Advanced Environmental Monitoring for Marine Renewable Energy

Emma Cotter

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Brian Polagye, Chair

Steven Brunton

Andrew Stewart

Program Authorized to Offer Degree:
Mechanical Engineering

University of Washington

Abstract

Advanced Environmental Monitoring
for Marine Renewable Energy

Emma Cotter

Chair of the Supervisory Committee:
Professor Brian Polagye
Mechanical Engineering

Marine renewable energy has the potential to provide clean, reliable power to coastal communities and offshore facilities. However, the effects that marine energy development might have on the environment are not yet well understood. One environmental risk of particular concern is that of collision between an animal and a marine energy converter, but conducting the requisite environmental monitoring to understand this risk has presented a challenge at marine energy sites around the world for several reasons. First, if collision does occur, it is likely to be a rare event, meaning that detection requires continuous monitoring over extended deployments. Second, there is no single sensor that can provide all of the necessary information, and a combination of active acoustic, passive acoustic, and optical sensors is required. Third, these sensors can rapidly accrue vast volumes of data (petabyte-scale), making it difficult to extract insight from collected data. Finally, waves and currents at marine energy sites complicate the deployment of any monitoring instrumentation.

Integrated instrumentation platforms that combine sensors into a single platform can address some of these challenges, because they can provide all of the necessary data and reduce deployment complexity. However, operation of such a platform must meet three directives in order to be most effective: 1) avoid biasing animal behavior through the use of instrumentation, 2) reliably detect rare events, and 3) avoid collection of unmanageable

volumes of data. In this thesis, it is demonstrated that it is possible to simultaneously meet all three of these directives. This is demonstrated using the Adaptable Monitoring Package (AMP), an integrated instrumentation platform that combines multibeam sonars, optical cameras, hydrophones, and an acoustic Doppler current profiler.

Artificial illumination is necessary to collect data from optical cameras when ambient light is not available. However, this light can either attract or repel animals. To minimize these effects (e.g., meet directive 1), the AMP uses detection, tracking, and classification of targets in the multibeam sonar data to restrict the use of artificial illumination to periods when a target of interest is present and might be detectable by the optical cameras. Information about target presence is also used to limit data archival to periods when a target of interest is present and avoid curation of data that does not contain any useful information (e.g., meeting directives 2 and 3). To benchmark this capability, real-time target detection and tracking are used to limit data archival to periods when any target of potential interest is present during a deployment of the AMP in Sequim Bay, WA. The target detection and tracking approach was found to have a true negative rate of 0.99 (e.g., an estimated 1% of targets of interest were not recorded), but 45% of recorded data did not contain a biological target.

To address this relatively high false positive rate, recorded data were used to train machine learning classification of tracked targets. Three machine learning algorithms, trained using varying parameters and features, were evaluated for this task. A random forest algorithm was found to perform best, and the resulting classification model was able to distinguish between biological targets (e.g., seals, fish) and non-biological targets (e.g., acoustic artifacts) with a true positive rate of 0.97 and a false negative rate of 0.13. This model was then implemented in real-time during a second deployment of the AMP and used to limit data acquisition to periods when biological targets were predicted to be present. The model achieved the same true positive rate and a false positive rate of 0.23 in real-time after re-training with site

specific data. From these results, general recommendations are made for implementation of real-time classification of biological targets in multibeam sonar data at new marine energy sites.

All active acoustic sensors used on the Adaptable Monitoring Package, including the multibeam sonar used for real-time classification, have operating frequencies above the upper limit of marine mammal hearing. However, high-frequency transducers can still produce sound at lower frequencies audible to marine mammals. A comprehensive evaluation of the acoustic emissions of four active acoustic transducers used on the Adaptable Monitoring Package was conducted to understand whether they might cause hearing damage or bias marine mammal hearing (e.g., violating directive 1). All four transducers were found to produce measurable sound below 160 kHz, the reported upper limit of marine mammal hearing. A spatial map of the acoustic emissions of each sonar was used to evaluate potential effects on marine mammal hearing if the transducer were continuously operated from a stationary platform. Based on the cumulative sound exposure level metric, the acoustic emissions from any of the the transducers are unlikely to cause hearing damage to marine mammals. However, the extent of audibility is estimated to be on the order of 100 m, and further research is needed to understand how this might affect marine mammal behavior.

In sum, this thesis provides a framework for effective environmental monitoring that can be used to reduce the the uncertainty surrounding the environmental effects of marine renewable energy. Further, many aspects are widely applicable to the ocean instrumentation community. Automatic classification of fauna in multibeam sonar data had not been previously demonstrated, and has applications in biological research. The methods developed for evaluation of the acoustic emissions of active acoustic sensors allow for effective comparison between transducers, which can be used to inform sensor selection and government regulation of their use.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

α: Absorption coefficient

μPA: Micropascals

τ: Pulse duration

A: Area

ADCP: Acoustic Doppler current profiler

AMP: Adaptable Monitoring Package

dB: Decibels

DFT: Discrete Fourier Transform

f: Frequency

FAST: Fundy Advanced Sensor Technology platform

FLOWBEC: FLOW and BEnthic ECology platform

FP: False positive

FPR_{BIN}: False positive rate for binary classification of biological and non-biological targets

GPS: Global positioning system

I: Intensity

kHZ: Kilohertz

KNN: k-nearest neighbors

m: Meters

MAJ: Major axis length

MEC: Marine energy converter

MIN: Minor axis length

MSL: Pacific Northwest National Laboratories Marine Sciences Laboratory

n: Number of pulses

N: Count

NMFS: US National Marine Fisheries Service

P: Power

p^2: Mean-square sound pressure spectrum

p_{REF}: Reference pressure

PSD: Power Spectral Density

R: Range

RF: Random forest

s: Seconds

S: Speed

SEL_{CUM}: Cumulative sound exposure level

SL: Source level (recieved level at a range of 1 meter)

SL_{DDEC}: Decidecade (one-third octave band) source levels

SVM: Support vector machine

T: Time

TP: True positive

TTS: Temporary threshold shift

TPR_C: True positive rate for a class, c

TPR_{BIN}: True positive rate for binary classification of biological and non-biological targets

TPR_{BIO}: True positive rate for all biological targets

UDP: User Datagram Protocol

ACKNOWLEDGMENTS

I have been incredibly lucky to be surrounded by the most supportive family, friends, colleagues, and mentors throughout the past five years. First and foremost, I would like to thank my parents and my brother, Harry, for their endless support in all of my endeavors, even when that means moving to the opposite coast. I'm looking forward to seeing them frequently next year.

My advisor and committee chair, Brian Polagye has demonstrated the type of advisor that I hope to become through his dedication to his students, breadth of knowledge, and openness to new ideas. It's been an incredible opportunity to work with him, and I look forward working (and climbing!) together in the future. I would also like to thank my committee members, Steve Brunton, Andrea Copping, and Andy Stewart, who have provided guidance on wide-ranging aspects of my research.

I've been lucky to work with a fantastic team of engineers in the Mechanical Engineering Department and at the Applied Physics Laboratory on the Adaptable Monitoring Package project, including Paul Murphy, James Joslin, Paul Gibbs, Corey Crisp, and Robert Cavagnaro. It's been a privilege to work with this group, and I've learned from all of them along the way. I've worked especially closely with Paul Murphy over the last few years, and I can't imagine a better partner-in-crime to build the AMP software with.

While testing the Adaptable Monitoring Package at the Pacific Northwest National Laboratory Marine Science Laboratory in Sequim, WA, I've had the chance to work John Vavrenic, Garrett Staines, Kate Hall, Sue Southard, Genevra Harker-Klimes, and Kailan Mackereth. They have all been integral in the success of the AMP project,

and I can't thank them enough.

The Marine Renewable Energy Laboratory has been an incredible place to work, I have continually learned from my labmates and made some fantastic friendships. I would especially like to thank Hannah Ross, who began her PhD at the same time as me, for her support and friendship over these past five years. Thank you to Trevor Harrison, who sat next to me for the last few years, for both the productive discussions and hilarious distractions. I'd also like to thank my labmates Dominic Forbush, Benjamin Strom, Curtis Rusch, Isabel Scherl, Jessica Noe, Kate VanNess, Carl Stringer, Hannah Aaronson, and Abigail Snortland for their collaboration and community.

I would also like to acknowledge Allan Hurst, who led an all-girls robotics team at my middle school in Cranston, Rhode Island. While I didn't know it at the time, the early exposure to engineering set me on the path I'm on now, and his dedication as a teacher has been a continual inspiration to me.

Finally, I would like to acknowledge the sources of funding for my research. I have been supported by a National Science Foundation Graduate Research Fellowship. Funding for my research has primarily come from the U.S. Department of Energy, as well as the Naval Facilities Engineering Command.

Chapter 1
INTRODUCTION

1.1 Environmental Monitoring for Marine Renewable Energy

Marine renewable energy has the potential to provide clean, reliable power to coastal communities and offshore facilities. However, the effects that marine energy development might have on the marine environment are not yet well understood [8]. This has proven to be a barrier to marine energy development around the world. For example, in northern Scotland, development of tidal energy was limited pending more information about the impacts that tidal turbines might have on harbor seals, a species of particular environmental concern in that area [57]. A better understanding of the environmental effects of marine renewable energy will enable sustainable development of the sector through either mitigation or effective regulation of any negative environmental effects found to be substantial.

One of the primary environmental concerns surrounding the development of marine renewable energy is the risk of collision between marine mammals, fish, or diving birds and a marine energy converter (MEC) [8]. There are many approaches to understanding this risk. Biological sampling has been used to understand the population densities of different species at marine energy sites, which provides insight into the probability that an animal will encounter a MEC (e.g., [6]). Studies of marine mammal behavioral responses to the sound produced by MECs have been conducted to better understand whether animals will be attracted to or avoid the sound (e.g., [29, 56]). These biological studies have provided input (e.g., population density) for models that predict the probability that collision will occur (e.g., [9, 25]). Modeling efforts have unanimously indicated that if collision between an animal and a turbine does occur, it will be a rare event, but time-resolved data monitoring

animal behavior around MECs is required to validate and improve these models. However, collecting these data has proven a perennial challenge [48, 8].

There are several unique challenges to conducting effective environmental monitoring at marine energy sites. First, there is no single oceanographic sensor that can provide all of the necessary information [48]. Optical cameras provide high-resolution imagery sufficient for species-level classification of marine fauna, but can only operate when there is sufficient light, and their detection range is highly dependent on water clarity [35]. Passive acoustic instruments, such as hydrophones, can detect and localize vocalizing animals at long ranges (on the order of thousands of meters in deep water [44]), but cannot provide information about non-vocalizing species. Active acoustic instruments, such as multibeam sonar, can detect targets tens of meters away without the use of light [28]. However, because of the relatively low-resolution of multibeam sonar images, species-level (or even genus-level) classification is frequently not possible without fusion with co-temporal data from other sensors, such as optical cameras [10].

Second, the characteristics of marine energy sites complicate the deployment, maintenance, and recovery of instrumentation. Strong waves or tidal currents limit time windows for marine operations. At wave energy sites, operations are limited to periods with relatively low wave activity, which may occur infrequently, especially during winter months [2]. At tidal energy sites, operations can be conducted during tidal exchanges when currents are at a minimum, but slack tide windows are brief. Once deployed, instrumentation systems can be subject to extreme forces from waves and currents, mandating the design of robust instrumentation platforms that can withstand these conditions [33]. High-energy conditions can also impede the interpretation of collected data. Bubbles and turbulence can produce high-intensity signals in active acoustic data [43], and flow noise can elevate sound levels at low frequencies in passive acoustic data when there is water motion relative to a hydrophone element [4].

Third, sound or light produced by sensors may bias the behavior of the animals being monitored. Artificial light (used to illuminate the fields of view of optical cameras) can either

attract or repel fish [39]. If active acoustic transducers produce sound within the range of animal hearing, it may cause harm to the animals in the form of temporary or permanent threshold shifts [45] or affect their behavior [27, 60].

Finally, if any interaction between marine fauna and a MEC occurs, it is likely to be a rare event. Therefore, data must be collected over long periods (on the order of months) to reliably capture these events. However, if monitoring sensors were continually operated for these time windows, they would quickly quickly accrue unmanageable volumes of data. For example, a single multibeam sonar can produce over a terabyte of data in a single month, creating obvious challenges for data storage, human review, and post-processing. As a result, existing data sets have been referred to as DRIP-y, or "data rich, information poor" [66].

These challenges can be distilled to three directives for an effective environmental monitoring platform:

1. Collect data without altering the environment through use of sensors (e.g., production of sound that biases animal behavior)

2. Capture rare events (e.g., presence of a marine mammal near a tidal turbine)

3. Minimize collection of data that contains no useful information

These directives are presented in order of priority. For example, if a platform is effective in capturing rare events, but does so in a way that alters the behavior of the fish or marine mammals being monitored, the collected data would not represent the environmental interactions that would occur if the monitoring instrumentation was not present. Similarly, collection of data from all sensors on a sparse duty cycle would minimize data volumes, but is likely to miss rare events of interest.

1.2 Integrated Instrumentation Platforms

Integrated instrumentation platforms that combine monitoring sensors into a single, streamlined package may be able to simultaneously meet all three of the directives outlined in the

previous section. When a variety of sensors are used in aggregate, they can overcome some of the challenges inherent to operation of ocean instrumentation. For example, if a target is detected concurrently by a multibeam sonar and optical camera, the target can be classified with high certainty at close range in the optical camera data, and its behavior can be monitored over a long range in the multibeam sonar data. In addition, deployment and recovery of integrated platforms is much more straightforward than deployment of individual sensors, reducing the challenges associated with conducting marine operations at marine energy sites.

Development of an integrated instrumentation platform that meets all three directives can be framed in a series of three generations. The first generation involves the development of the requisite hardware to deploy and power all sensors as a single package. A first generation platform may capture rare events if data are collected continuously (i.e., meeting directive 2), but is likely to accrue unmanageable volumes of data, and offers no mechanism to limit the use of instruments that might alter the environment (i.e., not meeting directives 1 and 3).

The second generation of development involves the development of software to centralize sensor control and data acquisition. This allows for synchronization of data streams and simplified control and configuration of sensors. Second generation development also provides a mechanism for collection of data on a predetermined duty cycle, which reduces data storage requirements and environmental impact, but is likely to miss rare events (i.e., meeting directives 1 and 3, but not directive 2). While there are advantages to second generation development over individual deployment of sensors, it typically requires discarding the manufacturer-provided sensor control and data acquisition software, and may require storage of data in formats that are not compatible with developer software for review and processing. In many cases, developer software licenses are a significant portion of the cost of an oceanographic instrument, and the costs of second generation integration may not outweigh the benefits.

Most integrated instrumentation platforms that have been deployed and tested to date are second generation platforms that either collect data continuously or on a duty cycle.

These platforms include the Fundy Advanced Sensor Technology (FAST) and FLOW and Benthic ECology (FLOWBEC) platforms. The FAST platform was developed for environmental monitoring at a tidal energy test site in the Bay of Fundy, Canada. FAST is a cabled platform (e.g., data are continuously streamed to shore) that includes active and passive acoustic sensors [1]. The FLOWBEC platform has been deployed at wave and tidal energy sites around the United Kingdom. FLOWBEC is an autonomous platform (i.e., powered by onboard batteries) that includes active acoustic sensors (a multibeam sonar, a multifrequency echosounder, and an acoustic Doppler velocimeter) and a fluorometer that measures chlorophyll and turbidity. Data from all FLOWBEC sensors are continuously recorded over a two-week deployment period (one spring-neap tidal cycle), after which the batteries must be recharged prior to redeployment [68]. Automatic target detection and tracking algorithms are used in post-processing to identify events of interest in the collected data [67].

First and second generation development pose significant challenges for both hardware and software development. However, this thesis is focused on third generation capabilities. A third generation system leverages the data streams available in real-time on an integrated instrumentation platform for automatic adaptive action. For example, real-time detection and classification of a school of fish can trigger data archival from any sensors that might be able to detect it. The use of sensors which may bias animal behavior can also be restricted to minimize their impact. For example, use of artificial illumination can be limited to periods when a target of interest is detected within the field-of-view of optical cameras on a multibeam sonar. In this way, a third generation system can simultaneously achieve all three directives for an effective monitoring platform. Prior to the research presented in this thesis, the third generation of integration had not been achieved by any monitoring platform.

Development of a third generation platform requires the ability to automatically detect and classify targets in the sensor data streams to inform automatic adaptive action and an understanding of how the monitoring sensors might bias the behavior of animals being monitored so that their environmental impact can be effectively minimized. These requirements are addressed in this thesis. Chapter 2 provides relevant background on the Adaptable

Monitoring Package (AMP), the integrated instrumentation platform used to demonstrate third generation capabilities. In Chapter 3, the capabilities of the AMP are assessed and the potential for third generation development of this system is evaluated. Chapter 4 details the development of real-time automatic target classification capabilities for a multibeam sonar integrated into the AMP. Finally, a comprehensive evaluation of acoustic emissions from active acoustic sensors used for environmental monitoring within the context of marine mammal hearing can be found in Chapter 5.

Chapter 2

BACKGROUND: THE ADAPTABLE MONITORING PACKAGE

The following chapters of this thesis detail the development of third generation capabilities for the Adaptable Monitoring Package (AMP), an integrated instrumentation platform that combines active acoustic, passive acoustic, and optical sensors for environmental monitoring at marine renewable energy sites. However, a basic understanding of the first and second generation development of the AMP platform is useful in interpreting results. This chapter provides an overview of the AMP platform, including the underlying hardware, the integrated sensors, and the core software that integrates these sensors.

2.1 The Adaptable Monitoring Package

The AMP refers to a set of electronics, sensors, and software (the AMP "backbone") that can be adapted to a variety of monitoring scenarios (see Figure 2.1). At the core of the AMP hardware is an integration hub that provides 48-volt, 24-volt, or 12-volt power to and enables communication with each sensor. The integration hub also enables real-time monitoring of the temperature, humidity, and current draw of each sensor. Because most oceanographic sensors use either serial or Ethernet communication, the integration hub accommodates both protocols. Integration of sensors that use serial communications is enabled by serial-to-Ethernet converters that convert data to Ethernet protocol. All data streams are converted from Ethernet to fiber media for transmission to shore. A second Ethernet/fiber-optic converter on shore converts data streams back to Ethernet protocol so that they can be read by a shore computer that handles data acquisition and control.

The core AMP software operates continuously on the shore computer. This software is

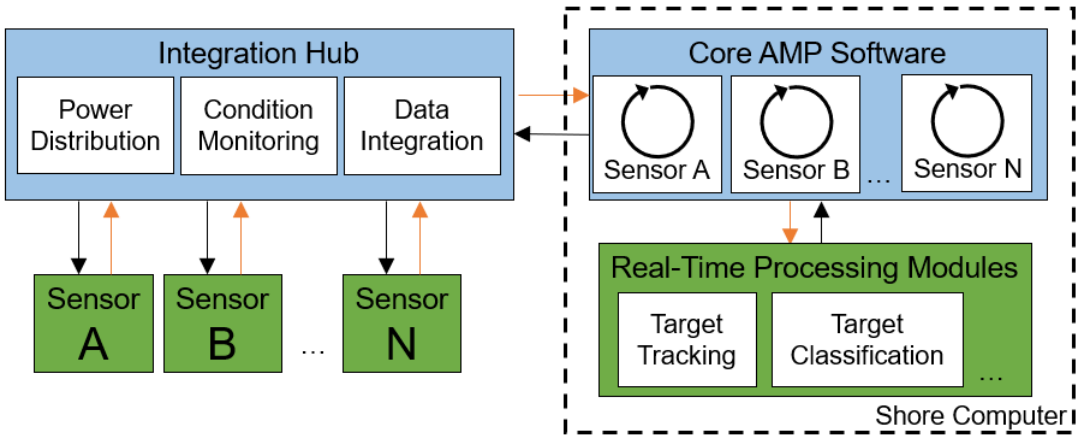

Figure 2.1: Diagram of overall AMP system architecture. Components in blue compose the AMP "backbone," while components in green are configured specifically for each deployment. Orange arrows indicate the flow of data, and black arrows indicate the flow of controls.

written in LabVIEW (National Instruments) and handles data acquisition and control of the integration hub and sensors. One central control program (LabVIEW "virtual instrument") handles communications with the integration hub - enabling power to individual sensors and monitoring current draw, temperature, and humidity. Each sensor is operated by an individual "virtual instrument" that handles configuration of that sensor (e.g., range of a multibeam sonar) and parses the sensor data stream. This modular software structure allows for relatively straightforward addition or removal of sensors.

Data from low-bandwidth sensors (e.g., current profiler) can be continually archived without accruing unmanageable volumes of data. Data from high-bandwidth sensors (e.g., optical camera or multibeam sonar) are stored in variable-length ring buffers (typically 45-60 seconds) in the computer's volatile memory. These ring buffers are written to disk when a command is generated either by a user, on a fixed duty cycle (e.g., archive one minute of data every two hours), or by an external real-time data processing module. Data archival commands from real-time processing modules are received over UDP. The ring buffer structure serves three purposes: 1) when an archival command is generated, contextual data

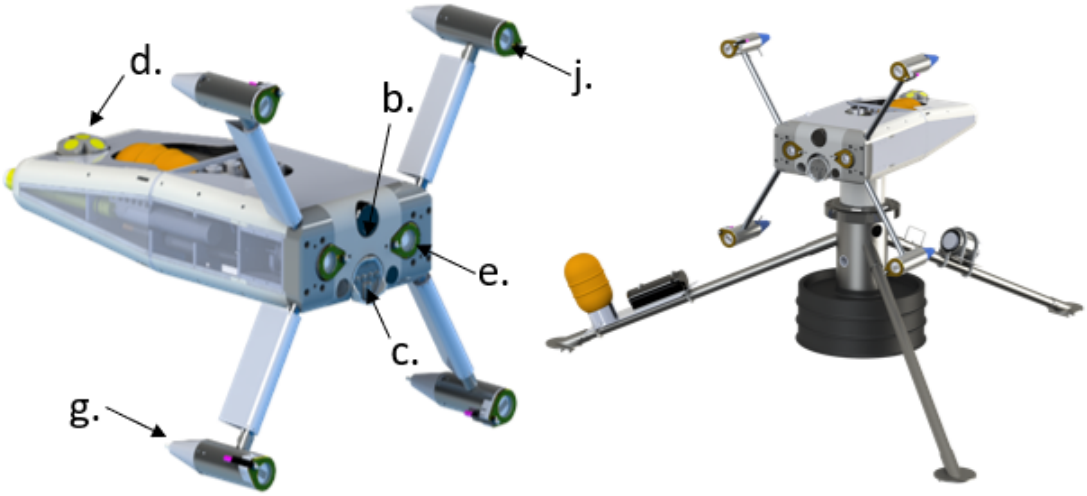

Figure 2.2: Renderings of the MSL-1 AMP. The AMP body is shown on the left, and AMP on its docking station is shown on the right. Instruments labels are referenced to Table 2.1. Renderings were created in SolidWorks (Dassault Systemes) by Paul Gibbs, University of Washington Applied Physics Laboratory.

before the event is available, 2) any real-time processing codes do not need to operate in true "real-time," as long as they operate at the same rate as data acquisition, and 3) data stored in memory can be used to establish background levels for target detection. Real-time processing modules are configured for a specific monitoring application.

The AMP backbone has been integrated into multiple hardware packages. Two distinct AMP systems were used to collect the data discussed in this thesis: the MSL-1 and MSL-2 AMPs. Both systems are named for their test site: the Pacific Northwest National Laboratory Marine Sciences Laboratory (MSL). Renderings of the MSL-1 and MSL-2 platforms can be found in Figure 2.2 and 2.3, respectively, and a comprehensive list of the sensors integrated into each platform can be found in Table 2.1. An overview of both platforms follows:

- **MSL-1**: The MSL-1 platform was the first version of the Adaptable Monitoring Package to be tested, and was first deployed 2017. All sensors were enclosed in a streamlined shell to minimize forces on the system, and the system is cabled to shore via a docking

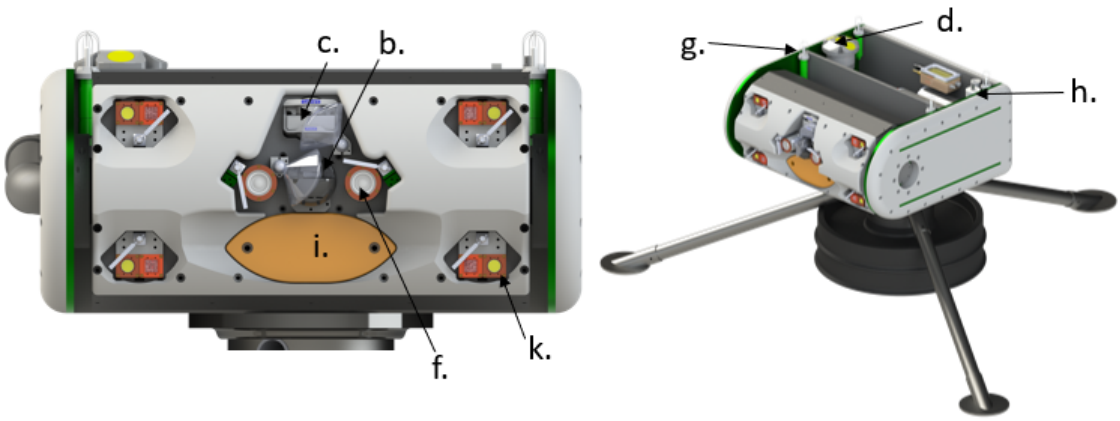

Figure 2.3: Renderings of the MSL-2 AMP. A zoomed-in view of the instrument head is shown on the left, and the entire platform as deployed is shown on the right. Instruments labels are referenced to Table 2.1. Renderings were created in SolidWorks (Dassault Systemes) by Paul Gibbs, University of Washington Applied Physics Laboratory.

station [55, 33]. Mechanical wipers (Zebra-Tech hydro-wipers) were used to prevent biofouling on the optical cameras and artificial illumination. During the 2017 deployment, a real-time processing module that detected and tracked targets in the multibeam sonar data was used to limit data acquisition to periods when targets were present. This system was used to establish the capabilities of the AMP backbone, evaluate the potential for third-generation development (see Chapter 3), and build a training data set for machine-learning classification of targets in multibeam sonar data (see Chapter 4).

- **MSL-2**: The MSL-2 AMP incorporated significant hardware and software upgrades from the MSL-1 AMP, and was first deployed in 2019. Major hardware upgrades from the MSL-1 AMP include the addition of a pan motor to rotate the instrument head and the addition or replacement of several sensors (see Table 2.1). The cross-sectional profile of the MSL-2 AMP shell was greatly reduced through re-design of the artificial illumination system, which removed the four struts that were used to separate strobe lights on the MSL-1 AMP. In addition to the Zebra-Tech Hydro-Wipers used

on the MSL-1 AMP, ultraviolet lights (AML UVC LED Lights) were used to prevent biofouling on the multibeam sonar and acoustic camera transducers. All active acoustic instruments, with the exception of the fisheries echosounder, were synchronized using a centralized hardware triggering system to prevent active acoustic crosstalk. During the 2019 deployment, a real-time processing module that classifies targets in the multibeam sonar data was used to control data acquisition and automatic adaptive action (see Chapter 4).

Ref.	Type	Instrument	Manfuacturer	AMP Systems	Acquisition Rate (Hz)
a.	Multibeam Sonar	M3	Kongsberg	MSL-1	5 Hz
b.	Acoustic Camera	M900-2250	Teledyne BlueView	MSL-1, MSL-2	5 Hz
c.	Multibeam Sonar	Gemini 720is	Tritech	MSL-2	10 Hz
d.	Acoustic Doppler Current Profiler	Signature 500	Nortek	MSL-1, MSL-2	1-8 Hz
e.	Optical Camera	Manta 307	Allied Vision	MSL-1	10 Hz
f.	Optical Camera	Manta 507	Allied Vision	MSL-2	10 Hz
g.	Hydrophone	icListen	OceanSonics	MSL-1, MSL-2	512000 Hz
h.	Light Scattering Meter	EcoBB	Seabird Scientific	MSL-2	1
i.	Echosounder	WBTMini with ES38-18/200-18C Transducer	Simrad	MSL-2	1 Hz at Maximum Range
j.	Xenon Strobe Lights	MVS-5000	Excelitas	MSL-1	n/a
k.	LED Strobe Lights	n/a	Custom	MSL-2	n/a

Table 2.1: Sensors included in AMP sensor suite for the MSL-1 and MSL-2 AMPs. Reference letters refer to Figures 2.2 and 2.3. For sensors with user-configurable acquisition rates, the acquisition rate used during AMP deployments is listed.

Chapter 3

BENCHMARKING SENSOR FUSION CAPABILITIES OF AN INTEGRATED INSTRUMENTATION SYSTEM

In this chapter, the feasibility of third generation integrated instrumentation is investigated and the AMP is used to collect a training data set for machine learning classification of targets in multibeam sonar data. The test described in this chapter was conducted using the MSL-1 AMP, which is described in detail in Chapter 2. For simplicity, this platform is referred to as "the AMP" throughout the chapter.

The contents of this chapter that follow were originally published in the *International Journal of Marine Energy* in September, 2017:

E. Cotter, P. Murphy, and B. Polagye. Benchmarking sensor fusion capabilities of an integrated instrumentation package. *Int. J. Mar. Energy*, 20:64−79, 2017.

This paper has been adapted for this thesis. The original publication includes a overview of integrated instrumentation systems, which is discussed in Chapter 1 of this thesis, an introduction to the AMP platform, which is discussed in Chapter 2 of this thesis, and a preliminary analysis of machine-learning classification of target tracks, which is significantly expanded upon in Chapter 4 of this thesis. To avoid redundancy, these sections have been removed here. Additional minor edits have been made for consistency with the terminology used in the rest of this thesis.

Paul Murphy led the evaluation of passive acoustic detection capabilities, with input from the author. Because vocalizing marine mammals were rare at the Sequim, WA test site, passive acoustic detection and classification are not investigated further in this dissertation.

3.1 Introduction

In this chapter, the AMP is used to evaluate the benefits of third generation integration. First, cooperative targets are used to benchmark individual sensor performance. Then, a real-time module is implemented that triggers data acquisition when "opportunistic" targets are detected and tracked in the multibeam sonar data streams. The effectiveness of this real-time module is evaluated by comparing these data to data acquired on a sparse duty cycle. Results confirm several of the hypothesized benefits of third generation integration, but also demonstrate the challenges to achieving this. Throughout this chapter, the following terms have specific meaning:

- *Detection*: the recognition of a target within the AMP field of view
- *Classification*: the assignment of a target to a particular class (e.g., diving bird)
- *Identification*: specific assignments within a class, potentially to the level of a species (e.g., identifying a diving bird as a *Cepphus Columba*).

3.2 Methods

3.2.1 Deployment Site and Configuration

The AMP was deployed at the Pacific Northwest National Laboratory Marine Sciences Laboratory in Sequim, WA from January to March 2017, with active acoustic and optical sensors oriented in a cross-channel direction (Figure 3.1). Water depth at the deployment site varied from 8-12 meters, depending on the stage of the tide, and peak currents were approximately 1.5 m/s. Harbor seals (*Phoca vitulina*) were commonly observed at the site. Historical trawls indicate as many as fifty species of fish present in the vicinity of Sequim Bay, the most abundant being Shiner Perch (*Cymatogaster aggregate*) (personal communication, Anna Kagley, National Marine Fisheries Service). Diving birds (primarily pigeon guillemots, *Cepphus columba*) were also commonly observed at the test site.

Throughout the deployment, data were collected on a duty cycle: 45 seconds of acquisition by the Kongsberg M3 multibeam sonar, Teledyne BlueView multibeam sonar, and OceanSonics icListen hydrophones every 15 minutes, and 15 seconds of optical camera data every 30 minutes (concurrent with other sensors). Strobe lights were only activated for 3 seconds during each data acquisition window to minimize potential for behavioral bias. In addition, a real-time module was used to trigger data acquisition when a potential target was automatically detected within 10 meter range of the AMP using the data streams from the multibeam sonars, as described in Section 3.2.4. Targets in these data (e.g., marine mammals, diving seabirds) are defined as opportunistic in the sense that their appearance and position could not be controlled. Because vocalizing marine mammals are not common in the area, opportunistic detection with the passive acoustic sensors was not possible.

3.2.2 Sensor Data Streams

Following acquisition by the LabVIEW core AMP software (see Chapter 2), data manipulation were performed by a real-time processing module implemented in MATLAB (Mathworks), except where otherwise specifically noted. For the evaluation described in this chapter, sensors and data acquisition were configured as described in the following sections.

Passive Acoustic Array

The four OceanSonics icListen HF hydrophones were connected through a pressure housing containing an Ethernet switch, power distribution, and common electrical connection between the unit designated as the master and those designated as slaves for synchronization. Data were acquired at 512,000 Hz, and stored in separate WAV files for each hydrophone.

Figure 3.1: AMP Test site in Sequim, WA.

Multibeam Sonars

The Kongsberg M3 ("M3") multibeam sonar was operated in "Imaging" mode[1], with a 15^o vertical swath, 120^o horizontal swath, and 50 meter horizontal range (see Figure 3.2). The multibeam sonars were synchronized via a hardware trigger, with the M3 acting as the "master." The data were stored as flat text files in polar format with 0.94^o beam spacing, 3 cm range resolution, and 32-bit intensity values with a minimum value of zero, and maximum values on the order of 100.

The Teledyne BlueView ("BlueView") M900-2250 multibeam sonar has two operating frequencies: 900 and 2250 kHz. The 2250 kHz transducer was used to provide higher resolution imagery to a range of 10 meters. The BlueView has a 20^o vertical swath and a 130^o horizontal swath (see Figure 3.2). Data were stored as 8-bit PNG images in Cartesian format, with a resolution of 2.5 cm/pixel and pixel intensity values ranging from 0-255. Five radial lines are continually visible in the BlueView image (see Figure 3.3), and are considered part of the image background. These lines result from the fact that the BlueView uses three separate transducers to create a single image.

Wave and Current Profiler

The Nortek Signature 500 ADCP collected one minute of data at 1 Hz immediately after data were acquired from the other sensors. The ADCP was configured to not collect data concurrently with the other AMP instruments to avoid interference with the M3, which also operated at 500 kHz (the ADCP does not have a hardware trigger when equipped with Ethernet communications) [13]. Data were reported with 0.5 meter bin sizes ranging from 0.5 meter above the instrument to the water surface.

---

[1]The "Enhanced Image Quality (EIQ)" mode (which provides higher image quality with a slower acquisition rate) was not utilized, due to measurements indicating that this mode produced the most sound within the range of marine mammal hearing [13].

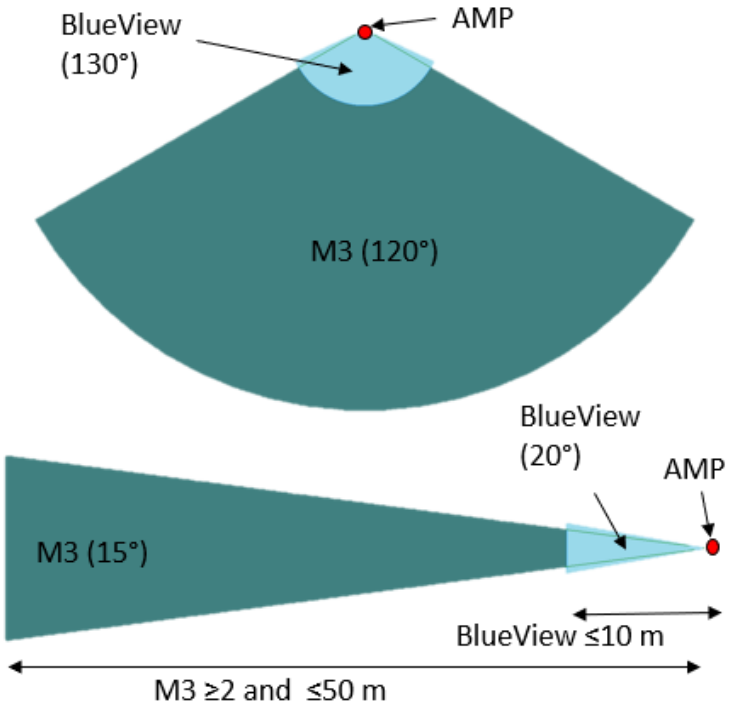

Figure 3.2: Fields of view for the two multibeam sonars, the Kongsberg M3 and the Teledyne BlueView. Top panel shows top view, and bottom panel shows side view.

Optical Camera

The camera system consisted of two Allied Vision G-304 Manta cameras and four Excelitas MVS-5000 strobe lights. Mechanical wipers (Zebra-Tech Hydro-Wipers) were used to prevent biofouling on optical ports during a deployment [34]. One optical camera was configured as master, and was used to trigger synchronous acquisition from the second camera with an overlapping field of view. Camera exposure was allowed to automatically adjust based on the available ambient light since this varied in time due to the relatively shallow water depth and fluctuations in turbidity. Data were stored as 16-bit greyscale PNG images.

3.2.3 Benchmarking of Sensor Capabilities

A variety of cooperative targets with known location and passive or active acoustic signatures were used to benchmark sensor detection capabilities and range. First, 14 cm diameter fishing floats were used as targets for the active acoustic instruments, which provided an acoustic return approximately equivalent to a medium sized swim-bladdered fish, such as a 14-16 cm Pacific Hake (personal communication, John Horne, University of Washington School of Aquatic and Fishery Sciences). These targets were suspended from a surface buoy and allowed to drift with currents through the field of view of both active acoustic sensors. These tests were also helpful to verify the AMP orientation on the seabed by comparing the target position in the multibeam sonar data with GPS estimates of surface buoy position [10]. Second, in a coarse approximation of a pinniped, a diver equipped with a rebreather (to minimize air bubbles) swam to a variety of predetermined distances and depths within the field of view.

To evaluate the detection capabilities of the hydrophone array, acoustic sources emitted sound at a variety of frequencies and ranges from the AMP. An OceanSonics icTalk HF acoustic projector was configured to produce narrow-band tones at seven frequencies (Table 3.1). Tones were one second in duration at each frequency and were separated by one second pauses. An additional six second pause was inserted at the end of the sequence resulting in a total sequence duration of 20 seconds, which was then repeated continuously. These tones were chosen to represent potential marine mammal vocalizations, such as whistles, that might be present at marine energy sites. Source levels were taken from manufacturer calibrations.

The acoustic projector was mounted at a depth of approximately 1 meter beneath a surface vessel, which was maneuvered to a range of up to one kilometer from the AMP to assess the upper limits of detectability as a function of range, source frequency, and hydrophone position on the AMP. A GPS logger (QStarz BT-Q10000eX) recorded vessel position.

Frequency (kHz)	**Source Level (dB re 1 μPa)**
10	126
20	140
40	143
60	145
80	144
100	144
120	145

Table 3.1: Acoustic projector source frequencies and levels. Source levels based on bandwidths of 6 Hz.

Hydrophone data were processed using the "Whistle and Moan" detector available in PAMGuard[2] to determine when, if ever, each tone was detected by each of the four AMP hydrophones. GPS data were linearly interpolated to determine the approximate vessel position at the time each sound was detected. Position estimates were slightly adjusted to account for transmission time assuming a direct path from source to receiver and uniform sound speed of 1500 m/s.

3.2.4 Target Detection And Tracking

Target detection and tracking were implemented for the multibeam sonar data streams. The same scheme was implemented in a real-time module and applied in post-processing. Target detection involves the discrimination of "targets" (objects in the foreground) from the image background. The foreground for each image was found by subtracting the median value for each pixel in a window preceding that frame. A 10 second window was used for the M3,

[2]PAMGuard is an open-source passive acoustic detection and classification software package (https://www.pamguard.org/).

and a 30 second window was used for the BlueView. Because the M3 images were higher bandwidth, computational limitations necessitated a shorter window to conduct background subtraction at the same speed as data acquisition.

A two-dimensional median filter was applied to each foreground frame to reduce noise in the image. Next, an intensity threshold was applied to each frame to create a binary image. For the M3, this threshold was varied with distance from the AMP from 1.5 to 3 (arbitrary units), such that sensitivity was higher at closest range, and decreased as noise became more prominent at greater range. A constant threshold of 20 (arbitrary units) was used to produce a binary image from the BlueView data. Any targets with overlapping bounding boxes were then merged to create a single target[3]. Objects in the binary foreground image larger than a threshold value (0.0125 m^2 and 0.06 m^2 on the BlueView and M3 data, respectively) were classified as candidate targets. These thresholds were selected based on the characteristics of targets identified through human review of acquired data.

A Kalman filter [41] was employed to track candidate targets based on the position of their centroid, under the assumption that true targets moved with constant velocity between each frame of data. This is reasonable, given that frames were acquired at an interval of 0.2 seconds by both sonars. The polar format M3 data were transformed to Cartesian coordinates, so that a linear Kalman filter could be used. An attempt was made to match each candidate target with an existing track by comparing detections in each frame with the locations of candidate targets predicted by the Kalman filter. Targets with centroid positions within 0.25 or 1.5 meters of the predicted location of a track (for the BlueView and M3 data, respectively) were associated with that track. In the case that there were multiple targets within this range, the closest target was used. Candidate targets that were not detected across at least five frames were discarded. A target track was declared to end after a specified time without a detection (0.3 or 1 seconds on the BlueView and M3, respectively). Like the detection thresholds, these values were determined by human review.

---

[3]This feature was only implemented in post-processing.

3.2.5 Triggered Data Acquisition

The target detection scheme described in the previous section was implemented in real-time to trigger data acquisition when sufficiently large and intense targets were present in the sonar field of view. Background subtraction was conducted in the core AMP software in LabVIEW, using the data already stored in ring buffers to calculate the image background. After background subtraction, foreground images were stored as temporary files and read sequentially by the MATLAB real-time module for target tracking. For testing and evaluation purposes, real-time target tracking was only conducted for targets within the first 10 meters of the M3 image (overlap with the BlueView image). A data acquisition trigger was relayed to the core AMP software over UDP whenever a target track longer than five frames (one second) with a 75^{th} percentile target area and mean intensity above predetermined thresholds was detected on either sonar. The core AMP software then triggered activation of strobe lights to illuminate any targets within the optical camera field of view for 3 seconds, and waited an additional 15 seconds to trigger data offload from all sensors. This ensured that the entire event was captured in the ring buffers. After a trigger was generated, a second trigger was not generated until at least one ring-buffer length (45 seconds) had elapsed to avoid overlap in the data, but still allow an event which lasted longer than 15 seconds to be recorded. 75^{th} percentile target area thresholds of 0.2 m^2 and 0.3 m^2, and mean intensity thresholds of 40 and 2, were used for the BlueView and M3 imagery, respectively. These thresholds were tuned to capture most diving seabirds, fish schools, and seals with an acceptable number of false detections, but were too high to detect individual fish or targets of other classes at certain orientations. Because the primary objective of real-time detection during this deployment was to obtain a sufficient number of target samples to train classification algorithms, this trade-off was deemed acceptable. The size, position, and average intensity of all detected targets were logged for later analysis.

The features used for automatic target detection (average target intensity and 75^{th} per-

[4]Sunrise/sunset times did not vary significantly over the eleven-day review period.

Feature(s)	**Units**	**Normalization Limits (min,max)**	**Description**
A_{75}, A_{50}, A_{mean}, A_{max}, A_{std}	m^2	A_{75}: (0,0.5), A_{50}: (0,0.5), A_{mean}: (0,0.5), A_{max}: (0,0.75) A_{std}: (0, 0.25)	the 75^{th} percentile (A_{75}), 50^{th} percentile (A_{50}), mean (A_{mean}), maximum (A_{max}) and standard deviation (A_{std}) of target area for all targets associated with a track
I_{75}, I_{50}, I_{mean}, I_{max}, I_{std}	Arbitrary	I_{75}: (0,150), I_{50}: (0,150), I_{mean}: (0,150), I_{max}: (0,255) I_std: (0, 75)	the 75^{th} percentile (I_{75}), 50^{th} percentile (I_{50}), mean (I_{mean}), maximum (I_{max}) and standard deviation (I_{std}) of intensity of targets associated with a track
Maj_{75}	m	(0,1.1)	the 75^{th} percentile major axis length for each target associated with a track
Min_{75}	m	(0,1.1)	the 75^{th} percentile major axis length for each target associated with a track
AR	m	(0,1)	the mean ratio of major and minor axis length for each target associated with a track
D	Number of frames	(5,30)	the duration of each track, represented as the number of frames in which it was detected
R	m	(0.2,10)	the mean range from the transducer for each track, in m
S	m/s	(0,1)	the average speed of the target in each track, calculated using an average of the velocity between every 4 target detections associated with a track
N_{50}, N_{90}	Number of targets	N_{50}: (0,5), N_{90}: (0,5)	the 50^{th} (N_{50}) and 90_{th} (N_{90}) percentile of the number of other tracked targets within a two-meter range of each target
dS	m/s	(-1,1)	the difference between the average track speed (S) and the current speed at the time of detection
C	m/s	(0,1.5)	the current speed (in m/s) at the time of target detection
T	hours	(0,12)	the time of target detection in absolute hours relative to midnight in local time (i.e., both 11:00 PM and 1:00 AM have a value of 1)[4].

Table 3.2: Features calculated for each target track

centile target area) were not sufficient to automatically classify targets. Therefore, a broader set of track features was developed to assess the feasibility of automatic classification in post-processing. These features are as summarized in Table 3.2.

3.2.6 Human Review and Classification

Human review was conducted for data acquired over an eleven-day period during which sensor settings and thresholds were held constant. Prior to this point in the deployment, automatic detection thresholds and processing codes were regularly adjusted to improve effectiveness. Human review was only conducted for the BlueView data, because the higher resolution allowed more accurate human classification of targets. The reviewer watched each recorded sequence with the results of the target detection and tracking algorithm overlaid, then classified each target track as a "seal", "school of fish", "diving bird", "bubble clouds", "false detection", or "other/unknown". "False detection" was used to classify any target track which was not a real target (e.g., high intensity regions resulting from high turbidity in the water or interference from other sensors). "Other/unknown" was reserved for targets that were obviously not a false detection, but also could not be classified with any confidence. Reflections resulting from high-intensity targets (i.e., high intensity regions appearing near a detected target that were not a part of the target itself, but artifacts of the sensor processing) were also classified as other. In addition to classifying each track, the reviewer indicated if there were any targets present which were not tracked and if there was an object obscuring the sonar field of view (e.g., a crab occluding the BlueView transducer). For all triggered data sequences, optical camera data were also reviewed and used to validate target track classification, when possible.

These results were used to benchmark the effectiveness of the detection algorithm. Data acquired on a duty cycle were used to estimate the false negative rate for the detection and tracking scheme (percentage of duty cycle data which contained targets for which a trigger was not generated), and the triggered acquisition data were used to determine the false positive rate (percentage of triggered data acquisition sequences which contained no clear

targets). In the event that a trigger was generated during a duty cycle acquisition, the event was considered to be a triggered acquisition, meaning that the remaining duty-cycle data set was a sparse representation of the negative case (periods when no trigger was generated). As discussed in Section 3.3.4, the chosen duty cycle captured relatively few targets of interest.

3.2.7 Duty Cycle Simulation

To evaluate the advantages of triggered data acquisition (third generation integration) over a fixed duty cycle (second generation integration), a series of duty cycles were simulated to determine what percentage of human-classified true targets could be captured without triggering. For each simulated duty cycle, a 45 second data acquisition window was used and the time interval between acquisition events was varied. For example, a 2.5% duty cycle acquired 45 seconds of data every 30 minutes. Each cycle (1-100%) was simulated 500 times with a randomly selected starting point within the first cycle period of the test period (i.e., a 2.5% duty cycle could start anywhere between 12:00 to 12:30 AM on the first day of the review period). A target track was considered to be "detected" if at least one second was contained within the duty cycle window (i.e., a human reviewer would be likely to correctly classify the target with this information). For each simulated duty cycle, the detection rate (percentage of all targets of interest in the automatically triggered acquisitions which would have been detected if operating on a duty cycle) and resulting data mortgage (volume of collected data containing no true targets) were calculated.

3.3 Results

3.3.1 AMP System Reliability

The AMP was fully operational for 94% of the 76-day deployment. The majority of system downtime is attributable to three causes: scheduled downtime for software upgrades (13% of downtime, or <1% of total test time), power outages at the test site (31% of downtime, or 2% of total test time), and software crashes, mainly related to transfer of data to long-

term storage (55% of downtime, or 3% of total test time). This indicates that integrated instrumentation systems should be able to perform with similar reliability to individual sensors.

3.3.2 Benchmarking of Sensor Capabilities

Within the range of the BlueView (10 meters), the drifting floats were used to compare appearance of the target and detection capabilities for the two sonars. Figure 3.3 shows two snapshots from the float drifting past the AMP. In case 1 (top), the target is clearly visible at the same location on both instruments. However, in the M3 image, the target appears larger and distorted (i.e., not circular, and larger than it appears in the BlueView image). In case 2 (bottom), the target is not visible as it passes through the intersection of the BlueView transducers, but is clearly visible in the M3 image.

The drifting float could not be used to assess the effective range of the M3 for two reasons: it was difficult to maneuver the float to the correct position, and, after the beam swath intersected the water surface, it was not possible to distinguish between the surface expression and the sub-surface target. Human divers proved a more effective cooperative target for determining the effective range of the sonar because they could easily maneuver to predetermined locations without any surface expression. Figure 3.4 shows detection of the same diver at approximately 5 meters (left) and 25 meters (right) range from the transducer. As expected, the target is clear at close range but is difficult to distinguish from the background at a range of 25 meters. The maximum intensity value of the intensified area resulting from the diver is 64 (arbitrary units) at a range of 5 meters and 13 at a range of 25 meters. Because the diver and oxygen tank represent a larger and more intense target than any fish or marine mammal present at the test site, target detection beyond a range of 25 meters would not likely be possible for the M3 data at this site.

Passive acoustic detection rates varied significantly as a function of range and source frequency (three examples are shown in Figure 3.5). Detection rates and ranges were similar at 20, 40, 60, 80, and 100 kHz. Within the fields-of-view of the active acoustic sensors,

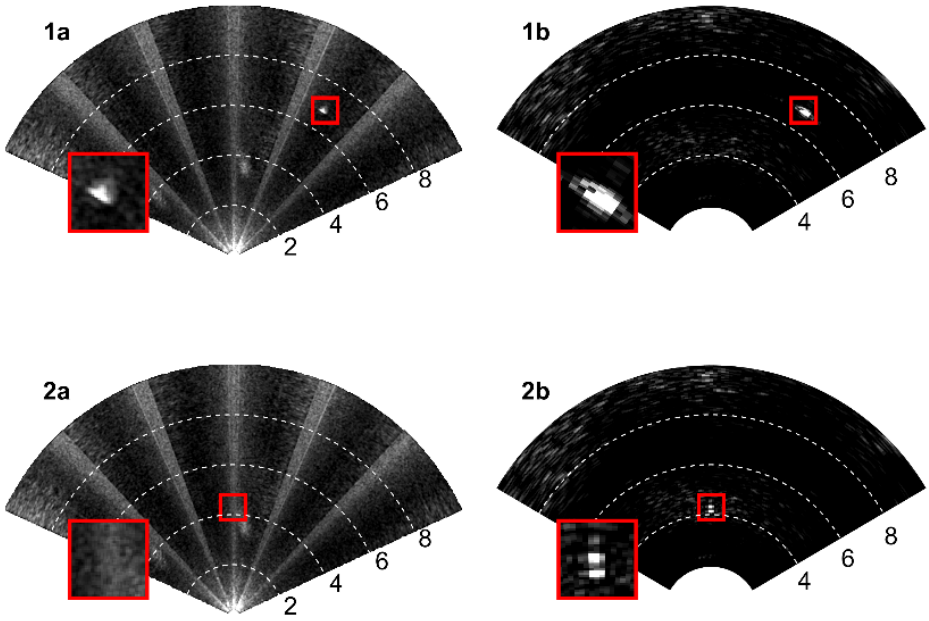

Figure 3.3: Two cases of simultaneous visualization of a fishing float by the BlueView (left) and M3 (right). In case 1 (top), the float is clearly visible at the same location on both sensors. In case 2 (bottom), the float is masked in the BlueView data. A 1x1 meter area around the float is magnified at the same position for both cases.

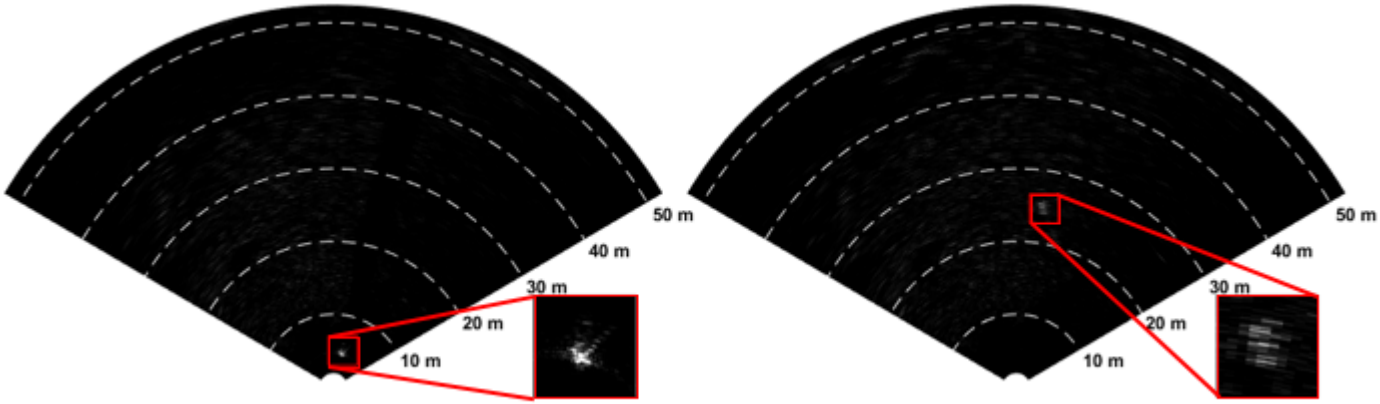

Figure 3.4: Comparison of the appearance of a diver in the M3 data at a range of approximately 5 meters and a range of approximately 25 meters. A 4x4 meter area around the target in each frame is magnified.

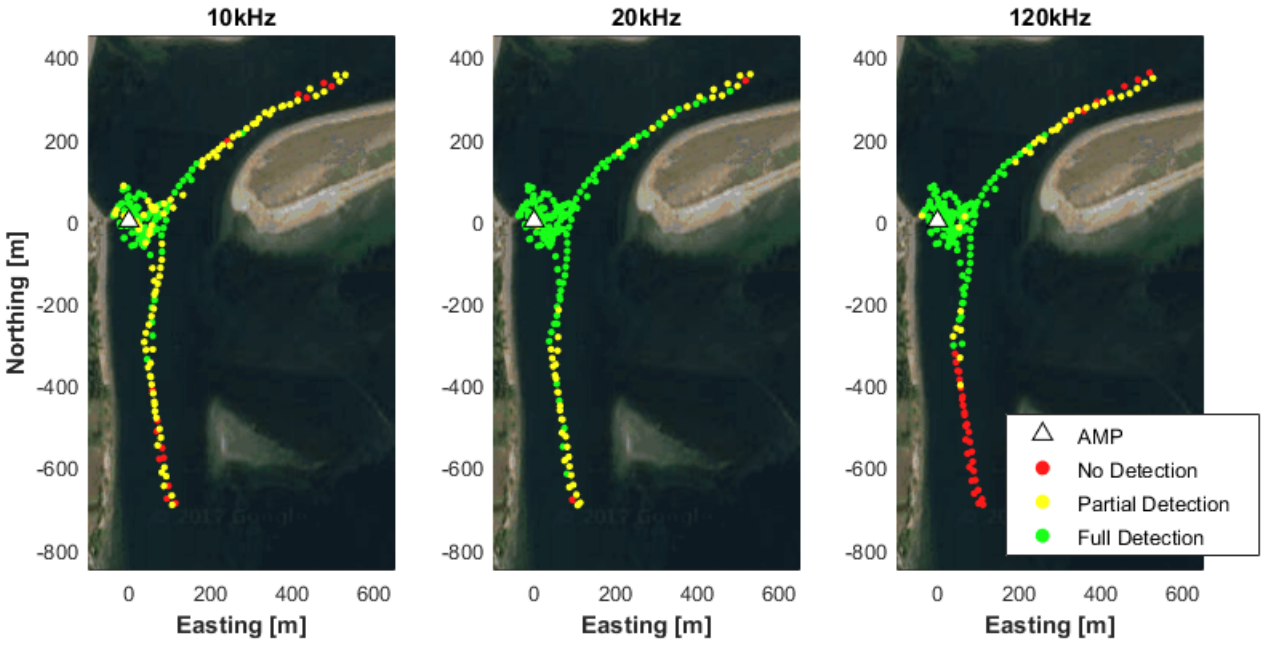

Figure 3.5: Passive acoustic detections for three frequencies produced by the acoustic projector (10, 60, and 120 kHz). "Full detection" indicates that all four AMP hydrophones detected the tone, and "partial detection" indicates that some, but not all, hydrophones detected the tone.

detection was reliable for all tested tones. The 10 kHz tone was more challenging to detect as a consequence of its relatively lower source level (Table 3.1) and relatively higher ambient noise levels. Poor detection of the 120 kHz tone may be a result of the relatively higher attenuation of sound by absorption at higher frequencies.

3.3.3 Target Detection and Tracking

During the human review period (eleven days), 989 duty cycle sequences and 479 triggered sequences were acquired. Figure 3.6 shows the distribution of targets for a representative day of reviewed data superimposed on the tidal current. Several trends are evident, which were consistent throughout the reviewed data: detections of all targets were most common during slack or near-slack tides, detections of diving birds were limited to daylight hours,

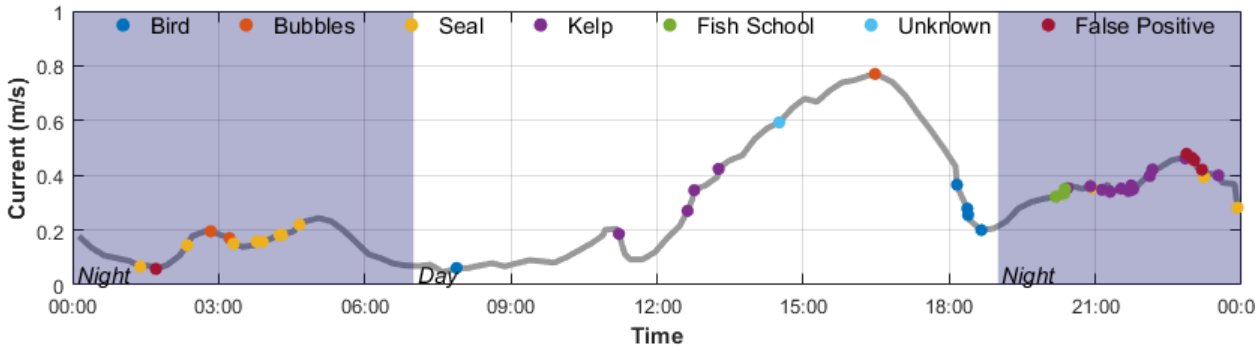

Figure 3.6: Distribution of triggered data offload events for a representative day of data, superimposed on the tidal current speed measured 1 meter above the AMP. Several trends can be observed which were consistent through all reviewed data: detections of birds are limited to daylight hours, detections of seals are limited to nighttime hours, and nearly all detections are during slack or near-slack tidal conditions.

and detections of seals were limited to nighttime hours. Figure 3.7 shows the probability distribution for a representative selection of the track features detailed in Table 3.2 for each target class.

Table 3.3 shows the distribution of target classes for which a trigger was generated, as well as the percentage of triggers generated by each sonar. Targets were considered to be"co-registered" if target tracks on both sonar exceeded trigger thresholds within five seconds of the generated trigger, because multiple triggers were not generated for targets detected within a single 45 second time window. 58% of triggered data offloads were triggered by a "real" target (seal, bird, kelp, cloud of bubbles, fish school, or other/unknown). Of the 42% of triggered data events that were caused by false targets, 45% were the result of an object (likely a piece of kelp or a crab) occluding the BlueView transducer for an extended period. Human review determined that the automatic detection algorithm missed nine targets of interest in the duty cycle data (i.e., 1% of duty cycle sequences contained a target of interest). A discussion of the characteristics of each class of target track follows.

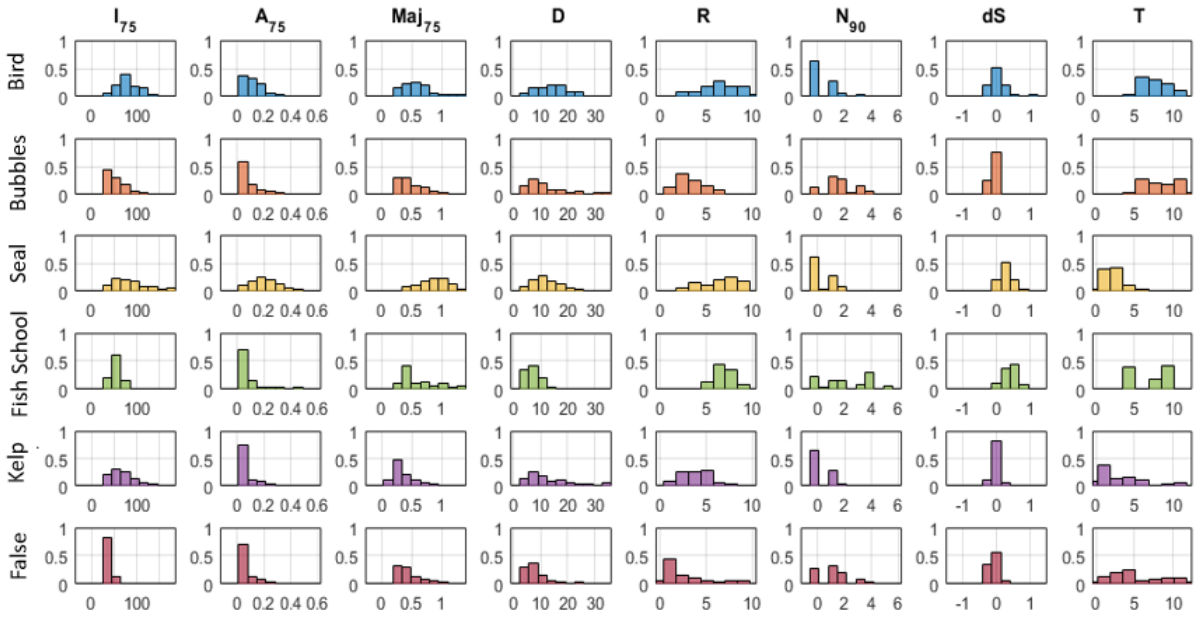

Figure 3.7: Probability distribution for a subset of track features: 75^{th} percentile intensity (I_{75}), 75^{th} percentile target area (A_{75}), 75^{th} percentile major axis length (Maj_{75}), track duration (D), mean track range (R), 90^{th} percentile of targets within a range of 2 meters (N_{90}), difference in speed relative to tidal current (dS), and time of day (T).

Target type	**Number of triggers**	**% of total**	**% co-registered**	**% detected only in acoustic camera data**	**% detected only in multibeam sonar data**
Seal	95	20%	13%	49%	13%
Bird	66	14%	2%	30%	64%
Kelp/Seagrass	52	11%	2%	50%	35%
Bubbles	42	9%	0%	95%	0%
Fish School	11	2%	0%	91%	0%
Other/Unknown	14	3%	7%	21%	50%
False Detection	199	42%	<1%	86%	1%
Total	**479**	**100%**			

Table 3.3: Breakdown of target classes in triggered data acquisition

Seals

Seals were responsible for the most automatic triggers, and were characterized as being relatively large and having a high intensity. When seals were detected swimming horizontally through the field of view, they presented as long, elliptical targets, and the hind flippers were often clearly distinguishable in BlueView imagery. When a seal was diving through the field of view vertically, it presented as a smaller target. Because seals passed through the field of view from many angles, the size of targets is broadly distributed (A_{75} in Figure 3.7). Seals were typically moving faster than the current speeds (dS), and not detected concurrently with a large number of other targets (N_{90}). All detections of seals were at slack or near-slack tide and during night-time hours. Most detections of seals were in the BlueView data, though in some cases seals appeared as more intense targets in the M3 data.

Birds

Diving birds were also responsible for a relatively large number of automatic triggers. Birds produced smaller targets than seals (A_{75}), but with similar intensity (I_{75}). Because of their diving trajectory, most birds were detected with minimal motion in the horizontal plane. However, on several occasions, birds were observed to swim horizontally, possibly while foraging. Based on optical camera data, birds were detected while diving and resurfacing, though it was not possible to distinguish between these two trajectories using the BlueView data. All detections of birds were at slack or near-slack tide and during day-time hours. A larger number of both birds and seals were detected at longer range from the instrument, because the conical shape of the BlueView swath ensonified a larger area at longer range (higher probability of a target being present in the sonar field of view). Most detections of birds were in the M3 data, because birds presented as larger and more intense targets in the M3 data relative to the BlueView data (similar to the cooperative target in Figure 3.3, case 1).

Bubbles

Clouds of bubbles were frequently detected and likely originate from a diving bird or seal (in both cases, air entrained from the surface during a dive). In many cases, a cloud of bubbles was detected when there was no other target observed, likely because a bird or seal was diving outside of the field of view. When many small bubbles presented acoustically as a single large target, they were sufficient to trigger data offload. However, they were not commonly detected beyond a range of approximately 5 meters due to reduced intensity caused by beam spreading (see distribution of R in Figure 3.7). Bubbles were most commonly detected during day-time hours (likely resulting from diving birds), but occasionally detected during night time hours (likely resulting from diving seals). In cases where a diving bird and the resulting cloud of bubbles were detected as a single track, the entire track was classified as a bird. Triggered events containing bubbles were exclusively generated by the BlueView

data, because the small size of bubbles made them appear as lower intensity (or not visible) targets in the lower resolution M3 data.

Kelp and Eelgrass

The appearance of kelp and eelgrass varied depending on the range and size of the target. Single pieces of kelp or eelgrass floating vertically in the water column at close range produced high-intensity targets in the data from both multibeam sonars. It should be noted that in cases without optical verification, human reviewers may have misclassified some of these targets as small individual fish. Detections of large pieces of kelp were typically lower intensity, but had a larger target area. Large pieces of kelp could be distinguished from seals by their slower velocity (dS) and variable shape as they moved with the currents.

Fish Schools

Fish schools were not common in triggered data sequences. Typically, triggers originated from a portion of the fish school detected as a single, large target, which was above area and intensity thresholds, and individual fish on the edges of the school were detected as multiple, smaller targets. All fish that were visually associated with the school were classified as a part of the school during human review. When multiple target tracks were associated with a fish school, they produced a longer tail in the N_{90} track feature, relative to other target classes (Figure 3.7). Fish schools were observed to scatter after activation of strobe lights, and were detected during both night time and day time hours. All fish school triggers were generated from the BlueView data.

False Positives (False Targets)

Three classes of false targets were identified: targets resulting from organic debris in the water column (typically at close range), targets resulting from electrical interference produced by the strobe lights (typically around a range of 8 meters, and too small to generate a

trigger), and targets resulting from an object occluding the sonar transducers (typically observed within a range of 5 meters). These false detections occurred almost exclusively in the BlueView data, because the higher resolution and smaller minimum area threshold allowed for detections of clouds of organic debris in the water. Additionally, the open nose-cone of the BlueView allowed objects to intermittently snarl on the transducer.

Missed Targets

Three types of false negatives were identified. The first occurred when a target was only partially ensonified (i.e., not entirely contained within the instrument swath), producing an atypically low intensity. The second occurred when schools of fish were not aggregated closely enough to be detected as a single target, and were detected as individual fish, with no target large enough to generate a trigger. The third case occurred when a target was not effectively tracked, either due to an error with the target tracking code (i.e., tracking lagged behind data acquisition and did not process every ping) or rapid changes in target direction inconsistent with the fundamental assumption of the Kalman filter.

3.3.4 Duty Cycle Simulation

Figure 3.8 shows the results of duty cycle simulation for cycles ranging from 1-100%. The presented detection rate includes all classes of true targets (seals, birds, bubble clouds, fish schools, and pieces of kelp/seagrass). Detection rates varied with the starting time for each cycle, so the standard deviation (darker grey) and maximum/minimum (lighter grey) detection rate calculated for each cycle are indicated.

As an example, the 5% duty cycle used during this deployment (45 seconds of data every 15 minutes) would be expected to capture, on average, 6% of targets of interest in triggered sequences, with minimum and maximum values of 3% and 9%, respectively. This cycle would be expected to accrue 170 GB of empty data each day, and, on average, 1.6% of sequences would contain a target of interest. For comparison, triggered data acquisition with

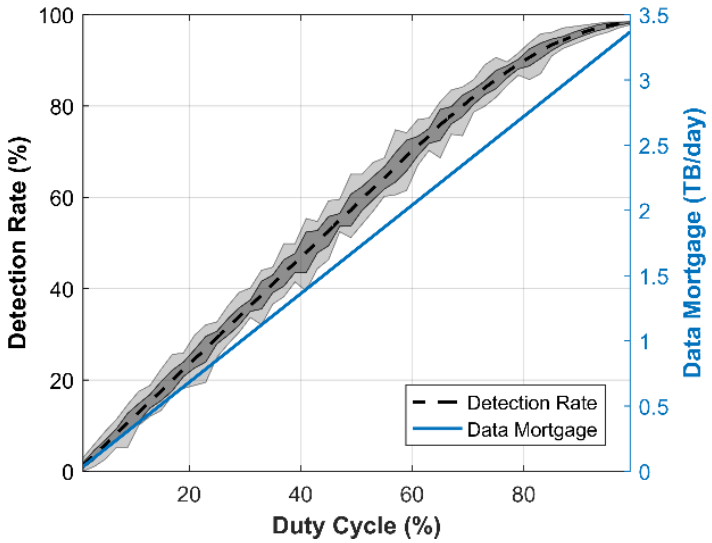

Figure 3.8: Mean detection rate for all simulated data cycles, with standard deviation (dark grey) and maximum/minimum (light grey) values indicated. The resulting data mortgage for each cycle (volume of data containing no useful targets) is indicated in blue. The detection rate does not exactly reach 100% for a 100% cycle: this is an artifact of simulation resulting from targets with a persistence of approximately one second that were split between adjacent "windows" and, therefore, not counted in either.

a 58% true positive rate accrued, on average, 28 GB/day of empty data, while capturing an estimated 99% of targets.

3.4 Discussion

3.4.1 Effectiveness of Cooperative Targets

While cooperative targets were critical for assessing sensor functionality and range, they were not used to train active acoustic automatic detection algorithms because the cooperative targets did not provide a sufficiently accurate representation of the real targets present at the site (e.g., seals, birds). Similarly, cooperative targets would likely not be effective for quantifying true positive of false negative rates for automatic classification algorithms. For passive acoustic monitoring, cooperative targets were essential in assessing target detection

capabilities, because the expected targets (fish tags and marine mammal vocalizations) could be more easily mimicked and these targets were not naturally present at the test site.

3.4.2 Quantifying System Effectiveness

This study quantified true positive and false negative rates for a real-time module that acquired data whenever a target of sufficient size and intensity was detected and tracked. This helped in interpreting the statistical significance of observations made around marine energy projects. However, there will be a large uncertainty in estimates of false negative rates unless 100% of data are collected and reviewed, which is not feasible for almost any integrated system. Calculation of true positive rates is more straightforward, because only data identified by automatic detection and/or classification algorithms must be reviewed to quantify effectiveness. This asymmetric effort mismatch is analogous to the philosophical difficulty of proving a negative to retire certain environmental risks.

3.4.3 Third Generation System Feasibility and Benefits

Results from the AMP endurance test demonstrate the feasibility and advantages of third generation integrated instrumentation. First, human classification of targets benefited from multiple, synchronous data streams. Specifically, optical camera data were often critical in interpreting the sonar data. Figure 3.9 shows detections of a small fish and a diving seal within 1 meter of the AMP. The targets appear similar in size and intensity in the BlueView data, and would likely not have been correctly classified without concurrent acquisition of the optical camera data. Similarly, clouds of bubbles were initially misclassified during human review as schools of fish or unknown targets before bubbles were observed on the optical camera. In addition, because simulated marine mammal vocalizations were detectable within the range of the sonar, it is expected that passive acoustic data would improve target classification or species-level identification at sites where vocalizing marine mammals are present.

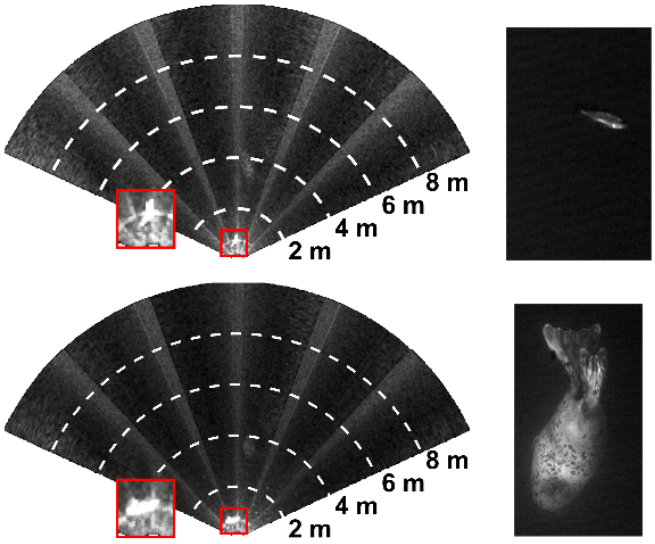

Figure 3.9: Comparison of a small fish (top) and diving seal (bottom) detected a range of approximately 1 meter. A 1x1 meter area around the target is magnified for both cases. Without optical camera data, the targets are difficult to distinguish at this range.

Second, real-time target detection in sonar data streams enabled the collection of a data set that can be used to train automatic classification algorithms. Based on the simulated duty cycle analysis, data collection on a 5% duty cycle would only have been expected to generate 6% of the target-containing sequences that were acquired using triggered acquisition. This reduced training data set would likely not have been sufficient to train classifiers with the performance demonstrated here. Further, human reviewers would have been required to sift through a large volume of empty data to detect and classify the targets.

The features described in Table 3.2 were descriptive of the characteristics of the observed target classes, as shown in Figure 3.7. However, classification of target tracks using manually-tuned thresholds using these features would be highly subjective and likely achieve relatively low true positive rates. This motivates the application of machine learning to the classification task.

Overall, automatic detection, followed by human classification, enabled intuitive representation of biological trends at a marine energy site (e.g., Figure 3.6). This can allow

regulators to more rapidly evaluate risk (e.g., are seals likely to be present during periods of strong tidal currents when collision with a turbine would be most severe?) or identify trends over time (e.g., are fish schools more prevalent as an artificial reef forms around a WEC foundation?). Additionally, automatic classification of detected targets could further streamline the process of extracting biological insight from these data streams in the future. Such insight generally requires third generation integration, as first generation systems are inefficient at collecting the synchronous data streams necessary for human classification of training data and second generation systems are likely to produce an unmanageable data mortgage in pursuit of training data for automatic post-processing. Consequently, third generation integration is likely necessary to automate monitoring programs that can cost-effectively address perennial challenges to marine energy development, such as the risk of collision. However, all third generation systems may not physically resemble the AMP discussed in this study. For example, a third generation system could consist of multiple sensor nodes distributed around a marine energy project and networked back to a single control and classification system.

3.5 Conclusions

The initial deployment of the Adaptable Monitoring Package demonstrated many benefits of third generation integrated instrumentation packages for environmental monitoring at marine renewable energy sites. Triggered data acquisition is shown to substantially improve the probability of detecting rare targets while decreasing the volume of low-value data. This has the potential to reverse the "DRIP" (data-rich, information-poor) paradigm in marine energy environmental monitoring by simultaneously reducing data volumes and reliably detecting important events [66]. In addition, it was demonstrated real-time target detection and tracking can be used to limit the use of sensors that might bias animal behavior. This suggests third generation systems like the AMP can satisfy all three directives for effective marine energy environmental monitoring.

Chapter 4

AUTOMATIC CLASSIFICATION OF FAUNA IN A TIDAL CHANNEL USING A MULTIBEAM SONAR

In Chapter 3, the benefits and capabilities of a third-generation integration instrumentation system were demonstrated and a real-time module that detected and tracked targets in multibeam sonar data was used to collect a training data set suitable for machine learning classification. In this chapter, that training data set is used to develop automatic classification of targets in multibeam sonar data. Automatic classification is then implemented in a real-time processing module and used to control sensors and data acquisition, furthering the third generation capabilities developed in Chapter 3.

Three subsets of training data are referred to in this chapter: 2017, 2019a, and 2019b. The 2017 data set refers to the data described in Chapter 3, which was collected using the MSL-1 AMP. The 2019a and 2019b data sets are first introduced in this chapter, and were collected using the MSL-2 AMP.

The contents of this chapter that follow were are currently under review for publication by the *Journal of Atmospheric and Oceanic Technology*:

E. Cotter and B. Polagye. Automatic Classification of Biological Targets in a Tidal Channel using a Multibeam Sonar. *Submitted to the Journal of Atmospheric and Oceanic Technology.* 2019.

Portions of the introduction have been removed to avoid redundancy with Chapter 1, and minor edits have been made for consistency with the rest of this thesis. Most notably, the discussion of the real-time implementation of automatic classification has been restructured using the terminology in Chapter 2.

4.1 Introduction

In Chapter 3, target size and intensity thresholds were used predict the presence of potential targets of interest in multibeam sonar data. Previous approaches for processing of multibeam sonar data at marine renewable energy sites have generally used similar methods, though target presence had not previously been used to control an instrumentation platform in real-time. In [67], a multibeam sonar was used to monitor a tidal turbine structure. Data were continually recorded for a period of two weeks, and targets in the data set were automatically tracked in post-processing using a nearest-neighbor algorithm. These target tracks were parameterized by morphometric measurements (e.g., size, intensity) and behavior (e.g., velocity, direction of travel) to aid in manual classification. Similarly, in [32], targets were detected and tracked using a nearest-neighbors approach, then classified based on manually-tuned target size thresholds. In addition, targets moving with a constant velocity were ignored under the assumption that they were passive objects advected by tidal currents. In [38], targets were automatically tracked in post-processing using a Kalman filter, and these target tracks were used to limit human review to periods when targets were present. [21] proposed a simple method for classifying targets in multibeam sonar imagery collected at a wave energy test site based on the length of a detected target in a single sonar image. Target length thresholds were used as the sole parameter to distinguish between marine mammal species (e.g., grey seals and harbor seals), but the accuracy of the method was not evaluated (e.g., true positive or false negative rates). Additionally, all targets were assumed to be fish or marine mammals, and the presence of non-biological targets that might have similar appearance in multibeam sonar data (e.g., kelp) [12] was not explicitly considered.

However, because the size, shape, and intensity of targets in multibeam sonar data vary with their position and orientation relative to the sonar [21], manually-tuned thresholds are subjective and unlikely to be sufficient for automatic classification. A more objective alternative is target classification using machine learning. While the author is unaware of any prior application of machine learning to classification of fauna in multibeam sonar data,

multiple data sets collected at marine energy sites have been proposed as training data for future machine learning models [67, 38]. Machine learning has also been applied to similar multibeam sonar classification problems. For example, in [16], a two-layer neural network was used to classify naval mines in multibeam sonar data using hand-engineered features that described the size and intensity of detected targets. The neural network was trained using the subset of features that resulted in an optimal classification rate and was able to classify mines with the same probability as an expert human reviewer. Machine learning has also been applied to multibeam sonar data for benthic habitat mapping. In both [26] and [30], it was found that a model using a random forest algorithm could classify seafloor substrates and biota (e.g., algae) with greater than 80% accuracy using hand-engineered features that described the intensity and texture of the image. In addition, machine learning has been used for automatic classification of birds in radar data collected at wind energy sites [53]. Radar imagery is similar to multibeam sonar in that it can be used to detect targets at longer ranges than optical cameras and does not require light, but the images are relatively low resolution. In [53], six machine learning algorithms were evaluated for classification of target tracks in radar data. A random forest algorithm performed best, achieving true positive rates over 90% for discrimination between bird and non-bird targets, and true positive rates between 81 and 83% for discrimination between specific groups of birds (e.g., herons and swallows). The classification models were, however, likely site-specific, as they were trained using data from only one site.

In this chapter, three machine learning algorithms are evaluated for automatic classification of marine fauna in multibeam sonar data. Manually reviewed target tracks (described in detail in Chapter 3) serve as training data, and two feature selection methods are evaluated for improvements to classification model performance. Classification models are evaluated using relevant metrics for environmental monitoring applications. Following this, classification performance is evaluated with varying volumes of training data to understand the requirements for adding new classes of target tracks to an existing model. Finally, classification is implemented in real-time at a tidal energy test site. Real-time performance is

presented and informs recommendations for implementation at new marine energy sites.

4.2 Classification Algorithms

There are a wide variety of machine learning algorithms for automatic classification. For our application of classifying fauna at a marine energy site, an algorithm must satisfy three requirements. First, the algorithm must be able to discriminate between more than two types of classes (multi-class). Multi-class classification can, however, be achieved for a binary (i.e., limited to two classes) classification algorithm using a combination of classification models specific to each class [17]. Second, the algorithm must perform well with a relatively small training data set, because some targets of interest (e.g., marine mammals) are relatively rare and annotation of training data is time intensive. Third, the algorithm must be able to predict the classification of a new target track relatively quickly to be suitable for real-time classification. Based on these requirements and prior investigations, three supervised machine learning algorithms were selected for evaluation: k-nearest neighbors, random forests, and support vector machines. Advantages and disadvantages of each algorithm, as they pertain to our application, are summarized in the proceeding sections.

4.2.1 k-Nearest Neighbors

A nearest neighbors algorithm is the simplest method of automatic classification. An unknown sample is assigned the class of the closest sample in the training data set: its "nearest neighbor". K-nearest neighbors expands on this approach by taking a "vote" of the k nearest neighbors, where k is a user-specified integer. These "votes" can be weighted by the distance of each neighbor from the unknown sample, making the algorithm relatively robust to outliers in the training data [18]. However, this algorithm can be sensitive to the value of the parameter k [70], and the computational requirement for classification of a new sample increases with the size of the training data and the value of k. A k-nearest neighbors algorithm can be used for classification of any number of classes.

4.2.2 Support Vector Machines

The support vector machine algorithm finds the hyperplane that best divides a data set into two classes [7]. A hyperplane is an n-dimensional division of the data set, where n is the number of features describing each point in the data set. For example, when $n = 2$, the hyperplane is the line that best divides the data into two classes (each class is on one side of the line). The support vector machine algorithm is a kernel method, meaning that it uses a kernel function to map data into a higher-dimension feature space to find the optimal hyperplane when a linear separation would not perform well [7]. While they have moderate computational requirements for training, support vector machines can classify a new data point with low computational cost. Support vector machines are inherently binary classification models, though there are several methods in the literature for achieving multi-class classification [17].

4.2.3 Random Forests

A random forest is comprised of multiple decision trees, a classification method that divides data into a set of "nodes," or decision points [5]. Single decision trees can perform well for classification, but are prone to over-fitting (i.e., producing a model that is too specific to the training data and performs poorly for data that was not contained in the training data set). However, when many decision trees are combined in a random forest, the method is more robust. Random forests have been shown to perform well with small training data sets [3], and can be used for any number of classes.

4.3 Methods

4.3.1 Training Data

The manually reviewed multibeam sonar target tracks described in Chapter 3 were used as training data to evaluate classification models in post-processing. Salient details of the data collection and human review are repeated here. Data were collected using the Teledyne

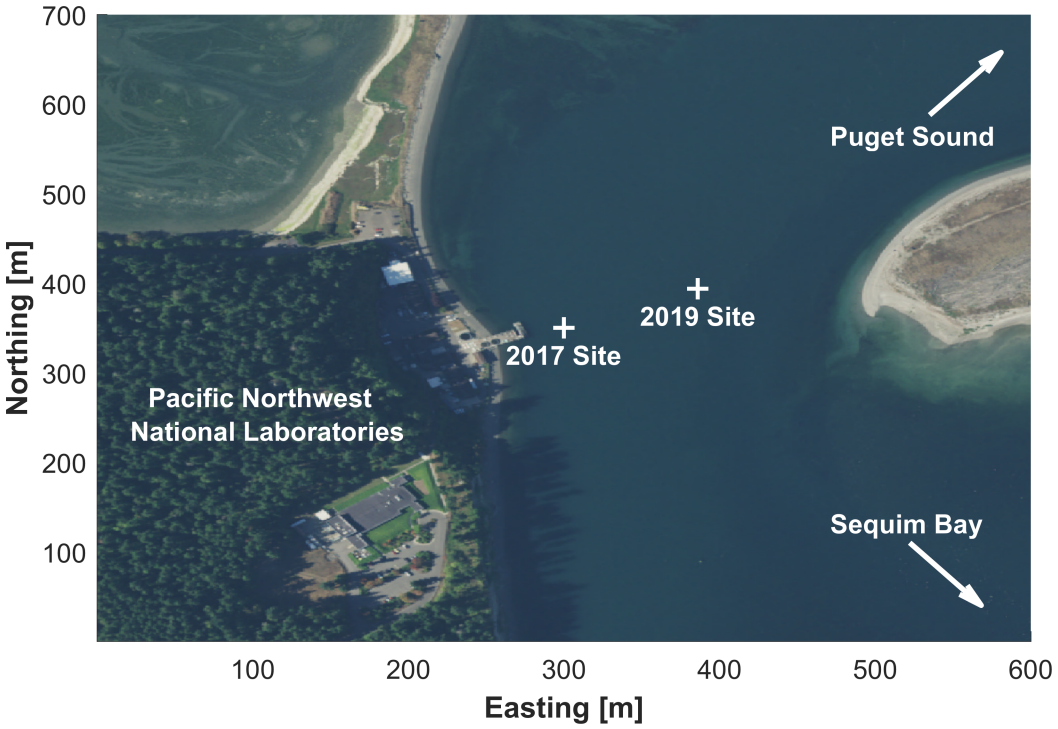

Figure 4.1: Sequim Bay test site. The deployment sites where the 2017 and 2019 data sets were collected are indicated.

BlueView M900-2250 multibeam sonar (2250 kHz) integrated into the Adaptable Monitoring Package. The multibeam sonar was positioned to look across a narrow tidal channel in Sequim Bay, WA (see Figure 4.1). At the deployment location, the water was approximately 10 meters deep and experienced peak tidal currents around 1.5 m/s. Targets were detected in the foreground of each multibeam sonar image and tracked using a Kalman filter. Target tracks were then classified by a human reviewer (the corresponding author of this study). Here, these targets are separated into five classes: non-biological targets (N), seals (S), diving birds (B), fish schools (F), and "small targets" (Sm). The latter class consisted of drifting kelp/eelgrass or individual fish which could not be reliably differentiated due to sonar resolution. The non-biological target class includes all target tracks that were not associated with plants or animals: sonar artifacts (e.g., from side lobes of the transducer [63]), bubbles, and turbulent backscatter. Reviewer annotations were verified with concurrent

Class	Abbv.	Number of Tracks		
		2017	2019a	2019b
Non-biological	N	2631	947	1988
Seals	S	147	46	11
Small Targets	Sm	302	60	135
Birds	B	81	6	34
Fish Schools	F	77	59	85

Table 4.1: Number of target tracks belonging to each class in each of the three data sets used to evaluate classification performance. The 2017 data set was used to evaluate classification models in post-processing, and the 2019 data sets were used to evaluate real-time performance.

optical camera imagery when water clarity/illumination allowed. Table 4.3.1 lists the total number of target tracks available for each class in this data set ("2017 data set").

Each target track was described by a set of 28 hand-engineered features, described in Appendix A. This set of features was adapted and expanded from the features used to assess the feasibility of classification in Chapter 3. Features include descriptions of target shape, target motion, target intensity, the characteristics of the sonar image that the target was detected in, and the environmental covariates. The values of all features for each track were denoted by a 1-by-m vector, $\vec{x}_n$, where n was the total number of features ($n = 32$, because one of the 28 features results in 4 separate values), and m was the total number of target tracks (3238). Each feature vector, $\vec{x}_n$, was normalized such that the the 10^{th} and 90^{th} percentiles ranged between 0 and 1, respectively, as:

$$\vec{x}_{n,norm} = \frac{\vec{x}_n - P_{10}(\vec{x}_n)}{P_{90}(\vec{x}_n) - P_{10}(\vec{x}_n)}, \tag{4.1}$$

where P_i represents the i^{th} percentile of $\vec{x}_n$, and $\vec{x}_{n,norm}$ is the vector containing the normalized data. This approach was selected over normalizing such that the maximum and minimum values equaled 0 and 1, respectively, to prevent outliers from skewing the data set.

Algorithm	KNN (ensemble)	KNN (multi-class)	SVM	RF
Paramater	k	k	Kernel Function	Number of Trees
Values	3,5,7	3,5,7	Radial Basis Function (RBF), Linear, 3^{rd}-degree Polynomial	40, 70, 100, 130, 160, 190

Table 4.2: Parameters tested for each machine learning algorithm. MATLAB defaults were used for all other parameters.

All normalized feature vectors, $\vec{x}_{norm}$, were then assembled into an m-by-n matrix $\mathbf{X}$. The reviewer-assigned class for each target track was stored in a 1-by-m vector, $\vec{y}$.

4.3.2 Classification Model Validation

The 2017 data set was used to evaluate the performance of the k-nearest neighbor (KNN), random forest (RF), and support vector machine (SVM) algorithms for target track classification. All models were implemented in MATLAB (Mathworks) using the Statistics and Machine Learning Toolbox. For each algorithm, sensitivity to a core parameter was evaluated, as described in Table 4.2, while MATLAB default values were used for all parameters not listed. For each set of KNN parameters, both a multi-class model and an ensemble of binary models were tested (i.e., a binary classification model was trained for each class, and each target track was assigned the class of the model with the highest probability of classification). The probability of classification using KNN models was calculated using the posterior probability. Because SVM is a binary classification algorithm, only ensembles of binary models were tested using this algorithm. The probability of classification for SVM models was calculated using Platt's method [47]. RF is inherently an ensemble approach, so

only multi-class configurations were tested.

The same validation process was used for each classification model. First, to address the disparity in the number of tracks belonging to each class in the 2017 data set, data were randomly subsampled (i.e., undersampling [37]) to provide an equal number of training data points from each class (77 samples from each class). The resulting subset of the 2017 data set was then randomly divided into training data (70% of the data) and validation data (30% of the data), with equal representation from each class in both data subsets. The training data were used to train the machine learning algorithm to produce a classification model, which was then used to predict the classes of the validation data.

Classification effectiveness can be described in many ways (e.g., true positive rate, overall classification rate). Here, a set of metrics selected to be informative for environmental monitoring are used. The true positive rate for each class, (TPR_c), also known as the sensitivity, was calculated using the predicted classes of the validation data as

$$TPR_c = \frac{TP_c}{N_c}, \tag{4.2}$$

where TP_c is the number of correct classifications belonging the class, c, (e.g., true positives), and N_c is the total number of target tracks in the validation data set belonging to the class, c. The true positive rate for all biological targets (plants and animals), TPR_{bio}, was calculated as:

$$TPR_{bio} = \frac{TP_S + TP_{Sm} + TP_B + TP_F}{N_S + N_{Sm} + N_B + N_F}. \tag{4.3}$$

The TPR_{bio} metric is important in our application because it is maximized when targets of biological interest are correctly classified. The correct classification of non-biological targets is of less significance for monitoring at marine energy sites because non-biological targets are inherently not of interest. The true positive rate for binary classification of biological targets, TPR_{bin}, was calculated as:

$$TPR_{bin} = \frac{TP_{bio}}{N_{bio}}, \tag{4.4}$$

where TP_{bio} is the number of validation target tracks belonging to biological targets (here, seals, fish schools, small targets, or birds) that were not predicted to be non-biological targets, but were not necessarily predicted to belong to the correct biological class, and N_{bio} is the total number of biological targets. This metric provides insight into the rate at which targets of interest would be rejected as non-biological targets. This is relevant to our application because one benefit of a real-time classification model is to restrict data acquisition to only periods when a target of interest is present. Finally, the false positive rate (i.e., specificity) for binary classification of biological targets, FPR_{bio}, was calculated as:

$$FPR_{bin} = \frac{FP_{bio}}{N_{nonbio}}, \tag{4.5}$$

where FP_{bio} is the number of non-biological targets incorrectly classified as biological targets (false positives), and N_{nonbio} is the total number of non-biological targets. This metric provides insight into the volume of data that does not contain targets of interest that would be recorded if the classification model was used to limit data acquisition to periods when a biological target was predicted to be present.

This validation process, beginning with data subsampling from the total pool of track data, was repeated 100 times for each classification model, and median and interquartile ranges for all metrics were used to evaluate model sensitivity to the specific target tracks selected for training and validation.

4.3.3 Effect of Feature Selection

In any classification problem, it is possible that the full set of features in the available training data will not provide optimal classification performance, and a subset of those features would perform better. There are two distinct types of methods for feature selection: filter methods and wrapper methods [36]. Filter methods take only the training data into

Percent of Training Data Retained	70	60	50	40	30	20	10
Number of Tracks per Class	54	46	38	31	23	15	8

Table 4.3: Data volumes used to evaluate the effect of reduced training data volumes. Note that data were trimmed to have equal representation from all classes before classification (77 tracks from each class). 30% of the available training data (23 target tracks from each class) were used for validation in each case.

account, and select features based on some measure of the information that they contain. Wrapper methods account for both the training data and the classification algorithm, and find the subset of features that provides optimal performance for a particular algorithm. To evaluate the consequence of the hand-engineered feature set on classification performance, each classification model described in Section 4.3.2 was tested with three sets of features: (1) the entire set of 32 features, (2) a subset of features selected using a filter method, and (3) a subset of features selected using a wrapper method. While there are numerous approaches to feature selection in the literature (e.g., [36, 59]), two approaches were selected to compare representative filter and wrapper methods and to provide insight into the performance of the hand-engineered feature set. The filter method selected the 16 features whose feature vectors, x_n, had the highest linear correlation coefficients with the class vector, $\vec{y}$. The wrapper method used a hill-climbing algorithm with the objective of optimizing TPR_{bio}, which was adapted from [16], and is detailed in Appendix B.

4.3.4 Extensions of Existing Classification Models

Given that the most time-intensive step in developing a classification model is the annotation of training data, a natural question is the difficulty inherent to extending an existing classification model to a new location or incorporating a new target class. This could be particularly problematic when training data includes environmental covariates. For example,

if a classification model for seals is effective at a tidal energy site, it might not be as effective at a wave energy site if the class is correlated with environmental conditions that never occur (e.g., strong tidal currents).

First, to investigate how much training data would be required to adapt a classification model to a new site, the analysis in Section 4.3.2 was repeated for the top-performing classification models using each algorithm (based on TPR_{bio}) with sequentially reduced volumes of training data (Table 4.3). The same fraction of training data (30% of the undersampled training data set) was used for validation in each case to avoid biasing results. Second, to evaluate the role of environmental covariates, the wrapper feature selection process was repeated with features based on environmental covariates excluded (time of day, current speed, current direction, and relative target speed). The tidal current was measured by a co-located acoustic Doppler current profiler. The top-performing classification models using each algorithm, based on TPR_{bio}, were then re-evaluated with the modified feature lists following the procedures in Section 4.3.2.

4.3.5 Real-time Implementation

Test Site

Target track classification was implemented in real-time during a second deployment of the Adaptable Monitoring Package in Sequim Bay, Washington in 2019. The BlueView multibeam sonar (2250 kHz) was again used for real-time target detection and tracking. This deployment site was approximately 110 meters northeast of the location where 2017 data set was collected (Figure 4.1). The 2019 location was shallower (average depth of approximately 6 meters) and experienced stronger currents (peak currents of approximately 2 m/s) than the 2017 deployment site. The data collected during this deployment are separated into two subsets - 2019a and 2019b, which are enumerated in Table 4.3.1.

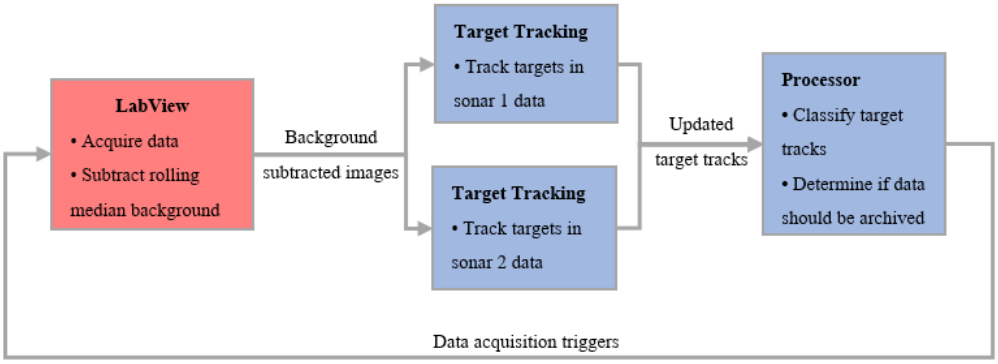

Figure 4.2: Real-time target tracking and classification software architecture. Grey arrows indicate data transfer. Processes in red take place in the LabVIEW core AMP software, the components of the real-time module are shown in blue. Note that this architecture could be extended to any number of sonars (depending on available processing power), though only two are shown here.

Software Architecture

Data acquisition was performed in the core AMP software (implemented in LabVIEW, National Instruments) and real-time processing modules were implemented in MATLAB (Mathworks). The overall architecture is shown in Figure 4.2. Data were stored in 60-second ring buffers in the core AMP software. The foreground of each sonar image was calculated by subtracting a rolling background using the median of the previous 10 seconds of data. To reduce real-time computational requirements, the background was updated once per second, while multibeam sonar data were acquired at 5 Hz. The foreground of each image was then stored in a temporary directory accessible by the MATLAB real-time processing modules.

The target tracking and classification modules each operated on a separate worker on a parallel pool in MATLAB (see Figure 4.2). Here, real-time target tracking and classification were only implemented for the BlueView M900-2250 multibeam sonar (2250 kHz transducer), but the same architecture could be extended to handle processing from multiple sensors. The target tracking module continually read in new foreground images produced by the core

AMP software. A range-varying intensity threshold (highest intensity threshold closest to the transducer) was applied to each image before detecting any targets in the image foreground and tracking them using a Kalman filter, as described in Chapter 3. Each new target track was assigned a unique identification number. After each image was processed, any target tracks that were updated in that image and had been detected in at least 4 previous images were sent to the classification module. If a target track was detected in more than 5 images, it was sent to the classification module again every time a new target was associated with the track. Any target tracks that were not updated for over 1.5 seconds were removed from memory. When the target tracking module lagged behind data acquisition by more than 10 seconds, data older than 10 seconds were ignored until target tracking was able to "catch up" to real-time acquisition. This "pressure relief" feature was needed to ensure software stability, but did not have a significant impact on performance, as discussed in Section 4.3.5.

The classification module was designed to be agnostic to the classification model, meaning that any of the models discussed in Section 4.3.2 could be used. For each target track that was received, all features listed in Appendix A were calculated. These features were then normalized based on the 10^{th} and 90^{th} percentiles of the training data (Equation 4.1) and used to predict the class of the target track. The predicted class, track features, track identification number, and the time of classification were logged so that tracks could be re-classified with a different model in post-processing, if desired. The time lag between data acquisition and target track classification, t_{lag}, is a useful parameter to assess real-time implementation. t_{lag} was calculated as the difference between the time stamp of the sonar image for the last target associated with a track and the time when that track was assigned a class.

The classification module then determined whether the track met criteria to send a trigger to the core AMP software to archive data. Trigger criteria were user configurable: decisions could be made based on the predicted class of a target (e.g., only archive data which is predicted to contain a seal), based on target track features (e.g., only archive data when a target track exceeds an area threshold), based on environmental covariates (e.g., only archive

data when current speeds exceed 1 m/s), or any combination of these criteria. Triggers were passed to the core AMP software via UDP.

Upon receiving a trigger, the core AMP software enabled artificial illumination for co-located optical cameras for three seconds when ambient light was insufficient to illuminate the image. The core AMP software then waited 30 seconds to center the event in the 60-second ring buffer and then archived the entire buffer to disk. The classification module did not generate a second trigger until 30 seconds had elapsed after the first, because such an event would already be contained in the archived data (albeit not centered in the recorded sequence).

Implementation

Based on the performance results presented in Section 4.4.1, a random forest classification algorithm using 130 trees and features selected by the hill-climbing algorithm was used for real-time classification. Initially, the model was trained using the 2017 data set and the processing code was configured to generate a trigger for any target track predicted to be a target of interest (e.g., a seal) or any target of potential interest that had a 75^{th} percentile area greater than 0.15 m^2 and intensity greater than 60 (sonar data were logged as 8-bit images, meaning that the maximum intensity value was 255).

110 one-minute sequences collected using this scheme were initially reviewed. The 1203 target tracks in these sequences formed the 2019a data set (Table 4.3.1), which was used to evaluate classification model performance. These data revealed that the classification model performed relatively poorly in real time (see Section 4.3.5). To address this, the 2019a data set was pooled with the 2017 data set, and the combined data set was used to retrain the classification model. The updated classification model was then implemented in real-time and an additional 326 one-minute sequences containing 2253 target tracks were manually reviewed. These target tracks formed the 2019b data set (Table 4.3.1), and were used to evaluate the updated classification model. Classification using only the 2019a data set for training was not attempted due to the relatively small volume of training data available for

some classes (e.g., diving birds).

When calculating the metrics presented in Section 4.3.2 to assess the performance of real-time classification, tracks were considered to be correctly classified if they were classified as belonging to the correct class for the majority of the track. This distinction is necessary because a track could be reclassified multiple times as new targets were detected and added to the track. For example, if a target track associated with a seal was tracked for 3 seconds, and sent to the classification module 10 times, it would be considered correctly classified (a true positive) if it were classified as a seal at least 5 times.

I note that the results presented in Section 4.4.4 are for targets that were detected and tracked in real-time. However, while a prediction of the class of each target was made in real-time, the results shown were re-calculated in post-processing because classification methods were refined over the course of the deployment, and the logs generated by the real-time processing code enabled re-classification of target tracks in post-processing by a model other than the one used to generate triggers. However, the models tested in post-processing should achieve the same results when implemented in real-time, because target detection, tracking, and calculation of the features used for classification occurred in real-time. Additionally, the same algorithm (RF) was used in real-time (albeit with varying training data and feature lists), and tracks of interest were reliably assigned a predicted class with $t_{lag} < 10$ seconds (see Section 4.3.5).

4.4 Results and Discussion

4.4.1 Classification Model Performance

Figure 4.3 shows the median value of TPR_{bio} for all classification models tested. Because 44 classification models were tested, Figure 4.4 shows the full set of classification metrics for only the best-performing model using each algorithm (KNN, SVM, and RF), based on TPR_{bio}. The full set of metrics for all tested classification models can be found in Appendix C and sensitivity to feature selection is discussion in Section 4.4.2. The top performing RF and KNN

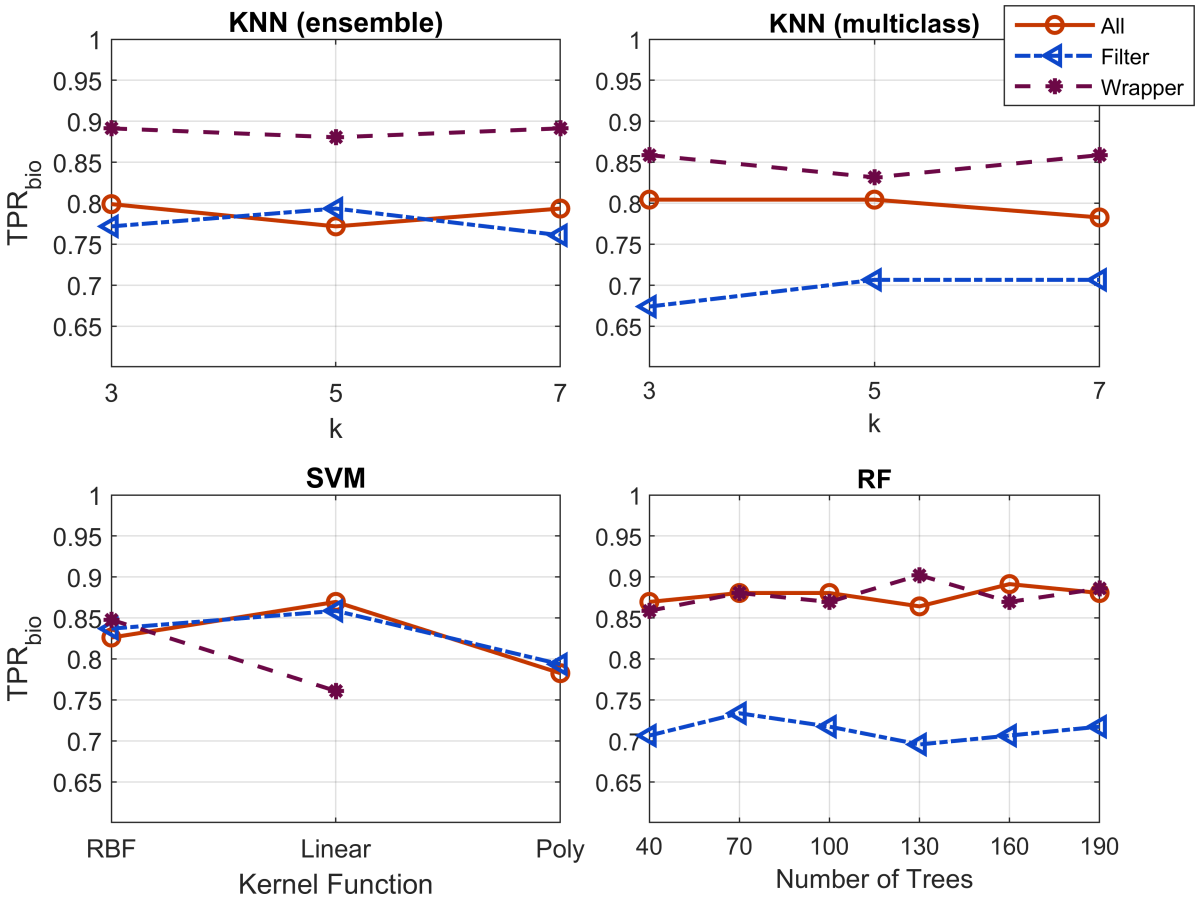

Figure 4.3: Median TPR_{bio} for each classification model tested, separated by algorithm and feature selection method. Note that the y-axis minimum is 0.6 (60% true positive rate for all biological targets).

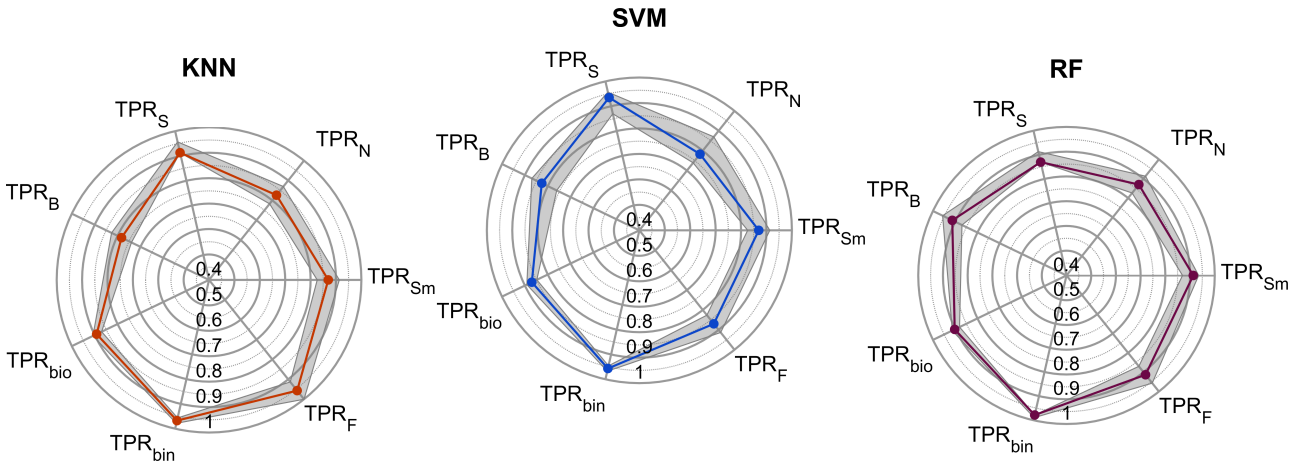

Figure 4.4: All classification metrics for the top-performing classification model using each machine learning algorithm. The colored line indicates the median value for 100 iterations of classification model validation, and the interquartile range is indicated by the grey shaded region. Note that the center of the plot is 0.4 (i.e., 40% classification rate).

models used sets of features selected by the wrapper feature selection algorithm while the top-performing SVM model used the full set of features. While there are meaningful differences between the best- and worst-performing classification models for each algorithm, some choice of model parameters produced relatively high performance for each algorithm. All three algorithms achieved binary classification rates, TPR_{bin}, greater than 0.96, indicating that targets of biological interest were infrequently classified as non-biological targets. This metric also showed the least variability with the subset of data selected for training (i.e., smallest inter-quartile range for performance metrics).

The RF model demonstrated superior performance by the most metrics, though each algorithm performed "best" by some metric. The RF model performed best for classification of diving birds and small targets ($TPR_B = 0.91$ and $TPR_{Sm} = 0.91$), while the SVM model performed best for classification of seals ($TPR_S = 0.93$), and the KNN model performed best for classification of schools of fish ($TPR_F = 0.96$). The RF model had the highest values of TPR_{bio} and TPR_{bin} (0.90 and 0.97, respectively), and the lowest value of FPR_{bin} (0.13, compared to 0.17 and 0.22 for KNN and SVM, respectively). Superior performance

of the random forest algorithm for this class of problem is consistent with prior research on benthic habitat classification using multibeam sonar data [26, 30] and radar target track classification [53].

4.4.2 Feature Selection

As shown in Figure 4.3, the wrapper method (hill-climbing algorithm) improved classification performance for most algorithms, with an average change in TPR_{bio} of $+0.03$ relative to classification using the full feature set, and never resulted in a significant decrease in performance. Conversely, the filter method (correlation) reduced performance for several algorithms relative to classification using the full feature set, with an average change in TPR_{bio} of -0.09. This disparity indicates that the hand-engineered feature set contains redundant features, which are removed by the hill-climbing algorithm but are reinforced by the correlation method.

The features contained in the subsets selected by the hill-climbing algorithm varied with classification model. The number of features selected varied as well, with a median list of 14 features and an interquartile range of 11.5. Figure 4.5 shows the percentage of the optimized feature lists that each feature appeared in. The time of day (T) feature appeared in the most feature lists (85%). This is unsurprising, because seals were almost exclusively present at night in the training data, and diving birds were almost exclusively present during daylight hours. Further, time of day is often included as an explanatory variable in ecological models, such as generalized additive models [54]. Interestingly, while human reviewers often relied on the shape of a target to annotate the training data, the shape features were used infrequently in the optimized feature sets. Finally, every candidate feature appeared in at least one optimized feature list.

The superior performance of classification models with feature lists optimized by the hill-climbing algorithm reinforces that wrapper feature selection methods that explicitly account for algorithm strengths and weaknesses are preferable to filter-based approaches based purely on the feature set [16]. For a given algorithm, the choice of parameter (e.g., number of

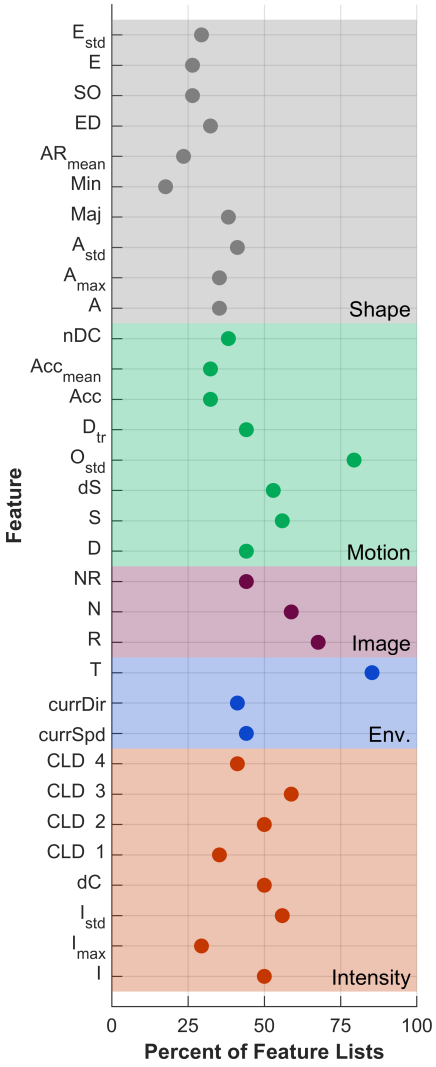

Figure 4.5: Features selected by the hill-climbing algorithm. The x-axis indicates the percentage of all optimized feature lists in which each feature is included. The shaded regions denote the general category of feature (i.e., shape, motion, image, environmental, and intensity descriptors). Features are defined in Appendix A.

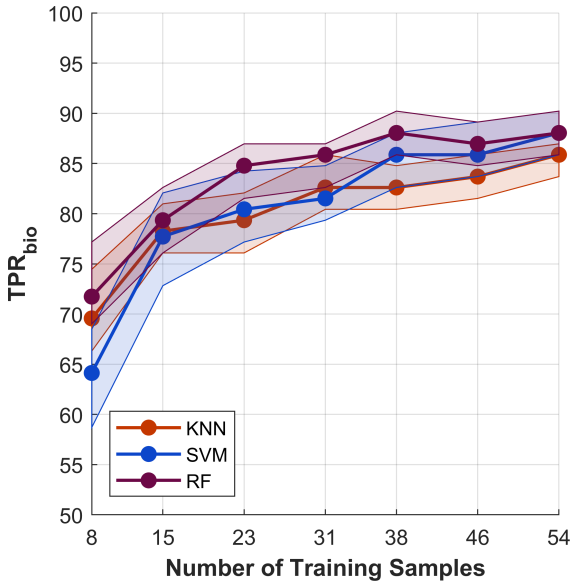

Figure 4.6: TPR_{bio} for the top-performing classification model using each algorithm with varying volumes of training data. For each model, the same number of samples from each target class was used (number of training samples). The solid line indicates the median value, and the shaded region indicates the interquartile range for 100 iterations of validation. Slight deviation from Figure 4.4 is a consequence of the randomized training data selection for each iteration.

nearest neighbors) has implications for the optimized feature set selected by hill-climbing. This approach is robust for hand-engineered feature sets, because it can remove features that are poorly defined or redundant and features that may match human intuition (e.g., target area), but do not aid in automatic classification. While the hill-climbing approach requires moderate computation time (up to 12 hours on a computer with an i7 processor and 16 GB of RAM), this is not expected to be a hindrance to its application to real-time classification. Calculation of an optimal feature only needs to occur at the time that the model is trained, so it will occur before real-time implementation.

4.4.3 Extensions of Existing Classification Models

Training Data Requirements

Figure 4.6 shows TPR_{bio} for the three classification models in Figure 4.4 with varying volumes of training data. As expected, performance generally improves with increasing volumes of training data. However, for all models, improvement begins to taper off as more training data are available. For the RF model, TPR_{bio} values are approximately stable after 50% of the available data (38 target tracks from each class) are used for training. Both the KNN and SVM models continue to show improvement with added training data beyond this point, but with diminishing returns. While the SVM model reaches approximately the same TPR_{bio} performance as the RF model with the full training data set (54 target tracks from each class), it does not perform as well with fewer samples. Overall, this indicates an advantage of the RF model: for automatic classification of infrequently appearing marine animals, high true positive rates can be achieved with relatively limited training data.

Reliance on Site Dependent Features

Figure 4.7 compares all classification metrics for the top performing classification model using each algorithm with and without environmental features excluded from the training data set. Several trends are apparent. First, classification performance is reduced for most metrics. However, slight increases are observed in the values of TPR_N, the true positive rate for non-biological targets. This is likely due to the fact that non-biological target tracks originate from a variety of sources (e.g., reflections from biological targets, turbulent backscatter during strong currents), and are therefore not correlated with environmental features. TPR_{bin} and TPR_F remain relatively constant for the SVM and RF models. The stability of TPR_{bin} indicates that while biological classes were classified with lower accuracy, they were still classified as biological targets and not as non-biological targets (e.g., some birds were misclassified as seals and vice versa). Significant reductions in performance are seen for TPR_B and TPR_S, the true positive rates for classification of birds and seals,

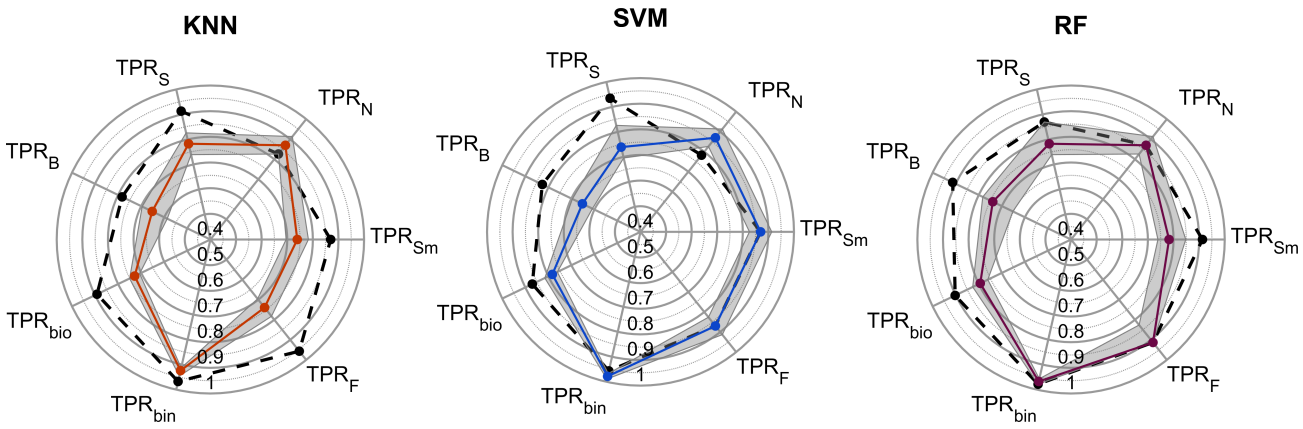

Figure 4.7: Results for the top-performing classification model using each machine learning algorithm trained without environmental features. The colored line indicates the median value for 100 iterations of classification model validation, and the interquartile range is indicated by the grey shaded region. The black dashed line indicates the results shown in Figure 4.4, with environmental features included in the training data set, for comparison. Note that the center of the plot is 0.4 (40% classification rate).

respectively. This is because birds and seals were only present in the training data during near-slack tidal conditions and during daytime and nighttime hours, respectively, resulting in strong correlations between presence and environmental covariates. As a result, TPR_{bio}, the overall assessment of biological classification rate, was significantly reduced without the inclusion of environmental features. Finally, the KNN model shows the most significant decreases in performance, indicating that it relies more heavily on environmental covariates for classification.

4.4.4 Real-time Implementation

Software Performance

The real-time target tracking, classification, and triggering software was stable over the course of the deployment, with no software crashes that were not the result of software upgrades (e.g., debugging after implementation of new features) or power outages. The tracking code was able to keep up with data acquisition reliably - tracking lagged behind

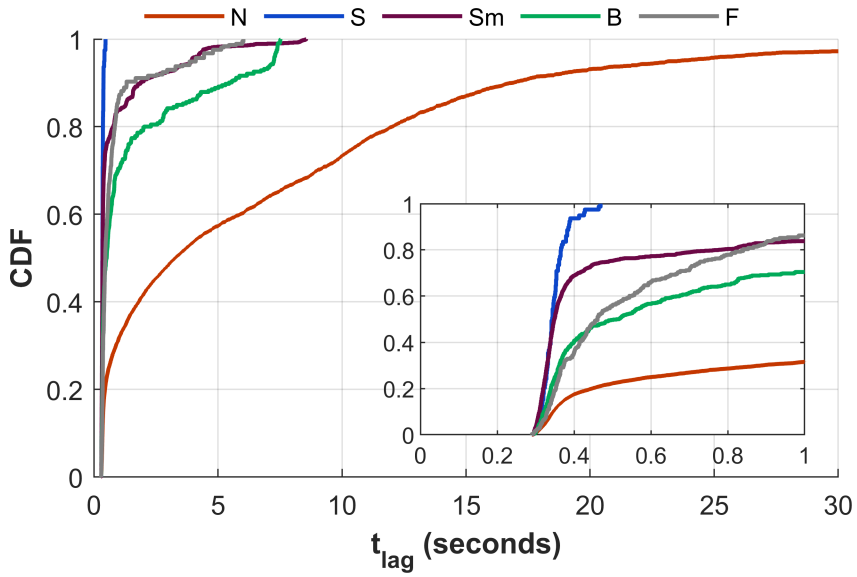

Figure 4.8: Cumulative density functions of the time to detect, track, and classify a target (t_{lag}) for each target class (non-biological (N), seals (S), small targets (Sm), diving birds (B), and fish schools (F)). The inset highlights the region of the plot where the time to track classification was less than one second.

data acquisition by more than 10 seconds on only 63 occasions over a 15 day period (an average of 4.2 times per day). On average, it took the target tracking software 9.5 seconds to recover and resume normal operation.

Target classification lagged behind target tracking due to the sequential nature of the real-time architecture. When more targets were present, the time required to track and classify each target track increased. Figure 4.8 shows the cumulative density functions of t_{lag} for each target class. The shortest t_{lag} was approximately 0.3 second, 100% of biological targets were classified within 10 seconds, and 100% of seals were classified within 0.5 seconds. 90% of non-biological target tracks were classified within 30 seconds, the maximum possible value of t_{lag} before the raw data containing the target track was no longer stored in the ring-buffer. Seals were consistently classified fastest because they typically were the only target in the field of view at the time of detection. Conversely, non-biological targets took longer to classify because they were frequently detected concurrently (i.e., many targets simultaneously

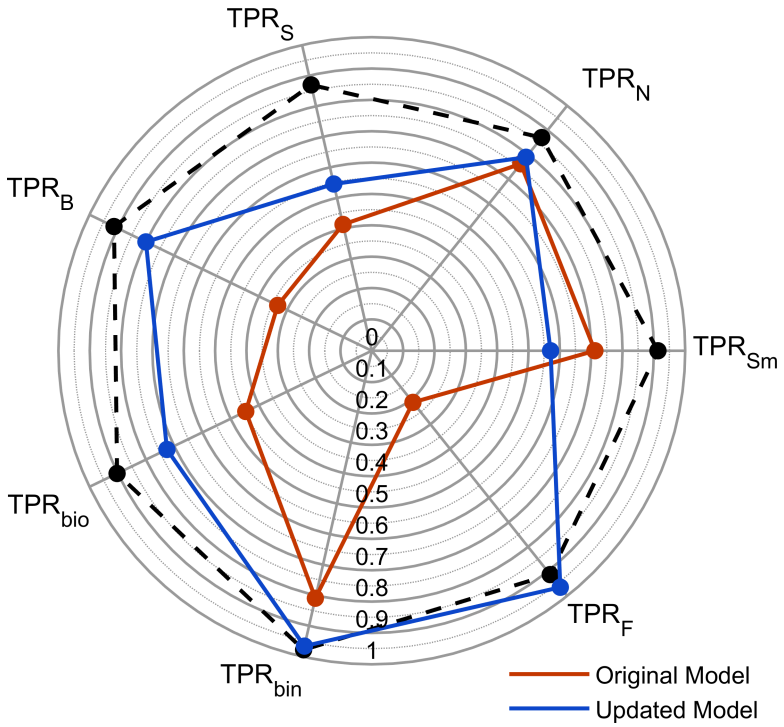

Figure 4.9: Real-time classification results, using the original classification model trained with only the 2017 data set and the updated classification model trained with the 2017 and 2019a data sets. The 2019a data set was used to validate the original model, and the 2019b data set was used to validate the updated model. The black, dashed line shows the post-processing performance of the RF model trained and validated using only the 2017 data set (from Figure 4.4).

present). These results are encouraging because non-biological targets are not of significant interest for environmental monitoring. It is important to note that t_{lag} would vary with the computational power of the control computer, but expect that the general trends would be consistent.

Classification Performance

Figure 4.9 compares real-time classification performance for the original classification model (trained using only the 2017 data set) and the updated classification model (trained using both the 2017 and 2019a data sets). Performance using the original model is quite poor when

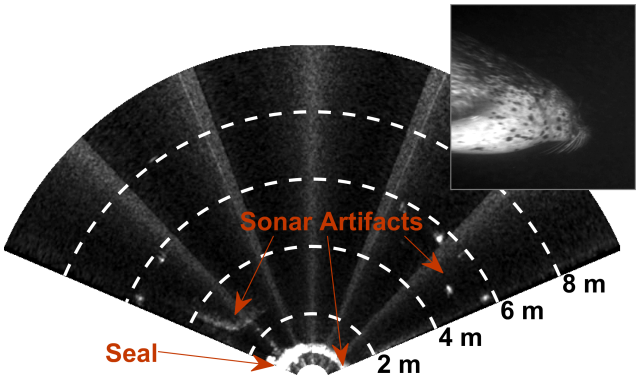

Figure 4.10: Seal detected at <1 meter range from the transducer. The seal is classified using concurrent optical data (top right), but difficult to distinguish in the sonar data due to high-intensity sonar artifacts.

compared to the results using the 2017 data for training and validation in post-processing shown in Figure 4.4. However, the updated classification model that included site-specific training data improved by nearly every metric. This is unsurprising, given that both 2019 data sets differed in several ways from the 2017 data set. First, fish schools were observed much more frequently in 2019. The 77 fish school target tracks in the original training data set were observed in only 11 sequences (e.g., 11 fish schools were tracked, but they were detected as multiple targets). It is likely that this small sample size resulted in overfitting of the original classification model for this class (e.g., classification of fish schools relied on trends that would not be present in a larger training data set). This resulted in an extremely low classification rate for fish schools using the original model ($TPR_F = 0.21$). However, after retraining, fish schools have the highest true positive rate of any biological class and are correctly classified 96% of the time ($TPR_F = 0.96$). Second, non-biological targets in 2019 had different characteristics — "clouds" of turbulent backscatter were more frequently observed at longer ranges, likely due to the stronger currents and shallower water depth. This resulted in a relatively low binary true positive rate ($TPR_{bin} = 0.81$) using the original model, because these targets were frequently misclassified as biological targets. However, TPR_{bin} increased to 0.97 using the updated classification model, a value nearly

equal to that achieved in post-processing using the 2017 data set. Finally, seals were more commonly observed at close range to the transducer in 2019. Specifically, 63% of seals in the 2019a and 2019b data sets were observed at a range less than 2 meters, while <10% of seals in the 2017 training data were observed inside this range (see Chapter 3). As shown in Figure 4.10, at this range, a seal produces high-intensity sonar artifacts, and accurate manual classification is only possible with concurrent optical camera data. This resulted in relatively low classification rates for seals using both models – TPR_S for both the original and updated classification models was below 0.6.

With the original model trained using the 2017 data set, 25% of non-biological targets were incorrectly classified as biological targets for the majority of the target track (FPR_{bin} = 0.25). In other words, if the original model was used to limit data acquisition to periods when a target was predicted to be a biological target for more than 50% of its target track, 25% of recorded target tracks would not be associated with a biological target. This metric remained relatively constant for the updated model, at 0.23. This represents a significant reduction in review effort when compared to human review of continuous data.

It is important to note that these classification results only pertain to targets that were successfully detected and tracked for more than one second, and it is likely that a small number of targets with shorter tracks were never classified. However, based on the analysis presented in Chapter 3, it can be expected that only 1% of discarded data contained targets of interest. Further, even a human reviewer would be challenged to classify targets that are only present in multibeam sonar data for less than one second and the environmental interactions of greatest interest for marine renewable energy will generally involve more sustained target presence.

4.4.5 Recommendations for Implementation at a Marine Energy Site

I have demonstrated that machine learning classification of targets in multibeam sonar data can achieve relatively high accuracy, with the exception of classification of large targets detected at close range (e.g., seals $\leq$ 2 meters from the transducer). However, classification

model effectiveness appears to be strongly site-specific, with more substantial differences observed for the same classification model across two locations than for a range of classification models at the same location. This is consistent with target classification using analogous radar imagery [53]. Site-specific variation is exacerbated at marine energy sites, particularly tidal energy sites, where high-energy conditions create dynamic environments with different characteristics at spatial scales on the order of 100 meters or less [49].

I have also demonstrated that binary classification (i.e., biological vs. non-biological) is more accurate than taxonomic classification (i.e., target belongs to "seal" class). This is also consistent with the findings of [53] for classification of birds tracked in radar data. Based on this, I recommend that real-time binary classification of biological targets be used to control data acquisition at marine energy sites, because the "cost" to store additional data is relatively low compared to the risk of missing a critical event. Using this approach, environmental interactions of potential interest can be archived with a relatively high true positive rate (over 95%). This method is also expected to record a manageable volume of data compared to continuous acquisition (true negative rate below 25%), which will enable rapid human review for timely decision-making. Because real-time taxonomic classification showed lower true positive rates, but still performed relatively well, these results could be used to control automatic adaptive action (e.g., enabling mitigation measures, such as fish deterrents), where the "cost" of an incorrect classification is lower.

Based on the results presented here, I recommend the following procedure to train a classification model at a new marine energy site:

- Implement real-time target tracking, using manually-tuned thresholds to trigger data acquisition (e.g., size and intensity). These thresholds should be set based on the smallest/least intense target of interest (e.g., area thresholds can be higher if marine mammals are the only targets of interest, and lower if detection of fish is desired). Tuning of thresholds may require iterative human review if either too many targets that are not of interest are being recorded or it is suspected that thresholds are too

high to record targets of interest.

- Have a human reviewer annotate recorded target tracks to the finest taxonomic level possible and use this information to build a binary classification model that distinguishes between biological and non-biological targets. I recommend using a random forest algorithm and a wrapper feature selection method, as this method was robust and required the least training data to achieve a high true positive rate. Regardless of the algorithm employed, feature optimization is recommended.

- Evaluate the model in post-processing using the methods outlined in Section 4.3.2. If the model achieves satisfactory performance for the application, implement in real-time to limit data acquisition to periods when a biological target is likely present, thereby minimizing the volume of data requiring curation or human review. If the model does not achieve satisfactory performance, continue to acquire additional training data to increase the volume of training data available.

- When finer taxonomic classification is desired, continue acquiring data and reviewing target tracks until approximately 40 target tracks are identified in each class of interest, then retrain the model with these classes defined. I recommend that this model be used to control automatic adaptive action in real-time or to guide human review in post-processing.

This approach can also be applied to existing data sets in post-processing, if the model is trained and validated using a subset of the available data and then used to automatically process the remaining data. Such a method could be used to rapidly evaluate temporal and spatial trends in class presence/absence. Additionally, the same target-tracking based approach to classification could be implemented with other multibeam sonars. However, sonar-specific training data would be required because the ranges of color-based (intensity) parameters will vary between sensors and this feature is often used in classification. Automated translation of classification models between sonars is a potential topic for future

research, as is using the predicted probability of classification to identify aberrations that may be indicative of a new class.

4.5 Conclusions

Automatic classification of seals, diving birds, small targets, fish schools, and non-biological targets in multibeam sonar data has been demonstrated in post-processing with a true positive rate of 0.90 for biological targets. Support vector machines, k-nearest neighbors, and random forests with varying parameters were evaluated for this task, and the random forest algorithm was found to perform best. Performance was improved through a wrapper feature selection method. The random forest classification model was implemented in real-time, and, after limited retraining with site-specific data, was able to distinguish between biological and non-biological targets with a true positive rate of 0.97 and a false negative rate of 0.23. Development of a classification model at a new test site is expected to be possible with relatively small training data sets (< 40 samples from each class). In aggregate, this method appears to be broadly applicable within the marine energy sector to gather information needed to retire or mitigate environmental risks.

Chapter 5

ACOUSTIC EVALUATION OF SENSORS FOR MARINE ENVIRONMENTAL MONITORING

Because artificial illumination is known to bias marine animal behavior [24, 39, 52], the Adaptable Monitoring Package uses active acoustic sensors to limit the use of strobe lights to periods when there is a target in close proximity to the system. However, the effects that these active acoustic sensors might have on animal hearing are not well studied. Potential impacts on marine mammal hearing are of particular interest, because marine mammals have relatively broad hearing ranges and use sound for communication and navigation. In this chapter, the acoustic emissions from four active acoustic transducers that have been integrated into the Adaptable Monitoring Package are investigated within the context of marine mammal hearing. Potential impacts on fish and invertebrate hearing are also of interest [46], but are not the focus of this work.

The contents of this chapter that follow were originally published in the *Marine Pollution Bulletin*:

E. Cotter, P. Murphy, C. Bassett, B. Williamson, and B. Polagye. Acoustic Characterization of Sensors for Marine Environmental Monitoring. *Marine Pollution Bulletin* (114) 205-215. 2019.

Minor modifications have been made for cohesion with the remainder of this thesis.

5.1 Introduction

Active acoustic sensors are used for a variety of scientific purposes, including fisheries stock assessment [20, 58], water current measurements [61, 64], hydrographic surveys (Lundblad

et al., 2006; Wilson et al., 2007), and marine mammal monitoring [28, 38, 50, 69]. The operational frequencies used by these sensors are driven by many factors, including scattering physics, acoustic attenuation rates, and study range. For example, in fisheries acoustics, lower frequency instrumentation (< 100 kHz) that allows for sampling at ranges of hundreds of meters is often used, while short-range imaging applications generally operate at higher frequencies. There are many advantages to using active acoustics in lieu of optical or passive acoustic sensors: they can operate when darkness, water conditions, or range may limit the use of optical cameras, and can detect marine animals when they are not vocalizing. However, if transducers produce sound within the range of marine mammal hearing, it may cause temporary or permanent hearing threshold shifts (NMFS, 2018) or affect their behavior if they are attracted to or avoid the source [27, 60].

The US National Marine Fisheries Service (NMFS) provides technical guidance to assess the effects of sound exposure on marine mammals, including definition of the hearing ranges of five groups of marine mammals (Table 5.1). Further, NMFS recommends that the cumulative sound exposure level (SEL_{cum}) be used to assess the risk of hearing damage from non-impulsive sounds, and that sonars be treated as non-impulsive sources because their pulse durations are sufficient to reach a steady-state [45]. SEL_{cum} uses an equal energy hypothesis and an auditory weighting function specific to each marine mammal hearing group to assess the cumulative impact that a source has on a marine mammal over time [62]. If SEL_{cum} exceeds a hearing group-specific threshold (Table 5.1), a temporary threshold shift (TTS) in hearing may be expected. In extreme cases, a permanent threshold shift may occur.

For many applications, active acoustic sensors are mounted to a moving vessel and deployed for short periods. In this scenario, exposure levels for any individual animal will likely be low because an individual animal is unlikely to remain within the ensonified area for a long period. However, when continuous monitoring of a site is required, sensors may be operated for extended periods from a stationary platform (e.g., monitoring of a tidal turbine, as in [67]). In these cases, a more careful consideration is required to evaluate exposure effects. If the source level of a transducer at maximum power input is known, then it can be calculated

at other power levels as

$$SL = SL_{max} + 10 \log_{10} \frac{P}{P_{max}} \tag{5.1}$$

where SL_{max}, in dB re 1 μPa, is the source level (received level at a range of one meter) on-axis within the beam at P_{max}, the maximum transmit power, and SL, in dB re 1 μPa, is the source level at another transmit power, P. SEL_{cum} for a receiver on-axis within the beam can then be calculated as

$$SEL_{cum} = SL_W - 20 \log_{10} R + 10 \log_{10} \tau + 10 \log_{10} n - \alpha R \tag{5.2}$$

where SL_W, in dB re 1 μPa, is the weighted source level calculated by applying the auditory weighting function for a specific marine mammal hearing group to SL; R is the range, in meters, from the transducer to the receiver; τ is the pulse duration, in seconds; n is the number of pulses within an evaluation window; and α is the coefficient of absorption, in dB/m. SEL_{cum} has units of dB re 1 μPa^2-s. Estimating SEL_{cum} when off-axis is more difficult as it requires either measurements of the beam pattern or an analytical model based on the geometry of the transducer [42]. For fisheries echosounders, if the nominal operating frequency lies within a marine mammals hearing range (Table 5.1), this calculation is straightforward because the source level, pulse duration, and beamwidth are known through calibration [14].

If the nominal operating frequency is above a marine mammals hearing range, there may still be acoustic emissions at lower frequencies [15, 26, 51]. This "out-of-band" sound can originate from sources such as "leakage" of the high-frequency signal to lower frequencies or sound associated with switching between transmit/receive modes or other sonar functions. Source levels for emissions at these frequencies, particularly as a function of position within the beam, are often not well-characterized, and out-of-band emissions may vary over the nominal pulse duration.

In this study, acoustic emissions from four active acoustic transducers are characterized: two multibeam sonars (one with two operating frequencies) and an acoustic Doppler current

Group	**Hearing range (kHz)**	**SEL_{TTS} (dB re 1 μPa^2-s)**
Low-frequency cetaceans (baleen whales)	0.007-35	179
Mid-frequency cetaceans (dolphins, toothed whales, beaked whales, bottlenose whales)	0.150-160	178
High-frequency cetaceans (porpoises, *Kogia*, river dolphins, *Lagenorhynchus criciger*, *L. australis*	0.275-160	153
Phocid pinnipeds (seals)	0.050-86	181
Otariid Pinnipeds (sea lions and fur seals)	0.060-39	199

Table 5.1: Hearing ranges and SEL_{cum} thresholds for TTS onset (SEL_{TTS}) for five marine mammal hearing groups (NMFS, 2018)

profiler (ADCP). As summarized in Table 5.1, the nominal operating frequencies of these transducers exceed the marine mammal auditory range. Acoustic emissions below 160 kHz, the upper limit of marine mammal hearing [45], are characterized in the along-swath and across-swath directions. From these measurements, SEL_{cum} is estimated throughout the sonar swath and the extent of audibility to each marine mammal hearing group is estimated. Acoustic emissions from one of the multibeam sonars (Tritech Gemini) have been previously characterized [27] but are not presented in a manner that would allow SEL_{cum} to be estimated. The author is not aware of acoustic characterizations in the public domain for the other multibeam sonar (BlueView M900-2250) or acoustic Doppler current profiler (Nortek Signature 500). The methods presented in this paper can be used to assess other active acoustic sensors in a manner that allows for effective comparison between transducers and for the results to be used for environmental assessments.

Type	**Manufacturer**	**Sensor**	**Nominal Operating Frequency (kHz)**	**Nearfield Estimate (m)**	**Along-swath**	**Across-swath**
Multibeam Sonar	Teledyne BlueView	M900-2250	900	3.4	130^o	20^o
Multibeam Sonar	Teledyne BlueView	M900-2250	2250	1.5	130^o	20^o
Multibeam Sonar	Tritech	Gemini 720is	720	5	120^o	20^o
Acoustic Doppler Current Profiler	Nortek	Signature 500	500	1.2	2.9^o	2.9^o

Table 5.2: Evaluated active acoustic sensors, their nominal operating frequencies, an estimate of the extent of their nearfield, and their swath dimensions. Nominal operation frequencies and swath dimensions are specified by the manufacturers. The dimensions for the Nortek Signature 500 are for a single beam (the sensor has five diverging beams with the same axisymmetric beamwidth). Nearfield extent was estimated at the nominal operating frequency (Appendix A). Swath dimensions are also specified at the nominal operating frequency and are not expected to match those of out-of-band measurements.

5.2 Methods

Evaluation of acoustic emissions was performed in two phases. First, an initial evaluation of each transducer was conducted to determine what, if any, sound it produced below 160 kHz at the center of the swath (i.e., on-axis) and how that sound varied with transducer operating mode (e.g., intended operating range, power level). Second, across-swath and along-swath sweeps were conducted for each transducer at the operating mode associated with highest amplitude sound. All activities were performed in Portage Bay, Seattle, WA (freshwater, approximately 21^{o}C) on R/V Henderson, a 21-meter steel-hull catamaran vessel. The water depth at the test site was 5.1 meters and the water was quiescent, with no currents or waves. Two Ocean Sonics icListen HF hydrophones (0.01-200 kHz frequency range, 512 kHz sample rate, standard GeoSpectrum hydrophone) were used to record sound from each transducer. The icListen uses a sigma-delta converter that samples at 16 MHz. Data are low-pass filtered (cut-off frequency of 250 kHz) to prevent aliasing and decimated to a sampling rate of 512 kHz (personal communication, Mark Wood, Ocean Sonics Ltd.).

An in-situ calibration of the hydrophones was conducted to determine their respective sensitivities in the direction of the active acoustic sources. The hydrophones were mounted in their measurement positions on R/V Henderson and an F41 acoustic projector with a traceable calibration was mounted at the same location as the active acoustic transducers under evaluation. Hydrophone sensitivities were estimated from 20 to 150 kHz in 0.5 kHz steps. Because the F41 transducer does not produce calibrated sound at frequencies below 20 kHz, sensitivities from 1 kHz to 20 kHz were taken from the manufacturer calibration, and sensitivities from 0.001 to 1 kHz were taken from a third-party calibration by Ocean Networks Canada. Although the calibrations conducted at Ocean Networks Canada did not account for directionality, the manufacturer calibration suggests directionality is not significant at frequencies less than 20 kHz. It was necessary to move the hydrophones between the two measurement phases, which resulted in 5^o degree azimuthal uncertainty in their orientation.

For the first phase of testing, each active acoustic sensor was mounted to a hydraulic ram

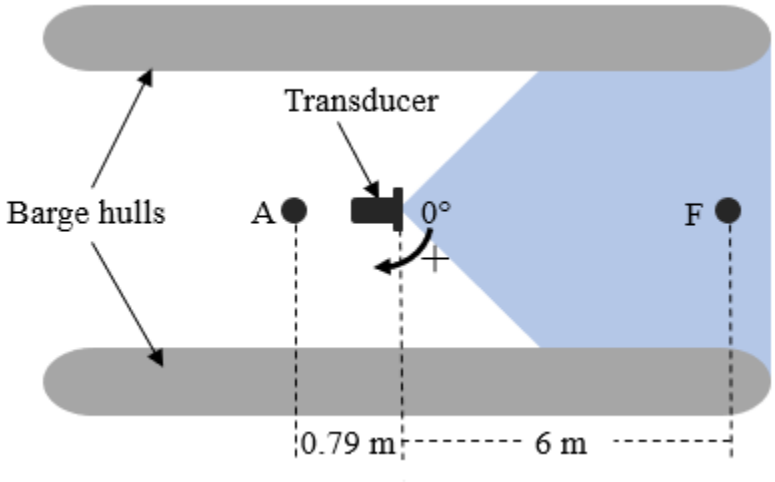

Figure 5.1: Test configuration (plan view; not to scale). "A" represents the aft hydrophone, and "F" represents the fore hydrophone. The blue swath represents the projection of a nominal swath from a transducer. The transducer and hydrophones were positioned below the barge hulls.

between the hulls of the vessel, oriented towards the bow of the vessel, and lowered to a depth of 2.1 meters. The transducers were below the catamaran hulls at this depth. The multibeam sonars were oriented such that the along-swath direction was parallel to the water surface. The Gemini was mounted with a 10-degree upward tilt to compensate for an internal 10-degree downward tilt. The Signature 500 was oriented such that the center (i.e., "vertical") beam was directly facing the hydrophone (all other beams were disabled during testing). The hydrophones were mounted to rigid posts and lowered below the hull of the vessel such that their transducers were at the same depth as the center of the beam. One hydrophone was mounted 6.0 meters directly in front of the active acoustic transducer ("fore hydrophone") and one hydrophone was mounted 0.79 meter directly behind the active acoustic transducer ("aft hydrophone") (Figure 5.1). The fore hydrophone was in the acoustic far-field of all transducers at their nominal operating frequencies (see Table 5.1). Because the distance to the far-field is proportional to frequency for a fixed aperture (Medwin and Clay, 1997), the near-field at the frequencies reported in this paper ($<$ 160 kHz) is expected to be smaller

than the values presented in Table 5.1 (see further discussion in Appendix D). The aft hydrophone was not positioned in the far-field of all transducers, so source levels could not be calculated for this orientation. The test depth was fixed by the length of hydrophone cable, and the distances from the active transducer were fixed by hardware mounting points available on R/V Henderson.

As summarized in Table 5.2, both BlueView transducers (900 and 2250 kHz) were cycled through multiple power settings. The range setting of the Gemini was varied instead of power level, as the Gemini software (SeaTec) does not allow the power level to be adjusted. The Signature 500 was tested at its maximum transmit power with only the center beam enabled to prevent reverberation from the diverging beams, under the assumption that the other four beams have comparable acoustic characteristics. Manufacturer defaults were used for all other settings. At least 60 pings were recorded at each operating mode. The hydrophone data were monitored in real-time to avoid collecting data when ambient levels were elevated due to intermittent anthropogenic noise (e.g., vessel traffic), and data were collected at night when ambient noise was at a diurnal minimum. At least 30 seconds of ambient noise (no transducer pinging) was recorded between each test of transducer operating modes.

Individual pings were isolated from the hydrophone data to characterize the sound generated by a transducer in a particular operating mode. For the first phase of testing, the first arrivals of 30 pings were isolated for each operating mode. This was done using a semi-automated process. A spectrogram of the recording window for each sonar mode was displayed to a human reviewer. From this, a relatively "clean" portion of the time series (e.g., no passing vessel traffic) was selected for processing. The reviewer then selected a frequency bin where the active transducer signal was clearly visible in the spectra to create a bandpass filter centered at the selected frequency with a passband and stopband that were 25 and 500 Hz wide, respectively. This filter was then applied to the voltage time series, and a threshold was manually selected to identify the approximate time of each ping in the filtered data, as shown in Figure 5.2a. All subsequent analysis was performed on the original, unfiltered data.

Sensor	**Operating mode**	**Range setting (m)**	**Pulse repetition rate (Hz)**
BlueView (2250 kHz)	25% power level	10	11.4
	50% power level	10	11.4
	75% power level	10	11.4
	100% power level	10	11.4
BlueView (900 kHz)	50% power level	50	3.3
	70% power level	50	3.3
	85% power level	50	3.3
	100% power level	50	3.3
Gemini	Default	10	10
	Default	50	10
	Default	100	10
ADCP	Single beam, Maximum Power Level	47	8

Table 5.3: Transducer operating modes. Note that the range setting indicates the sensor operating setting, not the range at which received levels were measured.

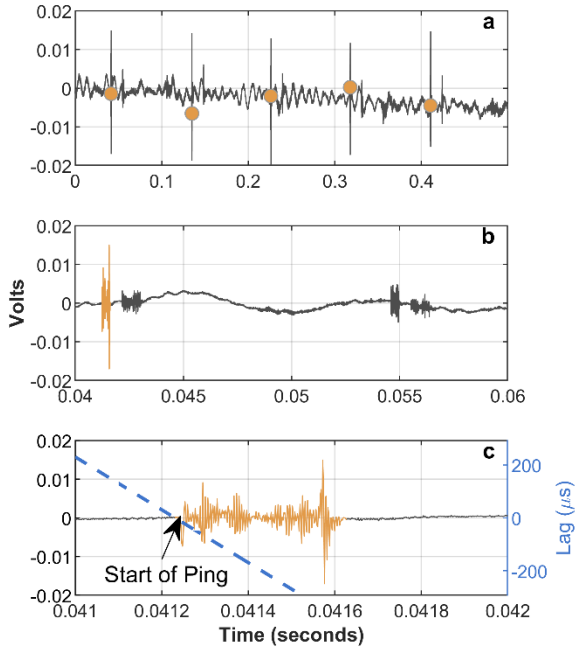

Figure 5.2: Sample voltage time series data from on-axis measurement of the 2250 kHz BlueView at 100% power. a) Approximate ping events detected in filtered data using an amplitude threshold (yellow circles). b) 20 ms window around one ping event, including the first arrival (yellow) and subsequent multi-path arrivals. Only the first arrival is used for subsequent analysis. c) Ping automatically detected using cross-correlation with the manually identified template signal. The dashed blue line shows the lag values at peak cross-correlation in the window.

A 2 ms time window around an identified ping was then displayed, and the reviewer manually selected the start and end time. This produced a "template ping" used to isolate other pings in the time series. As shown in Table 5.2, the pulse durations of the pings observed in the measurements did not match the manufacturer-specified pulse durations. The discrepancy is likely because the manufacturer-specified pulse durations apply to the operating frequency of each instrument, while these measurements are limited to the bandwidth of the hydrophones.

The template ping was used to isolate pings at each of the identified points in the filtered

Sensor		**Pulse duration (μs)**	
		Manufacturer-specified	**Observed**
BlueView (2250 kHz)		640	398
BlueView (900 kHz)		1,600	2,381
Gemini	10 meters	12.2	184
	50 meters	88.3	328
	100 meters	177	328
Signature 500		673	1,678

Table 5.4: Manufacturer-specified pulse duration at nominal operating frequency and estimated values from measurements.

data using cross-correlation. The time window with the highest cross-correlation with the template ping was identified, and the lag value at the peak cross-correlation was taken to be the start of that ping, as shown in Figure 5.2c. After automatic detection, each ping was manually reviewed to ensure that the beginning of the ping was accurately located. Because surface reflections resulted in multiple arrivals of each ping (Figure 5.2b), the ping was shown to the reviewer in the context of the time series to ensure that only first arrivals were included in subsequent analysis. The same template ping was used for transducer operating modes expected to have the same ping profile (e.g., the same transducer operating at 50% and 100% power level). A representative ping from each transducer, received by the fore hydrophone, is shown in Figure 5.3.

The first arrivals of pings in the aft hydrophone data were automatically isolated by shifting the time window for each detected ping by the time difference of arrival between the two hydrophones using an estimated sound speed of 1486 m/s [40] and the separation distance from the forward face of the transducer.

The frequency content of each ping was calculated using a discrete Fourier transform (DFT). Because of the short pulse duration (300-2000 μs), a window length equal to the size

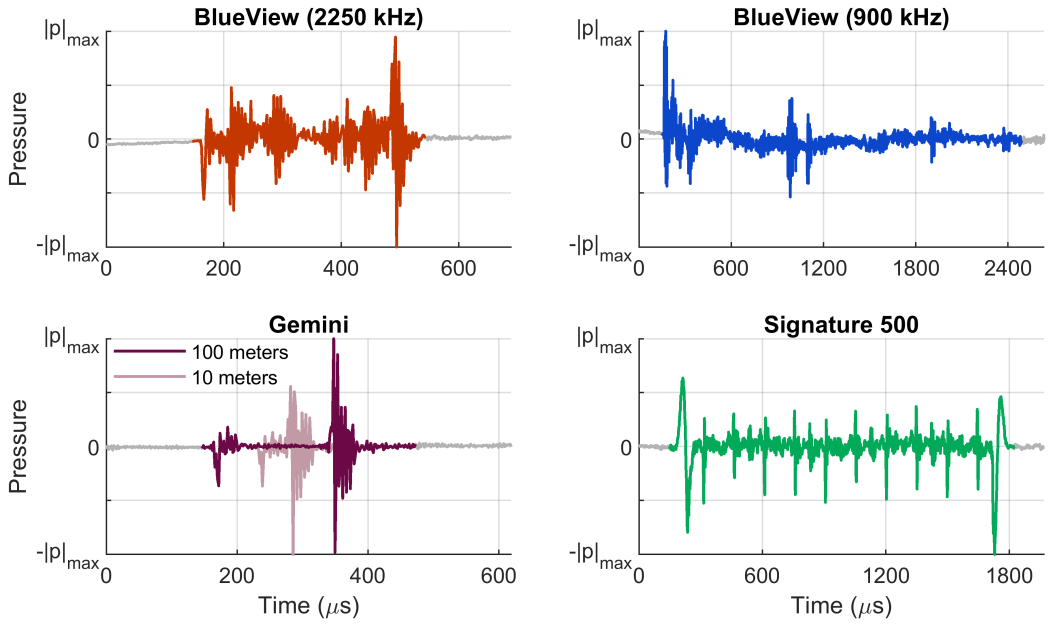

Figure 5.3: Representative on-axis pings from each transducer. For visualization, data are normalized by the peak absolute pressure in each time-series. 100% power modes are shown for both BlueView transducers, and both 10 and 100-meter operating ranges are shown for the Gemini. The structure of the pings were approximately constant with varying power level for both BlueView transducers, as well as for the 50 and 100-meter range modes for the Gemini. Note that the time scale varies between plots because of varying ping duration.

of the longest ping (the 900 kHz BlueView, at 1219 points) was used and a periodic Blackman taper applied. For pings with a pulse duration shorter than the window length, the data were zero-padded after tapering. This yielded the mean-square sound pressure, $p^2(f)$, in units of Pa2/Hz with a frequency resolution of 420 Hz. The median and interquartile range of $p^2(f)$ were calculated for the 30 identified pings.

Fifteen seconds of ambient data were processed using the same approach as the ping data to allow sound during a ping to be compared to ambient noise. The mean-square sound pressure spectral density (PSD) of each ping, in dB re 1 μPa^2/Hz, was calculated by subtracting the ambient mean-square sound pressure spectrum recorded closest to each ping, $p_a^2(f)$:

$$PSD_h(f) = 10 \log_{10} \frac{p^2(f) - p_a^2(f)}{p_{ref}^2} \tag{5.3}$$

$PSD_h(f)$ is the PSD measured at the hydrophone, and p2ref is the reference pressure (1 μPa2/Hz).

Fluctuations in ambient sound pressure did not contribute significantly to ping-to-ping variations in $PSD_h(f)$. Figure 5.4 compares the interquartile range of a 15-second window of ambient PSD recorded before each transducer was tested to the PSD measured immediately before each detected ping. This demonstrates that ambient PSD did not vary significantly during testing. The peak in the ambient PSD at 50 kHz is likely associated with depth sounders from a nearby marina. Because pings that contained the 50 kHz signal were discarded during manual review, this means that $PSD_h(f)$ is slightly over-corrected at this frequency. However, this is of limited consequence, as ambient PSD is at least a 10 dB below the source PSD of the active transducer pings for most frequencies (see Figure 5.5).

$PSD_h(f)$ for the first arrival was extrapolated to the PSD at the source (received levels at a range of 1 meter) as:

$$PSD(f) = PSD_h(f) + 20 \log_{10} R_h \tag{5.4}$$

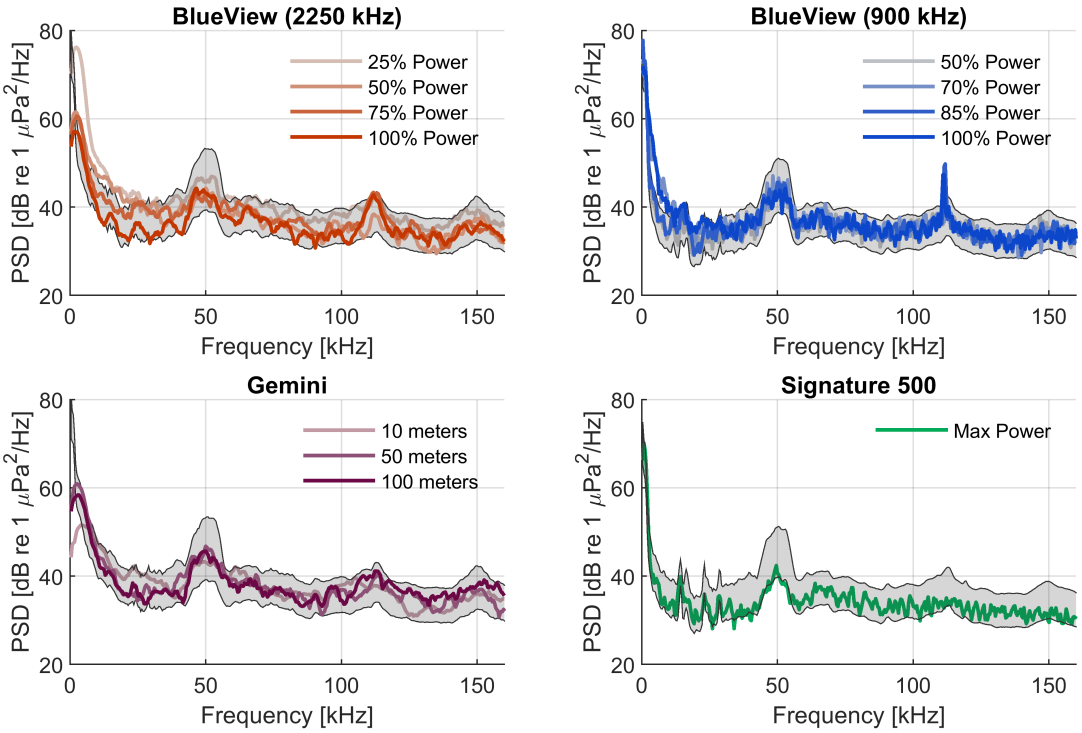

Figure 5.4: Variations in ambient sound during testing for each transducer. The grey shaded region depicts the interquartile range of the ambient PSD calculated for 15-second recordings before the test window. The colored lines indicate the median PSD calculated for the 3.9 ms (2000 points) preceding each detected ping.

$PSD(f)$ is the source PSD, in dB re 1 μPa2/Hz, and R_h is the range to the hydrophone (6 meters). Given that measurements were made in freshwater with a relatively short path length, absorption is negligible ($\approx$0). Specifically, absorption is approximately 5.4 dB/km in freshwater at 21^{o}C with a pH of 7.5 at the measurement depth of 2.1 meters, or 0.03 dB over the 6 meter measurement distance [22, 23], which is negligible in comparison to measurement accuracy. As previously discussed, a receiver at 1-meter range is in the far-field for all transducers at 160 kHz except the Gemini, which has an estimated nearfield range of 1.1 meters at this frequency (see Appendix A).

$PSD(f)$ can be inverted to yield a source pressure distribution as:

$$p_s^2(f) = p_{ref}^2 10^{\frac{PSD(f)}{10}} \tag{5.5}$$

Decidecade source levels ([31], equivalent to one-third octave band source levels, were calculated as:

$$SL_{ddec}(f) = 10\log_{10}\left(\int_{f1}^{f2} \frac{p_s^2(f)}{p_{ref}^2} df\right) \tag{5.6}$$

where f_1 and f_2 are the upper and lower frequencies of the band, respectively, and $SL_{ddec}(f)$ is in dB re 1 μPa.

The active acoustic sensors were rotated by the hydraulic ram to characterize variation in sound level in the across-swath and along-swath axes for the multibeam sonars, and across the beam of the Signature 500 (see Figure 5.1). To do this, the transducers and hydrophones were initially mounted in the same configuration as in the first phase of testing. The multibeam sonars were rotated 360 degrees in a clockwise direction, pausing every 5 degrees to record at least 30 pings. The ram was positioned with 0.02-degree precision. Within 10 degrees of the manufacturer-specified edge of the sonar swath, the angular resolution was doubled to 2.5 degrees to ensure that the edges of the swath were well-resolved. This test was repeated with the sonar swaths oriented parallel and perpendicular to the seafloor to characterize along-swath and across-swath variation, respectively. An error in hydrophone data acquisition

resulted in a loss of data between 20^o and 35^o for the Gemini across-swath sweep (outside of the nominal swath width). Because the Signature 500 beam is relatively narrow compared to the multibeam sonars swaths, 1-degree resolution was used for the Signature 500, with the resolution increased to 0.5 degrees within one degree of the edge of the beam (the ram was positioned with 0.01-degree precision). The Signature 500 beam pattern was assumed to be axisymmetric [42], so only one orientation was tested. The raw Signature 500 beam pattern showed the peak in received levels shifted 0.5 degrees to the right. However, it was assumed that this shift was due to inaccuracy in positioning the Signature 500 on the hydraulic ram, and reported values were shifted so that the peak received level is at zero degrees rotation.

For these measurements, pings were detected in the time series data by the peak cross-correlation with the template ping. If fewer pings were detected than anticipated (e.g., near the edge of the swath as received levels decreased), a bandpass filter was applied to attempt to isolate pings in the data, following the methodology used to isolate pings in the on-axis measurements. Additionally, when the ping structure was observed to change within the swath, the reviewer was able to update the template ping (maintaining the same pulse duration). All pings were manually reviewed to ensure accuracy in ping selection. Ten pings were selected for analysis at each angular position, and $PSD(f)$ and p_s^2(f) were calculated for each ping at each position using equations 5.3-5.5.

The source level (SL) of each ping, in dB re 1 μPa at 1 meter, was calculated from measurements as:

$$SL = 10\log_{10}\left(\int_0^{160kHz} \frac{p_s^2(f)}{p_{ref}^2} df\right) \tag{5.7}$$

SL for each angular position was defined as the median SL of the 10 pings at that position, calculated in pressure-squared space.

The PSD at each angular position was used to estimate SEL_{cum} for each marine mammal group as a function of range and direction from the transducer. The auditory weighting function for each marine mammal hearing group, $W(f)$, was applied to the PSD for each position in the across-swath and along-swath sweeps to produce a weighted PSD, PSDW.

$W(f)$ was applied in log-space as:

$$PSD_W(f) = W(f)PSD(f) \tag{5.8}$$

A weighted source level, SL_W, in dB re 1 μPa, was then calculated as:

$$SL_W = 10\log_{10}\left(\int_0^{160kHz} 10^{\frac{PSD_W}{10}} df\right) \tag{5.9}$$

These weighted SL_W values were used to calculate SEL_{cum}, in dB re 1 μPa^2-s, at each angle and range, R, in meters, from the transducer by applying Equation 5.2. The number of pulses, n, varies with the exposure time, T, as:

$$n = Tr \tag{5.10}$$

where r is the pulse repetition rate, in pulses per second, and T is time in seconds. For example, the 2250 kHz BlueView had a pulse repetition rate of 11.4 Hz with an observed pulse duration of 398 s. After $T = 1$ minute within the sonar swath, a marine mammal would be exposed $n = 684$ pings. Calculation of SEL_{cum} was repeated for exposure times from 0 to 104 minutes (approximately 7 days) and for ranges from 1 to 20 meters. It was assumed that an animal makes no auditory recovery between pings [45], and the absorption coefficient, α , was calculated for representative conditions in seawater (10^{o}C, 35 psu, pH 8) at a depth of 10 meters. These calculations assume an unobstructed acoustic path between source and receiver and neglect boundary interactions. In other words, these results would be accurate for a downward looking sonar in relatively deep water, but in shallower water or other sonar orientations, site-specific propagation modeling would be necessary to estimate the propagation loss. Results are presented in terms of the percentage of the TTS threshold (SEL_{TTS} in Table 5.1) that is exceeded at a given position and exposure time ($\frac{SEL_{cum}}{SEL_{TTS}}$), calculated in pressure-squared space). In addition, the exposure time, T_{TTS}, that an animal would need to spend at a given location in the sonar swath before exceeding SEL_{TTS} was

calculated. The tables contained in [45] were used for auditory weighting functions and TTS exposure thresholds.

Finally, $PSD(f)$ was used to estimate the maximum ranges at which marine mammals might be able to detect the sound produced by each sonar using the composite audiograms for each marine mammal hearing group derived in [45] and representative ambient noise levels at sea state one [65]. A simple propagation model was used to predict the received level at varying ranges from the sonar:

$$RL = PSD(f) - 20\log_{10} R - \alpha(f)R \tag{5.11}$$

It is assumed that integration times play no role in audibility to provide a conservative estimate of hearing ranges [19]. This calculation was repeated for the PSD measured at each angular position within the beam. The absorption coefficient for each frequency, was calculated for representative conditions in seawater (10^oC, 35 psu, pH 8) at a depth of 10 meters, the same conditions used for SEL_{cum} calculations. The sonar was considered to be no longer audible at the range where received levels at all frequencies and at all points in the beam fell below either the animals hearing threshold or the ambient noise level. Because of the relatively low ambient noise levels, animal hearing thresholds were the determining factor in audibility for all frequencies and marine mammal hearing groups.

5.3 Results

5.3.1 Acoustic Characteristics

Figure 5.5 shows the median source PSD and SL_{ddec} at 1 meter range from the transducer for pings measured by the fore hydrophone for each sonar operating mode at the center of the swath. The dashed, black line in each plot shows the ambient noise level during the test. All three sonars produced measurable sound within the range of marine mammal hearing in front of the transducer. For both BlueView transducers, the amplitude of the sound decreased non-linearly with power level setting, though the frequency content was independent of power

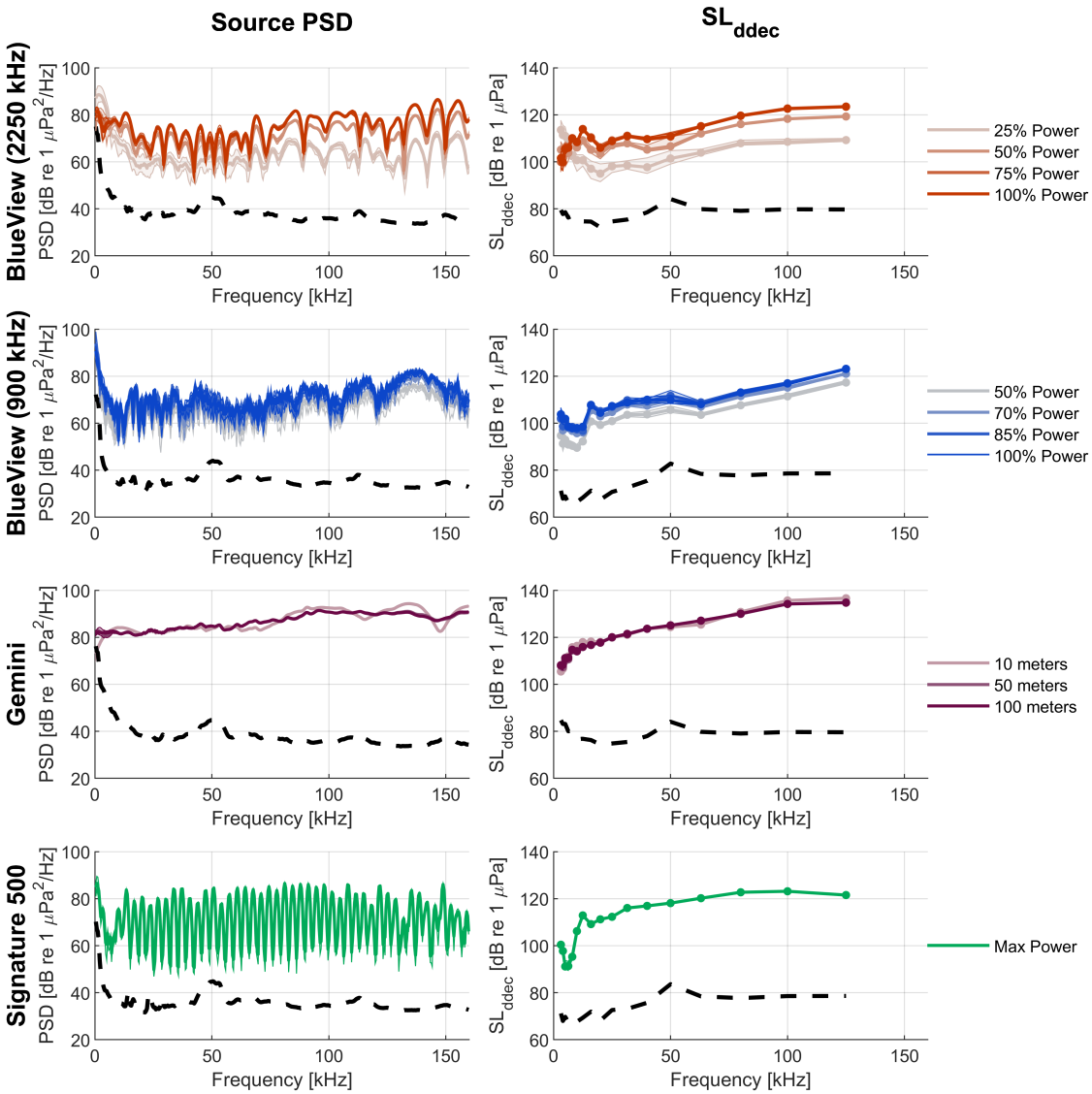

Figure 5.5: Median mean-square sound pressure spectral densities measured at the center of the sonar swath or Signature 500 beam for each transducer mode. The left-hand column shows the estimated PSD at one meter range from the transducer, and the right-hand column shows decidecade sound pressure levels at the same range. The shaded region indicates the interquartile range for all pings, and the dashed line indicates the ambient noise level recorded closest to each test.

level. For example, for the 2250 kHz transducer, the decrease in source PSD from 75% to 50% power is smaller than the decrease from 50% to 25% (average decreases of 4 dB and 8 dB, respectively), but the source PSD for the 75% and 100% power modes are similar. The same trend is observed for the 900 kHz BlueView transducer. This non-linear relationship between power level and source level is consistent with Equation 5.1.

The BlueView (both transducers) and Gemini also produced low-amplitude sound within the range of marine mammal hearing behind the transducer (aft hydrophone), most significantly for the 2250 kHz BlueView. As with the fore measurement, PSD increased with power level for both BlueView transducers. For the Gemini, the 50-meter and 100-meter range modes again had similar PSDs, while the 10-meter range was not detectable above ambient noise behind the transducer. The Signature 500 also was not detectable behind the transducer. For all transducers, the received PSD behind the transducer (0.79 meter range) did not exceed 20 dB above the ambient levels at any measured frequency.

5.3.2 Beam Patterns

Figure 5.6 shows the across-swath and along-swath variations in SL (Equation 5.7) for each sonar at the operating mode corresponding to the maximum SL. The dashed line indicates the median ambient level closest to the time of recording. Several observations can be made. First, for the BlueView and Gemini, SL at the edge of the nominal swath is lower in the along-swath direction, and rolls off more quickly outside of the nominal swath in the along-swath direction than in the across-swath direction. Second, distinct trends can be observed in the beam pattern of each instrument. For the Signature 500, the center is the highest-amplitude point and SL rolls off towards the edge of the beam. As a consequence of increased transducer complexity in the along-swath direction, the Gemini and BlueView have greater variability. For the Gemini, the maximum SL is observed near the edge of the swath. For the BlueView 2250 kHz, there are local maxima in SL throughout the swath, likely because three synchronized transducers are used to produce a single acoustic image. Asymmetry in the beam pattern is likely due to angular positioning error in sensor mounting (e.g., -5^o and

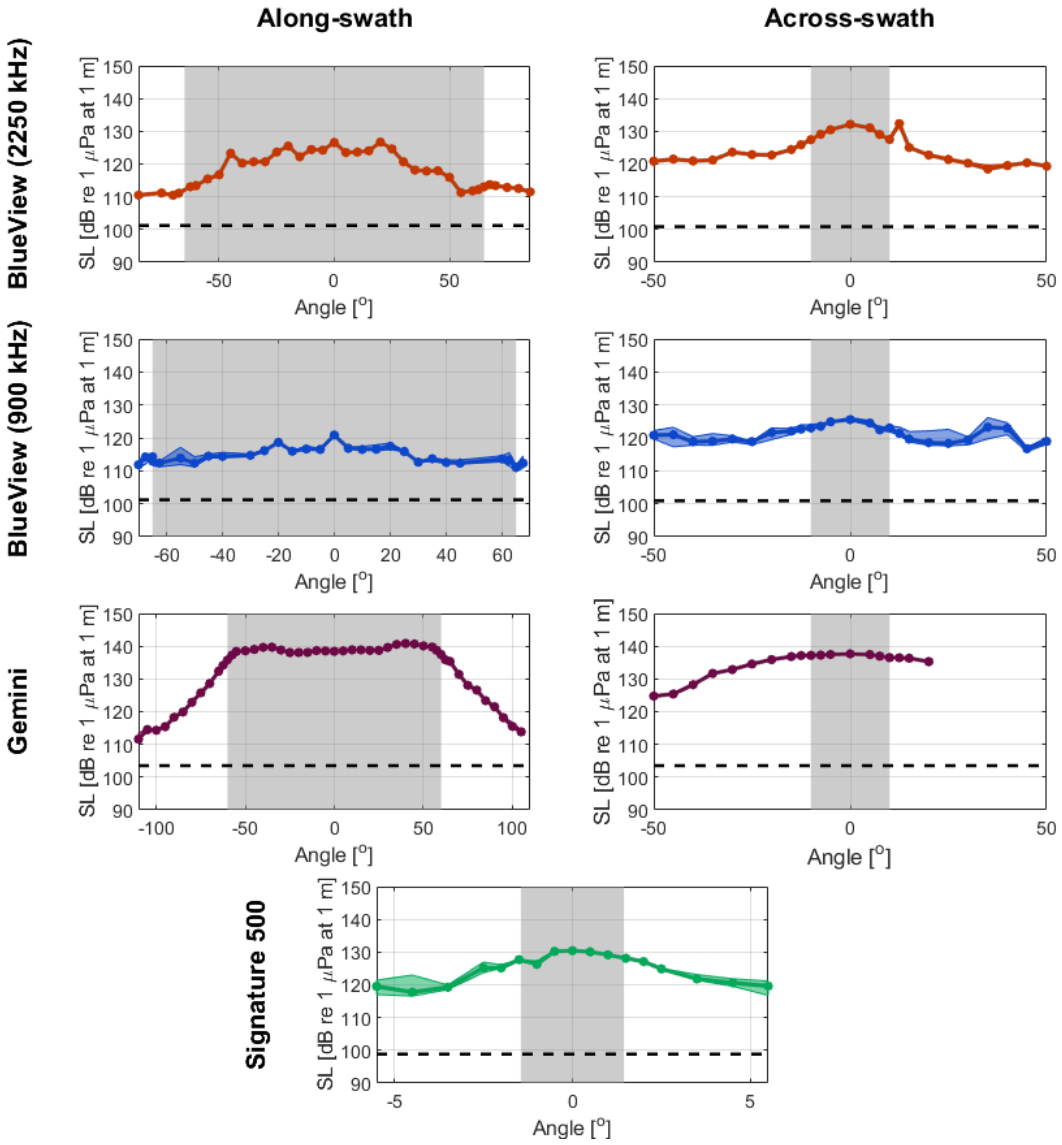

Figure 5.6: Median source level (unweighted, 0-160 kHz band) for ten pings recorded with the fore hydrophone at each orientation. The left-hand column shows the along-swath levels and the right-hand column shows the across-swath levels. The grey shaded region indicates the nominal swath or beamwidth of each transducer at the nominal operating frequency, and the colored shaded region indicates the interquartile range for the source level. The black dashed line indicates the ambient received level at the time of recording. The deviations between along-swath and across-swath measurements at 0^o are most likely attributable to alignment error in positioning the transducers (see Section 5.3.3).

5^o were not equidistant from the center axis) and the relatively coarse resolution in across-swath and along-swath sweeps, such that sharp gradients within the swath may not have been resolved.

5.3.3 Experimental Repeatability

Because the test set-up did not allow for precise sonar alignment, there is some inter-test variability. For example, the BlueView was removed and re-mounted to the hydraulic ram between the along-swath and across-swath characterization. This resulted in an increase in estimated SL at the beam center in the vertical measurements. This indicates that the zero-degree position in the along-swath direction was not perfectly aligned with the corresponding zero-degree position in the across-swath position. Similarly, the source PSD curves reported for the different operating modes are at the nominal beam center in the across and along swath direction. This suggests that the measurements are sensitive to the precise position in the beam and that some results here may slightly under-report the peak values within the swath. More precise measurements could be obtained by a rigid mounting system directly coupling the active acoustic sensor to the hydrophone.

5.3.4 Cumulative Sound Exposure Level

Figure 5.7 shows the percentage of SEL_{TTS} as a function of range and time spent at the angular position within the swath/beam with the maximum SL. It is emphasized that these results are for a case in which boundary interactions are negligible. Results are only shown for high-frequency cetaceans, because they were the only one of the five NMFS marine mammal hearing groups that exceeded the TTS threshold within T = 1 day of exposure (the maximum integration time recommended for a continuous source [45]). Of the sonars tested, the Gemini could cause a temporary threshold shift in high-frequency cetaceans first, at approximately T = 2.6 hours at 1 meter from the sonar. For context, at a range of 1 meter, the Gemini swath is only 35 cm in the along-swath direction, and therefore an animal larger than the ensonified area would need to remain stationary for 2.6 hours before experiencing TTS. The

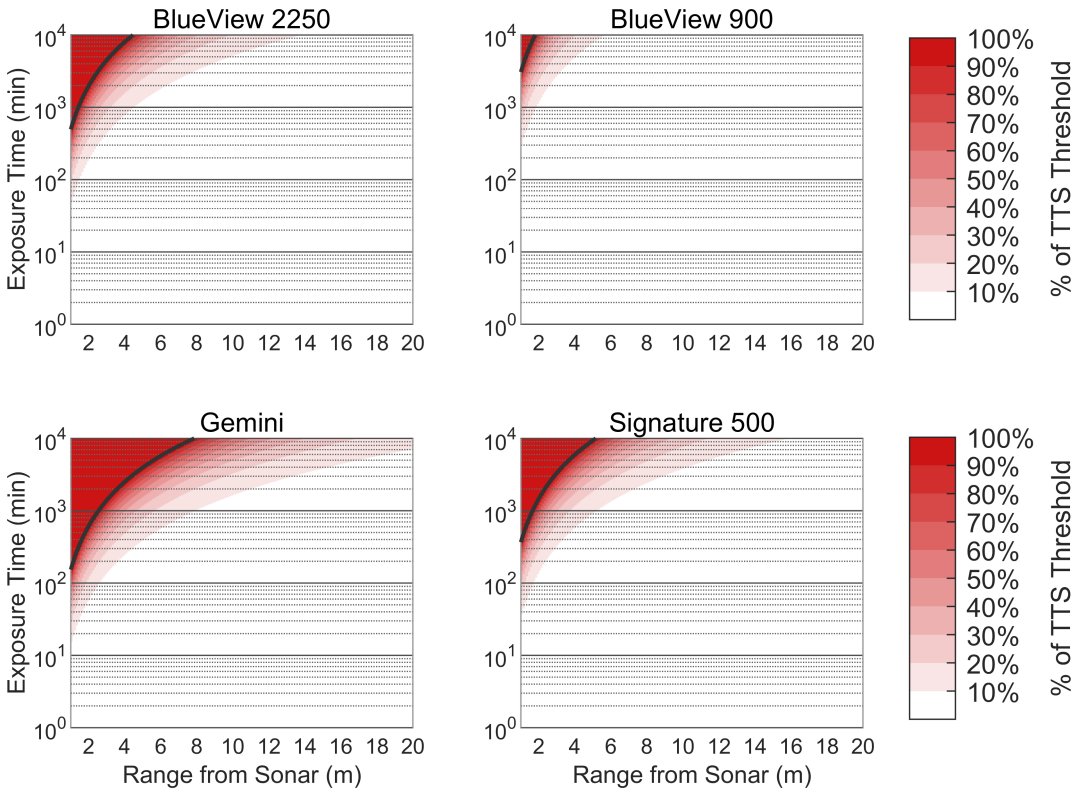

Figure 5.7: Percent of SEL_{cum} TTS threshold for high-frequency cetaceans, as a function of time and range from the sonar. The black line indicates T_{TTS}, the time at exceedance of the TTS threshold for a high-frequency cetacean in the portion of the swath with the highest source level.

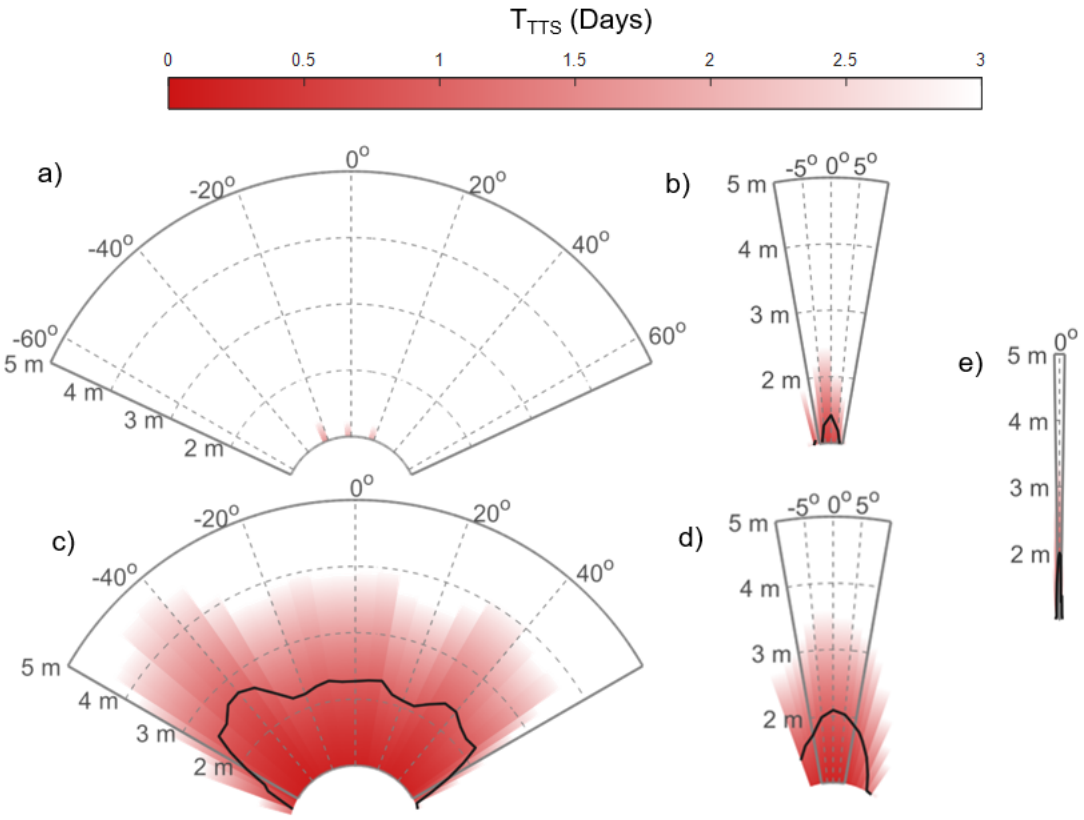

Figure 5.8: The exposure time, T_{TTS}, before the NMFS SEL_{cum} threshold for high-frequency cetaceans is exceeded as a function of position relative to the transducer. All transducer swaths are shown to the same scale. The black line indicates where T_{TTS} exceeds 24 hours. a) and b) show the along and across-swath direction of the 2250 kHz BlueView, respectively, c) and d) show the along and across-swath directions of the Gemini, respectively, and e) shows the symmetrical swath of the Signature 500. The 900 kHz BlueView is not shown because T_{TTS} exceeds 24 hours at all ranges and positions. As for Figure 5.6, the disparity between T_{TTS} in the across- and along-swath orientations for the BlueView is most likely a consequence of a minor alignment error.

Signature 500, 2250 kHz BlueView, and 900 kHz BlueView would exceed the TTS threshold at 1-meter range at approximately 6.1, 8.3, and $>$ 24 hours of exposure, respectively.

While all four of the transducers produced measurable sound within the range of marine mammal hearing, the SEL_{cum} metric indicates that in an operational context, these transducers are unlikely to affect hearing sensitivity for any of the five marine mammal hearing groups when operated continuously from a stationary platform. Figure 5.8 shows the time, T_{TTS}, that a high-frequency cetacean would need to spend at a given point in the sonar swath before TTS is expected. The black line indicates the range where T_{TTS} is equal to 1 day. The region of the swath where T_{TTS} would be expected within 24 hours is less than 3 meters for all sonars. For the Gemini, the transducer with the highest amplitude out-of-band emissions, the ensonified volume where SEL_{cum} exceeds the TTS threshold within one day is less than 5 m^3. Even when considering that the Signature 500 is typically operated with 4 or 5 diverging beams, the comparable volume is less than 0.03 m^3 because the individual beams are relatively narrow. While cumulative sound exposure levels are only presented for the most sensitive marine mammal hearing group, the supplemental material provided with [11] can be used to calculate the cumulative sound exposure level for any of the five marine mammal hearing groups at any point in the beam when surface reflections are negligible.

5.3.5 Audibility to Marine Mammals

Because the sound produced by these sonars is likely to be audible to marine mammals, is it possible that behavioral changes could occur, even when SEL_{cum} is below the TTS threshold. Consequently, if active acoustic sensors are used for environmental monitoring of marine mammals, the potential for animal behavior to be affected by sonar emissions should be considered during study design and when interpreting results.

Table 5.5 lists the maximum range at which a marine mammal from each hearing group might be able to detect each sonar, given hearing thresholds and relatively low ambient noise. The Gemini could be audible to high-frequency cetaceans up to a range of 108.5 meters, the longest range at which any of the tested sonars might be detectable by a marine

Sensor	**Low-frequency cetaceans**	**Mid-frequency cetaceans**	**High-frequency cetaceans**	**Otariid pinnipeds**	**Phocid pinnipeds**
BlueView (2250 kHz)	82.5	27.0	53.0	9.5	52.0
BlueView (900 kHz)	41.0	12.0	34.0	3.5	14.0
Gemini	20.0	29.5	108.5	4.5	21.5
Signature 500	61.5	33.0	68.0	7.0	30.0

Table 5.5: Maximum range, in meters, at which a marine mammal from each hearing group may be able to detect each sonar.

mammal. Counterintuitively, both BlueView transducers and the Signature 500 are audible to low-frequency cetaceans at a longer range than to mid-frequency cetaceans, and, in the case of the BlueView transducer, at a longer range than to high-frequency cetaceans. This can be attributed to the peaks in source level observed at low frequencies (< 20 kHz) for these transducers (see Figure 5.5). At these frequencies, the composite audiograms for low-frequency cetaceans suggest a higher sensitivity than for mid-frequency or high-frequency cetaceans. Further, for some hearing groups, the maximum range of audibility is similar to, or exceeds, the functional range of some of the multibeam sonars. In other cases, the maximum range of audibility is a small fraction of the sonar range.

These estimates are consistent with a previous study [27] which showed that harbor porpoises (a high-frequency cetacean) changed their surfacing patterns in a pool when a Gemini was operated, out to a range of 40 meters. Because audibility is necessary, but not sufficient to cause behavioral change, the values in Table 5 should be taken as conservative estimates of the maximum range for potential changes in behavior. Further, the ambient

noise levels in coastal environments, such as a marine energy site, would likely be higher than those considered here, resulting in reduced hearing ranges, and it is recommended that a site-specific propagation be used to evaluate hearing ranges for specific conditions. For example, [27] estimated that the Gemini could be audible to harbor porpoises (a high frequency cetacean) to a range of only 60 meters at a tidal energy site.

5.4 Conclusions

The active acoustic transducers characterized in this study all have nominal operating frequencies above 500 kHz. While all of the transducers produced out-of-band sound audible by marine mammals (i.e., at frequencies less than 160 kHz) at some range, cumulative sound exposure levels are unlikely to cause hearing threshold shifts for any marine mammal group. If acoustic emissions could be a confounding factor for a behavioral study, it has been shown that these may be decreased by adjusting the sonar operating mode. Overall, out-of-band sound should be considered when sonars are used to study marine mammals, but the measurements presented here show that these instruments should not raise regulatory concerns under 2018 NMFS guidelines [45].

Chapter 6

CONCLUSIONS AND FUTURE WORK

6.1 Conclusions

Effective methods for environmental monitoring are essential to the continued sustainable development of marine renewable energy. To be most effective, an environmental monitoring platform must meet three directives: 1) do not bias the animals being monitored through the use of sensors, 2) capture rare events, and 3) minimize the collection of data that does not contain useful information. The work presented in this dissertation has demonstrated that it is possible to develop an integrated instrumentation platform that simultaneously meets all three of these directives. This is achieved through the development of automatic, real-time data processing for multibeam sonars and careful analysis of the sensors used for monitoring.

6.1.1 Benefits of Third Generation Integrated Instrumentation

In Chapters 3 and 4, it was demonstrated that it is feasible to develop a third generation integrated instrumentation platform that meets all three directives of environmental monitoring, and that the benefits of such a platform outweigh the costs. This was demonstrated using the Adaptable Monitoring Package, an integrated instrumentation system that combines active acoustic sensors, passive acoustic sensors, and optical cameras. Real-time detection and tracking of targets in multibeam sonar data was used to limit data acquisition and the use of artificial illumination to periods when targets were present (i.e., meeting directives 1, 2, and 3). This approach to data acquisition significantly reduced the volume of data that was archived and subsequently reviewed by a human observer, while capturing an estimated 99% of targets of interest. As a result, it was possible to annotate a training data set for

machine learning classification of the tracked targets with relatively little effort compared to human review of continuously acquired data. Using these training data, a classification model was able to distinguish between biological and non-biological targets with a 97% true positive rate. This capability was demonstrated both in real-time and in post-processing. The classification model was largely site-specific, and new training data were required for effective classification at a second deployment site. However, when real-time target detection and tracking are used to control data acquisition, it is relatively straightforward to build a training data set for classification of targets in multibeam sonar data at a new marine energy site. In practice, third generation instrumentation platforms like the Adaptable Monitoring Package will enable effective collection of the data necessary to understand the environmental risks of marine renewable energy (e.g., collision of marine fauna with marine energy converters).

6.1.2 Audibility of Sonars to Marine Mammals

An in-depth analysis of the acoustic emissions of four active acoustic transducers used for environmental monitoring was conducted in Chapter 5. All four transducers had nominal operating frequencies well above the reported upper limit of marine mammal hearing (160 kHz), but were found to produce measurable sound below this limit. Acoustic emissions within the range of marine mammal hearing were analyzed using the US National Marine Fisheries Service technical guidance for assessing the effects of anthropogenic sound on marine mammals. It was found that while it is unlikely that any of the tested transducers will cause hearing damage to any marine mammals, they could be audible for tens of meters, potentially causing changes in animal behavior. Therefore, acoustic emissions from sonars with operating frequencies above the range of marine mammal hearing should be considered when conducting environmental monitoring of marine mammals, but should not be cause for regulatory concern. This evaluation provides the necessary foundation to develop an integrated instrumentation platform that meets the first directive of environmental monitoring.

6.2 Areas for Future Research

With the third generation capabilities developed in this thesis, the Adaptable Monitoring Package provides a platform for countless areas of future research. While an exhaustive list of potential applications of this system is not possible here, three possibilities are discussed in this section.

6.2.1 Classification of Fish using Multibeam Sonars

An obvious gap in the work presented in Chapter 4 is the classification of individual fish in multibeam sonar data. This was not possible at the Marine Science Laboratory (MSL) test site because the resolution of the multibeam sonar was frequently not sufficient for a human reviewer to distinguish individual fish from kelp, eelgrass, or other small objects in the water column, and the relatively turbid water conditions prevented reliable concurrent detection in the optical camera data. There are two possible approaches to obtain the data necessary to train a model for classification of individual fish using the Adaptable Monitoring Package. At a site where optical camera data is not reliable, like MSL, multibeam sonar data stream could be fused with data from a traditional fisheries echosounder, a sensor for which there are existing methods for fish detection and classification. Concurrent detection of fish on the echosounder and multibeam sonar could be used to build a training data set for fish classification in the multibeam sonar data. An echosounder has already been integrated into the MSL-2 AMP, though the data stream has not yet been integrated into any real-time processing modules. At a test site with high water clarity, such as the Wave Energy Test Site in Hawaii, the same objective could be achieved through the fusion of the multibeam sonar data with optical camera imagery. This approach would likely require more extensive human review or algorithm development, because methods for detection and classification of fish are relatively well established for echosounder data, but are active areas of research for optical camera data.

6.2.2 Removal of Sonar Artifacts

In Chapter 4, all non-biological targets are grouped into a single class. This includes targets related to the environment, such as "clouds" of sediment or turbulent backscatter, as well as artifacts of the multibeam sonar acoustic processing, such as high-intensity reflections that appear around targets with a strong acoustic return. While this approach worked relatively well, these two types of non-biological targets are fundamentally different, and false-positive rates would likely decrease if they were treated separately. Sonar artifacts can also impede target tracking when they are detected concurrently with targets of interest, reducing the quality of the training data used for classification. Therefore, if these artifacts could be removed before target detection and tracking, it would likely improve classification performance. Potential approaches to this problem include building a model to predict the position of sonar artifacts based on target position and intensity or using adaptive thresholding methods to remove artifacts that are lower intensity than the real target.

6.2.3 Animal Response to Active Acoustic Sensors

In Chapter 5, it was found that additional research is required to understand the potential effects of sound produced by active acoustic sensors on marine mammal behavior. The Adaptable Monitoring Package is an ideal platform to study these effects, because it enables concurrent deployment of multiple active acoustic sensors and can be used to automatically cycle their operation. For example, operation of the BlueView and Gemini multibeam sonars could be alternated, and the operating multibeam sonar (e.g., BlueView) could be used to observe any changes in marine mammal behavior that occur when the second multibeam sonar (e.g., Gemini) is disabled. The target detection, tracking, and classification methods described in this thesis could be used to rapidly quantify marine mammal presence, though interpretation of results will require biological expertise. The results of such a study could be used to inform selection of active acoustic sensors for environmental monitoring at marine energy sites where marine mammals are expected to be present.

BIBLIOGRAPHY

[1] FORCE completes two underwater monitoring platforms. *Tidal Energy Today*, 2015.

[2] R. Atan, J. Goggins, and S. Nash. A Detailed Assessment of the Wave Energy Resource at the Atlantic Marine Energy Test Site. *Energies*, 9(11):967, 2016.

[3] A. Baccini, M. A. Friedl, C. E. Woodcock, and R. Warbington. Forest biomass estimation over regional scales using multisource data. *Geophysical Research Letters*, 31(L10501), 2004.

[4] C. Bassett, J. Thomson, P. H. Dahl, and B. Polagye. Flow-noise and turbulence in two tidal channels. *The Journal of the Acoustical Society of America*, 135(4):1764–1774, 2014.

[5] L. Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001.

[6] M. Broadhurst and C. David L. Orme. Spatial and temporal benthic species assemblage responses with a deployed marine tidal energy device: A small scaled study. *MArine Environmental Research*, 99:76–84, 2014.

[7] C. J. C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 2:121–167, 1998.

[8] A. Copping, N. Sather, L. Hanna, J. Whiting, G. Zydlewski, G. Staines, A. Gill, I. Hutchinson, A. O'Hagan, T. Simas, J. Bald, C. Sparling, J. Wood, and E. Masden. Annex IV 2016 State of the Science Report: Environmental Effects of Marine Renewable Energy Development Around the World. Technical report, 2016.

[9] A. E. Copping and M. E. Grear. Applying a simple model for estimating the likelihood of collision of marine mammals with tidal turbines. *International Marine Energy Journal*, 1(1):27–34, 2018.

[10] E. Cotter, S. Matzner, J. Horne, P. Murphy, and B. Polagye. Benchmarking sensor fusion capabilities of an integrated monitoring package. In *Proceedings of the 4th Marine Energy Technology Symposium*, Washington, D.C., 2016.

[11] E. Cotter, P. Murphy, C. Bassett, B. Williamson, , and B. Polagye. Acoustic characterization of sensors used for marine environmental monitoring. *Marine Pollution Bulletin. In Review.*, 2019.

[12] E. Cotter, P. Murphy, and B. Polagye. Benchmarking sensor fusion capabilities of an integrated instrumentation package. *International Journal of Marine Energy*, 2017.

[13] E. Cotter, B. Williamson, and B. Polagye. Challenges To Integrating Active Acoustic Sensors. In *Proceedings of the 3rd Marine Energy Technology Symposium METS2015 April 27-29, 2015, Washington, D.C.*, pages 1–5, Washington, D.C., 2015.

[14] D. A. Demer, L. Berger, M. Bernasconi, E. Bethke, K. M. Boswell, D. Chu, R. Domokos, A. Dunford, S. Fässler, S. Gauthier, L. T. Hufnagle, J. M. Jech, N. Bouffant, A. Lebourges-Dhaussy, X. Lurton, G. J. Macaulay, Y. Perrot, T. Ryan, S. Parker-Stetter, S. Stienessen, T. Weber, and N. Williamson. *Calibration of acoustic instruments*, volume 326. International Council for the Exploration of the Sea, 2015.

[15] Z. D. Deng, B. L. Southall, T. J. Carlson, J. Xu, J. J. Martinez, M. A. Weiland, and J. M. Ingraham. 200 kHz commercial sonar systems generate lower frequency side lobes audible to some marine mammals. *PLoS ONE*, 9(4):1–6, 2014.

[16] G. J. Dobeck, J. C. Hyland, and L. Smedley. Automated Detection/Classification of Sea Mines in Sonar Imagery. In *Proceedings of SPIE*, volume 3079, pages 90–110, Orlando, FL, 1997.

[17] K. Duan and S. S. Keerthi. Which Is the Best Multiclass SVM Method? An Empirical Study. In N. C. Oza, R. Polikar, J. Kittler, and F. Roli, editors, *Multiple Classifier Systems, 6^{th} International Workshop*, pages 278–285, 2005.

[18] S. A. Dudani. The Distance-Weighted k-Nearest-Neighbor rule. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6(4):325–327, 1976.

[19] C. Erbe, C. Reichmuth, K. Cunningham, K. Lucke, and R. Dooling. Communication masking in marine mammals: A review and research strategy. *Marine Pollution Bulletin*, 103:15–38, 2016.

[20] P. G. Fernandes, F. Gerlotto, V. D. Holliday, O. Nakken, and E. J. Simmonds. Acoustic applications in fisheries science: the ICES contribution. *ICES Marine Science Symposia*, 215:483–492, 2002.

[21] F. Francisco and J. Sundberg. Detection of Visual Signatures of Marine Mammals and Fish within Marine Renewable Energy Farms using Multibeam Imaging Sonar. *Journal of Marine Science and Engineering*, 7(22):1–19, 2019.

[22] R. E. Francois and G. R. Garrison. Sound absorption based on ocean measurements: Part I: Pure water and magnesium sulfate contributions,. *Journal of the*, 72(3):869–907, 1982.

[23] R. E. Francois and G. R. Garrison. Sound absorption based on ocean measurements. Part II: Boric acid contribution and equation for total absorption. *Journal of the Acoustical Society of America*, 72(6):1879–1890, 1982.

[24] M. J. Hamel, M. L. Brown, and S. R. Chipps. Behavioral Responses of Rainbow Smelt to in situ Strobe Lights. *North American Journal of Fisheries Management*, 28:394–401, 2008.

[25] L. Hammar, L. Eggertsen, S. Andersson, J. Ehnberg, R. Arvidsson, M. Gullström, and S. Molander. A probabilistic model for hydrokinetic turbine collision risks: Exploring impacts on fish. *PLoS ONE*, 10(3):1–25, 2015.

[26] R. C. Hasan, D. Ierodiaconou, and J. Monk. Evaluation of Four Supervised Learning Methods for Benthic Habitat Mapping Using Backscatter from Multi-Beam Sonar. *Remote Sensing*, 4:3427–3443, 2012.

[27] G. Hastie. Tracking marine mammals around marine renewable energy devices using active sonar. *SMRU Ltd. report URN 12:D/328 to the Department of Energy and Climate Change.*, 2012.

[28] G. Hastie, D. Gillespie, J. Gordon, J. Macaulay, B. McConnel, and C. Sparling. Tracking Technologies for Quantifying Marine Mammal Interactions with Tidal Turbines: Pitfalls and Possibilities. In *Marine Renewable Energy Technology and Environmental Interactions*, chapter 10, pages 127–139. Springer Netherlands, 2014.

[29] G. D. Hastie, D. J. F. Russell, P. Lepper, J. Elliott, B. Wilson, S. Benjamins, and D. Thompson. Harbour seals avoid tidal turbine noise: Implications for collision risk. *Journal of Applied Ecology*, 55:684–693, 2018.

[30] K. Herkül, A. Peterson, and S. Paekivi. Applying Multibeam Sonar and Mathematical Modeling for Mapping Seabed Substrate and Biota of Offshore Shallows. *Estuarine, Coastal and Shelf Science*, 192:57–71, 2017.

[31] ISO/TC 43/SC 3 Underwater Acoustics. ISO 18405: Underwater Acoustics - Terminology. Technical report, International Organization for Standardization, 2017.

[32] P. Jepp. Target Tracking Using Sonars for Marine Life Monitoring around Tidal Turbines. In *Proceedings of the 12th European Wave and Tidal Energy Conference*, Cork, Ireland, 2017.

[33] J. Joslin, E. Celkis, C. Roper, A. Stewart, and B. Polagye. Development of an Adaptable Monitoring Package for Marine Renewable Energy Projects Part II: Hydrodynamic Performance. In *OCEANS*, San Diego, CA, 2013. IEEE.

[34] J. Joslin and B. Polagye. Demonstration of Biofouling Mitigation Methods for Long-Term Deployments of Optical Cameras. *Marine Technology Society Journal*, 49(1):88–96, 2015.

[35] J. Joslin, B. Polagye, and S. Parker-Stetter. Development of a stereo-optical camera system for monitoring tidal turbines. *Journal of Applied Remote Sensing*, 8(1):083633, 2014.

[36] R. Kohavi and G. H. John. Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2):273–324, 1997.

[37] S. Kotsiantis, D. Kanellopoulos, and P. Pintelas. Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering*, 30:25–36, 2006.

[38] L. Lieber, T. Nilsen, C. Zambrano, and L. Kregting. Optimising multiple multibeam sonars to assess marine life interactions with an underwater kite. In *Proceedings of the 12th European Wave and Tidal Energy Conference*, pages 1–6, Cork, Ireland, 2017.

[39] M. Marchesan, M. Spoto, L. Verginella, and E. A. Ferrero. Behavioural effects of artificial light on fish species of commercial interest. *Fisheries Research*, 73:171–185, 2005.

[40] W. Marczak. Water as a standard in the measurements of speed of sound in liquids. *Journal of the Acoustical Society of America*, 102(5):2776–2779, 1997.

[41] P. Maybeck. Introduction. In *Stocastic models, estimation, and control*, pages 1–19. Academic Press, 1979.

[42] H. Medwin and C. Clay. *Fundamentals of Acoustical Oceanography.* Academic Press, 1997.

[43] G. D. Melvin and N. A. Cochrane. Multibeam Acoustic Detection of Fish and Water Column Targets at High-Flow Sites. *Estuaries and Coasts*, 38(1):227–240, 2015.

[44] R. P. Morrissey, J. Ward, N. Dimarzio, S. Jarvis, and D. J. Moretti. Passive acoustic detection and localization of sperm whales (Physeter macrocephalus) in the tongue of the ocean. *Applied Acoustics*, 67:1091–1105, 2006.

[45] National Marine Fisheries Service. 2018 Revision to: Technical Guidance for Assessing the Effects of Anthropogenic Sound on Marine Mammal Hearing: Underwater Acoustic Thresholds for Onset of Permanent and Temporary Threshold Shifts. Technical report, 2018.

[46] Normandeau Associates, Inc. Effects of Noise on Fish, Fisheries, and Invertebrates in the U.S. Atlantic and Arctic from Energy Industry Sound-Generating Activities. A Workshop Report for the U.S. Dept. of the Interior, Bureau of Ocean Energy Management. Contract M11PC00031. Technical report, 2012.

[47] J. C. Platt. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In *Advances in Large Margin Classifiers*, pages 61–74. MIT Press, 1999.

[48] B. Polagye, A. Copping, R. Suryan, J. Brown-Saracino, and C. Smith. Instrumentation for monitoring around marine renewable energy converters: Workshop final report. Technical Report PNNL-23110, Pacific Northwest National Laboratory, Seattle, WA, 2014.

[49] B. Polagye and J. Thomson. Tidal energy resource characterization: methodology and field study in Admiralty Inlet, Puget Sound, WA (USA). *Proceedings of the Institution of Mechanical Engineers, Part A: Journal of Power and Energy*, 227(3):352–367, 2013.

[50] C. D. Pyć, M. Geoffroy, and R. R. Knudsen. An evaluation of active acoustic methods for detection of marine mammals in the Canadian Beaufort Sea. *Marine Mammal Science*, 32(1):202–219, 2016.

[51] D. Risch, B. Wilson, and P. Lepper. Acoustic Assessment of SIMRAD EK60 High Frequency Echo Sounder Signals (120 & 200 kHz) in the Context of Marine Mammal Monitoring. Technical Report 13, 2017.

[52] C. N. Rooper, K. Williams, A. De Robertis, and V. Tuttle. Effect of underwater lighting on observations of density and behavior of rockfish during camera surveys. *Fisheries Research*, 172:157–167, 2015.

[53] I. M. D. Rosa, A. T. Marques, G. Palminha, H. Costa, M. Mascarenhas, C. Fonseca, and J. Bernardino. Classification success of six machine learning algorithms in radar ornithology. *International Journal of Avian Science*, 158:28–42, 2016.

[54] P. Rothery and D. B. Roy. Application of generalized additive models to butterfly transect count data. *Journal of Applied Statistics*, 28(7):897–909, 2010.

[55] B. Rush, J. Joslin, A. Stewart, and B. Polagye. Development of an Adaptable Monitoring Package for Marine Renewable Energy Projects, Part 1: Conceptual Design and Operation. In *Proceedings of the 2nd Marine Energy Technology Symposium*, Seattle, WA, 2014.

[56] M. P. Schramm, M. Bevelhimer, and C. Scherelis. Effects of hydrokinetic turbine sound on the behavior of four species of fish within an experimental mesocosm. *Fisheries Research*, 190:1–14, 2017.

[57] SCOS. Scientific Advice on Matters Related to the Management of Seal Populations: 2013. Technical report, Scottish Natural Environment Research Council, 2013.

[58] J. Simmonds and D. MacLennan. *Fisheries Acoustics: Theory and Practice*. Blackwell Publishing Ltd, 2^{nd} edition, 2007.

[59] F. Song, Z. Guo, and D. Mei. Feature selection using principal component analysis. In *International Conference on System Science, Engineering Design and Manufacturing Informatization*, volume 1, pages 27–30. IEEE, 2010.

[60] B. L. Southall, S. L. Deruiter, A. Friedlaender, A. K. Stimpert, J. A. Goldbogen, E. Hazen, C. Casey, S. Fregosi, D. E. Cade, A. N. Allen, C. M. Harris, G. Schorr, D. Moretti, S. Guan, and J. Calambokidis. Behavioral responses of individual blue whales (Balaenoptera musculus) to mid-frequency military sonar. *Journal of Experimental Biology*, 222:1–15, 2019.

[61] J. Thomson, B. Polagye, V. Durgesh, and M. C. Richmond. Measurements of turbulence at two tidal energy sites in puget sound, WA. *IEEE Journal of Oceanic Engineering*, 37(3):363–374, 2012.

[62] J. Tougaard and M. Dähne. Why is auditory frequency weighting so important in regulation of underwater noise? *Journal of the Acoustical Society of America*, 142:415–420, 2017.

[63] P. Urban, K. Köser, and J. Greinert. Processing of multibeam water column image data for automated bubble / seep detection and repeated mapping. *Limnology and Oceanography: Methods*, 15:1–21, 2017.

[64] R. Vennell. Acoustic Doppler Current Profiler measurements of tidal phase and amplitude in Cook Strait, New Zealand. *Continental Shelf Research*, 14(4):353–364, 1994.

[65] G. M. Wenz. Acoustic Ambient Noise in the Ocean: Spectra and Sources. *Journal of the Acoustical Society of America*, 34(12):1936–1956, 1962.

[66] T. A. Wilding, A. B. Gill, A. Boon, E. Sheehan, J. Dauvin, J. Pezy, F. O'Biern, U. Janas, L. Rostin, and I. De Mesel. Turning off the DRIP ('Data-rich, Information-poor') - rationalising monitoring with a focus on marine renewable energy developments and the benthos. *Renewable and Sustainable Energy Reviews*, 74:848–859, 2017.

[67] B. Williamson, S. Fraser, P. Blondel, P. S. Bell, J. J. Waggit, and B. E. Scott. Multisensor Acoustic Tracking of Fish and Seabird Behavior Around Tidal Turbine Structures in Scotland. *IEEE Journal of Ocean Engineering*, PP(99):1–18, 2017.

[68] B. J. Williamson, P. Blondel, E. Armstrong, P. S. Bell, C. Hall, J. J. Waggitt, and B. E. Scott. A Self-Contained Subsea Platform for Acoustic Monitoring of the Environment Around Marine Renewable Energy Devices-Field Deployments at Wave and Tidal Energy Sites in Orkney, Scotland. *IEEE Journal of Oceanic Engineering*, 41(1):67–81, 2016.

[69] M. F. J. Wilson, B. O'Connell, C. Brown, J. C. Guinan, and A. J. Grehan. Multiscale Terrain analysis of Multibeam Bathymetry Data for Habitat Mapping on the Continental Slope. *Marine Geodesy*, 30(1-2):3–35, 2007.

[70] M. Zhang and Z. Zhou. A k -Nearest Neighbor Based Algorithm for Multi-label Classification. In *GnC'05: IEEE International Conference on Granular Computing*, pages 718–721, New York, 2005. IEE.

Appendix A

FEATURES USED FOR TRACK CLASSIFICATION

The tables that follow list the 28 features used for classification of target tracks in Chapter 4.

Feature	**Abbv.**	**Type**	**Description**
Extent	E	Shape	75^{th} percentile of the extent of each target associated with the track. Extent is calculated to the ratio of the total number of pixels in a target to the area of the target bounding box.
Extent Standard Deviation	E_{std}	Shape	Standard deviation of the extent of each target associated with the track.
Solidity	SO	Shape	75^{th} percentile of the solidity of each target associated with the track. Solidity is calculated as the ratio of the number of pixels in a target to the area of the smallest polygon that can contain the target.
Equivalent Diameter	ED	Shape	75^{th} percentile of the equivalent diameter of each target associated with the track.

Feature	**Abbv.**	**Type**	**Description**
Axis Ratio	AR	Shape	Mean of the ratio of the major axis length to the minor axis length for each target associated with the track.
Minor Axis Length	Min	Shape	75^{th} percentile of the minor axis length of each target associated with the track. Minor axis length is defined as the length of the minor axis of the ellipse that has the same second moments as the target.
Major Axis Length	Min	Shape	75^{th} percentile of the major axis length of each target associated with the track. Major axis length is defined as the length of the major axis of the ellipse that has the same second moments as the target
Area Standard Deviation	A_{std}	Shape	Standard deviation of the area of each target associated with the track.
Maximum Area	A_{max}	Shape	Maximum area of any target associated with the track.
Direction Changes	nDC	Motion	Number of times that the acceleration of the target crosses zero.
Acceleration	Acc	Motion	75^{th} percentile of the acceleration of the tracked target.
Mean Acceleration	Acc_{mean}	Motion	Mean acceleration of the tracked target.

Feature	**Abbv.**	**Type**	**Description**
Distance Traveled	D_{tr}	Motion	Point-to-point distance traveled by the target over the duration of the target track
Change in Orientation	O_{std}	Motion	Standard deviation of the orientation of each target associated with the track. Orientation is calculated as the angle between the major axis and the x-axis of the image.
Relative Speed	dS	Motion	Average speed of the target relative to the current speed.
Speed	S	Motion	Average track speed. Track speed is by dividing the distance traveled between every five targets associated with the target by the elapsed time.
Duration	D	Motion	Duration of the target track.
Proximal Targets	NR	Image	75^{th} percentile of the number of targets concurrently detected within a 1 meter radius of each target associated with a track.
Concurrent Targets	N	Image	75^{th} percentile of the number of targets concurrently detected in the same image as each target associated with a track.
Range	R	Image	Average range from the sonar at which each target associated with the track was detected.
Time	T	Env.	Time of day when the target track was detected, calculated as the number of hours from midnight (e.g., both 11 PM and 1 AM give a value of 1).
Current Direction	$currDir$	Env.	Direction of the tidal current at the time of detection, as measured by the AMP ADCP

Feature	**Abbv.**	**Type**	**Description**
Current Speed	$currSpd$	Env.	Magnitude of the horizontal velocity of the tidal current at the time of detection, as measured by the AMP ADCP
Color Layout Descriptor	CLD	Intensity	75^{th} percentile of each value of the 2x2 color layout descriptor \cite{Kasutani2001} values for each target associated with the track. This feature results in 4 separate feature vectors for each target track.
Weighted Centroid Offset	dC	Intensity	75^{th} percentile of the distance between the centroid and weighted centroid of each target associated with a track. The weighted centroid is calculated based on the intensity of each pixel associated with a target.
Intensity Standard Deviation	I_{std}	Intensity	Standard deviation of intensity values of all targets associated with a track.
Intensity Maximum	I_{max}	Intensity	Maximum intensity of any pixel in any target associated with a track.
75^{th} Percentile Intensity	I	Intensity	75^{th} percentile of intensity values of all targets associated with a track

Appendix B

FEATURE SELECTION ALGORITHM

The hill-climbing feature selection algorithm was implemented as follows:

1. Initialize an empty feature list.

2. Evaluate the performance of the classification model (Section 4.3.2) with each individual feature added to the existing feature list. On the first iteration, classification performance is tested for each candidate feature in isolation.

3. Add the candidate feature that produced the highest value of TPR_{bio} in the test cases to the feature list.

4. Return to step (2), and repeat steps (2) and (3) until all features have been added to the list in a ranked order.

5. Select the subset of the feature list that yields the highest value of TPR_{bio} (i.e., remove features at the end of the sorted feature list if performance begins to degrade as more features are added).

6. Test removal of each feature from the selected feature list to determine if higher performance can be achieved. If removal of any single feature results in a higher value of TPR_{bio}, remove that feature from the feature list.

7. Repeat (5) until no features can be removed from the feature list without reducing TPR_{bio}.

For the ensemble configurations of KNN and SVM, a separate feature list was produced for each binary classification model. The hill-climbing algorithm was not implemented for the SVM model with a polynomial kernel function, because this kernel function requires more than one feature to operate.

Appendix C

EXTENDED CLASSIFICATION RESULTS

Tables C.1, C.2, and C.3 list the full set of classification metrics for all k-nearest neighbors, support vector machine, and random forest classification models that were tested in Chapter 4, respectively.

k	**Feature Method**	TPR_{bio}	TPR_{bin}	FPR_{bio}	TPR_N	TPR_S	TPR_{Sm}	TPR_B	TPR_F
Multiclass Models									
3	All	0.8	0.97	0.22	0.78	0.83	0.91	0.7	0.83
3	Wrapper	0.86	0.98	0.22	0.78	0.91	0.83	0.87	0.83
3	Filter	0.67	0.96	0.22	0.78	0.57	0.78	0.65	0.74
5	All	0.8	0.96	0.26	0.74	0.83	0.87	0.67	0.83
5	Wrapper	0.83	0.97	0.39	0.61	0.87	0.83	0.83	0.83
5	Filter	0.71	0.97	0.17	0.83	0.65	0.83	0.61	0.74
7	All	0.78	0.95	0.22	0.78	0.83	0.87	0.65	0.83
7	Wrapper	0.86	0.98	0.26	0.74	0.87	0.83	0.87	0.87
7	Filter	0.71	0.97	0.17	0.83	0.65	0.87	0.61	0.74
Ensemble Models									
3	All	0.8	0.95	0.24	0.76	0.85	0.78	0.7	0.83
3	Wrapper	0.89	0.97	0.17	0.83	0.91	0.87	0.78	0.96
3	Filter	0.77	0.96	0.22	0.78	0.74	0.83	0.7	0.83
5	All	0.77	0.96	0.26	0.74	0.78	0.87	0.61	0.78
5	Wrapper	0.88	0.97	0.17	0.83	0.87	0.89	0.78	0.96
5	Filter	0.79	0.96	0.22	0.78	0.78	0.78	0.74	0.83
7	All	0.79	0.97	0.3	0.7	0.83	0.89	0.65	0.83
7	Wrapper	0.89	0.98	0.26	0.74	0.87	0.87	0.83	1
7	Filter	0.76	0.97	0.17	0.83	0.7	0.87	0.7	0.83

Table C.1: Classification metrics for each KNN model that was evaluated.

.

Kernel Function	**Feature Method**	TPR_{bio}	TPR_{bin}	FPR_{bio}	TPR_N	TPR_S	TPR_{Sm}	TPR_B	TPR_F
linear	All	0.87	0.96	0.22	0.78	0.93	0.87	0.83	0.87
linear	Wrapper	0.76	0.96	0.13	0.87	0.78	0.83	0.8	0.65
linear	Filter	0.86	0.96	0.13	0.87	0.91	0.87	0.83	0.83
polynomial	All	0.78	0.96	0.26	0.74	0.87	0.83	0.7	0.78
polynomial	Filter	0.79	0.97	0.22	0.78	0.83	0.87	0.65	0.83
rbf	All	0.83	0.97	0.26	0.74	0.87	0.83	0.78	0.91
rbf	Wrapper	0.85	0.96	0.17	0.83	0.91	0.78	0.78	0.91
rbf	Filter	0.84	0.96	0.17	0.83	0.83	0.83	0.78	0.91

Table C.2: Classification metrics for each SVM model that was evaluated.

.

NumTrees	**Feature Method**	TPR_{bio}	TPR_{bin}	FPR_{bio}	TPR_N	TPR_S	TPR_{Sm}	TPR_B	TPR_F
40	All	0.87	0.97	0.09	0.91	0.91	0.83	0.85	0.91
40	Wrapper	0.86	0.96	0.17	0.83	0.91	0.83	0.83	0.87
40	Filter	0.71	0.96	0.13	0.87	0.7	0.78	0.61	0.78
70	All	0.88	0.96	0.13	0.87	0.91	0.87	0.87	0.91
70	Wrapper	0.88	0.97	0.13	0.87	0.91	0.83	0.87	0.91
70	Filter	0.73	0.97	0.17	0.83	0.7	0.74	0.65	0.83
100	All	0.88	0.97	0.09	0.91	0.91	0.87	0.87	0.91
100	Wrapper	0.87	0.97	0.22	0.78	0.91	0.83	0.87	0.91
100	Filter	0.72	0.96	0.09	0.91	0.7	0.78	0.65	0.78
130	All	0.86	0.96	0.17	0.83	0.87	0.83	0.83	0.91
130	Wrapper	0.9	0.98	0.13	0.87	0.87	0.91	0.91	0.91
130	Filter	0.7	0.96	0.13	0.87	0.65	0.7	0.65	0.78
160	All	0.89	0.97	0.13	0.87	0.91	0.87	0.87	0.91
160	Wrapper	0.87	0.96	0.13	0.87	0.87	0.8	0.87	0.96
160	Filter	0.71	0.96	0.09	0.91	0.65	0.74	0.65	0.83
190	All	0.88	0.97	0.13	0.87	0.91	0.87	0.87	0.87
190	Wrapper	0.89	0.97	0.17	0.83	0.93	0.83	0.87	0.91
190	Filter	0.72	0.96	0.13	0.87	0.65	0.83	0.65	0.78

Table C.3: Classification metrics for each RF model that was evaluated.

Appendix D

NEARFIELD CALCULATIONS FOR ACTIVE ACOUSTIC SENSORS

The nearfield extent of a line array transducer (used in both the BlueView and Gemini multibeam sonars) can be estimated by

$$r = \frac{L^2}{\lambda}, \tag{D.1}$$

where r is the extent of the nearfield, in meters, L is the length or aperture of the transducer, in meters, and λ is the wavelength of the measured sound waves [42]. Substituting $f = c\lambda$, where f is the measured frequency, in Hz, and c is the speed of sound, in m/s,

$$r = \frac{L^2 f}{c}, \tag{D.2}$$

This equation is used to estimate the nearfield extent at both the transducer nominal operating frequency and the highest measured frequency (160 kHz) for the BlueView and Gemini in Table D.2. For the Gemini, the manufacturer provided the nearfield extent estimate at the nominal operating frequency rather than the transducer aperture. This value was used to estimate the transducer aperture for the calculation at 160 kHz.

Medwin and Clay [42] also provide an estimate for the nearfield of a circular piston transducer:

$$r = \frac{\pi a^2}{\lambda}, \tag{D.3}$$

which can be equivalently expressed as

$$r = \frac{\pi a^2 f}{c}. \tag{D.4}$$

Instrument	**Manufacturer Provided Information**	**Nearfield extent at operating frequency (m)**	**Nearfield extent at 160 kHz (m)**
BlueView (900 kHz)	Transducer aperture, L = 0.075 m	3.4	0.61
BlueView (2250 kHz)	Transducer aperture, L = 0.032 m	3.4	0.11
Gemini (720 kHz)	Nearfield extends 5 m from transducer at operating frequency	5	1.1
Signature (500 kHz)	Circular piston transducer with a = 0.033 m radius	1.2	0.37

Table D.1: Manufacturer provided information and calculation of nearfield extent estimations for each transducer at its operating frequency and at the highest frequency reported in this paper (160 kHz) based on [42].

The latter equation is used to estimate the nearfield of the Signature 500 in Table D.2. An estimated speed of sound of 1486 m/s is used for all calculations. Because the nearfield extent is proportional to the measured frequency, the nearfield extent for frequencies less than 160 kHz will be smaller than the values reported in Table D.2.