OF RATS AND MEN: UNDERWATER PASSIVE ACOUSTIC LOCALIZATION INVESTIGATIONS USING RELATIVE ARRIVAL TIMES AND BLIND CHANNEL ESTIMATION

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

OCEAN AND RESOURCES ENGINEERING

DECEMBER 2022

By

Brendan P. Rideout

Dissertation Committee:

Eva-Marie Nosal, Chairperson
Anders Høst-Madsen
Bruce M. Howe
Marc O. Lammers
Lora J. Van Uffelen

Keywords: underwater acoustics, localization, blind channel estimation, fin whale

But Mousie, thou art no thy lane,
In proving foresight may be vain:
The best laid schemes o' mice an' men
Gang aft agley,
An' lea'e ue nought but grief an' pain,
For promis'd joy!

—

Robert Burns, "To a Mouse, on Turning Her Up in Her Nest With the Plough, November, 1785"

ACKNOWLEDGMENTS

I am profoundly grateful to Eva-Marie Nosal, my advisor and mentor throughout the eleven years of my doctoral program. Your unwavering confidence in me and insight have guided me through both successes and struggles, and gives me an admirable example of scientific leadership I endeavor and aspire to live up to.

I thank my other committee members (Anders Høst-Madsen, Bruce Howe, Marc Lammers, and Lora Van Uffelen) for their guidance and interest throughout this long process.

I thank my supervisor and managers (Mark Trevorrow, Michel Couillard, Warren Connors, Dan Hutt, and Vince Myers) for their support and guidance as I juggled school and work responsibilities, and my colleagues for many interesting conversations.

Thanks to the ALOHA Cabled Observatory team for providing access to, and technical support for, their data archive.

I thank Robert Dunn for providing data from his 2018 air gun survey near ACO that was essential in demonstrating the performance of my single hydrophone localization method.

Thanks to Stephen Becker for assistance in configuring his NESTA software.

Funding for this work was provided by the National Science Foundation, the US Navy Office of Naval Research, the University of Hawaii, and the Canadian Armed Forces.

I thank my friends in the Marine Mammal Research Program here at the University of Hawaii for adopting me and answering my marine mammal ecology questions over chili, volleyball, and Halloween parties.

Finally, I thank James Harvey for his editorial assistance with this document, and my family & friends for supporting my scientific and engineering endeavors all these long years. It takes a whole village to raise a graduate student, and I feel privileged to have had the opportunity to study with, and learn from, you all.

ABSTRACT

Understanding the ecology of any organism requires an understanding of all its life stages. Underwater acoustics provides the ability to observe the submerged lives of marine mammals in ways not possible through visual means. The complexities of underwater acoustic propagation yield both challenges and opportunities to extract information from recorded data, of which the estimation of underwater location is one example. This dissertation presents two signal processing approaches related to passive underwater acoustic localization.

Blind channel estimation is a computational approach to estimating a set of impulse responses based on simultaneous recordings of the same unknown source by different receivers. The estimated impulse responses from this approach may simplify passive acoustic localization for some types of sound sources by making arrival times easier to identify. Differences among the received waveforms are interpreted as evidence of differences in the underlying channel impulse responses. Using a sparse assumption on these impulse responses, several different optimization approaches (OMP, CoSaMP, and NESTA) are applied to simulated ocean acoustic data. Estimated channels are a good fit to the true channel when the channels are static, but the introduction of a time-varying characteristic to the true channels negatively impacts channel recovery performance.

Single hydrophone passive acoustic localization is the practice of estimating location characteristics of an underwater sound source using acoustic recordings from a single receiver. This work develops an approach for estimating the horizontal range between a submerged source and receiver in a deep ocean environment without relying on modal dispersion. We develop a cost function and optimization approach that are robust to some significant sources of noise and environmental uncertainty. Results from simulations and ground-truthed measured data demonstrate the accuracy of this localization approach.

Underwater acoustic data recorded by the ALOHA Cabled Observatory (ACO) are processed using this single hydrophone localization approach. Acoustic recordings at ACO made between February 2007 - September 2017 yield 41,481,171 fin whale 20 Hz call detections, of which 3,445,568 detections remain after pruning away suspected sei whale and duplicate call detections. Fin whale 20 Hz calls are concentrated in the November-April time period near ACO. Estimated fin whale call parameters, including inter-call interval, are more clearly estimated from the recorded data following pruning to reduce the false positive rate.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Understanding the ecology of any organism requires an understanding of all life stages of this organism. Accurately assessing the effect of anthropogenic activities, and developing strategies for mitigating these effects, on an animal population requires, among other information, knowledge of their spatial and temporal extent (i.e., where this population tends to be found, and when the population can be found there). For terrestrial and airborne species, there are established tracking procedures to monitor latitude, longitude, and elevation. This tracking data informs the establishment of wildlife preserves whose boundaries reflect the home ranges of the different species in question, and can be used in Environmental Impact Assessments. Due to the unique challenges faced by underwater data collection, establishing the location of underwater organisms requires different technology and approaches to acquire the same information.

Marine mammals are an important constituent of the marine environment. The term 'marine mammal' is not a taxonomic grouping, but rather traditionally refers to the set of mammals that derive most or all of their diet from the ocean. In Jefferson et al. [2015], the set of marine mammals is said to include members of the Order Cetacea (i.e., toothed whales and baleen whales), some members of the Order Carnivora (including pinnipeds, some otter species, and polar bears) and members of Order Sirenia (e.g., manatees). Marine mammals are found throughout the worlds oceans, and live in an environment that can be disturbed by underwater anthropogenic noise sources (Jefferson et al. [2015]). Ambient acoustic background noise levels in the Northeast Pacific have increased over the last fifty years (McDonald et al. [2006]), which has reduced the communication space for many marine mammal species (Clark et al. [2009]) (i.e., the distance one animals calls can be reliably perceived by a conspecific). Oil exploration activities may have temporary yet important effects on the rates at which nearby bowhead whale calls are localized, likely due to animal deflection away from the noise source or a reduction in calling rates (Blackwell et al. [2013]). Even ocean acidification could impact marine mammal communication space by decreasing acoustic absorptivity, in particular below 10 kHz (Hester et al. [2008]), thereby increasing not only the distance animal calls travel underwater but the effective travel distance of noise sources, too. In summary, marine mammals live in an environment encroached on by human activity over a variety of different mechanisms that can have a measurable impact on behavior. Understanding the human role in these changes is necessary to developing and implementing strategies for mitigating the effects on marine mammal ecology in this changing environment.

In order for mitigation strategies to be effective, they must be informed by the behavior and ecology of the animals in question. Knowledge of their spatial distribution, and any seasonality in this distribution, is important information to have when setting shipping lanes, vessel speed limits in sensitive habitats (e.g., Davies and Brillant [2019]), and when establishing the spatial and temporal boundaries for marine sanctuaries. An important component in understanding marine mammal ecology is understanding their movement patterns on a variety of times scales. What times of year do they inhabit which parts of the ocean? Does their calling behavior change over the course of a day or a season? What are their daily movement patterns? Understanding where marine mammals spend their time is critical to assessing the impact of human activities on their life cycle.

Currently, measuring marine mammal location can be done using visual, tag, or acoustic means. Visual

observation techniques have a long and continuing history for the study of marine mammals, and can provide useful insight into marine mammal ecology. In modern times, visual observation using trained human observers is an often-used tool to study marine mammals (Gordon et al. [2004]), as well as to monitor for marine mammal presence during seismic surveys (Compton et al. [2008]). Trained observers are taught to recognize the characteristic shapes of whale exhalations (called blows), dorsal and pectoral fin shapes, and tail markings. In some cases, visual identification can be done to the species level, but for some species, such as killer whales (*orcanus orca*), identification of individuals is also carried out (Lockyer and Morris [1990]). In addition to using light in the visible spectrum, recent research has explored using infrared (IR) cameras to spot marine mammal dorsal fins and blows (i.e., the warm flesh and exhaled gases from the animals) against the backdrop of colder ocean water (Cuyler et al. [1992]). Recent implementations of IR-based methods for marine mammal monitoring include land-based (e.g., Perryman et al. [1999], Graber et al. [2011]), airborne (e.g., Burn et al. [2006]), and ship-based (Zitterbart et al. [2013]). A recent review paper (McCafferty [2007]) describes a compilation of literature on the use of infrared measurement to study mammals, including marine mammals.

Tags, of which DTAGs (Johnson and Tyack [2003]) are a commonly-used example, are another approach to measuring the presence of marine mammals. Tagging studies take place on a variety of time scales, from hours (e.g., Stimpert et al. [2007]) to months (e.g., Mate et al. [2011]), and from one animal at a time (e.g., Mate et al. [2011]) to a host of simultaneously tagged animals (e.g., Jay et al. [2010]). Tags can measure a variety of types of data (including acoustic, GPS position, and depth). Tag delivery and attachment techniques include pole-assisted suction cups (e.g., Johnson and Tyack [2003]), crossbow and air guns (e.g., Jay et al. [2006]), and other animal-mounted solutions (e.g., to a walrus tusk as in Wiig et al. [1993]). Tagging studies can focus on measuring animal depth over time (Wiig et al. [1993]), or also include acoustical measurement to monitor the calls of one (e.g., Stimpert et al. [2007]) or multiple nearby marine mammals (e.g., Chen et al. [2016]). In summary, tagging studies can take a wide range of forms and collect a variety of different types of data. However, tagging studies can be labor intensive and expensive and attaching the tag to the animal involves an unavoidable degree of disturbance of the animal.

Underwater acoustic monitoring of marine marine mammals is a large field, and includes studies that work on small to large spatial scales, short duration to multiple months or years, and that focus on a single species or on a variety of different species. The recording platform can be stationary (e.g., moored to the bottom as in Rideout et al. [2013]) or mobile (e.g., Lammers et al. [2006]). One interesting example of a mobile acoustic recording platform is described in Klinck et al. [2012], which uses a hydrophone mounted on a Seaglider to carry out wide scale ($\sim$ 390 km) monitoring over three weeks at sea off the coast of Hawaii. Sometimes underwater acoustics is used to monitor migration paths (e.g., Davies and Brillant [2019]), other times population density is estimated from the recorded acoustic data (e.g., Marques et al. [2009]).

Since humans began acoustically monitoring the ocean, techniques have been developed to make use of the large range observation capabilities available to underwater acoustic sensors. Locating submerged targets (also known as localization) is one such field which has received considerable investigative attention. Knowledge of underwater marine mammal position is important not only from a purely ecological perspective, but also provides useful guidance in cases where human and marine mammal activities may interact. For example, decisions on establishing protected sanctuaries for marine mammals are more effective when knowledge of the spatial and temporal extent of the animals is known and can be properly weighed against

(often) competing human interests in the area. Reliably minimizing marine mammal exposure to underwater anthropogenic noise sources, such as from oil exploration (Compton et al. [2008]), requires knowledge of their location.

With the exception of the polar bear (*Ursus maritimus*), all marine mammal species spend large parts of their lives underwater. Some (e.g., pinnipeds) spend time out of water (on ice or land) at certain stages of their lives. Whales and dolphins spend their entire lives at sea, and typically spend more time away from the ocean surface than at it. Even under optimal conditions, electromagnetic (EM) radiation in the wavelength range visible to the human eye (390-700 nm range) only penetrates on the order of 200 m into the ocean, and turbidity can reduce this distance to a small fraction of this value. Thus, when long distance monitoring of marine mammals, particularly underwater monitoring, is desired visual techniques will fail to observe all but the closest submerged animals. In contrast to underwater EM transmission, underwater sound transmission is relatively efficient (Frisk [1994]), resulting in a single underwater acoustic sensor being able to monitor a larger ocean space than a single visual sensor could. This has helped motivate the development of techniques for acoustically monitoring underwater marine mammals.

In addition to the immediate usefulness of knowing marine mammal position for the protection of these animals from loud oceanic anthropogenic activities, knowledge of underwater marine mammal movement patterns is an important marine mammal ecological constituent to understand due to the largely underwater lives these animals live. Marine mammals spend much of their lives underwater, where visual observation is largely ineffective, that an understanding of a given species habitat utilization, for example, based solely on the surface expression of that species would be incomplete. Knowledge of marine mammal location over time can help provide insight into feeding depths, swim speed, and other biologically and physiologically-relevant insights. In some cases, visual techniques may be used to gauge the presence of nearby marine mammals, but due to the virtual impossibility of collecting underwater visual observations over ecologically-relevant ranges, acoustic techniques are typically preferred for long-range studies.

This dissertation develops and applies passive underwater acoustic techniques, which are of particular interest to the marine mammal community as they do not subject the animal in question to additional acoustic noise (in the way that an active acoustics-based underwater surveillance technique would). Passive underwater acoustic techniques are those which use the sounds and vocalizations produced by the animals themselves to gather insight into the underwater world (typically, but not exclusively limited to, knowledge pertaining to the animal such as position and calling rate). For example, an active approach to underwater localization could be to produce a sound at a known location, listen for reflections of that sound off a submerged whale at a set of underwater acoustic receivers at known locations, and infer the position of that whale based on the time between sound creation and reception of the reflection at each individual receiver. A passive acoustics approach to locating the same animal would be to record the calls of the whale on the same set of underwater acoustic receivers and use the relative times of arrival of these calls at the known receiver locations to estimate whale position. Advantages of passive acoustic localization techniques over active acoustic methods are:

1. no risk of exposure of the animal to additional anthropogenic noise from the experiment (i.e., no change in behavior or hearing damage from the measurement);

2. simpler equipment requirements (no need for an underwater acoustic projector)

In contrast, use of active acoustic localization can afford greater control over the detection range of the

experiment (through variation in source level and characteristics), although they require more equipment. Active acoustic localization can also locate silent underwater targets, while passive acoustics studies rely on the subject to generate sound. Studies involving active acoustics can also require additional permitting and government oversight due to the protected status of many marine mammal species. Active acoustic localization has recently been used to study sperm whale presence near Hawaii Island (Giorli et al. [2018]).

Before either passive or active acoustic localization can take place, the relevant features from the recorded acoustic waveform must be extracted. As will be discussed in Chapter 3, one such feature is the set of direct and interface-reflected arrival times for a single call arriving at each receiver. For some types of calls, these arrival times can be visually apparent and established underwater acoustic signal processing techniques (e.g., cross-correlation) adapted to extracting these times. In other cases, though, these features may not be reliably extracted with these established practices. In some cases, the sought-after features may be obscured due to the nature of the underlying source (e.g., humpback whale songs) and/or the characteristics of the underwater acoustic propagation environment. To address such a case, Chapter 2 discusses a technique called Blind Channel Estimation (BCE), which aims to estimate waveguide impulse responses based on multiple recordings of a single unknown source. In other cases, the relevant features may generally be apparent for an individual call, but the sheer quantity of data requires a compromise be made between the quality and quantity of the extracted features (e.g., multi-year data sets of acoustic recordings containing seasonal fin whale presence). This second case is covered in Chapters 3 and 4, which cover the automation of marine mammal localization, including the automated extraction of localization data from recorded underwater acoustic recordings, and applied to a single-hydrophone multiyear data set recorded north of Hawaii.

1.1 Dissertation Overview

This dissertation presents the results of two distinct research projects, each connected to the field of passive acoustic underwater sound source localization. In **Chapter 2**, a technique for measuring ocean acoustic channel impulse responses called Blind Channel Estimation (BCE) is presented. Our implementation of BCE aims to assist in the identification of acoustic arrival times of an unknown source recorded by two or more spatially-separated hydrophones. These estimated acoustic arrival times could then be used to estimate source position through a subsequent analysis, including by the technique presented in Chapter 3. Ultimately, the performance of BCE in estimating simulated ocean acoustic impulse responses was found to be poor when ocean swell-induced variability in water depth was introduced into the simulation. This degree of time-variability degraded channel estimation performance to such a degree that the estimated channels were a poor match to the true channels.

The results of a second research project are presented in Chapters 3 and 4. In this second project, an approach for estimating the horizontal range between an underwater sound source and a single receiver is developed. A localization cost function is produced which is more robust to false positives and multiple simultaneous sound sources than contemporary localization approaches have been. Single hydrophone localization methods require less equipment, and simpler deployment procedures, than multiple hydrophone deployments. Our localization approach reduces the false positive rate through a process we call phantom pruning by using the localization result to limit each direct and interface-reflected arrival to locate no more than one call.

Our localization approach uses the time differences in direct and reflected arrival paths of a single source at a single receiver to estimate horizontal range. These relative arrival times (RATs) are the difference in arrival time between the earliest arrival and all subsequent higher order bounce paths. The RATs change in a predictable way as the horizontal range between source and receiver changes. By emulating the environment near the receiver using an acoustic propagation model, a set of model-predicted RATs can be estimated at the set of source ranges which are to be assessed as the best fit range, given the measured data.

For this work, the measured data are recorded the the ALOHA cabled observatory (ACO), located approximately 115 km north of Oahu. Figure 1.1 shows a ray trace plot for the vicinity of ACO using a water depth of 4667 m, a source depth of 50 m, and a sound speed profile collected at ACO in January 2012 (Figure 1.2). The BELLHOP ray tracing propagation model (Porter and Liu [1994]) was used for this plot. The maximum depth of the limiting ray (the red-highlighted ray in Figure 1.1) is approximately 4300 m, which is also called the conjugate depth. The conjugate depth is the minimum water depth at which convergence zones become present in a given environment and there are ray paths which do not interact with the bottom. Practically speaking, conjugate depth is of minimal importance to this work given that the data were collected at an ocean bottom hydrophone.

The bottom-limiting ray (blue-highlighted in Figure 1.1) reaches the closest point to the ocean bottom at a range of 31.6 km from the source. Thus, the direct-path cutoff range is expected to be approximately 30 km.

Two convergence zones are visible in Figure 1.1 and highlighted in green: one at 63 km and another at 126 km. These two ranges are the distance from a source, given a 50 m source depth, that a turning ray from the source will return to the surface.

Figure 1.3 shows the five eigenray paths, and their labels, used for the single hydrophone range estimation portion of this work.

In addition to a set of simulations, our single hydrophone localization approach is demonstrated using acoustic data collected by the ocean bottom hydrophone at the ALOHA cabled observatory. Recordings of an air gun survey, where the true range to each air gun shot is known *a priori*, demonstrate the reliability of our localization approach. Low frequency fin whale vocalizations are then localized in the data recorded at ACO between February 2007 and September 2017. Call parameters for these fin whale calls are presented, including peak frequency, minimum and maximum frequency, and inter-detection interval.

The remaining chapters in this dissertation are laid out as follows:

Chapter 2 discusses blind channel estimation (BCE) as a potential means to infer characteristics of the ocean acoustic environment. Background information on BCE, and its ties to compressed sensing, are presented. Next, a derivation for BCE governing equation is carried out and its assumptions discussed. A brief comparison of several optimization approaches to solve this governing equation is then made. Simulation and measured-data results will be shown to highlight some potential drawbacks to the use of BCE in a realistic underwater acoustical environment.

Chapter 3 develops and describes a single-hydrophone approach to estimating the range of a sound source from the hydrophone. A brief history of passive acoustic marine mammal localization approaches is presented, along with motivation for the development of single hydrophone techniques over those requiring multiple hydrophones. Each step of the localization signal processing sequence is described. A process for estimating the range-dependent range resolution is presented and applied to generate the set of candidate

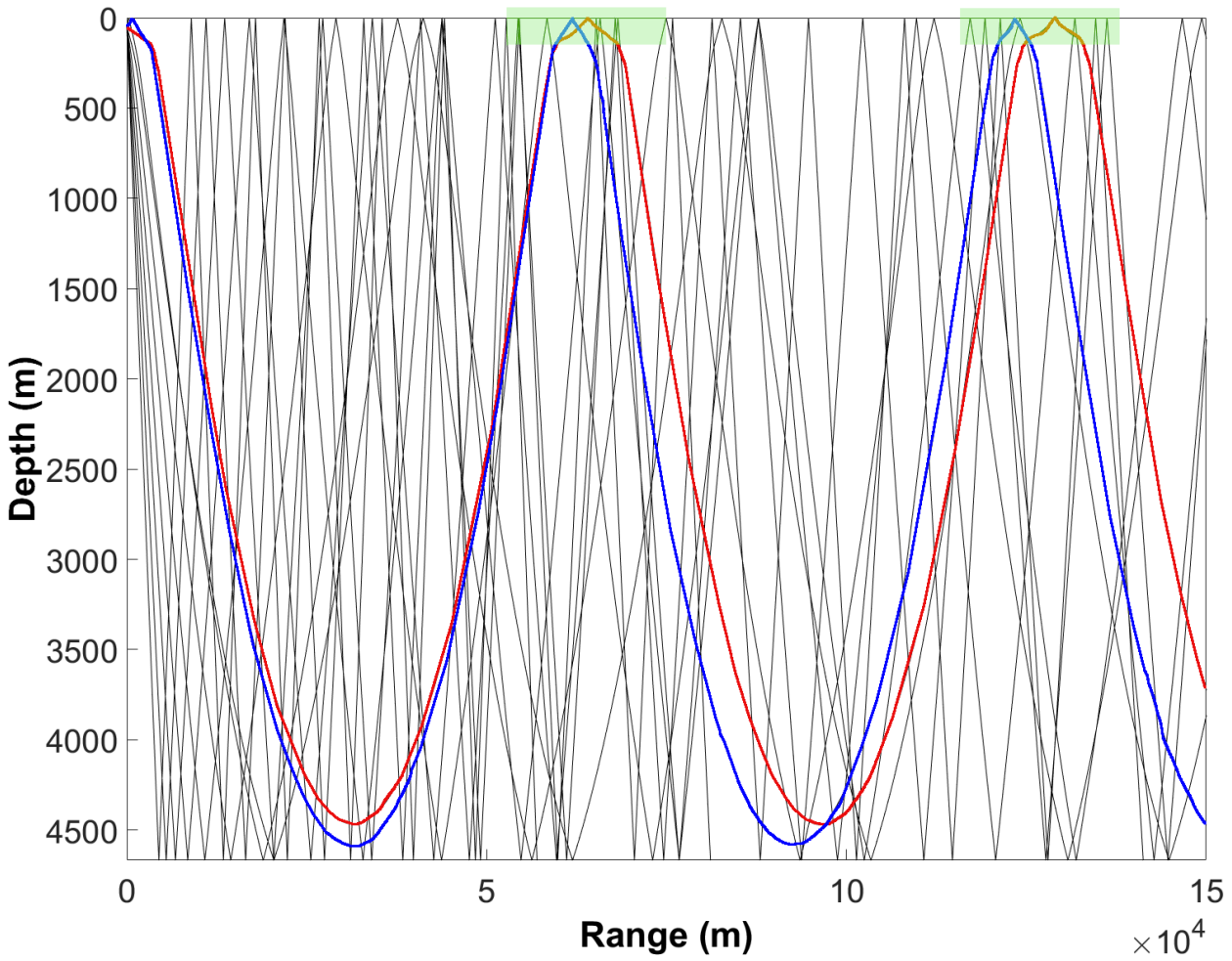

Figure 1.1: Ray trace plot for the conditions near the ALOHA cabled observatory (ACO) using the BELLHOP ray tracing acoustic propagation model. A water depth of 4667 m was used, and a sound speed profile collected in January 2012 at the ACO site. The source depth was set at 50 m to emulate the source depths used in Chapters 3 and 4. The red ray approximates the limiting ray, from which the conjugate depth for for this situation ($\approx$ 4300 m) can be estimated. The blue ray approximates the bottom-limiting ray, from which the range the direct-path arrival ($\approx$ 31 km) stops connecting can be seen. The two green highlight regions show two convergence zones at the ocean surface.

source ranges used for localization. Ground-truthed air gun results demonstrate the performance of the single hydrophone range estimation approach when the true source range is know.

Chapter 4 applies the localization approach from Chapter 3 to the task of estimating ranges to calling fin whales in the vicinity of Station ALOHA, an oceanographic data collection site located roughly 115 km north of Oahu. The history and configuration of the data collection efforts at Station ALOHA are described, as are the motivations for investigating fin whales at this site and historical fin whale passive acoustic studies around Hawaii. Fin whale calling rates as reported by the call detection algorithm are compared with those informed by the localization algorithm. A comparison of fin whale results from this work and those in McDonald and Fox [1999] and Thompson and Friedl [1982] is also made.

Chapter 5 summarizes the results presented in this work, draws some conclusions based on the results presented herein, and presents suggestions for avenues of future research.

Appendix A is a glossary of commonly used terms found in this dissertation.

Appendix B presents simulation results that demonstrate the performance of the approach in idealized and more realistic settings, including how multiple calling animals can influence localization performance.

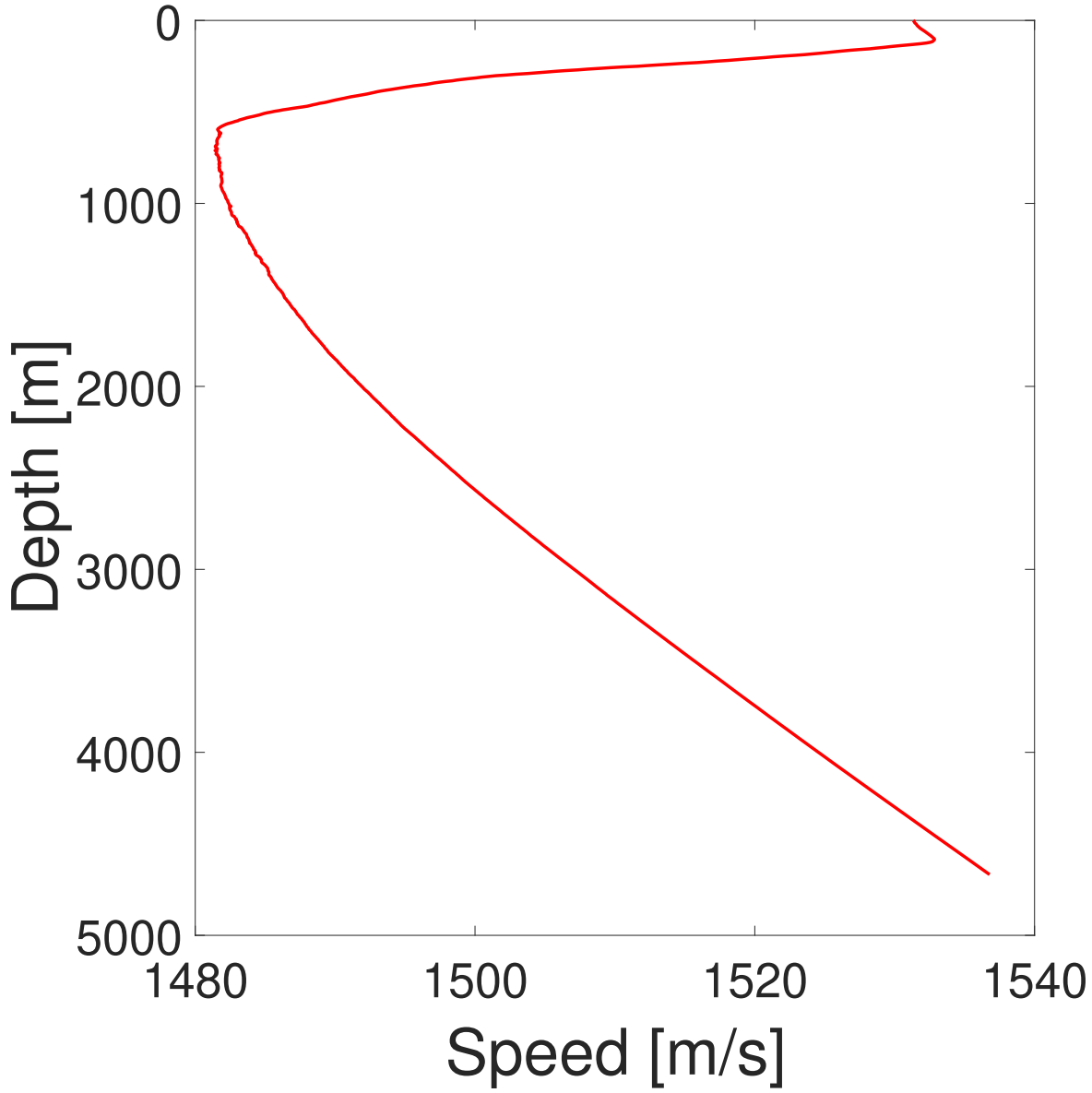

Figure 1.2: Sound speed profile calculated from CTD data collected during HOT cruise #239 and the Chen and Millero [1977] calculation approach near the ACO data collection site. HOT Website: https://hahana.soest.hawaii.edu/hot/cruises.html

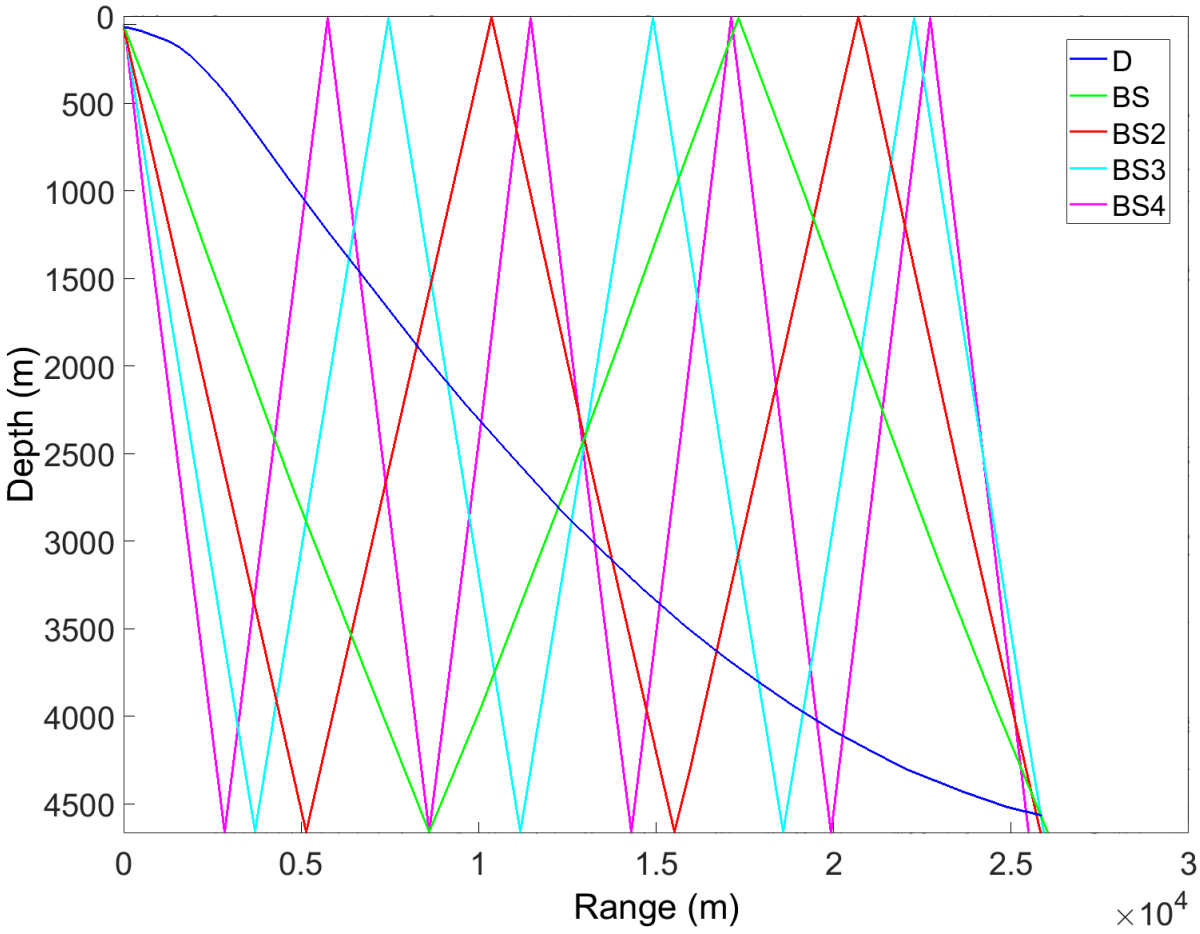

Figure 1.3: Eigenray paths for the five rays used for localization in this work: Direct (D), Bottom Surface (BS), Bottom Surface Bottom Surface (BS2), Bottom Surface Bottom Surface Bottom Surface (BS3), and Bottom Surface Bottom Surface Bottom Surface Bottom Surface (BS4). A source approximately 26 km from the ocean-bottom receiver, and at a depth of 50 m, is modeled here using the same environmental conditions used in Figure 1.1.

CHAPTER 2
BLIND CHANNEL ESTIMATION

Acoustical oceanography, defined as, "...the active or passive use of sound to study physical parameters and processes, as well as biological species and behaviors, at sea," (Medwin and Clay [1998]) dates at least as far back as 1912, when interest in locating large submerged objects following the sinking of the RMS *Titanic* spurred research into the use of sound to accomplish this task. Over time, applications of underwater sound have included tracking marine mammals which make impulsive (Watkins and Schevill [1972]) or non-impulsive vocalizations (Stafford et al. [1998]) and using acoustical thermometry (which is a specific application in the more general field of ocean acoustic tomography) to measure ocean temperature trends (Dushaw et al. [2009]). Localization and thermometry both involve a sound produced in one oceanic location being measured by hydrophones in another. The interaction of this sound with the ocean environment (e.g., through refraction and reflections off the ocean:air and ocean:bottom interfaces of the water column) acts on the source sound to produce the sound observed by the hydrophone. Effectively, this interaction between sound and environment encodes information about the source, receiver, and environmental characteristics into the sound that reaches the hydrophone. For example, consider Figure 2.1, where an impulsive sound source is produced at four different range:depth combinations relative to a single fixed ocean-bottom acoustic receiver in the absence of noise. As the source changes position relative to the receiver, the received sound changes to reflect the different environmental acoustical information encoded into the received sound. These plots are only meant to illustrate how the information content of the received waveform changes with changing source location, and leaves out important acoustical phenomena such as interface/volumetric scattering and background noise. Also, in practice, a realistic underwater source likely wouldn't have the perfect impulsive characteristics this source possesses.

Both underwater acoustical localization and thermometry are termed underwater inverse problems, where observations on the ocean environment are made using acoustic recordings. Underwater acoustical localization can, in general, be thought of as an application of matched field processing (Baggeroer et al. [1993]). In the underwater context, matched field processing (MFP) uses observed spatial variability in the measured underwater acoustic field to estimate parameters for a given acoustic propagation model (e.g., sound source location, water depth, and/or sound speed). Essentially, a set of model parameters is sought which produces a model-predicted acoustic field that is a good match to the observed field. In thermometry, acoustic ray travel times between known source and receiver locations are used to estimate mean sound speed and ocean temperature between source and receiver.

For cases where the source signal is controlled (Dushaw et al. [2009]) or relatively impulsive (Rideout et al. [2013]), features in the received waveform relevant to the inverse problem can be estimated relatively easily. In other cases, even if the source is unknown and non-impulsive enough features can be extracted to allow further analysis (Stafford et al. [1998], Valtierra et al. [2013]). However, in other cases there likely remains unused information in the received signal which is not apparent in either time or frequency domain analysis. For example, an underwater recording of a humpback whale will consist of not just the acoustic energy traveling directly between whale and hydrophone, but also higher-order eigenrays which reflect one or more times off the ocean surface and/or bottom before reaching the hydrophone. The arrival times of these

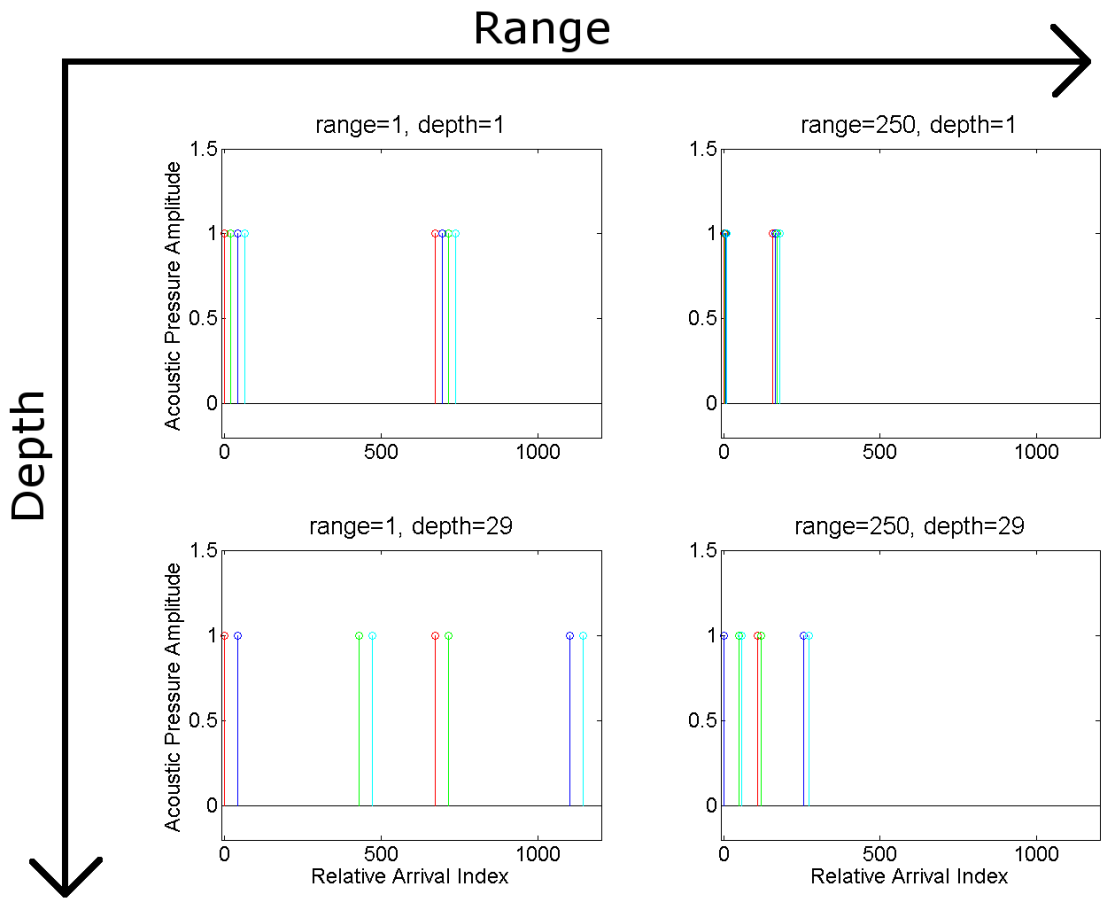

Figure 2.1: Eigenray arrival times for 4 different source locations. The x-axis is measured in samples after the arrival of the direct-path sound at the simulated hydrophone. Range and depth are dimensionless quantities meant to convey relative differences in source range and depth between the four cases shown in this figure. Only the first 8 direct and interface-reflected arrival paths are shown. If the source is impulsive, these plots can also be interpreted as approximations of the impulse responses for these four cases. The receiver is held fixed at a range and depth (r,z) of (0,39) with a water depth of 40. Sources are at (1,1), (250,1), (1,39), and (250,39). A range independent propagation environment is used.

higher order reflections can be difficult to pick out from the direct path due to the length of the original source signal. However, including more of these higher order paths could lead to higher performance localization of non-impulsive acoustic sources and better performance in acoustic tomography using uncontrolled sources, among other applications.

One potential approach to reduce the significance of the source waveform on the extraction on information from the received waveform is to seek to estimate the underlying channel impulse response rather than extract information from the received waveform directly. A more complete description of what a channel impulse response is in the underwater acoustics context is presented in Section 2.1, but in brief a channel impulse response encapsulates the totality of effects (including source position, receiver position, water depth/bathymetry, sound speed profile, surface wave characteristics, and relative source:receiver velocity)

that that act on the source waveform to create the signal observed by the receiver. To illustrate the advantage of estimating the impulse response, rather than using features measured directly from the received waveform, consider the following example. A near-surface marine mammal is vocalizing in the vicinity of a bottom-moored receiver (located 1 m from the ocean bottom in 30 m of water). The animal produces both impulsive (i.e., broadband & short duration) and tonal (i.e., narrowband & long duration) calls. Arrival times for impulsive calls are easier to pick out in the received data than for tonal calls. However, if the tonal received waveform could be processed in such a way to make it more closely resemble the impulsive received waveform, those same (simple) processing techniques that are used for the impulsive calls could also be used for the tonal calls. This is one of the prime motivations for this investigation of blind channel estimation in the context of passive marine mammal acoustical monitoring: seeking to process a received acoustic signal in such a way that the environmental information present in the underwater acoustic channel impulse response, and encoded into the received signal, is more readily available.

This chapter presents the development and testing of an approach to estimate underwater acoustic impulse responses without knowledge of the source signal and with few assumptions on the nature of these impulse responses. Simulations will be used to support potential explanations for diminished BCE performance using measured data. Ultimately, the estimated impulse responses are a poor match to the true channels (in simulation) when a reasonable degree of time variability is incorporated into the simulated environment. The work in this chapter is a distinct research project from the results shown in Chapters 3 and 4.

2.1 Underwater acoustic waveguide impulse response estimation via blind channel estimation and ℓ_1 optimization

In the underwater acoustics context, an acoustic waveguide approximately exists in the water between the ocean surface and bottom due to the large acoustic impedance contrast at these interfaces. These large impedance contrast interfaces serve to limit the amount of sound that is transmitted into the air and ocean bottom; when sound is incident at either surface most is reflected back into the water. This waveguide, and the underwater acoustical environment in general, modifies the source signal (e.g., a humpback whale or fin whale song) such that the sound field at a given location in that environment is a function of both the source waveform and the environmental parameters (e.g., the local bathymetry, source and receiver locations, and the sound speed profile).

Given that the environment dictates how the source signal is modified before arriving at the receiver, it is reasonable to assume that this modification encodes information onto the source waveform about the state of this environment. For example, if the receiver is moved to the middle of the water column (from its original position 1 m off the bottom), while keeping all other aspects of the environment the same, the received waveform will look different because the environment encoded different information onto the source waveform to create the received signal. There are a variety of techniques in acoustical oceanography to leverage the environmental 'information content' of the received signal to characterize an acoustical environment. One such technique is acoustical tomography, where a known sound signal is broadcast into the ocean and differences between the known source and measured received signal are used to infer properties of the ocean acoustic environment (e.g., large scale average ocean temperature). Another technique is underwater passive acoustic localization, where insight into sound source location (and occasionally a refinement of the measured environmental characteristics) is sought by leveraging the fact that the characteristics of the received signal

changes as the source moves underwater.

In both the tomography and localization cases, the effect of the environment on the source signal may be expressed as the impulse response of the environment, which is an important characteristic of a linear time-invariant (LTI) system. A system which obeys linearity is one where the superposition property holds (i.e., the output of a system to a sum of inputs is the same as the sum of the outputs from the individual inputs). Time invariance describes a system where the system output to a given input at time $t = 0$ is the same as at any other time. The impulse response ($h(t)$) of an LTI system is the response of the system when the input is a unit impulse, $\delta(t)$, which is an infinite-bandwidth infinitesimal time-duration signal. On the premise that any input signal can be thought of as a sequence of individual scaled unit impulses, the output from an LTI system for any given input signal can be predicted given $h(t)$. Of more pertinence to this study, analysis of the impulse response can provide insight into the environmental characteristics that went into shaping it. For example, a vocalizing whale swimming underwater while making a sequence of N identical calls can be modeled as the same source being acted on by a set of N impulse responses. Given that the source waveform is a unit impulse, Figure 2.1 can also be interpreted as showing the impulse response of four different acoustic environments.

Given recordings of an approximately impulsive sound source, it is possible to estimate the sound source position under certain conditions (Rideout et al. [2013]). For example, consider the case, shown in Figure 2.2, of three spatially-separated hydrophones deployed in 30 m of water to record Pacific walrus (*Odobenus rosmarus divergens*) knock vocalizations. Walrus knocks are short duration, wide bandwidth (i.e., approximately impulsive) vocalizations. Figure 2.3 shows a single knock recorded on each of the three hydrophones. These recordings show both the direct path sound (the sound that traveled from source to receiver without reflecting off either the surface or bottom) and interface-reflected sound (sound that reflected at least once off either ocean interface), and many of these arrivals were labeled so the walrus could be located. Due to the impulsive nature of the source waveform:

1. arrival times for interface-reflected sound are less obscured by earlier arrivals; and

2. the received waveforms approximate the underlying impulse responses.

For each source:receiver pair, there will be a different received waveform and thus a different approximated impulse response. By extracting information from these approximated impulse responses, we can estimate walrus position and refine our prior knowledge of some environmental parameters (e.g., receiver locations, water depth, sound speed, and inter-receiver time synchronization factors). Since this estimation can be made when the source is approximately impulsive, it should be possible to carry out the same estimation for a non-impulsive source signal if we can estimate the impulse responses from the received waveforms rather than recording them directly. This idea provides the motivating impetus for the work described in this chapter.

One approach to estimating the impulse response for an arbitrary system is called Blind Channel Estimation (BCE), which aims to calculate multiple impulse responses simultaneously given a set of received waveforms from a single source; no knowledge of this source waveform is needed. Dating at least as far back as 1975 (Sato [1975]), the aim of BCE was to improve wireless communication fidelity in time-varying environments (e.g., cell phone communication while traveling through urban canyons). In such environments, reflections of the RF signal interfere with signal reception. Given knowledge of the impulse response between

the source (e.g., the cell phone tower) and receiver (e.g., the phone), the received signal can be processed to compensate for environmental effects. Existing methods for operating in such environments required the sacrifice of valuable bandwidth (e.g., sending training signals such as might power a deconvolution-based correction protocol). Thus, a way to correct for these unknown and time-varying environmental effects was sought which don't require a training step.

A 1995 paper (Xu et al. [1995]) describes a problem formulation whereby multiple communication channels, exposed to the same source signal, could be identified without the source signal being known. The solution approach presented in this work required the channel order (i.e., the polynomial order of the Z-transform of the true impulse responses) be known *a priori*, but the authors suggested that it could be estimated during the solution process. A related approach is described in Moulines et al. [1995]. In a simulation study, both the least squares approach in Xu et al. [1995] and the subspace approach in Moulines et al. [1995] were found to yield worse results than several other sparse recovery approaches (Zeng et al. [2013]).

In the literature, the BCE problem is formulated as an ℓ_1 optimization problem, whereby a set of channel impulse responses is sought which are not only supported by the measured data but which consist of a minimum number of non-zero elements. Both the least squares (Xu et al. [1995]) and subspace (Moulines et al. [1995]) approaches employ an ℓ_1 optimization approach.

To improve upon the results obtained from the least squares and subspace approaches, particularly in the presence of noise and when channel order is unknown, a number of different optimization approaches were explored (e.g., Tropp and Gilbert [2007], Needell and Tropp [2009], Becker et al. [2011]) while working with a similar problem formulation to that described in Xu et al. [1995]. First, two greedy algorithms were implemented: Orthogonal Matching Pursuit (OMP, Tropp and Gilbert [2007]) and Compressive Sampling Matching Pursuit (CoSaMP, Needell and Tropp [2009]). Both are iterative and converge to a solution by taking locally optimal steps to minimizing the misfit between the measured data and an estimation of the model-predicted data. The number of non-zero values in the set of impulse responses, K, estimated by these algorithms is set *a priori*. In practice, these methods were found to be sensitive to the choice of K; underestimating K led to poor simulation results, and while slight over-estimation generally didn't cause problems grossly overestimating K seems to degrade results, too. This dependence on the value of K, which in practice will be difficult to set, led to a more robust optimization approach being sought.

Another class of algorithms for solving signal recovery problems is referred to as Basis Pursuit (BP, Chen et al. [2001]). NESTA (Becker et al. [2011]) is an example of a sparse recovery algorithm based on BP. BP-based algorithms aim to minimize the ℓ_1 norm of the recovered signal (or, in our case, the estimated impulse responses) while fitting to the measured data, thus preferring solutions with relatively few non-zero values. This approach is sometimes referred to as sparse estimation. While minimizing the ℓ_0 norm of the recovered signal would more precisely maximize the sparsity of the signal, the ℓ_1 norm approximation is used due to it being a convex optimization problem rather than the more difficult non-convex approach which an ℓ_0 minimization would require.

A 2013 paper (Zeng et al. [2013]) described simulation results intended to demonstrate that acoustic channels, both those in underwater and in-air environments, may be estimated using techniques such as those presented above. A comparison of results using the Least Squares (Xu et al. [1995]), Subspace (Moulines et al. [1995]), Matching Pursuit (Mallat and Zhang [1993]), Orthogonal Matching Pursuit (Tropp and Gilbert [2007]), and Basis Pursuit (Chen et al. [2001]) optimization approaches was presented. Simulation results

showed promise, but the choice in Zeng et al. [2013] to not use a standard underwater acoustic propagation model, and lack of measured data results, leaves room for improvement.

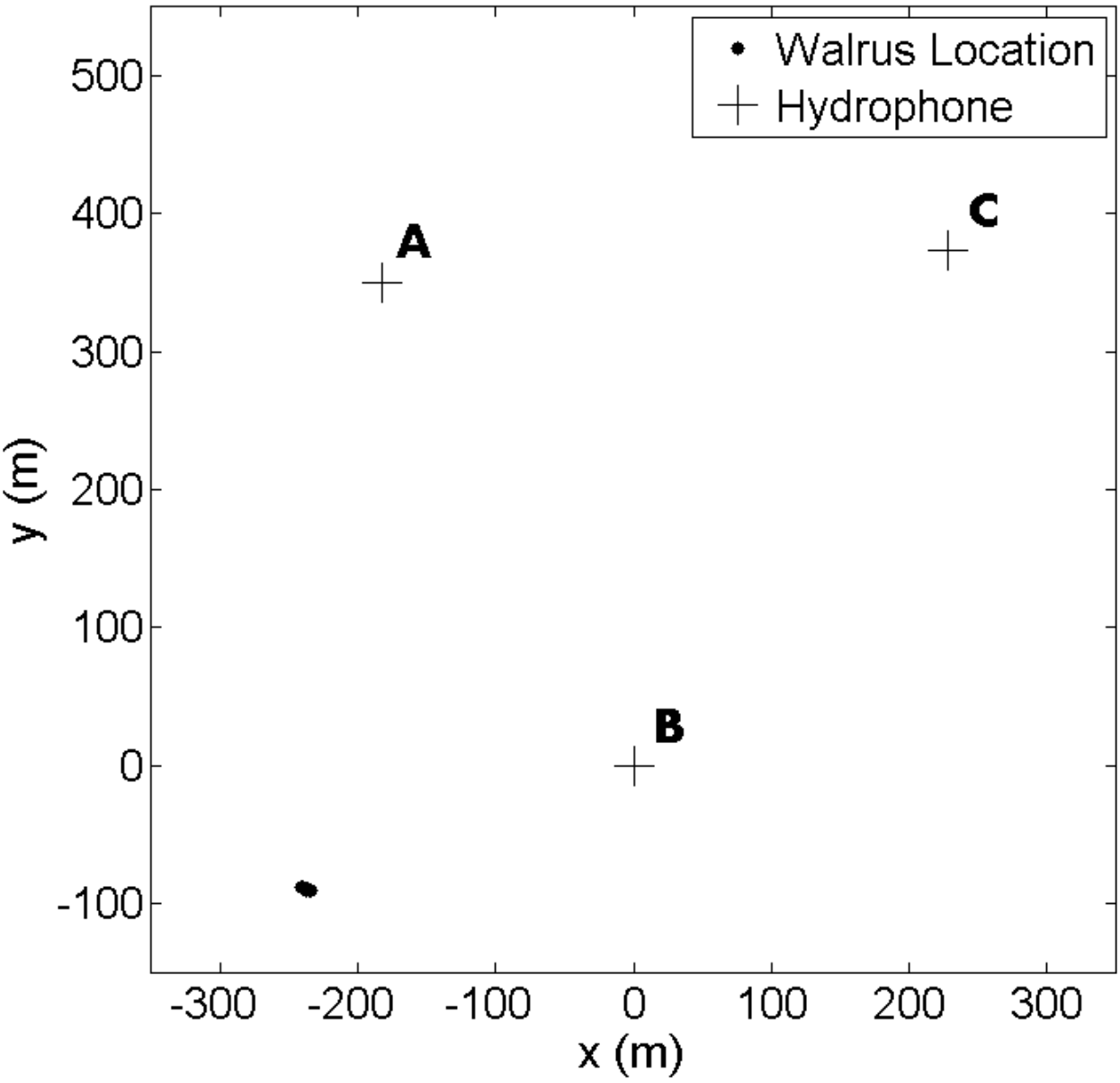

Figure 2.2: Estimated walrus locations relative to the estimated hydrophone positions. Hydrophones A, B, and C are labeled accordingly. Rideout et al. [2013]

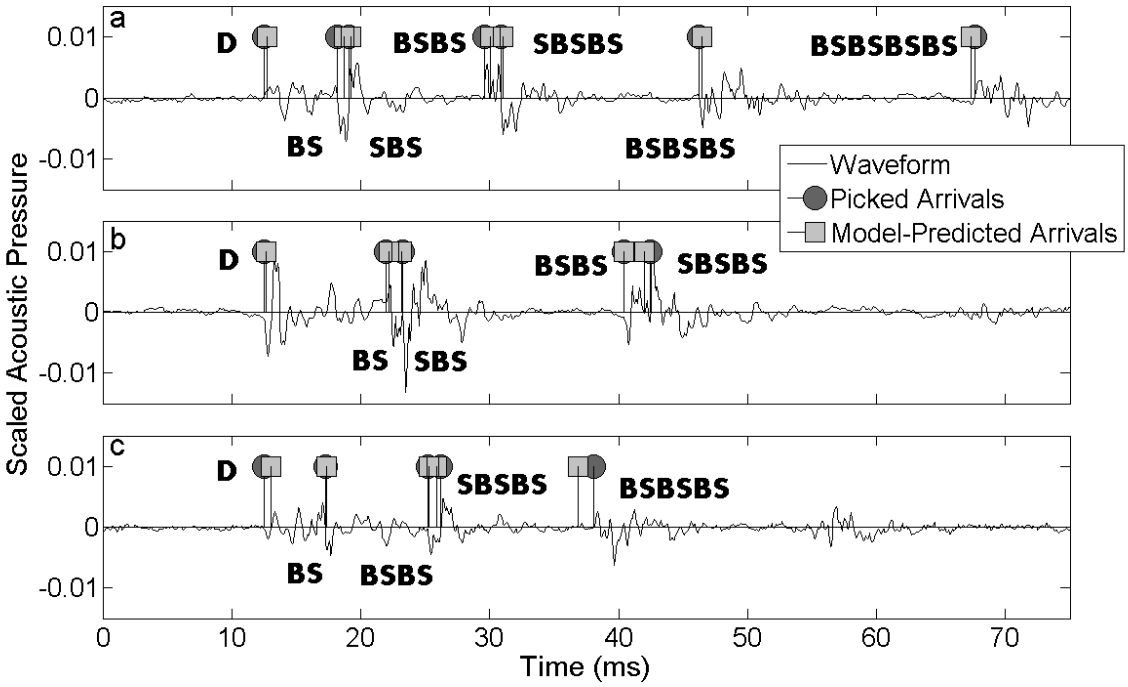

Figure 2.3: Picked and model-predicted arrival times along with the recorded acoustic time series for a single Pacific walrus knock recorded at (a) hydrophone A, (b) hydrophone B, and (c) hydrophone C (see Figure 2.2. Rideout et al. [2013]

2.2 BCE Problem Formulation

This section presents a development of the set of blind channel estimation governing equations. Vectors (e.g., $\mathbf{h}$) and matrices (e.g., $\mathbf{X}$) are shown in lower and uppercase boldface characters, respectively. The approach presented here could be extended to any number of receivers, and is based on the development presented in Xu et al. [1995].

For the one source/two receiver case, as described in Xu et al. [1995], the received (i.e., sampled) waveforms ($\mathbf{x}_i$ and $\mathbf{x}_j$) are expressed as

$$\mathbf{x}_i(k) = \mathbf{h}_i(k) \otimes \mathbf{s}(k) \tag{2.1}$$

$$\mathbf{x}_j(k) = \mathbf{h}_j(k) \otimes \mathbf{s}(k) \tag{2.2}$$

where $\mathbf{h}_i(k)$ is the impulse response for the i^{th} receiver, and $\mathbf{s}(k)$ is the source waveform. These two equations can be expressed as

$$\mathbf{h}_i(k) \otimes \mathbf{x}_j(k) = \mathbf{h}_i(k) \otimes \big(\mathbf{h}_j(k) \otimes \mathbf{s}(k)\big) \tag{2.3}$$

$$\mathbf{h}_i(k) \otimes \mathbf{x}_j(k) = \mathbf{h}_j(k) \otimes \big(\mathbf{h}_i(k) \otimes \mathbf{s}(k)\big) \tag{2.4}$$

$$\mathbf{h}_i(k) \otimes \mathbf{x}_j(k) = \mathbf{h}_j(k) \otimes \mathbf{x}_i(k) \tag{2.5}$$

$$\mathbf{h}_i(k) \otimes \mathbf{x}_j(k) - \mathbf{h}_j(k) \otimes \mathbf{x}_i(k) = 0 \tag{2.6}$$

This convolution relationship can be expressed as a system of linear equations:

$$\begin{pmatrix} \mathbf{X}_j(L) & -\mathbf{X}_i(L) \end{pmatrix} \begin{pmatrix} \mathbf{h}_i \\ \mathbf{h}_j \end{pmatrix} = 0 \tag{2.7}$$

$$\mathbf{X}(L)\mathbf{h} = 0 \tag{2.8}$$

where

$$\mathbf{X}_j(L) = \begin{pmatrix} x_j(L) & x_j(L+1) & \dots & x_j(2L) \\ x_j(L+1) & x_j(L+2) & \dots & x_j(2L+1) \\ \vdots & \vdots & \ddots & \vdots \\ x_j(N-L) & x_j(N-L+1) & \dots & x_j(N) \end{pmatrix} \tag{2.9}$$

,

$$\mathbf{h}_i \equiv [h_i(L), \dots, h_i(0)]^T \tag{2.10}$$

,

and $L+1$ is the length of the longest IR in the set.

Due to the ambiguities inherent to the problem formulation in Equation (2.8), BCE is unable to resolve the start time and scaling factor for the channels being estimated. In acoustical terms, there is magnitude and phase ambiguity in the solution. The literature presents at least two possibilities to resolve this ambiguity. In Xu et al. [1995], the constraint $|\mathbf{h}|_1 = 1$ is described. This addresses the scaling factor but not the phase ambiguity. In Zeng et al. [2013], the ℓ^{th} entry in $\mathbf{h}$ is defined to be equal to 1. All other values of $\mathbf{h}$ are

estimated relative to this value. To achieve this, the ℓ^{th} column of $\mathbf{X}$ is defined as $\mathbf{b}$ and ℓ is set equal to some constant (usually, $(L+1)/2$). The ℓ^{th} column of $\mathbf{X}$ is removed, and the set of remaining columns of $\mathbf{X}$ renamed to the measurement matrix $\mathbf{A}$. At the same time, the ℓ^{th} entry in $\mathbf{h}$ (having been defined as 1 to resolve the phase and amplitude ambiguity), is removed and the remaining vector renamed as $\mathbf{c}$. Thus, the set of governing equations becomes (for the two channel case) is now expressed as:

$$\mathbf{Ac} = \mathbf{b} \tag{2.11}$$

where

$$\mathbf{c} = [[h_1(L), \ldots, h_1(\ell+1), h_1(\ell-1), \ldots, h_1(0)], \mathbf{h}_2^T], \tag{2.12}$$

$\mathbf{b}$ is the ℓ^{th} column of $\mathbf{X}$, and $\mathbf{A}$ is obtained by removing the l^{th} column of $\mathbf{X}$. This is generalizable to any number of channels.

2.2.1 BCE Optimization Function Descriptions

In Section 2.2, the BCE problem formulation was presented, which yielded Equation (2.11) as the set of governing equations for this formulation. Thus, we can rephrase the objective of the blind channel estimation analysis as the calculation of $\mathbf{c}$ which satisfies Equation (2.11). We now require an optimization algorithm to solve Equation (2.11) for $\mathbf{c}$. Two early approaches to solving the blind channel estimation problem are described in Xu et al. [1995] and Moulines et al. [1995], and are referred to as the Least Squares (LS) and Subspace algorithms, respectively. A comparison between these two approaches is described in Zeng and Tong [1996]. In the LS method, the matrix $\mathbf{X}$ from Equation (2.8) is first formed by over-estimating the channel length L. Next, the singular value decomposition of $\mathbf{X} = \mathbf{U\Sigma V}^*$ is calculated. For a noise free case, the singular vector of $\mathbf{V}$ corresponding to the smallest non-zero singular values (i.e., $\mathbf{\Sigma}$) yields the set of channel estimates. In practice, though, for non-zero noise, the smallest non-zero singular value can indicate channel impulse response estimates that are more indicative of the underlying noise since there may not be any singular values equal to zero. According to Zeng et al. [2013], the Subspace method "... is based on the fact that the range space of the channel filtering matrix is orthogonal to the noise subspace". Both the LS and Subspace approaches require a large amount of data and are sensitive to knowing the length of the channels ahead of time. In practice, knowing channel length *a priori* can be difficult.

The choice of optimization approach is guided by the expectation that $\mathbf{c}$ for the underwater acoustic case should be a sparse vector. In Zeng et al. [2010], a sparse vector is defined as one with small, but nonzero, support due to the characteristics of ray propagation in an ocean acoustic waveguide. In an optics sense, a waveguide is a set of conditions which confines (with little energy loss) the expansion of a wave to one or two dimensions. For the case of underwater acoustic propagation, the large acoustic impedance contrasts that exist at the ocean:air and ocean:seabed interfaces (Frisk [1994]) tend to trap sound underwater with small, but nonzero, losses of acoustic energy into one of the flanking media when a sound reaches either of these two interfaces. The result of these reflections on the received acoustic signal are that this signal has discrete time intervals of source reception (either from acoustic energy traveling directly from source to receiver or reflecting one or more time off the surface and bottom of the waveguide) with intervening periods of no signal from the source arriving at the receiver. Figure 2.4 shows an example of the direct and reflected path arrivals of a signal arriving at a hydrophone without overlap. Eight individual sources, and their higher order

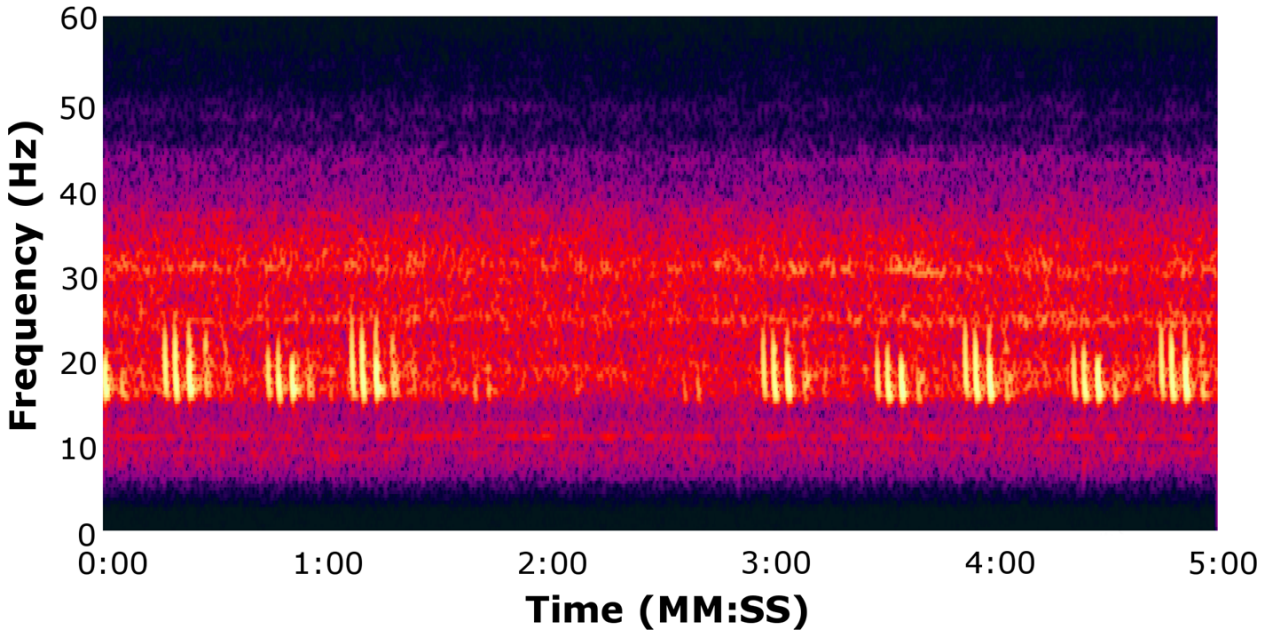

Figure 2.4: Eight individual fin whale calls recorded by the ALOHA cabled observatory. The source waveform is a frequency downsweep from $\approx 23 - 18$ Hz over a second. Deep water conditions at this receiver, and the source length, result in discrete bundles of acoustic energy reaching the hydrophone, with periods of comparatively low received level between these arrivals.

reflected arrivals, are shown in this figure. This ray-tracing model for acoustic propagation is the result of a high frequency approximation to the equations governing underwater acoustic propagation. In summary, under the assumption of a ray model being an appropriate description of the acoustic propagation conditions present, we expect the energy in the impulse response for the waveguide to be concentrated to relatively short time intervals whose total length is small compared to the overall length of the channels (i.e., sparse).

If the sparse channel assumption holds, then we will consider sparse representation approaches to solving for $\mathbf{c}$. Effectively, these approaches aim to estimate $\mathbf{c}$ such that most of the energy in $\mathbf{c}$ is concentrated in a small number of indices compared to the overall length of the vector. The search for a sparse representation for $\mathbf{c}$ would suggest that a $\mathbf{c}$ be sought which both fits to the measured data while having the minimum number of non-zero values (i.e., a minimum ℓ_0 constraint). However, this minimization problem is (NP)-hard (Zeng et al. [2010]). In many cases, a sparse representation of $\mathbf{c}$ can be calculated by instead applying a minimum ℓ_1 (i.e., the sum of the absolute value of the elements of $\mathbf{c}$) constraint, thus allowing the application of a broad set of algorithms (Baraniuk et al. [2010]). Zeng et al. [2013] and Baraniuk et al. [2010] classify these approaches as either convex optimization (e.g., Basis Pursuit [Chen et al. [2001]]) or greedy (e.g., CoSaMP [Needell and Tropp [2009]]) approaches. The approaches that fall into either of these two general categories have their own advantages and disadvantages.

Convex Optimization Approach

The convex optimization approach solves a well-understood minimization problem expressed in Boyd and Vandenberghe [2004] as:

$$\text{minimize } f_0(x) \tag{2.13}$$

$$\text{subject to } f_i(x) \leq 0, i = 1, \ldots, m, \tag{2.14}$$

$$a_i^T x = b_i, i = 1, \ldots, p \tag{2.15}$$

where the functions $f_0, \ldots, f_m$: $\mathrm{R}^n \rightarrow (R)$ are convex, i.e., satisfy

$$f_i(\alpha x + \beta y) \leq \alpha f_i + \beta f_i y. \tag{2.16}$$

or, more specifically for the blind channel estimation case, in Zeng et al. [2013] as

$$\min \parallel c \parallel_{\ell_1} \text{ subject to } \parallel \mathbf{Ac} - \mathbf{b} \parallel_{\ell_2} \leq \epsilon \tag{2.17}$$

where ϵ is proportional to the signal to noise ratio (SNR) on the data; as noise on the data increases, ϵ increases to give the sparse representation approach flexibility to fit to the underlying signal in the presence of increasing noise. Thus, we seek the vector $\mathbf{c}$ with the smallest ℓ_1 norm which fits to the measured data to the degree expected by the estimated SNR. Algorithms, such as Basis Pursuit, which solve this complex optimization approach can have high computational complexity compared to those approaches falling into the second, greedy category.

One example of a BP-based algorithm is called NESTA (Becker et al. [2011]), which stands for "Nesterov's algorithm", and is based on the smooth function convex optimization work in Nesterov [1983]. The authors state that the beneficial properties of NESTA, in addition to its accuracy and flexibility for a variety of different problems, are its computational efficiency (making it particularly useful for large scale problems which less efficient algorithms are less well suited for), and its robustness that comes from relying on only a few tuning parameters (largely just a smoothing parameter μ). A detailed description of the inner workings of NESTA are beyond the scope of this dissertation, and can be found in Becker et al. [2011] and Becker [2011].

Greedy Optimization Approach

Greedy approaches aim to solve Equation (2.11) in a less optimal, but frequently less computationally intensive, fashion than a rigorous convex optimization approach such as basis pursuit. Some examples of greedy approaches are Matching Pursuit (MP, Mallat and Zhang [1993]), Orthogonal Matching Pursuit (OMP, Tropp and Gilbert [2007]) and Compressive Sampling Matching Pursuit (CoSaMP, Needell and Tropp [2009]). Zeng et al. [2013] describes the MP and OMP algorithms as iterative approaches that both require the number of nonzero values in $\mathbf{c}$ be known *a priori*. One at a time, both the MP and OMP approaches select the index in $\mathbf{c}$ that best correlates with the residual vector (initialized to $\mathbf{b}$) from Equation (2.11). Next, the value of $\mathbf{c}$ at this index is calculated. For MP, this calculation is formulaic, whereas in OMP an additional, least squares, optimization step is performed (as per Equation 19 in Zeng et al. [2013]). Finally, the residual vector is updated to reflect the contribution of the newly calculated value in $\mathbf{c}$. These steps are repeated until the desired number of non-zero values is reached. Zeng et al. [2013] claims that the OMP

algorithm, due to the additional optimization step that it employs, is of higher performance than MP and is more comparable to a BP approach.

The CoSaMP algorithm is somewhat more complicated than either MP or OMP. It is an iterative approach based on OMP, and while it also requires the sparsity level (i.e., number of non-zero values) be known *a priori*, it claims to offer guarantees on execution speed and the quality of the recovered signal that OMP does not (Needell and Tropp [2009]). Unlike MP and OMP, CoSaMP estimates all the non-zero values at once, but still aims to select non-zero value indices that best match to the measured data (i.e., minimize $\mathbf{Ac} = \mathbf{b}$ from Equation (2.11)). The steps of the CoSaMP algorithm are laid out in Needell and Tropp [2009]. Translating the variables used in their algorithm depiction into those used in this dissertation, $\mathbf{a}$ is the impulse response vector $\mathbf{c}$, $\mathbf{\Phi}$ is the measurement matrix $\mathbf{A}$, and $\mathbf{u}$ is the vector $\mathbf{b}$.

Needell and Tropp [2009] lists three candidate "halting criteria" to assess when the CoSaMP algorithm has converged. Halting after a fixed number of iterations is the simplest candidate, and the one Needell and Tropp [2009] implies although this approach relies on the selection of an appropriate number of iterations to halt after, potentially either resulting in a set of channel estimates that are still converging or converged several iterations ago. The authors also suggest two alternative halting criteria which aim to assess the difference between measured and model-predicted data after each iteration and halt once a predefined similarity threshold is reached.

Channel and Source Suitability

It is reasonable to expect that not all sets of channels, and not all source waveforms, will be equally well suited for BCE. As is shown for the two channel case in Equation (2.8), BCE as it is formulated in this work interprets differences in the two received waveforms as differences in the channel impulse response between the single source and each of the two receivers. This assumes that each receiver is exposed to the same source waveform (e.g., that the source has no significant directionality) so that differences in received waveforms are due to channel impulse response differences. It also assumes that the two channels are sufficiently different. For example, consider a simple one source : two receiver case. For this example, the ocean surface and bottom are both assumed to be flat, the sound speed profile horizontally uniform, and background noise set to zero. The two receivers, and single source, are all omnidirectional, and the two receivers are the same range from the source and at the same depth. Under these conditions, regardless of the source characteristics, the received signals at the two receivers will be identical. Since there are no received signal differences, both sides of Equation (2.8) equal zero and no estimate of the two impulse responses is possible. In a more general sense, consider the channel identifiability criteria presented in Xu et al. [1995] and Rotili et al. [2012]:

- $p = Cs(n) \geq 2L_h - 1$; and
- The z-transforms of the channel impulse response vectors i.e., $h_i n$ are co-prime (i.e., no common factors between the impulse response vectors),

where Cs(n) is the linear complexity (Blahut [2012]) of the discrete-time source waveform $s(n)$, L+1 is the maximum number of samples in each of the channels (or, equivalently, L is the maximum order of the z-transform of the impulse response vectors $h_i(n)$ and M is the number of channels). The co-prime requirement arises from the fact that if the set of channel impulse responses are not co-prime there will be unrecoverable

features of the impulse responses that no BCE algorithm could estimate due to the structure of Equation (2.8).

The linear complexity test is related to the information content of the source waveform. For example, consider the case where $\mathbf{s}(k) = \sin(2 * \pi * f * k)$ (i.e., the source waveform consists of a sinusoid at a single frequency). In this case, the received waveforms represent the system response to the single frequency only. By contrast, consider the case where the input signal is a dirac delta (i.e., $\mathbf{s}(k) = \delta(k)$). By definition, the received waveforms from such a source waveform are equal to the respective impulse responses.

2.2.2 Proposed Optimization Approach

A number of different approaches to solving (2.11) have been explored, including Least Squares (LS, Xu et al. [1995]), Subspace (Sub, Moulines et al. [1995]), Orthogonal Matching Pursuit (OMP, Tropp and Gilbert [2007]), Compressive Sampling Matching Pursuit (CoSaMP, Needell and Tropp [2009]), and two different approaches under the NESTA umbrella (Becker et al. [2011]).

During testing, the Least Squares and Subspace approaches gave poor results in the presence of even low noise. The OMP and CoSaMP approaches, while yielding better results than either Least Squares or Subspace, required that both the channel order (i.e., the total number of samples in each IR being estimated) and the number of non-zero values to be known *a priori*. Estimating both these quantities by trial and error would have reduced the confidence we had in our estimated channels, in addition to increasing computational cost. After investigating a couple sparse recovery optimization approaches, NESTA was selected for further analysis.

Using Equation (2.11) as a starting point, the blind channel estimation problem can be formulated as a quadratically constrained ℓ_1-minimization problem (as per Equation 1.2 in Becker et al. [2011]):

$$\min \parallel c \parallel_{\ell_1} \text{ subject to } \parallel \mathbf{Ac} - \mathbf{b} \parallel_{\ell_2} \leq \epsilon \tag{2.18}$$

where $\mathbf{c} \epsilon R^p$ and $p = 2 * (L + 1) - 1$ (for the two channel case). The Lagrangian form (i.e., Equation 1.3 from Becker et al. [2011]) of this problem could also be solved. However, given the more intuitive relationship between the noise on the recorded acoustic data and ϵ the quadratic formulation is preferred.

The NESTA algorithm is based on a sampling theory known as compressed sensing (CS, Donoho [2006]). Essentially, CS is an approach to, "... condense the information in a compressible signal into a small amount of data, "(Becker [2011]), where a compressible signal is one in which most of the energy is concentrated in a small number of samples relative to the length of the signal. NESTA achieves this condensation through the ℓ_1 constraint shown in Eqn (2.18). NESTA is an iterative, sparse approximation approach which operates like a modified gradient descent optimization, and is based around ideas described in Nesterov [2005] which seek to improve the reliability of non-smooth optimization approaches while also speeding up the rate of convergence. The non-smooth optimization is achieved by substituting the ℓ_1 constraint (a non-smooth constraint) with a smoothed version. While this has the effect of solving an approximation to the problem described in Equation (2.18), the balance between the smoothed and non-smooth cost function can be tuned for each problem. The improved convergence rate is achieved through an approach called Continuation (Daubechies et al. [2008], Hale et al. [2007]), which adds a second layer of iteration on top of the original one. Continuation works by first converging to a solution to an approximation of Equation (2.18) (a second

Parameter	Value
Receiver 1 [m]	$(X, Y, Z) = (49.6, 0, 109.3)$
Receiver 2 [m]	$(X, Y, Z) = (161.1, 0, 355)$
Source [m]	$(X, Y, Z) = (68.2, 0, 150.3)$
Water Depth [m]	385
Sound Speed [m/s]	1530

Table 2.1: Environmental conditions shared by the static and time-varying channel simulations

approximation in addition to that made to smooth Equation (2.18)), using that converged solution as the starting model for an approximation closer to problem we're seeking to solve. This process continues until the desired difference between the approximated and desired cost function is reached.

In Becker et al. [2011], much of the emphasis is on cases where $\mathbf{A} * \mathbf{A}$ is an orthogonal projector. For cases where $\mathbf{A}$ is more fixed, this may be a reasonable assumption. However, when estimating underwater acoustic impulse responses using the formulation presented in this document the orthogonal projector requirement will not generally hold, even in the absence of noise. Thus, the singular value decomposition of $\mathbf{A}$ is passed to the NESTA algorithm. Additionally, the standard NESTA approach is extended to incorporate reweighting (§2.7.2.3 in Becker [2011]) and an overdetermined $\mathbf{A}$ (§3.7.3 in Becker [2011]). The overdetermined modification is particularly desirable for this work because we expect the data record to be much longer than the channels being estimated.

2.3 Impulse Response Channel Recovering Simulations

In this section, a set of simulations are shown which present the performance of several sparse recovery algorithms acting on the problem formulation in Equation (2.17).

2.3.1 Simulation - Static Channel

To explore how well the BCE approach could work in ocean acoustic conditions, a series of simulations was conducted based on near-shore ocean conditions. Features of these simulations are summarized in Table 2.1. Z is measured positive downward from the ocean surface. Source and receiver positions were chosen as though all were on a tilted vertical array. The objective of these simulations is to express the performance of three different sparse recovery algorithms (OMP, CoSaMP, and NESTA) when the true underlying channels are sparse (i.e., most of the IR values are equal to zero). The static channel simulation is done as follows. The true impulse response for both of the source:receiver pairs is calculated using a straight-line acoustic propagation model. Travel time and transmission loss along the first 8 eigenray paths (direct, bottom, surface, surface bottom, bottom surface, bottom surface bottom, surface bottom surface, and surface bottom surface bottom) are calculated using this model, yielding a pair of impulse responses; one for each source:receiver pair. This impulse response expresses the transformation the environment imparts on the source waveform to create the waveform observed at the receiver. The pair of impulse responses are normalized by the value of the greatest IR value, and leading zeros are removed. This yields a pair of minimum-length IRs whose largest value is 1 and whose first non-zero value occurs at the first sample. A synthetic 5-second pseudo

white Gaussian signal was chosen as the source waveform. This source signal is convolved with each of the calculated impulse response vectors to yield a pair of received waveforms. Additive white Gaussian noise is added to each received waveform to yield a signal to noise ratio of $\approx$ 47dB; zero-noise received data is not used here since some sparse recovery algorithms are numerically unstable with zero-noise data.

All three recovery algorithms require an over-estimate of the maximum length for the set of channels being estimated. For OMP and CoSaMP, an estimate of the sparsity level (i.e., the total number of non-zero values across both channels) is also required. For NESTA, the value for the tradeoff parameter λ is required. λ indirectly controls the sparsity level of the eventual solution; higher lambda values yield a sparser set of channels. A sparsity level of 30 was chosen for both OMP and CoSaMP. This value is greater than actual sparsity level for this simulation (16), so as to emulate the expectation that in a realistic setting the precise sparsity level of the channels under examination will be unknown. The default λ value of 1 is used for NESTA.

Figure 2.5 shows the true vs. estimated IR channels for the OMP, CoSaMP, and NESTA sparse recovery algorithms and the Receiver 1:Source IR. Figure 2.6 shows the same results, only with a threshold applied to the estimated channels such that only estimated values whose amplitude is greater than 0.01 are shown. OMP and CoSaMP provide good channel estimates even without an amplitude threshold. Since a sparsity level of 30 is chosen, small non-zero values are placed by these two algorithms in amongst the expected channel values. Before the threshold is applied, the large number of small amplitude samples makes picking out the larger samples difficult. After the amplitude threshold is applied, the largest non-zero values in the NESTA channel estimate are shown to be a close approximation of the true channel. Pre and post-amplitude threshold figures for the Receiver 2:Source IR are shown in Figures 2.7 and 2.8.

The good match between the true and estimated channels for the static channel case illustrates that all three of the recovery algorithms are capable of recovering the underlying channels when presented with high SNR data and sparse underlying channels. For OMP and CoSaMP, the extra non-zero values beyond the 16 required to completely characterize the channels are small values, which is the preferred outcome since the simulated data should not require large non-zero values other than the 16 found in the true IRs. For NESTA, the largest non-zero values in the estimated channels were a good match to the true channels.

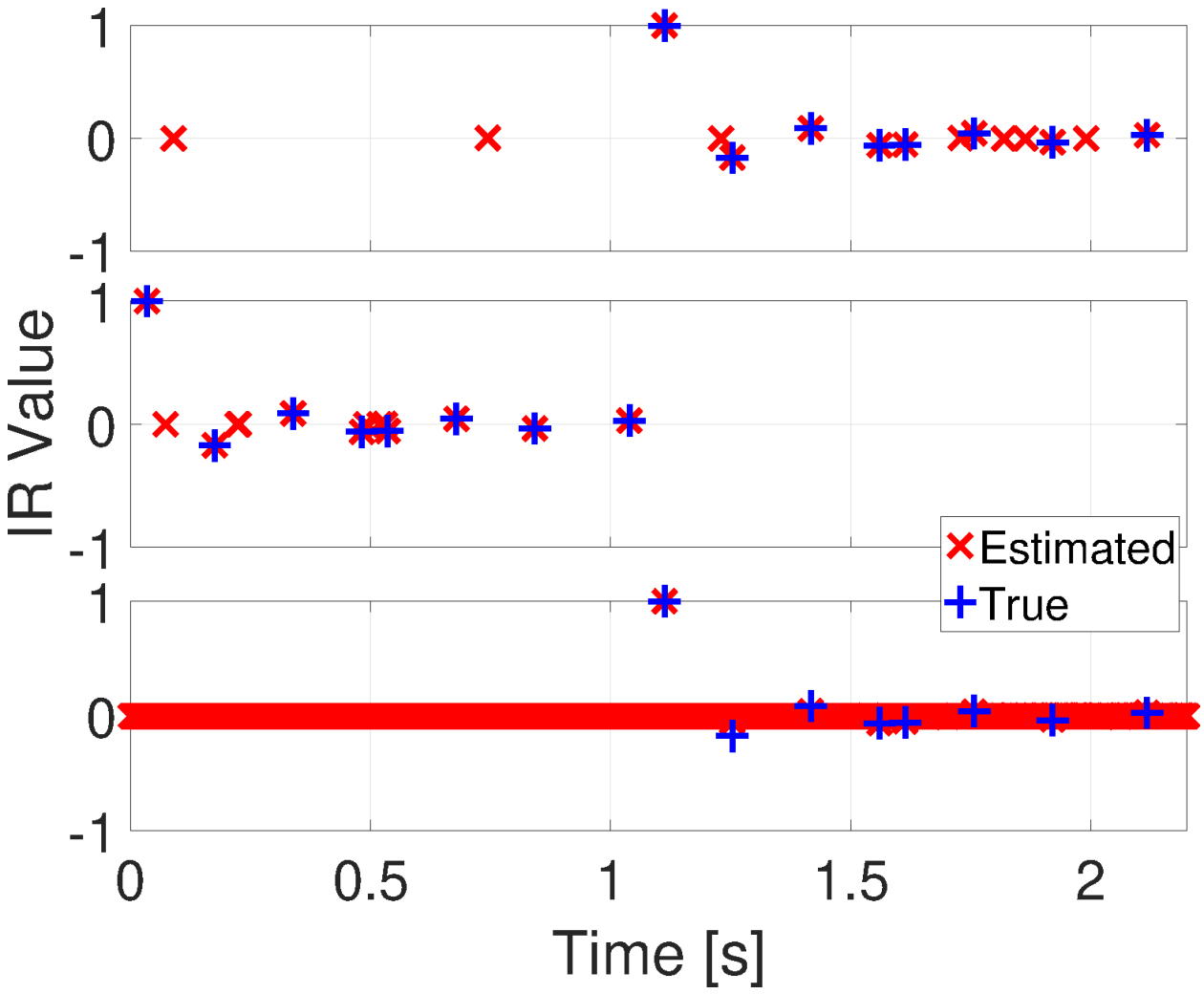

Figure 2.5: Estimated vs. true impulse response channels for the Receiver 1:Source pair. Results for OMP (top), CoSamP (middle) and, NESTA (bottom) are shown. A sparsity level of 30 is used for both OMP and CoSaMP, while the default λ value of 1 is used for NESTA.

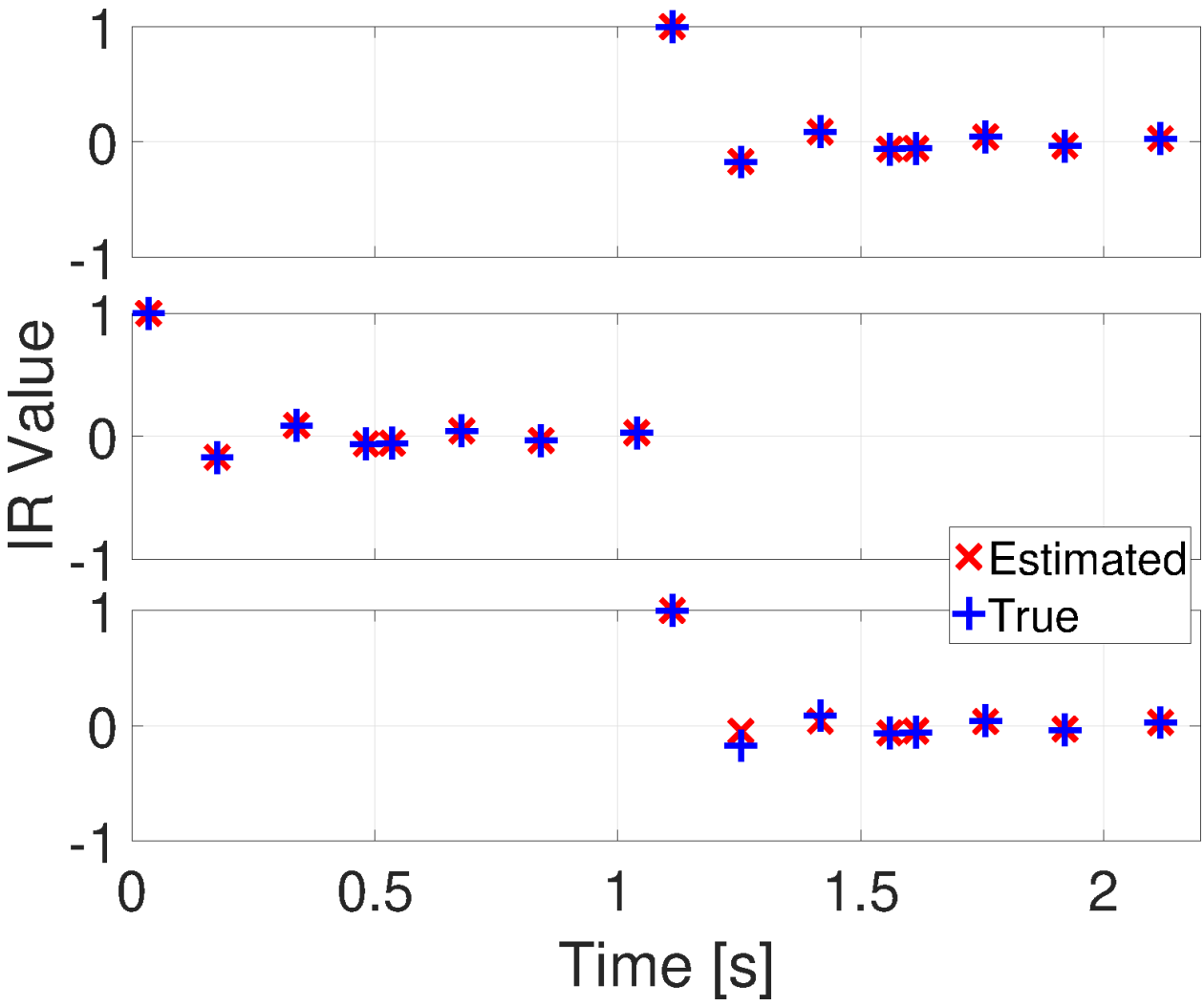

Figure 2.6: Estimated vs. true impulse response channels for the Receiver 1:Source pair. Results for OMP (top), CoSamP (middle) and, NESTA (bottom) are shown. A sparsity level of 30 is used for both OMP and CoSaMP, while the default λ value of 1 is used for NESTA. Only estimated channel values whose amplitude is greater than 0.01 are shown.

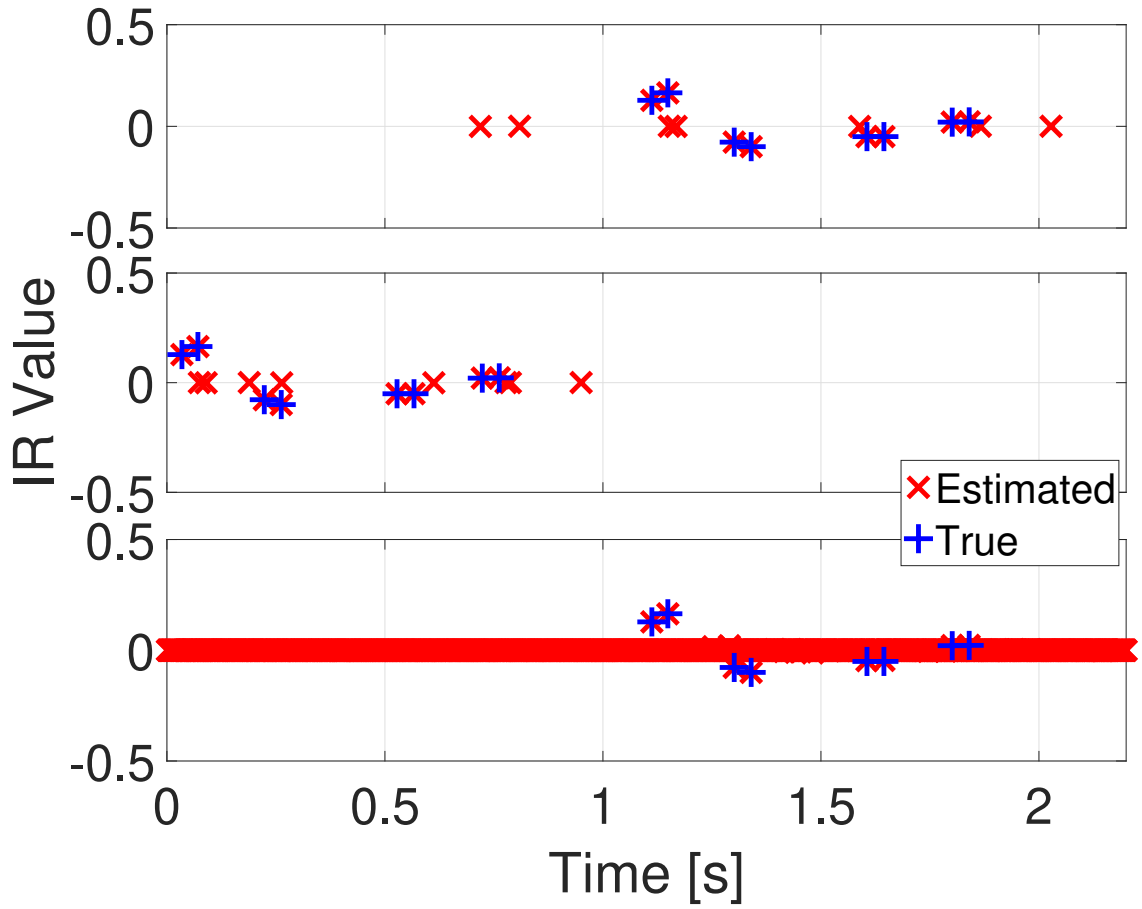

Figure 2.7: Estimated vs. true impulse response channels for the Receiver 2:Source pair. Results for OMP (top), CoSamP (middle) and, NESTA (bottom) are shown. A sparsity level of 30 is used for both OMP and CoSaMP, while the default λ value of 1 is used for NESTA.

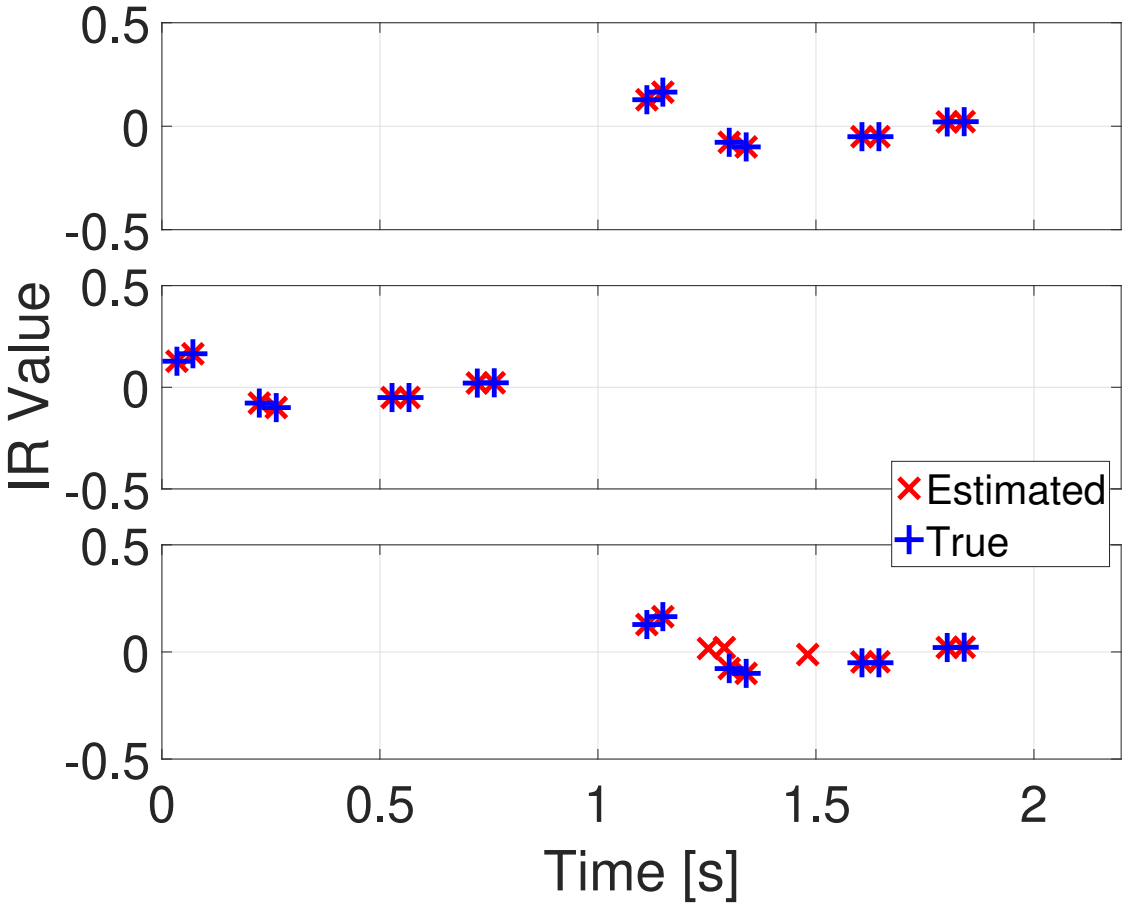

Figure 2.8: Estimated vs. true impulse response channels for the Receiver 2:Source pair. Results for OMP (top), CoSamP (middle) and, NESTA (bottom) are shown. A sparsity level of 30 is used for both OMP and CoSaMP, while the default λ value of 1 is used for NESTA. Only estimated channel values whose amplitude is greater than 0.01 are shown.

Parameter	Value
Peak-to-Peak Wave Height [m]	4
Wave Period [s]	6

Table 2.2: Environmental conditions used only for the time-varying channel simulation

2.3.2 Simulation–Time-Varying Channel

In Zeng et al. [2013], good channel recovery results are described for a one source:two receiver underwater acoustic channel estimation simulation using a selection of different sparse recovery algorithms. A limitation of their simulation is the static nature of the underlying channels. In the ocean acoustic environment, time variability can occur on different time scales. Gradual changes in deep water salinity can introduce deep sound speed profile variability over long time periods. Over shorter time intervals, surface weather conditions, including wind and surface temperature, can change shallow sound speed profile characteristics by changing the depth of the surface mixed layer or creating surface ducts. Additionally, surface waves (in particular, waves where gravity rather than capillary action is the restoring force) may change the effective water depth over the time scale of the wave period, thereby introducing time variability into the water depth and channel impulse responses. In this section, water depth variability through ocean surface gravity waves is introduced to the impulse response calculation to examine the effect of water depth variability on impulse response estimation.

Other than water depth variability, the environmental characteristics for this simulation are the same as those for the static channel simulation in the previous section. Source and receiver locations are the same, source waveform is the same, and the received data SNR is the same for both the static and dynamic channel cases. Surface waves are simulated by adding a sinusoidally-varying offset to the still water level. Table 2.2 lists the ocean surface wave properties used to create the time-varying channel. It is expected that a longer wave period, along with a longer source waveform, would yield similar channel recovery performance. Effectively, each sample in the source waveform sees a slightly different water depth than the sample before it. Rather than precisely modeling a range-dependent surface wave profile, the entire ocean surface is assumed to sinusoidally move up and down with the designated wave amplitude and period. A more nuanced surface elevation model that treats the ocean surface in the same way that some underwater acoustic models incorporate range dependent bathymetry into their calculation of eigenray travel times would yield somewhat different impulse responses, however for the purpose of investigating how surface wave-based channel time variability influences channel recovery, the simpler model was chosen.

Figure 2.9 shows how the locations of the non-zero values in the impulse response between the source and receiver 1 changes over the course of the source waveform. The source waveform is 5 s long, with a 900 Hz sampling frequency, yielding an overall source waveform length of 4500 samples. The maximum impulse response length is 909 samples long (after truncating leading and trailing zeros). Thus, there are 4500 columns and 909 rows in Figure 2.9. Transmission loss, as well as arrival time, for the eigenrays that reflect at least once off the surface also varies over the course of the source transmission due to variation in ray path length.

Figure 2.10 plots the estimated Receiver 1:Source IR for both the static (top) and dynamic (bottom) channel cases and the OMP sparse recovery algorithm. Figure 2.11 plots the results for the Receiver 2:Source.

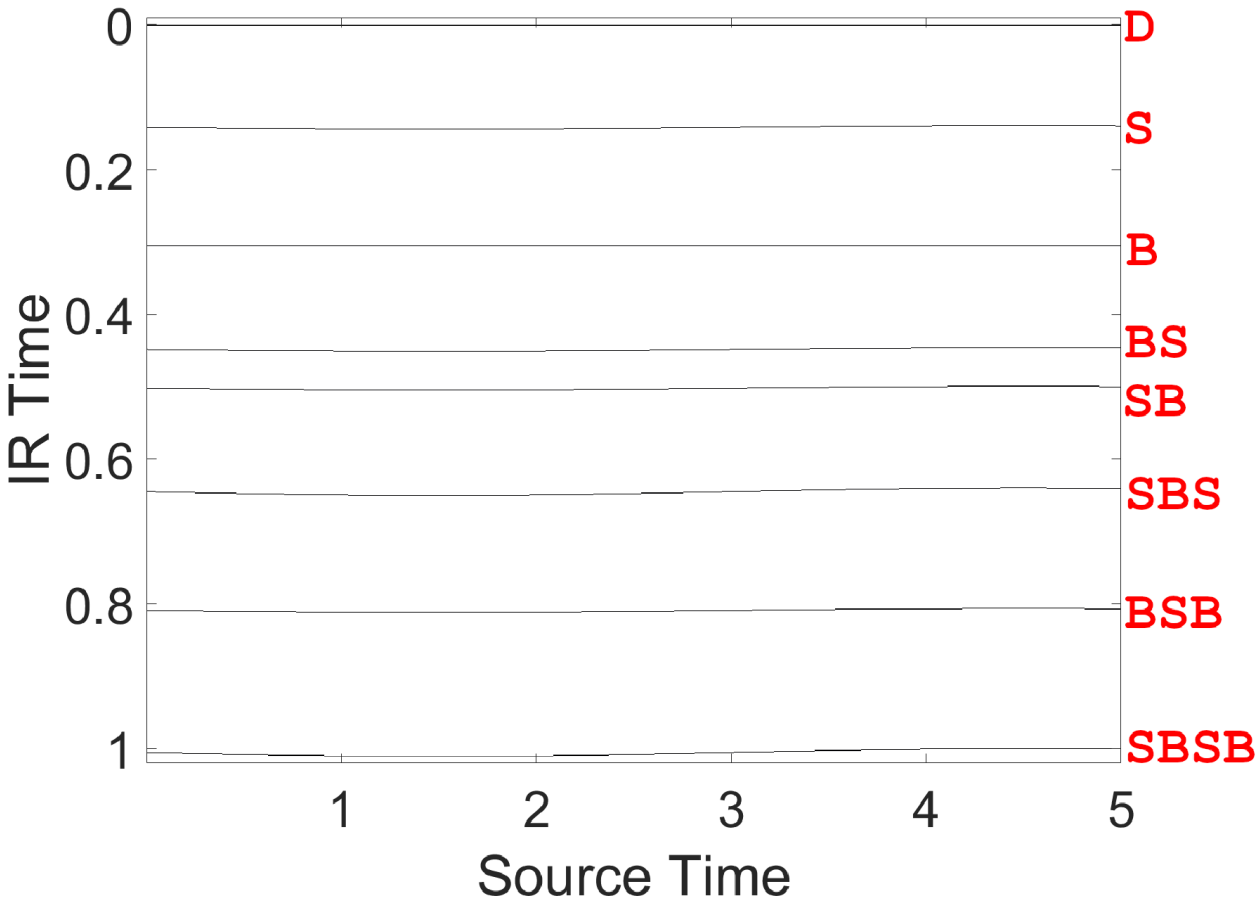

Figure 2.9: Variation in Receiver 1:Source IR non-zero values for each sample in the source waveform. Labels to the right of the plot indicate the eigenray path (D for direct, B for bottom bounce, S, for surface bounce, BS for bottom surface bounce, ...) for each line.

For the static channel case, the true IR is plotted. For the dynamic channel plots, the IR for the first source waveform sample (i.e., the first column in Figure 2.9) is plotted to give an approximate reference for where the non-zero values in the impulse response are expected to be found.

Figure 2.12 plots the estimated Receiver 1:Source IR for both the static (top) and dynamic (bottom) channel cases and the CoSaMP sparse recovery algorithm. Figure 2.13 plots the results for the Receiver 2:Source. For the static channel case, the true IR is plotted. For the dynamic channel plots, the IR for the first source waveform sample (i.e., the first column in Figure 2.9) is plotted to give an approximate reference for where the non-zero values in the impulse response are expected to be found.

Figure 2.14 plots the estimated Receiver 1:Source IR for both the static (top) and dynamic (bottom) channel cases and the NESTA sparse recovery algorithm. Figure 2.15 plots the results for the Receiver 2:Source. For the static channel case, the true IR is plotted. For the dynamic channel plots, the IR for the first source waveform sample (i.e., the first column in Figure 2.9) is plotted to give an approximate reference for where the non-zero values in the impulse response are expected to be found.

Since NESTA is prone to estimating a large number of non-zero values which are orders of magnitude

smaller than the largest values, an amplitude threshold is applied to the results shown in Figures 2.16 and 2.17. In other words, all estimated IR values for both the static and time-varying channel cases whose amplitude is less than 0.01 are not plotted for the results shown in Figures 2.16 and 2.17.

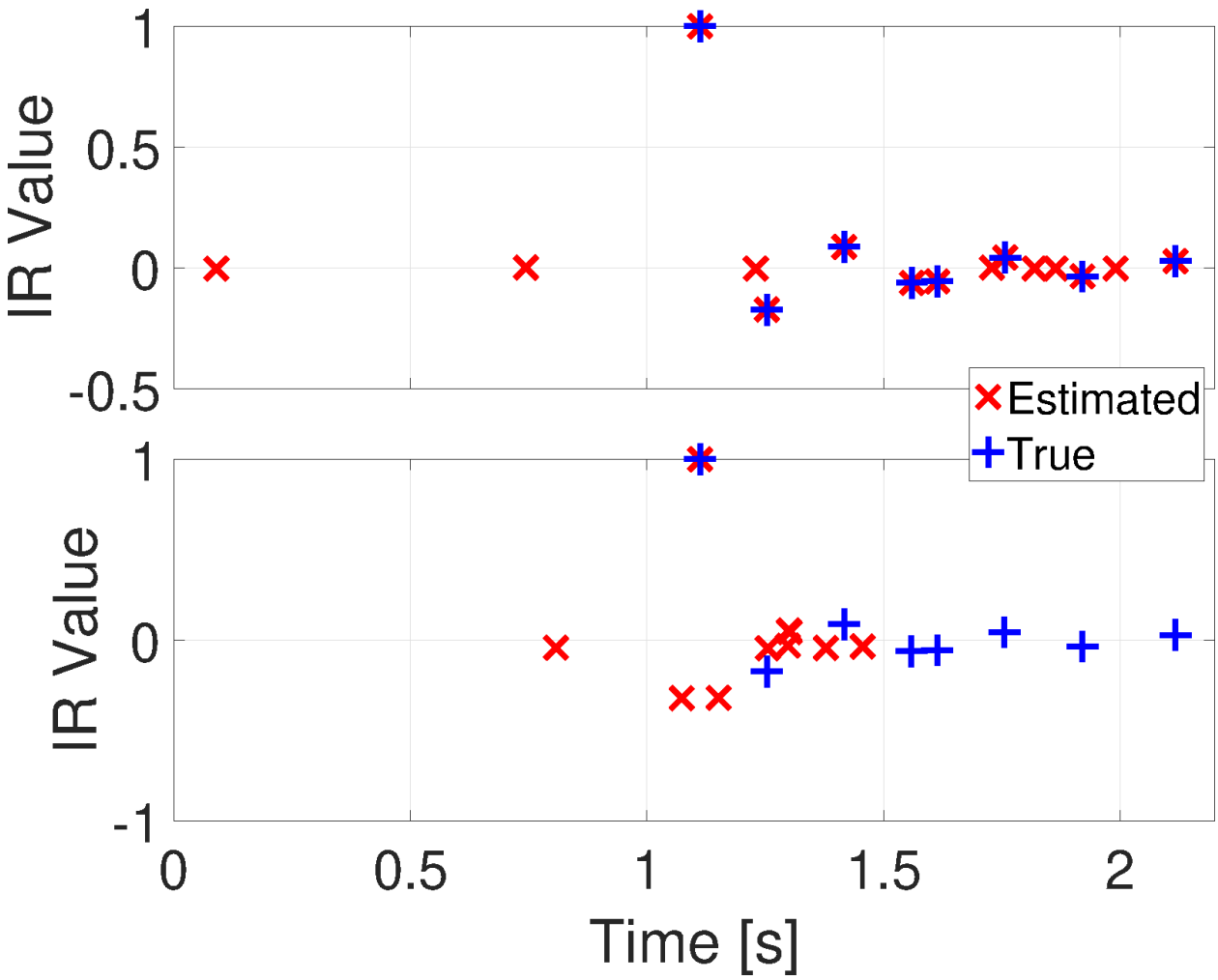

Figure 2.10: Static (top) vs. time-varying (bottom) impulse response estimation results for the Receiver 1:Source pair and the OMP sparse recovery approach. The true static channel is also shown in the top subplot, while the 'true' channel in the bottom plot is the IR seen by the first sample in the source waveform. A sparsity level of 30 is used.

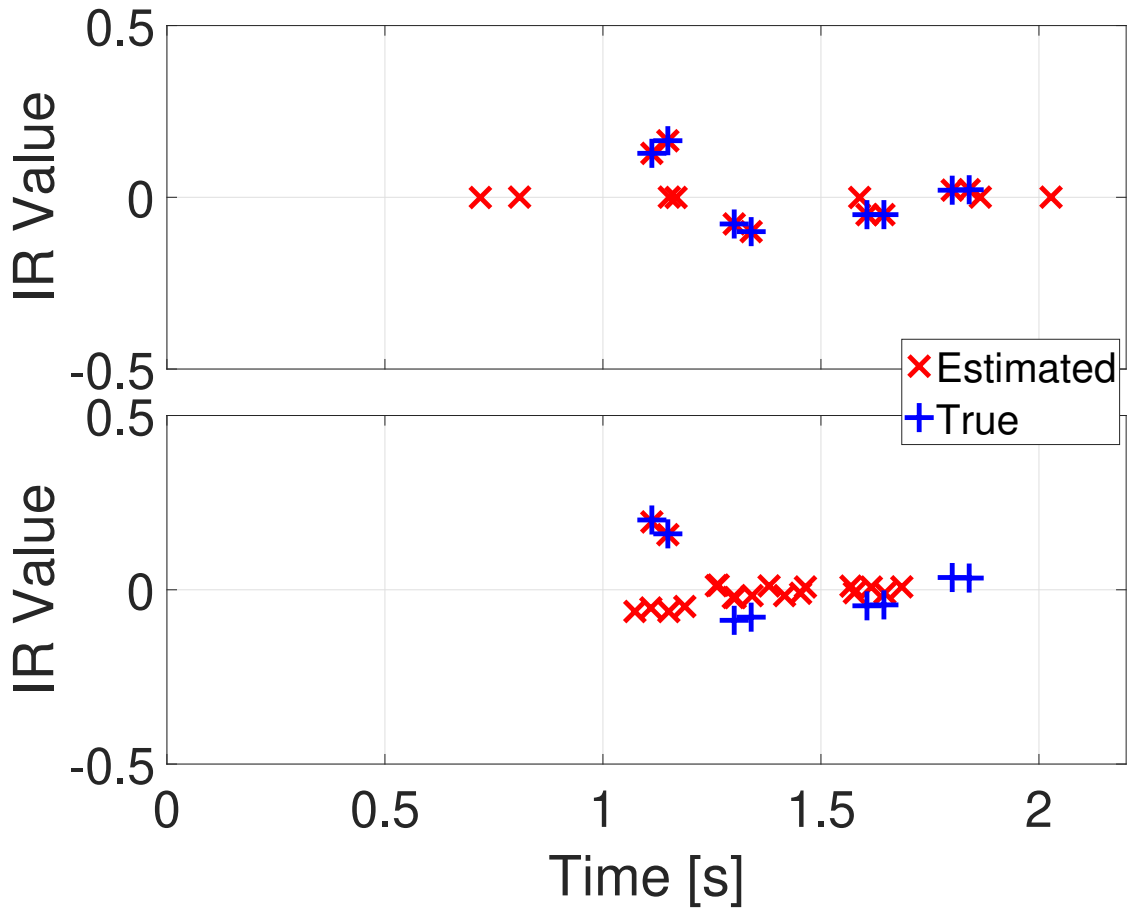

Figure 2.11: Static (top) vs. time-varying (bottom) impulse response estimation results for the Receiver 2:Source pair and the OMP sparse recovery approach. The true static channel is also shown in the top subplot, while the 'true' channel in the bottom plot is the IR seen by the first sample in the source waveform. A sparsity level of 30 is used.

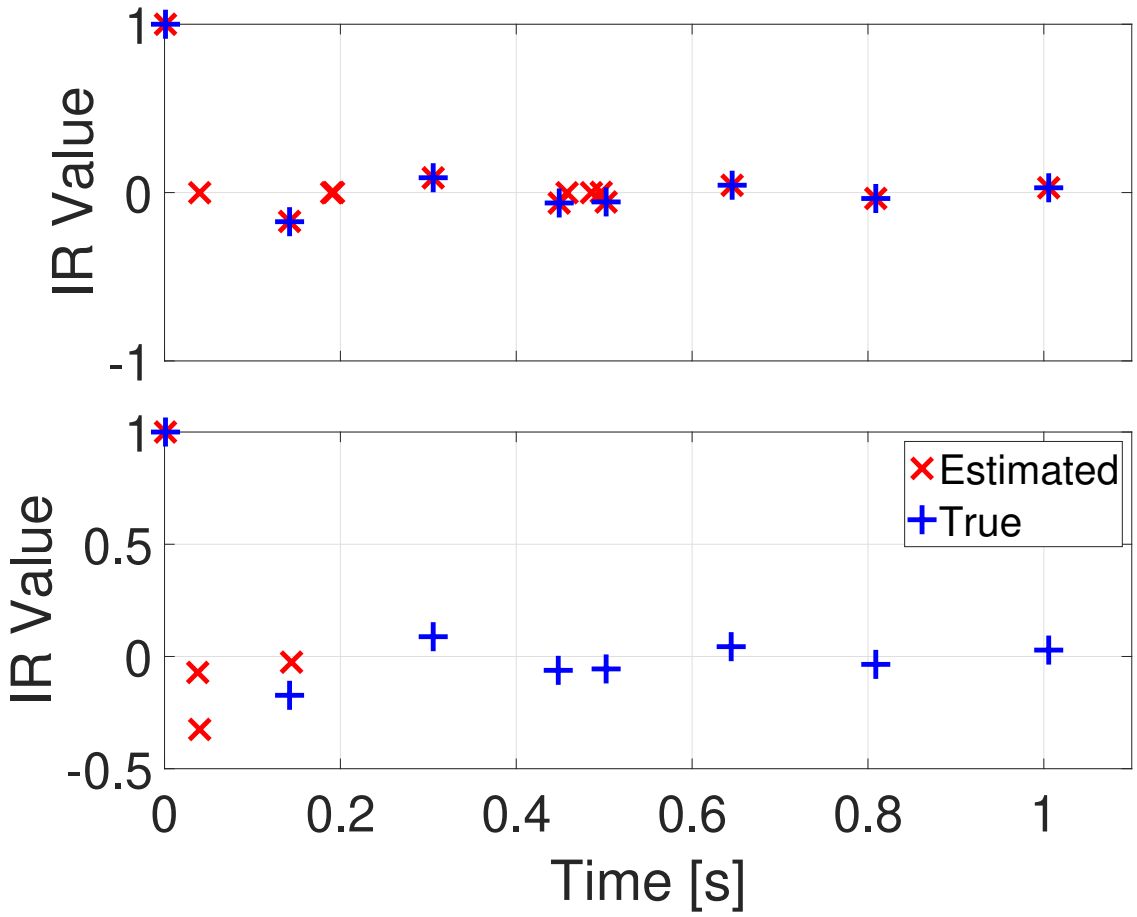

Figure 2.12: Static (top) vs. time-varying (bottom) impulse response estimation results for the Receiver 1:Source pair and the CoSaMP sparse recovery approach. The true static channel is also shown in the top subplot, while the 'true' channel in the bottom plot is the IR seen by the first sample in the source waveform. A sparsity level of 30 is used.

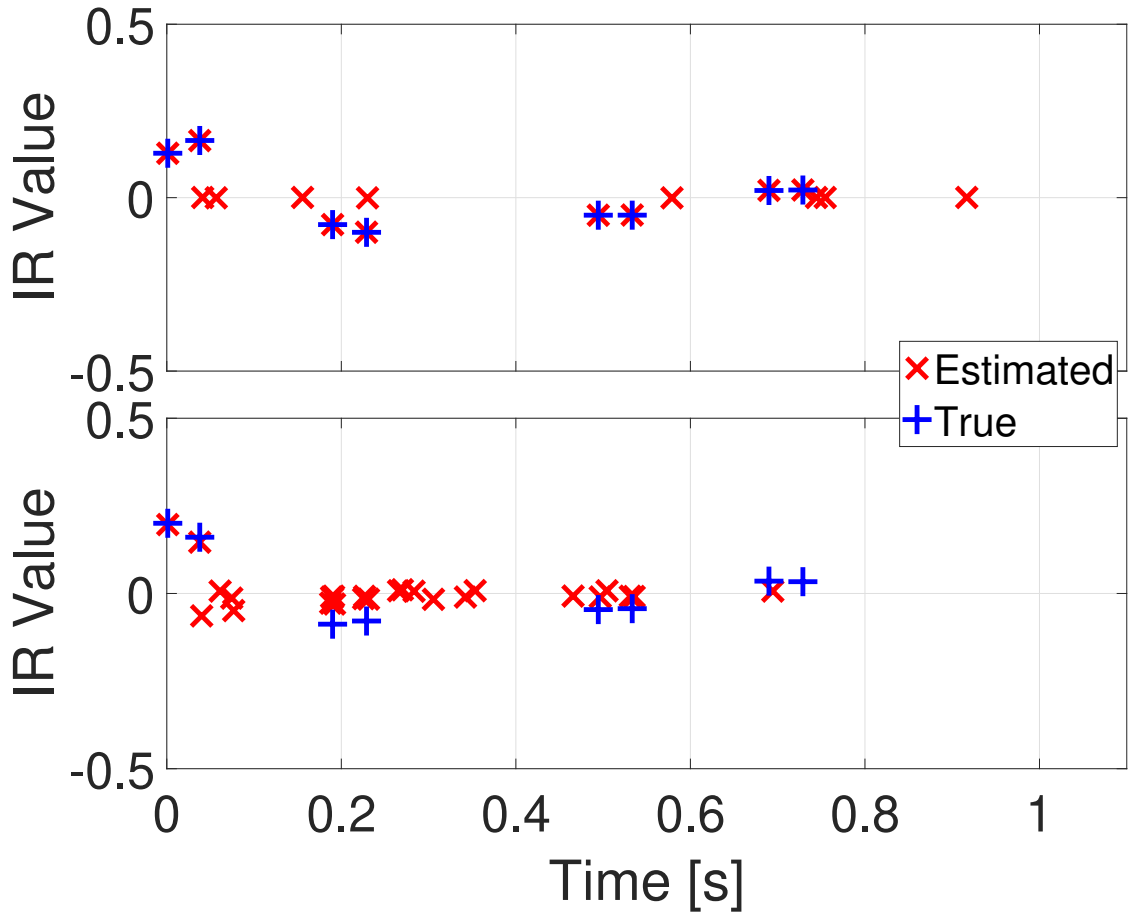

Figure 2.13: Static (top) vs. time-varying (bottom) impulse response estimation results for the Receiver 2:Source pair and the CoSaMP sparse recovery approach. The true static channel is also shown in the top subplot, while the 'true' channel in the bottom plot is the IR seen by the first sample in the source waveform. A sparsity level of 30 is used.

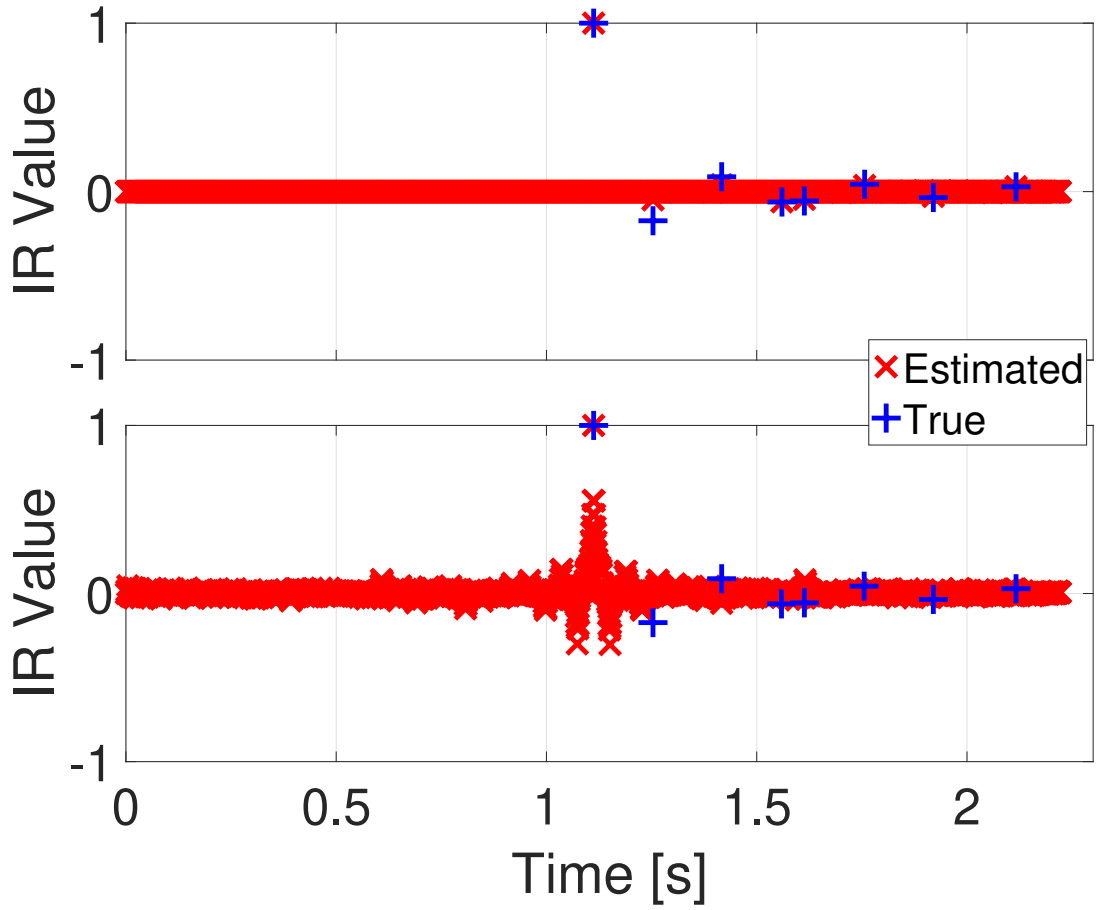

Figure 2.14: Static (top) vs. time-varying (bottom) impulse response estimation results for the Receiver 1:Source pair and the NESTA sparse recovery approach. The true static channel is also shown in the top subplot, while the 'true' channel in the bottom plot is the IR seen by the first sample in the source waveform. A lambda value of 1 is used.

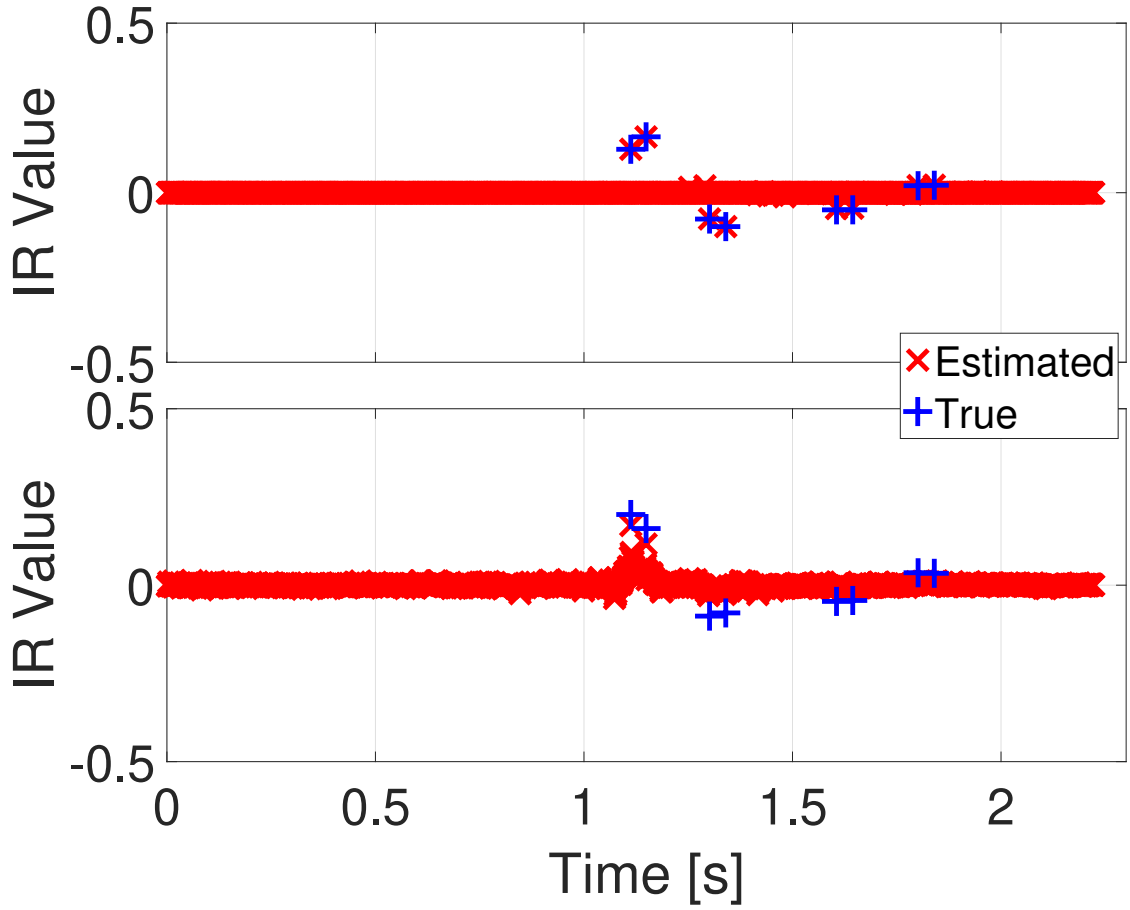

Figure 2.15: Static (top) vs. time-varying (bottom) impulse response estimation results for the Receiver 2:Source pair and the NESTA sparse recovery approach. The true static channel is also shown in the top subplot, while the 'true' channel in the bottom plot is the IR seen by the first sample in the source waveform. A lambda value of 1 is used.

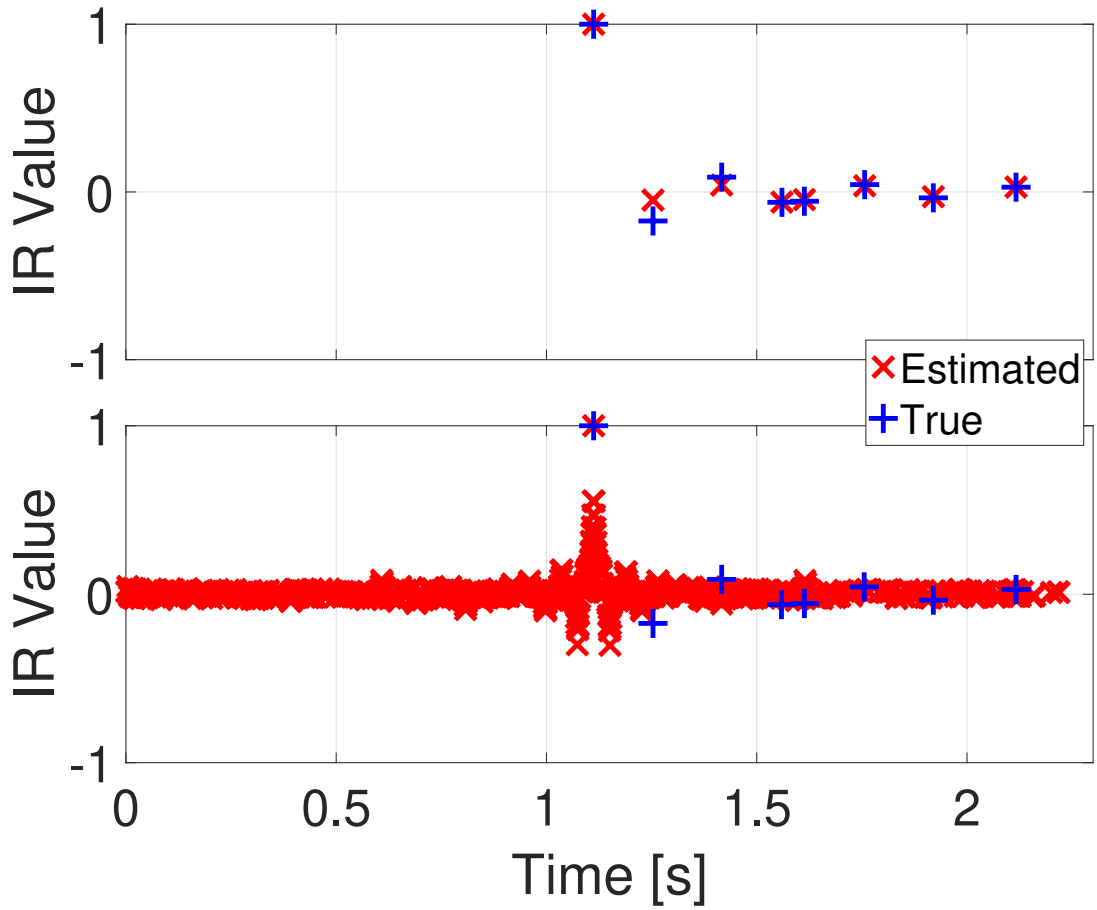

Figure 2.16: Static (top) vs. time-varying (bottom) impulse response estimation results for the Receiver 1:Source pair and the NESTA sparse recovery approach. An amplitude threshold of 0.01 is applied such that only estimated IR values whose amplitude exceeds 0.01 are plotted. The true static channel is also shown in the top subplot, while the 'true' channel in the bottom plot is the IR seen by the first sample in the source waveform. A lambda value of 1 is used.

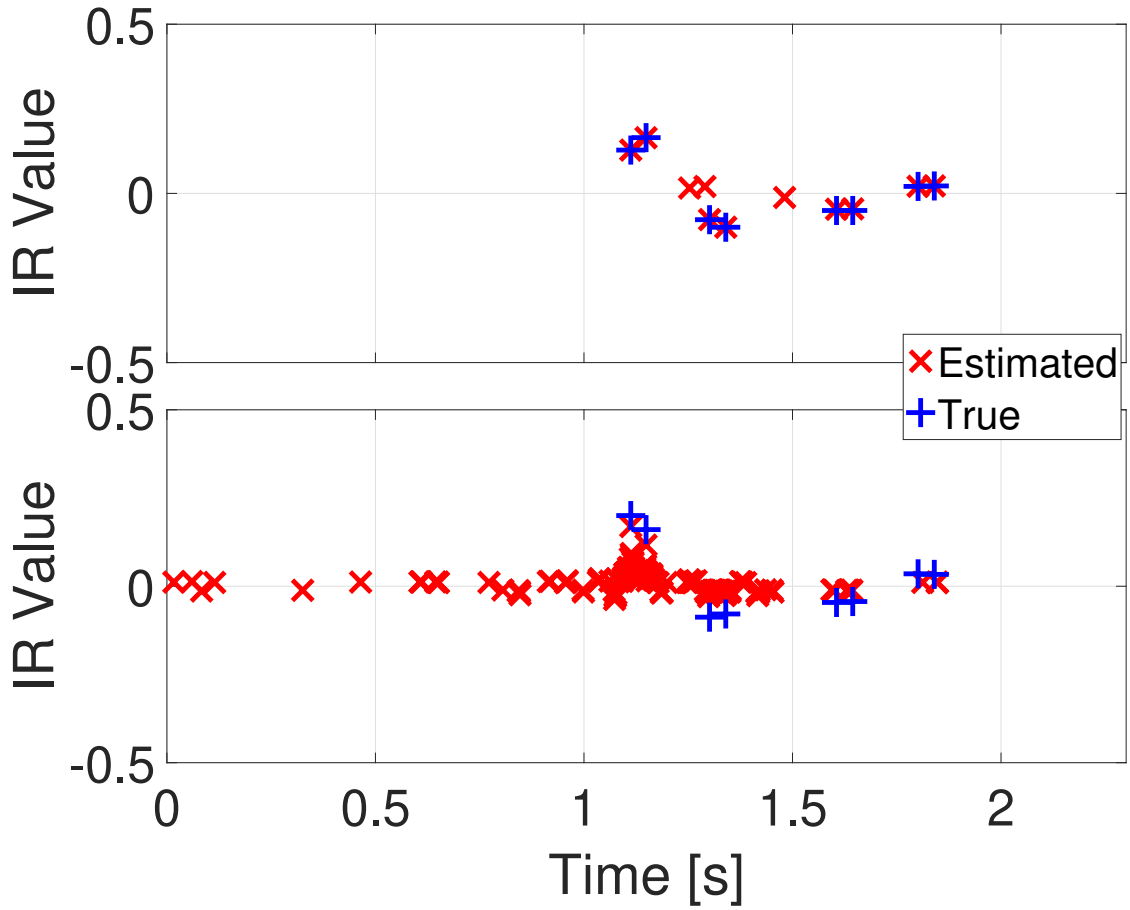

Figure 2.17: Static (top) vs. time-varying (bottom) impulse response estimation results for the Receiver 2:Source pair and the NESTA sparse recovery approach. An amplitude threshold of 0.01 is applied such that only estimated IR values whose amplitude exceeds 0.01 are plotted. The true static channel is also shown in the top subplot, while the 'true' channel in the bottom plot is the IR seen by the first sample in the source waveform. A lambda value of 1 is used.

2.4 Discussion

In Section 2.3, a series of simulation results are shown which demonstrate the performance of three sparse recovery approaches (OMP, CoSaMP, and NESTA) under static and time-varying underwater acoustic conditions. For the static channel cases (Section 2.3.1), the match between estimated and true channels was good based on how OMP, CoSaMP, and NESTA construct their estimated channels. For OMP and CoSaMP, the algorithms are instructed to estimate a fixed number of non-zero values across a set of channels whose length is also defined in advance (i.e., a sparsity level is a tuning parameter for both these approaches). Thus, these two approaches will return a set of estimated channels of the defined sparsity level, regardless of whether all the non-zero values contribute significantly to the solution. For the static channel case, the set of true channels has exactly 16 non-zero values. Since a sparsity level of 30 is selected, OMP and CoSaMP place some small non-zero values in addition to the 16 needed to fit to the simulated data. These spurious non-zero values are smaller than those that match to the true solution. The time offset difference between the OMP and CoSaMP solutions is due to a difference in which sample in the estimated IR is set to zero (in order to resolve the magnitude and phase ambiguity in the BCE solution; the CoSaMP sample is set to be 30 samples into the solution, while the OMP sample is set 1000 samples into the solution).

The NESTA approach also relies on the overall length of the set of channels being defined in advance, and uses a tuning parameter λ to indirectly control the sparsity level of the solution (rather than directly controlling it in the way OMP and CoSaMP do). Before a threshold is applied to the NESTA data (Figures 2.5 and 2.7), the large number of small amplitude non-zero values in the NESTA solutions mask the few significant values. After a 0.01 amplitude threshold is applied (Figures 2.6 and 2.8), the match between true and estimated channels for the NESTA case is visibly improved.

A simulation comparing static and time-varying channels is presented in Section 2.3.2. The same static channel results from the previous section are used, and a surface waves are simulated by adding an oscillation about the still water level which can increase or decrease the path length for a given eigenray path (depending on the phase angle for the surface wave at the time in question). For the 8 eigenray paths considered in this simulation, the location of the non-zero values in the true channel impulse responses changes by as many as 6 samples (with a 900 Hz sampling frequency). The figures in Section 2.3.2 show that the incorporation of a time-varying channel yields channel estimates that are a poor fit to any channel found in the time-varying IR set. This simulation supports the hypothesis that time variability in ocean acoustic channels can make channel recovery using BCE approaches more difficult.

CHAPTER 3
SINGLE HYDROPHONE RANGING IN DEEP WATER

The comprehensive study of marine bioacoustics requires that both data collection (DC) and data analysis (DA) take place so that genuinely useful ideas can be distinguished from plausible-sounding, yet ultimately fruitless, avenues of study. For the purposes of this document, DA is defined as the process of extracting information and insight (e.g., the range of a marine mammal from the hydrophone receiver) from data (e.g., recorded underwater acoustic data from a single hydrophone). DC and DA activities each pose unique challenges, requiring different skill sets to carry out successfully. Doing only one of these two activities well can yield an unusably-poor acoustic data set that even the most robust processing approaches would fail to extract useful information from, or a high quality data set nonetheless of limited bioacoustical quality (e.g., due to inadequate algorithmic support to extract the latent information from the data).

Well implemented DC includes the selection of low-noise acoustic recorders, design of recorder deployment & recovery systems (e.g., heave stabilization for hydrophones lowered over the side of a surface vessel and acoustic release layouts for moored recorder deployments), a comprehensive data management plan (with redundancy) so sufficient and reliable digital storage space is available for the duration of the data collection effort, selection of recorder bit depth (e.g., 16-bit) and sampling frequency (e.g., 96 kHz) so that calls from species of interest are reliably observable, and careful site selection so that prevailing ocean conditions (e.g., nearby container shipping lanes that will raise the local noise floor, acoustic shadow zones which limit the detection range, absence of nearby marine mammals which reduces the size of the data set that will eventually be analyzed) don't significantly reduce the quality or quantity of the recorded data. As well, the use of numerical modeling and simulation under realistic DC conditions helps contextualize DA results derived from measured data and may inform future DC initiatives. Additionally, preliminary DA activities, including modeling and simulation, can inform DC activities and improve outcomes.

Once DC has occurred, DA can begin in earnest. In underwater acoustical studies of marine mammals, it is common to have more data than resources to completely process and analyze it. This is hardly a problem, in the sense that an embarrassment of riches is hardly a problem, but it does pose computational challenges during the DA phase of a project. As with the DC phase of the experiment, care must be taken during the design and implementation of DA algorithms so that they accurately and reliably achieve the objectives of the experiment. For studies involving the detection, classification, and localization (DCL) of signals in noise (whether these signals are of animal or anthropogenic origin), DA algorithm design starts with an understanding of the local acoustic environment and the characteristics of nearby signal and noise sources. This local acoustic picture includes characteristics of the signals being sought after (including signal start and stop frequencies, signal duration and periodicity, source directivity, and source range and depth patterns in signalling behavior), consequences for missed/undetected (i.e., false negatives) and erroneous (i.e., false positive) signals, background noise characteristics (e.g., episodic shipping noise, wave breaking, rain noise, snapping shrimp presence), and the presence of interfering signals (e.g., sei whale calls when the purpose of the study is to investigate fin whale calls). All of these features contribute to the construction of a comprehensive picture of the local acoustic environment which can be used to design DA algorithms of sufficient sophistication to carry out DCL of the signals of interest.

Consider two different underwater passive acoustic localization tasks. In Rideout et al. [2013], the objective of the experiment was to record and locate Pacific walrus (*Odobenus rosmarens divergens*) knock vocalizations near a trio of bottom-moored acoustic recorders deployed in 30 m of water northwest of Alaska in the Chukchi Sea. This environment has very little nearby shipping traffic, and the quantity of recorded acoustic data made it feasible to perform much of the DCL analysis with little in the way of automated algorithms. The recorded walrus knocks were relatively impulsive and of high signal to noise ratio, and the recordings were free of similar conflicting calls, which meant that a simple energy detector (Balanda and MacGillivray [1988]) was sufficient to identify broad time intervals within the several months-long data record that contained probable walrus knocks. Visual analysis by a human operator was then used to precisely identify walrus knocks (within the identified broad time intervals of interest) and to carry out localization. This is an example of a DA situation in which minimal automation is required; walrus knocks from less than a minute of acoustic data were ultimately processed in that study, as the study objective was to demonstrate an existing localization approach in a novel situation rather than process a large quantity of walrus knocks. This DA approach would not be not suitable for a larger data set or more complicated acoustic environment because of the high degree of manual analysis required to generate the localization results; the localization approach in this study was able to extract a wealth of information from the recorded acoustic data at the expense of requiring a great deal of human operator time to arrive at these results.

In contrast, consider the calls shown in the waveform and spectrogram plots in Figure 3.1. This figure shows five minutes of acoustic time series and spectrogram data from the ALOHA cabled observatory north of Hawaii, and has 10 individual fin whale 20 Hz calls (and their associated reflected path arrivals). The energy detector from Balanda and MacGillivray [1988] may not be a sufficiently sophisticated detector for this data set due to the presence of multiple types of calls. The details of the design of the DA approach for this particular situation are covered in detail later in this chapter.

In an era in which continuously recording underwater hydrophone arrays are increasingly common, techniques need to be developed to aid in the interpretation and analysis of the recorded data while reducing (but not eliminating) the dependence on human operators. While time consuming, human-based detection and classification of recorded marine mammal vocalizations is a well-established practice, particularly for validation of computer-generated detection and classification; trained human operators remain the gold standard for these validation efforts, although it is a time-consuming and expensive process.

This dissertation presents an approach for calculating the horizontal range between a source and a single hydrophone. A selection of multiple, and single, hydrophone localization approaches are described in the next section. Techniques only requiring data from a single hydrophone offer a number of key advantages over those requiring data from multiple receivers. Single hydrophone deployments are logistically simpler, and oftentimes less expensive, than multiple hydrophone deployments. As well, data from a single underwater receiver can be used to carry out marine mammal density estimation using the point-transect approach (such as the single seismometer used in Harris et al. [2013]). Single hydrophone approaches are insensitive to timing mismatch between multiple hydrophones, since they do not rely on having precise relative timing information. If the same call can be identified in the data from multiple unsynchronized receivers a single hydrophone range estimation approach may be used to carry out 3D localization (Nosal and Frazer [2006]). If azimuth-dependent bathymetry is available, data from a single hydrophone may be used to estimate 3D position (Tiemann et al. [2006]). Even if only range can be estimated from the single-hydrophone data, there

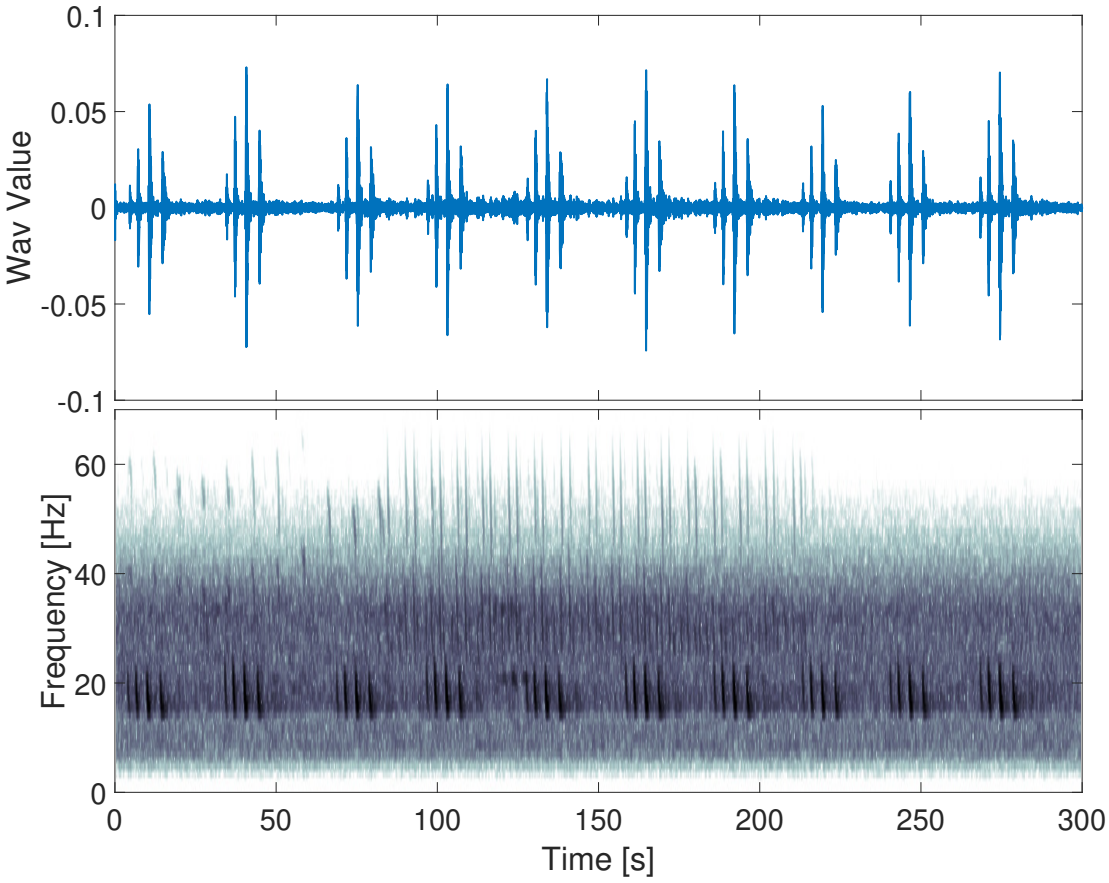

Figure 3.1: Fin whale 20 Hz calls, and other marine mammal calls, recorded between 06h00-06h05 on March 3 2014 at the ALOHA Cabled Observatory. Waveform (top) and spectrogram (bottom) plots are shown.

is still value in carrying out localization. For example, as is described in Section 3.4.7 and demonstrated in Section 4.4.4, performing a 1D localization provides an avenue to reducing the false positive rate beyond what is possible merely from a matched filter by requiring that each direct or reflected arrival of a given call be used no more than once to perform localization. Thus, single-hydrophone localization approaches can offer advantages over methods requiring multiple hydrophones in some situations.

3.1 A Partial History of Passive Acoustic Marine Mammal Localization

Passive acoustic marine mammal localization is defined as the estimation of the (oftentimes underwater) position of a marine mammal through the sensing of their natural sounds (sometimes referred to as calls or vocalizations). It dates back to at least the early 1960s when Walker [1963] carried out underwater passive acoustic underwater localization of what was later (Watkins [1981]) considered to be fin whales using a three-element array of ocean bottom hydrophone recorders deployed off the coast of Massachusetts. In that

study, relative arrival times of the same call at each hydrophone pair were used to estimate the horizontal position of each fin whale around the array. This section will present just a handful of examples of passive acoustic marine mammal localization approaches. For a recent overview of this subject, consult Au and Lammers [2016].

In the early 1970s (Watkins and Schevill [1971] and Watkins and Schevill [1972]), William Watkins and William Schevill (working at the Woods Hole Oceanographic Institution in Massachusetts) devised an approach to estimate underwater marine mammal position (either full three-dimensional position for nearby sources or bearing for more distant ones) using a floating, non-rigid array of four hydrophones. The premise behind their approach is called time difference of arrival, and other marine mammal localization approaches have subsequently used this approach. In short, this approach uses the difference in arrival time of a single source at multiple receivers to gauge the position of this source. As the source moves relative to the array, the arrival time differences change accordingly. With a minimum of four receivers, the three-dimensional position of the source can be calculated. Sounds from a variety of manmade (e.g., pingers and echosounders) and animal sources (right whales, fin whales, and white-beaked dolphins) were located. Their equipment and approach was later applied to estimating the bearing to diving sperm whales (Watkins and Schevill [1977]).

Spiesberger and Fristrup [1990] described a comprehensive overview of passive acoustic localization of both aquatic and terrestrial animals that also aims to map the local wind and sound-speed fields. Characteristics of acoustic propagation are discussed, the premise of using arrival time differences to estimate the position of a calling animal is presented (whereby the arrival time differences in a hydrophone pair describes a hyperboloid in three-dimensional space on which the sound source must lie and the intersection of multiple hyperboloids further constrains the position). An acoustic tomographic technique for estimating sound speed fluctuations and receiver positions using recorded animal calls is discussed, which leads into the framing of the passive acoustic localization problem as a geophysical inverse problem.

An approach similar to that described in Watkins and Schevill [1972] is used in Stafford et al. [1998] as part of a detection and localization approach for monitoring blue whales (*Balaenoptera musculus*) in the northeast Pacific ocean. Ocean-bottom hydrophones, rather than the floating hydrophones used by Watkins and Schevill [1972], within the U.S. Navy's SOund Surveillance System (SOSUS), recorded blue whale calls in August - September 1993. A set of calls was manually identified in the recordings from this period, and mathematical representations of the average of the strongest components of the calls in this set were calculated. A matched filter was made from the mathematical representations, and applied to recordings from a broader set of receivers in the SOSUS network to help associate the same call on multiple receivers. These associated calls were located by using the arrival time of the same call on multiple receivers and using a nonlinear least-squares approach to estimate source time, latitude, and longitude. This localization approach was validated using data collected from sonobouy deployments by U.S. Navy maritime patrol aircraft that used relative arrival times of the same call at multiple sonobouys. The SOSUS-based locations were found to be consistent with the sonobouy-based locations.

3.2 Single Hydrophone Localization

The complexities and costs associated with collecting acoustic data at sea have led to the development of localization approaches that require a minimum amount of gear be placed in the water. Battery pack

failure, pressure housing leaks, and acoustic release malfunctions are only some of the fates that can befall underwater acoustic recorders deployed for even hours at a time. Some techniques for passive acoustic localization require knowledge of the relative positions of the receivers in order to provide reliable position estimates (e.g., Walker [1963], Watkins and Schevill [1972], Stafford et al. [1998]). Techniques for estimating receiver relative positions, once deployment has been carried out, exist based on producing sound at known locations and using an approach similar to passive acoustic localization to estimate receiver positions using known source positions (Dosso and Sotirin [1999]), but could be difficult to implement in all situations (e.g., in deep water where known position sound sources would have to be loud and expensive in order to produce sufficient sound to be heard by sea bottom receivers). Calculating time synchronization between individual underwater acoustic recorders can also be challenging, particularly for widely-spaced arrays. Thus, passive acoustic localization approaches that only require data from a single receiver can have useful advantages over those that require multiple receivers.

Frazer and Pecholcs [1990] described an early single-hydrophone localization approach which builds upon the single-hydrophone approach in Clay [1987]. The localization approach in Clay [1987] requires that the source signal be known, in addition to the received signal, and amounts to finding a candidate source location that produces a model-predicted received waveform sufficiently similar to the measured received waveform. Frazer and Pecholcs [1990] described four new localization approaches which do not require that the source waveform be known and that make use of the relationship between the received signal and the impulse response of the acoustic waveguide in question. The performance of these approaches, however, are improved if the source signal is known *a priori*.

Jesus et al. [1998] presents a single hydrophone approach whereby the range to a 500 Hz bandwidth signal 5 km from the receiver in 130 m of water is estimated using a multipath-delay maximum-likelihood estimator. The impulse response which characterizes the environment under investigation is first estimated from the measured data, under the assumption that the signals traveling along different eigenray paths experience uncorrelated perturbations. If the number of non-zero samples in the impulse response is known *a priori*, the location and amplitudes of these impulse response samples can be estimated. To estimate horizontal range, the peaks in the estimated impulse response are taken to be representative of the eigenray arrival times in the measured data, and are compared against model-predicted eigenray arrival times calculated using the BELLHOP ray tracing acoustic propagation model (Porter and Liu [1994]). The source range and source depth combination yielding the best match between the estimated impulse response and the model-predicted eigenray arrival times is the estimated source location. A drawback of this approach lies in how to select the number of model-predicted arrival paths to calculate. In theory, this can be estimated from the measured data, but in practice the calculation of the optimal number of model-predicted eigenray paths can be problematic.

In 1999, McDonald and Fox [1999] used data from one of the same hydrophone installations used by Thompson and Friedl [1982] to estimate the range to calling fin whales north of O'ahu and, in turn, estimate a lower bound on fin whale population density around this site using what is called a point transect density estimation approach. Range estimation was accomplished through a combination of transmission loss and relative arrival time (RAT) modeling, although the authors found that the required effort of the relative arrival time method was significantly higher than the transmission loss-based approach. In this context, RATs are the time difference between the direct-path arrival time and each subsequent higher-order arrival

path (e.g., the time difference between the bottom-surface bounce and the direct path). McDonald and Fox [1999] also found that when the local bathymetry about the hydrophone had significant range or azimuth dependence, the RAT-based approach could allow horizontal position to be estimated (rather than just range from the hydrophone) but that bathymetric uncertainty could also introduce problems with the range estimation process. Aside for validating the transmission loss-based ranging results, the authors made limited use of the RAT-based approach. The transmission loss-based approach used the parabolic equations acoustic propagation model RAM using the environmental characteristics for the data collection site, the hydrophone depth, and an estimated fin whale calling depth of 50 m (Watkins et al. [1987]). By estimating that a fin whale passing directly overhead the hydrophone would produce a recorded SNR of 30 dB, the spherical spreading transmission loss predictions were transformed into expected fin whale call SNR as a function of horizontal range and the distance to a fin whale call could be made by measuring the SNR of the signal. Potential sources of error in this approach include variable background noise levels (e.g., due to weather at the survey site) and variation in fin whale call source level; as the background noise level increases, and/or the source level decreases, the animals estimated position will be biased toward longer ranges.

In Aubauer et al. [2000] impulsive vocalizations of Spinner dolphins (*Stenella longirostris*) are used to estimate the 2-D range (from the receiver) and depth for the calling animal. To do this, RATs are calculated by subtracting the bottom-bounce and surface-bounce arrival times from the direct-path arrival time. Closed form expressions for source range and source depth, and error estimates for these quantities, are presented using the calculated RATs, receiver depth, water depth, the depth-invariant sound speed in the water column, and error bounds on these four types of data.

A single hydrophone technique for estimating the pitch angle of a calling sperm whale (*Physeter macrocephalus*) is described in Laplanche et al. [2005]. As the pitch angle of a sperm whale changes (e.g., as it transitions from breathing at the surface to diving), the authors expect that the received level of the various reflected path arrivals will change. Thus, they expect that by comparing the received level of the direct and reflected path arrivals the pitch angle of the sperm whale can be estimated. The distance between the whale and its prey is also estimated using the interpulse interval (i.e., the time spacing between consecutive sperm whale echolocation clicks).

In 2006, Tiemann et al. [2006] described a technique whereby underwater sound sources (in this case sperm whales in the Gulf of Alaska) could be located solely using data from a single hydrophone and a knowledge of the nearby bathymetry. This approach uses Relative Arrival Times (RATs) to estimate sound source position. RATs consist of the difference in arrival time between the earliest arrival of a given call and subsequent, higher-order reflection paths at a single receiver. This phenomenon will be described in greater detail in Section 3.4, but in short RATs are the set of times resulting from the subtraction of the direct-path arrival time from all subsequent higher-order arrival times of the same source. These RATs change as the animal moves through the environment. In Tiemann et al. [2006], acoustic modeling is used to predict relative arrival times at a set of ranges from, and along a discrete number of bearing lines around, the hydrophone. Sperm whale clicks were identified in the recorded data, and RATs for each click estimated. These measured RATs are compared to the set of model-predicted RATs and the match found, indicating the estimated sound source location.

In 2012, Mouy et al. [2012] described an approach to estimate the range and depth of a calling Pacific walrus (*Odobenus rormarus divergens*). Their approach is similar to the method of Tiemann et al. [2006].

First, walrus knocks (which are impulsive vocalizations) were detected in the recorded data using a kurtosis criteria (Balanda and MacGillivray [1988]) that is sensitive to sudden changes in signal level. Next, arrival times for a fixed number of direct and reflected-path ray paths at each range and depth grid point were calculated using a ray-tracing acoustic propagation model. The Teager-Kaiser energy was then calculated for the detected walrus knocks to enhance the higher order reflected path arrivals in the recorded acoustic signal. Finally, the resulting enhanced signal is compared to the model-predicted signal at each candidate range and depth grid point. The grid point whose model-predicted signal best matches the measured signal is deemed the best-fit location for the calling walrus.

Duan et al. [2014] describes a single-hydrophone approach to estimating the depth, initial range, and speed of a moving source using the relative arrival time between the direct and surface-bounce rays. This approach is related to an extended Kalman filter, and was demonstrated using a towed synthetic source recorded by a bottom-moored (300 m above the bottom) hydrophone deployed in water approximately 4500 m deep. The synthetic source broadcast a repeating, 10 s long pseudorandom sequence between 700-800 Hz.

A holistic approach to carrying out passive detection and localization of fin whales using a single ocean bottom seismometer (OBS) is described in Matias and Harris [2015]. The authors used a modified matched filter approach to automatically detect fin whale calls in their OBS data; a high SNR call from their data set was used as the reference waveform, and the cross-correlation expression was modified to penalize waveforms that differed in amplitude. This was done to help distinguish direct-path from reflected-path arrivals, as their technique for estimating range yields spurious results when processing non-direct-path arrivals. Rejection of reflected-path arrivals is further aided by a windowing approach whereby localization is only carried out on the strongest cross-correlation peak in a given time period. The three components of the recorded OBS data for each direct-path fin whale call are used to estimate the azimuth and incident angle of arrival of the call at the sea bed, and is useful for calls within a defined range from the OBS. This range is calculated using the ratio of the p-wave speed in water (α_1) and in the ocean-bottom sediment (α_2), and for the authors study case fell in the $4.88 - 6.49$ km range.

In Weirathmueller et al. [2017], another fin whale localization approach is described, this one aimed at estimating the range to calling animals in the vicinity of an ocean bottom seismometer (approximately 5 km for this study). A matched filter is used to identify fin whale calls in the recorded data. Earthquakes are eliminated from the measured data set by looking for detections with more energy outside (specifically in the 5-15 Hz band) than within the fin whale calling frequency band (15-25 Hz). Detections were also forced to be at least 10 s apart to reduce the likelihood of multipath arrivals being included in the detection set. Ranging is carried out using both the RATs and relative amplitudes between the direct and first two reflected-path arrivals. The vertical channel of the OBS is ultimately used for ranging, and so a geoacoustic propagation model is used in conjunction with the BELLHOP ray tracing acoustic propagation model to predict how the different ray paths will be expressed on the vertical OBS channel as a function of source range. Next, for each matched filter detection, the relative arrival times and amplitudes for each multipath are calculated. This approach for ranging a fin whale call relies on being able to identify or assume the path for two or more rays (i.e., direct and reflected paths) for the call in question. At each candidate source range, the model-predicted RATs and relative arrival amplitudes are compared with the measured values. The candidate range yielding the best fit to the RATs and relative arrival amplitudes is deemed to be the estimated fin whale range. This range is only kept if there are at least 6 other calls within 1 km and 6

minutes of this call; this leverages the repetitive, almost metronomic, regularity of typical fin whale calling bouts and helps weed out false positives.

In Weirathmueller et al. [2017], the cost function a normalized timing residual between measured and model-predicted acoustic arrival times in ocean bottom seismometer data is calculated, as depicted in Equation 3.1,

$$t_{res}(x) = \Big(\sum_{i=1}^{n} \Big[\frac{O_i - P_i}{\sigma_i} \Big]^2 \Big)^{1/2}, \tag{3.1}$$

where i is the index of the arrival (e.g., direct, bottom-surface, etc.), O_i is the i^{th} measured travel time, P_i is the i^{th} model-predicted travel time, and σ_i is the travel time uncertainty for the i^{th} arrival.

3.2.1 Single Hydrophone Localization - Dispersion

In addition to the arrival time-based approaches discussed above, there are other single hydrophone range estimation methods that rely on a dispersive environment and the method of normal modes (Frisk [1994]). In a normal modes decomposition of an acoustic signal, there is a frequency-dependent transmission time between source and receiver whereby different frequencies in the source waveform are observed to reach the receiver at predictable yet different times. Different vibrational modes are frequently observed, resulting in the same frequency potentially expressing multiple distinct arrival times in the received signal. These times are related to the physical phenomenon whereby, given the source and receiver locations and other environmental parameters, there are times at which upward and downward-traveling acoustic energy at the receiver constructively interfere to yield a strong received signal, while at other times the phase difference between these upward and downward acoustic energies yields destructive interference (and thus, has a comparatively lower signal to noise ratio).

The KRAKEN normal modes propagation model (Porter [1992]) is an example of a "forward model" that, given source and receiver information and environmental parameters, can be used to calculate transmission loss for an acoustically-dispersive environment. Since the characteristics of these modes is partially dependent on the range between source and receiver, it is possible to develop an "inverse model" whereby the observed modal behavior in the received signal is used to predict the range that must have existed between source and receiver to produce such a modal observed signal. In this section, range estimation approaches that leverage a dispersive environment are described. The deep ocean conditions described later in this dissertation do not display modal behavior in any significant way, so these dispersive range estimation approaches are presented here solely in the interest of completeness.

In Wiggins et al. [2004], an approach for estimating range from the modal dispersion pattern in the data from a single hydrophone is described. North Pacific right whales are the topic of this study. To estimate range, modal distortion curves at a set of candidate source ranges are calculated using a synthetic undistorted call and a propagation model. The propagation model is used to calculate dispersion curves as a function of frequency for each mode expected in the measured data. A comparison between measured and model-predicted dispersion-distorted received waveforms can allow for an inverse approach to estimate range if other aspects of the environment are sufficiently well known. In addition to a qualitative comparison between these two sets of curves allowing for a manual estimation of range, the authors propose a more automated approach whereby the time difference of arrival of the same frequency in different modes could

be calculated for both the measured and model-predicted data. The candidate source range whose modal time differences of arrival best match the measured time differences would therefore be the estimated source range.

An approach to characterizing a shallow (< 400 m) waveguide using an impulsive low frequency (< 200 Hz) source recorded at a single hydrophone is proposed in Bonnel et al. [2009]. An adaptive signal processing approach termed "frequency warping" is described whereby knowledge of the environment is leveraged to transform a dispersion-distorted received signal back into a Dirac delta function in the time domain. They show that environmental mismatch (i.e., the modeled environment being a poor match to the actual one) will result in the frequency warped signal not resembling an impulsive source signal. Given this environmental mismatch sensitivity, the authors expect that frequency warping may be used to estimate environmental parameter values using the resemblance of the warped signal to a delta function. These parameters that could be estimated via such an inversion approach include the source range.

In Bonnel et al. [2010], the objective of the proposed inversion is to estimate the time synchronization between a source and receiver, assuming that the distance between them is known. Since modes overlap both in time and frequency when range is small, the authors propose a warping signal processing approach to isolate the waveform of a single mode in isolation from the others. Once the individual modes are isolated, subsequent processing is simpler. Warping, in the time domain, is also used in Thode and Bonnel [2015] to estimate vertical array tilt and modal shapes.

Two different range estimation approaches are described in Munger et al. [2011]: a time difference of arrival (TDOA) approach using acoustic recordings at three or more hydrophones, and a normal mode dispersion model (NMDM) approach using data from a single hydrophone. The NMDM approach is considered appropriate when the propagation distance is several times the water depth, and when the ratio of the water depth to the wavelength is not large (i.e., 4:1 or lower). The TDOA approach estimates the arrival time of the same call on multiple hydrophones and uses a nonlinear least-squares parameter estimation approach to estimate the best-fit horizontal sound source position. In contrast, the NMDM is an iterative approach that uses data from a single hydrophone, and builds up on work presented in Wiggins et al. [2004]. First, arrival times of the same frequency across multiple modes are manually identified in the measured data. Model-predicted group velocities at this frequency are then calculated. Arrival times and modal arrival times are used to estimate the horizontal range between source and receiver. This source range is used in a normal modes code to produce a model-predicted modal dispersion pattern based on a synthetic model of the initial call. The fit between measured and model-predicted is used to guide manual adjustments to the range and synthetic call contour parameters until the fit is deemed adequate.

In contrast to modal-based inversion methods that seek to compare measured and model-predicted modal dispersion curves (such as in Wiggins et al. [2004]), Bonnel et al. [2012] presents a method called "modal reversal" that aims to leverage environmental knowledge to undo the dispersion present in the received data. Modal reversal also does not require prior knowledge of the source signal. If the correct environmental parameters are used, the modally-reversed modes will be in phase and add up constructively. Environmental mismatch will yield out of phase reversed modes. In practice, this process is carried out on each mode individually by isolating the modes one at a time, performing the modal reversal, and summing over the set of modally-reversed modes. Maximizing the similarity between the reversed modes allows geoacoustic inversion to take place. This approach was found to be precise for high signal to noise ratios, but performed

more poorly at lower SNR values.

Bonnel et al. [2013] presents a single hydrophone Bayesian inversion approach that assumes an impulsive source, unlike Bonnel et al. [2012] which doesn't require prior source knowledge. This approach uses warping, similar to the approach described in Bonnel et al. [2010], to isolate individual modes that (particularly at close range) can overlap in both the time and frequency domains. Once the dispersion curves in the measured data are estimated via warping, a Bayesian inversion is carried out using a hybrid optimization approach known as adaptive simplex simulated annealing, using the extracted modal dispersion curves as the input data. Model-predicted dispersion curves are calculated during the inversion, and ultimately estimates for source range and time, sediment speed/density/depth, and basement speed and density (along with uncertainty bounds on these parameters) are calculated. Warner et al. [2015] builds upon this approach to carry out a trans-dimensional inversion for sound speed profile characteristics and ocean bottom sediment layer structure using a close range warping transformation of modal dispersion data.

In Thode et al. [2012], an automated approach for the detection, classification, and localization of bowhead whale calls in the presence of seismic airgun sounds is presented. They present an image processing and neural network-based data processing sequence to estimate bowhead whale locations in the presence of (sometimes dispersive) seismic airgun sounds. While their approach requires that multiple receivers record the same sound, they note that modal dispersion can complicate signal detection due to performance degradation of phase-based detection approaches (including matched filtering) observed in the presence of dispersion.

3.3 Data Collection Site Description

The ALOHA cabled observatory (ACO, Duennebier et al. [2008], Howe et al. [2011]) is located at (22.738772 N, 158.006186W), approximately 115 KM north of O'ahu. ACO is a collection of ocean-bottom oceanographic data collection equipment in 4728 m of water that began operation in February 2007. Its current data collection modalities include pressure, temperature, salinity, oxygen, current velocity, chlorophyll, turbidity, and acoustic. Data are recorded by hardware at the ACO and relayed back to O'ahu using the repurposed AT&T HAW-4 fibre optic cable. Data are archived on shore, and available for researchers using an online portal. Figure 3.2 shows an underwater picture of the ACO data collection site, while the topographic and bathymetric map in Figure 3.3 highlights the ACO location north of O'ahu. Also shown in Figure 3.3 is the expected radius to which sources can be ranged based on the methods developed in this dissertation (based on the range that bottom-surface-bottom-surface (BS2) arrivals are predicted to still reach the ACO hydrophone from).

Acoustic data at the ACO were recorded during two different time periods. An initial proof-of-concept recording period ran from February 18 2007 - October 22 2008. Recording resumed June 6 2011 and continued at least up to the present day (Summer 2022). Figure 3.4 shows a sample recording from the ACO hydrophone showing four water-borne (rather than sediment-borne) acoustic arrivals of a single air gun shot. The horizontal range from ACO to the air gun was approximately 12.2 km. The distinct ray pattern for this shot, as evidenced by the time separation between the water-borne arrivals, suggests that a ray-based localization approach could be used to estimate the range to this air gun, and potentially other sounds recorded by ACO; the deep water around ACO could make lower frequency calls (such as the 20 Hz calls of fin whales) reach the hydrophone as distinct ray arrivals as well.

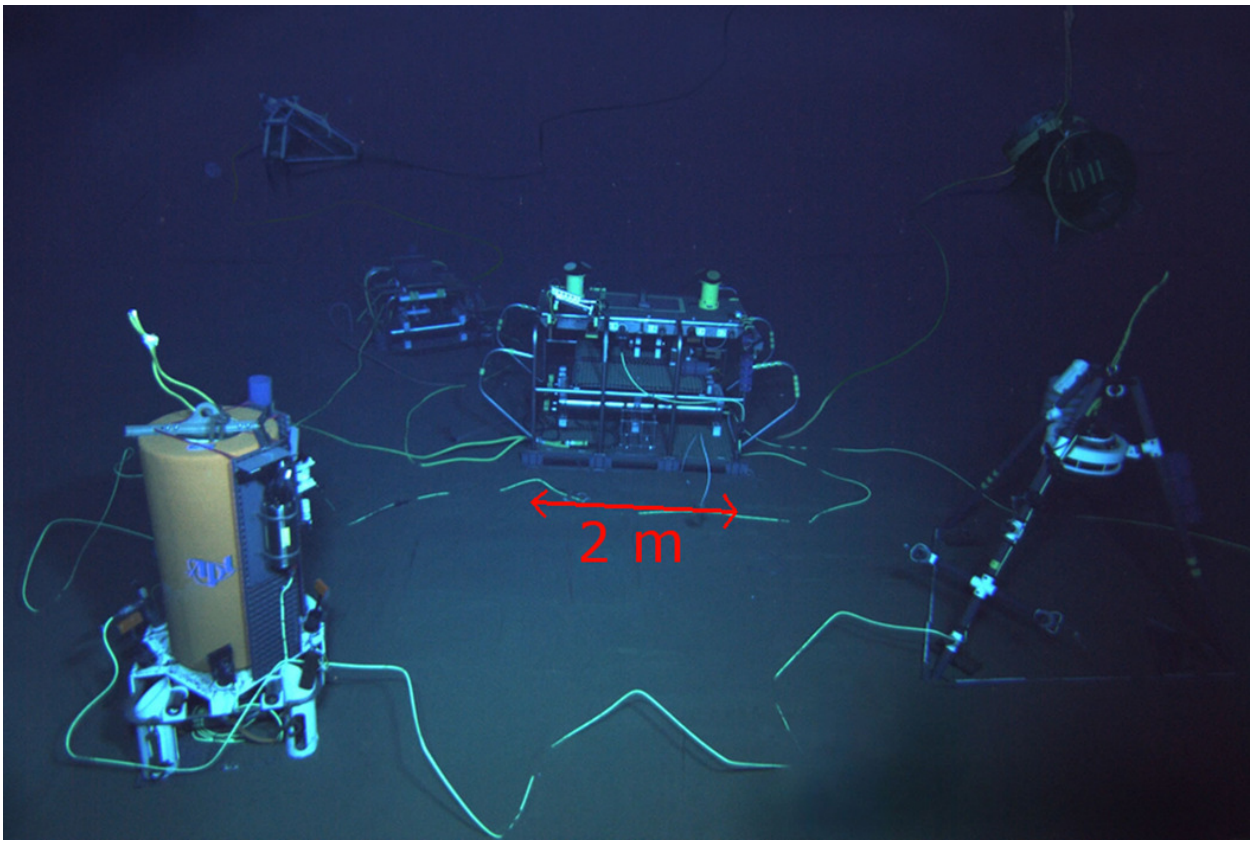

Figure 3.2: Underwater photograph of the ALOHA cabled observatory. The width of one of the data collection platforms is included to provide a sense of scale. Photo source: ALOHA Cabled Observatory [https://aco-ssds.soest.hawaii.edu]

The ACO is well suited for studying the low-frequency calls of mysticetes. The ACO is not close to major shipping lanes, which contributes to the low ambient acoustic background sound levels observed at the site. The fibre optic cable linking ACO to land allows for the continuous recording of acoustic data without the expense and uncertainty of a battery and on-board storage-based experimental setup that requires regular servicing from a surface vessel. Such a servicing plan for ACO would have been particularly complicated given the large water depth at the site.

Another advantage of ACO is that the large water depth facilitates long-distance low-frequency acoustic propagation. A localization approach utilizing direct and interface-reflected arrival times requires that the sound traveling along different reflection paths (e.g., direct, bottom-surface bounce, etc.) between source and receiver arrive at the receiver at distinguishably-different times. As the source range (i.e., the horizontal distance between source and receiver) increases, this distinguishably-different criteria is less likely to hold. The primary reason for this is that, as source range increases, refraction has an increasingly significant impact on underwater acoustic propagation; Snell's Law and the depth-dependent sound speed profile found in all but the shallowest ocean environments causes underwater sound to travel along curving paths that curve toward minima in the sound speed profile. A common mathematical model for describing this sort of acoustic propagation regime is ray tracing (e.g., the acoustic ray tracing propagation model BELLHOP).

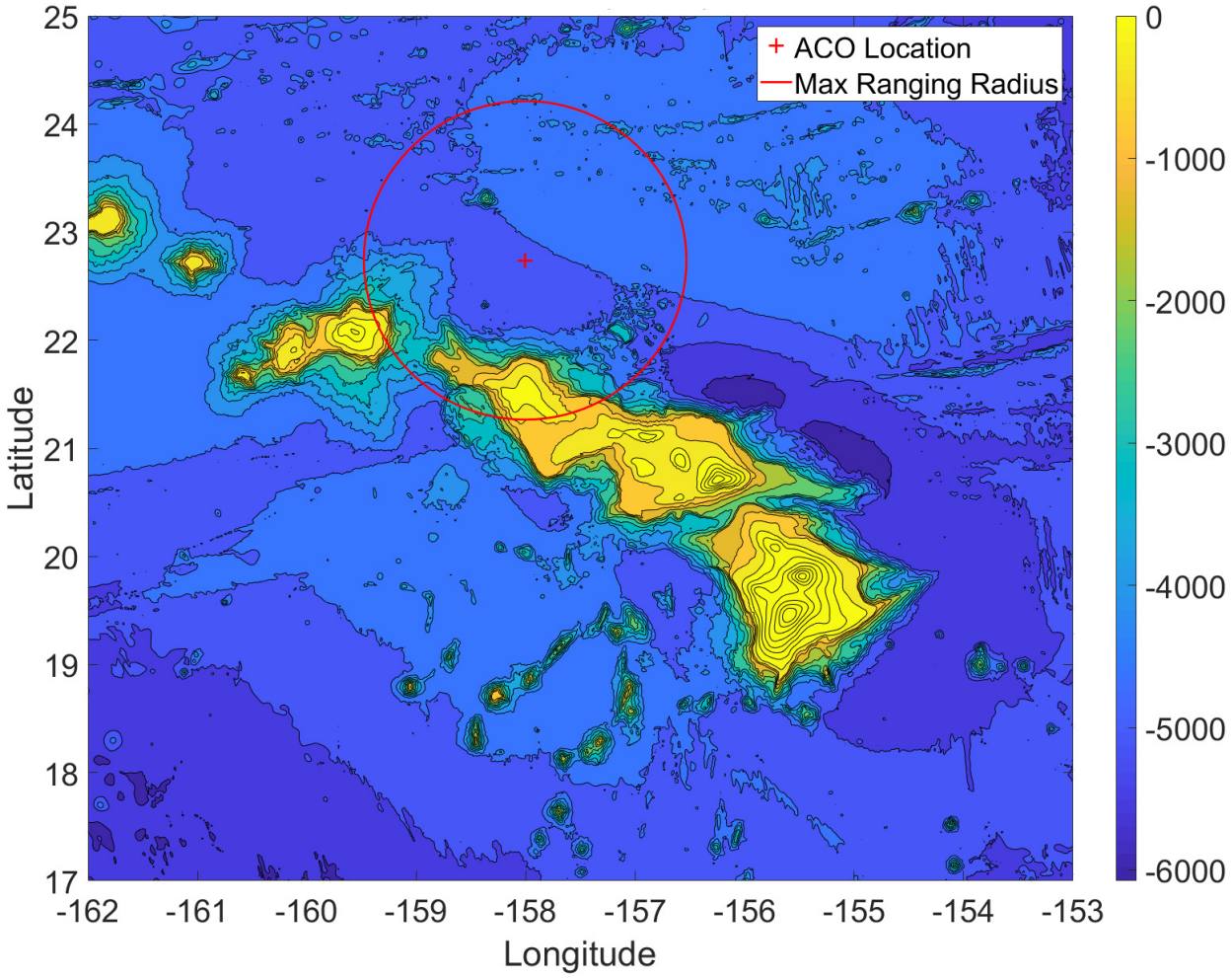

Figure 3.3: Bathymetry and topology near the Hawaiian islands. The ACO location is highlighted by the red + symbol, while the red circle highlights distance from ACO that BS2 arrivals are predicted to reach the ACO hydrophone from (other than in the landfall directions). Bathymetric and topographic data: Hawaii Mapping Research Group [http://www.soest.hawaii.edu/HMRG/cms/]

This refraction phenomenon eliminates some acoustic transmission paths as source range increases (typically starting with the transmission path directly linking source and receiver without reflecting off the surface and/or bottom). As source range continues to increase, refraction causes more ray paths to cease connecting between source and receiver. Eventually, at a sufficiently long range, the sounds reaching the receiver have reflected so many times that a ray tracing model is no longer the most appropriate approach to describe the acoustic propagation conditions, and other computational approaches (e.g., a normal modes propagation model such as KRAKEN) yield more reliable results. Thus, the ACO site is well-suited to collecting acoustic data in support of a RAT-based localization approach because the deep-water conditions create a large ocean footprint over which acoustic arrival times are distinguishably different-enough to permit reliable localization.

Another related aspect of the ACO that makes it well suited for this study is that the deep water lends itself well to studying low frequency mysticetes. The ray tracing approximation to the Helmholtz equation

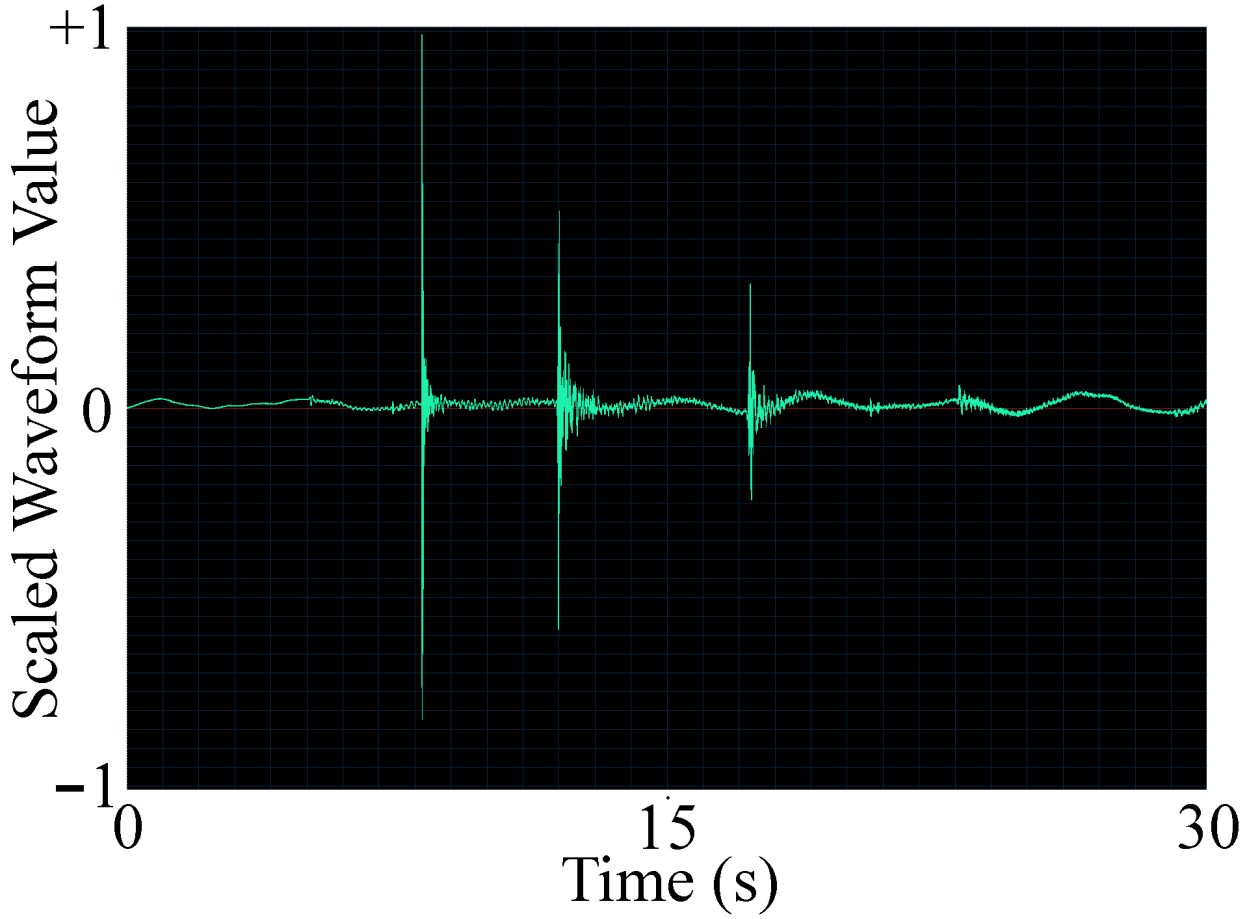

Figure 3.4: Direct and higher order arrival paths of a single airgun shot as recorded by the ACO hydrophone on October 6 2018 at 12h01:35.

is more applicable to low frequency calls as water depth increases, which makes the use of ray tracing-based acoustic propagation models more appropriate than they would be in shallower water. Additionally, the deep water conditions allow for long distance acoustic propagation given the reduced bottom interaction which should have the effect of increasing the range at which underwater sounds are detectable by the hydrophone at ACO. Also, the preference for deep ocean conditions by fin whales makes it more likely they will be recorded at ACO than at a shallower site.

The local bathymetry at the ACO is also advantageous. Figure 3.3 shows the bathymetry in the vicinity of the ACO. Water depth increases rapidly when traveling north from O'ahu, resulting in mid-ocean water depths relatively close to shore. As will be discussed in a later section, the relative absence of continental shelf north of O'ahu (compared to, for example, the typical continental shelf width on the west coast of the US mainland) simplifies, and reduces the computational burden, of carrying out localization using ACO data.

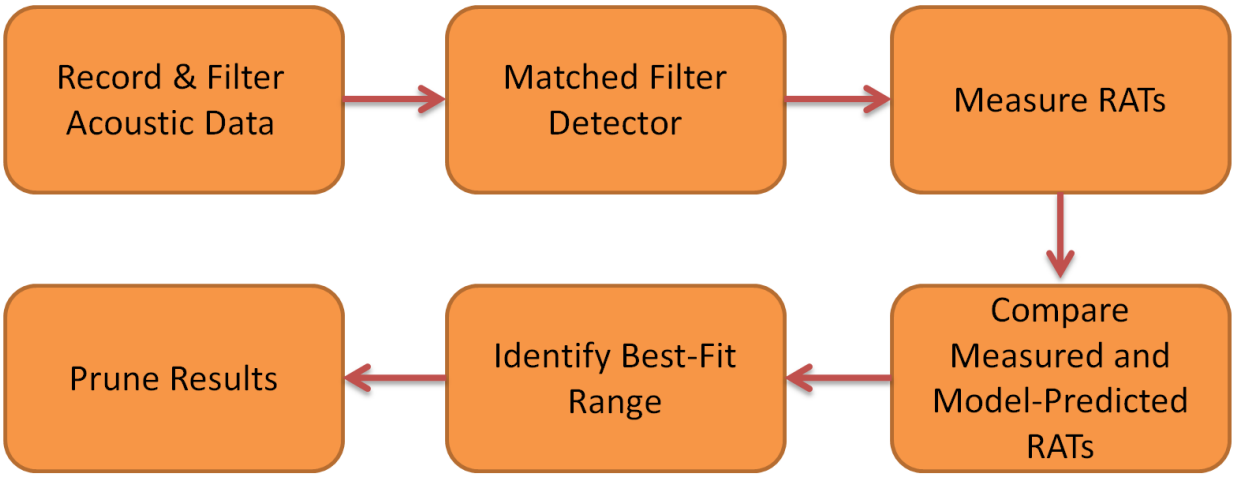

Figure 3.5: Data processing flow chart for the localization approach developed and used in this work.

3.4 Localization Algorithm Description

This section presents a detailed description of the proposed localization approach. The section begins with a high level overview of the data processing chain that begins with the collection of the underwater acoustic data and ends with the calculation of best-fit locations for each of the detected signals. A graphic depiction of the flow of information through our localization approach is presented in Figure 3.5.

After data collection, the first step in the localization approach is initial data processing. This includes bandpass filtering the acoustic data to include only the frequency band of interest and reducing the sampling frequency (through a proper downsampling approach that prevents aliasing from occurring). The bandpass filtering step increases the SNR of the recorded data by filtering out spurious signals (e.g., vocalizations from species not part of the current study and nearby anthropogenic noise sources). It also reduces the computational load of the subsequent data processing steps.

The bandpass filtered and downsampled acoustic data is subsequently processed using a computer program that detects the presence of marine mammal calls of interest. This type of program is typically called an auto-detector, since the objective is for almost all the work to be handled by the computer program with a degree of supervision by a human operator. For this work, the auto-detector is a matched filter with a reference waveform derived from an analysis of a set of recorded vocalizations. Effectively, the purpose of this detection phase is to identify times in the recorded acoustic data which are likely to contain fin whale calls.

Two important concepts related to auto-detection are false positives and false negatives. No auto-detection algorithm is perfect; in addition to (hopefully) detecting the presence of the signal of interest (in our case, fin whale 20 Hz calls), the detector will erroneously identify background noise, or other calls, as being the signal of interest (i.e., false positive), and will fail to identify the presence of some instances of the signal of interest (i.e., false negative). The false positive and false negative rates for a given detector are often inversely proportional and related through tuning parameters on the detector itself; relaxing the tuning parameters to make the detector more permissive (i.e., reduce the false negative rate) will typically

increase the false positive rate as more marginally-appropriate signals are deemed by the auto-detector to be signals of interest.

Following the detection of likely fin whale calls in the acoustic data, the information that will be used to estimate the location of each detection is extracted from the acoustic data. The deep ocean conditions at ACO do not lend themselves to creating modal dispersion that has previously been used to carry out low frequency single hydrophone localization (as described in Section 3.2). Thus, a different type of acoustic information is used to carry out localization. Due to the high acoustic impedance contrasts at the air:water and water:ocean bottom interfaces that bound the ocean acoustic waveguide, the sound from the fin whale 20 Hz call reaching the hydrophone may have reflected one or more times off the ocean surface and bottom, in addition to the acoustic energy that travels directly from the whale to the hydrophone (as long as the whale is close enough to the hydrophone for the direct path to not be so strongly refracted that it no longer reaches the hydrophone). If the conjugate depth at ACO is smaller than the water depth, these fin whale calls may be detectable at the surface at long range due to a convergence zone effect. However, since our hydrophone is near the bottom this effect should not be relevant.

The position of the whale relative to the hydrophone is expressed in the relative arrival times (RATs) of the acoustic energy reaching the hydrophone; as the whale moves away from the hydrophone at a constant depth, the RATs decrease (e.g., the difference in arrival time between the arrival of the direct-path acoustic energy and that reflecting once off the surface and bottom before reaching the hydrophone) even as the travel time for each arrival increases. Figure 3.6 was used in Chapter 2 to depict impulse response variation as a function of source location variability. Here, Figure 3.6 depicts RAT variation as source position changes to illustrate how knowledge of relative arrival times for an underwater sound may assist in estimating its position. As the source range increases (plots on the same row in Figure 3.6), the RATs decrease because the individual ray path lengths become more similar as source range increases. As the source depth increases, a different type of change is seen in the RATs; specifically, as source depth increases (plots in the same column in Figure 3.6), rays that bounce off the surface arrive later because they have further to travel as the source withdraws from the surface. These changes in RATs occur in predictable ways that are dependent on source and receiver positions, local bathymetry, and the sound speed profile (i.e., the depth and range-dependent phase speed of acoustic energy in the water). This predictability forms the basis of the inverse problem and localization approach proposed in this dissertation, in which source position as a function of RATs is estimated (this is the inverse of the forward modeling problem, in which RATs are calculated from source position).

The localization process presented herein is essentially a comparison between features of the received acoustic signal from each detected fin whale call and the expected, model-predicted call features at a set of candidate ranges that the animal could have been present at when it made the call. This comparison is carried out within a cost function that is described in Section 3.4.1. The candidate range whose model-predicted call features best match those of the measured call is deemed to be the best estimate for call range. More specifically, the measured RATs for each detected 20 Hz fin whale call are compared to model-predicted RATs at each candidate range. Thus, before localization can take place, these model-predicted RATs must be calculated using an underwater acoustic propagation model.

Once model-predicted RATs at each candidate range have been calculated, the comparison between each measured and model-predicted set of RATs can be made. The comparison specifically consists of calculating

the scaled sum of squared differences between corresponding measured and model-predicted RATs. This calculation yields a scalar that expresses how well the measured RATs match to the model-predicted RATs.

In subsequent sections of this chapter, each of these individual steps will be described in greater detail.

3.4.1 Cost Function Description

At the core of the localization approach proposed in this dissertation is the cost function that quantifies how well a set of measured data matches to a set of model-predicted data. The output of the cost function is an ambiguity vector with one ambiguity value for each candidate source range being searched over. The global maximum of the ambiguity vector is the estimated range for the sound in question. In other words, the location whose model-predicted data yields the best match to the measured data is deemed to be the estimated location for the sound source. In this work, the measured data are the estimated relative arrival times (RATs) for a detected sound, and the model-predicted data are the calculated RATs based on an acoustic propagation model simulating the oceanographic conditions for the study site.

The two primary contributions of our cost function involve the our treatment of false positives and different numbers of model-predicted arrival paths as range increases. Specifically, our cost function was designed to allow for different numbers of model-predicted RATs as a function of range (i.e., as refraction causes lower order arrivals to no longer connect source and receiver as range increases). The cost function also allows for false positives in the set of measured RATs due to the expectation that an automated RAT estimation code will yield some non-zero number of estimated RATs that do not correspond to an acoustic arrival from the source of interest (e.g., an estimated RAT due to background noise or calls from conspecifics in the acoustic data).

Our cost function can be thought of in the context of a residual error distribution in which errors are assumed to be Gaussian distributed. These errors are introduced due to the measurement process, mismatch between the true and assumed model characteristics, and the data processing steps. We introduce "ad-hoc" elements (i.e. a maximization over sets term and a normalization term) to the likelihood surfaces for position found elsewhere in the model-based localization field (e.g., Nosal and Frazer [2006], Rideout et al. [2013], Warner et al. [2017]). These "ad-hoc" elements make our cost fuction robust to false detections and multiple simultaneous sources, but they mean that the cost function is no longer a true likelihood surface and cannot be used for rigorous error analysis. However, our cost function formulation is reminiscent of the form used in these earlier works to retain a framework that might allow a more complete treatment of error in future work (i.e., it is reminiscent of a likelihood function for position):

$$C(r) = \max_{K(r)} \left[\prod_{j=1}^{N(r)} e^{D(r)} \right], \text{where} \tag{3.2}$$

$$D(r) = \left(\frac{-1}{2\sigma^2 N(r)} \right) \left(\Delta t_{K(r),j} - \Delta \hat{t}_j(r) \right)^2 \tag{3.3}$$

As we do not attempt error analysis in this dissertation, an equivalent form of the cost function, which may be more familiar to some readers, is obtained by taking the log of the argument inside the max of our

cost function to obtain a weighted least-squares version, Equation 3.4.

$$C_2(r) = \min_{K(r)} \left[\left(\frac{1}{2\sigma^2 N(r)} \right) \sum_{j=1}^{N(r)} \left(\Delta t_{K(r),j} - \Delta \hat{t}_j(r) \right)^2 \right]. \tag{3.4}$$

The role of the normalization term, $N(r)$, is more obvious in this weighted least-squares form: The normalization term, $N(r)$ in Equation 3.4 effectively serves to calculate the average squared residual at a given candidate source range. This is done to make cost function values calculated with different numbers of model-predicted RATs directly comparable. The $N(r)$ in 3.2 serves the same purpose. Maximizing $C(r)$ is equivalent to minimizing $C_2(r)$.

In these equations, r is the candidate source range whose cost ($C(r)$ or $C_2(r)$) is being evaluated, j is the number of the model-predicted RAT currently being compared against the set of model-predicted RATs, $N(r)$ is the number of model-predicted RATs that make up the RAT set at the current value of r, $K(r)$ is the set of valid combinations of measured RATs at range r, σ^2 is the RAT variance, $\Delta t_{K(r),j}$ is the j^{th} measured RAT in the $K(r)^{th}$ combination of RATs, and $\Delta \hat{t}_j(r)$ is the j^{th} model-predicted RAT at current range r. In this context, "valid" for $K(r)$ indicates that the cost function will search over the sets of combinations of measured RATs of the same size(s) as the number of model-predicted RATs. In this work, given that the model-predicted RATs have 4, 3, or 2 RATs (as range increases), the set $K(r)$ will have every combination of 4, 3, or 2 measured RATs.

For a perfect fit between measured and model-predicted RATs, $C(r)$ will have a value of 1. In all other cases, $0 < C(r) < 1$. In order not to bias the cost function toward longer range solutions (which have fewer model-predicted RATs, and thus encompass the product of fewer exponential functions in Equation 3.2 than a shorter candidate source range may have), $N(r)$ is included in the denominator of Equation 3.2. In this way, if the best fit set of measured RATs matches to each model-predicted RAT to the same degree (e.g., 0.01 s) at two different ranges, $C(r)$ at these two ranges will be the same.

The output of Equation 3.2, $C(r)$, is a scalar value that represents the best match between a subset of the set of measured RATs for the current detection and the model-predicted RAT set at each candidate range. A key difference between the cost function formulation in Equation 3.2 and previous approaches is the explicit and effective treatment of false peaks (i.e., those corresponding to noise rather than acoustic arrivals) in the measured data and changing numbers of model-predicted RATs as a function of range. The details of how the measured and model-predicted RATs are calculated is described later in this chapter (Sections 3.4.5 and 3.4.6). However, suffice it for now to say that the cost function was designed to allow for different numbers of model-predicted RATs as a function of range (i.e., as refraction causes lower order arrivals to no longer connect source and receiver as range increases). The cost function also allows for false positives in the set of measured RATs due to the expectation that an automated RAT estimation code will yield some non-zero number of estimated RATs that do not correspond to an acoustic arrival from the source of interest (e.g., an estimated RAT due to background noise or calls from conspecifics in the acoustic data).

3.4.2 Cost Function Rationale

Given the large dataset to be processed in Chapter 4, robust automation is a priority. Manually picking measured RATs may be an effective way to identify information for localization, but it is not an efficient way

to carry out this task for more than a nominal number of detections. Thus, the cost function is structured in such a way that it is robust to noise in the measured RATs. This is accomplished by instructing the cost function to search over every valid combination of measured RATs, the set of which is indicated by $K(r)$ in Equation 3.2. Requiring that the set of measured RATs is, in every case, larger than the set of model-predicted RATs makes the cost function flexible to the inevitable noise. This flexibility comes at the cost of additional computational complexity, since the cost function will perform many operations to calculate the eventual scalar value representing the best fit between a subset of the measured RATs and the model-predicted RATs for the candidate source range in question.

Another advantage of a large $K(r)$ set is to allow multiple simultaneous calling animals to be localized. If the multipath arrivals of a second call overlap with those of an earlier one, a superposition of the two calls will be recorded. By picking more measured RATs than the number of arrivals expected in a call, the set of measured RATs is more likely to contain all the relevant arrivals for the first call, in addition some spurious ones from the second call and background noise. The search for a relevant subset of $K(r)$ can rule out some of these spurious RATs and still have enough measured RATs from the call in question to locate the call. Then, after the first call is located, the second call can be located by aiming to rule out the spurious RATs resulting from the earlier call.

A further advantage to having more measured RATs than model-predicted RATs is that the size of the modeled set decreases as candidate source range increases. Due to refraction caused by the shape of the underwater sound speed profile, the direct path arrival is expected to be undetectable beyond a certain range. Higher order arrivals will also have their own ranges of undetectability. If the set of candidate source ranges includes one of these undetectable range values, the set of model-predicted RATs will either have to ignore some early arrivals (thereby discarding valuable information) or have different numbers of model-predicted RATs at different ranges. Thus, to avoid wasting information (e.g., avoid looking for D arrivals at all because they are not found beyond close range), the cost function will need to accurately compare ambiguity values for different candidate source ranges even when the model-predicted RATs sets at those ranges are of different sizes. As a result, at least at some ranges the measured RAT set will naturally be larger than the model-predicted RAT set.

The size of the measured RAT set is a compromise between range estimation reliability and computational burden. Too small a measured RAT set and reliable range estimates won't be possible since the relatively faint higher order arrivals may not reliably be found in the measured set, while keeping too many will result in too many combinations of RATs to search over (and, thus, resulting in excessively high computational burden). Experimentation with different measured RAT set sizes found that keeping 9 measured RATs for each detection yielded enough higher order arrivals to permit reliable range estimation. In different propagation conditions, such as if bottom losses were greater and fainter arrivals were sought in the measured data, more RATs than 9 might yield superior range estimates. However, for this dissertation, each detected call yields 9 measured RATs, with the expectation that only a subset of these 9 RATs will be based on direct or interface-reflected arrivals in the underlying acoustic data and that the rest will be noise or calls from conspecifics. At close range, each candidate range will have 4 model-predicted RATs and the first arrival is the direct-path (D) arrival. For sources at medium range, direct-path arrivals no longer reach the hydrophone and each candidate range will have 3 model-predicted RATs with the first arrival being a bottom-surface (BS) arrival. At far range, neither direct nor BS reach the hydrophone, leaving 2 model-predicted RATs

at each candidate source range with the first arrival being a bottom-surface-bottom-surface (BS2) arrival. Every candidate source range is predicted to have BS2, BS3, and BS4. Based on experience with the ACO data set, arrivals of BS5 and higher are not expected to reach the hydrophone with sufficient signal strength to be reliably used for localization.

In this work, for each detected call the cost function is evaluated at a set of candidate source ranges. These candidate source ranges are the ranges at which model-predicted RATs are calculated. Thus, the estimated range for a given source must be found in this set of candidate source ranges. The process for selecting the source ranges that make up the set of candidate source ranges is described in Section 3.4.6.

Given the properties of the measured and model-predicted RATs sets described above, ambiguity vector calculation using the cost function is carried out as follows. For a given detected sound, 9 measured RATs are identified. For each close candidate source range (i.e., each candidate source range where a direct-path arrival is predicted to connect source and receiver), 4 model-predicted RATs are present. This gives $C_4^9 = 126$ individual comparisons between a subset of four of the measured RATs and the full model-predicted RAT set at candidate range. Thus, for each close candidate range, 126 individual ambiguity values are calculated. The peak ambiguity value across this set of 126 is selected as the ambiguity vector entry for the candidate source range. At medium range, 3 model-predicted RATs are present (i.e., BS is the shortest eigenray) which gives $C_3^9 = 84$ individual comparisons for each candidate sources range. Finally, at far range, BS2 is the shortest eigenray path and 2 model-predicted RATs are present which gives $C_2^9 = 36$ individual comparisons for each candidate source range. Once all of these comparisons have been carried out and the ambiguity vector is fully populated, the range with the highest overall ambiguity value is selected as the estimated source range.

If the cutoff range for the D and BS arrivals is known, no further modification to the cost function evaluation procedure described above is necessary. However, we account for uncertainty in the eigenray cutoff range ranges by introducing "double option zones". These zones are intended to allow for uncertainty in the range at which the D and BS rays no longer connect. In these double options, each candidate source range will have two different sets of RATs. For example, around the D cutoff range, each candidate source range has two different RATs sets: one in which D is the earliest arrival, and a second in which BS is the earliest arrival (and D isn't detectable in the acoustic data). Similarly, in the vicinity of the BS cutoff range, each candidate source range has two different RATs sets: one in which BS is the earliest arrival, and one in which BS2 is the earliest arrival. During the ambiguity vector calculation step, the double option zones are combined to preserve the highest ambiguity value at each candidate source range. More detail on the double option zones is presented in Section 3.4.6.

There is more than one way to formulate a localization cost function, depending on the available data and the need for robustness from noise or errors in either the measured or model-predicted data. For example, in Rideout et al. [2013], the path (e.g., direct, surface bounce, bottom-surface bounce, etc.) for each measured arrival is known *a priori*, and a linearized local approximation of the underlying cost function is traversed to allow a large number of model parameters (including source and receiver locations) to be estimated without requiring a grid search.

3.4.3 Initial Data Processing

After raw acoustic data are collected, initial data processing is done to limit the scope of the data to the narrow bandwidth within which the 20 Hz fin whale calls are found. The primary motivation for this step is computational efficiency, as some of the processing steps to follow will have run times proportional to the sampling frequency of the acoustic data. To reduce the data transmittal, and archival storage, requirements on the ACO hardware, the acoustic data provided by the ACO group was stored in a custom binary format (i.e., *.HYD24BBpk) rather than a more customary format such as *.wav. Using custom software written by the ACO group, the binary files were read into Matlab with a sampling frequency of 24 kHz, yielding approximately 12 kHz of usable bandwidth. The eventual calls of interest for this work (fin whale 20 Hz calls, as described in Section 4.1) are frequency downsweeps from roughly 25-18 Hz. Thus, most of the 12 kHz bandwidth is unnecessary to reliably represent the fin whale calls of interest and the data were first resampled from a 24 kHz to 200 Hz sampling frequency.

Next, the data are bandpass filtered using a butterworth filter with the following characteristics:

- Passband: 9 - 31 Hz
- Stopband: <= 5 Hz and >= 40 Hz
- Minimum stopband attenuation: 20 dB
- Maximum passband ripple: 0.5 dB

The primary motivation for the filtering step is to eliminate some potential spurious signals, found outside the fin whale calling band, from the recorded data. The 20 Hz fin whale calls exhibit remarkable consistency in their frequency characteristics (i.e., minimum frequency, maximum frequency, and peak frequency) which allows for a tighter filter passband than would be required if the calls exhibited greater variability. Following bandpass filtering, the downsampled and filtered time series were written to *.wav files for subsequent processing.

3.4.4 Fin Whale Call Detection

As shown in Figure 3.5, once the acoustic data has been downsampled and bandpass filtered but before a fin whale call can be located, its presence in the data must first be observed. This process is called detection, and is a common task within marine mammal passive acoustic localization field and in the broader pattern recognition field. In this context, detection can be thought of as the process of assessing the presence or absence of a signal of interest within a broader dataset. Examples of marine mammal call detection approaches include power law processors (Nuttall [1994]) and spectrogram correlation (Mellinger and Clark [2000]). Many detection approaches are intended to assess the presence or absence of a potential signal of interest rather than to assess the source of the signal in question. This second (assessment) task is regularly called classification. The stereotyped nature of the 20 Hz fin whale calls means that simultaneous detection and classification is possible through an approach called matched filtering (Van Trees Jr et al. [2013]).

Effectively, matched filtering aims to find instances of a known signal in a noisy time series. It does this by comparing a reference waveform (i.e., the signal of interest) against a recorded time series. The output of the matched filter algorithm is a time series of the same length as the recorded time series whose value

Date	# of Manual Detections	# of Automated Detections
March 1, 2007	9294	20438
November 17, 2007	12247	21580
February 9, 2008	15487	25073

Table 3.1: Number of manual and automated 20 Hz fin whale detections on the three days of manual analysis. Automated results are shown without pruning.

represents the similarity of the time series to the reference waveform. After the comparison stage between the reference and recorded time series, the output of the matched filter is processed to identify times at which the presence of the reference waveform in the recorded signal is indicated. In this work, a Page's test (Page [1954]) is used to perform the identification step on the matched filter output. Effectively, the Page's test identifies abrupt changes in the matched filter output which are interpreted to correspond with start and stop times of matches in the recorded time series to the reference waveform. Following this Page's test, a median filter is applied to close small gaps in the Page's test output. Next, an edge detector is applied to the median filter output to identify detection start and stop times.

Figure 3.7 presents the detection process as it is implemented in this work. This detector takes as its input recorded acoustic data and reference waveforms for signals of interest, and outputs times in the acoustic data where sufficient evidence of signals of interest is found.

The choice of reference waveform is a crucial one in the design of a matched filtering approach. For this work, the reference waveform for the 20 Hz fin whale calls recorded by the ACO was described in Silver [2014], and consists of a synthesized 20 Hz fin whale call whose characteristics were calculated using a set of manually-identified 20 Hz calls from the ACO data set. This reference waveform was deemed appropriate for our localization work since it also comes from the ACO data set during the time interval considered in this work.

To test the performance of the automated detection approach described above, the detector output is compared against manually detected fin whale calls in a small subset of the ACO acoustic data record. The manual detection step was completed as part of the MSc research presented in Silver [2014]. All of the 20 Hz fin whale calls apparent in the recordings were identified by examining both time and frequency domain representations of the acoustic data. Table 3.1 lists the dates on which manual analysis took place and total number of detections on each of these days from Silver [2014], and also shows the number of automated detections on each of those days based on the analysis done in this dissertation. In Section 3.4.7, a pruning approach will be discussed which uses frequency and localization information to reduce the false positive rate for the detector.

Figure 3.8 from Silver [2014] shows a comparison between manual and automatic detection results for five minutes of acoustic data on November 17 2007. The spectrogram for the acoustic file in question is shown in the background, while automated and manual detections are denoted using blue and red vertical lines respectively. In general, the manually detected calls were reliably represented in the automated detection results, with more automated detections than manual ones. The excess of automated detections was a conscious decision, and desired outcome, when designing the detection scheme for this work. A perfect detector would identify all the 20 Hz fin whale calls in the recorded data and none of the background noise. In other words, the perfect detector has a probability of correct detection of 100% and a probability of false

Path Label	Path Abbreviation
Direct	D
Surface	S
Bottom	B
Surface Bottom	SB
Bottom Surface	BS
Bottom Surface Bottom Surface	BS2
Bottom Surface Bottom Surface Bottom Surface	BS3
Bottom Surface Bottom Surface Bottom Surface Bottom Surface	BS4

Table 3.2: Naming convention for eigenray paths used in this work.

alarm of 0%. Practically speaking, this perfect detector doesn't exist, and a compromise must be struck between the correct detection and false alarm probabilities. For this particular instance, the detector was designed to be intentionally permissive (i.e., have a high correct detection probability at the expense of a high false alarm rate) because data processing steps that come after the detection step will provide additional information (such as the peak frequency for each detection and localization information) that will be used to help distinguish true from false detections.

A key observation to make from Figure 3.8 is that the automated detector identifies not only the first arrival for each call, but multiple higher order arrivals too. This is due to the similarities between the signal characteristics among the different arrival paths. In other words, at least for the arrival paths considered in this work, the spectral content of a higher-order arrival such as a bottom-surface-bottom-surface-bottom-surface arrival (i.e., a BS3 arrival) is similar enough to that of lower-order arrivals (e.g., D or BS) that the same matched filter reference waveform capable of detecting the first arrival of a 20 Hz fin whale call is predisposed to detecting other arrivals, too. A consequence of the detector returning times for higher order arrivals, along with the earliest, is that the same call is included more than once in the set of detections. This can distort analyses carried out on this detection set, including measurements of calling rates and inter-call interval, so minimizing their impact can be useful. The process for reducing this false positive rate is called pruning in this work, and is described in greater detail in Section 3.4.7. For reference, Table 3.2 lists the eigenray naming conventions used in this work.

Once the matched filter detector has identified potential signals of interest in the recorded acoustic data, features in the recorded data that are relevant to estimating the horizontal range between the source and receiver are estimated, as described in Section 3.4.5.

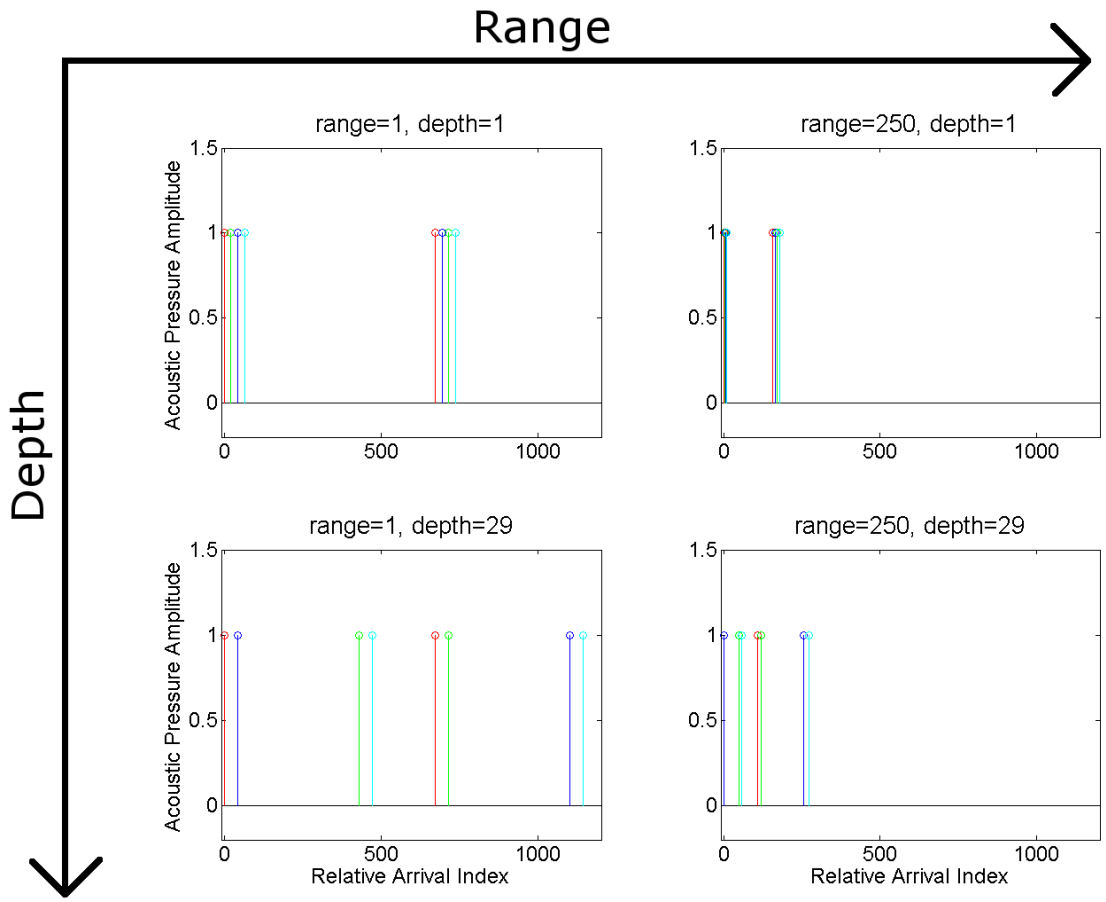

Figure 3.6: Eigenray arrival times for 4 different source locations. The x-axis is measured in samples after the arrival of the direct-path sound at the simulated hydrophone. Range and depth are dimensionless quantities meant to convey relative differences in source range and depth between the four cases shown in this figure. Only the first 8 direct and interface-reflected arrival paths are shown. If the source is impulsive, these plots can also be interpreted as approximations of the impulse responses for these four cases. The receiver is held fixed at a range and depth (r,z) of (0,39) with a water depth of 40. Sources are at (1,1), (250,1), (1,39), and (250,39). A range independent propagation environment is used.

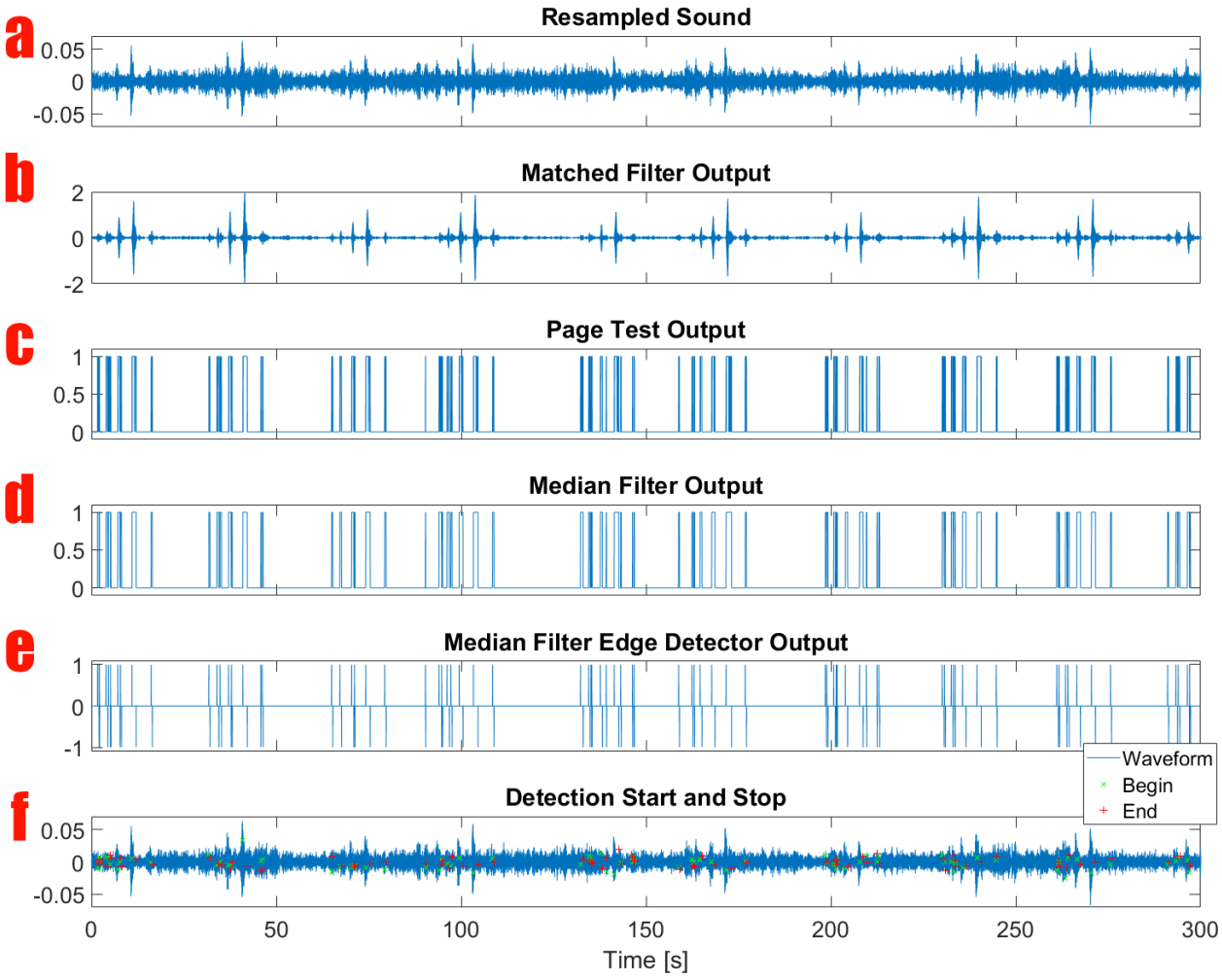

Figure 3.7: Automated fin whale call detector output for 0h00-0h05 on March 3 2014. Acoustic data recorded at the ALOHA cabled observatory. A sample recorded time series is shown in subplot **a**. Subplot **b** shows the matched filter output using the fin whale reference waveform and the data shown in subplot **a**. The Page's test output is shown in subplot **c**. Subplot **d** shows the median filter output, while subplot **e** shows the output of an edge detector on the median filter output. Finally, subplot **f** superimposes the detection start and end times on the waveform shown in **a**.

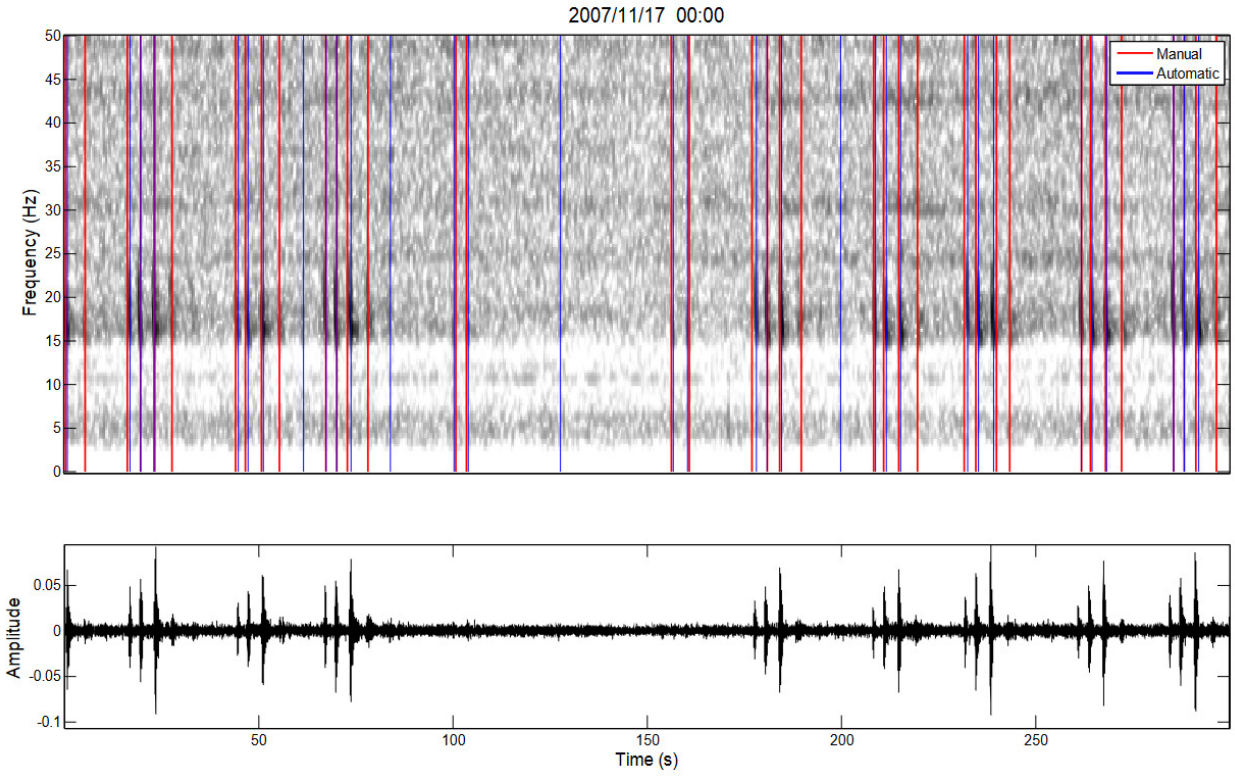

Figure 3.8: Manual vs automated detection results from November 17 2007, from Silver [2014]. Acoustic data recorded at the ALOHA cabled observatory.

3.4.5 Measured Relative Arrival Time Estimation

The information content that the localization algorithm leverages to estimate the range from the hydrophone to the calling fin whale is contained in the Relative Arrival Times (RATs) for the fin whale call in question. For this work, RATs are defined as the set of time differences between the earliest connecting arrival (i.e., the lowest order arrival to connect whale and hydrophone) and all apparent higher order arrivals. At short range, the earliest arrival to reach the hydrophone is the direct (D) path from source to receiver. Due to refraction, as range increases the direct path will cease to connect source and receiver. For a shallow source and ocean-bottom receiver, the surface (S), bottom (B), and surface-bottom (SB) arrivals will arrive will arrive very close in time to the D arrival, and will stop connecting at similar ranges. Thus, given this source:receiver geometry, if the D arrival is undetectable the next distinguishable arrival is expected to be the bottom surface (BS) arrival.

Figure 3.9 illustrates how RATs appear in the ACO fin whale data. Subplot **a** shows the recorded waveform for a single 20 Hz fin whale call and all its visually-apparent higher-order arrivals. A spectrogram of this waveform is shown in the middle subplot. The red $*$ symbols indicate times where the matched filter detected the presence of a fin whale call. In this 28.5 s time window 2 fin whale call detections were made, one for the earliest detection of this call and one for a faint fifth arrival for this call. This figure shows data where most of the false positive detections have already been pruned from the detection set. On this time scale, the arrival times for the first four arrivals can be manually identified from either the time or frequency domain plots as: 2.5, 5.06, 8.57, and 12.47 s. By subtracting the first arrival time (2.5) from the remaining 3 arrival times, the RAT set for this particular call can be manually calculated as: 2.56, 6.07, and 9.97. The previous example shows how RAT calculation can take place in a manual setting. While illustrative, this approach is not practically useful when dealing with the large volume of calls expected once analysis of the full ACO data set is undertaken. The greater the degree of automation that can reliably be brought to bear to the task of estimating the RATs for each fin whale call detection, the less time required to estimate the RATs for all the detected 20 Hz fin whale calls. In this work, the automated RAT calculation is implemented by calculating the Hilbert transformation (i.e., the envelope) of the cross-correlation of the recorded time series immediately about the detection time with the full analysis window for this detection. Specifically, the steps for this automated RAT calculation approach are:

1. Define the cross-correlation reference window as the set of recorded acoustic data within $\pm 1s$ of the detection time.

2. Define the cross-correlation analysis window as the data set that includes recorded data from 2.5 s before the current detection to 20 s after the current detection.

3. Calculate the cross-correlation between the reference and analysis windows.

4. Calculate the absolute value of the Hilbert transform of the cross-correlation ($H(t)$) calculated in the previous step.

5. Calculate the location of every peak (i.e., local maxima) in $H(t)$.

6. Identify the anchor peak; the peak in $H(t)$ corresponding to time when the reference window perfectly overlaps with its copy in the analysis window.

Relative Arrival Time
2.50
5.07
5.85
6.41
7.47
9.40
9.93
10.46
11.00

Table 3.3: Calculated RATs for the sample detection shown in Figure 3.9.

7. Except for the anchor peak, remove all peaks earlier than 1 s after the anchor peak.
8. From the remaining peaks, identify the times for the 10 largest peaks in $H(t)$.
9. Subtract the earliest arrival time from arrival time for each of the other 9 peaks from Step 8, yielding a set of 9 candidate RATs.
10. Repeat from the first step for every other detected 20 Hz fin whale call.

The steps outlined above are illustrated in Subplot **c** in Figure 3.9. In this figure, the cross-correlation between the reference and analysis windows is shown by the red line, and its envelope by the blue line. Disregarding the auto-correlation peak (corresponding to the perfect overlap of the reference window with itself at Time = 2.5 s), the 9 sufficiently-isolated envelope peaks are the 9 measured RATs for this detection. To avoid side lobes of a strong peak counting in this set of 9, envelope peaks are sorted from largest to smallest, and peaks within 0.5 s of a stronger peak are disregarded. This threshold was chosen as being smaller than the smallest model-predicted RAT given the environment that is used for the measured data in Chapter 4. Table 3.3 lists the relative arrival times indicated in subplot **c** in Figure 3.9, calculated by subtracting the arrival time for the auto-correlation peak from the 9 highest isolated peaks.

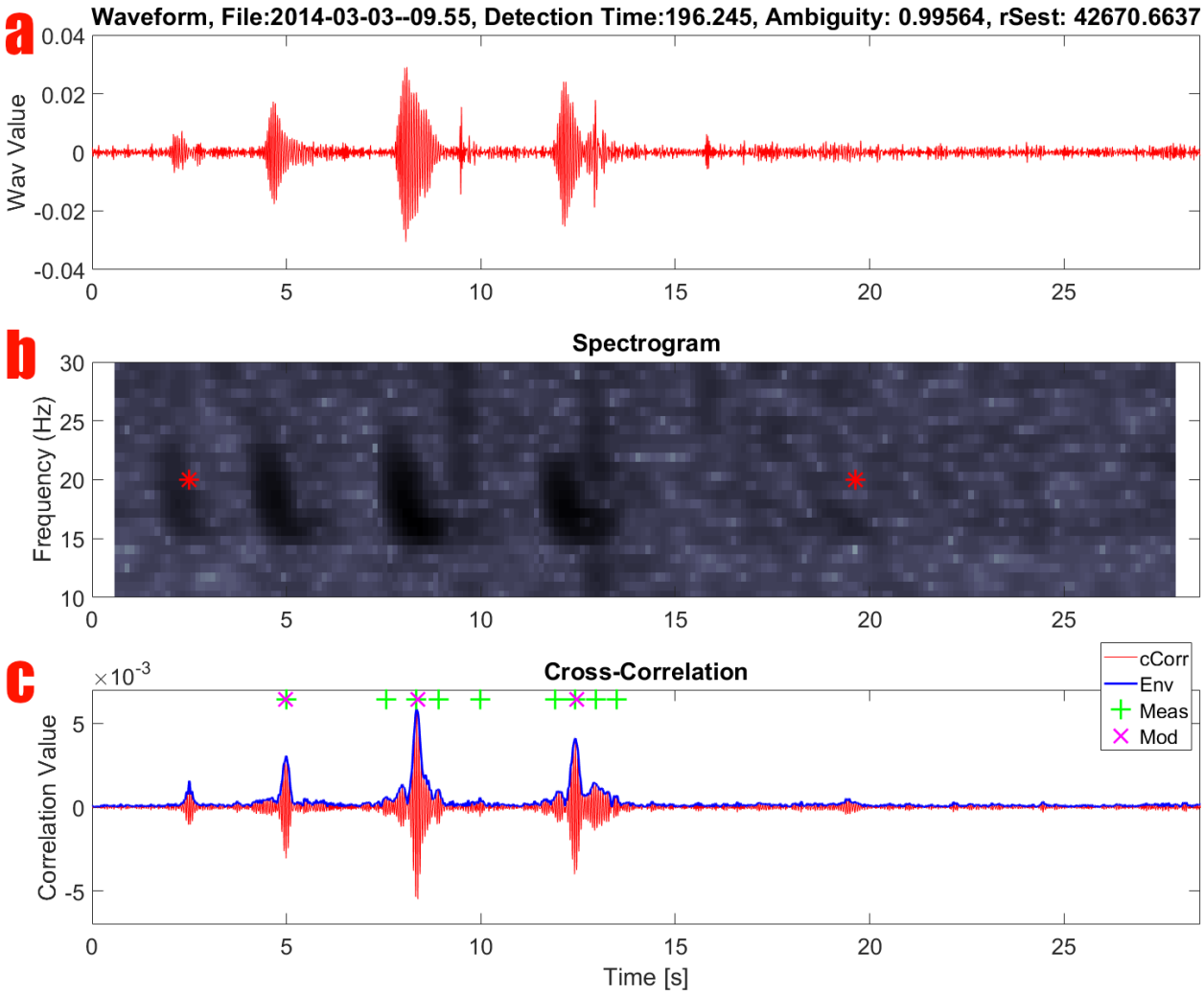

Figure 3.9: Time series depicting the measured relative arrival time estimation process. **a** Acoustic time domain data for a single fin whale 20 Hz call recorded 09h58:16.245 UTC on March 3 2014 at the ALOHA cabled observatory. **b** Spectrogram in the 10-30 Hz range, with red $*$ symbols highlighting the output from the automated detector after pruning was done. **c** Cross-correlation (red trend), and cross-correlation envelope (blue trend) between the earliest arrival waveform and the entire time series shown in the top subplot. Green + symbols highlight the 9 sufficiently time-separated envelope peaks, and purple X symbols show the best fit between a model-predicted set of RATs and this set of measured RATs.

3.4.6 Underwater Acoustic Propagation Modeling Approach

A critical step in the localization process is the calculation of model-predicted RATs (MPRATs) at each candidate source range. These MPRATs express the effect that environment, receiver depth, source depth, and distance between source and receiver has on the sound that reaches the hydrophone, and must be calculated for each new environment. The more accurately the model emulates the physical environment, the more reliable the resulting propagation modeling will be and the more useful the MPRATs will be at comparing against the RATs estimated from the acoustic data. For this work, the environment consists of the sound speed profile (i.e., the variation in sound speed as a function of depth) and the bathymetry (i.e., the shape of the ocean bottom as a function of latitude and longitude), given that source and receiver depths are held to be constant. If a poor model, or inaccurate environmental characteristics, are used, this localization result will still endeavor to find the optimum fit between MPRATs and RATs. However, as the adage "garbage in, garbage out" goes, a poor model will likely yield location estimates that poorly match to the actual source location. In this section, the underwater acoustic propagation modeling, and modeling conditions, used for this work are discussed.

The sound speed profile (SSP) used for this work is shown in Figure 3.10, and was calculated using the approach outlined in Chen and Millero [1977] (as promulgated in Fofonoff and Millard Jr. [1983]) and using Conductivity Temperature Depth (CTD) data collected during the Hawaii Ocean Time-series (HOT) Cruise #239 between 17-21 January 2012. This date was selected to be both near the mid-point of the ACO acoustic data record considered in this work and also lying in 2011-2012 fin whale calling season (approximately in the northern hemisphere winter months). Near-monthly CTD data is available from the HOT archive dating back to 1988. This single SSP is used for data from all days in this work due to the stability of the SSP at ACO and the depth of the ACO hydrophone, as previous studies near the ACO site have found the SSP to be relatively stable below the near-surface mixed layer (Au and Lammers [2016]). Given that the ACO hydrophone is very close to the ocean bottom, every eigenray will spend more of its travel time in regions of the SSP outside the shallow mixed layer. In other words, the similarity in the profiles, particularly at depths below the mixed layer, support the claim that SSP variation shouldn't have a great effect on ray travel times and that a single SSP be used to reliably calculate MPRATs throughout the measured data record.

The bathymetry for the ACO was measured by the Hawaii Mapping Research Group at the University of Hawaii at Manoa. Data of 1 km resolution is available for the region around ACO, and is shown in Figure 3.11. Slices through this bathymetry data along different bearings will be used to evaluate how much the MPRATs change with bearing angle from ACO.

Given the deep ocean conditions, and the low frequency (and thus long wavelength) of the fin whale calls under consideration, a ray-tracing propagation modeling approach is used. Ray tracing is not typically the propagation modeling approach of choice for such low frequencies, but given that the water depth at ACO is $4728/(1530/18) \approx 57$ times the call wavelength a ray-based acoustic propagation modeling approach is considered appropriate. In shallow water, a ray tracing propagation model would not be appropriate, but recent work on single hydrophone underwater acoustic modal separation may provide an alternative (Bonnel et al. [2020]).

The RAYFAST ray-tracing code (Dosso and Ebbeson [2006]) was used to calculate eigenray arrival times from which to calculate MPRATs. This model was selected over the more widely-used BELLHOP model (Porter and Liu [1994]) because RAYFAST is more effective at calculating eigenray travel times than

Path	BELLHOP Travel Time	RAYFAST Travel Time	% Difference
D	3.0469	3.0449	0.0666
BS	9.3669	9.3296	0.3997
BS2	15.7178	15.6148	0.6597
BS3	22.0616	21.9000	0.7381

Table 3.4: Eigenray travel times, and percent difference, for the RAYFAST and BELLHOP underwater acoustic propagation models computed using the same environmental conditions.

Arrival Path
Direct
BS
BS2
BS3
BS4

Table 3.5: Eigenray arrival times calculated at each candidate source range. Not all eigenrays connect at all candidate source ranges.

BELLHOP is. The two models have been found to yield similar ray travel times. Table 3.4 compares travel times for four eigenray paths computed for the same environmental conditions using the RAYFAST and BELLHOP models. The same SSP was used for both models. Water depth of 4728 m, a hydrophone depth of 4727 m, source depth of 150 m, and a source range of 100 m were also used.

The travel times for the eigenrays listed in Table 3.5 were sought for each candidate source position (i.e., horizontal distance between source and receiver). The source depth was set to 50 m, based on the fin whale calling depth in Watkins et al. [1987] and Weirathmueller et al. [2013], and the receiver depth was set to 2 m off the bottom (i.e., 4726 m). The set of candidate source ranges are defined according to the approach described in Section 3.4.6 (Range Grid Search Resolution Determination) below, rather than using a uniform spacing. In this table, BS refers to the eigenray that, after leaving the source, reflects first off the ocean bottom (B), then off the ocean surface (S), before reaching the receiver.

As part of the ongoing Hawaii Ocean Time-series data collection effort, monthly CTD casts have been collected in the vicinity of the ACO site since October 1988. Au and Lammers [2016] shows a collection of measured SSPs calculated from CTD casts at the ACO site collected monthly over a single year. Despite differences in the shallowest depths of these profiles, the deepest regions of the profiles display noteworthy similarity. Since the receiver is located close to the ocean bottom, and despite the fact that the fin whale sources are all near the surface, each eigenray traveling from source to receiver spends most of its time at depths where the SSPs are very similar. Due to the similarity between the deep water portions of these profiles, a single SSP will be used to calculate the model-predicted RATs.

The RAYFAST acoustic propagation model (Dosso and Ebbeson [2006]) is used to calculate eigenray arrival times at each candidate source range and assuming a constant source depth of 50 m for fin whales. A range-independent water depth of 4667 m, the sound speed profile in Figure 3.10, and a receiver 1 m above the bottom, describe the environmental characteristics passed to RAYFAST. A discussion on how this water depth value was selected, rather than using the measured water depth at ACO, is included in Section 3.4.6.

Using the modeling conditions described above, ray arrival times were calculated at each candidate source

Arrival Path	Cutoff Range [km]
Direct	30
BS	92
BS2	151
BS3	n/a
BS4	n/a

Table 3.6: Eigenray arrival times calculated at each candidate source range. Cutoff ranges for the BS3 and BS4 arrival paths are greater than that of BS2 and not relevant.

range using RAYFAST. Due to the shallow source depth, and the hydrophone being located close to the ocean bottom, the rays are assumed to arrive at the receiver in closely-spaced groups of four (e.g., the Direct, Surface (S), Bottom (B), and Surface-Bottom (SB) paths should arrive within a 0.1 s of each other depending on source range, followed by a time interval with no acoustic arrivals, followed by the Bottom-Surface (BS), Bottom-Surface-Bottom (BSB), Surface-Bottom-Surface (SBS), and Surface-Bottom-Surface-Bottom (SBSB) arrivals, followed by another gap and even higher order arrivals). These closely-spaced arrivals will likely not be individually resolvable in the measured data unless the source signal is sufficiently short and the sampling frequency sufficiently high. The recordings of fin whale calls considered in this work are not expected to satisfy these two criteria, thus only the earliest arrival in each group of four (i.e., D, BS, etc.), as listed in Table 3.5, will be modeled and sought from the measured data. Due to vertical refraction caused by variation in sound speed with depth, as range increases low order arrival paths will gradually stop connecting between source and receiver. Based on the modeling effort and conditions described above, Table 3.6 lists the approximate ranges at which rays in Table 3.5 stop connecting.

Once the eigenray arrival times are calculated at each candidate source range, the RATs are then calculated. For close ranges where D rays connect, the D arrival reaches the hydrophone first and its arrival time is subtracted from all higher order arrival times to yield a set of four RATs. Similarly, at medium ranges where the D arrival does not reach the hydrophone due to refraction and the BS arrival reaches the hydrophone first, the BS arrival time is subtracted from all higher-order ones. Finally, at long ranges, the BS2 arrival reaches the hydrophone first and its arrival time is subtracted from subsequent ones to calculate the RATs. Figures 3.12 and 3.13 show respectively the model-predicted arrival times, and relative arrival times, at each candidate source range.

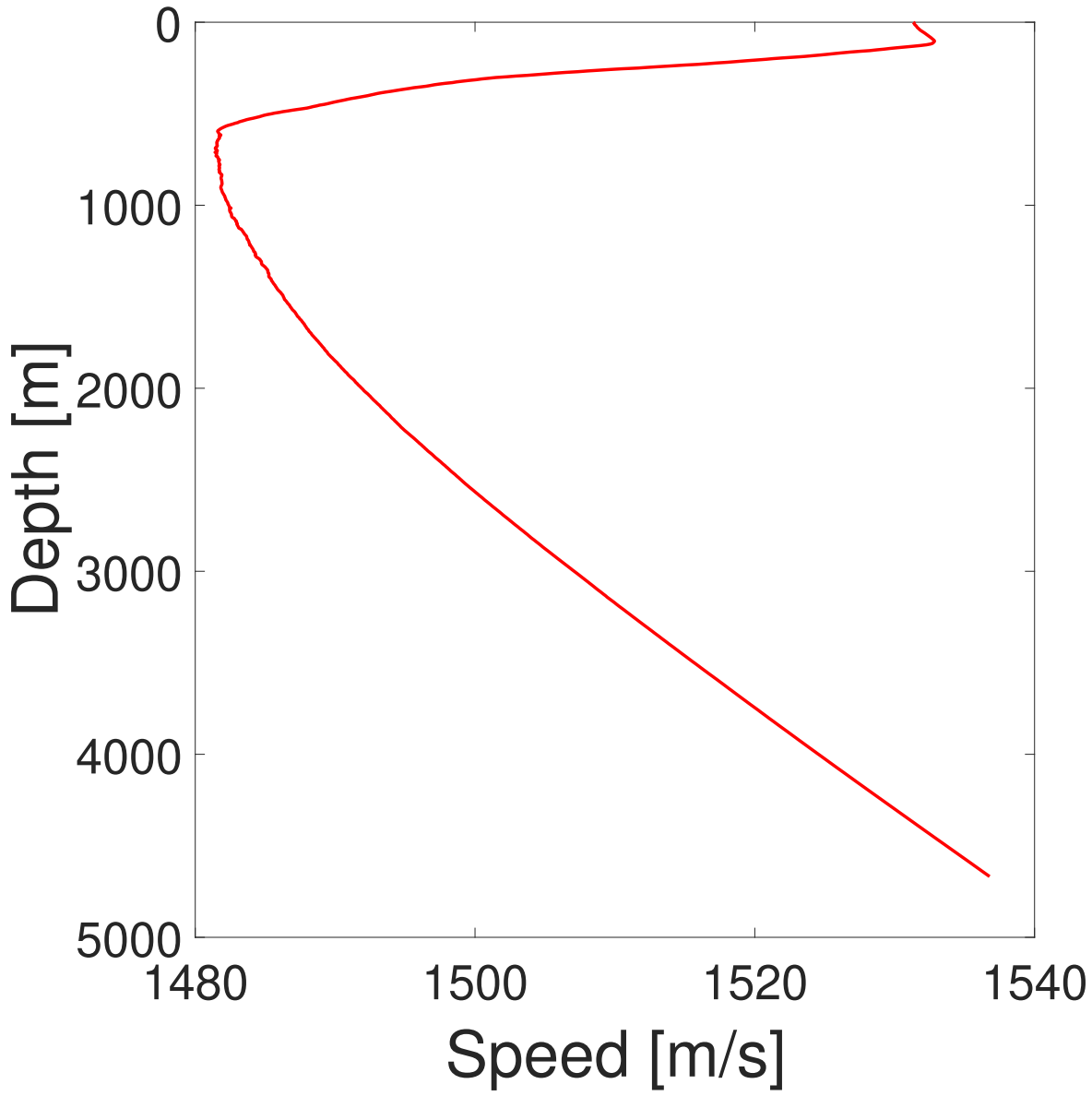

Figure 3.10: Sound speed profile calculated from CTD data collected during HOT cruise #239 and the Chen and Millero [1977] calculation approach. HOT Website: https://hahana.soest.hawaii.edu/hot/cruises.html

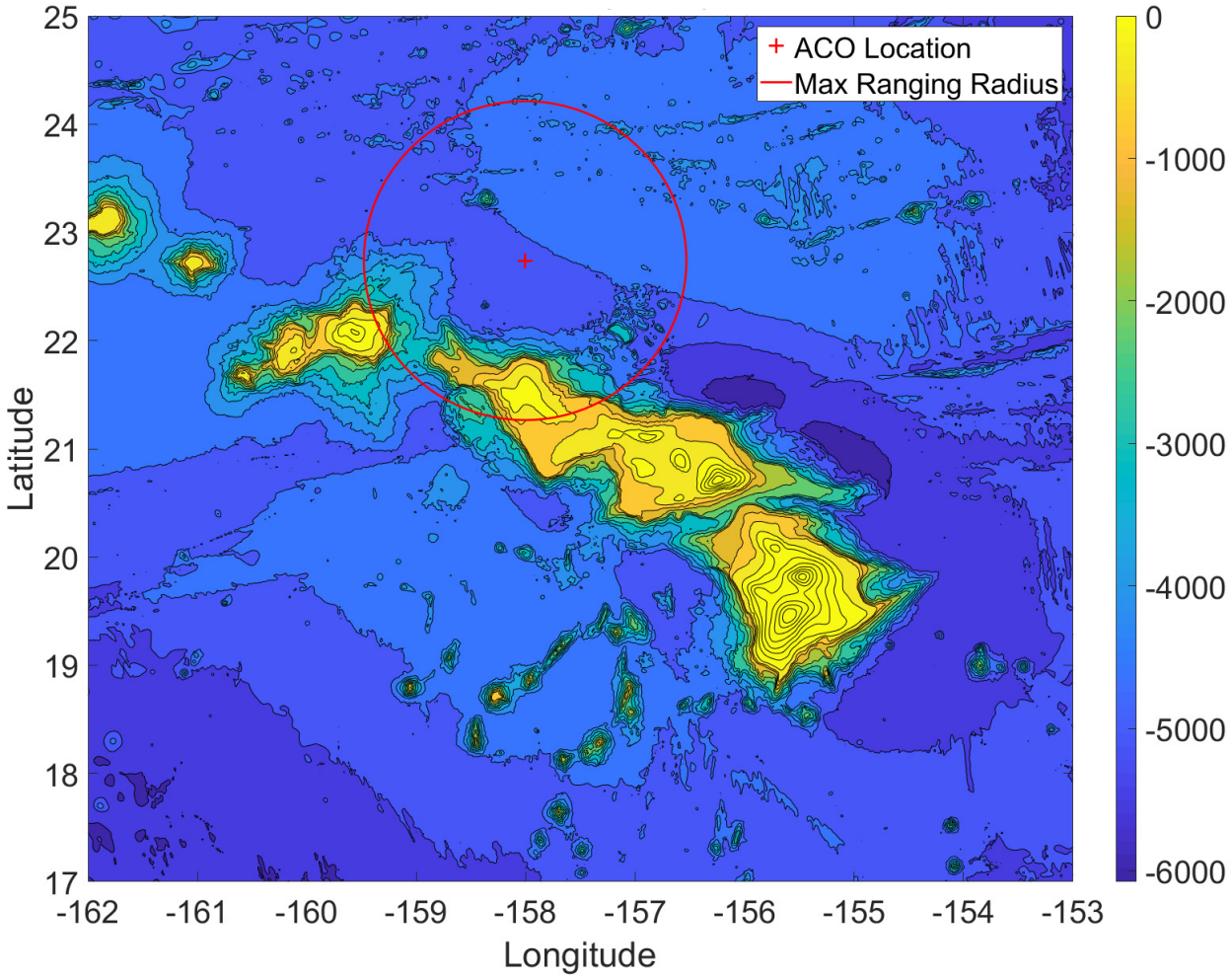

Figure 3.11: Bathymetry and topology near the Hawaiian islands. The ACO location is highlighted by the red + symbol, while the red circle highlights distance from ACO that BS2 arrivals are predicted to reach the ACO hydrophone from (other than in the landfall directions). Bathymetric and topographic data: Hawaii Mapping Research Group [http://www.soest.hawaii.edu/HMRG/cms/]

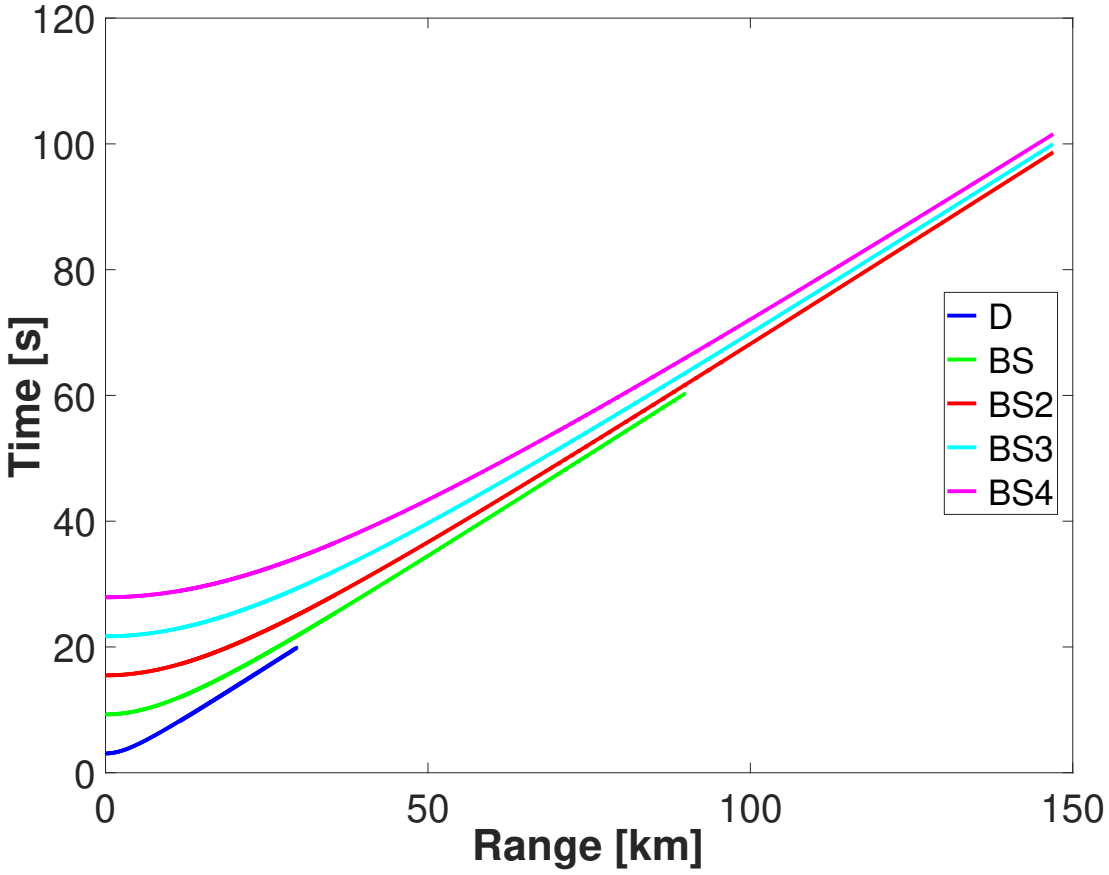

Figure 3.12: Model-predicted arrival times for the D, BS, BS2, BS3, and BS4 eigenrays for each candidate source range.

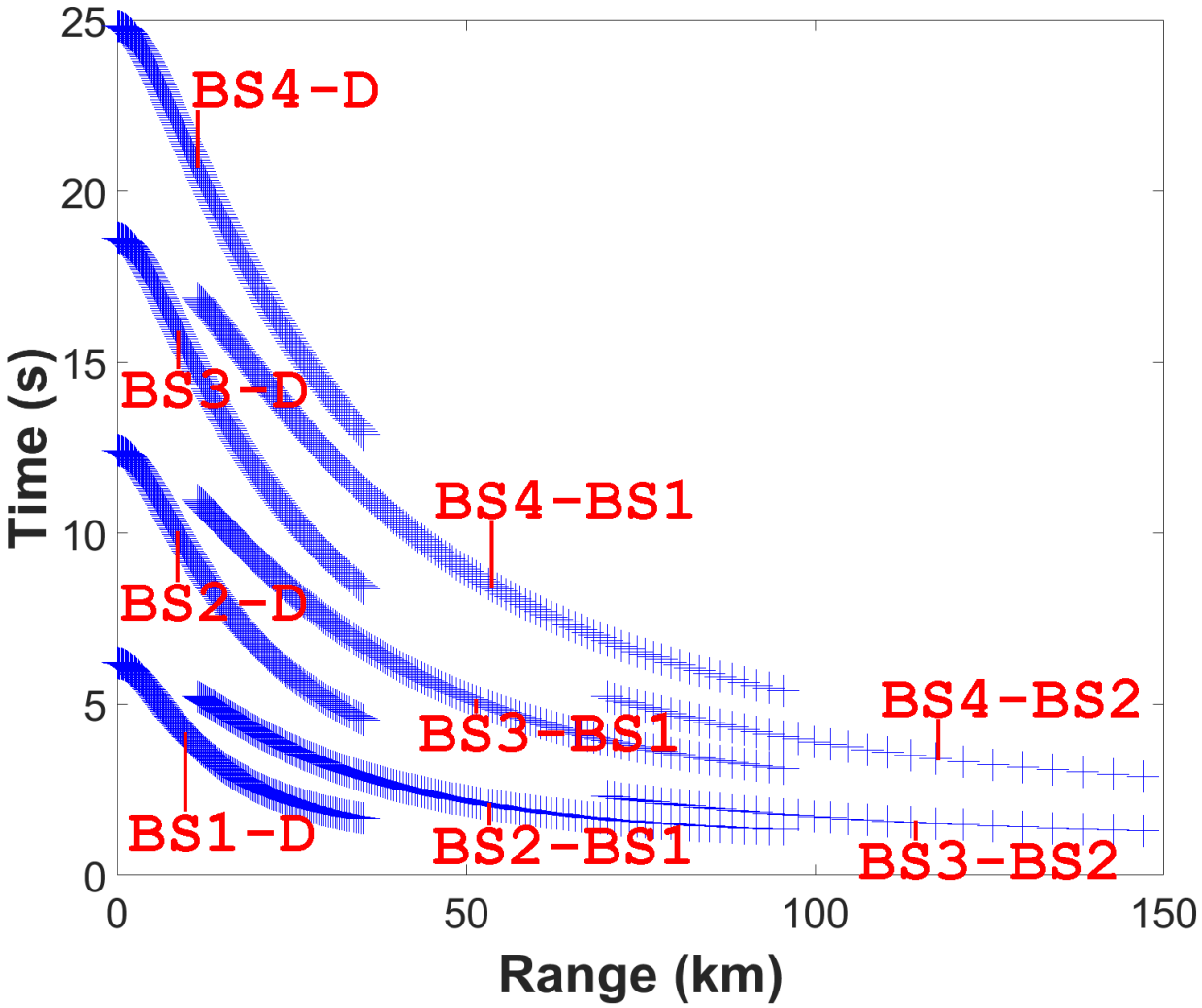

Figure 3.13: Model-predicted relative arrival times each candidate source range.

Range Grid Search Resolution Determination

The localization approach presented in this work is effectively a comparison between measured and model-predicted RATs. The process for estimating measured RATs from the acoustic data recorded by the ACO was described in Section 3.4.5. The steps for estimating the model-predicted RATs was described in the previous section. While using a uniform spacing between candidate source ranges may seem like a reasonable plan, it is expected that sources close to the hydrophone can be localized more precisely than more distant sources because RATs change more rapidly at close ranges than at more distant ones. Thus, choosing a range-dependent candidate source range grid spacing will allow fine resolution range estimation to take place close to the hydrophone where warranted while coarser resolution range estimation is done at longer ranges to relieve computational burden. A Monte Carlo simulation is conducted to test this range-dependent range resolution hypothesis.

First, RAT uncertainty is estimated. This number is estimated from the recorded data, specifically from the cross-correlation analysis used to estimate the RATs, and is reflective of the time uncertainty in the measured RATs. The width of the peaks in the cross-correlation (which is related to the bandwidth of the underlying call) is taken to represent the uncertainty in RAT estimation. Figure 3.14 shows the RAT estimation process for a measured fin whale call. The width of the cross-correlation peaks is approximately 0.1 s, which we take as the RAT uncertainty.

Next, model-predicted RATs are calculated for a set of candidate source ranges. Source depth is fixed at 50 m, and water depth is fixed at the ACO depth of 4728 m. A measured full water column sound speed profile is used with the RAYFAST propagation model. Gaussian-distributed noise with a standard deviation of 0.1 s is added to these calculated RATs at each of the source range values listed in column 1 of Table 3.7. At each range value, 2000 independent noisy sets of RATs are calculated, and the source range estimated from each of these sets. The standard deviation of the set of calculated locations (σ_{range}), for each range is then calculated. Twice this value ($2 * \sigma_{range}$) is taken to be the calculated range resolution at that range. These calculated range resolution values are summarized in Table 3.7.

Once $2 * \sigma_{range}$ is calculated for each range, an exponential curve is fit to the 18 data points listed in Table 3.7. Values calculated by this exponential curve fit are used to determine a set of candidate source ranges to calculate model-predicted RATs at according to the following procedure:

1. Define the the closest candidate source range to be 50 m.
2. Calculate the model-predicted range resolution at the current range using the exponential curve fit. For ranges greater than 120 km, use the same range resolution (4.288 km) to keep candidate source range points from being too far apart.
3. Add this range resolution to the current range, incorporate this new range into the set of candidate source ranges.
4. Repeat from Step 2 until the BS2 cutoff range of 151 km is reached.

This procedure yields 261 candidate source ranges between 50 m and 147 km. Figure 3.15 shows the calculated range resolution values from Table 3.7 alongside the exponential curve fit.

Range [km]	**Range Resolution** ($2 * \sigma_{range}$) **[km]**
5	0.163
10	0.155
15	0.171
20	0.212
25	0.238
35	0.416
40	0.485
45	0.523
50	0.620
55	0.718
100	2.827
105	3.127
110	3.480
115	3.717
120	4.245
125	4.530
130	5.110
135	5.518

Table 3.7: Source ranges and simulated range resolution used to calculate the spacing from one candidate source range and the next. This range-dependent range resolution set of candidate source ranges is used to calculate model-predicted RATs.

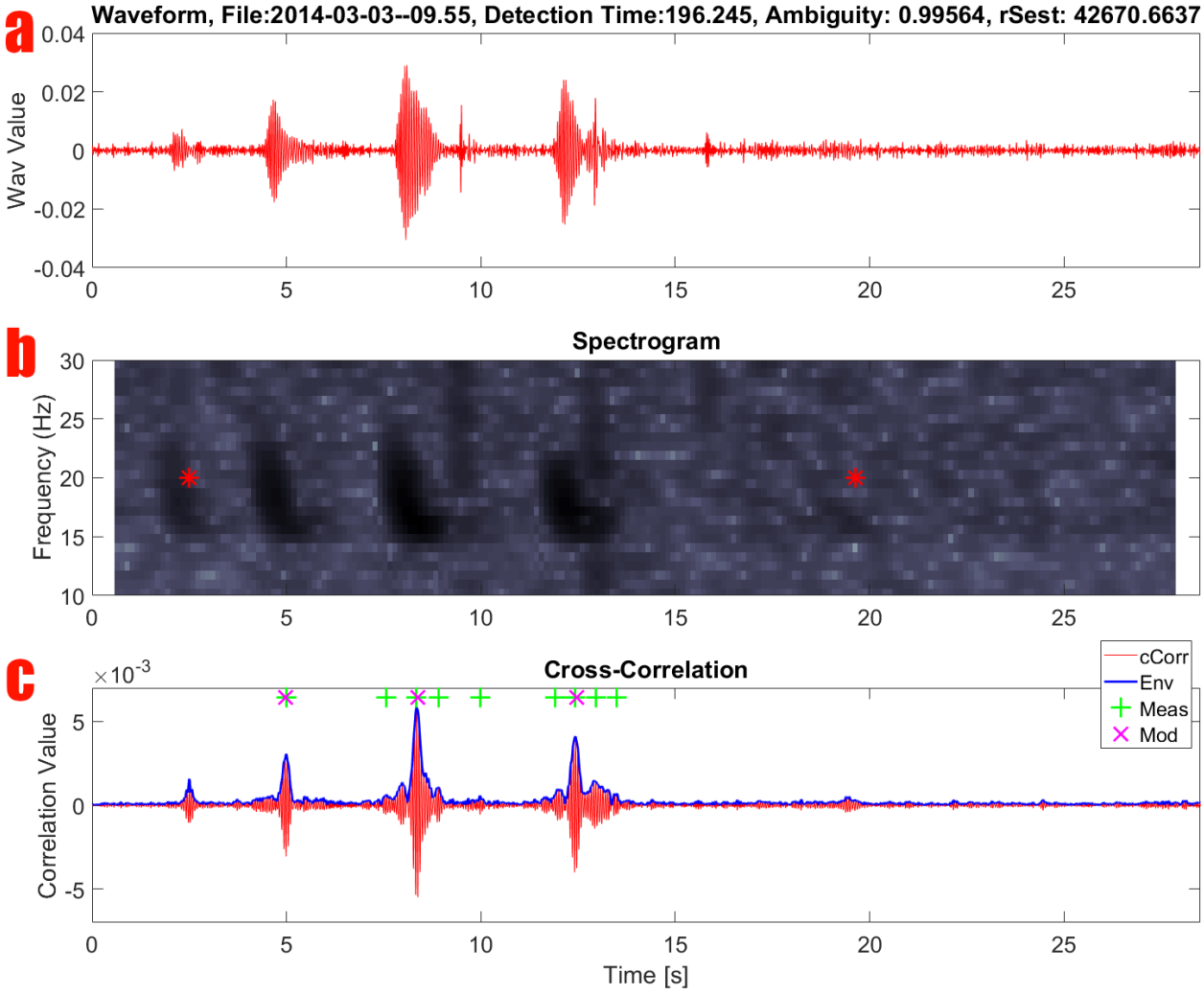

Figure 3.14: Time series depicting the measured relative arrival time estimation process. **a** Recorded time series waveform. **b** Spectrogram in the 10-30 Hz range, with red $*$ symbols highlighting automated detector output after pruning was done. **c** Cross-correlation (red trend), and cross-correlation envelope (blue trend) between the earliest arrival waveform and the entire time series shown in the top subplot. The width of the cross-correlation peaks is approximately 0.1 s.

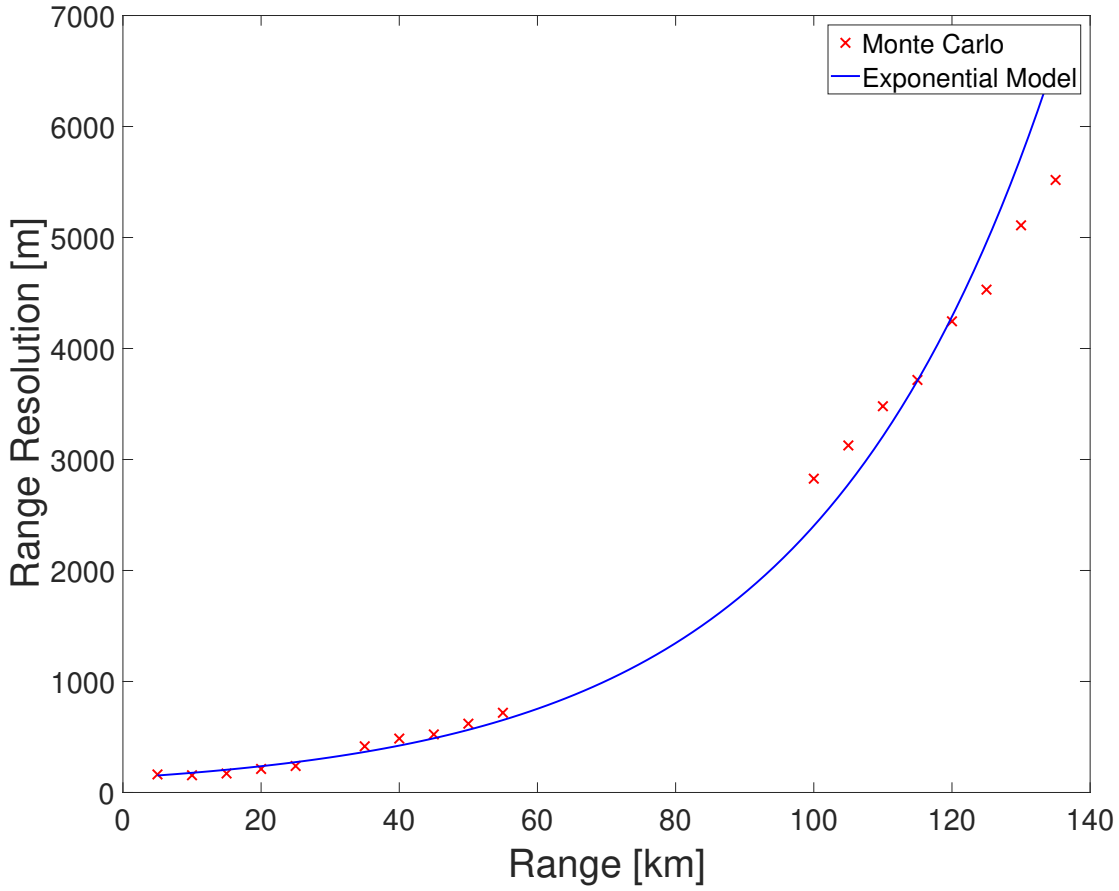

Figure 3.15: A comparison between the Monte Carlo-estimated range resolution (red x symbols) and the exponential curve fit. The set of candidate source ranges over which model-predicted RATs are calculated is calculated using this exponential curve fit to the Monte Carlo-estimated range resolution values.

Double Option Zones and Eigenray Cutoff Range Uncertainty

A characteristic of this range estimation approach is that it tolerates having more measured RATs due to noise on the measured data and side lobes from the cross-correlation used in the RAT estimation process. However, if the measured RAT set does not contain all the expected model-predicted RATs, an accurate range estimate is unlikely. For example, if the source is found at medium range, the model-predicted RAT set will include BS2-BS, BS3-BS, and BS4-BS. However, if the BS2-BS RAT is not found in the measured RAT set, the estimated range is not likely to be a reliable approximation of the true range.

One way in which a measured RAT can fail to yield the expected number of RATs is when the model-predicted eigenray received level (RL) is high enough for the propagation model to predict it connects, when in reality the RL is too low to be detectable. This phenomenon occurs near the ray cutoff ranges previously shown in Table 3.6. To account for this phenomenon, regions near ray cutoff ranges are designated within which reliable range estimation is possible regardless of potential differences between ray cutoff and ray detectability ranges. We term these regions "double option zones", and illustrate them in in Figure 3.16.

Figure 3.16 shows the same set of model-predicted RATs that were plotted earlier in Section 3.4.6, and highlights the RATs for two sets of candidate source ranges; one at the short:medium range threshold, and one at the medium:high range threshold. Within these zones, each candidate source range has two distinct sets of RATs; one which begins with a low eigenray, and one which begins with the next longest eigenray. For example, consider the short:medium double option zone near range= 30 km. Each candidate source range in the short:medium double option zone has the [BS-D, BS2-D, BS3-D, BS4-D] and [BS2-BS, BS3-BS, BS4-BS] RAT sets at each candidate source range. When the ambiguity value at each range is calculated, the measured RAT set is compared against each model-predicted RAT set within the double option zones and the best match between these two is taken as the ambiguity value at that candidate source range. In this way, continuing the previous example, if the source is located in a double option zone and D arrival isn't detectable in the measured data but BS is, an accurate range estimate is still possible.

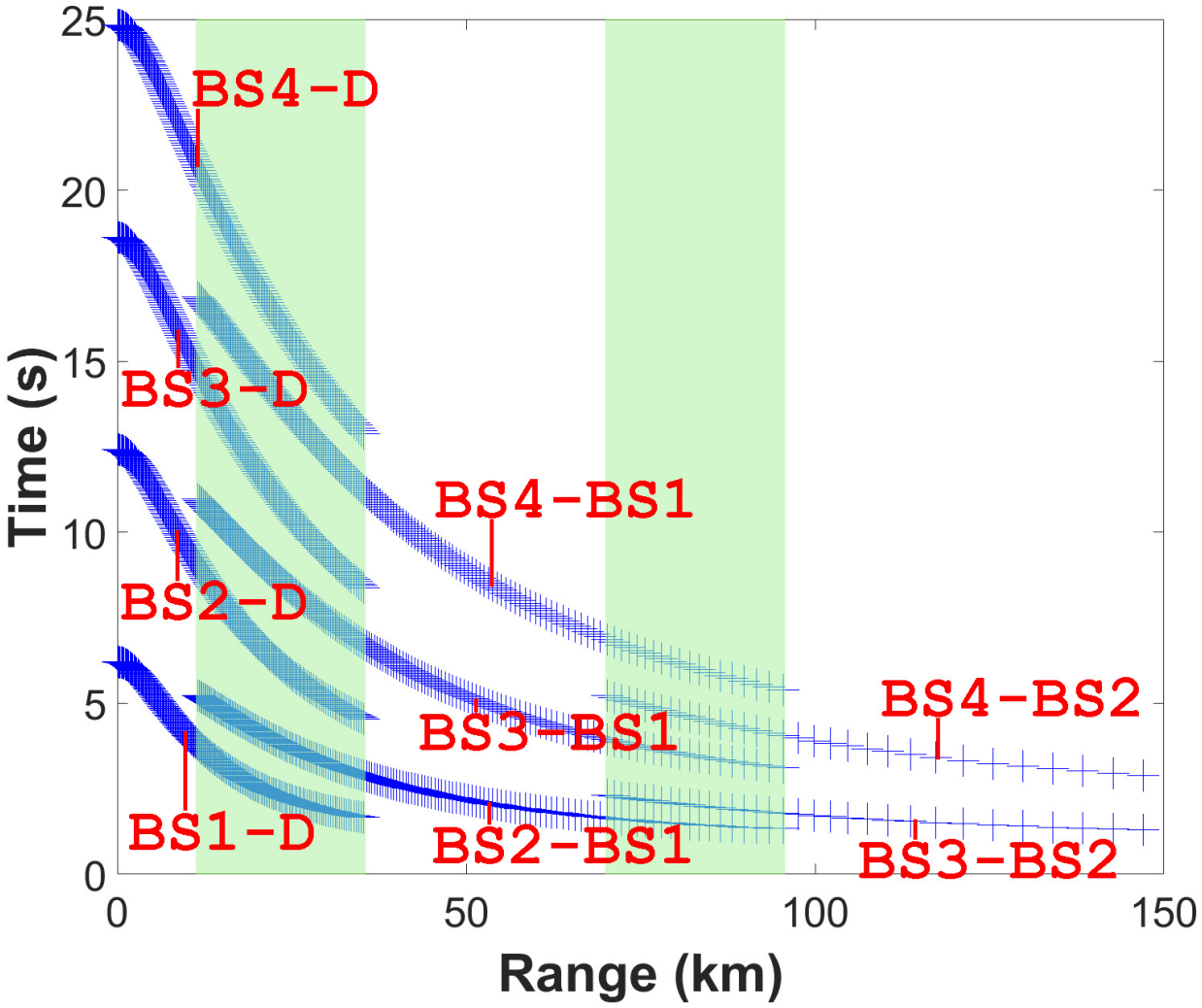

Figure 3.16: Model-predicted relative arrival times each candidate source range. Green regions highlight the short:medium and medium:high range double option zones.

Influence of Bathymetric Uncertainty on Model-Predicted RATs

Since bathymetry doesn't vary much (at a given range) with changing azimuth angle in the vicinity of ACO, we expect that the localization approach presented in this work will not be capable of estimating source azimuth. In other words because the water depth shows little variation with azimuth near ACO, we expect that measured RATs for an animal changing its azimuth angle to ACO (at constant range) will not change significantly enough in comparison to the underlying uncertainty in the measured RATs. Thus, we do not expect to be able to estimate azimuth angle from ACO data. Moreover, the gradual variation in water depth with changing range about ACO suggests that a range-independent bathymetric assumption is reasonable. In this section, this assumption is explored.

A representative water depth value was selected as follows. First, bathymetric data (of 1 km spatial resolution) for the ocean bottom surrounding ACO was downloaded from the Hawaii Mapping Research Group database. All water depth measurements within 90 km of ACO were identified, and the mean water depth within this bathymetric subset was calculated as approximately 4667.8 m (compared to the measured water depth at ACO of 4728 m). The 90 km range threshold was selected because beyond 90 km from ACO, depth along some azimuths get shallower than others, particularly toward O'ahu and Kauai. Also, given that water depth decreases along most azimuths from ACO, and that water depth has a direct bearing on model-predicted RATs, using a representative water depth to calculate model-predicted RATs will improve localization reliability.

Using the representative water depth of 4667.8 m, the ACO hydrophone depth is either 4665.8 m or 4657.8 m (since the ACO hydrophone depth changed from 2 m to 10 m off the bottom part way through its deployment). It was assumed throughout the analysis that hydrophone depth was fixed at 2 m above the bottom; this simplifying assumption is not expected to have a significant deleterious impact on localization because the differences in ray travel by incorrectly assuming the hydrophone is 8 m deeper than it actually is will be small compared to the expected measured RAT uncertainty..

Since the bathymetry around ACO is not perfectly flat, the range-independent assumption will impart a bias to estimated source range. For example, consider a whale 50 km from ACO in water 4800 m deep, but with water depth erroneously assumed to be 5000 m. Under such an assumption, even if the bathymetry between the whale and ACO were actually flat, model-predicted RATs for a source 50 km from ACO will appear larger than they should due to the increased water depth (all other things being equal, increasing water depth increases the ray arrival time, and the more times a given ray reflects off the bottom the greater the increase in arrival time). Section 3.5 presents results where this bias is demonstrated using recordings of 296 air gun shots by ACO for which true range is known. To briefly summarize these results here, at long range the actual water depth is shallower than the assumed depth, leading to source range being over-estimated. Similarly, at close range, the true water depth is deeper than the assumed depth, resulting in sources being estimated to lie closer to ACO than they actually do.

Using the same 90 km range cutoff used to calculate average water depth, 95% confidence bounds for water depth around ACO were established. Ultimately, these water depth confidence bounds will be used to assess the expected RAT variability due to water depth variability. The 95% confidence bounds consist of two water depth values based on the set of measured water depth values within 90 km of ACO: one which is shallower than 2.5% of measured 1 km grid spaced measured water depth values and one which is shallower than 97.5% of measured water depth values. This was accomplished by sorting the set of measured water

Confidence Interval	Water Depth [m]
2.5%	4044.3
Average	4667.7
97.5%	4963.7

Table 3.8: Calculated confidence interval water depth values within 90 km of ACO

Water Depth Condition	D Cutoff Range [km]	BS Cutoff Range [km]	BS2 Cutoff Range [km]
2.5%	24.68	73.37	121.3
Average	29.92	88.08	151.3
97.5%	30.55	93.45	142.7

Table 3.9: Model-predicted ray cutoff ranges for each water depth case

depth values from least to greatest and linearly-interpolating to establish the 2.5% and 97.5% values. Table 3.8 shows these water depth values resulting from this analysis, and the average water depth.

After establishing our representative water depth, model-predicted RATs were calculated using BELLHOP. BELLHOP was used rather than RAYFAST due to BELLHOPs greater ease of changing the environmental parameters vs. RAYFAST. The similarity of the eigenray travel times predicted by these two models suggests that this choice of propagation model did not significantly change the results shown in this section.

The measured ACO SSP was truncated or linearly-extrapolated (as appropriate) to account for water depths smaller or larger, respectively, than the water depth at ACO. Hydrophone depth was set to 2 m above the water depth (i.e., either 4042.29 m or 4961.71 m). The same candidate source ranges that were used to calculate the analysis RATs were used to calculate the confidence interval RATs. Figure 3.17 shows calculated RATs for the average (4667.7 m), 2.5% (4044.29 m), and 97.5% (4963.74 m) water depth cases across all candidate source ranges. Pink circles indicate 97.5% water depth RATs, red + symbols denote RATs for the average water depth, and blue x symbols denote 2.5% water depth RATs. As water depth increases, the ray cutoff point (e.g., the range at which direct-path arrivals no longer connects) increases. Table 3.9 summarizes the cutoff ranges for the first three rays. The higher BS2 cutoff range for the average case, compared to the 97.5% case, is likely due to the distribution of ray takeoff angles used for this particular modeling run.

To examine how RATs change with changing water depth, the difference between average RATs and each of the 2.5% and 97.5% water depth cases was calculated. To do so, ranges at which the same set of RATs across all three depth conditions were identified. Next, the absolute value of the difference between the average and percentile depth cases was calculated over this subset of ranges. Figure 3.18 plots these RATs differences for each range in the subset. RAT differences reach a maximum prior to the direct cutoff range, with the range corresponding to the peak for higher order RATs differences being higher for higher order RATs. This phenomenon is likely due to the interaction of RATs becoming more similar as range increases, but the effect of difference in water depth being more important at close ranges when range is not yet large compared to water depth.

Overall, Figure 3.18 demonstrates that, particularly at close range, mismatch between actual and assumed water depth can lead to RATs differences considerably larger than the 0.1 s uncertainty value due to RAT estimation uncertainty from the cross-correlation time series, as described in Section 3.4.6. At medium and large range, these differences diminish but remain significant compared to the RAT arrival time uncertainty.

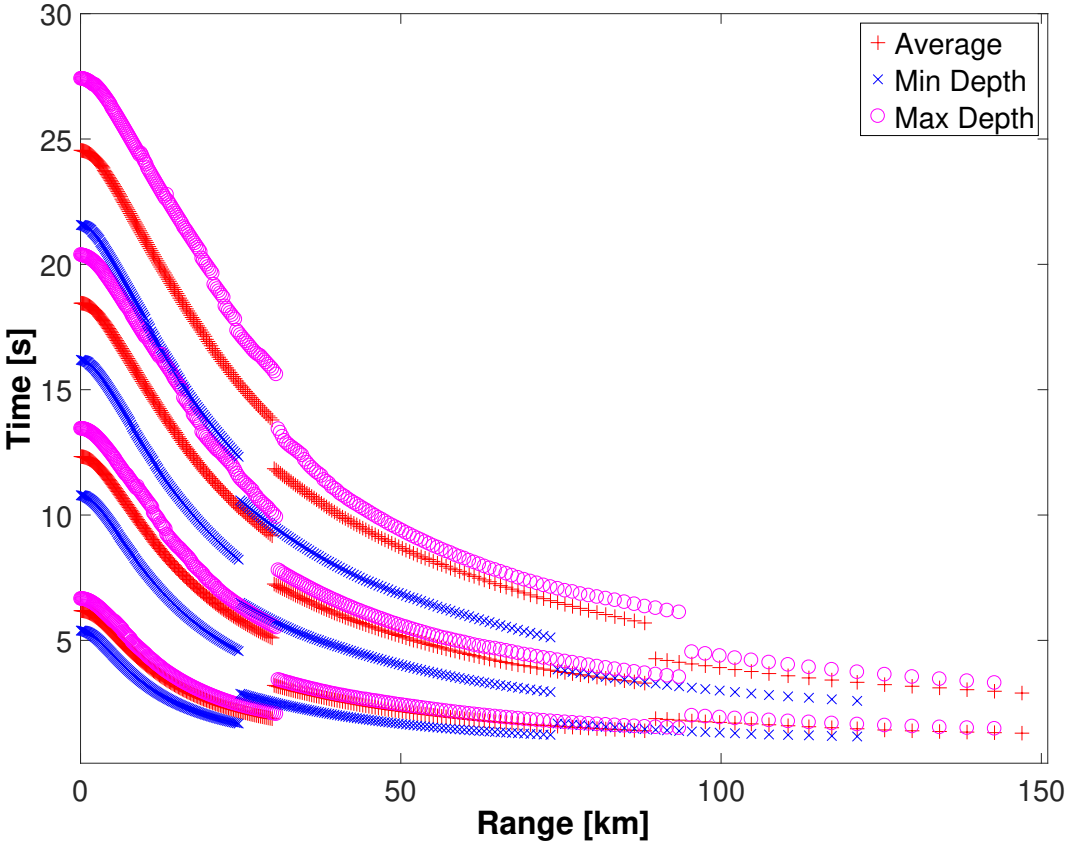

Figure 3.17: Model-predicted relative arrival times each candidate source range for three different range-independent water depths.

The dashed line in Figure 3.18 (1.47 s) is the mean absolute difference between the average and percentile water depth cases, averaged across all RAT orders and source ranges. These RAT difference values at short range, while large, do not express the actual variability expected in measured RATs at these close ranges. This is due to bathymetry less than 30 km from ACO, for example, varying by much less than the confidence interval values listed in Table 3.8. Thus, the results in Figure 3.18 present a pessimistic view on measured RAT variability at close range, although the reliability of this analysis is expected to improve as range increases.

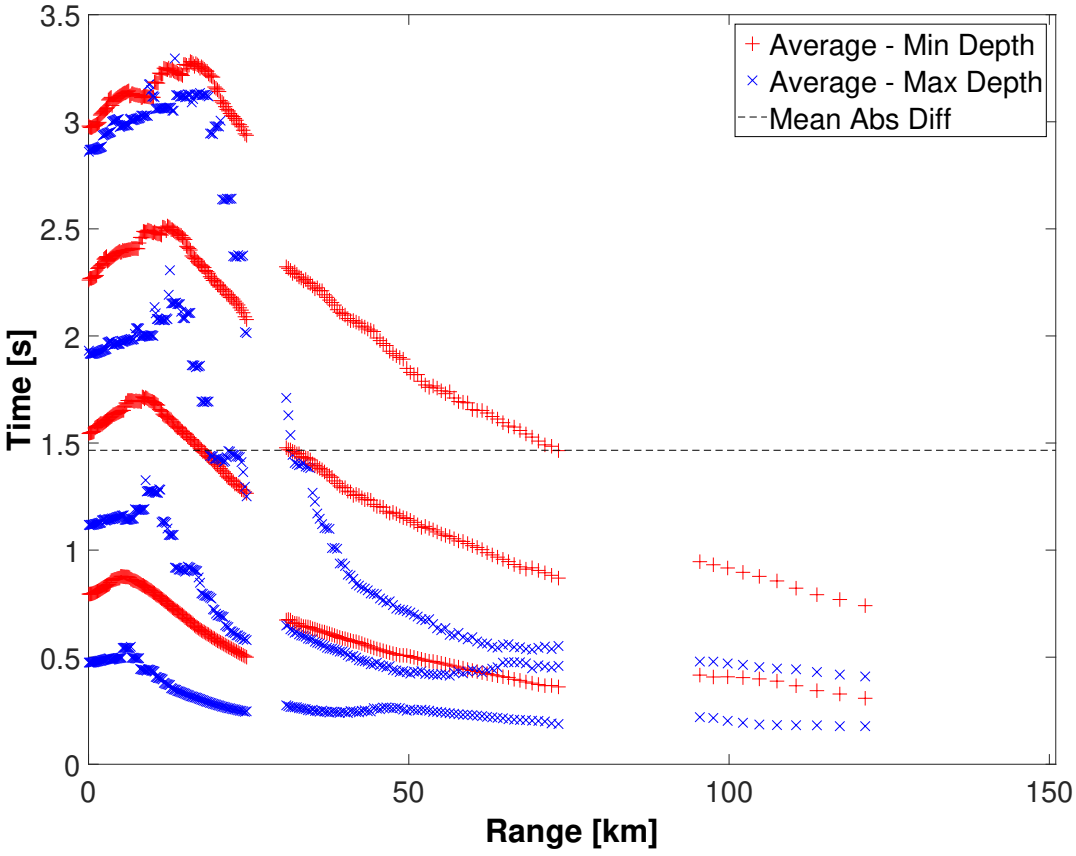

Figure 3.18: Magnitude of the differences between model-predicted relative arrival times each candidate source range for three different range-independent water depths. The Average - Min Depth points plot the magnitude of the difference between model-predicted RATs for the water depth = 4667.7 and 4044.3 m cases. The Average - Max Depth points plot the magnitude of the difference between model-predicted RATs for the water depth = 4667.7 and 4963.7 m cases.

3.4.7 Pruning

The final step in the localization process is a pruning step which helps eliminate false positives from the set of localized sounds. Two different types of pruning are implemented in this work: frequency and phantom. Frequency pruning uses info provided by the detector to distinguish calls of interest from other source waveforms which may yield high confidence locations but which have characteristics incompatible with the sought-after signals. In this work, peak frequency and maximum frequency are used to distinguish fin whale calls from sei whale calls and background noise.

Phantom pruning is used to eliminate calls that satisfy the frequency criteria, but which are part of a call which has already passed the frequency pruning criteria. Due to the similarities in the received waveforms from different eigenray paths, the matched filter detector is expected to identify higher order reflected arrivals in addition to the earliest arrival for a given call. Phantom pruning reduces the likelihood of a single call being present more than once in the final set of location estimates, which in turn makes estimates of calling rates and inter-call interval based on this set more accurate. This type of pruning is sometimes referred to as phantom-track pruning, given the predisposition of these superfluous detections to yield a smooth track at the wrong range which mirrors the true track.

Based on Silver [2014], frequency pruning is performed such that only calls whose frequency characteristics match both these criteria are kept:

- Peak frequency (f_{peak}, the frequency of highest power spectral density value energy within the detected call) less than 19.5 Hz
- Stop frequency (f_{stop}, the lowest frequency greater than f_{peak} whose PSD value is 6dB lower than the peak PSD value) less than 24 Hz

One of the primary motivations for frequency pruning is the observation that sei whale and fin whale vocalizations in Hawaiian waters can be similar (Rankin and Barlow [2007]). The f_{peak} and f_{stop} criteria under frequency pruning were selected to help discriminate between fin whale 20 Hz calls and sei whale 39-21 Hz calls.

Additionally, phantom pruning only keeps calls that both pass frequency pruning and whose arrival times are all more than 1.25 s different than all previous calls that passed frequency pruning. This is done both to reduce the false positive rate of the detection set, and to eliminate phantom tracks from the set of locations. The value of 1.25 s was selected because it is slightly smaller than the smallest model-predicted RAT value considered in this work. No two valid arrivals are expected to be closer than 1.25 s in the measured data. An example of a phantom track from the measured data results is seen in Figure 4.13 in Section 4.2.

To apply these pruning criteria, data from the detection and localization steps are used. Specifically, the measured f_{peak} and f_{stop} values for each detection are compared against the frequency pruning criteria. Those detections that pass these criteria are then evaluated using phantom pruning. During the localization step, the subset of measured RATs that best matches to a set of model-predicted RATs is stored. Only calls whose subset of measured RATs satisfy the phantom pruning criterion are considered to have passed both the frequency and phantom pruning steps.

Appendix B presents a simulation study demonstrating the effectiveness of these pruning approaches.

3.5 Validation Study: Airgun Survey with Known Ranges

In the previous section, we demonstrated the performance of the range estimation approach in simulated deep-ocean acoustic conditions. As useful as simulations are to help us examine the performance of a system in realistic ocean conditions, there is no substitute for real-world, measured, ground-truthed data. In this section, the performance of the RAT-based range estimation approach will be demonstrated using a controlled underwater acoustic source where the distance between source and receiver is known.

3.5.1 Measured Data Description

From September 11 - October 21 2018, a geophysical research cruise (Shillington et al. [2018]) took place on the R/V Marcus G. Langseth near Hawaii. On October 5-6, an operating airgun array was towed past the ACO in a roughly south-southwesterly course. From 21:06:47 (Hawaii Standard Time) on October 5 to 12:16:14 on October 6 the airguns fired approximately every 3 minutes, yielding 296 individual airgun shots. Figure 3.19 shows a map of the ocean region around the ACO with the portion of the Langseth's path containing these 296 shots highlighted.

The first shot considered in this analysis occurs 114.31 km from ACO, the closest shot to ACO occurred at 1201h (October 6) at a range of 12.24 km, and the final shot occurs 12.37 km from ACO. Start and end points for this course plot include the point of closest approach to the hydrophone along with the previous 15 hours of travel as the R/V Langseth approached the ACO.

During this time period, two hydrophones were installed at the ACO. One, an icListen acoustic recorder (Ocean Sonics, Great Village, Nova Scotia) recorded all 296 shots at a sampling frequency of 32 khz. Figures 3.20 and 3.21 show the spectrogram for two 5-minute recordings made by this icListen; a far range shot in Figure 3.20 and a close range shot in Figure 3.21. A highpass filter, with a cutoff frequency of 10 Hz, was applied to these data prior to plotting to eliminate low-frequency noise present in the recording.

Both Figure 3.20 and Figure 3.21 show that the airgun source signal arrives at the hydrophone as distinct eigenray arrivals with no apparent evidence of acoustic dispersion. The difference in the RATs between the far and close range airgun shots illustrates how RATs change with range. To more clearly visualize how RATs change with range, the received air gun waveforms are time-aligned and plotted. To prepare this figure, the lowest order arrival path shared by all 296 shots is identified. This determination requires knowledge of the maximum shot range, and acoustic propagation modeling to identify the range at which eigenrays stop connecting. Given that the maximum shot range for this transect is 114 km, and given that the BS2 arrival is predicted to have a cutoff range of 151 km, the BS2 arrival was identified as the lowest order arrival shared by all shots and thus the arrival which would be time-aligned across all 296 shots.

Next, the arrival path labels for each shot are determined. The closest shot to ACO has a range of 12.2 km. According to propagation modeling, the first arrival to reach the hydrophone from this shot the direct path. The near-surface source, ocean bottom receiver, and duration of the air gun source waveform indicate that the next arrival to reach the hydrophone will be BS, followed by BS2, BS3, and so on.

Once the first shot's arrival paths are labeled, the next closest shot's arrival paths are then labeled. The close time and geographic spacing of consecutive shots, and the relatively flat bathymetry about ACO, means that RATs will smoothly and slightly vary from one shot to the next. Thus, arrival path labels from one shot can help pick those of the next shot. In this fashion, arrival path labels for all 296 shots are identified.

Once BS2 arrival times are picked out for each shot, a time-series window around each BS2 arrival time is extracted. Each window starts 35 seconds before the BS2 arrival, and ends 80 seconds after. These bounds are selected to allow any geoacoustic arrivals to be plotted alongside the water-borne arrivals, and for very high order arrival paths to also be present. These extracted time series windows are stacked and time-aligned so that the BS2 arrival for each shot falls at Relative Time ≈ 0. Figure 3.22 shows the recorded time series for each of the 296 air gun shots as a function of the true shot range. There are 296 rows in this figure; one for each air gun shot. Arrival path labels are also included. As expected, RATs vary smoothly as range varies, supporting the assertion that by measuring RATs we can, in turn, estimate range. Note that the received levels for the D and BS arrivals decrease as range increases toward their eigenray cutoff ranges. This attenuation, brought about primarily by vertical refraction, creates conditions within which the received levels of these two eigenrays is less than some higher order arrivals. The permissive detector developed for this work aids in picking up these weak arrivals, thereby improving the overall reliability of the localization approach.

While not directly pertinent to this work, two additional features in the air gun shot figure deserve emphasis. Figure 3.23 highlights the areas in question. In region 1, geoacoustic arrivals are seen reaching ACO in advance of the waterborne arrivals. These are deemed sediment-borne arrivals, rather than water-borne, because as range increases their arrival time also increases (which would be observed if those arrivals spend significant time traveling through the higher-speed sediment). While there is information on both source location and the ocean bottom sediment layer structure encoded in these sediment-borne arrivals, the use of this information is beyond the scope of this work.

Region 2 highlights a set of high order arrivals with particularly high received levels. In principle, these high order arrivals could also be used to carry out localization, although the fact that different arrivals show particularly high received levels as range increases would complicate the process of incorporating them into the model-predicted RAT set. The band of relatively high received levels among high order eigenray paths may be caused by a reflection coefficient at the ocean:sediment interface that is particularly high for a certain subset of grazing angles.

The smooth variation in RATs as a function of range, and the faint arrivals shown in regions 1 and 2, are apparent in the combined shot plot in Figure 3.22, but these features are less apparent when considering one shot at a time.

Figure 3.19: Map showing the ALOHA cabled observatory location (green marker, "ACO") and the start and stop locations (yellow markers, "Dunn_Start" and "Dunn_End") for the air gun data used to validate the range estimation approach in this work. Map: Google Maps, 2022

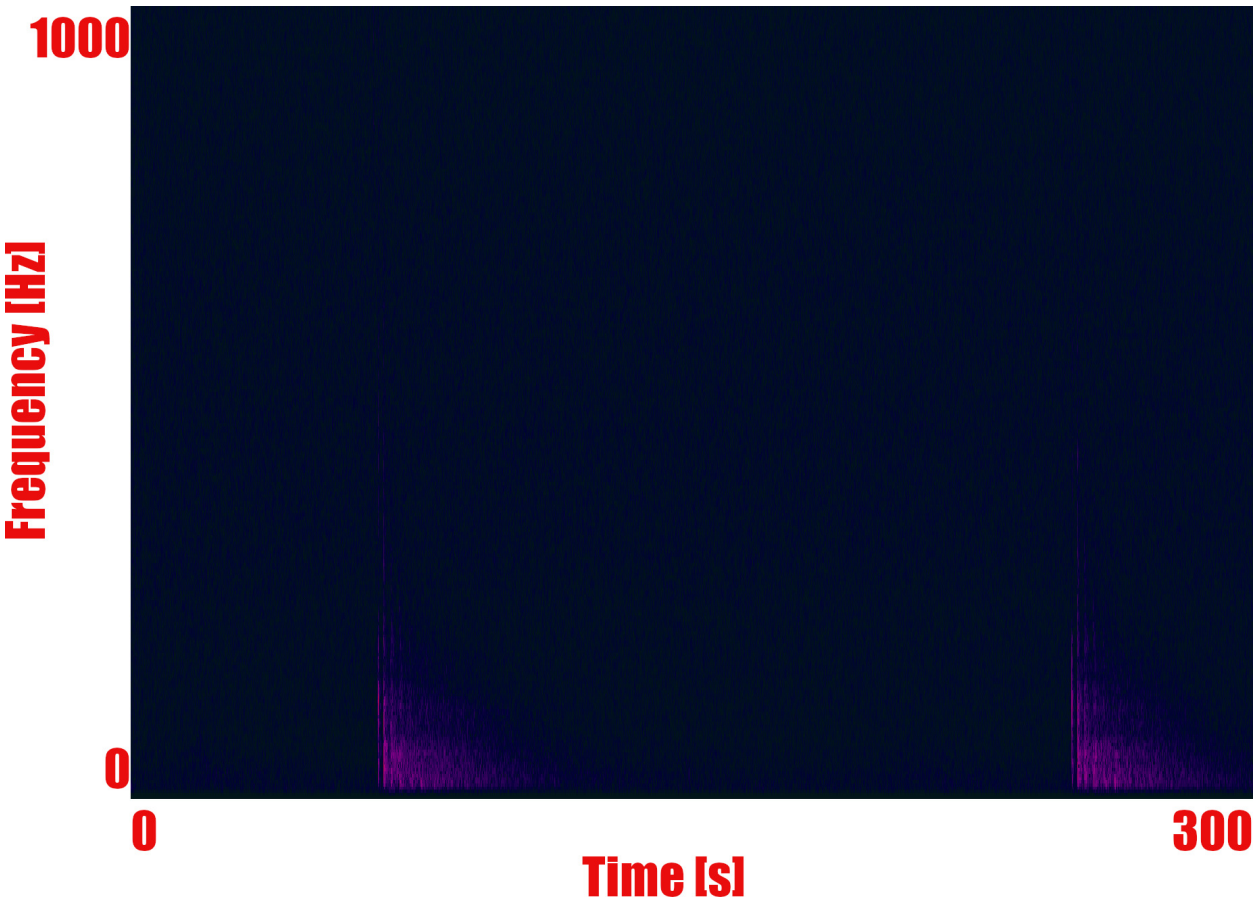

Figure 3.20: A spectrogram, in the 0-1000 Hz band, for a 5-minute underwater acoustic recording from ACO. The received signal from two consecutive air gun shots at approximately 110 km from ACO are shown.

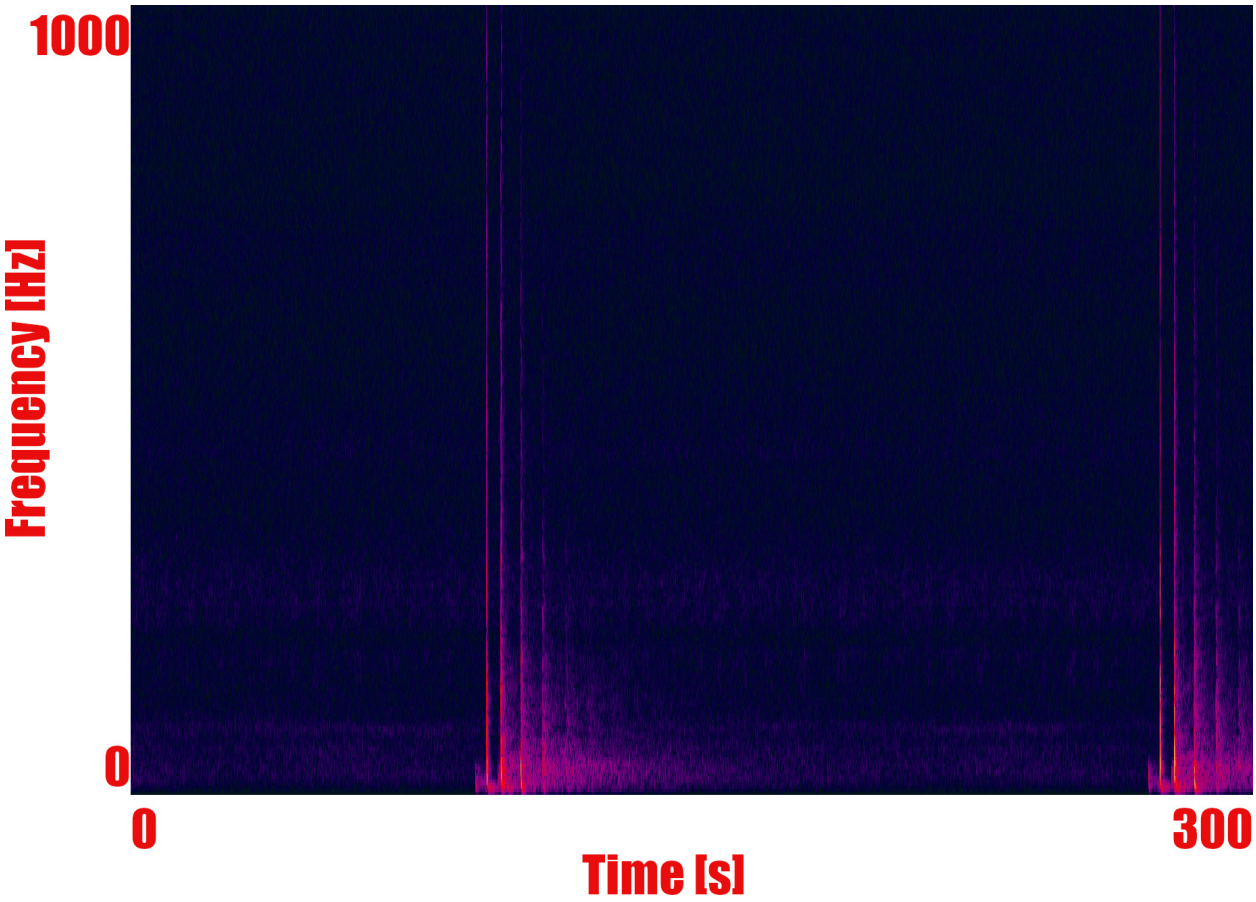

Figure 3.21: A spectrogram, in the 0-1000 Hz band, for a 5-minute underwater acoustic recording from ACO. The received signal from two consecutive air gun shots at approximately 12 km from ACO are shown.

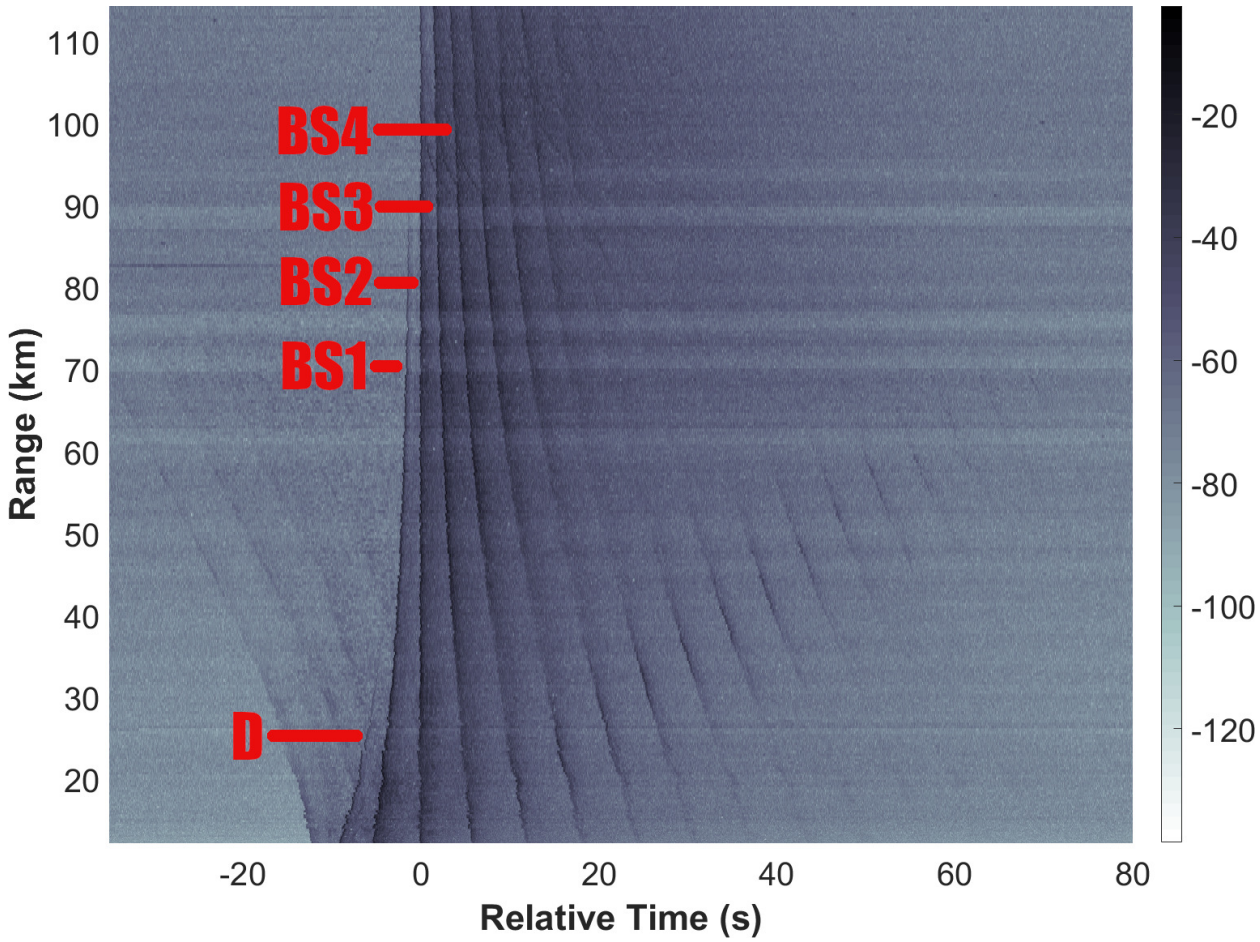

Figure 3.22: The magnitude of the received waveform for the 296 airgun shots, plotted against the horizontal range between the airgun array and ACO. The waveforms are plotted in log scale to more readily show eigenray received level variation as range changes.

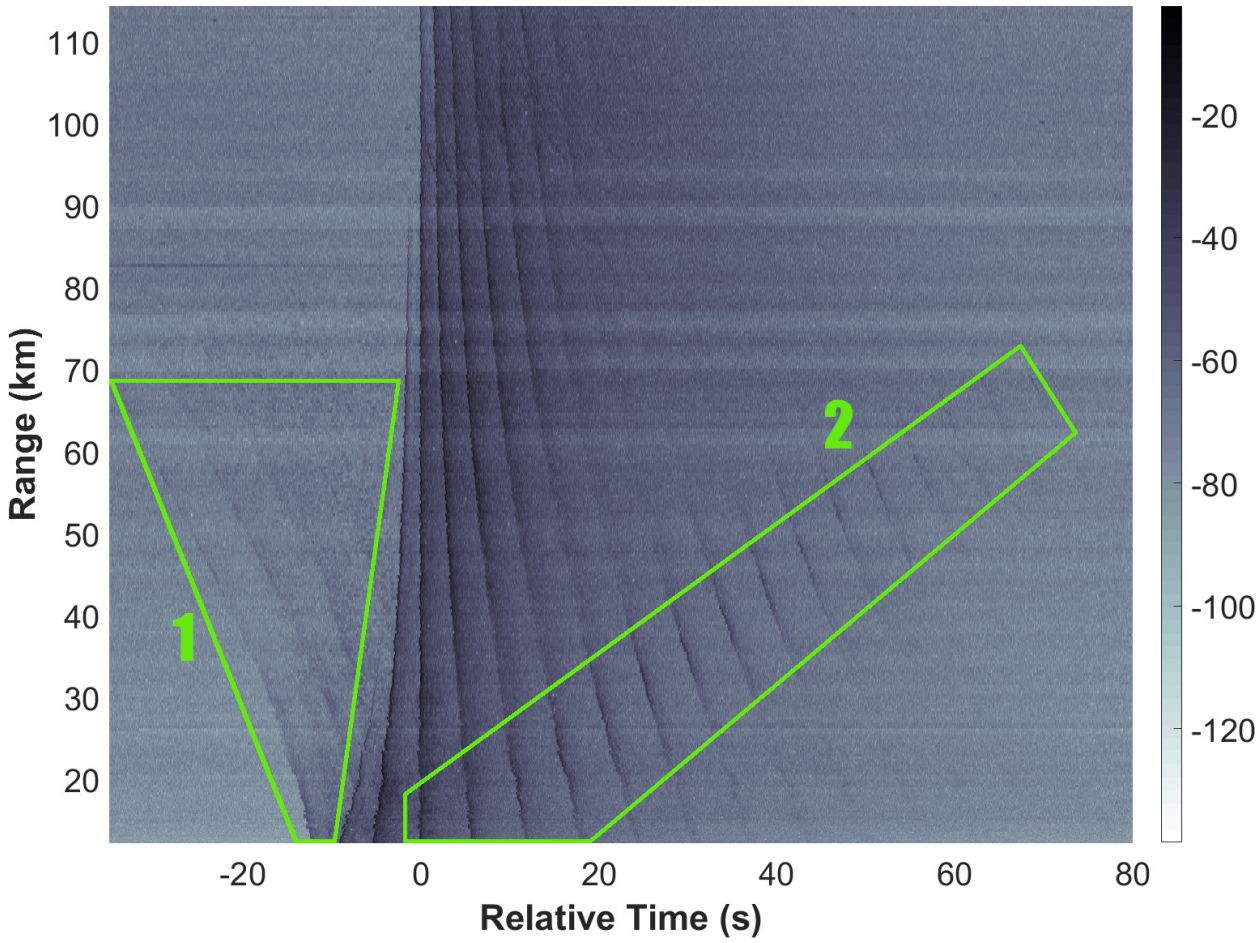

Figure 3.23: The magnitude of the received waveform for the 296 airgun shots, plotted against the horizontal range between the airgun array and ACO. The waveforms are plotted in log scale to more readily show eigenray received level variation as range changes. Region 1 shows acoustic energy reaching the hydrophone that partially traveled through the ocean bottom. Region 2 shows a reflection coefficient effect whereby the received level is particularly high for rays within a certain set of grazing angles with the ocean bottom.

Peak to Peak Received Level Variation with Range

As is described in Section 3.4.4, the automated detector used here was intentionally designed to be permissive. In other words, the detector has a high probability of detection while having a high false positive rate. An important reason for preferring a permissive detector comes from a visual examination of the recorded data. As figures in the previous section demonstrated, contrary to what would be expected from a purely geometric spreading perspective, the earliest arrival of the received airgun signals does not necessarily have the highest received level. Mcdonald et al. [1995] similarly recorded air gun shots using a seafloor recorder in deep water, and also observed cases where the received level of the earliest arrival is less than that for some higher-order arrivals. This phenomenon is attributed to refraction playing a more significant role in the received level of lower order arrivals than higher order ones, particularly as range increases toward the cutoff range for these eigenrays. Figure 3.24 shows a comparison between range and peak to peak received level for ACO air guns and air gun data described in Mcdonald et al. [1995]. The left subplot shows uncalibrated hydrophone data from the 296 airgun shots used in this work, while the right subplot shows calibrated hydrophone data previously described in Mcdonald et al. [1995]. The Mcdonald et al. [1995] results include D through BS3 arrival paths, while the ACO results include results from the D through BS4 arrival paths. For both subplots, the D arrival is loudest at close range, with the received level decreasing sharply as range increases. The use of uncalibrated acoustic data doesn't complicate the comparisons between the ACO and McDonald data because differences in source size and propagation conditions between the two experiments already preclude comparisons based on received levels. Instead, relative peak to peak levels (e.g., calculated as the range at which peak to peak received levels for different arrival paths are equal) can be compared between the two experiments.

Table 3.10 compares the range at which the received peak to peak values for two arrival paths cross in the results shown in Figure 3.24. For every pair of arrival paths in these tables, the McDonald crossing occurs at a closer range (or no crossing exists, for the cast of the BS2:BS3 pair) than the ACO crossing. Differences in propagation conditions between the two data collection sites may explain these differences. The ACO data was recorded by a seabed acoustic recorder in 4.7 km deep water 100-200km north of O'ahu, while McDonald's data were recorded at the seabed in 2.4 km deep water 500 km west of Oregon. Differences in bottom composition between the two sites may also help explain differences in attenuation for different ray paths; McDonald's site may have lower reflection coefficients. Another difference is that, for a given range and taking into account differences in water depth between these two sites, the eigenrays arriving at ACO will have a larger grazing angle than McDonald's, leading to differences in reflection coefficient even if the bottom composition is similar.

Since the earliest arrival path can not be reliably assumed to have the highest received level, and if we

	ACO			**McDonald**		
	D	BS	BS2	D	BS	BS2
BS	14.8			7.6		
BS2	23.3	34.3		10.0	15.9	
BS3	25.3	47.0	15.4	11.8	19.0	no cross

Table 3.10: Multipath peak to peak received level intersection ranges in km

assume that in any detection approach (either automated or manual) the probability of detecting an arrival is directly related to the received level, then the early, weak arrivals are less likely to be detected than later, stronger arrivals. As is discussed in Section 3.4, correct localization using the approach presented in this work requires that the earliest connecting arrival (i.e., the lowest order arrival path to reach the hydrophone that doesn't fail to connect due to refractory effects) in the model-predicted RATs also be among the detection results from the measured data. Thus, the detection threshold was designed to be permissive with the intention of creating a set of detected 20 Hz fin whale calls that reliably contain the first arrival for each call.

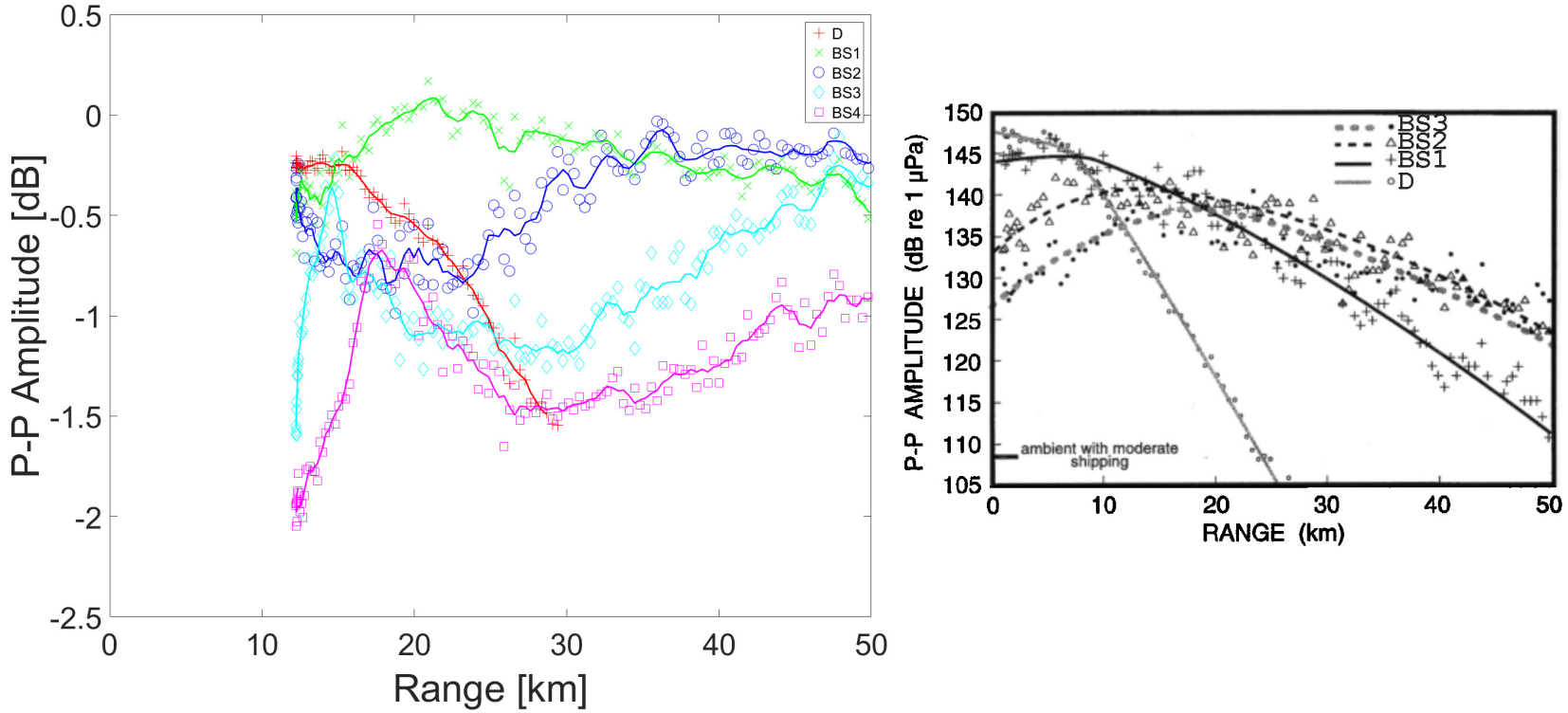

Figure 3.24: ACO (left) and Mcdonald et al. [1995] airgun peak to peak airgun amplitude as a function of range and arrival path.

3.5.2 Air Gun Range Estimation

The first step in demonstrating the validity of our proposed localization approach to calculate the true range between ACO and each air gun shot. The latitude and longitude for ACO is readily available from the ACO website, and the shot logs for the Langseth airgun trial contain GPS coordinates for the airguns for each of the shots. Horizontal great circle distance between each shot location and ACO are computed using the Haversine formula (Shylaja [2015]) and an assumed Earth radius of 6371 km.

3.5.3 Automated Detector and RAT Estimation for Validated Air Gun Analysis

Although the source spectra for the airgun shots and for the 20 Hz fin whale calls are not entirely dissimilar, the matched filter-based automated detector designed for these fin whale calls is not necessarily a reliable airgun shot detector. Additionally, the intention of this analysis of the airgun survey data is to test the ranging algorithm, rather than the entire data processing chain. Thus, to isolate ranging algorithm performance from that of detector and automated RAT estimation performance, RATs for each of the 296 airgun shots were manually identified through visual inspection of the waveforms and spectrograms. Care was taken to consistently identify the same waveform feature to reduce RAT variability due to this visual inspection process. Also, only RATs expected to align with the expected RATs at that particular range were manually identified.

In addition to working with the manually-identified RATs, the manually-identified detections will also be processed using the automated RAT estimation code in a separate analysis. Differences in localization performance between these two cases will be compared in order to explore the significance of automated RAT estimation on localization performance.

Having established that the RATs of the air gun shot waveforms vary smoothly with range, Sections 3.5.4 and 3.5.5 present analyses that explore the behavior of the localizer when high-quality RATs and high-quality detections, respectively, are used. No automated detector is used in either of these analyses, as the purpose of this analysis is to evaluate the performance of the localizer. Additionally, while a matched filter tuned to detect fin whales may also yield useful detection results for air gun shots (due to the overlap in their frequency bands), such an 'off label' use of the fin whale detector could introduce error into perceived localizer performance. As well, from the onset of this work it was expected that a matched filter would yield good results when seeking stereotyped fin whale 20 Hz calls, which made an investigation into the overall matched filter performance in seeking either fin whale or air gun sources less important and interesting than evaluating localizer performance.

3.5.4 Experiment #1: Known RATs

In Figure 3.25, the flow chart of the overall range estimation approach is depicted. Each box represents a different data processing step. For Experiment 1, the data have already been collected, signal detection and RAT estimation is performed manually, and manually-picked RATs are inserted into the sequence immediately before the fourth step "Compare Measured and Model-Predicted RATs". The objective of the experiment is to show how well the localizer works with measured data when the RATs expected by the localizer are known to have been correctly identified. To make this test more realistic, noisy RATs were added to each set of measured RATs prior to passing them to the localizer.

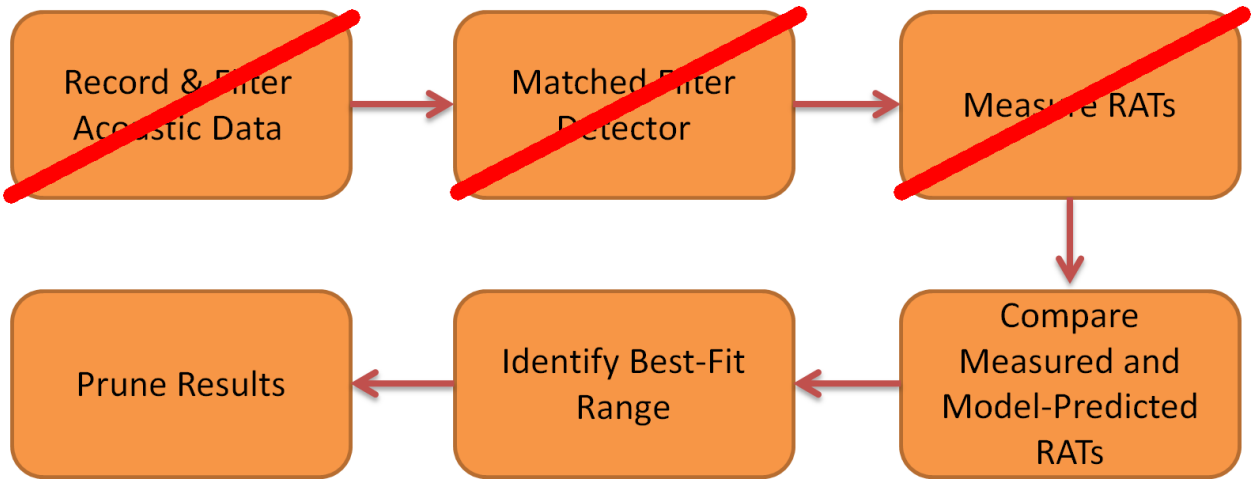

Figure 3.25: Processing steps used to demonstrate localization performance on the ACO airgun data when manually-picked RATs are used.

To manually estimate RATs, arrival times were visually and manually picked directly from the recorded acoustic recording, after which RATs were calculated based on the hand-picked arrival times. This is a labor-intensive, yet highly accurate, process. Nevertheless, only an afternoon's worth of work was needed to manually identify RATs for the 296 airgun shots. To make the test more realistic, noisy RATs (in addition to the true and expected RATs) are manually picked for each of the 296 airgun shots at the same time that the actual RATs are identified. Thus, for each shot, a total of 9 RATs were manually picked; those that would later be expected by the localizer, and enough noisy ones such that a total of 9 RATs are picked for each shot. For example, for an air gun shot near 60 km, the manual RAT estimation process includes the BS2-BS, BS3-BS, and BS4-BS4 RATs. The remaining six RATs picked for this shot are randomly picked (by the manual operator) within the time window that RATs are expected in the model-predicted RATs set.

After the measured RATs are identified, they are passed to the fifth step. Given that frequency pruning is done to separate fin and sei whale calls, and that phantom tracks only appear when higher order RATs are mistakenly processed as the earliest arrival of a given call, no pruning step was necessary.

Figure 3.26 plots true (red "x" symbols) and estimated (blue "+" symbols) for each of the 296 air gun shots. Good agreement between the estimated and true ranges demonstrates that the localizer effectively performed range estimation on the air gun shots. Given the spectral and temporal similarities between air gun shots and fin whale 20 Hz calls, and the fact that the air gun data was collected in the same environment as the fin whale calls were, these results support the claim that range estimation for fin whale calls can be effectively carried out using the method presented herein.

The slight difference in slope between the estimated and true range trends is due to a mismatch between the modeled and actual bathymetry. As the air gun traveled toward ACO, water depth decreased; at the beginning of the transect (114 km from ACO), water depth is approximately 4438 m, while at the transect's closest point of approach to ACO (12.23 km from ACO) the water depth is 4760 m. Meanwhile, the acoustic propagation model used to calculate the model-predicted RATs assumed a uniform, range-independent water depth of 4667 m. Effectively, for the air gun transect, water depth at long range was overestimated, while water depth at close range was underestimated. Thus, at long range measured RATs are closer together

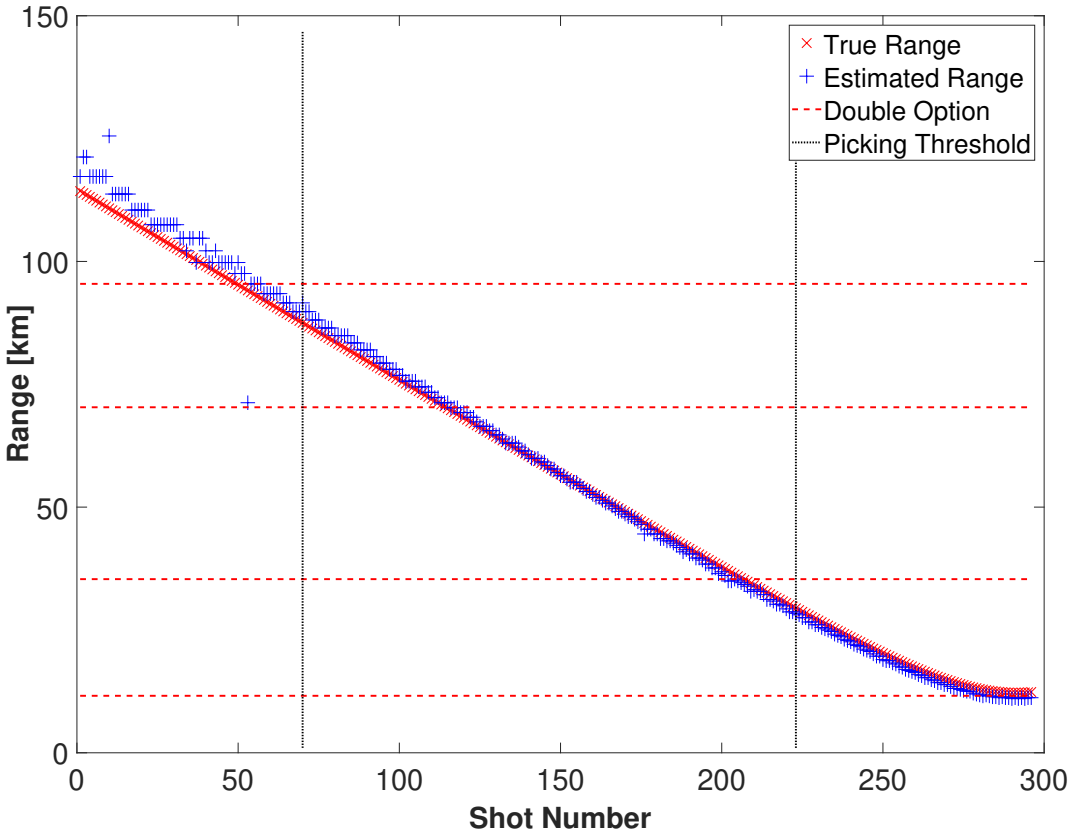

Figure 3.26: True (red "x" symbols) vs. estimated (blue "+" symbols) range for manually estimated RATs from the 296 air gun shot data set. Two pairs of dashed lines delineate the close:medium and medium:long double option zones.

than the model would predict at that range, while at short range measured RATs are further apart than expected. Since the differences between RATs (e.g., $(BS3 - D) - (BS2 - D) = (BS3 - BS2)$) decreases as range increases, the RAT biases due to water depth mismatch will result in overestimating range at long range and underestimating range at short range. If the uniform water depth used to create the model-predicted RATs were changed, it is expected that the range at which the transition from underestimation to overestimation will also change.

Figure 3.27 plots the difference between true and estimated range for each of the 296 air gun shots. True minus estimated ranges are shown by the red "x" symbols, while the mean range difference (-0.564 km) across all shots is shown by the blue line. Diagonal arrangements of range differences for low shot numbers are due to coarse resolution of the candidate source range set at long range; effectively, multiple air gun shots at different ranges are assigned to the same candidate source range during localization.

Figure 3.28 plots the ambiguity vector for each of the 296 shots. Shot #1 has a true range of 114.3 km, while Shot #296 has a true range of 12.37 km. Alternating regions of high and low ambiguity within the two double option zones are present because the consolidation step (which, for each unique candidate range

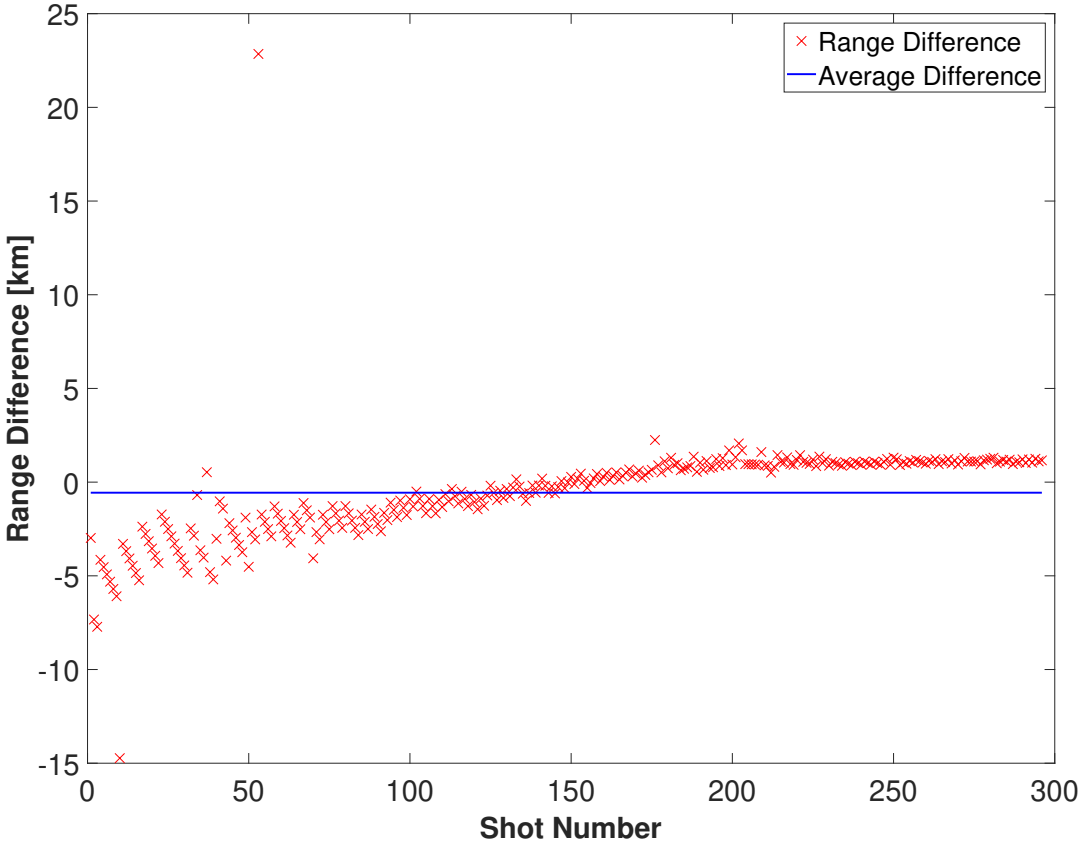

Figure 3.27: The difference between true and estimated range for each of the 296 airgun shots. The average range difference across all shots (-0.564 km, blue line) is also plotted.

in a double option zone, picks the highest overall ambiguity value) has not been performed on the ambiguity vectors. Figure 3.29 shows the ambiguity vectors after consolidation. Significant side lobes in the ambiguity vectors are due to partial matches with model-predicted RAT sets at consistent higher and lower ranges. These secondary peaks were lower than the expected peak for almost all shots.

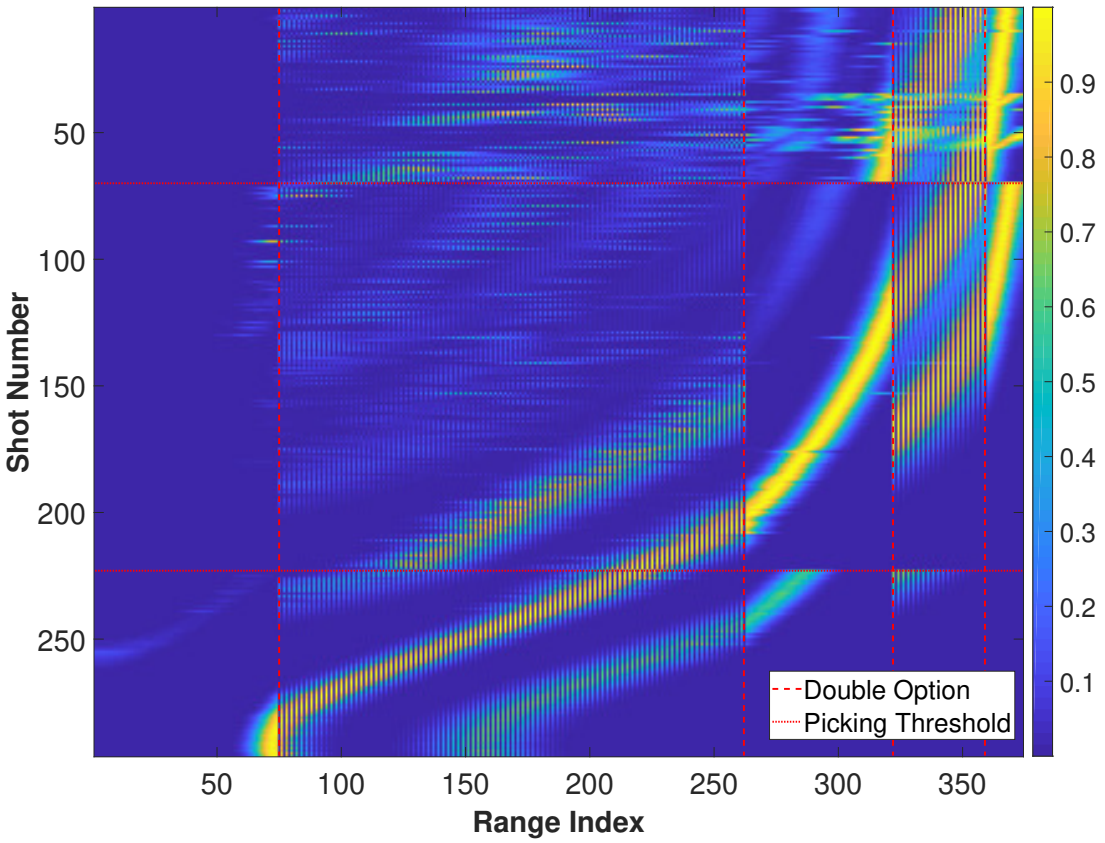

Figure 3.28: The ambiguity vector for each of the 296 airgun shots. Yellow indicates a better match between measured and model-predicted RATs. Pairs of vertical red dashed lines delineate the boundaries of the double option zones. Red dotted horizontal lines indicate changes in the number of RATs picked from the underlying shots: for shots 1-70, BS2 is the lowest order arrival path, for shots 71-223 BS is the lowest order, while for shots 224-296 B is the lowest order arrival.

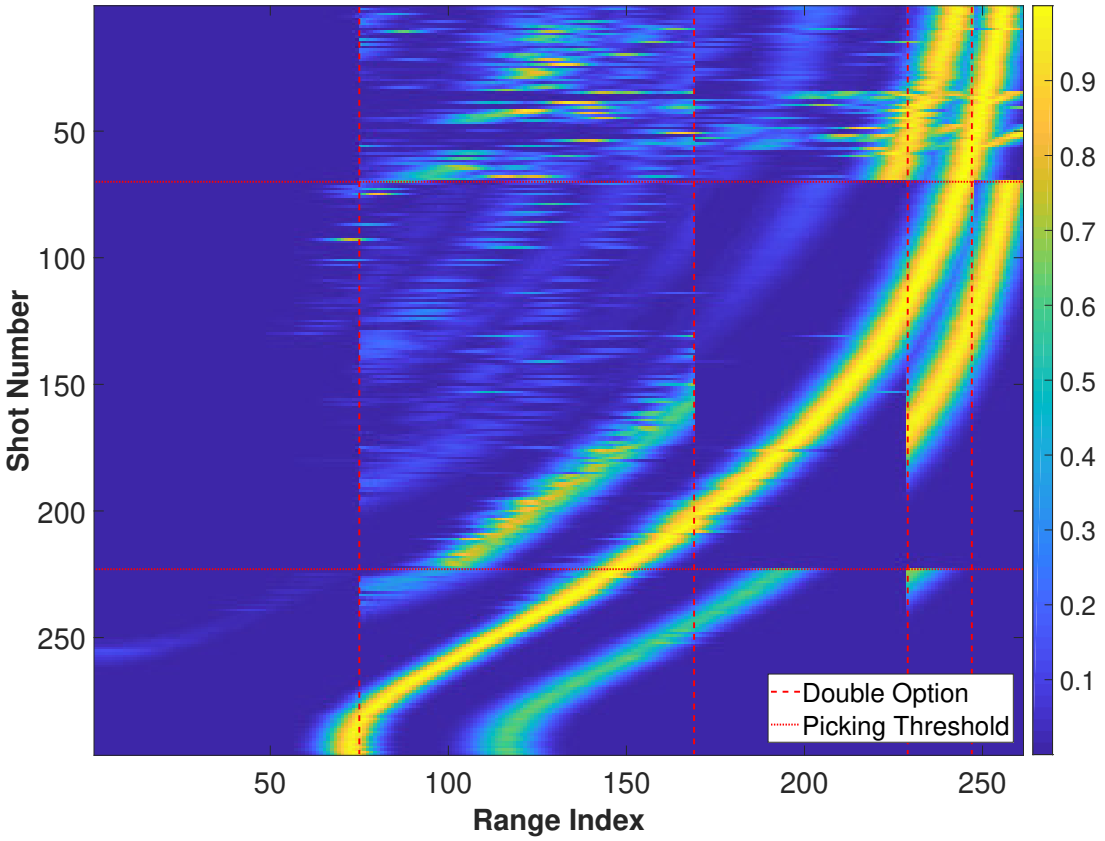

Figure 3.29: The ambiguity vector for each of the 296 airgun shots after the double option zones have been consolidated (i.e., one ambiguity vector for each unique candidate source range). Yellow indicates a better match between measured and model-predicted RATs. Pairs of vertical red dashed lines delineate the boundaries of the double option zones. Red dotted horizontal lines indicate changes in the number of RATs picked from the underlying shots: for shots 1-70, BS2 is the lowest order arrival path, for shots 71-223 BS is the lowest order, while for shots 224-296 D is the lowest order arrival.

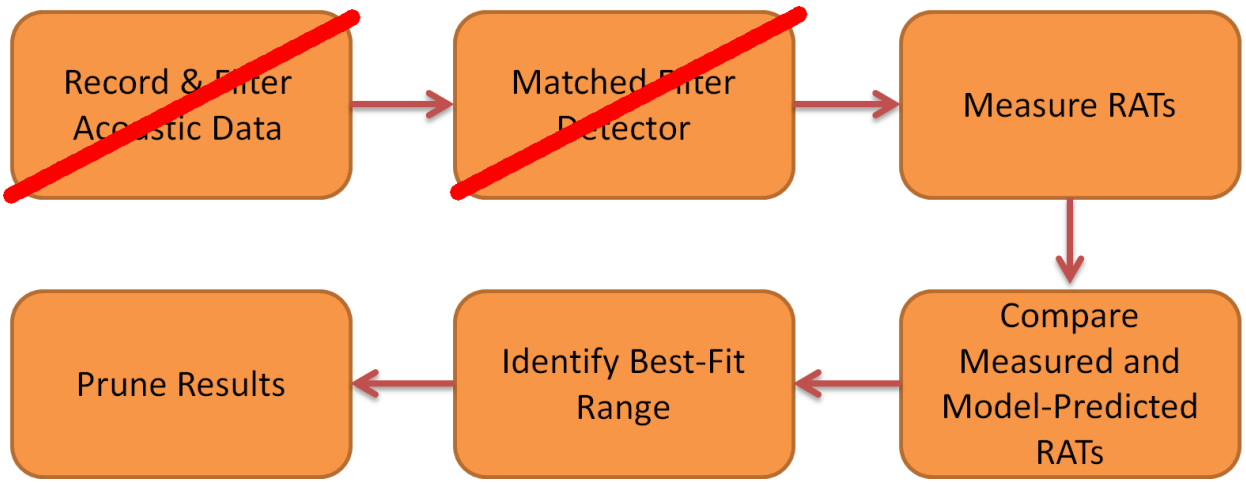

Figure 3.30: Processing steps used to demonstrate localization performance on the ACO airgun data when manually-picked detections are used.

3.5.5 Experiment #2: Known Detections

In Section 3.5.4, the localizer was shown to effectively perform range estimation when the RATs for each shot were known to incorporate all the relevant arrival paths for a given range plus some noisy ones. In this section, automated RAT estimation in addition to automated range estimation is carried out on the same acoustic data that is processed in the previous section. This is done to examine what errors the automated RAT estimation process incorporates into the range estimation process, since the only difference between the two experiments is that manual RAT estimation is performed in Experiment #1 whereas automated RAT estimation is done in Experiment #2.

To improve the consistency and comparability of the results from the two experiments, the manually-picked arrival times from Experiment #1 were used as the input to the automated RAT estimator. Effectively, these manually-picked times are treated as the output of the detection phase. Figure 3.30 shows the processing flow chart for this experiment, with the manual detections from Experiment #1 being inserted into the processing chain before Step 3. Thus, a total of 2960 detection times were passed to the automated RAT estimator; 10 for each shot. The earliest arrival for each shot is guaranteed to have a detection that could yield a useful range estimate for that shot, since in every case it is the time of arrival of the earliest arrival path of an airgun shot. As is seen with the automated fin whale detector output, higher order arrivals will also be present and could yield high ambiguity, yet false, range estimates due to being erroneously processed as an earliest arrival.

First Arrival Range Estimation Analysis

Since the first arrival for each shot has exactly one detection within the overall set of 2960 detections from the air gun data set, a direct comparison between the Experiment #1 results is possible. Figure 3.31 plots true (red "x") and estimated (blue "+") ranges for the detection corresponding to the first waterborne arrival for each airgun shot. This figure intentionally excludes the other 9 detections from each shot, which renders it a somewhat artificial comparison. However, by selectively looking at the same detections processed in Experiment #1, a clearer understanding of the role the automated RAT estimator plays on localization accuracy is possible than if all 2960 detections were considered here. Black vertical dashed lines separate regions with the same manual analysis picking behavior. In Region 1, BS2 is the lowest order arrival picked from the air gun data. Similarly, BS and D are the earliest picked arrivals in Regions 2 and 3, respectively. The change in picking behavior was designed to take place within a double option zones, so track continuity should not be effected. Double option zone boundaries are shown with horizontal red dashed lines.

At long ranges (ranges exceeding the medium:long range double option zone), estimated ranges show the same overestimation tendency shown in Experiment #1, with some isolated cases where the estimated range is lower than the true range. These long range discrepancies are due to noise in the automated RAT estimation process yielding a higher ambiguity at a false range than the closest candidate source range to the true range value. Despite these few discrepancies, the estimated range for each long-range shot remains close to the true value. One possible explanation for the greater proportion of shots at long range whose estimated ranges exhibit significant deviation from the true range is that RAT differences decrease, and the rate at which those differences decrease, as range increases. Thus, small perturbations in the locations of local maxima in the cross-correlation envelope function used for RAT estimation can lead to greater range errors than those same perturbations at closer range.

At medium range (ranges between the two double option zones), the range bias is small due to a better match between modeled and actual water depths. The match between true and estimated ranges is good at medium range.

Estimated range for shots in the close:medium double option zone can show an offset from the true range values. Refraction and reflection coefficient effects on the received levels of low and high order arrivals, respectively, for short range sources complicates the automated estimation of RATs for these short range shots. Figure 3.22 shows that as range increases, the direct-path received level gradually decreases until it is no longer visible. The abrupt change in true vs. estimated range difference when D begins to be picked (around range $\approx$ 29 km on Figure 3.31) is caused by a weak D arrival yielding weak cross correlation envelope peaks. This is shown in the bottom subplot of Figure 3.32. The difficulties in range estimation near 30 km from ACO are further exacerbated by a weak BS4 arrival due to reflection coefficient effects. Since the localizer step was designed to be resilient to noisy measured RATs, but will yield reliably poor results if the set of measured RATs does not include all the model-predicted RATs at the true source range, the combination of a weak D and weak BS4 arrival mean that the BS4-D RAT is difficult to pick out of the cross-correlation envelope time series. Thus, in some cases where D and BS4 received levels are weak, the localizer prefers imperfect 3 or 2-RAT matches at higher range over ranges near the true range. This is because the true range requires a BS4 arrival (the propagation model predicts it will be present) but BS4 is difficult to detect at some ranges.

At even closer ranges ($\approx$ 20 km and less), a related phenomenon makes range estimation problematic.

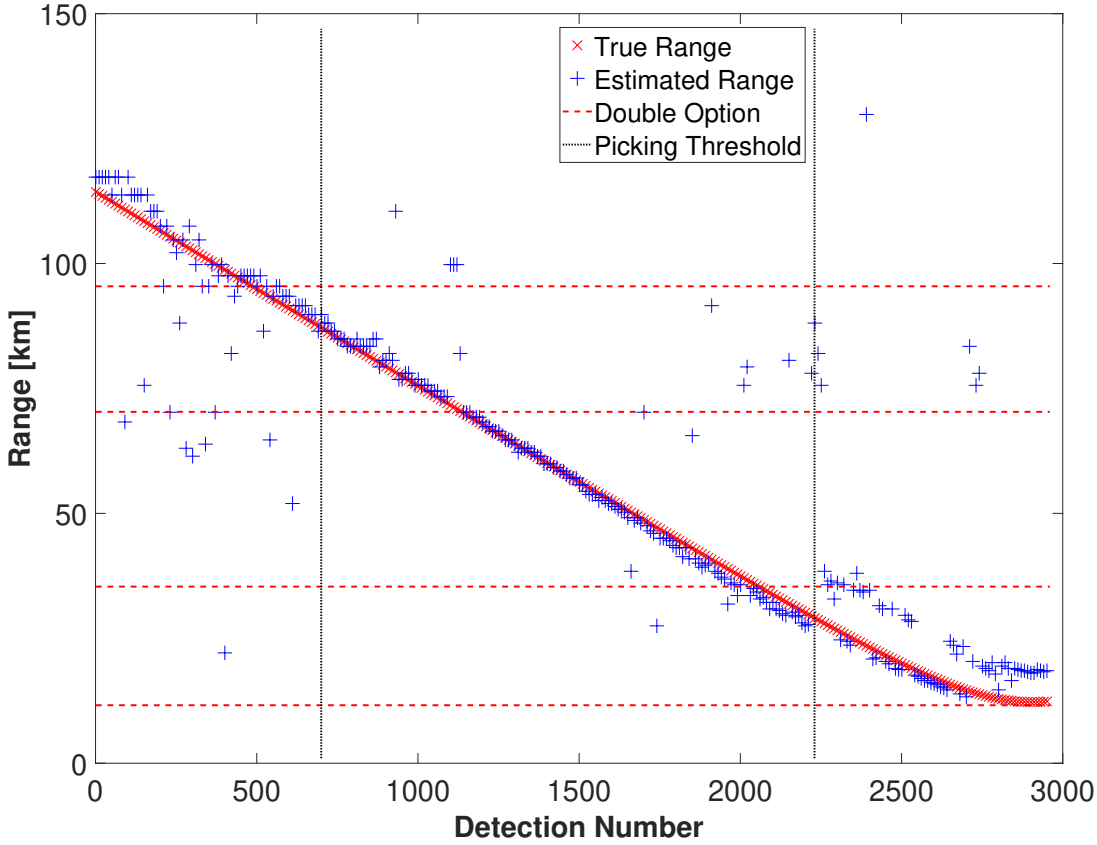

Figure 3.31: True (red "x" symbols) vs. estimated (blue "+" symbols) range for manually estimated detections from the 296 air gun shot data set. Two pairs of horizontal red dashed lines delineate the close:medium and medium:long double option zones. Two vertical black lines delineate regions with the same lowest order arrival path: for detections 1-710, BS2 is the lowest order arrival path, for shots 711-2240 BS is the lowest order, while for shots 2241-2960 D is the lowest order arrival.

Near the hydrophone, the D arrival has a sufficiently high received level to enable the cross correlation to pick out the first three RATs, but the BS4 arrival is too faint to be detected by the automated RAT estimation code. Figure 3.33 illustrates this problem. Note that, in the bottom plot, cross-correlation envelope peaks for D, BS, BS2, and BS3 are clearly visible, but no such peak is apparent for BS4. In this instance, the propagation model predicted that the BS4 arrival would reach the ACO hydrophone, but the model failed to take into account the effects of reflection/transmission losses, which results in a mismatch between modeled and actual acoustic behavior at close range. Therefore, lacking a usable BS4 cross-correlation peak, a bias is introduced into the range estimate.

Even with the short range bias in the results, across all 296 air gun shots, the average difference between true and estimated range, $\mu_{range} = mean(rStrue - rSest)$ is -2.29 km. In other words, on average the localizer estimates the source range 2.39 km further from the hydrophone than the actual range. Figure 3.34 plots this range difference for each shot, along with the average difference. To account for a mixture

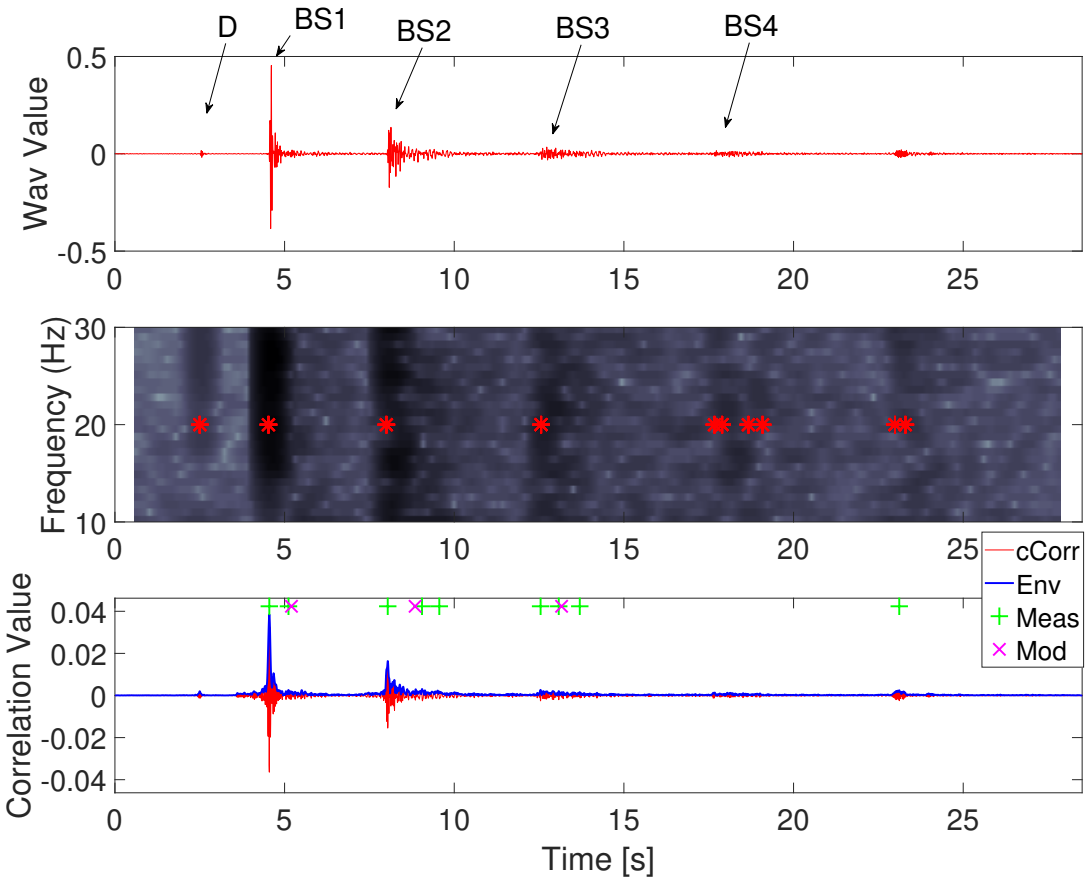

Figure 3.32: A depiction of the measured RAT estimation process for the direct-path arrival of airgun shot 227, with a true range of 28.0 km. The recorded waveform (top), spectrogram with detector output highlighted (middle), and cross-correlation (bottom) are shown. Arrival paths are labeled. The weak signal corresponding to the direct-path arrival at Time= 2.5 s yields overall low amplitude cross-correlation peaks, particularly for the D, BS3 (Time= 12.5 s), and BS4 (Time= 17.5 s) arrival paths.

of positive and negative biases potentially masking the variability in these range differences, Figure 3.35 plots the magnitude of this difference for each shot, along with the average difference magnitude (6.8 km). Excluding the 5% highest range differences (which, for this instance, are due to background noise yielding a high confidence range estimate at the wrong range), the average range difference magnitude drops to 4.17 km.

The short range bias due to faint D and/or BS4 arrivals starts when the true range equals 27.99 km. With the maximum range within the Dunn air gun transect being 114.30 km:

$$(114.30 - 27.99)/(114.30) = 75.51\% \tag{3.5}$$

of the transect is free from this short range bias. However, given that the observable area around ACO is circular, not linear, the area free of short range bias can also be estimated. The maximum sensing radius about ACO for this work is the range out to which BS2 arrivals are expected to still reach ACO; beyond

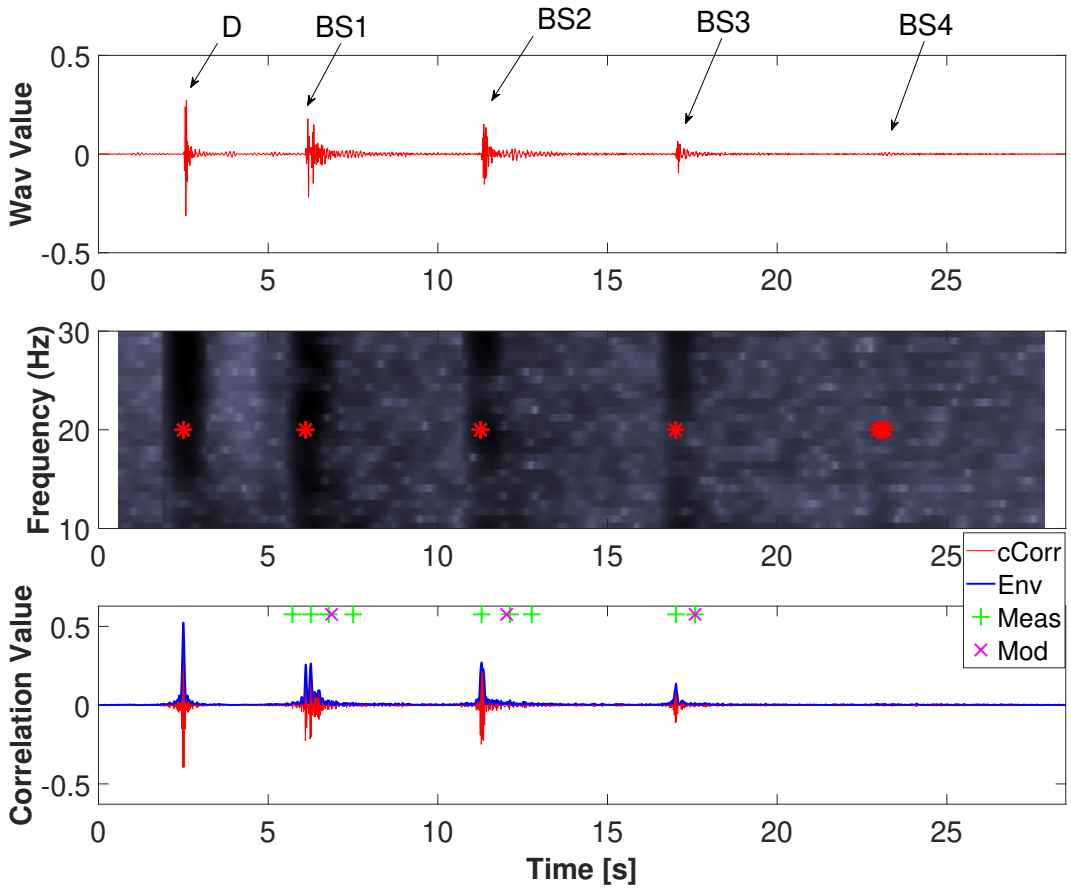

Figure 3.33: A depiction of the measured RAT estimation process for the direct-path arrival of airgun shot 278, with a true range of 13.3 km. The recorded waveform (top), spectrogram with detector output highlighted (middle), and cross-correlation (bottom) are shown. Arrival paths are labeled. The BS4 is particularly weak for this shot, while the D through BS3 arrivals are comparatively stronger.

this range, the propagation model predicts that the lowest order eigenray path will be BS3. Given that the highest order path reliably seen in the fin whale data is BS4, only having BS3 and BS4 arrivals yields only a single RAT with which to locate the call in question. Beyond approximately 90 km radius from ACO; beyond this radius, the bottom slopes up significantly toward Kauai and O'ahu from ACO. Taking this 90 km as a conservative approximation of the max sensing range from ACO given the underlying bathymetry, the percentage of the ocean area about ACO free of short range bias is:

$$(\pi * 90^2 - \pi * 27.99^2)/(\pi * 90^2) = 90\% \tag{3.6}$$

Thus, approximately 90% of the area immediately around ACO is free of short range bias. This percentage goes up to 94% if we consider the max Dunn transect range of 114.30 km, and up to 96.4% if we use the BS2 cutoff range of 146.98 km.

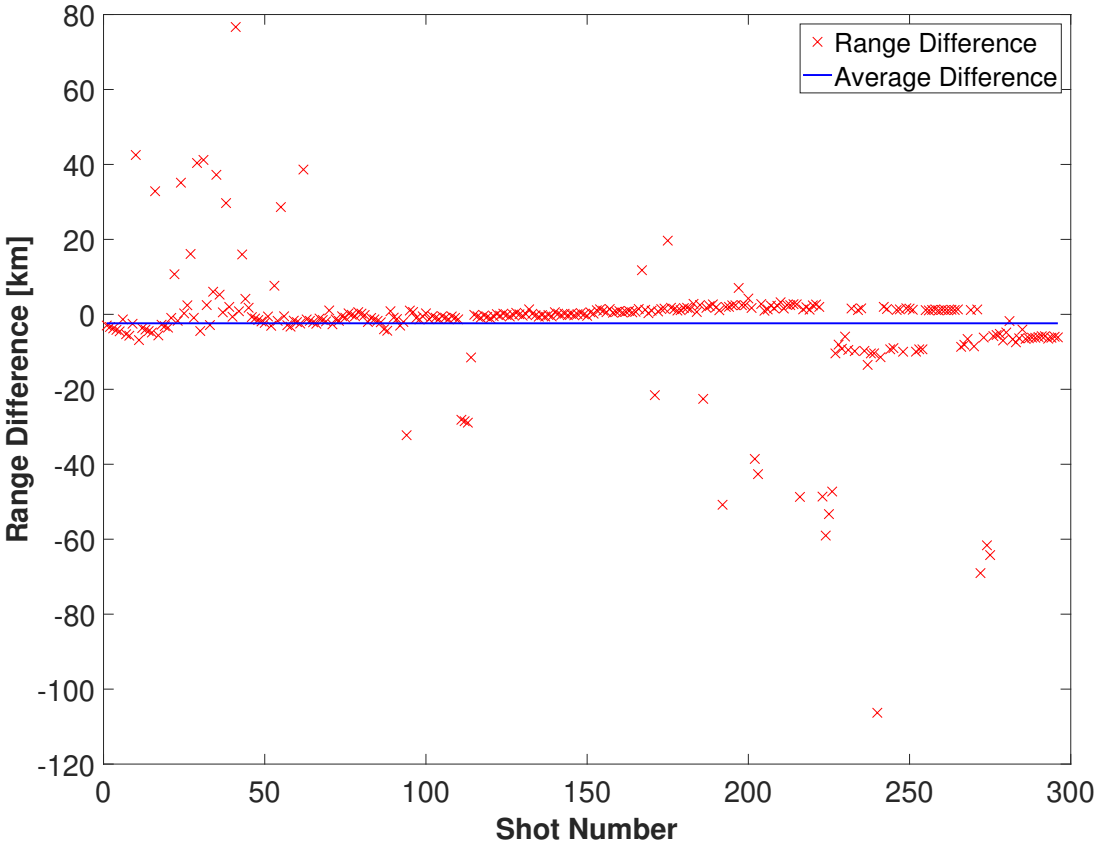

Figure 3.34: The difference between the true and estimated range for each shot when manually picking detections and allowing RATs to be automatically estimated.

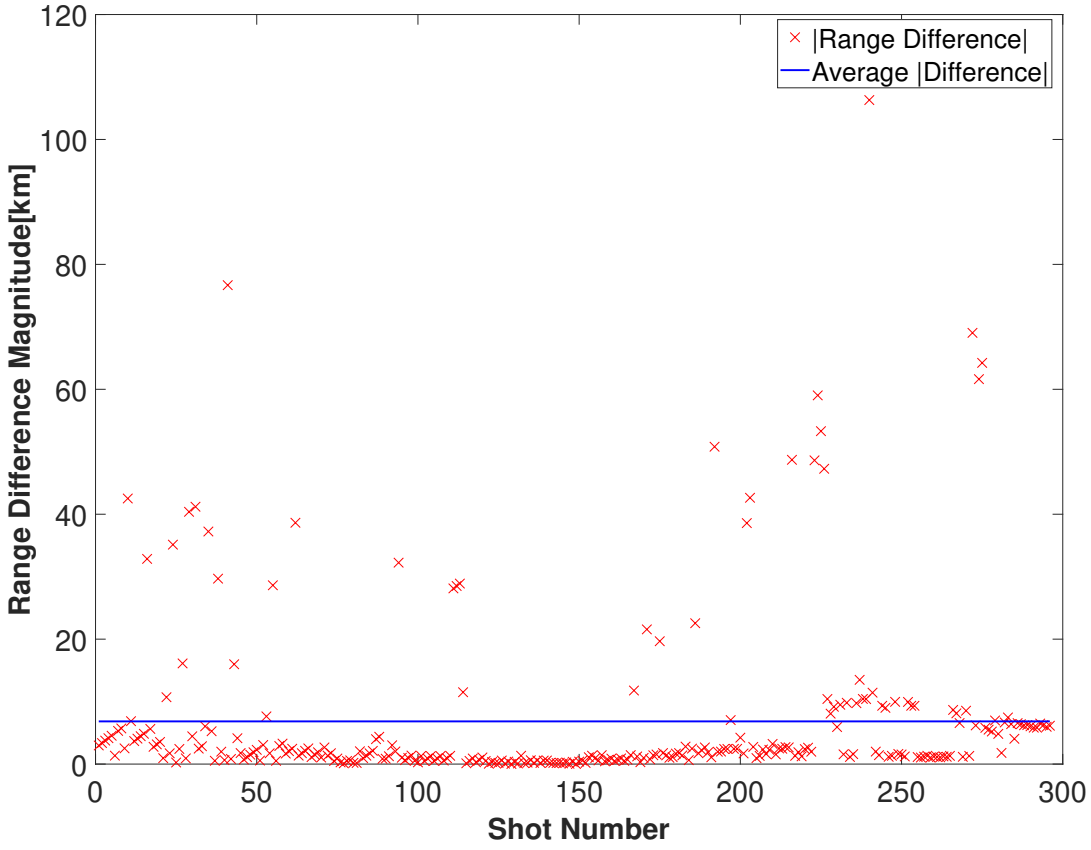

Figure 3.35: The magnitude of the difference between the true and estimated range for each shot when manually picking detections and allowing RATs to be automatically estimated.

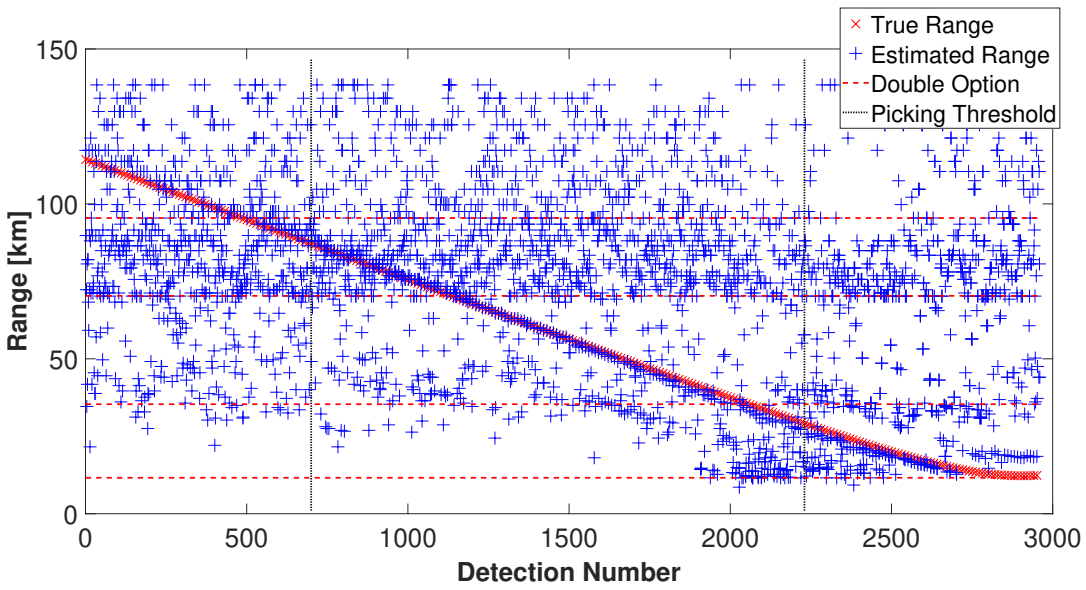

Figure 3.36: True (red "x" symbols) vs. estimated (blue "+" symbols) range for manually estimated detections from the 296 air gun shot data set (10 detections per shot). Two pairs of horizontal red dashed lines delineate the close:medium and medium:long double option zones. Two vertical black lines delineate regions with the same lowest order arrival path: for detections 1-710, BS2 is the lowest order arrival path, for shots 711-2240 BS is the lowest order, while for shots 2241-2960 D is the lowest order arrival.

All Detected Arrivals Range Estimation Analysis

In Section 3.5.5, the range for each detection corresponding to the earliest arrival for each shot was plotted. This was done to better compare with the range estimation results shown in Section 3.5.5 where RATs were manually picked. In this section, all ten detections from each shot, not just the first, are considered and plotted. This presents a more realistic example of localizer performance, since when fin whale calls are considered in the next chapter the subset of detections corresponding to the earliest arrival of those fin whale calls is not known *a priori*.

Figure 3.36 shows the estimated range for each of the 2960 individual detections within the Dunn air gun set compared to the true locations. No pruning was done here, so all detections are plotted regardless of ambiguity or phantom track status. Ninety percent of the detections represent either higher order reflected paths (i.e., not the earliest path of a shot) or spurious detections that are unrelated to an air gun shot.

Figure 3.37 plots the subset of the 2960 detections whose range ambiguity exceeds 0.8. Even at this seemingly high ambiguity threshold, it is difficult to identify the differences between the unpruned and pruned figures. At short and medium range, there is some evidence of a track, but as candidate source range spacing increases and the density of false detections increases, the track becomes less visually apparent. This figure illustrates that ambiguity value alone can be insufficient to adequately distinguish a track from noisy and higher order arrival detections.

Figure 3.38 shows the subset of the 2960 calls which match not only the ambiguity > 0.8, but also the phantom track, criteria. In this case, a call is pruned away if, and only if, its ambiguity is less than 0.8 or the minimum distance between its set of measured arrival times and the set of prior calls is less than 1.25 s.

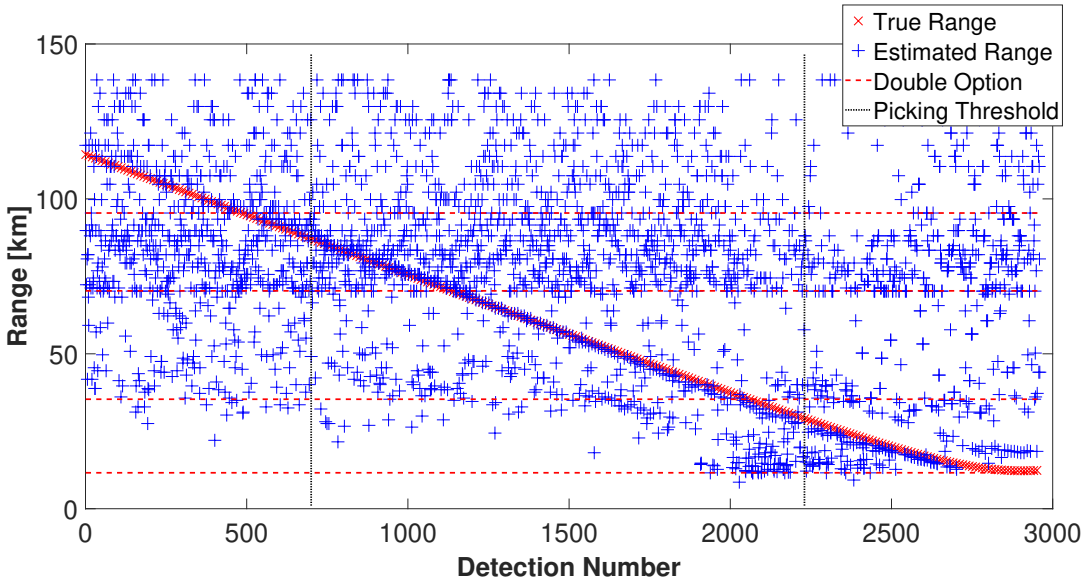

Figure 3.37: True (red "x" symbols) vs. estimated (blue "+" symbols) range for manually estimated detections from the 296 air gun shot data set (10 detections per shot). Only those detections whose ambiguity value exceeds 0.8 are plotted. Two pairs of horizontal red dashed lines delineate the close:medium and medium:long double option zones. Two vertical black lines delineate regions with the same lowest order arrival path: for detections 1-710, BS2 is the lowest order arrival path, for shots 711-2240 BS is the lowest order, while for shots 2241-2960 D is the lowest order arrival.

This process differs slightly from that used for pruning fin whale calls because for the fin whale case, before a call can be added to the set of prior calls against which a new calls arrival times are compared, the call's peak frequency must be less than 19.5 Hz and its stop frequency must be less than 24 Hz. Regardless of it's ambiguity, if a call fails either of those frequency pruning criteria, it is pruned and not used to judge the fitness of subsequent calls to pass the phantom pruning criteria. This makes it less likely that a spurious detection will cause subsequent calls to be incorrectly pruned away. These frequency criteria are not used for the pruning of the air gun data since the frequency characteristics of air gun shots are not precisely aligned with the fin whale calls whose characteristics were used to create the frequency criteria.

The striking difference between the phantom (Figure 3.38), and no phantom (Figure 3.37), pruning cases illustrates the utility of phantom pruning. The actual track is easier to distinguish at all ranges, with particular improvement in the mid-high double option zone and at high range, when phantom pruning takes place. Detections of higher order reflected paths regularly yield high ambiguity, yet spurious, location estimates. If the earliest arrival of a call is detected and localized, any calls whose arrival times are sufficiently similar to this earlier one are pruned by the phantom track criterion. This criterion helps distinguish localizations of physically meaningful detections from both localizations of higher order arrivals, and background noise, if either consists of sufficiently similar arrival times. The coarse candidate source range spacing at high range makes it more difficult to distinguish the actual track from noisy detections than at closer ranges.

Figure 3.39 shows the localization results if an ambiguity threshold of 0.9, rather than 0.8, is used in addition to phantom pruning. The higher ambiguity threshold succeeded in pruning away a small number of additional detections that passed the 0.8 ambiguity criterion, but this small improvement in noisy detection

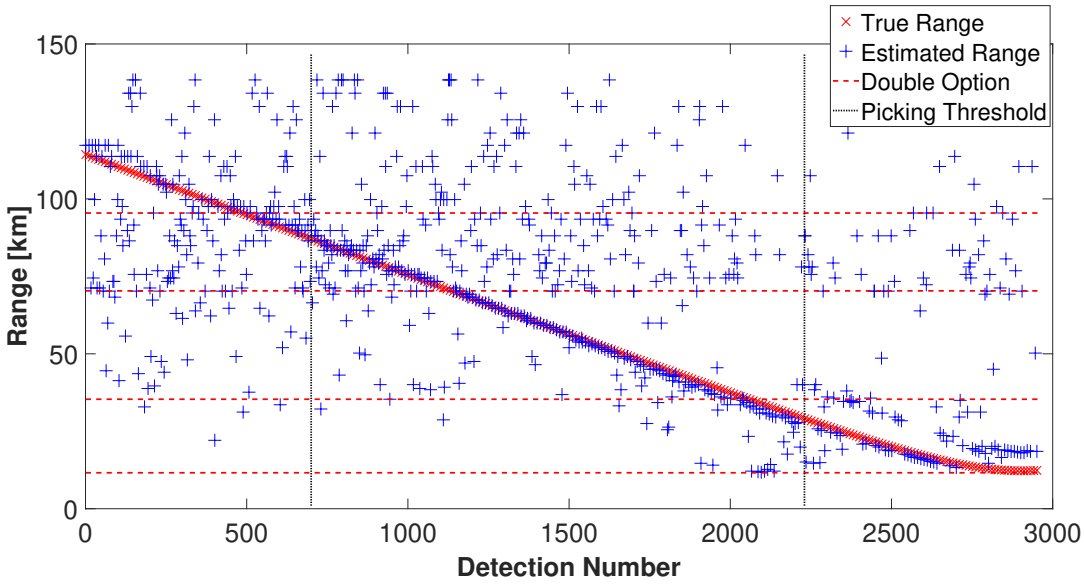

Figure 3.38: True (red "x" symbols) vs. estimated (blue "+" symbols) range for manually estimated detections from the 296 air gun shot data set (10 detections per shot). Only those detections whose ambiguity value exceeds 0.8, and which pass the phantom pruning criteria, are plotted. Two pairs of horizontal red dashed lines delineate the close:medium and medium:long double option zones. Two vertical black lines delineate regions with the same lowest order arrival path: for detections 1-710, BS2 is the lowest order arrival path, for shots 711-2240 BS is the lowest order, while for shots 2241-2960 D is the lowest order arrival.

rejection did not materially improve the clarity of the track with respect to the noisy and spurious detections.

3.5.6 Summary

The objective of this section is to demonstrate how well our localization algorithm works when recorded underwater acoustic data when the true source range is known. In particular, recordings of 296 individual air gun shots recorded October 5-6 2018 a the ALOHA cabled observatory are used. This set of shots start off approximately 114 km from ACO and has a closest point of approach of 12.2 km. To test the effectiveness of the localizer, RATs are manually identified from the acoustic data for each shot and, in turn, inputted to the localization algorithm. The close match between the true and estimated range sets demonstrates that the localizer can yield range estimates that closely match the true range from the ACO hydrophone. Secondly, to investigate the effect of automated RAT estimation on localization accuracy, air gun shot times are passed to the RAT estimation process. Following automated RAT estimation, range estimates are made for each of these manually-detected shots. Some deterioration in localization accuracy is seen when automated RAT estimation is incorporated into the data processing sequence, showing that there is room for improvement in the capacity of the RAT estimation process to yield RATs which in turn yield accurate and reliable range estimates. A comparison between phantom pruning and ambiguity value thresholding demonstrates that phantom pruning can be an effective means to reduce the incidence of false positives in a set of estimated ranges, albeit at the potential cost of rendering the overall approach less likely to accurately predict the presence of overlapping signals.

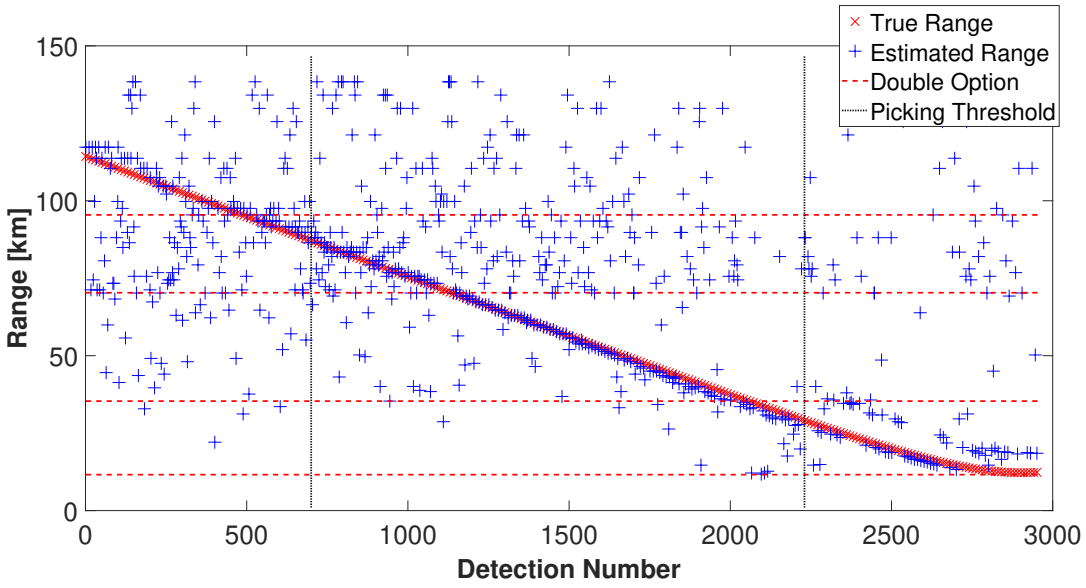

Figure 3.39: True (red "x" symbols) vs. estimated (blue "+" symbols) range for manually estimated detections from the 296 air gun shot data set (10 detections per shot). Only those detections whose ambiguity value exceeds 0.9, and which pass the phantom pruning criteria, are plotted. Two pairs of horizontal red dashed lines delineate the close:medium and medium:long double option zones. Two vertical black lines delineate regions with the same lowest order arrival path: for detections 1-710, BS2 is the lowest order arrival path, for shots 711-2240 BS is the lowest order, while for shots 2241-2960 D is the lowest order arrival.

CHAPTER 4
AUTOMATED FIN WHALE LOCALIZATION

Having developed and tested the fin whale range estimation approach in Chapter 3 using validated air gun data, Chapter 4 presents results from measured acoustic data collected at the ALOHA Cabled Observatory (ACO) that contain marine mammal (including fin whale) calls. In the previous chapter (Section 3.3) a description of the data collection apparatus at ACO was presented. Section 4.1 gives a brief description of fin whale ecology with an emphasis on the waters near Hawaii. Section 4.2 presents localization results from fin whale 20 Hz calls recorded on March 3, 2014 as an example of localization performance. In Section 4.3, fin whale detection rates as a function of pruning configuration, month:hour, and ambiguity level are shown. Section 4.4 discusses estimated fin whale call characteristics and how the estimation of these parameters is influenced by pruning configuration and ambiguity threshold. Finally, Section 4.5 presents a brief summary of this chapter.

4.1 Fin Whale Ecology in Hawaiian Waters

Fin whales (*Balaenoptera physalus*, also sometimes referred to as finback whales in the literature) are the second-largest species of whale in the world, are found in all the major oceans of the world, and are considered Endangered by the International Union for Conservation of Nature. Sometimes referred to as the greyhound of the sea, fin whales are thought to be capable of speeds up to 37 km/h. They are sometimes known to gather in pods of 2-7 or more, and hybrid fin whale/blue whale individuals have been documented. Fin whale life spans can reach 80-90 years, and their only known predator is the killer whale. Fin whale calves are born in sub-tropical waters during the winter months, but there is little evidence of distinct calving waters (as there is for humpback whales, for example). Most populations seem to exhibit a poleward migration in summer months for feeding, and a return to subtropical waters in the winter for breeding, although there are some notable examples of fin whales resident in some areas year-round (e.g., the Gulf of California and the East China Sea) (Jefferson et al. [2015]).

Acoustically, fin whales are known mostly for their 20-Hz calls, which are produced year round but emitted in structured bouts during the breeding season (Schevill et al. [1963], Jefferson et al. [2015]). Figure 4.1 shows a waveform and spectrogram for a sequence of ten 20-Hz calls recorded by the ocean-bottom hydrophone at the ALOHA Cabled Observatory. Historically, 20-Hz calls were identified as being produced by fin whales in Schevill et al. [1963], while Walker [1963] reported on recordings of "$\cdots$ intense, low-frequency, underwater sounds of wide geographic distribution, apparently of biological origin" which were later (Watkins [1981]) attributed to fin whales. An examination of the characteristics of 20-Hz calls is described in Watkins et al. [1987], which describes them as 1 s, frequency-downswept pulses from approximately 23 Hz down to 18 Hz with a source level of up to 186 dB re 1 μPa at 1 m. The interpulse interval (IPI) was between 7-26 s, and individual bouts were separated by more than 2 hours and lasted as long as 32.5 hours. Within a single bout, rests of 1-20 mins occurred every 15 minutes or so, with occasional gaps between 20-120 minutes also present in their data record.

The 20 Hz calls described in Watkins et al. [1987] were recorded near Bermuda, and were significantly more common in the winter months than during the summer (i.e., there was significant seasonality in the

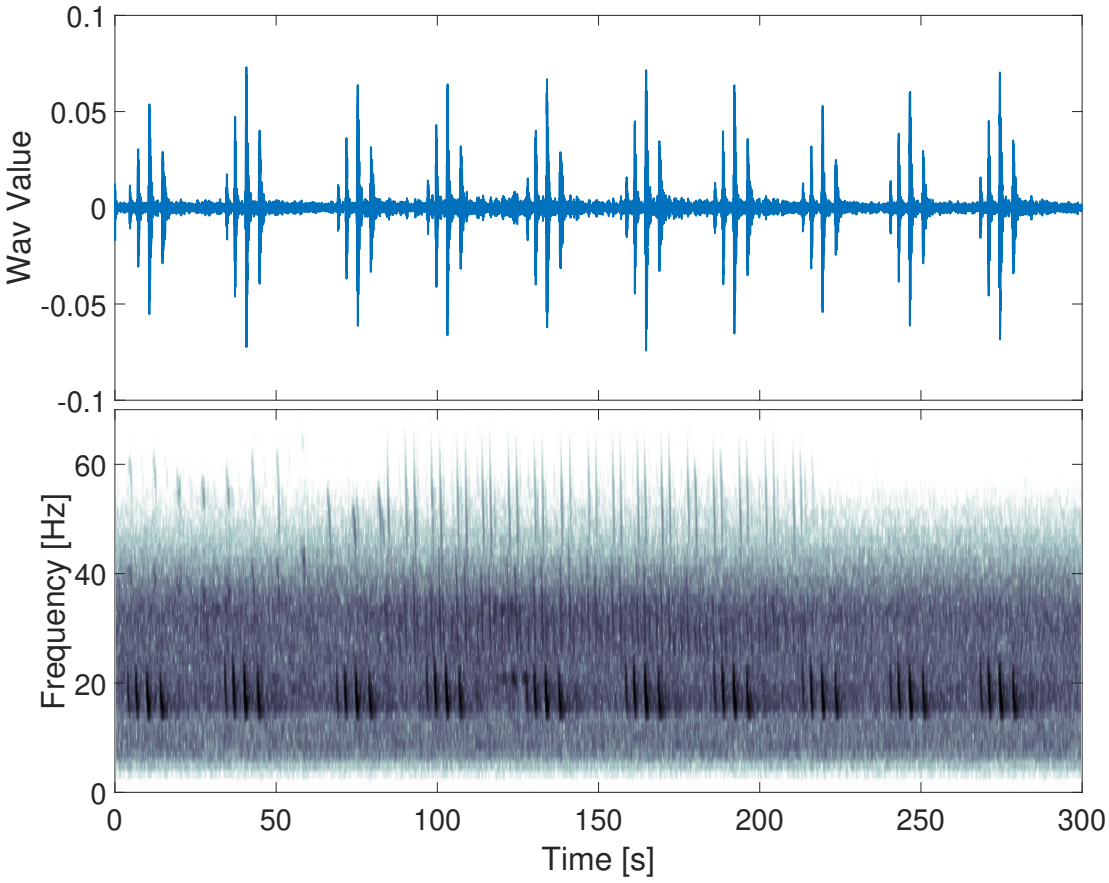

Figure 4.1: Fin whale 20 Hz calls, and other marine mammal calls, recorded between 06h00-06h05 on March 3 2014 at the ALOHA Cabled Observatory. Waveform (top) and spectrogram (bottom) plots are shown.

fin whale 20 Hz calling behavior). Figure 4.2 shows the percent of time that fin whale calls were present in their Bermuda data record for each month of the year. Fin whale calling bouts occupied a larger percentage of the data record during the winter months than during the summer. Figure 4.3 shows similar results for fin whale call occurrence near Hawaii and Midway Island.

In addition to their 20 Hz calls, fin whales north of Hawaii have also been reported to produce calls in the 20-35 Hz and 30-90 Hz frequency bands (McDonald and Fox [1999]), although the repetition interval of these calls is more irregular than that of the 20 Hz calls. More recently, sei whales were proposed (Rankin and Barlow [2007]) an alternative source for the 20-35 Hz calls described by (McDonald and Fox [1999]). Figure 4.4 shows examples of the 20 - 35 Hz irregular calls from four individuals off the coast of Southern California.

In Thompson and Friedl [1982], acoustic data from two ocean bottom hydrophones were collected from December 1978 to April 1981 in waters near O'ahu, Hawaii. The two hydrophones were 11.6 km apart and in 400 fathoms ($\approx$ 732 m) of water. Calls from five different whale species were identified in the data:

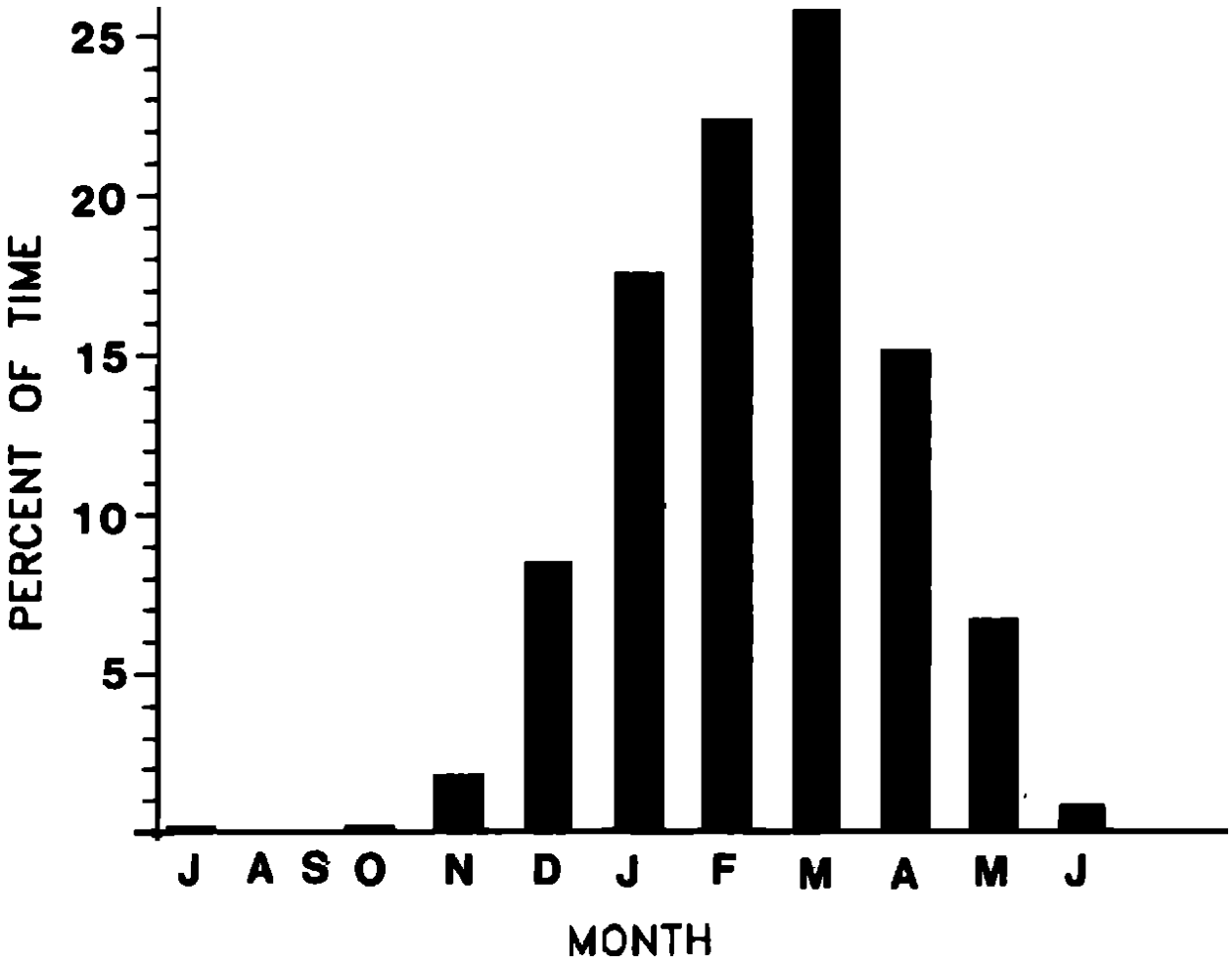

Figure 4.2: The percentage of the signal duration in each month, over four years of recording near Bermuda, within which fin whale 20 Hz calls were present (Watkins et al. [1987]).

humpback whale, fin whale, blue whale (*Balaenoptera musculus*), sperm whale, and pilot whale (*Globicephala macrorhynchus*). Additionally, so-called "boing" sounds were identified in the data that may be the same type of "boing" sounds identified in Winn and Perkins [1976] as belonging to minke whales (*Balaenoptera acutorostrata*). With regards specifically to fin whales, the authors observed that the 20 Hz fin whale calls were typically observed as doublets, whereby the interpulse interval (IPI) between consecutive calls alternated back and forth from 17-24 s and 40-52 s and the call frequency sweep alternated from 20 - 16 Hz and 30 - 16 Hz. The authors also ascribed two other call types (20 - 35 Hz pulses, and 30 - 90 Hz pulses) to fin whales based on the similarity of these pulses (other than their frequency range) to the 20 Hz fin whale calls, while noting that these two higher-frequency calls could have been made by another large whale such as the minke or sei whale. Considering all three of these call types, fin whale call presence showed significant seasonality, with the 20 Hz doublet pulses being most common in the winter months. Figure 4.5 shows the percent occurrence of calls from six marine mammal species in each month of the year; the authors speculate that the bimodal trend in the data may be due to fin whale migration, and that the three different fin whale call

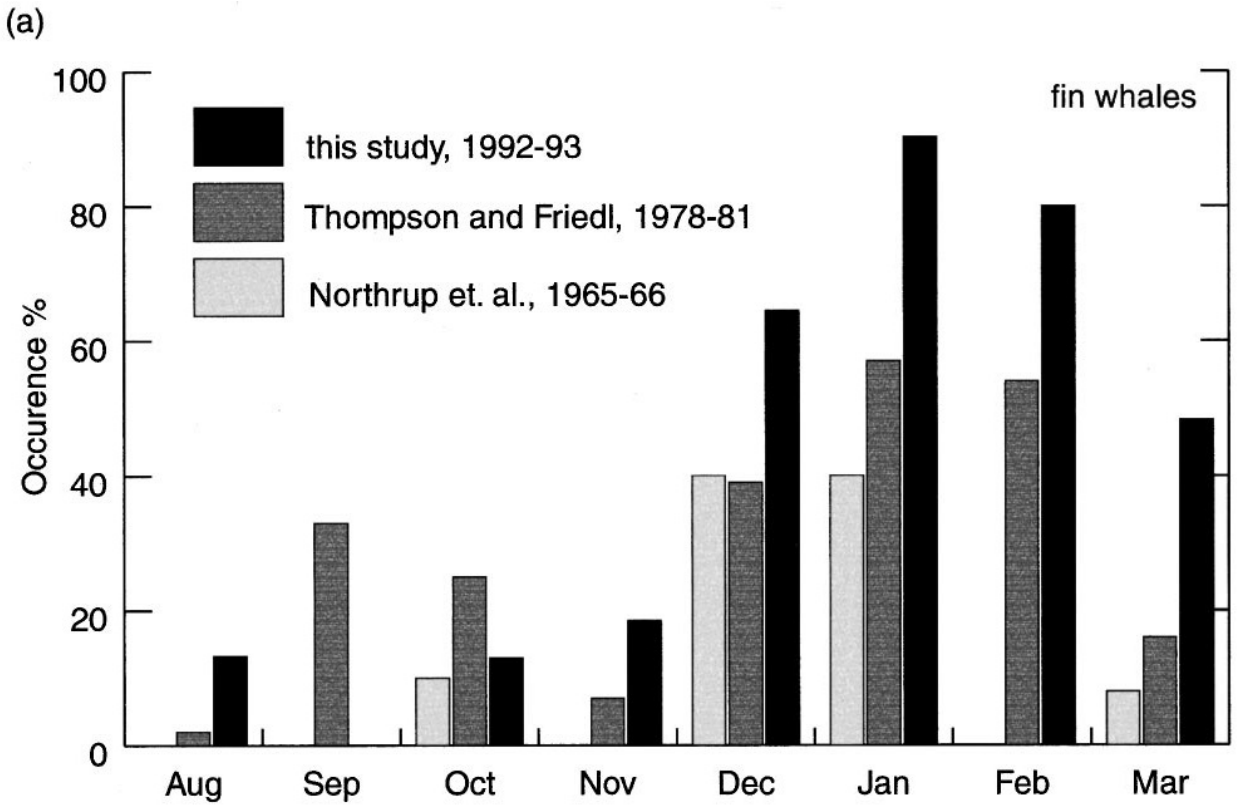

Figure 4.3: The percentage of recording days in each month within which fin whale calls were received in data near O'ahu (Black and dark grey bars) and near Midway Island (light grey bars). In this legend, "this study" refers to McDonald and Fox [1999], where this figure was originally published.

types may demonstrate an age and/or sex difference in fin whale calling behavior.

4.2 Fin Whale Range Estimation

In this section, example fin whale tracks from the full ACO data set are shown. In this context, a track is defined as a set of individual consecutive, or nearly consecutive, calls that could have come from the same source. Spatial uncertainties are sufficiently high that it is not possible to attribute consecutive calls to the same individual i.e. an individual "track" could be made up of calls from multiple closely-spaced calling fin whales rather than a single calling animal. Additionally, since fin whales are known to call in bouts with interspersed periods of silence (Watkins et al. [1987]), the animal(s) calling in one continuous part of a track may be different than those following a break. These statements are meant to express the limits of knowledge that we can infer from a fin whale track, rather than to hypothesize on the calling and group behavior of fin whales.

The data in this chapter are processed following the methods outlined in Chapter 3, and Figure 4.6 outlines the fully-automated data processing sequence. Acoustic data were continuously recorded by the moored ocean-bottom hydrophone at the ALOHA cabled observatory (ACO) approximately 115 km north

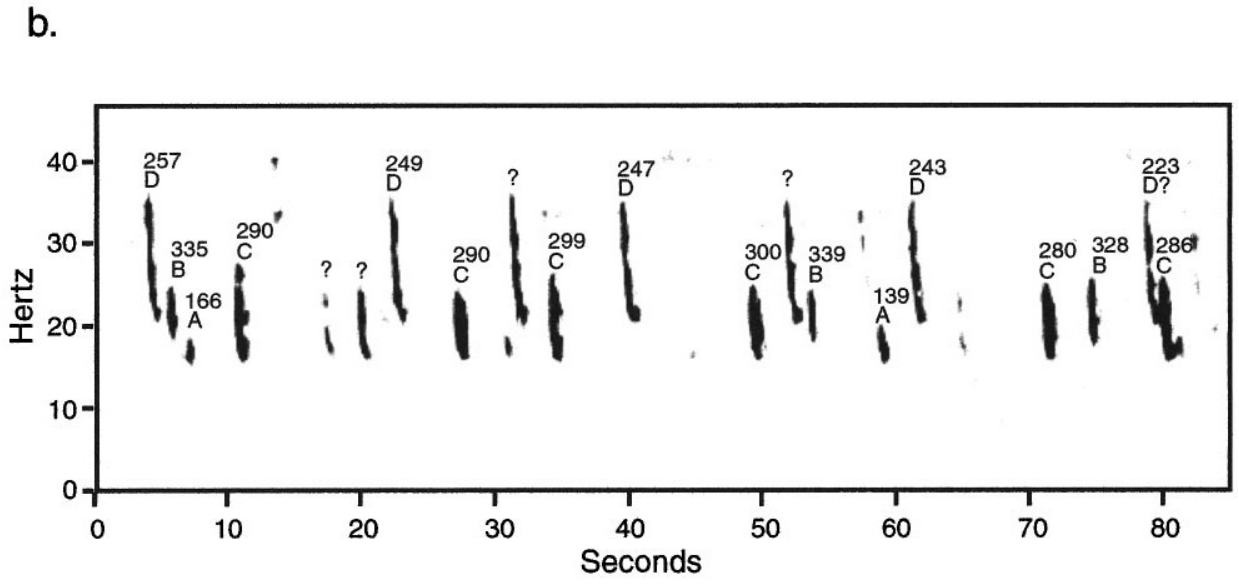

Figure 4.4: Examples of 20 to 35 Hz irregular fin whale calls from four different animals (labeled A-D) along with compass bearing for each call (McDonald and Fox [1999]).

of O'ahu. These data were recorded from February 18 2007 - September 30 2017 with a notable recording gap from October 23 2008 - June 6 2011. Data are available after the September 30 2017 cutoff date for this analysis, but are not included here. Data are bandpass filtered with a passband of 9-31 Hz and downsampled to a sampling frequency of 200 Hz. A matched filter is used to pick out sounds resembling fin whale 20 Hz calls. Measured relative arrival times (RATs) for each of the detected calls are estimated. These measured RATs are then compared against model-predicted RATs at a set of candidate source ranges using our cost function (Equation (3.2)). The best-fit range is the candidate source range whose model-predicted RATs most closely match to the measured RATs. Lastly, the set of localized calls are pruned using a set of frequency and phantom pruning criteria designed to reduce the false positive rate.

The data are processed using a desktop PC with an 8-core 3.6 GHz i7-9700K CPU and 32 GB of RAM. Overall, the acoustic data processing sequence shown in Figure 4.6 took approximately a month of continuous operation to analyse all the ACO acoustic data recorded between February 18 2007 - September 30 2017. Data were processed exclusively in Matlab, and many of the functions used were implicitly multi-threaded by Matlab.

Figure 4.7 shows 24 hours of fin whale calls from March 3 2014. No pruning was applied to the detections in this figure, and no ambiguity threshold was applied either. Thus, every detection made by the matched filter in this time period is registered as a single blue + symbol in this figure. Two subsequent figures show the same day of detections, with frequency pruning alone (Figure 4.8) and both frequency and phantom pruning (Figure 4.9). A series of individual calling bouts appears in the 40-45 km range from ACO starting near 02h00 and ending near 16h00. A shorter track starts at longer range at approximately 18h00 and finishes near 19h30. Gaps in time between individual calling bouts do not necessarily mean that the animal was (or animals were) silent in these times, although the permissive nature of the matched filter detector suggests that if calls were present in these times they likely would appear in the set of detections. Also,

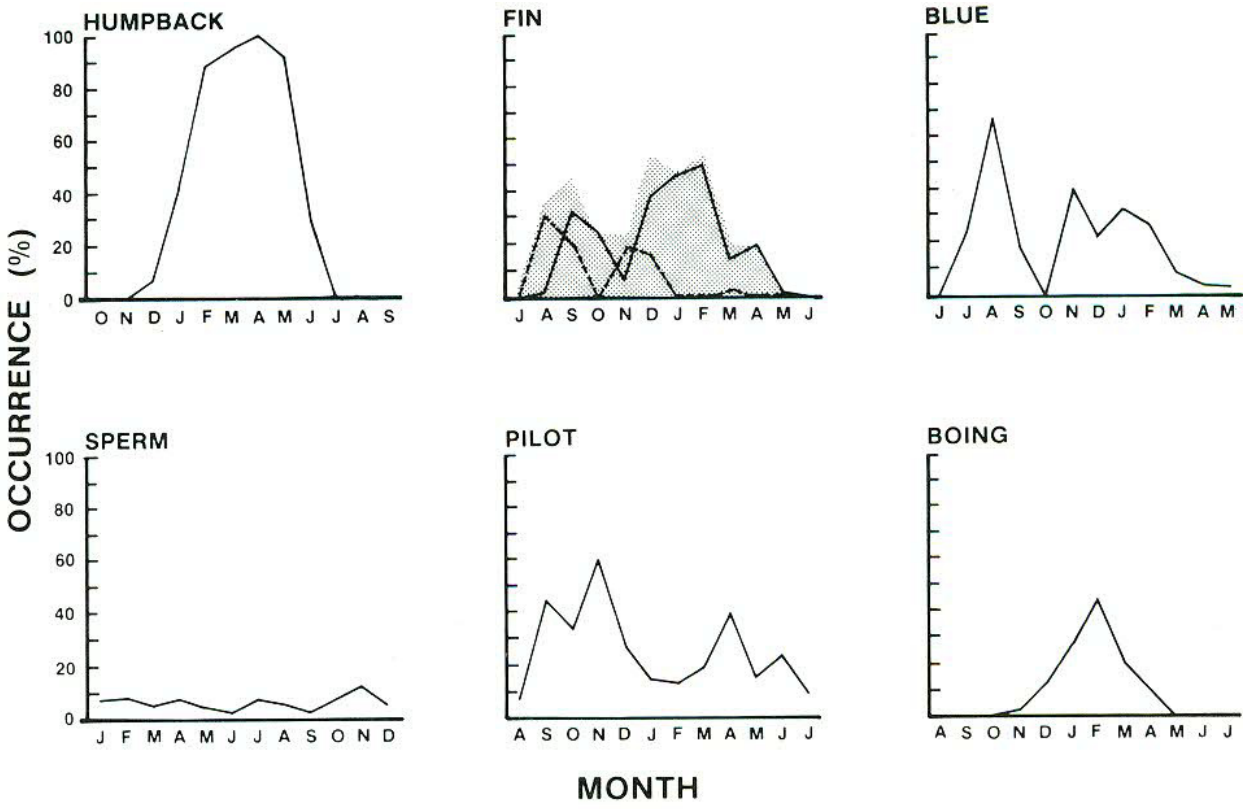

Figure 4.5: Yearly whale occurrence north of O'ahu, December 1978 - April 1981. Occurrence is the percent of recording days in each month that a call was received. For the fin whale plot, the solid line is for 20 Hz calls, the dashed line for irregular pulses attributed to fin whales, and the background stippling indicates the sum of these two call types. (Thompson and Friedl [1982])

there isn't necessarily just one animal calling during this bout. Only male fin whales are known to make 20 Hz calls (Croll et al. [2002]), and fin whales are known to sometimes travel in small groups (Jefferson et al. [2015]). Thus, this set of individual calling bouts represents one or more male fin whales producing 20 Hz calls, with no detected fin whale calls between the bouts.

Figure 4.10 plots the same day of data as Figures 4.7-4.9, but with all the location estimates with ambiguity less than 0.1 removed. Most of the noisy location estimates at minimum range (50 m) have been eliminated, but relatively few location estimates at other ranges were removed. Figure 4.11 shows the same day of range estimates after frequency pruning is applied in addition to an ambiguity threshold of 0.1. A number of calls from the apparent track near 01h00 and near 21h00 were pruned away, but the track is largely unaffected by the frequency pruning. This suggests that the track is more likely to have been a fin whale than sei whale. The gaps between the track segments (e.g., from 08h00-10h00) show superior noisy range estimate rejection than when an active track is present (e.g., 10h00-12h30). Figure 4.12 shows the resulting track estimates when both frequency and phantom pruning are applied, in addition to an ambiguity threshold of 0.1. Most of the noise not eliminated by the frequency pruning was removed by the phantom pruning. Notably, the fictitious track at reduced range (highlighted in Figure 4.13) is not present once

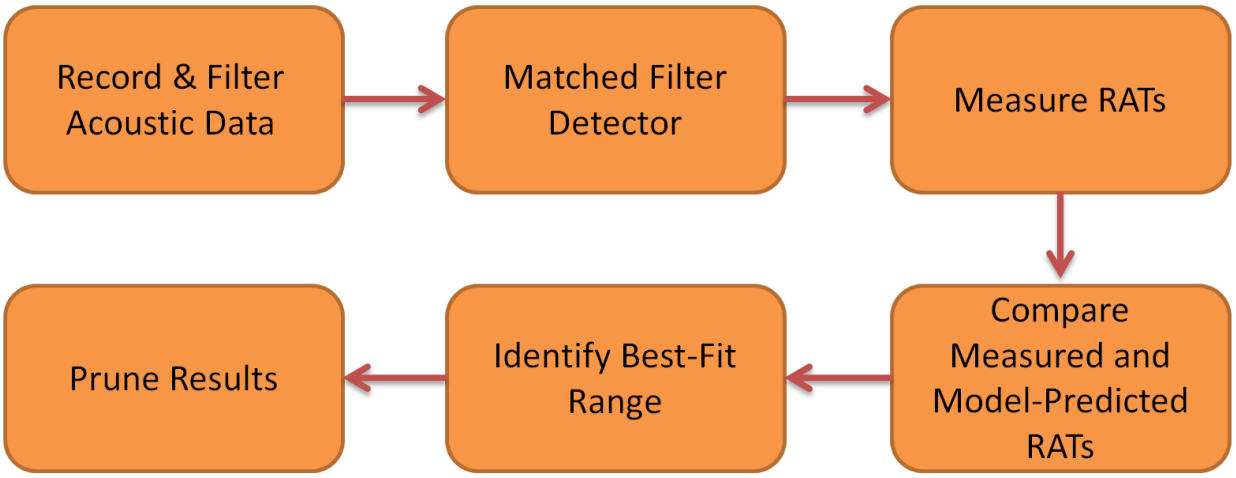

Figure 4.6: Data processing flow chart for the localization approach developed and used in this work.

phantom pruning is applied. This demonstrates that phantom pruning can eliminate false tracks from our set of range estimates, and that it also eliminates most noisy range estimates while leaving the integrity of a physically-meaningful track intact.

Figure 4.14 plots the localization info for the first call in the calling bout that begins at 10h00 in Figure 4.12. The recorded waveform is shown in subplot (a), while the spectrogram in the 10-30 Hz band is shown in subplot (b). Note that the two earliest arrivals (likely BS1 and BS2, given the estimated range of 42.67 km for this call) have lower received levels than the latter two arrivals for this call. Given the variability in arrival path received levels shown by the air gun results in Section 3.5.2, early arrivals having lower received levels than some later arrivals is not unexpected. The two detections that survived the frequency and phantom pruning regimes are denoted by red $*$ symbols in subplot (b). Lastly, the cross-correlation between the first arrival of this call, and the entire 27.5 s window about this detection, is shown in subplot (c). The largest peaks in the envelope of this signal line up well with the higher order arrivals clearly visible in both the time series and spectrogram subplots. The optimum match between the nine sufficiently-isolated peaks in the cross-correlation envelope (green + symbols) and the model-predicted RAT set is shown by the purple X symbols. The good agreement between these measured and model-predicted data (qualitatively shown in (c), and quantitatively indicated by the 0.99564 ambiguity score for this detection) indicate that the estimated range of 42.67 km for this call is well supported by the data.

Figure 4.15 plots the ambiguity vector for the call shown in Figure 4.14. The multiple local maxima in this ambiguity vector are typical for the ambiguity expression used in this work, and complicate the calculation of a simple uncertainty metric from this ambiguity vector. The two double option zones are delineated by dashed red lines. These double option zones have been consolidated, as they are in Figure B.3. The various ambiguity vector local maxima are due to different sets of model-predicted RATs yielding locally good fits to the set of measured RATs for this call, while the overall peak ambiguity for the candidate source range at 42.67 km is the estimated source range for this call.

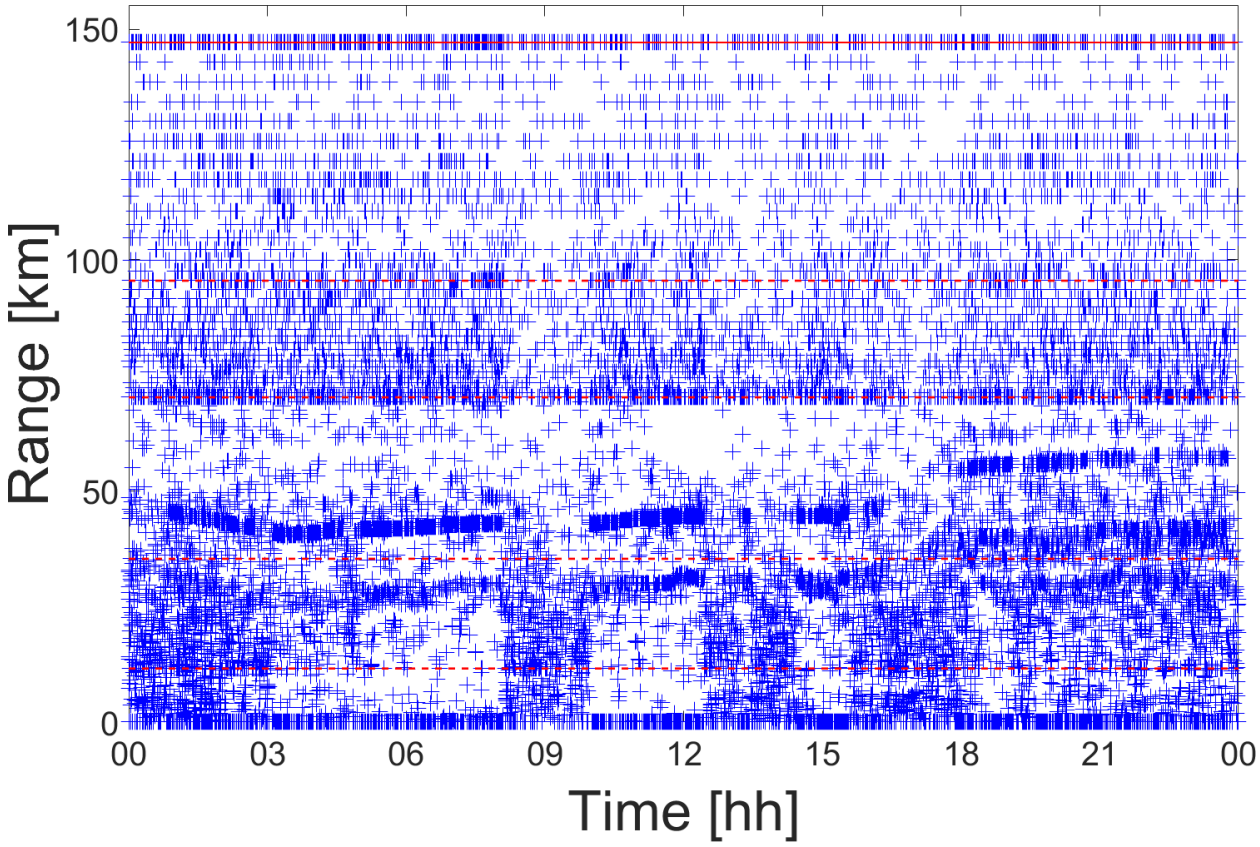

Figure 4.7: Estimated range (blue + symbols) for every detection made on March 3 2014. The edges of double option zones are indicated by dashed red lines. The max candidate source range is indicated by the solid red line.

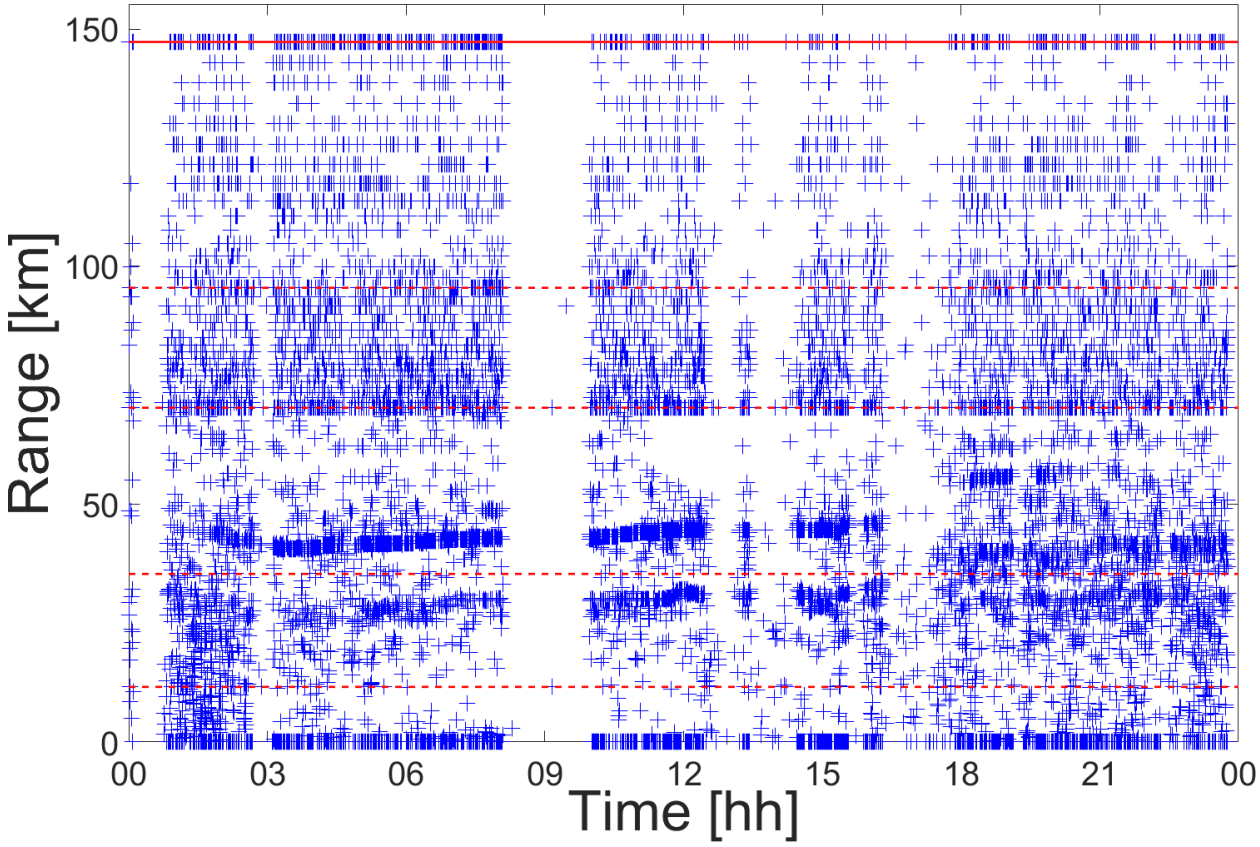

Figure 4.8: Estimated range (blue + symbols) for detections made on March 3 2014 that passed frequency pruning. The edges of double option zones are indicated by dashed red lines. The max candidate source range is indicated by the solid red line.

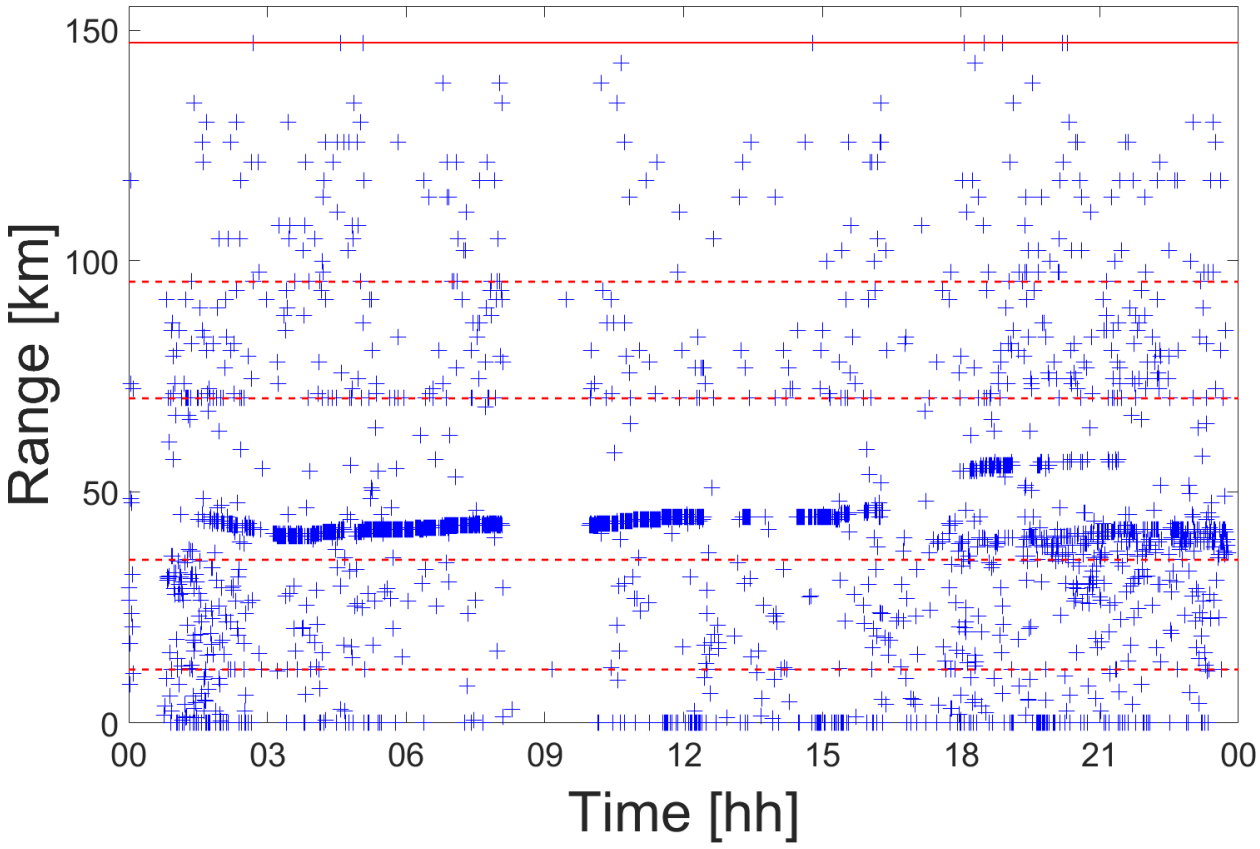

Figure 4.9: Estimated range (blue + symbols) for detections made on March 3 2014 that passed frequency and phantom pruning. The edges of double option zones are indicated by dashed red lines. The max candidate source range is indicated by the solid red line.

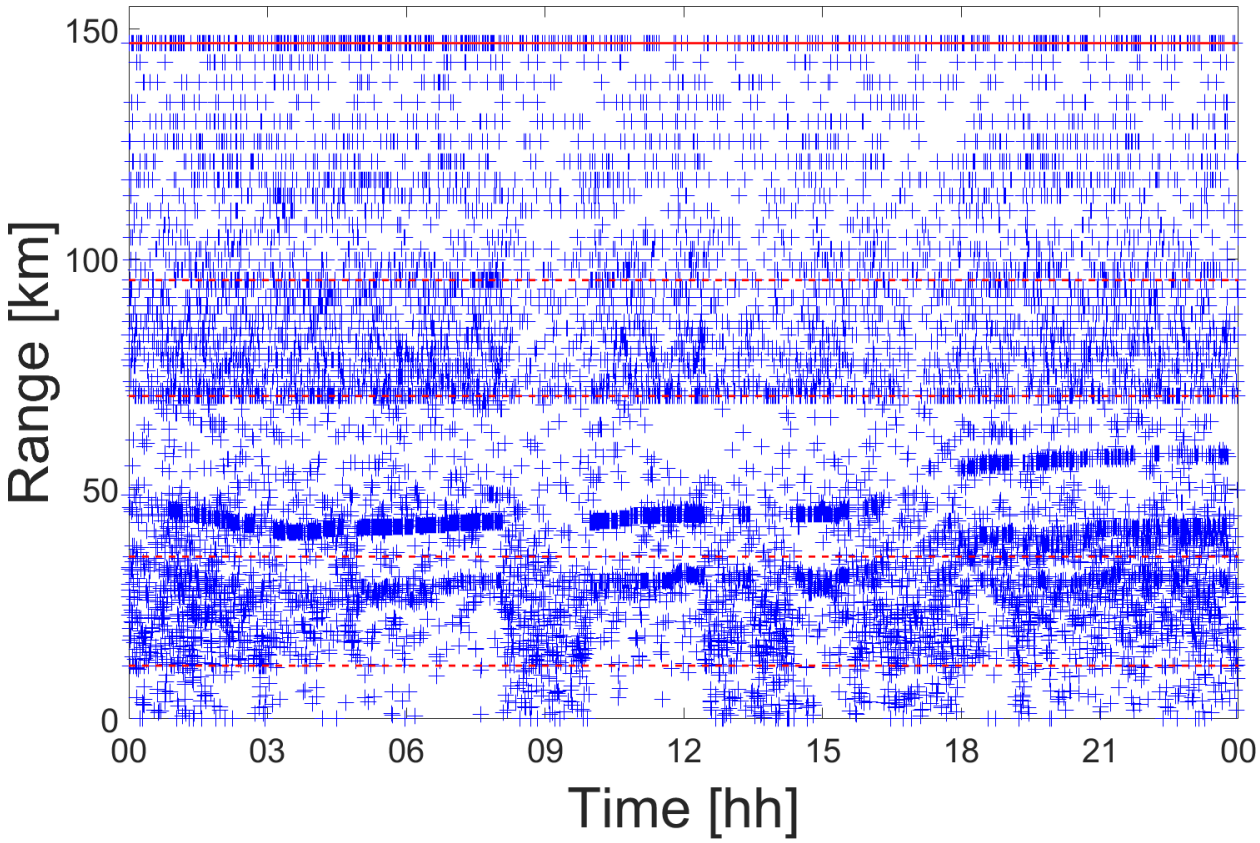

Figure 4.10: Estimated range (blue + symbols) for every detection made on March 3 2014 whose ambiguity exceeds 0.1. The edges of double option zones are indicated by dashed red lines. The max candidate source range is indicated by the solid red line.

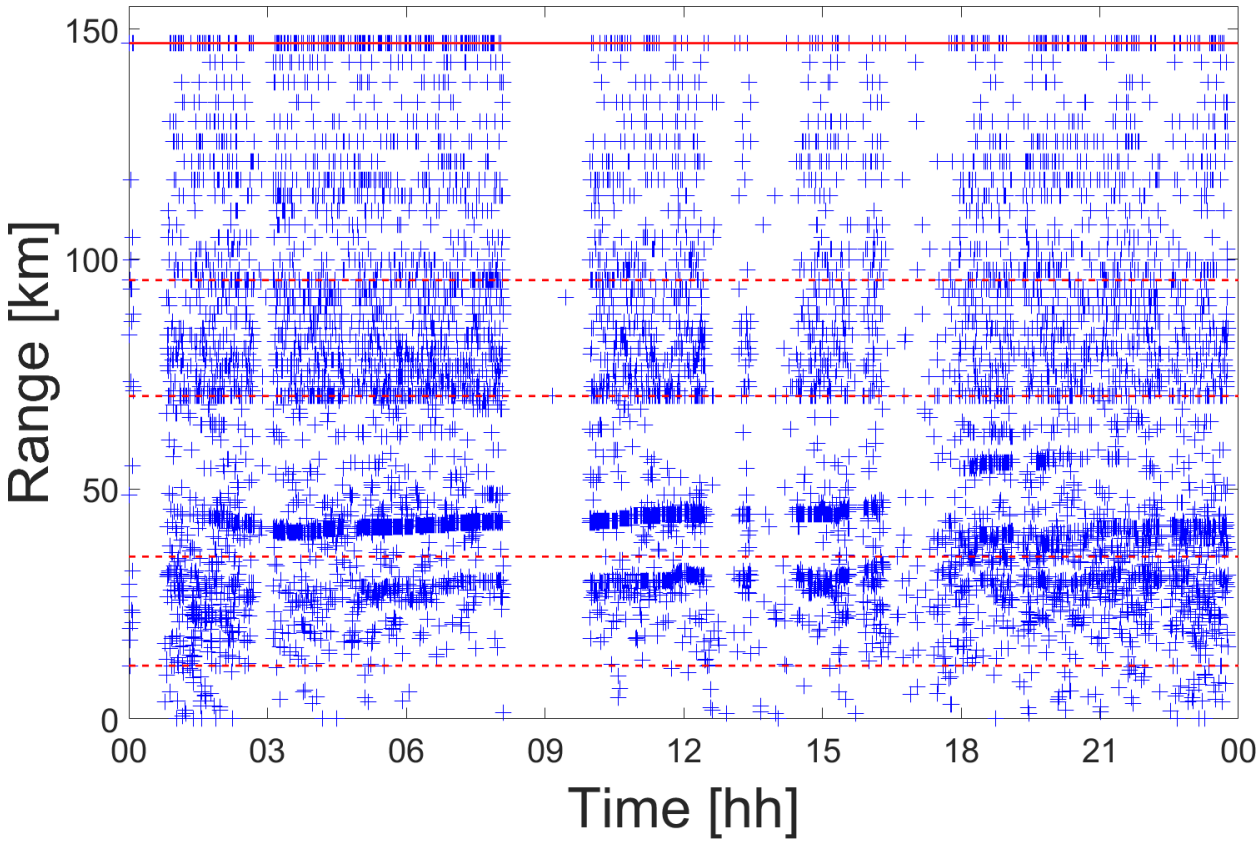

Figure 4.11: Estimated range (blue + symbols) for detections made on March 3 2014 that passed frequency pruning and whose ambiguity exceeds 0.1. The edges of double option zones are indicated by dashed red lines. The max candidate source range is indicated by the solid red line.

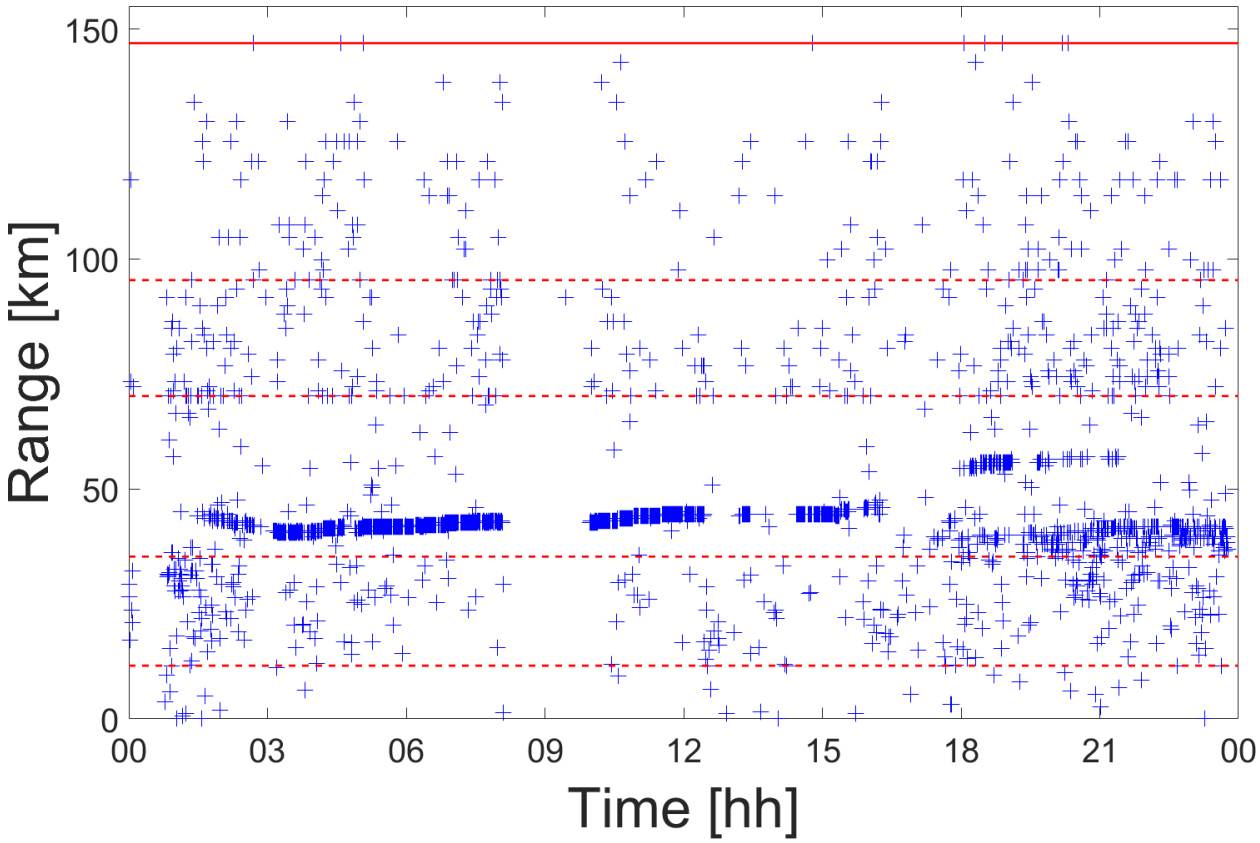

Figure 4.12: Estimated range (blue + symbols) for detections made on March 3 2014 that passed frequency and phantom pruning and whose ambiguity exceeds 0.1. The edges of double option zones are indicated by dashed red lines. The max candidate source range is indicated by the solid red line.

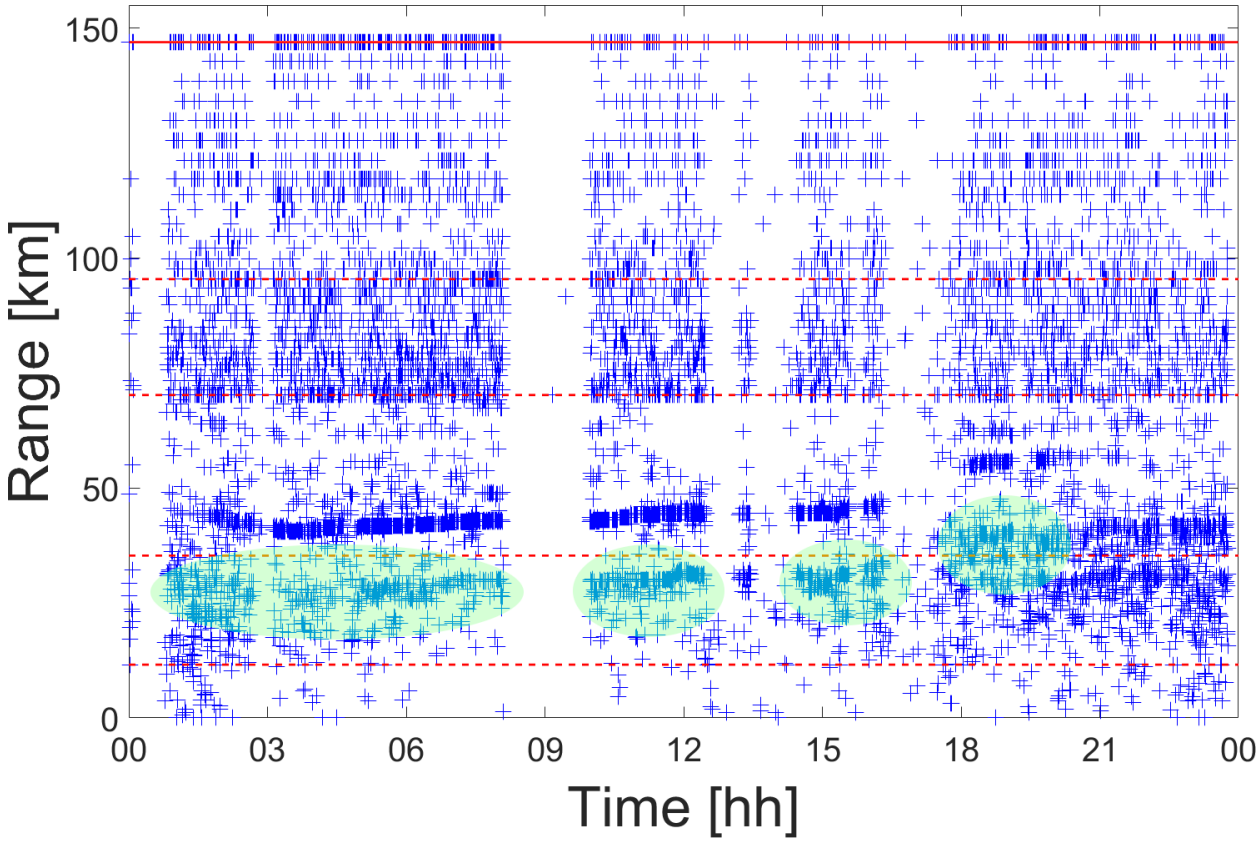

Figure 4.13: Estimated range (blue + symbols) for detections made on March 3 2014 that passed frequency and phantom pruning and whose ambiguity exceeds 0.1. The green regions highlight a phantom track at lower range to the actual track. The edges of double option zones are indicated by dashed red lines. The max candidate source range is indicated by the solid red line.

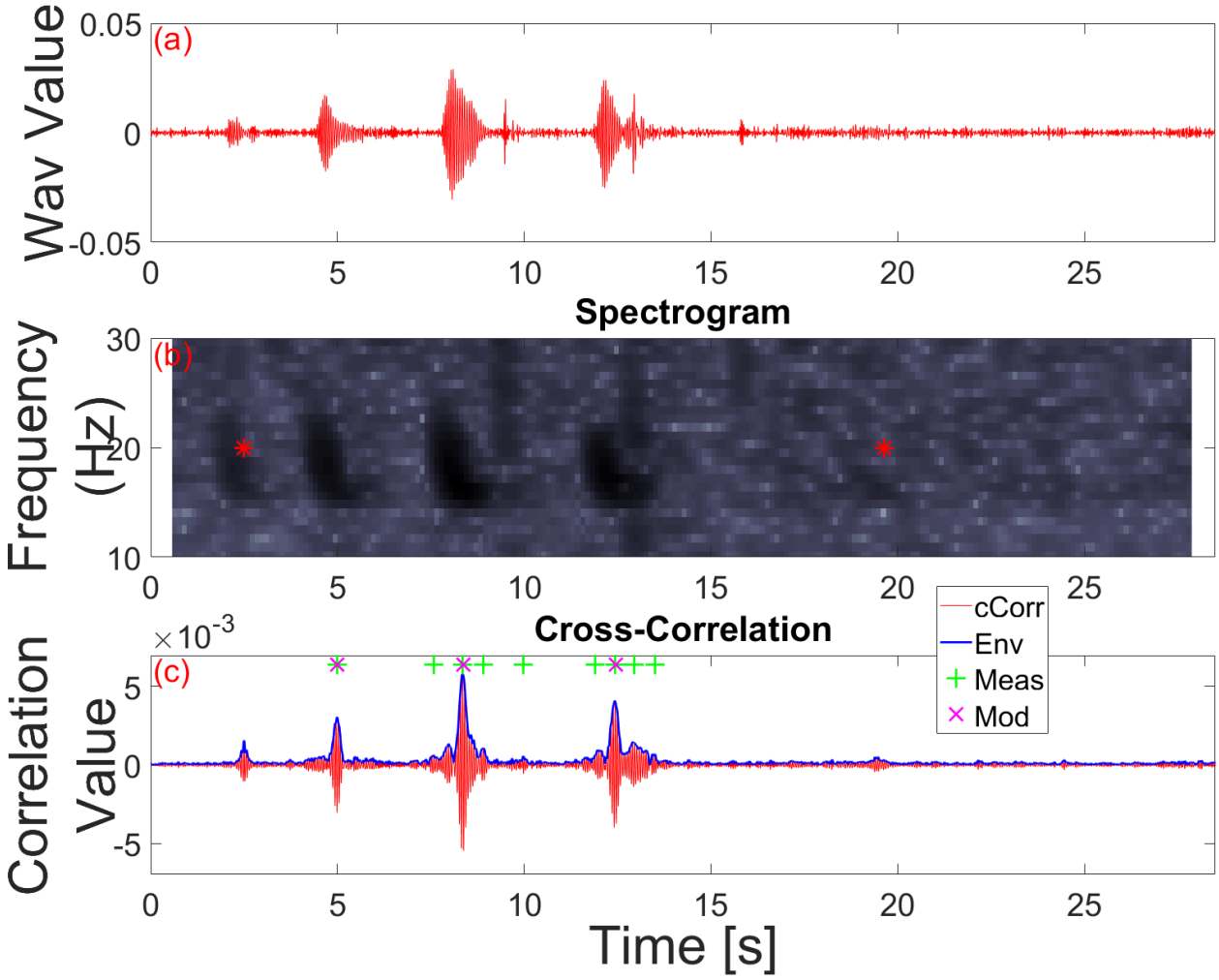

Figure 4.14: Information on the first call in the calling bout beginning at 10h00 in Figure 4.12. (a) Recorded waveform. (b) Spectrogram of the time series shown in the top subplot, with detections highlighted by red $*$ symbols. (c) Cross-correlation (red trend), cross-correlation envelope (blue trend), and measured (green $+$ symbols) and model-predicted (magenta x symbols) for the optimum match.

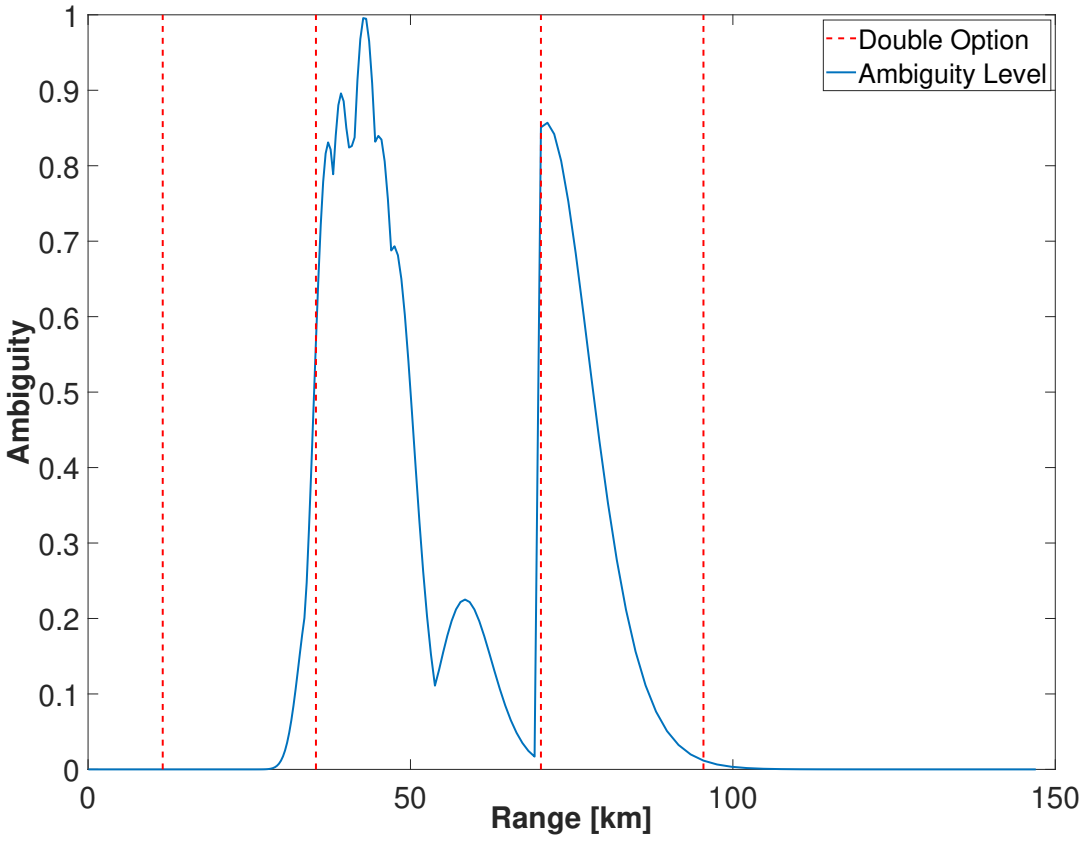

Figure 4.15: Ambiguity vector for the detection shown in Figure 4.14.

Pruning Status	# of Detections
Unpruned	41,481,171
Frequency	13,069,837
Frequency + Phantom	3,445,568

Table 4.1: Total fin whale 20 Hz call detections from 2007-2017 in three different pruning configurations.

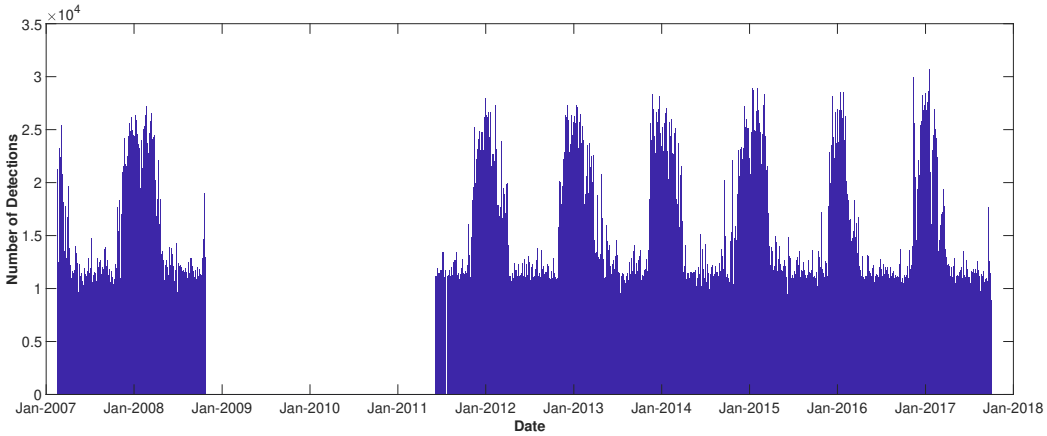

Figure 4.16: Number of fin whale 20 Hz call detections in acoustic data recorded by the ACO hydrophone. Results from February 2007 until September 2017 as recorded by the ACO hydrophone are plotted.

4.3 Fin Whale Call Detection Rates

In this section, variability in fin whale call detection rates over the ACO data set will be explored. Table 4.1 lists the total number of 20 Hz call detections across all years of data, the number of detections which satisfy the frequency pruning criteria, and those which satisfy both frequency and phantom pruning.

The number of detections in each day of data is shown in Figure 4.16 for the unpruned data set. A clear seasonality is present in this figure, with more detections during winter months than at other times of the year. Figure 4.17 shows daily detection totals after frequency pruning is applied. This pruning dropped the number of noisy detections between calling seasons down to a small fraction of what they were prior to pruning. The January 2016 and January 2017 calling seasons had a higher proportion of calls pruned away by the frequency pruning than other seasons.

Finally, daily detection totals after both frequency and phantom pruning are implemented are shown in Figure 4.18. Phantom pruning reduced the daily detection totals in all parts of the data, but most visibly during the fin whale calling seasons when higher order arrivals that satisfy the frequency pruning criteria are eliminated.

Figure 4.19 plots the number of detections, after frequency and phantom pruning, in each month and hour of the data. In this figure, each month between January 2007 and January 2018 has a single row. In each row, the total number of detections in each hour of the day are plotted as an individual histogram. As also shown in Figure 4.18, the reduced call totals in the January 2016 and January 2017 calling seasons is apparent in Figure 4.19. Additionally, there is no obvious tendency for more fin whale calls to be detected in one hour of the day over any other. These results agree with published accounts of fin whale call occurrence

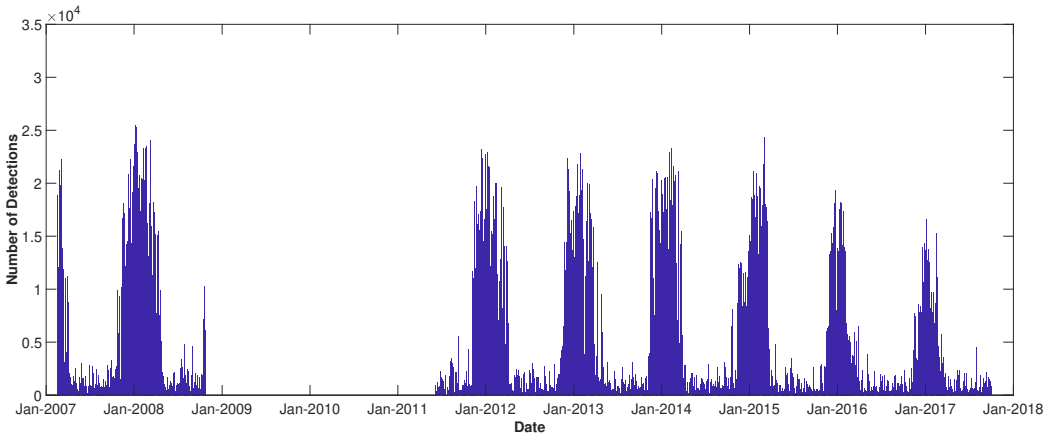

Figure 4.17: Number of fin whale 20 Hz call detections once frequency pruning has been applied. Results from February 2007 until September 2017 as recorded by the ACO hydrophone are plotted.

as a function of month near Bermuda, Hawaii, and Midway Island shown earlier in this chapter in Figures 4.2, 4.3, and 4.5.

Figure 4.20 plots total number of detections in each month of the ACO data set as a function of ambiguity threshold for the frequency and phantom pruning case. Thus, the far left column of this figure has the total number of detections in each month with no ambiguity threshold. This figure shows that a small ambiguity cutoff of 0.1 achieves a considerable rejection of detections over the case where no ambiguity cutoff is used.

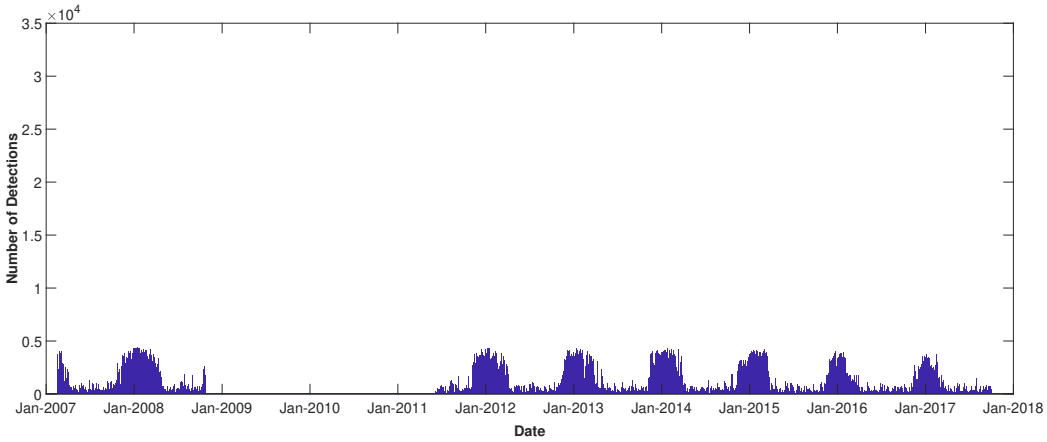

Figure 4.18: Number of fin whale 20 Hz call detections once frequency and phantom track-based pruning have been applied. Results from February 2007 until September 2017 as recorded by the ACO hydrophone are plotted.

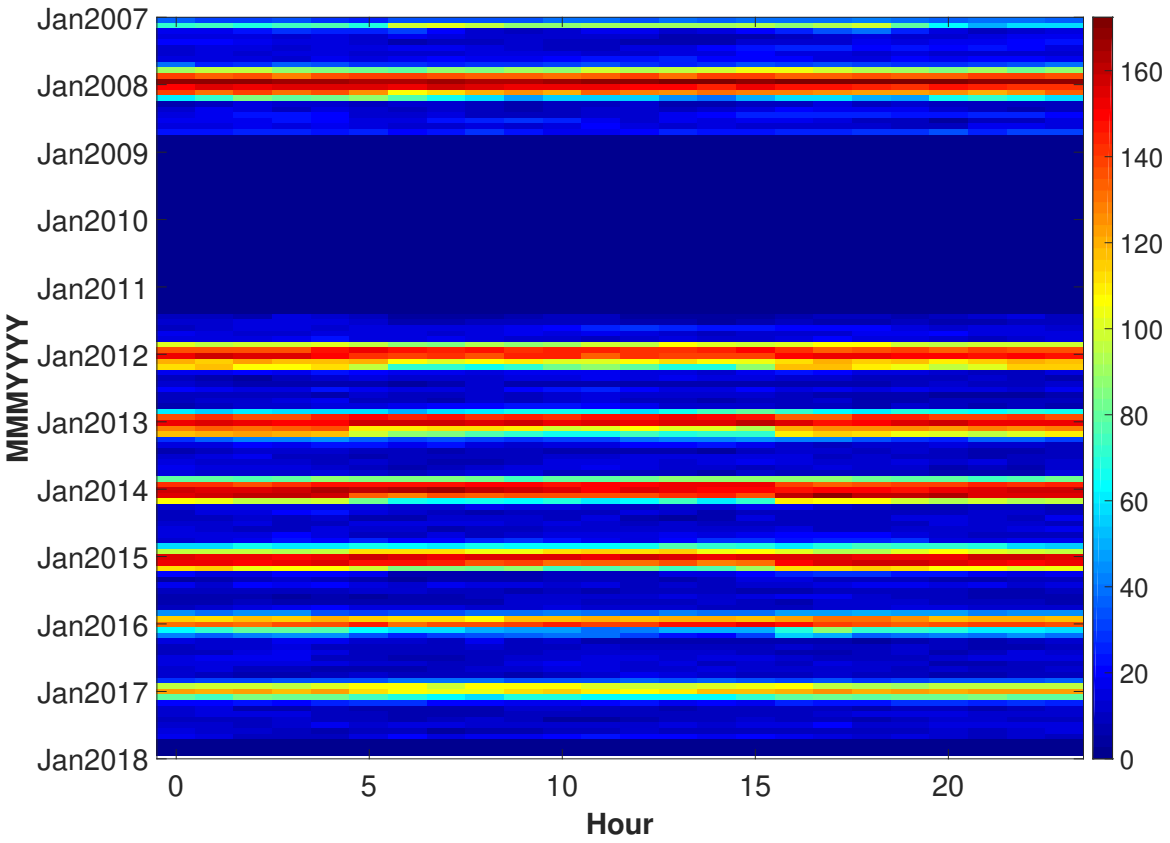

Figure 4.19: Number of fin whale 20 Hz call detections, once frequency and phantom track-based pruning have been applied, in each month and hour of the day (UTC). Results from February 2007 until September 2017 as recorded by the ACO hydrophone are plotted. Local sunrise and sunset times are approximately 16h00 and 04h00 (UTC), respectively.

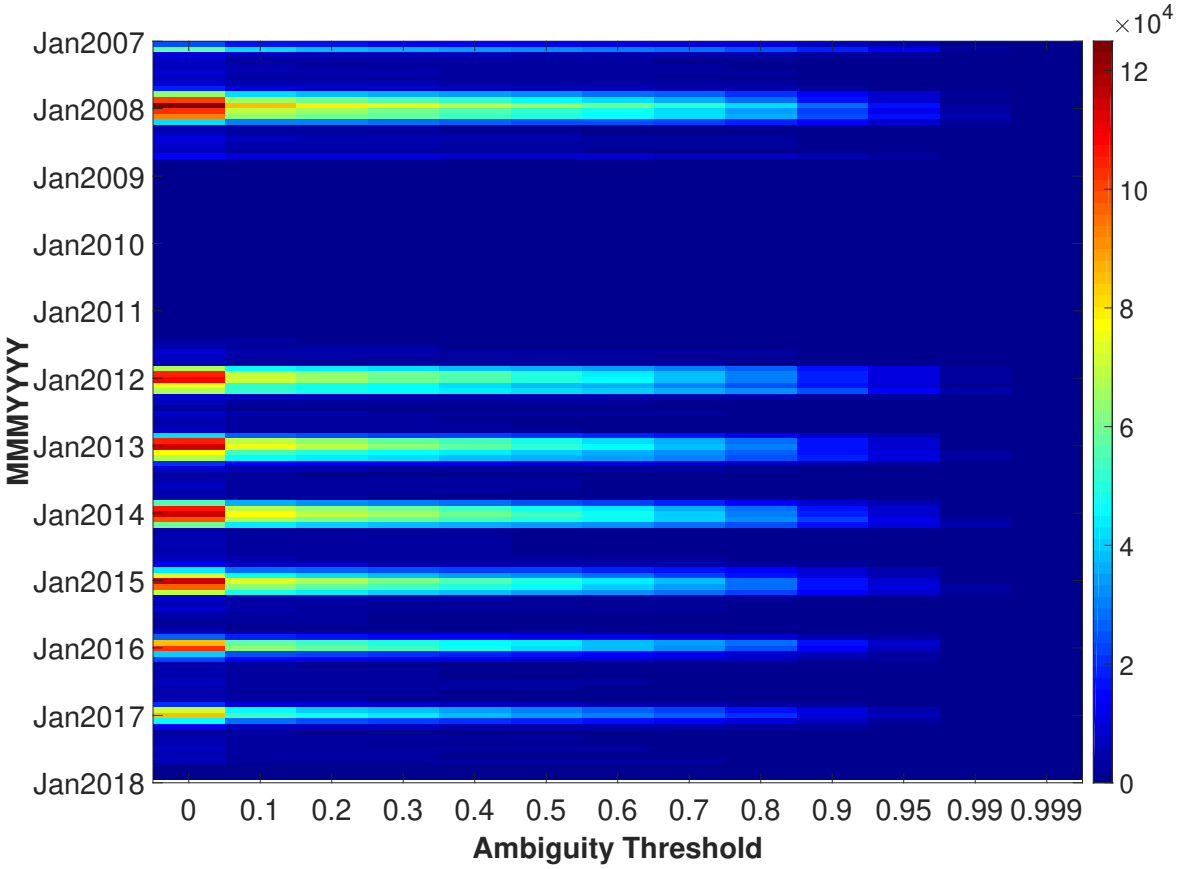

Figure 4.20: Number of fin whale 20 Hz call detections, once frequency and phantom track-based pruning have been applied, sorted by month and ambiguity level. Results from February 2007 until September 2017 as recorded by the ACO hydrophone are plotted.

4.4 Fin Whale Call Characteristics

In this section, selected fin whale 20 Hz call characteristics assessed over the 2007-2017 ACO data set are presented. Frequency characteristics are assessed using the power spectral density (PSD) for the 2.24 second waveform window about each detection. The call characteristics evaluated in this section are:

- Peak frequency (f_{peak}, the frequency of highest power spectral density value energy within the detected call)
- Start frequency (f_{start}, the highest frequency lower than f_{peak} whose PSD value is 6dB lower than peak PSD value)
- Stop frequency (f_{stop}, the lowest frequency greater than f_{peak} whose PSD value is 6dB lower than peak PSD value)
- Inter-call interval (ICI, the time spacing between consecutive calls in a calling sequence)

The composite histograms in the three sections that follow (Section 4.4.1 - 4.4.3) are normalized in the same fashion. Each column presents a histogram based on the detections from a single year. After calculating raw histogram bin counts, the set of counts for each year is individually normalized such that the sum of normalized bin counts in each year equals 1. Each year is individually normalized so that differences in number of detections between years do not change the relative shade of plotted histogram bins.

The figures without pruning are shown with no ambiguity threshold, and thus incorporate all detections regardless of fit to the measured data. The figures incorporating pruning, either frequency or frequency and phantom pruning, include an ambiguity threshold of 0.1. This serves to reduce the impact of noisy detections on call characteristic estimates.

4.4.1 Peak frequency

Peak frequency is assessed during the detection phase, and is calculated as the peak of the power spectral density (PSD) for a time series window about each detection. After the PSD is calculated, the PSD for a given detection is upsampled by a factor of ten before frequency characteristics of the PSD are extracted. This upsampling is done to increase the frequency resolution for the estimates of f_{peak}, f_{start}, and f_{stop}.

Figure 4.21 plots histograms of estimated f_{peak} across all detections and years in the pre-pruned data set. The horizontal banding is considered a processing artefact rather than a physically meaningful part of the underlying data. The frequency of each narrow dark blue horizontal band corresponds to a frequency found in the pre-interpolated PSD calculated by the fin whale call detector, and each of these pre-interpolated PSD frequencies has a dark blue band in this f_{peak} histogram. The largest proportion of detections from 2007 to 2015 have a peak frequency of 16.67 Hz. In 2016 and 2017 the frequency with the highest proportion of detections shifts to 30.06 Hz (which is close to the upper limit of the passband selected for the bandpass filter).

Figure 4.22 plots f_{peak} for every detection in the frequency-pruned data set. As with the unpruned figure, each year (column) is individually normalized so the histograms for years with fewer overall detections aren't unjustly influenced. The lower number of noisy detections in this set compared to the unpruned set makes the overall peak at 19.5 Hz more apparent, although this peak is less prominent in 2016 and 2017.

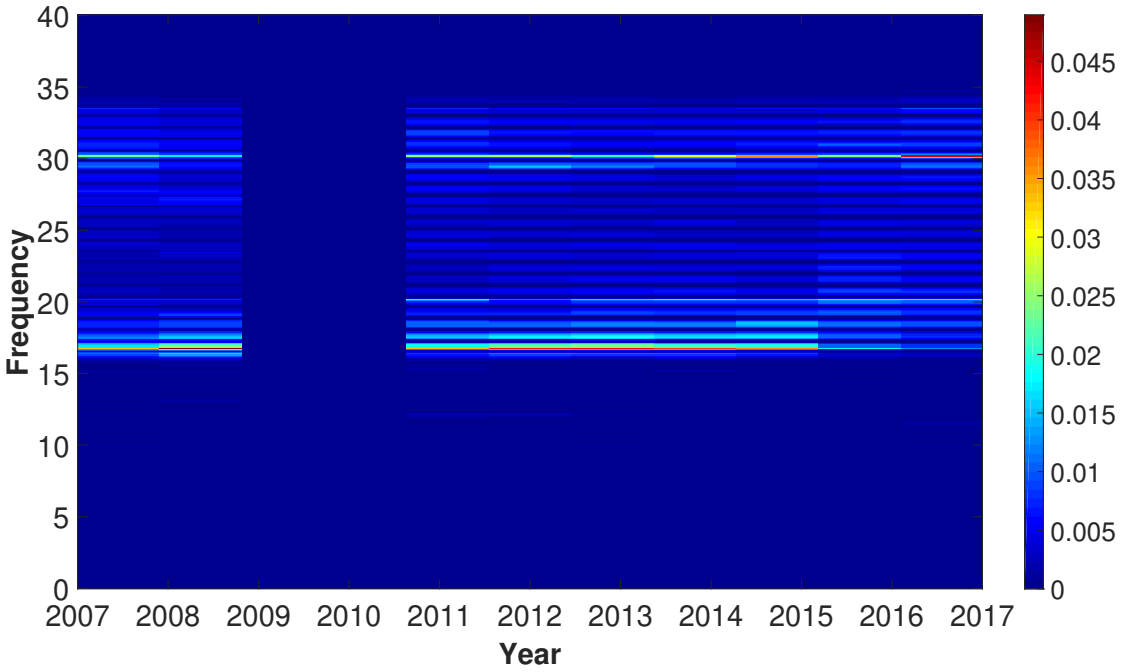

Figure 4.21: Histogram of estimated peak frequency in each year of data. No detection pruning is applied. Each year is individually normalized such that the sum of histogram bins in each year equals 1. No ambiguity threshold is applied.

Figure 4.23 plots f_{peak} for every detection in the frequency + phantom pruned set. The 16.67 Hz peak remains present in all years, although it is less apparent than for the pruning case where only frequency pruning was carried out. This could be because phantom pruning is particularly effective at eliminating detections of higher order arrivals from the set of detections. Thus, each individual fin whale 20 Hz call is expected to be represented by fewer overall detections in the frequency + phantom pruned set than in the frequency pruned set. Since these higher order arrivals are expected to have similar frequency characteristics to the earliest arrival of a given call, reducing the incidence of multiple detections of the same calls should make the peak associated with the peak frequency less prominent.

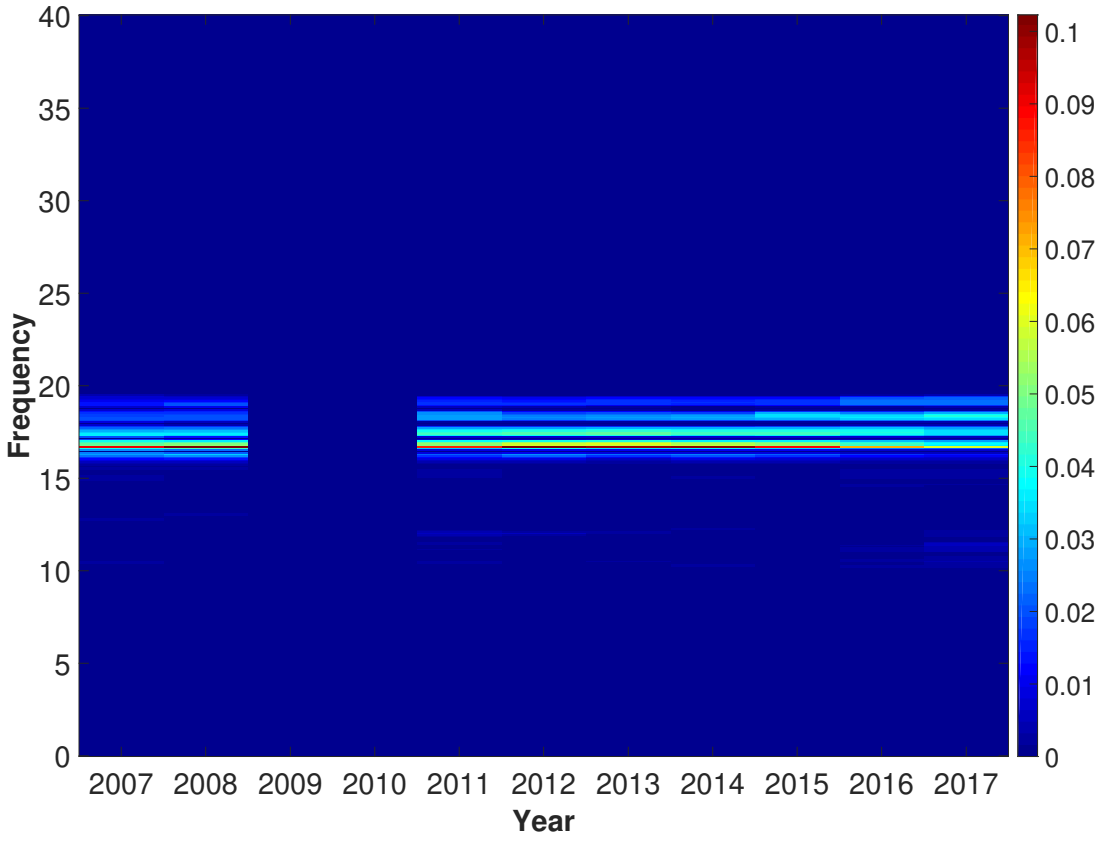

Figure 4.22: Histogram of estimated peak frequency in each year of data. Frequency pruning is applied. Each year is individually normalized such that the sum of histogram bins in each year equals 1. No ambiguity threshold is applied.

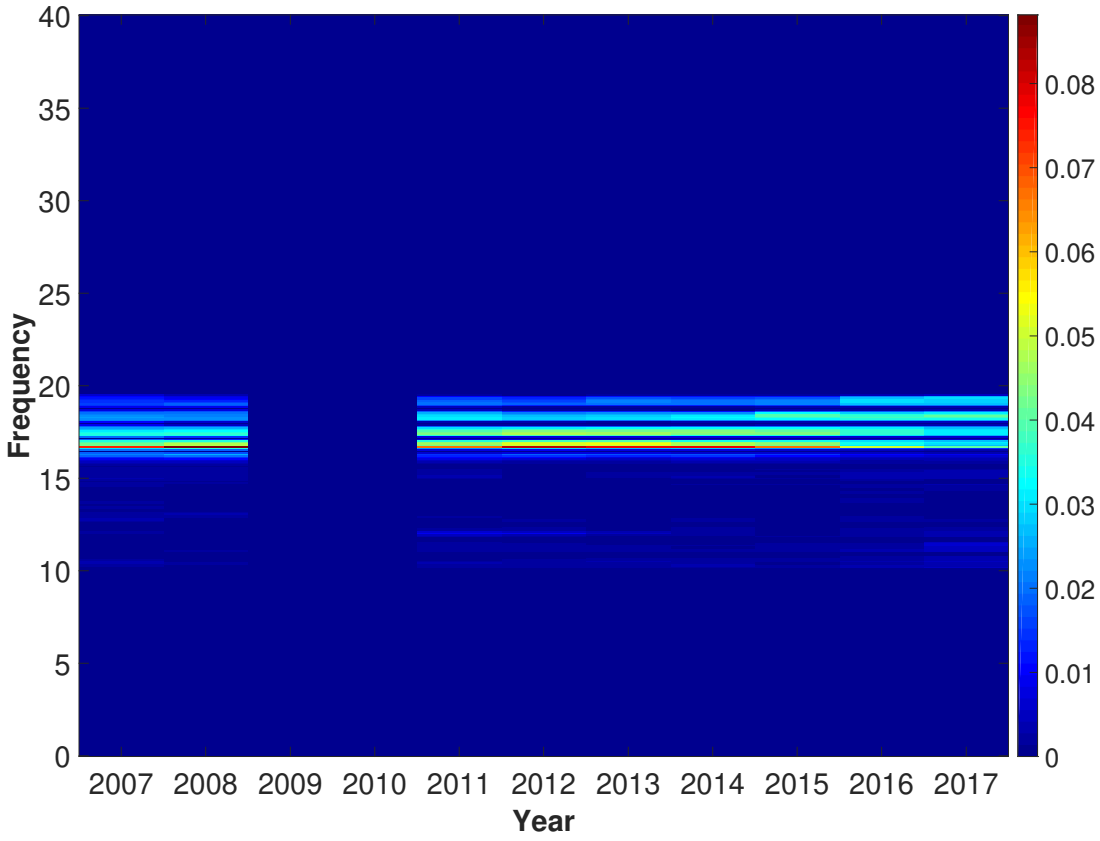

Figure 4.23: Histogram of estimated peak frequency in each year of data. Frequency and phantom pruning is applied. Each year is individually normalized such that the sum of histogram bins in each year equals 1. No ambiguity threshold is applied.

4.4.2 Start frequency

Figure 4.24 plots f_{start} across all years for the pre-pruned data set. As with peak frequency, all the figures in this section are individually normalized by year to make comparisons between years easier. The f_{start} histogram peak is at 14.83 Hz from 2007-2015, although as with the peak frequency analysis the peak for 2016 and 2017 is at a higher frequency (16.67 Hz) in 2016 and 2017. Faint banding is present in the 15-30 Hz frequency range, although it is less vivid than the peak frequency banding. This faint banding is likely caused by the peak frequency banding and the correlation between peak and start frequency. Since start frequency is defined as the frequency whose PSD level is 6dB lower than the peak frequency PSD value, introducing a bias into peak frequency by making some frequencies less likely than others to be picked as the peak frequency for a given call should also introduce a bias into start and stop frequency. These f_{start} and f_{stop} biases may be different sizes than the size of the f_{peak} bias as they will be influenced not only by the magnitude and direction of the f_{peak} bias but also the shape of the call spectrum. These dual influences yield banding in f_{start} and f_{stop} that is less prominent than that in f_{peak} due to the effect of call spectrum shape.

Figure 4.25 plots f_{start} across all years for the frequency-pruned data set. The diminished role of noise in this pruned data set compared to the unpruned set shown in Figure 4.24 yields a more consistent f_{start} peak across all years, with a significant secondary peak at 16.67 Hz and broader distribution of f_{start} yielding a less prominent overall peak in 2016 and 2017. Slight variation in the f_{start} peak between years is present, although the coarse frequency resolution of these data makes it unclear if this small variability is significant.

Figure 4.26 plots f_{start} across all years for the frequency- and location-pruned data set. As with the f_{peak} results in Figure 4.23, the histogram peak in each year is broader for the frequency and phantom pruning case than when frequency pruning alone is used. Additionally, the secondary histogram peak at 16.67 Hz remains present in these histograms.

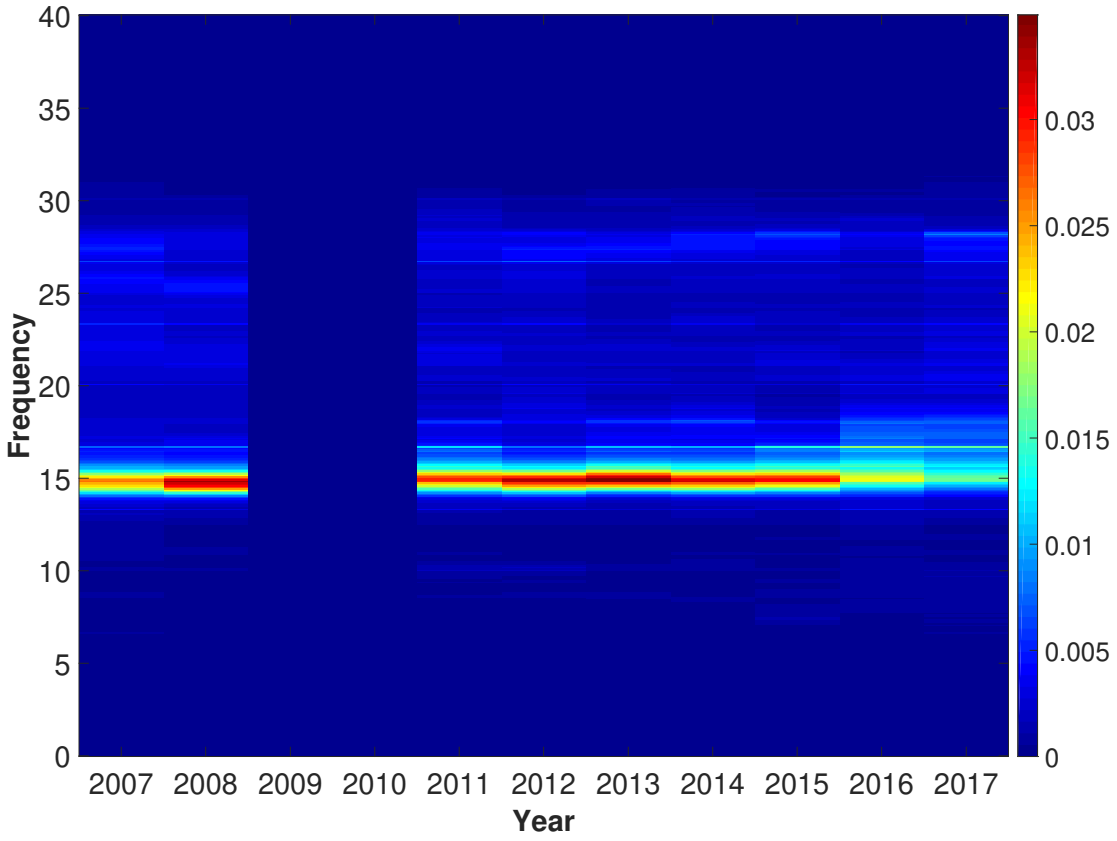

Figure 4.24: Histogram of estimated start frequency in each year of data. No detection pruning is applied. Each year is individually normalized such that the sum of histogram bins in each year equals 1. No ambiguity threshold is applied.

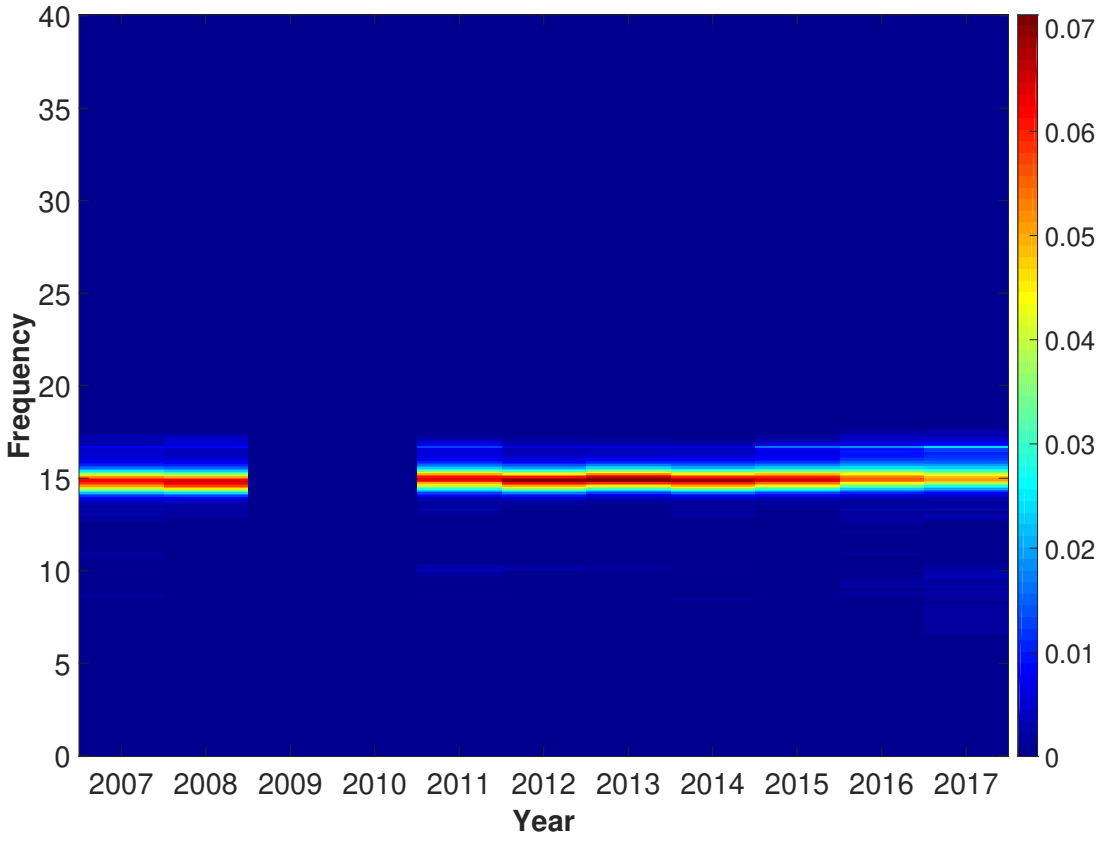

Figure 4.25: Histogram of estimated start frequency in each year of data. Frequency pruning is applied. Each year is individually normalized such that the sum of histogram bins in each year equals 1. No ambiguity threshold is applied.

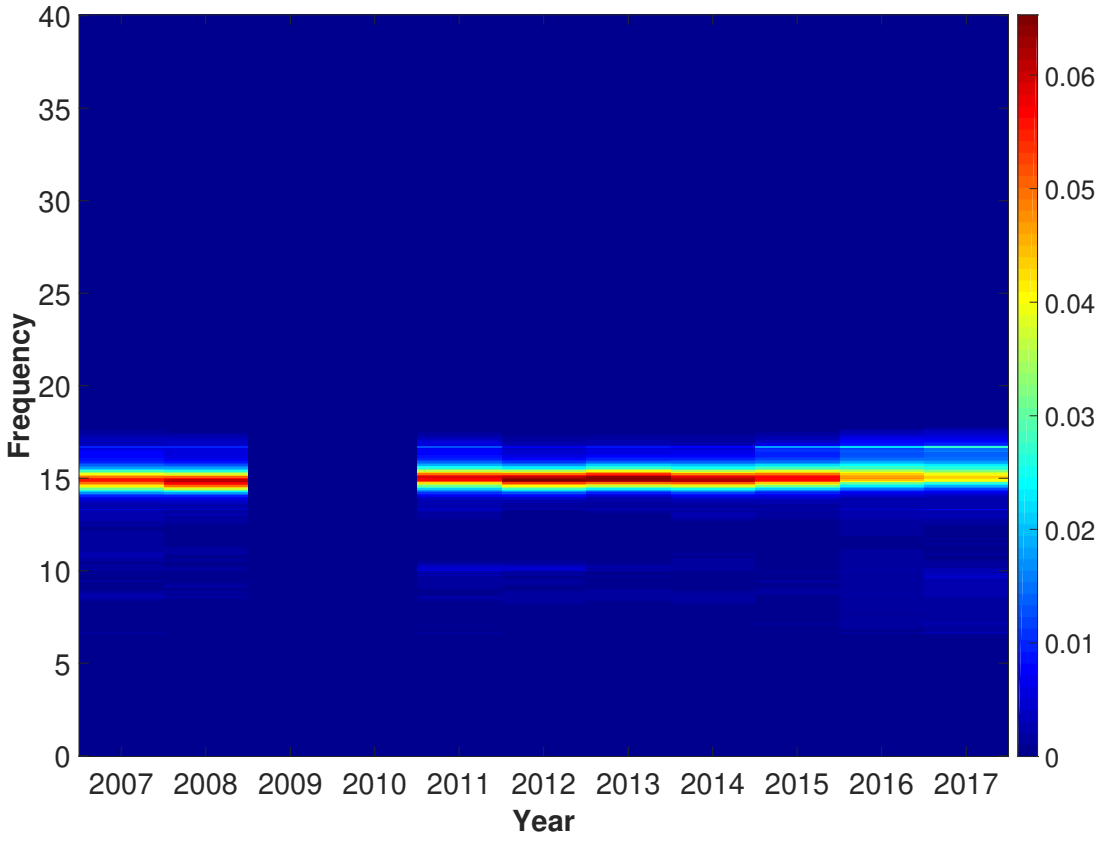

Figure 4.26: Histogram of estimated start frequency in each year of data. Frequency and phantom pruning is applied. Each year is individually normalized such that the sum of histogram bins in each year equals 1. No ambiguity threshold is applied.

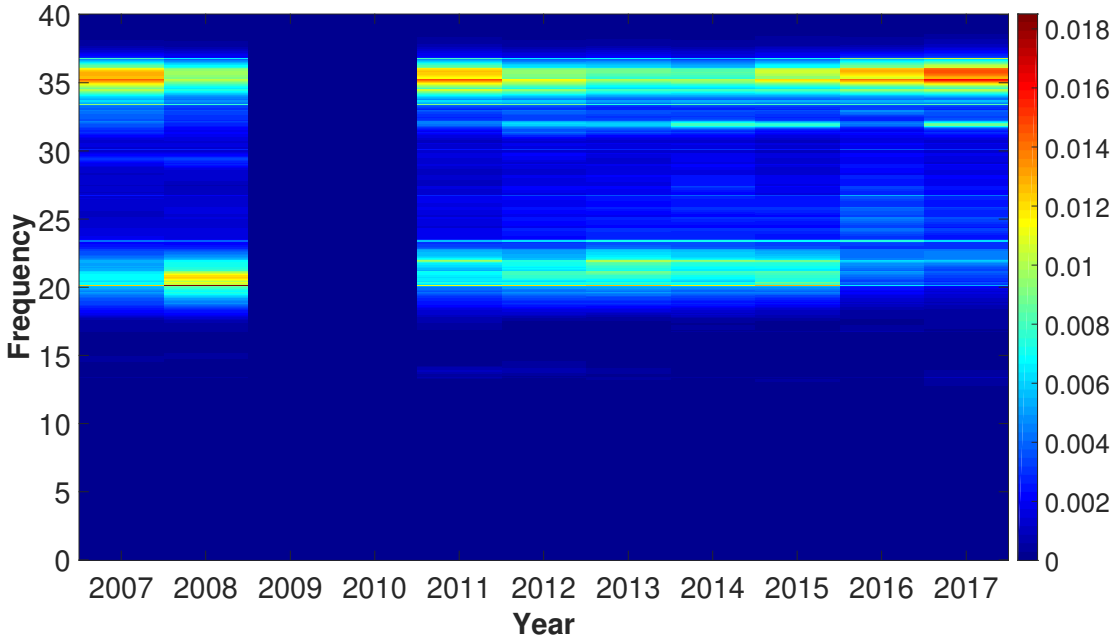

Figure 4.27: Histogram of estimated stop frequency in each year of data. No detection pruning is applied. Each year is individually normalized such that the sum of histogram bins in each year equals 1. No ambiguity threshold is applied.

4.4.3 Stop frequency

Figure 4.27 plots f_{stop} across all years for the pre-pruned data set. As with the f_{start} plots in Section 4.4.2, banding is present in each year of the data, and is likely caused by banding in the estimated peak frequency from call spectrum interpolation. The significant peaks reside in one of two bands: 18-22 Hz and 33-37 Hz.

Figure 4.28 plots f_{stop} across all years for the frequency-pruned data set. The collection of peaks near 35 Hz were suppressed by the frequency pruning, leaving only the set of peaks near 20 Hz. All calls whose f_{stop} exceeds 24 Hz were eliminated by the frequency pruning. A consistent f_{stop} peak of 20.04 Hz is found across all years.

Figure 4.29 plots f_{stop} across all years for the frequency and phantom pruned data set. Similar to what was observed with f_{start} in Section 4.4.2, the primary f_{stop} peak at 20.04 Hz is smaller compared to the surrounding histogram bin values when phantom pruning is implemented compared to frequency pruning alone. The banding is also more prominent when location pruning is implemented.

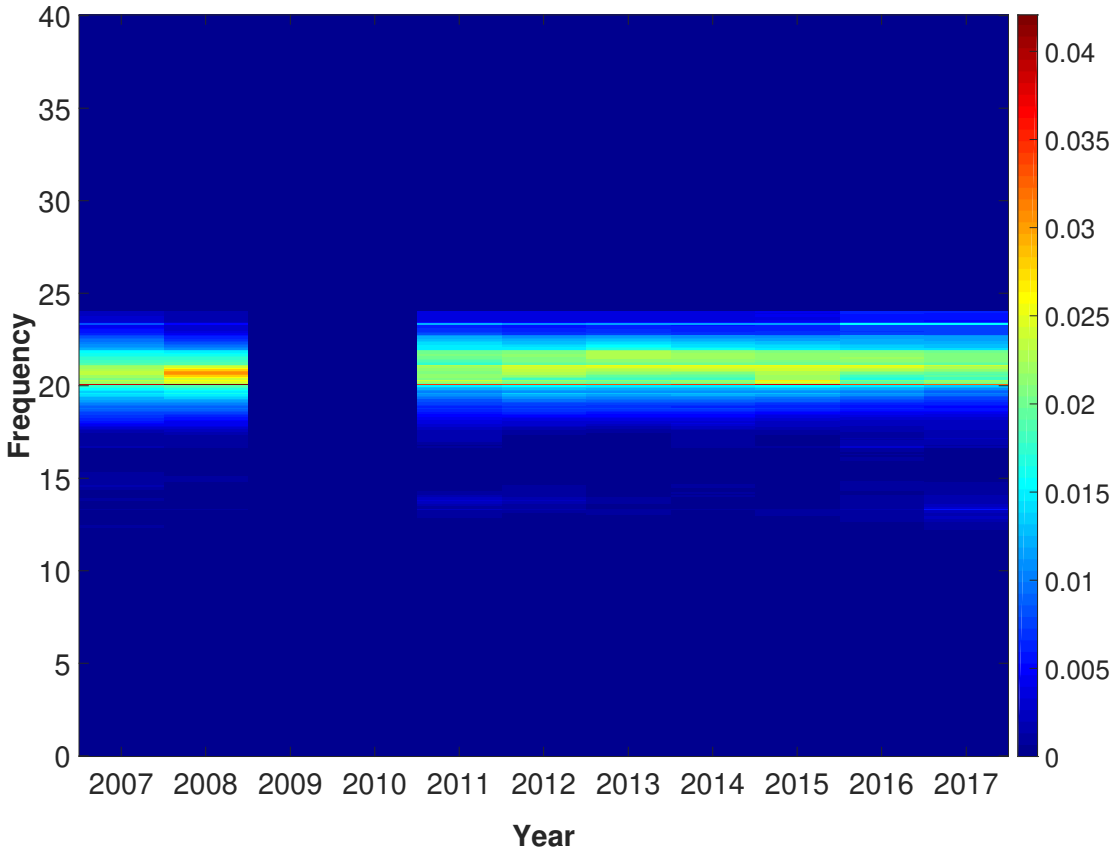

Figure 4.28: Histogram of estimated stop frequency in each year of data. Frequency pruning is applied. Each year is individually normalized such that the sum of histogram bins in each year equals 1. No ambiguity threshold is applied.

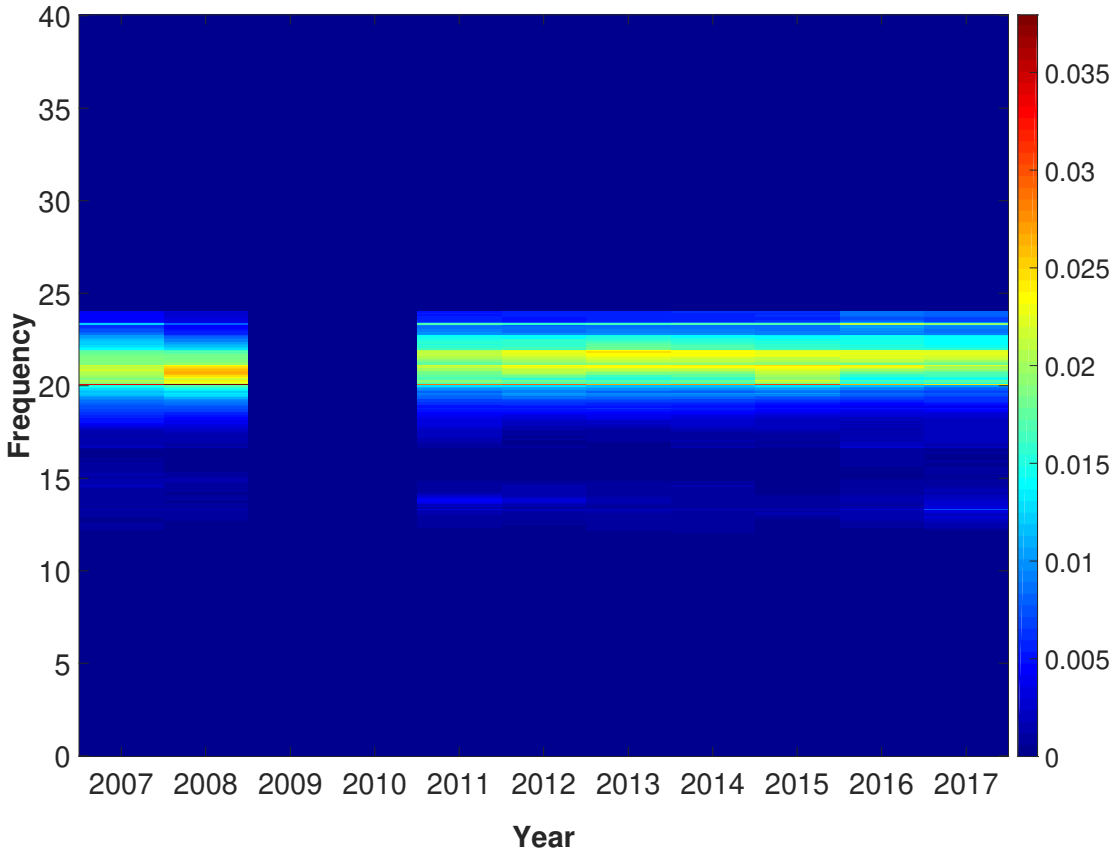

Figure 4.29: Histogram of estimated stop frequency in each year of data. Frequency and phantom pruning is applied. Each year is individually normalized such that the sum of histogram bins in each year equals 1. No ambiguity threshold is applied.

4.4.4 Inter call interval

Inter Call Interval (ICI) is defined as the time spacing between consecutive calls in a calling sequence. ICI is a useful bioacoustic parameter because it can be used to estimate number of calling animals from number of detections per unit time, which in turn can be used to estimate the animal density (animals per unit area).

Measuring ICI autonomously can be tricky, particularly in a reverberent environment as has been shown to exist about ACO. The permissive matched filter detector in this work yields a large number of false positive detections. As a result, simply calculating the time spacing between consecutive detections will yield Inter Detection Interval (IDI), but given that each call will typically have more than one detection (due to higher order arrival paths), estimating a reliable value for ICI based on IDI data can be problematic. The problem is further exacerbated by noisy arrivals. Figure 4.30 illustrates this phenomenon: For each month between January 2007 and January 2018, a histogram is calculated of the time spacing between consecutive detections that meet the current pruning and ambiguity threshold criteria. Thus, the more red the shade of a pixel, the more IDI values fall in that month. In this case, the frequency-pruned set (i.e., no calls whose f_{peak} exceeds 19.5 Hz, or whose f_{stop} exceeds 24 Hz) is used with an ambiguity threshold of 0 (i.e., no calls were excluded based on ambiguity). Effectively, this means that the data in this figure were only processed by a detector + classifier pair, with no localization information incorporated. In this figure, a consistent peak at IDI=0.6 s is present in each calling season alongside a shorter and broader peak around 2.8 s. The minimum model-predicted RAT value is 1.3 seconds (for a range of 147 km), so the 0.6 s peak is unlikely to be the time spacing between fin whale arrivals of the same call. It could be a different type of call that has frequency characteristics compatible with the frequency pruning, though. Figure 4.31 shows an example of this type of call. Arrows highlight two time intervals of interest on the spectrogram in the middle subplot. The first arrows show a spacing of 3.025 s between detections of this call type, while the second set of arrows shows a shorter spacing of 0.735 s. These values agree with the main and secondary peaks in Figure 4.30. These types of small-time-interval calls are found throughout the data set.

Figures 4.32 - 4.35 show IDI for the frequency pruning case with a selection of different ambiguity thresholds. Note the changing color bar scaling as ambiguity increases, which reflects the smaller data set size with increasing ambiguity threshold. As the ambiguity threshold increases, the main peak at 0.6 seconds remains prominent, while the secondary peak near 2.8 s fades away. Additionally, a tertiary peak, lower than the other two, moves toward high IDI as ambiguity increases. The wide time extent of this peak, and the lack of a consistent peak location with varying ambiguity threshold makes it difficult to draw conclusions based on this tertiary peak.

Figure 4.36 shows the set of calculated IDI values for each month when both frequency and phantom pruning are applied. No ambiguity threshold is used in this figure, so every call which satisfies both the frequency and phantom pruning criteria is included regardless of how good the fit between measured and model-predicted data is. As with the other IDI figures in this section, no normalization is applied here, which is why the peaks for January 2016 and January 2017 are noticeably lower (due to lower overall detection counts in those calling seasons) than the other calling seasons. Unlike Figure 4.30, there is no peak at IDI=0.6 s since the phantom pruning reliably eliminates the second call in each of the closely-spaced pairs of calls. The secondary peak at 2.8 s is also eliminated when phantom pruning is applied, possibly because the corresponding detections are used for localization of an earlier call and so are pruned away during phantom pruning.

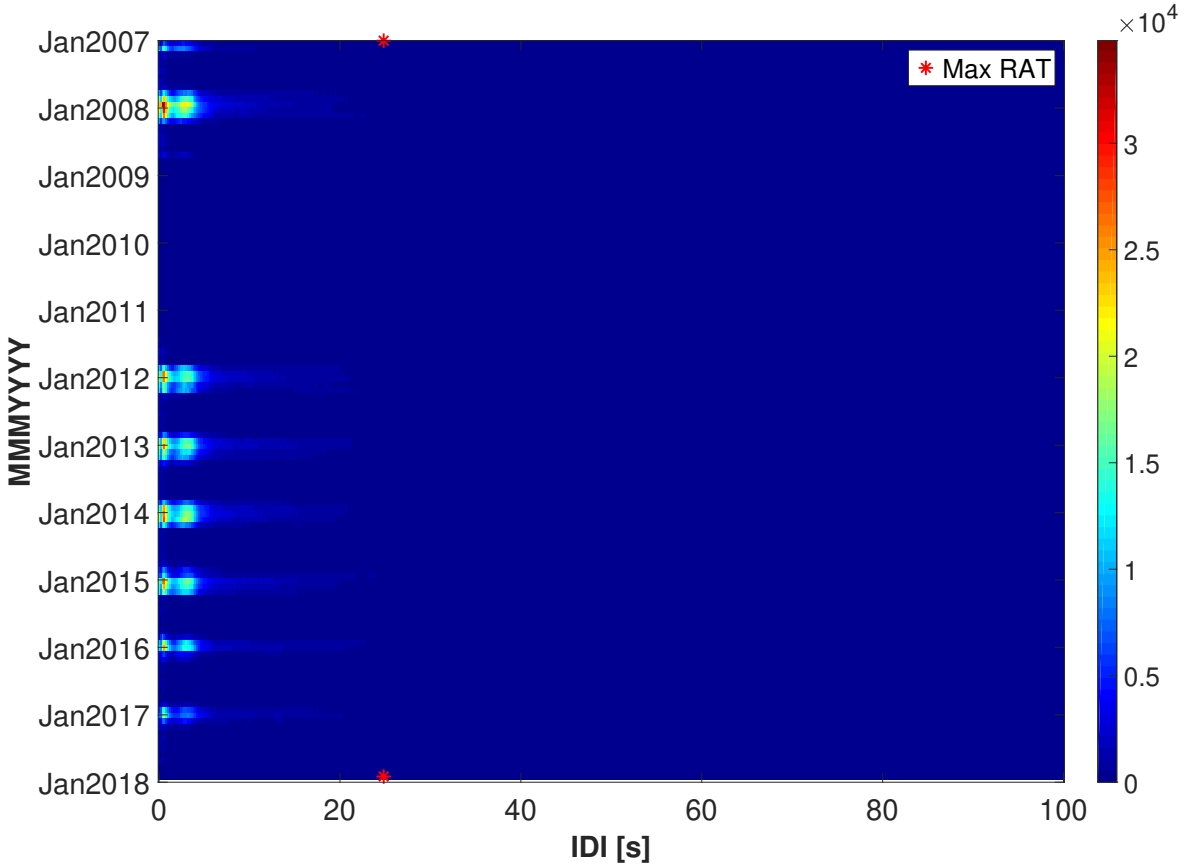

Figure 4.30: Calculated inter detection interval (IDI) values for each month of data between January 2007 and January 2018. One histogram of IDI values is calculated for each month. The more red a given pixel is, the greater the number of IDI values are found in that month and at that IDI value. Frequency pruning was applied to these data, no ambiguity threshold is used, and no normalization is carried out. The maximum model-predicted RAT value (24.83 s) is indicated by the pair of red * symbols.

The main peaks in Figure 4.36 are at approximately 6.8 s, 20.08 s, and 28.11 s, with the last peak being the largest during each calling season. One possibility for the peak at 6.8 s is that 6.8 s is slightly larger than the largest RAT from the 2-RAT range set (i.e., the set of ranges comprising the union of medium-high double option zone and the high range zone). The highest RAT in this set is the BS4-BS2 RAT (5.2 s) at a range of 75.6 km. Allowing for the phantom track blanking interval of 1.25 s, the earliest time after this RAT that a detection could fall and not be pruned by the phantom track criterion is 6.45 s, which is well within the IDI peak with the smallest IDI value (6.8 s).

The IDI peak around 20.08 s, and the lower IDI histogram values on either side, may be related to the interaction of noisy measured RAT sets (that are predisposed to being at high range due to it being easier to find a noisy match to 2 model-predicted RATs than 3 or 4 model-predicted RATs) and the non-uniform distribution of RATs as a function of RAT time (i.e., the spacing of one RAT to the nearest one is smallest at short range). Further investigation is required to investigate this hypothesis, possibly through a simulation

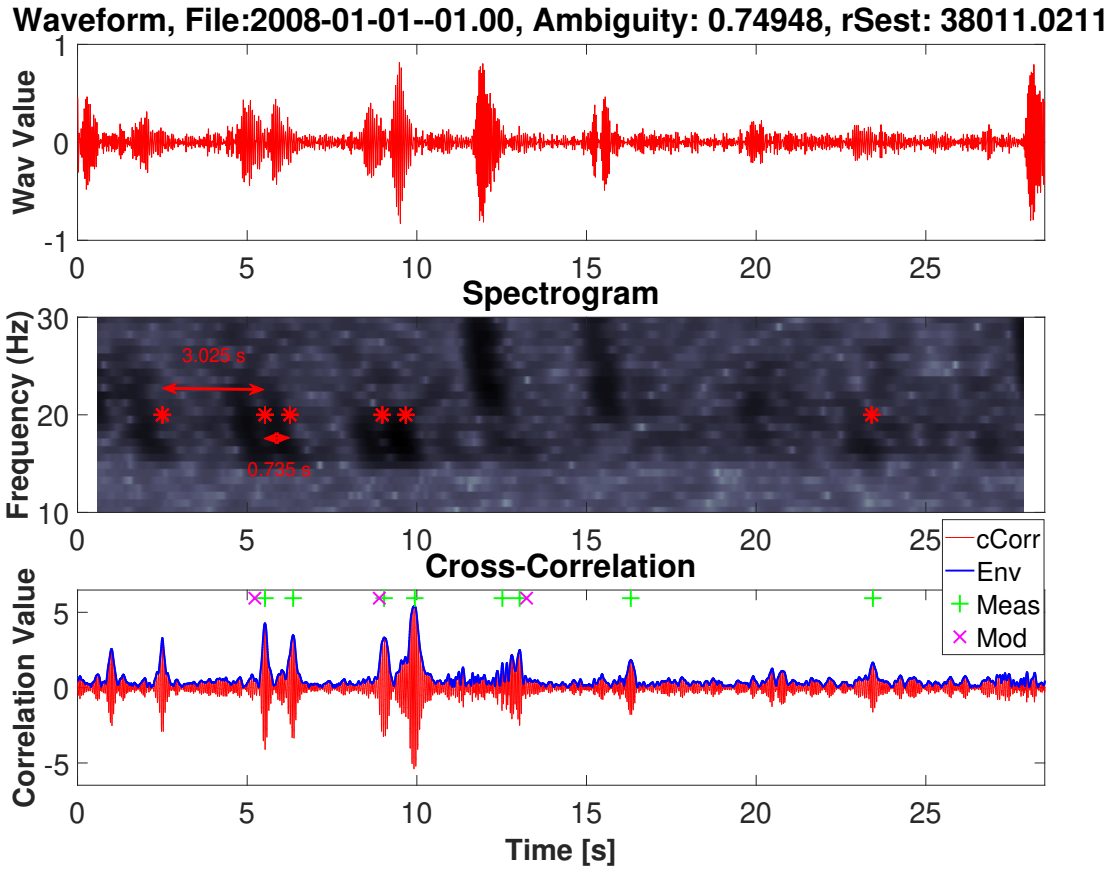

Figure 4.31: An example of the inter-detection intervals seen in Figure 4.30. The recorded waveform (top), spectrogram with detector output highlighted (middle), and cross-correlation (bottom) are shown. Red * symbols on the spectrogram are the output from the matched filter detector. Arrows on this spectrogram highlight two inter-detection intervals that match the peaks seen in the previous IDI figure.

study.

In Soule and Wilcock [2013], Northeast Pacific Ocean fin whale tracks were studied using a set of ocean bottom seismometers. Inter-call interval was evaluated for 154 individual tracks identified in over 800 hours of recordings. Figure 4.37 presents a histogram of inter-call interval across all of the tracks in Soule and Wilcock [2013]. More tracks had an inter-call interval of 24 seconds than any other value, with ICI values as low as 7 and as high as 35 observed. The shape of the Soule & Wilcock histogram in Figure 4.37 bears considerable similarity to the shape of our IDI histogram in Figure 4.36. Both histograms show considerable histogram counts from just below 10 s to approximately 35 s. The peak in the Soule & Wilcock histogram appears at 24 s, while Table 4.2 presents the single most common IDI value in any month for each of the 7 complete calling seasons present in the ACO data. For this work, a calling season is the period from November-April when most fin whale 20 Hz calls are observed at ACO. The peak histogram values in Figure 4.36 and Table 4.2 agree well with the results from Soule & Wilcock in Figure 4.37.

Figure 4.38 shows an enlargement of Figure 4.36 in the IDI= $18 - 35$ s range. In every calling season

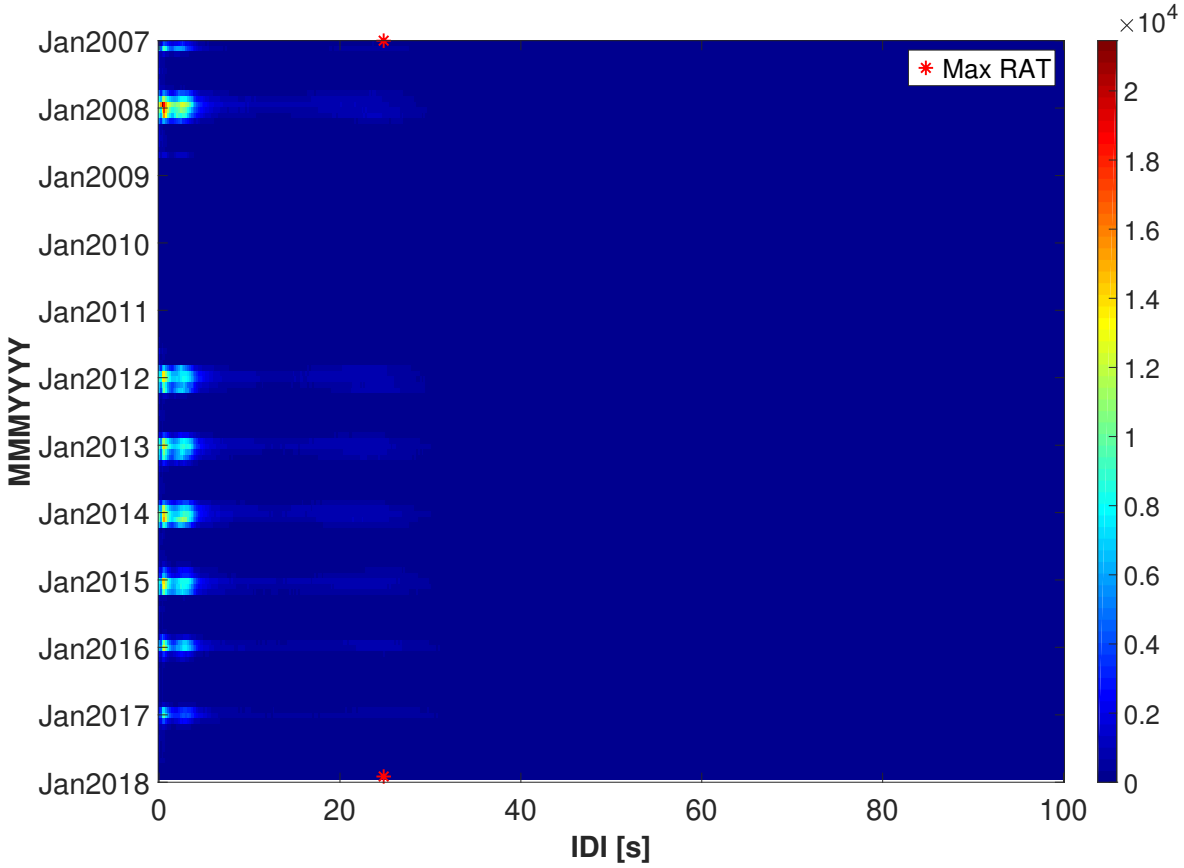

Figure 4.32: Calculated inter detection interval (IDI) values for each month of data between January 2007 and January 2018 and an ambiguity threshold of 0.3. One histogram of IDI values is calculated for each month. The more red a given pixel is, the greater the number of IDI values are found in that month and at that IDI value. Frequency pruning was applied to these data and no normalization is carried out. The maximum model-predicted RAT value (24.83 s) is indicated by the pair of red * symbols.

from January 2008 - January 2016 (excluding those in 2009, 2010, and 2011 for which there is no data), there appears to be a tendency toward longer IDI as the calling season progresses. The reduced number of calls in the January 2017 calling season makes extending this potential trend uncertain, however if inter call interval does change in a meaningful and systematic way for North Pacific fin whales, this could have implications on how fin whale density estimation is carried out in this region.

Figures 4.39 - 4.42 show IDI for the frequency + phantom pruning case for a selection of different ambiguity thresholds. The peak IDI just below 30 s is visible all all four of these ambiguity threshold levels, although the histogram match to the Soule & Wilcock histogram in Figure 4.37 is less obvious as ambiguity threshold increases. For the ambiguity threshold = 0.99 plot in Figure 4.42, the faint peaks near IDI = 60 and IDI = 90 could represent cases where one or two (respectively) calls in a bout fail to yield a high ambiguity location, thus increasing the measured IDI to 2 or 3 times the expected inter call interval.

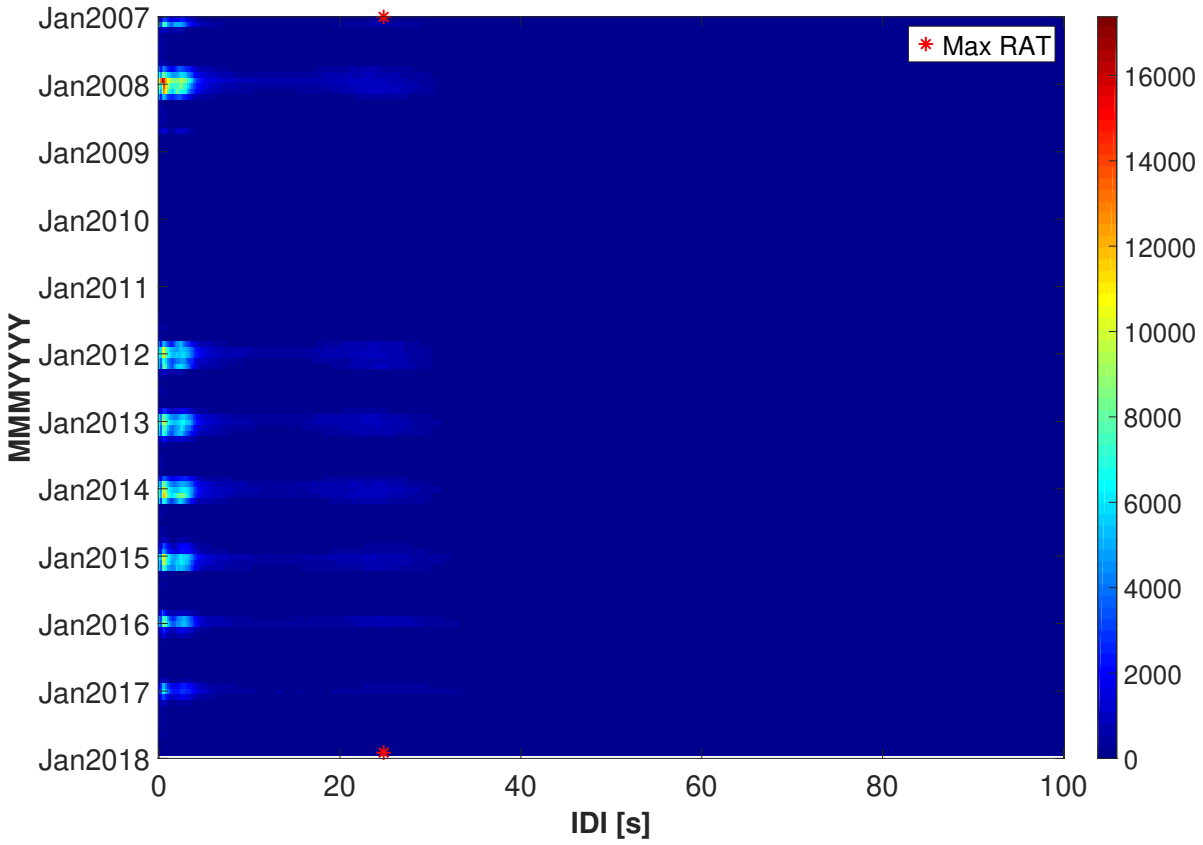

Figure 4.33: Calculated inter detection interval (IDI) values for each month of data between January 2007 and January 2018 and an ambiguity threshold of 0.6. One histogram of IDI values is calculated for each month. The more red a given pixel is, the greater the number of IDI values are found in that month and at that IDI value. Frequency pruning was applied to these data and no normalization is carried out. The maximum model-predicted RAT value (24.83 s) is indicated by the pair of red * symbols.

Calling Season	**Peak IDI** [s]
2007-2008	28.71
2011-2012	26.51
2012-2013	26.71
2013-2014	26.51
2014-2015	26.51
2015-2016	29.32
2016-2017	29.32

Table 4.2: Peak IDI value, from the frequency + phantom track-pruned data set, in any month of each calling season (November - April) as shown in Figure 4.36.

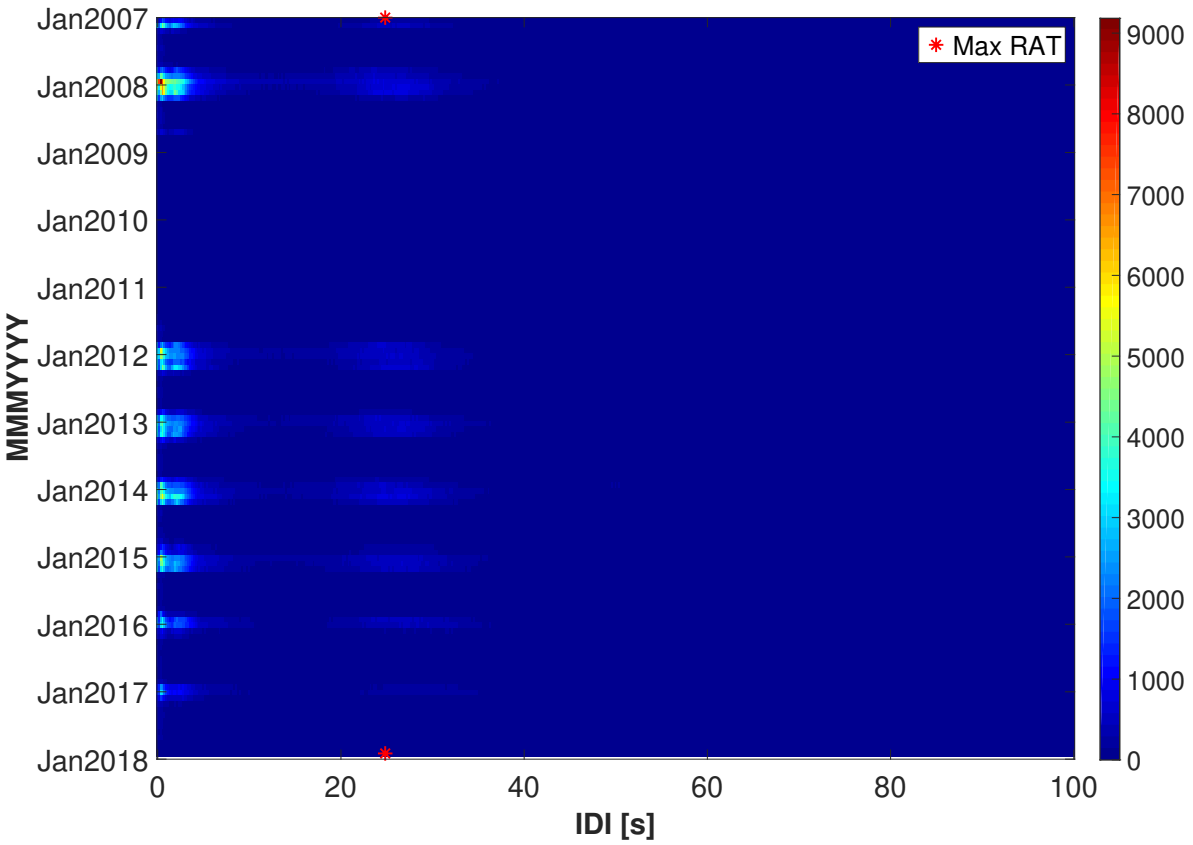

Figure 4.34: Calculated inter detection interval (IDI) values for each month of data between January 2007 and January 2018 and an ambiguity threshold of 0.9. One histogram of IDI values is calculated for each month. The more red a given pixel is, the greater the number of IDI values are found in that month and at that IDI value. Frequency pruning was applied to these data and no normalization is carried out. The maximum model-predicted RAT value (24.83 s) is indicated by the pair of red * symbols.

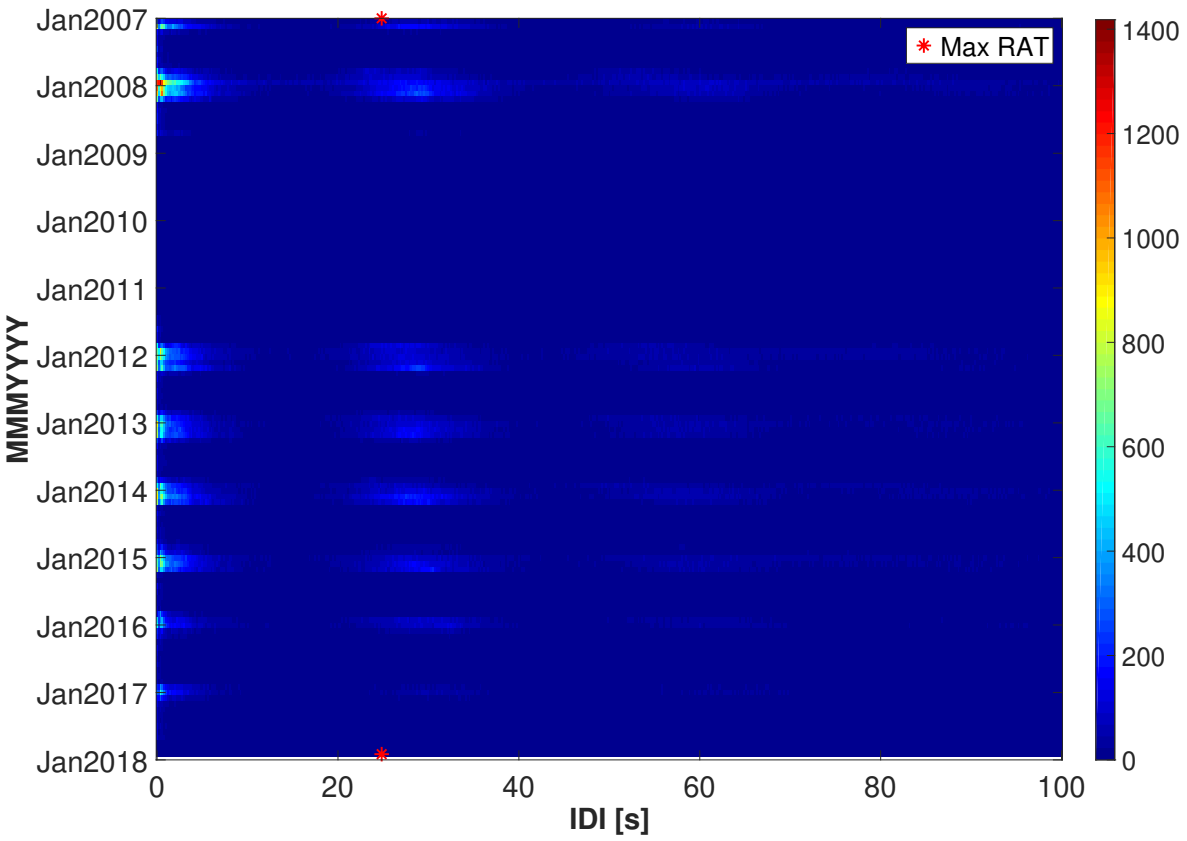

Figure 4.35: Calculated inter detection interval (IDI) values for each month of data between January 2007 and January 2018 and an ambiguity threshold of 0.99. One histogram of IDI values is calculated for each month. The more red a given pixel is, the greater the number of IDI values are found in that month and at that IDI value. Frequency pruning was applied to these data and no normalization is carried out. The maximum model-predicted RAT value (24.83 s) is indicated by the pair of red * symbols.

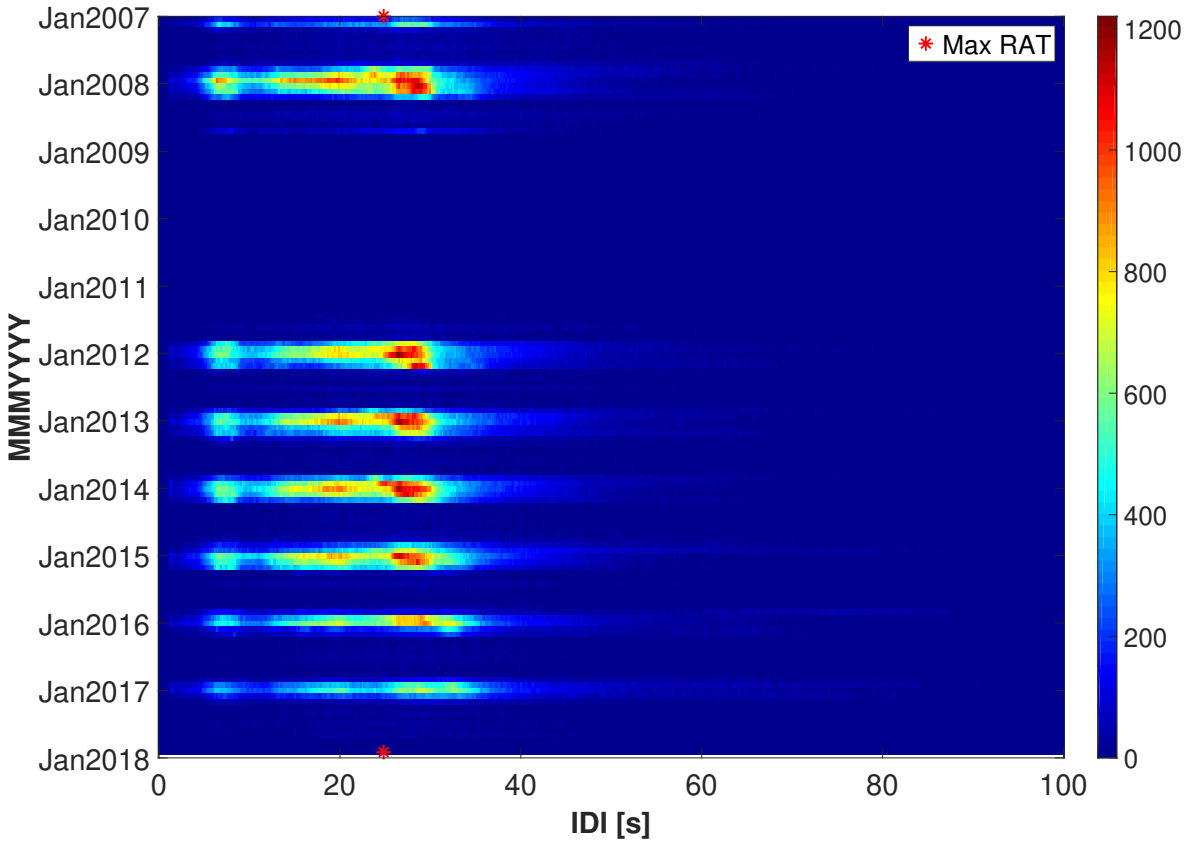

Figure 4.36: Calculated inter detection interval (IDI) values for each month of data between January 2007 and January 2018. One histogram of IDI values is calculated for each month. The more red a given pixel is, the greater the number of IDI values are found in that month and at that IDI value. Both frequency and phantom pruning are applied to these data and no normalization is carried out. The maximum model-predicted RAT value (24.83 s) is indicated by the pair of red * symbols.

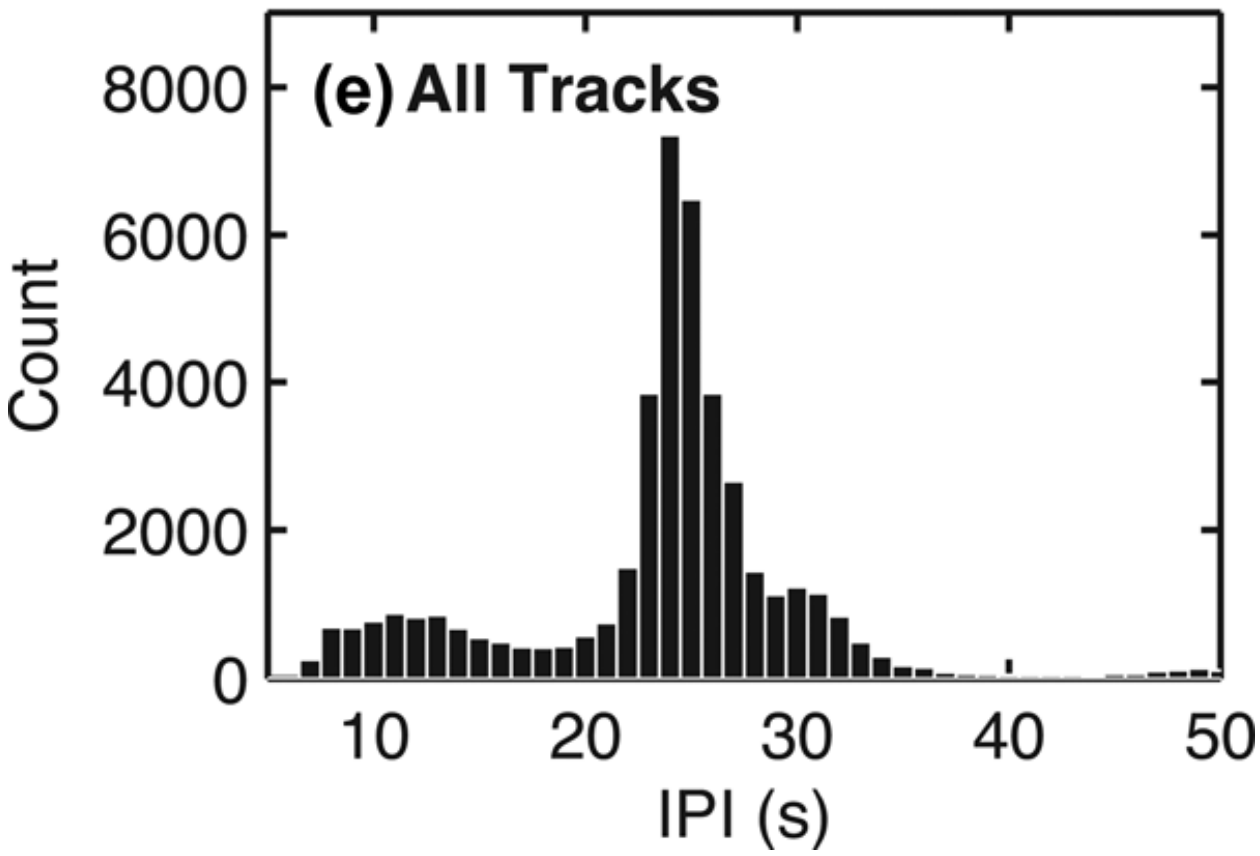

Figure 4.37: Observed inter-call interval (termed IPI in this figure) for 154 fin whale tracks recorded near the Juan de Fuca Ridge in the Northeast Pacific Ocean (Soule and Wilcock [2013]).

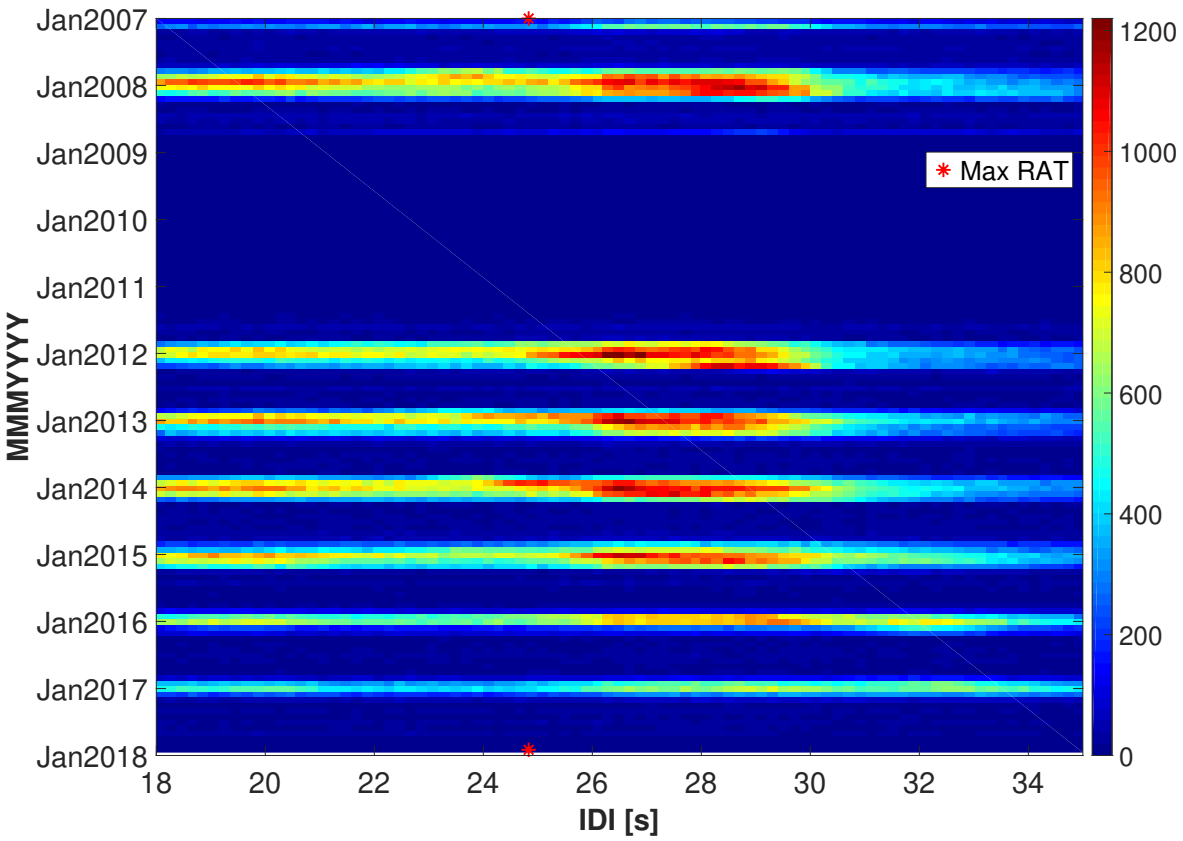

Figure 4.38: Calculated inter detection interval (IDI) values for each month of data between January 2007 and January 2018, zoomed in on the IDI$= 18 - 35$ s interval. One histogram of IDI values is calculated for each month. The more red a given pixel is, the greater the number of IDI values are found in that month and at that IDI value. Both frequency and phantom pruning are applied to these data and no normalization is carried out. The maximum model-predicted RAT value (24.83 s) is indicated by the pair of red * symbols.

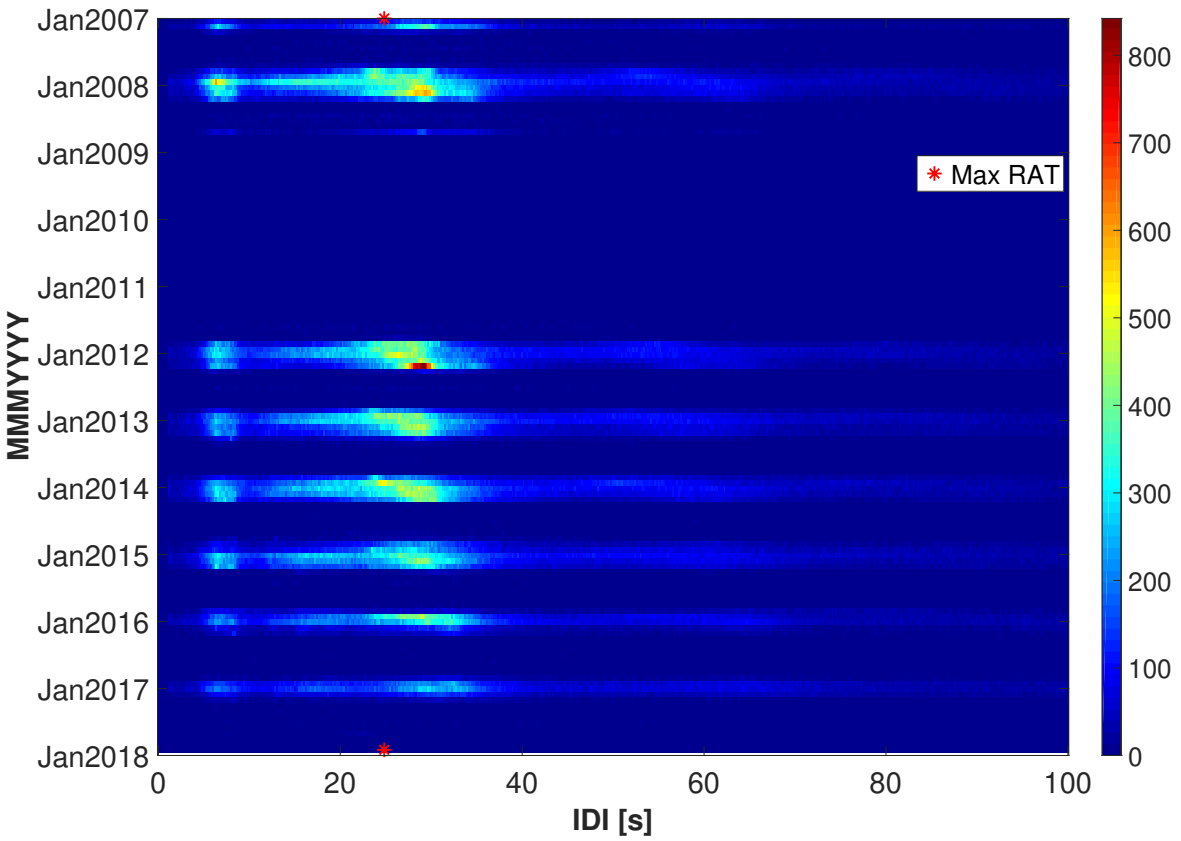

Figure 4.39: Calculated inter detection interval (IDI) values for each month of data between January 2007 and January 2018 and an ambiguity threshold of 0.3. One histogram of IDI values is calculated for each month. The more red a given pixel is, the greater the number of IDI values are found in that month and at that IDI value. Frequency and phantom pruning are applied to these data and no normalization is carried out. The maximum model-predicted RAT value (24.83 s) is indicated by the pair of red * symbols.

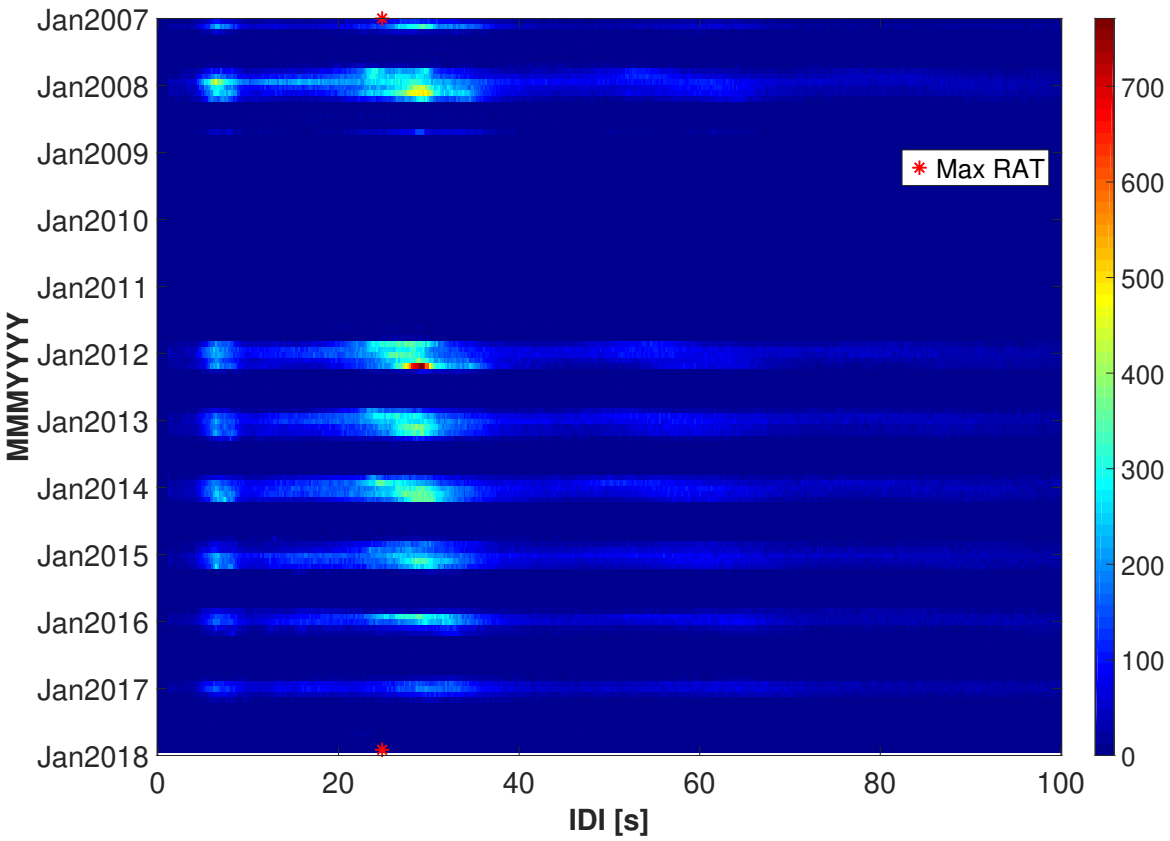

Figure 4.40: Calculated inter detection interval (IDI) values for each month of data between January 2007 and January 2018 and an ambiguity threshold of 0.6. One histogram of IDI values is calculated for each month. The more red a given pixel is, the greater the number of IDI values are found in that month and at that IDI value. Frequency and phantom pruning are applied to these data and no normalization is carried out. The maximum model-predicted RAT value (24.83 s) is indicated by the pair of red * symbols.

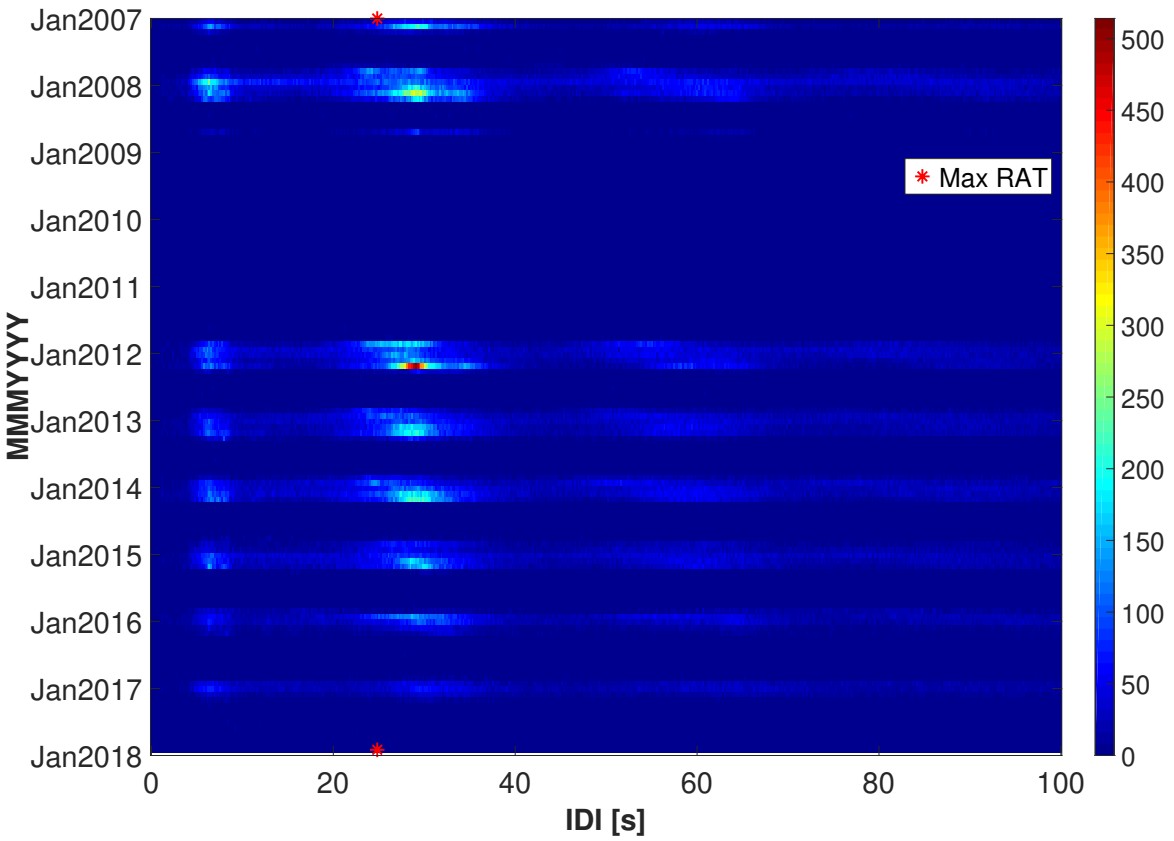

Figure 4.41: Calculated inter detection interval (IDI) values for each month of data between January 2007 and January 2018 and an ambiguity threshold of 0.9. One histogram of IDI values is calculated for each month. The more red a given pixel is, the greater the number of IDI values are found in that month and at that IDI value. Frequency and phantom pruning are applied to these data and no normalization is carried out. The maximum model-predicted RAT value (24.83 s) is indicated by the pair of red * symbols.

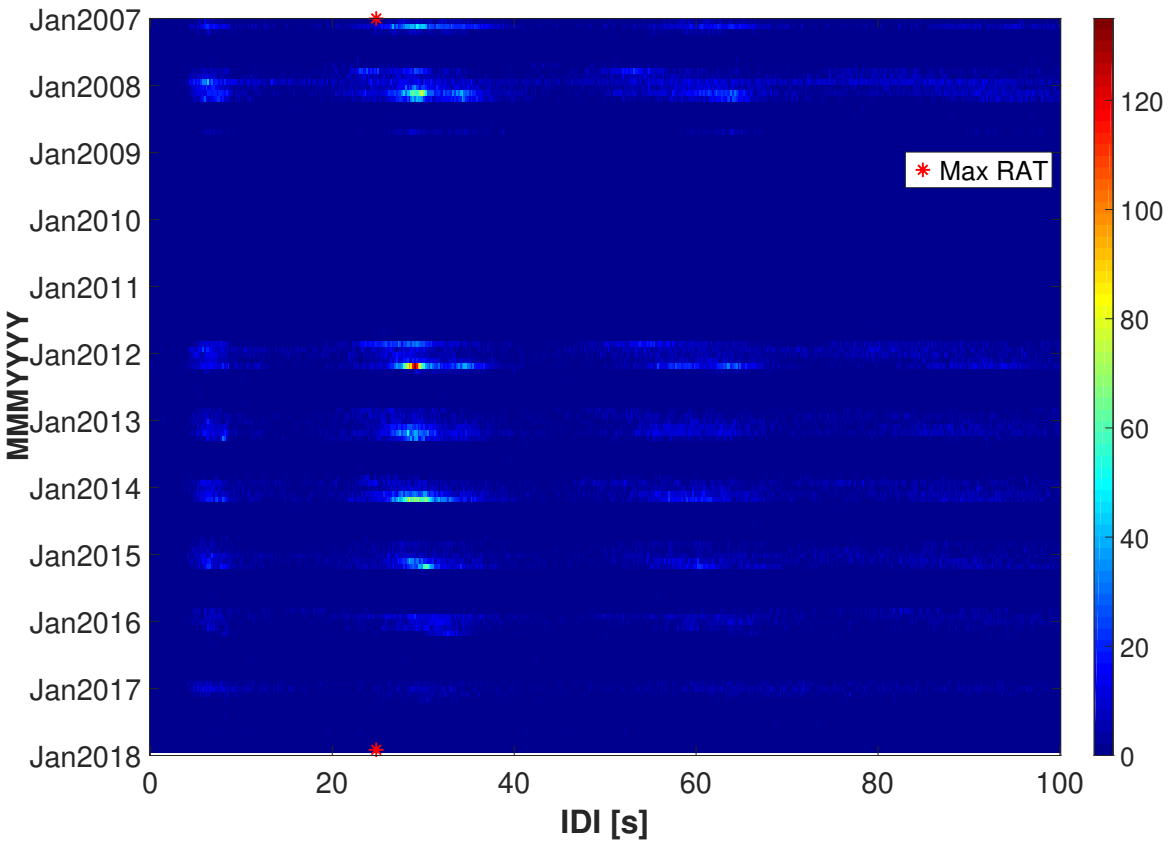

Figure 4.42: Calculated inter detection interval (IDI) values for each month of data between January 2007 and January 2018 and an ambiguity threshold of 0.99. One histogram of IDI values is calculated for each month. The more red a given pixel is, the greater the number of IDI values are found in that month and at that IDI value. Frequency and phantom pruning were applied to these data and no normalization is carried out. The maximum model-predicted RAT value (24.83 s) is indicated by the pair of red * symbols.

4.5 Summary

This chapter has presented measured data results using underwater acoustic data recorded at the ALOHA Cabled Observatory north of O'ahu. The objectives of this chapter are to demonstrate the performance of our localization approach on measured fin whale 20 hz calls and to show how estimation of fin whale call characteristics can be improved by carrying out frequency and phantom pruning.

The first objective is demonstrated using recordings of fin whales made between February 19 2007 - September 30 2017. Using a matched filter, 41,481,171 sounds resembling fin whale 20 Hz calls are detected. RATs are autonomously identified for each of these detections, and in turn localized. Following localization, pruning is done in two steps. First, frequency pruning is applied to distinguish fin whale 20 hz calls from sei whale calls and some noise, yielding a reduced set of 13,069,837 calls. The second pruning step, in which phantom tracks are eliminated from the frequency-pruned set, yield a final set of 3,445,568 detections which satisfy both the frequency and phantom pruning criteria. Detected calls from March 3 2014 are shown, including an intermittent medium-range track between 40-45 km from ACO that starts around 02h00 and finishes near 16h00. A strong seasonality to the fin whale call detection rates is shown in a set of multiyear plots; fin whale 20 Hz calls are detected in greater numbers during the winter months in O'ahu than in other times of the year.

To meet the second objective, call parameters are evaluated following the localization step to see if the addition of localization information improves the clarity of the estimated parameters. Peak frequency, start frequency, stop frequency, and inter-detection interval (IDI) are plotted for the unpruned, frequency pruned, and frequency & phantom pruned cases. The specific implementation of the peak frequency estimator introduced banding into the peak frequency plots that is also visible in the stop frequency plots. Frequency pruning made start and stop frequency easier to assess than in the unpruned data. Finally, calculated IDI better matched with the literature for North Pacific fin whale ICI when both frequency and phantom pruning are applied compared to frequency pruning alone. The improvement in ICI estimation when phantom pruning is applied demonstrates that localization can be a useful tool in evaluating marine mammal call parameters even when the localization information itself is not of immediate use.

CHAPTER 5
SUMMARY, CONCLUSIONS, AND FURTHER RESEARCH DIRECTIONS

5.1 Summary

This dissertation presented two different underwater acoustical signal processing approaches aimed at accessing and/or utilizing information contained in the recorded acoustic data. Particular emphasis was placed on the objective of passive acoustic localization of marine mammals, although the potential applications of these techniques are not limited to biological sound sources.

In Chapter 2, blind channel estimation (BCE) is presented as an approach to identify ocean waveguide impulse responses (IRs) using simultaneous recordings of the same unknown source by different geographically-spaced hydrophones. Given the greater clarity that arrival times may possess in a waveguide IR, compared to the recorded acoustic data, BCE is explored as a potential data processing approach to emphasize these key features in the data before subsequent analyses take place. A single impulse response is estimated between the source and each receiver. Differences between the recordings at different hydrophones are taken to reflect differences in the underlying channel IRs. Initial simulations using an underwater acoustic propagation model indicate that BCE may yield reasonable IR estimates under somewhat idealized conditions. However when a time-varying property to the underlying IRs is introduced, IR estimation deteriorates. This time-varying property resembles water depth variation due to progressive ocean surface waves on the same time scale as the duration of the source waveform. One possible explanation for this deterioration is that a time-varying characteristic to the IRs introduces differences in the received waveforms that are inconsistent with a static source and a set of static IRs. Thus, BCE seeks a more complicated set of estimated IRs than for the static IR case, resulting in estimated IRs that are a poor match to the original.

In Chapter 3, a novel algorithm that uses differences in eigenray arrival times to estimate the horizontal range between a single hydrophone and a submerged sound source is presented. Following bandpass filtering and downsampling, a matched filter-based automated detector is used to identify sounds of interest in the acoustic data. Next, the differences between the earliest eigenray arrival for each detection, and the next 2-4 arrivals, are measured. These so-called relative arrival times (RATs) identified from each detected sound are then compared to model-predicted RATs calculated using an underwater acoustic propagation model at a set of candidate source ranges. The closest fit between measured and model-predicted RATs indicates the estimated range for the source in question. Fit is assessed using a novel cost function that is robust to some unavoidable sources of false positives in the measured RAT sets and different numbers of model-predicted RATs at different ranges.

Chapter 3 then presents a series of simulations which demonstrate the performance of this localization approach during typical operation, as well as when the source range or depth go outside the established bounds. These simulations also explore how overlapping calls may lead to reduced performance. Finally, this chapter uses ground-truthed air gun data collected at the ALOHA cabled observatory (ACO) to show that estimated ranges to the 296 individual air gun shots match well with their true ranges.

Finally, Chapter 4 presents localization results using fin whale 20 Hz calls recorded at ACO between

2007-2017. The localization process presented in Chapter 3 is applied to this multi-year data set, yielding 41,481,171 localized detections. Frequency pruning is then applied, reducing the dataset to 13,069,837 calls with a lower false positive rate from noise and sei whale calls. Following phantom pruning, 3,445,568 detections remain with fewer duplicate detections per call due to multipath arrivals. Following the pruning steps, the impact of pruning on the estimation of fin whale 20 Hz call features is investigated. Peak frequency, stop frequency, start frequency, and inter-call interval are measured for the no pruning, frequency pruning, and frequency + phantom pruning cases. Frequency pruning was particularly helpful in identifying start, stop, and peak frequency values. Phantom pruning (in particular the phantom-track pruning, rather than an ambiguity threshold) yielded estimates of inter-call interval which are consistent with published values for North Pacific fin whales.

5.2 Conclusions

While the focus of this dissertation is to develop and demonstrate a set of data analysis approaches within the context of underwater acoustic signal processing, rather than carry out a study with a particular research question in mind, there are a number of observations that may be drawn from this work.

Time-varying underwater acoustic impulse responses may make blind channel estimation problematic. Indeed, given that the very existence of an impulse response requires the underlying system be both linear and time-invariant, any system (whether or not it be an underwater acoustic waveguide) whose impulse response varies on the same time scale as the duration of the signal being used to excite said system may not be well-suited to investigation using blind channel estimation.

The localization cost function presented in this dissertation is robust to the inevitable presence of noise in the measured acoustic data. One form this noise takes is in the measured RATs for each detected source. The cost function is specifically designed to expect that only a subset of the measured RATs will be physically meaningful for the source in question. The origins for the remaining RATs include spurious peaks in the cross-correlation envelope function due to side lobes from true arrivals and background noise. Simulations and validated measured data results show that the maximum ambiguity calculated using this cost function can accurately indicate source range in the presence of realistic ocean conditions.

The validated measured data also reveals how localization errors can be introduced by the measured RAT estimation process. The improved range estimation accuracy when RATs were manually picked, versus when detections were manually picked and RATs were autonomously picked, suggests there is room for improvement in how RATs are currently estimated in this approach. One such improvement would be improving the capacity of the underwater acoustic propagation model to predict eigenray detectability with the aim of reducing short range localization bias.

This work demonstrates that localization can be a useful signal processing step even if the estimated locations themselves are not of immediate use. The utility of a localization analysis is particularly apparent in the estimated inter-detection interval results when pruning is applied which relies on localization having taken place first. The good agreement between estimated fin whale inter-call interval in this work and published values from the literature, once phantom pruning is applied, demonstrates this indirect utility of localization. The elimination of higher order eigenray paths from a set of marine mammal detections through phantom pruning, for example, reduces the likelihood of counting the same call more than once.

This simplifies the identification of inter-call interval from an inter-detection interval analysis.

Phantom pruning gives an additional tool to discriminate spurious detections from physically meaningful ones. In some cases, these spurious detections may yield phantom tracks which match well to the measured data but at an incorrect range. Thus, ambiguity score alone is insufficient to identify these false tracks. Incorporating a pruning criterion whereby calls are eliminated whose measured arrival times match those of previous detected calls within a defined time interval can be more effective at eliminating these phantom tracks than ambiguity score. A downside of this phantom pruning is that multiple closely-spaced calls are less likely to each be localized correctly, since latter ones may be pruned away depending on how similar their arrival times are to the earlier arriving calls.

5.3 Further Research Directions

A useful next step to improve the localization approach introduced in this dissertation is the inclusion of an uncertainty analysis following the range estimation step. The area under the ambiguity vector calculated for each detection could be interpreted as a reflection of the range estimation uncertainty. However, the multiple significant local maxima typically found in the ambiguity vectors for this work will require careful thought in how best to represent this uncertainty in a reliable and insightful way.

Given that marine mammal density estimation (i.e., estimating the number of marine mammals per unit area) uses calls produced per unit time as one of its inputs, it may be possible to refine estimates of North Pacific fin whale density in the vicinity of ACO using local estimates of inter-call interval developed in this work.

Fin whale 20 Hz call detection rates are visibly lower in the January 2016 and 2017 calling seasons than in the four previous calling seasons. This phenomenon suggests a combination of two hypotheses:

- There are fewer 20 Hz fin whale calls in the 2016 and 2017 calling seasons; and
- The matched filter detector is be less effective in these later years than it was in earlier calling seasons.

Looking into which of these hypotheses are plausible would yield insight into whether fin whale calling behavior or presence north of Hawaii is changing. Alternatively, a sea surface temperature anomaly observed near Maui that influenced humpback whale chorusing levels (Kügler et al. [2020, 2021]) may also help explain the reduction in fin whale detection rates in these calling seasons.

When calculating model-predicted RATs, using a propagation model which better accounts for instances where reflection or refraction-based attenuation may make identification of certain arrival paths difficult in the measured data will improve localization performance (particularly at close range).

Currently, the localizer can handle having more measured RATs than expected. This was necessary to building the localizer in the first place, since the RAT estimator is never going pick out exclusively the physically meaningful RATs while perfectly disregarding noise. Building on this, the localizer could be made robust to predicting more measured arrivals than are actually present in the data. Effectively, the cost function could be modified to not only search for a subset of the set of measured RATs for each detection, but also search over a subset of the model-predicted RATs. This will come at the cost of significantly-higher computational cost, however.

The banding seen in call frequency characteristics histograms is likely a result of interpolation of the power spectral density within the matched filter detector. In other words, not all frequencies in the interpolated PSD are equally likely of containing a PSD local maximum due to the interpolation. Calculating frequency characteristics of each detection prior to downsampling will improve any bias in peak frequency values due to downsampling, at the cost of increased computational complexity of the detection phase of the overall localization approach. Looking into how much time is saved by processing downsampled, rather than the original, acoustic files may help in the decision on whether downsampling is justified in the first place.

Implementing a database to consolidate detection, measured RATs, and range estimates in a single spot (rather than multiple individual MATLAB directory structures as it is currently implemented) would simplify analysis of the localization information by making it easier to keep track of data as it moves from pre-pruned through the post-pruned data structures. Such a database would also reduce the storage requirements of this approach by reducing the needless duplication of detections and measured RAT sets between the pre and post-pruned cases.

The same data used in this dissertation may be of interest in other applications. The length of the acoustic data record at ACO, and the low-noise recording environment, make it a high quality data source. The localization work in this dissertation looked exclusively at data in the lowest 100 Hz, leaving most of the recording bandwidth unused. Additionally, in recent years, advances have been made in the application of machine learning approaches to underwater sound source localization (Bianco et al. [2019], Roch et al. [2021]). The validated air gun data used in this work to demonstrate our localization approach in Section 3.5.2 could potentially be used to tune a supervised machine learning-based localizer for the ocean environment near ACO.

APPENDIX A
GLOSSARY

This appendix lists and defines some commonly-used terms used in the body of this dissertation. In some cases (e.g., ACO), terms are used in more than one chapter. In these cases, terms are defined in the first chapter they appear in.

Chapter 2 - Blind channel estimation

ACO

The ALOHA Cabled Observatory - an underwater oceanographic data collection site and initiative located north of O'ahu

BCE

Blind Channel Estimation - the process of estimating a set of impulse responses using recorded data of the same source signal without knowledge of the characteristics of this source

BP

Basis Pursuit - The process of decomposing a signal into the superposition of an optimal set of simpler signals whereby the ℓ_1 norm over the set of decomposition coefficients is minimized, as presented in Chen et al. [2001]

Convex Optimization

An optimization approach whereby:

$$\min f_0(x) \tag{A.1}$$

$$\text{subject to } f_i(x) \leq b_i,\ i = 1, \ldots, m \tag{A.2}$$

$$\text{where } f_0, \ldots, f_m \text{ are convex (i.e., satisfy } f_i(\alpha x + \beta b) \leq \alpha f_i(x) + \beta f_i(y)), \tag{A.3}$$

as described in Boyd and Vandenberghe [2004]

CoSaMP

Compressive Sampling Matching Pursuit - A blind channel estimation algorithm presented in Needell and Tropp [2009]

Eigenray

An acoustic ray path connecting a source and receiver pair (e.g., the ray that leaves the source, bounces once off the bottom, once off the surface, then reaching the receiver). Sometimes referred to as an arrival path.

IR

Impulse response

LS

Least Squares - A blind channel estimation algorithm presented in Xu et al. [1995]

MP

Matching Pursuit - A blind channel estimation algorithm presented in Mallat and Zhang [1993]

NESTA

Nesterov's Algorithm - A basis pursuit-based convex optimization algorithm as presented in Becker et al. [2011]

OMP

Orthogonal Matching Pursuit - A blind channel estimation algorithm presented in Tropp and Gilbert [2007]

Passive Underwater Acoustic Localization

The process of estimating components of an underwater subjects location (including range and/or depth) using the underwater sounds of the subject

Sparse Recovery Algorithm

A computational approach to estimating a set of impulse responses where the majority of IR values are assumed to be zero

Subspace

A blind channel estimation algorithm presented in Moulines et al. [1995]

Chapter 3 - Single hydrophone ranging in deep water

Ambiguity

A quantitative measure of the match between measured and model-predicted RATs. The higher the ambiguity value, the better the match.

BELLHOP

A ray-tracing underwater acoustic propagation model as presented in Porter and Liu [1994]

BS

The eigenray path that leaves the source, reflects off the bottom first, then the surface, before reaching the receiver

BS2

The eigenray path that leaves the source, reflects off the bottom first, then the surface, then the bottom, then the surface before reaching the hydrophone

BS3

The eigenray path that leaves the source, reflects off the bottom first, then the surface, then the bottom, then the surface, then the bottom, then the surface before reaching the hydrophone

BS4

The eigenray path that leaves the source, reflects off the bottom first, then the surface, then the bottom, then the surface, then the bottom, then the surface, then the bottom, then the surface before reaching the hydrophone

Candidate source range

A range from ACO at which model-predicted RATs are calculated and the potential presence of a source signal is evaluated at

D

The eigenray path that travels from source to receiver without reflecting off either the surface or bottom

DA

Data Analysis

DC

Data Collection

DCL

Detection, Classification, Localization - In the context of this work, the set of passive underwater acoustic algorithms implemented for assessing the presence of a potential signal of interest (Detection), assigning a species-level label to these detections (Classification), and estimating parameters of the sound source location (Localization)

Depth

The vertical distance between the ocean surface and a point of interest

Double Option Zone

A set of candidate source ranges about an eigenray cutoff range which has two distinct set of RATs at each candidate source range: one where the eigenray in question is still detectable, and one in which it is not detectable

Frequency Pruning

The process of comparing measured peak and stop frequency values against thresholds intended to discriminate between fin and sei whale calls

Higher-order Arrival Path

An eigenray which arrives comparatively later at the hydrophone

Matched Filter

A signal processing technique, commonly used in detection and classification applications, whereby the cross-correlation between a signal of interest and a recorded waveform is used to assess the presence or absence of the signal of interest.

Measured Relative Arrival Times

The set of relative arrival times estimated from the recorded acoustic data

MPRATs

Model-Predicted Relative Arrival Times - The set of relative arrival times at each candidate source range which are calculated using an acoustic propagation model and compared against the measured relative arrival times

OBS

Ocean Bottom Seismometer

Peak Frequency

f_{peak}, the frequency of highest PSD within the detected call

Phantom Pruning

The process of comparing measured RATs for the current call against those of previous calls whose peak and stop frequencies satisfied the frequency pruning criteria. Intended to reduce the false positive rate by reducing the likelihood of multiple reflected arrivals of a single call being localized.

Phantom Track

A spurious track, paralleling the true track, arising from higher order arrivals of a set of calls being misinterpreted as lower order arrivals

Pruning

The process of carrying out frequency and phantom discrimination on the features of detected and localized calls with the aim of reducing the false positive rate

PSD

Power Spectral Density

Range

The horizontal distance between two points

RAT

Relative Arrival Time - The difference in arrival time at a hydrophone between the earliest arrival of an underwater sound and a higher-order arrival path. The direct-path arrival may, or may not, be the earliest arrival depending on local oceanographic conditions.

RAYFAST

A ray-based underwater acoustic propagation model specifically designed to calculate eigenray arrival times, as presented in Dosso and Ebbeson [2006]

Reference Waveform

In a matched filtering context, the signal whose presence is being assessed in the data

SNR

Signal to Noise Ratio

SSP

Sound Speed Profile

Stop Frequency

f_{stop}, the lowest frequency greater than f_{peak} whose PSD is 6dB lower than the peak PSD value

TDOA

Time Difference Of Arrival - The difference in arrival time of the same sound at multiple hydrophones. Not to be confused with Relative Arrival Times.

Chapter 4 - Automated fin whale localization

ICI

Inter-Call Interval, the time spacing between consecutive calls in a calling sequence

IDI

Inter-Detection Interval, the time spacing between consecutive call detections.

Start frequency

f_{start}, the highest frequency lower than f_{peak} whose PSD is 6dB lower than the peak PSD value

Track

A set of individual consecutive, or nearly consecutive, calls that could have come from the same source

APPENDIX B
SIMULATION STUDIES

In this appendix, simulation results are presented to demonstrate the performance of the localization approach in conditions emulating those of the ACO site. To better match with the results that will be presented in Chapter 4, the source waveforms for these simulations are a set of three measured fin whale 20 Hz call recorded by the ACO hydrophone.

These simulations investigate four different questions related to the application of this sound source range estimation approach to the ACO data set:

- **Simulation 1**: Can a source that smoothly travels through the close, medium, and far ranges near ACO be accurately tracked?
- **Simulation 2**: If source depth deviates from the expected value of 50 m, does the range estimate become less accurate?
- **Simulation 3**: If source range exceeds the maximum range included in the candidate source range set, what effect is seen in estimated range?
- **Simulation 4**: If multiple simultaneous tracks are present in the recorded acoustic data (with one of three different time offsets between the two tracks) what effect is seen in the estimated ranges?

To help answer these 4 questions, a series of 82 simulated source locations are defined. Using a water depth of 4667 m, a hydrophone 2 m off the bottom, the measured SSP used to generate the MPRATs, and the RAYFAST propagation model, model-predicted travel times for the D, BS, BS2, BS3, and BS4 arrival paths for each of the 82 sources are calculated. These travel times, in addition to the fin whale source waveform and call source times, are used to simulate the received waveforms for each of the 82 simulated source locations. In other words RAYFAST is used to calculate a simple impulse response for each of the 82 source locations based on the defined source, receiver, and environmental conditions. These impulse responses are then convolved with the same source waveform to create simulated received waveforms for each of the 82 simulated sources. These 82 simulated waveforms are then concatenated to create a single simulated recorded waveform. White Gaussian noise is added to the simulated waveforms to allow the automated detector to potentially find false positives in the background noise. The resulting concatenated simulated waveform is shown in Figure B.1. This process is repeated for two different recorded fin whale calls (also from the ACO recorded data set) to give some insight into how different source waveforms at the same physical location, and with the same background noise, may be localized differently. Once this concatenated waveform is calculated, the data processing sequence depicted in Figure 3.5 is applied to the data.

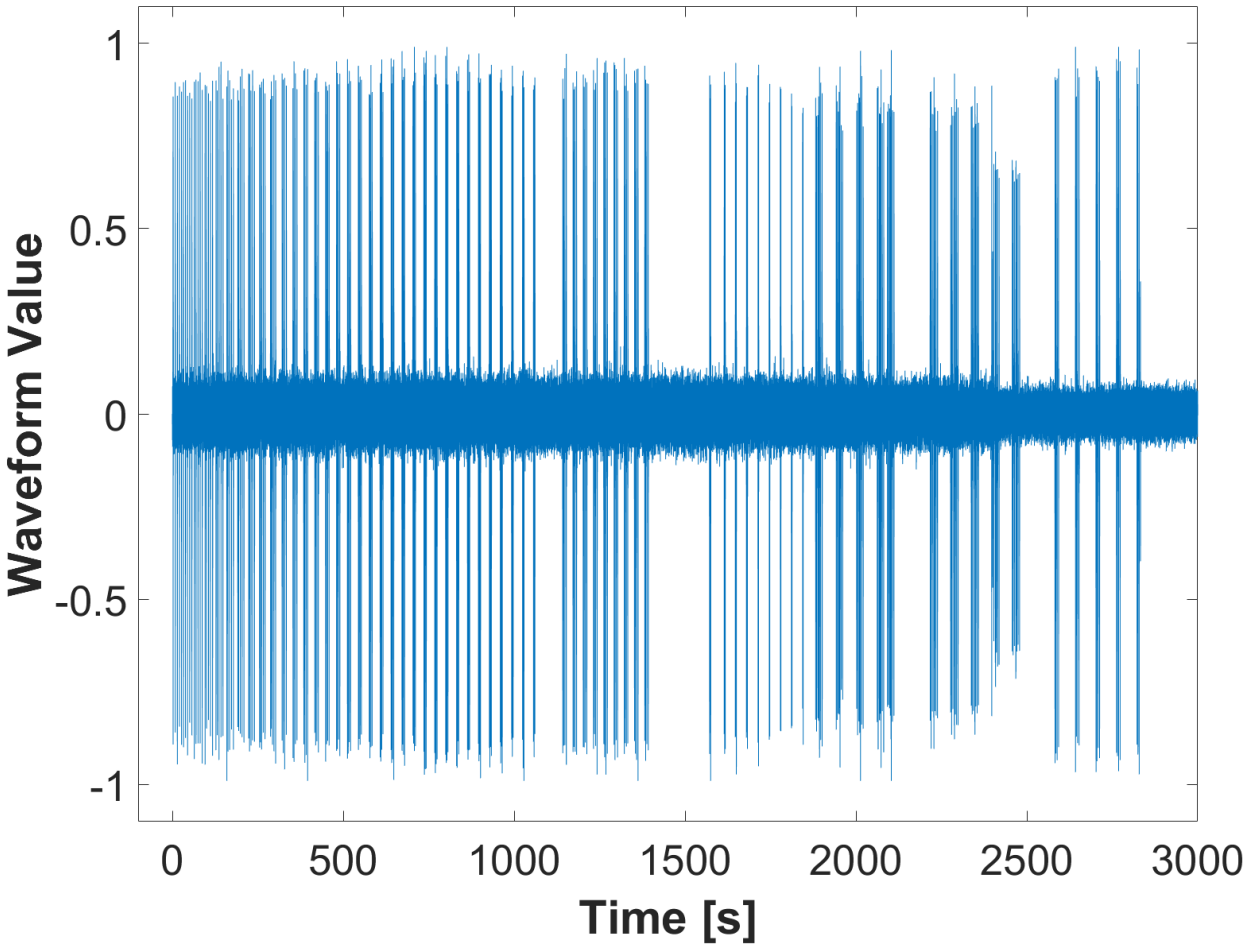

Figure B.1: Simulated waveform for 82 source locations in an environment emulating that of ACO.

B.1 Simulation 1: Source moving through close, medium, and far range

Simulation 1 uses 34 source ranges. The closest source is 500 m from the hydrophone, the next closest is 3 km away, and subsequent sources step progressively by 3 km in range to a maximum range of 99 km for the 34^{th} source. Each source is at a depth of 50 m, with 30 seconds between sources so that the received waveform for one call does not overlap with another.

Figure B.2 shows the ambiguity vector for the Simulation 1 source at 39km. The lower and upper limits for both double option zones are indicated by vertical dashed lines. Note that each candidate source range value in a double option zone has two different ambiguity values; one for each of the two different sets of model-predicted RATs at each of those double option zone ranges. The true source range (39 km) resides at medium range, indicating that there are exactly 3 RATs that the localization function ought to search for from this shot: BS2-BS, BS3-BS, and BS4-BS. The jaggedness of the cost function in these double option zones is due to model-predicted RAT sets consisting of 3 RATs yielding higher ambiguity values than either the 4 RAT (in the short:medium double option zone) and the 2 RAT (in the medium:long double option zones) model-predicted sets.

Figure B.3 shows the same ambiguity vector plotted in Figure B.2, only with the two double option zones consolidated. In other words, at each candidate source range in each of the two double option zones, only the peak ambiguity value is plotted in Figure B.3. These consolidated ambiguity values represent the overall ambiguity at each range, and is the value used when ascertaining the global maximum ambiguity value and corresponding source range estimate.

The global maximum ambiguity value in Figure B.3 is the closest candidate source to the true source range in this case. However, there are 6 other local maxima in the consolidated ambiguity vector. Each of these reflects the best-fit match between the same set of data and an incorrect set of model-predicted RATs. Figure B.4 illustrates the fit between the simulated data for a source at 39 km and the model-predicted RATs at the closest candidate source range. In particular, subplot C shows that the cross-correlation yields significant side lobes in addition to the largest peak for each arrival path. These secondary ambiguity peaks in the ambiguity vector in Figure B.3 represent cases where the model-predicted data matches to a subset of the true measured RATs or a match to some combination of noise and true peaks. In practice, these secondary ambiguity vector peaks are lower than the expected peak and typically do not interfere with accurate localization.

Figure B.5 shows the set of true and estimated ranges for Simulation 1 and Source 1. Simulation 1 results for the remaining two sources are shown in Figure B.6 and Figure B.7.

Across all three source waveforms, there is good match between the true and estimated range for the 34 source waveforms. A tendency toward a greater difference between true and estimated range as range increases is due to decreasing range resolution with increasing range.

The three different source waveforms gave different false positive rates in Figures B.5 through B.7. Figures B.8, B.9, and B.10 show the resulting set of estimated vs. true ranges after frequency and phantom pruning is applied. The pruning process successfully eliminates almost all of the false positives, at the expense of also pruning away several true positives.

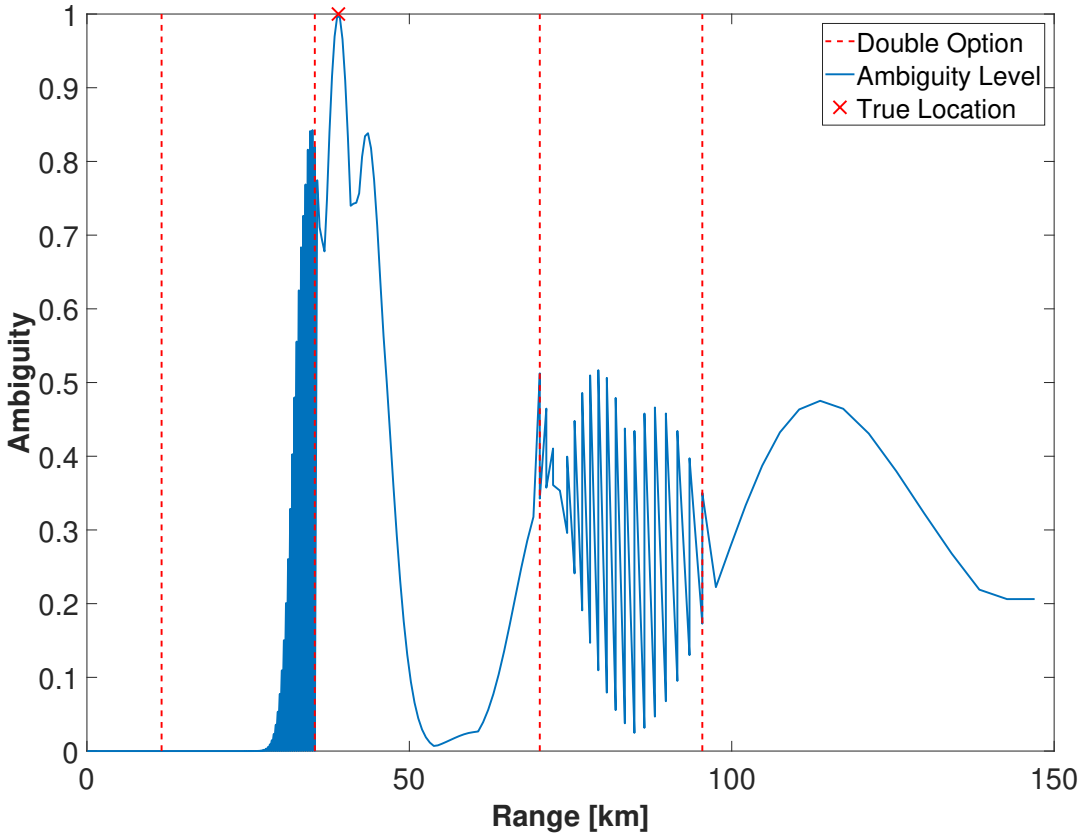

Figure B.2: Ambiguity vector (blue solid line) for the Simulation 1 source at 39 km. Double option zone boundaries are marked by vertical red dashed lines. The true source range is indicated by the red X symbol.

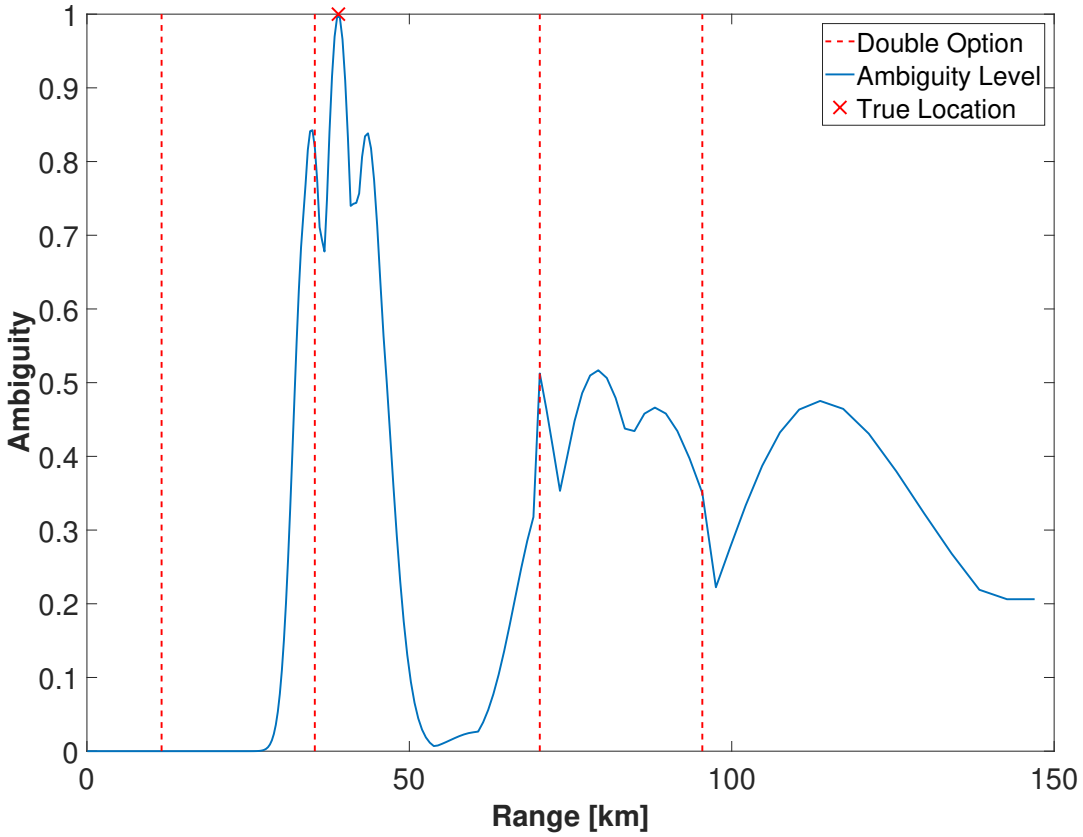

Figure B.3: Ambiguity vector (blue solid line) for the Simulation 1 source at 39 km. Double option zones have been consolidated such that the maximum ambiguity value at each range is plotted. Double option zone boundaries are marked by vertical red dashed lines. The true source range is indicated by the red X symbol.

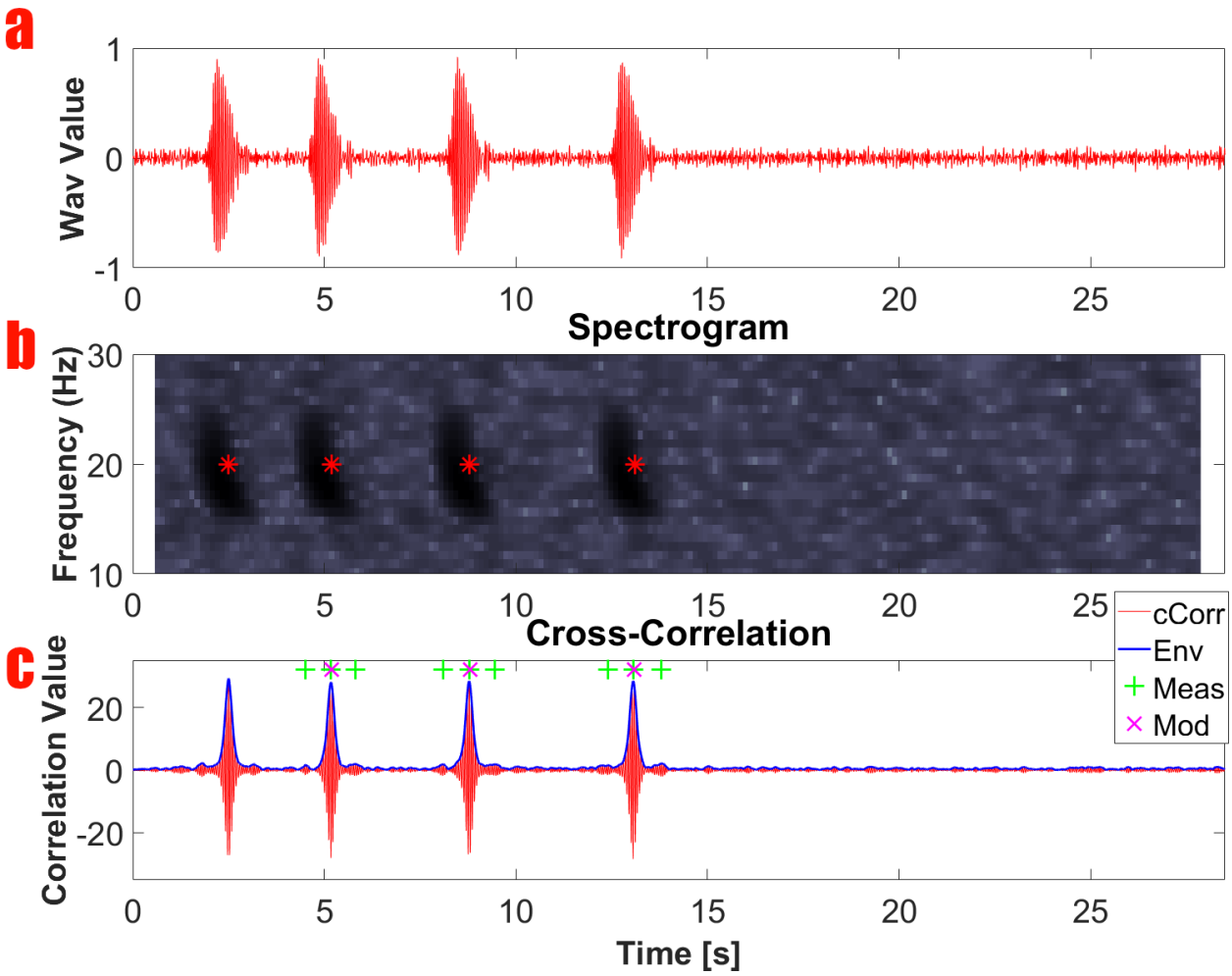

Figure B.4: Time series depicting the relative arrival time estimation process for the Simulation 1 source 39 km from the hydrophone. **a** Recorded time series waveform. **b** Spectrogram in the 10-30 Hz range, with red $*$ symbols highlighting automated detector output after pruning was done. **c** Cross-correlation (red trend), and cross-correlation envelope (blue trend) between the earliest arrival waveform and the entire time series shown in the top subplot. The width of the cross-correlation peaks is approximately 0.1 s. Green + symbols indicate the 9 highest cross-correlation envelope peaks other than those for the arrival waveform at time $= 2.5s$. Purple X symbols show the optimum fit between a model-predicted RAT set.

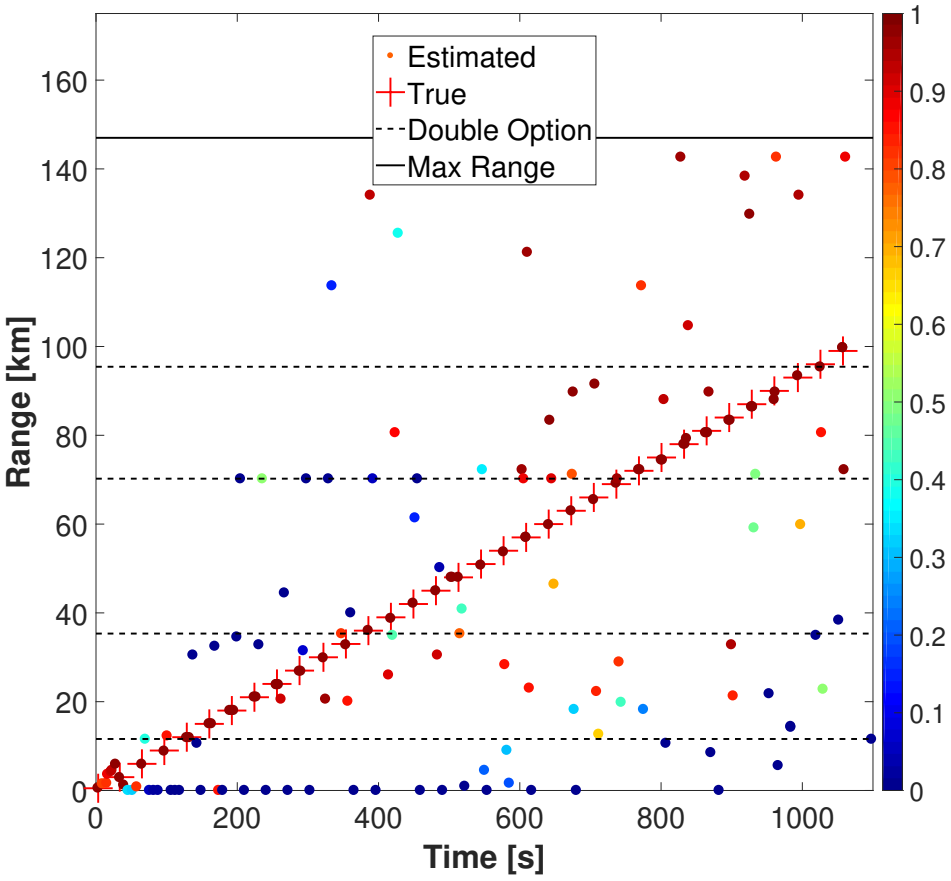

Figure B.5: True (red + symbols) vs. estimated (filled circles) source range for Simulation 1 and Source 1. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

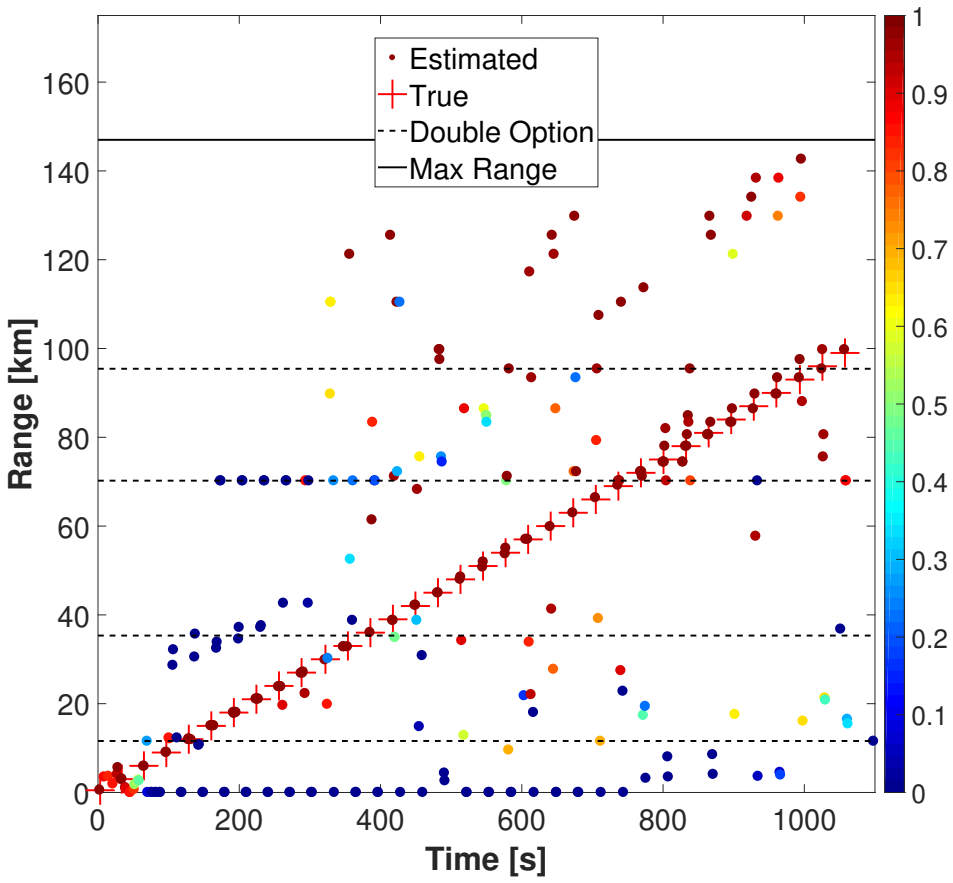

Figure B.6: True (red + symbols) vs. estimated (filled circles) source range for Simulation 1 and Source 2. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

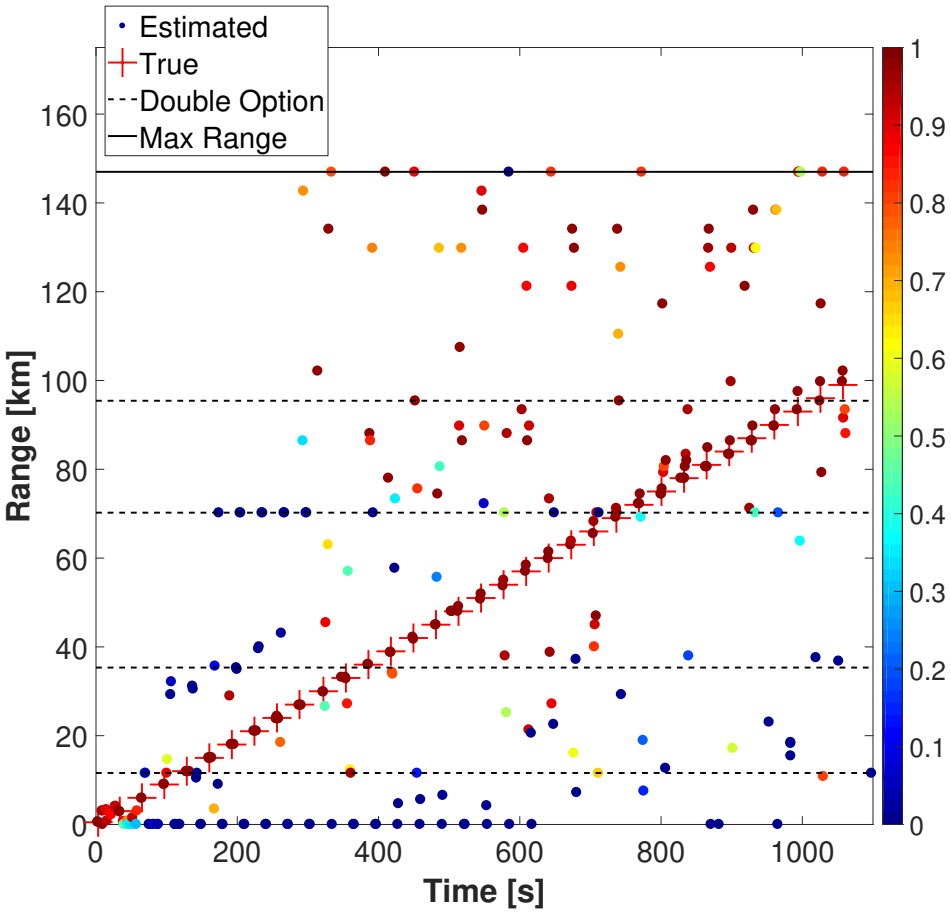

Figure B.7: True (red + symbols) vs. estimated (filled circles) source range for Simulation 1 and Source 3. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

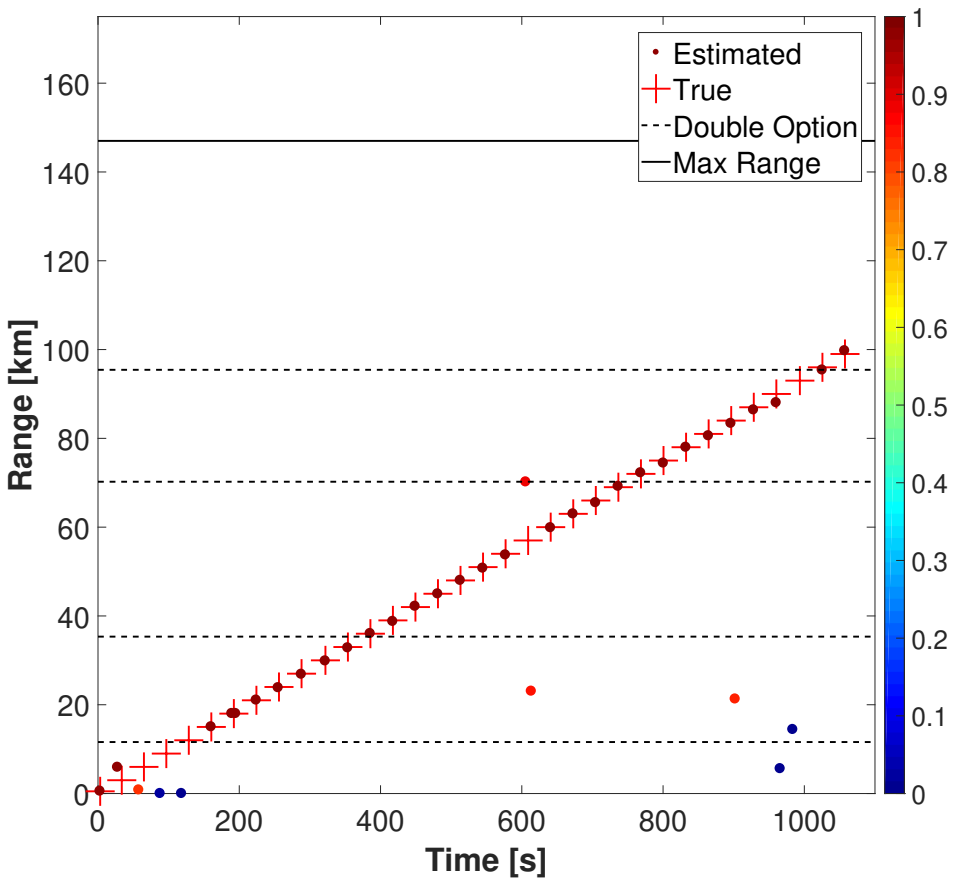

Figure B.8: True (red + symbols) vs. estimated (filled circles) source range for Simulation 1 and Source 1 after frequency and phantom pruning are applied. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

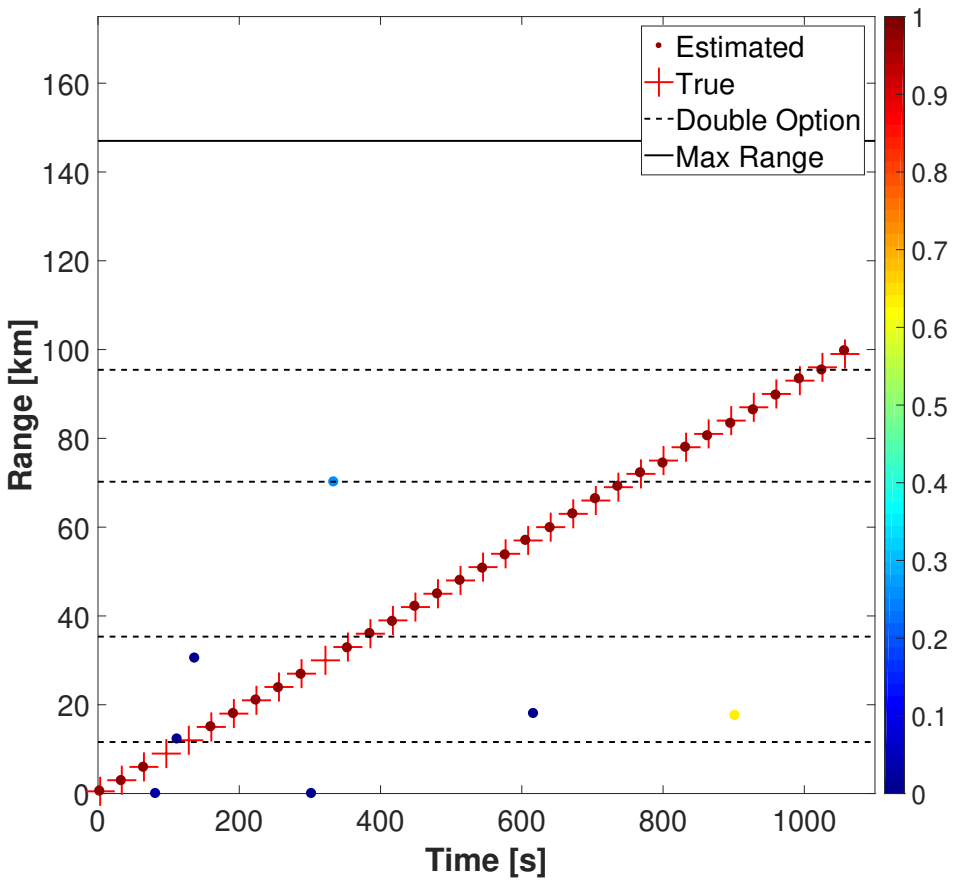

Figure B.9: True (red + symbols) vs. estimated (filled circles) source range for Simulation 1 and Source 2 after frequency and phantom pruning are applied. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

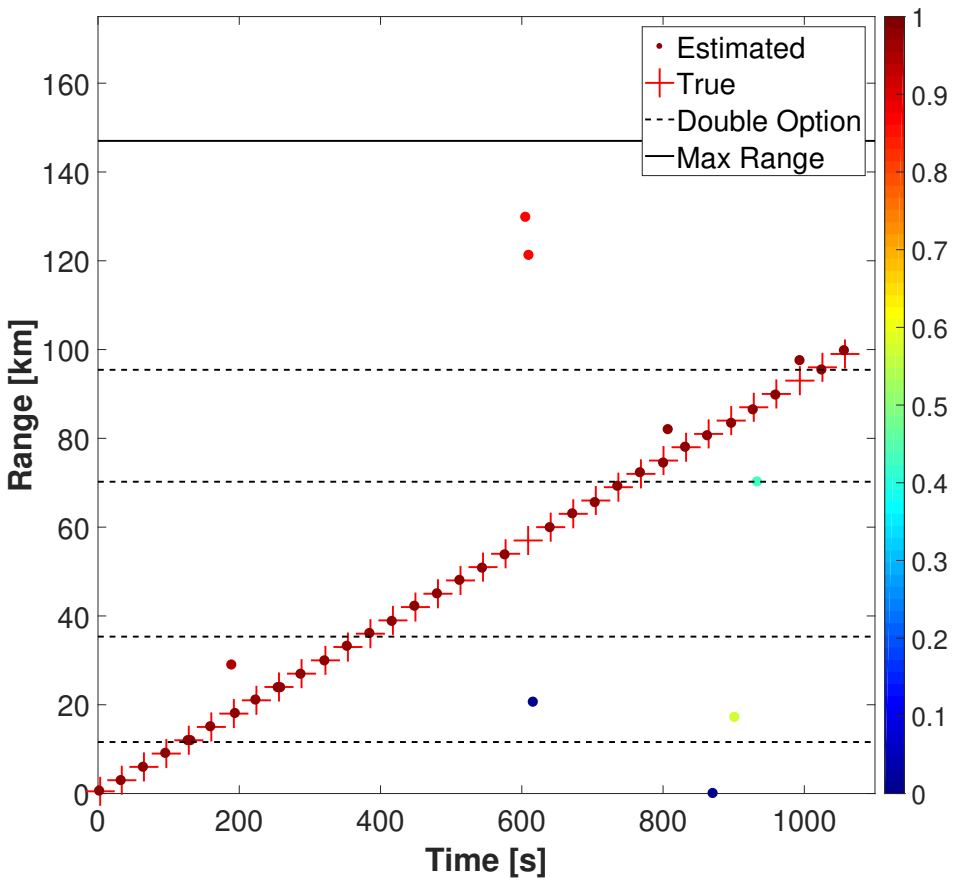

Figure B.10: True (red + symbols) vs. estimated (filled circles) source range for Simulation 1 and Source 3 after frequency and phantom pruning are applied. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

B.2 Simulation 2: Source depth deviates from 50 m

Simulation 2 explores the implications of incorrectly assumed source depth on horizontal range estimates. Nine different source depths are simulated, all at a range of 40 km. The set of simulated source depths are: (10, 30, 50, 70, 90, 200, 300, 400, 500) m, and the nine source transmissions are spaced 30 seconds apart. The results from the three sources didn't greatly differ, so only those from Source 1 are shown here in Figure B.11.

When source depth is within 40 m of the assumed depth (i.e., for the [10, 30, 50, 70, 90] m simulated source depth cases), the estimated range is the closest candidate source range to the true range. Thus, the range estimate is correct when the simulated and assumed source depth are both equal to 50 m (third source from the left in Figure B.11). For the next three deepest simulated source depths [200, 300, 400] m, the estimated range is within 500 m of the true range. Finally, for the deepest simulated source depth [500 m], the estimated range is 900 m greater than the true range. For the greatest difference between true (500 m) and assumed (50 m) source depth (the last source on the right), the estimated range is within 900 m of the true range.

Figure B.12 shows the results of Simulation 2 for Source 1 after pruning has been applied to the results. As with the Simulation 1 results, almost all of the false positives were removed, as was the estimated range for the simulated call whose true depth is 500 m. This case illustrates how a noisy detection can interfere with a true detection despite the use of both frequency and phantom pruning. Here, the noisy detection with an estimated range of 138 km matches the frequency pruning criteria and is close enough in time to the ninth simulated source that phantom pruning removes the true detection in favor of the earlier, noisy one.

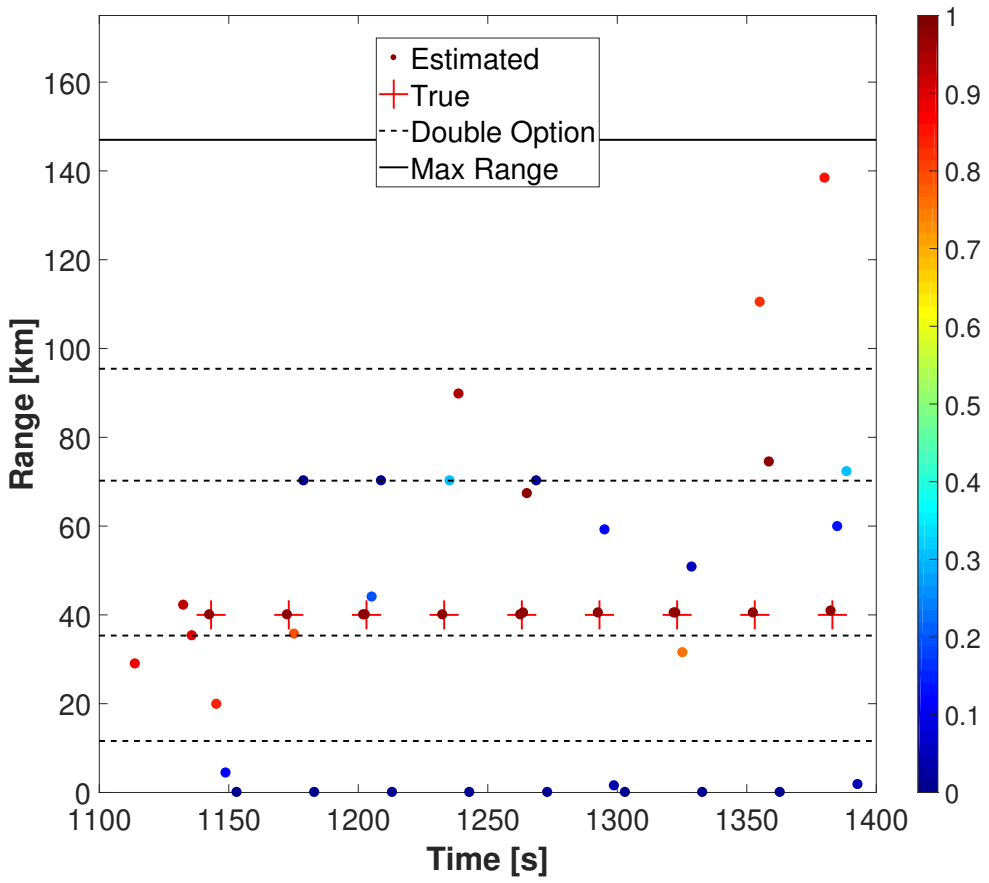

Figure B.11: True (red + symbols) vs. estimated (filled circles) source range for Simulation 2 and Source 1. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

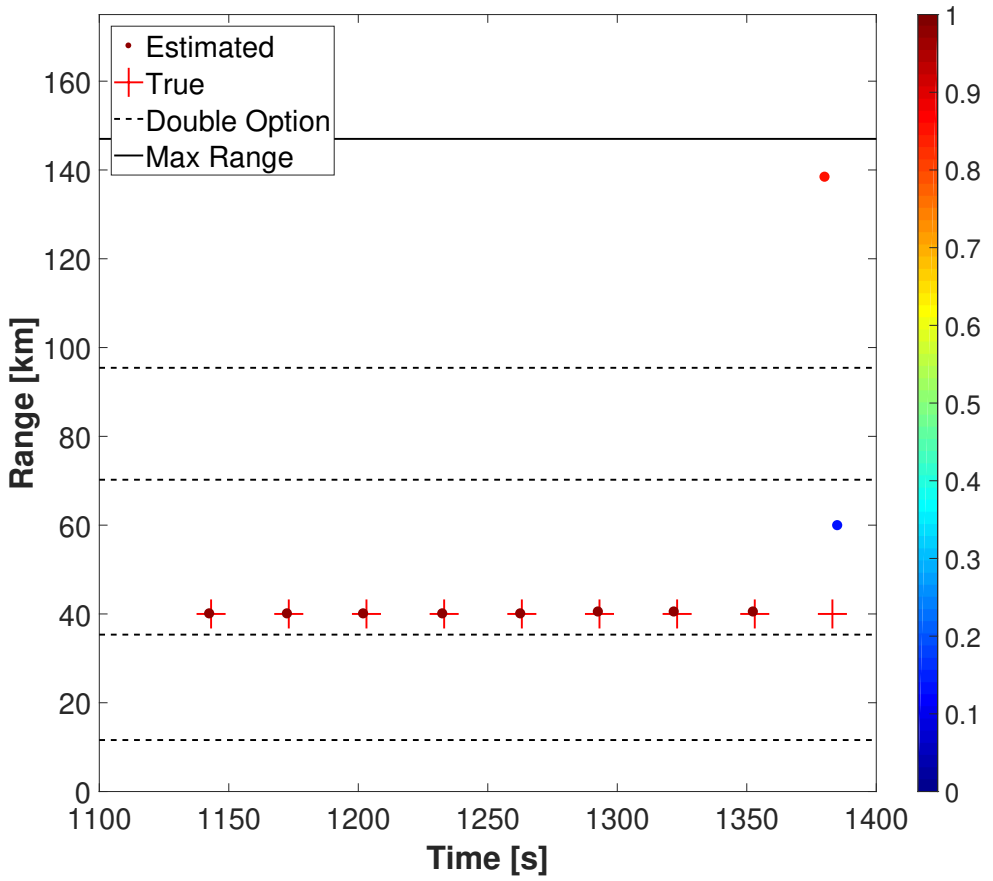

Figure B.12: True (red + symbols) vs. estimated (filled circles) source range for Simulation 2 and Source 1 after frequency and phantom pruning are applied. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

B.3 Simulation 3: Source range exceeds maximum candidate source range

In Simulation 3, nine different simulated calls were used to explore the effect on range estimates when the true call range exceeds the maximum range of the MPRATs. The track for Simulation 3 starts at a range of 138 km, with each subsequent call 4 km more distant from the hydrophone than the last out to a maximum range of 170 km. The maximum MPRAT range is 147 km. The results for the three different source waveforms are not significantly different, so only those for Source 1 are shown here in Figure B.13.

The first three sources are within the sensing range of the MPRATs, and localization yields reliable ranges estimates for these sources. The following five sources are beyond the maximum range. The fourth source yields an estimated range somewhat close to the true range, suggesting that if the source is just beyond the max sensing range the localizer will yield an estimated range at the maximum sensing range. However, the final five sources yield estimated ranges that are distant from the true range.

Figure B.14 shows the results of Simulation 3 for Source 1 after pruning has been applied to the results. In this case, phantom pruning eliminated the estimated range for the first call in this simulation, although the estimated ranges for the three subsequent calls remain.

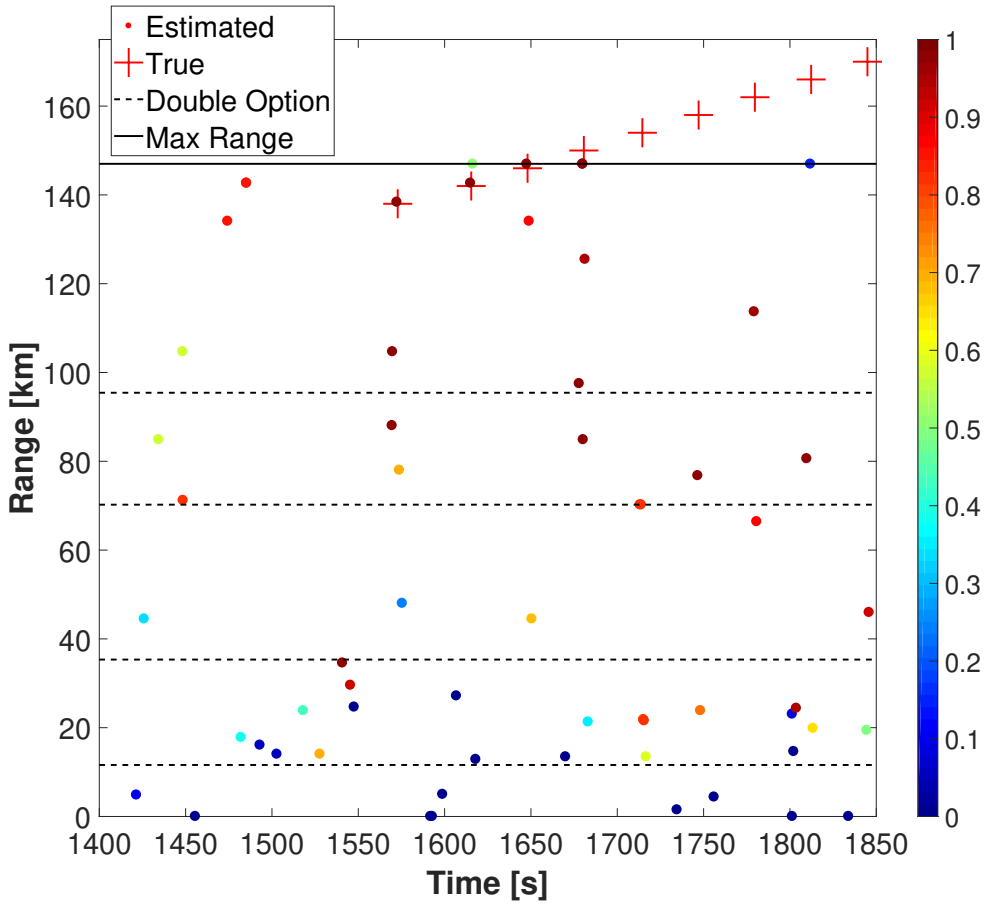

Figure B.13: True (red + symbols) vs. estimated (filled circles) source range for Simulation 3 and Source 1. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

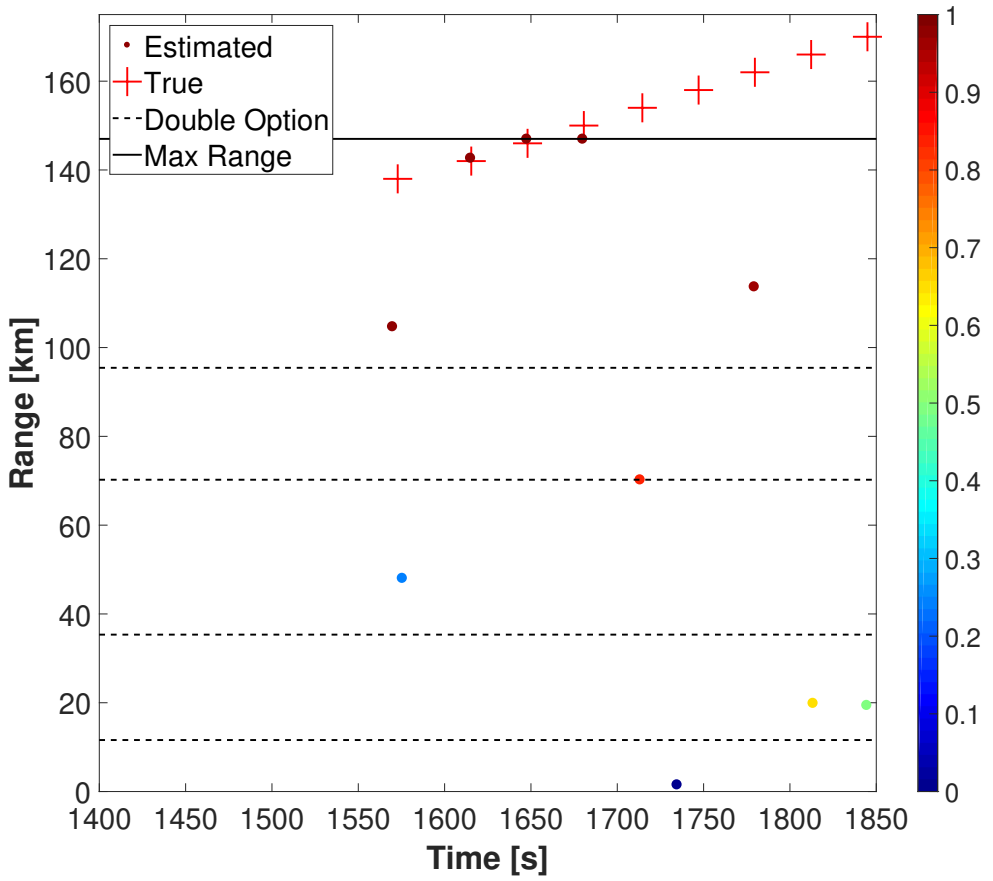

Figure B.14: True (red + symbols) vs. estimated (filled circles) source range for Simulation 3 and Source 1 after frequency and phantom pruning are applied. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

B.4 Simulation 4: Overlapping call sequences

To look at the effect of overlapping calls, pairs of tracks with the same source waveform and at the same source locations, but with a constant time offset between the calls in the two tracks, were used. In each case, two tracks (Track A and B), each composed of 5 individual sources, are used. The three time offsets between the tracks are: 8.5 s, 11 s, and 0.25 s. These three time offsets were specifically chosen to highlight three different degrees of obscurement of Track A by Track B.

B.4.1 Simulation 4a: No multipath overlap

The first overlap (8.5 s) was chosen so that the D arrival of Track B reaches the hydrophone between the BS2 and BS3 arrivals of Track A. The proximity of the D arrival in Track B to these higher-order arrivals from Track A mean that phantom pruning may make discerning Track B difficult. Prior to pruning (Figure B.15), all ten of the sources are correctly ranged. After pruning (Figure B.16), the first four calls in Track B are pruned away due to phantom pruning. However, the fifth call in Track A is also pruned away, allowing the fifth call in Track B to pass through the pruning process.

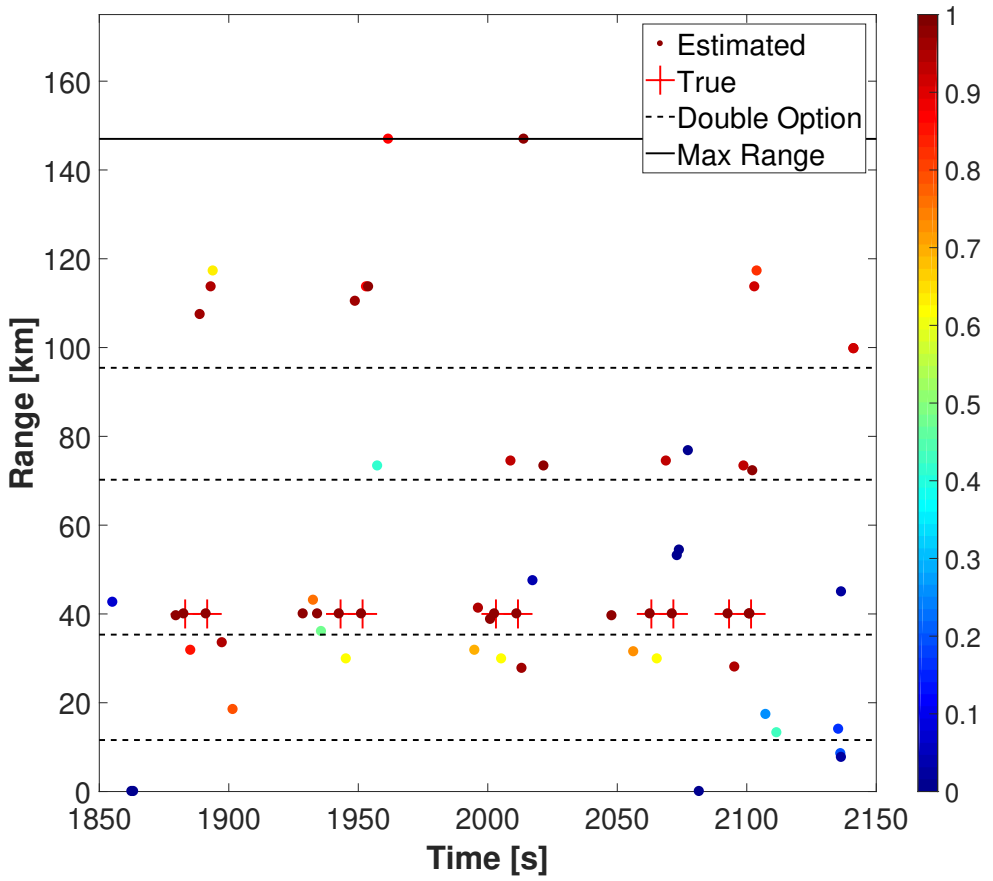

Figure B.15: True (red + symbols) vs. estimated (filled circles) source range for Simulation 4 and Source 1, with an offset of 8.5 s between the two tracks. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

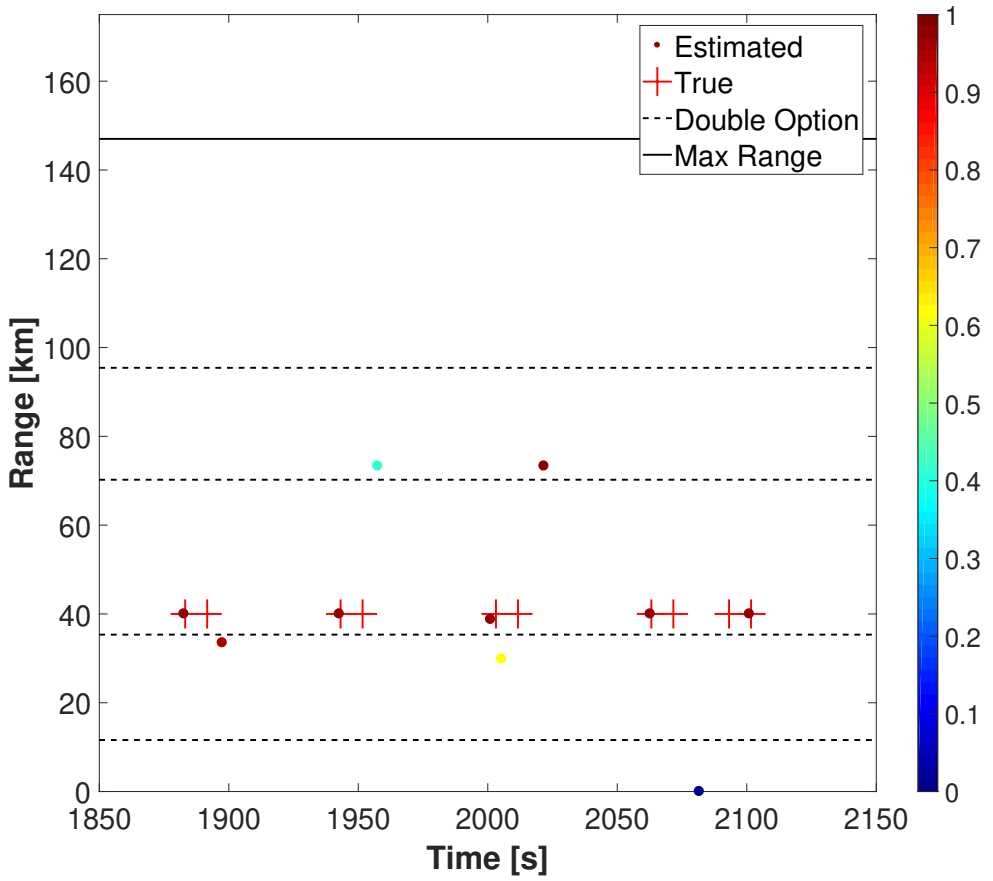

Figure B.16: True (red + symbols) vs. estimated (filled circles) source range for Simulation 4 and Source 1, with an offset of 8.5 s between the two tracks. Only post-pruning results are shown. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

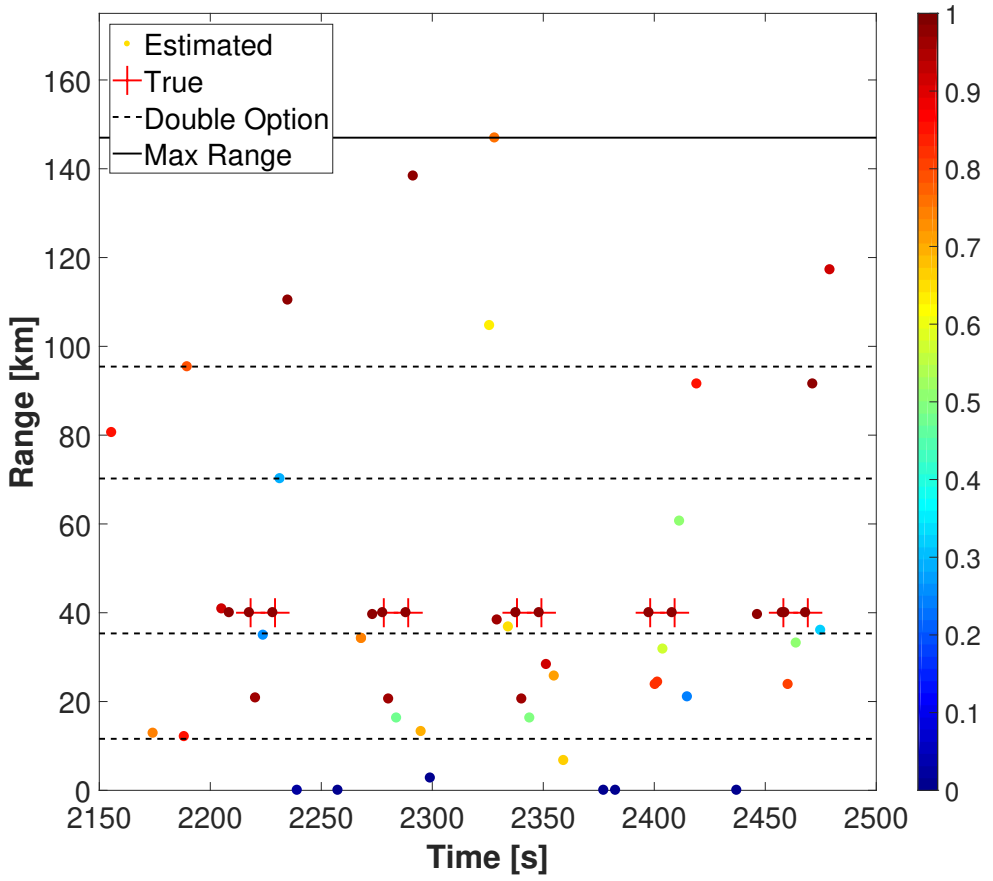

Figure B.17: True (red + symbols) vs. estimated (filled circles) source range for Simulation 4 and Source 1, with an offset of 11 s between the two tracks. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

B.4.2 Simulation 4b: Partial multipath overlap

The second overlap (11 s) was chosen so that the D arrival of Track B reaches the hydrophone at the same time as the BS4 arrival of Track A. As such, the sources in Track B are not expected to yield locations that survive the pruning process. Prior to pruning (Figure B.17), all ten sources across Tracks A and B are localized accurately, along with a collection of noisy range estimates. After pruning (Figure B.18), no sources from Track B are left, and most of the noisy range estimates have also been eliminated.

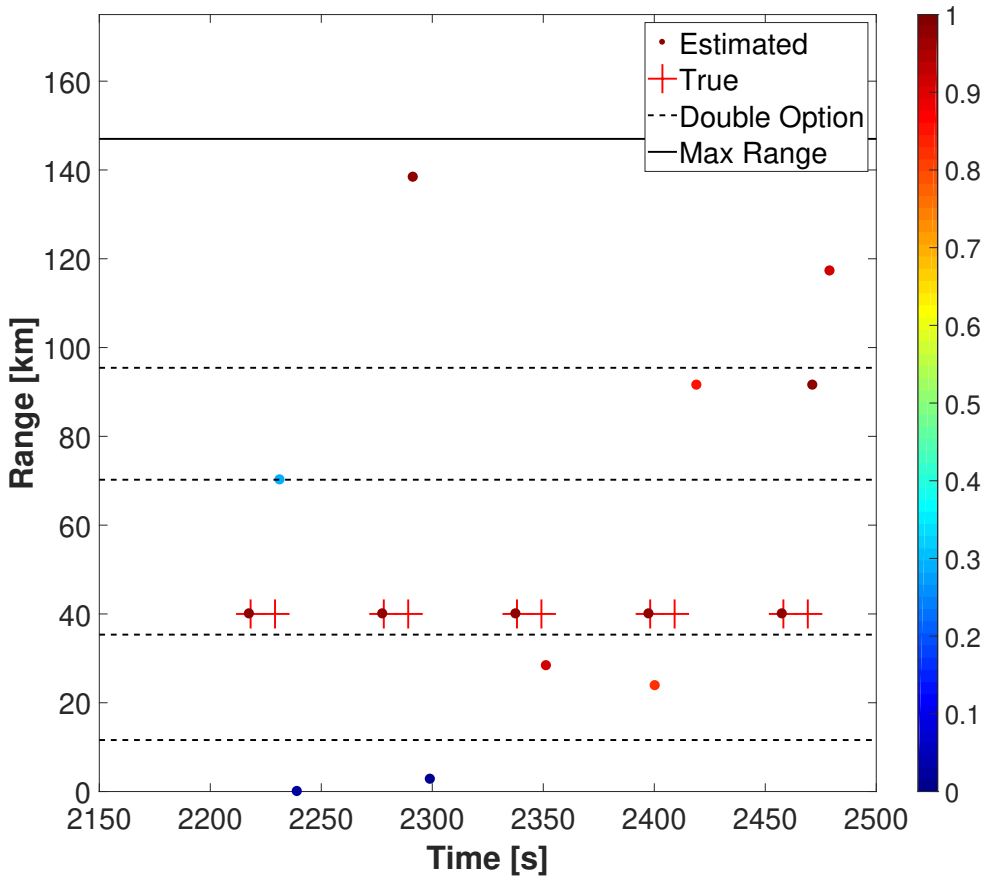

Figure B.18: True (red + symbols) vs. estimated (filled circles) source range for Simulation 4 and Source 1, with an offset of 11 s between the two tracks. Only post-pruning results are shown. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

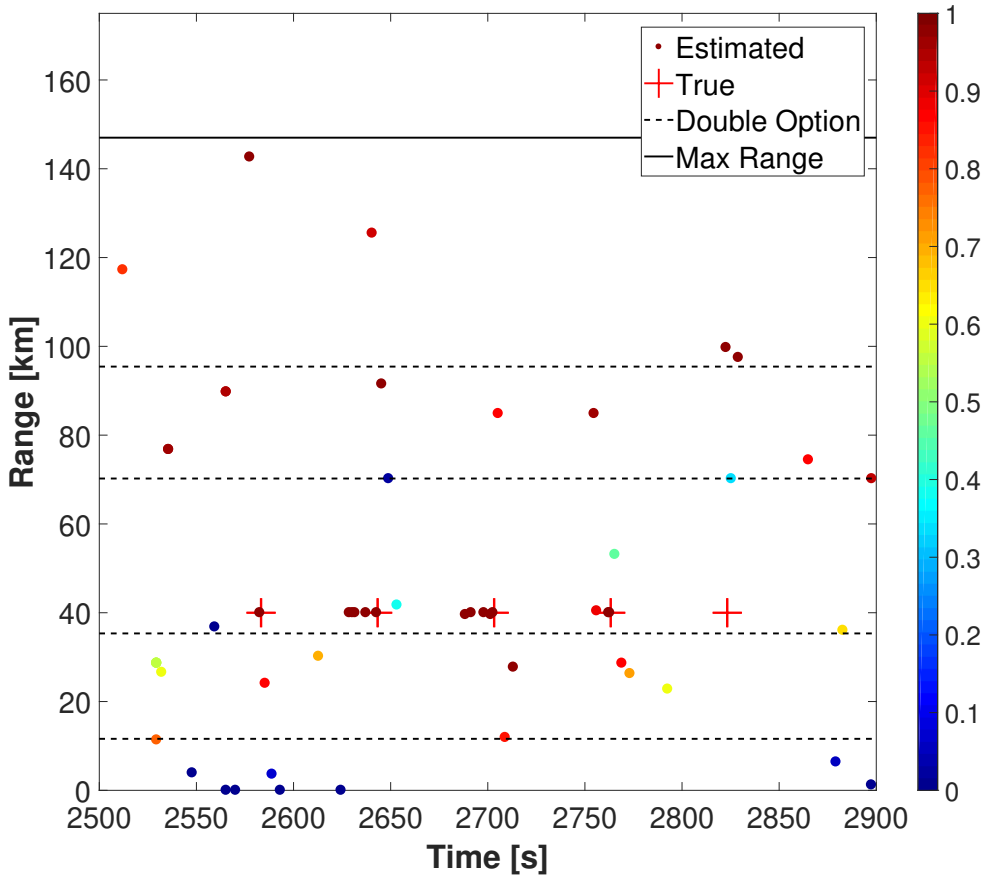

Figure B.19: True (red + symbols) vs. estimated (filled circles) source range for Simulation 4 and Source 1, with an offset of 0.5 s between the two tracks. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

B.4.3 Simulation 4c: Full multipath overlap

The third overlap (0.5 s) was chosen so that each arrival of Track B reaches the hydrophone half a second after the same arrival in Track A. With this degree of overlap, reliable localization of either Track A or Track B may be difficult, given that the waveform for each arrival is partially obscured. Prior to pruning (Figure B.19), the localizer yields good range estimates for the first 4 sources in Track A, but no sources from Track B. Following pruning (Figure B.20), no calls from either Track remain. For both figures, there is only 0.5 s between the Track A and Track B true source times, which explains why it appears that only 5 true source + symbols are present, compared to the 10 found in the other figures shown for Simulation 4.

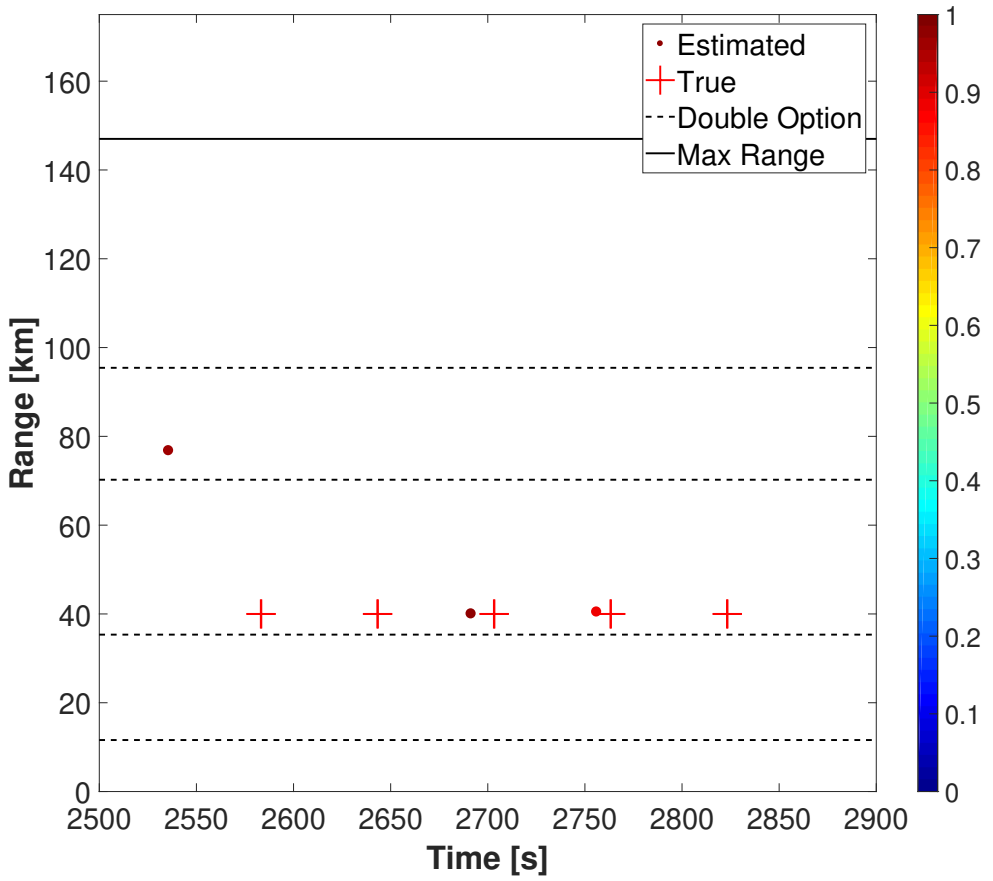

Figure B.20: True (red + symbols) vs. estimated (filled circles) source range for Simulation 4 and Source 1, with an offset of 0.5 s between the two tracks. Only post-pruning results are shown. Filled circle color indicates the ambiguity value for that estimated location on the color bar. Double option zone limits, and the range estimation limit, are shown as dashed and solid lines, respectively.

B.5 Discussion

In Simulation 1 (Section B.1), the ability of the localization approach to yield correct source ranges for a track that spans short:medium:far range is tested. The localization algorithm accurately estimates source range for all calls prior to pruning. Most of the false positive detections are eliminated during the pruning process, at the cost of also eliminating several true positives. While the loss of true positives is unfortunate, this is not seen as a significant flaw in the localization approach presented here. While a tracking algorithm is not implemented in this work, it is anticipated that such a tracking approach would be capable of gracefully dealing with small holes in an otherwise consistent track such as is used in Simulation 1, or would be able to overcome clutter caused by false positives. It is also shown in Simulation 1 that the three sources produced similar results. Because of this, only results using source 1 are shown for the remaining 3 simulations.

In Simulation 2 (Section B.2), the behavior of the localization approach if source depth is not precisely at 50 m is tested. Prior to pruning, the range for each source is within 900 m of the true range (40 km) when the actual source depth varies from 10 m to 500 m. As source depth increases, there is a slight tendency to over-estimate range, although this effect is small relative to source range.

In Simulation 3 (Section B.3), the effect of sources whose range exceeds the largest candidate source range was tested. The maximum range from the set of candidate source ranges is 147 km. In Simulation 3, nine simulated calls between 138 - 170 km are located. Prior to pruning, accurate range estimates are made for the three sources whose range is less than largest candidate source range. The source whose range is slightly longer than the maximum range still yields a plausible range estimate, but the remaining 5 sources in the track do not yield useful range estimates. This suggests that there is a source range (somewhat larger than the largest candidate source range) beyond which reliable range estimation is no longer possible. During the pruning for Simulation 3, if a noisy detection occurs just before the track begins, the first source in the track is pruned away by the phantom pruning criteria. This illustrates an occasional, but not particularly problematic, drawback of our localization approach wherein from time to time a noisy detection will satisfy the pruning criteria at just the right moment to obscure a call of interest.

Simulation 4 (Section B.4) investigated the effect of call overlap on range estimation. The figures in this section showed that the phantom pruning criteria can complicate range estimation of a closely-following call. In addition, when all the arrivals of the first call are partially obscured by those of the second (such as might happen when two calls at the same location are made one immediately following the other), the second call may not appear in the set of pre-pruned range estimates at all.

BIBLIOGRAPHY

Whitlow W. L. Au and Marc O. Lammers, editors. *Listening in the Ocean*. Springer New York, New York, NY, 2016. ISBN 978-1-4939-3175-0. doi: 10.1007/978-1-4939-3176-7.

Roland Aubauer, Marc O. Lammers, and Whitlow W. L. Au. One-hydrophone method of estimating distance and depth of phonating dolphins in shallow water. *The Journal of the Acoustical Society of America*, 107 (5 Pt 1):2744–2749, 5 2000. ISSN 0001-4966.

Arthur B. Baggeroer, William A. Kuperman, and Peter N. Mikhalevsky. An overview of matched field methods in ocean acoustics. *IEEE Journal of Oceanic Engineering*, 18(4):401–424, 1993. ISSN 03649059. doi: 10.1109/48.262292. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=262292`.

Kevin P. Balanda and H. L. MacGillivray. Kurtosis: a critical review. *The American Statistician*, 42(2): 111–119, 5 1988. ISSN 00031305. doi: 10.2307/2684482.

Richard G. Baraniuk, Emmanuel J. Candès, Michael Elad, and Yi Ma. Applications of sparse representation and compressive sensing [scanning the issue]. *Proceedings of the IEEE*, 98(6):906–909, 2010.

Stephen Becker. *Practical Compressed Sensing: modern data acquisition and signal processing*. PhD thesis, California Institute of Technology, 2011.

Stephen Becker, Jérôme Bobin, and Emmanuel J. Candès. NESTA: a fast and accurate first-order method for sparse recovery. *SIAM Journal on Imaging Sciences*, 91125:1–37, 2011.

Michael J. Bianco, Peter Gerstoft, James Traer, Emma Ozanich, Marie A. Roch, Sharon Gannot, and Charles-Alban Deledalle. Machine learning in acoustics: Theory and applications. *The Journal of the Acoustical Society of America*, 146(5):3590–3628, 2019. ISSN 0001-4966. doi: 10.1121/1.5133944.

Susanna B. Blackwell, Christopher S. Nations, Trent L. McDonald, Charles R. Greene, Aaron M. Thode, Melania Guerra, and A. Michael Macrander. Effects of airgun sounds on bowhead whale calling rates: Evidence for two behavioral thresholds. *Marine Mammal Science*, 29(4):342–365, 2013. ISSN 19326203. doi: 10.1371/journal.pone.0125720.

Richard E. Blahut. *Algebraic Methods for Signal Processing and Communications Coding*. Springer Science & Business Media, New York, 2012.

J. Bonnel, C. Gervaise, B. Nicolas, and J. I. Mars. Single-receiver geoacoustic inversion using modal reversal. *The Journal of the Acoustical Society of America*, 131(1):119–128, 2012. ISSN 0001-4966. doi: 10.1121/1.3664083.

Julien Bonnel, Barbara Nicolas, Jérôme I. Mars, and Dominique Fattaccioli. Frequency warping for waveguide characterization with a single hydrophone. In *IEEE OCEANS 2009*, number June 2014, pages 1–5, 2009. ISBN 9781424449606. doi: 10.23919/oceans.2009.5422406.

Julien Bonnel, Barbara Nicolas, Jérôme I Mars, and Dominique Fattaccioli. Single receiver inversion in shallow water. In *ECUA 2010 - 10th European Conference on Underwater Acoustics*, number April 2015, pages 1–5, 2010.

Julien Bonnel, Stan E. Dosso, and N. Ross Chapman. Bayesian geoacoustic inversion of single hydrophone light bulb data using warping dispersion analysis. *The Journal of the Acoustical Society of America*, 134 (1):120–130, 2013. ISSN 0001-4966. doi: 10.1121/1.4809678.

Julien Bonnel, Aaron Thode, Dana Wright, and Ross Chapman. Nonlinear time-warping made simple: A step-by-step tutorial on underwater acoustic modal separation with a single hydrophone. *The Journal of the Acoustical Society of America*, 147(3):1897–1926, 2020. ISSN 0001-4966. doi: 10.1121/10.0000937.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004. ISBN 9780511804441. doi: 10.1017/CBO9780511804441. URL `http://ebooks.cambridge.org/ref/id/CBO9780511804441`.

Douglas M. Burn, Marc A. Webber, and Mark S. Udevitz. Application of airborne thermal imagery to surveys of Pacific walrus. *Wildlife Society Bulletin*, 34(1):51–58, 2006. ISSN 0091-7648. doi: 10.2193/0091-7648(2006)34[51:AOATIT]2.0.CO;2.

Chen-Tung Chen and Frank J. Millero. The Sound Speed in Seawater at high pressures, 1977.

Jessica Chen, Adam A. Pack, Whitlow W. L. Au, and Alison K. Stimpert. Measurements of humpback whale song sound levels received by a calf in association with a singer. *The Journal of the Acoustical Society of America*, 140(5):4010–4015, 2016. ISSN 0001-4966. doi: 10.1121/1.4967444. URL `http://asa.scitation.org/doi/10.1121/1.4967444`.

Scott Shaobing Chen, David L. Donoho, and Michael A. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 43(1):129–159, 2001. URL `http://epubs.siam.org/doi/abs/10.1137/S1064827596304010`.

Christopher W. Clark, William T. Ellison, Brandon L. Southall, Leila Hatch, Sofie M. Van Parijs, Adam Frankel, and Dimitri Ponirakis. Acoustic masking in marine ecosystems: intuitions, analysis, and implication. *Marine Ecology Progress Series*, 395:201–222, 12 2009. ISSN 0171-8630. doi: 10.3354/meps08402. URL `http://www.int-res.com/abstracts/meps/v395/p201-222/`.

C. S. Clay. Optimum time domain signal transmission and source location in a waveguide. *The Journal of the Acoustical Society of America*, 81(3):660–664, 1987. doi: 10.1121/1.394834.

Ross Compton, Lissa Goodwin, Richard Handy, and Victor Abbott. A critical examination of worldwide guidelines for minimising the disturbance to marine mammals during seismic surveys. *Marine Policy*, 32 (3):255–262, 2008. ISSN 0308597X. doi: 10.1016/j.marpol.2007.05.005.

Donald A. Croll, Christopher W. Clark, Alejandro Acevedo, Bernie Tershy, Sergio Flores, Jason Gedamke, and Jorge Urban. Only male fin whales sing loud songs. *Nature*, 417:809–809, 2002.

L. C. Cuyler, R. Wiulsrød, and N. A. ØRitsland. Thermal Infrared Radiation From Free Living Whales. *Marine Mammal Science*, 8(2):120–134, 1992.

Ingrid Daubechies, Massimo Fornasier, and Ignace Loris. Accelerated Projected Gradient Method for Linear Inverse Problems with Sparsity Constraints. *Journal of Fourier Analysis and Applications*, 14(5-6):764–792, 2008. ISSN 1069-5869. doi: 10.1007/s00041-008-9039-8. URL http://link.springer.com/10.1007/s00041-008-9039-8.

Kimberley T.A. Davies and Sean W. Brillant. Mass human-caused mortality spurs federal action to protect endangered North Atlantic right whales in Canada. *Marine Policy*, 104(September 2018):157–162, 2019. ISSN 0308597X. doi: 10.1016/j.marpol.2019.02.019. URL https://doi.org/10.1016/j.marpol.2019.02.019.

David L. Donoho. Compressed Sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 4 2006. ISSN 0018-9448. doi: 10.1109/TIT.2006.871582. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1614066.

Stan E. Dosso and Gordon R. Ebbeson. Array Element Localization Accuracy and Survey Design. *Canadian Acoustics*, 34(4):3–13, 2006.

Stan E. Dosso and Barbara J. Sotirin. Optimal array element localization. *The Journal of the Acoustical Society of America*, 106(6):3445–3459, 1999. ISSN 00014966. doi: 10.1121/1.428198. URL http://link.aip.org/link/JASMAN/v106/i6/p3445/s1&Agg=doihttp://link.aip.org/link/JASMAN/v106/i6/p3445/s1{&}Agg=doi.

Rui Duan, Kunde Yang, Yuanliang Ma, Qiulong Yang, and Hui Li. Moving source localization with a single hydrophone using multipath time delays in the deep ocean. *The Journal of the Acoustical Society of America*, 136(2):EL159–EL165, 2014. doi: 10.1121/1.4890664.

Fred Duennebier, David Harris, and James Jolly. ALOHA Cabled Observatory Will Monitor Ocean in Real Time. *Sea Technology*, 49(2):51–54, 2008. ISSN 00933651.

Brian D. Dushaw, Peter F. Worcester, Walter H. Munk, Robert C. Spindel, James A. Mercer, Bruce M. Howe, K. Metzger Jr., T. G. Birdsall, R. K. Andrew, M. A. Dzieciuch, B. D. Cornuelle, and D. Menemenlis. A decade of acoustic thermometry in the North Pacific Ocean. *Journal of Geophysical Research: Oceans*, 114(7):1–24, 2009. ISSN 21699291. doi: 10.1029/2008JC005124.

Nicholas P. Fofonoff and R. C. Millard Jr. Algorithms for computation of fundamental properties of seawater. Technical report, UNESCO, 1983.

L. Neil Frazer and Peter I. Pecholcs. Single-hydrophone localization. *The Journal of the Acoustical Society of America*, 88(2):995–1002, 1990. doi: 10.1121/1.399750.

George V. Frisk. *Ocean and Seabed Acoustics*. Prentice-Hall, Englewood Cliff, NJ, 1994.

Giacomo Giorli, Jeffrey C. Drazen, Anna B. Neuheimer, Adrienne Copeland, and Whitlow W.L. Au. Deep sea animal density and size estimated using a Dual-frequency IDentification SONar (DIDSON) offshore the island of Hawaii. *Progress in Oceanography*, 160(June 2016):155–166, 2018. ISSN 00796611. doi: 10.1016/j.pocean.2018.01.002.

Jonathan Gordon, Douglas Gillespie, John Potter, Alexandros Frantzis, Mark P. Simmonds, René Swift, and David Thompson. A Review of the Effects of Seismic Surveys on Marine Mammals. *Marine Technology Society Journal*, 37:14–32, 2004. ISSN 00253324. doi: 10.4031/002533203787536998. URL http://www.okeanos-foundation.org/assets/Uploads/Gordon-et-al.-2004-Review-of-Seismic-Surveys-Effects-16-45-47.pdf.

Joseph Graber, Jim Thomson, Brian Polagye, and Andrew Jessup. Land-based infrared imagery for marine mammal detection. *Proceedings of SPIE - The International Society for Optical Engineering*, 8156:81560B–81560B–11, 2011. ISSN 0277786X. doi: 10.1117/12.892787. URL http://link.aip.org/link/PSISDG/v8156/i1/p81560B/s1&Agg=doi.

Elaine T. Hale, Wotao Yin, and Yin Zhang. A Fixed-Point Continuation Method for L1-Regularized Minimization with Applications to Compressed Sensing. *CAAM Technical Report*, (June 2015):1–45, 2007.

Danielle Harris, Luis Matias, Len Thomas, John Harwood, and Wolfram H. Geissler. Applying distance sampling to fin whale calls recorded by single seismic instruments in the northeast Atlantic. *The Journal of the Acoustical Society of America*, 134(5):3522–3535, 2013. ISSN 0001-4966. doi: 10.1121/1.4821207.

Keith C. Hester, Edward T. Peltzer, William J. Kirkwood, and Peter G. Brewer. Unanticipated consequences of ocean acidification: A noisier ocean at lower pH. *Geophysical Research Letters*, 35(19):1–5, 2008. ISSN 00948276. doi: 10.1029/2008GL034913.

Bruce M. Howe, Roger Lukas, Fred Duennebier, and David Karl. ALOHA Cabled Observatory Installation. *Oceans 2011*, pages 1–11, 2011. ISSN 00278424. doi: 10.1073/pnas.1113095109.

Chadwick V. Jay, Mads Peter Heide-Jorgensen, Anthony S. Fischbach, Mikkel V. Jensen, David F. Tessler, and Anders V. Jensen. Comparison of remotely deployed satellite radio transmitters on walruses. *Marine Mammal Science*, 22(1):226–236, 2006.

Chadwick V. Jay, Mark S. Udevitz, Ron Kwok, Anthony S. Fischbach, and David C. Douglas. Divergent movements of walrus and sea ice in the northern Bering Sea. *Marine Ecology Progress Series*, 407:293–302, 5 2010. ISSN 0171-8630. doi: 10.3354/meps08575. URL http://www.int-res.com/abstracts/meps/v407/p293-302/.

Thomas A. Jefferson, Marc A Webber, and Robert L. Pitman. *Marine Mammals of the World: A Comprehensive Guide to Their Identification*. Academic Press, London, second edition, 2015.

Sérgio. M. Jesus, Michael B. Porter, Yann Stéphan, Emanual Coelho, and Xavier Démoulin. Broadband source localization with a single hydrophone. In *IEEE Oceans 1998*, volume 2, pages 1078–1082, 1998. doi: 10.1109/oceans.1998.724402.

Mark P. Johnson and Peter L. Tyack. A digital acoustic recording tag for measuring the response of wild marine mammals to sound. *IEEE Journal of Oceanic Engineering*, 28(1):3–12, 2003. ISSN 03649059. doi: 10.1109/JOE.2002.808212.

Holger Klinck, David K. Mellinger, Karolin Klinck, Neil M. Bogue, James C. Luby, William A. Jump, Geoffrey B. Shilling, Trina Litchendorf, Angela S. Wood, Gregory S. Schorr, and Robin W. Baird. Near-real-time acoustic monitoring of beaked whales and other cetaceans using a Seaglider™. *PLoS ONE*, 7(5): 1–8, 2012. ISSN 19326203. doi: 10.1371/journal.pone.0036128.

Anke Kügler, Marc O. Lammers, Eden J. Zang, Maxwell B. Kaplan, and T. Aran Mooney. Fluctuations in Hawaii'S Humpback Whale Megaptera Novaeangliae Population Inferred from Male Song Chorusing Off Maui. *Endangered Species Research*, 43:421–434, 2020. ISSN 16134796. doi: 10.3354/ESR01080. URL http://dx.doi.org/10.3354/esr01080.

Anke Kügler, Marc O. Lammers, Eden J. Zang, and Adam A. Pack. Male Humpback Whale Chorusing in Hawai'i and Its Relationship With Whale Abundance and Density. *Frontiers in Marine Science*, 8 (December):1–16, 2021. ISSN 22967745. doi: 10.3389/fmars.2021.735664.

Marc O. Lammers, Michiel Schotten, and Whitlow W. L. Au. The spatial context of free-ranging Hawaiian spinner dolphins (Stenella longirostris) producing acoustic signals. *The Journal of the Acoustical Society of America*, 119(2):1244–1250, 2006. ISSN 00014966. doi: 10.1121/1.2151804.

Christophe Laplanche, Olivier Adam, Maciej Lopatka, and Jean-Francois Motsch. Male sperm whale acoustic behavior observed from multipaths at a single hydrophone. *The Journal of the Acoustical Society of America*, 118(4):2677–2687, 2005. doi: 10.1121/1.2033567. URL http://link.aip.org/link/?JASMAN/118/2677/1.

C. H. Lockyer and R. J. Morris. Individual Recognition of Cetaceans: Use of Photo-Identification and Other Techniques to estimate Population Parameters. *Report of the International Whaling Commission*, (12): 113–118, 1990. ISSN 0722-4060. doi: 10.1098/rsbl.2010.0638. URL http://www.vliz.be/imisdocs/publications/253951.pdf#page=203%0Ahttp://www.vliz.be/imisdocs/publications/253951.pdf%23page=343%5Cnhttp://www.vliz.be/imisdocs/publications/253951.pdf%23page=121.

Stephane G. Mallat and Zhifeng Zhang. Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, 41(12):3397–3415, 1993. URL http://ieeexplore.ieee.org/xpls/abs%7B_%7Dall.jsp?arnumber=258082http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=258082.

Tiago A. Marques, Len Thomas, Jessica Ward, Nancy DiMarzio, and Peter L. Tyack. Estimating cetacean population density using fixed passive acoustic sensors: an example with Blainville's beaked whales. *The Journal of the Acoustical Society of America*, 125(4):1982–1994, 4 2009. ISSN 1520-8524. doi: 10.1121/1.3089590. URL http://www.ncbi.nlm.nih.gov/pubmed/19354374.

Bruce R. Mate, Amanda Bradford, Grisha Tsidulko, Vladimir Vertyankin, and Valentin Ilyashenko. Late-feeding season movements of a western North Pacific gray whale off Sakhalin Island, Russia and subsequent migration into the Eastern North Pacific. In *Proceedings of the 19th Biennial Conference on the Biology of Marine Mammals*, page 193, Tampa, Florida, 2011. URL http://archive.iwcoffice.org/_documents/sci_com/SC63docs/SC-63-BRG23.pdf%5Cnpapers2://publication/uuid/43C1FB70-1DB8-47CD-838C-1C2A5140BE5A.

Luis Matias and Danielle Harris. A single-station method for the detection , classification and location of fin whale calls using ocean-bottom seismic stations. *The Journal of the Acoustical Society of America*, 138 (1):504–520, 2015. doi: 10.1121/1.4922706. URL http://dx.doi.org/10.1121/1.4922706.

Dominic J. McCafferty. The value of infrared thermography for research on mammals: Previous applications and future directions. *Mammal Review*, 37(3):207–223, 2007. ISSN 13652907. doi: 10.1111/j.1365-2907.2007.00111.x.

Mark A. McDonald and Christopher G. Fox. Passive acoustic methods applied to fin whale population density estimation. *The Journal of the Acoustical Society of America*, 105(5):2643–2651, 1999. ISSN 00014966. doi: 10.1121/1.426880. URL http://link.aip.org/link/JASMAN/v105/i5/p2643/s1%7B&%7DAgg=doi.

Mark A. Mcdonald, John A. Hildebrand, and Spahr C. Webb. Blue and fin whales observed on a seafloor array in the Northeast Pacific. *The Journal of the Acoustical Society of America*, 98(2):712–721, 1995. ISSN 00014966. doi: 10.1121/1.413565.

Mark A. McDonald, John A. Hildebrand, and Sean M. Wiggins. Increases in deep ocean ambient noise in the Northeast Pacific west of San Nicolas Island, California. *The Journal of the Acoustical Society of America*, 120(2):711–718, 2006. ISSN 0001-4966. doi: 10.1121/1.2216565. URL http://asa.scitation.org/doi/10.1121/1.2216565.

H. Medwin and C. S. Clay. *Fundamentals of Acoustical Oceanography*. Academic Press, New York, 1998.

David K. Mellinger and Christopher W. Clark. Recognizing transient low-frequency whale sounds by spectrogram correlation. Technical report, 2000. URL http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp.

Eric Moulines, Pierre Duhamel, J.-F. Cardoso, and Sylvie Mayrargue. Subspace methods for the blind identification of multichannel FIR filters. *IEEE Transactions on Signal Processing*, 43(2):516–525, 1995. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=348133http://ieeexplore.ieee.org/xpls/abs%7B_%7Dall.jsp?arnumber=348133.

Xavier Mouy, David E. Hannay, Mikhail Zykov, and Bruce Martin. Tracking of Pacific walruses in the Chukchi Sea using a single hydrophone. *The Journal of the Acoustical Society of America*, 131(2):1349–1358, 2012.

Lisa M. Munger, Sean M. Wiggins, and John A. Hildebrand. North Pacific right whale up-call source levels and propagation distance on the southeastern Bering Sea shelf. *The Journal of the Acoustical Society of America*, 129(6):4047–4054, 2011. ISSN 0001-4966. doi: 10.1121/1.3557060.

Deanna Needell and Joel A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied and Computational Harmonic Analysis*, 26:301–321, 2009. URL http://www.sciencedirect.com/science/article/pii/S1063520308000638.

Yurii Nesterov. A method for unconstrained convex minimization problem with the rate of convergence O (1/kˆ2). *Doklady AN USSR*, 269, 1983.

Yurii Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1):127–152, 2005. ISSN 00255610. doi: 10.1007/s10107-004-0552-5.

Eva-Marie Nosal and L. Neil Frazer. Track of a sperm whale from delays between direct and surface-reflected clicks. *Applied Acoustics*, 67(11-12):1187–1201, 11 2006. ISSN 0003682X. doi: 10.1016/j.apacoust.2006.05.005.

Albert H. Nuttall. Detection Performance of Power-Law Processors for Random Signals of Unknown Location, Structure, Extent, and Strength. Technical report, 1994.

Ewan S. Page. Continuous Inspection Schemes. *Biometrika*, 41(1):100–115, 1954.

Wayne L. Perryman, Meghan A. Donahue, Jeffrey L. Laake, and Thomas E. Martin. Diel variation in migration rates of eastern Pacific gray whales measured with thermal imaging sensors. 15(2):426–445, 1999.

Michael B. Porter. The KRAKEN Normal Mode Program. *Naval Research Laboratory*, pages 1–198, 1992. ISSN 1098-6596. doi: 10.1017/CBO9781107415324.004.

Michael B. Porter and Yong-Chun Liu. Finite-element ray tracing. *Theoretical and Computational Acoustics*, 2:947–956, 1994. URL `http://www.hlsresearch.com/personnel/porter/papers/ICTCPraypaper.pdfhttp://www.hlsresearch.com/personnel/porter/papers/ICTCPraypaper.pdf`.

Shannon Rankin and Jay Barlow. Vocalizations of the Sei whale Balaenoptera borealis off the hawaiian islands. *Bioacoustics*, 16:137–145, 2007. ISSN 0952-4622. doi: 10.1080/09524622.2007.9753572.

Brendan P. Rideout, Stan E. Dosso, and David E. Hannay. Underwater passive acoustic localization of Pacific walruses in the northeastern Chukchi Sea. *The Journal of the Acoustical Society of America*, 134 (3):2534–2545, 2013. doi: 10.1121/1.4816580.

Marie A. Roch, Peter Gerstoft, Bozena Kostek, and Zoi-Heleni Michalopoulou. How Machine Learning Contributes to Solve Acoustical Problems. *Acoustics Today*, 17(4):48–56, 2021. ISSN 15570215. doi: 10.1121/at.2021.17.4.48.

Rudy Rotili, Emanuele Principi, Stefano Squartini, and Bjorn Schuller. A Real-Time Speech Enhancement Framework in Noisy and Reverberated Acoustic Scenarios. *Cognitive Computation*, (August):2012, 2012. doi: 10.1007/s12559-012-9176-x.

Yoichi Sato. A method of self-recovering equalization for multilevel amplitude-modulation systems. *IEEE Transactions on Communications*, 23(6):679–682, 1975.

William E. Schevill, William A. Watkins, and R. H. Backus. The 20-cycle signals and Balaenoptera (Fin Whales). In W. N. Tavolga, editor, *Marine Bio-Acoustics*, pages 147–152. Pergamon Press Ltd, 1963.

Donna Shillington, Anthony Watts, Robert Dunn, Garrett Ito, and Paul Wessel. Cruise MGL1806 on RV Marcus G. Langseth: Hawaii Emperor Seamount Phase 1, 2018.

B. S. Shylaja. From navigation to star hopping: forgotten formulae. *Resonance*, 20(4):352–359, 2015. ISSN 0973712X. doi: 10.1007/s12045-015-0190-7.

Kara Silver. A Passive Acoustic Automated Detector for Sei and Fin Whale Calls. Technical report, University of Hawaii at Manoa, Honolulu, Hawaii, 2014.

Dax C. Soule and William S. D. Wilcock. Fin whale tracks recorded by a seismic network on the Juan de Fuca Ridge, Northeast Pacific Ocean. *The Journal of the Acoustical Society of America*, 133(3):1751–1761, 2013.

John L. Spiesberger and Kurt M. Fristrup. Passive Localization of Calling Animals and Sensing of their Acoustic Environment Using Acoustic Tomography. *The American Naturalist*, 135(1):107–153, 1990. ISSN 0003-0147. doi: 10.1086/285035.

Kathleen M. Stafford, Christopher G. Fox, and David S. Clark. Long-range acoustic detection and localization of blue whale calls in the northeast Pacific Ocean. *The Journal of the Acoustical Society of America*, 104 (6):3616–3625, 12 1998. ISSN 0001-4966.

Alison K. Stimpert, David N. Wiley, Whitlow W.L. Au, Mark P. Johnson, and Roland Arsenault. 'Megapclicks': Acoustic click trains and buzzes produced during night-time foraging of humpback whales (Megaptera novaeangliae). *Biology Letters*, 3(5):467–470, 2007. ISSN 1744957X. doi: 10.1098/rsbl.2007.0281.

Aaron M. Thode and Julien Bonnel. Experimental demonstration of time warping to invert for array tilt and mode shape on a vertical array in a shallow arctic environment. In *IEEE OCEANS 2015 - Genova*, number May, pages 1–7, 2015. ISBN 9781479987368. doi: 10.1109/OCEANS-Genova.2015.7271636.

Aaron M. Thode, Katherine H. Kim, Susanna B. Blackwell, Charles R. Greene, Christopher S. Nations, Trent L. McDonald, and A. Michael Macrander. Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys. *The Journal of the Acoustical Society of America*, 131 (5):3726–3747, 2012. ISSN 0001-4966. doi: 10.1121/1.3699247.

Paul O. Thompson and William A. Friedl. A Long Term Study of Low Frequency Sounds from Several Species of Whales Off Oahu Hawaii USA. *Cetology*, 45(45):1–19, 1982. ISSN 0097-031X.

Christopher O. Tiemann, Aaron M. Thode, Janice Straley, Victoria O'Connell, and Kendall Folkert. Three-dimensional localization of sperm whales using a single hydrophone. *The Journal of the Acoustical Society of America*, 120(4):2355–2365, 2006. ISSN 00014966. doi: 10.1121/1.2335577.

Joel A. Tropp and Anna C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 53(12):4655–4666, 2007. URL `http://ieeexplore.ieee.org/xpls/abs%7B_%7Dall.jsp?arnumber=4385788`.

Robert D. Valtierra, R. Glynn Holt, Danielle Cholewiak, and Sofie M. Van Parijs. Calling depths of baleen whales from single sensor data: Development of an autocorrelation method using multipath localization. *The Journal of the Acoustical Society of America*, 134(3 Pt. 2):2571–2581, 2013.

Harry L. Van Trees Jr, Kristine L. Bell, and Zhi Tian. *Detection, Estimation, and Modulation Theory - Part 1: Detection, Estimation, and Filtering Theory*. Wiley, New York, 2nd editio edition, 2013.

R. A. Walker. Some Intense, Low-Frequency, Underwater Sounds of Wide Geographic Distribution, Apparently of Biological Origin. *The Journal of the Acoustical Society of America*, 35(11):1816–1824, 1963. ISSN NA. doi: 10.1121/1.1918828.

Graham A. Warner, Stan E. Dosso, Jan Dettmer, and David E. Hannay. Bayesian environmental inversion of airgun modal dispersion using a single hydrophone in the Chukchi Sea. *The Journal of the Acoustical Society of America*, 137(6):3009–3023, 2015. ISSN 0001-4966. doi: 10.1121/1.4921284. URL `http://dx.doi.org/10.1121/1.4921284`.

Graham A. Warner, Stan E. Dosso, and David E. Hannay. Bowhead whale localization using time-difference-of-arrival data from asynchronous recorders. *The Journal of the Acoustical Society of America*, 141 (3):1921–1935, 2017. ISSN 0001-4966. doi: 10.1121/1.4978438. URL `http://dx.doi.org/10.1121/1.4978438`.

William A. Watkins. Activities and underwater sounds of fin whales. *Scientific Report of Whale Research Institure*, (33):83–117, 1981.

William A. Watkins and William E. Schevill. Four Hydrophone Array for Acoustic Three-Dimensional Location. Technical Report 71, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts, 1971.

William A. Watkins and William E. Schevill. Sound source location by arrival-times on a non-rigid three-dimensional hydrophone array. *Deep-Sea Research*, 19(10):691–706, 1972.

William A. Watkins and William E. Schevill. Spatial distribution of Physeter catodon (sperm whales) underwater. *Deep-Sea Research*, 24:693–699, 1977.

William A. Watkins, Peter Tyack, Karen E. Moore, and James E. Bird. The 20-Hz signals of finback whales (Balaenoptera physalus). *The Journal of the Acoustical Society of America*, 82(6):1901–1912, 1987. ISSN 0001-4966. doi: 10.1121/1.380594. URL `http://asa.scitation.org/doi/10.1121/1.380594`.

Michelle J. Weirathmueller, William S. D. Wilcock, and Dax C. Soule. Source levels of fin whale 20 Hz pulses measured in the Northeast Pacific Ocean. *The Journal of the Acoustical Society of America*, 133 (2):741–749, 2013. ISSN 0001-4966. doi: 10.1121/1.4773277.

Michelle J. Weirathmueller, William S. D. Wilcock, and Rose S. Hilmo. Estimating range to a vocalizing fin whale using the timing and amplitude of multipath arrivals. *The Journal of the Acoustical Society of America*, 142(4):2101–2120, 2017. doi: 10.1121/1.5005494.

Sean M. Wiggins, Mark A. McDonald, Lisa M. Munger, Sue E. Moore, and John A. Hildebrand. Waveguide propagation allows range estimates for north pacific right whales in the bering sea. *Canadian Acoustics - Acoustique Canadienne*, 32(2):146–154, 2004. ISSN 07116659.

Oystein Wiig, Ian Gjertz, David Griffiths, and Christian Lydersen. Diving patterns of an Atlantic walrus Odobenus rosmarus rosmarus near Svalbard. *Polar Biology*, 13(1):71–72, 1993.

Howard Elliott Winn and Paul J. Perkins. Distribution and sounds of the minke whale: with a review of mysticete sounds. *Biological Systems*, 1976.

Guanghan Xu, Hui Liu, Lang Tong, and Thomas Kailath. A least-squares approach to blind channel identification. *Signal Processing, IEEE*, 43(12):2982–2993, 1995. URL `http://ieeexplore.ieee.org/xpls/abs%7B_%7Dall.jsp?arnumber=476442`.

Hanks H. Zeng and Lang Tong. Connections between the least-squares and the subspace approaches to blind channel estimation. *IEEE Transactions on Signal Processing*, 44(6):1593–1596, 1996. ISSN 1053587X. doi: 10.1109/78.506629.

Wen-Jun Zeng, Xue Jiang, Xi-Lin Li, and Xian-Da Zhang. Deconvolution of sparse underwater acoustic multipath channel with a large time-delay spread. *The Journal of the Acoustical Society of America*, 127 (2):909–919, 2010. doi: 10.1121/1.3278604. URL `http://link.aip.org/link/?JASMAN/127/909/1`.

Wen-Jun Zeng, Xue Jiang, and Hing Cheung So. Sparse-representation algorithms for blind estimation of acoustic-multipath channels. *The Journal of the Acoustical Society of America*, 133(4):2191–2197, 4 2013. ISSN 1520-8524. doi: 10.1121/1.4792247. URL `http://www.ncbi.nlm.nih.gov/pubmed/23556588`.

Daniel P. Zitterbart, Lars Kindermann, Elke Burkhardt, and Olaf Boebel. Automatic Round-the-Clock Detection of Whales for Mitigation from Underwater Noise Impacts. *PLoS ONE*, 8(8):2–7, 2013. ISSN 19326203. doi: 10.1371/journal.pone.0071217.