithm on simulated data produces a localization estimate that is consistent with expectation, we apply the forward ambiguity map method to the simulated time series. Again, assuming our model is an accurate enough representation of the real environment, the localization estimate made from simulated data should return a position close to where the simulated source was placed (*i.e.*, [402, -246]). Such a result would verify that our localization algorithm is at least accurate for events simulated in our modeled environment. To apply the algorithm, we again assume the flexural wave propagation speed in the ice cover as 336 m/s at 16 Hz. Fig. 6.21 shows the localization result for this case. With the simulated data, the position estimate is [383, -209], which is only 41.6 m from the simulated source placement. This relatively small difference suggests our localization algorithm can indeed return a close-to-correct estimate for events generated in our modeled environment.

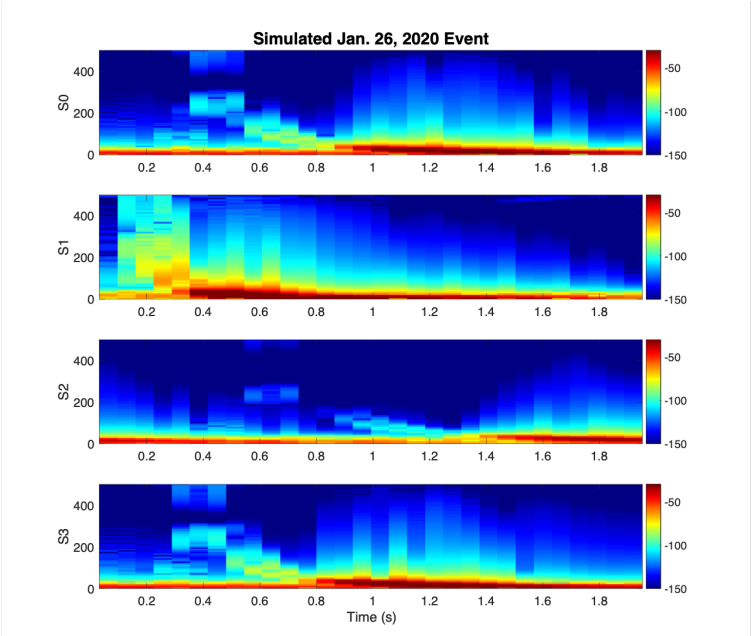
6.8 Summary & Lessons Learned

6.8.1 Main Results

In this chapter, we presented the deployment of a seismo-acoustic system of geophones and hydrophones to monitor ice cover activity and localize transient events. The system consists of two geophone arrays (one cabled, one stand-alone) and two hydrophone units. We primarily focused our analysis on the cabled array. As a demonstration of the array's capability to localize events, we applied three localization algorithms to a calibration dataset and compared their performances. The first two methods are forward algorithms that utilize MPD lines and TDoA hyperbolas to output a localization estimate or generate an ambiguity map for the event location. The third approach is an inverse method that predicts the event location by matching the measured TDoAs against a set of simulated templates. All three algorithms showed similar performance in localizing the calibration events. Localization error



(a)



(b)

Figure 6.20: Simulated time series and spectrograms of the January 26 Event. (a) Simulated time series on stand-alone (S0-S3) geophones; the y-axes are unnormalized amplitude values. (b) Simulated spectrograms on S0-S3; the y-axes are frequency in Hz.

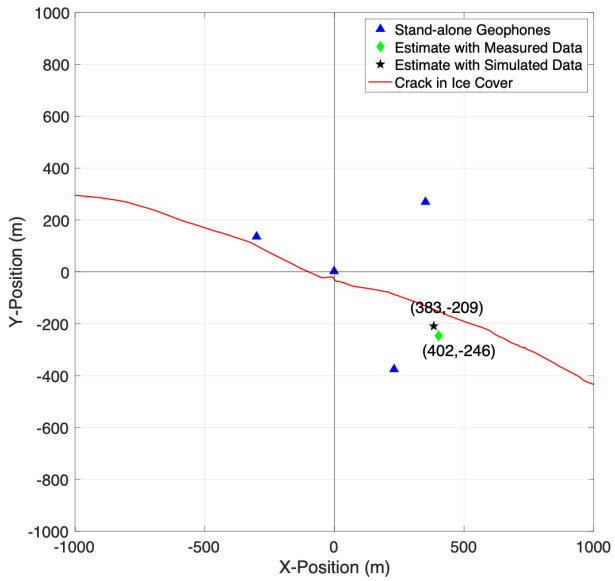


Figure 6.21: Localization estimate for the simulated January 26 event using modeled time series at the remote geophone locations. The location estimate produced with simulated data is relatively close to the estimate produced with measured data, which is where the simulated source is placed.

is typically less than 20 m within the perimeter of the deployed array but grows to much larger values outside of the array perimeter. This result demonstrates the importance of array coverage area to localization success as regardless of the method, the localization accuracy was much better within the array perimeter than outside of the perimeter.

We further demonstrated the utility of remote geophones for ice cover seismo-acoustic monitoring. Because of their cable-less nature, these units can be easily deployed to form an array that covers a much larger area than what is possible with cabled geophones. Using the four remote geophones deployed, we were able to provide a localization estimate for a natural event that occurred on January 26, 2020. While we do not have ground truth for this event, the modeled time series at the remote geophone locations assuming the event occurred at the estimated location are agreeable with the measured time series data. This result provides some support that the estimate for the natural event is fairly accurate.

6.8.2 Recommendations for Future Deployment

For future deployments of similar seismo-acoustics systems, we make the following recommendations based on lessons learned from our experiment. First, in order to have high accuracy in event localization, a large grid of geophones with less than 500 m spacing between each unit should be deployed to cover the entire area of interest. This spacing is selected so that an event occurring anywhere in the coverage area should be well recorded by at least 3 of the geophones in the array. To create such a large array, remote geophone nodes can be used as we have demonstrated their utility. In deploying the geophones, it is important to ensure that the orientation of the x-y-z axes are well documented if the MPD lines are to be used for localizing detected transient events. Second, if in-water hydrophones are also deployed as part of a system, they should be lowered to at least 10 m below the ice cover. Doing so would allow the ice cover flexural waves from transient events to dissipate so that waterborne pressure wave arrivals may be observed in the time series. The pressure wave arrivals can then be utilized in a combined processing of the hydrophone and geophone data, where they can provide a range estimate to further improve the localization of an event. Lastly, other approaches for event localization should be explored. Namely, machine learning may be a useful application here. For example, a calibration dataset can be used as a training set to condition the machine learning model. This approach could be beneficial in alleviating errors due to estimating TDoAs between geophone pairs and the flexural wave propagation speed in the ice cover. For higher quality dataset where the P and SH wave arrivals can be clearly observed in addition to the flexural wave arrival, both the arrival time differences between the different wave types and the time varying polarization property of the dispersive flexural wave can be leveraged to aid localization. For instance, if accurate modeling is accessible, both can be used in a "matched field" approach to invert for event locations.

Many of the above suggestions have been incorporated into the next SIDEx experiment, scheduled to take place in the Spring of 2021 in the Beaufort Sea. These include a larger remote geophone array and deeper hydrophone placements. Thus,

the lessons learned through our analysis in this chapter were valuable in driving the design of the next experiment.

Chapter 7

Conclusions and Future Work

The body of work presented in this thesis characterizes changes in the Beaufort Sea ambient soundscape and demonstrates how the changes are indicative of broader environmental shifts in the region. Along with data analysis, we have also developed approaches for noise range estimation, transient event detection & characterization, and seismo-acoustic event localization. In this chapter, we reiterate the main results of the thesis and outline some directions for future work.

7.1 Conclusions

7.1.1 Ambient Noise during the SIMI-94 Experiment

The environment during the SIMI-94 experiment is a representation of the historical Arctic environment. The underwater SSP monotonically increases with depth while the ice cover consists more of thicker, multi-year ice. The spectral shape of the ambient noise shows noise level generally decreases with increasing frequency with a peak near 20 Hz. The measured noise level is higher than previous Arctic recordings [57], although this discrepancy is likely caused by higher self-noise of the SIMI-94 array. The vertical directionality of the recorded noise shows a peak in noise level arriving from broadside. This result matches well with modeled noise directionality assuming

uniform noise generation by the ice cover, demonstrating that such a model is still a fitting description of ice cover noise generation during SIMI-94. A time domain, amplitude-based transient event detection algorithm is applied to study the temporal distribution of transient noise events during SIMI-94. This analysis demonstrates that transient occurrences may be described as a clustering process, where events occur in quick succession followed by a gap during which there are no transients. More events are detected at higher frequency octave bands, which indicates transient noise accounts for a larger portion of the total acoustic soundscape at higher frequencies.

7.1.2 Environmental Changes in the Beaufort Sea

More recently, the Arctic Ocean environment has undergone significant environmental changes, which are still continuing presently. These changes are observed in the Beaufort Sea in two forms. The first is a shift in the underwater SSP. An influx of warm Pacific water entering the region from the Bering Strait has increased water temperature between ∼40-80 m depth [69, 71], creating the "Beaufort Lens". Consequently, the traditional, monotonically increasing SSP is disrupted and a sound speed maximum now occurs at near 70 m depth. This change affects acoustic propagation in the region by creating a double duct environment - one at the surface as usual and another just below the sound speed local maximum. In the surface channel, acoustic waves encounter frequent interactions with the ice-water interface, resulting in severe attenuation. In contrast, the lower duct has been shown to promote long range propagation by effectively trapping sound above 300 Hz [13, 14, 75, 76, 77]. The form of the Beaufort Lens SSP is continuously evolving as new sound speed measurements from the region have demonstrated the possibility of two local maxima within the water column instead of just one.

The second component of environmental change is a shift in ice cover thickness and age. As a result of increasing average global temperature, the mean winter thickness of central Arctic ice cover has declined more than 10 cm/year between 1993-2007 [81]. Furthermore, the percentage of multi-year ice has dramatically reduced [45, 85], replaced by first-year ice that is more fragile and susceptible to ridging. Consequently,

ambient noise generation by the ice cover may no longer be adequately described by a uniform distribution of sources. Instead, we present the hypothesis that noise generation has become more spatially discrete and is better modeled with a single source positioned at a specific range.

7.1.3 Environment Induced Effects on Ambient Noise during the ICEX-16 Experiment

The environment during the ICEX-16 experiment exhibits the new Beaufort Sea environment with the double-ducted SSP and thinner, first-year ice cover. While the spectral shape of ambient noise during ICEX-16 remains similar to previous measurements - noise level decreases with increasing frequency, its spatial characteristics have changed. Specifically, the peak in noise vertical directionality occurs near $\pm$10-15 degrees with a noise notch at broadside. Comparisons with modeling results demonstrate the Beaufort Lens SSP is primarily responsible for the noise notch while discrete surface noise generation $\sim$30-50 km away from the experiment site explains the peak noise directionality angles. This noise generation range aligns with the location of a ridge formation in the ice cover, further supporting our hypothesis that noise during ICEX-16 was generated at discrete ranges.

The change in surface noise generation also affects the temporal characteristics of transient noise events during ICEX-16. Compared to event clusters during SIMI-94, the median time gaps between consecutive event clusters for the two datasets are similar; however, cluster lengths and the number of events in each cluster are greater during SIMI-94 than ICEX-16. These results demonstrate that when the ice cover was active, more transient events occurred during SIMI-94 than ICEX-16.

7.1.4 Model-based CNN Approach to Noise Range Estimation

As a more robust alternative for estimating the range of surface noise generation than conventional MFP, we present a model-based CNN approach. The advantage of

model-based machine learning instead of the usual data-driven framework is in the ease of generating a large quantity of training data, which may be difficult or costly to collect for underwater acoustics applications. We design two CNNs, one performs range estimation by classification (CNN-c) while the other by regression (CNN-r). These models are trained using simulated data and tested with both simulated and real data from experiments against mismatches in SSP and bottom depth. Compared with MFP, CNN-c shows performance improvements in both cases. On the other hand, CNN-r performs better than MFP against SSP mismatch but less so against bottom depth mismatch. The reason for CNN-r's inconsistent performance is likely due to its goal of lowering the overall MSE cost during training. This specification causes CNN-r to have less variability in its predictions, which lowers the overall error of the estimates but increases error on individual predictions compared to MFP.

We further explore how the model-based CNNs may be achieving their robustness by examining their intermediate outputs. We find that the CNN pre-prediction output vectors for neighboring source ranges are also near each other in Euclidean space. As a result, any slight change to the pre-prediction vectors due to environmental mismatch is less likely to cause the CNNs to output a prediction that is drastically different from the correct output. In contrast, MFP does not share this property.

7.1.5 Frequency Domain Transient Event Detection

While time domain, amplitude-based transient event detection algorithms are simple to implement and tune, they lack in providing a complete description of the detected events' spectral characteristics. As a result, time domain methods are less helpful in hypothesizing the mechanisms by which the detected transients may have been generated, as different ice cover mechanics can produce different spectral signatures [56, 58]. To address this issue, we present a frequency domain event detection algorithm which leverages image processing and h-clustering to identify transients in data spectrograms. Applying this method to noise data collected during ICEX-16, we are able to identify 1097 events and categorize them based on their bandwidth and duration. Of these, 773 are *stbb* transients, which are indicative of small, impulsive

ice fractures that may occur in the periphery of larger cracking events. 242 detections are *ltbb* events, which represent more prolonged periods of ice cracking. Lastly, 82 detections are *ltnb* events, which result from rubbing or shearing interactions between ice floes of different sizes. By examining the temporal distribution of the detected transients, we conclude that ice cracking during ICEX-16 likely happened on small temporal and spatial scales. Small fractures occurred regularly, whereas larger cracking events and ice floe interactions were more rare. When larger cracking events did occur, they tend to have happened in clusters, resulting in a sudden increase in transient noise generation.

7.1.6 Ice Cover Seismo-acoustic Monitoring during the SIDEx Experiment

Another approach to monitor ambient noise generated by ice cover activity is through a seismo-acoustic system consisting of both geophones and hydrophones. We describe such a system deployed during the SIDEx experiment in 2020. The two main components of the system are a cabled geophone array ($\sim$50 m aperture) and a larger ($\sim$1000 m aperture) remote geophone array. To localize a set of calibration events recorded on the cabled array, we present two localization algorithms. These methods utilize MPD lines and TDoA hyperbolas to output a localization estimate or generate an ambiguity map for the event location. They show similar performance in localizing the calibration events to an inverse TDoA matching technique. Localization error is typically less than 20 m within the perimeter of the deployed cabled array but grows to much larger values outside of the array perimeter. This result re-affirms the importance of array coverage area to localization success.

We also demonstrate the utility of remote geophones for ice cover seismo-acoustic monitoring. Because of their cable-less nature, these units can be easily deployed to form an array that covers a much larger area than what is feasible with cabled geophones. Using the remote geophone array, we provide a localization estimate for a natural event that occurred on January 26, 2020. Through comparison with

modeled time series at the remote geophone locations generated by assuming the event occurred at the estimated location, we find supporting evidence that the estimate for the natural event is fairly accurate.

7.2 Future Work

Based on the experiences and lessons learned from our work, we provide some intriguing avenues for future studies in this section.

7.2.1 Continued Monitoring of Beaufort Sea Environment

As noted, the Beaufort Sea environment, particularly its SSP, remains dynamic and thus invites continued observation. Newer measurements of the underwater sound speed may be obtained through the WHOI ITP program [73] to further document the rate and extent of its change. Noise modeling can also be used to investigate the effects of more recently observed shifts in the SSP, such as the emergence of two sound speed local maxima in the water column.

Another aspect of the Beaufort Sea SSP that should be examined is how slight variations to the profile would impact the modeling results presented in Chapter 3. When modeling the expected ambient noise vertical directionality to estimate the noise generation range, the SSP measured during the ICEX-16 experiment was used. However, while the timescale of the data recording period is relatively short, it is certainly possible that during that time, the actual SSP in the environment deviates from the measured profile, creating a mismatch when modeling. Thus, an important task for future work is to examine how slight variations to the measured ICEX-16 SSP would affect the modeling results. Doing so would establish an interval of acceptable SSP variability over which the conclusions for noise range generation from Chapter 3 are valid. In addition to temporal variation, the effect that the spatial variability of the SSP has on the ambient soundscape should also be studied. Noise modeling can be expanded from the range-independent assumption to a 3-D, range-dependent, environment. Using measured SSP data (from ITP, for example), an estimate of

the typical spatial variability of the Beaufort Lens can be made and included in modeling to investigate how its spatial variation affects the ambient noisefield. Then, comparison of the modeled results with measured noise data would provide insight into the amount of SSP variability present during the period of data recording.

More and longer ambient noise collections in the Beaufort Sea would also be instrumental in determining whether the results from the ICEX-16 data analysis are the norm or more specific to the experiment. Combining underwater ambient noise recording with ice cover dynamics monitoring using a seismo-acoustic system presented in Chapter 6 would be helpful in further verifying the connection between ice cover noise generation and observed ambient noise features.

Further information on ice cover properties such as thickness and roughness would further improve the modeling of the ice cover in simulation. In addition to physical properties, the inclusion of a slush layer between the ice and water column that is neither completely solid or liquid should be considered as such a layer can have an effect on acoustic propagation at higher frequencies.

7.2.2 Further Examination of CNN Approach

The CNN approach should be compared with more advanced MFP methods that extend beyond a simple matching between the observed data and the modeled replicas. Further insight into how the CNN approach achieves its performance can be gained by exploring whether there are underlying similarities between CNN and more advanced MFP approaches such as the multiple constraints method (MCM) [99], which also achieves more robustness to mismatch by broadening the mainlobe of the processor.

The performance of the CNN approach to noise range estimation may be improved by employing more sophisticated network architectures such as a deep residual network [195]. Such networks contain many more layers and parameters than the networks presented in this thesis; consequently, they are more computationally intensive to train and more care must be taken to prevent overfitting. One training method that may aid in the application of these networks is transfer learning [196], which partially pre-trains a network on a separate, widely available, general dataset to re-

fine some of the network's weights before transitioning the network to train on data relevant to the problem at hand. One can imagine a hybrid model where the network is initially trained using a large simulated dataset before being refined with a smaller real dataset if such data exist. The use of more sophisticated networks would also enable greater source and environmental estimate capabilities such as source depth, radiation pattern, or ocean bottom properties.

7.2.3 Less Empirical Parameters for Transient Categorization

Our frequency domain, transient event detection algorithm can be improved by designing a less empirical categorization metric for grouping features. For example, the metric can incorporate feature properties such as slope in addition to bandwidth and duration. The categorization metric can also be devised based on knowledge of the transients of interest's specific spectral signature, *e.g*, if it has the form of a chirp or follows certain characteristic specifications. Applying these improvements would allow for more specialized identification of events and further improvement of detection accuracy.

7.2.4 SIDEx Experiment in the Beaufort Sea

The next iteration of the SIDEx experiment has been planned for Spring of 2021 in the Beaufort Sea. Incorporating our recommendations for seismo-acoustic system deployment, this experiment will collect more ice motion data using a more comprehensive array of geophones and hydrophones. Thus, future work regarding this project includes continued data processing from the new experiment to localize transient ice cracking events. Furthermore, the application of machine learning approaches for event localization and ice characterization should be explored. For example, a model could be trained using collected calibration data to construct a mapping between arrivals on the deployed sensors to source location or other variables of interest. For datasets where the P and SH wave arrivals can be clearly observed in addition to

the flexural wave arrival, both the arrival time differences between the different wave types and the time varying polarization property of the dispersive flexural wave can be leveraged to aid localization. For instance, if accurate modeling is accessible, both can be used in a "matched field" approach to invert for event locations.

7.3 Concluding Remarks

Detailed characterization of ambient noise is an ever-growing need in the Arctic Ocean, particularly as the region grows even more in significance in the future. With the data analyses conducted in this thesis, we hope to have demonstrated the intrinsic link between its underwater ambient soundscape and the Arctic's rapidly changing environment. The region remains highly dynamic and thus invites continued observation and study. Through the noise characterization methods we proposed, we hope to have contributed to building a strong foundation for future work in this field.

Appendix A

Templates for Environmental Modeling using OASES

Here, we present a couple of input templates for environmental modeling with OASES, particularly with the OASN and OASP modules. These modules have a great deal of capability and their input files vary depending on the desired simulated environment and sources. Consequently, the reader should refer to [35] to have a full overview of the modules. Below, we present a template for generating the covariance matrix upon a line array using OASN and a template for generating the recorded time domain move-out on geophones using OASP. We also provide some MATLAB code for parsing in the modeled outputs from the two modules.

A.1 Template for Running OASN

The OASN module models the seismo-acoustic field on an arbitrary 3-dimensional array of hydrophones and geophones in the presence of surface noise and discrete sources in the water column. One option for the form of the resultant field output is the signal covariance matrix upon the recording array. To simulate surface noise, OASN applies the Kuperman-Ingenito model [61] presented in Chapter 2. Below is a template for the input to this module in the case of modeling the field recorded on a hydrophone line array due to surface noise and a single discrete source.

OASN Template

```
# Block I: Title - can be a descriptor to help identify what is modeled.
SIMI-94 Environment, Surface + Discrete Source, VLA

# Block II: Options
N J 0
#N - outputs the covariance matrix
#J - wavenumber integration contour offset
#0 - uncorrelated surface sources

# Block III: Frequency Sampling
80 100 3 0 #3 frequencies between 80-100 Hz, sampled equally.

# Block IV: Environment
13 #number of layers
0.0   0.0   0.0   0.0   0.0   0.0   0.0
0   3600 1800   0.216   0.648 0.9   0
2   3600   1800 0.216   0.648 0.9   -0.6  19.1  2.5
2   1431.5 0.0 0.0   0   1   0
100   1443  -999.999 0.0  0   1   0
200   1450  -999.999 0.0  0   1   0
300   1458  -999.999 0.0  0   1   0
450   1460  -999.999 0.0  0   1   0
600   1460.5 -999.999 0.0   0   1   0
700   1462  -999.999 0.0  0   1   0
750   1463  -999.999 0.0  0   1   0
3000  1500  -999.999 0.0  0   1   0
3000  2200  1500  0.5   0.5   2.9   0
#D; CC; CS; AC; AS; RO; RG; CL
#D - depth (m); CC - compressional speed (m/s); CS - shear speed (m/s); AC
```

```
    - compressional attenuation (dB/wavelength); AS - shear attenuation
    (dB/wavelength); RO - density (g/cm^3); RG - RMS interface roughness
    (m); CL - roughness correlation length (m)
#if a field is empty, it is assumed as 0.0
#setting CS = -999.999 is a flag for continuous SSP which sets the sound
    speed at the bottom of the layer equal to the speed specified for the
    top of the next layer below

# Block V: Receiver Array
32 #number of receivers
63 0 0 1 -1
70 0 0 1 -1
77 0 0 1 -1
84 0 0 1 -1
91 0 0 1 -1
98 0 0 1 -1
105 0 0 1 -1
112 0 0 1 -1
119 0 0 1 -1
126 0 0 1 -1
133 0 0 1 -1
140 0 0 1 -1
147 0 0 1 -1
154 0 0 1 -1
161 0 0 1 -1
168 0 0 1 -1
175 0 0 1 -1
182 0 0 1 -1
189 0 0 1 -1
196 0 0 1 -1
203 0 0 1 -1
210 0 0 1 -1
```

```
217 0 0 1 -1
224 0 0 1 -1
231 0 0 1 -1
238 0 0 1 -1
245 0 0 1 -1
252 0 0 1 -1
259 0 0 1 -1
266 0 0 1 -1
273 0 0 1 -1
280 0 0 1 -1
#Z; X; Y; T; G
#Z - depth (m); X - x-location (m); Y - y-location (m); T - receiver type
    (1 = hydrophone); G - receiver gain (dB)

# Block VI: Sources
60 0 0 1
#SL WL DL ND
#SL - surface source level (dB); WL - white noise level (dB); DL - deep
    noise level (dB); ND - number of discrete sources

# Surface Noise
1300 1E8 #phase velocity interval [cmin, cmax] (m/s)
-1 0 0 #wavenumber sampling in continuous, discrete, evanescent spectrum;
    -1 0 0 for automatic sampling.

# Discrete Sources
10 5 0 155 #source depth (m), x-range (km), y-range (km), strength (dB)

1300 1500 #phase velocity interval [cmin, cmax] (m/s)
-1 0 0 #wavenumber sampling in continuous, discrete, evanescent spectrum;
    -1 0 0 for automatic sampling.
```

A.2 MATLAB Code for Parsing .chk Files

The output covariance matrices from OASN are written in .chk files, one for each run of the module. The MATLAB code below can be used to parse in a directory of .chk files.

chkReader.m

```
array_size = 32; # number of receivers in array, change as needed
prefix = 'INSERT PATH TO .chk FILES HERE';
directory = dir([prefix '*.chk']); # directory of all .chk files

# initialize empty covariance matrix
oasn_cov = zeros(array_size, array_size, length(sorted_directory));

# read in each .chk file in directory
for k = 1:length(directory)
    filename = [prefix directory(k).name];
    fileID = fopen(filename);
    C =
        textscan(fileID,'\%*s\%s\%s\%s\%*s\%*s\%*s','HeaderLines',array_size+9);

    for i = 1:length(C{1,1})
        num(i) = str2double(C{1,1}(i));
        re(i) = str2double(C{1,2}(i));
        im(i) = str2double(C{1,3}(i));
    end

    # get rid of any potential NaNs
    keep = ~isnan(num);
    re = re(keep);
    im = im(keep);
    data = complex(re,im);
```

```
    # build covariance matrix
    counter = 1;
    for i = 1:array_size
        for j = i:array_size
            oasn_cov(i,j,k) = data(counter);
            if i ~= j
                oasn_cov(j,i,k) = conj(data(counter));
            end
            counter = counter + 1;
        end
    end

    fclose(fileID);
end
```

A.3 Template for Running OASP

The OASP module calculates the depth-dependent Green's function and determines the transfer function at any receiver position by evaluating the wavenumber integral. Once the transfer function is determined, the stress or particle velocity at any receiver due to sources in the modeled media can be calculated.

OASP Template

```
# Block I: Title - can be a descriptor to help identify what is modeled.
SIDEx Environment

# Block II: Options
J f V H 2
#J - wavenumber integration contour offset
```

```
#f - full Hankel function integration scheme, recommended for cases where
    nearfield is needed
#V - calculates vertical particle velocity
#H - calculates horizontal (radial) particle velocity
#2 - vertical point source with amplitude of 1 Newton

# Block III: Source Frequency
16 0 # source spectrum centered at 16 Hz

# Block IV: Environment
7 #number of layers
0.0 0.0 0.0 0.0 0.0 0.0 0.0
0.0 3000.0 1200.0 0.216 0.648 0.9 0.0
0.5 2700.0 900.0 0.216 0.648 0.9 0.0
1.0 2400.0 600.0 0.216 0.648 0.9 -0.2 19.1 2.5
1.0 1435.0 0 0.0 0.0 1.0 0.0
5.0 1435.0 0 0.0 0.0 1.0 0.0
5.0 1550.0 100.0 0.75 1.25 1.6 0.0
#D; CC; CS; AC; AS; RO; RG; CL
#D - depth (m); CC - compressional speed (m/s); CS - shear speed (m/s); AC
    - compressional attenuation (dB/wavelength); AS - shear attenuation
    (dB/wavelength); RO - density (g/cm^3); RG - RMS interface roughness
    (m); CL - roughness correlation length (m)
#if a field is empty, it is assumed as 0.0

# Block V: Source
0.1 #source depth (m)

# Block VI: Receiver Depths
0 0.5 3
#depth of first receiver (m); depth of last receiver (m); number of
    receivers
```

```
# Block VII: Wavenumber Sampling
100 1e8 #phase velocity interval [cmin, cmax] (m/s)
-1 1 1 -1 #number of wavenumber samples; first sampling point; last
    sampling point; frequency sample increment; (-1 1 1 -1) for auto

# Block VIII: Frequency and Range Samping
2048 5 500 0.001 0.2 0.01 81
#NT; FR1; FR2; DT; R1; DR; NR
#NT - number of time samples; FR1 - lower limit of frequency band; FR2 -
    upper limit of frequency band; DT - time sampling increment; R1 -
    first range; DR - range increment; NR - number of ranges
```

A.4 MATLAB Code for Parsing .asc Files

The output particle velocities from OASP are written in a single .asc file. The MATLAB code below can be used to parse in this file. Note, prior to running the code, it is necessary to rename the .asc as a .txt file. The result of running the code is a time table of particle velocities at each receiver location.

ascReader.m

```
filename = 'INSERT NAME OF FILE HERE.txt';
FS = 1024; # sampling frequency used for OASP, change as needed
ts_len = 2048; # length of time series simulated using OASP, change as
    needed
num_rec = 81; # number of receivers simulated using OASP, change as needed

# read in .asc file
fout = readASC(filename);
TT = array2timetable(fout,'SampleRate',FS);
```

```
for i = 1:length(TT.Properties.VariableNames)
    TT.Properties.VariableNames(i) = {[num2str((i-1)*10) 'm']};
end

# function for reading in .asc file
function [fout] = readASC(filename, ts_len, num_rec)
    f = readmatrix(filename,'NumHeaderLines',6,'CommentStyle',{'/#'});
    f(:,6:end) = [];
    f(any(isnan(f),2),:) = [];
    fout = reshape(transpose(f),ts_len, num_rec);
end
```

Appendix B

Code for Parsing ICEX-16 Data

Among the data analyzed in this thesis, all are stored in a file formate that can be easily parsed into MATLAB. The exception is data from ICEX-16, which are recorded into binary .DAT files. The code for reading in a directory of these files is presented below.

ICEXdataParse.m

```
# Set path to data
prefix = 'INSERT PATH TO DATA FOLDER HERE';
directory = dir([prefix 'ACO*.DAT']);

FS = 12000; # sampling frequency
NUM_SAMPLES = FS*2; # number of samples per .DAT file
NUM_CHANNELS = 32; # number of channels in ICEX-16 array

# first and last files to read in (change as needed)
first_file = 1;
last_file = 100;

# looping over ACO*.DAT files
counter=0;
```

```
for i = first_file:last_file
    counter=counter+1;
    filename = [prefix directory(i).name];
    fid = fopen (filename, 'r', 'ieee-le');
    if (fid <= 0)
        continue;
    end

    # read in single precision float acoustic data [samples X chn]
    for j = 1:NUM_CHANNELS
        data(((counter-1)*NUM_SAMPLES+1):(counter*NUM_SAMPLES),j) =
            fread(fid, NUM_SAMPLES, 'float32');
    end

    fclose(fid);
end
```

Bibliography

[1] IPCC. 2001: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change [Houghton, J.T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

[2] IPCC. 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M.Tignor and H.L. Miller (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

[3] F. Jensen, W. A. Kuperman, M. Porter, and H. Schmidt. *Computational ocean acoustics*. Springer, 2011.

[4] Bazile Kinda. *Acoustic remote sensing of Arctic sea Ice from long term soundscape measurements*. Doctoral thesis, Université de Grenoble, Grenoble, France, 2013.

[5] A. R. Milne and J. H. Ganton. Ambient noise under Arctic sea ice. *J. Acoust. Soc. Am.*, 36:855–863, 1964.

[6] James K. Lewis and Warren W. Denner. Arctic ambient noise in the Beaufort Sea: Seasonal space and time scales. *J. Acoust. Soc. Am.*, 82(3):988–997, 1987.

[7] James K. Lewis and Warren W. Denner. Higher frequency ambient noise in the Arctic Ocean. *J. Acoust. Soc. Am.*, 84(4):1444–1455, 1988.

[8] Robert S. Pritchard. Arctic Ocean background noise caused by ridging of sea ice. *J. Acoust. Soc. Am.*, 75(2):419–427, 1984.

[9] Mary Townsend-Manning. *Analysis of central Arctic noise events*. Master's thesis, Mass. Inst. of Tech., Cambridge, MA, 1987.

[10] Nicholas C. Makris and Ira Dyer. Environmental correlates of pack ice noise. *J. Acoust. Soc. Am.*, 79(5):1434–1440, 1986.

[11] Yuriy V. Dudko. *Analysis of seismo-acoustic emission from ice fracturing events during SIMI'94*. PhD thesis, Mass. Inst. of Tech., Cambridge, MA, 1999.

[12] Catherine Stamoulis. *Analysis of ice-induced acoustic events in the central Arctic*. Ph.D. dissertation, Mass. Inst. of Tech., Cambridge, MA, 1997.

[13] Andrew J. Poulsen and Henrik Schmidt. Acoustic noise properties in the rapidly changing Arctic Ocean. *Proceedings of Meetings on Acoustics*, 28(1):070005, 2016.

[14] H. Schmidt and T. Schneider. Acoustic communication and navigation in the new Arctic: A model case for environmental adaptation. In *2016 IEEE Third Underwater Communications and Networking Conference (UComms)*, pages 1–4, Aug 2016.

[15] Martin Jakobsson, Larry Mayer, Bernard Coakley, Julian A. Dowdeswell, Steve Forbes, Boris Fridman, Hanne Hodnesdal, Riko Noormets, Richard Pedersen, Michele Rebesco, Hans Werner Schenke, Yulia Zarayskaya, Daniela Accettella, Andrew Armstrong, Robert M. Anderson, Paul Bienhoff, Angelo Camerlenghi, Ian Church, Margo Edwards, James V. Gardner, John K. Hall, Benjamin Hell, Ole Hestvik, Yngve Kristoffersen, Christian Marcussen, Rezwan Mohammad, David Mosher, Son V. Nghiem, Maria Teresa Pedrosa, Paola G. Travaglini, and Pauline Weatherall. The international bathymetric chart of the Arctic Ocean (ibcao) version 3.0. *Geophysical Research Letters*, 39(12), 2012.

[16] T. Vihma. Effects of Arctic sea ice decline on weather and climate: A review. *Surv. Geophys.*, 25:1175–1214, 2014.

[17] J.A. Screen and I. Simmonds. The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature*, 464:1334–1337, 2010.

[18] Igor V. Polyakov, John E. Walsh, and Ronald Kwok. Recent changes of Arctic multiyear sea ice coverage and the likely causes. *Bulletin of the American Meteorological Society*, 93(2):145–151, 2012.

[19] J.A. Screen. Influence of Arctic sea ice on European summer precipitation. *Environ. Res. Lett.*, 8:044015, 2013.

[20] Miles G. McPhee, Timothy P. Stanton, James H. Morison, and Douglas G. Martinson. Freshening of the upper ocean in the Arctic: Is perennial sea ice disappearing? *Geophysical Research Letters*, 25(10):1729–1732, 1998.

[21] R. W. Macdonald, E. C. Carmack, F. A. McLaughlin, K. K. Falkner, and J. H. Swift. Connections among ice, runoff and atmospheric forcing in the Beaufort gyre. *Geophysical Research Letters*, 26(15):2223–2226, 1999.

[22] James Morison, Ron Kwok, Cecilia Peralta-Ferriz, Matt Alkire, Ignatius Rigor, Roger Andersen, and Mike Steele. Changing Arctic Ocean freshwater pathways. *Nature*, 481:66–70, 2012.

[23] Bruce J. Peterson, Robert M. Holmes, James W. McClelland, Charles J. Vörösmarty, Richard B. Lammers, Alexander I. Shiklomanov, Igor A. Shiklomanov, and Stefan Rahmstorf. Increasing river discharge to the Arctic Ocean. *Science*, 298(5601):2171–2173, 2002.

[24] Charles H. Greene, Andrew J. Pershing, Thomas M. Cronin, and Nicole Ceci. Arctic climate change and its impacts on the ecology of the north Atlantic. *Ecology*, 89(sp11):S24–S38, 2008.

[25] Martin C. Hänsel, Jörn O. Schmidt, Martina H. Stiasny, Max T. Stöven, Rudi Voss, and Martin F. Quaas. Ocean warming and acidification may drag down the commercial Arctic cod fishery by 2100. *PLOS ONE*, 15(4):1–14, 04 2020.

[26] Olav Sigurd Kjesbu, Bjarte Bogstad, Jennifer A. Devine, Harald Gjøsæter, Daniel Howell, Randi B. Ingvaldsen, Richard D. M. Nash, and Jon Egil Skjæraasen. Synergies between climate and management for Atlantic cod fisheries at high latitudes. *Proceedings of the National Academy of Sciences*, 111(9):3478–3483, 2014.

[27] Kenneth F. Drinkwater. The response of Atlantic cod (Gadus morhua) to future climate change. *ICES Journal of Marine Science*, 62(7):1327–1337, 01 2005.

[28] Grete K. Hovelsrud, Birger Poppel, Bob van Oort, and James D. Reist. Arctic societies, cultures, and peoples in a changing cryosphere. *AMBIO*, 40:100–110, 2011.

[29] Donald L. Gautier, Kenneth J. Bird, Ronald R. Charpentier, Arthur Grantz, David W. Houseknecht, Timothy R. Klett, Thomas E. Moore, Janet K. Pitman, Christopher J. Schenk, John H. Schuenemeyer, Kai Sørensen, Marilyn E. Tennyson, Zenon C. Valin, and Craig J. Wandrey. Assessment of undiscovered oil and gas in the Arctic. *Science*, 324(5931):1175–1179, 2009.

[30] Joshua Ho. The implications of Arctic sea ice decline on shipping. *Marine Policy*, 34(3):713–715, 2010.

[31] Eddy Bekkers, Joseph F. Francois, and Hugo Rojas-Romagosa. Melting ice caps and the economic impact of opening the northern sea route. *The Economic Journal*, 128(610):1095–1127, 2018.

[32] Lassi Heininen. Arctic geopolitics from classical to critical approach – importance of immaterial factors. *Geography, Environment, Sustainability*, 11:171–186, 03 2018.

[33] Andreas Østhagen. Geopolitics and security in the Arctic: What role for the EU? *European View*, 16(2):239–249, 2017.

[34] Elana Wilson Rowe. Analyzing frenemies: An Arctic repertoire of cooperation and rivalry. *Political Geography*, 76:102072, 2020.

[35] H. Schmidt. OASES: Ocean acoustic and seismic exploration synthesis. (Last viewed July 24, 2020). http://lamss.mit.edu/lamss/pmwiki/pmwiki.php?n=Site.Oases.

[36] Harry L. Van Trees. *Arrays and Spatial Filters*, chapter 2, pages 17–89. John Wiley & Sons, Ltd, 2002.

[37] Nicholas Rypkema. *Underwater & Out of Sight: Towards Ubiquity in Underwater Robotics*. Ph.D. dissertation, Mass. Inst. of Tech., Cambridge, MA, 2019.

[38] Michael B. Porter and Yong-Chun Liu. *Finite-Element Ray Tracing*, pages 947–956. world Scientific Publishing Co., 1994.

[39] V. Červený. *Ray tracing algorithms in three-dimensional laterally varying layered structures*. Reidel, Boston, MA, 1987.

[40] Michael B. Porter. The bellhop manual and user's guide, 2011.

[41] Henrik Schmidt and Finn B. Jensen. A full wave solution for propagation in multilayered viscoelastic media with application to gaussian beam reflection at fluid–solid interfaces. *J. Acoust. Soc. Am*, 77(3):813–825, 1985.

[42] Henrik Schmidt and Gerard Tango. Efficient global matrix approach to the computation of synthetic seismograms. *Geophysical Journal International*, 84(2):331–359, 1986.

[43] Gordon M. Wenz. Acoustic ambient noise in the ocean: Spectra and sources. *J. Acoust. Soc. Am.*, 34(12):1936–1956, 1962.

[44] David Marsan, Harry Stern, Ron Lindsay, and Jérôme Weiss. Scale dependence and localization of the deformation of Arctic sea ice. *Phys. Rev. Lett.*, 93:178501, 2004.

[45] James Maslanik, Julienne Stroeve, Charles Fowler, and William Emery. Distribution and trends in Arctic sea ice age through spring 2011. *Geophys. Res. Lett.*, 38(LI3502):1434–1440, 2011. doi:10.1029/2011GL047735.

[46] David Marsan, Jérôme Weiss, Eric Larose, and Jean-Philippe Métaxian. Sea-ice thickness measurement based on the dispersion of ice swell. *J. Acoust. Soc. Am*, 131(1):80–91, 2012.

[47] Julienne C. Stroeve, Mark C. Serreze, Marika M. Holland, Jennifer E. Kay, James Malanik, and Andrew P. Barrett. The Arctic's rapidly shrinking sea ice cover: a research synthesis. *Climatic Change*, 110:1005–1027, 2012.

[48] Robert J. Urick. *Ambient noise in the sea*. Naval Sea System Command, Department of the Navy, 1984.

[49] J. H. Ganton and A. R. Milne. Temperature- and wind-dependent ambient noise under midwinter pack ice. *J. Acoust. Soc. Am.*, 38(3):406–411, 1965.

[50] A. R. Milne, J. H. Ganton, and D. J. McMillin. Ambient noise under sea ice and further measurements of wind and temperature dependence. *J. Acoust. Soc. Am.*, 41(2):525–528, 1967.

[51] Michael V. Greening and Pierre Zakarauskas. Spatial and source level distributions of ice cracking in the Arctic Ocean. *J. Acoust. Soc. Am.*, 95(2):783–790, 1994.

[52] Pierre Zakarauskas and Jon M. Thorleifson. Directionality of ice cracking events. *J. Acoust. Soc. Am.*, 89(2):722–734, 1991.

[53] Ethan H. Roth, John A. Hildebrand, and Sean M. Wiggins. Underwater ambient noise on the chukchi sea continental slope from 2006-2009. *J. Acoust. Soc. Am.*, 131(1):104–110, 2012.

[54] Nicholas C. Makris and Ira Dyer. Environmental correlates of Arctic ice edge noise. *J. Acoust. Soc. Am.*, 90:3288–3298, 1991.

[55] G. Bazile Kinda, Yvan Simard, Cédric Gervaise, Jérôme I. Mars, and Louis Fortier. Under-ice ambient noise in eastern Beaufort Sea, Canadian Arctic, and its relation to environmental forcing. *J. Acoust. Soc. Am.*, 134(1):77–87, 2013.

[56] G. Bazile Kinda, Yvan Simard, Cédric Gervaise, Jérôme I. Mars, and Louis Fortier. Arctic underwater noise transients from sea ice deformation: Characteristics, annual time series, and forcing in Beaufort Sea. *J. Acoust. Soc. Am.*, 138(4):2034–2045, 2015.

[57] E. Ozanich, Peter Gerstoft, Peter F. Worcester, Matthew A. Dzieciuch, and Aaron Thode. Eastern Arctic ambient noise on a drifting vertical array. *J. Acoust. Soc. Am.*, 142(4):1997–2006, 2017.

[58] Yunbo Xie and David M. Farmer. The sound of ice break-up and floe interaction. *J. Acoust. Soc. Am.*, 91(3):1423–1428, 1992.

[59] Zhen Ye. Sound generation by ice floe rubbing. *J. Acoust. Soc. Am.*, 97(4):2191–2198, 1995.

[60] J.G. Veitch and A.R. Wilks. A characterization of Arctic undersea noise. Technical Report 12, Office of Naval Research, 1983.

[61] W. A. Kuperman and F. Ingenito. Spatial correlation of surface generated noise in a stratified ocean. *J. Acoust. Soc. Am.*, 67(6):1988–1996, 1980.

[62] Henrik Schmidt and W. A. Kuperman. Estimation of surface noise source level from low-frequency seismoacoustic ambient noise measurements. *J. Acoust. Soc. Am.*, 84(6):2153–2162, 1988.

[63] W. S. Liggett and M. J. Jacobson. Covariance of surface-generated noise in a deep ocean. *J. Acoust. Soc. Am.*, 38(2):303–312, 1965.

[64] J. Capon. High-resolution frequency-wavenumber spectrum analysis. *Proceedings of the IEEE*, 57(8):1408–1418, 1969.

[65] Gaute Hope, Hanne Sagen, Espen Storheim, Halvor Hobæk, and Lee Freitag. Measured and modeled acoustic propagation underneath the rough Arctic sea-ice. *J. Acoust. Soc. Am.*, 142(3):1619–1633, 2017.

[66] Pierre Zakarauskas, Carlton J. Parfitt, and Jon M. Thorleifson. Automatic extraction of spring-time Arctic ambient noise transients. *J. Acoust. Soc. Am.*, 90(1):470–474, 1991.

[67] Pierre Zakarauskas. Detection and localization of nondeterministic transients in time series and application to ice-cracking sound. *Digital Signal Processing*, 3:36–45, 1993.

[68] C. F. Chen. *Analysis of marginal ice zone noise events.* Ph.D. dissertation, Mass. Inst. of Tech., Cambridge, MA, 1990.

[69] J. M. Toole, M. L. Timmermans, D. K. Perovich, R. A. Krishfield, A. Proshutinsky, and J. A. Richter-Menge. Influences of the ocean surface mixed layer and thermohaline stratification on Arctic sea ice in the central Canada basin. *J. Geophys. Res.*, 115(C10018), 2010.

[70] M. L. Timmermans, J. M. Toole, and R. A. Krishfield. Warming of the interior Arctic Ocean linked to sea ice losses at the basin margins. *Science Advances*, 4(8), 2018.

[71] Jason D. Sagers, Megan S. Ballard, and Mohsen Badiey. Investigating the effects of ocean layering and sea ice cover on acoustic propagation in the Beaufort Sea. *Proceedings of Meetings on Acoustics*, 25(1):005003, 2015.

[72] G. A. Maykut and Miles G. McPhee. Solar heating of the Arctic mixed layer. *J. Geophy. Res.: Oceans*, 100(C12):24691–24703, 1995.

[73] Ice Tethered Profiler Program, 2001. Woods Hole Oceanographic Institution, http://www.whoi.edu/itp.

[74] R. Krishfield, J. Toole, A. Proshutinsky, and M-L. Timmermans. Automated ice-tethered profilers for seawater observations under pack ice in all seasons. *Journal of Atmospheric and Oceanic Technology*, 25(11):2091–2105, 2008.

[75] S. Carper. *Low frequency active sonar performance in the Arctic Beaufort Lens.* Master's thesis, Mass. Inst. of Tech., Cambridge, MA, 2017.

[76] L. Freitag, K. Ball, J. Partan, P. Koski, and S. Singh. Long range acoustic communications and navigation in the Arctic. In *OCEANS 2015 - MTS/IEEE Washington*, pages 1–5, 2015.

[77] T. Howe. *Modal analysis of acoustic propagation in the changing Arctic environment*. Master's thesis, Mass. Inst. of Tech., Cambridge, MA, 2015.

[78] Timothy F. Duda. Acoustic signal and noise changes in the Beaufort Sea pacific water duct under anticipated future acidification of Arctic Ocean waters. *J. Acoust. Soc. Am*, 142(4):1926–1933, 2017.

[79] H. R. Langehaug, F. Geyer, L.H. Smedsrud, and Y. Gao. Arctic sea ice decline and ice export in the cmip5 historical simulations. *Ocean Modelling*, 71:114 – 126, 2013. Arctic Ocean.

[80] Wadhams P. Arctic ice cover, ice thickness and tipping points. *Ambio*, 41(1):23–33, 2012.

[81] R. Kwok and D. A. Rothrock. Decline in Arctic sea ice thickness from submarine and icesat records: 1958-2008. *Geophys. Res. Lett.*, 36(L15501), 2009.

[82] Vladimir Kattsov, Vladimir Ryabinin, James Overland, Mark Serreze, Martin Visbeck, John Walsh, W. Meier, and Xiangdong Zhang. Arctic sea-ice change: A grand challenge of climate science. *Journal of Glaciology*, 56:1115–1121, 12 2010.

[83] Martin O. Jeffries, James E. Overland, and Donald K. Perovich. The Arctic shifts to a new normal. *Physics Today*, 66(10):35–40, 2013.

[84] Ingrid H. Onarheim, Tor Eldevik, Lars H. Smedsrud, and Julienne C. Stroeve. Seasonal and regional manifestation of Arctic sea ice loss. *Journal of Climate*, 31(12):4917 – 4932, 2018.

[85] J. A. Maslanik, C. Fowler, J. Stroeve, S. Drobot, J. Zwally, D. Yi, and W. Emery. A younger, thinner Arctic ice cover: Increased potential for rapid, extensive sea-ice loss. *Geophysical Research Letters*, 34(24), 2007.

[86] Maria Alejandra Parra-Orlandoni. *Target Tracking Onboard an Autonomous Underwater Vehicle: Determining Optimal Towed Array Heading in an Anisotropic Noise Field*. Master's thesis, Mass. Inst. of Tech., Cambridge, MA, 2005.

[87] M. Tschudi, G. Riggs, D. K. Hall, , and M. O. Román. VIIRS/NPP ice surface temperature 6-min l2 swath 750m, version 1. *Boulder, Colorado USA, NASA National Snow and Ice Data Center Distributed Active Archive Center*, 2017. (Last viewed July 24, 2020).

[88] H. B. Riley and J. A. Tague. Matched field processing: An overview. In *The 24th Southeastern Symposium on System Theory and The 3rd Annual Symposium on Communications, Signal Processing Expert Systems, and ASIC VLSI Design*, pages 180–184, 1992.

[89] A. B. Baggeroer, W. A. Kuperman, and P. N. Mikhalevsky. An overview of matched field methods in ocean acoustics. *IEEE Journal of Oceanic Engineering*, 18(4):401–424, 1993.

[90] Arthur B. Baggeroer and William A. Kuperman. *Matched Field Processing in Ocean Acoustics*, pages 79–114. Springer Netherlands, Dordrecht, 1993.

[91] A.Tolstoy. Applications of matched-field processing to inverse problems in underwater acoustics. *Inverse Problems*, 16:1655–1666, 2000.

[92] Henrik Schmidt and Arthur B. Baggeroer. *Physics-Imposed Resolution and Robustness Issues in Seismo-Acoustic Parameter Inversion*, pages 85–90. Springer Netherlands, Dordrecht, 1995.

[93] Michael I. Taroudakis and Maria G. Markaki. On the use of matched-field processing and hybrid algorithms for vertical slice tomography. *J. Acoust. Soc. Am*, 102(2):885–895, 1997.

[94] H. P. Bucker. Use of calculated sound fields and matched field detection to locate sound sources in shallow water. *J. Acoust. Soc. Am.*, 59:368–373, 1976.

[95] Claire Debever and W. A. Kuperman. Robust matched-field processing using a coherent broadband white noise constraint processor. *J. Acoust. Soc. Am*, 122(4):1979–1986, 2007.

[96] K. L. Gemba, W. S. Hodgkiss, , and P. Gerstoft. Adaptive and compressive matched field processing. *J. Acoust. Soc. Am*, 141:92–103, 2017.

[97] Kay L. Gemba, Santosh Nannuru, Peter Gerstoft, and William S. Hodgkiss. Multi-frequency sparse bayesian learning for robust matched field processing. *J. Acoust. Soc. Am*, 141(5):3411–3420, 2017.

[98] A. B. Baggeroer, W. A. Kuperman, and H. Schmidt. Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem. *J. Acoust. Soc. Am.*, 83:571–587, 1988.

[99] H. Schmidt, A. B. Baggeroer, W. A. Kuperman, and E. K. Scheer. Environmentally tolerant beamforming for high-resolution matched field processing: Deterministic mismatch. *J. Acoust. Soc. Am.*, 88:1851– 1862, 1990.

[100] Gihoon Byun, F. Hunter Akins, Kay L. Gemba, H. C. Song, and W. A. Kuperman. Multiple constraint matched field processing tolerant to array tilt mismatch. *J. Acoust. Soc. Am.*, 147:1231–1238, 2020.

[101] Stan E. Dosso and Michael J. Wilmut. Uncertainty estimation in simultaneous bayesian tracking and environmental inversion. *J. Acoust. Soc. Am*, 124(1):82–97, 2008.

[102] Stan E. Dosso and Michael J. Wilmut. Comparison of focalization and marginalization for bayesian tracking in an uncertain ocean environment. *J. Acoust. Soc. Am*, 125(2):717–722, 2009.

[103] Stan E. Dosso, Michael J. Wilmut, and Peter L. Nielsen. Bayesian source tracking via focalization and marginalization in an uncertain mediterranean sea environment. *J. Acoust. Soc. Am*, 128(1):66–74, 2010.

[104] Z. H. Michalopoulou, Andrew Pole, and Ali Abdi. Bayesian coherent and incoherent matched-field localization and detection in the ocean. *J. Acoust. Soc. Am.*, 146:4812–4820, 2019.

[105] William Mantzel, Justin Romberg, and Karim Sabra. Compressive matched-field processing. *J. Acoust. Soc. Am*, 132(1):90–102, 2012.

[106] Donald F. Gingras and Neil L. Gerr. Minimax robust matched-field processing. *J. Acoust. Soc. Am*, 93(5):2798–2808, 1993.

[107] L. T. Fialkowski, M. D. Collins, W. A. Kuperman, J. S. Perkins, L. J. Kelly, A. Larsson, J. A. Fawcett, and L. H. Hall. Matched-field processing using measured replica fields. *J. Acoust. Soc. Am.*, 107:739–746, 2000.

[108] Dag Tollefsen and Stan E. Dosso. Three-dimensional source tracking in an uncertain environment. *J. Acoust. Soc. Am*, 125(5):2909–2917, 2009.

[109] Cedric A. Zala and John M. Ozard. Matched-field processing in a range-dependent environment. *J. Acoust. Soc. Am*, 88(2):1011–1019, 1990.

[110] E. K. Westwood. Broadband matched-field source localization. *J. Acoust. Soc. Am.*, 91:2777–2789, 1992.

[111] Z. H. Michalopoulou and M. B. Porter. Matched-field processing for broadband source localization. *IEEE J. Ocean. Eng.*, 21:384–392, 1996.

[112] S. P. Czenszakand and J. L. Krolik. Robust wideband matched-field processing with a short vertical array. *J. Acoust. Soc. Am.*, 101:749–759, 1997.

[113] C. Soares and S. M. Jesus. Broadband matched-field processing: Coherent and incoherent approaches. *J. Acoust. Soc. Am.*, 113:2587–2598, 2003.

[114] Michael D. Collins, Laurie T. Fialkowski, W. A. Kuperman, and John S. Perkins. Environmental source tracking. *J. Acoust. Soc. Am*, 94(6):3335–3341, 1993.

[115] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge, MA, 2012.

[116] Andreas Köhler, Matthias Ohrnberger, and Frank Scherbaum. Unsupervised pattern recognition in continuous seismic wavefield records using self-organizing maps. *Geophysical Journal International*, 182(3):1619–1630, 2010.

[117] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

[118] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521:436–444, 2015.

[119] E. Çakır, G. Parascandolo, T. Heittola, H. Huttunen, and T. Virtanen. Convolutional recurrent neural networks for polyphonic sound event detection. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(6):1291–1303, 2017.

[120] Ananya Sen Gupta, Craig Kletzing, Robin Howk, William Kurth, and Morgan Matheny. Automated identification and shape analysis of chorus elements in the van allen radiation belts. *Journal of Geophysical Research: Space Physics*, 122(12):12,353–12,369, 2017.

[121] Qingkai Kong, Daniel T. Trugman, Zachary E. Ross, Michael J. Bianco, Brendan J. Meade, and Peter Gerstoft. Machine learning in seismology: Turning data into insights. *Seismo. Res. Lett.*, 90(1):3–14, 2018.

[122] Karianne J. Bergen, Paul A. Johnson, Maarten V. de Hoop, and Gregory C. Beroza. Machine learning for data-driven discovery in solid earth geoscience. *Science*, 363(6433), 2019.

[123] M. J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M. A. Roch, S. Gannot, and C.-A. Deledalle. Machine learning in acoustics: Theory and applications. *J. Acoust. Soc. Am.*, 146:3590–3628, 2019.

[124] Erin M. Fischell and Henrik Schmidt. Classification of underwater targets from autonomous underwater vehicle sampled bistatic acoustic scattered fields. *J. Acoust. Soc. Am*, 138(6):3773–3784, 2015.

[125] Erin M. Fischell and Henrik Schmidt. Environmental effects on seabed object bistatic scattering classification. *J. Acoust. Soc. Am*, 141(1):28–37, 2017.

[126] A. S. Gupta and I. Kirsteins. Disentangling sonar target features using braided feature graphs. In *OCEANS 2017 - Anchorage*, pages 1–5, 2017.

[127] Xanadu C. Halkias, Sébastien Paris, and Hervé Glotin. Classification of mysticete sounds using machine learning techniques. *J. Acoust. Soc. Am*, 134(5):3496–3505, 2013.

[128] Lior Shamir, Carol Yerby, Robert Simpson, Alexander M. von Benda-Beckmann, Peter Tyack, Filipa Samarra, Patrick Miller, and John Wallin. Classification of large acoustic datasets using machine learning and crowdsourcing: Application to whale calls. *J. Acoust. Soc. Am*, 135(2):953–962, 2014.

[129] Marielle Malfante, Jérôme I. Mars, Mauro Dalla Mura, and Cédric Gervaise. Automatic fish sounds classification. *J. Acoust. Soc. Am*, 143(5):2834–2846, 2018.

[130] H. Niu, E. Reeves, and P. Gerstoft. Source localization in an ocean waveguide using supervised machine learning. *J. Acoust. Soc. Am.*, 142:1176–1188, 2017.

[131] Haiqiang Niu, Emma Ozanich, and Peter Gerstoft. Ship localization in santa barbara channel using machine learning classifiers. *J. Acoust. Soc. Am*, 142(5):EL455–EL460, 2017.

[132] Y. Wang and H. Peng. Underwater acoustic source localization using generalized regression neural network. *J. Acoust. Soc. Am.*, 143:2321–2331, 2018.

[133] Jianyun Yangzhou, Zhengyu Ma, and Xun Huang. A deep neural network approach to acoustic source localization in a shallow water tank experiment. *J. Acoust. Soc. Am.*, 146:4802–4811, 2019.

[134] Z. Huang, J. Xu, Z. Gong, H. Wang, and Y. Yan. Source localization using deep neural networks in a shallow water environment. *J. Acoust. Soc. Am.*, 143:2922–2932, 2018.

[135] J. Chi, X. Li, H. Wang, D. Gao, and P. Gerstoft. Sound source ranging using a feed-forward neural network trained with fitting-based early stopping. *J. Acoust. Soc. Am.*, 146:EL258, 2019.

[136] H. Niu, Z. Gong, E. Ozanich, P. Gerstoft, H. Wang, and Z. Li. Deep-learning source localization using multi-frequency magnitude-only data. *J. Acoust. Soc. Am.*, 146:211–222, 2019.

[137] H. Niu, Z. Gong, E. Ozanich, P. Gerstoft, H. Wang, and Z. Li. Deep learning for ocean acoustic source localization using one sensor. *ArXiv e-prints*, 2019.

[138] W. Wang, H. Ni, L. Su, T. Hu, Q. Ren, P. Gerstoft, and L. Ma. Deep transfer learning for source ranging: Deep-sea experiment results. *J. Acoust. Soc. Am.*, 146:EL317, 2019.

[139] W. Liu, Y. Yang, M. Xu, L. Lv, and Y. Shi. Source localization in the deep ocean using a convolutional neural network. *J. Acoust. Soc. Am.*, 147:EL314, 2020.

[140] E. Ozanich, P. Gerstoft, and H. Niu. A feedforward neural network for direction-of-arrival estimation. *J. Acoust. Soc. Am.*, 147:2035–2048, 2020.

[141] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.*, 79(8):2554–2558, 1982.

[142] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.

[143] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter. Self-normalizing neural networks. *ArXiv e-prints*, 2017.

[144] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[145] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biol. Cybernetics*, 36:193–202, 1980.

[146] N. Srivastava, G. Hinton, and A. Krizhevsky. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15:1929–1958, 2014.

[147] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of The 32nd International Conference on Machine Learning*, page 448–456, Lille, France, July 6-11, 2015.

[148] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[149] Yann A. LeCun, Léon Bottou, Genevieve B. Orr, and Klaus-Robert Müller. *Efficient BackProp*, pages 9–48. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[150] D. P. Kingma and J. L. Ba. Adam: A method for stochastic optimization. *ArXiv e-prints*, 2014.

[151] S. Han, J. Pool, J. Tran, and W. Dally. Learning both weights and connections for efficient neural networks. *ArXiv e-prints*, 2015.

[152] Laurie T. Fialkowski, John S. Perkins, Michael D. Collins, Michael Nicholas, John A. Fawcett, and W. A. Kuperman. Matched-field source tracking by ambiguity surface averaging. *J. Acoust. Soc. Am*, 110(2):739–746, 2001.

[153] J. Murray and D. Ensberg. The SWellEx-96 experiment. (Last viewed Dec. 8, 2020). http://swellex96.ucsd.edu/.

[154] G. L. D'Spain, J. J. Murray, W. S. Hodgkiss, N. O. Booth, and P. W. Schey. Mirages in shallow water matched field processing. *J. Acoust. Soc. Am.*, 105(6):3245–3265, 1999.

[155] Rex Allen. Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, 72(6B):S225–S242, 1982.

[156] Mitchell Withers, Richard Aster, Christopher Young, Judy Beiriger, Mark Harris, Susan Moore, and Julian Trujillo. A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bulletin of the Seismological Society of America*, 88(1):95–106, 1998.

[157] A. Trnkoczy. *Understanding and parameter setting of STA/LTA trigger algorithm*, pages 1–20. GFZ, Potsdam, Germany, 2012.

[158] W. C. Cummings, O. I. Diachok, and J. D. Shaffer. *Acoustic transients of the marginal sea ice zone: A provisional catalog.* Naval Research Laboratory, Washington, D.C., 1989.

[159] K. L. Williams, M. L. Boyd, A. G. Soloway, E. I. Thorsos, S. G. Kargl, and R. I. Odom. Noise background levels and noise event tracking/characterization under the Arctic ice pack: Experiment, data analysis, and modeling. *IEEE Journal of Oceanic Engineering*, 43(1):145–159, 2018.

[160] Irwin Sobel. An isotropic 3×3 image gradient operator. *Presentation at Stanford A.I. Project*, 1968.

[161] ItSeez. The OpenCV Reference Manual. (Last viewed Feb. 15, 2021). http://opencv.org/.

[162] S.C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32:241–254, 1967.

[163] National Oceanic and Atmospheric Administration. HICEAS Dataset, 2017. (Last viewed August 8, 2020).

[164] S. Rankin, J. Barlow, Y. Barkley, and R. Valtierra. A guide to constructing hydrophone arrays for passive acoustic data collection during NMFS shipboard cetacean surveys, 2013. U.S. Dept. of Commerce, NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-511.

[165] Ildar R. Urazghildiiev and Christopher W. Clark. Acoustic detection of north Atlantic right whale contact calls using spectrogram-based statistics. *J. Acoust. Soc. Am*, 122(2):769–776, 2007.

[166] Mark F. Baumgartner and Sarah E. Mussoline. A generalized baleen whale call detection and classification system. *J. Acoust. Soc. Am*, 129(5):2889–2902, 2011.

[167] Jérôme Weiss and David Marsan. Scale properties of sea ice deformation and fracturing. *Comptes Rendus Physique*, 5(7):735–751, 2004.

[168] Jérôme Weiss. Intermittency of principal stress directions within Arctic sea ice. *Phys. Rev. E*, 77:056106, 2008.

[169] Jérôme Weiss, David Marsan, and Pierre Rampal. Space and time scaling laws induced by the multiscale fracturing of the Arctic sea ice cover. In *IUTAM Symposium on Scaling in Solid Mechanics*, pages 101–109, Dordrecht, 2009. Springer Netherlands.

[170] Jacqueline A. Richter-Menge, Donald K. Perovich, Bruce C. Elder, Keran Claffey, Ignatius Rigor, and Mark Ortmeyer. Ice mass-balance buoys: a tool for measuring and attributing changes in the thickness of the Arctic sea-ice cover. *Annals of Glaciology*, 44:205–210, 2006.

[171] P. Rampal, J. Weiss, D. Marsan, R. Lindsay, and H. Stern. Scaling properties of sea ice deformation from buoy dispersion analysis. *Journal of Geophysical Research: Oceans*, 113(C3), 2008.

[172] P. Rampal, J. Weiss, and D. Marsan. Positive trend in the mean speed and deformation rate of Arctic sea ice, 1979–2007. *Journal of Geophysical Research: Oceans*, 114(C5), 2009.

[173] Chris Polashenski, Don Perovich, Jackie Richter-Menge, and Bruce Elder. Seasonal ice mass-balance buoys: adapting tools to the changing Arctic. *Annals of Glaciology*, 52(57):18–26, 2011.

[174] Erland M. Schulson. Compressive shear faults within Arctic sea ice: Fracture on scales large and small. *Journal of Geophysical Research: Oceans*, 109(C7), 2004.

[175] F. Vincent, D. Raucoules, T. Degroeve, G. Edwards, and M. Abolfazl Mostafavi. Detection of river/sea ice deformation using satellite interferometry: limits and potential. *International Journal of Remote Sensing*, 25(18):3555–3571, 2004.

[176] H. L. Stern and R. W. Lindsay. Spatial scaling of Arctic sea ice deformation. *Journal of Geophysical Research: Oceans*, 114(C10), 2009.

[177] Dyre O. Dammann, Hajo Eicken, Andrew R. Mahoney, Franz J. Meyer, Jeffrey T. Freymueller, and Alexander M. Kaufman. Evaluating landfast sea ice stress and fracture in support of operations on sea ice using sar interferometry. *Cold Regions Science and Technology*, 149:51–64, 2018.

[178] D. O. Dammann, L. E. B. Eriksson, A. R. Mahoney, H. Eicken, and F. J. Meyer. Mapping pan-Arctic landfast sea ice stability using Sentinel-1 interferometry. *The Cryosphere*, 13(2):557–577, 2019.

[179] Bruce E. Miller and Henrik Schmidt. Observation and inversion of seismo-acoustic waves in a complex Arctic ice environment. *J. Acoust. Soc. Am*, 89(4):1668–1685, 1991.

[180] James K. Lewis, Walter B. Tucker III, and Peter J. Stein. Observations and modeling of thermally induced stresses in first-year sea ice. *Journal of Geophysical Research: Oceans*, 99(C8):16361–16371, 1994.

[181] T. C. Yang and G. R. Giellis. Experimental characterization of elastic waves in a floating ice sheet. *J. Acoust. Soc. Am*, 96(5):2993–3009, 1994.

[182] T. C. Yang and T. W. Yates. Flexural waves in a floating ice sheet: Modeling and comparison with data. *J. Acoust. Soc. Am*, 97(2):971–977, 1995.

[183] Peter J. Stein, Steven E. Euerle, and James C. Parinella. Inversion of pack ice elastic wave data to obtain ice physical properties. *Journal of Geophysical Research: Oceans*, 103(C10):21783–21793, 1998.

[184] Yuriy V. Dudko, Henrik Schmidt, Keith von der Heydt, and Edward K. Scheer. Edge wave observation using remote seismoacoustic sensing of ice events in the Arctic. *Journal of Geophysical Research: Oceans*, 103(C10):21775–21781, 1998.

[185] Ludovic Moreau, Pierre Boué, Agathe Serripierri, Jérôme Weiss, Daniel Hollis, Ildut Pondaven, Benjamin Vial, Stéphane Garambois, Éric Larose, Agnés Helmstetter, Laurent Stehly, Gregor Hillers, and Olivier Gilbert. Sea ice thickness and elastic properties from the analysis of multimodal guided wave propagation measured with a passive seismic array. *Journal of Geophysical Research: Oceans*, 125(4):e2019JC015709, 2020.

[186] Ludovic Moreau, Jérôme Weiss, and David Marsan. Accurate estimations of sea-ice thickness and elastic properties from seismic noise recorded with a minimal number of geophones: From thin landfast ice to thick pack ice. *Journal of Geophysical Research: Oceans*, 125(11):e2020JC016492, 2020.

[187] Maurice Ewing, A. P. Crary, and A. M. Thorne. Propagation of elastic waves in ice. part i. *Physics*, 5(6):165–168, 1934.

[188] Maurice Ewing and A. P. Crary. Propagation of elastic waves in ice. part ii. *Physics*, 5(7):181–184, 1934.

[189] S. E. Dosso, G. J. Heard, and M. Vinnins. Seismo-acoustic propagation in an ice-covered Arctic Ocean environment. In *MTS/IEEE Oceans 2001. An Ocean Odyssey. Conference Proceedings (IEEE Cat. No.01CH37295)*, volume 3, pages 1487–1491 vol.3, 2001.

[190] Stan E. Dosso, Garry J. Heard, and Michael Vinnins. Source bearing estimation in the Arctic Ocean using ice-mounted geophones. *J. Acoust. Soc. Am*, 112(6):2721–2734, 2002.

[191] Stan E. Dosso, Michael Vinnins, and Garry J. Heard. Arctic field trials of source bearing estimation using ice-mounted geophones (l). *J. Acoust. Soc. Am*, 113(6):2980–2983, 2003.

[192] J. E. White. Motion product seismograms. *Geophysics*, 29(2):288–298, 1964.

[193] Åke Björck. Least squares methods. In *Handbook of Numerical Analysis*, volume 1, pages 465–652. Elsevier, 1990.

[194] Douglas Schwarz. Fast and robust curve intersections. (Last viewed February 18, 2021). https://www.mathworks.com/matlabcentral/fileexchange/11837-fast-and-robust-curve-intersections.

[195] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.

[196] S. J. Pan and Q. Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10):1345–1359, 2010.