UC San Diego
UC San Diego Electronic Theses and Dissertations

Title

Passive acoustic localization and tracking using arrays and directional sensors

Permalink

https://escholarship.org/uc/item/7dx302j2

Author

Tenorio-Hallé, Ludovic

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Passive acoustic localization and tracking
using arrays and directional sensors

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Oceanography

by

Ludovic Tenorio-Hallé

Committee in charge:

      Aaron Thode, Chair
      Jay Barlow
      William Coles
      William Hodgkiss
      Peter Shearer

2021

The dissertation of Ludovic Tenorio-Hallé is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my adventurous parents, Miguel Angel Tenorio and Brigitte Hallé.

EPIGRAPH

Sometimes science is more art than science, Morty.
—Rick Sanchez C-137

Son of a b****, I'm sick of these dolphins.
—Captain Steve Zissou

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This dissertation marks the end of an important chapter in an odyssey that started many years ago. I hope the following paragraphs convey the gratitude I have for the many people who made this achievement possible.

First and foremost, I would like to thank my advisor, committee chair and mentor, Aaron Thode, for allowing me to embark on this doctoral journey. Thank you for believing in me, and for all the doors you opened for me along the way. For always taking the time to review and discuss my work, for your patience and your generosity. And for all the opportunities to conduct fieldwork and participate in conferences in places I had previously only dreamt of traveling to.

I would also like to acknowledge the rest of my doctoral committee – Bill Hodgkiss, Peter Shearer, Jay Barlow and Bill Coles – for their valuable revisions, comments and insights on my research.

My dissertation research would not have been possible without the support of many collaborators, whom I would like to recognize. Thank you to Bill Kuperman, Bill Hodgkiss, Chris Verlinden, Jit Sarkar and Jeff Tippmann from the Marine Physical Laboratory, for providing the dataset that was used in Chapter 2 of this dissertation. Regarding Chapter 4, I would like to thank Marc Lammers, from the Hawaiian Islands Humpback Whale National Marine Sanctuary, for reviewing my work and sharing his insights on humpback whale behavior, and to Katherine Kim and Alex Conrad from Greeneridge Sciences Inc. for providing and deploying the equipment used to collect the data.

Steven Swartz and Jorge Urbán, I sincerely thank you both for allowing me to be part of your research program. The time spent in the field in Laguna San Ignacio has not only provided me with the dataset used in Chapter 3 of this dissertation, but with countless experiences and

lessons that will forever inspire me, both in my professional and personal lives. I would also like to extend my deep appreciation to the entire community of Laguna San Ignacio for their continued support. Thank you to everyone working at Kuyimá and Baja Discovery for hosting us, and to all the pangeros for your help and companionship. Additional thanks to all the volunteers who assisted in the data collection: Alex Conrad, Ben Meissner, Héloïse Frouin-Mouy, Regina Lobo Barrera and Hunter Akins.

I would also like to thank Paul White from the University of Southampton, as well as Armando Jaramillo-Legorreta from CONANP in Mexico, who first introduced me to marine bioacoustics and inspired me to pursue research in this field.

While I am immensely grateful for the time I have spent in San Diego pursuing this PhD, this experience has come at the price of being far away from my friends and family in France, whom I miss and would like to acknowledge. To my parents Miguel and Brigitte, my brother Diego, and my grandparents Nicolas and Arlette, thank you for your unconditional support. Each one of you has influenced and inspired me in your own way, making this dissertation yours as much as it is mine. To one of my oldest friends, William Dassargues, thank you for visiting me in San Diego so many times; I'm grateful we've remained such good friends despite the distance and the years. I would also like to give a shoutout to my hometown Barfleur and the rest of the crew: Louis, Matthieu, Victor, Justine, Chispa, Darius, George le dauphin, etc.

Fortunately, I was able to make new friends in California who have kept me going in this journey. I would like to start by expressing my gratitude to my roommate and dear friend Andrew Wilde. These past few years would not have been the same without you. Thanks for all the good times, and thanks to your family: Natali, Danny, Katie, Andy, Georgia, Pete, etc. for always making me feel at home. Thank you, Sue and Dallas, my coolest San Diego friends, for sharing

your home when I first arrived, and for introducing me to Boneyards and its crew. To my uncle Gildas and my aunt Greta, thank you for your delicious Thanksgiving dinners and for teaching me how to brew beer. I would also like to give a very special shoutout to my cousin Vava, and all my other California friends: Elsa, Valeria, Alex, Mattia, Alessandro, Patrick Jantz, etc. And finally, I would like to thank my girlfriend, Diana, whose charm, warmth and support gave me the motivation I needed to finish my dissertation in this final and crucial year.

Some final acknowledgements to the Scripps community are in order. I would particularly like to thank some of my peers whom over the years have been friends, fellow students, colleagues, mentees, mentors, and surf buddies: Reuben Demirdjian, Hunter Akins, Travis Courtney, Travis Schramek, Daniel Yee, Andy Mullen, Bia Villas Bôas, Matt Pendergraft, Gabriel Castro, RJ, Mike Bianco, Yassir Eddebbar, Nick Pizzo, Will Chapman, Hugh Runyan, Chase James, Kerri Seger, Chris Verlinden, Jit Sarkar, Emma Reeves, Pichaya Lertvilai and Gino Passalacqua.

Chapter 2, in full, is a reprint of the material as it appears in the Journal of the Acoustical Society of America. Tenorio-Hallé, L., Thode, A. M., Sarkar, J., Verlinden, C., Tippmann, J., Hodgkiss, W. S., and Kuperman, W. A., (2017). A double-difference method for high-resolution acoustic tracking using a deep-water vertical array. *J. Acoust. Soc. Am.*, 142(6), 3474-3485. doi: https://doi.org/10.1121/1.5014050. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in the Journal of the Acoustical Society of America. Tenorio-Hallé, L., Thode, A. M., Swartz, S. L., and Urbán R., J. (2021). Using anisotropic and narrowband ambient noise for continuous measurements of relative clock drift between independent acoustic recorders. *J. Acoust. Soc. Am.*, 149(6), 4094-4105. doi:

VITA

2011   Master of Engineering, University of Southampton (United Kingdom)

2013   Master Degree, Institut Polytechnique de Grenoble (France)

2021   Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Tenorio-Hallé, L., Thode, A. M., Sarkar, J., Verlinden, C., Tippmann, J., Hodgkiss, W. S., and Kuperman, W. A., (2017). A double-difference method for high-resolution acoustic tracking using a deep-water vertical array. *J. Acoust. Soc. Am.*, 142(6), 3474-3485. doi: https://doi.org/10.1121/1.5014050.

Frouin-Mouy, H., **Tenorio-Hallé**, L., Thode, A., Swartz, S., and Urbán, J. (2020). Using two drones to simultaneously monitor visual and acoustic behaviour of gray whales (Eschrichtius robustus) in Baja California, Mexico. *Journal of Experimental Marine Biology and Ecology, 525*, 151321. doi:https://doi.org/10.1016/j.jembe.2020.151321.

Tenorio-Hallé, L., Thode, A. M., Swartz, S. L., and Urbán R., J. (2021). Using anisotropic and narrowband ambient noise for continuous measurements of relative clock drift between independent acoustic recorders. *J. Acoust. Soc. Am.*, 149(6), 4094-4105. doi: https://doi.org/10.1121/10.0004996.

ABSTRACT OF THE DISSERTATION

Passive acoustic localization and tracking
using arrays and directional sensors

by

Ludovic Tenorio-Hallé

Doctor of Philosophy in Oceanography

University of California San Diego, 2021

Aaron Thode, Chair

In bioacoustics, passive acoustic localization and tracking plays an important role in studying marine mammals and other organisms that produce underwater sounds. However, the implementation of such techniques faces many practical challenges, such as lack of environmental data for accurately modeling acoustic propagation, uncertainties in sensor position, time-synchronization of autonomous instruments, and logistical constraints due to large arrays. The three research chapters of this dissertation cumulatively address these hurdles.

Chapter 2 develops a reformulation of the "double-difference" method for long-range tracking of acoustic sources. Originally developed for high-resolution localization of earthquakes across a network of widely distributed sensor, the double-difference approach is here adapted to

exploit acoustic multipath on a vertical array, deployed in a deep-water waveguide. Results are shown to provide high-precision relative depth and range tracks of sources on the order of 50 km away, by compensating for biases caused by underdetermined array tilt and sound speed model. The method is demonstrated on both a towed acoustic source and a sperm whale (*Physeter macrocephalus*).

Chapter 3 presents a passive time-synchronization technique for independent autonomous acoustic recorders. This approach relies on the coherent ambient noise sources maintaining the same statistical angular distribution around the instruments. Under this assumption, the temporal evolution of the cross-correlation function between sensor pairs reveals their relative time drift. This method enables continuous measurements of clock offset, including small-scale non-linear fluctuations of the drift, otherwise unobservable with standard time-synchronization techniques. Data from a field study in San Ignacio Lagoon, Mexico, is used to demonstrate this technique which is here applied to low frequency pulses, most likely originating from croaker fish (*Sciaenidae* family).

Chapter 4 uses acoustic vector sensor data to track multiple sources simultaneously. The method is demonstrated on singing humpback whales (*Megaptera novaeangliae*) off western Maui. Here, the directional capabilities of vector sensors are exploited to identify and match azimuthal tracks from multiple sources between sensors, yielding localized whale tracks in terms of latitude and longitude over time. This approach shows potential for further applications such as tracking boats and analyzing the directional properties of ambient noise field.

Chapter 1

Introduction

Passive acoustic localization (PAL) uses the sound emitted by an acoustic source to infer information about its spatial location. Early work on this topic began during World War I, followed by substantial advances in locating and tracking enemy submarines throughout World War II and the Cold War (Holler, 2014). While navy applications remain relevant today, the ubiquitous prevalence of sound in the marine environment has made PAL a useful tool in a variety of oceanography-related topics.

Marine mammals are known to produce underwater vocalizations for a variety of purposes. These calls provide a unique opportunity for applying PAL techniques to study various aspects of the behavior and ecology of marine mammals. For example, precise localization and tracking of individual animals can yield information about their swimming kinematics and its relation to their acoustic behavior (Thode *et al.*, 2000; Zimmer *et al.*, 2003; Soule and Wilcock, 2013; Thode *et al.*, 2016; Guazzo *et al.*, 2017; Thode *et al.*, 2017; Henderson *et al.*, 2018; Varga *et al.*, 2018; Wiggins and Hildebrand, 2020; Hendricks *et al.*, 2021). PAL and subsequent tracking may also provide information on habitat use (Frankel *et al.*, 1995; Henderson *et al.*, 2018; Hendricks *et al.*, 2021), or the response of various species to anthropogenic stressors or other environmental changes (Hastie *et al.*, 2014; Blackwell *et al.*, 2015). Furthermore, PAL of calls enables estimation of their source levels and thus the detection range (Cato, 1998; Sirovic *et al.*, 2007; Mathias *et al.*, 2013; Thode *et al.*, 2016; Bouffaut *et al.*, 2021), which are required parameters for conducting passive acoustic density and abundance surveys (Marques *et al.*, 2013). PAL techniques also have the potential to be used for real-time conservation management by mitigating exposure to

underwater noise or ship strikes (Compton *et al.*, 2008; Dolman *et al.*, 2009; Van Parijs *et al.*, 2009; Klinck *et al.*, 2012; Baumgartner *et al.*, 2013; Kowarski *et al.*, 2020).

Bio-logging technologies (tags) have enabled observation of fine-scale movements of marine mammals and address some of the research topics mentioned above (Schmidt *et al.*, 2010; Goldbogen *et al.*, 2014; Stimpert *et al.*, 2020). However, in addition to being intrusive in comparison to PAL, tagging marine mammals remains logistically expensive and yields small sample sizes. Thus, PAL complements bio-logging in the sense that while it has lower-resolution localization capabilities, it allows localizing large numbers of animals, providing a more holistic understanding of the behavior and ecology of marine mammal species.

Beyond marine mammals, interest is growing in other organisms that also produce underwater sound, such as many species of fish and certain invertebrates (Popper *et al.*, 2001; Slabbekoorn *et al.*, 2010; Vermeij *et al.*, 2010; Hawkins and Popper, 2017; Ladich, 2019; Butler *et al.*, 2021). PAL can also be used to study these species as well as their ecosystems by tracking individual animals (Mouy *et al.*, 2018), characterizing the spatiotemporal distribution of sounds and the directional properties of the ambient noise field (Thode *et al.*, 2021). In addition to biological signals, PAL may also be used for marine surveillance by tracking illegal fishing vessels and other unreported offshore activities (Sorensen *et al.*, 2010; Kline *et al.*, 2020).

To infer information on the location of an acoustic source, PAL generally requires some knowledge – or assumptions – about the position of the receiver(s) and the propagation environment between the source and the receiver(s). This relationship between the source, the receiver(s), and the environment results in many PAL principles being applicable to analogous oceanographic topics. For example, in a well-characterized environment the principle of acoustic reciprocity implies the position of a receiver can be determined using sources whose position is

known. These acoustic element localization techniques have practical applications in ocean engineering for tracking the position of underwater sensors and vehicles, which cannot use Global Positioning System (GPS) like land-based and airborne systems (Van Uffelen *et al.*, 2016). Similarly, many underwater inversion problems are also related to PAL, allowing environmental features to be inferred from sources and receivers whose position is known. This concept forms the basis of ocean acoustic tomography (Munk *et al.*, 2009) and geoacoustic inversion (Collins *et al.*, 1992; Lindsay and Chapman, 1993; Gerstoft, 1994; Bonnel and Chapman, 2011).

Over the past few decades, advancements in computer power, acoustic propagation models, and data processing and storage techniques have enabled PAL to be applied to new problems. However, implementing PAL often still faces practical hurdles:

1. Lack of data and measurements for accurately modeling propagation environments (e.g., sound speed profile, bathymetry, seafloor composition).

2. Uncertainties in the position of the receivers (e.g., autonomous underwater vehicles, deployment location uncertainty, array tilt and shift over time).

3. Internal clock offsets and drift between sensors with independent data acquisition systems.

4. Logistical constraints due to the cumbersome nature of many PAL systems.

This dissertation focuses on how linear arrays (multiple sensors) and directional sensors (acoustic vector sensors) can alleviate some of these challenges, with an emphasis on applications for studying marine mammals and other biological sources. Chapter 2 addresses points 1 and 2, while Chapters 3 and 4 both address points 3 and 4. In this introductory chapter, theoretical concepts and previous literature relevant to the rest of the dissertation are reviewed, followed by an overview of the three research chapters. Section 1.1 reviews TDOA localization, a standard approach for many marine mammal PAL applications. Sections 1.2 and 1.3 review topics related

to Chapter 2, Section 1.4 reviews noise interferometry, a topic fundamental to Chapter 3, and Section 1.5 reviews acoustic vector sensors, the technology that lies at the heart of Chapter 4.

1.1 TDOA localization

Time-difference-of-arrival (TDOA) localization is a standard approach for applying PAL to marine mammals. This basic approach assumes the signal travels in a straight line from the source to the receiver in a medium of constant sound speed. In a two-dimensional plane, the locus of points that produces a given TDOA between two sensors is a hyperbola. By measuring the TDOA between all sensor pairs, the intersection of the resulting hyperbolas yields the location of the source (Watkins and Schevill, 1972; Schau and Robinson, 1987). Some work has examined the influence of spatial variation in sound speed on TDOA localization (Spiesberger and Fristrup, 1990). The accuracy and precision of the localization depends on the bandwidth of the signal in question, its received signal-to-noise ratio (SNR), the validity of the constant sound speed assumption, and the relative locations of the source and the sensors. When multiple animals are calling simultaneously, or when significant multipath or other propagation effects degrades the coherent structure of a signal, it can become difficult to identify the same call on different sensors.

Localization of marine mammals over large areas, in longitude and latitude, is most often performed using this method (Sirovic *et al.*, 2007; Simard *et al.*, 2008; Baggenstoss, 2011; Helble *et al.*, 2015; Hendricks *et al.*, 2021). In addition to the sensors needing to be time-synchronized, the main requirement for localizing within a two-dimensional plane is that the source should be detected on at least three spatially distinct sensors. In principle, it is possible to extend this technique for localizing sources in three-dimensions by adding at least a fourth sensor. However, this approach only works at short ranges (Mouy *et al.*, 2018), as the vertical separation of the

4

sensors required for long-range three-dimensional localization is not compatible with the depth of the ocean.

1.2 Beamforming

If sensors are clustered closely together so that their spacing is less than half the acoustic wavelength, the TDOA (or phase difference) of a signal between sensors can be used to estimate its angle of arrival, relative to the array, under the assumption that the signal is propagating in the form of a plane wave. This approach, known as "beamforming", is usually applied to multi-element line arrays as it also improves the SNR of the received signal, thus increasing the detection range of the source along with features estimated from the signal. If the background noise is uncorrelated between sensors, the SNR is improved by $10 log N$ decibels (dB), where N is the number of sensors. The angular resolution of the beamforming estimate is related to the aperture of the array, with a wider (longer) array providing a higher angular resolution. While beamforming can also be performed in the time-domain, frequency-domain beamforming allows using adaptive processing techniques may improve the resolution of the estimate against environmental uncertainty (Cox *et al.*, 1987; Debever and Kuperman, 2007; Jensen *et al.*, 2011).

1.3 Multipath localization using vertical line arrays

Acoustic multipath occurs when reflection and/or refraction mechanisms yield different propagation paths between a source and receiver. At short ranges and high frequencies, a direct arrival can be followed by time-separated reflections from the surface and the bottom. Over longer ranges in shallow water, low-frequency multipath reflections from ocean boundaries can be

mathematically expressed in terms of dispersive normal mode propagation. At longer ranges in deep water, multipath can arise from refraction due to the depth-dependent sound speed profile.

The time delay between acoustic multipath arrivals provides additional information that may allow localization, even from a single sensor (Thode *et al.*, 2002; Tiemann *et al.*, 2006; Mouy *et al.*, 2012; Bonnel *et al.*, 2014). Beamforming on a vertical array also provides the vertical elevation angle of each arrival relative to the horizontal, providing further localization information. Using the appropriate acoustic propagation model along with knowledge of the array depth and the environment, acoustic multipath measured on a vertical array can be used to localize the source (Thode *et al.*, 2000; Thode, 2004; Mathias *et al.*, 2013). This approach is the basis of matched-field processing (MFP), in which measured data from a vertical array is compared to a modeled acoustic field given a series of candidate source positions to find the maximum-likelihood solution (Fizell and Wales, 1985; Baggeroer *et al.*, 1993; Newhall *et al.*, 2012). When using range-independent propagation models, as it is most often the case, these methods usually allow finding the two-dimensional position of the source in cylindrical coordinates (i.e., depth, and range from the array). If azimuthal information is available, the source may be localized in three-dimensions (Tiemann *et al.*, 2006).

1.4 Noise interferometry

In wave physics, the Green's function describes how a signal is modified by the environment as it propagates from point A to point B. Specifically, it represents the signal that would be received at point B if a unit amplitude impulse of infinitely short duration is produced at point A. Knowledge of the Green's function allows deriving the received signal structure from any arbitrary source signal.

Both theoretical and experimental studies have shown that in the presence of a diffuse ambient noise field, the time-domain Green's function can be extracted from the time-averaged cross-correlation function between two sensors (Lobkis and Weaver, 2001; Weaver and Lobkis, 2001; Campillo and Paul, 2003; Malcolm *et al.*, 2004; Brooks and Gerstoft, 2007; Gouedard *et al.*, 2008). The field of study that exploits this property is often referred to as "noise interferometry". In underwater acoustics, noise interferometry has been applied to ambient noise for passive tomography and geoacoustic inversion purposes (Ren and Hermand, 2013; Godin, 2018), synchronization of clocks on autonomous underwater instruments and self-localizing acoustic sensors (Sabra *et al.*, 2005c). Chapter 3 extends the results of Sabra *et al.* (2005c) to cases where the ambient noise field violates the conditions needed for standard noise interferometry, but still contains enough information to time-synchronize two sensors.

1.5 Vector sensors and source triangulation

Most underwater acoustic studies measure only pressure, as this quantity is straightforward to measure underwater with a pressure sensor (hydrophone). However, in addition to fluctuations in pressure, the propagation of an acoustic wave also generates a local motion of the particles comprising the medium. This vector quantity can be formulated in terms of displacement, velocity, or acceleration, and provides information on an acoustic wave's direction of travel. Acoustic vector sensors are instruments that have been developed to measure both pressure and particle motion at a single point (Greene *et al.*, 2004; Martin *et al.*, 2016; Raghukumar *et al.*, 2020). This ability allows them to infer directional properties of the acoustic field such as the dominant direction of incoming sound (D'Spain *et al.*, 1991; D'Spain *et al.*, 2006; Thode *et al.*, 2019). While these sensors have been used for navy purposes since the 1960s (Holler, 2014), their use in bioacoustics has only

become recognized over the past 20 years (Greene *et al.*, 2004; McDonald, 2004; Nedelec *et al.*, 2016).

One main advantage of vector sensors regarding PAL is that they do not require precise time-synchronization, as beamforming or TDOA localization techniques do. The location of an acoustic source can be determined with vector sensors using triangulation from a minimum of two estimated bearings. Marine mammals have been localized in two-dimensions from this approach using vector sensors (Greene *et al.*, 2004; Thode *et al.*, 2012). Triangulation localization can also be applied to TDOA-estimated bearings estimated from nested hydrophone arrays to obtain three-dimensional tracks (Gassmann *et al.*, 2015).

1.6 Dissertation overview

This dissertation is comprised of three research chapters. In Chapter 2, a "double-difference" localization approach is adapted from seismology for long-range tracking of acoustic sources using a single vertical array in deep water. Chapter 2 draws inspiration from the concept of noise interferometry to time-synchronize independent sensors using an ambient noise field whose properties violate the azimuthal symmetry requirements of standard approaches. In Chapter 4, vector sensor data is used for tracking multiple singing humpback whales in two-dimensions off the western coast of Maui. The novel contribution of this chapter lies in the processing of the vector sensor directional data. Cumulatively, these three chapters provide additional avenues for solving the hurdles to implementing PAL that opened this introduction.

Chapter 2

A double-difference method for high-resolution acoustic tracking using a deep-water vertical array

Ray-tracing is typically used to estimate the depth and range of an acoustic source in refractive deep-water environments by exploiting multipath information on a vertical array. However, mismatched array inclination and uncertain environmental features can produce imprecise trajectories when ray-tracing sequences of individual acoustic events. "Double-difference" methods have previously been developed to determine fine-scale relative locations of earthquakes along a fault (Waldhauser and Ellsworth, 2000). This technique translates differences in travel times between nearby seismic events, recorded at multiple widely separated stations, into precise relative displacements. Here, this method for acoustic multipath measurements on a single vertical array of hydrophones is reformulated. Changes over time in both the elevation angles and the relative arrival times of the multipath are converted into relative changes in source position. This approach is tested on data recorded on a 128-element vertical array deployed in 4 km deep water. The trajectory of a controlled towed acoustic source was accurately reproduced to within a few meters at nearly 50km range. The positional errors of the double- difference approach for both the towed source and an opportunistically detected sperm whale are an order of magnitude lower than those produced from ray-tracing individual events.

2.1 Introduction

Passive acoustic monitoring offers a non-invasive approach for localizing and tracking acoustic sources such as marine mammals, and has numerous academic, industrial, and naval applications. A standard approach for localizing impulsive signals is to measure the difference in detected arrival times of a signal across multiple hydrophones, assuming a spatially homogenous sound speed and conducting hyperbolic localization (Watkins and Schevill, 1972). However, the separation between the sensors generally needs to be at least 20% of a source's range for this technique to work. Thus, in order to locate a source over any significant region, multiple hydrophone stations need to be deployed across at least several kilometers range to create the required aperture.

More recently, methods for long-range tracking of marine mammals in deep water using a short aperture array of hydrophones have been developed (Tiemann *et al.*, 2006; Mathias *et al.*, 2013). The term "short" refers to the aperture of the array relative to the water depth, not relative to an acoustic wavelength, and the term "deep water" refers to water depths that greatly exceed the acoustic wavelength of the lowest frequency component in a signal.

These tracking techniques are mostly based on ray theory, exploiting the multipath arrival information [elevation angle and relative time of arrival (RTOA)] present in a signal's refracted and reflected ray arrivals across the array aperture. By modeling the back-propagation, or "ray-tracing," detected ray arrivals into the waveguide, the source's location can be determined by measuring where the rays converge (Fig. 2.1). This method has been defined as either ray-tracing or matched-field processing (MFP; Fizell and Wales, 1985), the latter term covering situations where RTOA information has been incorporated into the ray-propagation model. Sperm whales have been the focus of several of these tracking studies because of the nature of their vocalizations,

which can be characterized as a series of loud and impulsive broadband clicks (Goold and Jones, 1995), which are convenient for multipath detection (Thode *et al.*, 2002; Zimmer *et al.*, 2003; Thode, 2004).
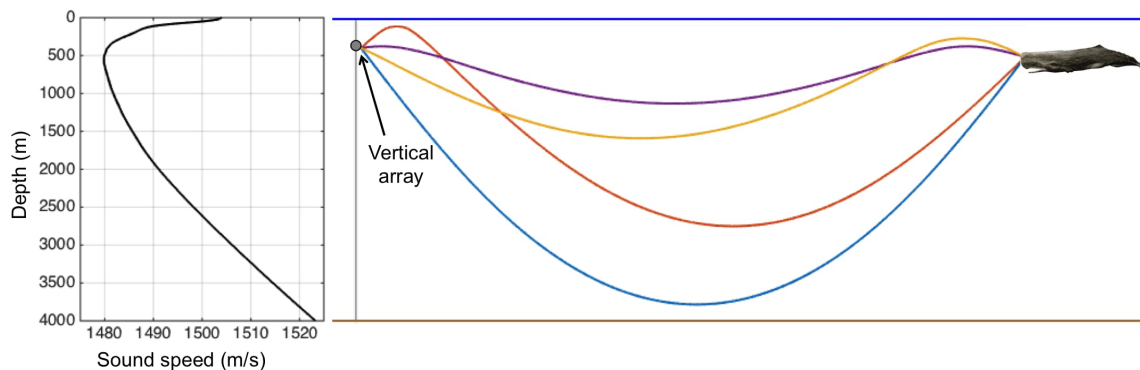
Figure 2.1: Illustration of 2D ray-tracing using a vertical array. Rays detected at the array are back-propagated into the waveguide, whose sound speed profile is known (left of the image), in order to determine the depth and range of the vocalizing sperm whale (right of the image).

While ray-tracing methods allow localizing at long ranges relative to the array aperture, they often yield significant uncertainties in the position of the source due to several factors. Because of the lack of extensive field measurements, environmental propagation models are typically assumed to be range independent, even though the underlying environment varies with range. The resulting mismatch in the modeled sound speed profile can therefore be an important source of error when applying ray-tracing, especially at longer ranges. Another potential source of error is the array inclination, or tilt, which affects the measured elevation angles of rays. Because of these errors, tracking a moving source by individually ray-tracing a series of acoustic events often yields a trajectory with large error estimates.

Various adaptive filtering and state-state techniques like Kalman filtering can be adapted for smoothing acoustic trajectories, essentially by incorporating information from previous localizations into the next estimate. However, this approach requires making assumptions about

the behavior of the system, such as speed and heading, for example (Evensen, 2009). Here, we are interested in techniques that rely entirely on measured data and not on interpolative or smoothing approaches.

Methods for high-resolution relative localization of earthquakes are common in seismology whenever seismic events in close spatial proximity are detected at multiple sensors (Poupinet *et al.*, 1984; Ito, 1985; Fremont and Malone, 1987; Shearer, 1997). These methods are based on measuring the differences in travel times between sets of events to determine the relative location of these events. A popular and efficient formulation of this approach is the "double-difference" method by Waldhauser and Ellsworth (2000), which will be subsequently labeled as "WE." The technique minimizes the residuals between measured and modeled travel times for clusters of nearby earthquakes recorded across common stations. This approach assumes the propagation paths of sets of nearby earthquakes are similar enough that they share the same systematic errors in the modeled travel times caused by environmental perturbations and/or instrument timing error. The double-difference method has been adapted for tracking fin whales in the northeast Pacific using a network of seafloor seismic stations (Wilcock, 2012), but to our knowledge has not been previously applied to multipath tracking from a single array deployment.

This paper presents how double-difference methods can be adapted to a deep water waveguide to track a moving acoustic source at long ranges, using multipath on a single vertical hydrophone array, instead of separate direct arrivals on a seismic network of widely spaced recorders. This approach, which usually requires relatively high signal-to-noise ratio (SNR) signals, is shown to yield much higher precision than standard ray-tracing results, and can reduce systematic errors arising from environmental mismatch and array tilt uncertainties. The method is tested on a towed acoustic source whose position is independently measured throughout its

deployment. Results from applying this technique to opportunistically measured sperm whales are also presented and discussed.

2.2 Theory

2.2.1 Nomenclature

Consider a two-dimensional waveguide (a vertical range/depth slice of the ocean) in which multiple acoustic events from a moving source are received on a vertical array of hydrophones. Figure 2.2 schematically illustrates two acoustic ray paths a and b reaching the vertical array, each propagating from two time-separated events i and j. In theory, the events need not be generated by the same source, but must be sufficiently spatially and temporally close to each other such that the difference in propagation environments between two events is nearly homogenous, and that at least two ray paths from each event traverse similar propagation paths to the receiver. If T_α^i represents the travel time along a particular ray path a from a given acoustic event i to the array, then the RTOA between two ray paths α and β is defined here as

$$\Delta T_\alpha^i = T_\alpha^i - T_\beta^i$$

(2.1)

and the difference between the RTOAs of the two events i and j becomes

$$\Delta T_{\alpha\beta}^{ij} = \left(T_\alpha^i - T_\beta^i\right) - \left(T_\alpha^j - T_\beta^j\right)$$

(2.2)

Note that Eq. 2.2 effectively removes the effects of range-dependent sound speed fluctuations between the array and the two events, because the time perturbations generated by these fluctuations affect both events and are thus subtracted out in Eq. 2.2. Let the "elevation angle" of

a ray across the vertical array be defined as the arrival angle measured relative to the horizontal with positive angles pointing toward the seafloor and negative angles pointing toward the surface. The elevation angle of ray path a from an acoustic event i is then noted θ_α^i. Following the same logic as for the ray travel times, Eq. 2.2 can be rewritten in terms of the relative differences in elevation angles (RDEAs) between rays α and β and events i and j,

$$\Delta\theta_{\alpha\beta}^{ij} = \left(\theta_\alpha^i - \theta_\beta^i\right) - \left(\theta_\alpha^j - \theta_\beta^j\right)$$

(2.3)

Note that Eq. 2.3 removes any measurement bias arising from incorrect estimates of array inclination, assuming no changes in array inclination occur between events i and j.
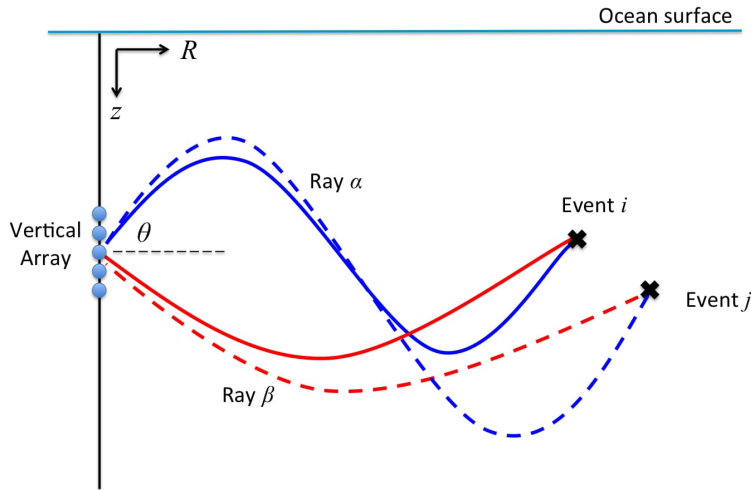
Figure 2.2: Schematic diagram of two nearby acoustic events i and j, each producing two acoustic rays α (blue) and β (red). A given ray path follows a slightly different trajectory from one event to the other and therefore arrives at the vertical array with slightly different elevation angles and RTOA for each event. However both experience the same range-dependent fluctuations in sound speed between the events and the vertical array.

2.2.2 Double-difference equations for a vertical array

The double-difference method minimizes residuals between differences in measured and modeled travel times for clusters of nearby events recorded on the same sets of sensors (WE). The

technique assumes that the quantities in Eqs. 2.2 and 2.3 remove or reduce systematic errors in modeling or recording. Differences in relative travel times between event pairs are then translated into positional changes, permitting high-resolution relative localizations of the events. The original double-difference equations in WE can be rewritten in terms of the RTOA (Eq. 2.2) and RDEA (Eq. 2.3) instead of the absolute travel times. For a pair of events i and j, whose ray paths α and β are detected on a vertical array of sensors, the resulting "double difference" equations are as follows:

$$\frac{\partial \Delta T_{\alpha\beta}^i}{\partial R} \Delta R^i + \frac{\partial \Delta T_{\alpha\beta}^i}{\partial z} \Delta z^i - \frac{\partial \Delta T_{\alpha\beta}^j}{\partial R} \Delta R^j - \frac{\partial \Delta T_{\alpha\beta}^j}{\partial z} \Delta z^j = \left(\Delta T_{\alpha\beta}^{ij}\right)^{meas} - \left(\Delta T_{\alpha\beta}^{ij}\right)^{mod}$$

(2.4)

$$\frac{\partial \Delta \theta_{\alpha\beta}^i}{\partial R} \Delta R^i + \frac{\partial \Delta \theta_{\alpha\beta}^i}{\partial z} \Delta z^i - \frac{\partial \Delta \theta_{\alpha\beta}^j}{\partial R} \Delta R^j - \frac{\partial \Delta \theta_{\alpha\beta}^j}{\partial z} \Delta z^j = \left(\Delta \theta_{\alpha\beta}^{ij}\right)^{meas} - \left(\Delta \theta_{\alpha\beta}^{ij}\right)^{mod}$$

(2.5)

where $(\ )^{meas}$ and $(\ )^{mod}$ denote the measured and modeled RTOAs and RDEAs, respectively, in Eqs. 2.2 and 2.3, and $\left(\Delta R^i, \Delta z^i\right)$ and $\left(\Delta R^j, \Delta z^j\right)$ represent the range and depth adjustments required for events i and j in order for the model to better fit the data. For N events with a total of M measurements (both RTOA and RDEA), Eqs. 2.4 and 2.5 can be compactly expressed in matrix form as

$$\mathbf{Gm} = \mathbf{d}$$

(2.6)

where $\mathbf{G}$ is a $M \times 2N$ matrix of partial derivatives, $\mathbf{m}$ is a $2N$ vector of positional adjustments, and $\mathbf{d}$ is the set of M residuals on the right-hand sides of Eqs. 2.4 and 2.5. Note that Eqs. 2.2-2.5 differ slightly from those presented in WE. The equations presented here use α and β to represent acoustic ray paths, instead of individually placed seismic stations. Furthermore, the residuals on

the right-hand sides of Eqs. 2.4 and 2.5, defined as "double-differences" in the original paper, have technically become "triple-differences" here, as they represent differences between measurements and models of differences between two events of RTOAs or RDEAs (which are themselves differences of arrival times or angles between two ray paths). A consequence of this reformulation is that the absolute times at which events i and j are generated, which were additional unknown variables in the original expressions, have been eliminated.

For a given event pair i and j, Eqs. 2.4 and 2.5 are first solved using estimates of the absolute event locations as an initial model, which can be provided by initial ray-tracing of one of the events. Figure 2.3a shows how the resulting components of **m** are then used to readjust event positions. This process is then re-iterated using the updated modeled positions of each event, until the residuals on the right-hand sides of Eqs. 2.4 and 2.5 are minimized. The final value of the residuals provides an estimate of the localization errors for the events, as will be discussed in Sec. 2.2.3.

The partial derivatives (or "ray derivatives") in these equations are calculated for the current modeled position of the source. From basic ray theory, it can be shown that

$$\frac{dT}{dR} = \frac{1}{c(z)}\cos{(\theta)} \equiv p$$

(2.7)

$$\frac{dT}{dz} = \frac{1}{c(z_s)}\sin{(\theta)} \equiv \eta(z_s)$$

(2.8)

where $c(z)$ is the depth-dependent sound speed and z_s denotes the depth of the source. p is known as the ray parameter, or horizontal slowness, which is constant along a given ray path in a range-

independent propagation environment, and $\eta(z)$ is known as the vertical slowness on the medium (Cornuelle, 1985; Frisk, 1994; Shearer, 1999).
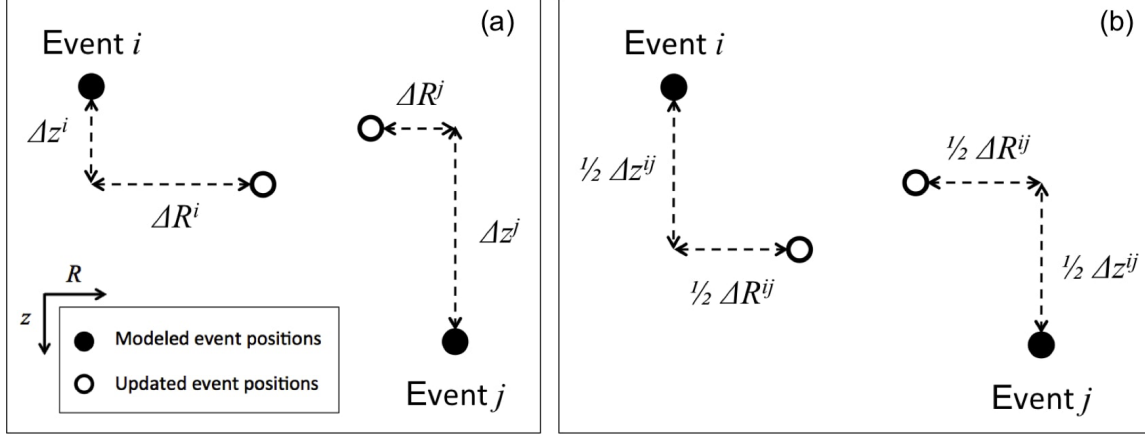
Figure 2.3: Schematic diagram demonstrating the repositioning of events using the double-difference: (a) using Eqs. 2.4 and 2.5, (b) assuming a homogenous environment using Eqs. 2.9 and 2.10.

2.2.3 Double-difference equations for a homogeneous environment

It is possible for an environment to be range dependent over long distances (e.g., on the order of tens of kilometers between an event cluster and receiver), but still be treated as effectively homogenous over the region connecting two consecutive events (e.g., tens of meters for consecutive events detected from a moving source). Further simplifications in Eqs. 2.4 and 2.5 can be then made by assuming that the sound speed gradients are effectively identical inside the spatial region containing events i and j. Equations 2.4 and 2.5 then reduce to

$$\frac{\partial \Delta T_{\alpha\beta}^{ij}}{\partial R} \Delta R^{ij} + \frac{\partial \Delta T_{\alpha\beta}^{ij}}{\partial z} \Delta z^{ij} = \left(\Delta T_{\alpha\beta}^{ij}\right)^{meas} - \left(\Delta T_{\alpha\beta}^{ij}\right)^{mod}$$

$$(2.9)$$

$$\frac{\partial \Delta \theta_{\alpha\beta}^{ij}}{\partial R} \Delta R^{ij} + \frac{\partial \Delta \theta_{\alpha\beta}^{ij}}{\partial z} \Delta z^{ij} = \left(\Delta \theta_{\alpha\beta}^{ij}\right)^{meas} - \left(\Delta \theta_{\alpha\beta}^{ij}\right)^{mod}$$

where the derivatives above are the average of the derivatives evaluated at the locations of events i and j. Note that Eqs. 2.10 and 2.11 now solve for ΔR^{ij} and Δz^{ij}, the adjustments in range and depth separation between events i and j (Fig. 2.3b), instead of ΔR^i, Δz^i, ΔR^j, and Δz^j, which adjust the range and depth of each event individually, as seen in Fig. 2.3a. This simplification reduces the number of variables in Eq. 2.6 from $2N$ to $2(N-1)$ and provides an option for Eqs. 2.9 and 2.10 to be solved independently from other event pairs. The movements of a source can therefore be estimated one event pair at a time, instead of having to solve for the relative position of each event in the entire track all at once, as would be the case for the full "inhomogeneous" problem.

For each pair of events there are two equations (Eqs. 2.9 and 2.10) and two unknowns (ΔR^{ij} and Δz^{ij}), so a minimum of two shared ray paths between the model and the measurements is required for every iteration. If this requirement is not met, the system becomes under-determined and the separation between the pair of events cannot be determined. This method therefore requires that pairs of events occur close enough together in space for them to share similar ray paths.

It is possible for the model and the measurements of an event pair to share more than two ray paths in common. In this case, the system becomes over-determined, which gives an option to solve the system using only Eq. 2.9, ignoring the RDEA elevation angle measurements. This approach is defined here as the "RTOA-only" double-difference, as opposed to the "full" double-difference approach that uses the RDEA measurements. Unless otherwise specified, subsequent use of the term double-difference method below refers to the full reformulated double-difference.

2.2.4 Error estimates for double-difference computations

Error estimates for double-difference computations can be computed using bootstrap or jackknife resampling of residual values, or singular-value decomposition (SVD; WE). The small numbers of events used in this paper makes the use of the SVD approach straightforward and practical (Press, 1992). The $M \times 2(N-1)$ matrix $\mathbf{G}$ defined in Eq. 2.6 has a SVD of $\mathbf{G} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^T$, where $\mathbf{\Lambda}$ is a diagonal matrix of singular values, $\mathbf{U}$ is an $M \times M$ matrix with orthogonal columns, and $\mathbf{V}$ is a $2(N-1) \times 2(N-1)$ matrix with orthogonal columns.

Mimicking the notation of WE, the error estimate σ_i of the ith component of the vector $\mathbf{m}$ in Eq. 2.6 becomes

$$\sigma_i = \sqrt{C_{ii} \cdot \sigma_d^2}$$

(2.11)

where C_{ii} is the ith diagonal element of the covariance matrix $\mathbf{C} = \mathbf{V} \mathbf{\Lambda}^{-2} \mathbf{V}^T$, and σ_d^2 is the variance of the final residuals (elements of $\mathbf{d}$ in Eq. 2.6)

$$\sigma_d^2 = \frac{1}{2N-3} \sum_{k=1}^{2(N-1)} (d_k - \bar{\mathbf{d}})$$

(2.12)

where $\bar{\mathbf{d}}$ is the mean of $\mathbf{d}$ (WE).

In the full case where both RTOA and RDEA measurements are made (and thus the components of d have different units), Eq. 2.12 is computed independently for the RTOA and RDEA residuals, and then each row of $\mathbf{G}$ is rendered unitless by dividing it by the appropriate value of $\sigma_{d,RTOA}$ or $\sigma_{d,RDEA}$. The SVD is performed, and Eq. 2.11 then reduces to $\sigma_i = \sqrt{C_{ii}}$.

2.3 Methods

2.3.1 Equipment

Acoustic data were collected on the mid-frequency noise array (MFNA), which was designed and built by the Marine Physical Laboratory (MPL) at the Scripps Institution of Oceanography (Fig. 2.4). This drifting vertical array consists of 128 elements with 0.1m spacing (7.5kHz half wavelength), creating an aperture of 12.7 m. Vertical inclinometers were attached on the top and bottom elements of the assembly to monitor the array tilt. Data from each channel were sampled at 25kHz and stored autonomously with 16-bit resolution. The system also had the capability of transmitting data via an 802.11 wireless connection close to a monitoring vessel.

An ITC-2040X acoustic source built by International Transducer Corporation (ITC, Santa Barbara, CA) was used as a sound source during the experiment. It was programmed to produce one second long, frequency-modulated (FM) linear sweeps from 3 to 9kHz at 140dB re 1 μPa at 1m, which were repeated every second. A depth sensor attached to the source recorded the depth of the tow to within 1-m resolution once every 30 s.

128 ELEMENT MF NOISE ARRAY
MFN1 DEPLOYMENT 2

COMM BUOY: (BUOY 1)
- 900 MHZ, 2.4GHZ RADIO MODEMS
- IRIDIUM RELAY
- GPS
- FLASHER, RADAR REFLECTOR
- IRIDIUM BEACON
- ALKALINE BATTERY PACKS

SURFACE
FLOAT

LIFT LOOP
(2m)

3-4 SUBMERGED

24m SURFACE
LINE

UMBILICAL (4TSP)

10 38# SPHERES
SPACED @ 2m

75# TOTAL

200# NOTCHED ELLIPSOID FLOAT

270# TOTAL

YALE GRIP

319m UMBILICAL EXTENSION

210# TOTAL

2 x 200# ELLIPSOID FLOATS
STACKED TOGETHER

~330m

600# TOTAL

TOP STRUCTURE WITH
RECORDING ELECTRONICS (RAVA7)
AND 4 LITHIUM BATTERY PACKS

BUNDLED ARRAY CABLES

REAL TIME TILT

ELEMENT 1

TILT 04

4m

REAL TIME DEPTH

12.7m APERTURE

ARRAY
-128 HYDROPHONES @ 0.1m SPACING (CENTER)
- 2 STRENGTH MEMBERS (L,R)
- LIFT LOOPS @ 2m SPACING

ELEMENT 128

TILT 05

BOTTOM STRUCTURE

ACOUSTIC RELEASE

ANCHOR
425# AIR
~360# WATER

DRAWING NOT TO SCALE

2/21/14

Figure 2.4: Diagram of the Mid-Frequency Noise Array (MFNA) deployment configuration.

2.3.2 Experiment

The MFNA was deployed at 330 m depth off the coast of San Diego, CA, in a region of flat bathymetry 4 km deep. The deployment occurred at 22:00 UTC on February 17, 2014 at 31.72° N and 122.12° W and was recovered at 14:00 UTC on February 19 at 31.7° N and 122.01° W, corresponding to a drift of ~10km. The inclinometer data on February 18 showed that the tilt at top of the array stayed fairly constant, just below 1 degree, whereas at the lower end of the array, the tilt showed slow fluctuations between 1 and 2 degrees throughout the day, suggesting a slight catenary.

Figure 2.5 shows the measured depth and range of the towed acoustic source with respect to the array, throughout the deploy- ment. The source was deployed at 32.12° N and 122.18° W, around 17:00 UTC on February 18, from R/V Melville. Between 17:00 and 19:00 UTC, it was lowered to almost 300 m depth, with pauses every 100 m. The source was then towed away from the array at a depth slightly less than 200 m, over about 7 km, from 20:00 to 22:00 UTC. After that, the source was lowered to almost 300 m depth and then brought back up to the surface, once again pausing every 100 m, until it was recovered shortly before 0:00 UTC on February 19, at 32.17° N and 122.2° W. The drift of the array during the entire deployment of the acoustic source was about 1.5km. The horizontal range between the source and the array was assumed to be the distance between the array and the vessel. Since the source is being towed ~200 m behind the vessel, this assumption would yield an absolute offset between –200 m and 200 m from the true position, depending on the vessel's heading. However, the inclination of the towed array cable remained steady throughout the tow, indicating that the relative range separations between towed array positions was the same as the relative range shift of the ship measured via the Global Positioning System (GPS).

In addition to the FM sweeps broadcasted from the towed source, sperm whale clicks were opportunistically recorded over the same time period in the data collected on the vertical array. Occasional humpback whale and dolphin signals were also encountered.
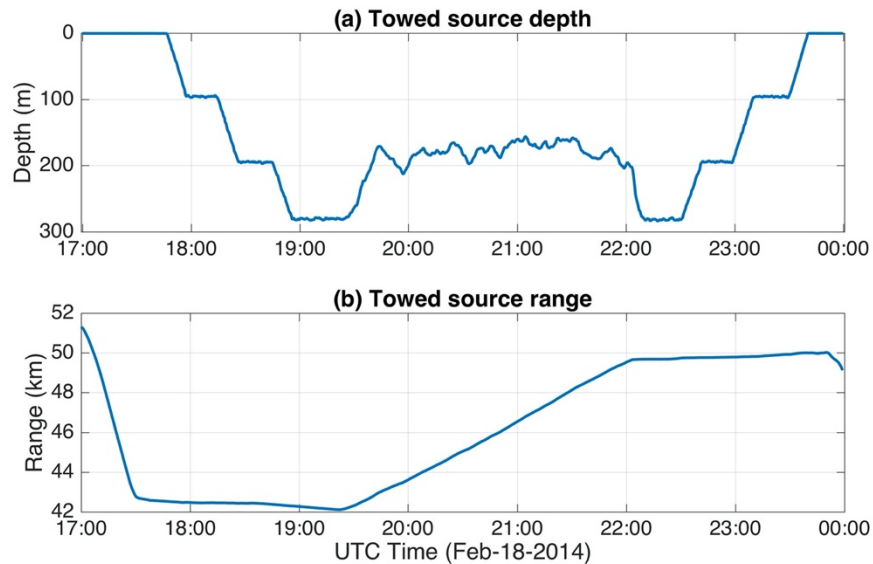
Figure 2.5: Depth (a) and range (b) trajectories of the towed acoustic source from the MFNA array during its deployment from 17:00 UTC on February 18, 2014 to 0:00 UTC on February 19, 2014.

2.3.3 Sound speed profile

Figure 2.6a shows the sound speed profile used in this analysis. Over the course of the experiment 15 CTD (conductivity temperature–depth) measurements were conducted to a depth of 2 km, from which an average sound speed profile was derived for the upper portion of the water column. These measured data are consistent with historical data for this region in February (Locarnini *et al.*, 2013) and shows a minimum sound speed close to 500 m depth. The lower portion of the profile down to the seafloor was then extrapolated based on historical models. A decaying exponential function was used to smoothly transition from the measured profile to the extrapolated linear function.
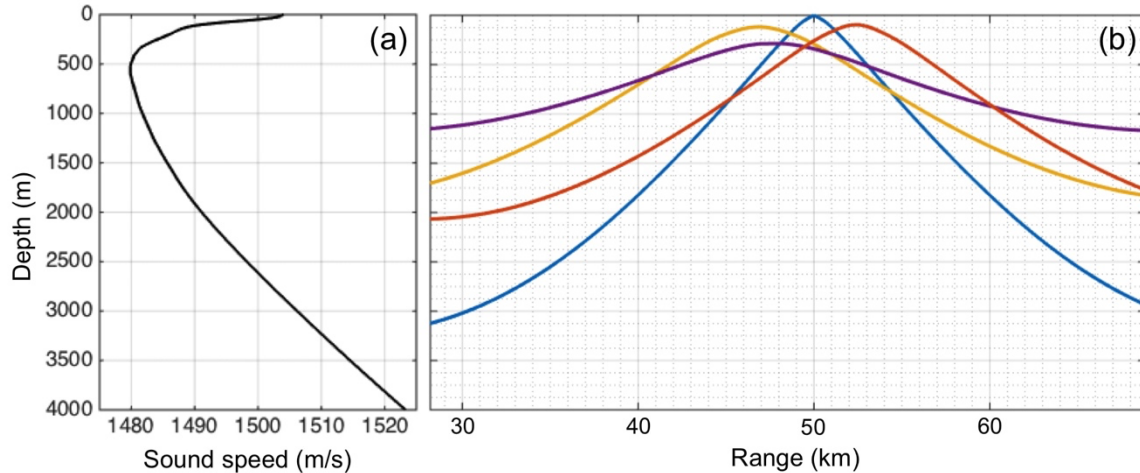
Figure 2.6: Example of propagation modeling. (a) Sound speed profile used in this analysis. The top 2000 m are based on averaged CTD measurements conducted during the experiment. The lower portion of the profile down to the seafloor at 4000 m was extrapolated. (b) Example of simple ray-trace propagation modeling of a towed source event. Appendix 2.8.2 details how the best-fit position is estimated from ray paths that do not converge at a single point.

2.3.4 Data processing: Measuring elevation angle and RTOA of rays

In order to identify and measure the RTOA and RDEA of individual ray arrivals, acoustic events (towed source signals or sperm whale clicks) are plotted as two-dimensional (2D) migration plots, which display arrival angle versus time. These plots are computed by applying time-delay and sum beam- forming to the data with an angular resolution of 0.1 degrees.

For the towed source, the FM nature of the signals it produces makes individual rays difficult to identify and RTOA measurements inaccurate if the migration plot is produced directly from the raw signal. Because the characteristics of these signals are known precisely, matched filtering was used to increase the SNR and yield more precise measurements (Skolnik, 1962). This process converts the FM signals from the source into single impulses, allowing RTOAs between rays to be measured more precisely. Figure 2.7 shows an example of a migration plot from a

matched filtered towed source signal. Note that for sperm whale clicks, it was not necessary to apply a matched filter as their signals are already broadband impulses.

The arrival times and angles for each ray path were measured by taking the local maxima of a manually selected region on the averaged migration plots.
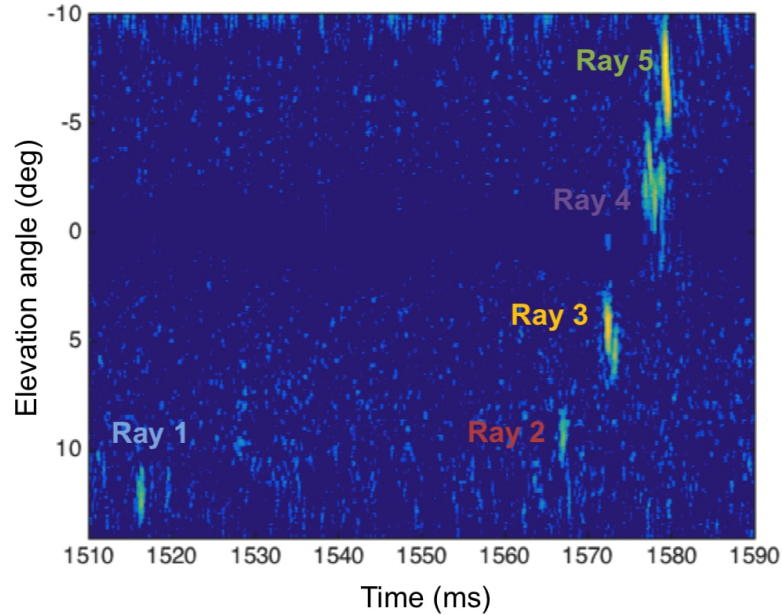
Figure 2.7: Migration plot of a matched filtered FM sweep signal from the towed source at 22:36 UTC on February 2014. Five main rays, arriving at distinct times and elevation angles, can be identified.

2.3.4 Implementing the reformulated double-difference algorithm for a vertical array

The modeled RTOA and RDEA differences in the right- hand sides of Eqs. 2.9 and 2.10 were computed using the ray-tracing program BELLHOP (Porter and Reiss, 1984; Porter, 1991). Eigenrays were computed using a fan of 3000 rays from –16 to 16 degrees, and a step size of 10 m, yielding an estimated "miss" criteria of ~10 m for logging an eigenray. Each modeled ray is then matched with a particular measured ray, based on the smallest difference in elevation angle

25

between the model and measurement. If the minimum difference exceeded 1.5°, data associated with that measured ray path were discarded from the iteration.

The ray derivatives in Eq. 2.9 were computed analytically in terms of the ray parameter and the vertical slowness, using Eqs. 2.7 and 2.8. The ray derivatives in Eq. 2.10, while also having analytical solutions (see Appendix 2.8.1), were instead calculated numerically from BELLHOP by modeling sources with small offsets in range and depth from a reference position and obtaining the angular shifts. Equations 2.9 and 2.10 are iterated until the value of $\mathbf{m}$ either converges or settles into an alternating cycle. In the latter case, the maximum value of $\mathbf{m}$ obtained during the cycle is used to compute the final residuals $\mathbf{d}$.

To generate a complete trajectory, the first set of events is assigned initial modeled positions (i.e., initial guess for the absolute position) based on either ray-tracing or the known position. For subsequent events in the track, the initial modeled positions are then based on the result of the previous double-difference computation, as discussed above. Positional errors for the double-difference estimates are computed as shown in Sec. 2.2.4.

Figure 2.6b shows a typical example where rays do not converge at a single point. The algorithm used to produce ray-tracing position estimates as well as their associated errors is presented in Appendix 2.8.2.

2.4 Results

The vertical array double-difference algorithm is first tested on the towed acoustic source by comparing the results to the measured trajectory of the source during two distinct phases of the source deployment: a horizontal source tow and a vertical haul during the source recovery. For each case, trajectories obtained from individually ray-tracing each event in the track are also

shown. A comparison between the full double-difference and the RTOA-only approach is also presented for the vertical haul trajectory. Finally, the full double-difference technique is applied to a 25-min sperm whale track.
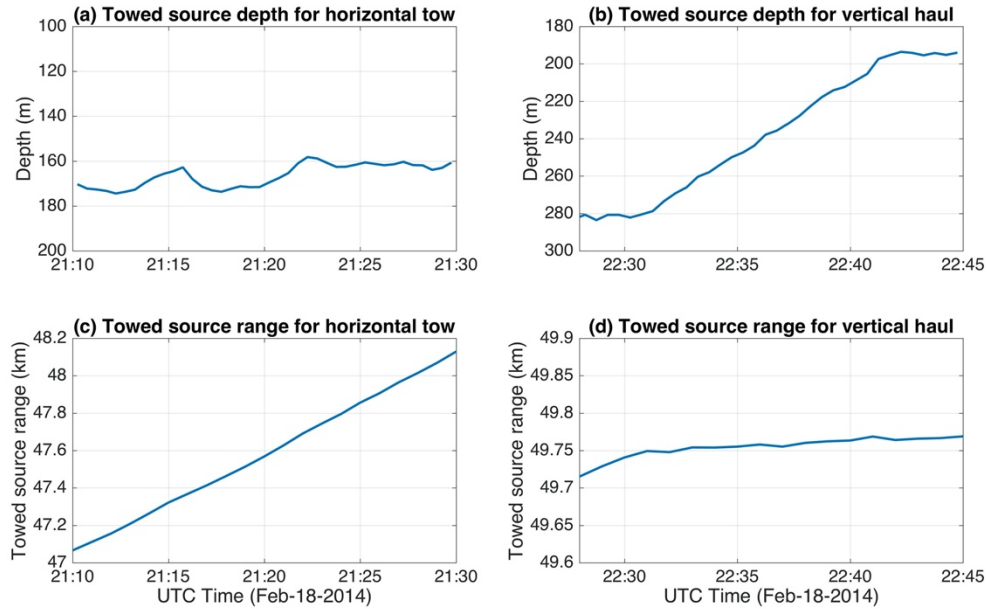
Figure 2.8: Depth and range trajectories of the acoustic source for the horizontal tow, from to 21:10 to 21:30 UTC (a)(c), and the vertical haul, from 22:30 to 22:45 UTC (b)(d), on February 18, 2014.

2.4.1 Towed acoustic source

2.4.1.1 Horizontal tow

Figures 2.8a and 2.8c display the depth and range, respectively, of the towed source over a 20 min period, when the source is being towed from 47 to 48 km range from the array, bouncing between depths of 160 and 180 m. Figures 2.9a and 2.9c plot RTOA and elevation angle measurements, respectively, from nine events, which are 2-4 min apart, where three rays were identified in each event. The angular measurements show that ray 1 increases by 1 degree over the 20 min period at a fairly constant rate, whereas rays 2 and 3 fluctuate over a 1-2 degree window

and do not show a clear trend. The RTOA of rays 2 and 3 both show a fairly steady increase of 8 and 4 ms, respectively, with respect to ray 1, chosen here as the reference ray.

Figure 2.10 shows the corresponding double-difference results, plotted both in terms of absolute (Figs. 2.10a and 2.10c) and relative (Figs. 2.10b and 2.10d) depth and range (with the latter being defined relative to the location of the first event). Also plotted are the measured positions of the source, as well as localization estimates based on ray-tracing individual events. Vertical bars show the uncertainties associated with each technique. The initial modeled positions for the first pair of events were set to 46.8km in range and 150m in depth. Figures 2.10b and 2.10d illustrate how the ray-tracing results diverge from the measured range of the source, while also yielding larger uncertainties for both range (6100–500 m) and depth (610–40 m), as seen in Figs. 2.10b and 2.10d. By contrast, the double-difference method accurately reconstructs the relative trajectory of the source within a few meters of the measured depth trajectory and less than 100 m for the range trajectory, with uncertainties approximately an order of magnitude smaller than for the ray-tracing.

2.4.1.2 Vertical haul

Over this 20 min period the vessel was stationary at around 49.75km range, while hauling the source back toward the surface, from 280 to 200 m depth. Figures 2.8b and 2.8d show the depth and range trajectories, respectively, of the source between 22:28 and 22:40 UTC. Over that same period, Figs. 2.9b and 2.9d show the RTOA and elevation angles, respectively, measured from seven events, spaced in 2 min intervals. Four to five rays were identified for each event. The elevation angles of all the rays remain fairly steady over this time period, but the RTOA of rays 2 to 5 (relative to ray 1) begin decreasing at a fairly constant rate around 22:32 UTC, which is when

the source begins its vertical ascent. The decrease of RTOA spans from about 5 ms to almost 20

ms depending on the ray.

Figures 2.10e-2.10h show the measured, ray-traced, and double-difference position

estimates for each event, in terms of both absolute (Figs. 2.10e and 2.10g) and relative (Figs. 2.10f

and 2.10h) depth and range. The initial modeled positions for the first pair of events were set to

49.75km in range and 270 m in depth. Figure 2.10f shows that both the double-difference and the

ray-tracing produce accurate depth trajectories, within a few meters of the measured position, but

Fig. 2.9h shows that ray-tracing is less accurate at reproducing the range trajectory, with biases on

the order of 200 m, versus the 100 m for the double-difference. Even more striking are the relative

uncertainties in position estimates between the two methods: ray-tracing produces uncertainties on

the order of tens of meters in depth and hundreds of meters in range (Fig. 2.10g), in comparison

with the uncertainty in depth and range (Fig. 2.10h) for the double difference results, which are an

order of magnitude smaller.
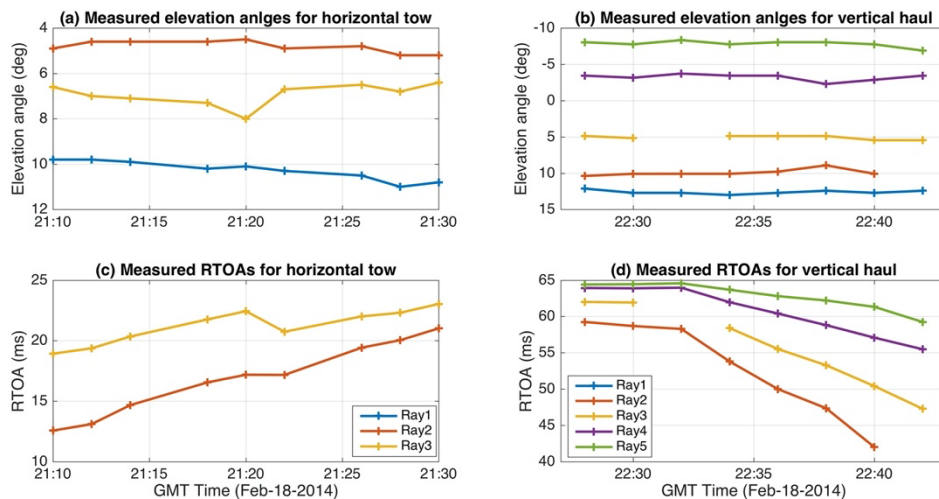
Figure 2.9: Elevation angle and RTOA measurements from the acoustic source signals for the horizontal tow, from to 21:10 to 21:30 UTC (a)(c), and the vertical haul, from 22:28 to 22:40 UTC (b)(d), on February 18, 2014.
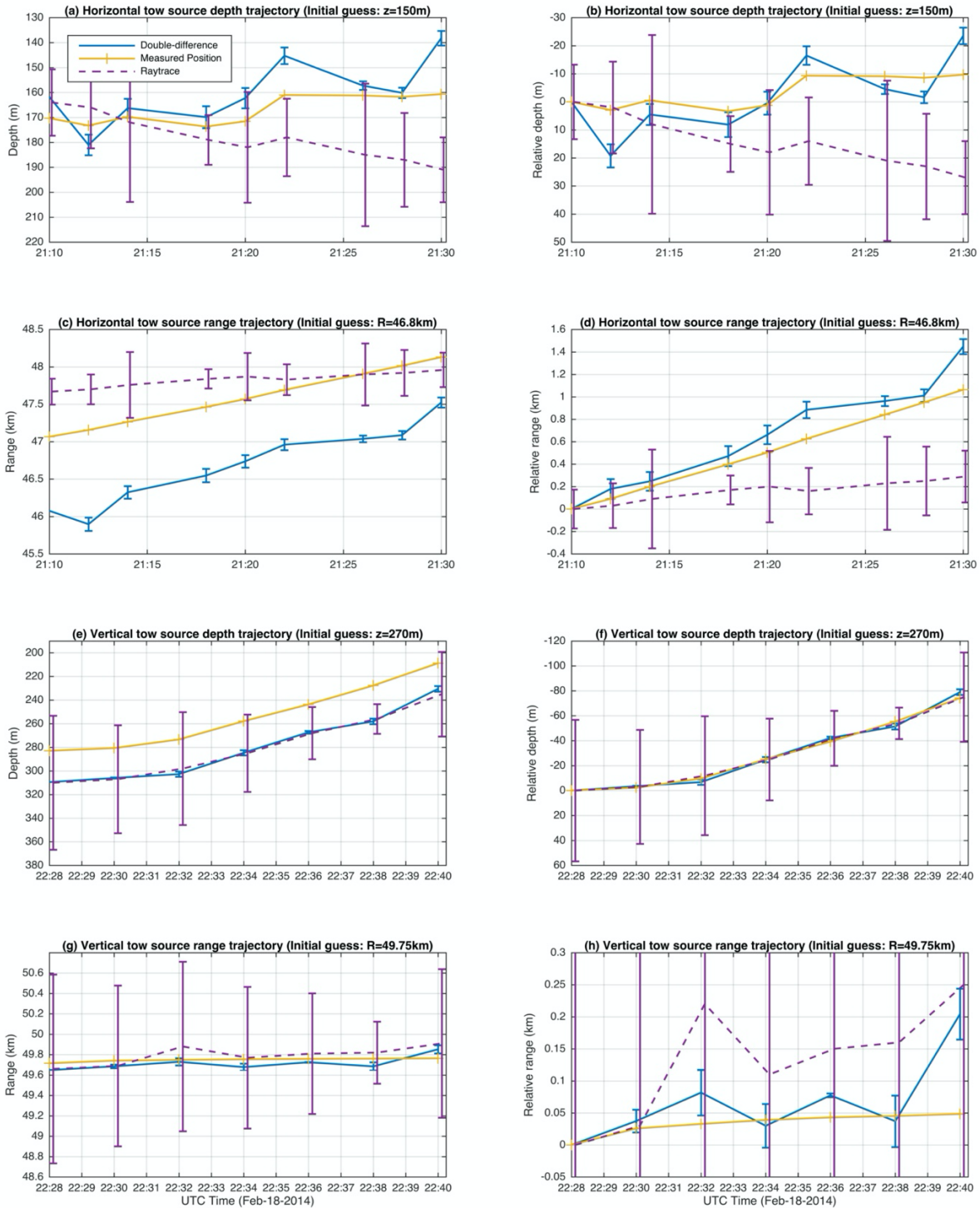
Figure 2.10: Depth and range trajectories of the towed acoustic source: double-difference results (solid blue line), ray-tracing results (dashed purple line) and measured position (dashed yellow line). The top four plots show the horizontal tow and the bottom four plots show the vertical haul. The left and right columns show the trajectories in term of the absolute and relative positions respectively. Note that the vertical scale of (h) is different than the vertical scale of (g).

Figure 2.11 compares the full double-difference results for the depth trajectory of the vertical haul with an RTOA-only double-difference computation. This particular scenario was selected for this comparison because it was one of the few situations where sufficient ray paths existed per event to provide a solution in terms of RTOA only. The results show that the RTOA-only double-difference performs very similarly to the full double-difference approach, but fails to converge for the last few events of the trajectory, yielding an incomplete track. The uncertainties of the RTOA-only approach are also greater than the full solution.
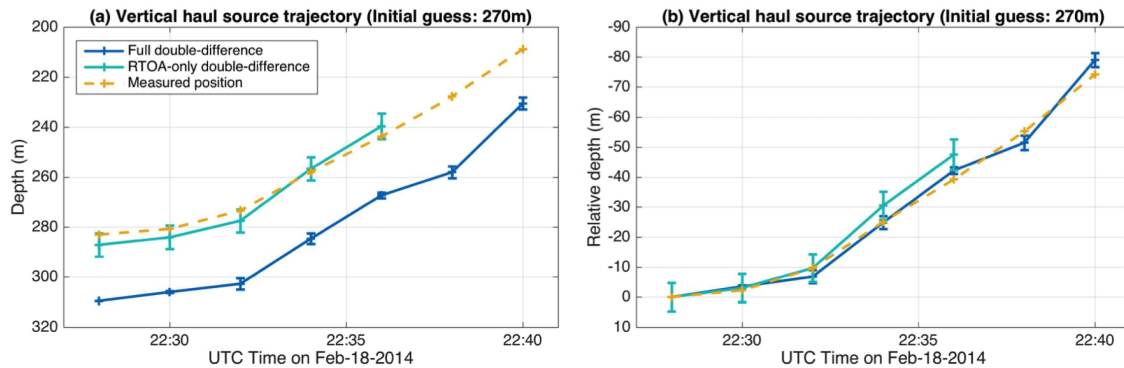
Figure 2.11: Absolute (a) and relative (b) depth trajectories of the towed acoustic source: "full" double-difference results (solid blue line), "RTOA-only" double-difference results (solid turquoise line) and measured depth (dashed yellow line).

2.4.2 Sperm whale track

RTOA and elevation angle measurements were taken from 15 sperm whale click events over a 25 min period at intervals of 1, 2, and 4 min. During its dive the whale's inter-click interval (average time between clicks in a series of clicks) stayed nearly constant at around 0.7s. Figure 2.12 indicates three consistent ray paths were identified throughout the track. The angular measurements show that ray 1 increases by about 1 degree at a relatively steady rate, whereas rays 2 and 3 remain fairly constant. RTOA measurements of these rays show a clear overall increase of about 15–20 ms (relative to ray 1) throughout the track.

The true position of the animal is unknown in this case, but results from applying the double-difference method can still be compared to ray-tracing estimates from the individual events in this track. The initial modeled positions for the first pair of events were set to 46 km in range and 350 m in depth, based on initial ray-tracing estimates.

The resulting trajectories (Fig. 2.13) show that the sperm whale remains at a fairly constant depth, between approximately 360 and 380 m, and travels from 45 to 47 km away from the vertical array. The double-difference and ray-tracing produce similar trajectories, but once again the double-difference approach yields relative uncertainties that are almost negligible (6–22 m in range and 1 m in depth) compared to ray-tracing uncertainties that are around 80–250 m in range and 5–20 m in depth.

Since the azimuth of the whale cannot be measured on the vertical array, the exact speed and heading of the whale cannot be calculated. However, a minimum and maximum possible speed can be estimated based on the geometry of the problem and assuming constant velocity. The minimum speed assumes zero heading, i.e., the animal is swimming directly away from the array, whereas the maximum speed assumes of heading of 90 degrees, with the whale swimming tangentially to the array from 45 to 47km. Taking into account the 90 m drift of the array during time period, the possible speed of the whale ranges from 2.4 to 17.9 knots.
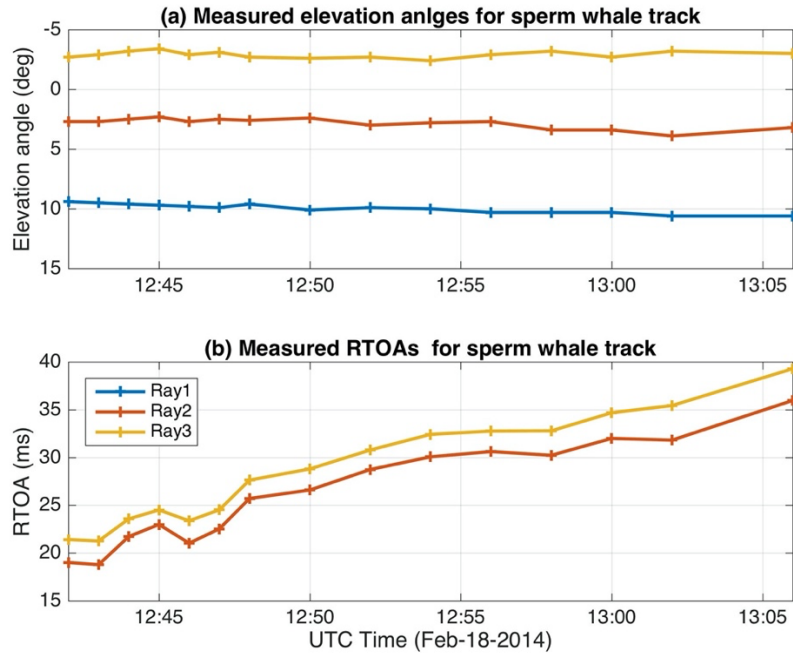
Figure 2.12: (a) Elevation angle and (b) RTOA measurements from the sperm whale track, between 12:42 and 13:06 UTC on February 18, 2014.
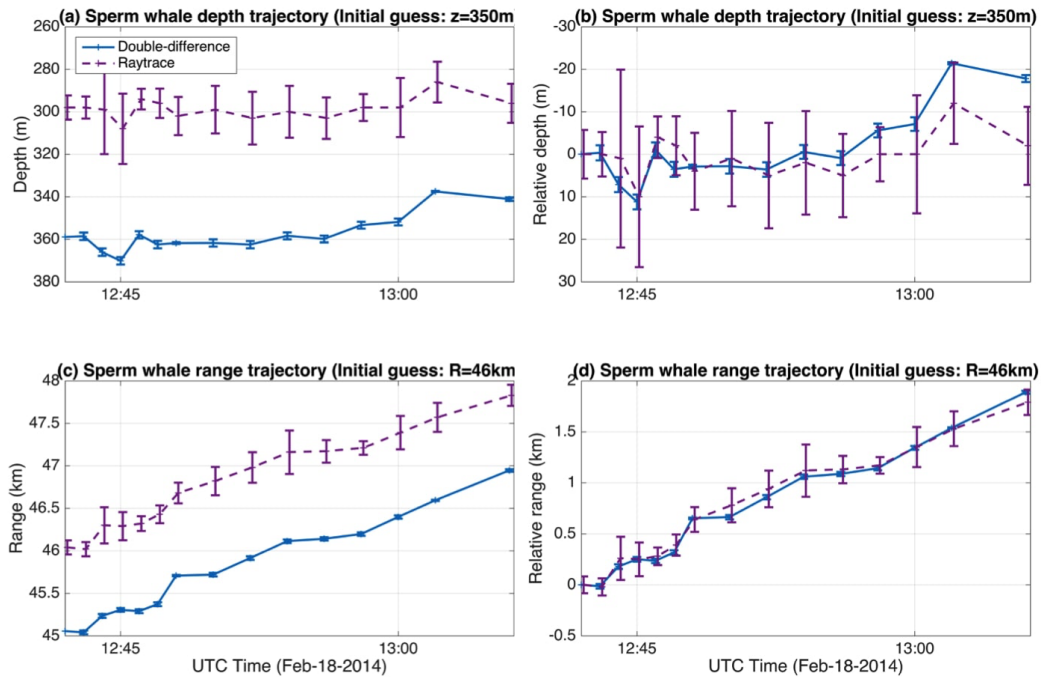
Figure 2.13: Sperm whale track between 12:42 and 13:06 UTC on February 18, 2014: double-difference results (solid blue line) and ray-tracing results (dashed purple line). The left and right columns show the trajectories in term of the absolute and relative positions respectively.

2.5 Discussion

Applying the reformulated double-difference algorithm to data gathered from a field experiment shows how depth and range trajectories of both a towed acoustic source and a sperm whale can be reconstructed. This approach yields trajectories that are sometimes slightly offset from the absolute position of the source, but accurately reproduces the positions of acoustic events relative to each other (relative trajectory), as seen throughout Fig. 2.10. These double-difference trajectories are also shown to be more precise than those obtained using standard ray-tracing, with localization uncertainties an order of magnitude smaller than the ray-tracing uncertainties. The improved performance of the double-difference versus ray-tracing was particularly obvious for the range estimates of the controlled acoustic source, as seen in Figs. 2.10d and 2.10h.

Both the full and RTOA-only double-difference methods were tested on the same data (see Fig. 2.11). The outcome of this analysis demonstrates that both approaches perform similarly, as long as there are enough matching rays between the model and the measurements for a given pair of events. This requires not only that the ray structure between two events be similar, but also that these rays be identified and measured in the data. If this requirement is not met, the double-difference equations become under-determined. This can be observed in Fig. 2.11, which shows an incomplete track for the RTOA-only double-difference that resulted from insufficient numbers of rays to permit a least-squares solution. Even though four to five rays were identified in this trajectory (see Figs. 2.9b and 2.9d), the results suggest only two of these rays were successfully matched with the model, which permitted solutions for the full double-difference but not for the RTOA-only approach. An advantage however, of the RTOA-only algorithm is that it does not need to repeatedly run BELLHOP to compute the angular derivatives in Eq. 2.11, making it less computationally expensive. The RTOA-only double-difference can therefore be a more efficient

approach as long as, for each iteration, there are at least three matching rays between the model and the measurements of a given event pair. This is sometimes a challenging requirement, so the full double-difference approach, which requires only two matching rays and also produces smaller errors, is generally considered more robust.

The importance of matching ray arrivals between the model and the data is also clear whenever testing the robustness of the double-difference method to uncertainties in array tilt and perturbations in the sound speed profile. As long as sufficient rays are matched, the double-difference appears to be fairly robust to uncertainties in array tilt and sound speed. However, small perturbations can sometimes cause the modeled ray structure to change substantially, reducing the number of matching rays between events and making the system under-determined. Another key factor that can have a major impact on the results is the initial modeled position of the source. Small changes (sometimes on the order of a few meters for the depth) in the initial modeled position can cause the ray structure to change enough for the double-difference to produce an incomplete track. For each track shown here, a series of trial initial event locations was tested to determine the positions that generated enough matching ray paths across the entire sequence of events.

These analyses, including the sensitivity to initial position, suggest that the depth-dependent sound speed profile variability is an important factor when considering the robustness of the approach. Indeed, the depths analyzed here are all below 400 m, a region where the sound speed changes rapidly with depth, as seen in Fig. 2.6. The ray structure can therefore vary rapidly with depth in this shallow region, which would explain why the double-difference results are sometimes sensitive to small changes in initial modeled position and other perturbations such as array tilt and sound speed. If true, this hypothesis predicts that double-difference approaches could

be even more robust for deeper sources in the water column, where the variability of the sound speed profile is smaller. This depth-dependent environmental variability might also explain the similarity in performance between the double-difference and the ray-tracing results for the vertical haul, as seen in Figs. 2.10b and 2.10f.

The double-difference was implemented here assuming a locally "homogenous" environment (Eqs. 2.9 and 2.10), which permitted the separations between any given pair of events to be solved separately from all other events in a trajectory. Attempts to implement the "inhomogeneous" version of the problem (Eqs. 2.4 and 2.5), where the relative positions of all events in a trajectory are computed simultaneously in one single matrix inversion (WE), were unsuccessful because it was challenging to discover initial modeled positions that yielded the required number of matching rays throughout the entire sequence of events. Had we made RTOA or RDEA measurements between all possible combinations of events, and not just events adjacent in time, we might have made more progress with the inhomogeneous formation. However, Fig. 2.10 shows that the homogenous formulation of the double-difference equations produced satisfying results, showing that shifts in source position on the order of a few meters could be measured accurately at ranges of nearly 50 km, even in a portion of the water column with a strong sound speed gradient.

Results also show that this method could track a sperm whale over 25 min, as seen in Fig. 2.13. The double-difference analysis reveals that the animal is holding a fairly steady depth within 10 m of its original position, while steadily increasing its range. Observations this precise would have been impossible with the high-uncertainty trajectory measurements resulting from standard ray-tracing. The estimated speed window of 2.4–17.9 knots spans known sperm whale swimming speeds found in the literature (Miller *et al.*, 2004; Aoki *et al.*, 2007).

2.6 Conclusion

The double-difference method, initially developed to localize earthquakes by WE and recently applied to fin whales (Wilcock, 2012), has been reformulated here for localizing sources using multipath arrival information (RDEA and RTOA) detected on an array of sensors. This reformulation has been applied for long-range tracking of moving acoustic sources on an array whose aperture is short relative to the water depth.

Results show that relative depth and range trajectories of a towed acoustic source, located about 50 km from the array, can be accurately recovered using this method. Comparing these results to a standard ray-tracing algorithm found that the double-difference method yielded more accurate range trajectories of moving sources, and produced range and depth error estimates at least an order of magnitude smaller than ray-tracing approaches, even those that incorporated least-square fits of RTOA data. This method was also used to track the movements of a sperm whale over a 25 min period, yielding information on an animal's swimming and diving behavior.

At least two common ray paths must be identified between the model and the measurements for pairs of events, which can be challenging in an environment in which the sound speed changes rapidly with depth, but becomes feasible as more events become available to process, permitting RTOA and RDEA measurements from more combinations of events.

2.7 Acknowledgements

deployment of the array and acoustic source. The MFNA experiment was sponsored by the Office of Naval Research. This work was also partially supported by Grant No. N000141410403 from the Office of Naval Research Marine Mammal Program.

Chapter 2, in full, is a reprint of the material as it appears in the Journal of the Acoustical Society of America. Tenorio-Hallé, L., Thode, A. M., Sarkar, J., Verlinden, C., Tippmann, J., Hodgkiss, W. S., and Kuperman, W. A., (2017). A double-difference method for high-resolution acoustic tracking using a deep-water vertical array. *J. Acoust. Soc. Am.*, 142(6), 3474-3485. doi: https://doi.org/10.1121/1.5014050. The dissertation author was the primary investigator and author of this paper.

2.8 Appendix

2.8.1 Ray derivatives for elevation angle

The analytical forms of the derivatives in Eqs. 2.5 and 2.10 are given by

$$\frac{d\theta}{dR} = \frac{1}{\eta(z_s)}\left(\frac{d^2\tau}{dp^2}\right)^{-1}$$

(2.13)

$$\frac{d\theta}{dR} = p\frac{1}{\eta(z_s)\eta(z_r)}\left(\frac{d^2\tau}{dp^2}\right)$$

(2.14)

where τ is known as the delay time, p is the ray parameters, g is the vertical slowness, and z_r and z_s are the receiver and source depths, respectively (Frisk, 1994; Shearer, 1999). In practice, the evaluation of these derivatives is vulnerable to numerical instabilities, so for practical applications

these derivatives were simply numerically evaluated by perturbing BELLHOP runs, as explained in Sec. 2.3.4.

2.8.2 Ray-tracing localization algorithm

Acoustic events were localized based on a least-squares approach, which was equivalent to a maximum-likelihood approach for this dataset. For each individual event, measured elevation angles were first propagated back into the waveguide from the array using BELLHOP. A rectangular region was then selected around the most likely convergence zone and the eigenrays were computed for each point in the grid (depth interval: 1 m, range interval: 10 m). For each grid point that produced enough rays to be matched with the measurements, the log-likelihood function was computed based on the following expression:

$$L = -\frac{1}{2\sigma_\theta^2}\sum_{i=1}^{N}\left(\theta_i^{meas} - \theta_i^{mod}\right)^2 - \frac{1}{2\sigma_T^2}\sum_{i=1}^{N}\left(T_i^{meas} - T_i^{mod}\right)^2$$

(2.15)

where θ_i and T_i are the elevation angle and travel time of the ith ray, T_0 is the travel time of the reference ray, σ_θ^2 and σ_T^2 are the estimated variances in the measured elevation angles and travel times, respectively, using the migration plots shown in Fig. 2.7. The depth and range of the event was then determined by locating the minimum of the log-likelihood over the search grid. The error was then computed by running the localization algorithm 50 times with random perturbations in both the elevation angle and RTOA measurements, following a normal distribution with variances σ_θ^2 and σ_T^2, respectively. The error bars for the ray-tracing displayed in Figs 2.10 and 2.13 are the standard deviations of the localization results from these 50 iterations.

Chapter 3

Using anisotropic and narrowband ambient noise for continuous measurements of relative clock drift between independent acoustic recorders

Relative clock drift between instruments can be an issue for coherent processing of acoustic signals, which requires data to be time-synchronized between channels. This work shows how cross-correlation of anisotropic narrowband ambient noise allows continuous estimation of the relative clock drift between independent acoustic recorders, under the assumption that the spatial distribution of the coherent noise sources is stationary. This method is applied to two pairs of commercial passive acoustic recorders deployed up to 14 m apart at 6 and 12 m depth respectively, over a period of 10 days. Occasional calibration signals show that this method allows time-synchronizing the instruments to within ± 1 ms. In addition to a large linear clock drift component on the order of tens of milliseconds per hour, the results reveal for these instruments, non-linear excursions of up to 50 ms that standard methods would not be able to measure, but are crucial for coherent processing. The noise field displays the highest coherence between 50 to 100 Hz, a bandwidth dominated by what are believed to be croaker fish, which are particularly vocal in the evenings. Both the passive and continuous nature of this method provides advantages over time-synchronization using active sources.

3.1 Introduction

Most modern data acquisition systems use crystal oscillator clocks, which are subject to drift against a more precise reference time (e.g. GPS). Unlike land-based systems, the internal clocks of underwater instruments cannot be corrected for drift via GPS. In this case, data can be time-synchronized by making clock measurements before and after an instrument is deployed, and assuming a linear drift. This method is commonly used for passive acoustic localization using multiple acoustic recorders spaced on the order of kilometers apart (Sirovic *et al.*, 2007; Simard and Roy, 2008). However, when conducting coherent processing (e.g. beamforming) on smaller arrays with receivers on the order of tens of meters apart, time-synchronization within a millisecond becomes necessary, and any potential non-linear clock drifts may not be negligible. If all channels are sampled by a central acquisition system, time-synchronization is not an issue since any potential drift is the same across all channels. In contrast, if array channels are sampled independently (i.e. separate instruments), relative clock drift between channels can be problematic. In this case, regular acoustic calibration signals, transmitted from a known location relative to the receivers, can be used to time-synchronize the data (Mathias *et al.*, 2013). However, this approach can be logistically constraining since it requires additional equipment. Furthermore, active acoustic sources can impact bioacoustic measurements and violate local underwater noise regulations.

Multiple studies have demonstrated, theoretically and experimentally, that the cross-correlation function between two receivers in a diffuse ambient noise field is related to the structure of the time-domain Green's function between the two receivers (Lobkis and Weaver, 2001; Weaver and Lobkis, 2001; Campillo and Paul, 2003; Malcolm *et al.*, 2004; Wapenaar, 2004; Weaver and Lobkis, 2005; Campillo, 2006; Gouedard *et al.*, 2008). This property has been used in both seismology and underwater acoustics for inferring clock errors between instruments (Sabra

et al., 2005a; Sabra *et al.*, 2005c; Thode *et al.*, 2006; Stehly *et al.*, 2007; Sens-Schonfelder, 2008; Hannemann *et al.*, 2014; Hable *et al.*, 2018), including non-linear clock drift between ocean-bottom mounted seismometers (Gouedard *et al.*, 2014). In order to obtain a high-quality Green's function estimate, these methods require the noise field to be coherent over a large bandwidth and be isotropic in the plane of the two receivers (Sabra *et al.*, 2005b).

Here, we experimentally demonstrate that ambient noise cross-correlations can be used to continuously track the clock drift of two acoustic receivers in a complex noise field that violates these ideal theoretical assumptions of isotropic broadband noise. Although the narrow coherent bandwidth and the directional characteristics of a local noise field might not allow estimating the Green's function, the relative clock drift between the acoustic recorders can still be tracked, assuming the spatial distribution of the ambient noise sources is sufficiently stationary. From a single independent measurement of the time offset between the clocks, the two receivers are time-synchronized throughout the deployment using the estimated clock drift. The data were collected on instrument pairs separated by up to 14m and at depths of 6 and 12 m in February 2018 in Laguna San Ignacio, a shallow-water lagoon on the Pacific coast of Baja California, Mexico.

3.2 Theory

The cross-correlation function between two time series $x_1(t)$ and $x_2(t)$ of duration T from sensors 1 and 2 respectively, is given by

$$C_{12}(\tau) = \int_0^T x_1(t)x_2(t-\tau)\, dt$$

(3.1)

where τ is the delay time.

As demonstrated in previous work, in a broadband noise field that is isotropic around the sensors, the time derivative of their cross-correlation function can be approximated as

$$\frac{dC_{12}(\tau)}{d\tau} \approx -\frac{\delta(L/c + \tau)}{L} + \frac{\delta(L/c - \tau)}{L}$$

(3.2)

where L and c are the separation and average sound speed respectively between sensors 1 and 2 (Sabra *et al.*, 2005a). The RHS terms of Eq. 3.2 result from both the positive and negative travel-times $\pm L/c$ between sensors 1 and 2, and correspond to the forward and backward (causal and acausal) time-domain Green's function. For a time-synchronized system this function is symmetric and yields two peaks at $\tau = \pm L/c$. If a clock offset is present, it can thus be measured from the asymmetry of the two peaks about $\tau = 0$ (Sabra *et al.*, 2005c; Stehly *et al.*, 2007; Gouedard *et al.*, 2014).

If the bandwidth of the correlated components of the spectrum narrows, and the distribution of sound sources becomes anisotropic around the receivers, the cross-correlation function will no longer have the dual-peak structure expected from Eq. 3.2. Nevertheless, in a noise field of constant bandwidth and directional properties that remain statistically stationary, any changes in the structure of the cross-correlation function over time should still reflect fluctuations in both the travel-time and the clock offset between the sensors. Thus, if the travel-time between the sensors remains constant, and as long as the exact clock offset can be measured at one point in time, the clock offset $\tau_{off}(t)$ can then be tracked backwards and forward in time throughout the deployment from any given persistent feature $p(t)$ (e.g. peak) within the cross-correlation structure as

$$\tau_{off}(t) = p(t) - p(t_0) + \tau_0$$

(3.3)

where τ_0 is the clock offset measured at reference time t_0.

3.3 Methods

3.3.1 Acoustic deployment

Acoustic data used in this study were collected in Laguna San Ignacio, a shallow coastal lagoon (25 m maximum depth) along the Pacific coast of Baja California, Mexico. The lagoon is approximately 30 km long with an average width of 5 km (Fig. 3.1). It is known primarily for being both a breeding and calving site for gray whales (*Eschrichtius robustus*) in the winter months of their annual migration. The area was designated a UNESCO World Heritage site in 1993 and ecotourism has since then developed into one of the main drivers of the local economy. In order to avoid any disturbance to the local marine environment, it is therefore preferable to avoid the use of active acoustic sources for underwater monitoring.

The lagoon's dynamic acoustic environment arises from a variety of biological and physical sources. In addition to gray whale and bottlenose dolphins (*Tursiops truncatus*), the bioacoustic component of the ambient noise is dominated by snapping shrimp and fish (Dahlheim, 1987). Small boat traffic (from both ecotourism and fishing), wind and tides also contribute to this complex and dynamic acoustic environment (Seger *et al.*, 2015).

Two linear acoustic arrays were deployed off Punta Piedra (26.789160°N, 113.243530°W), the narrowest point of Laguna San Ignacio (Fig. 3.1). Each array consisted of two anchors connected by a 28 m segment of polypropylene rope, upon which were attached two acoustic recorders and an acoustic transponder. The recorders were separated by 14 m, with the transponder located halfway in between (Fig. 3.2).

The acoustic instruments were SoundTrap 300 STD (Ocean Instruments, New Zealand) that recorded 16-bit samples at 96 kHz (the data were subsequently downsampled to 9.6 kHz for

this analysis). The internal clocks of these instruments are OV-7604-C7 Low Power Clock Oscillator 32.768 kHz (Micro Crystal Switzerland). The SoundTraps were color-coded as Red, Orange, Blue and Green; henceforth, the two arrays will be designated as the Red/Orange (RO) and Blue/Green (BG) arrays. The transponders, acoustic release units model AR-60-E (Sub Sea Sonics), provided two-way communication with a portable ARI-60 transponder box and transducer using high frequency pings, centered around 38.462 kHz. The transponder pings provided independent measurements of clock drift at a few fixed times during the deployment. The BG array was also equipped with a HOBOware pressure & temperature logger (Onset, Massachusetts) that sampled data at 5 min intervals.

Both arrays were deployed on February 13, 2018 off a small fishing boat by dropping one anchor, traveling in the direction of the desired orientation, and dropping the second anchor once the line was taut, while taking GPS points at both drops. Note that in practice, the line may not have been perfectly taut, and the spacing of the SoundTraps may have been less than 14 m. Immediately after the deployment, the boat circled each array several times, creating an acoustic signature to estimate the instruments' clock offset, separation and orientation on the seafloor, as described in Appendix A. The RO and BG arrays were deployed at approximately 6 and 12 m depth respectively and were separated by roughly 200 m from one another (Fig. 3.1). The arrays were recovered on February 22, 2018.
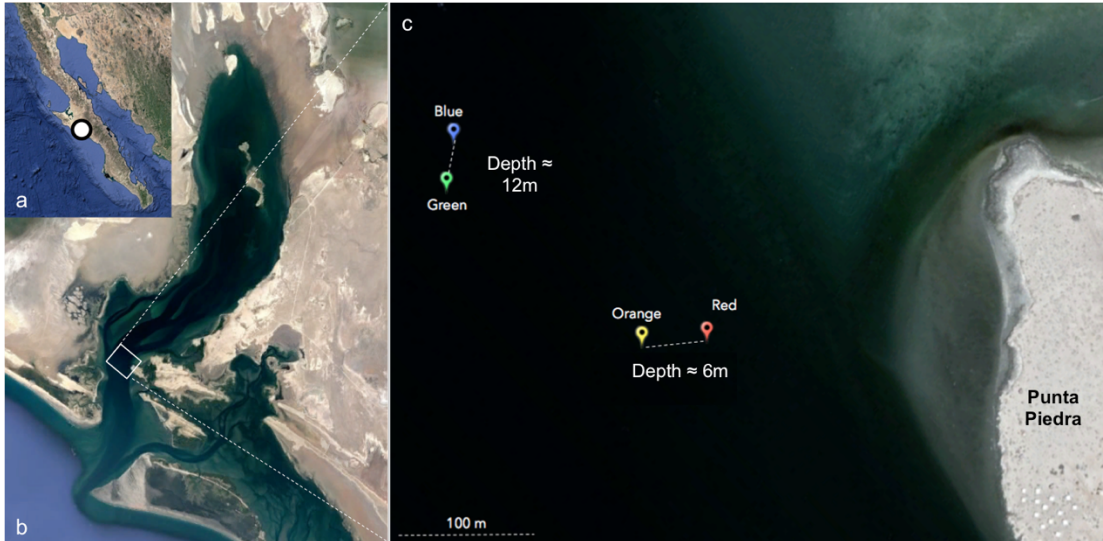
Figure 3.1: Deployment site. (a) Map of Baja California, Mexico, showing the location of Laguna San Ignacio. (b) Satellite image of Laguna San Ignacio. (c) Location of the deployment of the two arrays off Punta Piedra. The points on the map represent the GPS locations where the anchors from each array were dropped. Their color is representative of the closest SoundTrap.
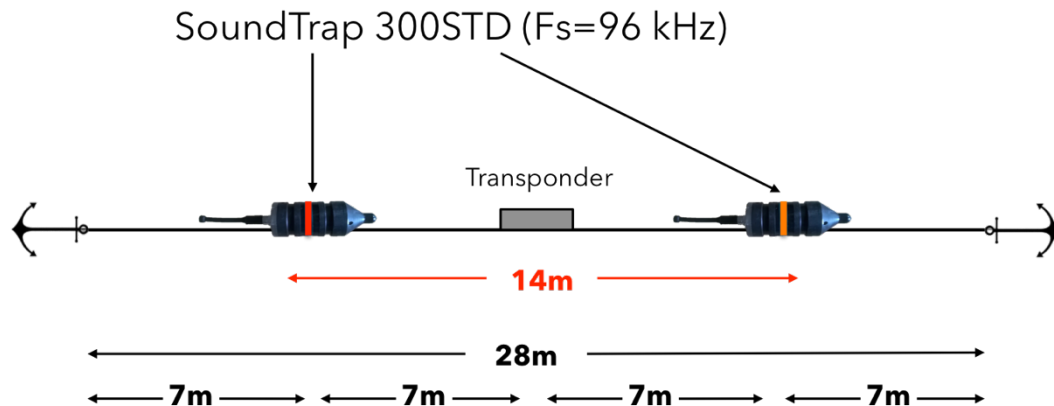
Figure 3.2: Schematic diagram of the acoustic arrays. Two SoundTrap 300STD sampling at 96 kHz (the data were subsequently downsampled to 9.6 kHz for this analysis) were attached to a 28 m polypropylene rope that had an anchor at each end. The acoustic recorders were separated by 14 m of rope, with a transponder placed halfway between the two, to provide occasional independent estimates of clock drift.

3.3.2 First-order time-synchronization (removing linear clock drift)

As shown in the results, the clock drift is dominated by a linear component. In order to measure and correct the small-scale non-linear clock drift, it is thus convenient to apply a first-order time-synchronization by estimating and removing the linear clock drift and initial offset τ_0. Let the total clock offset be modeled as

$$\tau_{off}(t) = \tau_L(t) + \tau_{NL}(t)$$

(3.4)

where $\tau_L(t)$ and $\tau_{NL}(t)$ are the linear and non-linear components of the clock offset. The linear correction term $\tau_L(t)$ can be defined as

$$\tau_L(t) = T_{drift} * (t - t_0) + \tau_0$$

(3.5)

where T_{drift} is the linear clock drift rate.

The exact clock offset τ_0 at the time of the deployment t_0 can be measured after the deployment of the instruments from boat engine noise, as shown in Appendix 3.8.1. For logistical reasons, the same measurement could not be made when recovering the arrays. However, an approximate clock offset (within the travel time between the instruments) τ_{end} at time t_{end} at the end of the deployment can still be estimated by visually aligning the time series of any transient acoustic event picked up at both sensor. The relative linear clock drift rate between sensor pairs can then be approximated as

$$T_{drift} = \frac{\tau_{end} - \tau_0}{t_{end} - t_0}$$

(3.6)

Let a "snapshot" be a short time series such that any clock drift is negligible over its duration. This first-order time-synchronization can then be applied to any two snapshots by time-adjusting the non-reference snapshot by $\tau_L(t)$ (Eq. 3.5). Removing this large linear component of relative clock drift between the instruments allows observation of the small (but significant in the context of coherent processing) residual non-linear drift across the entire deployment. This non-linear component of clock offset can be expressed by combining Eqs. 3.3 and 3.4 as

$$\tau_{NL}(t) = p(t) - p(t_0) - T_{drift}(t - t_0) = p_{NL}(t) - p_{NL}(t_0)$$

(3.7)

so that $\tau_{NL}(t_0) = 0 \approx \tau_{NL}(t_{\text{end}})$. Here, $p_{NL}(t)$ represents a persistent feature (e.g. peak) in the cross-correlation structure once the first-order time-synchronization has been applied.

This first-order time-synchronization also provides important practical advantages of enabling shorter cross-correlation snapshots when processing data in bulk. Indeed, if two snapshots are poorly time-synchronized, longer-duration snapshots are necessary in order to capture time-overlapped data and obtain a meaningful cross-correlation function. By allowing shorter snapshots, this first-order time-synchronization reduces computational cost and permits time averaging over more snapshots, which reinforces the coherent contribution of the noise spectrum.

3.3.3 Weighted cross-correlation of ambient noise field

The cross-spectral density of a signal is defined by

$$G_{12}(f) = E[X_1^*(f)X_2(f)]$$

(3.8)

where f is the natural frequency, X_1 and X_2 are the Fast Fourier Transforms (FFT) of the time series x_1 and x_2, and * denotes the complex conjugate. Here, the operator E denotes the expected

value. Assuming the signal is ergodic, an estimate $\hat{G}_{12}(f)$ of $G_{12}(f)$ can be obtained by time averaging $X_1^*(f)X_2(f)$ over N snapshots. The cross-correlation function (Eq. 3.1) can be computed using the inverse Fourier Transform of G_{12}

$$C_{ij}(\tau) = \int_{-\infty}^{\infty} G_{12}(f)e^{i2\pi f\tau}\, df$$

(3.9)

In order to emphasize frequency components that are highly correlated across the two sensors, the integrand in Eq. 3.9 can be weighted by

$$\psi(f) = \frac{|\hat{\gamma}_{12}(f)|^2}{|\hat{G}_{12}(f)|[1 - |\hat{\gamma}_{12}(f)|^2]}$$

(3.10)

where $\hat{\gamma}_{12}$ denotes the estimated coherence at a given time between the two channels as a function of frequency

$$\hat{\gamma}_{12}(f) = \frac{\hat{G}_{12}(f)}{\sqrt{\hat{G}_{11}(f)\hat{G}_{22}(f)}}$$

(3.11)

and $\hat{G}_{11}$ and $\hat{G}_{22}$ are the estimated auto-spectral densities of x_1 and x_2 respectively. The coherence, whose magnitude lies between -1 and 1, must be estimated by time averaging over multiple snapshots because the coherence of a single FFT snapshot between sensors is always 1. As seen in Eq. 10, the weight factor $\psi(f)$ ranges from 0 to infinity, based on the estimated coherence of that frequency component.

The time averaged coherence is then used to compute the final weighted cross-correlation for each individual snapshot

$$C_{12}^{\psi}(\tau) = \int_{-\infty}^{\infty} \psi(f)X_1^*(f)X_2(f)e^{i2\pi f\tau}\, df$$

This approach has been shown to be a maximum-likelihood estimator for determining time-delay between two signals (Knapp and Carter, 1976). Equation 3.12 allows the cross-correlation procedure to adaptively and automatically adjust its effective bandwidth to create the sharpest possible maxima in the time-domain cross-correlation estimate.

3.3.4 Stacked cross-correlation plots

In order to observe and measure the non-linear clock drift, it is convenient to stack the cross-correlation functions with respect to absolute time. First, the coherence is estimated by (1) downsampling the data from 96 kHz to 9.6 kHz, (2) dividing 10 minutes of data from each receiver into non-overlapping 1-sec snapshots ($N = 600$), (3) applying the first-order time-synchronization, (4) computing the 16384-point FFT of each snapshot (9600 samples, zero-padded), and (5) applying Eq. 3.11. The clock drift is assumed to be negligible over this 10-minute window after the first-order linear correction. $C_{12}^{\psi}(\tau)$ (Eq. 3.12) is then computed for every FFT snapshot pair over that same 10-minute interval, using the estimated coherence for that window.

One of the dynamic aspects of Laguna San Ignacio's acoustic environment is the presence of loud transient signals from boat noise and whales. Because the directional properties of these signals are non-stationary, the cross-correlation of snapshots containing such transients is not representative of the stationary component of the noise field, and thus a single strong transient within a 10-minute window can destroy the valid cross-correlation structure. In order to remove these outliers, the median of M adjacent estimates of $C_{12}^{\psi}(\tau)$, evaluated at each lag point τ is computed to obtain $\tilde{C}_{12}^{\psi}(\tau)$. This differs from most previous work, in which standard averaging is usually applied. Here, $M = 600$ with a 50% overlap is used, meaning $\tilde{C}_{12}^{\psi}(\tau)$ is estimated every 5

minutes over a 10-minute time window. The midpoint of each window is the absolute time associated with each measurement of $\tilde{C}_{12}^{\psi}(\tau)$. In principle, the number of snapshots used to estimate the coherence N and the median of the cross-correlation M can differ, but in this case they are both set to cover 10 minutes.

In order to emphasize the primary cross-correlation peak when plotting results over long periods of time, referred to here as "stacked" cross-correlation plots, each estimate of $\tilde{C}_{12}^{\psi}(\tau)$ is normalized by its maximum value such that

$$\tilde{C}_{12}^{\psi,n}(\tau) = \frac{\tilde{C}_{12}^{\psi}(\tau)}{\max(\tilde{C}_{12}^{\psi}(\tau))}$$

(3.13)

3.3.5 Measuring the non-linear clock drift

The precision required by the separation of the instrument pairs is such that any non-linear clock drift – on the order of milliseconds – is non-negligible and needs to be measured (Eq. 3.7). An estimate $\hat{p}_{NL}(t)$ of $p_{NL}(t)$ can be determined from the time delay associated with the maximum value of each $\tilde{C}_{12}^{\psi}(\tau)$. Additional outliers (e.g. cross-correlation side peaks) are manually removed by visually inspecting scatter plots of $\hat{p}_{NL}(t)$. Finally, in order to estimate a continuous function $p_{NL}(t)$ function, a cubic smoothing spline ('csaps' MATLAB function from curve fitting toolbox) is applied to the remaining points of $\hat{p}_{NL}(t)$. The spline algorithm's interpolation interval is 3 min and the smoothing parameter for the spline fit was $P = 0.9999$ ($P = 0$ being a least-squares straight line fit, and $P = 1$ is the natural cubic spline interpolant). The resulting function can then be used as $p_{NL}(t)$ to estimate the non-linear clock offset function $\tau_{NL}(t)$ (Eq. 3.7) and the total offset function $\tau_{off}(t)$ (Eq. 3.4).

3.3.6 Time-synchronization validation using transponder pings

It is important to confirm that the observed fluctuations of the cross-correlation structure over time are a result of clock drift and not changes in the spatial distribution of the ambient noise sources – or at least that any effect of potential redistributions in the noise field on the correlation structure are negligible in comparison to the effects of non-linear clock drift.

The transponders from each array (Fig. 3.2) provide an independent means of validating the clock drift measurements using ambient noise cross-correlation. Indeed, once the time-synchronization has been applied, the time delay between the transponder pings should remain constant throughout the deployment. If the arrays were deployed in a perfect line, the time delay between pings should be exactly zero since the transponder would be equidistant from each acoustic recorder. Even if an imperfect deployment had placed the transponder slightly closer to one of the recorders, a successful time-synchronization should yield constant time delays between pings throughout the experiment.

The transponders from each array were activated on four distinct occasions during the experiment. Whenever a transponder was pinged, it would transmit back a series of 5-10 pings at 38.462 kHz that could be detected by the acoustic recorders of the corresponding array. For each individual transponder ping, the time delay was measured manually from the signal envelopes (Hilbert Transform) of the time-synchronized time series. The median time delay was then computed for each series of pings, as well as the interquartile range (IQR) between the 25th and 75th percentiles.

3.4 Results

Figure 3 displays examples of normalized stacked cross-correlations (Eq. 3.13) and coherence (Eq. 3.11) for the RO array over 24 hours following the reference time t_0. Figures 3.3a and 3.3b show the stacked cross-correlations before and after applying the first-order time-synchronization. The cross-correlation structure in Fig. 3.3a shifts over time on the order of tens of milliseconds per hour, according to the total relative clock drift $\tau_{off}(t)$. Once the linear component of clock drift has been removed (Fig. 3.3b), the peaks' timing only shifts by about 10 ms over this time window: this residual shift is related to the non-linear clock drift $\tau_{NL}(t)$.

The cross-correlation structure shows two distinct regimes. During nighttime, between approximately 19:00 and 5:00 the next day, the cross-correlations can be described as a series of evenly spaced peaks whose amplitudes decay symmetrically about a main peak. In the daytime, the same main peak can still be observed most of the time, but the side lobes are less regular over time, giving the cross-correlations a less predictable and noisier structure. These two distinct regimes can be related to the structure of the stacked coherence (Fig. 3.3d). The 50-100 Hz band shows consistent evidence of coherence throughout the entire 24-hour time window. However, the nighttime coherence becomes much higher and extends up to 400 Hz, better defining the cross-correlation peaks. This same 24-hour cycle persists in both the cross-correlation and coherence structures throughout the entire deployment for both arrays. A one-minute spectrogram of the transient impulsive signals that increase the coherence in the 50-400 Hz band during nighttime is shown in Fig. 3.4. Croaker fish (*Sciaenidae* family) are believed to produce these pulses, as further discussed in Sec. 3.5.2.

In this dataset, taking the median of cross-correlation snapshots $C_{12}^{\psi}(\tau)$ rather than their average is shown to improve the sharpness of the cross-correlation main peak during the daytime,

when the coherence is lower. On the other hand, little difference is noticeable during nighttime when the coherence is high. To illustrate the difference, Fig. 3.3c shows an example of averaging the snapshots $C_{12}^{\psi}(\tau)$ in the stacked cross-correlations.

Table 3.1 lists the clock offset τ_0 measurements and the resulting linear relative clock drift rate T_{drift} estimates between the two instruments, for each array. The GB array is shown to drift linearly by 46.22 ms/h, which is about three times more than the drift of the RO array, which drifts by 14.54 ms/h.
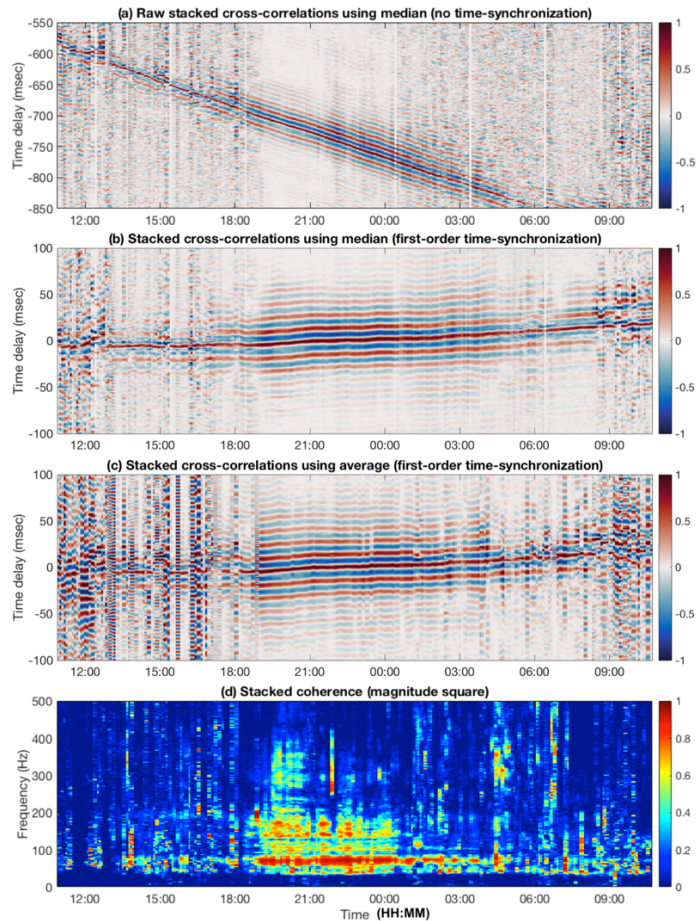
Figure 3.3: Stacked cross-correlations and coherence for the RO array over 24 hours following the reference time $(t_0 = 02/13\ 10{:}50)$ with (a) no time-synchronization, (b) first-order time-synchronization (removing linear component of clock drift and initial offset). Panel (c) is the same as (b) but averaging the snapshots $C_{12}^{\psi}(\tau)$ instead of taking the median. Panel (d) displays the squared magnitude of the coherence.

54
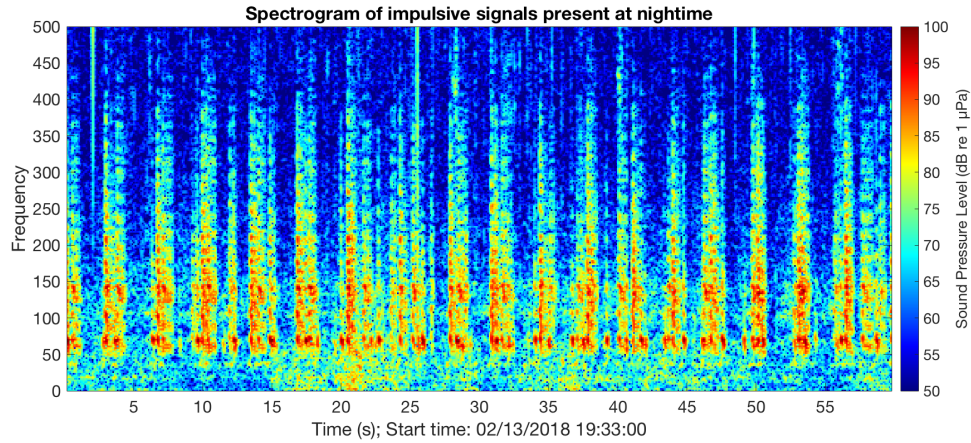
Figure 3.4: Spectrogram of the transient impulsive signals that increase the coherence in the 50-400 Hz band during nighttime. These are believed to be from a fish species of the *Sciaenidae* family. The spectrogram was plotted using the down-sampled data at 9.6 kHz, with 75% overlapping FFT windows of length 4096 samples.

Table 3.1: Clock offset measurements and calculated linear clock drifts rates for each array.

Array	t_0 (Date/time)	τ_0 (s)	t_{end} (Date/time)	τ_{end} (s)	T_{drift} (ms/h)
RO	02/13 10:50	-0.578	02/22 16:00	-3.794	-14.54
GB	02/13 10:35	-1.343	02/22 16:00	-11.577	-46.22

After removing the dominant linear clock drift component, the small-scale non-linear fluctuations can be observed from the noise cross-correlation structure over longer periods of time, as seen in Fig. 3.5, which shows stacked cross-correlations for both arrays over the entire deployment. These fluctuations are different between the two arrays. The RO array shows a larger excursion of up to about 50 ms, while the GB array shows smaller fluctuations between ± 20 ms.

Using the RO array as an example, Fig. 3.6 shows the steps involved in estimating the continuous interpolated fit for $p_{NL}(t)$. The resulting modeled function follows the cross-

correlation peak to within a millisecond. Figure 3.7 then shows the non-linear component of clock

offset $\tau_{NL}(t)$ for both deployments, which is computed from $p_{NL}(t)$ using Eq. 3.7.
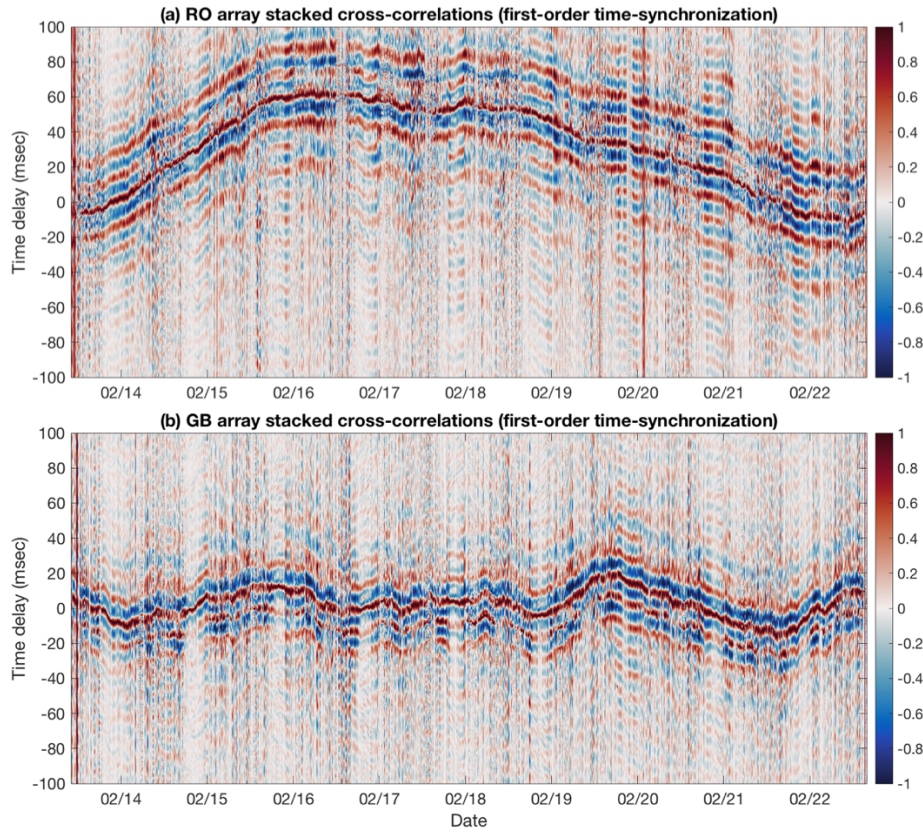
Figure 3.5: Stacked cross-correlations for the (a) RO and (b) GB arrays over the entire deployment, after the first-order time-synchronization. The visible shifts of the ambient noise cross-correlation structure are related to the non-linear clock drift $\tau_{NL}(t)$.
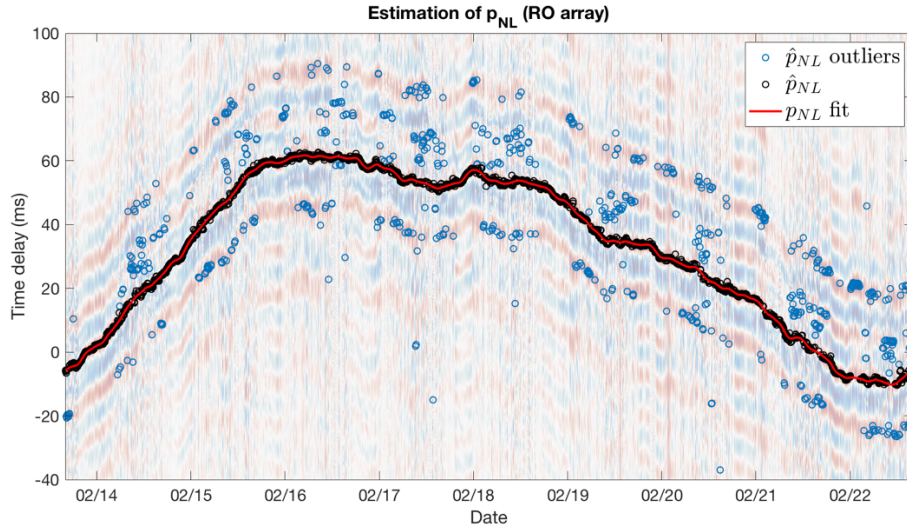
Figure 3.6: Estimation of $p_{NL}(t)$ for the RO array, overlaid over the stacked cross-correlations (same as Fig. 3.3a). Outliers in $\hat{p}_{NL}(t)$ (blue circles) are manually removed before applying a spline interpolation to the remaining data points of $\hat{p}_{NL}(t)$ (black circles), to obtain a continuous function for $p_{NL}(t)$ (red line).
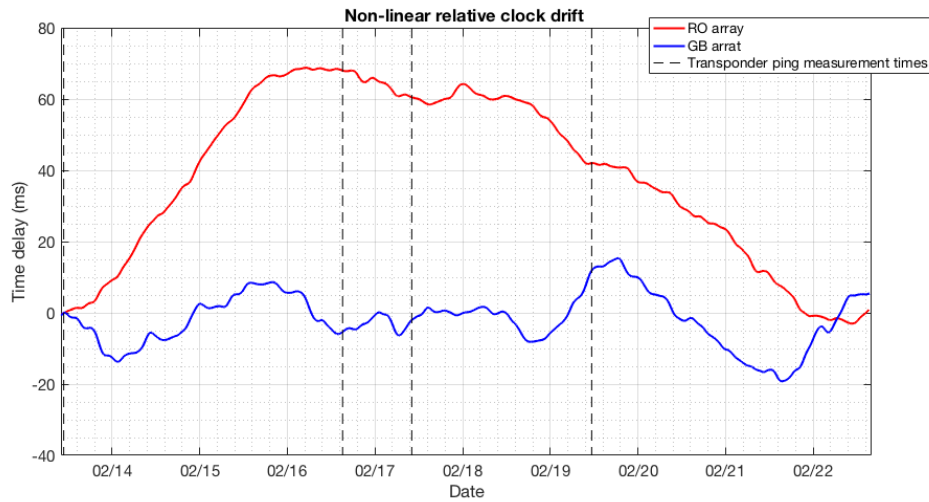
Figure 3.7: Non-linear clock offsets $\tau_{NL}(t)$ for both the RO and GB arrays obtained by applying Eq. 3.7 to the interpolated $p_{NL}(t)$ functions. The transponder ping measurement times are also plotted as vertical dashed lines.

Measurements of the transponder ping time delays were performed four times (shown in Fig. 3.7) during the experiment. Figure 3.8 shows that for all measurements the time delays of the transponder pings remain within ± 1 ms, even though the non-linear time drift show time offsets from –20 to 70 ms. The transponder pinrg measurements thus independently confirm the predictions of the ambient noise time-synchronization. Additionally, the fact that the transponder ping time delays lie so close to zero confirms that the transponders of each array were indeed nearly halfway between the two instruments. A detailed list of the transponder ping measurements is shown in Appendix 3.8.2.
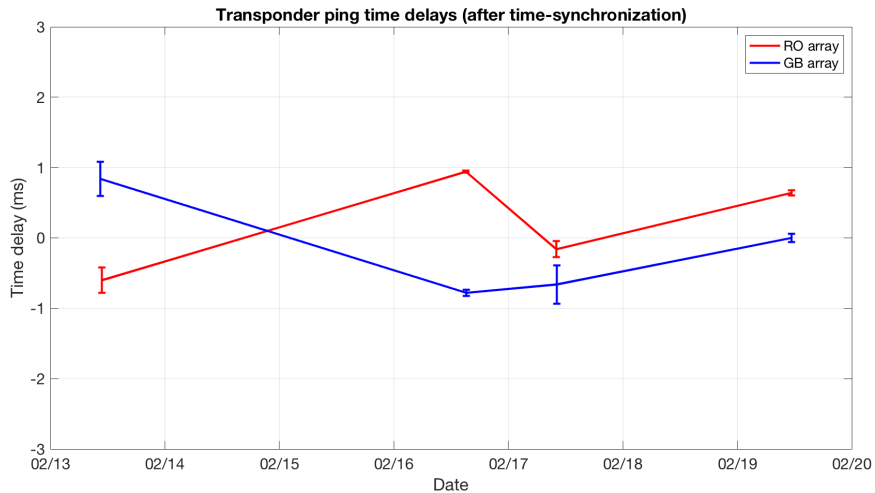
Figure 3.8: Median time delays of transponder pings from each array measured on four different days throughout the experiment. These measurements were made after the full time-synchronization was applied. The error bars represent the IQR between the 25th and 75th percentiles of each series of pings from which the median was computed.

3.5 Discussion

3.5.1 Evaluation of ambient noise time-synchronization

The approach presented in this paper relies on one main assumption: that the spatial coherence of the ambient noise field remains stationary over time. This assumption requires that the travel time between instruments remains steady, and that the statistical angular distribution of the sources that comprise the ambient noise field, is also stationary.

The travel time is dependent on both the separation L of the instruments, and the average sound speed c. Here, the instruments are assumed not to move since they are weighted and lying on the bottom. This assumption has been verified in subsequent datasets by checking the instrument separation with boat noise (as shown in Appendix 3.8.1) several times throughout the deployment. Additionally, for closely spaced instruments, changes in sound speed have negligible impact on the travel time between instruments. For example, at 20°C a 1°C change in seawater temperature causes approximately a 0.3% change in sound speed, which results in a change of travel-time of 0.03 ms between two sensors 14 m apart (Mackenzie, 1981). In this case, the temperature fluctuations during this experiment (Appendix 3.8.3) will have a negligible effect – less than 0.1 ms – on the travel-time.

However, the assumption that the angular distribution of the ambient noise sources remains stationary raises the question as to whether the observed shifts in $\tau_{NL}(t)$ are a result of clock drift alone, or changes in the spatial distribution of sources themselves. The independent transponder ping time delay measurements provide the most compelling evidence that the underlying coherent ambient noise field is sufficiently stationary to restrict any shifts in time delay to within ±1 ms. For example, if the bulk of the ambient noise sources arrive broadside of the 14 m array, then the distribution shift must be less than ±6.2° in order to generate a shift of less than ±1 ms in the

ambient cross-correlation peak. Were the bulk of the ambient noise sources arriving endfire of the array, then the distribution shift must be less than ±27° to generate less than a ±1 msec shift. Any changes in the noise source spatial distribution beyond those angles would have caused shifts in the time delays of the pings throughout the deployment. The transponder ping measurements thereby confirm that any changes in the source angular distributions are sufficiently small such that the observed changes in the structure of the cross-correlation are dominated by clock drift, and not a redistribution of the noise sources.

The magnitude of the non-linear clock drift (Fig. 3.7) provides additional evidence that the observed non-linear time offsets arise from clock drift, and not redistributions of ambient noise sources. Assuming the instruments are 14 m apart and a sound speed of 1500 m/s, physically implies all time-delays must lie within the range ± 9.33 ms. Figure 3.7, by contrast, demonstrates that the time offsets in $\tau_{NL}(t)$ easily exceed these values, and thus instrumental clock drift must be responsible for the majority of the drift observed, and not shifts in the spatial distribution of the noise sources. Comparing $\tau_{NL}(t)$ between the two arrays also shows no temporally correlated shifts, which would have been expected if changes in the ambient noise field were responsible for the observed shifts.

3.5.2 Origin of coherent ambient noise signal

The results show that throughout the deployment, noise components between 50 and 100 Hz tend to be highly coherent (Fig. 3.3d). Note that this bandwidth is coherent with the spacing of the peaks in the stacked cross-correlation plots (Fig. 3.3b for example), which is between 10 and 15 ms. In the evenings, the coherence increases in both magnitude and bandwidth, extending these daily events to up to 400 Hz, and permitting the adaptive cross-correlation function (Eq. 3.12) to sharpen the cross-correlation peaks (Figs. 3.3a-c).

The spectrogram (Fig. 3.5) reveals that a cacophony of short pulses dominate the ambient noise field at these specific times and bandwidth. Despite being at lower frequencies than what is described in the literature, based on their regular daily cycle, these pulses are believed to be fish calls from the *Sciaenidae* family, commonly known as the croakers (Locascio and Mann, 2011; Erisman and Rowell, 2017; Rice *et al.*, 2017). Laguna San Ignacio is inhabited by four species from this family: *Bairdiella icistia*, *Menticirrhus undulates*, *Umbrina roncador* and *Umbrina xanti* (Barjau Gonzalez, 2008), but it is not known which one(s) of these species are producing the measured pulses.

In Sabra, *et al.* (2005c), croaker fish were also found to be the main source of coherent ambient noise. Cross-correlations were averaged over an 11 min time window between 300-700 Hz. The omnidirectionality of the field and the coherence at higher frequencies allowed extracting features of the Green's function itself. In contrast, because most of the coherent energy presented here is at lower frequencies and a narrower bandwidth (50-100 Hz), the ambient noise cross-correlation does not allow extracting features of the Green's function. However, it is shown here that the noise structure is sufficient to infer clock drift information to within ±1 ms, over long periods of time.

The assumption that fish with a diel acoustic activity pattern are producing these pulses raises the question whether there is a diel cycle to their spatial distribution. However, a Fourier analysis of $\tau_{NL}(t)$ shows no sign of a regular oscillation in the clock offset, suggesting there is no diel oscillation present in the ambient noise spatial distribution. Combined with the independent transponder ping measurements, this observation provides further evidence that any potential changes in the angular distribution of these noise sources is sufficiently small to restrict changes in the cross-correlation peaks to ±1 ms.

3.5.3 Signal processing parameters for the ambient noise cross-correlation

Previous literature has shown that averaging the cross-correlation function over longer periods of time can improve the Green's function estimates, since longer averaging times suppress the uncorrelated components while making the sampled field more isotropic (Sabra *et al.*, 2005b). However, averaging for longer than 10 minutes did not show improvements in the results of this study, in terms of coherent bandwidth or timing precision.

Instead of averaging the cross-correlation functions over multiple snapshots as was done in similar studies, taking the median was found to yields better results (Fig. 3.3). This approach was particularly effective during the daytime when the coherence is lower and there is more potential to pick up loud transient outliers that are not representative of the stationary noise field.

In this specific study the improvement in performance of the weighted cross-correlation (Eqs. 3.9 and 3.10) over standard band-pass filtering was negligible. However, this approach had the practical advantage of allowing to the cross-correlation frequency bandwidth to be chosen automatically and adaptively, removing the need for altering the bandwidth manually to optimize coherence in the noise field.

3.5.4 Potential causes of clock drift

The measured relative clock drifts of the SoundTrap pairs are very high, and would make coherent processing impossible within minutes of being deployed. The large predominant linear clock drift of the instruments can be attributed to small manufacturing variations in the oscillating crystals that give each clock its unique linear drift rate. Indeed, subsequent deployments show the same value of T_{drift} for a given instrument pair.

The exact cause of the smaller-scale non-linear fluctuations remains unknown, but is most likely caused by fluctuations in the internal temperature or pressure of the instrument. Note that certain modern oscillator circuits (known as TCXOs) use temperature compensation to minimize temperature-related effects on the speed of the clock, but this was not the case of those of the SoundTraps used in this study.

However, relative clock drifts do not show any sign of correlation between the simultaneous deployments of the RO and GB arrays (Fig. 3.7), which rules out the possibility that changes in the water temperature are driving the clock fluctuations. Furthermore, the temperature record from the BG array is not correlated with observed relative clock drift (see Appendix 3.8.3).

3.6 Conclusion

Ambient noise cross-correlations in Laguna San Ignacio permit tracking the clock drift between two acoustic recorders to within ±1 ms, despite the fact that the coherent part of the ambient noise field is narrowband and is anisotropic. Most coherent components of the noise field are shown to be transient pulses, which are particularly dominant daily from 18:00 to midnight. Fish from the croaker family are believed to be the dominant source of this coherent noise.

This analysis assumed that the spatial distribution of the sources generating the ambient noise field between 50 and 100 Hz remained stationary throughout the experiment. Independent timing measurements from a transponder on both arrays confirmed that changes in the structure of the ambient noise cross-correlation were dominated by nonlinear clock drift, and any changes in the angular distribution of the noise sources were less than 27° for endfire sources and less than 6.2° for broadside sources.

3.7 Acknowledgements

3.8 Appendix

3.8.1 Using boat noise to determine clock offset and instruments separation

After the deployment of each array, the boat was driven in a circle of about 100 m radius around the deployment location. Figure 3.9a shows the boat track after the GB array deployment as an example.

The exact clock offset between the instruments at the time of the boat circles is then inferred from cross-correlating snapshots of the boat noise. As the boat circles around the array, the time delay associated with the cross-correlation peak varies from its minimum to maximum values as it goes from one endfire of the array to the other. The time delay associated with the cross-correlation peak of the boat track is shown in Fig. 3.9b, along with the time delay modeled based on the GPS track and the separation and orientation of the instruments. The mid-point between the maximum and minimum time delay represents the clock offset at that specific time. By trial and error, the exact clock offset can be found by adjusting the mid-point to 0, as is the case in Fig. 3.9b.

The difference between the time delay maxima and minima corresponds to the travel time between the instruments. Based on the sound speed, the separation can then be computed. Measurements from subsequent years show that the instrument separation remains constant throughout these deployments.

In order to emphasize the boat noise in the cross-correlation, the data were band-pass filtered from 400 to 5200 Hz and the snapshots length was 0.5 s.
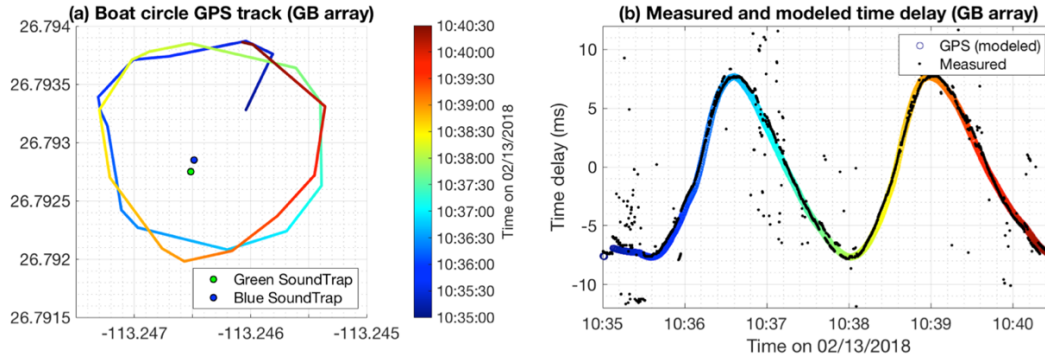
Figure 3.9: (a) GPS track of the boat circle around the GB array done immediately after its deployment and (b) time delay measured using boat noise cross-correlation (black dots) and modeled from the GPS track and separation and orientation of the instruments (colored circles).

3.8.2 Transponder ping transmissions

Table 3.2: Details for each transponder transmission (series of pings) used to validate the time-synchronization between the instruments. These measurements are plotted in Fig. 3.7.

Array	Date/time	Median time delay (ms)	IQR (ms)
RO	02/13 10:45:50	-0.6	0.18
RO	02/17 10:04:50	0.94	0.02
RO	02/17 10:14:25	-0.16	0.11
RO	02/17 10:20:20	0.64	0.04
GB	02/19 11:19:45	0.84	0.25
GB	02/19 11:27:20	-0.78	0.05
GB	02/19 11:31:40	-0.66	0.27
GB	02/16 15:11:20	0	0.06

3.8.3 BG array temperature record

Figure 3.10: Temperature record measured at the GB array by the HOBOware sensor over the entire experiment.

Chapter 4

Multi-target 2D tracking method for singing humpback whales using vector sensors

Acoustic vector sensors allow estimating the direction of travel of an acoustic wave at a single point by measuring both acoustic pressure and particle motion on orthogonal axes. In a two-dimensional plane, the location of an acoustic source can thus be determined by triangulation using the estimated azimuths from at least two vector sensors. However, when tracking multiple acoustic sources simultaneously it becomes challenging to identify and link sequences of azimuthal measurements between sensors to their respective sources. This work illustrates how two-dimensional vector sensors, deployed off the coast of western Maui, can be used to generate azimuthal tracks from individual humpback whales singing simultaneously. Incorporating acoustic transport velocity estimates into the processing generates high-quality azimuthal tracks that can be linked between sensors by cross-correlating features of their respective azigrams, a particular time-frequency representation of sound directionality. Once the correct azimuthal track associations have been made between instruments, subsequent localization and tracking in latitude and longitude of simultaneous whales can be achieved using a minimum of two vector sensors. Two-dimensional tracks and positional uncertainties of six singing whales are presented, along with swimming speed estimates derived from a high-quality track.

4.1 Introduction

Each winter, humpback whales (*Megaptera novaeangliae*) congregate in their respective breeding grounds in tropical waters around the globe. At these gatherings, males produce long sequences of structured vocalizations referred to as songs (Payne and McVay, 1971; Au *et al.*, 2006). Song is believed to play an important function in the mating system of humpback whales by signaling to females and mediating male-male interactions (Darling *et al.*, 2006; Herman, 2017; Cholewiak *et al.*, 2018). Despite having been studied since the early 1970s, many aspects of humpback whale singing remain poorly understood and the exact way in which song mediates whale interactions is still unclear.

While the recent advancements of bioacoustic tags have allowed documenting fine-scale acoustic behavior of specific individuals in many marine mammal species (Schmidt *et al.*, 2010; Goldbogen *et al.*, 2014; Stimpert *et al.*, 2020), tagging remains logistically expensive, yields low sample sizes and doesn't always allow attribution of detected calls to the tagged whale. Passive acoustic localization and tracking is a complementary approach to tagging as it is non-invasive and allows addressing ongoing research questions on a larger scale (Noad *et al.*, 2004; Schmidt *et al.*, 2010; Stanistreet *et al.*, 2013; Helble *et al.*, 2015; Helble *et al.*, 2016; Guazzo *et al.*, 2017; Henderson *et al.*, 2018). If few whales are present at a time, a network of widely spaced hydrophones can be used to track their position using time-of-arrival techniques on calls that are detected on at least three sensors (Schau and Robinson, 1987; Spiesberger, 2001). However, when large numbers of singing humpback whales are present, such as on their breeding grounds off Hawaii (Au *et al.*, 2000), their songs dominate the ambient noise field. Because the overlap in time and in frequency hinders signal extraction from individual whales, conventional localization techniques (e.g., time-domain cross-correlation between sensors) become impractical.

Developed during the second half of the 20th century, acoustic vector sensors were originally used in navy operations for detecting and localizing submarines, primarily through their use of DIFAR sonbuoys (Holler, 2014). In recent years, they have been developed into commercial recording packages and are used for a variety of passive acoustic monitoring applications (Greene *et al.*, 2004; Raghukumar *et al.*, 2020). In addition to measuring acoustic pressure like conventional hydrophones, vector sensors also measure particle motion, allowing them to estimate the dominant direction of travel of acoustic energy from a single point (D'Spain *et al.*, 1991; D'Spain *et al.*, 2006; Martin *et al.*, 2016). This ability permits triangulation to localize acoustic sources using multiple vector sensors. In addition to requiring a minimum of only two sensors, this approach is also highly advantageous when compared to time-of-arrival techniques required by conventional hydrophones, as triangulation does not require precise time-synchronization between independent autonomous sensor packages. However, while triangulation is relatively straightforward to execute and automate in the presence of few sources producing stereotyped signals (Greene *et al.*, 2004; Thode *et al.*, 2012; Thode *et al.*, 2021), it becomes challenging to apply when trying to track multiple non-stereotyped sources such as humpback whales.

Data from vector sensors can be processed in a variety of ways, including additive beamforming of the pressure and velocity channels (McDonald, 2004), or by multiplicative processing, where the pressure and velocity channels are multiplied together to model the acoustic intensity (D'Spain *et al.*, 1991). The speed of the latter approach makes it possible to display the dominant directionality of an ambient noise field as a function of time and frequency, a display that has been exploited many times and has recently been nicknamed an "azigram" when applied to biological studies (Thode *et al.*, 2019). One of the key advantages of the azigram versus a conventional spectrogram is that it allows individual sources to be distinguished based on their

location. Another key metric, the acoustic transport velocity, also describes the directional properties of an acoustic field and can be computed using vector sensors. This quantity, which is later defined in further detail, gives insight into the spatial distribution of the acoustic sources generating an ambient acoustic field. It allows differentiation between directional sources such as a whale or a boat, and diffuse background ambient sound which arises from numerous simultaneous sources from multiple directions (D'Spain *et al.*, 1991).

In this study we exploit the unique directional capabilities of vector sensors to track multiple singing humpback whales simultaneously, despite substantial overlap in their songs. We demonstrate this method on data collected off the coast of western Maui during the late breeding season in 2020, using modified DIFAR vector sensors (which measure horizontal particle velocity in two orthogonal directions) that enable two-dimensional (latitude and longitude) localization and tracking of whales. Section II presents the theory related to vector sensors relevant to this work, including the key metrics of dominant directionality and transport velocity. Section III gives an overview of the instruments, the deployment, and the demonstration dataset. Section IV details the tracking method along with examples and results illustrating the different steps of the algorithm. Finally, Section V presents tracking results for six whales, the track-derived swimming speed for one singer, and discusses the results as well as further potential applications.

4.2 Vector sensor theory

Different quantities can be used to describe an acoustic field. The most prevalent metric is acoustic pressure, which is easily measured underwater with a conventional hydrophone. This scalar quantity represents the amount of compression between particles of the medium and is also the quantity measured by both terrestrial and marine mammal ears. However, measurements of

acoustic pressure alone do not suffice in fully describing the physics of an acoustic wave traveling through a point. Particle motion on the other hand is a vector that quantifies the movement of the particles as sound travels through the medium, expressed in terms of displacement, velocity or acceleration. Vector sensors are designed to measure both acoustic pressure and particle velocity along two or three orthogonal axes.

The instantaneous acoustic intensity along a given axis k is defined as

$$I_k = pv_k$$

(4.1)

where p and v_k are the time series of acoustic pressure and particle velocity respectively. If the acoustic field is comprised by a single plane wave arriving from a distant, dominant, and spatially compact source, the magnitude of the particle velocity is proportional to and in phase with the pressure. Equation 4.1 then reduces to a form where the squared pressure alone can measure the intensity magnitude. However, since vector sensors measure pressure and particle motion independently, they provide direct measurements of the true underlying acoustic intensity, even in circumstances where the acoustic field is not dominated by a single plane wave.

The frequency-domain acoustic intensity S_k can be estimated at time-frequency bin (T, f) as

$$S_k(T,f) = \langle P(T,f)V_k^*(T,f)\rangle \equiv C_k(T,f) + iQ_k(T,f)$$

(4.2)

where P and V_k are short-time Fast Fourier Transforms (FFTs) of p and v_k respectively (Mann *et al.*, 1987). The symbol $*$ denotes the complex conjugate of a complex number, and $<>$ represents the ensemble average of a statistical quantity. If a time series can be considered to be statistically ergodic over a given time interval, this ensemble average can be obtained from time-averaging

consecutive FFTs (D'Spain *et al.*, 1991). In reality the ambient acoustic fields are often highly nonstationary, but a short enough time interval can typically be found where the ergodicity assumption is valid. In Eq. 4.2, C_k and Q_k are defined as the active and reactive acoustic intensities respectively and comprise the in-phase and in-quadrature components of the pressure and particle velocity. The active intensity C_k comprises the portion of the field where pressure and particle velocity are in phase and thus transporting acoustic energy through the measurement point. The reactive intensity Q_k comprises the portion of the field where pressure and particle velocity are 90° out of phase and arises whenever a spatial gradient exists in the acoustic pressure (Mann *et al.*, 1987). For the rest of this paper, we ignore the reactive component of intensity, and use the active component to define two directional metrics: the dominant azimuth and the normalized transport velocity (NTV).

In the case of a two-dimensional vector sensor that measures particle velocity along the x and y axis, the dominant azimuth from which acoustic energy is arriving φ, can be defined as

$$\varphi(T,f) = tan^{-1} \frac{C_x(T,f)}{C_y(T,f)}$$

(4.3)

where φ is expressed in geographical terms: increasing clockwise and starting from the y axis. The dominant azimuth can be displayed as a function of both time and frequency, as an image: the azigram (Thode *et al.*, 2019).

It is important to note that Eq. 4.3 estimates only the *dominant* azimuth, since acoustic energy may be arriving from different azimuths simultaneously at the measurement point. Equation 4.3 effectively represents an estimate of the "center of mass" of the arriving energy but provides no information about its angular distribution around the sensor. The normalized transport velocity (NTV) is a quantity that provides this second order information about the acoustic field.

For the same two-dimensional vector sensor assumed for Eq. 4.3, the NTV is defined by the ratio between the active intensity and the energy density of the field

$$U(T,f) = \frac{2\rho_0 c \, \langle [C_x^2(T,f) + C_y^2(T,f)]^{1/2} \rangle}{\rho_0^2 c^2 \, \langle |V_x(T,f)|^2 + |V_y(T,f)|^2 \rangle + \langle |P(T,f)|^2 \rangle}$$

(4.4)

where ρ_0 and c are the density and sound speed in the medium respectively (Mann *et al.*, 1987; D'Spain *et al.*, 1991). Equation 4.4 is normalized such that the NTV is comprised between 0 and 1. Although the NTV should be computed using particle velocity measurements along all three spatial axes, when measuring low-frequency sound in a shallow-water acoustic waveguide, only a small fraction of the total acoustic energy is transported vertically (along the z axis) into the ocean floor. As a result, a relatively accurate NTV can be obtained on a two-dimensional sensor using only particle velocity measurements along the horizontal axes. A NTV close to 1 implies that most of the acoustic energy traveling through the measurement point is clustered around the dominant azimuth. Such would be the case for a single azimuthally compact source such as a whale or a ship whose signal-to-noise ratio (SNR) is high. By contrast, a NTV of 0 indicates that no net acoustic energy is being transported through the measurement point, which implies either no acoustic energy is present at all, or equal amounts of energy are being propagated from opposite directions, as is the case for a standing wave. Thus, low transport velocity occurs in the presence of noise fields that are either isotropic or azimuthally symmetrical.

4.3 Illustrative dataset: Maui 2020

A DASAR (Directional Autonomous Seafloor Acoustic Recorder) Model "C" is an autonomous underwater recording package equipped with a DIFAR (Directional Frequency Analysis Recording) vector sensor, which is itself composed of an omnidirectional pressure sensor (149 dB re 1 μPa/V @ 100 Hz sensitivity) and two particle motion sensors capable of measuring the x and y components of particle velocity (Greene *et al.*, 2004; Thode *et al.*, 2012). The signals measured on each of the three channels were sampled at 1 kHz with sensors that have a maximum measurable acoustic frequency of 450 Hz.

The sensitivity of the directional channels, when expressed in terms of plane-wave acoustic pressure (-243.5 dB re m/s equates to 0 dB re 1 μPa), is 146 dB re 1 μPa/V @ 100 Hz. The sensitivity of all channels increases by +6 dB/octave (e.g., the sensitivity of the omnidirectional channel is 143 dB re 1 μPa/V @ 200 Hz), since the channel inputs are differentiated before being recorded. These values were measured from two DASARs calibrated at the U.S. Navy's underwater acoustic test facility TRANSDEC in San Diego in 2008. A finite impulse response (FIR) equalization filter was applied to recorded data to recover the original spectrum.

Between March and July 2020, three DASARs labeled A, B and C were deployed along the south facing coast of western Maui, capturing the last couple of months of the humpback whale breeding season. The instruments were spaced by approximately 3 km in a line running from the northwest (DASAR A) to the southeast (DASAR C) as shown in Fig. 4.1 at depths of approximately 20 m. The DASARs were lowered to the ocean floor from a small vessel using a rope, and thus the orientation of the package on the ocean floor could not be controlled and had to be measured acoustically. Using the same calibration technique as Thode *et al.* (2021), the small vessel was driven clockwise and counterclockwise around each DASAR after its deployment. From the GPS position of the boat and the associated estimated azimuths, the clock offset between

the GPS and the sensor data could be derived, along with the seafloor orientation of the sensors'

particle velocity axes. Additionally, this procedure was used to measure the uncertainty of the

dominant azimuth estimates (7.61° median error). The details of the calibration are shown in

Appendix 4.9.

In this study, 24 hours of data starting from midnight on April 18, 2020, are presented and

analyzed. This time window, which has fewer whales present than earlier in the breeding season,

was chosen for clarity of the presented method.
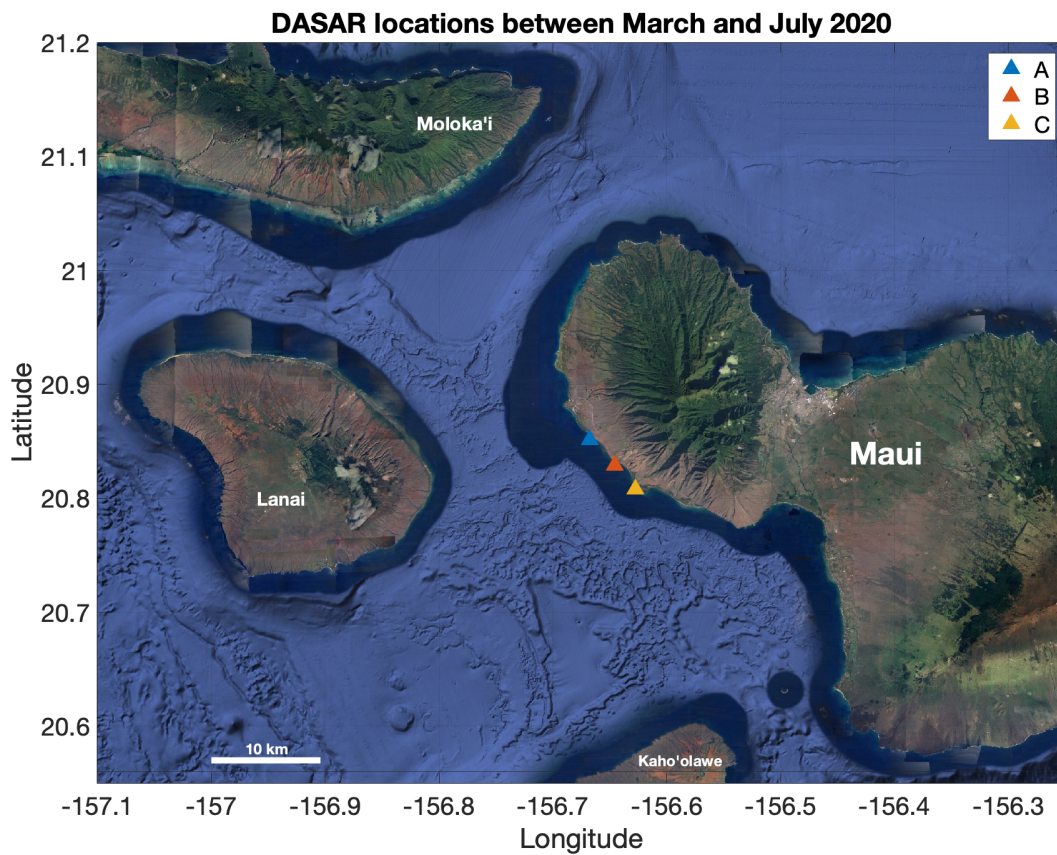
Figure 4.1: Satellite image indicating the position of DASARs A, B and C, deployed
between March and July 2020 off the western coast of Maui.

4.4 Tracking algorithm

4.4.1 Time-frequency representation of directional metrics

As shown in Eqs. 4.3 and 4.4, both dominant azimuth φ and NTV can be computed for each time-frequency bin (T, f) of a signal, allowing these quantities to be displayed as an image. In the dominant azimuth representation – the azigram – the color of each pixel is associated with a given geographical azimuth. In the NTV representation, the color of each pixel corresponds to a value between 0 and 1. Figure 4.2 shows the spectrogram, azigram and NTV of 30 seconds of data from DASAR B, starting at 2:24 UTC-10 on April 18, 2020. While the spectrogram suggests the presence of multiple humpback whales singing simultaneously, it does not allow straightforward association of songs units to individual whales (Fig. 4.2a). The azigram display, however, reveals distinct individual whales based on their color/azimuth (Fig. 4.2b). In this particular plot, the color scale is restricted to the 100° and 350° azimuth range, which points away from shore (see Fig. 4.1). Furthermore, the NTV time-frequency representation shows that whale calls have high NTV values (Fig. 4.2c).
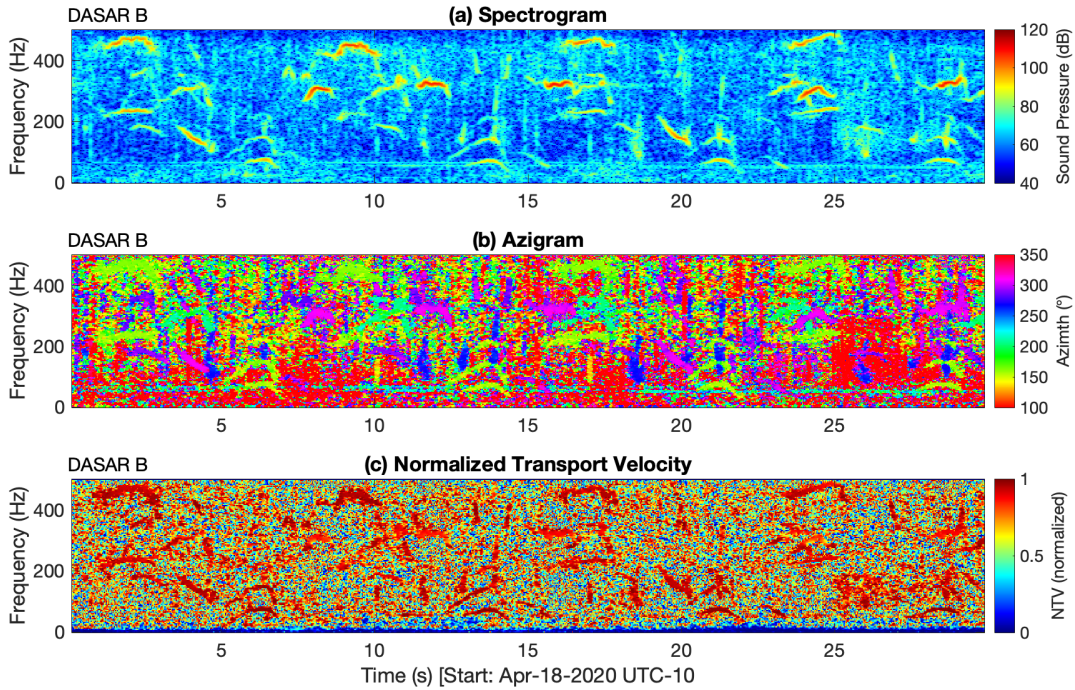
Figure 4.2: Spectrogram (a), azigram (b), and normalized transport velocity (c) over a 30-second time window starting at 2:24 UTC-10 on DASAR B. The color scale for (a) is in terms of power spectral density (dB re uPa^2/Hz), while the color scale for (b) is in terms of azimuth relative to geographic north. All three subplots are computed using window and FFT lengths of 256 samples with 90% overlap, and no time-averaging of the FFTs.

4.4.2 Identifying azimuthal tracks over long intervals

The number of singing whales and their azimuths to a DASAR can be estimated at any given time from the statistical distribution of φ. Let h_θ be defined as a histogram that counts the number of observations of $\varphi(T, f)$ that fall within azimuthal bin of center θ and width $d\theta$, within a time interval ΔT_h. h_θ estimates the distribution of azimuths measured across all time-frequency bins in the azigram. Note that the histogram time window ΔT_h should be long enough for the azigram to include whale calls from all currently singing whales, and short enough for any changes in the animal's azimuth to be negligible. For identifying humpback whale songs in this dataset, a time window $\Delta T_h = 60\ s$ was found to be sufficient. To minimize contributions to h_θ of the

background noise and other non-directional sources, a NTV threshold can be applied so that any observation $\varphi(T, f)$ associated with a NTV below a value γ_U is discarded. For example, the histogram computed from Fig. 4.2b would only be computed with azimuthal values whose NTV is above γ_U in Fig. 4.2c. The resulting filtered histogram H_θ emphasizes azimuths than are associated with highly directional compact sources such as whales and boats. H_θ is normalized by its maximum value such that the most probable azimuth is always 1. Figure 4.3 displays sample histograms, before and after applying the NTV threshold and normalization, illustrating how this filtering enhances the azimuths associated with four distinct whales.
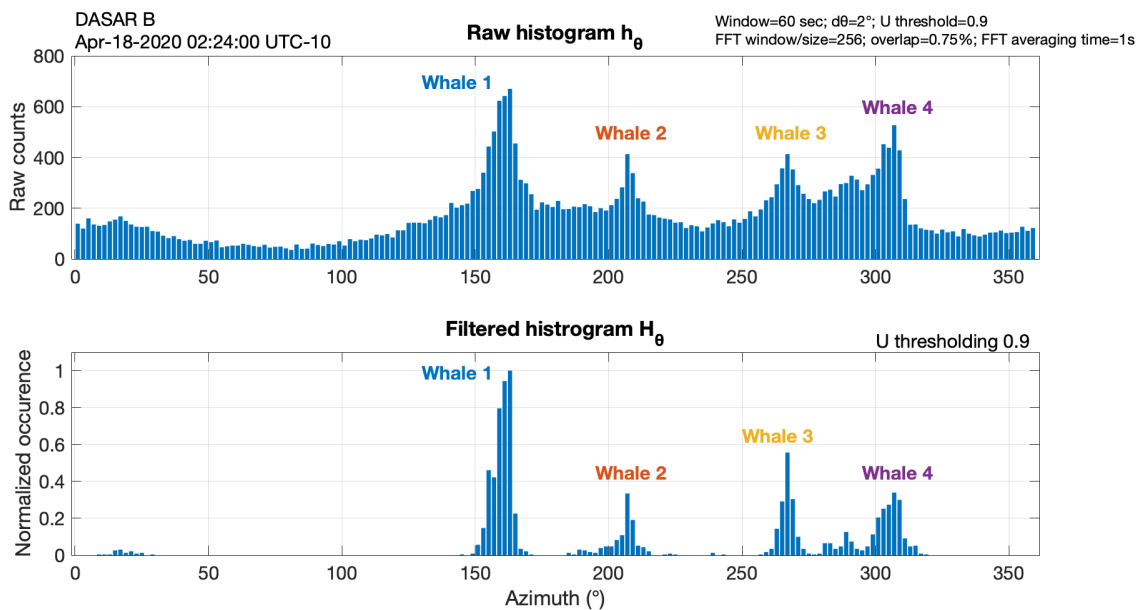
Figure 4.3: Raw histogram h_θ and filtered histogram H_θ at 2:24 UTC-10. These histograms estimate the distribution of azimuths over a time window $\Delta T_h = 60\ s$ using bin width $d\theta = 2°$, and illustrate how NTV thresholding ($\gamma_U = 0.9$) enhances the azimuthal peaks associated with (at least) four distinct humpback whales. The azigrams used to compute these histograms have window and FFT lengths of 256 samples with 75% overlap, and 1s time-averaging of the FFTs.

H_θ can be plotted as a function of both time and azimuth by stacking histograms into an image which will be referred to as "azimuthal histogram display" (AHD). This representation allows identification of continuous tracks from discrete sources such as whales and boats over longer periods of time. These "azimuthal tracks" are not to be confused with the resulting 2D localized tracks which correspond to a sequence of positions (latitude and longitude) for a given whale. Figure 4.4 displays the AHDs for all three DASARs over a 24-hour time window starting at midnight on April 18, 2020. Also shown in Fig. 4.4 is the unfiltered AHD from DASAR C, illustrating how the NTV filtering and normalization improves the quality of the azimuthal tracks. Whales and small boats are easily distinguishable by the slope (azimuthal rate) of their azimuthal tracks. Because they are generally louder and thus typically more distant from the sensors, the azimuthal rate of whales is low, on the order of several degrees per hour. By contrast small boats generally travel faster and need to be closer to a DASAR to be detected, thus displaying higher azimuthal rates and producing tracks that are much shorter (on the order of minutes) and steeper (nearly vertical lines in Fig. 4.4). The AHDs in Fig. 4.4 also indicate that whale song only arrives from between approximately 100 and 350° (i.e., away from shore), while boat tracks are visible from all directions, including from the direction of the coastline.

To localize multiple individual whales over time, azimuthal tracks from the same whale need to be linked on at least two DASARs. The deployment configuration is such that if a whale is only detected on two of the three DASARs, as it is sometimes the case, DASAR B (the middle instrument) will always be involved. The practical problem to solve here is thus to determine which, if any, of the azimuthal tracks from DASAR B are associated with tracks from DASARs A and/or C.
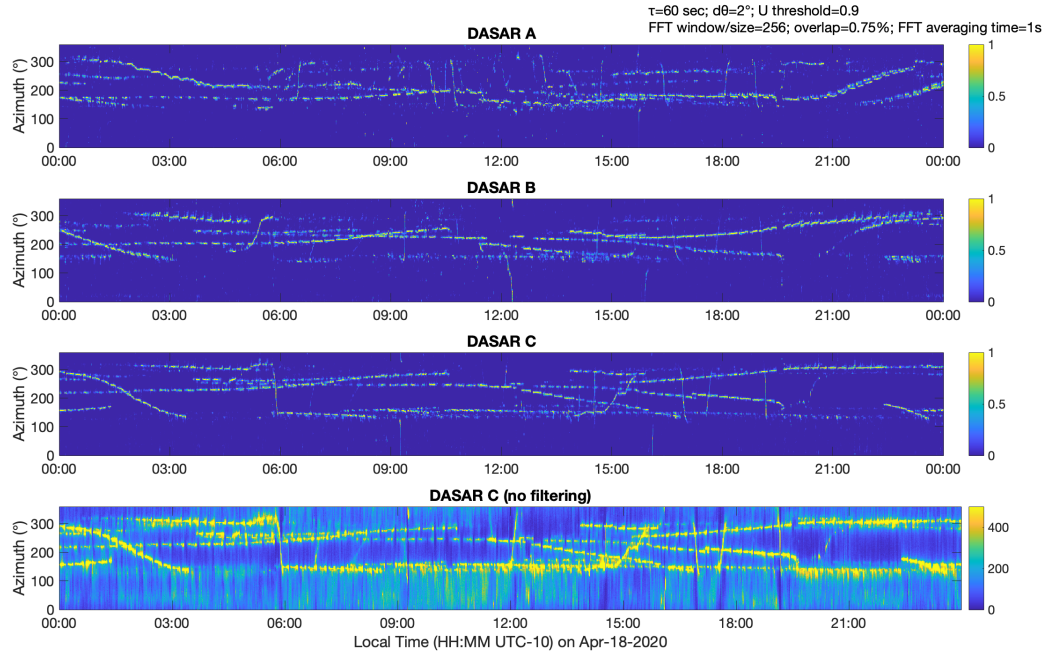
Figure 4.4: Azimuthal histogram displays for DASARs A, B and C on April 18, 2020. These plots reveal azimuthal tracks of both whales (longer and smoother) and boats (quasi-vertical lines). The bottom plot shows the unfiltered AHD for DASAR C, illustrating how the filtering improves the visibility of the azimuthal tracks. The parameters used here are the same as those used to compute the histograms in Fig. 4.3.

4.4.3 Manual selection of azimuthal tracks

While the goal is to eventually automate the azimtual track selection process, the purpose of this study is to demonstrate the azimuthal track linking technique. Here, tracks are traced manually from the AHDs of all three DASARs between midnight and 3:00 UTC-10 on April 18, 2020. Let the nth azimuthal track on DASAR α be denoted $\Theta_{n\alpha}(t)$. Figure 4.5 shows both the AHDs and resulting azimuthal tracks. These sample tracks have been manually traced with *a priori* knowledge of their associations between sensors so that those sharing the same number are from the same whale (i.e., n is both the track and whale number). The tracks show six distinct whales, five of which are detected on all three DASARs, and one whale which is detected on DASARs B and C only (whale 6). The time resolution of the azimuthal tracks is 1 minute.

81
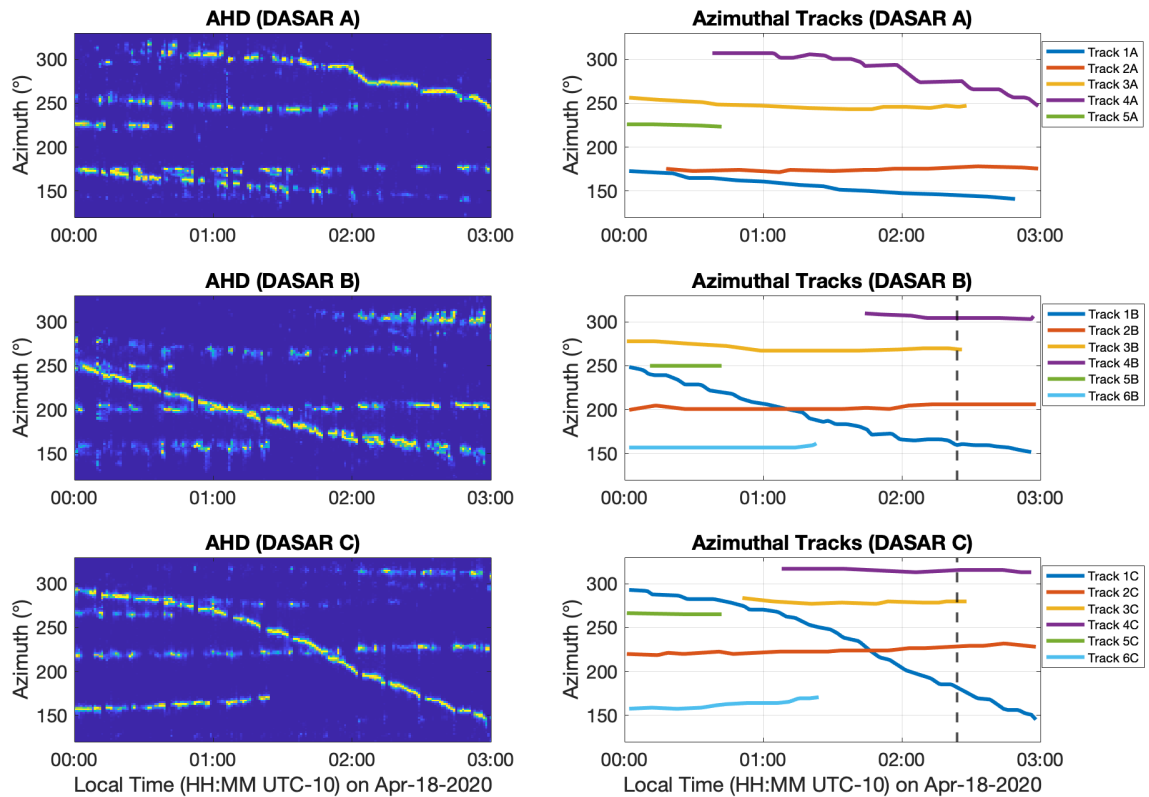
Figure 4.5: AHDs and manually selected azimuthal tracks from DASARs A, B and C between midnight and 3:00. The azimuthal tracks shown here have been selected such that tracks who share the same number are from the same whale. Their time resolution is 1 minute.

4.4.4 Azigram thresholding

Image thresholding can be applied to any azigram to create a binary image based on an azimuthal sector of width $d\varphi$, thus allowing whale calls arriving from a specific direction to be isolated on different DASARs. This process defined here as "azigram thresholding", works as long any two singing whales are separated by at least an angle $d\varphi/2$. To enhance the thresholded image, a 2D 3 by 3-pixel median filter (Huang *et al.*, 1979) is applied to remove speckle components of the image.

At any given time and DASAR, the azimuth associated with track $\Theta_{n\alpha}(t)$ (as shown in Fig. 4.5) can be used to threshold the corresponding azigram on sensor α and isolate the song units of whale n. As an example Fig. 4.6 shows four whale songs extracted from DASARs B and C. The 30-second time window presented here corresponds to the first half of the histograms shown in Fig. 4.3, and thus shows the same four whales.
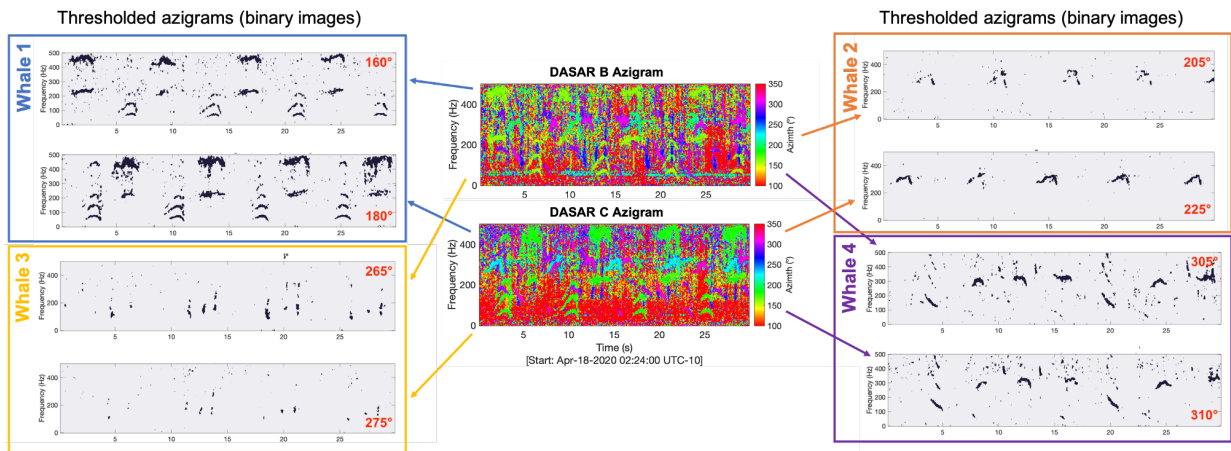
Figure 4.6: Azigrams and associated binary images obtained by applying azigram thresholding to DASARs B and C to isolate calls from four distinct whales. The azigrams were computed using the same parameters as used for Fig. 4.2. The center of the azimuthal sectors of width $d\varphi = 15°$ is displayed in red on each binary image. The 30-second time window presented here corresponds to the first half of the histograms shown in Fig. 4.3, and thus shows the same four whales.

4.4.5 Matching azimuthal tracks between DASARs

The goal is now to compare the thresholded azigrams, with no prior knowledge of their associations between DASARs, to determine which azimuthal tracks are from the same whale.

Let $B_\alpha(T, f)$ and $B_\beta(T, f)$ be two binary images covering a same time window of length ΔT_r, obtained from applying azigram thresholding to DASARs α and β respectively. The similarity of these two images can be quantified by taking the maximum value of the cross-correlation between B_α and B_β along time, expressed as

$$R = \max_\tau \left\{ \sum_f \sum_T B_\alpha(T, f) B_\beta(T + \tau, f) \right\}$$

(4.5)

where τ is the cross-correlation time delay. R can be normalized into a "cross-correlation score" as

$$\bar{R} = \frac{100}{\max\{[P_\alpha, P_\beta]\}} R$$

(4.6)

where P_α and P_β are the total number of positive pixels in B_α and B_β respectively. Equation 4.6 normalizes the cross-correlation score between any two images to lie between 0 and 100. Cross-correlating binary images is conceptually similar to "spectrogram correlation" methods used to detect stereotyped baleen whale calls (Mellinger and Clark, 2000).

For any time window that reports azimuthal tracks on two DASARs, the likelihood of these tracks being related can be assessed by computing their cross-correlation. The time window used for the cross-correlation ΔT_r should be long enough to include songs from all the whales being tracked, and short enough for their azimuths to remain constant to within the azimuthal sector $d\varphi$.

For comparing humpback whale calls in this dataset, a time window of $\Delta T_r = \Delta T_h = 60\ s$ and an azimuthal sector of $d\varphi = 15°$ (which corresponds to the azimuthal uncertainty derived in the appendix) was sufficient. By computing the median score of each cross-track combination across the portion of the two tracks that overlap temporally, the likelihood of any two azimuthal tracks being from the same whale can be identified. As an example, Fig. 4.7 shows the cross-correlation scores between Θ_{1C} and all six tracks from DASAR B. The median scores clearly indicate that Θ_{1B} is the most likely track associated with the reference track Θ_{1C}. The regular troughs in the cross-correlation score occur when the whale stops singing while it surfaces to breathe, approximately every 15 min.
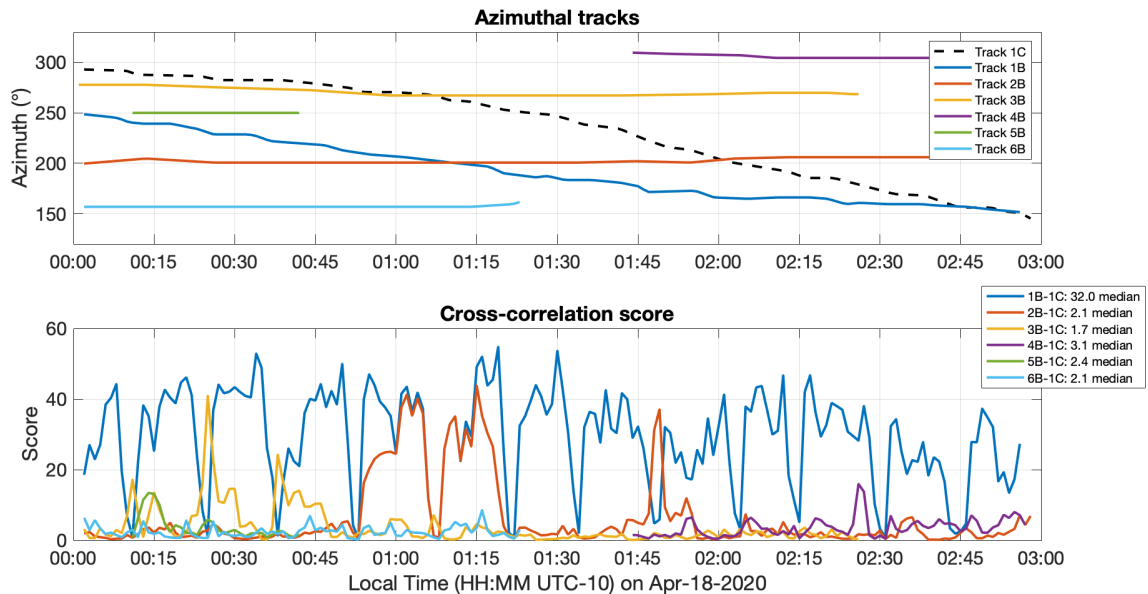
Figure 4.7: (top) Reference azimuthal track Θ_{1C} (dashed black line) compared to all six tracks from DASAR B (solid lines). These are the same manually traced tracks shown in Fig. 4.5. (bottom) Cross-correlation scores (Eq. 4.6) for overlapping times of any two tracks. The median scores of each comparison, which are displayed in the bottom legend, suggests that Θ_{1B} is the best match with reference track Θ_{1C}, as it has the highest median score. The cross-correlation time window and azimuthal sector width used here are $\Delta T_r = 60\ s$ and a $d\varphi = 15°$ respectively.

Following this procedure, the median scores for all combinations of tracks between DASAR B and DASARs A and C can be used to create confusion matrices, as illustrated in Fig. 4.8. As expected from the manual selection of the tracks, the correct associations (along the diagonal of the confusion matrices) consistently produce the highest median scores. Note that when comparing track Θ_{6B} to tracks from DASAR A (bottom row of left panel in Fig. 4.8), all median scores are low, which is expected since whale 6 is not detected on DASAR A (i.e., Θ_{6B} has no match on DASAR A). In this idealized example, any score above 15 is associated with a correct match between tracks.
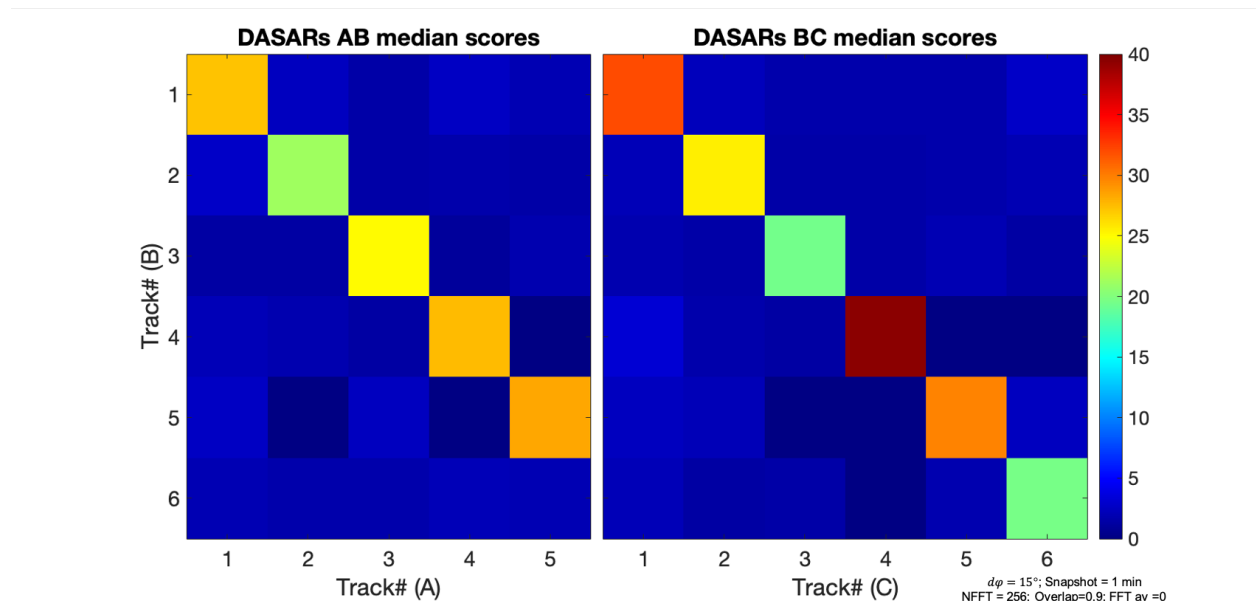
Figure 4.8: Confusion matrices for all combination of azimuthal tracks between DASAR B, and DASARs A and C. Each grid point in these matrices represents a track combination and is computed as the median value of their normalized cross-correlation scores (Eq. 4.6). The correct associations of azimuthal tracks along the diagonal consistently show the highest scores. The parameters used here at the same as those from Fig. 7.

4.5 Localization and tracking results

The whales whose azimuthal tracks were extracted and matched in the previous section are used to demonstrate the 2D tracking results. The 2D localization method used in this study relies on triangulation from at least two DASAR azimuthal measurements and allows estimating a localization precision if at least three measurements are available. This approach is based on (Lenth, 1981) and has already been used to track bowhead whales in several studies involving DASARs (Greene *et al.*, 2004; Blackwell *et al.*, 2007; Thode *et al.*, 2012). Here, a 2D "localized track" refers to a sequence of latitudes and longitudes derived from the linked azimuthal tracks of a same whale between distinct vector sensors. Individual localizations are therefore not linked to specific song units but rather they have the same 1-minute temporal resolution as the azimuthal tracks extracted from the AHDs.

Up to three DASARs are available in this dataset. Thus, at any given time the number of available azimuth measurements N (from the azimuthal tracks) for a whale lies between zero and three, $N \in [0, 3]$. A 2D whale track starts whenever $N \geq 2$ and ends if $N < 2$. Over time periods where $N = 3$, the localization uncertainty is also be computed for each localization. Figure 4.9 shows the 2D localized tracks of all six whales with a localization interval of 1 minute. Using the method from Greene *et al.* (2004), the 90% confidence ellipse is computed and displayed every 10 minutes over time periods where $N = 3$. The location uncertainty can be computed over part of all tracks, with the exception of whale 6, which is only detected on DASARs B and C.

The track from whale 1 is a good example which shows the whale traveling towards the southeast. The direction of travel can be inferred from the azimuthal tracks in Fig. 4.6. During the first part of the track, the whale is directly in front of DASAR B (about 2.5 kilometers off the coast) and is detected on all three instruments. As the whale travels southeast, the range

uncertainties grow larger since triangulated bearings cross at shallower angles. Eventually, at 2:28 when the whale is about 10 km away from DASAR A, it is only detected on DASARs B and C, and no localization uncertainty can be estimated. Using the first hour of the localized 2D track from whale 1 (when the animal is directly in front of DASAR B), the swimming speed is derived and shown in Fig. 4.10. The reported average swim speed of approximately 2 km/h is consistent with previous studies of signing whales (Frankel *et al.*, 1995; Noad and Cato, 2007). The azigram cross-correlation score (Eq. 4.6) between the tracks from whale 1 on DASARs B and C is also displayed in Fig. 4.10 to show when the whale is singing (high score) and when it is surfacing to breathe (low score). Comparing the swim speed to the dive cycle reveals that the two appear to be correlated, with the whale remaining mostly stationary while it sings.
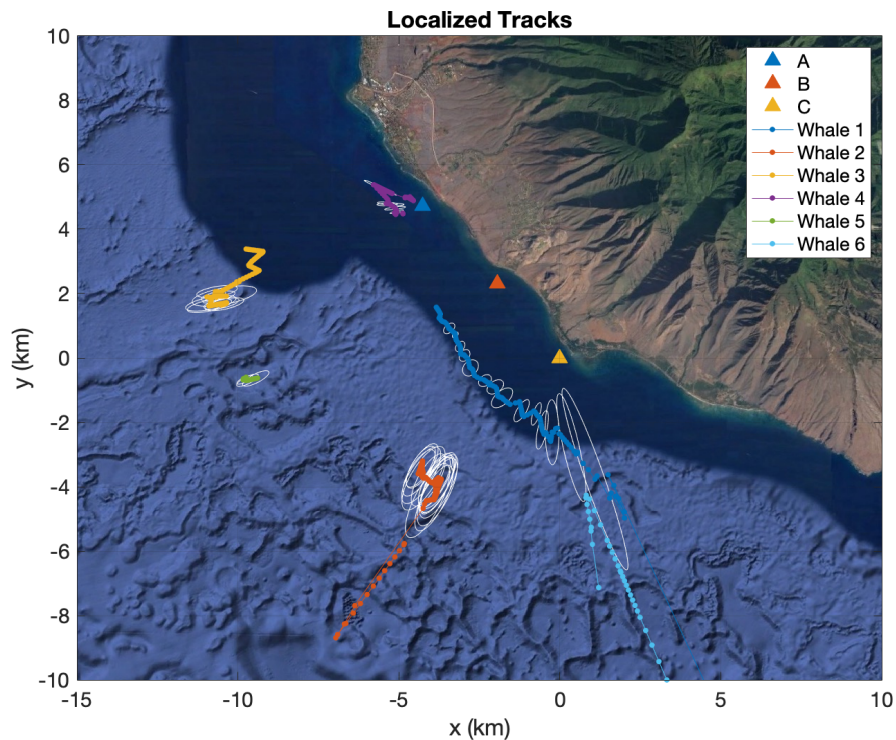
Figure 4.9: 2D localized tracks of the six whales whose azimuthal tracks were extracted and matched in Section IV. The whale 2D location estimates are computed every minute whenever at least two azimuthal measurements are available ($N \geq 2$). A 90% confidence ellipse is plotted every 10 minutes over time periods where three DASARs are available for the localization ($N = 3$).
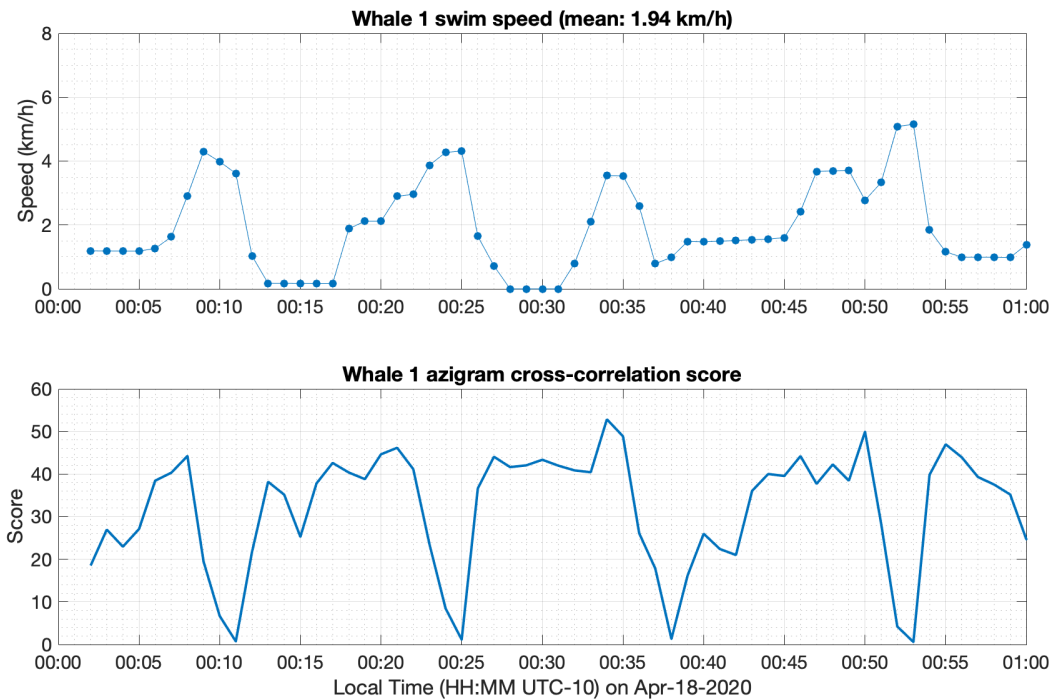
Figure 4.10: Swim speed of whale 1 derived from the first hour of its 2D localized track, when the whale is in front of DASAR B (top), and azigram cross-correlation score between the tracks from whale 1 on DASARs B and C (bottom). Here, the cross-correlation score indicates when the whale is singing (high score) and when it is surfacing to breathe (low score). Comparing the two plots shows that the swimming speed of the whale is correlated with the singing and diving cycle of the whale.

4.6 Discussion

4.6.1 Localization performance and limitations

Because of the intrinsic uncertainty in the dominant azimuth estimates, the performance of the 2D localization technique is highly dependent on the relative location of the source to the vector sensors and the deployment geometry itself. For this specific configuration with the DASARs along the coast, locations that are a few kilometers to the southwest of DASAR B produce the best results because (1) whales are detected on all three DASARs and (2) the azimuthal beams from all three DASARs intersect at steep angles. This is the case for approximately the first hour of the 2D

89

localized track from whale 1 in Fig. 4.9. As a whale moves further offshore, the SNR of its songs decreases along with the precision of the azimuth estimates, as well as the intersection angle of the beams and thus the localization precision. Similarly, when close the axis of this nearly linear array (termed "endfire"), the location uncertainties also become larger since the beams also intersect at shallow angles. Furthermore, the array aperture is such that whales that are endfire are often only picked up on the two closer DASARs. While the coastline deployment has practical advantages in terms of deployment logistics (e.g., shallow water depth and proximity to a harbor), the quality of the 2D localizations could be greatly improved by placing one or multiple DASARs farther away from the coast in deep water, or on the coast of Lanai, to define a triangular array perimeter.

4.6.2 Azimuthal track matching

The method used to compare and match azimuthal tracks across DASARs relies on cross-correlation of binary images obtained from azigram thresholding. While the example presented here is idealized (each azimuthal track from DASAR B has a unique matching track on at least one of the other two DASARs), identifying each matching track is straightforward (Fig. 4.8). The main potential weakness of this approach is its inability to differentiate between two whales that have the same azimuth on a given DASAR. Taking Fig. 4.7 as an example, the local cross-correlation around 1:10 UTC-10 shows high scores for both tracks Θ_{1B} and Θ_{2B}, which intersect at this time. Alternatively, two matching azimuthal tracks can produce low cross-correlation scores at times when the whale stops singing while it surfaces. These troughs, which can be seen approximately every 15 minutes in both the cross-correlation scores (to track 1B in Fig. 4.7) and the AHDs (Fig. 4), occasionally cause cross-correlation scores with the wrong azimuthal track to become high (as seen from the two cross-correlation spikes for track Θ_{3B}, around 00:24 and 00:40 UTC-10). Despite these potential limiting factors, when considering the entire azimuthal tracks and their

median scores, the track from the correct whale (track Θ_{1B}) produces the highest median score by far. Therefore, this technique can successfully link two tracks from a same whale as long as (1) it is singing throughout the majority of the azimuthal track and (2) the majority of its tracks are azimuthally unique on each DASAR.

An alternative method for matching azimuthal tracks from a same whale on multiple DASARs would be to attempt localizing every combination of tracks, and making a decision based on the uncertainty of the resulting 2D localized tracks. Indeed, the localization for the correct set of azimuthal tracks should produce physical 2D locations with lower uncertainties than potential localizations resulting from the wrong combination of tracks. However, this approach requires azimuthal tracks to be present on all three DASARs in order to compute the location uncertainties. While this approach may be faster and effective in accurately matching azimuthal tracks in conditions where whales have clearly distinguishable trajectories, the azigram cross-correlation technique presented here is more robust as it exploits additional features in the time-frequency domain to link azimuthal tracks.

4.6.3 Automation and real-time implementation

In order to process the entire dataset in a way that is both efficient and systematic, further automation of two key steps of the tracking algorithm is required: (1) extracting azimuthal tracks from the AHDs and (2) making association decisions between azimuthal tracks from different DASARs based on the median score confusion matrices (Fig. 4.8).

In this study, the matching problem was simplified by having a unique track for each whale on each DASAR. In practice, depending on how the azimuthal tracks are obtained, the track from one whale may be split into multiple segments (duplicate tracks). In this case, rather than tracks being matched one-to-one, duplicate tracks from one DASAR need to be attributed to a single track from

another. Alternatively, an azimuthal track from one DASAR may have no match on the other two (orphan track). Automating the whale tracking method would thus require implementing a decision-making algorithm (e.g., using thresholding) to determine whether azimuthal tracks are related.

The issues about duplicate azimuthal tracks are strongly related to the initial extraction of the azimuthal tracks. Indeed, if a given extraction algorithm is prone to producing duplicate tracks, the matching algorithm will have to be designed to allow matching these succesfully. Furthermore, the performance of a track extraction algorithm will also determine the ability of the 2D tracking algorithm to deal with large number of whales at the peak of the breeding season, when the ambient noise field is dominated by humpback whale songs. If the azimuthal tracks cannot be successfully extracted from the AHDs, 2D tracking using the method presented in this study becomes impossible.

Several techniques, such as Probability Hypothesis Density (PHD) filtering (Gruden and White, 2020) and graph-based approaches (Vo *et al.*, 2010; Meyer *et al.*, 2018) exist that would allow automatically extracting the azimuthal tracks from the AHDs. Note that these extraction techniques can also be implemented in real-time. By using only current and previous information to match azimuthal tracks, the 2D whale tracking algorithm presented here could potentially be achieved in real-time (or near real-time).

4.6.4 Processing parameters and potential applications

The parameters used here were chosen specifically for successfully identifying and matching humpback whale songs. The histogram and azigram cross-correlation time windows, ΔT_h and ΔT_r respectively, have similar requirements in that they both should be long enough to include song units from all currently singing whales. However, they need to be short enough to

account for both azimuthal changes of the source and the measurement uncertainty, with respect to the histogram bin width $d\theta$ and azimuthal sector $d\varphi$. The time window of 60 s used here for both the AHDs and azigram cross-correlation is suited for the rate at which the azimuth of whales' changes with time. In principle this same approach could be used to track boats. To do so, the length of the time windows would have to be reduced to account for the rapid changes in azimuth.

Note that the FFT averaging (1 second) was only used when computing the histograms. This allows reducing the variance of the azimuth estimates, which in turns makes the azimuthal tracks in the AHDs sharper. Because FFT averaging affects the shape of individual song units in the azigram image, the cross-correlation produces better results when no FFT averaging is applied.

The high NTV threshold used here to filter the AHDs is designed enhance the contributions from compact sources such as whales and boats. Alternatively, the ambient noise itself can be examined by removing samples of $\varphi(T, f)$ that have a high NTV and generating AHDs that only show contributions from low NTV sources. This approach would allow studying the directional characteristics of the ambient noise field over time.

In the subset of data presented in this study, the number of singing whales is such that individual whales can be distinguished. Earlier in the breeding season however, the number of whales is much higher, and distinguishing individual tracks in the AHDs becomes difficult. While this makes tracking individual whales less feasible, the overwhelming amount of song becomes an opportunity to treat humpack whale song as noise whose statistics can be analyzed (Seger *et al.*, 2016). Instead of tracking individual whales, the DASARs can be used to localize the "center of mass" of the whales. Furthermore, using longer time windows to compute the NTV, the azimuthal distribution of whales can be assessed to determine if whales are clustered or widely distributed.

4.7 Conclusion

This study presents a method for multi-target 2D tracking of acoustic sources using vector sensors. The technique, which relies on vector sensors' ability to measure directional quantities of the acoustic field, is demonstrated on simultaneously singing humpback whales off western Maui using three sensors. The extraction of azimuthal tracks from individual whales on each sensor is possible using a histogram representation of the azigrams (AHDs) to which a transport velocity threshold is applied to enhance their quality. Subsequently, the azimuthal tracks are compared across vector sensors by cross-correlating thresholded azigrams, allowing tracks from the same whale to be matched. Once the azimuthal tracks are correctly matched, individual whales can be localized and tracked in 2D (latitude and longitude). The localization is based on triangulation from each sensor, allowing a position uncertainty to be estimated whenever three azimuthal measurements are available. The method was demonstrated by tracking the position of six singing whales. The derived swimming speed of one whale track showed that the swimming speed of the whale is related to its dive cycle.

4.8 Acknowledgements

4.9 Appendix

Figure 11 shows the results of the optimized boat calibration, along with the residual difference between the acoustically derived and GPS-derived azimuths derived from boat noise. Only the narrow bandwidth between 350 and 400 Hz was available for boat noise calibrations due to the strong interference of overlapping humpback song. The median absolute difference between the GPS and acoustic estimates is 7.61°, and so the uncertainty of the acoustic bearings was set to ±7.5° in the azigram thresholding (azimuthal sector width, $d\varphi = 15°$).
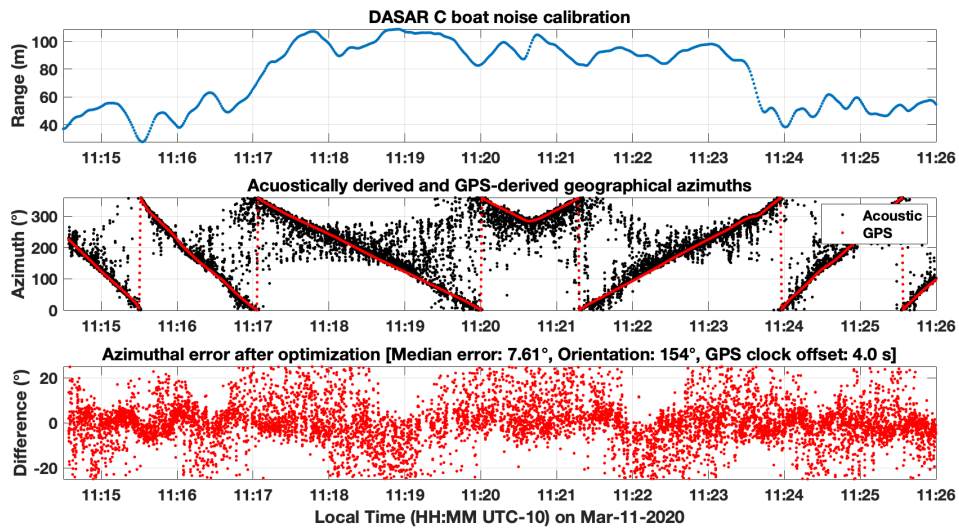
Figure 4.11: Boat noise calibration for DASAR C. GPS-derived range of boat (top), acoustically derived and GPS-derived geographic azimuths (middle) and azimuthal error, after optimization (bottom). The acoustically derived azimuths were estimated between 350 and 400 Hz. The optimization results show that the median error of the DASARs is 7.61°. For DASAR C, the x-velocity axis is 154° clockwise relative to true north and the recorder timing offset from GPS is 4 seconds.

Chapter 5

Conclusion

All three research chapters of this dissertation relate to passive acoustic localization and tracking using arrays and directional sensors and emphasize the application of these methods to signals of biological origin. Below are summarized the key points and findings of each chapter.

5.1 Double-difference localization for long-range tracking of acoustic sources

Initially developed in the field of seismology for high-precision earthquake localization, the concepts of the double-difference method (Waldhauser and Ellsworth, 2000) were reformulated in Chapter 2 for long-range tracking of acoustic sources using multipath arrivals measured on a single vertical array of sensors (as opposed to its original application to widely-distributed sensor networks). The improvement in precision of this approach relies on reducing systemic errors from mismatched modeling of the propagation environment, as well as array tilt.

Based on an approximate initial location in cylindrical coordinates (absolute depth and range from the array), this approach allows obtaining a high-precision relative depth and range trajectories of the source. Application of this method to a towed acoustic source, located 50 km in away from the array, yielded accurate relative trajectories that were precise to within a few meters in depth and up to 100 m in range. Both the accuracy and precision of the results were improved in comparison to localizing the source using a ray-trace model.

The reformulated vertical array double-difference was also applied to a sperm whale, which was tracked in both range and depth. Ray-trace modeling indicated that the animal was close to 50 km in range and about 300 m in depth. During the 25 min time window over which the whale was tracked, the high-precision double-difference results revealed the animal's range increased steadily by 2 km, while holding its depth to within about 20 m.

5.2 Time-synchronizing clocks from independent sensors using directional ambient noise

As mentioned in the introduction, time-synchronization can be a major hurdle for coherent processing when using independent autonomous acoustic recorders. In Chapter 3, it was shown that the ambient noise in Laguna San Igancio, Mexico, allows time-synchronizing independent autonomous acoustic recorders over a period of 10 days to within ± 1 ms. Measuring the relative clock drift between the instruments was achieved by tracking changes in the structure of their ambient noise cross-correlation function over time. The main advantage of this passive approach is that it allows observing small-scale non-linear fluctuations of clock drift, which other time-synchronization schemes would not permit. The technique relies on the assumption that the spatial distribution of coherent ambient noise sources remains constant. An interesting result of this work is that the coherent ambient noise that permits the time-synchronization is believed to originate from croaker fish, that are particularly vocal in the evenings. The statistical stationarity of the fish is confirmed by performing an independent calibration using signals from a known location to measure the accuracy of the time-synchronization results.

5.3 Simultaneous tracking of multiple singing humpback whales using vector sensors

Chapter 4 presents a method for multi-target localization and tracking (in latitude and longitude) of acoustic sources using vector sensors. The technique, which relies on vector sensors' ability to measure directional quantities of the acoustic field, is demonstrated on simultaneously singing humpback whales off western Maui. Extraction of azimuthal trajectories is made possible by displaying the data from each vector sensor as histograms of the acoustic field's azimuthal distribution over time, using the normalized transport velocity metric to restrict the measurements to those arising from discrete compact sources. The azimuthal tracks are matched between vector sensors for subsequent localization by cross-correlating time-frequency features of the humpback whales' songs. Localization is performed by triangulating the linked tracks between the sensors.

The performance of this vector sensor tracking approach shows potential for being applied to other acoustic sources such as boats. The azimuthal histogram display also provides a new visualization method for vector sensor data in which different aspects of the directionality of the acoustic field can be observed. The use of the normalized transport velocity provides another potential avenue for detecting and classifying transient acoustic signals which would not require estimating the signal-to-noise ratio (SNR) of the transient.

Bibliography

Aoki, K., Amano, M., Sugiyama, N., Muramoto, H., Suzuki, M., Yoshioka, M., Mori, K., Tokuda, D., Miyazaki, N., and Ieee (**2007**). "Measurement of swimming speed in sperm whales," 2007 Symposium on Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies, Vols 1 and 2, 467-471.

Au, W., Pack, A., Lammers, M., Herman, L., Deakos, M., and Andrews, K. (**2006**). "Acoustic properties of humpback whale songs," Journal of the Acoustical Society of America **120**, 1103-1110.

Au, W. W. L., Mobley, J., Burgess, W. C., Lammers, M. O., and Nachtigall, P. E. (**2000**). "Seasonal and diurnal trends of chorusing humpback whales wintering in waters off western Maui," Marine Mammal Science **16**, 530-544.

Baggenstoss, P. M. (**2011**). "An algorithm for the localization of multiple interfering sperm whales using multi-sensor time difference of arrival," Journal of the Acoustical Society of America **130**, 102-112.

Baggeroer, A. B., Kuperman, W. A., and Mikhalevsky, P. N. (**1993**). "An overview of matched field methods in ocean acoustics," Ieee Journal of Oceanic Engineering **18**, 401-424.

Barjau Gonzalez, E. (**2008**). *Estructura De La Ictiofauna Asociada a Fondos Blandos En Laguna San Ignacio, Baja California Sur,Mexico*.

Baumgartner, M. F., Fratantoni, D. M., Hurst, T. P., Brown, M. W., Cole, T. V. N., Van Parijs, S. M., and Johnson, M. (**2013**). "Real-time reporting of baleen whale passive acoustic detections from ocean gliders," Journal of the Acoustical Society of America **134**, 1814-1823.

Blackwell, S. B., Nations, C. S., McDonald, T. L., Thode, A. M., Mathias, D., Kim, K. H., Greene, C. R., and Macrander, A. M. (**2015**). "Effects of Airgun Sounds on Bowhead Whale Calling Rates: Evidence for Two Behavioral Thresholds," Plos One **10**, 29.

Blackwell, S. B., Richardson, W. J., Greene, C. R., and Streever, B. (**2007**). "Bowhead whale (Balaena mysticetus) migration and calling behaviour in the Alaskan Beaufort sea, Autumn 2001-04: An acoustic localization study," Arctic **60**, 255-270.

Bonnel, J., and Chapman, N. R. (**2011**). "Geoacoustic inversion in a dispersive waveguide using warping operators," Journal of the Acoustical Society of America **130**, EL101-EL107.

Bonnel, J., Thode, A. M., Blackwell, S. B., Kim, K., and Macrander, A. M. (**2014**). "Range estimation of bowhead whale (Balaena mysticetus) calls in the Arctic using a single hydrophone," Journal of the Acoustical Society of America **136**, 145-155.

Bouffaut, L., Landrø, M., and Potter, J. R. (**2021**). "Source level and vocalizing depth estimation of two blue whale subspecies in the western Indian Ocean from single sensor observations," The Journal of the Acoustical Society of America **149**, 4422-4436.

Brooks, L. A., and Gerstoft, P. (**2007**). "Ocean acoustic interferometry," Journal of the Acoustical Society of America **121**, 3377-3385.

Butler, J., Pagniello, C., Jaffe, J. S., Parnell, P. E., and Sirovic, A. (**2021**). "Diel and Seasonal Variability in Kelp Forest Soundscapes Off the Southern California Coast," Frontiers in Marine Science **8**, 14.

Campillo, M. (**2006**). "Phase and correlation in 'Random' seismic fields and the reconstruction of the green function," Pure and Applied Geophysics **163**, 475-502.

Campillo, M., and Paul, A. (**2003**). "Long-range correlations in the diffuse seismic coda," Science **299**, 547-549.

Cato, D. H. (**1998**). "Simple methods of estimating source levels and locations of marine animal sounds," Journal of the Acoustical Society of America **104**, 1667-1678.

Cholewiak, D. M., Cerchio, S., Jacobsen, J. K., Urban-R, J., and Clark, C. W. (**2018**). "Songbird dynamics under the sea: acoustic interactions between humpback whales suggest song mediates male interactions," Royal Society Open Science **5**, 14.

Collins, M. D., Kuperman, W. A., and Schmidt, H. (**1992**). "Nonlinear inversion for ocean-bottom properties," Journal of the Acoustical Society of America **92**, 2770-2783.

Compton, R., Goodwin, L., Handy, R., and Abbott, V. (**2008**). "A critical examination of worldwide guidelines for minimising the disturbance to marine mammals during seismic surveys," Marine Policy **32**, 255-262.

Cornuelle, B. D. (**1985**). "Simulations of acoustic tomography array performance with untracked or drifting sources and receivers," Journal of Geophysical Research-Oceans **90**, 9079-9088.

Cox, H., Zeskind, R. M., and Owen, M. M. (**1987**). "Robust adaptive beamforming," Ieee Transactions on Acoustics Speech and Signal Processing **35**, 1365-1376.

D'Spain, G. L., Hodgkiss, W. S., and Edmonds, G. L. (**1991**). "Energetics of the deep oceans infrasonic sound field," Journal of the Acoustical Society of America **89**, 1134-1158.

D'Spain, G. L., Luby, J. C., Wilson, G. R., and Gramann, R. A. (**2006**). "Vector sensors and vector sensor line arrays: Comments on optimal array gain and detection," Journal of the Acoustical Society of America **120**, 171-185.

Dahlheim, M. E. (**1987**). "Bio-acoustics of the gray whale (Eschrichtius robustus)."

Darling, J. D., Jones, M. E., and Nicklin, C. P. (**2006**). "Humpback whale songs: Do they organize males during the breeding season?," Behaviour **143**, 1051-1101.

Debever, C., and Kuperman, W. A. (**2007**). "Robust matched-field processing using a coherent broadband white noise constraint processor," Journal of the Acoustical Society of America **122**, 1979-1986.

Dolman, S. J., Weir, C. R., and Jasny, M. (**2009**). "Comparative review of marine mammal guidance implemented during naval exercises," Marine Pollution Bulletin **58**, 465-477.

Erisman, B. E., and Rowell, T. J. (**2017**). "A sound worth saving: acoustic characteristics of a massive fish spawning aggregation," Biology Letters **13**, 5.

Evensen, G. (**2009**). *Data Assimilation: The Ensemble Kalman Filter, Second Edition* (Springer, Berlin).

Fizell, R. G., and Wales, S. C. (**1985**). "Source localization in range and depth in an Arctic environment," The Journal of the Acoustical Society of America **78**, S57-S58.

Frankel, A. S., Clark, C. W., Herman, L. M., and Gabriele, C. M. (**1995**). "Spatial distribution, habitat utilization, and social interactions of humpback whales, *Megaptera novaeangliae*, off Hawaii determined using acoustic and visual techniques," Canadian Journal of Zoology-Revue Canadienne De Zoologie **73**, 1134-1146.

Fremont, M. J., and Malone, S. D. (**1987**). "High precision Relative locations of earthquakes at Mount St. Helens, Washington," Journal of Geophysical Research-Solid Earth and Planets **92**, 10223-10236.

Frisk, G. V. (**1994**). *Ocean and seabed acoustics : a theory of wave propagation* (Prentice Hall, Englewood Cliffs, N.J.).

Gassmann, M., Wiggins, S. M., and Hildebrand, J. A. (**2015**). "Three-dimensional tracking of Cuvier's beaked whales' echolocation sounds using nested hydrophone arrays," Journal of the Acoustical Society of America **138**, 2483-2494.

Gerstoft, P. (**1994**). "Inversion of seismoacoustic data using genetic algorithms and a posteriori probability distributions," Journal of the Acoustical Society of America **95**, 770-782.

Godin, O. A. (**2018**). "Acoustic noise interferometry in a time-dependent coastal ocean," Journal of the Acoustical Society of America **143**, 595-604.

Goldbogen, J. A., Stimpert, A. K., DeRuiter, S. L., Calambokidis, J., Friedlaender, A. S., Schorr, G. S., Moretti, D. J., Tyack, P. L., and Southall, B. L. (**2014**). "Using accelerometers to determine the calling behavior of tagged baleen whales," Journal of Experimental Biology **217**, 2449-2455.

Goold, J. C., and Jones, S. E. (**1995**). "Time and frequency-domain characteristics of sperm whale clicks," Journal of the Acoustical Society of America **98**, 1279-1291.

Gouedard, P., Seher, T., McGuire, J. J., Collins, J. A., and van der Hilst, R. D. (**2014**). "Correction of Ocean-Bottom Seismometer Instrumental Clock Errors Using Ambient Seismic Noise," Bulletin of the Seismological Society of America **104**, 1276-1288.

Gouedard, P., Stehly, L., Brenguier, F., Campillo, M., de Verdiere, Y. C., Larose, E., Margerin, L., Roux, P., Sanchez-Sesma, F. J., Shapiro, N. M., and Weaver, R. L. (**2008**). "Cross-correlation of random fields: mathematical approach and applications," Geophysical Prospecting **56**, 375-393.

Greene, C. R., McLennan, M. W., Norman, R. G., McDonald, T. L., Jakubczak, R. S., and Richardson, W. J. (**2004**). "Directional frequency and recording (DIFAR) sensors in seafloor recorders to locate calling bowhead whales during their fall migration," Journal of the Acoustical Society of America **116**, 799-813.

Gruden, P., and White, P. R. (**2020**). "Automated extraction of dolphin whistles-A sequential Monte Carlo probability hypothesis density approacha)," Journal of the Acoustical Society of America **148**, 3014-3026.

Guazzo, R. A., Helble, T. A., D'Spain, G. L., Weller, D. W., Wiggins, S. M., and Hildebrand, J. A. (**2017**). "Migratory behavior of eastern North Pacific gray whales tracked using a hydrophone array," Plos One **12**, 29.

Hable, S., Sigloch, K., Barruol, G., Stahler, S. C., and Hadziioannou, C. (**2018**). "Clock errors in land and ocean bottom seismograms: high-accuracy estimates from multiple-component noise cross-correlations," Geophysical Journal International **214**, 2014-2034.

Hannemann, K., Kruger, F., and Dahm, T. (**2014**). "Measuring of clock drift rates and static time offsets of ocean bottom stations by means of ambient noise," Geophysical Journal International **196**, 1034-1042.

Hastie, G. D., Gillespie, D. M., Gordon, J. C. D., Macaulay, J. D. J., McConnell, B. J., and Sparling, C. E. (**2014**). "Tracking Technologies for Quantifying Marine Mammal Interactions with Tidal Turbines: Pitfalls and Possibilities," in *Marine Renewable Energy Technology and Environmental Interactions*, edited by M. A. Shields, and A. I. L. Payne (Springer, Dordrecht), pp. 127-139.

Hawkins, A. D., and Popper, A. N. (**2017**). "A sound approach to assessing the impact of underwater noise on marine fishes and invertebrates," Ices Journal of Marine Science **74**, 635-651.

Helble, T. A., Henderson, E. E., Ierley, G. R., and Martin, S. W. (**2016**). "Swim track kinematics and calling behavior attributed to Bryde's whales on the Navy's Pacific Missile Range Facility," Journal of the Acoustical Society of America **140**, 4170-4177.

Helble, T. A., Ierley, G. R., D'Spain, G. L., and Martin, S. W. (**2015**). "Automated acoustic localization and call association for vocalizing humpback whales on the Navy's Pacific Missile Range Facility," Journal of the Acoustical Society of America **137**, 11-21.

Henderson, E. E., Helble, T. A., Ierley, G. R., and Martin, S. W. (**2018**). "Identifying behavioral states and habitat use of acoustically tracked humpback whales in Hawaii," Marine Mammal Science **34**, 1-17.

Hendricks, B., Keen, E. M., Shine, C., Wray, J. L., Alidina, H. M., and Picard, C. R. (**2021**). "Acoustic tracking of fin whales: Habitat use and movement patterns within a Canadian Pacific fjord system," The Journal of the Acoustical Society of America **149**, 4264-4280.

Herman, L. M. (**2017**). "The multiple functions of male song within the humpback whale (Megaptera novaeangliae) mating system: review, evaluation, and synthesis," Biological Reviews **92**, 1795-1818.

Holler, R. A. (**2014**). "The evolution of the sonobuoy from World War II to the cold war," (Navmar Applied Sciences Corp Warminster PA).

Huang, T., Yang, G., and Tang, G. (**1979**). "A fast two-dimensional median filtering algorithm," IEEE Transactions on Acoustics, Speech, and Signal Processing **27**, 13-18.

Ito, A. (**1985**). "High-resolution relative hypocenters of similar earthquakes by cross-spectral analysis method," Journal of Physics of the Earth **33**, 279-294.

Jensen, F. B., Kuperman, W. A., Porter, M. B., and Schmidt, H. (**2011**). "Computational Ocean Acoustics, Second Edition," Computational Ocean Acoustics, Second Edition, 1-794.

Klinck, H., Mellinger, D. K., Klinck, K., Bogue, N. M., Luby, J. C., Jump, W. A., Shilling, G. B., Litchendorf, T., Wood, A. S., Schorr, G. S., and Baird, R. W. (**2012**). "Near-Real-Time Acoustic Monitoring of Beaked Whales and Other Cetaceans Using a Seaglider (TM)," Plos One **7**, 8.

Kline, L. R., DeAngelis, A. I., McBride, C., Rodgers, G. G., Rowell, T. J., Smith, J., Stanley, J. A., Read, A. D., and Van Parijs, S. M. (**2020**). "Sleuthing with sound: Understanding vessel activity in marine protected areas using passive acoustic monitoring," Marine Policy **120**, 13.

Knapp, C., and Carter, G. (**1976**). "The generalized correlation method for estimation of time delay," IEEE Transactions on Acoustics, Speech, and Signal Processing **24**, 320-327.

Kowarski, K. A., Gaudet, B. J., Cole, A. J., Maxner, E. E., Turner, S. P., Martin, S. B., Johnson, H. D., and Moloney, J. E. (**2020**). "Near real-time marine mammal monitoring from gliders: Practical challenges, system development, and management implications," The Journal of the Acoustical Society of America **148**, 1215-1230.

Ladich, F. (**2019**). "Ecology of sound communication in fishes," Fish and Fisheries **20**, 552-563.

Lenth, R. V. (**1981**). "On finding the source of a signal," Technometrics **23**, 149-154.

Lindsay, C. E., and Chapman, N. R. (**1993**). "Matched-field inversion for geoacoustic model parameters using adaptive simulated annealing," Ieee Journal of Oceanic Engineering **18**, 224-231.

Lobkis, O. I., and Weaver, R. L. (**2001**). "On the emergence of the Green's function in the correlations of a diffuse field," Journal of the Acoustical Society of America **110**, 3011-3017.

Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Baranova, O. K., Zweng, M. M., Paver, C. R., Reagan, J. R., Johnson, D. R., Hamilton, M., Seidov, D., Levitus, S., National Oceanographic Data Center (U.S.). Ocean Climate Laboratory, and United States. National Environmental Satellite Data and Information Service (**2013**). *World ocean atlas 2013. Volume 1, Temperature* (National Oceanographic Data Center, Ocean Climate Laboratory, and United States National Environmental Satellite Data and Information Service, Silverspring, MD).

Locascio, J. V., and Mann, D. A. (**2011**). "Diel and seasonal timing of sound production by black drum (Pogonias cromis)," Fishery Bulletin **109**, 327-338.

Mackenzie, K. V. (**1981**). "Nine-term equation for sound speed in the oceans," The Journal of the Acoustical Society of America **70**, 807-812.

Malcolm, A. E., Scales, J. A., and van Tiggelen, B. A. (**2004**). "Extracting the Green function from diffuse, equipartitioned waves," Physical Review E **70**, 4.

Mann, J. A., Tichy, J., and Romano, A. J. (**1987**). "Instantaneous and time-averaged energy-transfer in acoustic fields," Journal of the Acoustical Society of America **82**, 17-30.

Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Moretti, D. J., Harris, D., and Tyack, P. L. (**2013**). "Estimating animal population density using passive acoustics," Biological Reviews **88**, 287-309.

Martin, B., Zeddies, D. G., Gaudet, B., and Richard, J. (**2016**). "Evaluation of Three Sensor Types for Particle Motion Measurement," Effects of Noise on Aquatic Life Ii **875**, 679-686.

Mathias, D., Thode, A. M., Straley, J., and Andrews, R. D. (**2013**). "Acoustic tracking of sperm whales in the Gulf of Alaska using a two-element vertical array and tags," Journal of the Acoustical Society of America **134**, 2446-2461.

McDonald, M. A. (**2004**). "DIFAR hydrophone usage in whale research," Canadian Acoustics **32**, 155-160.

Mellinger, D. K., and Clark, C. W. (**2000**). "Recognizing transient low-frequency whale sounds by spectrogram correlation," Journal of the Acoustical Society of America **107**, 3518-3529.

Meyer, F., Kropfreiter, T., Williams, J. L., Lau, R. A., Hlawatsch, F., Braca, P., and Win, M. Z. (**2018**). "Message Passing Algorithms for Scalable Multitarget Tracking," Proceedings of the Ieee **106**, 221-259.

Miller, P. J. O., Johnson, M. P., Tyack, P. L., and Terray, E. A. (**2004**). "Swimming gaits, passive drag and buoyancy of diving sperm whales Physeter macrocephalus," Journal of Experimental Biology **207**, 1953-1967.

Mouy, X., Hannay, D., Zykov, M., and Martin, B. (**2012**). "Tracking of Pacific walruses in the Chukchi Sea using a single hydrophone," Journal of the Acoustical Society of America **131**, 1349-1358.

Mouy, X., Rountree, R., Juanes, F., and Dosso, S. E. (**2018**). "Cataloging fish sounds in the wild using combined acoustic and video recordings," Journal of the Acoustical Society of America **143**, EL333-EL339.

Munk, W., Worcester, P., and Wunsch, C. (**2009**). *Ocean acoustic tomography* (Cambridge university press).

Nedelec, S. L., Campbell, J., Radford, A. N., Simpson, S. D., and Merchant, N. D. (**2016**). "Particle motion: the missing link in underwater acoustic ecology," Methods in Ecology and Evolution **7**, 836-842.

Newhall, A. E., Lin, Y. T., Lynch, J. F., Baumgartner, M. F., and Gawarkiewicz, G. G. (**2012**). "Long distance passive localization of vocalizing sei whales using an acoustic normal mode approach," Journal of the Acoustical Society of America **131**, 1814-1825.

Noad, M. J., and Cato, D. H. (**2007**). "Swimming speeds of singing and non-singing humpback whales during migration," Marine Mammal Science **23**, 481-495.

Noad, M. J., Cato, D. H., and Stokes, M. D. (**2004**). "Acoustic tracking of humpback whales: measuring interactions with the acoustic environment," in *Proceedings of ACOUSTICS*, pp. 353-358.

Payne, R. S., and McVay, S. (**1971**). "Songs of humpback whales," Science **173**, 585-597.

Popper, A. N., Salmon, M., and Horch, K. W. (**2001**). "Acoustic detection and communication by decapod crustaceans," Journal of Comparative Physiology a-Neuroethology Sensory Neural and Behavioral Physiology **187**, 83-89.

Poupinet, G., Ellsworth, W. L., and Frechet, J. (**1984**). "Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras fault, California," Journal of Geophysical Research **89**, 5719-5731.

Press, W. H. (**1992**). *Numerical recipes in FORTRAN : the art of scientific computing* (Cambridge University Press, Cambridge).

Raghukumar, K., Chang, G., Spada, F., and Jones, C. (**2020**). "A Vector Sensor-Based Acoustic Characterization System for Marine Renewable Energy," Journal of Marine Science and Engineering **8**, 15.

Ren, Q. Y., and Hermand, J. P. (**2013**). "Acoustic interferometry for geoacoustic characterization in a soft-layered sediment environment," Journal of the Acoustical Society of America **133**, 82-93.

Rice, A. N., Soldevilla, M. S., and Quinlan, J. A. (**2017**). "Nocturnal patterns in fish chorusing off the coasts of Georgia and eastern Florida," Bulletin of Marine Science **93**, 455-474.

Sabra, K. G., Gerstoft, P., Roux, P., Kuperman, W. A., and Fehler, M. C. (**2005a**). "Extracting time-domain Green's function estimates from ambient seismic noise," Geophysical Research Letters **32**, 5.

Sabra, K. G., Roux, P., and Kuperman, W. A. (**2005b**). "Emergence rate of the time-domain Green's function from the ambient noise cross-correlation function," Journal of the Acoustical Society of America **118**, 3524-3531.

Sabra, K. G., Roux, P., Thode, A. M., D'Spain, G. L., Hodgkiss, W. S., and Kuperman, W. A. (**2005c**). "Using ocean ambient noise for array self-localization and self-synchronization," Ieee Journal of Oceanic Engineering **30**, 338-347.

Schau, H. C., and Robinson, A. Z. (**1987**). "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," Ieee Transactions on Acoustics Speech and Signal Processing **35**, 1223-1225.

Schmidt, V., Weber, T. C., Wiley, D. N., and Johnson, M. P. (**2010**). "Underwater Tracking of Humpback Whales (Megaptera Novaeangliae) With High-Frequency Pingers and Acoustic Recording Tags," Ieee Journal of Oceanic Engineering **35**, 821-836.

Seger, K. D., Thode, A. M., Swartz, S. L., and Urban, J. (**2015**). "The ambient acoustic environment in Laguna San Ignacio, Baja California Sur, Mexico," Journal of the Acoustical Society of America **138**, 3397-3410.

Seger, K. D., Thode, A. M., Urban, J., Martinez-Loustalot, P., Jimenez-Lopez, M. E., and Lopez-Arzate, D. (**2016**). "Humpback whale-generated ambient noise levels provide insight into singers' spatial densities," Journal of the Acoustical Society of America **140**, 1581-1597.

Sens-Schonfelder, C. (**2008**). "Synchronizing seismic networks with ambient noise," Geophysical Journal International **174**, 966-970.

Shearer, P. M. (**1997**). "Improving local earthquake locations using the L1 norm and waveform cross correlation: Application to the Whittier Narrows, California, aftershock sequence," Journal of Geophysical Research-Solid Earth **102**, 8269-8283.

Shearer, P. M. (**1999**). *Introduction to seismology* (Cambridge University Press, Cambridge ; New York).

Simard, Y., and Roy, N. (**2008**). "Detection and localization of blue and fin whales from large-aperture autonomous hydrophone arrays: A case study from the St. Lawrence estuary," Canadian Acoustics **36**, 104-110.

Simard, Y., Roy, N., and Gervaise, C. (**2008**). "Passive acoustic detection and localization of whales: Effects of shipping noise in Saguenay-St. Lawrence Marine Park," Journal of the Acoustical Society of America **123**, 4109-4117.

Sirovic, A., Hildebrand, J. A., and Wiggins, S. M. (**2007**). "Blue and fin whale call source levels and propagation range in the Southern Ocean," Journal of the Acoustical Society of America **122**, 1208-1215.

Skolnik, M. I. (**1962**). *Introduction to radar systems* (McGraw-Hill, New York).

Slabbekoorn, H., Bouton, N., van Opzeeland, I., Coers, A., ten Cate, C., and Popper, A. N. (**2010**). "A noisy spring: the impact of globally rising underwater sound levels on fish," Trends in Ecology & Evolution **25**, 419-427.

Sorensen, E., Ou, H. H., Zurk, L. M., Siderius, M., and Ieee (**2010**). "Passive Acoustic Sensing for Detection of Small Vessels," Oceans 2010, 8.

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

Spiesberger, J. L. (**2001**). "Hyperbolic location errors due to insufficient numbers of receivers," Journal of the Acoustical Society of America **109**, 3076-3079.

Spiesberger, J. L., and Fristrup, K. M. (**1990**). "Passive Localization of Calling Animals and Sensing of their Acoustic Environment Using Acoustic Tomography," American Naturalist **135**, 107-153.

Stanistreet, J. E., Risch, D., and Van Parijs, S. M. (**2013**). "Passive Acoustic Tracking of Singing Humpback Whales (Megaptera novaeangliae) on a Northwest Atlantic Feeding Ground," Plos One **8**, 10.

Stehly, L., Campillo, M., and Shapiro, N. M. (**2007**). "Traveltime measurements from noise correlation: stability and detection of instrumental time-shifts," Geophysical Journal International **171**, 223-230.

Stimpert, A. K., Lammers, M. O., Pack, A. A., and Au, W. W. L. (**2020**). "Variations in received levels on a sound and movement tag on a singing humpback whale: Implications for caller identification," Journal of the Acoustical Society of America **147**, 3684-3690.

Thode, A. (**2004**). "Tracking sperm whale (Physeter macrocephalus) dive profiles using a towed passive acoustic array," Journal of the Acoustical Society of America **116**, 245-253.

Thode, A., Bonnel, J., Thieury, M., Fagan, A., Verlinden, C., Wright, D., Berchok, C., and Crance, J. (**2017**). "Using nonlinear time warping to estimate North Pacific right whale calling depths in the Bering Sea," Journal of the Acoustical Society of America **141**, 3059-3069.

Thode, A., Mellinger, D. K., Stienessen, S., Martinez, A., and Mullin, K. (**2002**). "Depth-dependent acoustic features of diving sperm whales (Physeter macrocephalus) in the Gulf of Mexico," Journal of the Acoustical Society of America **112**, 308-321.

Thode, A. M., Blackwell, S. B., Seger, K. D., Conrad, A. S., Kim, K. H., and Macrander, A. M. (**2016**). "Source level and calling depth distributions of migrating bowhead whale calls in the shallow Beaufort Sea," Journal of the Acoustical Society of America **140**, 4288-4297.

Thode, A. M., Conrad, A. S., Ozanich, E., King, R., Freeman, S. E., Freeman, L. A., Zgliczynski, B., Gerstoft, P., and Kim, K. H. (**2021**). "Automated two-dimensional localization of underwater acoustic transient impulses using vector sensor image processing (vector sensor localization)," Journal of the Acoustical Society of America **149**, 770-787.

Thode, A. M., D'Spain, G. L., and Kuperman, W. A. (**2000**). "Matched-field processing, geoacoustic inversion, and source signature recovery of blue whale vocalizations," Journal of the Acoustical Society of America **107**, 1286-1300.

Thode, A. M., Gerstoft, P., Burgess, W. C., Sabra, K. G., Guerra, M., Stokes, M. D., Noad, M., and Cato, D. H. (**2006**). "A portable matched-field processing system using passive acoustic time synchronization," Ieee Journal of Oceanic Engineering **31**, 696-710.

Thode, A. M., Kim, K. H., Blackwell, S. B., Greene, C. R., Nations, C. S., McDonald, T. L., and Macrander, A. M. (**2012**). "Automated detection and localization of bowhead whale sounds in the presence of seismic airgun surveys," Journal of the Acoustical Society of America **131**, 3726-3747.

Thode, A. M., Sakai, T., Michalec, J., Rankin, S., Soldevilla, M. S., Martin, B., and Kim, K. H. (**2019**). "Displaying bioacoustic directional information from sonobuoys using "azigrams"," Journal of the Acoustical Society of America **146**, 95-102.

Tiemann, C. O., Thode, A. M., Straley, J., O'Connell, V., and Folkert, K. (**2006**). "Three-dimensional localization of sperm whales using a single hydrophone," Journal of the Acoustical Society of America **120**, 2355-2365.

Van Parijs, S. M., Clark, C. W., Sousa-Lima, R. S., Parks, S. E., Rankin, S., Risch, D., and Van Opzeeland, I. C. (**2009**). "Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales," Marine Ecology Progress Series **395**, 21-36.

Van Uffelen, L. J., Howe, B. M., Nosal, E. M., Carter, G. S., Worcester, P. F., and Dzieciuch, M. A. (**2016**). "Localization and Subsurface Position Error Estimation of Gliders Using Broadband Acoustic Signals at Long Range," Ieee Journal of Oceanic Engineering **41**, 501-508.

Varga, L. M., Wiggins, S. M., and Hildebrand, J. A. (**2018**). "Behavior of singing fin whales Balaenoptera physalus tracked acoustically offshore of Southern California," Endangered Species Research **35**, 113-124.

Vermeij, M. J. A., Marhaver, K. L., Huijbers, C. M., Nagelkerken, I., and Simpson, S. D. (**2010**). "Coral Larvae Move toward Reef Sounds," Plos One **5**, 4.

Vo, B. N., Vo, B. T., Pham, N. T., and Suter, D. (**2010**). "Joint Detection and Estimation of Multiple Objects From Image Observations," Ieee Transactions on Signal Processing **58**, 5129-5141.

Waldhauser, F., and Ellsworth, W. L. (**2000**). "A double-difference earthquake location algorithm: Method and application to the northern Hayward fault, California," Bulletin of the Seismological Society of America **90**, 1353-1368.

Wapenaar, K. (**2004**). "Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation," Physical Review Letters **93**, 4.

Watkins, W. A., and Schevill, W. E. (**1972**). "Sound source location by arrival-times on a non-rigid 3-dimensional hydrophone array," Deep-Sea Research **19**, 691-706.

Weaver, R. L., and Lobkis, O. I. (**2001**). "Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies," Physical Review Letters **87**, 4.

Weaver, R. L., and Lobkis, O. I. (**2005**). "Fluctuations in diffuse field-field correlations and the emergence of the Green's function in open systems," Journal of the Acoustical Society of America **117**, 3432-3439.

Wiggins, S. M., and Hildebrand, J. A. (**2020**). "Fin whale 40-Hz calling behavior studied with an acoustic tracking array," Marine Mammal Science **36**, 964-971.

Wilcock, W. (**2012**). "Tracking fin whales in the northeast Pacific Ocean with a seafloor seismic network," Journal of the Acoustical Society of America **132**, 2408-2419.

Zimmer, W. M. X., Johnson, M. P., D'Amico, A., and Tyack, P. L. (**2003**). "Combining data from a multisensor tag and passive sonar to determine the diving behavior of a sperm whale (Physeter macrocephalus)," Ieee Journal of Oceanic Engineering **28**, 13-28.