

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Passive Acoustic Localization and Tracking in Marine Environments

Permalink

<https://escholarship.org/uc/item/8v27f19g>

ISBN

9798293839230

Author

Jang, Junsu

Publication Date

2025-09-16

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Passive Acoustic Localization and Tracking in Marine Environments

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

in

Oceanography

by

Junsu Jang

Committee in charge:

Florian Meyer, Chair
Nikolay A. Atanasov
Simone Baumann-Pickering
William S. Hodgkiss
Aaron M. Thode

2025

Copyright

Junsu Jang, 2025

All rights reserved.

The Dissertation of Junsu Jang is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2025

TABLE OF CONTENTS

Dissertation Approval Page	iii
Table of Contents	iv
List of Figures	viii
List of Tables	xii
Acknowledgements	xiii
Vita	xv
Abstract of the Dissertation	xvi
Introduction	1
Chapter 1 Detection and Directional Processing of Odontocete Echolocation Clicks Using Beamforming and TDOA Methods	4
1.1 Introduction	4
1.1.1 State-of-the-art	6
1.1.2 Contribution	8
1.2 Signal Model	9
1.2.1 Time Domain Signal Model	9
1.2.2 Frequency Domain Signal Model	10
1.3 Beamforming	11
1.3.1 Conventional Beamforming	11
1.3.2 Frequency Difference Beamforming	13
1.4 Time Difference of Arrival	15
1.4.1 Cross-correlation	15
1.4.2 Generalized Cross-correlation	16
1.4.3 DOA Computation from TDOA Measurements	17
1.5 Dataset	19
1.5.1 Noise Sources and Pre-processing	19
1.6 Implementation Aspects	21
1.6.1 Beamforming	21
1.6.2 TDOA: GCC-WIN	21
1.7 Methods	23
1.7.1 Overview of Methods Compared	23
1.7.2 Detection Thresholding	24
1.7.3 DOA Estimation Methods Comparison	25
1.7.4 TDOA Detection Rate: Low SNR Case Study	26
1.7.5 TDOA Detection Rate: Simulation Study	26
1.7.6 Information Loss from TDOA-to-DOA Conversion	27

1.8	Results	28
1.8.1	DOA Estimates	28
1.8.2	TDOA Detection Rate Analysis	30
1.8.3	Simulation Study	30
1.8.4	Information Loss in TDOA-to-DOA Conversion	32
1.9	Conclusions	35
Chapter 2	Bayesian Tracking of Odontocetes in 3-D from Their Echolocation Clicks	38
2.1	Introduction	38
2.1.1	Automated 3-D Tracking of Whales: Prior Work	40
2.1.2	Contributions, Chapter Organization, and Notation	43
2.2	Multi-target Tracking	45
2.2.1	MTT With Perfect Measurement-to-target Associations	45
2.2.2	MTT With MOU and Known Number of Targets	46
2.2.3	MTT With MOU and Unknown Number of Targets	49
2.3	The Proposed Data Processing Chain	52
2.3.1	Echolocation Click Detection and TDOA Estimates	53
2.3.2	MTT of Echolocating Odontocetes	54
2.4	Real Data Application	58
2.4.1	Implementation Aspects	58
2.4.2	Results	60
2.5	Simulation Studies	63
2.6	Discussion	69
2.7	Conclusion	71
Chapter 3	Ranging of a Moving Ship Using a Single Acoustic Receiver in Shallow Water	73
3.1	Introduction	73
3.1.1	Waveguide Invariant	74
3.1.2	Ship Sound Characteristics	77
3.1.3	Contribution and Chapter Organization	77
3.2	Waveguide Invariant-based Ranging	79
3.2.1	Spectrogram to Range-Frequency Surface Plot	80
3.2.2	Parameter Space for WI-Based Ranging	80
3.2.3	Estimating the WI and the Range Rate	82
3.2.4	Method Overview	83
3.3	Signal Model	84
3.3.1	Model Overview and Notation	84
3.3.2	Received Signal Model	86
3.3.3	Statistical Model of Acoustic Intensity	87
3.3.4	The Green's Function (Channel Model)	88
3.3.5	Signal Model in Striation-Frequency Domain	89
3.4	The Proposed Range Estimation Method	90
3.4.1	Broadband-Plus-Noise Variance Estimation	92

3.4.2	Scaling Factor Estimation	94
3.4.3	Broadband and Tonal Components Parameter Estimation	95
3.5	Dataset: SBCEX17	97
3.5.1	Event Overview	97
3.5.2	Acoustic Data Processing	98
3.5.3	Potential Acoustic Interference	99
3.6	Implementation Aspects and Comparison Methods	100
3.6.1	Implementation Aspects	100
3.6.2	Comparison to Existing Methods	102
3.7	Results	104
3.8	Conclusions and Future Work	106
Chapter 4	Mode-pairwise Waveguide Invariant Analysis in the New England Mud Patch	109
4.1	Introduction	109
4.1.1	Contributions and Chapter Organization	110
4.2	Background	112
4.2.1	Modal Interference and Striation Formation	112
4.2.2	Waveguide Invariant	113
4.2.3	Mode-pairwise Waveguide Invariant	114
4.3	Simulation and Validation Against Observed Modal Interference Patterns	115
4.3.1	Method	115
4.3.2	Results	117
4.4	Mode-pairwise WI Analysis	122
4.4.1	Method	122
4.4.2	Spectrogram Processing and WI Estimation Implementation	124
4.4.3	Results	126
4.5	Conclusions	131
Appendix A	Theoretical Justification and Derivation of the Statistical Parameter Estimators	133
A.1	Background	134
A.2	Broadband-plus-noise Variance Estimation	135
A.2.1	Broadband Scale Parameter α_k	136
A.2.2	Received Broadband Signal Variance $\underline{v}_{l,k}$	136
A.2.3	Broadband-plus-noise Variance $\underline{\sigma}_{l,k}^2$	137
A.3	Noncentrality Parameter Estimation	138
Appendix B	Correlation-based Ranging of a Ship in Shallow Water Using a Single Acoustic Receiver	142
Appendix C	Other Ship Ranging Results	145
C.1	Scenario 2: ZIM NEW YORK	145
C.2	Scenario 3: TEAM CORRIDO	147

Bibliography	150
--------------------	-----

LIST OF FIGURES

Figure 1.1.	Schematic of the passive acoustic localization setup using a volumetric hydrophone array. The array consists of four hydrophones located at positions $\mathbf{q}^{(1)}$ through $\mathbf{q}^{(4)}$ forming a small 3-D aperture as indicated in the dashed box.	6
Figure 1.2.	(a) A snapshot of time-domain signals recorded across the four hydrophones, capturing an echolocation click from a goose-beaked whale (<i>Ziphius cavirostris</i>). Each colored-trace corresponds to a different sensor.	12
Figure 1.3.	An example spectrogram of acoustic data from a single hydrophone with echolocation clicks and noise. The identified noise sources are the acoustic Doppler current profiler (ADCP), the instrument noise from the data storage system, and the undersea environment.	20
Figure 1.4.	An example spectrogram of the instrument noise that repeats periodically, recorded by a hydrophone on a HARP EE on July 1 st , 2018.	22
Figure 1.5.	Example of a simulated odontocete echolocation click signal embedded in real background noise.	27
Figure 1.6.	Elevation angle estimates obtained using five different DOA estimation methods applied to the same recordings of goose-beaked whales' echolocation clicks on July 1 st , 2018.	29
Figure 1.7.	Comparison of TDOA estimates between hydrophone sensors 3 and 4 using two methods: traditional cross-correlation (XCORR, red circles) and the proposed GCC-WIN (black dots), after applying a 5-median absolute deviation ($5 \times \text{MAD}$) threshold.	31
Figure 1.8.	TDOA estimates between sensors 3 and 4 using (a) standard cross-correlation (XCORR) and (b) the proposed generalized cross-correlation method (GCC-WIN), after applying a 2-median absolute deviation (2MAD) threshold, lower than the threshold used in Fig. 1.7.	32
Figure 1.9.	(a) Average detection rate of echolocation clicks with correct TDOA estimates for three methods (GCC-WIN, XCORR and GCC-PHAT) as a function of signal-to-noise ratio (SNR), based on simulated clicks injected into real noise (see Fig. 1.5).	33
Figure 1.10.	Detected TDOA values (black dots) for all hydrophone pairs using the GCC-WIN method.	34

Figure 1.11.	(a) Elevation and (b) azimuth angle estimates obtained by combining TDOA measurements from all sensor pairs using the GCC-WIN method in Fig. 1.10. Black dots represent DOA estimates for all detections.	35
Figure 2.1.	An example of acoustic source localization in 3-D from TDOA measurements. The cross is the location of the acoustic source. Each pair of dots and triangles represents a hydrophone pair that produces a TDOA measurement.	39
Figure 2.2.	The block diagram of the proposed data processing chain for detecting and tracking odontocetes in 3-D from their echolocation clicks. The TDOAs of echolocation clicks are extracted in the signal processing module.	53
Figure 2.3.	Example TDOA measurements (dots) and corresponding TDOA tracking results (solid lines) from one TDOA sensor of HARP EE (top) and one TDOA sensor of HARP EW (bottom).	61
Figure 2.4.	A comparison between a track generated from the hand-annotated data (dashed line) and estimated tracks provided by the proposed MTT approach (solid lines) in 3-D.	62
Figure 2.6.	Example TDOA measurements (dots) and corresponding TDOA tracking results (solid lines) from one TDOA sensor of HARP EE (top) and one TDOA sensor of HARP EW (bottom).	64
Figure 2.7.	A comparison between the tracks generated from the hand-annotated data (dashed lines) and estimated tracks provided by the proposed MTT approach (solid lines) in 3-D. Two closely-space goose-beaked whales are simultaneously diving into deeper waters.	65
Figure 2.8.	A comparison between the track generated from the hand-annotated data (dashed line) and estimated tracks provided by the proposed MTT approach (solid lines). Each axis of the 3-D domain is shown individually. The data from July 1 st , 2018, are considered.	66
Figure 2.9.	The average tracking RMSE across 200 Monte Carlo runs versus number of simultaneously present odontocetes for three different 3-D tracking approaches. The error bars denote the 75th percentile of the measured RMSE.	67

Figure 2.10.	The average RMSE versus time for three different 3-D tracking approaches. 200 Monte Carlo runs of a scenario with four odontocetes are considered. (SBT) and (MTT) perform sequential processing and thus lead to reduced RMSE values over time.	68
Figure 3.1.	Spectrogram of data recorded on March 24, 2017 (UTC) during SBCEX17 using a single element of a vertical line array. The large commercial ship MSC KALAMATA was moving away from the receiver.	76
Figure 3.2.	Examples of the r - f surface plots with projected striations (red lines) overlaid using a single reference frequency $f' = 45.5$ Hz and Eq. (3.1), for three candidate source ranges, which correspond to the final snapshot of the spectrogram.	79
Figure 3.3.	The joint log-likelihood function $\ell(r, \dot{r}, \beta)$ is evaluated over the parameter spaces: range $r \in \mathcal{S}_r$, range rate $\dot{r} \in \mathcal{S}_{\dot{r}}$, and WI $\beta \in \mathcal{S}_{\beta}$. It is shown as functions of (a) β and $\dot{r}$, (b) r and $\dot{r}$, and (c) r and β . In each plot, the third variable is fixed at its true value.	81
Figure 3.4.	The tracks of the ships between 18:58 and 19:45 on March 24 th , 2017 (UTC), in the vicinity of the VLA2 overlayed on the bathymetry map during the SBCEX17.	97
Figure 3.5.	(a) The spectrogram of the processed acoustic signals over the time- and frequency-window of interest. (b) PSDs of the acoustic recording every 15 minutes after 19:10 on March 24 th , 2017 (UTC) during the considered scenario.	98
Figure 3.6.	Range error (in percent) for five WI-based ranging methods: (G) proposed approach, (S) slope-based method, (T) statistical approach using only tonal signals, (B) statistical approach using only broadband signals, and (C) correlation-based method (see Ch. 3.6.2 for details).	105
Figure 4.1.	The tracks of the ships between 18:32 and 19:45 on March 24 th , 2017 (UTC), in the vicinity of the vertical line array (VLA2) overlayed on the bathymetry map during the Seabed Characterization Experiment 2017 (SBCEX17) [1].	111
Figure 4.2.	Comparison of simulated and measured spectrograms in the 38–50 Hz frequency band, shown as functions of source–receiver (ship–VLA2) range.	118
Figure 4.3.	Modal properties computed from the SC2 geoacoustic model in the 35–55 Hz frequency range. (a) Phase speeds (v_p , dashed blue lines) and group speeds (v_g , solid black lines) for the first three modes. ...	119

Figure 4.4.	Simulated mode-pairwise WI values as a function of frequency for three mode pairs 1-2 (black solid), 2-3 (blue dashed), and 1-3 (red dash-dotted) using geoacoustic environments at site SC2.	120
Figure 4.5.	Sensitivity of the mode-pairwise WI β_{12} for mode pair 1-2 to changes in mud properties for the SC2 at 42 Hz (solid lines) and 49 Hz (dashed lines).	121
Figure 4.6.	Relative dispersion curve (RDC) surface plots derived from acoustic recordings during the ship's transit relative to VLA 2 at three different times, showing prominent vertical lines corresponding to the RDCs of mode pairs 1-2, 2-3, and 1-3.	126
Figure 4.7.	(a) Normalized relative dispersion curve (RDC) surface plot computed from the spectrogram (18:32–19:07, March 24 th , 2017) during the transit of MSC KALAMATA.	128
Figure 4.8.	(a) The RDC surface plot depicting the slope of the RDC corresponding to the mode pair 1-2 in the log-log domain of wavenumber difference and frequency. It is obtained from acoustic recordings between 18:58 and 19:45. The slope is directly related to the WI of the mode pair 1-2.	129
Figure C.1.	Vessel tracks of ZIM NEW YORK (solid red) moving east (right) and MSC KALAMATA (dashed blue) relative to the VLA2 location (red triangle).	146
Figure C.2.	(a) Spectrogram of acoustic data recorded during ZIM NEW YORK's passage, showing striated interference patterns between 32–39 Hz beyond 30 km. (b) Range estimation error as a function of distance using WI = 1.18 and range rate = 9.4 m/s.	147
Figure C.3.	Vessel tracks of TEAM CORRIDO (solid red) moving east (right) and MSC KALAMATA (dashed blue) relative to the VLA2 location (red triangle). TEAM CORRIDO maintained a consistent heading along the same shipping lane during its eastbound transit.	148
Figure C.4.	(a) Spectrogram of acoustic data recorded during TEAM CORRIDO's passage, showing diffuse broadband energy between 42–49 Hz. No strong tonal features are evident.	149

LIST OF TABLES

Table 2.1.	Hyperparameters used for the tracking of multiple goose-beaked whales in the TDOA and the 3-D domain.	59
Table 4.1.	The established parameter space for the simplified geoacoustic model at SC2 [2]. The columns represent the layer thickness h , the sound speed at the top c_t and the bottom c_b of the layer, the density ρ , and the attenuation α	116

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Prof. Florian Meyer, chair of my doctoral committee, for his invaluable guidance, encouragement, and trust throughout my PhD. His mentorship has profoundly shaped the direction and quality of this work. I would also like to thank my committee members — Professors Nikolay Atanasov, Simone Baumann-Pickering, William Hodgkiss and Aaron Thode — for their insightful feedback, encouragement, and support throughout the course of my doctoral research.

Chapters 1 and 2 are adapted and extended in part with permissions from: J. Jang, F. Meyer, E. Snyder, S. Wiggins, S. Baumann-Pickering, and J. Hildebrand, “Bayesian detection and tracking of odontocetes in 3-D from their echolocation clicks,” *J. Acoust. Soc. Am.*, 2023 (Copyright 2023, Acoustical Society of America), and J. Jang, F. Meyer, E. R. Snyder, S. M. Wiggins, S. Baumann-Pickering and J. A. Hildebrand, ”Passive Acoustic Tracking of Whales in 3-D,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, June 2023 (Copyright 2023, IEEE). Some notation and structure have been revised for consistency and clarity within this dissertation. The dissertation author was the primary investigator and lead author of this work.

Chapters 3 and Appendix A are being prepared for submission as a Journal of the Acoustical Society of America manuscript: J. Jang, W. S. Hodgkiss, and F. Meyer, “Ranging of a Moving Ship Using a Single Acoustic Receiver in Shallow Water.” Portions of Chapter 3 are based on an earlier conference paper: J. Jang and F. Meyer, ”A New Statistical Model for Waveguide Invariant-Based Range Estimation in Shallow Water,” in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Oct. 2024 (Copyright 2024, IEEE). The dissertation author was the primary investigator and lead author of this work.

Chapter 4 is being prepared for submission as a JASA Express Letters manuscript: J. Jang and F. Meyer, ”Mode-pairwise Waveguide Invariant Analysis in the New England Mud Patch.” The dissertation author was the primary investigator and lead author of this work.

Appendix B is adapted in part with permission from: J. Jang and F. Meyer, "Navigation in Shallow Water Using Passive Acoustic Ranging," in *Proc. Int. Conf. Inf. Fusion*, June 2023. The dissertation author was the primary investigator and lead author of this work.

I would like to thank the students and staff of the Scripps Marine Bioacoustics Collaborative for their contributions to fieldwork, data curation, and preparation. Special thanks to Mr. Wenyu Zhang for his implementation of the particle-flow-based multi-target tracking algorithm used in Chapter 2. For the work presented in Chapters 3 and 4, I am grateful to Drs. Heechun Song, Donghyeon Kim, Gihoon Byun, and Sandrine Rakotonarivo for their insightful discussions on waveguide-invariant-based ranging, and to Prof. William Hodgkiss for his thoughtful critiques through multiple iterations of the manuscript.

This research was supported by the following funding sources:

- Chapters 1 and 2: National Science Foundation (NSF) CAREER Award No. 2146261, Office of Naval Research (ONR) YIP Award No. N00014-15-1-2587, and the Cooperative Ecosystems Study Unit under Cooperative Agreement No. N62473-18-2-0016 (U.S. Navy, U.S. Pacific Fleet).
- Chapters 3 and 4: Defense Advanced Research Projects Agency (DARPA) Grant No. D22AP00151 and ONR Grant No. N00014-23-1-2284.

Finally, I would like to thank my family, friends, labmates, and the SIO community for their unwavering support, camaraderie, and inspiration. My PhD journey has been shaped not only by the contents of this dissertation but also by the many invaluable experiences beyond it—from student-led cruises and summer fieldwork with MPL interns, to stimulating conversations with brilliant scientists across the globe.

VITA

2018	Bachelor of Science in Electrical and Computer Engineering, Carnegie Mellon University
2020	Master of Science in Media Arts and Sciences, Massachusetts Institute of Technology
2022	Master of Science in Oceanography, University of California San Diego
2025	Doctor of Philosophy in Oceanography, University of California San Diego

ABSTRACT OF THE DISSERTATION

Passive Acoustic Localization and Tracking in Marine Environments

by

Junsu Jang

Doctor of Philosophy in Oceanography

University of California San Diego, 2025

Florian Meyer, Chair

Passive acoustic monitoring (PAM) enables the study of marine mammals and anthropogenic sources without active sound transmission, offering a non-intrusive means to observe underwater environments. Yet, in practical deployments, PAM systems must operate under resource constraints—sparse hydrophone geometries, limited power, and noisy, non-cooperative sources—requiring robust signal processing methods to extract spatial information.

This dissertation presents two distinct contributions to passive source localization. The first addresses the challenge of localizing and tracking odontocetes using their echolocation clicks, which are impulsive, broadband signals (Chs. 1 and 2). Through

simulation and at-sea experiments with sparse volumetric arrays, we demonstrate that time-difference-of-arrival (TDOA) methods outperform beamforming approaches—including frequency-difference beamforming—in terms of accuracy and robustness. Building on this finding, we develop a particle filter-based multi-target tracking framework that localizes multiple animals in 3-D from the TDOA measurements while accounting for multiple targets, false alarms, and intermittent detections. The method is validated using acoustic recordings of echolocation click signals of from goose-beaked whale (*Ziphius cavirostris*), yielding automated movement tracks of deep-diving whales.

The second contribution focuses on passive range estimation of large commercial ships in shallow water, where acoustic propagate as modes. Using single-hydrophone recordings, we apply waveguide invariant (WI) theory to analyze striation patterns as a result of modal interference in the low frequency band. A statistical model is developed to characterize the received tonal and broadband acoustic signature of ship noise, enabling maximum likelihood-based range estimation. Applied to real data from the Seabed Characterization Experiment 2017 (SBCEX17), this method yields accurate range estimates out to 45 km. Building on this result, Ch. 4 examines the validity and limits of the WI approach when only a small number of modes propagate in a low frequency band. Through simulation and observational analysis, we show that range estimation remains effective when interference is dominated by a single mode pair. This finding highlights the practical viability of long-range passive ranging in environments based on WI.

Collectively, these results advance the use of PAM in both ecological and operational contexts by enabling precise source localization in scenarios where sensor density, acoustic conditions, or signal structure pose major challenges. The techniques developed here contribute to scalable monitoring of marine mammals and improved acoustic situational awareness.

Introduction

This dissertation develops and validates methods for passive acoustic source localization in two distinct applications: multi-target tracking (MTT) of odontocetes (toothed whales) using volumetric hydrophone arrays (Chs. 1 and 2), with an emphasis on processing their echolocation click signals, and passive ranging of ships in shallow-water environments using a single hydrophone, where low-frequency ship noise exhibits modal interference (Chs. 3 and 4). The proposed methods are validated and demonstrated using real acoustic data collected at sea. Each chapter builds upon theoretical and empirical tools to contribute to the broader goal of robust PAM in operational and ecological applications.

In Chapter 1, we examine direction-of-arrival (DOA) estimation strategies based on time-difference-of-arrival (TDOA) and beamforming for impulsive, broadband signals such as odontocete echolocation clicks. Through empirical evaluation using simulated and real data collected at sea, we demonstrate that TDOA methods outperform beamforming approaches—including frequency-difference beamforming (FDBF) and incoherent conventional beamforming (CBF). This validates the robustness of TDOA-based localization for impulsive, broadband signals in high-frequency PAM applications, particularly for echolocation clicks recorded on sparse volumetric arrays.

Chapter 2 presents a scalable MTT framework for localizing odontocetes using echolocation clicks. The algorithm addresses the challenges of high-dimensional state spaces, clutter, and missed detections by using particle filtering to track 3D positions and velocities of multiple whales simultaneously. We demonstrate automated tracking of goose-beaked whales (*Ziphius cavirostris*) from real data collected with high-frequency

acoustic recording packages (HARPs), showing the method’s effectiveness for studying elusive deep-diving species.

Chapters 3 and 4 focus on shallow water environments, where acoustic propagation is governed by complex interactions with the seabed and sea surface. In such environments, the ocean acts as an acoustic waveguide supporting discrete modal propagation. Interference between modes produces spatial striation patterns in range-frequency space, which can be interpreted using waveguide invariant (WI) theory. We introduce a statistical model for ship noise in shallow water that captures both tonal and broadband acoustic signatures in a low frequency band. Using this model, we develop a passive ranging algorithm capable of estimating ship distances up to 45 km using a single acoustic receiver. The method is validated on real-world data from three ships during the Seabed Characterization Experiment 2017 (SBCEX17), demonstrating accurate long-range performance.

Chapter 4 investigates the practical limits of applying waveguide invariant (WI) theory for passive acoustic ranging in shallow water when only a small number of modes propagate, i.e., when a low-frequency band is considered for long-range range estimation. Using data from the SBCEX17 experiment, the study analyzes striation patterns formed by modal interference in the 42–49 Hz range. To interpret these patterns, the signals are processed to isolate contributions from individual mode pairs. Both numerical simulations and field data show that when interference is dominated by a single mode pair—particularly mode pair 1–2—and the corresponding mode-pairwise WI remains approximately constant over the considered frequency band, the resulting striation slope can be accurately described by a scalar WI value. These results identify the conditions under which WI-based ranging remains robust at long distances (greater than 15 km), while also highlighting the challenges that arise at close ranges near the closest point of approach, where multiple interfering modes and frequency-dependent effects complicate estimation.

Collectively, this dissertation makes the following contributions: (1) empirical validation that TDOA-based localization outperforms beamforming approaches for direc-

tional estimation of impulsive signals on sparse volumetric arrays; (2) development of an automated, high-dimensional multi-target tracking algorithm for odontocetes using volumetric arrays; (3) a statistical model of shallow-water ship noise that integrates modal interference structure and spectral variability; (4) a passive ranging algorithm based on that model, validated on multiple ships to distances exceeding 45 km; and (5) identification of the conditions under which the WI-based ranging remains accurate at long distances. These advances support more accurate, scalable, and environmentally aware PAM analyses, contributing to both marine ecological studies and broader maritime monitoring.

Chapter 1

Detection and Directional Processing of Odontocete Echolocation Clicks Using Beamforming and TDOA Methods

1.1 Introduction

Passive acoustic monitoring (PAM) is a nonintrusive and efficient approach for studying and monitoring acoustically active animals, especially species that are challenging to observe visually. PAM enables detection, localization, and tracking of those animals and is therefore well-suited for studying their abundance [3, 4], behavior [5, 6, 7], and response to anthropogenic activities [8, 9]. With continuously increasing human activities in the ocean [10, 11, 12, 13], consistent monitoring and assessment of the population, behavior, phenology, and physiology of marine animals are necessary to make informed conservation plans and management policies [14]. Of particular monitoring interest are cetaceans (whales) since they are apex predators and environment sentinels [15]. Their density and geographic location can help understand complex environmental changes, e.g., caused by anthropogenic disturbances.

Cetaceans are known to produce various types of sounds for communication, navigation, and foraging [16]. They are divided into two suborders: *Odontoceti* (toothed whales

or odontocetes) and *Mysticeti* (baleen whales or mysticetes). Odontocetes predominantly use high-frequency whistles or burst pulses to communicate [17], while mysticetes produce low-frequency tonal calls, which when used in a pattern is considered song [18]. To locate prey and relevant features of the environment, odontocetes also frequently emit echolocation clicks [16], which are intense and directional short pulses, underwater. Since the echolocation clicks are impulsive and broadband in frequency [19], they are promising signals for researchers to process to localize and track the echolocating odontocetes with PAM.

Various PAM technologies suitable for studying whales have been developed. Promising sensing approaches for PAM include towed arrays [20], fiber optic cables [21], mobile hydrophone recorders [22, 23], and bottom-mounted hydrophone arrays [5, 24]. Emerging accessible and inexpensive PAM technologies are expected to provide acoustic datasets that are orders of magnitude larger than datasets provided by conventional technologies [25]. Thus, establishing effective algorithmic solutions for data processing, data management, data analysis, and performance evaluation is crucial.

To effectively utilize these sensing technologies for monitoring odontocetes, signal processing methods must address several challenges posed by the properties of echolocation clicks and the deployment constraints of hydrophone arrays. These challenges include the impulsive and broadband nature of the signals, high background noise levels, sparse and irregular array geometries, and the need for automated detection and localization across large volumes of acoustic data. This chapter evaluates and compares signal processing techniques—specifically beamforming- and time-difference-of-arrival (TDOA)-based methods—for detecting and estimating the direction of arrival (DOA) of odontocete echolocation clicks using sparse volumetric hydrophone arrays.

We focus on data collected from a four-element volumetric array, where the element spacing is significantly larger than the wavelengths of the signals of interest. A schematic to describe the general scenario is presented in Fig. 1.1, whose mathematical notations

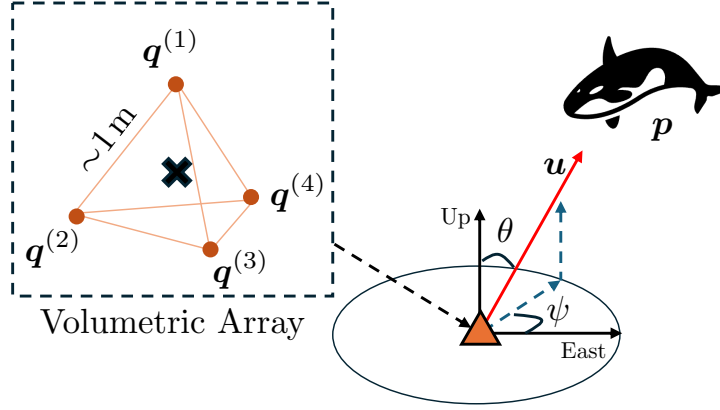


Figure 1.1. Schematic of the passive acoustic localization setup using a volumetric hydrophone array. The array consists of four hydrophones located at positions $\mathbf{q}^{(1)}$ through $\mathbf{q}^{(4)}$ forming a small 3-D aperture as indicated in the dashed box. The ‘X’ indicates the centroid of the array, which is also the reference position, i.e., the origin. East-North-Up (ENU) coordinate frame is used. An impulsive acoustic signal, e.g., an echolocation click, is emitted from a source (odontocete) at position $\mathbf{p}$. The unit direction-of-arrival (DOA) vector $\mathbf{u}$ is parameterized by elevation θ and azimuth ψ in the spherical coordinate system. This geometric configuration is used for estimating directional information for downstream localization and tracking.

will be introduced in Ch. 1.2 In particular, the analysis focuses on acoustic recordings of echolocation clicks produced by goose-beaked whales (*Ziphius cavirostris*). In this sparse-array scenario, we demonstrate that TDOA-based processing outperforms beamforming, yielding more precise DOA estimates and higher detection rates for impulsive, high-bandwidth echolocation clicks. Additionally, we introduce a TDOA-processing method that includes a noise-filtering step to suppress instrument noise, and we show that this enhancement further improves detection performance in both real and simulated datasets.

1.1.1 State-of-the-art

A well-established method for estimating the direction of arrival (DOA) of acoustic signals using an array is conventional beamforming (CBF), also known as delay-and-sum beamforming [26]. CBF estimates the direction of a sound source by time-shifting the signals received at each sensor so that those arriving from a specific direction align in time and add constructively, while signals from other directions remain misaligned. However,

for the dataset considered in this study, the volumetric hydrophone array is relatively sparse, with an element spacing of approximately 1 m. Since a secondary spectral peak of goose-beaked whale echolocation clicks, corresponding to the expected lowest excited frequency band, occurs near 17 kHz [27], equivalent to a wavelength of less than 0.1 m in water, spatial aliasing is expected. Spatial aliasing leads to multiple potential directions for which the same signals aligns, resulting in ambiguous or false DOA estimates. Although summing beamformer output incoherently across frequencies can help reduce aliasing, we find it does not reliably suppress these artifacts. Additionally, beamforming can be computationally demanding, limiting its practicality for real-time or large-scale processing.

Frequency-difference beamforming (FDBF) [28, 29, 30, 31] provides an alternative that is better suited for sparse arrays. By beamforming at the frequency difference between high-frequency spectral components, FDBF synthesizes a much lower effective frequency, thereby reducing sensitivity to spatial aliasing. This method has been successfully applied in a variety of acoustic domains, including blind deconvolution [28], underwater DOA estimation [32, 33], earthquake rupture imaging [34], and ultrasound tomography [35]. Motivated by these applications, we evaluate the use of FDBF for processing odontocete echolocation clicks recorded by a four-sensor volumetric hydrophone array.

A complementary approach for processing impulsive signals such as clicks is based on TDOA estimation [16]. This method typically cross-correlates time series from hydrophone pairs to detect signal arrivals, and then identifies peaks in the resulting cross-correlation functions to extract TDOA measurements. While standard cross-correlation [36] performs well under high signal-to-noise ratio (SNR) conditions, the generalized cross-correlation (GCC) [37] introduces frequency-dependent weighting (filtering) that suppresses noise and improves detection robustness in noisier environments.

1.1.2 Contribution

The primary goal of this chapter is to evaluate and compare signal processing methods for detecting and estimating the DOA of odontocete echolocation clicks using sparse volumetric hydrophone arrays. We evaluate beamforming- and TDOA-based methods to process acoustic recordings of echolocation click signals of goose-beaked whale made at sea using a four-element volumetric array.

To improve robustness under low SNR conditions and strong instrument noise, we introduce a new TDOA estimation method based on a generalized cross-correlation that whitens the instrument noise (GCC-WIN). This method enhances the detectability of echolocation clicks while suppressing interference of the instrumental self-noise. We also assess how DOA estimates obtained from TDOAs can lose information—especially when only a subset of sensor pairs yield usable data for downstream processing, e.g., tracking whales in 3-D from the extracted directional estimates.

The key contributions of this chapter are summarized as follows:

- FDBF is applied to echolocation clicks of odontocete, demonstrating that it mitigates spatial aliasing on sparse arrays when CBF is applied instead.
- A custom generalized cross-correlation, referred to as GCC-WIN, is proposed for TDOA detection and estimations. This method whitens instrument noise based on precomputed noise power spectral density (PSD) templates, improving click detections in the TDOA domain under low SNR.
- A comparative analysis is conducted between beamforming- and TDOA-based methods for DOA estimation, showing the superior performance of TDOA methods in noisy environments.
- A study is presented showing that converting TDOAs into DOAs leads to information loss, particularly in a low SNR scenarios. This motivates the use of TDOAs directly in

3-D tracking algorithms for increased detection rates and improved track continuity.

The chapter is structured as follows: Ch. 1.2 introduces the signal model and array geometry. Chs. 1.3 and 1.4 reviews the principles of beamforming and TDOA estimations. Ch. 1.7 describes the dataset and preprocessing, and Ch. 1.6 discusses some implementation aspects. Finally, Ch. 1.8 presents the evaluation results, and Ch. 1.9 concludes with a discussion of implications for 3-D tracking in the context of sparse hydrophone arrays.

1.2 Signal Model

Echolocation clicks from odontocetes are high-frequency, impulsive signals with durations on the order of hundreds of microseconds. An example time series containing a goose-beaked whale click, as recorded across four hydrophone sensors with arrival time differences due to array geometry and the position of the whale with respect to the array, is shown in Fig. 1.2(a). Given the known sensor geometry, sound speed, and observed arrival delays, the direction of the click can be inferred via beamforming or TDOA estimation.

For directional processing of the click signals, we model incoming signals as plane waves from far-field sources; the detected whales are typically hundreds of meters away. A spherical coordinate system is used, where the elevation angle $\theta = 0^\circ$ corresponds to the zenith (positive z-axis) and $\theta = 180^\circ$ to the nadir (negative z-axis). The azimuth angle ψ is defined such that $\psi = 0^\circ$ points East and increases counterclockwise, with North at $\psi = 90^\circ$ and West at $\psi = 180^\circ$.

1.2.1 Time Domain Signal Model

Assume that the acoustic source (whale) at discrete time step n is located at

$$\mathbf{p}_n = r_n \begin{bmatrix} \sin(\theta_n) \cos(\psi_n) & \sin(\theta_n) \sin(\psi_n) & \cos(\theta_n) \end{bmatrix}^T = r_n \mathbf{u}_n, \quad (1.1)$$

where r_n , θ_n , and ψ_n denote the range, elevation, and azimuth relative to the center of the array (used as the reference point), and $\mathbf{u}_n$ is the unit DOA vector. The relevant geometries and introduced notations are summarized in Fig. 1.1. Assuming far-field propagation, the signal arrives as a plane wave along the direction $-\mathbf{u}_n$. The array consists of N_h hydrophone sensors, where the position of sensor $h = 1, \dots, N_h$ is denoted $\mathbf{q}^{(h)} \in \mathbb{R}^3$.

The received echolocation click signals at a pair of sensors h_1 and h_2 can be modeled in discrete time step n as

$$\begin{aligned} x_n^{(h_1)} &= w_n + z_n^{(h_1)}, \\ x_n^{(h_2)} &= w_n + \tau + z_n^{(h_2)}, \end{aligned} \tag{1.2}$$

where $w_n \in \mathbb{R}$ is the true acoustic signal (click), $z_n^{(s)}$ is additive noise at sensor s , and τ is the TDOA between the sensors h_1 and h_2 , expressed in samples. The TDOA τ would not be an integer in practice, but we assume the sampling frequency F_s is sufficiently high to model the delay without significant discretization error. The signal is segmented into short snapshots N samples. These snapshots are processed either in the time domain (e.g., for TDOA estimation using cross-correlation) or transformed to the frequency domain (e.g., for generalized cross-correlation and beamforming). We assume at most one echolocation click occurs in each snapshot and that the direct-path signal dominates over multipath arrivals.

1.2.2 Frequency Domain Signal Model

For certain processing techniques such as beamforming, it is more convenient to analyze signals in the frequency domain using the short-time Fourier transform (STFT). For snapshot k and frequency bin l (with frequency f_l and corresponding angular frequency $\omega_l = 2\pi f_l$), the signal is modeled as

$$\mathbf{X}_{k,l} = \mathbf{V}_{k,l}W_{k,l} + \mathbf{Z}_{k,l}, \tag{1.3}$$

where $\mathbf{X}_{k,l} = [X_{k,l}^{(1)}, \dots, X_{k,l}^{(N_h)}]^T \in \mathbb{C}^{N_h}$ is the observed signal vector, $\mathbf{Z}_{k,l} = [Z_{k,l}^{(1)}, \dots, Z_{k,l}^{(N_h)}]^T \in \mathbb{C}^{N_h}$ is the additive noise vector, $W_{k,l} \in \mathbb{C}$ is the source signal, and $\mathbf{V}_{k,l} \in \mathbb{C}^{N_h}$ is the array manifold vector defined as

$$\mathbf{V}_{k,l} = \left[e^{j\frac{\omega_l}{c}\mathbf{u}_k^T \mathbf{q}^{(1)}}, \dots, e^{j\frac{\omega_l}{c}\mathbf{u}_k^T \mathbf{q}^{(N_h)}} \right]^T. \quad (1.4)$$

During a snapshot, the source is assumed to be stationary such that $\mathbf{u}_k$ and thus $\mathbf{V}_{k,l}$ are constant over the considered N samples. The TDOA (in samples) between sensors h_1 and h_2 can be expressed as

$$\tau = \frac{F_s}{c} \mathbf{u}_n^T (\mathbf{q}^{(h_1)} - \mathbf{q}^{(h_2)}), \quad (1.5)$$

where c is the sound speed in water and F_s is the sampling frequency.

1.3 Beamforming

In this section, we review CBF, i.e., delay-and-sum method, and the challenge of spatial aliasing that arises with sparse arrays, followed by a discussion of FDBF as a mitigation strategy. Beamforming is applied independently to each snapshot because the echolocation click duration is too short, which precludes covariance-based methods such as the multiple signal classification (MUSIC) [38, 39] algorithm. Instead, we use CBF, where array outputs are coherently combined across sensors and then incoherently averaged across frequency bins to estimate source direction.

1.3.1 Conventional Beamforming

For a given look direction defined by elevation θ and azimuth ψ , we define the unit direction vector $\mathbf{b}(\theta, \psi) = [\sin(\theta) \cos(\psi), \sin(\theta) \sin(\psi), \cos(\theta)]^T$, and the corresponding steering vector at frequency bin l is given by

$$\mathbf{a}_l(\theta, \psi) = \left[e^{-j\frac{\omega_l}{c}\mathbf{b}(\theta, \psi)^T \mathbf{q}^{(1)}}, \dots, e^{-j\frac{\omega_l}{c}\mathbf{b}(\theta, \psi)^T \mathbf{q}^{(N_h)}} \right]^T. \quad (1.6)$$

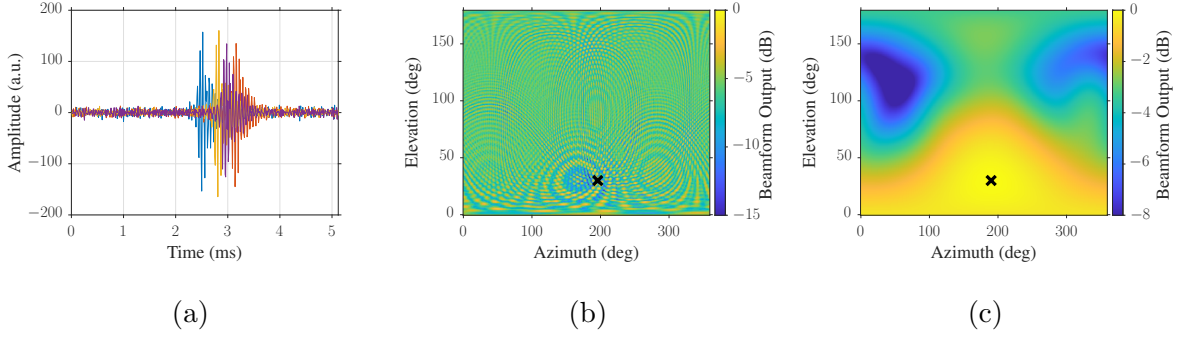


Figure 1.2. (a) A snapshot of time-domain signals recorded across the four hydrophones, capturing an echolocation click from a goose-beaked whale (*Ziphius cavirostris*). Each colored-trace corresponds to a different sensor. The time-differences-of-arrival (TDOA), i.e., the phase coherence, across sensors enable DOA estimation. (b) Normalized output of conventional beamforming (CBF, delay-and-sum), obtained by incoherently averaging beamformer power across selected frequency bins. The cross indicates the estimated DOA. Multiple peaks due to spatial aliasing are evident. (c) Normalized output of frequency-difference beamforming (FDBF) applied to the same data. FDBF suppresses spatial aliasing by operating at a lower synthetic frequency and requires computing a single sensing matrix, resulting in substantially faster processing (approximately $46\times$ faster than CBF in this case).

At each snapshot k and frequency bin l , the CBF output is computed as

$$Y_{k,l}(\theta, \psi) = |\mathbf{a}_l(\theta, \psi)^T \mathbf{X}_{k,l}|^2. \quad (1.7)$$

To leverage the broadband nature of the signal, the outputs across L frequency bins are incoherently averaged, i.e.,

$$\bar{Y}_k^{\text{FDBF}}(\theta, \psi) = \frac{1}{L} \sum_{l=1}^L Y_{k,l}(\theta, \psi). \quad (1.8)$$

For each snapshot, FDBF was applied and the DOA was estimated by locating the beamformer peak:

$$\hat{\theta}_k^{\text{CBF}}, \hat{\psi}_k^{\text{CBF}} = \arg \max_{\theta \in \Theta, \psi \in \Psi} \bar{Y}_{k,\Delta l}^{\text{CBF}}(\theta, \phi). \quad (1.9)$$

In our implementation, elevation and azimuth search grids were defined as $\Theta =$

$\{0^\circ, 1^\circ, \dots, 179^\circ\}$ and $\Phi = \{0^\circ, 1^\circ, \dots, 359^\circ\}$.

CBF becomes problematic when applied to sparse arrays at high frequencies, such as those encountered in odontocete echolocation clicks. In particular, when the sensor spacing exceeds half the wavelength of the signal, spatial aliasing causes multiple ambiguous peaks in the beamformer output, as shown in Fig. 1.2 (b). In this figure, the CBF output at multiple frequency bins have been incoherently averaged. The presence of strong aliased sidelobes can lead to incorrect estimates, as demonstrated later in Ch. 1.8. While there is no ground truth DOA of the whale available, the FDBF and TDOA based methods agreed closely, while the CBF result deviated by approximately 10° in the azimuth. In our dataset, the inter-sensor spacing is approximately 1 m, while the wavelengths corresponding to secondary spectral peaks of beaked whale clicks (e.g., at 17 kHz) are below 0.1 m. This mismatch leads to aliasing artifacts that complicate DOA estimation.

1.3.2 Frequency Difference Beamforming

Unlike CBF, which directly forms beams across frequency bins, FDBF instead processes pairs of frequency components to operate at a synthetic, much lower frequency. FDBF [29] offers a solution to CBF's spatial aliasing challenge when processing arrays with larger aperture by synthesizing a lower effective frequency through phase manipulation of signal components at two nearby frequency bins. Since the synthetic frequency is much lower than the original component frequencies, FDBF operates below the spatial Nyquist limit of the array.

Consider the array measurements $\mathbf{X}_{k,l} \in \mathbb{C}^{N_h}$ and $\mathbf{X}_{k,l+\Delta l} \in \mathbb{C}^{N_h}$ at frequency bins l and $l + \Delta l$ respectively. Here, Δl is the difference frequency. The autoprodut for FDBF at snapshot k is defined and denoted as [29]

$$\mathbf{X}_{k,(l,\Delta l)} = \mathbf{X}_l^* \circ \mathbf{X}_{l+\Delta l}, \quad (1.10)$$

where $\circ$ denotes the Hadamard (element-wise) product and $*$ is the complex conjugate. The autoprodut effectively mimics a signal at the synthetic frequency $\omega_{\Delta l} = \omega_{l+\Delta l} - \omega_l$, which is typically much lower than either component frequency.

Let $\mathbf{a}_{\Delta l}(\theta, \psi)^T \in \mathbb{C}^{N_h}$ denote the steering vector constructed for this synthetic difference frequency, analogous to Eq. (1.6) but evaluated at Δl . The FDBF output for the frequency pair $(l, l + \Delta l)$ is computed as

$$Y_{k,(l,\Delta l)}(\theta, \psi) = |\mathbf{a}_{\Delta l}(\theta, \psi)^T \mathbf{X}_{k,(l,\Delta l)}|^2. \quad (1.11)$$

To obtain a broadband estimate, we fix Δl and define $\mathcal{Q}$ as the set of frequency indices l such that both ω_l and $\omega_{l+\Delta l}$ lie within the excited frequency band of the source. The averaged FDBF output over all valid frequency pairs is then given by

$$\bar{Y}_{k,\Delta l}^{\text{FDBF}}(\theta, \psi) = \frac{1}{|\mathcal{Q}|} \sum_{l \in \mathcal{Q}} Y_{k,(l,\Delta l)}(\theta, \psi), \quad (1.12)$$

where $|\mathcal{Q}|$ is the cardinality of $\mathcal{Q}$, i.e., the number of valid frequency pairs used. For each snapshot, FDBF was applied and the direction of arrival was estimated by locating the beamformer peak of $\bar{Y}_{k,\Delta l}^{\text{FDBF}}(\theta, \psi)$, i.e.,

$$\hat{\theta}_k^{\text{FDBF}}, \hat{\psi}_k^{\text{FDBF}} = \arg \max_{\theta \in \Theta, \psi \in \Psi} \bar{Y}_{k,\Delta l}^{\text{FDBF}}(\theta, \phi). \quad (1.13)$$

The same search grid as that used for CBF, i.e., $\Theta = \{0^\circ, 1^\circ, \dots, 179^\circ\}$ and $\Phi = \{0^\circ, 1^\circ, \dots, 359^\circ\}$ is used.

FDBF offers a means to suppress spatial aliasing by performing beamforming at a lower synthetic frequency. This enables DOA estimation even when the original signal components lie well above the spatial Nyquist limit of the array, as long as a sufficiently small Δl is available. An example FDBF output is shown in Fig. 1.2(c),

where spatial aliasing is effectively suppressed and sidelobes are not apparent. However, FDBF is susceptible to reduced angular resolution due to its operation at lower synthetic frequencies.

1.4 Time Difference of Arrival

In contrast to beamforming-based approaches, TDOA-based methods rely on measuring relative arrival times between sensors and offer an alternative strategy for DOA estimation, particularly well-suited for impulsive signals like echolocation clicks.

1.4.1 Cross-correlation

TDOA measurements can be extracted using cross-correlation between signals recorded on a spatially separated hydrophone pair (h_1, h_2) . For a k^{th} snapshot of length N samples, the unbiased sample cross-correlation as a function of integer lag $d \in \{-N + 1, \dots, N - 1\}$ is defined as

$$\phi_{k,d}^{(h_1,h_2)} = \frac{1}{N - |d|} \sum_{n=0}^{N-|d|-1} x_n^{(h_1)}, x_{n+d}^{(h_2)}, \quad (1.14)$$

where $x_n^{(h_1)}$ denotes the time-series signal received at sensor h_1 , as introduced in Eq. (1.2). Signals are zero-padded outside the observation window, and the normalization factor ensures that the correlation estimate remains unbiased, particularly near the edges of the window.

To constrain the estimation to physically plausible delays, the search for valid lags is restricted to a finite interval determined by the maximum possible propagation delay N^{max} in samples, based on the sensor separation and assumed sound speed. The TDOA estimate is then given by the lag d that maximizes the cross-correlation within this

admissible range:

$$\hat{\tau}_k^{\text{XCORR}} = \arg \max_{d \in \{-N^{\max}, \dots, N^{\max}\}} \phi_{k,d}^{(h_1, h_2)}. \quad (1.15)$$

1.4.2 Generalized Cross-correlation

While standard cross-correlation performs well when signals are recorded at high SNR, its performance degrades in noisy environments. To improve robustness in such conditions, generalized cross-correlation (GCC) methods introduce frequency-domain weighting functions that emphasize frequency components dominated by signal energy while de-emphasizing noisy or uninformative bands. By applying appropriate weighting, GCC enhances the sharpness and reliability of the cross-correlation peak, improving the accuracy of TDOA estimation under low SNR.

Given the frequency domain signal for sensor pair h_1 and h_2 as $X_{k,l}^{(h_1)}$ and $X_{k,l}^{(h_2)}$ in the k^{th} snapshot, where $l = 0, \dots, N-1$ is the discrete frequency, the GCC as a function of discrete time delay d is

$$\hat{\phi}_{k,d}^{(h_1, h_2)} = \frac{1}{N} \sum_{k=0}^{N-1} \Gamma_{k,l} \Phi_{k,l}^{(h_1, h_2)} e^{j2\pi dl/N}, \quad (1.16)$$

where $\Phi_{k,l}^{(h_1, h_2)} = X_{k,l}^{(h_1)} (X_{k,l}^{(h_2)})^*$ is the cross power spectral density (CPSD) estimate of the received signals, and $\Gamma_{k,l}$ is the frequency weighting of the GCC. Then, the TDOA estimate is found at the location that maximizes the GCC as

$$\hat{\tau}_k^{\text{GCC}} = \arg \max_{d \in \{-N^{\max}, \dots, N^{\max}\}} \hat{\phi}_{k,d}^{(h_1, h_2)} \quad (1.17)$$

Frequency weighting is performed according to specific optimization criteria. In applications where the noise PSD is unknown, popular choices of frequency weighting factors result in the smoothed coherence transform (SCOT) [40] and phase transform (PHAT) [37]. The SCOT normalizes the PSDs of the individual signals to unit magnitude,

i.e., $\Gamma_{k,l}^{\text{SCOT}} = 1/(R_{k,l}^{(h_1)} R_{k,l}^{(h_2)})^{1/2}$, where $R_{k,l}^{(h_1)}$ denotes the PSD estimate of $x_n^{(h_1)}$ at snapshot k . Alternatively, the PHAT normalizes the CPSD to unit magnitude, i.e., $\Gamma_{k,l}^{\text{PHAT}} = 1/|\Phi_{k,l}^{(h_1,h_2)}|$.

In addition, we adopt GCC by utilizing the time-varying but periodic PSD estimates $G_{k,l}^{(h_1)}$ and $G_{k,l}^{(h_2)}$ of the respective noise $z_n^{(h_1)}$ and $z_n^{(h_2)}$, which are observed to be mainly dominated by the instrument noise. The noise PSDs can be measured or precomputed and is thus considered known. The optimal frequency weighting for whitening the instrument noise is given by

$$\Gamma_{k,l}^{\text{WIN}} = 1/(G_{k,l}^{(h_1)} G_{k,l}^{(h_2)})^{1/2}. \quad (1.18)$$

The resulting GCC-WIN can be interpreted as (i) applying the noise whitening filters $1/(G_{k,l}^{(h_1)})^{1/2}$ and $1/(G_{k,l}^{(h_2)})^{1/2}$ to the signals $x_n^{(h_1)}$ and $x_n^{(h_2)}$, respectively, and (ii) performing a conventional cross-correlation. In Ch. 1.5, we describe the dataset and noise sources. Further details on the implementation of GCC-WIN is provided in Ch. 1.6.2.

1.4.3 DOA Computation from TDOA Measurements

TDOA measurements between pairs of hydrophones can be used to compute the DOA of a click signal. For each hydrophone pair (h_1, h_2) , the corresponding TDOA defines a hyperboloid of potential source locations. Under the far-field assumption, each hyperboloid can be well approximated by a conical surface aligned along a direction vector. This means that each TDOA measurement constrains the DOA to lie on a cone whose axis depends on the relative sensor geometry.

For an array of N_h hydrophones, there are $N_s = \binom{N_h}{2}$ unique sensor pairs, and thus N_s such conical constraints. In the far-field limit, the DOA vector $\boldsymbol{\nu} = [\nu_x, \nu_y, \nu_z]^T \in \mathbb{R}^3$ (in Euclidean coordinate system) is assumed constant across sensor pairs and satisfies the following linearized system derived from the propagation delays:

$$\mathbf{q}^{(h_1,h_2)T} \boldsymbol{\nu} = c \cdot \tau^{(h_1,h_2)}, \quad (1.19)$$

where $\mathbf{q}^{(h_1, h_2)} = \mathbf{q}^{(h_1)} - \mathbf{q}^{(h_2)} \in \mathbb{R}^3$ is the relative position vector between sensors h_1 and h_2 , $\tau^{(h_1, h_2)} \in \mathbb{R}$ is the measured TDOA between the two sensors (in seconds), and c is the sound speed in water.

Stacking all pairwise equations yields the overdetermined system:

$$A\boldsymbol{\nu} = \mathbf{b}, \quad (1.20)$$

where $A \in \mathbb{R}^{N_s \times 3}$ has rows $\mathbf{q}^{(h_1, h_2)\text{T}}$, and $\mathbf{b} \in \mathbb{R}^{N_s}$ contains the scaled TDOA values. Let the index $s = 1, \dots, N_s$ correspond to the s^{th} unique sensor pair, such that $b_s = c \cdot \tau_s$, where τ_s is the measured TDOA for the s^{th} pair. The mapping from sensor pairs (h_1, h_2) to index s follows a lexicographical ordering of all combinations, i.e., $(1, 2), (1, 3), \dots, (N_s - 1, N_s)$. The DOA vector $\boldsymbol{\nu}$ can then be estimated via least squares:

$$\hat{\boldsymbol{\nu}} = (A^T A)^{-1} A^T \mathbf{b}. \quad (1.21)$$

To ensure $\hat{\boldsymbol{\nu}}$ is a unit vector, we normalize it, i.e., $\hat{\mathbf{u}} = \hat{\boldsymbol{\nu}} / \|\hat{\boldsymbol{\nu}}\|$.

Given the spherical coordinate system used in this thesis, the elevation and azimuth angles are then computed from the unit DOA vector $\hat{\mathbf{u}} = [\hat{u}_x, \hat{u}_y, \hat{u}_z]^T$ as:

$$\phi = 180^\circ - \cos^{-1}(\hat{u}_z), \quad (1.22)$$

$$\psi = \tan^{-1}(\hat{u}_y / \hat{u}_x), \quad (1.23)$$

where $\phi \in [0^\circ, 180^\circ)$ is the elevation angle and $\psi \in [0^\circ, 360^\circ)$ is the azimuth angle. In practice, the `atan2d` function in MATLAB is used to compute the azimuth, with values below 0° adjusted by adding 360° .

1.5 Dataset

Both Chs. 1 and 2 use acoustic signals measured on two high frequency acoustic recording packages (HARPs) [41], each of which is equipped with four hydrophones. The sampling frequency is 100 kHz, and the resultant TDOA resolution is 10 μ s. The hydrophones on the HARP are 1 m apart and arranged in a tetrahedral shape (see Fig. 2 in [5] for the actual array schematic) to form a small-aperture array. The hydrophones on each HARP share the same clock and are thus time-synchronized. They were deployed off the coast of California ($32^{\circ} 39' 31.4''$ N, $119^{\circ} 28' 37.6''$ W) at a depth of ~ 1330 m, and the arrays were approximately 1 km apart. There was also a separate single-sensor HARP in the vicinity, but its data were not considered for this study. The arrays located east and west are referred to as HARP EE and HARP EW, respectively, and they recorded from March 15th, 2018 to July 18th, 2018 (UTC). In this chapter, the results from processing the data recorded by HARP EE on June 11th and July 1st in 2018 are presented.

In the considered scenario, the water depth is far below the thermocline, with minimal change in the sound velocity as a function of depth. Hence, an acoustic wave propagation with spherical spreading in an isovelocity medium is assumed. The sound velocity is estimated to be 1490 m s^{-1} . There was no interference from other marine animals in the frequency band of the used datasets, and hence, a classification step was not required. Upon manual inspection, the echolocation clicks followed the characteristics of those from the goose-beaked whales described in [19].

1.5.1 Noise Sources and Pre-processing

Three noise sources are identified (see Fig. 1.3): a pulse signal from the co-present acoustic Doppler current profiler (ADCP), an instrument noise on HARP highly correlated among hydrophone measurements, and ambient noise. Note that the instrument noise is harmonic and occupies a wide band of frequency, thereby giving rise to multiple high

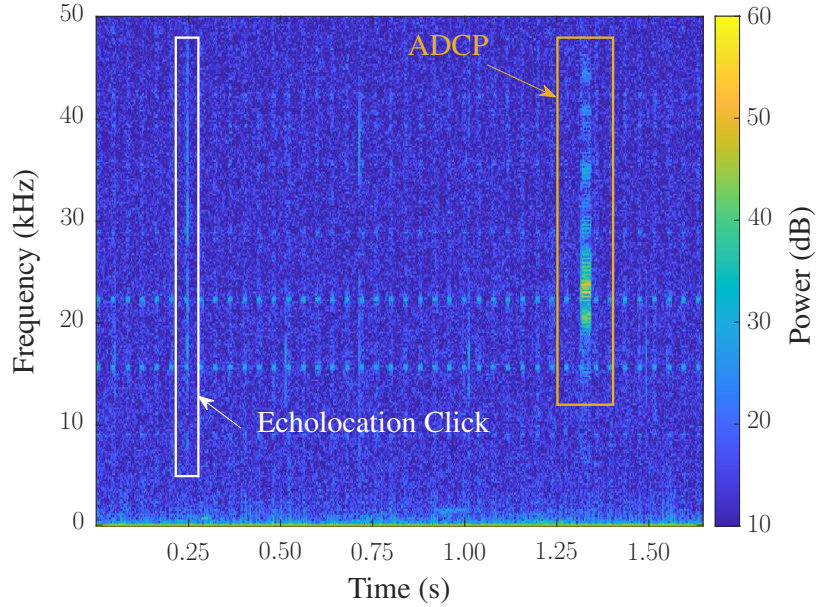


Figure 1.3. An example spectrogram of acoustic data from a single hydrophone with echolocation clicks and noise. The identified noise sources are the acoustic Doppler current profiler (ADCP), the instrument noise from the data storage system, and the undersea environment. The echolocation clicks are broadband signals whose center frequency is approximately 36 kHz. The ADCP signal is also a broadband signal with a center frequency near 25 kHz. The instrument noise is a harmonic narrowband signal that is repeated every 31 s. Its template spectrogram is displayed in Fig. 1.4.

amplitude peaks at wrong time delay locations when these signals are beamformed or cross-correlated. This would pose challenges for any further processing using the directional estimates, e.g., tracking whales in 3-D space. The ambient noise is dominant below 2 kHz, while the instrument noise is the major source of noise above 2 kHz. The echolocation clicks of a goose-beaked whale are known to have a peak frequency at 40.2 kHz, a center frequency at 35.9 kHz and a -10 dB bandwidth of 10.9 kHz [19]. Hence, the signals above 15 kHz are considered for this study.

The signal is first pre-filtered with a high pass filter based on a Parks-McClellan optimal FIR filter design [42] which is applied with a zero-phase shifting digital filter [43]. The stopband frequency and passband frequency are 13 kHz and 15 kHz, respectively. Then, the ADCP signal is identified and removed. A nearby ADCP generated a short

dominant pulse recurring approximately every 54 s, which completely flooded the acoustic measurements. Its center frequency is at 75 kHz, but since the Nyquist frequency of the instrument is 50 kHz, it is aliased and present at 25 kHz. The ADCP signal was too powerful that it was not fully suppressed by the anti-aliasing filter onboard. It is nonetheless readily identified and nulled based on its high energy and center frequency characteristics.

For techniques that require directional processing using frequency domain signals (CBF, FDBF, GCC-PHAT and GCC-WIN), the time series data have been converted into STFTs and spectrograms. They were computed using the snapshot length of 512 samples (corresponding to 5.12 ms), 50% overlap and Hamming window. This configuration yields an FFT bin width of 195.3 Hz.

1.6 Implementation Aspects

1.6.1 Beamforming

For both CBF and FDBF, we included most FFT bins from 15 kHz to 50 kHz, excluding those that were empirically observed to degrade beamformer performance. The echolocation clicks signals of the goose-beaked whales are known to have nulls [19]. Based on the array geometry and an assumed sound speed of 1490 m s^{-1} , the highest frequency at which spatial aliasing is not expected was calculated to be approximately 725.6 Hz. Accordingly, a difference of three frequency bins (781.3 Hz), i.e., $\Delta l = 3$ was selected for FDBF processing.

1.6.2 TDOA: GCC-WIN

To perform the GCC-WIN, a model of the noise in the form of a PSD is required. With the considered HARP instrument, the PSD of the noise is time-varying but periodic with a period of approximately 31 s. The spectrogram of the noise model is referred

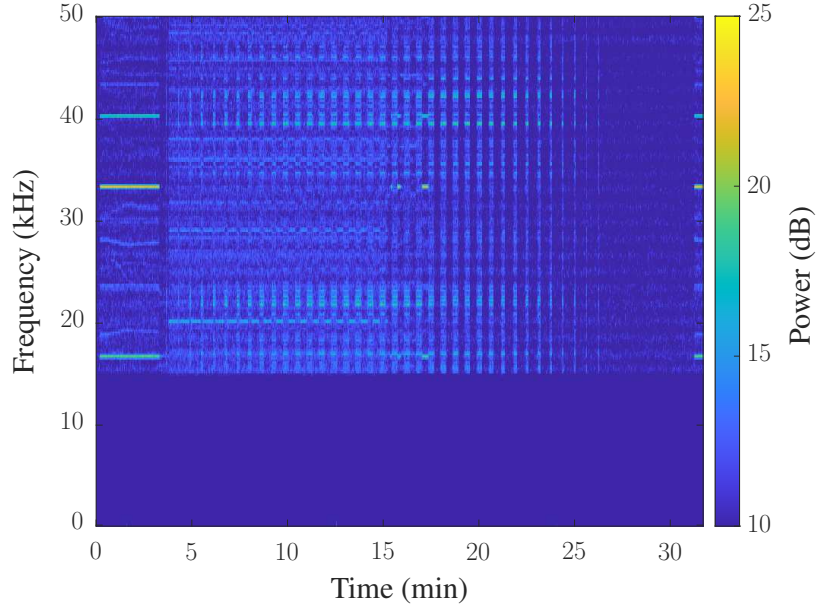


Figure 1.4. An example spectrogram of the instrument noise that repeats periodically, recorded by a hydrophone on a HARP EE on July 1st, 2018. Each column of the spectrogram serves as an estimated noise power spectral density (PSD) for the proposed generalized cross-correlation that whitens the instrument noise (GCC-WIN). For each hydrophone, an individual spectrogram is extracted.

to as the noise PSD template (an example in Fig. 1.4). The noise PSD template is estimated for each hydrophone individually from 20-min long signals before the appearance of echolocation clicks in the data. This segment is chosen manually by inspecting the Long-Term Spectral Averaging (LTSA) from the MATLAB-based program *Triton* [41]. The 20-min long acoustic recording is further divided into shorter segments that are 31.65-s long. The spectrograms computed from each segment are then averaged to estimate the noise PSD template for each hydrophone. Separate noise PSD templates are estimated for different days of data since the noise statistics vary throughout the deployment. The main source of instrument noise is believed to be the mechanical noise of the hard disks from the data storage system.

Before performing GCC-WIN, a critical step is to align the noise PSD template with the instrument noise of the signals of interest. This alignment step is based on the

averaged noise template in the time domain that is also computed for each hydrophone individually from the 20-min long signals used for estimating the noise PSD template. First, any pulse-like components, e.g., echolocation clicks or ADCP signals, that could interfere with the alignment process are removed. Then, the signal is cross-correlated with the noise template in the time domain to identify the locations of the peaks. These locations are the time steps at which the noise PSD template is aligned with the noise in the signal of interest. The peaks of the cross-correlation are expected to repeat approximately 31 s, which is equal to the period of the noise PSD. Once the noise template and the signals are aligned, the GCC-WIN technique is performed as described in Ch. 1.4.2 with the noise PSD template used as the frequency weights provided in Eq. 1.18.

1.7 Methods

This section presents the methodological framework used to evaluate the detection and localization performance of beamforming-based and TDOA-based methods for odontocete echolocation clicks. Our objective is to assess how each method performs under varying SNRs, both on real and simulated data, using common criteria for detection and DOA estimation.

1.7.1 Overview of Methods Compared

We evaluate five signal processing methods for detecting and estimating the DOA of odontocete echolocation clicks. These methods fall into two categories: beamforming-based and TDOA-based.

- **CBF (Conventional Beamforming)**: A delay-and-sum approach applied across frequency bins and averaged to form a broadband DOA estimate.
- **FDBF (Frequency-Difference Beamforming)**: A technique that suppresses spatial aliasing by beamforming at synthetic lower frequencies generated from closely

spaced frequency bin pairs.

- **XCORR (Cross-Correlation)**: A time-domain method that estimates TDOA by cross-correlating time series between sensor pairs.
- **GCC-PHAT**: A generalized cross-correlation method that applies phase-based weighting to improve TDOA estimation under noisy conditions.
- **GCC-WIN (Whitened Instrumental Noise)**: A proposed variant of GCC that whitens the cross-power spectrum using precomputed PSD templates to suppress narrowband instrumental noise.

These methods produce either a beamformer power surface (CBF, FDBF) or a set of TDOA estimates (XCORR, GCC-PHAT, GCC-WIN), from which DOAs are inferred. The following subsections describe how detections are thresholded and how performance is evaluated.

1.7.2 Detection Thresholding

All methods rely on thresholding to determine whether a given snapshot contains an echolocation click. For each method, the peak amplitude—either from the beamformer output or from the cross-correlation function—is extracted for all K snapshots, denoted as $\mathcal{I} = \{I_1, \dots, I_K\}$. In the case of DOA estimation based on TDOA measurements, the amplitudes of the TDOA peaks across all sensor pairs are summed for each snapshot to form a single I_k for the k^{th} snapshot. Thresholds are determined using the median absolute deviation (MAD), defined as

$$\text{MAD} = \text{median}(|I - \tilde{I}|), \quad I \in \mathcal{I} \quad (1.24)$$

where $\tilde{I}$ is the median of $\mathcal{I}$. A detection is declared if the peak amplitude I_k exceeds $\tilde{I} + \alpha \times \text{MAD}$, with values for α being empirically selected based on visual inspection of

detections.

The MAD is employed to set detection thresholds due to its robustness to outliers and skewed distributions, which is essential in this study given the impulsive nature of echolocation clicks. Echolocation clicks often produce peak amplitudes that are several times higher than the background, leading to heavy-tailed distributions in $\mathcal{I}$. This significantly inflates the standard deviation, often making it noticeably larger than the mean and rendering mean- or variance-based thresholds less reliable. In contrast, the MAD provides a more stable estimate of scale under such conditions, ensuring that the thresholding process remains sensitive to true signal peaks while being less affected by extreme outliers or non-Gaussian instrument self-noise.

1.7.3 DOA Estimation Methods Comparison

Using the five methods described in Ch. 1.7.1, DOA estimates are computed from real acoustic recordings collected by HARP EE from 19:10-19:30 on July 1st, 2018, where two goose-beaked whales were observed simultaneously.

For beamforming-based approaches (CBF and FDBF), DOA corresponds to the azimuth-elevation angle pair that maximizes the beamformer output. Since it was computationally untractable to compute the CBF output for every snapshots of the STFT for a 20-min long recording (approximately 470,000 snapshots, 64800 look angle pairs and 45 frequency bins), we first make detections using FDBF and use the same snapshots that yielded detection in FDBF to compute the CBF to investigate the effect of spatial aliasing. Although there could have been DOA detections made by CBF that were not possible with FDBF or TDOA-based approaches, it was sufficient to demonstrate that spatial aliasing would lead to incorrect estimates.

For TDOA-based methods (XCORR, GCC-PHAT, GCC-WIN), the estimated TDOAs are used to reconstruct the unit DOA vector using a least-squares solver described in Ch. 1.4.3. Each method is applied independently to the same dataset to compare

detection rates and localization consistency. For DOA comparison, the detection threshold was set where the MAD scaling factor is $\alpha = 10$ for FDBF, $\alpha = 5$ for XCORR and GCC-WIN, and $\alpha = 3$ for GCC-PHAT.

1.7.4 TDOA Detection Rate: Low SNR Case Study

To study detection performance under low-SNR conditions, we analyze a real encounter recorded by HARP EE from 10:53-11:13 on June 11th, 2018, during which a short track of goose-beaked whale click signals were identified. Here, we focus on using the acoustic recording from the hydrophone sensors 3 and 4. These clicks are weak and embedded in noise, making them ideal for comparing the detection performance of GCC-WIN and XCORR. Both methods are applied with identical thresholding criteria for $\alpha = 5$ and $\alpha = 2$, and the resulting detections are compared qualitatively in terms of detection continuity.

1.7.5 TDOA Detection Rate: Simulation Study

To isolate and quantify detection performance across a range of SNRs among the TDOA estimation methods (GCC-WIN, XCORR and GCC-PHAT), we conduct a simulation study in which artificial click signals are injected into a segment of real data that is free from naturally occurring echolocation clicks. A synthetic linear chirp signal is used to mimic an odontocete’s clicks. The simulated chirp is 1 ms in duration, sweeping from 20–40 kHz, and is injected every 256 samples over a 3-min window (18:35–18:38, July 1st, 2018). The time delay for each sensor is adjusted to yield different but constant TDOA values across each pair of hydrophones.

To avoid false detections from real data, snapshots that coincidentally produce a TDOA matching the simulated delay is excluded before injecting the artificial click signals. The resulting dataset contains approximately 60,500 snapshots per sensor pair. Simulated click amplitudes are scaled across five SNR levels (1 to 5 in steps of 1), using

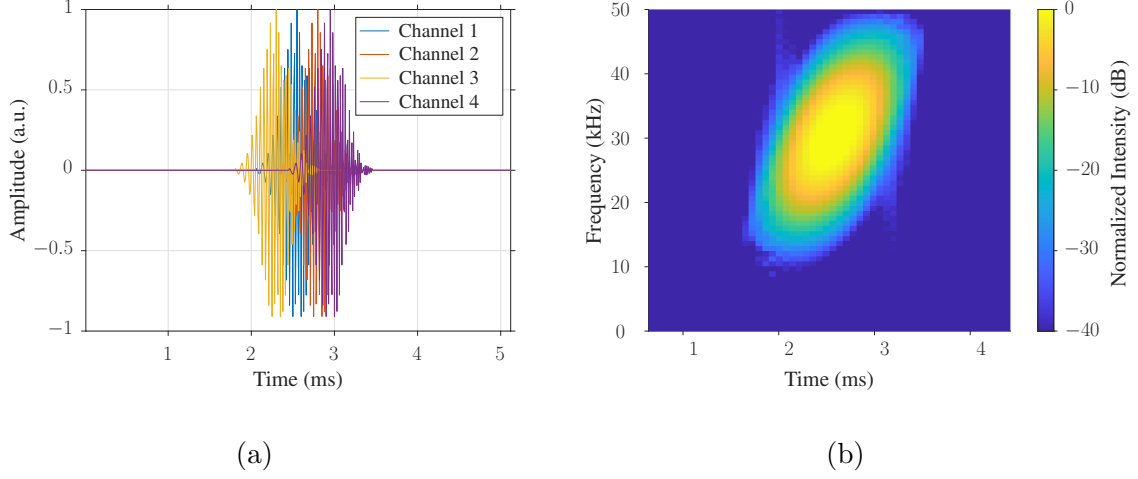


Figure 1.5. Example of a simulated odontocete echolocation click signal embedded in real background noise. (a) Time-domain waveforms across four hydrophone sensors, showing consistent impulsive structure with slight arrival time differences; (b) corresponding spectrogram of one sensor, highlighting the broadband nature of the click and its temporal localization. This simulated signal serves as a controlled input for evaluating the performance of TDOA estimation methods (XCORR, GCC-PHAT and GCC-WIN) presented in this chapter.

sensor-wise background noise power estimated from the real data. While the ambient noise is non-stationary, the use of identical noise realizations for both methods ensures a fair comparison.

1.7.6 Information Loss from TDOA-to-DOA Conversion

In practice, converting TDOAs to DOAs introduces potential information loss, particularly when some TDOA estimates are unreliable or missing due to low SNR; thus, even in the presence of some valid TDOA estimates no accurate DOA estimate can be computed. This motivates the question of whether TDOAs should be used directly for 3-D tracking rather than first converting them to DOAs. To investigate this, we analyze again the sequence of clicks in the recordings made from 10:53-11:13 on June 11th, 2018, where certain sensor pairs exhibit degraded TDOA estimates. We visualize both the raw TDOA measurements and the resulting DOAs to assess the effect of missing or noisy TDOAs on

angular consistency, and to highlight scenarios where using raw TDOAs would provide greater robustness.

1.8 Results

1.8.1 DOA Estimates

Fig. 1.6 presents elevation angle estimates from five DOA estimation methods, applied to the July 1st, 2018 dataset. Only signals exceeding their respective detection thresholds were included. Although some false positives are evident due to noise and threshold settings, two clear and contiguous tracks corresponding to goose-beaked whale echolocation clicks are observed across all methods except for that of CBF.

Among the TDOA-based approaches, both XCORR and GCC-WIN (Fig. 1.6 (c) and (e)) yielded high detection rates with coherent tracks, indicating precise localization. Although XCORR appears to exhibit tighter elevation angle clustering than GCC-WIN, this is primarily because GCC-WIN detects a larger number of weaker clicks under low SNR conditions. These low-SNR detections inherently yield less precise DOA estimates, resulting in a wider spread in elevation, yet they contribute to a more complete reconstruction of the click sequence over time. This is corroborated by the simulation study comparing these two methods in Ch. 1.8.3. GCC-PHAT also produced sharp tracks, but with noticeable gaps—clicks missed at higher thresholds were only recoverable at lower thresholds, which introduced more false positives.

The FDBF results followed the general trend of the TDOA-based results, but with a wider spread in elevation estimates. This imprecision becomes particularly problematic when two whales are in close proximity, e.g., at the beginning and end of the time considered dataset. The CBF results, based on clicks first detected using FDBF to reduce computational cost, show spatial aliasing artifacts. Detections appeared in discontinuous segments across the elevation axis.

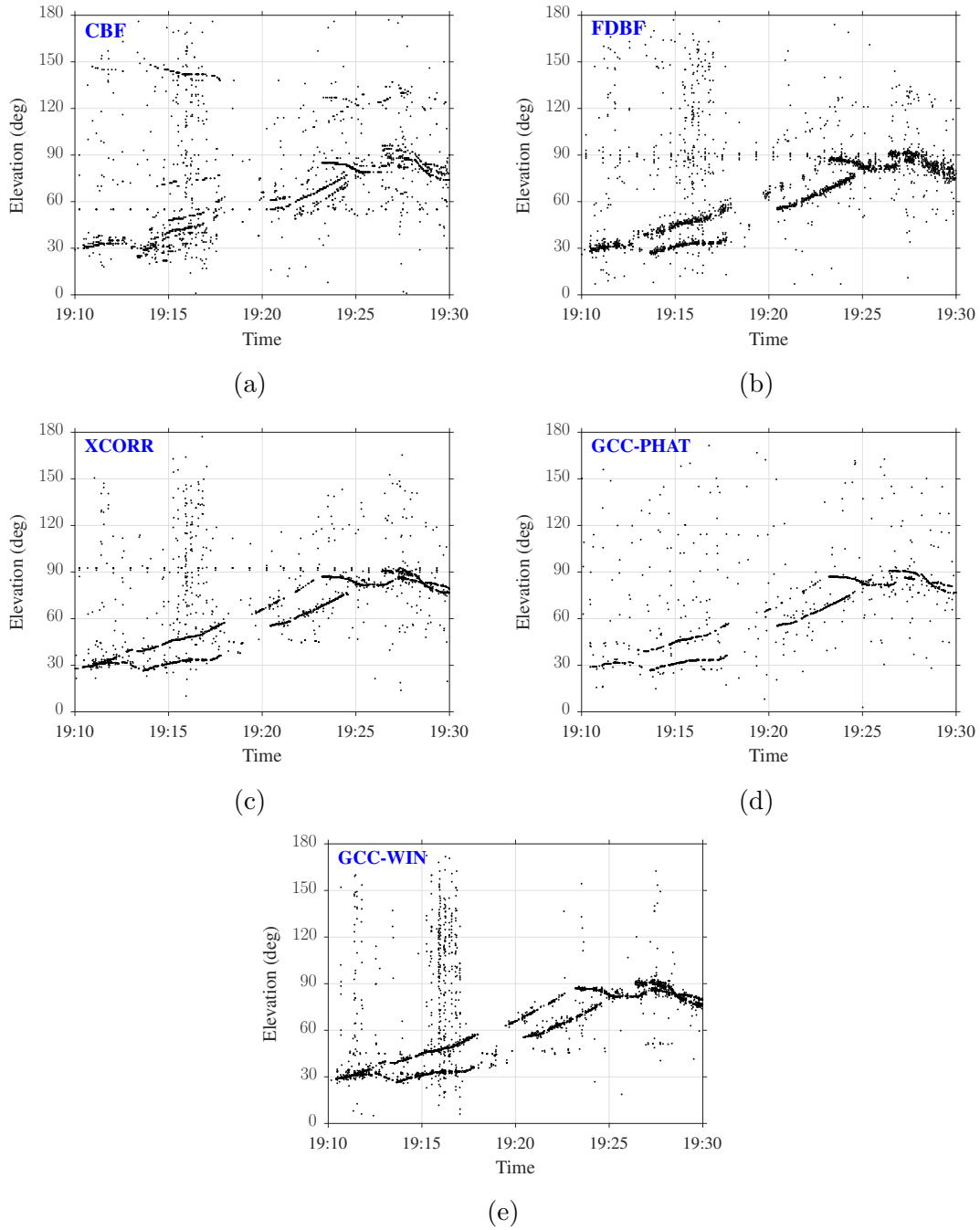


Figure 1.6. Elevation angle estimates obtained using five different DOA estimation methods applied to the same recordings of goose-beaked whales' echolocation clicks on July 1st, 2018. (a) Conventional beamforming (CBF); (b) Frequency-difference beamforming (FDBF); (c) Cross-correlation-based TDOA estimation (XCORR); (d) Generalized cross-correlation (GCC) with phase transform (PHAT); (e) GCC for whitening the instrument noise (GCC-WIN). While all methods reveal two main whale click tracks over time, the TDOA-based approaches (c) and (e) yield more continuous and precise DOA estimates.

1.8.2 TDOA Detection Rate Analysis

Given the superior performance of TDOA-based methods for impulsive, high-bandwidth signals like echolocation clicks, we further compared the impact of instrument noise filtering by analyzing detections from XCORR and GCC-WIN on a low SNR dataset (hydrophone sensors 3 and 4; June 11th, 2018).

Fig. 1.7 shows the results using a moderate threshold of $\alpha = 5$. The proposed GCC-WIN method revealed two temporally coherent and discernible TDOA tracks, while XCORR detected only a short segment of these tracks. Manual inspection of the time-domain waveforms confirmed that GCC-WIN detections corresponded to actual echolocation clicks, whereas many XCORR detections appeared spurious or misaligned.

At a lower threshold ($\alpha = 2$), Fig. 1.8 illustrates the limitations of standard cross-correlation. The XCORR method yielded dense clusters of detections, dominated by coherent instrument noise that masked the true signal structure. In contrast, GCC-WIN continued to suppress irrelevant peaks, isolating a more continuous and physically plausible track. This highlights the improved noise resilience of GCC-WIN and its ability to maintain detection quality under relaxed thresholding—an essential property for robust DOA estimation and tracking in practical field conditions.

1.8.3 Simulation Study

To further quantify the benefit of GCC-WIN, a simulation study was performed by injecting synthetic chirp signals, representing echolocation clicks, into noise-only segments of real data. The chirp swept 20–40 kHz over 1 ms and was inserted at known delays for each sensor pair. Detection rates and root-mean-square error (RMSE) were compared across a range of SNRs. Note that the SNR is not in the units of decibel.

As shown in Fig. 1.9, GCC-WIN consistently outperformed XCORR and GCC-PHAT, especially at low SNR. Notably, GCC-WIN detected approximately twice as many

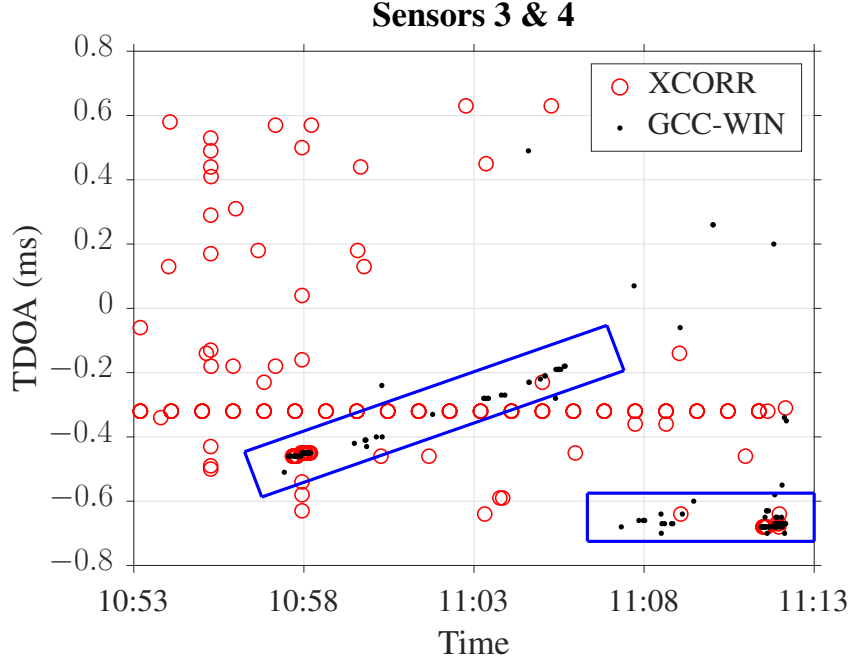


Figure 1.7. Comparison of TDOA estimates between hydrophone sensors 3 and 4 using two methods: traditional cross-correlation (XCORR, red circles) and the proposed GCC-WIN (black dots), after applying a 5-median absolute deviation ($5 \times \text{MAD}$) threshold. XCORR yields frequent spurious peaks and clusters at implausible values (the consistent detections at -0.3 ms are due to the coherent instrument noise). On the other hand, GCC-WIN produces two distinct, temporally consistent detection sequences (highlighted in blue rectangles, each corresponding to a different whale track). This demonstrates the increased robustness of GCC-WIN to noise and its improved ability to suppress irrelevant peaks in the cross-correlation function.

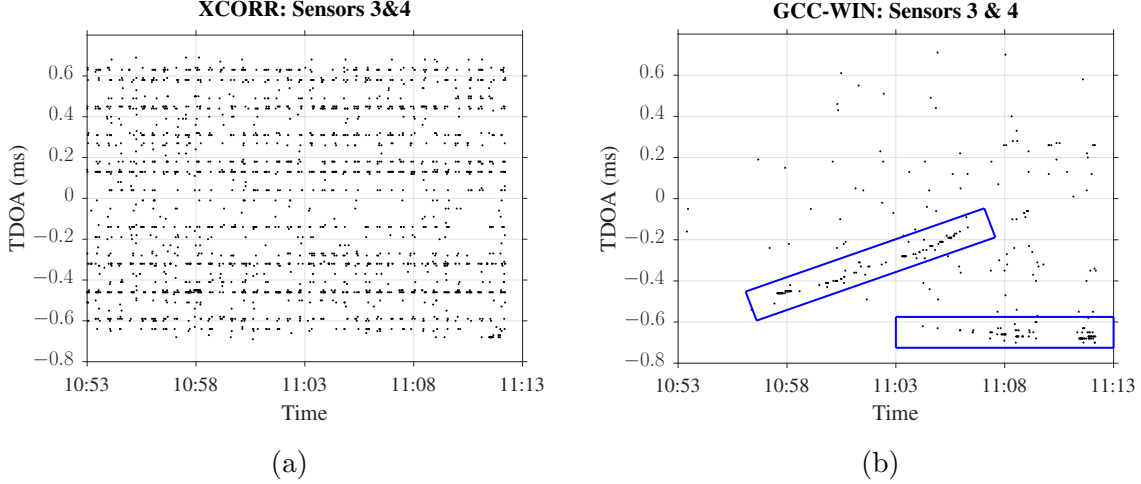


Figure 1.8. TDOA estimates between sensors 3 and 4 using (a) standard cross-correlation (XCORR) and (b) the proposed generalized cross-correlation method (GCC-WIN), after applying a 2-median absolute deviation (2MAD) threshold, lower than the threshold used in Fig. 1.7. The XCORR method yields dense clusters of spurious detections due to coherent instrument self-noise, masking the true temporal structure of the click arrivals. In contrast, GCC-WIN detects more clicks that form more continuous TDOA tracks, highlighting its improved noise resilience and suitability for downstream DOA and/or tracking applications. The two whale tracks identified by GCC-WIN are pointed out within the blue rectangles.

clicks as XCORR when $\text{SNR} < 2$, and it also converges to zero RMSE at $\text{SNR} = 3$, confirming its robustness in challenging conditions. The relatively poor performance of GCC-PHAT further underscores the importance of tailoring frequency-domain weighting to the specific noise characteristics; applying generic weighting without accounting for dominant noise sources can degrade TDOA estimation performance.

1.8.4 Information Loss in TDOA-to-DOA Conversion

To assess whether converting TDOA to DOA results in information loss, we analyzed a low SNR scenario using GCC-WIN with a $\alpha = 2$ threshold. Fig. 1.10 displays TDOA detections across all sensor pairs. Two coherent track segments—highlighted by red circles (sensors 2 and 3) and blue squares (sensors 3 and 4)—are clearly visible in some sensor pairs but absent or diffuse in others, indicating uneven detectability across the array.

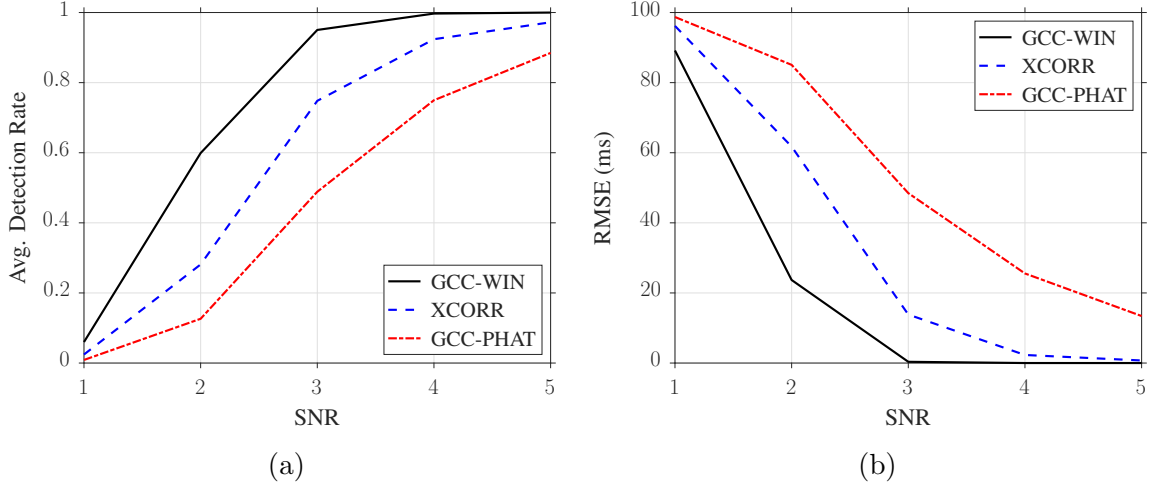


Figure 1.9. (a) Average detection rate of echolocation clicks with correct TDOA estimates for three methods (GCC-WIN, XCORR and GCC-PHAT) as a function of signal-to-noise ratio (SNR), based on simulated clicks injected into real noise (see Fig. 1.5). (b) Root-mean-square error (RMSE) of the corresponding TDOA estimates using the same simulation data. The proposed GCC-WIN method (black solid line) outperforms conventional cross-correlation (XCORR, red dashed line) and GCC-PHAT (blue dotted line), particularly at low SNR. These results demonstrate GCC-WIN’s robustness in detecting impulsive signals under realistic noise conditions and its improved accuracy for downstream localization.

When these TDOA detections are combined to estimate DOA (Fig. 1.11), the resulting elevation and azimuth angles (black dots) exhibit a wide spread, despite the presence of high-quality TDOA segments. Although the red and blue overlaid points (from reference TDOA tracks) remain tightly clustered, the overall dispersion highlights how noisy or spurious TDOA inputs from other sensor pairs can corrupt the final DOA estimate.

This illustrates a key limitation of converting TDOA measurements to DOA in low SNR conditions: a small number of noisy or spurious TDOAs can distort the resulting direction estimates. In Fig. 1.11, the dispersion of black DOA points contrasts with the more coherent reference tracks (red and blue), suggesting that the true directional structure is obscured. Such degradation in DOA quality poses a significant challenge for downstream localization and tracking processing, as tracking algorithms rely on temporally and spatially consistent direction estimates.

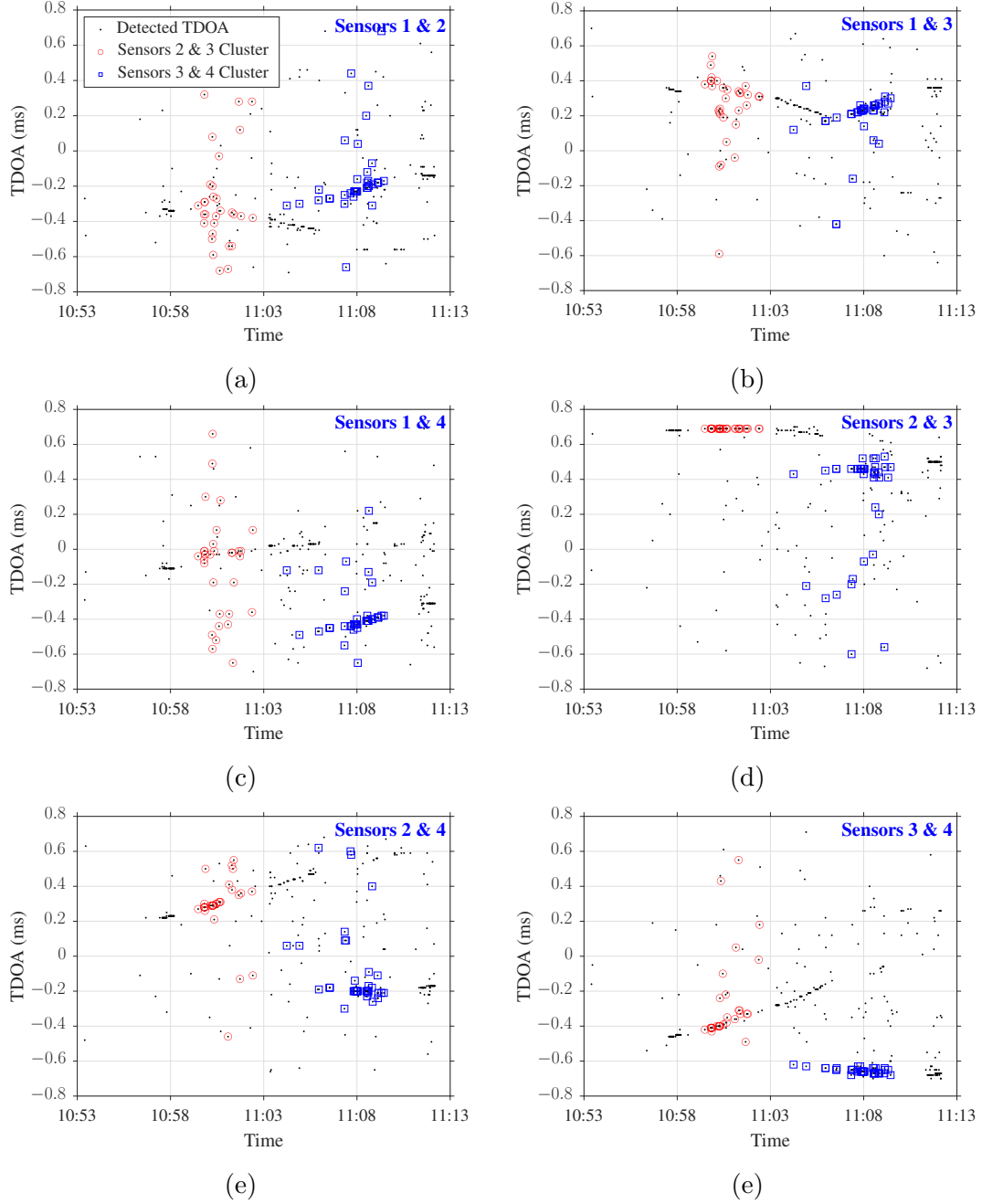


Figure 1.10. Detected TDOA values (black dots) for all hydrophone pairs using the GCC-WIN method. Red circles and blue squares are overlaid to indicate detections from two reference pairs—sensors 2 and 3 (d) and sensors 3 and 4 (f), respectively—that exhibit temporally coherent TDOA tracks. For each reference track, the corresponding time indices and detections are marked on all other sensor pairs (a–f) to examine whether similarly coherent TDOA detections are present. The dispersion of these highlighted points across non-reference sensor pairs shows that, under this low SNR scenario, false positives or inconsistent delays are common. Incorporating such noisy detections into DOA estimation (see Fig. 1.11) can lead to significant directional errors.

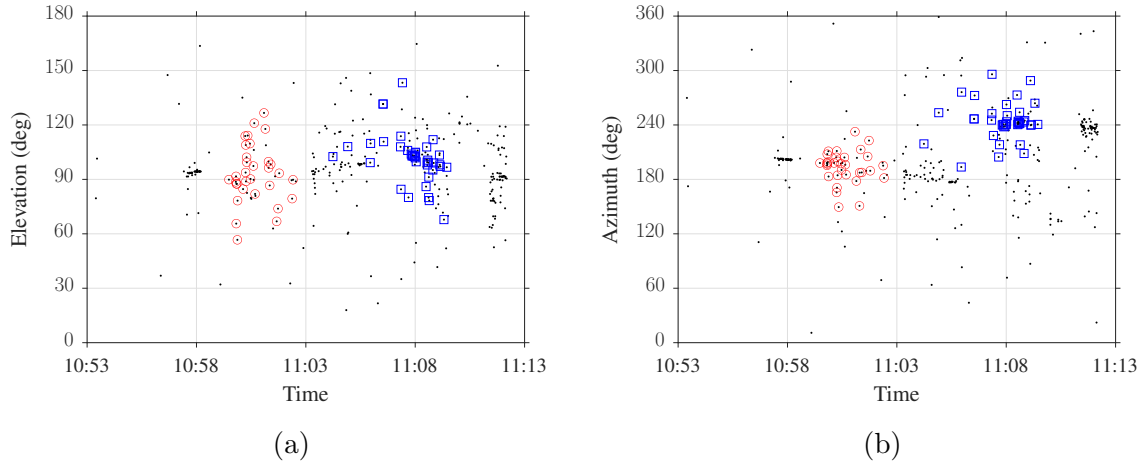


Figure 1.11. (a) Elevation and (b) azimuth angle estimates obtained by combining TDOA measurements from all sensor pairs using the GCC-WIN method in Fig. 1.10. Black dots represent DOA estimates for all detections. Red circles and blue squares are overlaid to mark the DOA results corresponding to the coherent TDOA tracks previously identified in sensor pair 2 & 3 and 3 & 4, respectively (see Fig. 1.10 (d) and (f)). The spread and inconsistency of these overlaid tracks illustrate how noisy TDOA inputs can distort directional estimates, even when some TDOA sensor pairs contain high-quality detections.

1.9 Conclusions

This chapter evaluated signal processing methods for detecting and estimating the direction of odontocete echolocation clicks recorded on sparse volumetric hydrophone arrays. We demonstrated that FDBF can be applied to impulsive, broadband echolocation clicks, and that it effectively mitigates the spatial aliasing issues inherent in conventional beamforming. However, due to its operation at low synthetic frequencies, FDBF suffers from poor angular resolution and higher measurement noise, which limits its utility for precise localization, especially in multi-source scenarios.

In contrast, methods based on TDOA yielded superior detection rates and directional estimates. Among these, the proposed GCC-WIN method, which applies a whitening filter derived from periodic instrument noise PSD templates, improved detection robustness under low SNR conditions. It consistently outperformed standard cross-correlation and

generalized cross-correlation with phase transform in both simulated and real datasets. Nevertheless, due to imperfect noise suppression, GCC-WIN remains susceptible to residual noise artifacts; hence, for high SNR signals, standard time-domain cross-correlation remains preferable due to its simplicity and precision.

We also showed that converting TDOA measurements into DOA estimates via least-squares fitting introduces notable information loss—particularly in cases where only a subset of hydrophone pairs yield reliable detections. This limitation is critical when dealing with sparse arrays and intermittent signals, as it can lead to spatially diffuse or erroneous DOA estimates even when some sensor pairs provide high-quality measurements. These findings motivate the direct use of raw TDOA measurements for downstream localization and tracking tasks, such as the multi-target tracking framework discussed in the next chapter, where maintaining temporal and spatial coherence across sensor pairs is essential for accurate track reconstruction.

Overall, this chapter underscores the importance of choosing processing methods that align with the nature of the signal (impulsive, broadband) and the limitations of the sensor geometry (sparse, volumetric). TDOA-based methods, particularly when enhanced with noise-adaptive preprocessing, offer a practical and robust foundation for the automated detection and tracking of echolocating odontocetes in passive acoustic monitoring systems.

Acknowledgments:

Chapters 1 and 2 are adapted and extended in part with permissions from: J. Jang, F. Meyer, E. Snyder, S. Wiggins, S. Baumann-Pickering, and J. Hildebrand, “Bayesian detection and tracking of odontocetes in 3-D from their echolocation clicks,” *J. Acoust. Soc. Am.*, 2023 (Copyright 2023, Acoustical Society of America), and J. Jang, F. Meyer, E. R. Snyder, S. M. Wiggins, S. Baumann-Pickering and J. A. Hildebrand, “Passive Acoustic Tracking of Whales in 3-D,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*,

June 2023 (Copyright 2023, IEEE). Some notation and structure have been revised for consistency and clarity within this dissertation. The dissertation author was the primary investigator and lead author of these works.

Chapter 2

Bayesian Tracking of Odontocetes in 3-D from Their Echolocation Clicks

2.1 Introduction

Accurate 3-D localization and tracking of echolocating odontocetes enable researchers to reconstruct underwater movement paths, estimate dive profiles, and examine behavioral patterns such as foraging, group coordination, and responses to anthropogenic noise [44, 6, 7, 45]. This is particularly important for deep-diving species like goose-beaked whales (*Ziphius cavirostris*), whose elusive behaviors are poorly understood through visual or surface-based monitoring alone. Therefore, developing robust methods for automated 3-D tracking from passive acoustic measurements is essential for advancing behavioral and ecological studies of odontocetes.

Ch. 1 introduced and compared PAM data processing methods that extract directional measurements—e.g., direction-of-arrival (DOA) and time-difference-of-arrival (TDOA)—of echolocating odontocetes from volumetric hydrophone arrays. It was shown that using a generalized cross-correlation (GCC) method that whitens instrument noise (GCC-WIN) yields higher detection rates and improved robustness against false positives in the presence of self-noise from the instruments. These directional measurements can be used to localize echolocating odontocetes in 3-D when appropriately combined across sensor pairs.

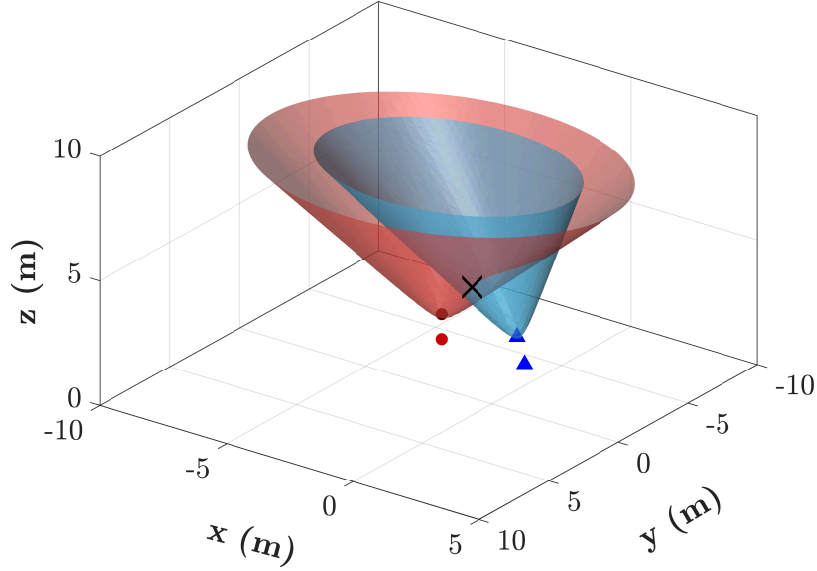


Figure 2.1. An example of acoustic source localization in 3-D from TDOA measurements. The cross is the location of the acoustic source. Each pair of dots and triangles represents a hydrophone pair that produces a TDOA measurement. Each TDOA measurement gives rise to a hyperboloid. The line resulting from the intersection of the two hyperboloids describes potential source locations. To obtain a unique source location, further hydrophone pairs are needed.

If the TDOA measurements are extracted from four hydrophone pairs and the positions of the hydrophones are known, the 3-D location of a single whale can, in principle, be computed by solving a nonlinear optimization problem [16]. In 3-D space, each TDOA measurement gives rise to a hyperboloid (Fig. 2.1). With more than two receivers, multiple TDOA measurements can be extracted, and all hyperboloids ideally intersect in a single point at the source location.

Multiple hydrophones typically form rigid volumetric arrays to facilitate deployment as well as data recording, storage, and processing. For geometric reasons, such arrays primarily provide bearing information for whales in the far field (see Ch. 1.4.3). Therefore, two or more arrays are often deployed for 3-D acoustic source localization. For example, in [5] and [46], a sequence of 3-D locations of echolocating beaked whales was estimated

from recordings of two high-frequency acoustic recording packages (HARPs) that host tetrahedron-shaped hydrophone arrays. For each snapshot of the acoustic signal, potential DOAs relative to each HARP were computed from the TDOA measurements [16]. Each whale was localized in 3-D using the least squares method based on manually selected DOAs likely generated by that whale. The sequence of each whale’s 3-D locations formed a track. Ch. 1, the analysis showed that in low signal-to-noise ratio (SNR) scenarios, degraded TDOA measurements can lead to inaccurate DOA estimates when using least squares. This motivates the development of a process that directly localizes whales in 3-D from the TDOA measurements.

In practice, the process of identifying and associating TDOA measurements corresponding to the same whale is complicated by noise, missed detections, and the presence of multiple simultaneously vocalizing individuals. Consequently, this task typically requires human operators to manually inspect acoustic data, determine the presence of whales, and select reliable measurements [47, 48, 5]. As a result, annotating and combining directional measurements remains labor-intensive and prone to human error.

These challenges necessitate a fully automated solution that can perform data association, manage uncertainty, and track multiple individuals over time. To this end, we turn to multi-target tracking (MTT) methods—Bayesian algorithms developed to infer the number and states of dynamic targets from sequences of noisy measurements provided by one or multiple sensors. In this chapter, a novel data processing chain is proposed that automatically detects and tracks odontocetes in 3-D from their echolocation clicks. The capabilities of the method are demonstrated through the tracking of goose-beaked whales near the coast of California.

2.1.1 Automated 3-D Tracking of Whales: Prior Work

Multi-target tracking (MTT) methods are sequential Bayesian estimation techniques that automatically infer the number and states of multiple targets from sequences of

measurements provided by one or multiple sensors, i.e., without the involvement of human operators. In the context of whale tracking, the states of interest can be the 3-D locations of the whales, the TDOA of whales for a particular hydrophone pair [20], or the frequency of narrow-band whale whistles in the spectrogram [49]. The MTT methods can succeed in tracking scenarios with false positive detections, missed detections, and data association uncertainty between measurements and targets. Traditional MTT methods such as the joint probabilistic data association filter [50] and multiple hypothesis tracker (MHT) [51], model measurements and target states as random vectors. Newer methods like the probabilistic hypothesis density (PHD) filter [52] and multi-Bernoulli filter [53, 54], are derived in the formalism of random finite set [53]. Recently, a graph-based MTT method that is highly scalable in the number of targets, measurements, and sensors has been proposed [55]. This approach uses particle-based computations to perform operations that can not be evaluated in closed form due to nonlinearities in the system model [56]. Graph-based MTT methods for localization and tracking from TDOA measurements in 2-D have been introduced in [57] and [58].

Some MTT methods have been successfully employed for whale tracking. In [59], the MHT is applied to tracking beaked whales in 3-D. Here, potential 3-D locations of beaked whales are preprocessed from TDOA measurements, which are associated across hydrophone pairs based on click characteristics [60]. The preprocessed 3-D locations are used as measurements for an MHT that determines the number of beaked whale tracks, performs data association of 3-D locations with tracks, and runs a Kalman filter for each track. The suboptimum computing of 3-D locations is necessary because the MHT is limited to linear measurement models or mildly nonlinear measurement models. Further inherent challenges of the MHT are its computational complexity and memory requirements [61]. In addition, a Gaussian mixture PHD (GM-PHD) filter [62] for the tracking of whales in the TDOA domain is introduced in [20]. Both echolocation clicks and whistles of false killer whales are exploited to compute TDOA measurements for MTT. The data were

acquired during line-transect surveys with towed hydrophone arrays. The introduced method extends the original GM-PHD filter by updating the existing and new whale tracks separately and by incorporating amplitude information to support the initialization of new whale tracks and to better reject false positive detections. However, the GM-PHD filter also relies on linear measurement models or mildly nonlinear measurement models and is therefore limited to tracking in the TDOA domain. Finally, [63] compares the graph-based MTT and the GM-PHD filter for tracking multiple singing humpback whales in azimuth using acoustic vector sensor data, demonstrating their respective performance on real recordings off Maui.

Various methods for the localization and tracking of whales from their acoustic signals have been developed. A common approach for the 3-D localization of whales is a grid-search [64, 65, 5, 66, 48], wherein an ambiguity surface is generated on a 3-D grid of potential whale locations by comparing the expected modeled TDOA measurements with the actual measurements. In [65] and [66], multiple whales are localized by determining local maxima of the ambiguity surface using the Simplex [67] or the Metropolis-Hastings [68] algorithm. The method in [65] iteratively finds whale locations in 3-D by selecting the maximum peak that corresponds to the best match of the TDOA measurements, removing the TDOA measurements corresponding to this maximum peak, and selecting the maximum peak that corresponds to the best match of the remaining TDOA measurements. In [66], Kalman filtering is used to estimate whale tracks. In the case of simultaneously present whales, multiple Kalman filters run in parallel, and the Hungarian algorithm [69] is used for associating local maxima of the ambiguity surface to Kalman filters. Alternative approaches perform 3-D localization of whales by first estimating the DOAs of acoustic signals from hydrophone arrays [22, 46, 70]. The method in [22] subsequently tracks individual whales using nonsequential estimation based on Gibbs sampling [71].

Common challenges of TDOA-based localizing and tracking are finding the correct combination of measurements across hydrophone pairs that correspond to the same

acoustic source and initializing whale tracks accordingly. Typically, there are multiple TDOA measurements per hydrophone pair due to the presence of noise, echoes, and simultaneous vocalization of multiple whales. This challenge is often referred to as a multisensor data association [56, 72]. To address this, existing techniques either employ human operators to select and combine measurements manually [5], rely on the local maxima corresponding to incorrectly matched TDOA measurements being significantly lower than those corresponding to true whale locations [65], compute potential whale locations in a brute-force manner, i.e., based on all possible combinations of TDOA measurements [66], or weigh grid points based on the number of TDOA measurements that are consistent with the corresponding potential whale location [60]. All existing methods for TDOA-based localizing and tracking of whales in 3-D either rely on human operators or heuristics to combine TDOA measurements and to initialize whale tracks.

2.1.2 Contributions, Chapter Organization, and Notation

The fundamental problem addressed in this chapter is establishing an algorithmic solution for the tracking of echolocating odontocetes in 3-D. The goal is to develop a data processing method that fully automatically determines the number of odontocetes in the environment and estimates their 3-D tracks from acoustic measurements. The proposed method will make it possible to (i) study deep-diving echolocating odontocetes more objectively and efficiently compared to approaches that rely on human operators and to (ii) reveal key insights on behaviors of odontocetes underwater that otherwise would remain unexplored.

The proposed data processing chain extracts TDOA measurements of echolocation clicks from the raw acoustic signals using a GCC. Introduced Ch. 1, a variant of the GCC referred to as whitening instrument noise (GCC-WIN) algorithm is utilized. This technique aims to suppress the instrument noise that interferes with the echolocation clicks. The peaks of the TDOAs above a certain amplitude threshold are used to estimate the

parameters of interest, i.e., locations and velocities of the odontocete in time. Odontocetes are first tracked in the TDOA domain using an MTT method based on the framework of factor graphs and the sum-product algorithm (SPA) [73]. A second MTT stage estimates odontocete tracks in 3-D by consistently combining (“fusing”) estimated TDOAs of all hydrophone pairs provided by the first stage. The increased detection rate of the GCC-WIN algorithm results in a lower probability of missing an echolocation click and, in turn, improved tracking performance. The first tracking stage aims to reject false positive TDOA measurements and at resolving longer gaps of missing TDOAs. This first MTT stage significantly improves the performance of estimating odontocete tracks in 3-D as performed by the second MTT stage.

Tracking whales in 3-D from TDOA measurements is further complicated because the underlying measurement model is nonlinear, and the state space is high-dimensional. To address this challenge, a SPA that embeds particle flow [74] is utilized. Here, the particles are actively migrated towards high likelihood regions, making it possible to obtain good target detection and tracking performance in high dimensions [75]. Contrary to existing methods for detecting and tracking whales in 3-D, the proposed signal processing chain systematically reduces the instrument noise and uses a statistical model for multisensor data association and initializing whale tracks. This is expected to improve detection and tracking performance, especially in scenarios with low SNR and a significant number of false positive measurements.

This chapter establishes a data processing chain that automatically detects and tracks odontocetes from acoustic measurements of their echolocation clicks. The key contributions of this chapter are summarized as follows.

- Two stages of graph-based MTT are established. The two MTT stages reject false positives, perform data association, determine the number of odontocetes, and estimate odontocete tracks in 3-D.

- The capabilities of the proposed data processing chain are demonstrated. For performance evaluation, recordings of echolocation clicks from two goose-beaked whales are considered.

Notation: Random variables are displayed in sans serif and upright fonts; their realizations in serif, italic fonts. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. For example, a random variable and its realization are denoted by $\mathbf{x}$ and x , respectively, and a random vector and its realization by $\mathbf{x}$ and $\mathbf{x}$, respectively. $\mathbf{x}_{0:k}$ is short for $[\mathbf{x}_0, \dots, \mathbf{x}_k]^T$. Furthermore, $\|\mathbf{x}\|$ and $\mathbf{x}^T$ denote the Euclidean norm and the transpose of vector $\mathbf{x}$, respectively; $\propto$ indicates equality up to a normalization factor; $f(\mathbf{x})$ denotes the probability density function (PDF) of random vector $\mathbf{x}$ (this is a short notation for $f_{\mathbf{x}}(\mathbf{x})$); $f(\mathbf{x}|\mathbf{y})$ denotes the conditional PDF of random vector $\mathbf{x}$ conditioned on random vector $\mathbf{y}$ (this is a short notation for $f_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y})$); $|\mathcal{S}|$ denotes the cardinality of set $\mathcal{S}$; $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. The operator $*$ denotes the complex conjugate, and $j = \sqrt{-1}$ is the imaginary unit.

2.2 Multi-target Tracking

A key challenge of MTT from sequences of measurements provided by one or multiple sensors is that the origin of measurements is typically unknown, i.e., it is not clear which target originated which TDOA measurement. This problem is referred to as measurement-origin-uncertainty (MOU). Moreover, since the number of targets is also unknown, it has to be estimated directly from the data.

2.2.1 MTT With Perfect Measurement-to-target Associations

Assuming that the origin of each measurement is perfectly known, i.e., measurement-to-target associations are either provided by a perfect human operator or a data association algorithm, the MTT problem can be split up into multiple parallel single target tracking

problems. Here, a sequential Bayesian estimation or Bayes filter [76, 77] is typically employed to estimate the state of each target individually and recursively. Define the target state and its associated measurements for all N_s sensors and at a discrete time step k , as random vectors $\mathbf{x}_k$ and $\mathbf{z}_k = [\mathbf{z}_{k,1}^T, \dots, \mathbf{z}_{k,N_s}^T]^T$, respectively. The target state typically consists of the target's position and motion-related parameters.

The objective is to estimate the target state, $\mathbf{x}_k$, from the available measurements up to time k , $\mathbf{z}_{1:k}$. Given the conditional PDF of the state given the measurements, $f(\mathbf{x}_k|\mathbf{z}_{1:k})$, the minimum mean-squared error (MMSE) estimate of the state of a single target, $\mathbf{x}_k$, can be found as [78]

$$\hat{\mathbf{x}}_k^{\text{MMSE}} = \int \mathbf{x}_k f(\mathbf{x}_k|\mathbf{z}_{1:k}) d\mathbf{x}_k. \quad (2.1)$$

To obtain $f(\mathbf{x}_k|\mathbf{z}_{1:k})$, one could naïvely marginalize the available joint PDF $f(\mathbf{x}_{0:k}|\mathbf{z}_{1:k})$. This approach, however, suffers from the curse of dimensionality since the dimension of $\mathbf{x}_{0:k}$ grows with each time step, and, as a result, the computational complexity of naïve marginalization increases exponentially and becomes intractable. The Bayes filter exploits that a first-order Markov process can describe the statistical model of single target tracking to reduce computations. At each time k , a prediction and N_s update steps are performed, and the resulting sequential processing schemes yields a computational complexity that is linear with time k [76, 77]. MTT methods are sequential Bayesian estimation methods that also consider MOU and the unknown number of states to be estimated.

2.2.2 MTT With MOU and Known Number of Targets

Consider an MTT problem with multiple sensors where the number of targets is known, but measurements are subject to MOU. In addition, there are false positives, i.e., measurements that have not been generated by any target, and missed detections,

i.e., present targets may not generate a measurement. It is assumed that there are $i \in \{1, \dots, N_t\}$ targets. At time k , the state of the i^{th} target is denoted as $\mathbf{x}_k^{(i)}$. For future reference, the notation $\mathbf{x}_k = [\mathbf{x}_k^{(1)}, \dots, \mathbf{x}_k^{(N_t)}]^T$ is introduced. Each target state evolves independently according to the Markovian state transition PDF, i.e., $f(\mathbf{x}_k | \mathbf{x}_{k-1}) = \prod_{i=1}^{N_t} f(\mathbf{x}_k^{(i)} | \mathbf{x}_{k-1}^{(i)})$.

Each sensor $s \in \{1, \dots, N_s\}$ produces $m_{k,s}$ TDOA measurements $\mathbf{z}_{k,s} = [\mathbf{z}_{k,s}^{(1)}, \dots, \mathbf{z}_{k,s}^{(m_{k,s})}]^T$. It is assumed that each measurement either originates from the target or is a false positive and that a target generates at most one measurement at each sensor. Measurement generation of target i at sensor s is modeled by a Bernoulli experiment characterized by the probability of detection $p_d^{(s)}(\mathbf{x}_k^{(i)})$. If the target with state $\mathbf{x}_k^{(i)}$ generates a measurement $\mathbf{z}_{k,s}^{(m)}$, the measurement is distributed according to $f(\mathbf{z}_{k,s}^{(m)} | \mathbf{x}_k^{(i)})$. The number of false positives at each time step is Poisson distributed with a mean $\mu_{\text{fp}}^{(s)}$. False positives are independent of the measurements that have originated from the targets. They are also independent and identically distributed (iid) according to the PDF $f_{\text{fp}}^{(s)}(\mathbf{z}_{k,s}^{(m)})$.

At time k and sensor s , the unknown association between measurements and targets is modeled by the latent random vector $\mathbf{a}_{k,s} = [\mathbf{a}_{k,s}^{(1)}, \dots, \mathbf{a}_{k,s}^{(N_t)}]^T$ that is composed of random variables, $\mathbf{a}_{k,s}^{(i)}$, defined as [50]

$$\mathbf{a}_{k,s}^{(i)} = \begin{cases} m \in \{1, \dots, m_{k,s}\} & \text{at time } k \text{ and sensor } s, \text{ target } i \\ & \text{generates measurement } m \\ 0 & \text{otherwise.} \end{cases} \quad (2.2)$$

For future reference, the joint vectors $\mathbf{a}_k = [\mathbf{a}_{k,1}, \dots, \mathbf{a}_{k,N_s}]^T$ and $\mathbf{z}_k = [\mathbf{z}_{k,1}, \dots, \mathbf{z}_{k,N_s}]^T$ are introduced.

The restriction that there can be at most one measurement associated with a target

at every time step can be checked by the following indicator function

$$\psi(\mathbf{a}_{k,s}) = \begin{cases} 0 & \exists i, i' \in \{1, \dots, N_t\} \text{ such that} \\ & i \neq i' \text{ and } a_{k,s}^{(i)} = a_{k,s}^{(i')} \neq 0 \\ 1 & \text{otherwise.} \end{cases} \quad (2.3)$$

The marginal posterior PDF, $f(\mathbf{x}_k^{(i)} | \mathbf{z}_{1:k})$, used for state estimation in Eq. 2.1 could be found by marginalization, i.e., $f(\mathbf{x}_k^{(i)} | \mathbf{z}_{1:k}) = \int \sum_{\mathbf{a}_{1:k}} f(\mathbf{x}_{0:k}, \mathbf{a}_{1:k} | \mathbf{z}_{1:k}) d\mathbf{x}_{\sim k}^{(i)}$ where $\mathbf{x}_{\sim k}^{(i)}$ is equal to $\mathbf{x}_{0:k}$ with state $\mathbf{x}_k^{(i)}$ removed. However, the computational complexity of this naïve marginalization would again be infeasible due to the reasons discussed above. To reduce computational complexity, one can again exploit the fact that the posterior PDF $f(\mathbf{x}_{0:k}, \mathbf{a}_{1:k} | \mathbf{z}_{1:k})$ follows a first-order Markov process, i.e.,

$$\begin{aligned} & f(\mathbf{x}_{0:k}, \mathbf{a}_{1:k} | \mathbf{z}_{1:k}) \\ & \propto f(\mathbf{x}_0) \prod_{k'=1}^k f(\mathbf{x}_{k'} | \mathbf{x}_{k'-1}) f(\mathbf{z}_{k'}, \mathbf{a}_{k'} | \mathbf{x}_{k'}) \\ & = f(\mathbf{x}_0) \prod_{k'=1}^k f(\mathbf{x}_{k'} | \mathbf{x}_{k'-1}) \prod_{s=1}^{N_s} f(\mathbf{z}_{k',s}, \mathbf{a}_{k',s} | \mathbf{x}_{k'}). \end{aligned} \quad (2.4)$$

Here, $f(\mathbf{x}_k | \mathbf{x}_{k-1})$ is the state-transition PDFs discussed above, $f(\mathbf{x}_0)$ is an arbitrary prior at time $k=0$, and $f(\mathbf{z}_{k,s}, \mathbf{a}_{k,s} | \mathbf{x}_k)$ is the conditional PDF that models the MOU measurement generation process. The last line of Eq. 2.4 uses the fact that the measurement generation given the target states is independent across the sensors. Note that $f(\mathbf{z}_{k,s}, \mathbf{a}_{k,s} | \mathbf{x}_k)$ is a function of $p_d^{(s)}(\mathbf{x}_k^{(i)})$, $\mu_{\text{fp}}^{(s)}$, $f_{\text{fp}}^{(s)}(\mathbf{z}_{k,s}^{(m)})$, and $f(\mathbf{z}_{k,s}^{(m)} | \mathbf{x}_k^{(i)})$ (see [50] for details). Based on the factorization of the statistical model in Eq. 2.4, e.g., a sequential Bayesian estimation approach referred to as probabilistic data association filter [50] can be developed.

2.2.3 MTT With MOU and Unknown Number of Targets

In real-world scenarios, such as the whale tracking problem, the number of targets is time-varying and unknown. To account for this, potential target (PT) states can be introduced [55]. Consider PTs with indexes $j \in \{1, \dots, j_k\}$, where j_k is the number of PTs at time k . A binary variable, $\mathbf{r}_k^{(j)} \in \{0, 1\}$, indicates the existence of the PT j , where $\mathbf{r}_k^{(j)} = 1$ if and only if PT j exists. The augmented state of PT j is given by $\mathbf{y}_k^{(j)} = [\mathbf{x}_k^{(j)\top} \mathbf{r}_k^{(j)}]^\top$ where $\mathbf{x}_k^{(j)}$ consists of the target's position and further motion-related parameters. There are two types of PTs:

- *New PTs* represent targets that, for the first time, have generated a measurement. Their states are denoted by $\bar{\mathbf{y}}_{k,s}^{(j)} = [\bar{\mathbf{x}}_{k,s}^{(j)\top} \bar{\mathbf{r}}_{k,s}^{(j)}]^\top$. At time k and sensor s , a new PT is introduced for each measurement $m \in \{1, \dots, m_{k,s}\}$.
- *Legacy PTs* represent targets that already have generated at least one measurement at a previous time step $k' < k$ or previous sensor $s' < s$. Their states are denoted by $\underline{\mathbf{y}}_{k,s}^{(j)} = [\underline{\mathbf{x}}_{k,s}^{(j)\top} \underline{\mathbf{r}}_{k,s}^{(j)}]^\top$.

The vectors that consist of all legacy and new PT states are denoted by $\underline{\mathbf{y}}_k$ and $\bar{\mathbf{y}}_k$, respectively, and the vector that consists of all PT state at time k is represented as $\mathbf{y}_k = [\underline{\mathbf{y}}_k \bar{\mathbf{y}}_k]^\top$.

At each time k , the number of targets that for the first time have generated a measurement at sensors s are Poisson distributed with mean $\mu_n^{(s)}$. The states of these newly detected targets are iid with the PDF $f_n^{(s)}(\mathbf{x}_k^{(j)})$. Newly detected targets are statistically independent of existing targets. A PT j that existed at time $k-1$ continues to exist at time k with survival probability $p_{\text{su}}(\mathbf{x}_k^{(j)})$. All PTs at time $k-1$ become legacy PTs at time k .

To reduce computational complexity, one can again exploit structure in the factorization of the posterior PDF $f(\mathbf{y}_{0:k}, \mathbf{a}_{1:k} | \mathbf{z}_{1:k})$. In particular, using common Markov

assumptions [55], $f(\mathbf{y}_{0:k}, \mathbf{a}_{1:k} | \mathbf{z}_{1:k})$ factorizes according to

$$\begin{aligned} f(\mathbf{y}_{0:k}, \mathbf{a}_{1:k} | \mathbf{z}_{1:k}) &\propto f(\mathbf{z}_{1:k}, \mathbf{a}_{1:k}, \mathbf{y}_{0:k}) \\ &= f(\mathbf{y}_0) \prod_{k'=1}^k f(\mathbf{z}_{k'}, \mathbf{a}_{k'}, \mathbf{y}_{k'} | \mathbf{y}_{k'-1}). \end{aligned} \quad (2.5)$$

Upon explicitly distinguishing between the legacy and new PTs and by exploiting the chain rule for PDFs, the conditional PDF of current measurements, current association variables, and current states given the previous states can be expanded as

$$\begin{aligned} f(\mathbf{z}_k, \mathbf{a}_k, \mathbf{y}_k | \mathbf{y}_{k-1}) \\ &= f(\mathbf{z}_k, \mathbf{a}_k, \underline{\mathbf{y}}_k, \overline{\mathbf{y}}_k | \mathbf{y}_{k-1}) \\ &= f(\underline{\mathbf{y}}_k | \mathbf{y}_{k-1}) f(\mathbf{z}_k, \mathbf{a}_k, \overline{\mathbf{y}}_k | \underline{\mathbf{y}}_k, \mathbf{y}_{k-1}). \end{aligned} \quad (2.6)$$

Now, one can use the fact that given the legacy PT states at time k , (i) the new PT states at time k are conditionally independent of the previous PT states at time $k-1$ and that (ii) the measurements, association variables, and new PTs are statistically independent across the N_s sensors, i.e.,

$$\begin{aligned} f(\mathbf{z}_k, \mathbf{a}_k, \mathbf{y}_k | \mathbf{y}_{k-1}) \\ &= f(\underline{\mathbf{y}}_k | \mathbf{y}_{k-1}) f(\mathbf{z}_k, \mathbf{a}_k, \overline{\mathbf{y}}_k | \underline{\mathbf{y}}_k). \\ &= f(\underline{\mathbf{y}}_k | \mathbf{y}_{k-1}) \prod_{s=1}^{N_s} f(\mathbf{z}_{k,s}, \mathbf{a}_{k,s}, \overline{\mathbf{y}}_{k,s} | \underline{\mathbf{y}}_{k,s-1}), \end{aligned} \quad (2.7)$$

where $\underline{\mathbf{y}}_{k,0} = \underline{\mathbf{y}}_k$ and $\underline{\mathbf{y}}_{k,s} = [\underline{\mathbf{y}}_{k,0}^T, \overline{\mathbf{y}}_{k,1}^T, \dots, \overline{\mathbf{y}}_{k,s-1}^T]^T$.

In addition, targets move in time independently, and therefore, the state-transition

PDF of the legacy PTs reads

$$f(\underline{\mathbf{y}}_k | \mathbf{y}_{k-1}) = \prod_{j=1}^{j_{k-1}} f(\underline{\mathbf{y}}_k^{(j)} | \mathbf{y}_{k-1}^{(j)}). \quad (2.8)$$

The state-transition model $f(\underline{\mathbf{y}}_k^{(j)} | \mathbf{y}_{k-1}^{(j)})$ is a function of the survival probability $p_{\text{su}}(\mathbf{x}_k^{(j)})$.

By plugging Eq. 2.8 into Eq. 2.7 and, in turn, Eq. 2.7 into Eq. 2.5, and by making use of the functional form of $f(\mathbf{z}_{k,s}, \mathbf{a}_{k,s}, \bar{\mathbf{y}}_{k,s} | \mathbf{y}_{k,s-1})$ (see [55] for details), the joint posterior PDF of $\mathbf{y}_{1:k}$ and $\mathbf{a}_{1:k}$ given an observed $\mathbf{z}_{1:k}$ becomes

$$\begin{aligned} & f(\mathbf{y}_{1:k}, \mathbf{a}_{1:k} | \mathbf{z}_{1:k}) \\ & \propto \prod_{k'=1}^k \left(\prod_{j'=1}^{j_{k'}-1} f(\underline{\mathbf{y}}_{k'}^{(j')} | \mathbf{y}_{k'-1}^{(j')}) \right) \prod_{s=1}^{N_s} \psi(\mathbf{a}_{k',s}) \\ & \times \left(\prod_{j=1}^{j_{k',s}} q(\underline{\mathbf{x}}_{k',s}^{(j)}, \underline{r}_{k',s}^{(j)}, a_{k',s}^{(j)}; \mathbf{z}_{k',s}) \right) \\ & \times \prod_{m=1}^{m_{k',s}} v(\bar{\mathbf{x}}_{k',s}^{(m)}, \bar{r}_{k',s}^{(m)}, a_{k',s}^m). \end{aligned} \quad (2.9)$$

Here, the factors $q(\underline{\mathbf{x}}_{k,s}^{(j)}, \underline{r}_{k,s}^{(j)}, a_{k,s}^{(j)}; \mathbf{z}_{k,s})$ and $v(\bar{\mathbf{x}}_{k,s}^{(m)}, \bar{r}_{k,s}^{(m)}, a_{k,s}^m)$ are functions of $p_{\text{d}}^{(s)}(\mathbf{x}_k^{(j)})$, $\mu_{\text{fp}}^{(s)}$, $f_{\text{fp}}^{(s)}(\mathbf{z}_{k,s}^{(m)})$, and $f(\mathbf{z}_{k,s}^{(m)} | \mathbf{x}_k^{(j)})$ (see [55] for details).

For target detection and estimation, the marginal PDFs $f(\mathbf{x}_k^{(j)}, r_k^{(j)} | \mathbf{z}_{1:k}) = f(\mathbf{y}_k^{(j)} | \mathbf{z}_{1:k})$ are required. In particular, target detection is performed by introducing a threshold p_{th} that is compared with the existence probability $p(r_k^{(j)} = 1 | \mathbf{z}_{1:k})$, i.e., PT $j \in \{1, \dots, j_k\}$ is declared to exist if $p(r_k^{(j)} = 1 | \mathbf{z}_{1:k}) > p_{\text{th}}$. Note that $p(r_k^{(j)} = 1 | \mathbf{z}_{1:k}) = \int f(\mathbf{x}_k^{(j)}, r_k^{(j)} = 1 | \mathbf{z}_{1:k}) d\mathbf{x}_k^{(j)}$. For PTs declared to exist, state estimation is performed by computing the minimum-mean-squared error estimate [78] as

$$\hat{\mathbf{x}}_k^{(j)} \triangleq \int \mathbf{x}_k^{(j)} f(\mathbf{x}_k^{(j)} | r_k^{(j)} = 1, \mathbf{z}_{1:k}) d\mathbf{x}_k^{(j)}, \quad (2.10)$$

where

$$f(\mathbf{x}_k^{(j)} | r_k^{(j)} = 1, \mathbf{z}_{1:k}) = \frac{f(\mathbf{x}_k^{(j)}, r_k^{(j)} = 1 | \mathbf{z}_{1:k})}{p(r_k^{(j)} = 1 | \mathbf{z}_{1:k})}. \quad (2.11)$$

For direct computation of $f(\mathbf{y}_k | \mathbf{z}_{1:k})$, one could again naïvely marginalize the available joint PDF $f(\mathbf{y}_{1:k}, \mathbf{a}_{1:k} | \mathbf{z}_{1:k})$. This marginalization can be performed efficiently by the framework of factor graphs and the SPA. The complete system model and the SPA for MTT can be found in [55].

The nonlinear measurement model and high-dimensional state space impose further challenges to tracking in 3-D from TDOA measurements, and therefore, a SPA that embeds particle flow is employed. To perform the SPA effectively, particles are migrated towards regions of high likelihood based on the solution of a partial differential equation. This makes it possible to obtain good target detection, and tracking performance in 3-D [75].

2.3 The Proposed Data Processing Chain

The proposed data processing chain performs two main tasks: (i) signal processing and (ii) parameter estimation (Fig. 2.2). In the signal processing module, pre-filters are first applied to the raw acoustic signal. Then, the GCC-WIN technique introduced in Ch. 1.4.2 is performed to extract time delay peaks with the GCC-WIN amplitudes above a threshold.

The parameter estimation module estimates the locations and velocities of the echolocating odontocetes. The odontocetes are first tracked in the TDOA domain and then in 3-D. In both cases, the tracking algorithm based on the SPA described in Ch. 2.2.3 is applied. In the TDOA domain, both the motion and the measurement models are linear. Tracking first in the TDOA domain makes it possible to reduce the number of false positives significantly. A low number of false positives is essential for successful tracking in 3-D. Finally, the output of the TDOA tracker is used as input for odontocete tracking in 3-D.

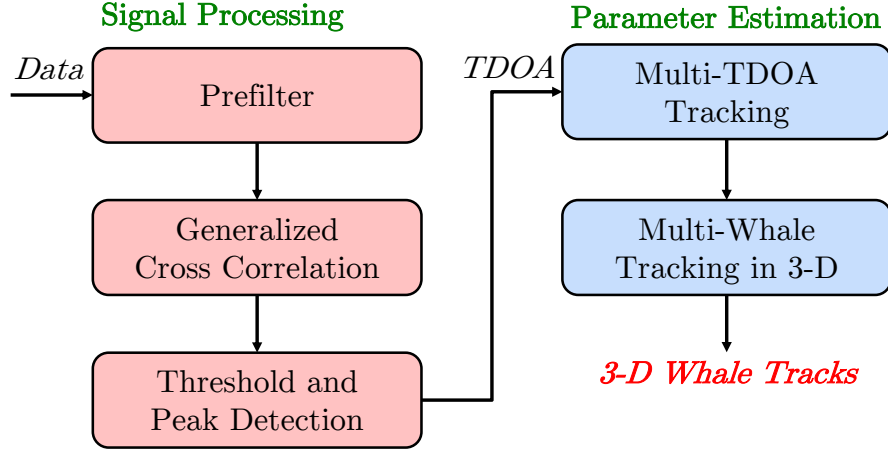


Figure 2.2. The block diagram of the proposed data processing chain for detecting and tracking odontocetes in 3-D from their echolocation clicks. The TDOAs of echolocation clicks are extracted in the signal processing module. The 3-D odontocete tracks are computed in the parameters estimation module. The parameter estimation module performs sequential Bayesian estimation for MTT in two stages.

2.3.1 Echolocation Click Detection and TDOA Estimates

The echolocation clicks are detected via performing GCC-WIN across pairs of hydrophones and considering TDOA peaks whose amplitudes are above a threshold. To enhance the probability of detection, TDOA measurements can be either (i) computed based on a sequence of click trains [65, 59, 44] or (ii) obtained by extracting TDOAs of individual echolocation clicks and combining them into a single measurement. In the presence of highly-correlated noise, the former approach is unsuitable; hence, the latter approach is considered.

To combine the detected TDOAs of individual click signals into a single measurement, it is assumed that the odontocete is stationary over a time interval in samples, T_m . Let T_g be the interval between each snapshot at which the GCC-WIN is performed, and we choose T_m that is longer than T_g , i.e., $T_m > T_g$. The accumulation duration T_m will be the duration of the time between the discrete time step k of the considered tracking algorithms. The peaks of multiple observation intervals of length T_g are then accumulated

over a time interval of duration T_m . The resulting set of TDOAs $\mathbf{z}_{k,s}^{(m)}$, $m \in \{1, \dots, m_{k,s}\}$, where $m_{k,s}$ is the number of measurements at time step k and sensor s , is considered as TDOA measurements of echolocation clicks generated by odontocetes. The TDOA measurements of all sensors are used as input for MTT.

Accumulating TDOAs, however, can lead to multiple TDOAs corresponding to one odontocete per time step k due to noise. Hence, a clustering of measurements is necessary to enforce the assumption that each odontocete generates at most one measurement. The following clustering procedure led to the best tracking results. First, large clusters are formed by finding TDOAs that are N_c samples apart. Within each large cluster, TDOAs are further separated into smaller clusters based on the local minima of the GCC-WIN amplitudes. The TDOAs within each small cluster are weighted based on their amplitudes and merged to generate a single TDOA measurement per small cluster.

2.3.2 MTT of Echolocating Odontocetes

In this section, details on how the MTT framework presented in Ch. 2.2.3 is applied to odontocete tracking are provided. In particular, the statistical models of the two MTT stages are discussed. The first stage operates in the TDOA domain and is applied to each hydrophone pair, i.e., a TDOA sensor, in parallel. The second stage operates in a 3-D Cartesian coordinate system and fuses the results provided by the first stage.

MTT in the TDOA Domain

For the tracking in the TDOA domain, at time step k and sensor s , the state of odontocetes j is given by $\mathbf{d}_{k,s}^{(j)} = [\mathbf{d}_{k,s}^{(j)} \dot{\mathbf{d}}_{k,s}^{(j)}]^T$, where $\mathbf{d}_{k,s}^{(j)}$ is the true TDOA of the echolocation clicks emitted by the odontocete, and $\dot{\mathbf{d}}_{k,s}^{(j)}$ is its rate of change. A nearly constant velocity

motion model [76] is considered, i.e.,

$$\mathbf{d}_{k,s}^{(j)} = \begin{bmatrix} 1 & T_m \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{d}_{k-1,s}^{(j)} \\ \dot{\mathbf{d}}_{k-1,s}^{(j)} \end{bmatrix} + \mathbf{u}_{k,s}^{(j)} \quad (2.12)$$

where the driving noise $\mathbf{u}_{k,s}^{(j)} \in \mathbb{R}^2$ is a zero-mean multivariate Gaussian random vector with covariance matrix

$$\Sigma_{\mathbf{u}} = \begin{bmatrix} \frac{T_m^3}{3} & \frac{T_m^2}{2} \\ \frac{T_m^2}{2} & T_m \end{bmatrix} \sigma_{\mathbf{u}}^2 \quad (2.13)$$

and driving noise standard deviation (STD) $\sigma_{\mathbf{u}}$.

The additive-Gaussian measurement model is given as follows. Let the TDOA measurement $\mathbf{z}_{k,s}^{(m)}$ be originated from the odontocete with index j at sensor s . The measurement model then reads

$$\mathbf{z}_{k,s}^{(m)} = \mathbf{d}_{k,s}^{(j)} + \mathbf{v}_{k,s}^{(m)}. \quad (2.14)$$

Here, $\mathbf{v}_{k,s}^{(m)}$ is a zero-mean Gaussian measurement noise with STD σ_v . The PDF that characterizes false positives, $f_{\text{fp}}^{(s)}(z_{k,s}^{(m)})$, is uniform on the interval $[-T_s^{\max}, T_s^{\max}]$, where $T_s^{\max}$ is the maximum time delay that can be measured by sensor s , i.e.,

$$T_s^{\max} = \|\mathbf{q}_{s_1} - \mathbf{q}_{s_2}\|/c \quad (2.15)$$

where $\mathbf{q}_{s_1} \in \mathbb{R}^3$ and $\mathbf{q}_{s_2} \in \mathbb{R}^3$ are the positions of the hydrophone pair (s_1, s_2) that defines sensor s , and c is the sound velocity.

In the considered linear MTT problem, there is MOU, and the number of odontocetes is unknown and time-varying. Thus, the SPA-based MTT method described in Ch. 2.2.3 is employed. The results of this first stage are sets of TDOA estimates $\hat{d}_{k,s}^{(j)}$, $j \in \{1, \dots, j_{k,s}\}$ for

each time step k and each sensor s . These TDOA estimates are then used as measurements in the 3-D MTT tracking stage. In what follows, the aforementioned set of TDOA estimates are denoted as $\hat{d}_{k,s}^{(m)}$, $j \in \{1, \dots, m_{k,s}\}$ to indicate that they are now used as measurements in the second MTT stage. For simplicity, it is assumed that the measurements provided by different TDOA sensors are statistically independent of each other. Note, however, that dependencies among the measurements provided by TDOA sensors do exist. This is because a TDOA sensor consists of a pair of hydrophones, and a hydrophone is involved in multiple TDOA sensors. Thus, the acoustic data of a single hydrophone is used in the measurements of multiple TDOA sensors.

MTT in 3-D

With the TDOA estimates available across all sensors, odontocetes are tracked in 3-D. The 3-D tracking method performs multisensor data association and track initialization. At time k , the state of the odontocete j is denoted as $\mathbf{p}_k^{(j)} = [\mathbf{p}_{k,x}^{(j)} \ \mathbf{p}_{k,y}^{(j)} \ \mathbf{p}_{k,z}^{(j)} \ \dot{\mathbf{p}}_{k,x}^{(j)} \ \dot{\mathbf{p}}_{k,y}^{(j)} \ \dot{\mathbf{p}}_{k,z}^{(j)}]^\top$, where $\mathbf{p}_{k,x}^{(j)}$, $\mathbf{p}_{k,y}^{(j)}$, and $\mathbf{p}_{k,z}^{(j)}$ are the position of the odontocete in a 3-D Cartesian coordinate system and $\dot{\mathbf{p}}_{k,x}^{(j)}$, $\dot{\mathbf{p}}_{k,y}^{(j)}$ and $\dot{\mathbf{p}}_{k,z}^{(j)}$ are the respective velocities.

The motion model follows the nearly constant velocity model [76]

$$\mathbf{p}_k^{(j)} = \begin{bmatrix} \mathbf{I}_3 & T_m \mathbf{I}_3 \\ \mathbf{0}_3 & \mathbf{I}_3 \end{bmatrix} \mathbf{p}_{k-1}^{(j)} + \begin{bmatrix} \frac{T_m^2}{2} \mathbf{I}_3 \\ T_m \mathbf{I}_3 \end{bmatrix} \mathbf{w}_k^{(j)} \quad (2.16)$$

where $\mathbf{w}_k^{(j)} \in \mathbb{R}^3$ is a zero-mean Gaussian driving noise with covariance $\mathbf{I}_3 \sigma_{\mathbf{w}}^2$ and driving noise STD $\sigma_{\mathbf{w}}$.

Suppose that the odontocete j generated the TDOA $\hat{\mathbf{d}}_{k,s}^{(m)}$ at time k and sensor s . The corresponding TDOA model is given by

$$\hat{\mathbf{d}}_{k,s}^{(m)} = (\|\mathbf{p}_k^{(j)} - \mathbf{q}_{s1}\| - \|\mathbf{p}_k^{(j)} - \mathbf{q}_{s2}\|)/c + \mathbf{b}_{k,s}^{(m)} \quad (2.17)$$

where $\mathbf{b}_{k,s}^{(m)}$ is zero-mean Gaussian measurement noise with STD σ_b . The PDF that characterizes false positives, $f_{\text{fp}}^{(s)}(z_{k,s}^{(m)})$, is again uniform on the interval $[-T_s^{\max} \ T_s^{\max}]$.

Note that the nonlinear measurement model in Eq. 2.17 is underdetermined, i.e., the position of the odontocete is 3-D while the TDOA is only 1-D. Therefore, for estimating the 3-D position, TDOAs provided by multiple sensors have to be fused. In particular, for a fixed TDOA $\hat{d}_{k,s}^{(m)}$, Eq. 2.17 describes potential odontocete locations on a hyperboloid. If there were no MOU and no noise, one could estimate the 3-D odontocete location by computing the intersection of multiple hyperboloids provided by multiple TDOA sensors. However, due to the presence of MOU and unknown number of odontocetes, reliable state estimation can only be performed sequentially. In particular, the MTT approach reviewed in Ch. 2.2.3 is used again. Due to the nonlinear measurement model Eq. 2.17, a particle-based implementation of this MTT approach is considered. However, since the conventional particle filter suffers from weight degeneracy due to the curse of dimensionality [79] and is thus only suitable for low dimensional state spaces, the particle flow variant recently proposed in [75] is used.

Particle flow techniques are attractive methods that address the weight degeneracy issues [74]. For each odontocete j and time k , particles are migrated iteratively to the posterior distribution. Particle motion follows a stochastic process described by an ODE that expresses a Bayesian update step. The flow equations and the particles after the flow describe a proposal distribution that can be used for importance sampling [80]. By embedding particle flow into SPA-based MTT, the weight degeneracy is avoided, and MTT in 3-D from TDOA measurements can be performed with a reasonable number of particles and computational complexity [75].

2.4 Real Data Application

Using the same dataset introduced in Ch. 1.5, four encounters of goose-beaked whales are processed, and in this chapter, the results from processing the data measured on June 11th and July 1st in 2018 are presented. The encounter durations on those two dates are 52 min long and 20 min long, respectively. These encounters were detected using the long-term spectral average (LTSA) from the MATLAB-based program *Triton* [41].

Since the hydrophones on each HARP share the same clock, they are time-synchronized. The two arrays (HARP EE and EW), however, are not synchronized, and therefore, the TDOA measurements are extracted only from the hydrophone pairs on the same arrays. Nonetheless, as long as the echolocation clicks are measured on both HARPs within a time window in which the goose-beaked whale can be considered stationary, precise synchronization is unnecessary [59, 22].

The tracking capability of the proposed data processing chain is demonstrated on acoustic datasets containing echolocation clicks from goose-beaked whales. In the signal processing step, their echolocation clicks are detected, and the corresponding TDOA measurements are computed using the GCC-WIN algorithm described in Ch. 1.4.2. In the parameter estimation step, the whales are tracked first in the TDOA domain and then in the 3-D domain using the implementation of the MTT framework described in Ch. 2.3.2. The 3-D tracking results are compared to the tracks generated from hand-annotated DOA measurements following the approach in [46].

2.4.1 Implementation Aspects

Signal Processing Module: TDOA Extraction

The peaks from the GCC-WIN, whose amplitudes are more significant than $A_{\text{tdoa}} = 0.15$, are extracted and saved along with their amplitude information. Note that since a low amplitude threshold is applied to maximize the detection rate, both the instrument and

Table 2.1. Hyperparameters used for the tracking of multiple goose-beaked whales in the TDOA and the 3-D domain.

Hyperparameters	TDOA	3-D
Detection probability, p_d	0.80	0.80
Survival probability, p_{su}	0.90	0.99
Mean number of false positives, μ_{fp}	10	1
Mean number of newly appearing whales	1.0×10^{-4}	1
Measurement noise STD, σ_v & σ_b	1.0×10^{-5}	3.0×10^{-5}
Driving noise STD, σ_u & σ_w	1.5×10^{-7}	1.0×10^{-2}
Number of particles	30,000	100,000
Minimum track length	20	5

white noise in the background can give rise to multiple false positive TDOA measurements. When the GCC-WIN peak amplitude is more than 10, i.e., an echolocation click likely in the direct line of sight is detected, any pulse-like signals within the next 40 ms are ignored since they are likely to be multipath signals.

The TDOA measurements are merged and binned with a longer discrete time step length of $T_m=7$ s and clustering distance $N_c=2$ samples, following the method described in Ch. 2.3.1. Given that inter-click-intervals (ICIs) of goose-beaked whales range between 0.3 to 0.9 s [5], the probability of detection is increased by inspecting over a longer time window. This also helps the tracker to be robust against the irregular nature of the ICIs as the goose-beaked whale is diving [5]. Moreover, using a step length of 7 s instead of a step length of 5.12 ms, which is the step length of performing one GCC-WIN, reduces the processing time by a factor of approximately 1,300. Since the average speed of the goose-beaked whale is 1.2 ms^{-1} , [5], it would have moved approximately 8.4 m within 7 s. Assuming that the goose-beaked whales are primarily hundreds of meters away from the arrays, the corresponding TDOA measurements are unlikely to change significantly over this period.

Parameter Estimation

Details for obtaining the tracks of the goose-beaked whales are provided. Since the TDOA is computed between a pair of hydrophones, there are $\binom{4}{2}=6$ TDOA sensors per array and $N_s=12$ sensors total. As described earlier, the TDOA measurements are accumulated over $T_m=7$ s. The Cartesian coordinate system follows the east, north and up (ENU) convention, where the x-, y-, and z-axes are positive along the east, north, and up directions. The origin is between the two arrays and at the sea surface. The hyperparameters used for TDOA tracking and 3-D tracking are summarized in Table 2.1. For TDOA tracking, the distribution of newly appearing whales is chosen to be uniformly distributed between the minimum and maximum possible TDOA of a hydrophone pair. For 3-D tracking, the distribution of newly appearing whales is chosen to be uniformly distributed on the 3-D region of interest.

A final pruning step is used to remove extra and unreasonable tracks. The average swim speed (Euclidean norm of the estimated velocity) of a goose-beaked whale is 1.2 m s^{-1} [22], and its horizontal speed could range from 1 to 3 m s^{-1} [5]. Consequently, if the median speed of the whale track is more significant than 2 m s^{-1} or the track length is shorter than five time steps, the track is discarded. Furthermore, at every time step, if its estimated speed is faster than 3.5 m s^{-1} , the state at that time step is ignored.

2.4.2 Results

The tracks in the TDOA and 3-D domains using the datasets from two different dates are presented. The results using the proposed method are compared to the tracking results from hand-annotated data using the framework proposed in [46]. In [46], the detected TDOAs are used to compute the azimuth and elevation angles of the goose-beaked whale relative to each hydrophone. They are tracked manually in the azimuth and elevation angles domain, whose tracks are fused between two hydrophones to estimate

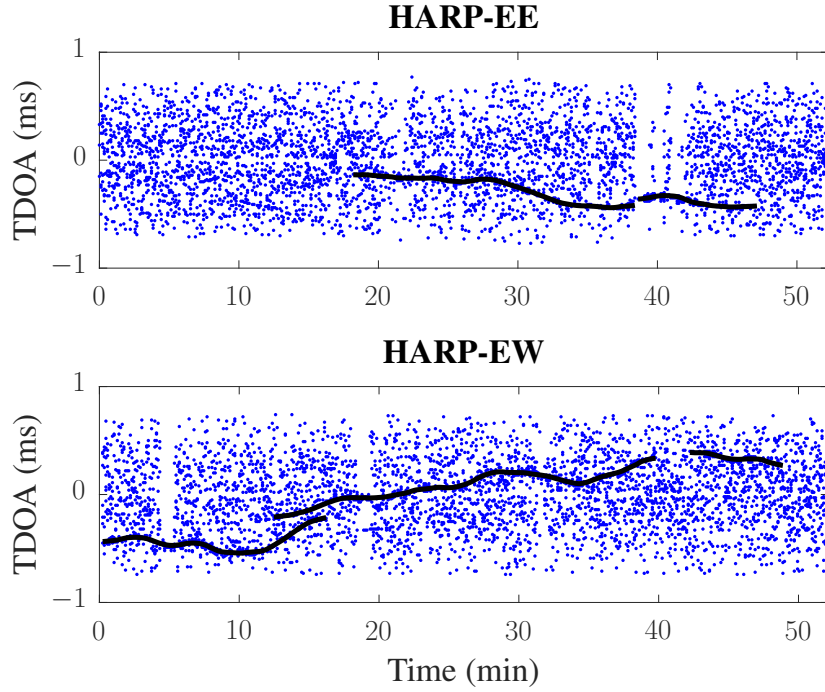


Figure 2.3. Example TDOA measurements (dots) and corresponding TDOA tracking results (solid lines) from one TDOA sensor of HARP EE (top) and one TDOA sensor of HARP EW (bottom). Two goose-beaked whales are observed at HARP EW, while only one is observed at HARP EE. As can be seen in Fig. 2.4, both goose-beaked whales are closer to HARP EW. The considered acoustic data were collected on June 11th, 2018.

their 3-D locations.

Two goose-beaked whales are identified and tracked from the data recorded on June 11th, 2018. While echolocation clicks from only one goose-beaked whale are detected from the TDOA data of HARP EE, those from two whales are detected from the TDOA data of HARP EW (Fig. 2.3). Based on the 3-D tracks (Fig. 2.4 and Fig. 2.5), the goose-beaked whale that first appear in the data is far from HARP EE, resulting in the lack of its detected echolocation clicks on HARP EE. To verify further, the corresponding TDOA measurements from the first goose-beaked whale 3-D track are computed using the TDOA model, and they are in accordance with the tracked TDOA. The manual tracking method could not generate the track for the first goose-beaked whale because it requires that the DOAs exist on both arrays simultaneously.

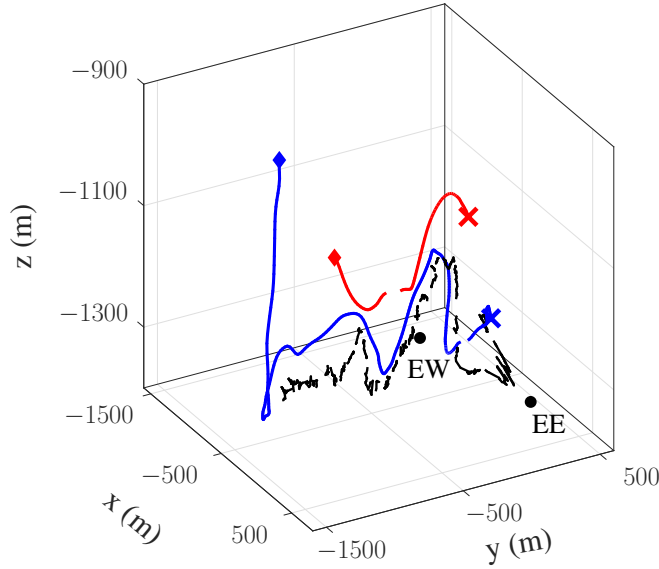


Figure 2.4. A comparison between a track generated from the hand-annotated data (dashed line) and estimated tracks provided by the proposed MTT approach (solid lines) in 3-D. With MTT, an additional track is extracted from the TDOA data of HARP EW, and the diving behavior of a goose-beaked whale can be explored. The diamond and the cross indicate each track’s start and end, respectively. The data from June 11th, 2018, are considered.

In the data collected on July 1st, 2018, two goose-beaked whales are observed again (Fig. 2.6, Fig. 2.7, and Fig. 2.8). Their diving behaviors are detected at the initial depths of approximately 450 m. The two tracks start close to each other, but the tracker is capable of separating them. More echolocation clicks are detected using the GCC-WIN algorithm. As a result, the proposed method yields longer 3-D tracks (approximately by 5 min) than the tracks from the hand-annotated data. To verify the reliability of the results, the 3-D tracker is run again with the TDOA data in the reverse time steps. Since it is easier for the tracking algorithm to identify two goose-beaked whales if they are spatially distinct in the beginning, the tracks from the reverse order would be more reliable in this scenario. The reverse order tracking results are indeed similar to the tracking presented here, verifying the correctness of the proposed tracker.

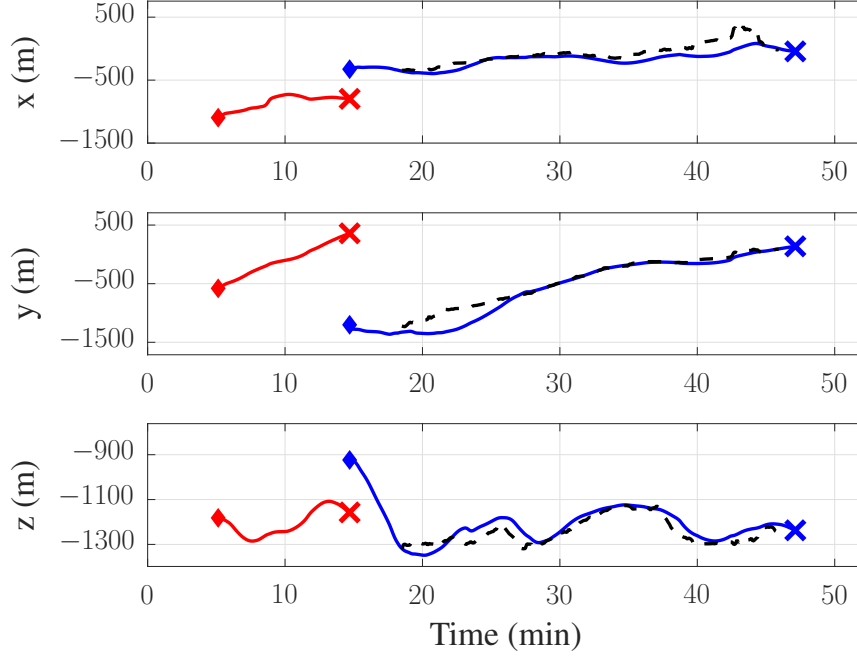


Figure 2.5. A comparison between the track generated from the hand-annotated data (dashed line) and estimated tracks provided by the proposed MTT approach (solid lines). Each axis of the 3-D domain is shown individually. The data from June 11th, 2018, are considered.

2.5 Simulation Studies

Intrinsic challenges of tracking marine animals with PAM are the lack of ground truth tracks and that the manual tracks are imperfect since the data annotation process is subjective. Therefore, the 3-D tracking performance of different benchmark approaches from simulations are compared to motivate the proposed 3-D tracking approach.

Four sets of 200 Monte Carlo simulations of odontocete tracks are generated and tracked using three different approaches. Each set of Monte Carlo simulation has an increasing number of simultaneously present odontocetes, ranging from 1 to 4. The odontocetes' starting positions are placed uniformly on a circle of radius 1000 m on a xy-plane and at a depth of 1000 m. TDOA measurements are generated based on the same array geometry described in Ch. 1.5 and from the model described in Eq. 2.17.

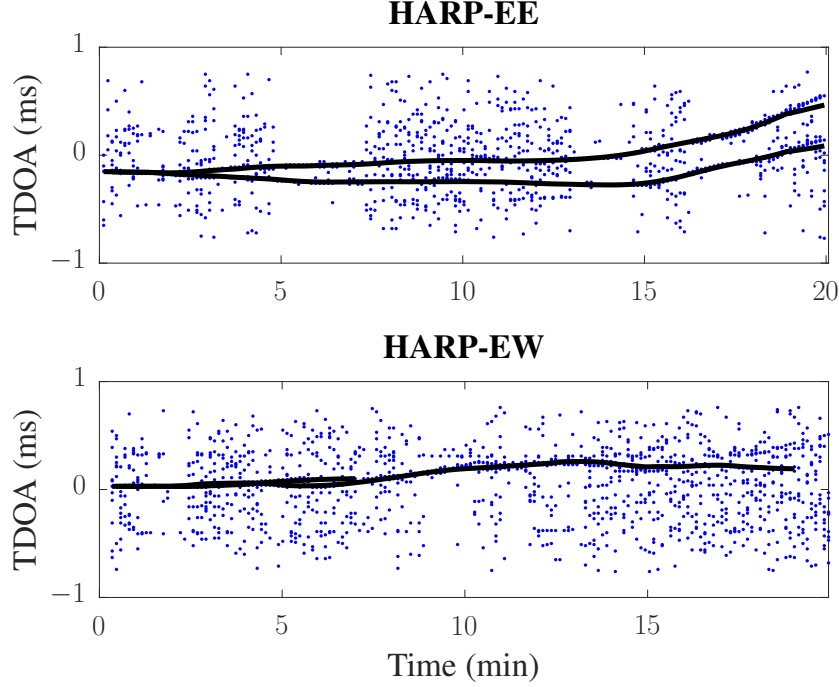


Figure 2.6. Example TDOA measurements (dots) and corresponding TDOA tracking results (solid lines) from one TDOA sensor of HARP EE (top) and one TDOA sensor of HARP EW (bottom). One of the goose-beaked whales is not successfully tracked towards the end by HARP EW because of long gaps of missing TDOA. The considered acoustic data were collected on July 1st, 2018.

False positives and missed detections are also generated based on the model discussed in Ch. 2.3.2. The same hyperparameters from Table 2.1 are used to generate the simulated data. Each simulation consists of 85 discrete time steps long, and the time step has a duration of 7 s. An odontocete is present for 50 time steps. In simulations with multiple odontocetes, an odontocete is introduced every ten steps.

Three tracking approaches are simulated: (i) an approach based on nonsequential tracking [46] (NST), (ii) a single Bernoulli tracker [81] (SBT), and (iii) the proposed MTT approach described in Ch. 2.3.2. In what follows, the approaches (i), (ii), and (iii) will be referred to as (NST), (SBT), and (MTT), respectively. (NST) computes odontocete positions by combining the DOAs of the echolocation clicks of the odontocete computed at each array but does not filter the results, i.e., odontocete positions are computed at each

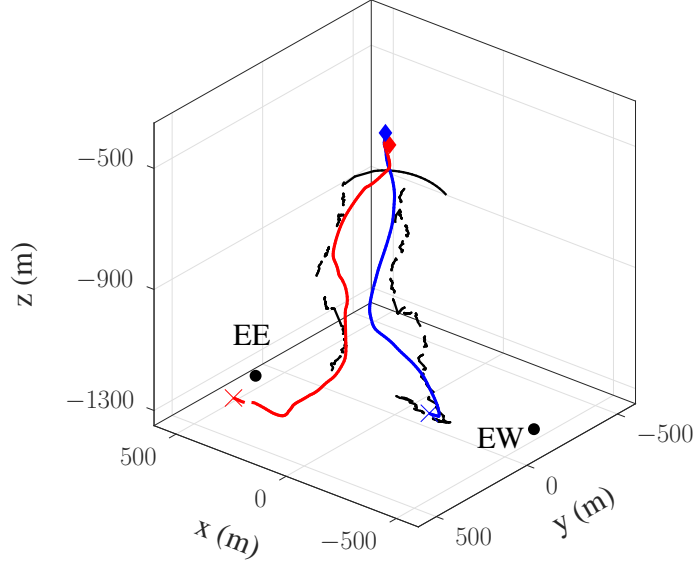


Figure 2.7. A comparison between the tracks generated from the hand-annotated data (dashed lines) and estimated tracks provided by the proposed MTT approach (solid lines) in 3-D. Two closely-space goose-beaked whales are simultaneously diving into deeper waters. The diamond and the cross indicate each track's start and end, respectively. The proposed MTT approach can successfully distinguish the two individuals. The data from July 1st, 2018, are considered.

discrete time step individually. Moreover, both the localization and DOA computations are performed within the least squares framework. (SBT) is a recursive Bayesian filter that can detect and track a single target in the presence of missed detections and false positives. Both (NST) and (SBT) cannot track multiple targets in an automated way, and thus, (NST) and (SBT) are provided with the correct data association solution and are applied to each odontocete track individually. (NST) and (SBT) only use the true measurements corresponding to a single odontocete subject to missed detections. If there is a missed detection, (NST) performs interpolation to obtain a measurement for each odontocete at each TDOA sensor. (SBT) relies on the same statistical model and parameters as (MTT). On the other hand, (MTT) does not know the correct association and thus performs data association automatically. It is provided with the TDOA measurements generated for all

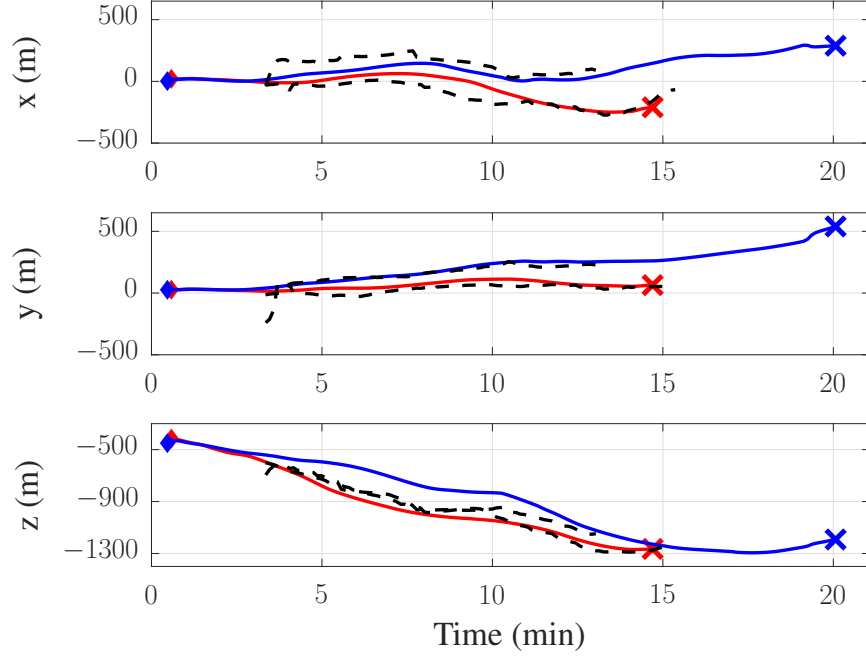


Figure 2.8. A comparison between the track generated from the hand-annotated data (dashed line) and estimated tracks provided by the proposed MTT approach (solid lines). Each axis of the 3-D domain is shown individually. The data from July 1st, 2018, are considered.

present odontocetes following the model for missed detections and false positives discussed in Ch. 2.3.2. It is worth noting that the correct data association solution is only available if TDOA measurements are synthetically generated. In a real-world scenario, the correct data association solution is unavailable and can only be approximated by a highly-trained human operator.

Each track’s root-mean-square error (RMSE) is computed for each approach mentioned above (Fig. 2.9). (MTT) and (SBT) sometimes do not detect the odontocete or have a significant error due to false positives or missed detections. In that case, the RMSE is set to an error value of 110 m, which is approximately twice the average RMSE related to (NST). Based on the simulation results, (NST) has the highest RMSE since it does not perform any filtering. On the other hand, (SBT) yields the lowest RMSE because it relies on the correct data association solution and makes use of a statistical model for

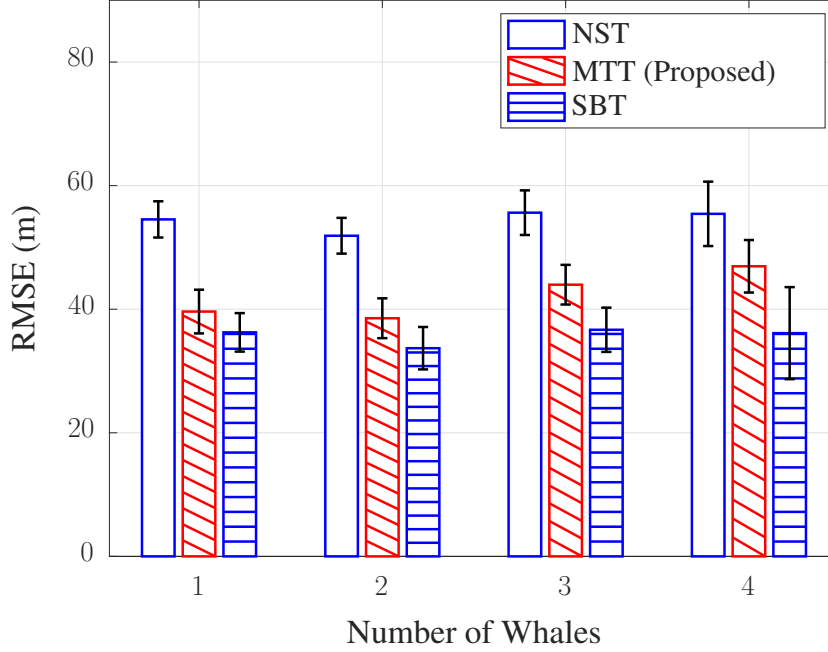


Figure 2.9. The average tracking RMSE across 200 Monte Carlo runs versus number of simultaneously present odontocetes for three different 3-D tracking approaches. The error bars denote the 75th percentile of the measured RMSE. While the correct data association solution is available to (NST) and (SBT), it is unknown to the proposed (MTT) method. In real data processing, the correct association solution is unavailable and can only be approximated by a human operator.

measurement noise, missed detection, and false positives. (MTT) yields an RMSE between those from (NST) and (SBT). (MTT) makes use of the same statistical model as (SBT) but also has to solve the data association problem; hence, it is expected to perform worse than (SBT). However, since (MTT) performs filtering, it still performs better than (NST). Furthermore, the RMSE of (MTT) increases with the number of odontocetes because the presence of multiple odontocetes makes the data association problem more challenging. Recall that (NST) and (SBT) require manual annotations in real-world problems where the correct data association solution is unavailable, while (MTT) performs data association in an automated manner. Thus, (MTT) is well-suited for tracking multiple odontocetes automatically.

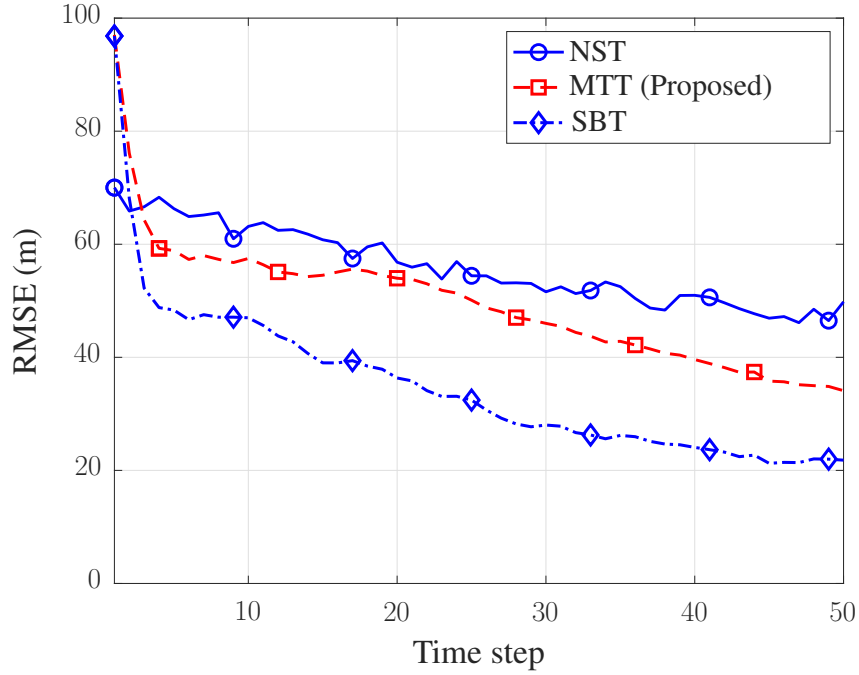


Figure 2.10. The average RMSE versus time for three different 3-D tracking approaches. 200 Monte Carlo runs of a scenario with four odontocetes are considered. (SBT) and (MTT) perform sequential processing and thus lead to reduced RMSE values over time.

Sometimes a different number of odontocete tracks are generated by (MTT). Either a single track is broken into two tracks, or extra tracks are generated from false positives. The percentage of the number of simulations in which extra numbers of tracks are generated are 0%, 1%, 11% and 19% for scenarios with one odontocete, two odontocetes, three odontocetes and four odontocetes, respectively. Additional tracks can indeed be formed with the real data, but in the final stage, either a human operator or an algorithm would join broken tracks or prune unlikely tracks.

In Fig. 2.10, the average RMSE over 800 tracks is shown as a function of the time step for the scenario with four odontocetes. The RMSE of the initialization phase of (MTT) and (SBT) are high, but after a few steps, they converge to a smaller error value compared to (NST). The real data tracks are also longer than 50 time steps; therefore, the total error is expected to be even smaller for longer tracks. The simulation outcome shows

that the proposed approach can generate trustworthy odontocete tracks.

2.6 Discussion

A data processing chain for automatically detecting and tracking multiple odontocetes from their echolocation clicks is developed. It successfully tracks multiple goose-beaked whales in 3-D from their echolocation click recordings made on a pair of volumetric hydrophone arrays. No human operators or heuristics are needed to initialize the tracks and combine the TDOA measurements corresponding to the individual goose-beaked whales. The graph-based MTT method efficiently solves the data association problem to fuse the TDOA measurements among missed detections and false positives and tracks odontocetes in a computationally tractable manner. In addition, new and more extended tracks of the goose-beaked whales could be extracted using the GCC-WIN algorithm, which normalizes the CPSD by the estimated instrument noise PSD to whiten the instrument noise.

A few factors need to be considered when adapting the processing chain for applications with another set of bioacoustic data of odontocetes. Since the tracks are distinguished based on spatial information, the data processing chain is species-agnostic. The pre-processing step needs careful customization, e.g., T_g , frequency filters, etc., so that the echolocation clicks from species of interest are processed. A manual or automated classification step would be desirable for scientific purposes. In addition, the GCC-WIN technique is only recommended if the instrument noise is dominant and can be estimated; otherwise, other GCC adaptations, such as SCOT and PHAT, are potentially better suited for the estimation of TDOAs.

The data processing chain could be further extended to yield more accurate and longer tracks of odontocetes. As seen in the second scenario (Fig. 2.7), closely spaced tracks can pose a challenge. To mitigate, the future model in the MTT framework could incorporate statistics of ICIs to distinguish among multiple echolocating odontocetes. It is

observed that, in the presence of closely spaced tracks, signals from multiple whales are occasionally merged into one by the clustering method described in Ch. 2.3.1. This results in the loss of one of the tracks, hindering tracking performance.

Another challenge is that the model based on a constant probability of detection is not entirely reflective of reality. Even though the probability of detection is affected by the SNR of the bioacoustic signal, their irregularity, i.e., occurrences of a burst of echolocation clicks followed by long gaps of silence, needs to be considered. For example, there is approximately a 2 min long gap of TDOA detections starting at 40 min in the data related to HARP EW in Fig. 2.3. This gap is likely due to the goose-beaked whale looking away from the array. With prior information on the echolocation click directionality and the motion of the odontocetes at each given time, the detector’s performance could be modeled more accurately. Alternatively, a time-varying detection probability could be estimated together with the target states [82].

Furthermore, although the used nearly constant velocity motion model of the odontocetes yielded good tracking results in this work, a tracking algorithm that selects the most suitable motion model for the odontocete motion during runtime is desirable. By incorporating interacting multiple models [83], the tracker would not only capture time-varying behaviors or motions of the odontocetes more accurately but also be more suitable for tracking different species of odontocetes.

Finally, some species of odontocetes are highly sociable and thus sometimes move together in close spatial proximity [84, 85]. In such a case, it would be challenging to track an individual odontocete, but subgroups could be tracked [20]. The question of the feasibility of the MTT frameworks to follow pods of odontocetes instead of individuals poses an interesting research problem.

2.7 Conclusion

Tracking acoustically active marine animals using passive acoustics can provide better understanding of their behaviors underwater that are difficult to observe otherwise. However, human operators are often required to annotate the bioacoustic data to associate the acoustic recordings with the marine animals that generated the signals. With increasing amount of available underwater bioacoustic recordings and an ultimate goal of real-time monitoring, tracking processes that involve human operators, who are imperfect, are not sustainable. Hence, it is crucial to research automated methods that are computationally tractable and accurate for passive acoustic tracking of marine animals. This will not only generate tracks that are more objective but also facilitate scientists to study the marine animals more efficiently.

In this chapter, a data processing chain for the fully automated detection and tracking of odontocetes in 3-D from echolocation clicks is developed. The detection rate of the echolocation clicks is improved by utilizing a generalized cross-correlation algorithm designed to whiten the instrument noise. Multiple odontocetes are detected and tracked simultaneously by applying two stages of a graph-based tracking method that efficiently solves the data association problem. Graph-based detection and tracking are first performed for each hydrophone pair individually in the time-difference-of-arrival domain and subsequently in the 3-D domain. The ability to track multiple odontocetes simultaneously without manual data selection by a human operator is demonstrated based on real acoustic data provided by two volumetric hydrophone arrays. In particular, tracking results in scenarios with two echolocating goose-beaked whales (*Ziphius cavirostris*) are presented. In addition, the simulation results suggest that the presented processing chain can be used to track a larger number of whales in a scalable manner and is thus particularly appealing for future passive acoustic monitoring systems. These results show how the proposed data processing chain can simplify scientists' steps to study

the deep-diving echolocating odontocetes.

Acknowledgments:

Chapters 1 and 2 are adapted and extended in part with permissions from: J. Jang, F. Meyer, E. Snyder, S. Wiggins, S. Baumann-Pickering, and J. Hildebrand, “Bayesian detection and tracking of odontocetes in 3-D from their echolocation clicks,” *J. Acoust. Soc. Am.*, 2023 (Copyright 2023, Acoustical Society of America), and J. Jang, F. Meyer, E. R. Snyder, S. M. Wiggins, S. Baumann-Pickering and J. A. Hildebrand, ”Passive Acoustic Tracking of Whales in 3-D,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, June 2023 (Copyright 2023, IEEE). Some notation and structure have been revised for consistency and clarity within this dissertation. The dissertation author was the primary investigator and lead author of these works.

Chapter 3

Ranging of a Moving Ship Using a Single Acoustic Receiver in Shallow Water

3.1 Introduction

Range information of a moving object is critical for maritime situational awareness. Passive acoustic ranging is an important task in maritime surveillance [86, 13] or in unmanned underwater vehicle (UUV) operations [87]. In addition, passive acoustic ranging can be used for self-localization by small autonomous platforms with limited navigation capabilities but equipped with a single hydrophone or a small hydrophone array. Such platforms include drifters [88], gliders [89, 90] and propeller-driven UUVs [91]. While these systems rely on acoustic signals for localization, estimating the range of a non-cooperative, broadband acoustic source, such as a transiting commercial ship, using a single hydrophone in shallow water remains a significant challenge due to multipath propagation and modal interference.

Among single-hydrophone-based ranging methods, various approaches face distinct limitations. Matched field processing [92, 93] relies either on detailed *a priori* environmental knowledge on seabed properties and sound-speed profiles or the availability of signal replicas. Modal filtering-based methods [94, 95] are restricted to impulsive or frequency-modulated

signals, limiting their applicability to broadband ship noise. Cepstrum-based methods [96, 97] estimate time differences of arrival by extracting propagation path pairs. However, they are typically effective only for short-range estimations, on the order of a few kilometers. Deep learning approaches [98] offer data-driven solutions but struggle with generalization and require extensive training data. In contrast, waveguide-invariant (WI)-based methods estimate range by processing interference patterns (striations) in spectrograms, enabling long-range passive tracking without detailed environmental knowledge. The WI, therefore, provides a powerful framework for passive acoustic ranging using a single receiver [99, 100, 101, 102, 103, 104, 105, 106].

This chapter extends WI-based methods by introducing a statistical framework that jointly models broadband and tonal ship noise, with the potential of enhancing robustness against intensity variations and environmental uncertainty.

3.1.1 Waveguide Invariant

In shallow water waveguides, sound propagates as discrete modes [107]. When an acoustic source moves relative to a receiver, interference between these modes produces striation patterns in a range-frequency representation of the received acoustic intensity. In an idealized setting, e.g., uniform source spectrum and negligible background noise, these striations manifest as interference-induced curves along which the acoustic intensity may follow local maxima, minima, or intermediate levels. These patterns, shaped by the waveguide environment, carry information about the acoustic source’s range, which can be extracted by making use of the WI provided that mode interference is sufficiently strong and stable under appropriate conditions [108, 109, 110, 111, 112].

The WI, often denoted as β , is a parameter that describes how frequency-dependent modal interference relates to the source-receiver range. It arises from the functional relationship between the phase and group speeds of a group of modes [113]. It is typically defined for frequency bands sufficiently above mode cutoff frequencies, since dispersion

properties change rapidly near the cutoff. While β is typically treated as a single, mode-independent (scalar) value, in certain cases, e.g., when signals with low frequencies are used and few modes propagate, or when groups of modes with distinct dispersion characteristics are present, it may be better represented as a distribution derived from pairwise modal interactions [114, 115]. For a comprehensive discussion on WI theory and practical considerations, see [107] and [113].

Since the striations are functions of range and frequency, their shape can be used to infer the source's range relative to the receiver. The WI establishes a mathematical relationship between frequency and range, governing the shape of the striations. This relationship is given by

$$\frac{f}{f'} = \left(\frac{r}{r'} \right)^\beta, \quad (3.1)$$

where (r', f') and (r, f) are different points along a striation, and β is the WI. Since striations follow this relationship, a known WI allows for the projection of candidate striations for different candidate source ranges directly onto a spectrogram. The comparison of projected striations with observed striations in the spectrogram provides a direct method for inferring the source range. While a known WI allows projection of candidate striations for different source ranges onto a spectrogram, this requires mapping time to range using an assumed range rate; the details of this mapping are provided in Ch. 3.2.1.

An example dataset for range estimation is shown in Fig. 3.1. It is a spectrogram of real acoustic data recorded during Seabed Characterization Experiment 2017 (SBCEX17) [1], where a large commercial ship was moving away from the acoustic receiver in a 75 m deep water column (This data will be discussed in more detail in Ch. 3.5). The tonal (horizontal lines of high intensities) and broadband components of ship-radiated noise are visible, along with striations formed by modal interference patterns. The red lines represent projected striations generated using Eq. (3.1) with the correct source ranges reported by the Automated Identification System (AIS) and the established WI value of $\beta = 1.18$.

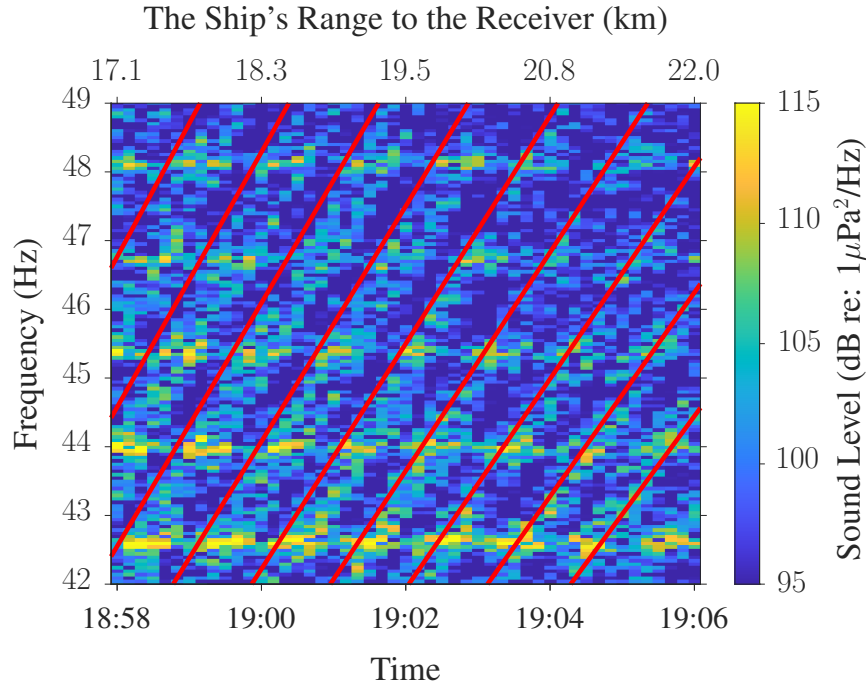


Figure 3.1. Spectrogram of data recorded on March 24, 2017 (UTC) during SBCEX17 using a single element of a vertical line array. The large commercial ship MSC KALAMATA was moving away from the receiver. Slightly diffuse tonal lines and broadband radiated noise from the ship are visible, along with modal interference patterns. The red lines show projected striations based on Eq. (3.1) using a waveguide invariant $\beta = 1.18$. Striations refer to interference-induced curves along which the acoustic intensity may follow local minima, maxima, or intermediate levels.

(Further discussion on the established value of β is provided in Ch. 3.6.1.) The red lines indicate striations, i.e., interference-induced curves in the spectrogram along which acoustic intensity remains approximately constant across frequency and time (range), with local maxima and minima most visually apparent. If incorrect ranges were used for projecting the line, the lines would no longer follow the intensity pattern in the spectrogram.

This chapter introduces a statistical approach that quantifies how well the observed intensities along projected striations match modeled intensities provided by our statistical models of the received ship-radiated noise. Since the curvature of projected striations depends on the assumed source range, our estimation method selects the candidate range

whose projected striations yield the best statistical agreement with the observed data.

3.1.2 Ship Sound Characteristics

The underwater noise radiated by transiting ships contributes significantly to the ocean ambient sound at low frequencies [116, 117]. It has been shown that, in the vicinity of the ship, high-intensity noise from 10 Hz to 1 kHz is emitted by the normal operation of the ship [118]. Considering the relatively low attenuation at lower frequencies [119] and acoustic propagation in shallow water [120], low-frequency ship noise can be detected at ranges of tens of kilometers [121].

The sound emitted from modern cargo ships can be considered as the superposition of the tonal and broadband acoustic signals [119] (see Fig. 3.1). The narrowband tonals correspond to the line spectrum that result from the periodicity of the machinery, such as the engine and the shaft [122]. The broadband source signal stems from the propeller cavitation noise and flow noise around the hull, which are complex hydrodynamic processes and challenging to model accurately [123, 124]. Depending on the physical properties of the ships, the mode of operations, the oceanographic conditions, and the aspect ratio, the sound level characteristics can vary [125]. This chapter models statistically and processes the received tonal and broadband radiated signature from a moving ship to estimate its range.

3.1.3 Contribution and Chapter Organization

This chapter introduces a new statistical model and method for estimating the range of a moving ship in shallow water using its radiated broadband and tonal acoustic signature, recorded by a single stationary hydrophone. By leveraging the WI, ranging is performed with minimal environmental knowledge through the analysis of the spectrogram of the received signal. A key assumption, as in prior work, is that the environment is range-independent, ensuring that the WI remains constant over range. Additionally, the

WI of the considered environment and the ship’s range rate, which is needed to convert the spectrogram’s time-axis into range for each candidate range, are assumed to be known.

The proposed method addresses key limitations of existing WI-based ranging techniques. Deterministic methods [100, 102, 126] that rely on directly extracting the slopes of the striations are sensitive to noise and require striations that exhibit relatively uniform intensity. The proposed statistical framework accounts for measurement uncertainties and variability in intensity along striations, improving robustness. Additionally, existing statistical methods focus on utilizing either the broadband [106] or the tonal [103] components of the ship noise. By incorporating both, the method increases the amount of usable data, reducing estimation uncertainty and improving range accuracy. The major contributions of the presented research are as follows:

- **Statistical Models:** We develop models for broadband and tonal ship sound in a range-independent shallow water environment.
- **Estimation Method:** We introduce a statistical signal processing approach that utilizes the WI for range estimation from spectrogram data.
- **Experimental Validation:** We demonstrate the estimation of ranges up to 45 km with less than $\pm 3\%$ error using real data from SBCEX17.

The chapter is organized as follows: The general strategy for range estimation using the WI is introduced in Sec.3.2. The proposed statistical signal model for broadband and tonal noise components is established in Ch. 3.3. Ch. 3.4 describes the proposed method and algorithm for range estimation using broadband and tonal noise components. Data and implementation details are outlined in Ch. 3.5 and Ch. 3.6, respectively. Finally, results are discussed in Ch. 3.7, and Ch. 3.8 concludes the chapter.

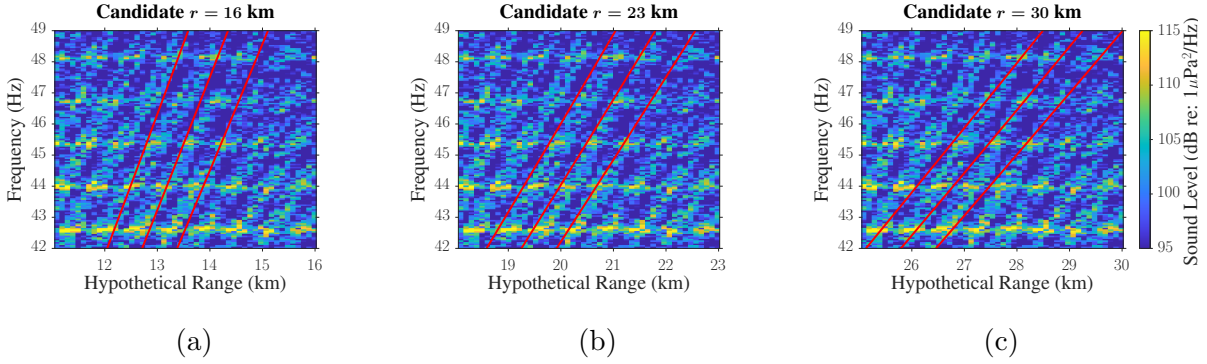


Figure 3.2. Examples of the r - f surface plots with projected striations (red lines) overlaid using a single reference frequency $f' = 45.5$ Hz and Eq. (3.1), for three candidate source ranges, which correspond to the final snapshot of the spectrogram. The candidate values are (a) $r = 16$ km, (b) $r = 23$ km, and (c) $r = 30$ km. All plots are based on the same spectrogram shown in Fig. 3.1. For each case, only three striations are shown for clarity; in practice, a denser set of projected striations that span the full frequency band are included (see Secs. 3.3.5 and 3.6.1 for selection criteria). The constant range rate $\dot{r} = 10.2$ m/s and the WI $\beta = 1.18$ are used. The true range is 23 km, and the projected striations are most well aligned for (b), where the candidate range matches the true value.

3.2 Waveguide Invariant-based Ranging

Our general strategy for estimating the source range is to systematically evaluate how well the expected striation patterns based on a set of candidate source ranges match those observed in a spectrogram. Given a known range rate and a candidate source range, the time axis of the spectrogram is converted to range axis, producing a range-frequency (r - f) surface plot. The source range to be estimated corresponds to the final snapshot of the spectrogram. Here, a snapshot refers to a Short-Time Fourier Transform frame, representing the spectral content of a windowed segment of the time-series signal.

For each candidate range, along with the conversion of the original spectrogram into a r - f surface, projected striations are generated using Eq. (3.1), allowing for direct comparison with the observed striation patterns. This is done by fixing a reference frequency f' and iterating over reference range values r' along the range axis. Each pair (r', f') defines a striation curve that is constructed by computing the corresponding

(r, f) values that satisfy Eq. (3.1) across frequencies in the band of interest. Only those projected striations that span the full frequency band are considered ‘valid’ and retained for processing. In general, multiple reference frequencies f' would be required to achieve full coverage of the considered r - f surface plot, but a single value is used in this work for simplicity. (See Fig. 3.2 for examples and Secs. 3.3.5 and 3.6.1 for practical details). The projected striations are compared to the observed patterns in the spectrogram, and the candidate range that yields the best overall alignment is selected as the estimated source range.

In this section, we establish how the r - f surface plot is obtained, define the parameter space for range estimation, address the parameter ambiguities, and provide a summary of the proposed method.

3.2.1 Spectrogram to Range-Frequency Surface Plot

We define the range at the last snapshot of the spectrogram as the range to be estimated. Given a spectrogram with N snapshots and denoting the range at the last snapshot as r_N , the range r_n at snapshot $n \leq N$ is

$$r_n(r_N; \dot{r}, t_\Delta) = r_N - (N - n)\dot{r}t_\Delta, \quad (3.2)$$

where t_Δ is the time step between snapshots. For simplicity, we assume a constant range rate $\dot{r}$, though the method can be extended to scenarios where $\dot{r}$ varies with time, provided that the time-dependent $\dot{r}$ is known.

3.2.2 Parameter Space for WI-Based Ranging

To accurately estimate the source range r , well-defined search intervals and prior knowledge of key governing parameters are essential. In particular, the shape of the striation pattern in the range-frequency domain is determined by three fundamental

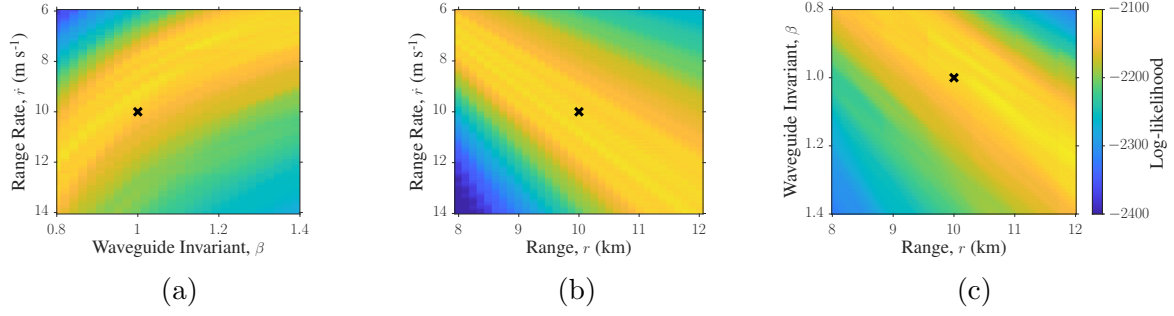


Figure 3.3. The joint log-likelihood function $\ell(r, \dot{r}, \beta)$ is evaluated over the parameter spaces: range $r \in \mathcal{S}_r$, range rate $\dot{r} \in \mathcal{S}_{\dot{r}}$, and WI $\beta \in \mathcal{S}_\beta$. It is shown as functions of (a) β and $\dot{r}$, (b) r and $\dot{r}$, and (c) r and β . In each plot, the third variable is fixed at its true value. The likelihood is computed using synthetic intensity spectrogram data from a Pekeris waveguide ($\beta \approx 1$) with a source moving at $\dot{r} = 10$ m/s, with a range of 10 km at the final snapshot. The cross marks the true parameter values. The resulting 2-D likelihood functions are non-unimodal, implying that accurate estimation of one parameter requires narrow search intervals for the other two, e.g., to accurately estimate the range r , the search intervals $\mathcal{S}_{\dot{r}}$ and $\mathcal{S}_\beta$ must be small.

parameters: the WI β , the range rate $\dot{r}$, and the source range ($r = r_N$). These parameters arise from the characteristics of the acoustic environment and the motion of the source. The search intervals for the key parameters with respect to r , $\dot{r}$ and β are: $\mathcal{S}_r = [r_{\min}, r_{\max}]$, $\mathcal{S}_{\dot{r}} = [\dot{r}_{\min}, \dot{r}_{\max}]$, and $\mathcal{S}_\beta = [\beta_{\min}, \beta_{\max}]$, respectively. For future reference, we also introduce the joint search interval denoted as $\mathcal{S} = \mathcal{S}_r \times \mathcal{S}_{\dot{r}} \times \mathcal{S}_\beta$ and the joint parameter vector $\mathbf{q} = [r, \dot{r}, \beta]^T \in \mathcal{S}$

To estimate any one of the parameters in $\mathbf{q}$, the other two must be known due to the coupling of these parameters in Eq. (3.1) and Eq. (3.2). Since striation patterns are smooth functions of range and frequency, multiple parameter combinations can produce similar-looking striation patterns, leading to ambiguity in estimation. This makes it essential to have prior knowledge of at least two parameters in $\mathbf{q}$ when estimating the third. In particular, to perform range estimation, knowledge of $\dot{r}$ and β is required. Alternatively, if r and $\dot{r}$ are known, the WI β can be estimated by evaluating the striation intensities over a set of candidate β values.

The importance of constraining these parameters is demonstrated in Fig. 3.3, which

highlights how ambiguity arises when two of the three parameters are unknown. The likelihood function $\ell(r, \dot{r}, \beta)$, defined in Ch. 3.4, is evaluated across three 2D parameter subspaces, while the third variable is held constant at its true value. The likelihood function can be interpreted as a statistical characterization of the match between the observed and expected striations for different parameter values. Higher likelihood values indicate stronger agreement between the projected and observed striation patterns.

Non-unimodal behavior of the likelihood functions is evident across all three cases, meaning that multiple combinations of parameters yield similar likelihood values. This ambiguity confirms that accurate estimation of one parameter requires the other two to be well-constrained. The next section discusses how β and $\dot{r}$ can be estimated to resolve ambiguity in range estimation.

3.2.3 Estimating the WI and the Range Rate

Approaches for obtaining knowledge on the WI β and the range rate $\dot{r}$ to perform range estimation are reviewed in what follows.

WI: In shallow water, the WI β can often assumed to be $\beta = 1$ when the dominating group of modes are surface- and bottom-reflected [113]. Alternatively, the value of β can be obtained in a calibration step by using ships of opportunity that share their position and velocity vector via the AIS [127]. Based on the shared position and velocity vector, the range $r = r_N$ and range rate $\dot{r}$ can be extracted and used for the estimation of β .

Range Rate: For the present study, we adopted a constant range rate based on AIS-reported ship speed. However, a potential approach for estimating source range rate is to process acoustic recordings near the closest point of approach (CPA) [128, 101, 104], if stable tonal components can be identified. Since the Doppler shift is zero at the CPA, the nominal tone frequencies can be extracted and used to estimate range rate based on the observed Doppler shifts before and after the CPA [129]. Our dataset includes CPA recordings, and preliminary inspection revealed a few tonal components between

300–800 Hz that are persistent and exhibited frequency shifts consistent with expected Doppler shifts. Although Doppler shift-based range rate estimation by exploiting CPA was not considered in the present work, this preliminary inspection suggests that it is a viable direction for future analysis.

3.2.4 Method Overview

The general range estimation framework is summarized in this discussion. The proposed method estimates the range of a moving ship in shallow water using a statistical framework that leverages the WI to process the striations in the spectrograms. The input data is the spectrogram of the recorded acoustic signal, and the method is based on the following key assumptions:

- The WI β and the range rate $\dot{r}$ are either estimated or approximated and therefore treated as known.
- The environment is range-independent, ensuring that β remains constant over the observed range.

Errors in β or $\dot{r}$ can introduce bias in the range estimate, as different parameter combinations may yield similar striation patterns, leading to ambiguity (Fig. 3.3).

Computing a range estimate for fixed β and $\dot{r}$ is discussed next. For each range in a set of candidate source ranges, $r \in \mathcal{S}_r$, the following steps are performed.

1. **Spectrogram Processing:** Using r and $\dot{r}$ with Eq. (3.2), the time axis of the spectrogram is converted into a range axis, producing a hypothetical r - f surface plot.
2. **Striation Projection:** For that hypothetical r - f surface plot, the expected intensities along WI-projected striations using β and Eq. (3.1) are interpolated on a regular range-frequency grid. The choice of the reference range and frequency for striation projection is discussed in Ch. 3.6.1.

3. Likelihood Function Computation: The joint likelihood function, parameterized by $\mathbf{q}$, is evaluated. The likelihood function quantifies how well the observed intensity measurements in the striation-frequency domain match the projected striations based on $\mathbf{q}$ within the proposed statistical model.

The most probable range is found by maximizing the joint likelihood function over all candidate ranges.

Fig. 3.2 shows examples of projected striation patterns overlaid on the same spectrogram for three different candidate source ranges. These r - f surface plots illustrate how the alignment between projected striations and the observed intensity patterns varies with the assumed source range. Among the three candidates, the projection at $r = 23$ km exhibits the best alignment (Fig. 3.2 (b)), consistent with the true source range. While the general striation patterns are visible in the spectrograms, due to the presence of the high intensity tonal signals and randomness in the broadband component as well as the background noise, the intensities are not uniform along each striation. These components are statistically modeled to enable evaluation of match between the projected and observed striation through the likelihood-framework.

The next section (Ch. 3.3) establishes the statistical model for the measured intensities, laying the foundation for the proposed likelihood function in Ch. 3.4.

3.3 Signal Model

3.3.1 Model Overview and Notation

This section introduces the statistical model of the received acoustic signal while a ship passes by an acoustic receiver in a range-independent shallow water environment. The tonal components of the ship radiated sound are treated as continuous sinusoidal signals with well-defined frequencies, whereas the broadband component is modeled as a random process due to its stochastic nature. Only a scenario where a single ship passes

by the receiver is considered, and it is assumed that ambient noise sources are relatively distant and incoherent.

The model is provided for the range-frequency domain representation of the acoustic signal, i.e., the r - f surface plot. The n^{th} snapshot corresponds to the acoustic signal at range r_n for $n \in \mathcal{N} = \{1, \dots, N\}$, where N is the number of snapshots in the r - f surface plot. A discrete set of K frequency bins, $\mathcal{F} = \{f_1, \dots, f_K\}$ with $f_1 < f_2 < \dots < f_K$, is processed, where the k^{th} frequency bin is $f_k \in \mathcal{F}$ for $k \in \mathcal{K} = \{1, \dots, K\}$.

The radiated acoustic signature of a moving ship is modeled as a superposition of broadband and tonal signals. The frequency bins are thus divided into two subsets: $\mathcal{F}^{\text{b}} \subseteq \mathcal{F}$, containing only broadband components, and $\mathcal{F}^{\text{bt}} \subset \mathcal{F}$, containing both broadband and tonal components. The broadband signal spans the entire band $\mathcal{F}$, while $\mathcal{F}^{\text{bt}}$ consists of J frequencies of narrowband tonal signals. For future reference, we define the corresponding index sets: $\mathcal{K}^{\text{b}} = \{k \mid f_k \in \mathcal{F}^{\text{b}}\}$ and $\mathcal{K}^{\text{bt}} = \{k \mid f_k \in \mathcal{F}^{\text{bt}}\}$ with $\mathcal{K} = \mathcal{K}^{\text{b}} \cup \mathcal{K}^{\text{bt}}$.

In this chapter, we will use the following notation. Random variables are displayed in sans serif, upright fonts, while their realizations are in serif, italic fonts. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. For example, a random variable and its realization are denoted by $\mathbf{x}$ and x ; a random vector and its realization by $\mathbf{x}$ and $\mathbf{x}$; and a random matrix and its realization are denoted by $\mathbf{X}$ and $\mathbf{X}$. The probability density function (PDF) and the expectation of random vector $\mathbf{x}$ are denoted as $\mathbb{E}[\mathbf{x}]$ and $f(\mathbf{x})$, respectively. Furthermore, $\mathcal{N}(u; \mu, \sigma^2)$ denotes a real Gaussian PDF (of random variable u) with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}$, $\mathcal{CN}(x; \mu_c, \sigma_c^2)$ denotes a complex Gaussian PDF (of random variable x) with mean $\mu_c \in \mathbb{C}$ and variance $\sigma_c^2 \in \mathbb{R}$ such that each real and imaginary part has variance $\sigma_c^2/2$, $\chi^2(y; d, \lambda)$ denotes the noncentral chi-squared PDF (of random variable y) with degrees of freedom d and noncentrality parameter λ , $\text{Exp}(z; \theta)$ denotes the exponential PDF (of random variable z) with scale parameter θ .

3.3.2 Received Signal Model

At range r_n and frequency $f_k \in \mathcal{F}$, the broadband and tonal source signals radiating from the moving ship are denoted as $\mathbf{s}_{n,k}^b \in \mathbb{C}$ and $s_{n,k}^t \in \mathbb{C}$, respectively. The broadband source signal $\mathbf{s}_{n,k}^b$ is modeled as $\mathcal{CN}(s_{n,k}^b; 0, (\sigma_k^b)^2)$, with frequency-dependent [122] variance $(\sigma_k^b)^2$. Nonrandom tonal source signals radiating from the moving ship are modeled by the complex amplitude $s_{n,k}^t = m_k^t e^{j\phi_{n,k}}$, where $m_k^t = |s_{n,k}^t|$ is an unknown, range-independent source magnitude that varies with frequency and $\phi_{n,k}$ is an unknown phase term.

The received acoustic signal is modeled as

$$\mathbf{z}_{n,k} = g_{n,k}(\mathbf{s}_{n,k}^b + s_{n,k}^t) + \mathbf{u}_{n,k}. \quad (3.3)$$

For a range-frequency index pair (n, k) , $g_{n,k} \in \mathbb{C}$ is the Green's function, i.e., the channel transfer function, evaluated at (n, k) , and $\mathbf{u}_{n,k} \in \mathbb{C}$ is a random additive background noise following the PDF $\mathcal{CN}(u_{n,k}; 0, (\sigma_k^u)^2)$. The channel model will be further discussed in Ch. 3.3.4. The background noise is assumed to be independent and identically distributed (iid) across n and k , and its variance $(\sigma_k^u)^2$ is deterministic and frequency-dependent.

The model in Eq. (3.3) is simplified and distinguished for two subsets of frequencies $\mathcal{F}^b$ and $\mathcal{F}^{bt}$ as

$$\mathbf{z}_{n,k} = \begin{cases} \mathbf{p}_{n,k}^b + \mathbf{u}_{n,k} & k \in \mathcal{K}^b \\ \mathbf{p}_{n,k}^b + p_{n,k}^t + \mathbf{u}_{n,k} & k \in \mathcal{K}^{bt}, \end{cases} \quad (3.4)$$

where $\mathbf{p}_{n,k}^b = g_{n,k} \mathbf{s}_{n,k}^b$ and $p_{n,k}^t = g_{n,k} s_{n,k}^t$. The received broadband signal, $\mathbf{p}_{n,k}^b$, is random and follows the PDF $\mathcal{CN}(p_{n,k}^b; 0, (\sigma_k^b)^2 |g_{n,k}|^2)$, while the received tonal signal, $p_{n,k}^t = g_{n,k} s_{n,k}^t$, is nonrandom and unknown. Due to range-dependency of $g_{n,k}$, the statistics of $\mathbf{p}_{n,k}^b$ are also range-dependent.

Across all frequency bins, $k \in \mathcal{K}$, the broadband-plus-noise signal $\mathbf{p}_{n,k}^b + \mathbf{u}_{n,k}$ is distributed according to $\mathcal{CN}(\mathbf{z}_{n,k}; 0, \sigma_{n,k}^2)$ with variance $\sigma_{n,k}^2$. This variance $\sigma_{n,k}^2$ will be

referred to as the broadband-plus-noise variance, which can be expressed as

$$\sigma_{n,k}^2 = (\sigma_k^b)^2 |g_{n,k}|^2 + (\sigma_k^u)^2. \quad (3.5)$$

In addition, it is assumed that σ_k^u and σ_k^b vary smoothly with frequency index k . This implies that σ_k^u is approximately equal to $(\sigma_{k+1}^u + \sigma_{k-1}^u)/2$ and σ_k^b is approximately equal to $(\sigma_{k+1}^b + \sigma_{k-1}^b)/2$.

3.3.3 Statistical Model of Acoustic Intensity

To perform range estimation, we are interested in the acoustic intensity defined as

$$\mathbf{x}_{n,k} = |\mathbf{z}_{n,k}|^2. \quad (3.6)$$

Consequently, for frequency bins with only the broadband component ($k \in \mathcal{K}^b$), the intensity $\mathbf{x}_{n,k}$ follows an exponential distribution with scale parameter $\theta_{n,k} = \sigma_{n,k}^2$, i.e., $\mathbf{x}_{n,k}^b \sim \text{Exp}(x_{n,k}^b; \theta_{n,k})$. The intensity $\mathbf{x}_{n,k}^b$ is statistically independent across n and k , and its scale parameter $\theta_{n,k}$ is both range- and frequency-dependent. Based on the functional form of $\text{Exp}(x_{n,k}^b; \theta_{n,k})$, the PDF of $\mathbf{x}_{n,k}^b$ is

$$f_{\text{Exp}}(x_{n,k}^b; \theta_{n,k}) = \frac{1}{\theta_{n,k}} \exp\left(-\frac{x_{n,k}^b}{\theta_{n,k}}\right). \quad (3.7)$$

For future reference, let $\mathbf{X} \in (\mathbb{R}^+)^{N \times K}$ be the measurement matrix, i.e., the r - f surface plot, that consists of elements $x_{n,k}$, where $n \in \mathcal{N}$ and $k \in \mathcal{K}$.

For frequency bins with broadband and tonal component, $k \in \mathcal{K}^{\text{bt}}$, based on the assumptions above, the received acoustic signal follows a circularly symmetric complex Gaussian distribution, $\mathbf{z}_{n,k} = \mathbf{p}_{n,k}^b + p_{n,k}^t + \mathbf{u}_{n,k} \sim \mathcal{CN}(z_{n,k}; p_{n,k}^t, \sigma_{n,k}^2)$, with mean $p_{n,k}^t$ and variance $\sigma_{n,k}^2$ (cf. Eq. (3.5)). Note that both $p_{n,k}^t$ and $\sigma_{n,k}^2$ are range- and frequency-

dependent. Let us introduce the normalized intensity defined as

$$y_{n,k}^{\text{bt}} = \frac{x_{n,k}^{\text{bt}}}{\sigma_{n,k}^2/2}, \quad (3.8)$$

where $x_{n,k}^{\text{bt}}$ is the intensity at frequency index $k \in \mathcal{K}^{\text{bt}}$. The normalized intensity follows a noncentral- χ^2 PDF with two degrees of freedom [130], i.e., $d = 2$, and noncentrality parameter $\lambda_{n,k} = |p_{n,k}^{\text{t}}|^2/(\sigma_{n,k}^2/2)$, i.e., $y_{n,k}^{\text{bt}} \sim \chi^2(y_{n,k}^{\text{bt}}; 2, \lambda_{n,k})$. Note that $|p_{n,k}^{\text{t}}| = m_k^{\text{t}}|g_{n,k}|$. The normalized intensity is assumed statistically independent across n and k . Based on the functional form of $\chi^2(y_{n,k}^{\text{bt}}; 2, \lambda_{n,k})$, the PDF of $y_{n,k}^{\text{bt}}$ can be written as

$$f_{\chi^2}(y_{n,k}^{\text{bt}}; \lambda_{n,k}) = \frac{1}{2} \exp\left(-\frac{y_{n,k}^{\text{bt}} + \lambda_{n,k}}{2}\right) I_0\left(\sqrt{y_{n,k}^{\text{bt}} \lambda_{n,k}}\right), \quad (3.9)$$

where $I_0(\cdot)$ is a modified Bessel function of the first kind and zeroth order.

To perform range estimation with the given signal model, however, the values of the introduced statistical parameters must be known. It is assumed that the statistics of the ambient noise are available from the acoustic measurements when the ship was not present, i.e., the background noise variance $(\sigma_k^{\text{u}})^2$ is known *a priori* at the time of processing. To compute the normalized intensity measurements $y_{n,k}^{\text{bt}}$, we need to obtain the broadband-pluse-noise variance $\sigma_{n,k}^2$, which will be computed from the measurement matrix $\mathbf{X}$. To evaluate the PDFs of $y_{n,k}^{\text{bt}}$, we also need to know the noncentrality parameter $\lambda_{n,k}$, which will also be computed from the measurement matrix $\mathbf{X}$. The computations of $\lambda_{n,k}$ and $\sigma_{n,k}^2$ will be discussed in Ch. 3.4.1.

3.3.4 The Green's Function (Channel Model)

Without loss of generality, let (n, k) and (n', k') , with $n, n' \in \mathcal{N}$ and $k, k' \in \mathcal{K}$, denote the range and frequency index pairs that, according to Eq. (3.1), lie on the same striation. The key assumption used in WI-based ranging is that the magnitude

of the Green's function only varies slowly along the striations and that the variation is independent of the considered striation (see, e.g., [131] and [132]). Based on this assumption, we introduce the channel model

$$|g_{n,k}| = \gamma_k |g_{n',k'}|, \quad (3.10)$$

where the scaling factor $\gamma_k \in \mathbb{R}^+$ is independent of the considered striation and only frequency-dependent.

3.3.5 Signal Model in Striation-Frequency Domain

To facilitate statistically principled range estimation, we transform the received data from the range-frequency domain to the striation-frequency domain, where acoustic intensities are sampled along projected striation curves defined by Eq. (3.1). The striation-frequency domain serves as a physics-informed coordinate system in which the statistics of intensity and normalized intensity measurements along each striation are analytically tractable for inference.

The transformation is carried out using a candidate parameter vector $\mathbf{q} = [r, \dot{r}, \beta]^T$. In the transformed domain, the original range axis of the spectrogram is replaced by a striation index $l \in \mathcal{L} = \{1, \dots, L\}$, which corresponds to projected striations. Specifically, for a fixed reference frequency $f' \in \mathcal{F}$ and each reference range r' along the range axis, a striation curve is generated by computing range r across frequencies $f_k \in \mathcal{F}$ that satisfy Eq. (3.1). Only those striations that span the full 42–49 Hz frequency band are retained, and the striation axis in the transformed domain is defined by these valid projections. (See Fig. 3.2 for a visual illustration of the projected striations, which define the striation axis in the transformed domain. Practical details are further provided in Sec.3.6.1.)

We define the intensity vector for a given striation as $\mathbf{x}_l(\mathbf{q}) = [\underline{x}_{l,1}(\mathbf{q}), \dots, \underline{x}_{l,K}(\mathbf{q})]^T$, where $\underline{x}_{l,k}(\mathbf{q})$ is obtained by interpolating measurements from the r - f surface plot along

the l^{th} striation. We assume that after this transformation, intensities and normalized intensities remain statistically independent across k and l , i.e., signal models in Ch. 3.3 also hold for n replaced by l . Furthermore, Eq. (3.10) can be re-expressed in the transformed domain as

$$|\underline{g}_{l,k}| = \gamma_k |\underline{g}_{l,K}|. \quad (3.11)$$

For future reference, a variable with an underline notation, i.e., $\underline{x}$, indicates that the variable, i.e., x , has been transformed using Eq. (3.1). More details on the transformation of the intensities from range-frequency domain to striation-frequency domain will be provided in Ch. 3.6.1.

3.4 The Proposed Range Estimation Method

We aim to estimate the source range r by maximizing a joint likelihood function that, for a candidate parameter vector $\mathbf{q}$, describes the match between modeled intensities and measured intensities statistically. This framework also allows for the estimation of WI β when the range and range rate are known, or range rate $\dot{r}$ when the range and WI are known. We therefore define a general likelihood function in terms of a joint parameter vector $\mathbf{q} = [r, \dot{r}, \beta]^T$. This formulation allows us to solve for range or WI or constant range rate within the same estimation framework.

We evaluate the PDFs of both the intensity measurements $\underline{x}_{l,k}^b(\mathbf{q})$ at frequencies $f_k \in \mathcal{F}^b$ with only the broadband components and the normalized intensity measurements $\underline{y}_{l,k}^{\text{bt}}(\mathbf{q})$ at frequencies $f_k \in \mathcal{F}^{\text{bt}}$ with both broadband and tonal components across all striations $l \in \mathcal{L}$. The joint likelihood function of $\mathbf{q}$ given the intensity measurements $\underline{x}_{l,k}^b(\mathbf{q})$ at frequencies $f_k \in \mathcal{F}^b$ is given by

$$\ell^b(\mathbf{q}) = \prod_{k \in \mathcal{K}^b} \prod_{l \in \mathcal{L}} f_{\text{Exp}}(\underline{x}_{l,k}^b(\mathbf{q}); \hat{\underline{\theta}}_{l,k}(\mathbf{q})), \quad (3.12)$$

and the joint likelihood function of $\mathbf{q}$ given the normalized intensity measurements $\underline{y}_{l,k}^{\text{bt}}(\mathbf{q})$ at frequencies $f_k \in \mathcal{F}^{\text{bt}}$ is defined as

$$\ell^{\text{bt}}(\mathbf{q}) = \prod_{k \in \mathcal{K}^{\text{bt}}} \prod_{l \in \mathcal{L}} f_{\chi^2}(\underline{y}_{l,k}^{\text{bt}}(\mathbf{q}); \hat{\lambda}_{l,k}(\mathbf{q})). \quad (3.13)$$

The scaling factor $\hat{\theta}_{l,k}$ and the noncentrality parameter $\hat{\lambda}_{l,k}$ are estimated from the measurements. The functional form of $f_{\text{Exp}}(\cdot)$ and $f_{\chi^2}(\cdot)$ are provided in Eqs. (3.7) and (3.9), respectively. Finally, the joint likelihood function of $\mathbf{q}$ given both intensity and normalized intensity measurements in their respective frequencies is

$$\ell(\mathbf{q}) = \ell^{\text{b}}(\mathbf{q}) \ell^{\text{bt}}(\mathbf{q}). \quad (3.14)$$

We aim to find the parameter vector $\mathbf{q}$ that best aligns the measured intensities and normalized intensities with their modeled distributions. To estimate $\mathbf{q}$, we seek the most probable set of parameters that maximize the likelihood. This follows a standard ML estimation framework, where the optimal estimate is given by

$$\hat{\mathbf{q}}^{\text{ML}} = \arg \max_{\mathbf{q} \in \mathcal{S}} \ell(\mathbf{q}). \quad (3.15)$$

In practice, the logarithm of the likelihood, i.e., the log-likelihood, is maximized for numerical stability.

As discussed in Ch. 3.2.2, unambiguous estimation of $\mathbf{q} = [r, \dot{r}, \beta]^T$ is only possible if the search interval for two of the three parameters is small. To estimate range based on Eq. (3.15), it is assumed that the range rate $\dot{r}$ and the WI β in the joint parameter vector $\mathbf{q}$ are known. In practice, we estimate them using separate methods (see Ch. 3.2.3) and use their estimates $\hat{\beta}$ and $\hat{\dot{r}}$ in the likelihood function. When performing range estimation,

this leads to the practical formulation $\mathbf{q} = [r, \hat{r}, \hat{\beta}]^T$, and the estimate is given by

$$\hat{r}^{\text{ML}} = \arg \max_{r \in \mathcal{S}_r} \ell(\mathbf{q} = [r, \hat{r}, \hat{\beta}]^T). \quad (3.16)$$

Similarly, if the range and the range rate are treated as known (estimated as $\hat{r}$ and $\hat{\dot{r}}$), the WI is estimated as

$$\hat{\beta}^{\text{ML}} = \arg \max_{\beta \in \mathcal{S}_\beta} \ell(\mathbf{q} = [\hat{r}, \hat{\dot{r}}, \beta]^T). \quad (3.17)$$

Note that likelihood evaluation is performed in the striation-frequency domain corresponding to a candidate parameter vector $\mathbf{q}$. The key challenge is obtaining the statistical parameters necessary to evaluate the likelihood function. We outline their computation in the following sections. Key steps are provided in Algorithm 1. We also introduce the normalized intensity matrix $\underline{\mathbf{Y}}^{\text{bt}}$ of size $L \times J$ that is composed of the measurements $\underline{y}_{l,k}^{\text{bt}}(\mathbf{q})$.

A preliminary step involves identifying and separating $\mathcal{F}^{\text{b}}$ and $\mathcal{F}^{\text{bt}}$. Given a frequency band of interest $\mathcal{F}$, the set of frequencies with tonal components ($\mathcal{F}^{\text{bt}}$) consists of the peak frequencies of the power spectral density (PSD) estimates, and $\mathcal{F}^{\text{b}} = \mathcal{F} \setminus \mathcal{F}^{\text{bt}}$. The measurement matrix is then separated into $\underline{\mathbf{X}}^{\text{b}}$ and $\underline{\mathbf{X}}^{\text{bt}}$ accordingly. The practical details on how $\mathcal{F}$, $\mathcal{F}^{\text{bt}}$ and $\mathcal{F}^{\text{b}}$ were selected are provided in Secs. 3.5.2 and 3.6.1. In what follows, to simplify notation, the dependency on $\mathbf{q}$ is omitted when clear from the context.

3.4.1 Broadband-Plus-Noise Variance Estimation

The broadband-plus-noise variance $\underline{\sigma}_{l,k}^2$ at striation with index $l \in \mathcal{L}$ and a frequency with index $k \in \mathcal{K}^{\text{b}}$, defined in Eq. (3.5), is required to compute the normalized intensity measurements which are then used within the likelihood function. This variance is estimated by determining the scale parameters of the intensity measurements using the

Algorithm 1: ML Range Estimation

Data: A spectrogram $\mathbf{X}$ of size $K \times N$

Input : $\mathcal{S}_r, \hat{\beta}, \hat{\mathbf{r}}, \mathcal{F}^b, \mathcal{F}^{bt}, (\sigma^u)^2$

Output: $\hat{r}^{\text{ML}}$

for $r \in \mathcal{S}_r$ **do**

```

     $\mathbf{q} \leftarrow [r, \hat{\beta}, \hat{\mathbf{r}}]^T$ ;
     $\underline{\mathbf{X}} \leftarrow \text{SampleIntensitiesAlongStriations}(\mathbf{X}, \mathbf{q})$ ;
     $\underline{\mathbf{X}}^b \leftarrow \text{GetBroadbandComponents}(\underline{\mathbf{X}}, \mathcal{F}^b)$ ;
     $\underline{\mathbf{X}}^{bt} \leftarrow \text{GetBroadbandTonalComponents}(\underline{\mathbf{X}}, \mathcal{F}^{bt})$ ;
    /* Process measurements in  $\mathcal{F}^b$  (Ch. 3.4.1) */
     $\hat{\sigma}^2 \leftarrow \text{EstimateBroadbandVariances}(\underline{\mathbf{X}}^b, (\sigma^u)^2)$ ;
    /* Process measurements in  $\mathcal{F}^{bt}$  (Ch. 3.4.3) */
     $\underline{\mathbf{Y}}^{bt} \leftarrow \text{ComputeNormalizedIntensity}(\underline{\mathbf{X}}^{bt}, \hat{\sigma}^2)$ ;
     $\hat{\underline{\lambda}} \leftarrow \text{Estimate}\chi^2\text{Parameters}(\underline{\mathbf{Y}}^{bt})$ ;
    /* Compute Log-likelihood */
     $LL(r) \leftarrow \text{Loglikelihood}(\underline{\mathbf{X}}^b, \underline{\mathbf{Y}}^{bt}, \hat{\sigma}^2, \hat{\underline{\lambda}})$ ;

```

$\hat{r}^{\text{ML}} = \arg \max_{r \in \mathcal{S}_r} LL(r)$;

relationship established in Ch. 3.3, i.e.,

$$\underline{\theta}_{l,k} = \underline{\sigma}_{l,k}^2. \quad (3.18)$$

We leverage the fact that, under the transformation defined by the correct parameter vector $\mathbf{q}$, the scale parameters of the exponentially distributed intensities within a striation (across frequency indices $k \in \mathcal{K}^b$) follow a structured relationship dictated by the Green's function (Ch. 3.3.4). Within our statistical model, at a reference frequency $f_{k'} \in \mathcal{F}$, the scale parameter varies across striations (cf. Eqs. (3.5) and (3.18)). At any other frequency $f_k \in \mathcal{F} \setminus \{f_{k'}\}$, the scale parameter for a given striation is a scaled version of the corresponding scale parameter at $f_{k'}$. The scaling factor is the same for all striations at a given frequency, as it depends only on the frequency-dependent variations in the magnitude of the source amplitude and the channel transfer function (cf. Eqs. (3.3), (3.10), and (3.11)). These scaling factors can be estimated by comparing measured intensities

across different frequencies.

In particular, computing the broadband-plus-noise variance $\underline{\sigma}_{l,k}^2$ can be mathematically described as follows. For a given frequency index $k \in \mathcal{K}^b$, we define $\eta_k = \sigma_k^b / \sigma_{k'}^b$ ($\eta_k > 0$) as the ratio of source broadband signal standard deviations at f_k and $f_{k'}$, where $k' \in \mathcal{K}^b$ is a fixed reference frequency index chosen arbitrarily. Consequently, the source signal can be modelled as $\mathbf{s}_{n,k}^b \sim \mathcal{CN}(0, (\sigma_{k'}^b)^2 \eta_k^2)$. Based on Eqs. (3.5) and (3.18),

$$\underline{\theta}_{l,k} = (\sigma_k^b |\underline{g}_{l,k}|)^2 + (\sigma_k^u)^2. \quad (3.19)$$

Using the definition of η_k and Eq. (3.11), it follows that

$$\underline{\theta}_{l,k} = (\eta_k \gamma_k)^2 (\sigma_{k'}^b |\underline{g}_{l,k'}|)^2 + (\sigma_k^u)^2, \quad (3.20)$$

where k' is again the reference frequency index. We introduce the received broadband scaling factor $\alpha_k = (\eta_k \gamma_k)^2$ and the received broadband signal variance $\underline{v}_{l,k} = (\sigma_{k'}^b)^2 |g_{n,k}|^2$, for $k \in \mathcal{K}^{bt}$. Then, the scale parameter $\underline{\theta}_{l,k}$ can be expressed as

$$\underline{\theta}_{l,k} = \alpha_k \underline{v}_{l,k'} + (\sigma_k^u)^2. \quad (3.21)$$

3.4.2 Scaling Factor Estimation

To estimate $\underline{\theta}_{l,k}$, we first compute the estimates $\hat{\alpha}_k$ of α_k and $\hat{\underline{v}}_{l,k'}$ of $\underline{v}_{l,k'}$. The received broadband scaling factor α_k is estimated as

$$\hat{\alpha}_k = \frac{\frac{1}{M} \sum_{l \in \mathcal{L}} \underline{x}_{l,k}^b - (\sigma_k^u)^2}{\frac{1}{M} \sum_{i \in \mathcal{L}} \underline{x}_{i,k'}^b - (\sigma_{k'}^u)^2}. \quad (3.22)$$

The a detailed derivation and justification of this estimator is provided in App. A. Additionally, the received broadband signal variance $\underline{v}_{l,k'}$ is estimated as

$$\hat{\underline{v}}_{l,k'} = \frac{1}{\frac{1}{K-J} \sum_{k \in \mathcal{K}^b} \hat{\alpha}_k} \left(\frac{1}{K-J} \sum_{k \in \mathcal{K}^b} \underline{x}_{l,k}^b - (\sigma_k^u)^2 \right). \quad (3.23)$$

This estimator is also derived in detail in App. A. Finally, based on Eq. (3.21), Eq. (3.22) and Eq. (A.11), an estimate of $\underline{\theta}_{l,k}$ is given by

$$\hat{\underline{\theta}}_{l,k} = \hat{\alpha}_k \hat{\underline{v}}_{l,k'} + (\sigma_k^u)^2. \quad (3.24)$$

Recall that the background noise variance $(\sigma_k^u)^2$ is known. Then, based on Eq. (3.18), an estimate of the broadband-plus-noise variance $\underline{\sigma}_{l,k}^2$ is given by

$$\hat{\underline{\sigma}}_{l,k}^2 = \hat{\underline{\theta}}_{l,k}. \quad (3.25)$$

In what follows, it is assumed that the estimate $\hat{\underline{\sigma}}_{l,k}^2$ for $k \in \mathcal{K}^b$ is perfect (no error), i.e., we use the variance estimate $\hat{\underline{\sigma}}_{l,k}^2$ as the true variance $\underline{\sigma}_{l,k}^2$.

3.4.3 Broadband and Tonal Components Parameter Estimation

The Normalized Intensity Computation

At striation with index $l \in \mathcal{L}$ and frequency index $k \in \mathcal{K}^{\text{bt}}$, let $\underline{y}_{l,k}^{\text{bt}} = \underline{x}_{l,k}^2 / (\underline{\sigma}_{l,k}^2 / 2)$ be the normalized intensity. We compute the samples of the normalized intensity using the estimates in Eq. (3.25) at the neighboring frequencies of $f_k \in \mathcal{F}^{\text{bt}}$. By making use of the assumption that the standard deviation of the broadband component varies smoothly with frequency index k as discussed in Ch. 3.3, an estimate of $\underline{\sigma}_{l,k}^2$ for $k \in \mathcal{K}^{\text{bt}}$ is obtained as

$$\hat{\underline{\sigma}}_{l,k}^2 = \frac{1}{2} \left(\hat{\underline{\sigma}}_{l,k-\Delta k}^2 + \hat{\underline{\sigma}}_{l,k+\Delta k}^2 \right). \quad (3.26)$$

Here, the frequency index offset Δk is chosen such that $k \pm \Delta k \in \mathcal{K}^b$, and thus, the estimates in Eq. (3.25) can be utilized. In what follows, it is assumed that the estimate $\hat{\sigma}_{l,k}^2$ for $k \in \mathcal{K}^{\text{bt}}$ is perfect (no error). Thus, the computed normalized intensity values $\underline{y}_{l,k}^{\text{bt}} = \underline{x}_{l,k}^2 / (\hat{\sigma}_{l,k}^2 / 2)$ are samples of the PDF $\chi^2(\underline{y}_{l,k}^{\text{bt}}; 2, \underline{\lambda}_{l,k})$ with the noncentrality parameter $\underline{\lambda}_{l,k} = |\underline{p}_{l,k}^{\text{t}}|^2 / (\hat{\sigma}_{l,k}^2 / 2)$ as discussed in Ch. 3.3.

Noncentrality Parameter Estimation

To compute the likelihood function using the samples of normalized intensity, we need to estimate the noncentrality parameters $\underline{\lambda}_{l,k}$ of the normalized intensity. The numerator of the noncentrality parameter $|\underline{p}_{l,k}^{\text{t}}|^2 = (m_k^{\text{t}} |\underline{g}_{l,k}|)^2$ is the intensity of the tonal source signal modulated by the channel transfer function.

Consistent with the variance estimation approach, under the transformation defined by the correct parameter vector $\mathbf{q}$, the noncentrality parameter within a striation (across frequency indices $k \in \mathcal{K}^{\text{bt}}$), is determined by the frequency-dependent propagation effects of the channel and the magnitude of the source amplitude (cf. definition of $|\underline{p}_{l,k}^{\text{t}}|^2$ above as well as Eqs. (3.3), (3.10), and (3.11)). While the noncentrality parameters $\underline{\lambda}_{l,k}$ vary across striations, they are scaled versions of a common striation-dependent parameter at each frequency, meaning their relative change with frequency is consistent across all striations. This structure makes it possible to estimate the noncentrality parameters $\underline{\lambda}_{l,k}$, based on normalized intensity measurements across both frequency and striation indices.

We estimate the noncentrality parameter $\underline{\lambda}_{l,k}$ as

$$\hat{\underline{\lambda}}_{l,k} = \frac{\left(\frac{1}{J} \left(\sum_{i \in \mathcal{K}^{\text{bt}}} \underline{y}_{l,i}^{\text{bt}}\right) - 2\right) \left(\frac{1}{M} \left(\sum_{j \in \mathcal{L}} \underline{y}_{j,k}^{\text{bt}}\right) - 2\right)}{\frac{1}{JM} \left(\sum_{i \in \mathcal{K}^{\text{bt}}} \sum_{j \in \mathcal{L}} \underline{y}_{i,j}^{\text{bt}}\right) - 2}. \quad (3.27)$$

A detailed derivation of this estimator is provided in App. A.

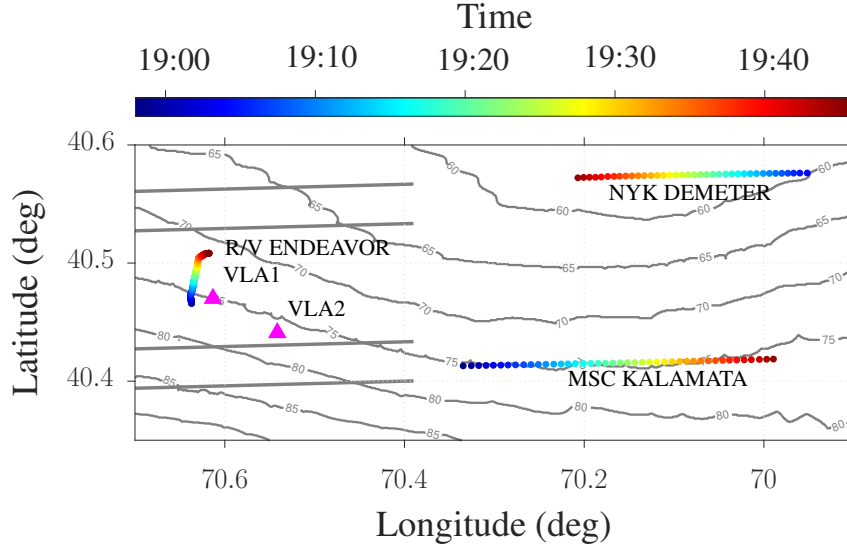


Figure 3.4. The tracks of the ships between 18:58 and 19:45 on March 24th, 2017 (UTC), in the vicinity of the VLA2 overlayed on the bathymetry map during the SBCEX17. The straight, gray lines indicate the shipping lanes, and the gray contour lines indicate the bathymetry (water depth) with 5 m intervals.

3.5 Dataset: SBCEX17

To evaluate the performance of the proposed range estimation method, underwater acoustic data from a single hydrophone were analyzed.

3.5.1 Event Overview

The selected event occurred from 18:58 to 19:45 UTC on March 24, 2017, during SBCEX17 [1]. During this period, the large commercial ship MSC KALAMATA (MMSI: 477510600) traveled in a straight trajectory at an average reported speed of 19.9 kn (≈ 10.2 m/s) along a shipping lane southeast of two vertical line arrays (VLAs) deployed by the *Marine Physical Laboratory*. The event began 26 minutes after CPA at 3.1 km relative to VLA2 (40.442°N, 70.527°W). This time window was selected because the range rate remained nearly constant, differing from the reported ship speed by less than 0.2 m/s. The environmental conditions were relatively range-independent, with seafloor depth variations within 5 m (75–80 m) and an approximately constant sound speed of 1473 m/s, measured

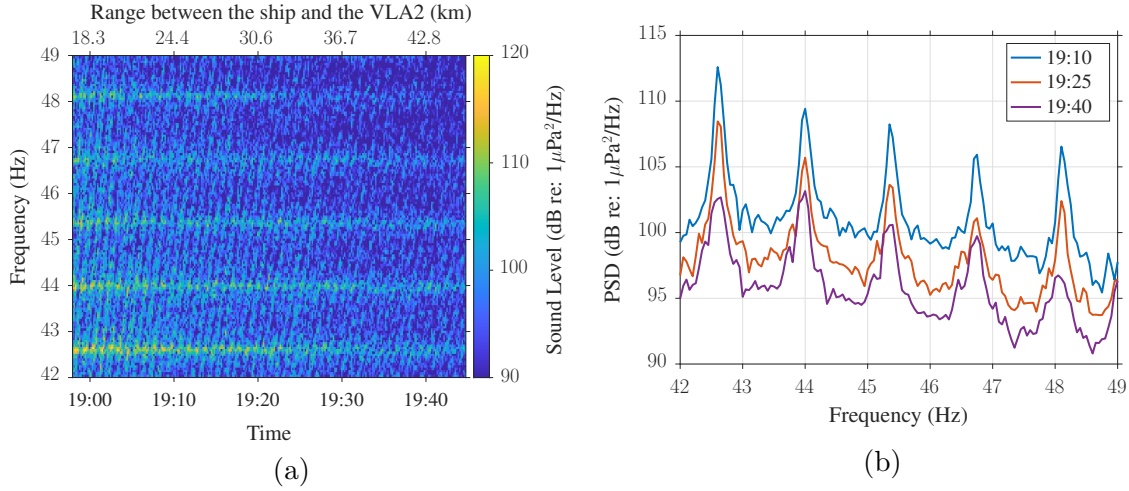


Figure 3.5. (a) The spectrogram of the processed acoustic signals over the time- and frequency-window of interest. (b) PSDs of the acoustic recording every 15 minutes after 19:10 on March 24th, 2017 (UTC) during the considered scenario. Both figures are results from processing the acoustic data recorded by the 8th element of the VLA2 (See Ch. 3.5.2 for details on the computations of the PSDs and the ambient noise.)

near the VLA2. We assumed that the SSPs varied mildly at larger distances at which the ship’s ranges were estimated. Range estimations were performed every two minutes, starting at 19:10 (Fig. 3.4). To estimate the range at 19:10, a spectrogram between 18:58 and 19:10 was needed.

3.5.2 Acoustic Data Processing

The eighth element of VLA2 (depth ≈ 33 m above the seafloor) recorded acoustic data at a sampling rate of 25 kHz, which was down-sampled to 2.5 kHz after applying a low-pass filter. The spectrogram used for range estimation was computed using 20-second snapshots, 50% overlap, and a Hamming window (null-null mainlobe width of 0.2 Hz), resulting in FFT bin centers spaced at 0.05 Hz, with 10-second intervals between snapshots. These parameters were chosen to minimize spectral leakage from the stronger tonal components of the ship. The spectrogram of the entire event is shown in Fig. 3.5(a), and a detailed view of the striations on the spectrogram is provided in Fig. 3.1 over the first 8 minutes.

Processing was performed in the 42–49 Hz frequency band to ensure a constant number of propagating modes within the processing time window, i.e., frequencies remained between two mode cut-off frequencies, preventing discontinuities in striation patterns, which are visible in the spectrogram of the acoustic data related to times when the ship is close to the CPA (not shown). The absence of discontinuities in the spectrogram was confirmed by visual inspection.

To track signal variations over time, PSDs were computed using Welch’s method [133] at 15-minute intervals beginning at 19:10 (Fig. 3.5(b)). Each PSD estimate used 90 consecutive snapshots (15-minute time windows), following the same processing parameters as used for the computation of spectrograms.

Comparisons with ambient noise were performed using acoustic data from 11:00–11:30, a period manually identified as having minimal ship activity based on AIS data. The maximum PSD during the quiet period in the 42–49 Hz band was 82 dB re: $1 \mu\text{Pa}^2/\text{Hz}$, while the minimum PSD during the event exceeded 90 dB re: $1 \mu\text{Pa}^2/\text{Hz}$, confirming that MSC KALAMATA’s broadband emissions dominated over background noise.

3.5.3 Potential Acoustic Interference

AIS data identified two nearby vessels that could have contributed to acoustic interference (Fig. 3.4). R/V Endeavor, moving at an average speed of 3.4 kn ($\approx 1.7 \text{ m/s}$), was closest to VLA2 at 6.0 km. Given its low speed, its contribution to interference in the frequency band of interest was assumed to be minor. Another large commercial ship, NYK DEMETER, was traveling in the opposite shipping lane at 15.5 kn and was 31 km away from VLA2 at 19:45, i.e., closer than MSC KALAMATA, which was at 46 km. However, further analysis showed that NYK DEMETER began slowing down around 19:40 and that its acoustic signature only began to dominate the spectrogram after 20:30, well after the analysis window used for range estimation. This suggests that NYK DEMETER

would have needed to be significantly closer and moving faster to noticeably affect the spectrogram during the estimation period. Thus, despite multiple potential interferers, MSC KALAMATA’s range was estimated accurately.

3.6 Implementation Aspects and Comparison Methods

Starting at 19:10, spectrograms are processed every two minutes for range estimation. Here, the time corresponds to the last snapshot of the spectrogram, i.e., the range $r = r_N$. The range estimation performance of the proposed method is compared to the representative versions of the existing methods. Since the vessel is 300 m long, it is expected that a certain bias is involved in the error calculation. Nonetheless, all methods, including the WI computation, use the same AIS data. All methods are thus affected the same way by the potential bias. The code of the proposed method can be found in our Github repository [134].

3.6.1 Implementation Aspects

The search interval: Every two minutes, the minimum and maximum candidate range values were set to $r_{\min} = 0.6 \cdot r_{\text{true}}$ and $r_{\max} = 1.4 \cdot r_{\text{true}}$, where r_{true} is the AIS-derived range. As a result, the set of candidate ranges $\mathcal{S}_r$ was set to a grid between $r_{\min}$ and $r_{\max}$ with a resolution of 10 m. The range rate was assumed to be constant at 10.2 m/s, equivalent to the reported ship speed. The proposed method and all comparison methods used $\beta = 1.18$.

WI for the environment: The spectrogram considered for range estimation, starting at 19:10, was also used to obtain the WI ($\beta = 1.18$) for the environment and the considered frequency band. To estimate the WI, the proposed approach was used to evaluate the likelihood function for a set of candidate WI values by fixing the range to reference values computed from information provided by the AIS (Eq. (3.17)). The search

range of β , i.e., the interval defining $\mathcal{S}_\beta$, was set between 0.8 and 1.3, centered around the canonical shallow-water value of 1.0. A coarse manual comparison of projected striations at various β values helped refine this range. The value $\beta = 1.18$ was assumed constant during the considered event, i.e., a range-independent environment was assumed during the considered processing window.

Given that the environment had almost a constant SSP and a flat bottom, the obtained value of 1.18 is worth further investigation because, typically, in such environments, one would expect $\beta = 1.0$ [113]. Our separate research and analysis in Ch. 4 validated this value. Given a water depth of 75 m and the considered frequency band centered at 45.5 Hz, the waveguide can only support a few propagating modes. It has been determined that striations observed at distances between 25–45 km resulted from modal interference between two dominant modes, whose mode pairwise WI value was close to 1.18 [115]. A detailed pairwise modal decomposition supporting this value is provided in Ch. 4.

Frequencies for $\mathcal{F}^b$ and $\mathcal{F}^{bt}$: Every two minutes, tonal center frequencies were identified as PSD peaks in the processed spectrogram. They were centered around $\mathcal{F}^{bt} = \{42.65, 44.00, 45.35, 46.75, 48.05\}$ Hz. Slight temporal variations occurred due to intrinsic frequency modulations present in the tonal signals and spectral discretization effects.

The broadband frequency set $\mathcal{F}^b$ was constructed to avoid overlap with tonal frequencies by selecting bins spaced at least 0.35 Hz away from each tonal. Specifically, three broadband frequency bins (spaced 0.35 Hz apart) were placed between each pair of adjacent tonal lines, along with additional bins located 0.35 Hz below the lowest tonal and 0.35 Hz above the highest tonal, yielding 14 broadband frequencies in total. The empirical spacing of 0.35 Hz was chosen to minimize spectral leakage while ensuring a sufficient number of broadband samples for variance estimation. Empirically, using a denser broadband bin spacing did not lead to meaningful accuracy gains on the considered dataset.

Spectrograms for range estimation: At each processing time, the spectrogram length N is chosen to ensure that $L = 30$ valid striations, i.e., those that span the entire 42 – 49 Hz band, can be projected using the maximum candidate source range in $\mathcal{S}_r$. This choice accounts for the fact that striations become flatter, i.e., less steep, at longer distances, requiring longer temporal windows to span the full frequency band. As a result, the required spectrogram duration increases with time as the ship moves away. This trend is evident in Fig. 3.5(a), which shows the full spectrogram for the considered event. Consequently, spectrogram durations increase with time as the ship moves away, e.g, from 12 minutes when estimating ship range at 19:10 to 22 minutes when estimating ship range at 19:45.

Striations are projected using the reference ranges $r' = r_n$ for $n \in \mathcal{N}$, at a fixed reference frequency $f' = 45.5$ Hz. The same number of valid striations is used across all candidate source ranges to ensure that an equal number of intensity measurements are evaluated for range estimation.

3.6.2 Comparison to Existing Methods

To assess the performance of the proposed statistical method, we compare it against four existing approaches used for WI-based passive ranging. Each group of methods is labeled with their acronym, and the proposed method is referred to as method (G).

Slope-based methods (S) Assuming the presence of a broadband signal, a class of methods [100, 127, 126] employs a two-dimensional discrete Fourier transform to analyze the range-frequency structure of acoustic striations, extracting the dominant slope as a key parameter for range estimation. These methods assume that striations are approximately linear within a short range-frequency window, enabling local slope approximation. The extracted slope, $df/dr = \beta f/r$, directly relates to the source range through the differential form of the WI equation.

However, these methods are sensitive to intensity variations, especially in the

presence of tonal components, and require uniform intensity along striations. Additionally, when $\beta \neq 1$, striation curvature introduces errors, necessitating careful selection of the window length to maintain accuracy [100]. In our study, we implemented the method presented in [126].

Tonal-based statistical method (T): Another class of WI-based ranging methods statistically processes prominent ship noise features [135, 136, 103, 106]. Young *et al.*[103] proposed a statistical model and search-based estimation method for tonal signals radiated by a moving ship in shallow water. While the general estimation framework is equivalent to the proposed method (G), it assumes that the background noise dominates over the broadband component of the ship noise, making the intensity statistics in non-tonal frequency bands effectively range-independent. Thus, the background noise level used for intensity normalization and noncentrality parameter $\underline{\lambda}$ estimation are computed directly from the off-tonal broadband frequency bins within the spectrogram. Additionally, method (T) requires estimating the noncentrality parameters $\underline{\lambda}$ for each striation via a search over candidate values, which increases computational cost. In contrast, the proposed method (G) computes these parameters directly from the data using a closed-form expression (cf. Eq. (3.27)), resulting in substantial computational savings. In our implementation of (T), the noncentrality parameter was searched from 0 to 100 in steps of 2.5. Empirically, the search step size smaller than 2.5 did not provide significant accuracy gain.

Broadband-based statistical method (B): Jang and Meyer [106] developed a statistical approach that processes the broadband component of received ship noise. The general estimation framework is equivalent to the proposed method (G), but the likelihood functions are evaluated using the PDFs of the broadband intensities only, without incorporating tonal frequencies.

Correlation-based method (C): Jang and Meyer [105] also proposed a deterministic approach that estimates range by analyzing intensity variations of tonal pairs in the spectrogram. For each pair of tonal frequencies, f_1 and f_2 , the method computes

their correlation across a set of candidate ranges. Using the WI equation (Eq. (3.1)), the intensity values at f_2 are sampled along the striations that intersect the candidate ranges at f_1 . The correlation between the intensity profile at f_1 and the sampled values at f_2 is then evaluated, and the candidate range that maximizes this correlation across all tonal pairs is selected as the estimate. This pairwise computation allows the use of striations that do not span all frequencies, increasing the available data. Mathematical formulation of the use of the correlation is provided in App. B.

In this study, methods (C), (T), and (B) use the same striation-frequency domain representation as method (G), since all methods evaluate acoustic intensities along projected striations within the same transformed spectrograms (Ch. 3.3.5). Accordingly, the spectrogram lengths vary with source range for all methods, and identical FFT parameters and frequency bands are used throughout.

3.7 Results

The five ranging methods from Sec.3.6 were evaluated by comparing their errors to AIS-derived ranges after 19:10. The range errors (in percent) are shown in Fig. 3.6. Percentage error is used as a range-normalized metric. In WI-based ranging, uncertainty in measured intensities along striations introduces range errors that grow approximately linearly with distance, but percentage errors remain relatively constant, offering a consistent measure across ranges.

All methods except for the slope-based approach demonstrated robust range estimation up to 45 km, with errors within $\pm 6\%$ (see Fig. 3.6 (b)), highlighting the potential of WI-based ranging at ranges of tens of kilometers (One anomalous result for (T) near 45 km is excluded.) The root mean square errors (RMSEs) for methods (G), (S), (T), (B), and (C) were 1.2%, 9.5%, 2.0%, 2.1%, and 1.9%, respectively. Fig. 3.6 (b) summarizes the statistical performance across methods (G), (T), (B), and (C), which yielded relatively

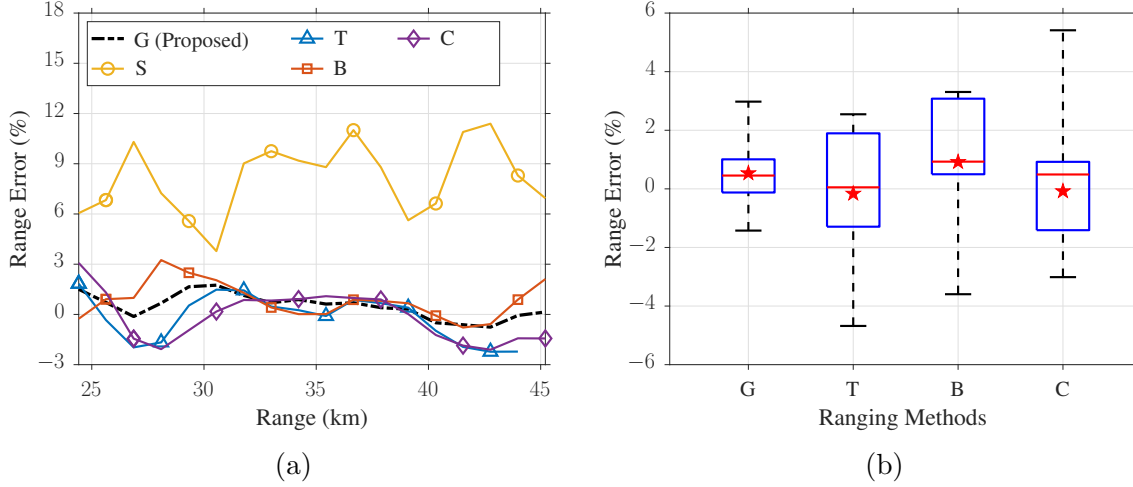


Figure 3.6. Range error (in percent) for five WI-based ranging methods: (G) proposed approach, (S) slope-based method, (T) statistical approach using only tonal signals, (B) statistical approach using only broadband signals, and (C) correlation-based method (see Ch. 3.6.2 for details). (a) Range errors over time (range) using AIS data as ground truth. Estimates are computed every two minutes and smoothed with a centered moving average (window size 3). (b) Box plot of the range errors for methods (G), (T), (B), and (C), which yielded low mean errors. In each box, the red line within indicates the median, the box edges mark the 25th and 75th percentiles, and the red star denotes the mean. Whiskers extend to the most extreme data points included in the analysis. One anomalous error near 45 km in method (T) was excluded from the analysis for clarity.

unbiased estimates, with both mean and median errors close to 0%.

The proposed method (G) outperformed all other methods across all considered statistical metrics, achieving the lowest RMSE (1.2%) and narrowest error spread, indicating both accuracy and consistency. In contrast, the slope-based method (S) exhibited the poorest performance, with a mean error of 9.5%, a standard deviation of 5.3%, and a maximum error of 17.1%, which is not fully shown in Fig. 3.6 (a) due to smoothing. This underperformance is likely due to intensity variation across frequencies and the method’s sensitivity to striation curvature. For $\beta > 1$, striations are curved, and linear slope-based methods that match only part of the curve would underestimate the true slope, leading to range overestimation.

The tonal-based (T), broadband-based (B) and correlation-based (C) methods

performed comparably. While method (T) leverages high-SNR tonal components, it also models the broadband energy as range-independent background noise, which is used both for normalizing the tonal intensity and for estimating the noncentrality parameter. However, this assumption is violated in our dataset due to modal interference of the radiated broadband signature of the ship, which introduces range-dependent fluctuations in broadband intensity, i.e., striation patterns, even outside the tonal bins. Method (B) explicitly models and exploits these broadband structures, which likely contributed to its strong performance. Method (C) benefits from using partially valid striations, increasing usable data and offering a simple, practical alternative.

A runtime analysis was conducted across methods, with runtimes normalized per minute of data to account for spectrogram lengths ranging from 12 to 22 minutes. The resulting average runtimes for estimating range from one minute of spectrogram data were 0.22, 1.87, 0.29, 0.23, and 0.02 seconds for methods (G), (S), (T), (B), and (C), respectively. All processing was conducted on a laptop with an *Apple M1 Pro* chip and 32 GB of memory. The correlation-based method (C) was the fastest due to its simplicity. In contrast, the slope-based method (S) was the slowest, primarily due to its computationally intensive procedure for evaluating a large number of candidate angles of the striations. The tonal-based method (T) required slightly more computation than methods (G) and (B), as it performs an exhaustive search over the noncentrality parameter during likelihood evaluation.

3.8 Conclusions and Future Work

This chapter proposes a statistical signal processing method for estimating the range between a stationary single hydrophone and a moving ship in range-independent shallow water, leveraging the spatial-interference pattern on the spectrogram of the ship’s radiated signature. This pattern is characterized by the WI. The proposed method was

validated using real acoustic recordings from a vertical line array element deployed during SBCEX17, where a large commercial ship transited along a nearby shipping lane. The proposed approach achieved range estimation errors below $\pm 3\%$ up to 45 km, outperforming reference methods. Furthermore, range estimation results on two additional commercial ships (presented in App. C) that transited similar tracks through the experiment region on different days further support the method’s effectiveness. By integrating the strengths of statistical methods that process either broadband or tonal components, the proposed method achieved the lowest estimation error among statistical approaches considered.

While the results demonstrate promising performance, certain assumptions, and experimental constraints warrant further investigation. The statistical parameters used for likelihood evaluation, such as the broadband scaling factors, noncentrality parameters, and ambient background noise statistics, were estimated from the data and assumed to be perfect during inference. Errors related to these parameter estimates are currently not captured by our statistical model, which may affect the robustness of our approach. While the proposed method demonstrated high estimation accuracy on the considered dataset, further testing and a comparison with reference methods across diverse environmental conditions, would be valuable to assess general applicability.

Areas for further investigation include extending the approach to range-dependent environments, where bathymetric and/or sound speed variations may affect the modal propagation and, consequently, the striation structure. Incorporating range-dependent WI formulations would enhance the applicability in a larger variety of ocean settings [109, 137, 138]. In addition, the method was evaluated in the relatively low-frequency band of 42–49 Hz, where a WI of 1.18 was observed. At higher frequencies, the approximation of $\beta = 1$ may be more applicable [113]. Applying and validating the method in higher frequency bands could help assess its broader utility. Finally, while the constant range rate was assumed based on AIS data, developing methods to estimate the range rate directly from the acoustic signal would remove the current

need to obtain this information from an additional sensor. Although previous work has proposed range-rate estimation based on phase differences across ranges, e.g., [102], our application of this method on ship-radiated noise did not yield reliable estimation results. Alternatively, a CPA-based range rate estimation approach—tracking multiple tones to extract Doppler shift information [55, 139]—could be explored in future studies.

Acknowledgments:

Chapters 3 and Appendix A are being prepared for submission as a Journal of the Acoustical Society of America manuscript: J. Jang, W. S. Hodgkiss, and F. Meyer, “Ranging of a Moving Ship Using a Single Acoustic Receiver in Shallow Water.” Portions of Chapter 3 are based on an earlier conference paper: J. Jang and F. Meyer, “A New Statistical Model for Waveguide Invariant-Based Range Estimation in Shallow Water,” in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Oct. 2024 (Copyright 2024, IEEE). The dissertation author was the primary investigator and lead author of this work.

Chapter 4

Mode-pairwise Waveguide Invariant Analysis in the New England Mud Patch

4.1 Introduction

This chapter investigates the mode-pairwise structure of the waveguide invariant (WI) in a shallow-water environment, focusing on how modal interference patterns evolve with frequency and range, and how these patterns are shaped by seabed properties and modal attenuation. A scalar WI value of $\beta = 1.18$ was previously estimated in Ch. 3 for the 42–49 Hz band using spectrograms recorded during a ship (MSC KALAMATA) passage in the Seabed Characterization Experiment 2017 (SBCEX17) [1]. This value significantly exceeds the canonical shallow-water approximation of $\beta \approx 1$, typically assumed in many realistic shallow water environments [113].

The findings in this chapter highlight that WI-based passive range estimation at low frequencies requires care: the effective WI cannot be assumed to be unity and may vary with frequency (especially near the mode cutoffs), propagation range (due to modal attenuation), and environmental factors such as sediment layering and sound speed profile below the seafloor. When only a few modes propagate, the interference pattern—and thus the WI—can differ across mode pairs. In such cases, β is better described as a distribution

rather than a fixed scalar [114, 140]. Nonetheless, in our dataset, modal attenuation beyond 16 km leaves only the first two modes active, justifying the use of a scalar WI at long range in the range estimation performed in Ch. 3.

To validate this interpretation, we conduct a combined simulation and data-driven analysis. First, we perform simulations using the normal mode program KRAKEN [141] with geoacoustic parameters derived from the SC2 site [2], located approximately 200 m from the vertical line array (VLA2) used in this study. The general study area (New England Mud Patch) is known for its interbedded mud and sand layers, and is well-studied in a significant number of geoacoustic inversions [142, 2, 143, 144]. Although the precise sediment properties at long ranges, over which a commercial ship’s range was estimated in Ch. 3, from the receiver are not known in detail, the SC2 model provides a realistic baseline for exploring the dispersion characteristics of the modes and how changes in environmental factors, such as mud thickness and sound speed, affect the WI estimation. The simulations reveal that only three modes propagate in the 42–49 Hz band, with the third mode attenuating significantly by 16–20 km. This motivates a mode-pairwise WI framework, where WI is estimated for each dominant mode pair to better capture the evolving interference structure.

We then apply this framework to acoustic recordings from a ship passage by MSC KALAMATA, which traveled from 3.2 km to 45.8 km from VLA2 over a 73 min interval (see Fig. 4.1). The real data are used in two ways: (1) to assess whether the simulated modal interference patterns and transitions match observed spectrograms (Ch. 4.3); and (2) to estimate mode-pairwise WI values from the recorded striations and determine whether the previously reported $\beta = 1.18$ reflects dominant modal interactions at long range (Ch. 4.4).

4.1.1 Contributions and Chapter Organization

This chapter makes the following contributions:

- A mode-pairwise framework [140] for estimating the WI β for each pair of observed

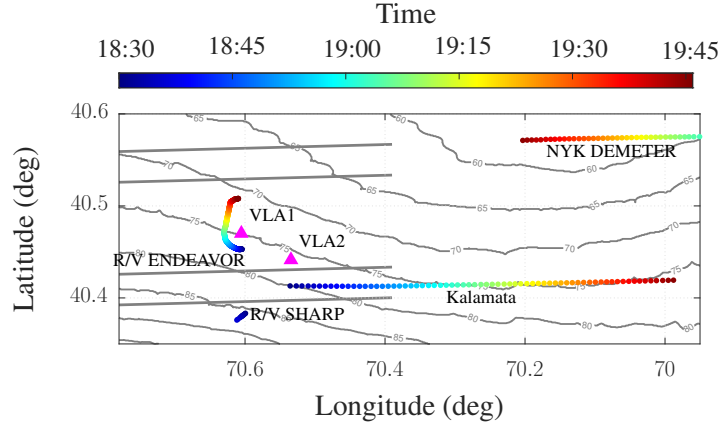


Figure 4.1. The tracks of the ships between 18:32 and 19:45 on March 24th, 2017 (UTC), in the vicinity of the vertical line array (VLA2) overlayed on the bathymetry map during the Seabed Characterization Experiment 2017 (SBCEX17) [1]. The site SC2 at which the geoacoustic properties are estimated in [2] and used in this chapter is located only 200 m away from VLA2 location. The straight, gray lines indicate the shipping lanes. Notice that the considered event in this text starts near the closest point of approach (CPA) instead of at 18:58 for the range estimation in Ch. 3.

modes is applied to real shallow-water acoustic data recording during SBCEX17.

- The observed $\beta = 1.18$ is primarily caused by the interference between the first two modes, with the third modes fading at long range.
- The sensitivity of WI values to environmental variability is examined by simulating changes in mud layer thickness, sound speed, and density based on the geoacoustic model at the site SC2 .
- The modal interference pattern and WI estimates are confirmed to be robust at long range and supports the use of a scalar WI for passive acoustic range estimation in the 42–49 Hz band beyond 16 km.

The chapter is organized as follows: Section 4.2 reviews modal interference theory and the concept of the WI. Ch. 4.3 presents simulation results using SC2 site-based geoacoustic models to characterize the propagating modes and their corresponding mode-pairwise WI sensitivity to environmental conditions and range. Ch. 4.4 studies the evolution

of propagating modes as a function of range and estimate the corresponding pairwise WI using the real acoustic data collected at SBCEX17.

4.2 Background

The WI β has been introduced as a parameter that encapsulates modal dispersion in a given waveguide and characterizes the resulting interference pattern (striations) in the range-frequency (r - f) surface plot in Ch. 3 [107]. In this section, we first review the modal interference, and then, we provide a more detailed explanation of the WI in two different contexts: (i) when a group of modes are propagating (at high frequencies) and (ii) when there is few modes propagating (at low frequencies).

4.2.1 Modal Interference and Striation Formation

In shallow-water waveguides, sound propagates as a superposition of discrete normal modes [107]. The number of propagating modes depends on frequency and the environmental conditions. Each mode is associated with a horizontal wavenumber $k_n(f)$ (for the n^{th} mode and at frequency f), which governs its phase evolution as it propagates in range.

For a range-independent environment, where a source at range r emitting an acoustic signal with power spectral density $P(f)$ at frequency f , the received acoustic intensity follows [107]

$$I(r, f) \propto \frac{P(f)}{r} \sum_{n \neq m} B_m B_n \cos(\Delta k_{mn}(f)r), \quad (4.1)$$

where $\Delta k_{mn}(f) = k_m(f) - k_n(f)$ represents the difference in horizontal wavenumbers between modes m and n , and the terms B_m and B_n represent depth- and mode-dependent amplitude coefficients that scale the contribution of each mode in the received intensity. The cosine term inside the summation oscillates with spatial frequency $\Delta k_{mn}(f)$, resulting

in constructive and destructive interference as a function of range and frequency. Terms omitted in Eq. (4.1) contribute to a non-oscillatory background intensity that do not affect the interference patterns. These interference patterns form striations in the r - f surface plot, where the local slope of the striations depends on the modal dispersion properties.

4.2.2 Waveguide Invariant

The WI β is a parameter that characterizes how modal interference patterns evolve with range and frequency. It is formulated as a group-wise parameter [145], capturing the shared dispersion trends of a set of propagating modes. At high frequencies, typically on the order of hundreds of Hertz depending on the water depth, a group of adjacent modes may propagate well above their cutoff frequencies. In this regime, both their phase and group velocities vary only mildly with frequency. As a result, their interference patterns can be described collectively as a group, rather than as the sum of individual mode-pair contributions. This leads to the WI, defined as [107]

$$\begin{aligned}\beta &\equiv \frac{r}{f} \frac{\delta f}{\delta r} \\ &= -\frac{d(1/v)}{d(1/u)},\end{aligned}\tag{4.2}$$

where v and u are phase and group speeds of the modes in the considered group of modes. This mode-independent formulation allows the use of a single effective WI β in passive acoustic ranging [100, 103]. By taking the integral of the differential form in Eq. (4.2) using separation of variable, we obtain

$$\frac{f}{f'} = \left(\frac{r}{r'} \right)^\beta,$$

equivalent to Eq. (3.1) in Ch. 3. This equation enables projection of striation on the r - f surface plot.

The value of β varies depending on the properties of the waveguide and the considered modes, but certain approximations hold under practical considerations. In an ideal waveguide with both pressure-release surface and bottom, the analytical expression of the WI is $\beta = \cos^2(\psi)$, where ψ is the grazing angle of the mode. For a small ψ (approximately $\psi \leq 20^\circ$), $\beta \approx 1$ which is a reasonable approximation for long-range propagation [107]. In realistic shallow water environments, such as a Pekeris waveguide with a penetrable half-space bottom or a non-uniform sound speed profile (SSP), $\beta \approx 1$ remains a valid approximation if the dominant propagating modes are surface- and bottom-reflected [113]. These modes share similar dispersion characteristics with those in an ideal waveguide, making $\beta = 1$ a commonly used assumption in source ranging applications in shallow water environments [128, 100, 104].

4.2.3 Mode-pairwise Waveguide Invariant

On the other hand, when multiple groups of modes with distinct dispersion characteristics or only a few modes propagate, the WI may be better described as a distribution. Rouseff and Spindel [114] showed that in a waveguide whose dominant propagating modes have weak or negligible energy near the ocean surface—such as in cases where modes are trapped beneath a thermocline— β is better modeled as a distribution rather than a single, constant value. To theoretically support their argument, they extended the WI concept to pairs of modes. Specifically, consider a pair of modes m and n ; the pair’s local WI is defined as [114]

$$\beta_{mn} \equiv -\frac{v_m^{-1} - v_n^{-1}}{u_m^{-1} - u_n^{-1}}, \quad (4.3)$$

where v and u are the phase and group velocities of the corresponding modes indicated by the subscript. Variability in β_{mn} values for different mode pairs was demonstrated with simulated data in [115], particularly at low frequencies where only a few modes propagate. If dispersion characteristics for the considered pair of modes are similar over a narrow frequency band, then β_{mn} is approximated as a constant [115], and the functional

relationship for striations remains the same as in Eq. (3.1) where β is replaced by β_{mn} . However, since the interference pattern is mode-pair specific, it may not be observable at certain depths where the mode function has a null, resulting in little to no modal contribution at those depths.

The assumption that β is a scalar—a single value independent of mode numbers as defined in Eq. (4.2)—holds only if the WI values for mode pairs (β_{mn}) with significant energy contributions to the acoustic field are approximately equal. A trivial case is when only one propagating mode pair propagates, which is the case with the acoustic data used for range estimation at long ranges in Ch. 3. To demonstrate that it is the case, we analyze both simulation data and the acoustic data from Ch. 3 using this perspective of mode-pairwise WI.

4.3 Simulation and Validation Against Observed Modal Interference Patterns

4.3.1 Method

In this section, we investigate the relationship between geoacoustic structure and modal interference. To demonstrate that only a few modes propagate in the considered environment and frequency band, we simulate acoustic propagation using the normal mode program KRAKEN with geoacoustic parameters derived from site SC2, located near the receiver (VLA2). These properties were estimated through inversion in Ref. [2]. The corresponding simplified seabed model is summarized in Tab. 4.1. For the simulations, some thin inter-mud layers with slightly different properties were interpolated and combined into continuous layers with linearly varying sound speed as well as weighted averaged densities and attenuation values. Comparison between the original and simplified models showed no appreciable difference in the resulting modal structures or simulated acoustic fields, justifying the use of the simplified model.

Table 4.1. The established parameter space for the simplified geoacoustic model at SC2 [2]. The columns represent the layer thickness h , the sound speed at the top c_t and the bottom c_b of the layer, the density ρ , and the attenuation α .

Layer	Label	h (m)	c_t (m/s)	c_b (m/s)	ρ (kg/m ³)	α (dB/m-kHz)
1	Mud1	4.0	1466	1488	1.78	0.040
2	Mud2	6.8	1488	1499	1.84	0.040
3	Mud-Sand	0.9	1499	1701	2.04	0.120
4	Sand	1.1	1735	1735	2.16	0.420
5	Deep	1.6	1882	1882	2.29	0.100
6	Basement	∞	1969	—	2.35	0.890

In the presented study, we use the geoacoustic parameters in Tab. 4.1, along with a water depth of 75 m at VLA2 and a source depth of 9 m (the ship’s draft was reported as 14 m). The average water SSP measurements in the vicinity of VLA2 was used as the water SSP whose value was almost isovelocity around 1470 m/s. Since these parameters represent conditions near the VLA2, discrepancies would exist in the actual environment at longer ranges. A noticeable difference is that the measured mud thickness along the commercial ship MSC KALAMATA track is approximately 7 m by projecting the position over the map provided in [143]. Nonetheless, assuming mild variations, our simulation provides useful insights into validating the observed interference pattern.

Our simulations serve three primary purposes:

1. **Qualitative validation:** Simulated spectrograms are compared with the measured spectrograms in two range intervals (3.5–10 km and 16–24 km), to confirm that modal interference patterns and transitions are consistent with observations.
2. **Modal characterization:** The modal properties—including group speed, phase speed, attenuation, and mode-pairwise WI values—are analyzed to interpret the physical basis of the observed striations. Three propagating modes are observed in the 42-49 Hz band near the CPA.
3. **Sensitivity analysis:** We evaluate how changes in mud layer thickness, sound

speed, and density affect the WI values for the mode pair of the first two modes, especially at long ranges where the third mode is attenuated and only the first two modes remain dominant.

4.3.2 Results

To assess how well simulations reflect observed acoustic propagation, we compare modeled modal interference patterns with real data from a ship passage event near the VLA2 (see Fig. 4.1). The spectrograms generated from real data are grouped into close-range (3.5–10 km) and far-range (16–24 km) intervals and evaluated alongside simulated r - f surface plots for the SC2 site. Here, the range refers to the range between the considered ship (MSC KALAMATA) and the VLA2.

Close-range Comparison (3.5–10 km)

In the close range, direct comparison of simulated and measured spectrograms (Fig. 4.2(a) and (b)) reveals moderate agreement in modal features but also key discrepancies. The real spectrograms are partially obscured by the ship’s tonal and broadband noise signatures as well as background ambient noise, complicating visual interpretation. Nevertheless, a consistent transition in the striation pattern emerges near 41–42 Hz. The cut-off frequency was computed approximately as 36.2 Hz, and thus, we will refer to the 41–42 Hz as the transition frequency range. Simulations based on SC2 geoacoustic parameters predict that mode 3 is present below 41 Hz at close ranges (< 5 km) but rapidly attenuates with increasing range. Although the real data are noisy, the modal transition is observable and suggests that environmental conditions near VLA2 are better approximated by a transition frequency range around 41–42 Hz. Minor bathymetric variations (from 80 m to 77 m over the 3.5–10 km track) may partially explain the offset between simulation and observation, as shallower depths tend to increase the same mode’s cutoff frequency [107].

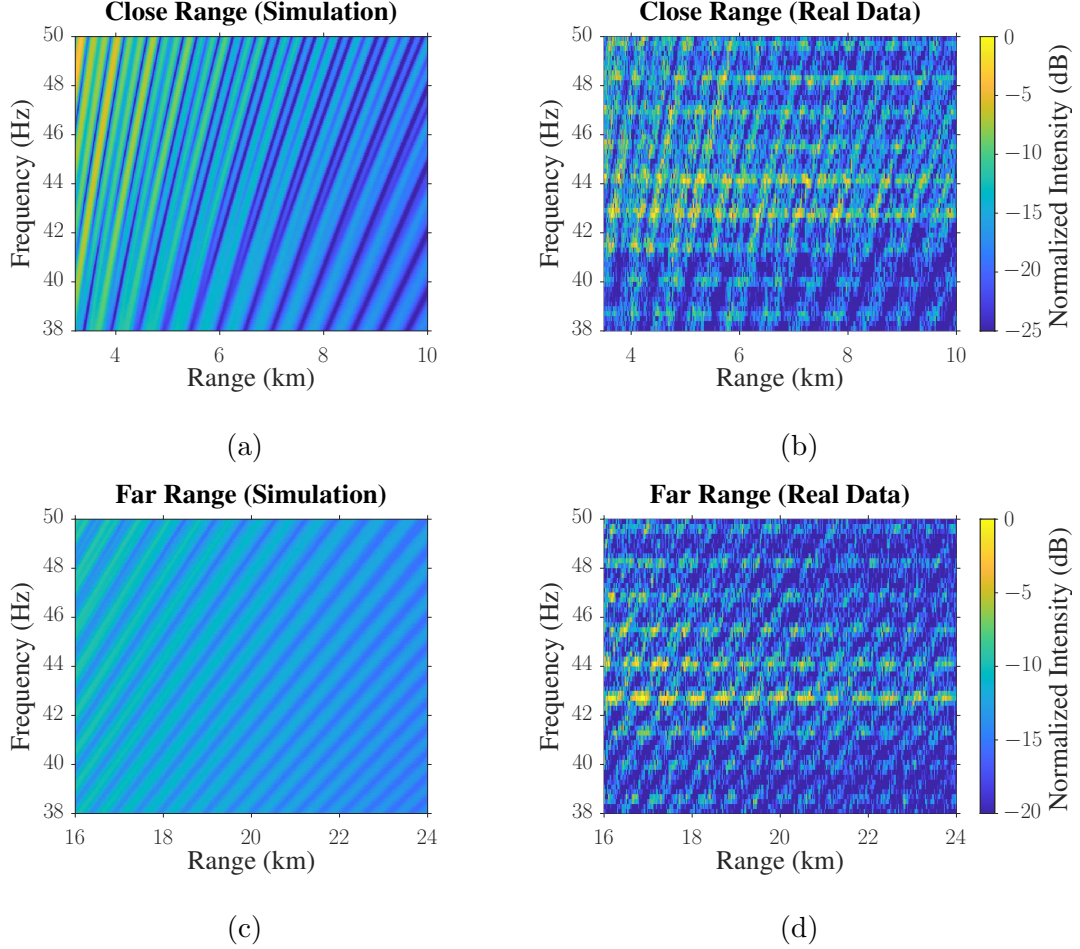


Figure 4.2. Comparison of simulated and measured spectrograms in the 38–50 Hz frequency band, shown as functions of source–receiver (ship–VLA2) range. (a) and (b) correspond to the close-range interval (3.5–10 km), while (c) and (d) correspond to the far-range interval (16–24 km). (a) and (c) are the simulated spectrograms based on the SC2 geoacoustic model using normal mode program KRAKEN. (b) and (d) are real data recorded by an element of VLA2 during the MSC KALAMATA ship passage on Mar 24th, 2017. At close range ((a) and (b)), three interfering modes produce complex striation patterns. At far range ((c) and (d)), the third modes attenuates, and the striations become dominated by interference between the first two modes. The simulation captures the qualitative transition in modal interference structure.

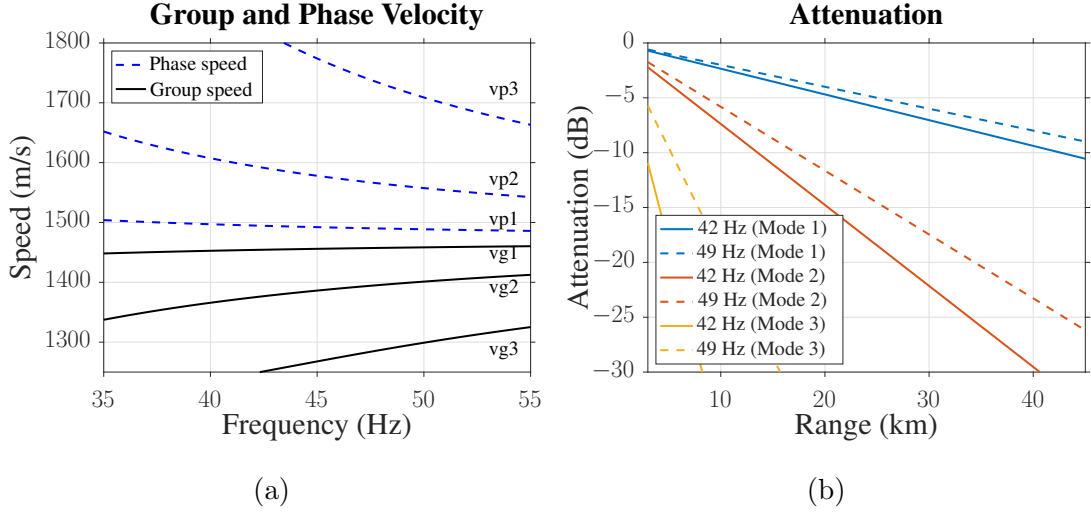


Figure 4.3. Modal properties computed from the SC2 geoacoustic model in the 35–55 Hz frequency range. (a) Phase speeds (v_p , dashed blue lines) and group speeds (v_g , solid black lines) for the first three modes. The subscript of the phase and group speed notations indicate the corresponding modes. The group velocity curves increase with frequency, while phase velocities decrease slightly, illustrating frequency-dependent dispersion. (b) Modal attenuation as a function of range for 42 Hz (solid lines) and 49 Hz (dashed lines). Higher-order modes attenuate more rapidly than lower-order modes, with mode 3 attenuating by more than 30 dB beyond 16 km. This behavior supports the dominance of the mode pair 1–2 interference at long ranges in the observed data.

Far-Range Comparison (16–24 km)

At longer ranges (Fig. 4.2 (c) and (d)), both measured and simulated spectrograms show that the third mode is significantly attenuated. The striation patterns are dominated by interference between the first two modes, resulting in simpler interference structures. This suggests that, beyond 16 km, the acoustic field can be effectively described by a single mode-pair interaction (mode 1 and 2), supporting the use of a scalar WI for range estimation.

This inference is further supported by modal attenuation estimates in Fig. 4.3 (b), which show that mode 3 suffers > 30 dB attenuation by 16 km at both 42 and 49 Hz, effectively removing it from the interference pattern. These attenuation values are

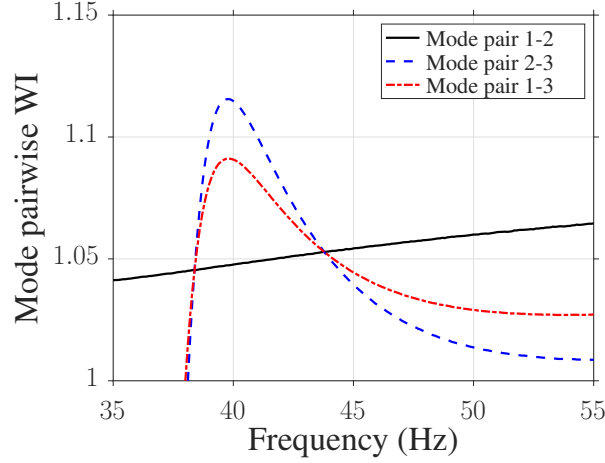


Figure 4.4. Simulated mode-pairwise WI values as a function of frequency for three mode pairs 1-2 (black solid), 2-3 (blue dashed), and 1-3 (red dash-dotted) using geoacoustic environments at site SC2. The mode pair 1-2 shows weak frequency dependence relative to the pairs 2-3 and 1-3.

computed from modal attenuation coefficients α_n in 1/m using KRAKEN:

$$\text{Attenuation}_n(r, f) = \exp(2\alpha_n(f)r), \quad (4.4)$$

where n refers to the n^{th} mode, r is the range and f is the frequency. The factor of 2 accounts for intensity being proportional to the square of pressure. Despite the potential environmental mismatch at farther ranges, the dominance of the first two modes justifies simplifying the WI as a single scalar value used in range estimation.

WI Variability Near Cut-Off Frequencies

Fig. 4.4 shows frequency-dependent WI values β_{mn} for various mode pairs. Near the cutoff frequency and the transition frequency band (38-45 Hz), the WI values of the mode pairs involving the third mode (1-3 and 2-3) become more variable than mode pair 1-2. This variability challenges the use of a single scalar WI to describe the full band at close ranges. The WI for mode pair 1-2 remains relatively stable across both sites and frequency bands, however, reinforcing its suitability for long-range range estimation.

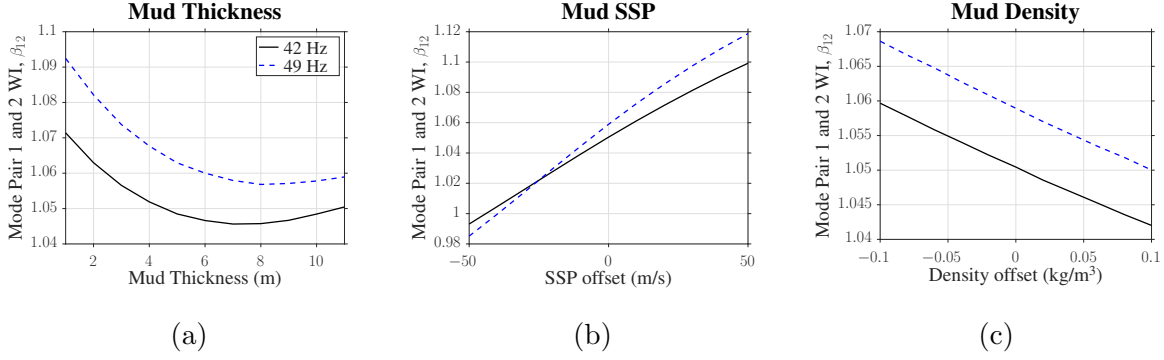


Figure 4.5. Sensitivity of the mode-pairwise WI β_{12} for mode pair 1-2 to changes in mud properties for the SC2 at 42 Hz (solid lines) and 49 Hz (dashed lines). The properties that are changed are (a) mud layer thickness (scaled relative to nominal values), (b) uniform shift in sound speed profile, (c) uniform shift in mud density. The sensitivity to sound speed is notably stronger than to either thickness or density variations.

Thus, while the assumption of a constant WI is technically violated near cutoff frequencies and at close ranges, it remains a useful approximation when higher modes are attenuated. Notably, despite this assumption, the range estimation method applied in earlier chapters at long ranges performed robustly, indicating that the dominant modal interference (from modes 1 and 2) governed the observed striations.

WI Sensitivity to Mud Properties

The acoustic environment in the study area exhibits spatially varying sediment characteristics. To evaluate how such variations impact the WI, we conduct simulations examining the sensitivity of the mode-pairwise WI value β_{12} —associated with the dominant mode pair 1-2 at long ranges—to perturbations in the mud layer’s properties. Specifically, we investigate the effects of (i) total mud thickness, (ii) sound speed profile (SSP), and (iii) density. These properties are varied individually while keeping the other two fixed at their nominal values specified in Tab. 4.1.

The mud is modeled as composed of two layers, and when varying thickness, all sublayers’ thicknesses within the mud layer are scaled proportionally. Sound speed and density perturbations are implemented by applying global additive shifts to the respective

values of the layer properties.

In addition to their influence on the mode-pairwise WI values, changes in mud properties could also affect mode cut-off behavior. However, although detailed results are omitted here, simulations show that at site SC2, there was no noticeable impact of the third mode's cutoff frequency on the interference pattern in the 42-49 Hz band.

Fig. 4.5 summarizes the simulation results. Among the three tested parameters, sound speed has the strongest influence on β_{12} . In SC2 environments, varying the sound speed profile by $\pm 50 \text{ m s}^{-1}$ leads to shifts in β_{12} of up to 0.10, indicating a moderate sensitivity to velocity perturbations. In contrast, changes in mud thickness and density produce relatively minor effects. The WI β_{12} remains within ± 0.01 of the baseline value, even for $\pm 50\%$ thickness changes and $\pm 0.1 \text{ kg/m}^3$ density shifts.

This analysis suggests that, for this location, accurate knowledge of the sediment sound speed is more critical for modeling mode-pair interference than precise estimates of mud thickness or density. However, it also confirms that β_{12} is relatively robust to small-to-moderate environmental uncertainties, which supports the viability of using WI-based ranging methods under imperfect geoacoustic knowledge.

4.4 Mode-pairwise WI Analysis

4.4.1 Method

To validate the WI estimation of the real data, we employ an additional approach based on relative dispersion curves (RDCs) [115]. RDCs provide a means to visualize the contributions of dominant mode pairs by analyzing how modal interference patterns evolve over range and frequency (Ch. 4.2.1). This approach is particularly useful in environments where only a few modes are actively propagating, as suggested by the normal mode simulation study discussed in Ch. 4.3. The below descriptions of the method for estimating the mode-pairwise WI closely follows the approach developed in [140].

Extracting Modal Contributions using RDCs

We first normalize the intensity data by multiplying each snapshot by its corresponding range to compensate for geometric spreading effects, as seen in Eq. (4.1). Then, To identify the modal contributions to the interference pattern, we perform a discrete Fourier transform (DFT) along the range axis at each frequency in the spectrogram (see Fig. 4.6). The resulting plot is referred to as the RDC surface plot. This transformation shifts the representation from range to the horizontal wavenumber k domain, where peaks correspond to dominant modal interactions. Since mode pairs exhibit interference at characteristic wavenumber differences (cf. Eq. (4.1)), the x-axis of the resulting RDC surface plot is labeled as the wavenumber difference Δk in Fig. 4.6. The summary of the operation is as follows:

$$\tilde{I}(\Delta k, f) = \text{DFT}_r[r \cdot I(r, f)], \quad (4.5)$$

where $\text{DFT}_r[\cdot]$ denotes the DFT operation along the range-axis and $I(r, f)$ represents the received acoustic intensity function at the corresponding range-frequency coordinates. Each energy peak in the Δk domain corresponds to the contribution of a specific mode pair interference, and the RDC represents the trend of these energy peaks across frequency.

Estimating Mode-Pairwise WI (β_{mn})

To compute β_{mn} in Eq. (4.3) per mode pair m - n , we inspect the RDCs in the logarithmic domain of the variables Δk and f , where the slope of each RDC satisfy the relationship [115]

$$\log(\Delta k_{mn}(f)) = \frac{-1}{\beta_{mn}} \cdot \log(f) + \log(\gamma_{mn}), \quad (4.6)$$

where γ_{mn} is a mode-dependent constant. This equation reveals that the slope of the RDC in the log-log domain is directly related to the WI of the mode pair β_{mn} .

To estimate β_{mn} , we apply the Radon transform [146] to the RDCs in the log-log domain of Δk and f . The Radon transform integrates function values along straight lines at different angles, effectively identifying the dominant slope corresponding to each mode pair's WI. In this context, the angle of integration lines relative to the coordinate axes corresponds to the slope of the RDCs. The line that maximizes the Radon transform output is used to determine the WI. The best-aligned line with the observed RDC yields the highest integration response, and its angle is used to compute the slope (i.e., $-1/\beta_{mn}$) in Eq. (4.6)). This process ensures that the dominant WI value per mode pair is extracted systematically.

Fig. 4.7 illustrate the processing of the r - f surface plot to obtain the mode-pairwise WI with real acoustic data. The example uses a spectrogram from 18:32-19:07 when the ship was near the CPA and thus three propagating modes were observed. The RDC surface plot and its linearization in the log-log domain are shown in Fig. 4.7 (a) and (b), respectively. The Radon transform of the linearized RDC surface plot and the Radon transform corresponding to each mode pair to estimate the mode-pairwise WI are shown in Fig. 4.7 (c) and (d), respectively.

This approach offers a detailed characterization of the WI compared to assuming a single constant value. Specifically, it reveals how individual mode pairs shape the observed interference structure and contribute to WI variability, particularly in environments and frequency bands where only a few modes propagate.

4.4.2 Spectrogram Processing and WI Estimation Implementation

For WI analysis, spectrograms were computed from the acoustic recordings using different temporal windows depending on the application. To analyze modal evolution,

three 15-min long spectrogram segments, during the transit of MSC KALAMATA, were selected: 18:32–18:47 (3.2–10.5 km), 18:55–19:10 (15.3–24.4 km), and 19:30–19:45 (36.7–45.8 km) on Mar 24th, 2017 (UTC). The first segment corresponds to the period immediately following CPA and is expected to show the least modal attenuation. The second aligns with the beginning of the range estimation interval used in Ch. 3, and the third captures the end of the event, when the ship was farthest from VLA2. To achieve higher temporal (and hence spatial) resolution, each spectrogram was computed using a 5-second Hamming window with 50 % overlap. These were then transformed into rf surface plot using AIS-derived source positions, enabling direct visualization of striation patterns over range and space.

For mode-pairwise WI estimation, two longer spectrograms were used. The first spans 35 minutes from 18:32 to 19:07, a time window selected to capture the full presence of the third acoustic mode before its attenuation. The second spans 18:58 to 19:45, during which the third mode was no longer reliably observable. Both spectrograms assume stationary modal properties during their respective intervals and were computed using 20-second snapshot (50 % overlap and Hamming windowed) to balance temporal and spectral resolutions.

The lengths of these spectrograms (35 min and 47 min) were selected to reduce uncertainty in estimating the slopes of each RDC due to discretization effects in the log-wavenumber domain. Specifically, discretization imposes a lower bound on slope accuracy, stemming from the finite pixel resolution. We assess the maximum slope deviation by considering a worst-case scenario in which the log-wavenumber values at the lowest and highest frequencies are each offset by half a pixel. The observed noise in the acoustic data and potential range-dependence of modal properties are expected to introduce additional variability, compounding the total slope estimation error beyond this discretization limit.

A key implementation detail is the use of frequency-wise normalization in the r - f surface plots. By ensuring that each frequency band has equal total energy, this normaliza-

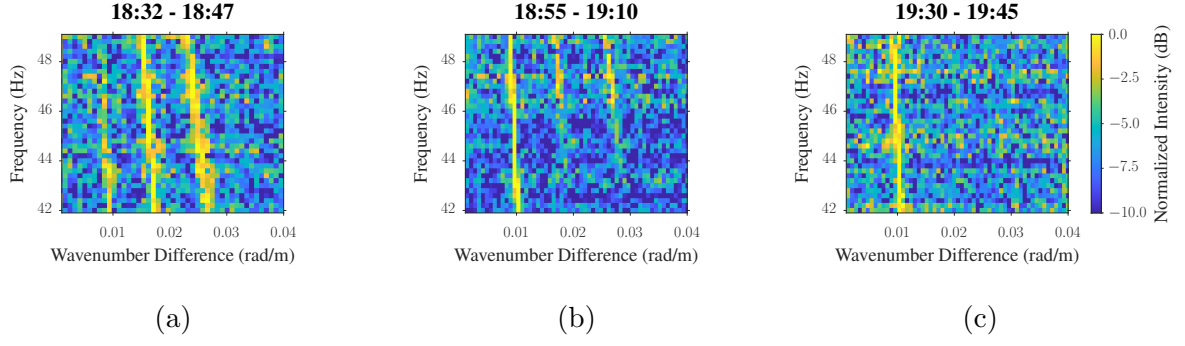


Figure 4.6. Relative dispersion curve (RDC) surface plots derived from acoustic recordings during the ship’s transit relative to VLA 2 at three different times, showing prominent vertical lines corresponding to the RDCs of mode pairs 1-2, 2-3, and 1-3 (from left to right within the RDC surface plot). The processed are spectrograms for the range intervals (a) 3.2–10.5 km, (b) 14.6–24.4 km, and (c) 32.3–45.6 km and for the time between (a) 18:32–18:47, (b) 18:55–19:10 and (c) 19:30–19:45. (a) correspond to the RDCs near the CPA, and while three RDC curves are observed, mode pairs 2-3 and 1-3 are more pronounced than mode pair 1-2. (b) is the time at which the range estimation is first performed in Ch. 3, and mode pair 1-2 is dominant over mode pairs 2-3 and 1-3. (c) At the end of the considered event, the mode pairs 2-3 and 1-3 have completely faded and there is only mode pair 1-2.

tion mitigates bias introduced by persistent tonal components that could otherwise distort the interference striations and, consequently, the slope estimation used for WI analysis.

4.4.3 Results

RDC Analysis

The RDCs extracted from the acoustic data over time provide insight into the relative modal contributions during the ship’s transit. Fig. 4.6 (a) shows the presence of three distinct vertical high intensity curves, each corresponds to a mode pair RDC, during the time window near the CPA. The ordering of the mode pairs as 1-2, 2-3, and 1-3, based on increasing Δk , follows the order found in simulations.

As the ship moves farther from VLA2, the third mode (associated with mode pairs 1-3 and 2-3) exhibits decreasing energy. In Fig. 4.6 (b), corresponding to mid-range propagation (14.6–24.4 km), mode pair 1-2 dominates, while the others fade. This trend

continues in Fig. 4.6 (c), where only mode pair 1-2 remains observable beyond 32 km. Interestingly, the dominant mode pair shifts over time. In Figure 4.6(a), mode pair 1-2 is relatively weak, while mode pairs 2-3 and 1-3 dominate. However, the contribution from the third mode diminishes over time, leading to increased prominence of mode pair 1-2 in the later part of the event. This temporal variability complicates range estimation near CPA, where multiple mode pairs interfere.

These observations confirm that only a small number of modes are excited and propagate effectively in the 42–49 Hz frequency band during the event. Notably, modal attenuation limits the contribution of the third mode at longer ranges. Moreover, as we observed in the simulations (Fig. 4.4), there is only a relatively mild change in the mode-pairwise WI value (β_{12}) for the mode pair 1-2 as a function of frequency. The interference patterns used for range estimation in Ch. 3 are thus primarily governed by the mode pair 1–2, which justifies the use of a single WI value in that analysis. At closer ranges, higher-mode contributions are present, but their limited persistence at far ranges allows a scalar WI to sufficiently capture the dominant striation behavior.

Estimated WI

We estimate the WI values for each mode pair using data from two intervals: 18:32–19:07 (when all three modes are present) and 18:58–19:45 (when the third mode has largely attenuated). Slope-based estimation in the linearized RDC surface plot of the data from 18:32–19:07 in the log-log domain yields WI values of approximately 1.09, 1.31, and 1.18 for mode pairs 1-2, 2-3, and 1-3, respectively.

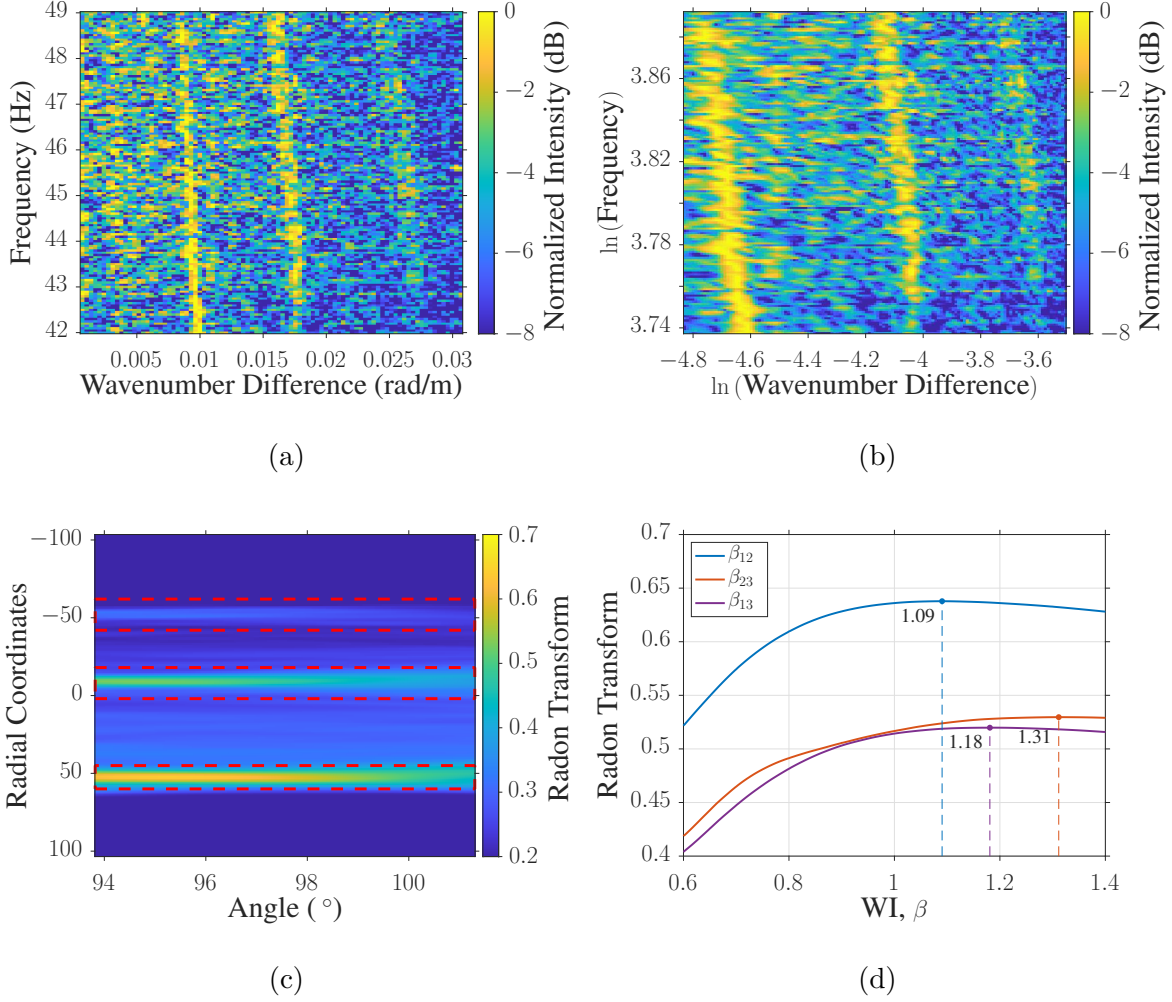


Figure 4.7. (a) Normalized relative dispersion curve (RDC) surface plot computed from the spectrogram (18:32–19:07, March 24th, 2017) during the transit of MSC KALAMATA. The three vertical bands of high intensity (the third being relatively faint) correspond to the interference patterns of three dominant mode pairs. (b) The same RDC surface plot, transformed into the log–log domain of wavenumber difference and frequency, where the slope of each feature is inversely related to the mode-pairwise WI. (c) Radon transform applied to (b), with angle on the x-axis corresponding to slope. The red dashed boxes denote horizontal strips centered at each mode pair. (d) Radon transform response as a function of hypothesized WI values, computed from each box in (c). Peak locations indicate the estimated WI for each mode pair: $\beta_{12} = 1.09$, $\beta_{23} = 1.31$, and $\beta_{13} = 1.18$ (from top to bottom in (c), respectively).

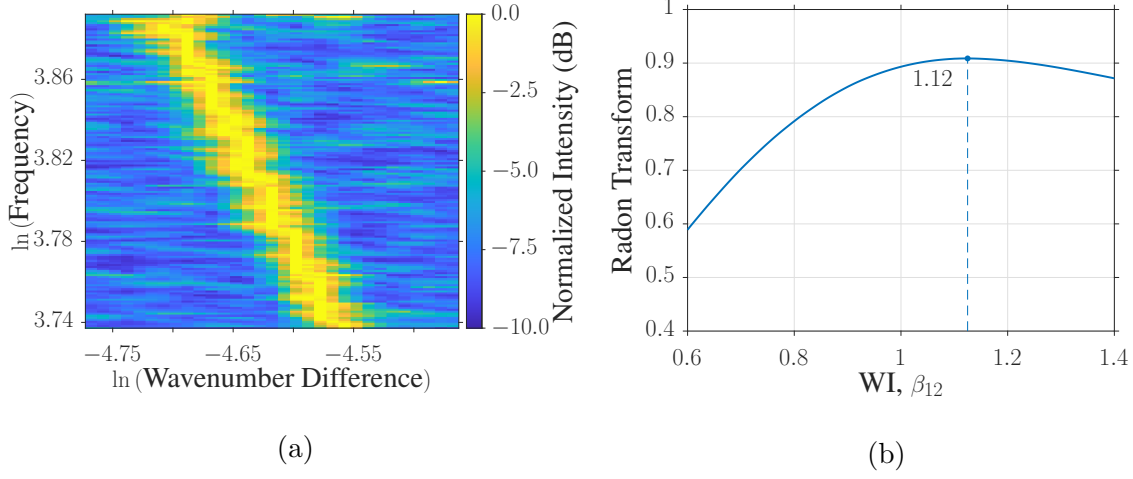


Figure 4.8. (a) The RDC surface plot depicting the slope of the RDC corresponding to the mode pair 1-2 in the log-log domain of wavenumber difference and frequency. It is obtained from acoustic recordings between 18:58 and 19:45. The slope is directly related to the WI of the mode pair 1-2. (b) Radon transform output as a function of hypothetical mode-pairwise WI values (β_{12}), showing the maximum response at $\beta = 1.12$, indicating the dominant mode pair's WI value.

To ensure consistency with the range estimation analysis presented in Ch. 3, we estimate a single scalar WI value from the spectrogram spanning 18:58 to 19:45, during which mode pair 1-2 remains the dominant contributor. As previously shown in Fig. 4.6 (b), the third mode gradually fades starting around 18:55, making this interval appropriate for extracting a stable mode-pairwise WI associated with the mode pair 1–2 interference. Processing the normalized Radon transform, we obtain a peak at $\beta = 1.12$, as illustrated in Fig. 4.8 (b), corresponding to the best-fit slope of the RDCs in the log–log domain. The underlying RDC used in this estimation is shown in Fig. 4.8 (a), where only the striation corresponding to mode pair 1-2 is extracted.

In contrast, the WI value of 1.18 was used for range estimation in Ch. 3, estimated from a shorter spectrogram (18:58–19:10) with the statistical approach. For the value of 1.12 obtained via mode-pairwise framework using the 47-min long spectrogram, the uncertainty in β_{12} arising from discretization in the $\log(\Delta k)$ axis is approximately ± 0.15 .

Therefore, the estimated value of 1.18 falls within the uncertainty range. Furthermore, this uncertainty reflects only the limits imposed by transform resolution; additional sources of error such as ambient noise, imperfect mode separation, and temporal variability are not accounted for in this bound.

In addition, the earlier 35-min long spectrogram (18:32–19:07) yielded $\beta_{12} = 1.03$ for the mode pair 1-2, which also falls within the expected error range. However, the discrepancy between this earlier value (1.03) and the later estimate (1.12) may also reflect changes in the environment, such as mild bathymetric variations or differences in mud layer properties.

Overall, our results confirm that modal interference in the 42–49 Hz band is primarily governed by the 1–2 mode pair at longer distances, justifying the use of a single scalar WI for passive ranging applications in this environment. It is also validated that the WI value can be higher than unity in this scenario. However, closer to CPA, where higher-mode contributions are non-negligible, a full mode-pairwise analysis may be necessary to account for temporal and spatial variability in striation structure. As a result, the WI-based ranging may have to consider modelling the WI β as a distribution instead of a single scalar value.

Limitations and Sources of Uncertainty

Several factors may have contributed to the observed discrepancies in WI estimation, particularly during the later stages of the transit. The resolution of the wavenumber-axis was not high enough, as the RDCs spanned at most three wavenumber pixels. This affects the precision of slope estimates, particularly for weaker mode pairs.

Both the WI estimation method proposed in Ch. 3 and the pairwise mode analysis involve interpolation steps that may introduce bias [115]. The proposed method in Ch. 3 required interpolating intensities when transforming the data into striation-frequency domain, while the pairwise mode analysis involved interpolating $I(\Delta k, f)$ into the log-log

domain, where data smearing occurred. Despite these limitations, the pairwise mode analysis confirms that modal interference in this environment was primarily governed by three mode pairs up to 19:10, after which only one dominant mode pair remained.

To verify that the mode-pairwise WI estimation method and its implementation are valid, the simulated data based on the geoacoustic properties of site SC2 (3.5 to 10.0 km) are also processed. The mode-based WI values were computed as 1.03, 1.06, and 1.02 for β_{12} , β_{23} and β_{13} respectively. The estimated values are close to the theoretically calculated values using Eq. (4.3) as shown in Fig. 4.4 (a), although there are some visible variabilities. [140] acknowledged that this method of computing the mode-pairwise WI based on the slopes of the linearized RDCs would yield a mean of the values across all frequencies, so the observed results are agreeable in that sense.

4.5 Conclusions

This chapter presented a detailed analysis of the modal interference patterns observed in the low-frequency band (42–49 Hz) during a commercial ship passage in the New England Mud Patch, a shallow-water environment characterized by layered sediment structures. Numerical simulations revealed that three acoustic modes propagate near the ship’s CPA, motivating a mode-pairwise framework to interpret the observed interference patterns.

By combining normal-mode simulations with real acoustic data, it was shown that long-range interference in this frequency band is primarily governed by the 1–2 mode pair. From processing the real data, the WI associated with this pair was estimated as $\beta_{12} = 1.12$ using the mode-pairwise framework, and as $\beta = 1.18$ using the statistical method introduced in Ch. 3. While these values differ slightly, the results consistently indicate that when only a few modes are present, the effective WI can exceed the canonical shallow-water value of unity.

The simulation study also highlighted important considerations for WI-based ranging at low frequencies. Sensitivity analysis showed that geoacoustic parameters, particularly the sediment sound speed, can affect WI values. Additionally, variability in the WI near the CPA, where multiple mode pairs contribute, suggests that modeling WI as a distribution rather than a fixed scalar may improve estimation accuracy in complex environments.

Despite these challenges, the analysis demonstrated that at far ranges (beyond 16 km), only the lowest two modes persist, and their interference can be effectively described using a single WI value. This supports the validity of using a scalar WI for long-range passive acoustic ranging, as applied in Ch.3.

In practical terms, the canonical assumption of $\beta \approx 1$ remains suitable for high-frequency applications and relatively short ranges [100, 113]. However, ranging at longer distances using low frequencies excited by commercial ships requires attention to modal content and environmental structure. When conditions support propagation of only a single dominant mode pair, as in the far-range segment of this study, WI-based ranging can still be performed reliably using a simplified scalar model—even at low frequencies.

Acknowledgments:

Chapter 4 is being prepared for submission as a JASA Express Letters manuscript: J. Jang and F. Meyer, "Mode-pairwise Waveguide Invariant Analysis in the New England Mud Patch." The dissertation author was the primary investigator and lead author of this work.

Appendix A

Theoretical Justification and Derivation of the Statistical Parameter Estimators

This appendix provides a detailed probabilistic justification and derivation of the estimators required to compute the proposed likelihood function in Ch. 3.4. The signal models are considered in the striation-frequency domain representation of the acoustic intensities. Striations and frequencies are indexed by $l \in \mathcal{L}$ and $k \in \mathcal{K}$, respectively. The total number of frequency bins is $K = |\mathcal{K}|$. The subset of frequencies containing only broadband components of ship-radiated noise is denoted by $\mathcal{K}^b$, while the subset containing both tonal and broadband components denote by $\mathcal{K}^{bt}$. There are J bins in $\mathcal{K}^{bt}$.

For $k \in \mathcal{K}^b$, the intensity follows an exponential distribution, i.e.,

$$\underline{x}_{l,k}^b \sim \text{Exp}(\underline{x}_{l,k}^b; \underline{\theta}_{l,k} = \underline{\sigma}_{l,k}^2)$$

where $\underline{\theta}_{l,k}$ is the scaling factor. In the main manuscript, the scaling factor was shown to have the same value as the broadband-plus-noise variance $\underline{\sigma}_{l,k}^2$. For $k \in \mathcal{K}^{bt}$, the normalized intensity follows a noncentral- χ^2 distribution, i.e.,

$$\underline{y}_{l,k}^{bt} \sim \chi^2(\underline{y}_{l,k}^{bt}; d = 2, \underline{\lambda}_{l,k})$$

where $d = 2$ is the degrees of freedom, and $\underline{\lambda}_{l,k}$ is the noncentrality parameter. We assume that intensities are statistically independent across striations and frequencies. For additional notation and signal model details, see Ch. 3.3 of the main manuscript.

To obtain the maximum likelihood estimate (ML) of the ship's range, we estimate the scaling factor $\underline{\theta}_{l,k}$ and the noncentrality parameter $\underline{\lambda}_{l,k}$ using Kolmogorov's Strong Law of Large Numbers (SLLN).

A.1 Background

To estimate the statistical parameters governing the probability density functions (PDFs) of the intensities and normalized intensities, we employ the Strong Law of Large Numbers (SLLN) and Kolmogorov's criterion [147]. Specifically, Kolmogorov's SLLN states that if $\{X_n\}$ is a sequence of mutually independent random variables with common distribution and finite mean $\mathbb{E}[X_n] = \mu$, then the sample mean converges almost surely to μ , i.e.,

$$\frac{1}{N} \sum_{n=1}^N X_n \xrightarrow{\text{a.s.}} \mu \text{ as } n \rightarrow \infty. \quad (\text{A.1})$$

Kolmogorov's criterion extends this result to sequences of mutually independent but not necessarily identically distributed random variables. If X_n has finite mean $\mathbb{E}[X_n] = \mu_n$ and variance $\sigma_n^2 = \text{Var}(X_n)$, and if the sequence satisfies

$$\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty$$

then the SLLN holds in the extended form, i.e.,

$$\frac{1}{N} \sum_{n=1}^N X_n \xrightarrow{\text{a.s.}} \frac{1}{N} \sum_{n=1}^N \mu_n \text{ as } N \rightarrow \infty. \quad (\text{A.2})$$

In the application considered in this paper, we can assume that the variance of the intensities, $\underline{x}_{l,k}^b$, is bounded across both striations and frequencies. This assumption is

supported by (i) the behavior of the channel transfer function, whose magnitude decays with range due to modal attenuation and spreading losses in a shallow water waveguide [107], and (ii) the finite energy of the ship's radiated noise and the presence of a constant (finite) background noise variance.

A.2 Broadband-plus-noise Variance Estimation

To compute the normalized intensities, we estimate the broadband-plus-noise variance $\underline{\sigma}_{l,k}^2$. As shown in the main manuscript, the broadband-plus-noise variance $\underline{\sigma}_{l,k}^2$ is equal to the scaling factor $\underline{\theta}_{l,k}$. Based on Eq. (3.19) in the main manuscript, we can write

$$\underline{\theta}_{l,k} = (\sigma_k^b |\underline{g}_{l,k}|)^2 + (\sigma_k^u)^2.$$

Note that the background noise $(\sigma_k^u)^2$ is assumed known. For a given frequency index $k \in \mathcal{K}^b$, we defined $\eta_k = \sigma_k^b / \sigma_{k'}^b$ ($\eta_k > 0$) as the ratio of broadband signal standard deviations at f_k and a reference frequency $f_{k'}$, where $k' \in \mathcal{K}^b$ is arbitrarily chosen. Based on the frequency-dependent Green's function relation $|\underline{g}_{l,k}| = \gamma_k |\underline{g}_{l,k'}|$ as presented in Eq. (3.11) of the main manuscript, we obtain

$$\underline{\theta}_{l,k} = (\eta_k \gamma_k)^2 (\sigma_{k'}^b |\underline{g}_{l,k'}|)^2 + (\sigma_k^u)^2. \quad (\text{A.3})$$

We define the received broadband scale parameter $\alpha_k = (\eta_k \gamma_k)^2$ and the received broadband signal variance $\underline{v}_{l,k} = (\sigma_k^b)^2 |\underline{g}_{l,k}|^2$. By using these definitions of α_k and $\underline{v}_{l,k}$ in Eq. (A.3), as well as flipping the terms on the right-hand side, we obtain

$$\underline{\theta}_{l,k} = (\sigma_k^u)^2 + \alpha_k \underline{v}_{l,k'}. \quad (\text{A.4})$$

Next, we compute an estimate of $\underline{\theta}_{l,k}$, by separately computing estimates of α_k and $\underline{v}_{l,k}$.

A.2.1 Broadband Scale Parameter α_k

Since $\mathbb{E}[\underline{x}_{l,k}^b] = \underline{\theta}_{l,k}$, applying the SLLN with Kolmogorov's criterion in Eq. (A.2), the sample mean satisfies

$$\frac{1}{L} \sum_{l \in \mathcal{L}} \underline{x}_{l,k}^b \xrightarrow{\text{a.s.}} \frac{1}{L} \sum_{l \in \mathcal{L}} \underline{\theta}_{l,k}. \quad (\text{A.5})$$

Let us average the scaling factor $\underline{\theta}_{l,k}$ across striations (cf. Eq. (3.21)), i.e.,

$$\frac{1}{L} \sum_{l \in \mathcal{L}} \underline{\theta}_{l,k} = (\sigma_k^u)^2 + \alpha_k \frac{1}{L} \sum_{l \in \mathcal{L}} v_{l,k'}. \quad (\text{A.6})$$

For a sufficiently large number of striations, we can make use of Eq. (A.5) such that, after rearranging Eq. (A.6) for α_k , we obtain

$$\alpha_k = \frac{\frac{1}{L} \sum_{l \in \mathcal{L}} \underline{x}_{l,k}^b - (\sigma_k^u)^2}{\frac{1}{L} \sum_{l \in \mathcal{L}} v_{l,k'}}.$$

Since $\alpha_{k'} = 1$ at the reference frequency $f_{k'}$, for the frequency index $k = k'$, the average in Eq. (A.6) reads

$$\frac{1}{L} \sum_{l \in \mathcal{L}} v_{l,k'} = \frac{1}{L} \sum_{l \in \mathcal{L}} \underline{x}_{l,k'}^b - (\sigma_{k'}^u)^2.$$

By substituting the denominator of Eq. (A.7) with the relationship in Eq. (A.7), an estimate of the broadband scale parameter is obtained as

$$\hat{\alpha}_k = \frac{\frac{1}{L} \sum_{l \in \mathcal{L}} \underline{x}_{l,k}^b - (\sigma_k^u)^2}{\frac{1}{L} \sum_{l \in \mathcal{L}} \underline{x}_{l,k'}^b - (\sigma_{k'}^u)^2}. \quad (\text{A.7})$$

A.2.2 Received Broadband Signal Variance $v_{l,k}$

The broadband signal variance is estimated using a similar approach. Assuming independence across frequency bins and bounded variance, the sample mean at striation

index $l \in \mathcal{L}$ converges, i.e.,

$$\frac{1}{K-J} \sum_{k \in \mathcal{K}^b} \underline{x}_{l,k}^b \xrightarrow{\text{a.s.}} \frac{1}{K-J} \sum_{k \in \mathcal{K}^b} \underline{\theta}_{l,k}. \quad (\text{A.8})$$

Next, we average the scaling factor $\underline{\theta}_{l,k}$ across frequencies (cf. Eq. (3.21)), i.e.,

$$\frac{1}{K-J} \sum_{k \in \mathcal{K}^b} \underline{\theta}_{l,k} = (\sigma_k^u)^2 + \underline{v}_{l,k'} \frac{1}{K-J} \sum_{k \in \mathcal{K}^b} \alpha_k. \quad (\text{A.9})$$

For a sufficiently large number of measurements across frequencies, we can use Eq. (A.8) in Eq. (A.9), i.e.,

$$\frac{1}{K-J} \sum_{k \in \mathcal{K}^b} \underline{x}_{l,k}^b = (\sigma_k^u)^2 + \underline{v}_{l,k'} \frac{1}{K-J} \sum_{k \in \mathcal{K}^b} \alpha_k.$$

After solving this expression for $\underline{v}_{l,k'}$, we obtain

$$\underline{v}_{l,k'} = \frac{1}{\sum_{k \in \mathcal{K}^b} \alpha_k} \left(\sum_{k \in \mathcal{K}^b} \underline{x}_{l,k}^b - (\sigma_k^u)^2 \right). \quad (\text{A.10})$$

By substituting the broadband scale parameter α_k in Eq. (A.10) with its estimate $\hat{\alpha}_k$ in Eq. (A.7), an estimate of the received broadband-plus-noise variance is finally calculated as

$$\hat{\underline{v}}_{l,k'} = \frac{1}{\sum_{k \in \mathcal{K}^b} \hat{\alpha}_k} \left(\sum_{k \in \mathcal{K}^b} \underline{x}_{l,k}^b - (\sigma_k^u)^2 \right). \quad (\text{A.11})$$

A.2.3 Broadband-plus-noise Variance $\underline{\sigma}_{l,k}^2$

Based on Eqs. (3.21), (A.7) and (A.11), an estimate of $\underline{\theta}_{l,k}$ is obtained as

$$\hat{\underline{\theta}}_{l,k} = \hat{\alpha}_k \hat{\underline{v}}_{l,k'} + (\sigma_k^u)^2.$$

Finally, the estimate of the broadband-plus-noise variance $\underline{\sigma}_{l,k}^2$, is given by $\hat{\underline{\sigma}}_{l,k}^2 = \hat{\underline{\theta}}_{l,k}$, with $\hat{\underline{\theta}}_{l,k}$ as presented in Eq. (3.18) of the main manuscript. In our practical implementation,

we set $\widehat{\underline{\sigma}}_{l,k}^2 = 0$ in case $\widehat{\underline{\sigma}}_{l,k}^2$ is negative. We also apply a moving average filter to $\underline{\sigma}_{l,k}^2$ across striations at each frequency. This helps to reduce the variance of the estimated parameters, leading to an improved overall performance.

A.3 Noncentrality Parameter Estimation

We estimate the noncentrality parameter $\underline{\lambda}_{l,k}$ for the normalized intensity $\underline{y}_{l,k}^{\text{bt}}$ at striation $l \in \mathcal{L}$ and frequency index $k \in \mathcal{K}^{\text{bt}}$. The expectation of $\underline{y}_{l,k}^{\text{bt}}$ is denoted by $\underline{\mu}_{l,k}^{\text{bt}} = \mathbb{E}[\underline{y}_{l,k}^{\text{bt}}]$ and expressed as

$$\underline{\mu}_{l,k}^{\text{bt}} = d + \underline{\lambda}_{l,k} \quad (\text{A.12})$$

where $d = 2$ is the degrees of freedom [130]. Following the model in Ch. 3.4.3, we assume that $\underline{\lambda}_{l,k}$ can be expressed as a product of a striation-dependent parameter and a frequency-dependent scaling factor,

$$\underline{\lambda}_{l,k} = \underline{\lambda}_l \rho_k \quad (\text{A.13})$$

where $\rho_k > 0$ is a frequency-dependent scale factor, and $\underline{\lambda}_l$ is the striation-wise noncentrality parameter.

For clarity in the derivation below, we use i and j instead of l and k when summing over striations and frequencies, respectively. This helps distinguish the indices being summed over from the particular index of interest, denoted as (l, k) .

Assuming that the normalized intensities satisfy Kolmogorov's criterion, the sample mean across striations at fixed frequency index k converges to the corresponding expected mean, i.e.,

$$\frac{1}{L} \sum_{i \in \mathcal{L}} \underline{y}_{i,k}^{\text{bt}} \xrightarrow{\text{a.s.}} \frac{1}{L} \sum_{i \in \mathcal{L}} \underline{\mu}_{i,k}^{\text{bt}}. \quad (\text{A.14})$$

Similarly, at a fixed striation index l , the sample mean across frequencies converges to the

expected average, i.e.,

$$\frac{1}{J} \sum_{j \in \mathcal{K}^{\text{bt}}} y_{l,j}^{\text{bt}} \xrightarrow{\text{a.s.}} \frac{1}{J} \sum_{j \in \mathcal{K}^{\text{bt}}} \mu_{l,j}^{\text{bt}}. \quad (\text{A.15})$$

Furthermore, by using Eq. (A.13) in Eq. (A.12), and by averaging $\underline{\mu}_{l,k}^{\text{bt}}$ over striations at fixed frequency k leads to

$$\frac{1}{L} \sum_{i \in \mathcal{L}} \underline{\mu}_{i,k}^{\text{bt}} = 2 + \rho_k \frac{1}{L} \sum_{i \in \mathcal{L}} \underline{\lambda}_i \quad (\text{A.16})$$

while averaging $\underline{\mu}^{\text{bt}}$ over frequencies at fixed striation l results in

$$\frac{1}{J} \sum_{j \in \mathcal{K}^{\text{bt}}} \underline{\mu}_{l,j}^{\text{bt}} = 2 + \underline{\lambda}_l \frac{1}{J} \sum_{j \in \mathcal{K}^{\text{bt}}} \rho_j. \quad (\text{A.17})$$

For a sufficiently large number of measurements across striations and across frequencies, we can apply Eqs. (A.14) and (A.15) to Eqs. (A.16) and (A.17), respectively. In this way, we obtain

$$\frac{1}{L} \sum_{i \in \mathcal{L}} y_{i,k}^{\text{bt}} = 2 + \rho_k \frac{1}{L} \sum_{i \in \mathcal{L}} \underline{\lambda}_i$$

and

$$\frac{1}{J} \sum_{j \in \mathcal{K}^{\text{bt}}} y_{l,j}^{\text{bt}} = 2 + \underline{\lambda}_l \frac{1}{J} \sum_{j \in \mathcal{K}^{\text{bt}}} \rho_j.$$

Similarly, the full average of the normalized intensities across all striations and frequencies is given by

$$\begin{aligned} \frac{1}{JL} \sum_{j \in \mathcal{K}^{\text{bt}}} \sum_{i \in \mathcal{L}} y_{i,j}^{\text{bt}} &= d + \frac{1}{JL} \sum_{j \in \mathcal{K}^{\text{bt}}} \sum_{i \in \mathcal{L}} \underline{\lambda}_i \rho_j \\ &= 2 + \frac{1}{JL} \sum_{i \in \mathcal{L}} \underline{\lambda}_i \sum_{j \in \mathcal{K}^{\text{bt}}} \rho_j. \end{aligned} \quad (\text{A.18})$$

By rearranging Eqs. (A.16) and (A.17), we obtain

$$\rho_k = \frac{\frac{1}{L} \sum_{i \in \mathcal{L}} \underline{y}_{i,k}^{\text{bt}} - 2}{\frac{1}{L} \sum_{i \in \mathcal{L}} \underline{\lambda}_i} \quad (\text{A.19})$$

$$\underline{\lambda}_l = \frac{\frac{1}{J} \sum_{j \in \mathcal{K}^{\text{bt}}} \underline{y}_{l,j}^{\text{bt}} - 2}{\frac{1}{J} \sum_{j \in \mathcal{K}^{\text{bt}}} \rho_j}. \quad (\text{A.20})$$

Substituting Eqs. (A.19) and (A.20) into Eq. (A.18) yields

$$\lambda_{l,k} = \frac{\left(\frac{1}{J} \left(\sum_{j \in \mathcal{K}^{\text{bt}}} \underline{y}_{l,j}^{\text{bt}}\right) - 2\right) \left(\frac{1}{L} \left(\sum_{i \in \mathcal{L}} \underline{y}_{i,k}^{\text{bt}}\right) - 2\right)}{\frac{1}{JL} \left(\sum_{i \in \mathcal{L}} \lambda_i\right) \left(\sum_{j \in \mathcal{K}^{\text{bt}}} \rho_j\right)}. \quad (\text{A.21})$$

Note that the denominator of the above equation is still a function of unknown variables ρ_j and λ_i . To address this, we first rearrange Eq. (A.18) as follows

$$\frac{1}{JL} \sum_{i \in \mathcal{L}} \lambda_i \sum_{j \in \mathcal{K}^{\text{bt}}} \rho_j = \frac{1}{JL} \sum_{j \in \mathcal{K}^{\text{bt}}} \sum_{i \in \mathcal{L}} \underline{y}_{i,j}^{\text{bt}} - 2. \quad (\text{A.22})$$

Next, we replace the denominator of Eq. (A.21), by the right hand side of Eq. (A.22). In this way, we obtain the following estimate of the noncentrality parameter

$$\hat{\lambda}_{l,k} = \frac{\left(\frac{1}{J} \left(\sum_{j \in \mathcal{K}^{\text{bt}}} \underline{y}_{l,j}^{\text{bt}}\right) - 2\right) \left(\frac{1}{L} \left(\sum_{i \in \mathcal{L}} \underline{y}_{i,k}^{\text{bt}}\right) - 2\right)}{\frac{1}{JL} \sum_{j \in \mathcal{K}^{\text{bt}}} \sum_{i \in \mathcal{L}} \underline{y}_{i,j}^{\text{bt}} - 2}$$

also is presented in Eq. (3.27) in Ch. 3.4.2.

At low SNR, the subtraction by two may be negative, and the denominator may be numerically unstable. Thus, when its value is negative or unrealistically large, the noncentrality parameter is set to zero, i.e., $\hat{\lambda}_{l,k} = 0$. In addition, a moving average is applied to the noncentrality parameter along the striation axis at each frequency.

Acknowledgments:

Chapters 3 and Appendix A are being prepared for submission as a Journal of the Acoustical Society of America manuscript: J. Jang, W. S. Hodgkiss, and F. Meyer, “Ranging of a Moving Ship Using a Single Acoustic Receiver in Shallow Water.” The dissertation author was the primary investigator and lead author of this work.

Appendix B

Correlation-based Ranging of a Ship in Shallow Water Using a Single Acoustic Receiver

This appendix presents the mathematical formulation of the correlation-based method (C) for estimating the range of a moving ship using a single acoustic receiver in shallow water discussed in Ch. 3.6.2. Unlike statistical approaches that rely on likelihood functions, this method uses deterministic correlation between pairs of tonal frequency components to estimate the source range.

Following the methodology introduced in [105], we utilize a transformation that maps the intensity measurements into the striation-frequency domain (See Ch. 3.3.5). The key idea is that the loci of intensity maxima and minima become aligned when the transformation is applied using the correct parameter vector $\mathbf{q} = [r, \dot{r}, \beta]^T$, where r is the candidate range, $\dot{r}$ is the range rate, and β is the WI. We assume that $\dot{r} = \hat{\dot{r}}$ and $\beta = \hat{\beta}$ have been estimated *a priori* (See Ch. 3.2.3).

Let $\underline{\mathbf{x}}_{l,k}^{\text{bt}}$ represent the intensity measurement at striation indexed by l at frequency bin $k \in \mathcal{K}^{\text{bt}}$. Furthermore, let $\mathcal{D}$ denote the set of all valid pairs of frequency indices (i, j) with $i \neq j$. For each pair, the correlation between the corresponding striation vectors—transformed using $\mathbf{q}$ —is expected to be maximized. The objective function for range estimation using the measured intensity matrix $\mathbf{X}^{\text{bt}}$ is then given by the normalized

dot product:

$$h_C(\mathbf{X}^{\text{bt}}|\mathbf{q}) = \sum_{(i,j) \in \mathcal{D}} \frac{1}{\Psi_{i,j}(\mathbf{q})} \sum_{l \in \mathcal{L}_{i,j}} (\underline{x}_{l,i}^{\text{bt}}(\mathbf{q}) - \bar{\underline{x}}_i^{\text{bt}}(\mathbf{q})) (\underline{x}_{l,j}^{\text{bt}}(\mathbf{q}) - \bar{\underline{x}}_j^{\text{bt}}(\mathbf{q})), \quad (\text{B.1})$$

where $\mathcal{L}_{i,j}(\mathbf{q})$ denotes the set of striation indices l for which both frequency components i and j contain valid measurements after transformation with $\mathbf{q}$, and $\bar{\underline{x}}_i^{\text{bt}}(\mathbf{q})$ is the sample mean of vector $\underline{x}_i^{\text{bt}}(\mathbf{q})$ over the set $\mathcal{L}$ at frequency index $i \in \mathcal{K}^{\text{bt}}$. The normalization term $\Psi_{i,j}(\mathbf{q})$ is given by:

$$\Psi_{i,j}(\mathbf{q}) = \left[\sum_{l \in \mathcal{L}_{i,j}(\mathbf{q})} (\underline{x}_{l,i}^{\text{bt}}(\mathbf{q}) - \bar{\underline{x}}_i^{\text{bt}}(\mathbf{q}))^2 \sum_{l \in \mathcal{L}_{i,j}} (\underline{x}_{l,j}^{\text{bt}}(\mathbf{q}) - \bar{\underline{x}}_j^{\text{bt}}(\mathbf{q}))^2 \right]^{1/2}. \quad (\text{B.2})$$

The final range estimate is obtained by maximizing the objective function h_C over the search space or range $\mathcal{S}_r$:

$$\hat{r}_C = \arg \max_{r \in \mathcal{S}_r} h_C(\mathbf{X}^{\text{bt}}|\mathbf{q} = [r, \hat{r}, \hat{\beta}]^T). \quad (\text{B.3})$$

Similarly, if the range and the range rate are treated as known (estimated as $\hat{r}$ and $\hat{\dot{r}}$), the WI is estimated as

$$\hat{\beta}_C = \arg \max_{\beta \in \mathcal{S}_\beta} h_C(\mathbf{X}^{\text{bt}}|\mathbf{q} = [\hat{r}, \hat{\dot{r}}, \beta]^T). \quad (\text{B.4})$$

This approach allows for range estimation using incomplete striations, i.e., it does not require that all tonal frequencies exhibit valid striations simultaneously. While this flexibility increases the robustness of the method to missing data, it also results in a combinatorial increase in computational load with the number of tonal frequencies considered, since all pairwise combinations are evaluated. Further application of this method to real acoustic data from the SWellEx-96 experiment [148], as well as its potential application for UUV self-localization, is detailed in [105].

Acknowledgments:

Appendix B is adapted in part with permission from: J. Jang and F. Meyer, "Navigation in Shallow Water Using Passive Acoustic Ranging," in *Proc. Int. Conf. Inf. Fusion*, June 2023. The dissertation author was the primary investigator and lead author of this work.

Appendix C

Other Ship Ranging Results

To demonstrate the broader applicability of the proposed waveguide-invariant (WI)-based ranging method, this appendix presents range estimation results for two additional ships traversing the same shipping lane as MSC KALAMATA. These serve to validate that the previously estimated WI value of $\beta = 1.18$ remains effective for other vessels operating in the same environment. In both cases, far-field conditions (ranges beyond 20 km) are considered, and the results confirm that the method consistently achieves absolute range estimation errors below 10%—even beyond 45 km—despite differences in vessel type and slight deviations in course. The same signal processing procedure described in Ch. 3 is applied, with adaptations for each scenario as needed.

C.1 Scenario 2: ZIM NEW YORK

This scenario examines the eastbound transit of the commercial container ship ZIM NEW YORK (MMSI: 477634700), which occurred between 11:25 and 12:05 UTC on March 28th, 2018. The vessel traveled at a reported speed of 18.3 kn (9.4 m/s). Figure C.1 shows the track of ZIM NEW YORK alongside that of MSC KALAMATA, plotted relative to the VLA2 array. The average range rate was computed as 9.4 m/s; however, the vessel made a slight northerly course correction midway through its trajectory, resulting in an estimated change of 0.3 m/s in effective range rate.

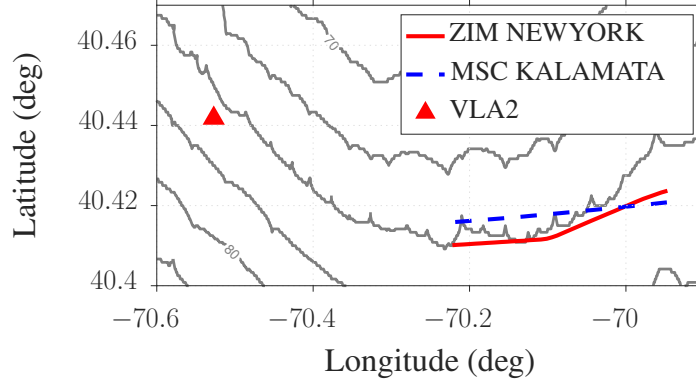


Figure C.1. Vessel tracks of ZIM NEW YORK (solid red) moving east (right) and MSC KALAMATA (dashed blue) relative to the VLA2 location (red triangle). Both ships transited east along the same shipping lane, with ZIM NEW YORK exhibiting a slight northward course change mid-transit. The gray contour lines indicate the bathymetry (water depth) with 2.5 m intervals.

The spectrogram of ZIM NEW YORK, shown in Fig. C.2 (a), reveals tonal features and modal interference patterns, albeit more diffuse compared to MSC KALAMATA. The 42–49 Hz band used previously showed weak interference patterns for this ship, so the 32–39 Hz band was used instead. Clear striations are visible at ranges beyond 30 km (closer-range data were not available for this event).

Using the same WI value of $\beta = 1.18$ and a nominal range rate of 9.4 m/s, the estimated range errors as a function of distance are shown in Fig. C.2 (b). Despite overall consistency, a slight bias in range estimates is observed, likely due to a mismatch between the assumed and actual WI values, compounded by the initial 1 km offset between the ships' tracks. Additionally, the mid-course change in heading introduced growing discrepancies in the range rate, contributing to the increase in error at longer ranges. Nonetheless, absolute errors remained under 10% even at distances exceeding 45 km.

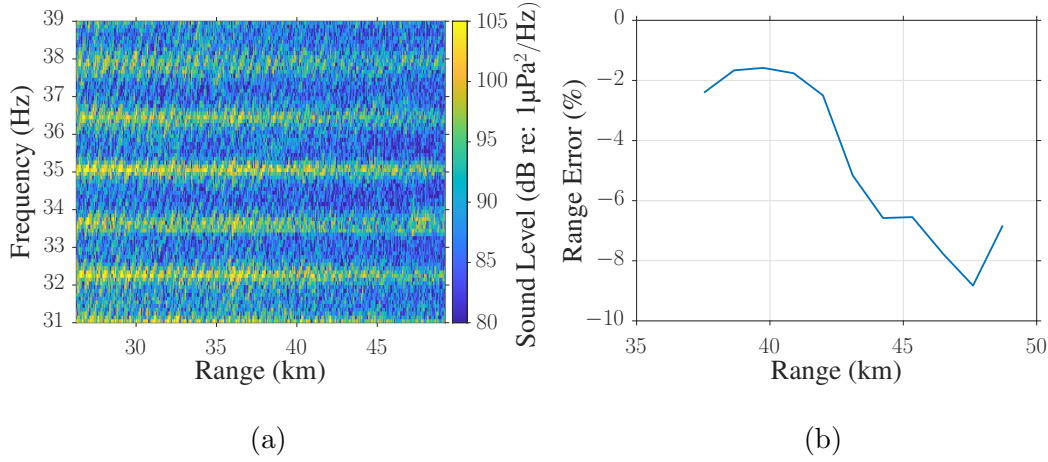


Figure C.2. (a) Spectrogram of acoustic data recorded during ZIM NEW YORK’s passage, showing striated interference patterns between 32–39 Hz beyond 30 km. (b) Range estimation error as a function of distance using $WI = 1.18$ and range rate = 9.4 m/s. A slight increase in error beyond 45 km reflects bias due to the ship’s course change and initial track offset.

C.2 Scenario 3: TEAM CORRIDO

This scenario evaluates the eastbound transit of the oil tanker TEAM CORRIDO (MMSI: 636014623) from 05:00 to 06:15 UTC on March 25th, 2018. The vessel maintained a consistent course and speed of 14.6 kn (7.5 m/s), with an estimated average range rate of 7.4 m/s. The corresponding track is shown in Fig. C.3, again plotted relative to MSC KALAMATA and the VLA2 location.

Unlike commercial container ships, TEAM CORRIDO did not produce distinct tonal features. Instead, its acoustic signature was characterized by diffuse energy, i.e., no strong tonals, as seen in Fig. C.4 (a). Due to the lack of prominent tonals, a broadband-based estimation method was employed instead, treating all frequency components as random broadband noise, following the approach in [106].

Range estimation was conducted over a segment spanning from 20 km to nearly 55 km, during which the track of MSC KALAMATA and that of TEAM CORRIDO remained within 500 m of each other. The results, shown in Fig. C.4(b), reveal low bias

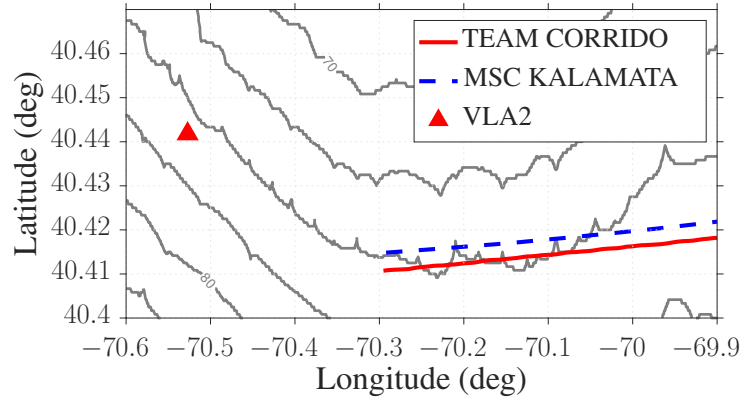


Figure C.3. Vessel tracks of TEAM CORRIDO (solid red) moving east (right) and MSC KALAMATA (dashed blue) relative to the VLA2 location (red triangle). TEAM CORRIDO maintained a consistent heading along the same shipping lane during its eastbound transit. The gray contour lines indicate the bathymetry (water depth) with 2.5 m intervals.

and strong agreement with ground truth, aided by the ship's steady heading and velocity. The absolute range estimation error remained below 4% throughout.

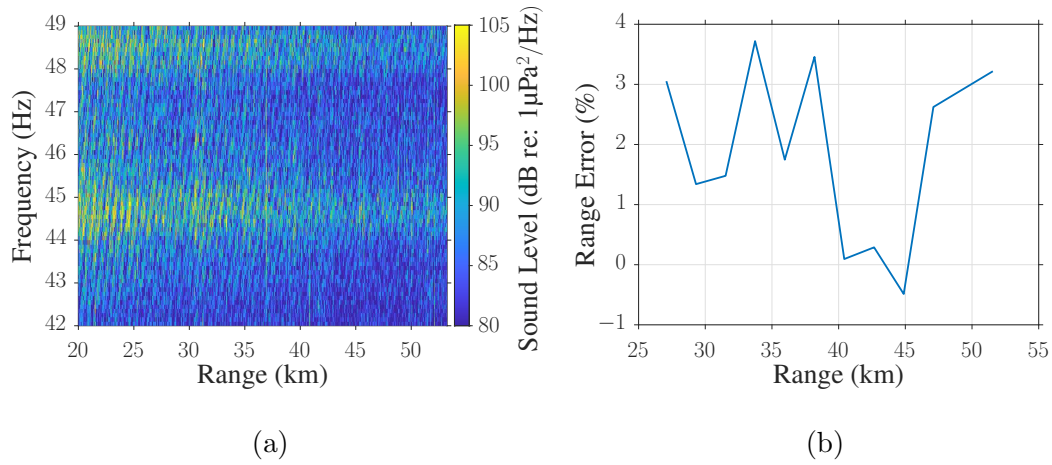


Figure C.4. (a) Spectrogram of acoustic data recorded during TEAM CORRIDO’s passage, showing diffuse broadband energy between 42–49 Hz. No strong tonal features are evident. (b) Range estimation error using a broadband method with $\text{WI} = 1.18$ and constant range rate of 7.4 m/s. The error remains below 4% across the entire 20–55 km range.

Bibliography

- [1] P. S. Wilson, D. P. Knobles, and T. B. Neilsen, “Guest editorial an overview of the Seabed Characterization Experiment,” *IEEE J. Ocean. Eng.*, vol. 45, no. 1, pp. 1–13, 2020.
- [2] Y.-M. Jiang, C. W. Holland, S. E. Dosso, and J. Dettmer, “Depth and frequency dependence of geoacoustic properties on the New England Mud Patch from reflection coefficient inversion),” *J. Acoust. Soc. Am.*, vol. 154, pp. 2383–2397, 10 2023.
- [3] T. A. Marques, L. Thomas, S. W. Martin, D. K. Mellinger, J. A. Ward, D. J. Moretti, D. Harris, and P. L. Tyack, “Estimating animal population density using passive acoustics,” *Biol. Rev.*, vol. 88, no. 2, pp. 287–309, 2013.
- [4] J. A. Hildebrand, S. Baumann-Pickering, K. E. Frasier, J. S. Trickey, K. P. Merkens, S. M. Wiggins, M. A. McDonald, L. P. Garrison, D. Harris, T. A. Marques, and L. Thomas, “Passive acoustic monitoring of beaked whale densities in the Gulf of Mexico,” *Sci. Rep.*, vol. 5, p. 16343, 2015.
- [5] M. Gassmann, S. M. Wiggins, and J. A. Hildebrand, “Three-dimensional tracking of Cuvier’s beaked whales’ echolocation sounds using nested hydrophone arrays,” *J. Acoust. Soc. Am.*, vol. 138, no. 4, pp. 2483–2494, 2015.
- [6] C. M. Schoenbeck, A. Solsona-Berga, P. J. S. Franks, K. E. Frasier, J. S. Trickey, C. Aguilar, I. D. Schroeder, A. Širović, S. J. Bograd, G. Gopalakrishnan, and S. Baumann-Pickering, “*Ziphius cavirostris* presence relative to the vertical and temporal variability of oceanographic conditions in the southern california bight,” *Ecol. Evol.*, vol. 14, no. 7, p. e11708, 2024.
- [7] M. N. Alksne, A. C. M. Kok, A. Agarwal, K. E. Frasier, and S. Baumann-Pickering, “Biogeographic patterns of pacific white-sided dolphins based on long-term passive acoustic records,” *Divers. Distrib.*, vol. 30, no. 9, p. e13903, 2024.
- [8] A. M. Thode, S. B. Blackwell, A. S. Conrad, K. H. Kim, T. Marques, L. Thomas, C. S. Oedekoven, D. Harris, and K. Bröker, “Roaring and repetition: How bowhead whales adjust their call density and source level (Lombard effect) in the presence of natural and seismic airgun survey noise,” *J. Acoust. Soc. Am.*, vol. 147, no. 3, pp. 2061–2080, 2020.

- [9] A. Krumpel, A. Rice, K. E. Frasier, F. Reese, J. S. Trickey, A. E. Simonis, J. P. Ryan, S. M. Wiggins, A. Denzinger, H.-U. Schnitzler, and S. Baumann-Pickering, “Long-term patterns of noise from underwater explosions and their relation to fisheries in Southern California,” *Front. Mar. Sci.*, vol. 8, 2021.
- [10] E. A. Falcone, G. S. Schorr, S. L. Watwood, S. L. DeRuiter, A. N. Zerbini, R. D. Andrews, R. P. Morrissey, and D. J. Moretti, “Diving behaviour of Cuvier’s beaked whales exposed to two types of military sonar,” *R. Soc. Open Sci.*, vol. 4, no. 8, p. 170629, 2017.
- [11] R. T. Buxton, M. F. McKenna, D. Mennitt, K. Fristrup, K. Crooks, L. Angeloni, and G. Wittenmyer, “Noise pollution is pervasive in U.S. protected areas,” *Sci.*, vol. 356, no. 6337, pp. 531–533, 2017.
- [12] C. Erbe, S. A. Marley, R. P. Schoeman, J. N. Smith, L. E. Trigg, and C. B. Embling, “The effects of ship noise on marine mammals—A review,” *Front. Mar. Sci.*, vol. 6, 2019.
- [13] V. M. ZoBell, J. A. Hildebrand, and K. E. Frasier, “Comparing pre-industrial and modern ocean noise levels in the Santa Barbara Channel,” *Mar. Pollut. Bull.*, vol. 202, p. 116379, 2024.
- [14] E. Pirotta, C. G. Booth, D. P. Costa, E. Fleishman, S. D. Kraus, D. Lusseau, D. Moretti, L. F. New, R. S. Schick, L. K. Schwarz, S. E. Simmons, L. Thomas, P. L. Tyack, M. J. Weise, R. S. Wells, and J. Harwood, “Understanding the population consequences of disturbance,” *Ecol. Evol.*, vol. 8, no. 19, pp. 9934–9946, 2018.
- [15] E. L. Hazen, B. Abrahms, S. Brodie, G. Carroll, M. G. Jacox, M. S. Savoca, K. L. Scales, W. J. Sydeman, and S. J. Bograd, “Marine top predators as climate and ecosystem sentinels,” *Front. Ecol. Environ.*, vol. 17, no. 10, pp. 565–574, 2019.
- [16] W. M. X. Zimmer, *Passive Acoustic Monitoring of Cetaceans*. Cambridge, UK: Cambridge University Press, first ed., 2011.
- [17] W. W. Au and M. O. Lammers, *Cetacean Acoustics*, pp. 843–875. New York, NY: Springer New York, 2014.
- [18] C. W. Clark, *Acoustic Behavior of Mysticete Whales*, pp. 571–583. Boston, MA: Springer US, 1990.
- [19] S. Baumann-Pickering, M. A. McDonald, A. E. Simonis, A. Solsona Berga, K. P. B. Merckens, E. M. Oleson, M. A. Roch, S. M. Wiggins, S. Rankin, T. M. Yack, and J. A. Hildebrand, “Species-specific beaked whale echolocation signals,” *J. Acoust. Soc. Am.*, vol. 134, no. 3, pp. 2293–2301, 2013.
- [20] P. Gruden, E.-M. Nosal, and E. Oleson, “Tracking time differences of arrivals of multiple sound sources in the presence of clutter and missed detections,” *J. Acoust. Soc. Am.*, vol. 150, no. 5, pp. 3399–3416, 2021.

- [21] L. Bouffaut, K. Taweesintananon, H. Kriesell, R. Rørstadbotnen, J. Potter, M. Landro, S. E. Johansen, J. Brenne, A. Haukanes, O. Schjelderup, and F. Storvik, “Eavesdropping at the speed of light: distributed acoustic sensing of baleen whales in the Arctic,” *Front. Mar. Sci.*, vol. 9, 2022.
- [22] J. Barlow, E. T. Griffiths, H. Klinck, and D. V. Harris, “Diving behavior of Cuvier’s beaked whales inferred from three-dimensional acoustic localization and tracking using a nested array of drifting hydrophone recorders,” *J. Acoust. Soc. Am.*, vol. 144, no. 4, pp. 2030–2041, 2018.
- [23] S. Fregosi, D. V. Harris, H. Matsumoto, D. K. Mellinger, J. Barlow, S. Baumann-Pickering, and H. Klinck, “Detections of whale vocalizations by simultaneously deployed bottom-moored and deep-water mobile autonomous hydrophones,” *Front. Mar. Sci.*, vol. 7, 2020.
- [24] E. Henderson, S. Martin, R. Manzano-Roth, and B. Matsuyama, “Occurrence and habitat use of foraging Blainville’s beaked whales (*Mesoplodon densirostris*) on a U.S. navy range in Hawaii,” *Aquat. Mamm.*, vol. 42, pp. 549–562, 12 2016.
- [25] R. Gibb, E. Browning, P. Glover-Kapfer, and K. E. Jones, “Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring,” *Methods Ecol. Evol.*, vol. 10, no. 2, pp. 169–185, 2019.
- [26] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. USA: Simon & Schuster, Inc., 1992.
- [27] J. Barlow, S. Fregosi, L. Thomas, D. Harris, and E. T. Griffiths, “Acoustic detection range and population density of Cuvier’s beaked whales estimated from near-surface hydrophones,” *J. Acoust. Soc. Am.*, vol. 149, pp. 111–125, 01 2021.
- [28] S. H. Abadi, H. C. Song, and D. R. Dowling, “Broadband sparse-array blind deconvolution using frequency-difference beamforming,” *J. Acoust. Soc. Am.*, vol. 132, pp. 3018–3029, 11 2012.
- [29] B. M. Worthmann and D. R. Dowling, “The frequency-difference and frequency-sum acoustic-field autoproductions,” *J. Acoust. Soc. Am.*, vol. 141, pp. 4579–4590, 06 2017.
- [30] A. S. Douglass, H. C. Song, and D. R. Dowling, “Performance comparisons of frequency-difference and conventional beamforming,” *J. Acoust. Soc. Am.*, vol. 142, pp. 1663–1673, 09 2017.
- [31] A. S. Douglass and D. R. Dowling, “Frequency-difference beamforming in the presence of strong random scattering,” *J. Acoust. Soc. Am.*, vol. 146, pp. 122–134, 07 2019.
- [32] J.-H. Lee, Y. Park, and P. Gerstoft, “Compressive frequency-difference direction-of-arrival estimation,” *J. Acoust. Soc. Am.*, vol. 154, pp. 141–151, 07 2023.

- [33] D. Kim, G. Byun, and J. Kim, “Direction-of-arrival estimation based on frequency difference–wavenumber analysis for sparse vertical array configuration,” *Sensors*, vol. 23, no. 1, 2023.
- [34] J. C. Neo, W. Fan, Y. Huang, and D. Dowling, “Frequency-difference backprojection of earthquakes,” *Geophys. J. Int.*, vol. 231, pp. 2173–2185, 08 2022.
- [35] R. Ali, T. Mitcham, I. Owolabi, S. McConnell, and N. Duric, “Frequency-differencing strategy to kickstart full-waveform inversion without cycle skipping,” *JASA Express Lett.*, vol. 5, p. 012001, 01 2025.
- [36] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York, NY, 10020: McGraw-Hill, Inc., second ed., 1984.
- [37] C. Knapp and G. Carter, “The generalized correlation method for estimation of time delay,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, 1976.
- [38] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, 1986.
- [39] Y. Park, P. Gerstoft, and J.-H. Lee, “Difference-Frequency MUSIC for DOAs,” *IEEE Signal Process. Lett.*, vol. 29, pp. 2612–2616, 2022.
- [40] C. G. Carter, A. H. Nuttall, and P. G. Cable, “The smoothed coherence transform,” *Proc. IEEE*, vol. 61, no. 10, pp. 1497–1498, 1973.
- [41] S. M. Wiggins and J. A. Hildebrand, “High-frequency acoustic recording package (HARP) for broad-band, long-term marine mammal monitoring,” in *Proc. IEEE Int. Symp. Underw. Technol.*, pp. 551–557, Apr. 2007.
- [42] L. Rabiner, J. McClellan, and T. Parks, “FIR digital filter design techniques using weighted Chebyshev approximation,” *Proc. IEEE*, vol. 63, no. 4, pp. 595–610, 1975.
- [43] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Englewood Cliffs, New Jersey 07632: Prentice Hall, first ed., 1989.
- [44] E. R. Snyder, A. Solsona-Berga, S. Baumann-Pickering, K. E. Frasier, S. M. Wiggins, and J. A. Hildebrand, “Where’s Whaledo: A software toolkit for array localization of animal vocalizations,” *PLoS Comput Biol.*, vol. 20, p. e1011456, 05 2024.
- [45] L. Baggett, “Long-term monitoring of cuvier’s beaked whale diving behavior in southern california using 3d tracking from fixed hydrophone arrays,” Master’s thesis, University of California San Diego, 2023.
- [46] S. M. Wiggins, B. J. Thayre, J. S. Trickey, S. Baumann-Pickering, and J. A. Hildebrand, “MPL technical memorandum 631: Beaked whale passive acoustic tracking offshore of Cape Hatteras 2017,” Tech. Rep. N62470-15-D-8006, MPL, La Jolla, CA, Oct. 2018.

- [47] L. Tenorio-Hallé, A. M. Thode, M. O. Lammers, A. S. Conrad, and K. H. Kim, “Multi-target 2D tracking method for singing humpback whales using vector sensors,” *J. Acoust. Soc. Am.*, vol. 151, no. 1, pp. 126–137, 2022.
- [48] Y. M. Barkley, E.-M. Nosal, and E. M. Oleson, “Model-based localization of deep-diving cetaceans using towed line array acoustic data,” *J. Acoust. Soc. Am.*, vol. 150, no. 2, pp. 1120–1132, 2021.
- [49] P. Gruden and P. R. White, “Automated extraction of dolphin whistles—a sequential Monte Carlo probability hypothesis density approach,” *J. Acoust. Soc. Am.*, vol. 148, no. 5, pp. 3014–3026, 2020.
- [50] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. Storrs, CT: Yaakov Bar-Shalom, 2011.
- [51] D. B. Reid, “An algorithm for tracking multiple targets,” *IEEE Trans. Autom. Control*, vol. 24, pp. 843–854, Dec. 1979.
- [52] R. Mahler, “Multitarget Bayes filtering via first-order multitarget moments,” vol. 39, no. 4, pp. 1152–1178, 2003.
- [53] R. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Norwood, MA: Artech House, 2007.
- [54] A. A. Saucan, M. Coates, and M. Rabbat, “Multi-sensor multi-Bernoulli filter,” *IEEE Trans. Signal Process.*, vol. 65, pp. 5495–5509, Oct. 2017.
- [55] F. Meyer, T. Kropfreiter, J. L. Williams, R. Lau, F. Hlawatsch, P. Braca, and M. Z. Win, “Message passing algorithms for scalable multitarget tracking,” *Proc. IEEE*, vol. 106, pp. 221–259, Feb. 2018.
- [56] F. Meyer, P. Braca, P. Willett, and F. Hlawatsch, “A scalable algorithm for tracking an unknown number of targets using multiple sensors,” *IEEE Trans. Signal Process.*, vol. 65, pp. 3478–3493, July 2017.
- [57] F. Meyer, A. Tesei, and M. Z. Win, “Localization of multiple sources using time-difference arrival measurements,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 3151–3155, Mar. 2017.
- [58] A. Tesei, F. Meyer, and R. Been, “Tracking of multiple surface vessels based on passive acoustic underwater arrays,” *J. Acoust. Soc. Am.*, vol. 147, no. 2, pp. EL87–EL92, 2020.
- [59] P. M. Baggenstoss, “A multi-hypothesis tracker for clicking whales,” *J. Acoust. Soc. Am.*, vol. 137, no. 5, pp. 2552–2562, 2015.
- [60] P. M. Baggenstoss, “Separation of sperm whale click-trains for multipath rejection,” *J. Acoust. Soc. Am.*, vol. 129, no. 6, pp. 3598–3609, 2011.

- [61] T. Kropfreiter, F. Meyer, S. Coraluppi, C. Carthel, R. Mendrzik, and P. Willett, “Track Coalescence and Repulsion: MHT, JPDA, and BP,” in *Proc. Int. Conf. Inf. Fusion*, pp. 1–8, Nov. 2021.
- [62] B. N. Vo and W. K. Ma, “The Gaussian mixture probability hypothesis density filter,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4091–4104, 2006.
- [63] P. Gruden, J. Jang, A. Kügler, T. Kropfreiter, L. Tenorio-Hallé, M. O. Lammers, A. Thode, and F. Meyer, “Automating multi-target tracking of singing humpback whales recorded with vector sensors,” *J. Acoust. Soc. Am.*, vol. 154, no. 4, pp. 2579–2593, 2023.
- [64] C. O. Tiemann, M. B. Porter, and L. N. Frazer, “Localization of marine mammals near Hawaii using an acoustic propagation model,” *J. Acoust. Soc. Am.*, vol. 115, no. 6, pp. 2834–2843, 2004.
- [65] E.-M. Nosal, “Methods for tracking multiple marine mammals with wide-baseline passive acoustic arrays,” *J. Acoust. Soc. Am.*, vol. 134, no. 3, pp. 2383–2392, 2013.
- [66] J. Macaulay, J. Gordon, D. Gillespie, C. Malinka, and S. Northridge, “Passive acoustic methods for fine-scale tracking of harbour porpoises in tidal rapids,” *J. Acoust. Soc. Am.*, vol. 141, no. 2, pp. 1120–1132, 2017.
- [67] J. A. Nelder and R. Mead, “A simplex method for function minimization,” *Comput. J.*, vol. 7, pp. 308–313, 1965.
- [68] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines,” *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [69] H. W. Kuhn, “The Hungarian method for the assignment problem,” *Nav. Res. Logist. Q.*, vol. 2, no. 1-2, pp. 83–97, 1955.
- [70] E. Snyder, S. M. Wiggins, S. Baumann-Pickering, and J. A. Hildebrand, “Cuvier’s beaked whale tracks in Southern California,” *J. Acoust. Soc. Am.*, vol. 146, no. 4, pp. 2939–2939, 2019.
- [71] S. Geman and D. Geman, “Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, pp. 721–741, 1984.
- [72] F. Meyer and J. L. Williams, “Scalable detection and tracking of geometric extended objects,” *IEEE Trans. Signal Process.*, vol. 69, pp. 6283–6298, Oct. 2021.
- [73] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, “Factor graphs and the sum-product algorithm,” *IEEE Trans. Inf. Theory*, vol. 47, pp. 498–519, Feb. 2001.

- [74] F. Daum and J. Huang, “Nonlinear filters with log-homotopy,” in *Proc. Soc. Photo-Opt. Instrum. Eng.*, pp. 423–437, Aug. 2007.
- [75] W. Zhang and F. Meyer, “Graph-based multiobject tracking with embedded particle flow,” in *Proc. IEEE Radar Conf.*, pp. 1–6, May 2021.
- [76] Y. Bar-Shalom, T. Kirubarajan, and X.-R. Li, *Estimation with Applications to Tracking and Navigation*. New York, NY: Wiley, 2002.
- [77] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking,” *IEEE Trans. Signal Process.*, vol. 50, pp. 174–188, Feb. 2002.
- [78] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.
- [79] F. Daum and J. Huang, “Curse of dimensionality and particle filters,” in *Proc. IEEE Aerospace*, vol. 4, pp. 1979–1993, Mar. 2003.
- [80] Y. Li and M. Coates, “Particle filtering with invertible particle flow,” *IEEE Trans. Signal Process.*, vol. 65, pp. 4102–4116, Aug. 2017.
- [81] B. Ristic, B.-T. Vo, B.-N. Vo, and A. Farina, “A tutorial on Bernoulli filters: Theory, implementation and applications,” *IEEE Trans. Signal Process.*, vol. 61, pp. 3406–3430, Apr. 2013.
- [82] G. Soldi, F. Meyer, P. Braca, and F. Hlawatsch, “Self-tuning algorithms for multisensor-multitarget tracking using belief propagation,” *IEEE Trans. Signal Process.*, vol. 67, pp. 3922–3937, Aug. 2019.
- [83] H. A. P. Blom and Y. Bar-Shalom, “The interacting multiple model algorithm for systems with markovian switching coefficients,” *IEEE Trans. Autom. Control*, vol. 33, no. 8, pp. 780–783, 1988.
- [84] R. C. Connor, J. Mann, P. L. Tyack, and H. Whitehead, “Social evolution in toothed whales,” *Trends Ecol. Evol.*, vol. 13, no. 6, pp. 228–232, 1998.
- [85] K. McHugh, “Odontocete social strategies and tactics along and inshore,” in *Ethology and Behavioral Ecology of Odontocetes*, pp. 165–182, Cham, Switzerland: Springer, 2019.
- [86] R. Dreo, A. Trabattini, P. Stinco, M. Micheli, and A. Tesei, “Detection and localization of multiple ships using acoustic vector sensors on buoyancy gliders: Practical design considerations and experimental verifications,” *IEEE J. Ocean. Eng.*, vol. 48, no. 2, pp. 577–591, 2023.
- [87] K. R. Kita, S. Randeni, D. DiBiaso, and H. Schmidt, “Passive acoustic tracking of an unmanned underwater vehicle using bearing-Doppler-speed measurements,” *J. Acoust. Soc. Am.*, vol. 151, pp. 1311–1324, 02 2022.

- [88] J. S. Jaffe, P. J. S. Franks, P. L. D. Roberts, D. Mirza, C. Schurgers, R. Kastner, and A. Boch, “A swarm of autonomous miniature underwater robot drifters for exploring submesoscale ocean dynamics,” *Nat. Commun.*, vol. 8, p. 14189, Jan. 2017.
- [89] C. Graupe, L. J. Van Uffelen, P. F. Worcester, M. A. Dzieciuch, and B. M. Howe, “An automated framework for long-range acoustic positioning of autonomous underwater vehicles,” *J. Acoust. Soc. Am.*, vol. 152, pp. 1615–1626, 09 2022.
- [90] R. A. S. Gehrman, D. R. Barclay, H. Johnson, N. Shajahan, V. Nolet, and K. T. A. Davies, “Ambient noise levels with depth from an underwater glider survey across shipping lanes in the Gulf of St. Lawrence, Canada,” *J. Acoust. Soc. Am.*, vol. 154, pp. 1735–1745, 09 2023.
- [91] V. Pallayil, Y.-T. Lin, E. Fischell, A. L. Kukulya, D. P. Knobles, and P. S. Wilson, “Integration and testing of a low-profile hydrophone array with REMUS 100 AUV for seabed characterization and marine mammal detection application,” in *Proceedings of 2019 IEEE Underwater Technology (UT)*, pp. 1–8, 2019.
- [92] S. Jesus, M. Porter, Y. Stephan, X. Demoulin, O. Rodriguez, and E. Coelho, “Single hydrophone source localization,” *IEEE J. Ocean. Eng.*, vol. 25, no. 3, pp. 337–346, 2000.
- [93] M. Cheney and I. Kirsteins, “Resolution of matched field processing for a single hydrophone in a rigid waveguide,” *J. Acoust. Soc. Am.*, vol. 152, pp. 3186–3197, 12 2022.
- [94] J. Bonnel and A. Thode, “Range and depth estimation of bowhead whale calls in the arctic using a single hydrophone,” in *Proceedings of 2014 IEEE Sensor Systems for a Changing Ocean (SSCO)*, pp. 1–4, 2014.
- [95] Y. Le Gall, F.-X. Socheleau, and J. Bonnel, “Performance analysis of single-receiver matched-mode localization,” *IEEE J. Ocean. Eng.*, vol. 44, no. 1, pp. 193–206, 2019.
- [96] A. Trabattoni, G. Barruol, R. Dréo, and A. Boudraa, “Ship detection and tracking from single ocean-bottom seismic and hydroacoustic stations,” *J. Acoust. Soc. Am.*, vol. 153, pp. 260–273, 01 2023.
- [97] L. Watkins, P. Stinco, A. Tesei, and F. Meyer, “Sequential source localization in shallow water by single-receiver probabilistic focalization,” *J. Acoust. Soc. Am.*, vol. 157, pp. 2819–2834, 04 2025.
- [98] H. Niu, Z. Gong, E. Ozanich, P. Gerstoft, H. Wang, and Z. Li, “Deep-learning source localization using multi-frequency magnitude-only data,” *J. Acoust. Soc. Am.*, vol. 146, pp. 211–222, 07 2019.
- [99] H. Tao, G. Hickman, J. L. Krolik, and M. Kemp, “Single hydrophone passive localization of transiting acoustic sources,” in *Proc. OCEANS*, pp. 1–3, 2007.

- [100] K. L. Cockrell and H. Schmidt, “Robust passive range estimation using the waveguide invariant,” *J. Acoust. Soc. Am.*, vol. 127, no. 5, pp. 2780–2789, 2010.
- [101] A. Turgut, M. Orr, and D. Rouseff, “Broadband source localization using horizontal-beam acoustic intensity striations,” *J. Acoust. Soc. Am.*, vol. 127, no. 1, pp. 73–83, 2010.
- [102] S. T. Rakotonarivo and W. A. Kuperman, “Model-independent range localization of a moving source in shallow water,” *J. Acoust. Soc. Am.*, vol. 132, no. 4, pp. 2218–2223, 2012.
- [103] A. H. Young, H. A. Harms, G. W. Hickman, J. S. Rogers, and J. L. Krolik, “Waveguide-invariant-based ranging and receiver localization using tonal sources of opportunity,” *IEEE J. Ocean. Eng.*, vol. 45, no. 2, pp. 631–644, 2020.
- [104] K. Sun, D. Gao, X. Zhao, D. Guo, W. Song, and Y. Li, “Estimation of target motion parameters from the tonal signals with a single hydrophone,” *Sensors*, vol. 23, no. 15, 2023.
- [105] J. Jang and F. Meyer, “Navigation in shallow water using passive acoustic ranging,” in *Proc. Int. Conf. Inf. Fusion*, (Charleston, SC, USA), pp. 1–8, 2023.
- [106] J. Jang and F. Meyer, “A new statistical model for waveguide invariant-based range estimation in shallow water,” in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, pp. 1–5, 2024.
- [107] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt in *Computational Ocean Acoustics*, ch. 2.4 and 5.14, Springer Publishing Company, Incorporated, 2nd ed., 2011.
- [108] S. Chuprov, “Interference structure of a sound field in a layered ocean,” *Ocean Acoustics, Current State*, pp. 71–99, 1982.
- [109] G. L. D’Spain and W. A. Kuperman, “Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth,” *J. Acoust. Soc. Am.*, vol. 106, no. 5, pp. 2454–2468, 1999.
- [110] K. L. Cockrell and H. Schmidt, “A relationship between the waveguide invariant and wavenumber integration,” *J. Acoust. Soc. Am.*, vol. 128, pp. EL63–EL68, 06 2010.
- [111] K. L. Cockrell and H. Schmidt, “A modal Wentzel-Kramers-Brillouin approach to calculating the waveguide invariant for non-ideal waveguides,” *J. Acoust. Soc. Am.*, vol. 130, pp. 72–83, 07 2011.
- [112] C. H. Harrison, “The relation between the waveguide invariant, multipath impulse response, and ray cycles,” *J. Acoust. Soc. Am.*, vol. 129, pp. 2863–2877, 05 2011.

- [113] G. Byun and H. C. Song, “The waveguide invariant for a Pekeris waveguide,” *J. Acoust. Soc. Am.*, vol. 151, pp. 846–851, 02 2022.
- [114] D. Rouseff and R. C. Spindel, “Modeling the waveguide invariant as a distribution,” *AIP Conf. Proc.*, vol. 621, pp. 137–150, 06 2002.
- [115] Y. Le Gall and J. Bonnel, “Passive estimation of the waveguide invariant per pair of modes,” *J. Acoust. Soc. Am.*, vol. 134, pp. EL230–EL236, 07 2013.
- [116] S. M. Haver, J. D. Adams, L. T. Hatch, S. M. Van Parijs, R. P. Dziak, J. Haxel, S. A. Heppell, M. F. McKenna, D. K. Mellinger, and J. Gedamke, “Large vessel activity and low-frequency underwater sound benchmarks in United States waters,” *Front. Mar. Sci.*, vol. 8, 2021.
- [117] G. M. Wenz, “Acoustic ambient noise in the ocean: Spectra and sources,” *J. Acoust. Soc. Am.*, vol. 34, pp. 1936–1956, 12 1962.
- [118] M. McKenna, D. Ross, S. Wiggins, and J. Hildebrand, “Underwater radiated noise from modern commercial ships,” *J. Acoust. Soc. Am.*, vol. 131, pp. 92–103, 01 2012.
- [119] R. J. Urick in *Principles of Underwater Sound*, ch. 5 and 9, Peninsula Publishing, 3rd ed., 2013.
- [120] F. B. Jensen and W. A. Kuperman, “Optimum frequency of propagation in shallow water environments,” *J. Acoust. Soc. Am.*, vol. 73, pp. 813–819, 03 1983.
- [121] P. Scrimger and R. M. Heitmeyer, “Acoustic source-level measurements for a variety of merchant ships,” *J. Acoust. Soc. Am.*, vol. 89, pp. 691–699, 02 1991.
- [122] C. Zhu, T. Gaggero, N. C. Makris, and P. Ratilal, “Underwater sound characteristics of a ship with controllable pitch propeller,” *J. Mar. Sci. Eng.*, vol. 10, no. 3, 2022.
- [123] D. Ross, *Mechanics of underwater noise / Donald Ross*. New York: Pergamon Press, 1976.
- [124] S. C. Wales and R. M. Heitmeyer, “An ensemble source spectra model for merchant ship-radiated noise,” *J. Acoust. Soc. Am.*, vol. 111, pp. 1211–1231, 03 2002.
- [125] M. F. McKenna, S. M. Wiggins, and J. A. Hildebrand, “Relationship between container ship underwater noise levels and ship design, operational and oceanographic conditions,” *Sci. Rep.*, vol. 3, p. 1760, May 2013.
- [126] Y. Yao, C. Sun, X. Liu, and G. Jiang, “Robust range estimation based on the slope of the two-dimensional Fourier transform ridge using radon transform,” *IEEE Access*, vol. 9, pp. 24093–24104, 2021.
- [127] C. M. A. Verlinden, J. Sarkar, B. D. Cornuelle, and W. A. Kuperman, “Determination of acoustic waveguide invariant using ships as sources of opportunity in a shallow water marine environment,” *J. Acoust. Soc. Am.*, vol. 141, pp. 102–107, 02 2017.

- [128] H. Tao, G. Hickman, J. L. Krolik, and M. Kemp, “Single hydrophone passive localization of transiting acoustic sources,” in *Proc. IEEE OCEANS*, pp. 1–3, 2007.
- [129] W. Burdic in *Underwater Acoustic System Analysis*, ch. 8.4.2, Peninsula Publ., 2nd ed., 2002.
- [130] M. Evans, N. Hastings, and B. Peacock in *Statistical distributions*, ch. 9, John Wiley & Sons, 3rd ed., 2000.
- [131] H. C. Song and G. Byun, “Extrapolating Green’s functions using the waveguide invariant theory,” *J. Acoust. Soc. Am.*, vol. 147, pp. 2150–2158, 04 2020.
- [132] D. Kim, D. Kim, G. Byun, J. Kim, and H. Song, “Improvement of a Green’s function estimation for a moving source using the waveguide invariant theory,” *Sensors*, vol. 24, no. 17, 2024.
- [133] P. Welch, “The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms,” *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, 1967.
- [134] J. Jang, “Shallow-water-ship-ranging.” <https://github.com/junsujang/Shallow-Water-Ship-Ranging>. Github Repository (2025).
- [135] A. Harms, J. L. Odom, and J. L. Krolik, “Ocean acoustic waveguide invariant parameter estimation using tonal noise sources,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, (South Brisbane, QLD, Australia), pp. 4001–4004, 2015.
- [136] A. H. Young, J. Soli, and G. Hickman, “Self-localization technique for unmanned underwater vehicles using sources of opportunity and a single hydrophone,” in *Proc. OCEANS*, (Anchorage, AK, USA), pp. 1–6, 2017.
- [137] G. Byun, H. C. Song, and C. Cho, “Adaptive array invariant in range-dependent environments with variable bathymetry,” *J. Acoust. Soc. Am.*, vol. 149, pp. 1363–1370, 02 2021.
- [138] D. Kim, G. Byun, J. Song, J. S. Kim, and H. C. Song, “Generalized Taylor expansion of the waveguide invariant for a homogenous waveguide with variable bathymetry,” *J. Acoust. Soc. Am.*, vol. 157, pp. 2473–2476, 04 2025.
- [139] J. Jang, F. Meyer, E. R. Snyder, S. M. Wiggins, S. Baumann-Pickering, and J. A. Hildebrand, “Bayesian detection and tracking of odontocetes in 3-D from their echolocation clicks,” *J. Acoust. Soc. Am.*, vol. 153, pp. 2690–, 05 2023.
- [140] Y. Le Gall and J. Bonnel, “Separation of moving ship striation patterns using physics-based filtering,” in *Proc. Meet. Acoust.*, vol. 19, p. 070073, 2013.
- [141] M. Porter, “The KRAKEN normal mode program,” Tech. Rep. NRL/MR/5120-92-6920, Naval Research Lab, Washington, DC, May 1992.

- [142] D. P. Knobles, P. S. Wilson, T. B. Neilsen, and W. S. Hodgkiss, “Influence of seabed on very low frequency sound recorded during passage of merchant ships on the New England shelf,” *J. Acoust. Soc. Am.*, vol. 149, pp. 3294–3300, 05 2021.
- [143] M. S. Ballard, D. D. Garcia, K. M. Lee, G. R. Venegas, A. R. McNeese, P. S. Wilson, and J. D. Chaytor, “Direct measurements of sediment geoacoustic properties in the New England Mud Patch and shelf break,” *J. Acoust. Soc. Am.*, vol. 156, pp. 2889–2902, 10 2024.
- [144] S. E. Dosso, P. S. Wilson, D. P. Knobles, and J. Bonnel, “Comparison and combination of matched-field and modal-dispersion inversion for seabed geoacoustic profiles at the new england mud patcha),” *J. Acoust. Soc. Am.*, vol. 157, pp. 4157–4174, 06 2025.
- [145] H. C. Song and G. Byun, “Simultaneous localization of a surface ship and a submerged towed source,” *J. Acoust. Soc. Am.*, vol. 144, pp. 2238–2241, 10 2018.
- [146] R. N. Bracewell, *Two-dimensional imaging*, pp. 505–537. USA: Prentice-Hall, Inc., 1995.
- [147] W. Feller, *An introduction to probability theory and its applications*, ch. 10.7. New York: Wiley, 1950.
- [148] J. Murray and D. Ensberg, “Swelllex-96: S5 event.” <http://swellex96.ucsd.edu/s5.htm>, 1996. Online; Accessed on 28-Feb-2023.