Contextualized Monitoring in the Marine Environment

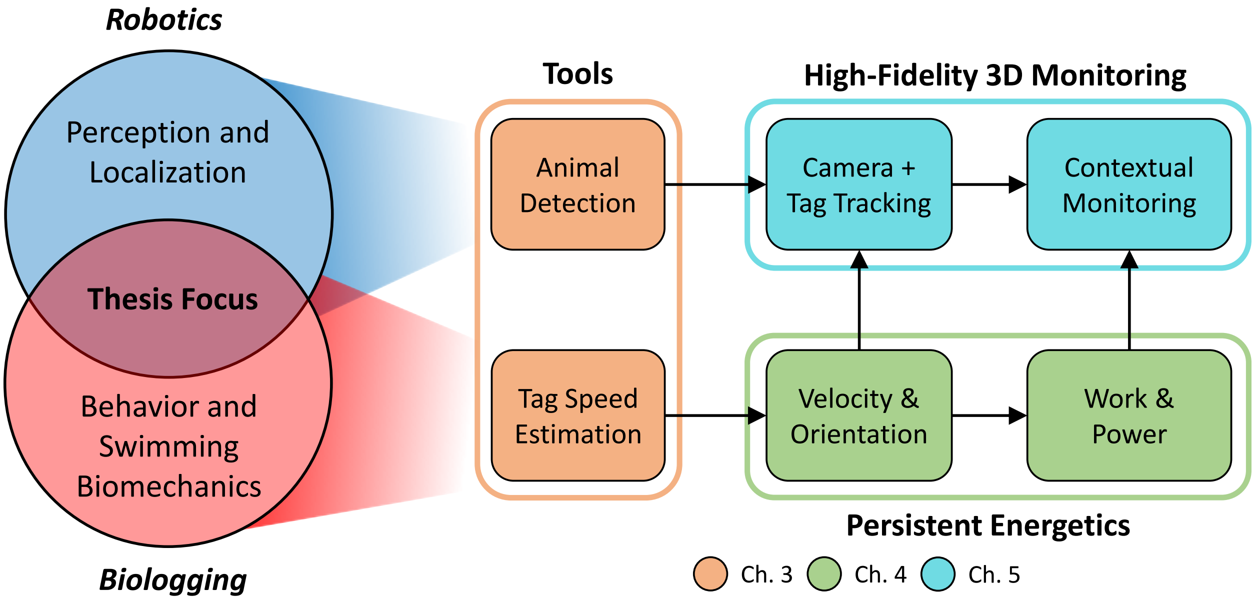
by

Joaquin T. Gabaldon

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Robotics)
in the University of Michigan
2021

Doctoral Committee:

Associate Professor Kira Barton, Co-Chair
Assistant Professor K. Alex Shorter, Co-Chair
Associate Professor Matthew Johnson-Roberson
Associate Professor Ramanarayan Vasudevan


Robotics
Perception and Localization
Thesis Focus
Behavior and Swimming Biomechanics
Biologging
Tools
Animal Detection
Tag Speed Estimation
High-Fidelity 3D Monitoring
Camera + Tag Tracking
Contextual Monitoring
Velocity & Orientation
Work & Power
Persistent Energetics
Ch. 3
Ch. 4
Ch. 5

Joaquin T. Gabaldon

gabaldon@umich.edu

ORCID iD: 0000-0003-3775-6154

So long, and thanks for all the fish.

ACKNOWLEDGMENTS

This dissertation represents the efforts of dozens of people, from collaborators, to advisors, to those who were with me for moral support. This would not have been possible without all of you.

The research performed in this dissertation would not have come about without funding from the University of Michigan Departments of Mechanical Engineering and Robotics, the Office of Naval Research, the Chicago Zoological Society, and Dolphin Quest Oahu. I am also grateful to the members of my dissertation committee, whose support and guidance were invaluable: Kira Barton, K. Alex Shorter, Matthew Johnson-Roberson, and Ramanarayan Vasudevan.

The research presented here encompasses the combined endeavors of multiple groups of people, both within the University and without, and I am especially thankful for the vast amount of support they have lent me to help this research come to pass. Throughout this journey, we have had collaborations with researchers from the Brookfield Zoo, Dolphin Quest Oahu, Duke University, Woods Hole Oceanographic Institute, and Loggerhead Instruments. These people were there every step of the way, from experimental formulation, to deployment, to interpretation and analysis. I have a special thank-you for my co-chairs, Profs. Barton and Shorter, who guided me through the development of what this dissertation would become and saw it through from beginning to end. Throughout my time at UMich, I have been a part of three different lab groups. All of them — DROP, ESTAR, and the Barton Lab Group — were welcoming, enthusiastic, and supportive, helping me enjoy my time as we toiled together in our respective research. In particular, I would like to give a special acknowledgment to Ding Zhang, my primary collaborator within the Barton Research Group, with whom the core components of this research were defined, built, and refined.

Finally, I would like to thank my family, friends, and especially my significant other, for their encouragement throughout the years. Without them by my side, figuratively and literally, this long and winding road would have presented naught but forks and dead-ends.

Thank you all.

PREFACE

Monitoring animals is key to understanding their behavior. Without this understanding, taking steps to improve animal welfare, both wild and not, becomes a guess-and-check process. Insights into their behavior give us an idea of metrics concerning habitat use, food requirements, and even the impacts of humans on their populations. For wild animals, this knowledge has the potential to influence how humans interact with these animals, and reduce the negative consequences of our activities. For institutionally-managed animals, this can result in directed, continuous welfare improvements as their caregivers better understand their needs. Unfortunately, for marine animals, monitoring is hindered by the environment itself: ocean water limits visibility, attenuates radio transmissions, and in the wild, makes physical access problematic.

For marine mammals, these limitations have historically resulted in studies being dependent on observations or interactions with the animals when they are near the surface. Commonly, this would involve aerial or surface-vehicle surveys to perform observations or direct interactions with smaller animals through mark-recapture studies [1, 2, 3]. Seafloor-mounted hydrophones have been used to monitor animal communications, however this is only viable for those with longer-range vocalizations [4, 5]. In the past few decades, recoverable biologging tags have been used to extend communication monitoring, and have enabled persistent kinematic data collection and some of the first continuous localization methods [6, 7, 8]. Figure P.1 provides an illustration of the various marine animal monitoring options in use to-date.

Despite the improvements tagging has provided, there are still significant gaps in monitoring methods for these animals. Animal energetics have not been thoroughly explored, due to both hardware limitations and incomplete computation methods, which impedes our understanding of their caloric requirements and activity levels. Additionally, localization techniques developed for tracking animals in the wild do not function well in a managed environment, making it difficult to investigate animal habitat use and how it relates to other behavior metrics.

At their core, these gaps all reflect limitations in hardware and processing methods for sensing applications. Any improved approach must be able to accurately track kinematics and perform localization in a highly dynamic and uncertain setting. This degree of capability has already been demonstrated in the field of robotics. High-fidelity underwater localization has been explored in the field of autonomous underwater vehicle tracking [9]. Robust dynamics and orientation monitoring has been demonstrated for use with Inertial Measurement Unit (IMU) sensors, in multiple subfields of the discipline [10, 11, 12, 13, 14]. Additionally, accurate and efficient object-tracking has been developed for use with computer-vision systems [15]. However, while these methods are powerful for their specific use cases, they do not directly translate for implementation in marine animal tracking. On their own, these techniques are tools that can provide crucial components to fill these sensing gaps, but additional frameworks and tools must be built around and alongside them to provide the necessary information for them to function and to interpret their results.

The goal of this dissertation is to provide both the means and the methods for contextualizing cetacean swimming behavior according to habitat use in a managed setting, through the implementation and extension of techniques originally built in the field of robotics. This is done through the introduction of additional tagging hardware and a new monitoring framework for high-fidelity persistent localization, kinematics, and energetics estimates of the animals. The performance of the tag hardware and framework are demonstrated through the monitoring of managed bottlenose dolphins (*T. truncatus*). They are further used in a long-term study to explore the potential of the approach in yielding new knowledge on energetics and localization for these animals.

There are three core contributions in this work. The first is the advancement of animal-borne sensing for marine animals. This includes the introduction and implementation of a speed sensor for use with biologging tags, and the production of an improved method for estimating animal energetics using this sensor. The second contribution focuses on the introduction of a new framework for marine animal localization in a managed setting. This involves the fusion of continuous tag-based animal kinematics and deep-learning object-tracking techniques to automate the high-fidelity reconstruction of an animal's trajectory and pose. The third contribution combines the previous two by producing new insights into how the dynamics and energetics of the monitored animals relate to their habitat use. This is done by contextualizing the tag-based activity metrics via the localization information, presented as not only new knowledge into their behavior, but also as an example of the potential provided by location-related behavior monitoring in better understanding these animals.

While this dissertation focuses on managed cetaceans, the techniques and hardware can be extended beyond the use cases of this research. The hardware developed to enable animal energetics measurements is suitable for use with wild animals, and the method can be transferred by modifying animal-specific parameters. Further, the localization framework is capable of use in other managed habitats with the appropriate hardware, and its core principles can be applied to wild-animal tracking. As a result, the approach described in this dissertation is presented as a specific implementation of a more general approach for enhancing marine animal monitoring through robotics in not only managed settings, but also potentially in the wild.

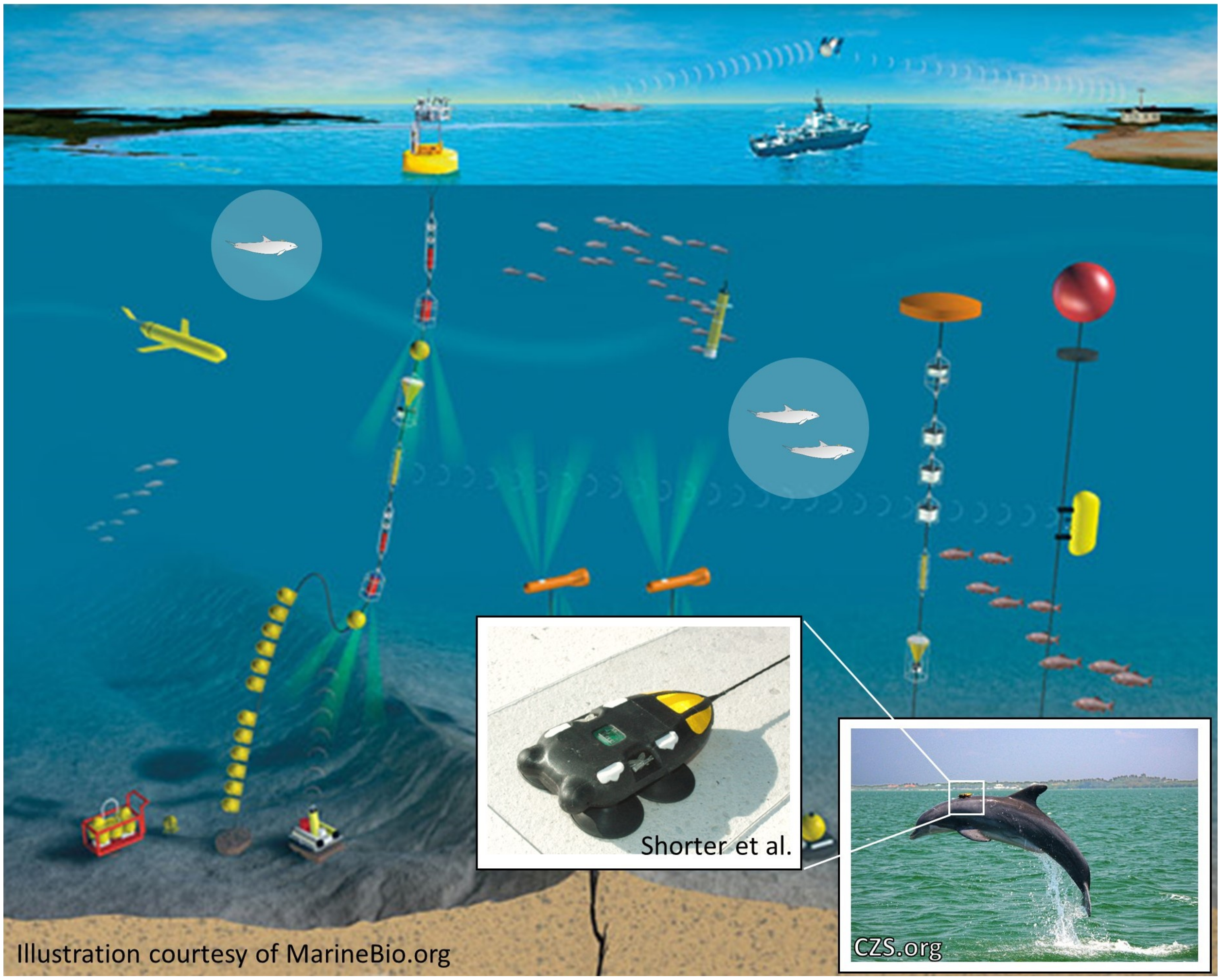


Figure P.1: Illustration composite of marine environment sensing methods. This figure provides context of the current monitoring approaches in use for marine mammals. External sensing methods range from surface observation platforms, to sub-surface hydrophones and autonomous underwater vehicles, which can record population data, animal communications, and localization information. Animal kinematics are generally recorded using on-body biologging tags (bottlenose dolphin: inset, bottom-right), which can have an array of depth, orientation, acceleration, and acoustic sensing, as well as position estimation via satellite communications when at the surface. The DTAG-3, a kinematic and acoustic recording device used primarily on cetaceans (inset, bottom-center [16]), is shown as an example. The primary figure is a modified illustration obtained from MarineBio.org, and the leaping dolphin inset is from the Chicago Zoological Society.

TABLE OF CONTENTS

Chapter

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACS Animal Care Specialist

CNN Convolutional Neural Network

CI Confidence Interval

CV Coefficients of Variation

COM Center of Mass

COT Cost of Transport

FFT Fast-Fourier Transform

FOV Field of View

GPS Global Positioning System

ICP Iterative Closest Point

ITS In Training Session

IMU Inertial Measurement Unit

IoU Intersection over Union

iSAM Incremental Smoothing and Mapping

K-S Kolmogorov-Smirnov

LIDAR Light Detection and Ranging

MEMS Micro-Electromechanical System

ODBA Overall Dynamic Body Acceleration

OTS Out of Training Session

PDF Probability Density Function

PIV Particle Image Velocimetry

PReLU Parametric Rectified Linear Unit

ReLU Rectified Linear Unit

RPN Region Proposal Network

RoI Region of Interest

RMR Resting Metabolic Rate

RMS Root Mean Square

RMSE Root Mean Square Error

SLAM Simultaneous Localization and Mapping

TDOA Time Delay of Arrival

TDT Total Distance Traveled

TTFF Time to First Fix

VGG Visual Geometry Group

YOLO You Only Look Once

ABSTRACT

Marine mammal monitoring has seen improvements in the last few decades with advances made to both the monitoring hardware and post-processing computation methods. The addition of tag-based hydrophones, Fastloc Global Positioning System (GPS) units, and an ever-increasing array of IMU sensors, coupled with the use of energetics proxies such as Overall Dynamic Body Acceleration (ODBA), has led to new insights into marine mammal swimming behavior that would not be possible using traditional secondary-observer methods. However, these advances have primarily been focused on and implemented in wild animal tracking, with less attention paid to the managed environment. This is a particularly important gap, as the cooperative nature of managed animals allows for research on swimming kinematics and energetics behavior with an intricacy that is difficult to achieve in the wild.

While proxy-based methods are useful for relative inter-or-intra-animal comparisons, they are not robust enough for absolute energetics estimates for the animals, which can limit our understanding of their metabolic patterns. Proxies such as ODBA are based on filtered on-animal IMU data, and measure the aggregate high-pass acceleration as an estimate for the magnitude of the animal's activity level at a given point in time. Depending on its body structure and locomotive gait, tag placement on the animal and the specific filtering techniques used can significantly impact the results. Any relation made to energetics is then strictly a mapping: a relation that may apply well to an individual or group under specific experimental conditions, but is not generalizable. To address this gap, this dissertation presents new tag-based hardware and data processing methods for persistently estimating cetacean swimming kinematics and energetics, which are functional in both managed and wild settings.

Unfortunately, localization techniques for managed environments have not been thoroughly explored, so a new animal tracking method is required to spatially contextualize information on swimming behavior. State-of-the-art wild cetacean localization operates via sparse GPS updates upon animal surfacings, and can be paired with biologging-tag-based odometry for a continuous track. Such an approach is hindered by the structure and scale of managed environments: GPS suffers from increased error near and within buildings, and current odometry methods are insufficiently precise for habitat scales where locations of interest might be separated by meters, rather than kilometers (such as in the wild). There is then a need for a tracking method that uses an alternate source of absolute animal locations that can achieve the high precision necessary for meaningful results given the spatial scale. To this end, this dissertation presents a novel animal localization framework, based on tracking and sensor filtering techniques from the field of robotics that have been tailored for use in this setting.

To summarize, this dissertation targets two main gaps: 1) the lack of persistent, absolute estimates of animal swimming energetics and kinematics, and 2) the lack of a robust, precise localization method for managed cetaceans. To address these gaps, the hardware and animal tracking methods developed to enable the rest of the dissertation are first defined. Next, a physics-based approach to directly monitor cetacean swimming energetics is both presented and implemented to study animal propulsion patterns under varying effort conditions. Finally, a high-fidelity 3D monitoring framework is introduced for tracking institutionally-managed cetaceans, and is applied alongside the energetics estimation method to provide a first look at the potential of spatially-contextualized animal monitoring.

CHAPTER 1

Introduction

1.1 Motivation: Marine Mammal Monitoring and Tracking

There are two primary method contributions in this work: the advancement of animal energetics/kinematics estimation and development of a new localization framework. To understand the scope of the contributions from this dissertation, an evaluation of the research motivation and historical methods is necessary. Chapter 2 presents the background description to provide this context.

Historically, both have been studied for cetaceans to some degree, with the extent of each dependent on the environmental condition (wild or managed), as well as specific limitations on the hardware or methods used. To understand the context in which the work of this dissertation is performed, Section 2.1 describes the general approaches and motivations for marine mammal monitoring studies as performed through external observers (i.e. non-instrumented animal studies). This section explores the range of methods used in both wild and managed settings, from historical examples to modern approaches. In general, three types of external observer monitoring methods are used, for both settings: direct observation/interaction, video/image analysis, and acoustic recording analysis. These methods are best suited for long-term population, health, and communication analyses.

Tagging provides more in-depth information on individual animals, and its benefits and limitations are addressed in Section 2.2. Biologging tags offer persistent rather than intermittent animal monitoring, unlike external observation methods, and it is through the introduction of this hardware that makes it possible to quantitatively understand animal kinematics. This helps present the motivation for contextualizing monitoring information with respect to animal location: the breadth of new knowledge enabled by persistent monitoring, much of it through tagging, can be more fully utilized by understanding the environmental stimuli an animal was exposed to when engaged in particular behavior modes.

As much of the research in this dissertation is built around animal localization, Sections 2.3 and 2.4 describe existing marine localization techniques and modern neural network tracking methods.

Highlighted in Section 2.3, marine animal tracking has seen strong progress in the last two decades with the advent of quick-acquisition Global Positioning System (GPS) techniques and tag-enabled animal-borne dead-reckoning. However, this dissertation focuses on managed environment localization, which the aforementioned methods are not suited for due to their inherent imprecision having been developed for deployment in the wild. To perform high-fidelity localization in a managed habitat, tracking techniques require much lower position uncertainties to account for the scale of the region. For this reason, Section 2.4 details state-of-the-art computer-vision object detection, which is currently built on neural network frameworks. To provide the relevant technical background, both to understand the structures presented in Section 2.4 and validate those deployed in Chapters 2 and 4, relevant neural network sub-structures are also detailed.

Finally, Section 2.5 details the historical and current methods for marine mammal energetics monitoring. While state-of-the-art techniques do provide direct measurements of animal energy use, they are either non-persistent or physically limit an animal's kinematic flexibility. Further, persistent estimations via proxy are not generally applicable across animals, either inter-or-intra-species. As such, there is potential for a physics-based estimation method for robust and persistent energetics tracking, whose background theory is presented in this section alongside the direct measurement and proxy techniques.

1.2 Aim 1: Advancing Tools in Marine Monitoring

Marine animal monitoring has primarily focused on enabling research in the wild, which is represented in the tools developed for the task. The marine environment is hostile towards most types of sensing and communications protocols, impeding vision and radio transmissions, and subjecting submerged hardware to destructive pressures. As a result, external observation techniques (i.e. not on-animal) rely on visual monitoring only at the surface [3, 4, 17, 18], direct animal contact [1, 2], or longer-range acoustic sensing deep underwater [5, 19, 20, 21]. On-animal biologging and satellite localization tags are built to be small yet physically robust [6, 22], and have relatively constrained sensor options due to the size, packaging, and battery life limitations [23].

The contributions in Chapter 3 are intended to extend our capabilities in marine mammal monitoring. As a portion of the research in this dissertation focuses on localizing animals in managed environments, the benefits and drawbacks of operating in such a setting must be addressed. Current state-of-the-art animal tracking methods in the wild depend on quick-acquisition GPS technology [8, 24]. While localizations can only be performed when the animal has surfaced, this approach takes advantage of the unobstructed view of the sky present in the open ocean. In contrast, managed settings are often indoors or near buildings, inducing localization errors due to multipathing effects (satellite signals bouncing off of environmental structures). As GPS precision is at best on the order of tens of meters, these errors further increase animal position uncertainties to larger than the size of the managed habitats, rendering them insufficient for localization in this setting.

Managed environments do have particular benefits to aid in animal observation: the shorter depth scales and generally higher water clarity offer much better visual monitoring opportunities than in the wild. So far, this has involved hand-tracked analysis of recorded video [25, 26, 27], though in recent years work has been done to automate this process through handcrafted image processing tools [28, 29]. However, handcrafted methods have been surpassed in performance by tools more commonly found in the field of robotics, in the form of neural network object detectors [30]. As such, there is room for significant improvement in both the precision and robustness of automated visual animal tracking in managed settings. Section 3.1 defines the neural network tracking structure developed, details the specific accommodations required for it to function in an experimental setting, and demonstrates the opportunities offered by such a system.

While biologging tags offer a host of useful sensing capabilities, animal speed estimation is a problem that still requires additional attention. To-date, tag-mounted speed sensing has either required unwieldy hardware [31, 32], or involved methods prone to noise-inducing disturbances [33, 34]. Section 3.2 presents new tag-based speed sensing hardware, explains the methods used for its calibration and verification, and demonstrates its capabilities in tracking animal swimming distances in live experiments. The speed sensor detailed in this section is intended to provide the necessary animal kinematics information required by the energetics estimation method in Chapter 4 and the localization framework in Chapter 5. Despite the application-focused nature of this speed sensor, such a device can be useful for understanding other animal swimming behaviors (e.g. foraging event analysis [35]), although this extension is not realized in this dissertation.

1.3 Aim 2: Persistent Marine Mammal Energetics

One primary component of this dissertation focuses on enabling energetics monitoring of tagged animals. Current research employs a combination of three types of methods: direct measurements of physical parameters, proxy mappings, and physics-based estimations. Direct measurements involve animal heart rate or oxygen consumption analyses [36, 37, 38, 39]. While these tend to be the most accurate approaches, they cannot be used to provide continuous-time (persistent) measurements. On the other hand, proxy methods are not standalone, and involve mapping biologging tag data to direct measurements [39, 40]. The state-of-the-art proxy method, Overall Dynamic Body Acceleration (ODBA), involves filtering accelerometer data to obtain proxy estimates of animal activity levels, and then comparing these time-averaged data to direct measurements (generally oxygen consumption). As tag data is persistent, proxy mappings can provide continuous estimates of energy use for individual animals. However, ODBA estimates can vary according to on-body tag location and the filtering approach used, which limits their robustness for comparisons between separate studies, as well as for inter-and-intra species analyses.

Physics-based methods attempt to predict animal propulsive power through fluid-dynamics models [25, 27, 41]. However, these require measurements of animal kinematics to function, and so far have relied upon speed and acceleration estimates from hand-annotated video recordings, rather than using on-animal measurements, which can be time-consuming to collect and have the potential to suffer from human error. The speed sensor presented in Section 3.2 is used in Chapter 4 to address this deficiency through its direct integration on a biologging tag, and is deployed alongside a modified version of a well-established physics-based propulsive power model [25]. Six bottlenose dolphins (*T. truncatus*) were tagged while they completed high-energy lap swimming tasks in a managed institution. Animal propulsive powers were estimated both for full laps as well as during steady-state (constant speed) travel, and their swimming efficiencies were calculated to observe the behavioral propulsive effort differences between the animals. Additionally, to demonstrate the capabilities of the method in long-term kinematics and energetics monitoring, one animal was tagged for a full 24-hour session. The animal's propulsive effort during its active (non-resting) phase was contrasted to its lap trial results to compare its behavior during moderate versus high-energy swimming. Further, its estimated full-day energy consumption was compared to its measured caloric intake to evaluate the method's viability for long-term metabolics tracking.

In summary, Chapter 4 presents two contributions: 1) the energetics monitoring framework and all necessary data processing methods to enable it, and 2) the new knowledge on dolphin swimming behavior made possible by the framework.

1.4 Aim 3: High-Fidelity 3D Monitoring

As the difficulties in localization presented by the marine environment tend to inhibit continuous animal tracking, one solution to this data sparsity problem is to directly track the animal when possible, and estimate its positions through dead-reckoning otherwise. Dead-reckoning requires subject orientation and speed, which are numerically integrated to provide best-guess animal pose information [7]. However, such methods have inherent drift, so despite preserving the overall shape of an animal's movements the position error will increase over time. By using absolute position information (i.e. on-tag GPS acquisitions) to anchor a continuous dead-reckoning track, it has been possible to reduce the error in tagged marine animal tracking both at the acquisition points and between them by stretching the continuous track to coincide with the absolute positions [8, 31, 32, 42]. However, tag-only position estimates provide minimum errors in the range of over 10 meters [24, 43], which are not compatible with the physical scales of managed environments, without even taking into account multipathing effects.

The contributions in Chapter 5 are intended to extend and enhance the absolute position with dead-reckoning localization approach through the integration and modification of techniques more commonly found in robotics. The sensor presented in 3.2 serves to provide accurate direction-of-travel speed measurements, which when employed alongside Inertial Measurement Unit (IMU)-estimated orientations [10], yield a high-quality dead-reckoning track for an animal. The neural-network objection detection method in Section 3.1 then provides the absolute position estimates when paired with tag-measured depth by processing video recordings of the animal in its habitat. Finally, the dead-reckoning track and absolute position estimates are combined through the Incremental Smoothing and Mapping (iSAM) optimization method [44], a pose filtering technique that takes into account the relative uncertainties in each position estimate (dead-reckoning vs. absolute) at each time step. This approach was chosen for its high performance in detailed subject tracking tasks in unconstrained environments [9]. To demonstrate the capabilities of this method in contextualizing marine mammal monitoring, the localization framework was deployed for the purpose of tracking a tagged bottlenose dolphin during a period of free-swimming. The animal's locomotive performance was calculated using the energetics estimation technique detailed in Chapter 4, which was then correlated to the animal's habitat use.

To summarize, Chapter 5 also presents two contributions: 1) the full localization framework and the additional sub-structures required for it to function, and 2) preliminary results on contextualized animal energetics monitoring to provide an example of the new opportunities afforded by the approach.

1.5 Contribution Interactions

The tools developed in Chapter 3 provide the backbone for the rest of the monitoring methods in this dissertation. The speed sensor hardware serves as the foundation for the energetics estimation techniques in Chapter 4, the components of which, when combined with the neural network animal detector, enable the 3D monitoring framework in Chapter 5. Figure 1.1 presents a diagram of how the separate contributions relate to each other.

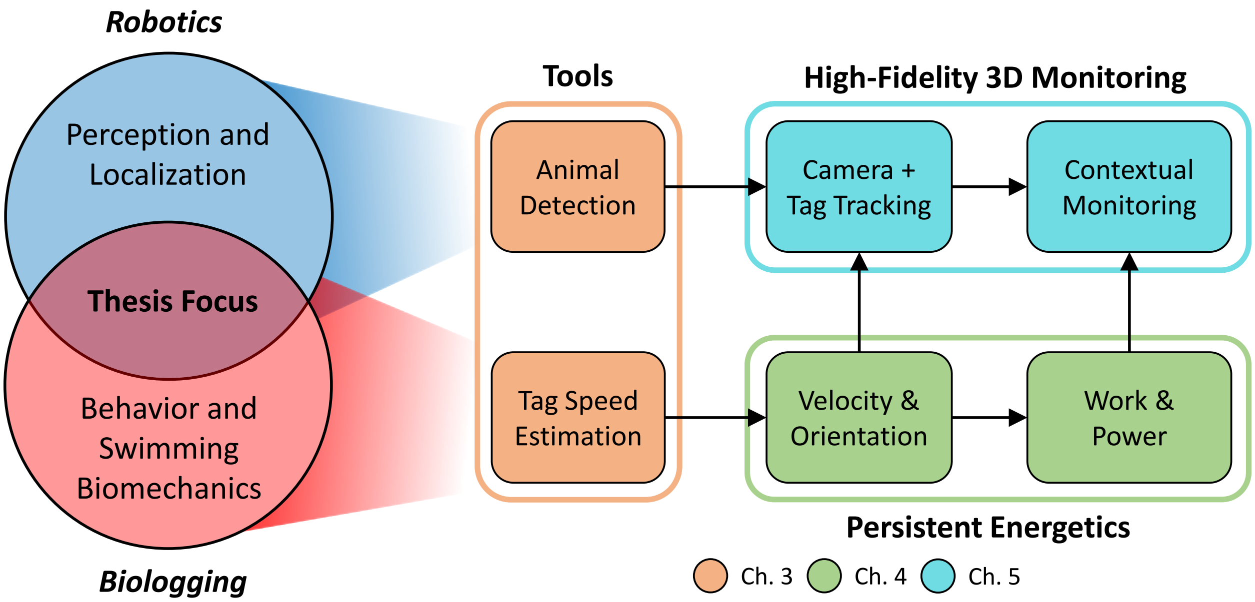


Figure 1.1: Contribution Method Flowchart. The tools developed in Chapter 3 (orange) stem from one of two core fields that encompass this dissertation: object perception and localization techniques (blue), grounded in robotics; and animal behavior and swimming biomechanics (red), part of marine animal biologging. Each tool supports an additional contribution, with speed estimation providing the necessary information for the physics-based energetics computations in Chapter 4 (green), and neural network in-video animal detection enabling the high-fidelity 3D monitoring framework in Chapter 5 (teal) when combined with both the methods and results from Chapter 4. As shown in the flowchart, the contextualized monitoring method in Chapter 5 represents the culmination of the work in this dissertation.

CHAPTER 2

Marine Mammal Monitoring and Localization

2.1 External-Observer Monitoring

Marine mammal monitoring has traditionally been performed using external observers. Before the inexorable miniaturization of electronics made tagging animals possible, observations were primarily viable through direct data-collection approaches. These included vision and acoustic sensing, as well as capture-recapture methods, which vary in range, precision, and depth capabilities depending on the method used. While the vast majority of these monitoring procedures were built for use in the wild, vision and acoustic sensing have been used in managed settings as well. This chapter section describes external-observer monitoring methods as they have been used to collect data on marine mammals in both wild and managed environments.

2.1.1 External Observers: Wild Habitats

Visual observation methods have historically been the most commonly used in the wild, due to their flexibility and variety of implementations. These methods use human-logged qualitative and quantitative measures, and are most effective when surveying from mobile platforms such as aerial vehicles or ships [3, 4, 17, 18]. Visual surveys primarily yield information pertaining to animal populations, as their data are limited to manual logging of individuals within a general area, or video/photographic records that are later examined. While the former can only give an estimate of population number and general makeup (e.g. gender, estimated age), the latter does offer the potential for re-identification of individual animals through bodily markings. This makes it possible to more precisely identify ingress and egress of individuals from a group, and more reliably monitor population changes over time for the surveyed region of the environment. Because visual observation methods focus on regional monitoring, they are limited in their ability to obtain detailed information on individuals that can travel outside their operational range. Additionally, observations are restricted to when an animal has surfaced due to the opacity of seawater. This method is mostly non-invasive as there are no direct interactions, however in the case of ship-based monitoring the noise from the vehicle can be sufficient to disturb the animals.

For smaller marine mammals, more direct capture-recapture methods can be used to measure additional information beyond what purely visual methods can provide. In these instances, animals are captured via nets, marked using either brands or a physical indicator, and released [1, 2]. During the period of captivity, animal morphometric measurements and blood draws can be collected for information on animal physical size distributions, current health state, genetic relationships, and reproductive conditions. At later dates, when animals are recaptured, the same measurements can be performed to monitor long-term (multi-year scale) trends. While these studies can be invaluable for tracking population size and health over a number of years, such methods are not equipped to provide information on animal swimming behavior or energetics, which are needed to understand foraging patterns and food intake requirements.

Cetaceans are well-known for their long-distance vocalizations. These are used for communication as well as echolocation emissions for foraging [45]. While higher-frequency emissions can travel in seawater for at most hundreds of meters, communication vocalizations can propagate up to hundreds of kilometers for the largest animals. This long-range vocalization makes it viable to monitor cetacean populations and communications through underwater acoustic detectors. Historically, individual and arrays of hydrophones (fixed location underwater microphones), and sonobuoys (expendable temporary hydrophones dropped from aircraft) have been used to this end [5, 19]. With multiple sensors, the range and bearing of vocalizations can be triangulated for signals with identifiable patterns, even allowing for localization. Further, recent research with hydrophone-equipped autonomous oceanic vehicles has allowed for odontocetes tracking via the animals' higher-frequency, shorter-range echolocation emissions, providing more flexible monitoring [20, 21]. While acoustic monitoring does provide more frequent information on the animals than other external observation methods, acoustic methods still cannot provide continuous information on animal behavior due to the potential infrequency of their vocalizations.

2.1.2 External Observers: Managed Habitats

Visual observations have commonly been used to directly monitor cetacean populations in managed settings, and have been implemented in studies concerning animal welfare and energetics research. Data collection tended to be performed real-time, with researchers recording information on the animals in shifts. Observations in these studies consisted of estimates of animal behavior patterns, focusing on subjects such as swimming pattern analysis in the search of neurotic behaviors [46, 47], and animal stress level monitoring in order to promote better welfare practices in managed facilities [48, 49, 50]. Given the simplified access to the animals involved in the research, these studies also could benefit from animal blood analysis and physiological trend measurements [47, 48, 50]. These laid the groundwork for understanding the mental and physical conditions of managed cetaceans, however as these studies involved measurements taken through human observation, results obtained from visual analysis of animal behaviors in these conditions tended to be relegated to qualitative estimates. One study was able to perform a quantitative kinematics and energetics analysis without using recorded video [37]; however, this was only possible by timing animal swim durations along prescribed paths to obtain speed information.

Studies using video recordings tended to have a heavier focus on animal kinematics and propulsion. Such recordings have allowed researchers to produce more quantitative analyses of animal activity levels. Tracking animals with scale marks (markings placed a known distance apart on the animal's body) through manual frame-by-frame video tracking has enabled speed analysis of bottlenose dolphins for a portion of a resting behavior monitoring study [51]. A similar method was also used to estimate the steady-state swimming power and drag profiles of multiple species of cetaceans [25, 26, 27]. The use of video recordings also enabled more precise analysis when used in social interaction studies [52], given the ability to recheck animal identities when observing animal pairing behaviors. Video analysis has allowed for more robust and repeatable estimations of the quantitative aspects of cetacean swimming behavior; however, one inherent drawback with these methods was the post-processing time: as automated computer-vision tracking methods have not matured by this period, animal tracking could only be done by hand.

Acoustic studies using environmentally-mounted hydrophones performed in managed settings have primarily focused on animal communications, although some work has been done in using vocalizations for animal localization. As cetaceans emit a complicated set of sounds for communication, special attention has been paid to how these sounds change dependent on the animals' social interactions. These have ranged from research on individualized animal vocalizations [53], to specific analyses of the communication patterns in mother-and-calf interactions [54, 55]. Research on managed cetacean localization through acoustics is limited; however, animals have been able to be localized via passive sensors (as opposed to active sonar) by triangulating their vocalizations [56]. Similar to in the wild, the use of passive sensing in this form of acoustic localization limits measurements to time instances where the animals choose to emit sounds, and cannot provide perpetual location data.

2.2 Marine Mammal Biologging Tagging

The marine environment is generally hostile to both sensing and communications protocols. In the best cases, visibility is limited to tens of meters, while cetaceans are capable of diving in excess of one thousand [57]. These depths also hinder the ability of humans to directly observe the animals in a significant portion of their active habitat. Additionally, non-visible light frequencies are heavily attenuated in water, making data transmissions to and from animal-attached sensors limited to the period when the sensor is at the water's surface, hindering satellite tracking methods. Acoustic monitoring is the only environmentally-mounted method that operates with any long-range capability underwater. Unfortunately, due to the migration capabilities of larger animals (whose vocalizations can be detected at distances of multiple kilometers), their effective habitat ranges can span thousands of kilometers making environment instrumentation difficult [58]. Animal-mounted biologging tags provide solutions to some of these problems: instead of waiting for an animal to come near a sensor, the sensor is taken to the animal.

2.2.1 Tagging in the Wild

Given the difficulty of instrumenting the marine environment, animal-borne sensing has proven invaluable in obtaining vocalization, kinematics, and location information on cetaceans in the wild. Some of the earliest research using animal-borne sensing on marine mammals involved recording diving profiles of Weddell seals (*L. weddelli*) in 1963-64 [22]. This arrangement used an analog recording needle inscribing pressure transducer readings onto a smoked glass disc, set to rotate via attachment to a kitchen timer. Such measures were necessary as depth-pressure recording devices had not been miniaturized sufficiently for animal attachment, which is one of the hurdles when developing tagging equipment: while the required sensors may exist, fitting them into a package that minimally disturbs the animal can be a problem. This has been aided significantly with the advent of solid-state electronics in on-device data parsing and storage, as well as Micro-Electromechanical System (MEMS) sensor suites, drastically reducing component sizes on all fronts. Electronic miniaturization has led to tags being able to include accelerometers, magnetometers, gyroscopes, depth and temperature sensors, hydrophones, and GPS units [6, 23, 59].

There are two primary types of devices within the domain of biologging tags: archival and transmitting. Archival tags are attached to the individuals to be monitored and must be recovered to obtain the recorded data. This requires either an additional encounter with the tagged animal to remove the device, or the tag must release itself from the animal to be recovered via radio telemetry while floating on the water's surface [23, 60]. However, the reduced power usage provided by archival tags allows recording sessions to extend for multiple days between tagging and recovery. Additionally, as archival tags are intended to be recovered, the volume of data collected can be much higher than when deploying transmitting tags. As such, archival tags have been used to perform extended, continuous studies of animal depth profiles, vocalizations, and kinematics [23, 35, 61, 62, 63]. Further, with the introduction of post-processed methods such as Fastloc-GPS, archival tags can take advantage of satellite-based telemetry for ocean-surface localization [8].

Transmitting tags cover some of the same use cases as archival units; however, they are primarily deployed for localization tasks. Tags of this type are attached to the animal and are not intended to be recovered, instead sending out signals when the animal is at the surface. Early implementations of transmitting tags (1960s) acted as radio beacons and were attached to larger cetaceans, allowing researchers to pursue the tagged animals as long as they were within radio range for the duration of the tag's battery life [59, 60]. In the past several decades, transmitting tags have seen success in long-distance intermittent localization tasks. Animals were tracked through tag communications with Argos satellites, which in turn sent the estimated location information to researchers [64, 65]. Tag effective durations can be extended by limiting the high-power transmissions to surfacing events, triggered by a submersion sensor, allowing for multi-month localization studies [65]. This form of tracking is specifically oriented towards large-scale travel monitoring, such as animal migration analyses, as the imprecision of even the most advanced satellite tags do not allow for fine-scale (sub 1-meter) position information for swimming behavior analysis [43].

2.2.2 Tagging in Managed Habitats

Historically, as access to animals in managed habitats is rather straightforward versus in the wild, direct observation or environmentally-mounted sensing methods have taken precedence in this setting. As a result, archival biologging tags have only been deployed on managed cetaceans in limited capacities, as the ease of recovery ensures data loss is infrequent. So far, these studies have either focused on answering specific biomechanics questions, such as the additional metabolic cost of archival tag attachment [38], the wave drag effects on a swimming animal [66], and how fluking gaits trend throughout a typical day [67].

This limited deployment set represents a significant opportunity in unveiling new knowledge on marine mammals in a managed setting, as tagging offers specific benefits either direct observation or environmentally-mounted sensing do not. First and foremost, archival tagging has the capability of persistent monitoring: data from the animal are collected continuously, whereas external methods tend to be intermittent or opportunistic. As the tag is attached to the animal, recordings are not interrupted regardless of where the animal is in its habitat, whereas direct observations require both line-of-sight and sufficient lighting to perform (limiting night monitoring). Second, tag sensors provide strictly quantitative data collection. This is particularly important for behavior analysis, as human observers provide qualitative assessments and concordance checks must be performed to ensure data consistency when using multiple observers [49]. Third, the quantitative data provided by a tag offers the potential for the automation of analysis. As studies increase in scale, data collection and processing efforts that are completed manually must also increase at a similar rate, hindering researchers' ability to extend the scope of their work. Automated tag data processing can mitigate this effect, as tag deployments are less intensive than direct observation, and parsing the data only requires additional computation time. This has the potential to streamline fields such as behavioral observation, as modern biologging data processing techniques have the potential to classify behavioral states that would otherwise require qualitative assessment [68].

2.3 Localization

In order to better analyze animal behavior, it is important to understand the stimuli that are presented to the animal. One component of this is reconstructing where in its environment the animal has visited, and how its time is divided among these locations. With knowledge of an animal's habitat use, it is more feasible to identify the factors in the environment that attract its attention. Further, when coupled with behavior metrics concurrently obtained from additional sensors (e.g. biologging tags, acoustic monitoring), knowledge of the animal's location can allow for the contextualization of these behaviors. As such, robust localization can provide new information on how animals both use and interact with their environments.

2.3.1 Depth Monitoring

In the marine environment, depth tracking is the most straightforward of localization methods due to the direct relation between environmental pressure and water column depth. As the pressure transducers used for sensing depth are relatively uninhibited by the marine environment, unlike visual or radio-based localization, depth measurements can be continuously recorded, yielding high-density vertical location estimates. For marine mammals, depth monitoring provides information on both general statistics concerning water column use as well as the relation of depth to specific behavior events. One early tagging study on marine mammals was performed specifically to estimate the general depth profiles of Weddell seals [22], and similar research has been done on other species of marine mammals [57, 60, 61, 64]. Animal depth profiles have been combined with IMU and hydrophone data for monitoring behavior in the context of partial location information. This has allowed researchers to analyze animal dynamics and vocalization patterns present during foraging events [35, 62, 63, 69], yielding detailed information on how animal vertical habitat use relates to feeding patterns.

2.3.2 Transmitting/Receiving Tag Telemetry

Radio tags produce electromagnetic signals in frequency bands > 100 MHz, serving as beacons leading to the tagged animals [60]. These tags could be tracked through directional multi-element antennas, and required that researchers pursue the tagged animals at distances on the order of tens of kilometers to remain within signal range [59]. While these initial models of transmitting tags served their purpose of aiding researchers in tracking animal locations, they were replaced by satellite-linked transmitting models. Satellite tags operate on radio frequency bands in the range of 401.6 MHz ±30 kHz and communicate with Argos satellites in low polar orbit [43]. Transmissions from the satellites are then sent to ground receiving stations. Localizations are performed by analyzing the Doppler shifts in the tag-transmitted signals as perceived by the satellites, combining the information from multiple messages sent in succession for ocean surface position estimates. Satellite tags offer the distinct advantage over radio tags in that localizations can be performed globally without the need for manually following the animals via a moving instrumented platform; however, the localization error of the Argos system is at best in the 100s of meters [43, 64].

More recently, advancements in GPS technology have allowed for global localization on tags using incoming satellite communications. Unlike Argos positioning, GPS-based technologies involve measuring the time-of-flight of 1.23 and 1.58 GHz radio signals sent by GPS satellites to a receiver, and using these timings to calculate their distances to the receiver [70]. With known satellite positions, the latitude, longitude, and altitude of the receiver can be estimated with an accuracy of < 2 meters for high-performance equipment [71]. However, traditional GPS animal tracking does not perform well in the marine environment for highly dynamic animals. As signal acquisitions cannot occur at depth due to radio-wave attenuation by seawater, the time windows for these events are limited to only several seconds when an animal breaches the surface. When deployed on animals, units are also powered down in between captures to conserve battery power [70]. As a result, the unit's Time to First Fix (TTFF), where a "fix" is a localization, is crucial for animals that are only exposed to the satellites for this short period of time. Traditional GPS units require time scales on the order of 10s of seconds, which is not viable for marine tracking [24]. To reduce TTFF, snapshot rapid fix receiver technologies, such as Fastloc, perform only a fraction of the position estimation process on-device and require additional computation in-post to obtain the final location data after tag recovery. This partial computation approach allows for snapshots to be captured in less than 1 second, which are then stored on the device's internal memory. Even with limited-window data captures, Fastloc GPS drastically improves upon Argos accuracy when 6 or more satellites (minimum of 4 to function) are visible to the receiver (50th percentile accuracy: 18 meters, 95th percentile: 71 meters) [24].

Satellite-aided localization methods are capable of providing latitude and longitude estimates for animals in the wild, however the information provided is limited. Most importantly, there is no capability for animal tracking when below the water's surface. These methods provide general positioning estimates, allowing for migration and habitat use analysis [65], however data logging frequencies are intermittent with periods between samples dependent on animal surfacing requirements (e.g. respiration), resulting in the majority of the animal's movements not being reported.

2.3.3 Marine Acoustic and Video Localization

Acoustic localization is enabled by the estimating the range and bearing of animal vocalizations with respect to hydrophone locations. This can be done with a variety of sensor arrangements and computation methods. With only a single hydrophone, it is possible to estimate the range of an animal vocalization through exploitation of the physics involving underwater acoustic reflections and how they relate to the surrounding geography's bathymetric profile (elevation structure). With knowledge of the geometry of a hydrophone's environment, it is possible to use ray-tracing techniques (back-calculation of a signal's origin by estimating reflection angles) on multipathed vocalizations, as the hydrophone will receive several signals from a single vocalization: the direct signal, the reflection off of the underside of the water surface, and reflections from the sea floor [72, 73]. Using measurements of the time differences between reflection arrivals, this leads to range estimates for the animal using only a single sensor [5].

When using multiple hydrophones, it is possible to compute both range and bearing for vocalizations. Individual arrival sound signatures are identified for each hydrophone by cross-correlating impulse patterns in the recordings, which are used to estimate each Time Delay of Arrival (TDOA) of a received vocalization between each sensor pair in the array [74]. In this case, rejecting false correlations due to multipathed reflections is crucial to obtaining accurate TDOA sets [74, 75]. The range and bearing of a vocalization from the array can then be computed by minimizing the disagreement in the set of predicted distances from source to receiver for all sensors. This is performed by either solving for a system of equations, where the total time delay from source to array and the source positions are unknown [56, 74], or by finding the maximum in a set of surfaces representing the likelihoods of each source location [75]. This type of localization approach has been performed using static positioned hydrophones in both wild [74, 75] and managed [56] settings, as well as on mobile platforms in the wild [20, 21].

While acoustic localization has potential in tracking vocalizing animals underwater for both long and short ranges, there are various drawbacks with the method depending on use case. Primarily, this type of localization requires the animals to actively emit sounds, which can be more or less frequent depending on the type of animal being tracked as well as the type of activity being performed, and results in sporadic localizations. Additionally, these methods require thorough knowledge of the environment's bathymetry, which is not always feasible in the wild. Even with this knowledge, multipathing effects due to environmental geometry can also be strong enough to break the method, such as in managed settings with high numbers of reflective surfaces [56].

Video localization is very rarely used for marine mammal tracking due to its inherent limitations in water, although some examples do exist. For use in wild settings, work has been done to enable autonomous aerial vehicles to track larger marine animals through handcrafted computer-vision tracking methods [76]. This method uses image hue and saturation thresholding to remove the background and identify patches of the image that correspond to the tracked animal. Such an approach is sufficient to track a single animal that is primarily close to the surface, although once the animal dives for any significant period of time re-acquisition will be difficult. In the managed environment, research that has involved automated video-tracking of marine mammals has also been relegated to handcrafted computer vision techniques. In one study, both robust principle component analysis and a mixture of Gaussians were separately used to subtract image backgrounds and extract animal profiles using underwater cameras; however, these only provided in-image locations and were not mapped to a general coordinate frame [28]. In a more recent study, image backgrounds were removed through a variety of image-subtraction methods to obtain animal profiles, and were converted to 3D habitat-frame coordinates (as opposed to 2D image-frame) through stereo vision transformations [29]. While both methods were accurate for their particular use cases, performance can still be improved through more modern detection methods that can be made robust to changes in lighting conditions and water-surface noise, such as neural network object tracking.

2.3.4 Dead-Reckoning

Dead-reckoning uses information on a tracked subject's pose and velocity to generate a continuous estimate of its trajectory with respect to its initial position [77]. This is contrasted starkly with the previously detailed methods, as although localization through dead-reckoning is guaranteed to be continuous, it is prone to drift over time, while satellite, acoustic, and video localization are not guaranteed to be continuous but are also not subject to drift. As marine mammals almost exclusively swim in the forward direction, odometry can be generated using an animal's estimated speed and 3D orientation (roll, pitch, yaw) [7]. Additionally, vertical velocity components are not required to be computed as on-tag depth sensing provides absolute water column localization.

2.3.4.1 Continuous Speed Estimation

To reduce the complexity of biologging tag hardware, it is advantageous to obtain additional metrics on animal kinematics by only using existing sensors. This keeps power draw consistent, and does not add additional bulk with the inclusion of another physical component. This has been accomplished through analysis of both depth and acoustic data. Depth measurements can be computed to speed through trigonometry: differentiated depth information provides the vertical component of velocity, and represents the sine (with respect to animal pitch) of the total speed [35]. This method is robust for high magnitude pitch, but once the animal levels off, the accuracy of the estimate decreases. On the other hand, fluid flow noise levels detected by a tag's hydrophone can be mapped to its speed, and this estimate is consistent across all orientations [33]. The mapping can be performed on-animal, as depth-estimated vertical speed during steep dives can be compared to the hydrophone baseline noise levels. These data points are then used to fit an expression that outputs forward speed dependent on flow noise. Unfortunately, flow noise is not always stable, leading to fluctuations in the speed estimate that must be filtered out, and this approach does not perform well at low speeds (exponential magnitude dropoff as speed $\rightarrow 0$), or at shallow depths (flow noise is drowned out by surface noise).

While they do require additional hardware and power consumption, dedicated speed sensors do offer the potential benefit of reliability. Thus far, sensors such as impeller transducers [31], paddle wheels [32], and flexible paddles [34], have been used to approach this problem. One particular point that was addressed with these sensing methods was the ability to accurately estimate very low speeds, as tagged animals are not always in a high-activity state. Unfortunately, these structures can be physically large [31, 32], which has the potential to induce unnecessary drag on the tagged animal during higher speed locomotion, or are susceptible to external noise [34], reducing the accuracy. Because of these limitations, there is the need for on-tag speed sensing hardware that is robust while maintaining a minimal size profile, that also has a wide range of detectable speeds.

2.3.4.2 Orientation Estimation

Early animal orientation estimations involved using a gimballed flux-gate compass to estimate tag orientation with respect to the Earth's magnetic field [31]. This approach suffered errors if the animal rolled or pitched over 95°, which in the marine environment is rather common. A solution to this problem was presented through the filtering of MEMS-based accelerometer and magnetometer data [6]. This approach first has a user estimate the pitch, roll, and heading of the tag with respect to the animal, either through visual inspection upon placement, photographs, or analysis of accelerometer data during surfacing events (the animal generally needs to have zero pitch and roll in order to breathe). Roll and pitch are then determined from the relations between accelerometer channels and their angles with respect to gravity. This operates under the assumption that animal-induced accelerations are trivial in magnitude with respect to gravitational acceleration, which does not always hold for high-dynamics actions. Heading is then estimated by using animal pitch and roll to transform the magnetometer measurement onto a plane perpendicular to gravity, and computing the angle with respect to true north. While this approach does offer a non-physically-constrained estimate of animal orientation, it can still suffer from gimbal-locking due to its 3D angle representation, as opposed to quaternions or rotation matrices that do not have this problem. Additionally, the use of acceleration data alone for pitch and roll limits the sensitivity of the estimate, as accelerometers have inaccurate tracking of high angular rates [78].

Until recently, gyroscopes have had power consumption levels that were too high to include in tags and still retain acceptable battery durations [23]. The inclusion of gyroscopes alongside accelerometers and magnetometers in the orientation sensor set drastically improves the accuracy and robustness of the estimation. While gyroscopes are prone to drift (being rate sensors), they perform best at high angular rates where accelerometer and magnetometer readings might lag or yield noisy measurements. This is exploited in modern estimation methods through sensor fusion techniques such as high-speed gradient-descent [10], high-accuracy Kalman filtering [11, 13], and magnetic-disturbance rejecting extended complementary filtering [12, 14]. The addition of gyroscopes as common sensors in biologging tags therefore offers opportunities to improve orientation accuracy estimates, which can not only enhance dead-reckoning, but has ramifications for gait and behavioral dynamics analyses as well.

2.3.5 Path Optimization

Combining both dead-reckoning and absolute position estimates (such as from GPS) provides the continuous tracking benefit of revealing fine-scale animal movement patterns, while anchoring the track to reduce the effects of drift [32]. The most straightforward method of combining dead-reckoning and GPS information is to first take the errors between the estimated dead-reckoning positions and the GPS localizations whenever acquisitions are made. Next, the estimated track is stretched evenly between acquisitions to anchor it to the GPS localizations [31, 42]. This does not take into account the error profiles of the data streams being fused, whereas "trusting" each stream according to its estimated error at the time of fusion can provide a more accurate track. This has only recently been performed for marine animal tracking, using a state-space model framework with relative "trust" defined for each data stream according to their position uncertainties at the moments of absolute position acquisitions (GPS and visual) [8].

The approach presented in [8] offers performance improvements in marine animal tracking, although more advanced path optimization methods exist in the field of robotics that can be ported over for use in animal tracking tasks. State-of-the-art pose-graph optimization methods record the position uncertainty at every step in a tracked object's dead-reckoned sequence through a linked set of poses [9, 44]. When there is an absolute position update, the uncertainties in the pose and the position update at that time are compared to effect a location shift in the pose, and this shift is propagated throughout the entire set of poses according to their relative uncertainties. This more optimally represents a tracked object's position and orientation, and its robustness can be useful for localization tasks in managed settings where animal trajectories can become highly complex.

2.4 Neural Network Object Detection

Neural networks are built to operate as universal function approximators. Their uses range from pattern identification in temporal data, to object detection in images, to even emulation of highly nonlinear mathematical functions. Their varying functionalities come from the way they are structured to process and retain information, and their robustness (or lack thereof) have to do with their abilities' direct dependence on how they learn to accomplish their given tasks.

In the most general description, a neural network is a highly customized, multi-level weighted sum. Each node in a level (ℓ_i, i-th level) is connected to a subset of the nodes in the previous level (ℓ_{i-1}), and takes as inputs the weighted sum of the outputs of its connected nodes in ℓ_{i-1}. Weights are assigned on a per-connection basis, with values set via training on known input and output data: an input signal is passed through the network, the output is measured against an expected ground-truth, and weights are updated according to the resulting error. This is repeated multiple times for a set of input and output signals until the error (preferably) converges to an asymptotic lower bound. Core to the functionality of neural networks are activation functions, which along with other nonlinear data collation and subsampling methods, introduce nonlinearities in the weighted sums [79]. Without any induced nonlinearities, neural networks would only serve as overdetermined linear operators. However, while their performance can be state-of-the-art, this structural complexity does result in neural networks being treated as black-boxes.

2.4.1 Convolutional Neural Network Structure

For this dissertation, we are primarily focused on neural networks that aid in computer-vision object detection and tracking. For our use, Convolutional Neural Network (CNN) modules are particularly beneficial for identifying specific structures in images [80]. CNNs operate by applying convolutions — discrete integral transforms that use small image kernels to apply simple filters to images — in sequential banks to extract structural information from an input image in the form of "features" [81]. In this case, the primary modifiers being computed are the kernels themselves: a single convolution takes as input an image patch the size of the kernel, and outputs a sum of the weighted pixel-to-pixel products. The output of a convolution applied to an input image is then the total set of outputs produced by convolving all patches in the input (each patch is centered about one pixel in the input image), arranged in the output image according to the original patch locations in the input. Each bank of convolutions takes the filtered image outputs from the previous convolution bank as inputs (the first bank takes in the original input image), which can allow for highly complex spatial features to be identified within an image. These features are generally fed into fully-connected neural network layers, which interpret the outputs of the convolution layers: which features are detected within an image, and how strongly they are identified, can signify how to classify the input image.

2.4.1.1 Activation Functions

Activation functions are particularly useful in determining whether a neural network node should be deprioritized based on its overall response to the input ("activation") [79]. As CNNs are intended to identify specific object classes within an image, if an input does not elicit a strong response in a node, that node is less relevant towards correctly classifying the object in the image. Deprioritizing such nodes is useful in reducing noise in the overall output, and using only prioritized nodes results in a more confident classification. Activation functions are placed in CNNs after each convolution bank, and this subsection presents several of the more commonly-used examples.

Early activation functions consisted of nonlinear expressions that bounded node activations. Examples of this are the sigmoid (otherwise known as the logistic function), which bounds activations to $[0, 1]$:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2.1}$$

and the hyperbolic tangent (tanh) function, which bounds activations to $[-1, 1]$:

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.2}$$

While these activation functions ensure that individual nodes do not activate too strongly, this effect does cause other problems. When employed in a deep network, activations tend to saturate, which slows down the learning speed of the network during training [82]. This causes deficiencies in both convergence and inference accuracy compared to results that use non-saturating functions.

More modern activation functions tend to use the Rectified Linear Unit (ReLU) format, which are sometimes allowed to saturate activations to a floor of 0, but tend to have no upper bound [83, 84]. The most common and simple form of this format is the original ReLU function:

$$f(x) = \max\{0, x\} \tag{2.3}$$

which nullifies negative node activations and passes through positive ones. In some cases, reduced negative activations can aid in network performance, which are allowed in "Leaky" ReLU [85]:

$$f(x) = \begin{cases} x & if\ x > 0 \\ \alpha x & if\ x \leq 0 \end{cases} \tag{2.4}$$

where α is a constant multiplier < 1. To further develop on this idea, a form of Leaky ReLU with a trainable α is presented in the form of the Parametric Rectified Linear Unit (PReLU) [86]. This allows for self-configurable negative activation multipliers, which can lead to enhanced precision in the network. Otherwise, α values are user-defined (in standard Leaky ReLU), which requires multiple retraining passes to determine which α is most performant. Depending on the use case, the modified ReLU activation functions are able to outperform the original [87].

2.4.1.2 Pooling

Pooling methods are beneficial to CNN object detection in that they subsample nodes as a method of noise removal and dimensionality reduction. Images are informationally dense, and not all pixel information is relevant to object classification. Pooling aids in reducing the unnecessary extra dimensionality when processing images, which reduces computational complexity and retains only general information on the activation of the network at that layer. This aids in preventing overfitting (situation where a neural network is too sensitive to specific training examples, limiting robustness), as this leads the network to perform classifications with only the most relevant node activations [88]. In general, pooling functions operate by taking a $m \times n$-sized window of spatially sequential nodes, and computing an aggregate of their activations. This is performed on a node array by computing the pooled values at each step with a sliding window ($m \times n$) across its rows and columns. Pooling operations do not require a stride of 1 (i.e. that the sliding window traverse every single row/column), and this range of strides allows for varying window overlap arrangements. In CNNs, a pooling layer is generally only placed after an activation function layer, and this subsection presents two commonly used examples.

Average pooling is the most conceptually straightforward method of dimensionality reduction: values in each window are averaged and outputted for use in the next computation layer [89]. Applied to an entire node array, this subsamples the input while not explicitly discarding any information, as all values in a window contribute to the pooled result. However, while average pooling reduces the dimensionality of a network, the averaging process has the potential to under-represent highly activated nodes if they are in the same window as minimally activated ones. This can limit strong activations, resulting in less confident classifications.

Conversely, maximum (max) pooling intentionally prioritizes strong node activations over others: the maximum value in each window is outputted and all others are discarded. While this does increase susceptibility to overfitting, classification performance is enhanced overall. Stronger node activations result in wider activation magnitude ranges across nodes, allowing for more discrimination between possible classifications for an input image.

2.4.2 CNN Object Classifiers

While the vast majority of CNN-based object classifiers use this general format, they differ heavily in overall structure, depth (number of convolution banks), types of convolutions used, activation functions, and subsampling methods [90]. They have become continuously more sophisticated over time as researchers have worked to understand the limitations and benefits of different modifications to their sub-structures. Originally, CNNs were not widespread due to the inherent limitation in raw compute power when they were first introduced in 1989 [80]. Since then, locally-available computation power has increased by multiple orders of magnitude, enabling the use of CNNs for image classification tasks in tractable time scales.

AlexNet, a network that succeeded in the 2012 ImageNet Challenge, was one early example that demonstrated the performance potential of deep neural networks [30]. This network used sequential banks of progressively smaller convolutions (initial bank: 11×11; final bank: 3×3) with ReLU activation functions following each convolution layer, and several max-pooling layers interleaved in-between. Three fully connected layers were appended to the end to serve as the final classifiers. This network represented the state-of-the-art upon its reveal; however, its use of larger convolutions resulted in heavy computation requirements due to the required number of weights.

Very deep networks consisting of only small (3×3) convolutions were proposed to improve on the performance of large convolution-size networks in the form of the Visual Geometry Group (VGG) nets (named after the groupd that developed them) [91]. Sequential stacks of 3×3 convolutions were found to be able to emulate the effects of the larger filters with lower computational complexity and higher classification performance. The general structure proposed was as follows: five sequential stacks of 3×3 convolution layers, each stack followed by a max-pooling layer, with three fully-connected layers at the end for final classification. Each stack had a varying number of convolution banks, with progressively more convolutions per bank as the stacks progressed deeper into the network. Six networks were demonstrated with this general format, with varying computational complexity. For the larger networks, 1×1 convolution layers were placed before each pooling layer to induce additional nonlinearity in the network. While performance was increased with respect to older networks, the depths of the networks still resulted in high overall computational complexity.

The Inception CNN architecture was developed to reduce computational complexity without sacrificing performance [92]. This architecture differs significantly from the other structures in that it makes heavy use of varying-size convolution banks, from 1×1 to 7×7, in sub-network modules of parallelized convolutions and pooling layers. Each module terminates with a layer that concatenates the outputs of the parallelized operations, ensuring there is not a dimensionality explosion problem by limiting the number of internal connections from module to module. Subsequent Inception architecture versions sought to further limit the computation complexity by factorizing the larger convolutions into functionally similar sequential sub-convolutions [93].

While deeper neural networks have enhanced potential for increasingly complex classification tasks, they also become increasingly difficult to train. This generally manifests itself in longer time to training convergence, although in the worst cases overall performance can be negatively impacted with higher classification error. The ResNet CNN architecture was introduced to allow for networks with high depth while bypassing this particular limitation [94]. While the layout of the convolution banks is sequential, similar to the VGG networks, the primary contribution of this architecture is to insert shortcut connections between stacks of convolution banks (e.g. a connection from layer ℓ_i to ℓ_{i+3}). These shortcuts bypass several layers in the network and add the activations from the originating layer to the output of the target. This results in increased network depth being unable to harm the performance: network layers that are not performing critical operations for classification and could harm the training process can be bypassed. The only drawback with ResNet networks is that their high depth tends to require high computation capability. This has been mitigated by combining both Inception and ResNet structures which has allowed for highly complex networks with relatively lower computation costs [95].

2.4.3 CNN Object Detection

The structures described in Section 2.4.2 are not sufficient to perform object detection (localization) within an image; on their own they can only classify the contents. This leads to the requirement for methods that allow CNN classifiers to operate on subsets of an image input in order to both classify and estimate locations of the objects contained within. While it would be possible to exhaustively scan through an image with varying patch sizes, this quickly becomes computationally inefficient. Modern object detection techniques have therefore prioritized optimization.

The current state-of-the-art method, Faster R-CNN, consists of two primary modules: a Region Proposal Network (RPN) and a Fast Region-based CNN (R-CNN) detector [15, 96]. The RPN is a network that shares convolutional features with the classifier used in the Fast R-CNN network, and serves to propose potential Region of Interest (RoI) bounding boxes for full classification. The Fast R-CNN network then takes the set of potential regions proposed by the RPN and processes them through an additional set of convolutions and fully-connected classification layers to identify which specific RoIs correlate most strongly to a class. Since convolutional features are shared between the RPN and Fast R-CNN modules, these can be trained concurrently on image sets with ground-truth object bounding box and class annotations. Additionally, as thousands of RoIs can be generated per input image, the resulting bounding boxes of successful detections are able to tightly enclose the objects in the image, providing highly accurate localizations.

For images with high object density and occlusions, it is useful to identify not only the general bounding boxes, but also the specific pixel segments (masks) that correspond to the classified objects. Mask R-CNN builds on the capabilities of Faster R-CNN by performing additional computations in parallel to the class and bounding box estimation steps to generate object masks for each detection [97]. This is done by computing potential masks for all object classes for each RoI, and then outputting the mask that corresponds to the identified object class, discarding the others. While Mask R-CNN is highly flexible in segmentation tasks, one of the downsides (as with all segmentation networks) is that training data masks must be manually defined, heavily increasing the preparation time for new segmentation implementations. As a result, an object detector of the capability of Mask R-CNN is not always necessary if simple localization is sufficient.

In the case where extreme speed is necessary, single-shot detectors can be useful. The current state-of-the-art detectors in this field are based on the You Only Look Once (YOLO) architecture [98]. The core concept of YOLO is that object detection is performed by a single network, and bounding boxes for all object classes are predicted in an image simultaneously. The image is broken up into an $S \times S$ grid, potential bounding boxes are generated for each grid cell (box centers are placed within a cell, but the bounds can exceed the size of the cell), and the most probable bounding boxes and their respective classes are determined by using the output of the classifier network that was run on the entire input image. As the complete image is processed at once, rather than processing individual regions such as in the R-CNN methods, YOLO is highly capable in avoiding false-positive detections in the background of the image (regions containing no objects). Additional updates to the architecture further improved the performance [99, 100]. However, while the speed and general accuracy of the YOLO detectors are high, they do have difficulty detecting smaller objects within images.

2.5 Marine Mammal Energetics

Measuring animal energetics is crucial to understanding their welfare in both wild and managed settings. These provide estimates of an animal's activity levels, behavior, health, and caloric requirements. In terrestrial animals, experiments that compare the efficiency of different movement conditions can be conducted in controlled experimental settings where the motion of the body and ground reaction forces can be measured directly [101, 102]. In contrast, it has been challenging to create fully instrumented controlled experimental environments for large, swimming animals.

2.5.1 Direct Measurements

Direct measurements of animal energetics rely on the collection of physiological data in response to varying activity conditions. Early energetics studies involved tracking multi-week carbon-dioxide production through the doubly-labeled water method [103, 104]. This involves the ingestion of water marked through uncommon isotopes of oxygen and hydrogen, and measuring their elimination rate over time in regularly-drawn blood samples. While doubly-labeled water tracking is a relatively robust method that captures overall metabolic cost over a period of time, it is unable to provide metabolic measurements on a short time scale. As a result, it is insufficient to track energetics use of specific swimming patterns or behaviors, and is best served for longer-term studies.

Similar to terrestrial studies, heart-rate monitoring is also possible with swimming animals [36, 105]. This allows researchers to track animal metabolic rates with high temporal fidelity, as oxygen consumption can be tracked in real-time through physiological knowledge of its relation to animal heart rate. Additionally, with simultaneous measurements of travel speed, animal Cost of Transport (COT) can be estimated, which measures the energy per meter traveled required by the animal to propel itself. Crucially, COT is a measure of propulsive efficiency, and can yield valuable insights into how animals optimize their travel through different swimming behaviors. However, this generally requires the animal to wear a form of harness as the electrodes for the heart rate monitor must be connected to a receiver moves alongside the animal. These can hinder animal mobility, which constrains energetics monitoring to swimming tasks with limited speed variation or spatial range. So far, heart-rate monitoring hardware has also been too cumbersome to implement as a standalone package (i.e. controlled by a biologging tag).

Respirometry bridges the gap between these two methods by providing non-invasive measurements of metabolic rates for time scales on the order of seconds. Animals breathe into special apparatuses that take samples of their inhalations and exhalations and send them to a gas analyzer for partial-pressure decomposition [37, 38, 39, 106]. The flow rate and gas composition are converted into oxygen consumption per breath, which can be directly mapped to caloric use [107]. While this does not allow for real-time energetics estimation, per-activity metabolic rates can still be computed as long as the animals exclusively breathe through the respirometer. This makes the method particularly effective in managed settings given the higher likelihood of animal cooperation in experiments. For COT estimates, distances traveled must be computed between each animal breath, so in this case localization methods or high-accuracy speed sensing (for integration into distance) are key.

2.5.2 Proxy Estimation

One stands to reason that if it is possible to obtain a general metric for animal activity levels through the analysis of persistent biologging tag data, it may be possible to map this metric to measured metabolic rates. This is the rationale behind the use of ODBA as a proxy for energetics estimation [40]. ODBA operates on the concept that any acceleration signals from a biologging tag below a frequency threshold (generally the Nyquist frequency of the animal's nominal gait) correspond to gravitational accelerations and centripetal effects from turning actions, and signals above this threshold correspond to movements induced by the animal. In experimental sessions, ODBA estimates are computed alongside respirometry measurements for prescribed animal tasks, and a mapping is then computed between the two [39]. This allows researchers to then estimate continuous metabolic costs of animal activities through tag monitoring, by using the ODBA → energy relation on accelerometer data from tagging sessions that do not also include any direct measurement method. Unfortunately, this approach is not generalizable for inter-or-intra-species comparisons, as ODBA is highly sensitive to both tag placement on the animal and the specific filtering methods used to perform the high-pass cutoff.

2.5.3 Physics-Based Estimation

Forces on an animal, such as thrust for propulsion or drag acting on the moving body, are difficult to measure directly in the marine environment. This has led to a body of work performed specifically to understand the fluid dynamics surrounding cetacean propulsion. Marked animals were requested to swim past high-speed cameras, and the recorded videos were manually analyzed in post to extract high-accuracy speed and acceleration [25, 26, 27]. These data were used in physics-based hydrodynamic models coupled with simulation estimates of animal physical structure to compute drag profiles and propulsive power as it relates to swimming speed. Animal propulsive force was more intricately explored through digital particle image velocimetry, which mapped the vortical flow patterns produced by animal fluke strokes in a controlled environment [41]. So far, the majority of research in this field has focused on steady-state (consistent gait) animal propulsion, leaving a great deal of work to be done in understanding variable-gait swimming behavior.

Physics-based methods offer a platform for first-principles estimates of animal energetics, and provide a springboard for persistent metabolics monitoring using biologging tags. By combining tag-estimated animal kinematics with physics-based energetics models, it will be possible to have fully continuous power expenditure tracking, so long as animal morphometrics are well understood (these inform the energetics model sub-parameters). The models require three tag data streams to function: animal speed, acceleration, and depth. Current biologging tags are capable of reliable depth estimation, and acceleration can be obtained from on-tag IMUs or by differentiating speed measurements. As such, there is a general need for robust tag-based speed estimation that does not induce excessive parasitic drag, in order to preserve the original swimming patterns of the tagged animal [106]. This can expand energetics monitoring in both wild and managed settings, allowing for a new chapter of research on the relations between swimming behaviors and metabolics.

CHAPTER 3

Advancing Tools in Marine Monitoring: Localization and Kinematics

3.1 Neural Network Animal Tracking

The contribution in this section investigates day-scale swimming kinematics using a neural network-based computer-automated framework to quantify the positional states of multiple animals simultaneously in a managed environment. Video recordings of the animals from a two-camera system were analyzed using CNN object-detection techniques and were post-processed via Kalman filtering to extract animal kinematics. The resulting kinematic states were used to quantify bottlenose dolphin habitat usage, kinematic diversity, and movement profiles during daily life. Section 3.1.1 defines the structure itself, and Section 3.1.2 describes the post-processing methods required to obtain the results described in Section 3.1.3, and later used in the framework detailed in Chapter 5. This represents a limited form of contextualized animal monitoring, as animal identification and full 3D localization are not possible with only camera-based information in the given setting.

Content of this chapter also published as:

D. Zhang, J. Gabaldon, L. Lauderdale, M. Johnson-Roberson, L. J. Miller, K. Barton, and K. A. Shorter, "Localization and Tracking of Uncontrollable Underwater Agents: Particle Filter Based Fusion of On-Body IMUs and Stationary Cameras," *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 6575-6581, doi: 10.1109/ICRA.2019.8794141. ©IEEE 2019. Reprinted with permission.

J. Gabaldon, E. L. Turner, M. Johnson-Roberson, K. Barton, M. Johnson, E. J. Anderson, and K. A. Shorter, "Integration, Calibration, and Experimental Verification of a Speed Sensor for Swimming Animals," in *IEEE Sensors*, vol. 19, no. 10, pp. 3616-3625, May 2019, doi: 10.1109/JSEN.2019.2895806. ©IEEE 2019. Reprinted with permission.

Content of this chapter also was submitted as:

J. Gabaldon, D. Zhang, L. Lauderdale, L. Miller, M. Johnson-Roberson, K. Barton, and K. A. Shorter, "Vision-based monitoring and measurement of bottlenose dolphins' daily habitat use and kinematics," in *PLOS One*. Copyright may be transferred without notice, after which this version may no longer be accessible.

3.1.1 CNN Object Detector Structure

The overarching structure employed the Faster R-CNN objection detection platform [15]. This was chosen as Faster R-CNN is a state-of-the-art method that can enable image classifiers to detect a wide range of object scales (unlike the YOLO class [98]). Exact animal profiles were not needed, only their positions, avoiding the extreme annotation time required for a more in-depth segmenting detector such as Mask R-CNN [97].

All modules used in the implementation were present in the MATLAB Deep Learning Toolbox excepting the PReLU activation function, which was defined with a custom neural network layer per directions in the MATLAB online documentation [86, 108]. The CNN classification structure used in the Faster R-CNN framework is as follows. For the input layer, the size was chosen to be similar to the smallest bounding boxes in a set of manually scored animal profiles, in the format of $(l,l,3)$, where l is 2× the side length of the smallest bounding box major axis. This can change according to the experimental conditions (e.g. camera angle and distance from the environment), and the sizes of the animals being tracked. The input layer had a third dimension of 3 as input images were in the RGB colorspace. The feature extraction layers were developed with two separate sub-structures:

(i) Four sets of 2D 3×3 convolution layers, each followed by batch normalization, PReLU activation, and 2×2 max pooling (stride 2) layers, in that order. The four convolution layers had, in order: 64, 96, 128, and 128 filters.

(ii) Four blocks of the same layout but different numbers of convolution filters. Each block consisted of two consecutive 3×3 convolution layers, then a batch normalization and a PReLU activation layer. A 2×2 max pooling layer (stride 2) followed each block. Each convolution layer within the main four blocks had, respectively, 64, 96, 128, and 128 filters.

The first sub-structure was developed for initial animal detection tasks, and deployed in the experimental research described throughout Section 3.1. The second sub-structure is a refinement of the first, and was deployed for the later research performed in Chapter 5. Each convolution was executed with one layer of zero padding along the edges of the inputs to avoid discounting the corners/edges. The classification layers used the extracted features from the previous layers to identify an image region as either an animal or the background. They consisted of: 1) A fully connected layer, length 512, to extract features from the final convolution layer, followed by a PReLU activation; 2) A fully connected layer, length 2, to determine non-scaled classification activations; 3) A softmax function layer to convert these into the final probabilities of the image region's classification, as the final activations can be arbitrarily large in magnitude:

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \text{ for } i = 1, \ldots, K; \ \mathbf{z} = (z_1, \ldots, z_K) \in \mathbb{R} \tag{3.1}$$

The highest probability from the softmax layer corresponded to the most likely classification for the region, and the magnitude of this probability indicated the confidence of the classification.

This classifier follows along the lines of the VGG structure [91], and was chosen for its low computational complexity to enable more efficient animal detection processing for longer recording sessions. While more intricate classifiers such as ResNet or Inception may offer marginally better accuracy [92, 94], this was not necessary given that the larger networks (including the deeper versions of VGG) were originally intended to discern between a wide variety of classes, while this implementation must only classify one type of animal. While this can be extended for multiple classes (animals or other dynamic objects) with more complex classifiers, computation performance will suffer as a result.

3.1.2 Experimental Deployment

The CNN object detector was used to monitor the behavior of a group of marine mammals both qualitatively and quantitatively in a managed setting. Camera-based animal position data were used to quantify habitat usage, as well as where and how the group of animals moved throughout the day. The position data were decomposed into kinematics metrics, which were used to discriminate between two general movement states — static and dynamic — using the velocities of the tracked animals. A general ethogram of the animals' behaviors monitored in this research is presented in Table 3.1. The kinematics metrics were further used to refine our understanding of the behavioral states the animals experienced both in and out of training sessions through a combination of Kolmogorov-Smirnov (K-S) statistical analyses and joint differential entropy computations. The study protocol was approved by the University of Michigan Institutional Animal Care and Use Committee and the Brookfield Zoo.

Approximately 99.5 hours of data from two cameras were collected for this experiment, resulting in $\sim$ 14 million individual frames of data. To extract spatial information about habitat use and swimming kinematics, we first needed to identify animals in the frames. These detections were filtered and associated with short trajectories (tracklets) from individual animals. Kinematic data (position, velocity and heading) from the tracklets were then used to parameterize and form probability distributions that were used to identify tendencies in animal motion during In Training Session (ITS) and Out of Training Session (OTS) swimming.

Table 3.1: Behavior condition ethogram of dolphins under professional care

Category	Behavior	Definition
ITS	Animal Care Session	Time period in which animal care specialists work with the dolphins to learn new behaviors or practice known behaviors without public audience.
ITS	Formal Presentation	Time period in which animal care specialists work with the dolphins in front of an audience to present educational information to the public.
OTS	Static	Animal movement state with little to no active fluking at a rate of speed less than $0.5\,\mathrm{m\,s^{-1}}$.
OTS	Dynamic	Animal movement state with active fluking at a rate of speed greater than $0.5\,\mathrm{m\,s^{-1}}$.

3.1.2.1 Experimental Environment

Seven bottlenose dolphins with an average age of 17 ± 12 yrs and length of 247 ± 17 cm were observed using a dual-camera system in the Seven Seas building of the Brookfield Zoo, Brookfield IL. The complete environment consists of a main habitat with public viewing, two smaller habitats behind the main area, and a medical habitat (not shown) between the two smaller habitats (Fig. 3.1). The main habitat is 33.5 m across, 12.2 m wide, and 6.7 m deep. The back habitats have circular diameters of 10.7 m and are 4.3 m deep, and the medical area is 7.6 m in diameter and 2.4 m deep. The habitats are connected through a series of gates. During formal training sessions in the main habitat, animal care specialists primarily engage with the animals on the island between the gates to the other areas. There are underwater observation windows for the viewing public on the far side of the main habitat from the island (not shown), and smaller windows looking into the offices of the animal care specialists on the island and next to the right gate (Fig. 3.2, top). Recordings of the main habitat took place across multiple days (between Feb. 6 and March 27, 2018), for varying portions of each day, for a total of 99.5 hours over 20 recordings. Data collection began at the earliest at 07:41 and ended at the latest at 16:21. During the recorded hours, the dolphins participated in four formal training sessions according to a regular, well-defined Animal Care Specialist (ACS)-set schedule.

A formal training session consisted of time in which the ACSs work with the dolphins to learn new behaviors or practice known behaviors. At the beginning of each formal training session, the dolphins were asked to maintain positions directly in front of the ACS (formally known as “stationing”). The animal care specialists then presented discriminative stimuli or gestures that indicated which behaviors they requested each dolphin produce. If a dolphin produced the desired behavior, they received a reward (i.e., reinforcement). If they chose not to produce the behavior, a specialist

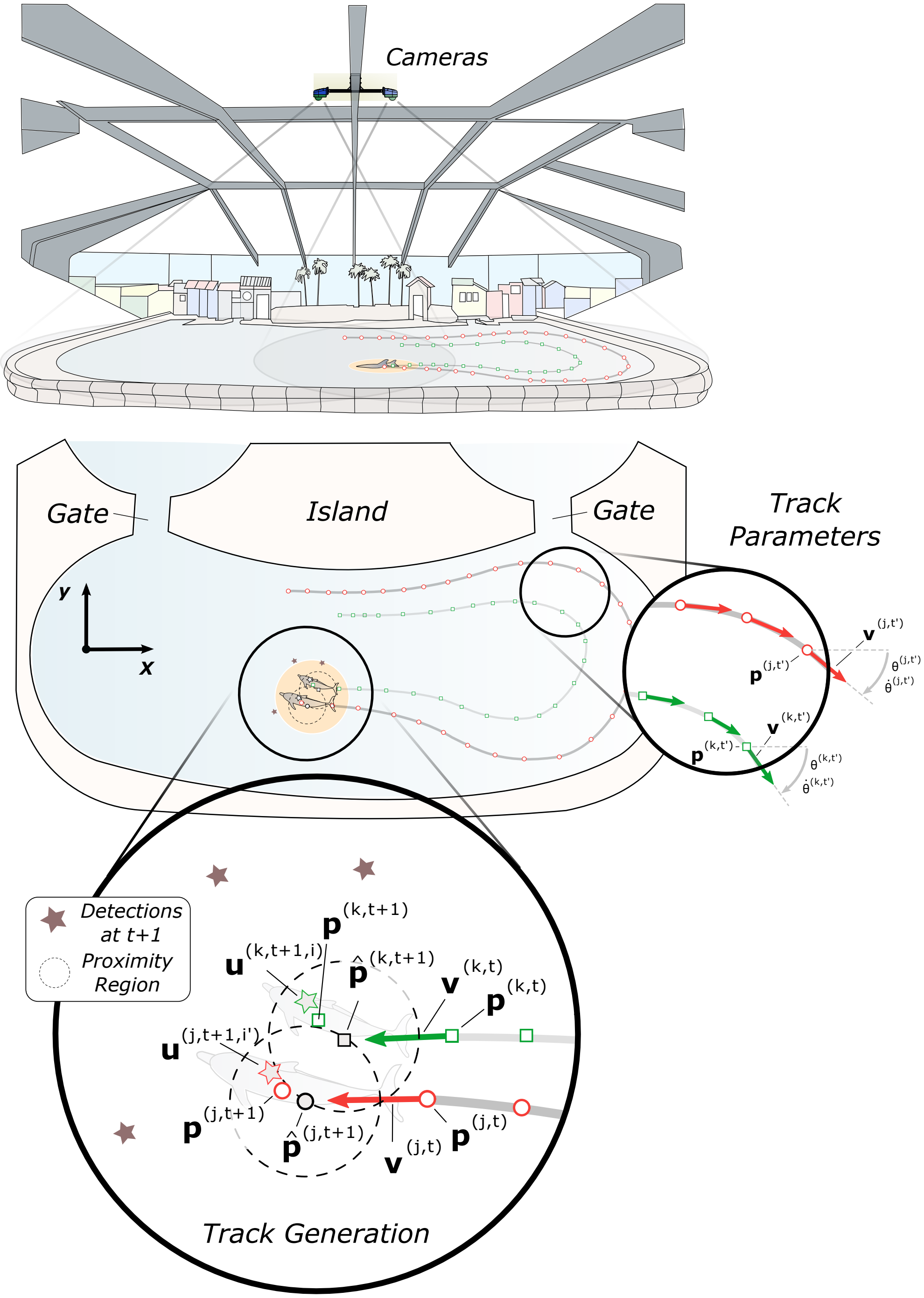


Figure 3.1: Diagram of the experimental setup. TOP: Illustration of the main habitat, with camera placements (blue enclosures) and fields of view (gray cones). BOTTOM: x-y view of example tracklets (red and green on gray lines) of two dolphins (highlighted light orange), which are also visible in the top of this figure. BOTTOM-ZOOM (RIGHT): Vector illustrations of the two example tracks. Example notation for tracklet j (red): position ($\mathbf{p}^{(j,t')}$), velocity ($\mathbf{v}^{(j,t')}$), yaw ($\theta^{(j,t')}$), and yaw rate ($\dot{\theta}^{(j,t')}$). BOTTOM-ZOOM (LEFT) Illustration of tracklet generation, with detections (stars) and tracklet proximity regions (dashed). Example notation for tracklet j (red): position ($\mathbf{p}^{(j,t)}$), velocity ($\mathbf{v}^{(j,t)}$), Kalman-predicted future position ($\hat{\mathbf{p}}^{(j,t+1)}$), true future position ($\mathbf{p}^{(j,t+1)}$), and future animal detection ($\mathbf{u}^{(j,t+1,i')}$).

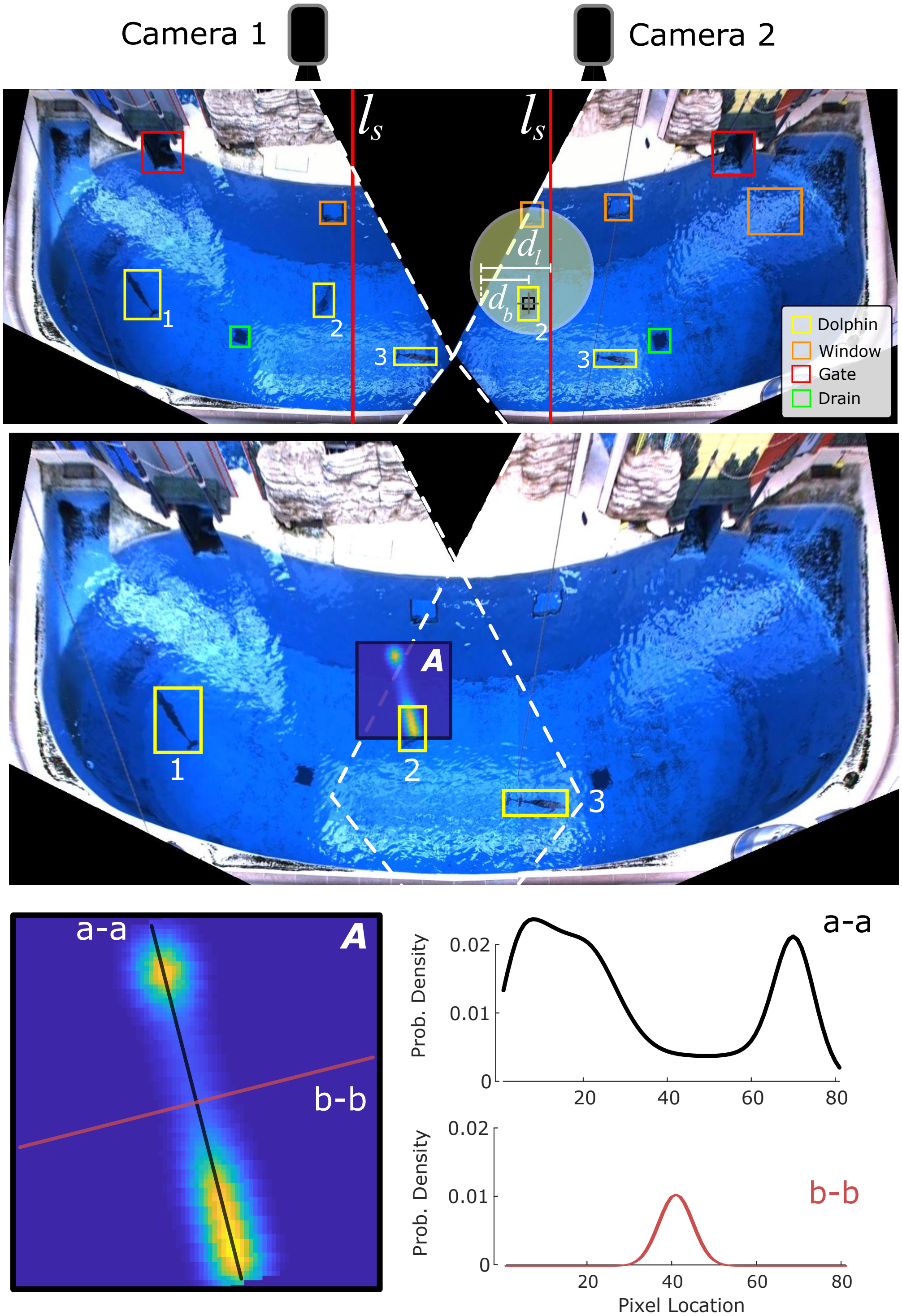


Figure 3.2: Combined figure demonstrating camera overlap, bounding box meshing, and animal position uncertainty. TOP: Transformed individual camera views, with objects in the habitat marked. Yellow – Dolphin bounding boxes, Green – Drains, Red – Gates between regions, Orange – Underwater windows (3 total). Correlated bounding boxes are indicated by number, and the habitat-bisecting lines (l_s) for each camera frame in solid red. Distances from Box 2 to the closest frame boundary (d_b) and the boundary to the bisecting line (d_l) are highlighted in yellow. MIDDLE: Combined camera views including dolphin bounding boxes (yellow), with the location uncertainty distribution (A) overlaid for Box 2. BOTTOM: 2D uncertainty distribution (A) with major (a-a, black) and minor (b-b, red) axes labeled and separately plotted.

Table 3.2: Block time intervals

	Time Interval	
Block	OTS	ITS
1	08:00 – 09:30	09:30 – 10:00
2	10:00 – 11:30	11:30 – 12:00
3	12:00 – 13:00	13:00 – 13:30
4	13:30 – 14:30	14:30 – 15:00
5	15:00 – 16:00	N/A

The ITS blocks (1 and 3) are animal care sessions, and the OTS blocks (2 and 4) are formal presentations.

may request the same behavior again or move on to a different behavior. When the animals were in a formal training session (abbreviated ITS), they experienced two formats of training during the data collection period: non-public animal care sessions and formal public presentations. Time outside of formal training sessions (abbreviated OTS) was defined as when the animals were not interacting with ACSs. During the OTS time periods, the ACSs would provide enrichment objects for the animals to interact with and select which parts of the habitat the animals could access using gates on either side of the main island. The time intervals for the OTS and ITS blocks are displayed in Table 3.2.

3.1.2.2 Experimental Equipment

Two AlliedVision Prosilica GC1380C camera sensors with Thorlabs MVL5M23 lenses were separately mounted in Dotworkz D2 camera enclosures, which were attached to 80/20 T-slotted aluminum framing. On the frame, the cameras were spaced approximately 2m apart. The frame was mounted to a support beam directly above the main habitat, with the cameras angled to give full coverage of the area when combined. Figure 3.1, top, illustrates the habitat, camera placement, and field of view coverage. For data collection, the cameras were connected through the Gigabit Ethernet protocol to a central computer with an Intel i7-7700K CPU. Recordings were executed using the MATLAB Image Acquisition Toolbox, in the RGB24 color format at a frame rate of 20Hz. Each camera was connected to a separate Ethernet port on an internal Intel PRO/1000 Pt PCIe card. A separate computer system was used for detection inference, and was outfitted with an Intel i7-8700K processor clocked to 4.8 GHz and a Nvidia Titan V graphics processing unit in Tesla Compute Cluster mode.

3.1.2.3 Network Training

Ground truth data were scored by a trained observer by manually defining bounding boxes that identified the locations of the dolphins in the training/testing frames (Fig. 3.2, A). These ground truth data were selected over a range of lighting conditions and dolphin locations to ensure robustness of the detection network. For each camera, 100 frames were extracted from each of 11 separate recordings, with evenly spaced time intervals between frames. The recordings were collected in May 2017, and February, March, and August 2018. Over 940 frames from each of the left and right cameras were found to contain usable dolphin locations, i.e. human-detectable dolphin profiles. Each usable dolphin location in the selected frames was manually given a bounding box tightly enclosing the visible profile. The detector for the left camera was trained on 1564 profiles and tested on 662, and the detector for the right camera was trained on 1482 profiles and tested on 662. The dolphin detectors were trained using the MATLAB implementation of Faster R-CNN, employing the first version of the previously-defined CNN structure as the classification method.

3.1.2.4 Detection Post-Processing

Detections were performed over all 99.5 hours of recorded data from both cameras, at 10Hz intervals (total of 7.16×10^6 frames), using a 95% minimum confidence threshold to ensure accuracy. The fields of view of the two cameras overlap for a portion of the habitat, resulting in some dolphins being detected simultaneously by both cameras. This yielded multiple sets of conflicting detection bounding boxes spanning the two fields of view, which necessitated associating the most likely left/right box pairs. Before conflict identification was performed, the detection boxes were first transformed into a common plane of reference termed the world frame. Using known world point coordinates, homographies from each camera to the world frame were generated using the normalized Direct Linear Transform method [109]. These homographies were used to convert the vertices of the bounding boxes to the world frame using a perspective transformation. Intersecting boxes were identified by evaluating polygonal intersections, and Intersection over Union (IoU) metrics were computed for intersecting boxes to measure how well they matched. Associations were identified between pairs of left/right intersecting boxes with the highest mutual IoU values.

Associated boxes' world frame centroid locations were meshed using a weighted mean. First, the boundaries of each camera's field of view were projected into the world frame, allowing us to obtain the line in the world frame y-direction defining the center of the overlap region, denoted $l_s = x_{mid}$ (Fig. 3.2, middle). x_{mid} is the x-coordinate in the world frame midway between the physical placement of the cameras. For each detection (u), the distance (d_b) in the x-direction from u to the nearest projected camera boundary line (b_n) was then determined. Next, the distance (d_l) in the x-direction from line l_s through u to b_n was found. Finally, the weight for the camera corresponding to b_n was calculated as $w_n = d_b/2d_l$, with the weight for the other (far) camera as $w_f = 1 - w_n$. This ensured that if detection u was on l_s, then $w_n = w_f = 0.5$, and as u moved closer to b_n, we would have $w_n \rightarrow 0$ and $w_f \rightarrow 1$.

In specific circumstances, the shapes of the drains at the bottom of the habitat were warped by the light passing through rough surface water, and resulted in false dolphin detections. Separate (smaller) image classifiers for each camera were trained to identify these false positive drain detections, and were run on any detections that occurred in the regions of the video frames containing the drains. These detectors were strictly CNN image classifiers and were each trained on over 350 images and tested on over 150 images. For the drain detector, the input layer size had the format of $(l_d, l_d, 3)$, where l_d is the mean side length of the detection bounding boxes being passed through the secondary classifiers. The feature detection layers had the same general structure as the Faster R-CNN classifier network, except in this case the convolution layers had, in order: 32, 48, 64, and 64 filters each. In the classification layers, the first fully connected layer had a length of 256.

3.1.2.5 Tracklet Formation

Each experimental session involved the detection of multiple animals throughout their habitat. However, animal detections were done independently for each frame of the video. To extract kinematic information from the animals in the video, the detection associations needed to be preserved across frames. Short continuous tracks (i.e. *tracklets*) were generated for each detected animal instance by identifying the most likely detection of that animal in the subsequent frame. To generate multiple individual tracklets in series of video frames, an iterative procedure of *prediction* and *association* was conducted under a Kalman filter framework with a constant velocity model. This is not necessary when only a single animal is present in the environment, or if the animal can be individually identified by the classifier. However, given the resolution and distance from the habitat of the cameras, identification was rendered impossible, requiring the use of tracklets as a result. *The method for generating tracklets from animal detections was devised and executed by D. Zhang*[1]*, and is included in this dissertation for completeness.*

The position of the i-th detected animal in one video frame at time t is denoted as $\mathbf{u}^{(t,i)} = [u_x^{(t,i)}, u_y^{(t,i)}]$. Each detection, $\mathbf{u}^{(t,i)}$ was either associated with a currently existing tracklet or used to initialize a new tracklet. To determine which action was taken, for each tracklet, denoted as $\mathbf{T}^{(k)}$ for the k-th tracklet, this process first predicted the state of the tracked animal in the next frame ($\widehat{\mathbf{T}}^{(k,t+1)}$) based on the current state information of the animal $\mathbf{T}^{(k,t)}$.

---

[1] In: J. Gabaldon, D. Zhang, L. Lauderdale, L. Miller, M. Johnson-Roberson, K. Barton, K. A. Shorter, "High-fidelity bottlenose dolphin localization and monitoring in a managed setting," To be published in *Science Robotics*.

$$\begin{aligned}
\mathbf{T}^{(k,t)} &= [\mathbf{p}^{(k,t)}, \mathbf{v}^{(k,t)}] && (3.2)\\
&= [p_x^{(k,t)}, p_y^{(k,t)}, v_x^{(k,t)}, v_y^{(k,t)}] && (3.3)\\
\widehat{\mathbf{T}}^{(k,t+1)} &= [\hat{\mathbf{p}}^{(k,t+1)}, \hat{\mathbf{v}}^{(k,t+1)}] && (3.4)\\
&= [\hat{p}_x^{(k,t+1)}, \hat{p}_y^{(k,t+1)}, \hat{v}_x^{(k,t+1)}, \hat{v}_y^{(k,t+1)}] && (3.5)
\end{aligned}$$

where $\mathbf{p}^{(k,t)} = [p_x^{(k,t)}, p_y^{(k,t)}]$ denotes the filtered position of the animal tracked by the k-th tracklet at time t and $\mathbf{v}^{(k,t)} = [v_x^{(k,t)}, v_y^{(k,t)}]$ is the corresponding velocity. Under a constant velocity model, the predicted next frame position $\hat{\mathbf{p}}^{(k,t+1)} = [\hat{p}_x^{(k,t+1)}, \hat{p}_y^{(k,t+1)}]$ was obtained by integrating the current velocity over one frame period and summing this to the current frame position. The predicted velocity remained constant.

$$\begin{aligned}
\hat{p}_x^{(k,t+1)} &= p_x^{(k,t)} + v_x^{(k,t)}\Delta t && (3.6)\\
\hat{p}_y^{(k,t+1)} &= p_y^{(k,t)} + v_y^{(k,t)}\Delta t && (3.7)\\
\hat{v}_x^{(k,t+1)} &= v_x^{(k,t)} && (3.8)\\
\hat{v}_y^{(k,t+1)} &= v_y^{(k,t)} && (3.9)
\end{aligned}$$

Using the predicted position, the k-th tracklet checked whether there existed a closest detection in the next frame that was within the *proximity region* of the predicted position. If true, that detection, denoted as $\mathbf{u}^{(k,t+1,i)}$ for the i-th detection in frame $t+1$ associated with the k-th tracklet, was used as the reference signal of the Kalman filter to update the state (position and speed) of tracklet $\mathbf{T}^{(k)}$. If false, the unassociated tracklet continued propagating forward, assuming the animal maintained a constant velocity. If a tracklet continued to be unassociated for 5 consecutive frames (empirically determined), it was considered inactive and was truncated at the last confirmed association. All information related to the k-th tracklet was saved after its deactivation:

$$\mathbf{T}^{(k)} = [\mathbf{T}^{(k,t_{start})}, \ldots, \mathbf{T}^{(k,t-1)}, \mathbf{T}^{(k,t)}, \mathbf{T}^{(k,t+1)}, \ldots, \mathbf{T}^{(k,t_{end})}]^T \qquad (3.10)$$

As illustrated in Figure 3.1, the tracklet formation operation linked each animal's individual detections ($\mathbf{u}$) over consecutive frames and returned not only the positions ($\mathbf{p}$) of the animals, but also the forward speed (v), yaw (heading, θ), and turning rate ($\dot{\theta}$), which could then be used to parameterize the positional states of the animals.

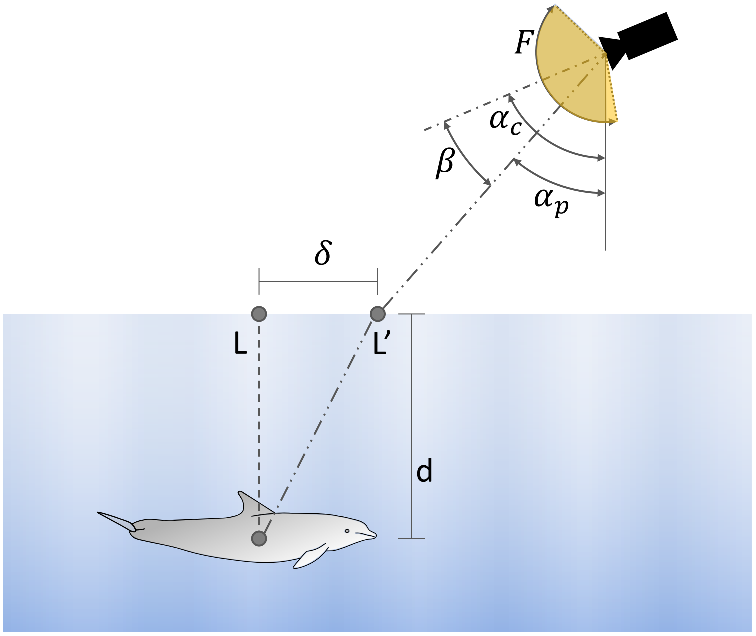


Figure 3.3: Diagram of projection and refraction effects on estimated dolphin location. L' is the location of the dolphin image apparent to the camera when converted to world-frame coordinates, and L is the true position. Refraction and camera perspective effects then cause the dolphin at depth d to be perceived at position L' with an offset of δ. The yellow region represents the camera FOV.

3.1.2.6 Position Uncertainty

There was a general position uncertainty for each animal detection due to noise in the Faster R-CNN detections. This was caused by a combination of limited camera resolution and distortion of an animal's image from waves and ripples on the surface of the water. Additionally, since the homographies used to transform detections from camera to world frame are computed with respect to the plane of the water's surface, a detection corresponding to an object not on the surface would be incorrectly "shifted" along this 2D plane. This is due to refraction and perspective effects (Fig. 3.3), where an agent's observed 2D position is shifted from L to L'.

For accurate 3D localization, it is important to find the "shifts" for an animal when it is below the surface, i.e. correctly project the "shifted" 2D image position to a 3D world position. For a known depth, a position in top-down view 2D (Fig. 3.2, middle) is projected into 3D via the following steps:

(i) Before any shift computations, the detections are projected back into their respective camera frames. This ensures the camera resolution and Field of View (FOV) values are used in the correct frame of reference.

(ii) Assuming the camera orientation is known, α_c, the angle of the camera image center with respect to the world-frame z-axis (world vertical), can be extracted. Using this convention,

α_c is always ≥ 0. The angle offset of a detection w.r.t. the camera image center is given by:

$$\beta = \tan^{-1}\left[(2p/r - 1)\tan(F/2)\right] \tag{3.11}$$

with $\beta \in [-F/2, F/2]$. p is the pixel value of the object in the image, r is the full frame resolution in pixels, and F is the FOV in radians. The angle of the detection with respect to the world vertical is then given by:

$$\alpha_p = \beta + \alpha_c \tag{3.12}$$

(iii) Next, use the detection angle to compute the necessary shift with respect to the camera orientation. Using Snell's Law, the shift is given by:

$$\delta = d \tan\left[\sin^{-1}\left(\frac{n_a}{n_w}\sin(\alpha_p)\right)\right] \tag{3.13}$$

where d is the depth of the tag in meters, $n_a = 1.0003$ is the refraction coefficient of air, and $n_w = 1.330$ is the refraction coefficient of water.

(iv) Apply steps (ii) and (iii) for both x and y image axes. These are the shifts $(\delta_{x,c}, \delta_{y,c})$ w.r.t. the camera.

(v) The camera orientation also yields the yaw with respect to the world frame, and this can be used to rotate the computed δ values into the world frame through a 2D rotation matrix to obtain δ_w. The true detection x-y location, L, is then obtained with:

$$L = L' + \delta_w \tag{3.14}$$

where L' is the original location estimate.

Despite the use of two cameras, lens FOVs did not allow for full stereo imaging, so animal depth could not be measured with this experimental setup. This induced uncertainties in the world-frame x-y location estimates that could not be corrected. The detection uncertainty was therefore represented as a 2D Probability Density Function (PDF), whose size and shape depended on the location of the detection with respect to the cameras (Fig. 3.2, bottom). The short (minor) axis, D_1, was a Gaussian uncertainty distribution defined according to a heuristically estimated error in the camera detections (~ 0.2 m), and represented the general position uncertainty in the Faster R-CNN detections (Fig. 3.2, bottom, b-b). The long (major) axis of the distribution, D_2, represented the position uncertainty caused by the perspective and refraction effects (uncertainty from unknown depth). A 1D PDF was defined according to previously measured animal depth data (total of 9.8 hours during separate OTS time blocks), obtained via non-invasive tagging, which represented the general distribution of depths occupied by the animals. This was convolved with D_1 to produce the general shape of D_2 (Fig. 3.2, bottom, a-a). The x-axis length scale for D_2 for a particular detection was obtained from the maximum position error in the detection's x-y location. This was the magnitude of the x-y position difference (original versus corrected x-y position) if the detection happened to be at maximum depth (~ 7 m). This magnitude varied dependent on the world-frame original location of the detection.

3.1.2.7 Mapping Animal Kinematics to Habitat

Heatmaps of dolphin position and speed were used to map animal positional state to the habitat. The dolphins were defined to be static or minimally mobile (drifting) when they were traveling at speeds below $0.5\,\mathrm{m\,s^{-1}}$, and dynamic otherwise. To generate the positional heatmaps, a blank 2D pixel map of the main habitat, M, was first created. Then, for each pixel representation p of a detection u, the maximum possible magnitude of location error due to depth was determined, defined as e_m (pixels, scale 1 pix = 5 cm), along with the orientation of the error propagation, ψ_m (radians). The perimeter of the habitat served as a hard constraint on the location of the animals, thus e_m was truncated if the location of the point with the maximum possible shift, $[p_x + e_m \cos(\psi_m), p_y + e_m \sin(\psi_m)]$, fell outside this boundary. The minor axis of the 2D uncertainty distribution, D_1, was a 1D PDF in the form of a Gaussian kernel with $\sigma_{gauss} = 0.2s$ (0.2 meters scaled to pixels by scaling factor $s = 20$). Next, the depth PDF was interpolated to be e_m pixels long, and was convolved with D_1 (to account for measurement uncertainty in the camera detections). This yielded the major axis 1D PDF, D_2. The 2D (unrotated) occupancy PDF, $E = D_1^\top D_2$, was then computed, where D_1, D_2 were horizontal vectors of the same length. The 2D rotated occupancy PDF, F, was calculated by rotating E by an angle of ψ_m through an interpolating array rotation. The MATLAB implementation of `imrotate` was used for this calculation. F was then normalized to ensure the distribution summed to 1. Finally, F was locally summed into M, centered at location $[x_u, y_u] = [p_x + 0.5e_m \cos(\psi_m), p_y + 0.5e_m \sin(\psi_m)]$, to inject the occupancy probability distribution for u into map M. This process was then repeated for all detections. For the sake of visibility, all heatmaps were sub-sampled down to the scale of 1 pix = 1 meter.

A similar process was used to form the speed heatmaps. In a speed heatmap, the values of F are additionally scaled by the scalar speed of the animal, v, that corresponds to detection u, and then locally summed into a separate map, N (sum $F \cdot v$ into N centered at $[x_u, y_u]$). Element-wise division of N by M was performed to generate S, a map of the average speed per location.

Lastly, the direction of motion of the animals throughout the monitored region was described using a quiver plot representation. To formulate the quiver plot, two separate heatmaps were generated, Q_x and Q_y, one each for the x and y components of the animals' velocities. Q_x was created using a similar method to the speed heatmap, but in this case F was scaled by the x-

component of the animal's velocity (sum $F \cdot v \cos(\theta)$ into Q_x centered at $[x_u, y_u]$), where θ was the heading of the animal corresponding to detection u. Similarly for Q_y, F was scaled by the y-component of the animal's velocity (sum $F \cdot v \sin(\theta)$ into Q_y centered at $[x_u, y_u]$). The vector components Q_x and Q_y combined represented the general orientation of the animals at each point in the habitat.

3.1.2.8 Probability Distributions of Metrics and Entropy Computation

For each time block of OTS and ITS, the PDFs of speed ($\mathrm{m\,s^{-1}}$) and yaw (rad) were numerically determined. These were obtained by randomly extracting 10^5 data samples of both metrics from each time block of OTS and ITS, and producing PDFs for each metric and time block from these data subsets.

Additionally, the joint differential entropies of speed and yaw were computed for each time block of OTS and ITS. In this case, the joint entropy of animal speed and yaw represents the coupled variation in these metrics for the animals. This indicates that speed-yaw joint entropy can be considered a proxy for measuring the diversity of the animals' kinematic behavior. To compute the joint entropy h for one time block, the randomly sampled speed (continuous random variable $\mathbf{S}$) and yaw (continuous random variable $\mathbf{\Psi}$) data subsets (S and Ψ, respectively) of that time block were used to generate a speed/yaw joint PDF: $f(s, \psi)$, where $s \in S$, $\psi \in \Psi$. f was then used to compute h with the standard method:

$$h(\mathbf{S}, \mathbf{\Psi}) = -\int_{S,\Psi} f(s, \psi) \ln[f(s, \psi)] ds d\psi \tag{3.15}$$

3.1.2.9 Kolmogorov-Smirnov Statistics

To evaluate the statistical differences in animal dynamics between time blocks, the two-sample K-S distances (Δ_{ks}) and their significance levels (α) were computed for each of the following metrics: speed ($\mathrm{m\,s^{-1}}$), yaw (rad), yaw rate ($\mathrm{rad\,s^{-1}}$), and the standard deviations of each [110]. These were done by comparing randomly-sampled subsets of each time block, with each subset consisting of 10^4 data samples per metric. Only time blocks of similar types were compared (i.e. no ITS blocks were compared to OTS blocks, and vice-versa). The computations were performed using the MATLAB statistics toolbox function `kstest2`.

3.1.3 Results

3.1.3.1 Detector and Filter Performance

During evaluation, the Faster R-CNN detectors for the left and right cameras achieved Average Precision scores of 0.76 and 0.78, respectively. The CNN drain classifiers for the left and right cameras achieved respective accuracy scores of 92% and 94%. Processing all 99.5 hours of recordings yielded 5.92×10^6 detections for the left camera and 6.35×10^6 detections for the right. The initial set of detections took ~ 8.4 days to compute when performed on the Titan V computer system. Of these, 3.83×10^4 (0.65%) detections from the left camera and 3.02×10^4 (0.48%) detections from the right camera were found to be drains misclassified as dolphins. After removing the misclassified detections, meshing the left and right detection sets yielded a total of 1.01×10^7 individual animal detections within the monitored habitat. The tracklet generation method used in this work associated animal track segments containing gaps of up to 4 time steps. As a result, the prediction component of the tracklet generation method's Kalman filter was used to fill in short gaps in the tracking data. Generating tracklets from the meshed detections yielded a total of 1.24×10^7 estimated dolphin locations, from 3.44×10^5 total tracklets.

3.1.3.2 Spatial Distribution — Position

During OTS, the tracked animals were found to be in a dynamic swimming state ~ 77% of the time and a static state for ~ 23% of the time. The static OTS behavior tended to be associated with particular features of their habitat: the gates that lead to the other areas of the habitat or at the underwater windows that offered views of the animal care specialist staff areas (Fig. 3.4). When swimming dynamically during OTS, the dolphins tended to spend more time near the edges of their habitat, with the most time focused on the island side with the gates and the windows (Fig. 3.5, left column). This was especially true during Block 5, with additional weight placed along the edge of the central island.

Throughout ITS, the dolphins were asked to engage in dynamic swimming tasks ~ 62% of the time, and were at station (in front of the ACSs) for the remaining ~ 38% of the time. During ITS, the dolphins had a heavy static presence in front of the central island, where the animals were stationed during formal training programs. Less emphasis was placed on the edges, contrasted to their locations during OTS (Fig. 3.6, left column). During ITS, the ACSs presented discriminative stimuli or gestures corresponding to specific animal behavior, which defined the spatial distributions of the dolphins' movements during these time blocks. Additionally, there were spatial distribution similarities between training sessions of similar type, e.g. Blocks 1, 3 were animal care and husbandry sessions, and 2, 4 were formal public presentations. Note the structure of the spatial distributions across the top of their habitat, where during the care sessions (Blk. 1, 3) the dolphins'

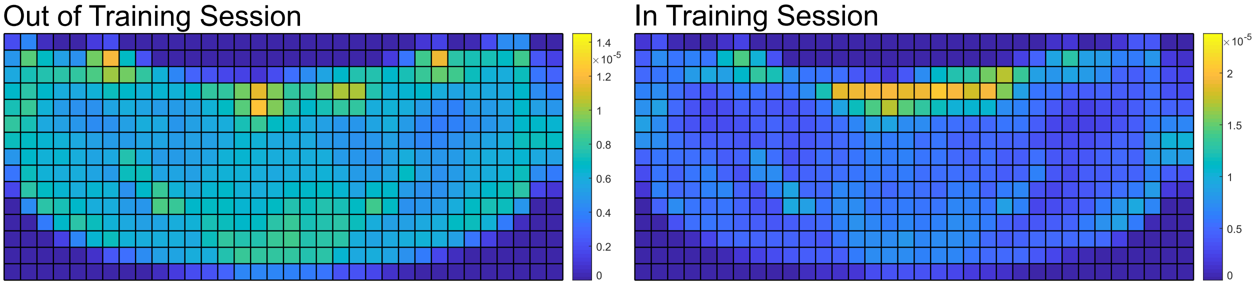


Figure 3.4: Static position distributions for OTS and ITS. A note on the format of the training sessions: Dolphins spent more time stationed at the main island during public presentations than non-public animal care sessions. During formal public presentations, ACSs spend a higher portion of the training session on the main island because it is within view of all of the public attending the presentation. Non-public animal care sessions are more fluid in their structure than public sessions. ACSs often use the entire perimeter of the habitat throughout the session.

positions were focused on specific points in the area, while during the presentations (Blk. 2, 4) their positions were distributed across the edge of the central island. This captured the formation used during presentations with animals distributed more uniformly across the island.

3.1.3.3 Spatial Distribution — Speed/Quiver

In Block 1 of OTS, the dolphins had relatively low speeds (mean $1.30\,\mathrm{m\,s^{-1}}$) across their habitat, and based on the vector field of the quiver plot for the block, were engaged in large, smooth loops along the edges of the habitat (Fig. 3.5, right column). This was contrasted with Block 2, which saw a higher general speed (mean $1.57\,\mathrm{m\,s^{-1}}$) as well as diversified movement patterns, with the right half exhibiting counter-clockwise chirality while the left half maintained the clockwise motion pattern. Blocks 3-5 exhibited higher mean speeds (Blk. 3: $1.45\,\mathrm{m\,s^{-1}}$, Blk. 4: $1.41\,\mathrm{m\,s^{-1}}$, Blk. 5: $1.43\,\mathrm{m\,s^{-1}}$) than Block 1, and lower than 2, with the dolphins' movement patterns shifting between each OTS block.

During ITS, the care blocks' (Blk. 1, 3) speed distributions and vector fields qualitatively demonstrated similar structures, while those of the presentations (Blk. 2, 4) were more mixed, with more similarities along the left and right far sides, but fewer in the center (Fig. 3.6, right column). The mean speeds did not share particular similarities between blocks of similar type (Blk. 1: $1.39\,\mathrm{m\,s^{-1}}$, Blk. 2: $1.45\,\mathrm{m\,s^{-1}}$, Blk. 3: $1.44\,\mathrm{m\,s^{-1}}$, Blk. 4: $1.39\,\mathrm{m\,s^{-1}}$).

3.1.3.4 Statistical Comparison of Metrics

Figure 3.7, top, displays the overlaid PDFs of the speed and yaw metrics during OTS, and Figure 3.7, middle, displays the PDFs during ITS. The K-S distances for all six metrics were reported in Table 3.3, with all values rounded to 3 digits of precision. For OTS, we saw from the K-S results

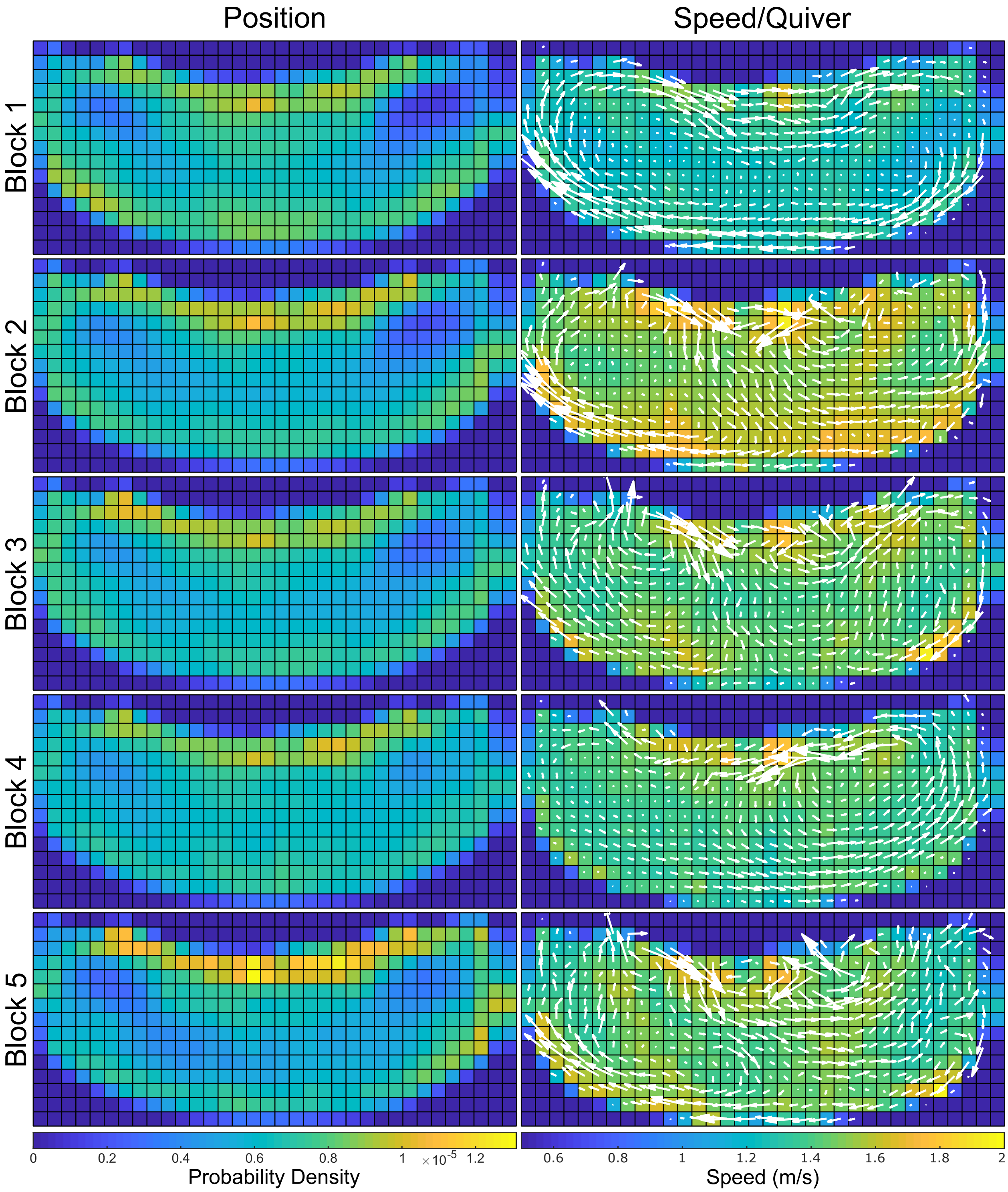


Figure 3.5: Spatial distributions for dynamic OTS, with position distributions along the first column and speed distributions/quiver plots along the second column. Prior to the first full training session of the day at 9:30 a.m., the dolphins were engaged in low intensity (resting) swimming clockwise around the perimeter of the habitat, with the highest average OTS speeds recorded after the 9:30 sessions. From there, speeds trail off for the subsequent two time periods. The 1:30-2:30 p.m. time block is characterized by slower swimming in a predominantly counterclockwise pattern. There is an increase in speed and varied heading pattern during the 3:00-4:00 time block.

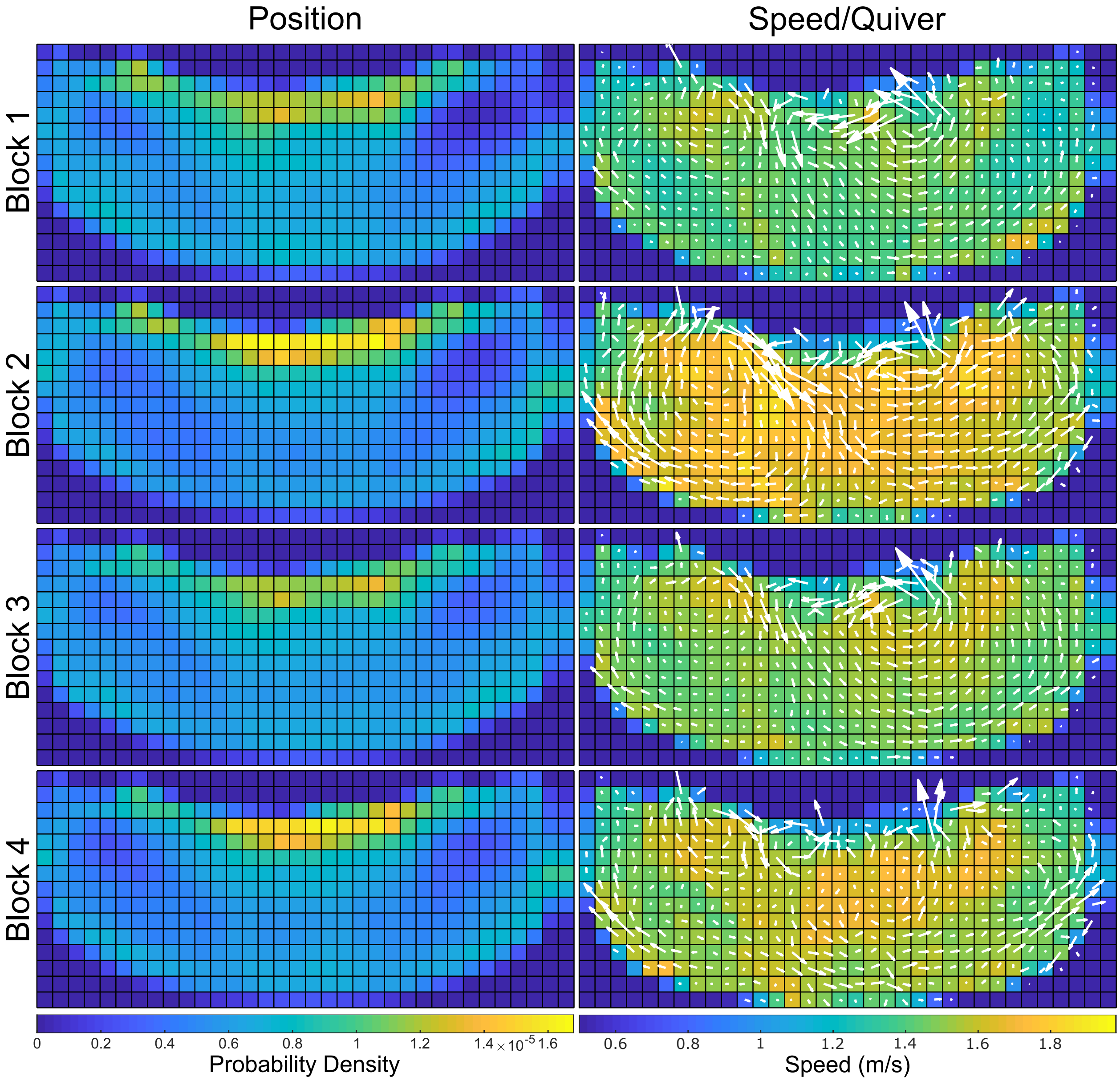


Figure 3.6: Spatial distributions for dynamic ITS, with position distributions along the first column and speed distributions/quiver plots along the second column. Speeds across the entire habitat are higher during public presentations than non-public animal care sessions because high-energy behaviors (e.g., speed swims, porpoising, and breaches) are typically requested from the group several times throughout the presentation. Though non-public presentations include high-energy behaviors, non-public animal care sessions also focus on training new behaviors and engaging in husbandry behaviors. Public presentations provide the opportunity for exercise through a variety of higher energy behaviors, and non-public sessions afford the ability to engage in comprehensive animal care and time to work on new behaviors.

that Blocks 1 and 2 varied the most with respect to the others in terms of speed, which was observed in Figure 3.7, top, while the yaw values were not generally significantly different, again observed in Figure 3.7, middle, (given the high number of samples used to generate the K-S statistics, we were able to compare the significance levels to a stronger threshold of $\alpha_{crit} = 0.001$). Across the board, Block 2 generally differed significantly from the rest of the OTS blocks for the most metrics, with Block 1 following close behind. In contrast, Blocks 3-5 differed the least significantly from each other, indicating similarities in the dolphins' dynamics patterns for Blocks 3-5.

For ITS, we note that the significant differences in metrics generally followed the structure type of each ITS block: comparisons between Blocks 1 vs. 3, and 2 vs. 4, were found to be significantly different the least often. This was to be expected, given Blocks 1 and 3 were animal care sessions, and 2 and 4 were presentations. Of particular note are the yaw std. dev. and yaw rate std. dev. metrics, with entire order of magnitude differences in K-S distances when comparing similar vs. different types of ITS blocks.

3.1.3.5 Speed and Yaw Joint Entropy

The joint differential entropies of speed and yaw per time block are displayed in Figure 3.7, bottom, with values reported in Table 3.4. The time blocks in this figure were presented in chronological order, and with that in mind we observed that the first blocks of each OTS and ITS had the least joint entropy (variation in speed and yaw throughout the time block), followed immediately by a peak in the second block of each. Subsequent time blocks for both OTS and ITS then yielded lower entropies that were sustained. Overall, ITS blocks were observed to have higher speed-yaw joint entropy than OTS blocks in similar time windows.

3.1.4 Discussion

3.1.4.1 Automatic Dolphin Detection

This contribution presents a framework that enables the persistent monitoring of managed dolphins through external sensing, performed on a scale that would otherwise require a prohibitively high amount of human effort. Both the Faster R-CNN dolphin detection and CNN drain detection methods displayed reliable performance in testing, and enabled large-scale data processing at rates not achievable by humans. Given that the total duration of video processed was $\sim$ 199 hours (2 cameras $\times$ 99.5 hours each), an inference time of $\sim$ 202 hours (1.013×) represents at minimum an order-of-magnitude increase in processing speed when compared to human data annotation. This estimate was obtained from prior experience in manual animal tracking, which could take over 10 hours of human effort per hour of video (frame rate of 10 Hz) annotated for a *single* animal. As such, the performance of this detection framework presents new opportunities in long-term animal

Table 3.3: Kolmogorov-Smirnov session comparison

			Speed		**Yaw**		**Yaw Rate**	
	Blk.		Δ_{ks}	α	Δ_{ks}	α	Δ_{ks}	α
OTS	1	2	0.187	< 0.001	0.028	< 0.001	0.047	< 0.001
	1	3	0.095	< 0.001	0.021	0.025	0.034	< 0.001
	1	4	0.080	< 0.001	0.019	0.049	0.057	< 0.001
	1	5	0.079	< 0.001	0.021	0.027	0.035	< 0.001
	2	3	0.096	< 0.001	0.028	< 0.001	0.017	0.099
	2	4	0.111	< 0.001	0.025	0.003	0.028	< 0.001
	2	5	0.110	< 0.001	0.023	0.012	0.016	0.148
	3	4	0.026	0.002	0.019	0.046	0.025	0.004
	3	5	0.026	0.003	0.022	0.012	0.010	0.685
	4	5	0.018	0.093	0.013	0.403	0.030	< 0.001
ITS	1	2	0.059	< 0.001	0.028	< 0.001	0.022	0.017
	1	3	0.021	0.019	0.020	0.039	0.008	0.871
	1	4	0.059	< 0.001	0.028	0.001	0.021	0.020
	2	3	0.061	< 0.001	0.023	0.009	0.028	< 0.001
	2	4	0.043	< 0.001	0.010	0.638	0.008	0.940
	3	4	0.068	< 0.001	0.029	< 0.001	0.028	< 0.001

			Speed σ		**Yaw σ**		**Yaw Rate σ**	
	Blk.		Δ_{ks}	α	Δ_{ks}	α	Δ_{ks}	α
OTS	1	2	0.047	< 0.001	0.035	< 0.001	0.076	< 0.001
	1	3	0.012	0.434	0.026	0.002	0.053	< 0.001
	1	4	0.025	0.004	0.029	< 0.001	0.062	< 0.001
	1	5	0.014	0.249	0.015	0.222	0.040	< 0.001
	2	3	0.047	< 0.001	0.031	< 0.001	0.033	< 0.001
	2	4	0.065	< 0.001	0.039	< 0.001	0.043	< 0.001
	2	5	0.051	< 0.001	0.048	< 0.001	0.043	< 0.001
	3	4	0.025	0.005	0.016	0.153	0.014	0.264
	3	5	0.008	0.889	0.026	0.002	0.026	0.002
	4	5	0.025	0.003	0.032	< 0.001	0.035	< 0.001
ITS	1	2	0.033	< 0.001	0.108	< 0.001	0.092	< 0.001
	1	3	0.027	0.001	0.012	0.423	0.016	0.139
	1	4	0.040	< 0.001	0.096	< 0.001	0.086	< 0.001
	2	3	0.046	< 0.001	0.103	< 0.001	0.100	< 0.001
	2	4	0.014	0.303	0.014	0.264	0.026	0.003
	3	4	0.056	< 0.001	0.093	< 0.001	0.095	< 0.001

Table 3.4: Speed and yaw joint differential entropy

	OTS					**ITS**			
Block	1	2	3	4	5	1	2	3	4
Entropy	2.358	2.599	2.543	2.508	2.541	2.521	2.675	2.584	2.605

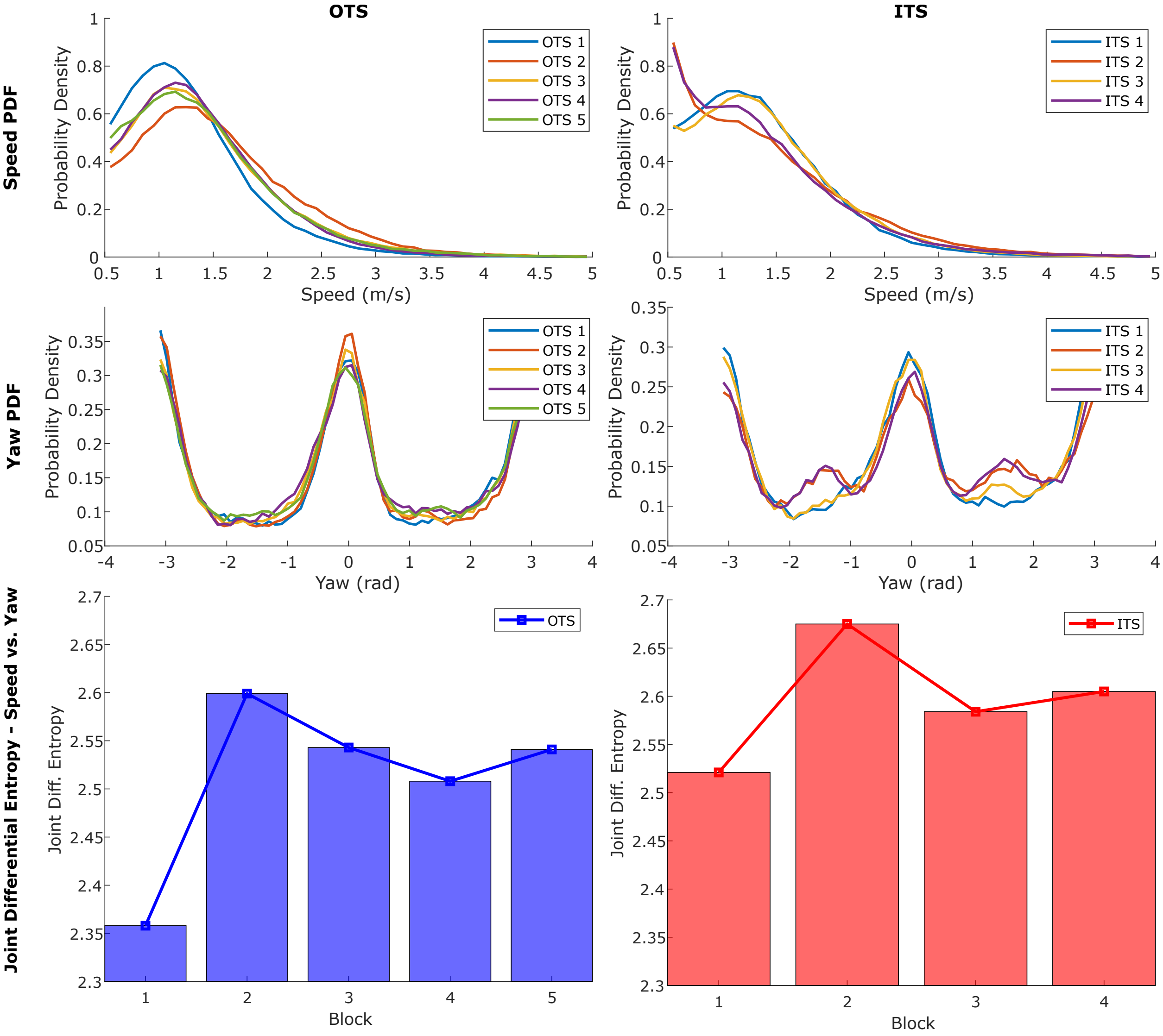


Figure 3.7: Speed and yaw probability distributions and joint differential entropies, respective to time block. TOP: Probability density functions of animal speed ($\mathrm{m\,s^{-1}}$) for OTS (left) and ITS (right). MIDDLE: Probability density functions of yaw (rad) for OTS (left) and ITS (right). BOTTOM: Joint differential entropy of speed and yaw for each block of OTS (left) and ITS (right), with limited-range y-axes to more clearly show value differences.

monitoring, and enables the automated processing of longer duration and more frequent recording sessions. For Section 3.1, the detection framework provided the large-scale animal position and dynamics data necessary to yield insights into animal kinematic behavior and spatial use.

3.1.4.2 Kinematic Diversity

Joint dynamic entropy was used to quantify differences in animal kinematic diversity throughout the day to explore how temporal changes in the dolphins' habitat would result in modified kinematic diversity levels (Fig. 3.7, bottom). While the use of entropy as a proxy for kinematic diversity has been applied in the past to characterize prey motion unpredictability for predator evasion, in this work it serves to provide a measure of animal engagement [111]. We observed the lowest kinematic diversity in the mornings as the ACSs were arriving at work and setting up for the day. The highest kinematic diversity when not interacting with animal care specialists then occurred immediately after the first ITS time block. In general, the first time blocks of both OTS and ITS showed the lowest kinematic diversity of their type, the second of each showed the highest, and the following blocks stabilized between the two extremes. The speed/quiver plots (Figs. 3.5-3.6, right) provide a qualitative understanding of the entropy results. For example, in Block 1 of OTS (Fig. 3.5, top-right) the dolphins engaged in slow swimming throughout their habitat in smooth consistent cycles along the environment edge, yielding the lowest joint entropy. Joint entropy then increased during both the morning ITS and OTS blocks and remained elevated for the rest of the day, representing higher animal engagement through the middle of their waking hours.

This is consistent with previous research on animal activity and sleep patterns, which reports a diurnal activity cycle for managed animals [51]. However, it is interesting to note that changes in animal kinematic diversity throughout the day during OTS are not gradual: the OTS time block displaying the minimum value is immediately followed by the block displaying the maximum, and are only separated by the first training session (30 minute duration). This sudden shift may not be fully explained by only the dolphins' diurnal activity cycle, and may be related to the fact that their first daily interactions with the ACSs occur between these two OTS time blocks.

3.1.4.3 Habitat Use

The kinematic data also enabled the investigation into how features in the habitat influenced animal behavior and spatial use, particularly during OTS. The animals tended to have a general focus on the area between the gates along the edge of the central island (Fig. 3.5, left). Additionally, throughout the OTS position plots (including static, Fig. 3.4, left) four animal-preferred locations were observed. The two hot spots to the left and right of the central island are gates (Fig. 3.1, middle, Fig. 3.2, top), where the dolphins could communicate with conspecifics when closed or pass through to other areas of their habitat when open. Conversely, the two hot spots nearer the middle of the island edge corresponded to underwater windows that led to an ACS work area (two central windows in Fig. 3.2, top/middle). Through these windows the dolphins may observe the ACSs, view conspecifics in one of the back habitats (through an additional window, not shown in Fig. 3.2), or observe enrichment occasionally placed on the other side of the glass (mirrors, videos, etc.). Regions of the habitat in proximity to these two windows experienced some of the highest occupancy in all OTS position plots, both static and dynamic. This indicates that particular attractors for the dolphins' attention were observable through those windows, whether they were the ACSs, conspecifics, or enrichment.

These attractors also influenced the dolphins' kinematics and activity levels. Of all the regions in the environment, only the positions in front of the central windows consistently recorded peak or near-peak location-specific animal swimming speeds for all OTS time blocks (Fig. 3.5, right). When combined with the results from the position distributions (Fig. 3.5, left), this implies that these dolphins not only focused their attention on these attractors, their presence correlated to higher activity levels in the dolphins when swimming in their vicinity.

3.1.4.4 Behavior Classification from Dynamics Metrics

During ITS blocks, ACSs asked for specific behaviors from the dolphins and these behaviors were often repeated. Elements of public educational presentations (ITS 2/4) were varied to include a mixture of both high and low energy segments, and this blend resulted in similar dynamic patterns for the public sessions. In contrast, the non-public animal husbandry and training sessions (ITS 1/3) were less dynamic overall, and yielded similar dynamic patterns for these sessions. Qualitative similarities in the pairs of animal training sessions were observable in both the position and speed/quiver plots in Figure 3.6, and the probability density functions presented in Figure 3.7.

The K-S statistics were used to quantify the similarities and differences between time blocks within both OTS and ITS. As the ACSs requested similar behaviors during ITS blocks of the same type, we expected similarities in the dynamics metrics for Blocks 1 vs. 3 and Blocks 2 vs. 4, and differences between the metrics for blocks of different types. The pattern displayed by the K-S statistics in Table 3.3 (particularly in the std. devs.) showed by far the most significant differences between time blocks of different types, and the fewest for blocks of the same type. Without prior knowledge of the block types, it would be possible to use this pattern to identify that Blocks 1 and 3 were likely the same type, as were 2 and 4. This demonstrated that the presented method of obtaining and analyzing dolphins' dynamics metrics was sufficient to differentiate between general behavior types.

This was useful for analyzing the OTS results, as the position and speed/quiver plots in Figure 3.5 only showed patterns in the animals' location preferences within their habitat. In contrast, the

K-S statistics gave a clearer view of the differences between OTS time blocks. Block 2 separated itself significantly from all other time blocks in nearly every metric, while Block 1 was in a similar position (though not as pronounced). Blocks 3-5 showed few significant differences for metrics comparisons between each other. This indicated that the dolphins had more distinct dynamics for Blocks 1 and 2, and maintained similar dynamics patterns throughout Blocks 3-5. When combined with the joint differential entropy values, these results indicated there may be three general OTS behavior types for the dolphins in this dataset (in terms of kinematic diversity [KD]): “Low KD” at the beginning of the day (Block 1), “High KD” immediately after the first training session (Block 2), and “Medium KD” for the remainder of the day (Blocks 3-5).

3.2 Persistent Biologging Tag Speed Sensing

The contribution in this section integrates a uni-directional turbine-based fluid speed sensor into a bio-logging tag and presents a systematic evaluation and calibration of the resulting system. The sensor configuration used in each experiment is defined in Section 3.2.1. In Section 3.2.2, the performance of the integrated system is evaluated in both steady and variable flow speeds using particle image velocimetry (PIV) to measure the local flow at the sensor and in the free stream [112, 113]. The characterized system was then used in Section 3.2.3 to make measurements of speed and total distance traveled from bottlenose dolphins in a managed environment during a controlled swimming task. Finally, Section 3.2.4 describes the uncertainty analysis that was performed to evaluate the inter-and-intra-sensor reliability of four custom made tag housings with integrated sensors. The inclusion of the sensor on a biologging tag is intended to provide a platform for energetics estimation (Chapter 4), and improve localization estimates of free-swimming animals (Chapter 5).

3.2.1 Configurations

The physical structure of the speed sensor is an ellipsoidal 5 mm diameter micro-turbine with 2 curved blades, connected to a fin through both ends of a wire passing through its central axis. The wire is secured to a plastic fin that is attached to a surface using adhesive. The turbine was originally developed by Coxmate Pty. Ltd. to measure the through water speed of racing shells, and is marketed to the public as a “Micro Impeller.” As with the racing shells, cetacean movement can be assumed to be in the direction of motion.

For two experiments in this chapter, the off-the-shelf ellipsoid/fin assembly was secured directly to an existing DTAG3 bio-logging tag (Configuration A, Fig. 3.8, bottom-left), and oriented such that the sensitive direction corresponds with the forward direction of the instrument (Fig. 3.8, top). The turbine contains a magnet, whose field strength is recorded by a magnetometer in the tag housing (Fig. 3.9). Moving fluid rotates the turbine, creating a magnetometer signal that fluctuates at the turbine rotation frequency. The frequency of the turbine signal is then correlated with the fluid speed in the vicinity of the sensor. Figure 3.8, top, also presents experimentally derived speed data from a bottlenose dolphin during a continuous swimming task (black). As the animal makes propulsive fluke strokes, the speed oscillates around an average value (u_{ave}) with a given frequency (f) and amplitude (a_{peak}) that are dependent on the animal's gait and speed. The ability of a uniaxial speed sensor to capture the full signal is dependent on the location of the tag and the amount of out-of-plane motion the tag will experience on the swimming animal.

A separate experiment used a custom housing (Configuration B) designed specifically to integrate the ellipsoidal speed sensor (Fig. 3.8, bottom-right). The new housing was 3D-printed on a Formlabs (Somerville, MA) Form 2 printer, using Durable resin. The housing measures $15.0 \times 7.4 \times 4.7$ cm. The micro-turbine was removed from its original plastic mounting fin, and directly inserted into the integrated mounting fin on the printed tag body via Nitinol wire. The use of the 3D printed body improved the repeatability of the manufacturing process and enhanced the structural strength of the exposed fin and rotating turbine, important considerations for an instrument used with large marine mammals. In the new design, turbine rotations are directly counted via a Panasonic AN48846B Hall-effect sensor, capable of detecting rotations of up to $\sim$ 1500 Hz, which are written to a data file at 0.2 second intervals.

3.2.2 Calibration Experiment - Flume

3.2.2.1 Method

Sensor Configuration A was first calibrated by measuring flow speeds and turbine rotational frequency in steady flow using a flow chamber. A cylinder was then placed in front of the tag to create oscillating flows to test the sensor's response to a dynamic input. For both experiments, the flow over the tag and turbine was visualized in a temperature controlled recirculating flume (Engineering Laboratory Design, Inc., freshwater, 20°C) using Particle Image Velocimetry (PIV) (Fig. 3.9). The test section of the flume was 1730 mm in the streamwise direction, 450 mm in the cross-stream direction, and 450 mm in depth. The tag, which is 175 mm long, 100 mm wide, and 38 mm tall, was mounted on the downward facing side of a flat plate, suspended horizontally in the flume test section. The plate spanned the full test section length and width, and its bottom surface was 230 mm above the bottom of the flume. The top of the tag was approximately 190 mm from the bottom of the flume. The leading edge of the plate was machined to a 10°knife edge to control boundary layer formation. The tag was centered in the cross-stream direction and the front end of the tag was 372 mm behind the leading edge of the flat plate (i.e., the inlet of the test section). Flat

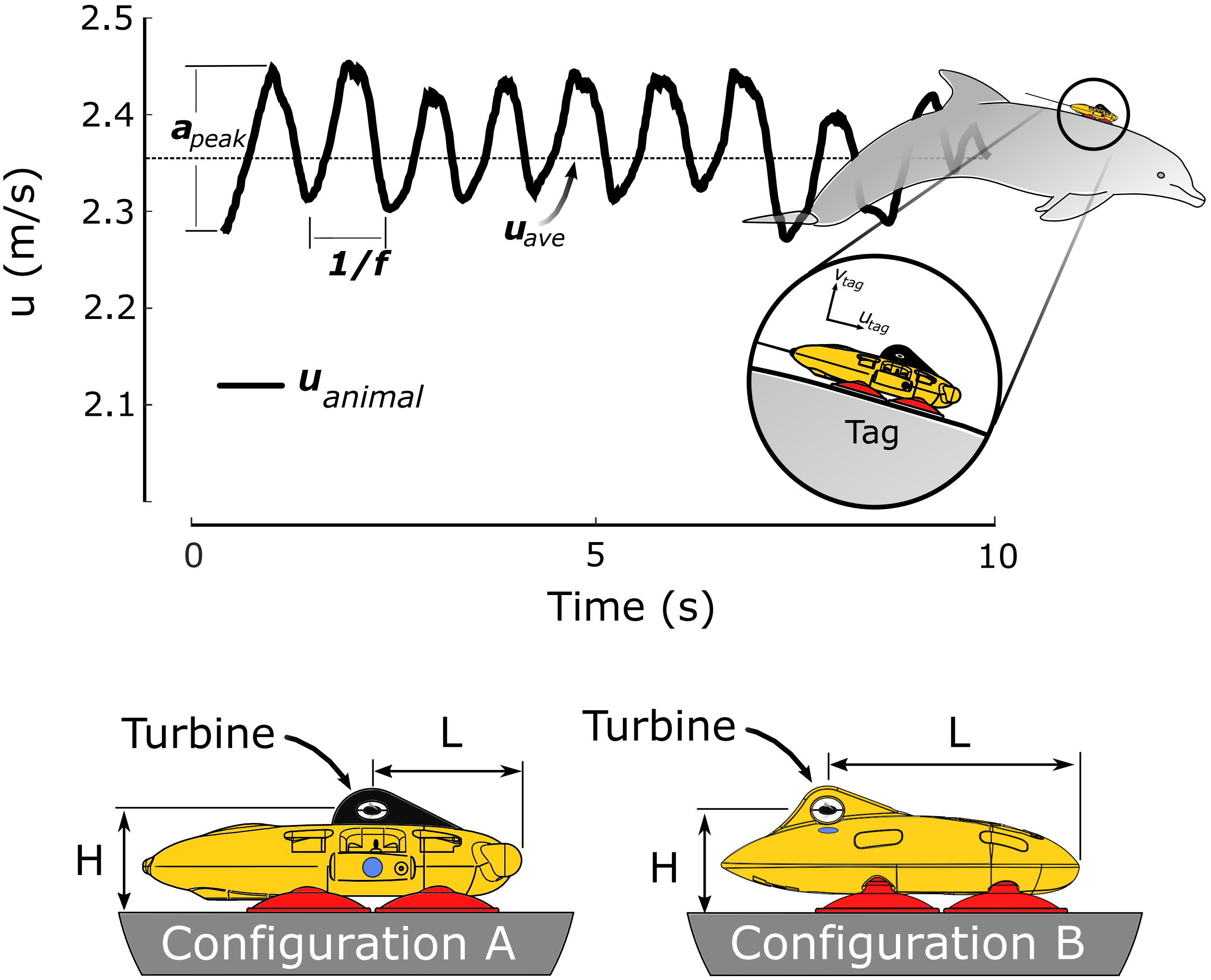


Figure 3.8: TOP: A representative 10s speed profile from a dolphin during a bout of swimming, calculated from experimental data. The black line corresponds to the dolphin's forward speed. The variable speed is characterized by an average speed (u_{ave}), a frequency of oscillation (f) about the average, and magnitude of excursion (a_{peak}) from the average. BOTTOM: Sensor configurations used in this research. Configuration A: 61 mm from nose to turbine (L), 42 mm from mounting surface to turbine (H). Configuration B: 105 mm from nose to turbine (L), 42 mm from mounting surface to turbine (H).

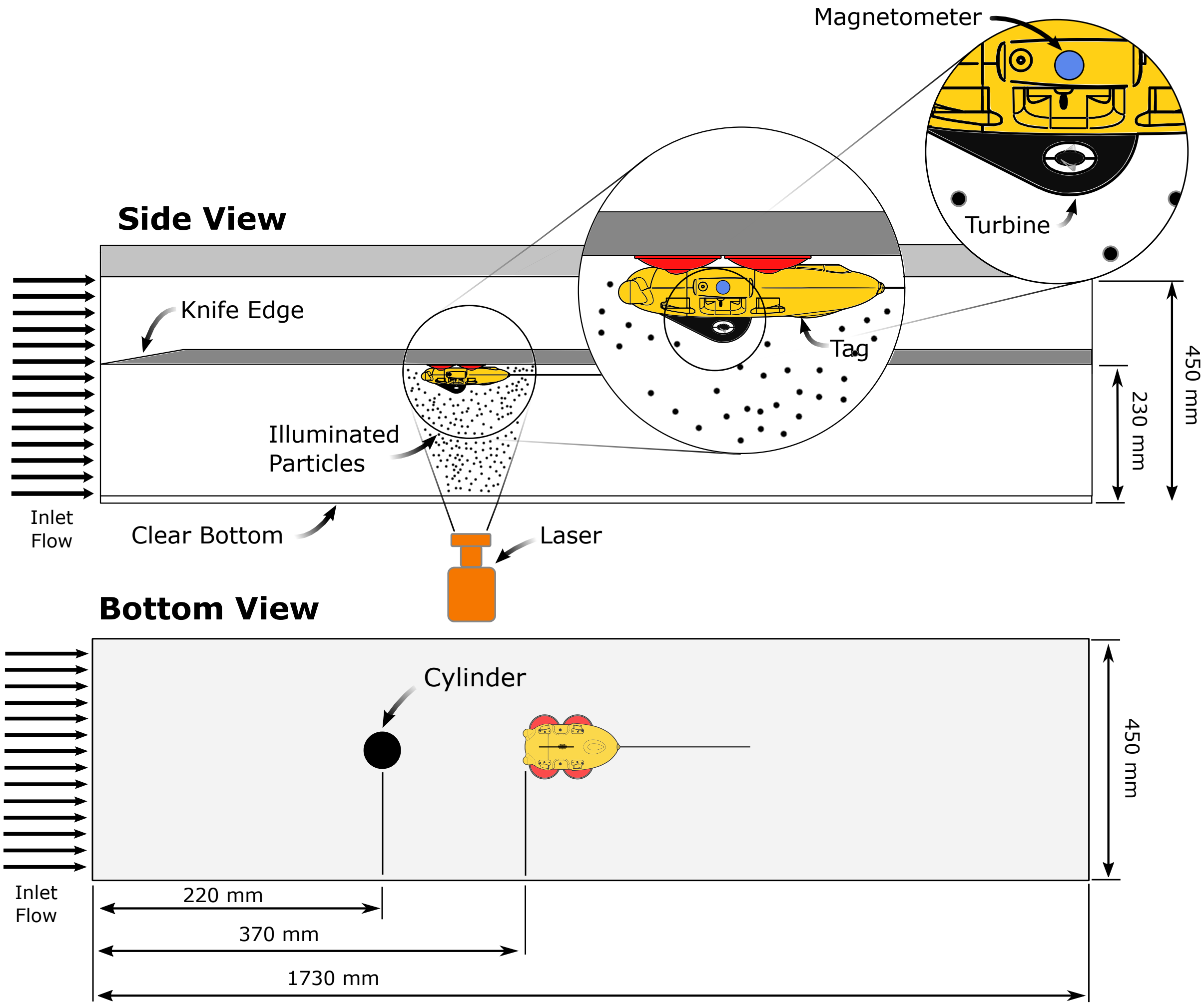


Figure 3.9: Experimental setup in the flume test section. Inlet flow speeds ranged from 0.28 – 1.1 m/s representing "length Reynolds numbers," Re_x, at the tag of $1 - 4 \times 10^5$. TOP: Side view of the steady flow case (i.e. no upstream cylinder — "unobstructed"). The tag is attached with suction cups to a large flat plate with a 10° "knife edge" at the leading edge to promote the development of roughly canonical flow over and idealized flat plate. A laser sheet was used to illuminate particles added to the flow. The particles were used to determine the fluid velocity field in a vertical plane along the centerline of the tag, which includes the turbine. BOTTOM: View of the oscillatory flow case from below the flume test section. Flow oscillations were generated using vortex shedding behind a cylinder placed upstream of the tag.

plate boundary layer theory, Blasius for laminar flow and 1/7th power profile for turbulent [114], predicts the boundary layer thickness on the plate to be in the range of 3 – 14 mm at the location of the tag for all of our experiments, taking into account the potential of the boundary layer flow being laminar or turbulent. Therefore, the turbine — spanning 43 to 52 mm above the flat plate — was not within the boundary layer of the flat plate. A rough estimate of the boundary layer on the tag body from the same flat plate theories is 1 – 4 mm. Since the turbine spans 5 – 14 mm above the tag surface, it is not expected to be in the tag boundary layer either. Most importantly, this fact was supported by the PIV data.

The turbine was subjected to a steady free-stream flow for 33 minutes, incrementally down-stepping flow for 16 minutes, and up-stepping flow for 17 minutes, where each step in flow speed was observed to reach steady-state before transitioning to subsequent steps. This portion of the experiment provided distinct, extended sections of consistent flow, at varying flow rates, used in the subsequent turbine rotation frequency/free-stream speed correlation analyses. Mean streamwise free-stream flow speeds, U, ranged from 0.28 to 1.1 m/s representing length (or local) Reynolds numbers, Re_x, of $1 - 4 \times 10^5$ at the front of the tag ($x = 372$ mm).

Next, the performance of the sensor Configuration A in variable flow was examined. To create an oscillating flow at the sensor, a vertical cylinder with a diameter, D, of 48 mm and a height of 200 mm was mounted to the plate approximately 150 mm ahead of the tag (Fig. 3.9, bottom). At the Re_D of these experiments ($2-5\times10^4$), it is well known that vortices are shed from the cylinder at a frequency, f, that can be predicted using the Strouhal number:

$$St = \frac{fD}{U} \tag{3.16}$$

Vortex shedding has been found to result in a relatively universal Strouhal number of about 0.20 over a wide range of Re_D including our range of Re_D. Therefore, for a given flow speed and cylinder diameter, the frequency of the fluid oscillations can be calculated, and will increase with free-stream speed. For the oscillatory flow test, the cylinder and turbine-tag arrangement was first exposed to a steady free-stream flow rate for 9 minutes. Next, the speed was stepped down over 12 minutes until the flow was too slow to turn the turbine. Then, the flow was stepped upwards over a period of 6 minutes. Ultimately, 7 distinct mean free-stream flow speeds, U, were tested, ranging from 0.43 to $1.1\,\mathrm{m\,s^{-1}}$. Note that speeds measured for the oscillatory flow trials refer to the flow speeds measured in the wake of the cylinder by PIV just upstream from, or "near", the turbine, u_{near}, and directly adjacent the turbine, u_{sens_nr}, and are computed using the calibration from the steady flow trials. Those flow speeds are lower ($0.3 - 0.7\ \mathrm{m\,s^{-1}}$) than the mean free-stream flow speeds, U, impinging on the cylinder and turbine-tag arrangement.

The set-up and parameters for PIV were as follows. 70 g of white nylon particles (Vestosint

1164, mean diameter 50 µm, density 1.06 $\mathrm{g\,cm^{-3}}$, Evonik Industries) were added to the flume (volume 5.74 $\mathrm{m^3}$) and illuminated them with a pulsed laser sheet 1-2 mm in thickness and wide enough to visualize a region of flow 160 mm in the streamwise direction. The laser (Firefly, Oxford Lasers, 808 nm, 0.2 – 30 mJ/pulse) has built-in optics to generate the laser sheet. It was mounted on a 3-axis motorized traverse under the transparent bottom of the flume and aimed upward toward the upside-down tag (Fig. 3.9, Side View). The sheet was oriented parallel to the streamwise direction and fell along the centerline of the tag. This allowed for the illumination of the flow in the plane of the turbine sensor and in the free stream.

The particles were imaged with a monochrome digital camera with a resolution of 1024 × 1024 8-bit pixels (Fastcam SA3, Photron, 50 mm AF Nikkor lens, Nikon) mounted perpendicular to the light sheet on a 2-axis motorized traverse next to the test section. The motorized traverses, constructed of linear slides (Techno-Isel) and controlled by software and hardware from National Instruments (LabVIEW, legacy FlexMotion toolbox and PCI-7344 motion controllers), allowed the camera and laser sheet to be precisely positioned and focused. They were also used to move the camera FOV up-and-downstream to create streamwise mosaics of the flow around the tag (Fig. 3.11, top). FOVs were 110 mm×110 mm for the PIV data used to calibrate and compare the turbine velocity readings, and 157 mm×157 mm for the mosaics. The mosaics were not used to calibrate the turbine or determine velocities in oscillatory flow experiments, they were simply used to visualize the time-averaged flow field around the entire tag. Image fields of view were auto-calibrated (DaVis 8.2, La Vision, Inc.) from pixels to millimeters using an image of a grid of 2.0 mm diameter dots on 1.00 cm centers.

The camera acquired images at 60 Hz, but the experiments were carried out in a dark laboratory and the laser pulses were timed to straddle the exposure periods of pairs of successive images, producing pairs of strobe images of particles in the flow separated by 0.4 – 2.0 msec (shorter for faster flow to restrict particle displacement in pixels to single digits for PIV algorithm accuracy). This results in a flow field visualization rate of 30 Hz. The velocity fields were calculated from the strobe pairs using a multi-pass 2D Fast-Fourier Transform (FFT) approach (DaVis 8.2, La Vision, Inc.), which essentially cross-correlates the particle patterns in corresponding subwindows of an image pair to find the mean particle displacement in each subwindow location. We performed one FFT pass for 64×64 pixel subwindows and 2 passes for 32×32 pixel subwindows, all with 50% overlap of one subwindow with another as the algorithm moved through an image pair. This resulted in velocity fields with a 64×64 grid of 2D vector components over the image FOV. Regions of the image showing the tag were masked from the algorithm.

The density of particles determined from the images was about 540 particles $\mathrm{cm^{-3}}$. This is roughly in keeping with the mass of particles added to the flume and distribution of particle diameter about the mean of 50 µm. More importantly, sample images from the experiment gave a mean of 12.7 particles per 32 × 32 subwindow with a standard deviation of 2.9. This degree of density and homogeneity of seeding resulted in very good velocity vector calculations. The uncertainties in the PIV velocities in the steady flow trials were only 0.9% – 1.9%. In the oscillatory flow cases — with the upstream cylinder —the uncertainties were higher (7.4% – 18.0%) as expected due to increased velocity gradients in the flow behind the cylinder causing greater deformation of the particle patterns in strobe image pairs. Nevertheless, time-averaged velocities and power spectra of the velocities determined by PIV and the turbine sensor in the cylinder experiments showed good agreement. The uncertainties in the PIV velocities were determined using LaVision's PIV software, DaVis 8.4. The software uses a method that compares the actual correlation of particle patterns in subwindow pairs to the ideal correlation that assumes no deformation of the particle pattern [115].

PIV image data was taken for 10 seconds at each flow speed used in the experiments. This resulted in 300 PIV velocity fields per flow speed. In 10 seconds, 23 – 100 separate volumes of fluid passed through the FOV given our range of oncoming flow speeds, U. The free-stream flow, U or u_{fs}, and flow near the turbine, u_{near}, were determined by sampling sub-regions of the PIV fields at roughly the starred positions on Figure 3.10, top. The sub-region for determining u_{fs} was a rectangle 40 mm long and 20 mm tall, 110 mm from the flat plate, which contained 276 velocity vector grid points. The sub-region for determining u_{near} was a rectangle 25 mm long and 10 mm tall, but rotated 22° so it could be set very close to the upstream edge of the angled turbine mount. It was positioned outside of the tag boundary layer and contained 85 grid points. A streamwise velocity vs. time at each of these positions was generated from the spatial average of streamwise velocity over the sub-regions in each of the 300 velocity fields in a 10 sec sequence. The time-averages of those velocity time series were used to determine overall average velocities as in Figure 3.10, middle-right. Standard deviations in velocity (u_{fs} and u_{near}) in these time series, in the steady flow cases, were only 0.01 – 0.03 $\mathrm{m\,s^{-1}}$ for means of 0.28 – 1.17 $\mathrm{m\,s^{-1}}$, respectively.

Mosaics of the velocity field around the tag were created at flow speeds of U = 0.36, 0.73, and 1.1 $\mathrm{m\,s^{-1}}$. The traverses carrying the camera and laser were moved to obtain four partially overlapping streamwise FOVs resulting in a velocity field 515 mm in the streamwise direction and 157 mm in the vertical. The field begins 179 mm upstream of the tag. At each FOV position and free-stream speed, U, the 300 velocity fields were averaged to produce a time-averaged velocity field. These were then knit together using position information from the motorized traverse, creating a new grid over the 515 mm×157 mm area, and using MATLAB's linear interpolation function, `interp2`. This was sufficient given that the mosaics were used primarily for a qualitative look at flow over the tag. Figure 3.10 (top) shows a portion of the mosaic for U = 1.1 $\mathrm{m\,s^{-1}}$.

The raw turbine magnetometer signal was sampled at 200 Hz. To calculate the frequency of the measured signal, a spline interpolant was used to locate zero-crossings in the sinusoidally

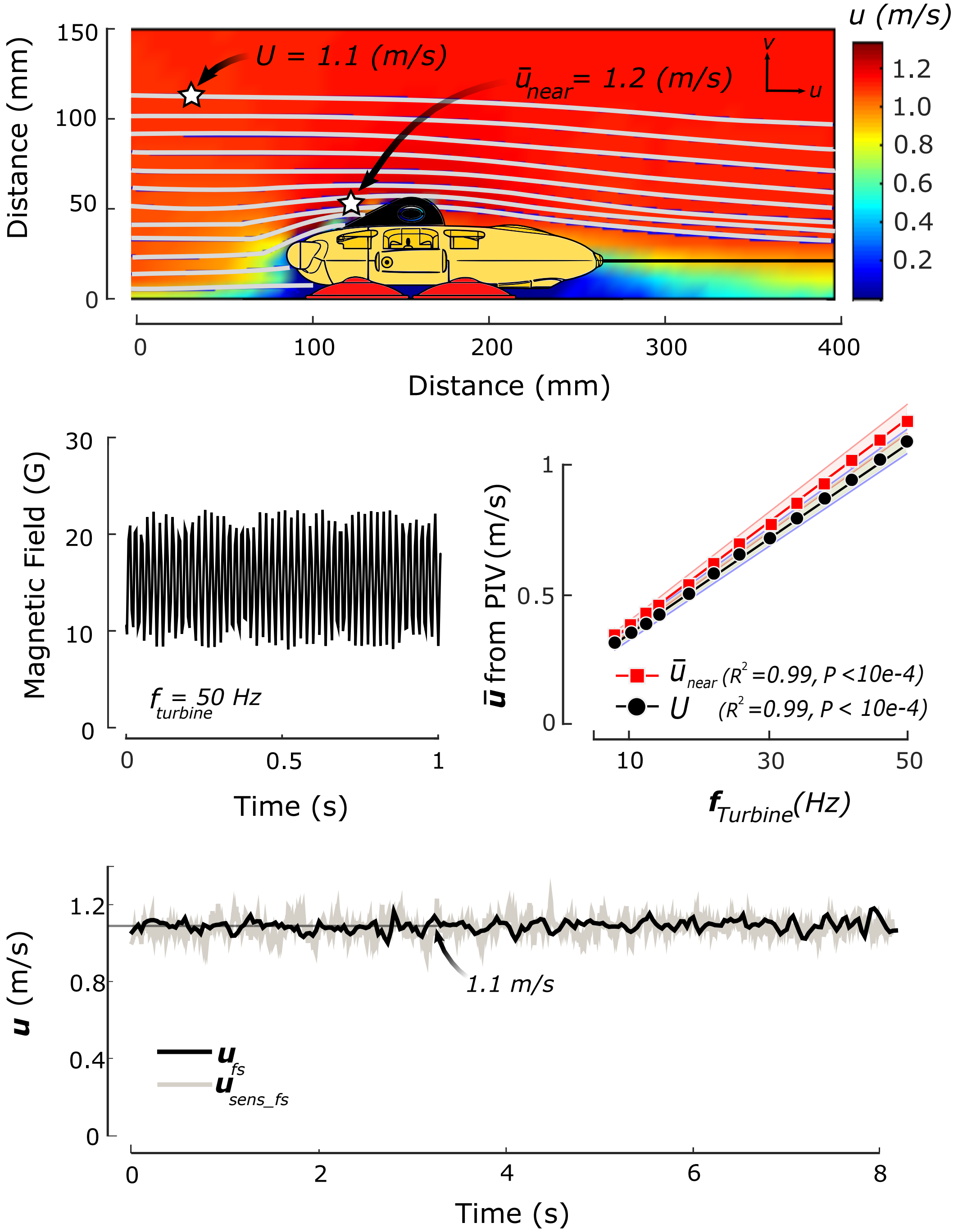


Figure 3.10: Results from the steady (i.e. unobstructed) flow tests. TOP: PIV calculated flow field during the fastest condition ($U = 1.1\ \mathrm{m\,s^{-1}}$). As expected, the shape of the tag increases the speed of the fluid as it moves over the front of the body at the sensor location, $\bar{u}_{near} = 1.2\ \mathrm{m\,s^{-1}}$. MIDDLE: Measurements of the changing magnetic field created by the spinning turbine made by the tag magnetometer during the $U = 1.1\ \mathrm{m\,s^{-1}}$ condition are shown on the left. The calibration generated during the experiment near the sensor($\bar{u}_{near}$, red squares), and the free-stream (U, black circles), along with the linear fits to the data, are shown on the right. The secondary lines represent the 95% confidence interval. BOTTOM: A comparison of the free-stream speed measurements during the $U = 1.1\ \mathrm{m\,s^{-1}}$ condition made using PIV ($u_{f\,s}$), and the turbine ($u_{sens\text{-}f\,s}$), using the turbine calibration from the mean free-stream flow (middle-right of this figure, black).

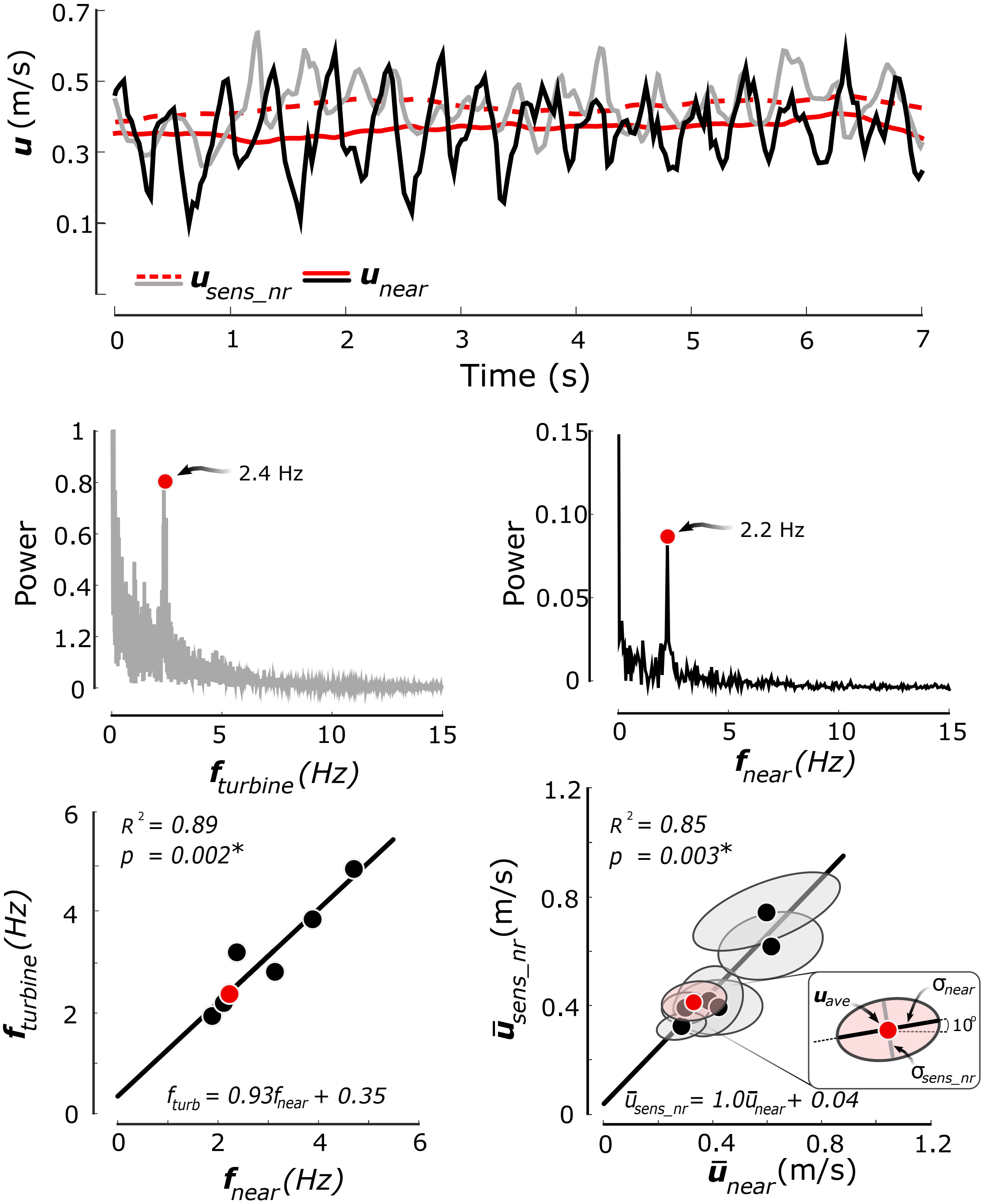


Figure 3.11: Results from the oscillatory flow tests. TOP: Streamwise flow speeds in the wake of a cylinder by the turbine (u_{sens_nr}, gray) and PIV (u_{near}, black) exhibit a similar high frequency oscillation with comparable low frequency speeds (red solid and dashed; free-stream flow, $U = 0.51\ \mathrm{m\,s^{-1}}$). The magnitudes do not track precisely, but a power spectrum of the two time traces shows that the oscillations do have essentially the same frequency (MIDDLE). BOTTOM: A comparison of dominant frequencies of oscillation from the power spectra (left) and the mean fluid speed (right) measured by the turbine vs. the measured PIV data near the sensor. Covariance ellipses for the velocity data are also shown.

fluctuating signal using built-in MATLAB spline interpolation functions. To ease computation requirements, these interpolation analyses were performed for 5 second intervals of the data, and the times of the zero crossings of each interpolated interval were extracted and assembled in a vector. Each zero-crossing represents the starting point of a one-half period of rotation of the turbine; the rotation periods of the turbine were therefore obtained by determining the time differences between every other zero-crossing. These period durations were calculated in seconds, and inverted to give the turbine's instantaneous frequency, $f_{turbine}$, in Hertz. The running standard deviation of the frequency for each test was calculated using a sliding centered window of 801 samples. The steady-state standard deviation in turbine frequency was calculated for each flow rate by manually determining the steady-state regions in each test and taking the standard deviation. These steady-state regions were determined by observing where the flow would settle for each test, and heuristically picking regions within these bounds.

Linear regressions between PIV-generated mean flow speeds and the corresponding mean $f_{turbine}$ were calculated to display the correlation between the $f_{turbine}$ and flow speeds at two locations: the free stream and near the sensor. In this section, instantaneous streamwise free-stream flow speed is represented as u_{fs}, mean streamwise free-stream flow speed as U (i.e., the time average of u_{fs}), instantaneous streamwise flow near the sensor as u_{near}, and the time average of the latter as $\bar{u}_{near}$. To clarify, u_{fs}, U, u_{near}, and $\bar{u}_{near}$, all correspond to to PIV-measured flow speeds. Each 95% Confidence Interval (CI) of each regression are shown as secondary lines of lower intensity. P-values were calculated to assess statistical significances. u_{sens_fs} is used to represent the instantaneous flow *as estimated by the turbine*, when the turbine is calibrated using the PIV-measured free-stream steady flow. Similarly, u_{sens_nr} is used to represent the turbine-estimated flow when the turbine is calibrated using the PIV-measured steady flow near the sensor.

3.2.2.2 Results

PIV data were used to reconstruct flow field data for the recirculating flume experimental conditions. Figure 3.10, top, presents a representative flow field during the $U = 1.1\ \mathrm{m\,s^{-1}}$ steady flow condition, and illustrates the increased fluid velocity, $\bar{u}_{near} = 1.2\ \mathrm{m\,s^{-1}}$, due to the acceleration of flow over the tag geometry. Figure 3.10, middle-left, shows raw data recorded by the magnetometer during the $U = 1.1\ \mathrm{m\,s^{-1}}$ trial that demonstrates the sinusoidal signal created by the spinning magnetic bead. Calibration curves relating $f_{turbine}$ to fluid speeds u_{fs} and u_{near} were calculated using mean free-stream flow ($U = \bar{u}_{fs}$, black) and mean flow near the sensor ($\bar{u}_{near}$, red), respectively (Fig. 3.10, middle-right, and Table 3.5). Figure 3.10, bottom, shows a comparison of the time series of u_{fs}, and the speeds predicted by the sensor calibrated by the mean free-stream flow, u_{sens_fs}. Not surprisingly, the time averages of both u_{fs} and u_{sens_fs} are close to $1.1\ \mathrm{m\,s^{-1}}$ due to the calibration, and as the flow is being sampled at two different locations, velocity fluctuations in

Table 3.5: Correlation analysis of uniform and oscillatory flow tests

Fit	Coefficient	Offset	R^2	P
Free-stream Speed (U vs. $f_{turbine}$)	0.019	0.17	0.99	$< 10^{-4}$
Flow Speed at Sensor ($\bar{u}_{near}$ vs. $f_{turbine}$)	0.020	0.19	0.99	$< 10^{-4}$
Frequency ($f_{turbine}$ vs. f_{near})	0.93	0.35	0.89	0.002
Flow Speed (behind cylinder, $\bar{u}_{sens\_nr}$ vs. $\bar{u}_{near}$)	1.03	0.04	0.85	0.003

u_{fs} and u_{sens_fs} exhibit no obvious correlation. In general, it was found that the turbine required a minimum flow speed of $U \approx 0.25$ m s^{-1} to turn; data captured below this limit were unreliable.

During the oscillating flow trials, the calibrated turbine captured the mean local velocity, $\bar{u}_{near}$ (red line, Fig. 3.10, middle-right), and frequency of fluid oscillation (f_{near}) well, but did not track the large changes in amplitude precisely. Representative data (Fig. 3.11, top) from a trial in which $U = 0.51$ m s^{-1} illustrates this (note that mean streamwise flows in the wake of the cylinder, $\bar{u}_{near}$ and $\bar{u}_{sens\_nr}$, were about 0.3 – 0.4 m s^{-1}). Sliding-window means of u_{near} and u_{sens_nr} (red solid and dashed lines, respectively) are comparable, but amplitude tracking (gray and black lines) is poor. Sliding-window means were computed using Savitzky-Golay smoothing, with windows of four times the respective update rate for each sensing method (i.e. windows of 800 points for the turbine, and 120 points for the PIV). This results in a smoothed mean computed over a sliding four-second window for each data set. The mismatch of amplitudes was particularly noticeable at low flow speeds, although the turbine was able to capture the frequency content of the measured PIV signal (Fig. 3.11, middle). For PIV validation purposes, in the $U = 0.51$ m s^{-1} case, flow fluctuation frequency was measured to be ~ 2.3 Hz, yielding a Strouhal number (Eq. 3.16) of 0.22, close to the value expected ($St \approx 0.20$) in vortex shedding from a cylinder.

For the seven free-stream speeds tested in the oscillating flow trials ($U = 0.43 - 1.1$ m s^{-1}), there was good linear correlation between the frequencies of oscillation ($f_{turbine}$ vs. f_{near}), and between the mean speeds in the wake of the cylinder ($\bar{u}_{sens\_nr}$ vs. $\bar{u}_{near}$, Fig. 3.11, bottom-left, and Table 3.5). The covariance ellipses plotted with the mean velocity data illustrate the ability of the sensor to track the magnitude of oscillation at the different flow speeds (Fig. 3.11, bottom-right). The longer major axes of the covariance ellipses occurring at higher U graphically illustrate a decrease in turbine tracking performance as the magnitude and frequency of the flow oscillations increased.

3.2.3 Validation Experiment

3.2.3.1 Method

To demonstrate the efficacy of sensor Configuration A, controlled swimming trials were conducted with an institutionally-managed bottlenose dolphin, at Dolphin Quest Oahu (Oahu HI). A new turbine (same model, separate unit) was attached to a DTAG3 unit with a higher sampling rate (625 Hz) than the original used in the flow chamber (200 Hz), allowing for higher detectable rotation frequencies of the turbine. The tag and turbine were secured to the animal using an array of the same minimally invasive suction cups used to secure the tag in the flume. For experimental verification, four dolphins were outfitted with the turbine-tag module, over a total of six sessions. During the trials, the dolphins performed a straight-line swimming task between set points at a self-selected pace. These six sessions produced a total of 41 recorded sections of animal travel, of lengths varying between 5.9 and 76.3 meters. All experimental protocols were approved by the Institutional Animal Care and Use Committee at the University of Michigan (Protocol #PRO00006909).

To verify the measurements made with the tag, the dolphin's position was also hand-tracked using an overhead camera setup. Tracking was performed manually rather than using the method described in Section 3.1 for the sake of expediency, as few videos had to be processed and an entirely new neural network would have been required given the separate experimental environment. The overhead video recordings were made using a GoPro HERO5 camera, which was mounted on the edge of the first-floor balcony of the building overlooking the lagoon. Videos of the sessions were recorded at a resolution of 1920×1440, at NTSC 60 Hz, and downsampled in post to 10 Hz to expedite the manual tracking process. The OCamCalib omnidirectional camera calibration toolbox was used to find the fisheye distortion model for the camera, by passing 128 distinct checkerboard images through the toolbox [116].

Specific locations on the perimeter of the lagoon were mapped, through tape measurement and triangulation, with respect to a common origin. These served as the world-frame perimeter points. Their corresponding pixel locations in the image frame were undistorted with the computed model, and used with the world-frame perimeter points to compute the homography transform that maps undistorted pixel locations to world-frame locations. Through this, the undistortion model and the homography transform make it possible to convert pixel locations in the original video to world-frame locations on the surface of the lagoon. This assumption is used due to the structure of the lagoon, as the testing region was shallow (< 3 meters) enough to ensure the dolphins' movements were primarily planar.

To examine the performance of sensor Configuration A with respect to the established ground-truth of the camera, the distance traveled estimates of both were compared. The camera estimate of distance traveled was computed by directly summing the distances between camera-detected dol-

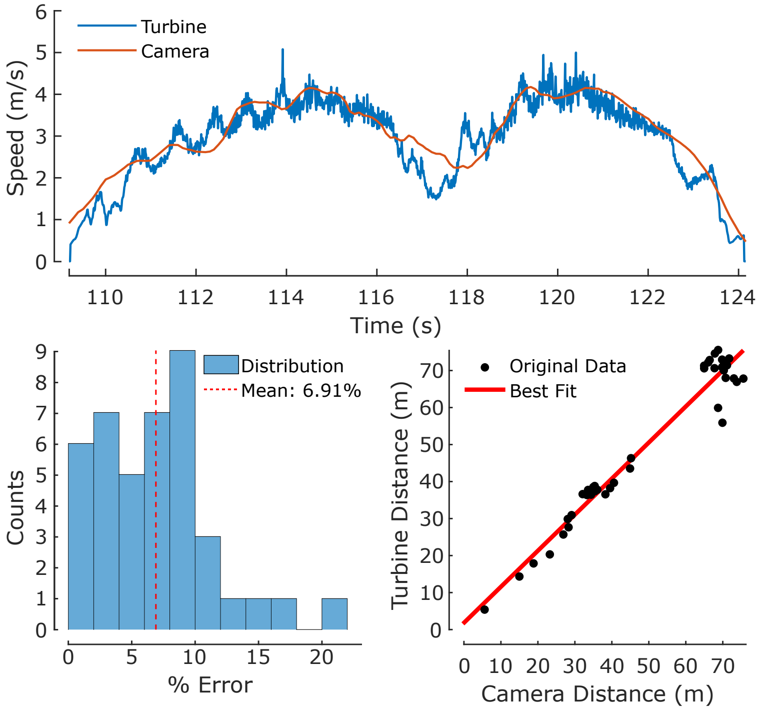


Figure 3.12: Results of the dolphin swimming experiment. TOP: Sample speed estimates for one trial lap, comparing turbine (blue) to camera (red) data. BOTTOM: The histogram of % error in distance traveled of the turbine estimate (assuming camera data represents the ground-truth) is shown on the left. The correlation between camera and turbine distance traveled estimates is shown on the right.

phin positions, while the turbine estimate was computed through a trapezoidal discrete integration of the measured speed. Both full-session and individual-lap distances were compared, to examine drift over time as well as short-term accuracy. To reduce the differentiation noise in the camera speed estimate (generated for visualization), the camera speeds were smoothed with a 1-second span moving average, while the turbine speed estimate was left unfiltered.

3.2.3.2 Results

A sample section of animal speed from the validation experiments, derived from camera distance data, is plotted alongside the corresponding turbine measurement with PIV calibration, Figure 3.12, top. For the experimental verification, the percentage errors in distance traveled for each of the 41 trials were computed using the final distance traveled values for each recording section, and resulted in a final mean percentage error of 6.91%, with a standard deviation of 4.67%. The full distribution of percentage error can be seen in Figure 3.12, bottom-left. A linear regression comparing the turbine and camera distance estimates was also used to examine the performance of the tag-based sensor, Figure 3.12, bottom-right. This resulted in a regression coefficient of 0.97, and an offset of 1.01 meters. The R^2 coefficient of the regression was found to be 0.95.

3.2.4 Calibration Experiment - Basin

3.2.4.1 Method

A final set of experiments to assess inter-and-intra-sensor variation of the speed sensor was conducted in the physical modeling basin of the Marine Hydrodynamics Laboratory at the University of Michigan. These experiments were conducted with sensor Configuration B, and employed the range of speeds observed during the animal trials (1-4 $\mathrm{m\,s^{-1}}$). To perform this experiment, the tags were mounted to a metal plate that was pulled through a large stationary body of water ($109.7 \times 6.7 \times \sim 3.1$ meters) by a moving carriage. The plate, dimensions $18 \times 24 \times 0.5$ inches, was machined from 6061-T651 alloy aluminum, with a 0.25 inch radius rounded front and 7° double-sided chamfered rear edge. This was vertically mounted to another 6061-T651 aluminum plate, which served as the interface between the tag mounting plate and a steel hydrodynamic strut attached to the moving carriage. The carriage was then returned to its starting position at a speed of 1 $\mathrm{m\,s^{-1}}$, and held stationary for a minimum of 30 seconds before the next run was initialized. The carriage contained a set of encoders that directly output the carriage's current position, updated at a rate of 2000 Hz.

Two tags at a time with sensor Configuration B were attached to the mounting plate via suction cups, one tag centered on either side of the plate, and were aligned with the direction of flow by ensuring the tag orientations were parallel to the edge of the plate via a level. The steel strut was lowered sufficiently into the water to submerge the tags by approx. 1 meter, to avoid free-surface effects. Four tags with sensor Configuration B were tested in two separate sessions. In each experimental session, the tags were subjected to four separate speeds (1, 2, 3, 4 m/s), with six runs per speed, in a randomized order. The maximum of 4 $\mathrm{m\,s^{-1}}$ was chosen as the water basin was not long enough to support steady-state run speeds faster than 4 $\mathrm{m\,s^{-1}}$ for longer than a few seconds. To perform a run, the carriage was accelerated to the target speed over a period of 10 seconds, held at the steady-state speed for as long as possible, and decelerated to 0 m/s over a period of 10 seconds. This resulted in usable steady-state durations ranging from 64 seconds for 1 $\mathrm{m\,s^{-1}}$, down to as short as 5 seconds for 4 $\mathrm{m\,s^{-1}}$.

The measurements made using sensor Configuration B were compared using, 1) the linearity of the sensors' responses, and 2) analysis of inter-and-intra-sensor variation. For the linearity analysis, the turbine spin data and carriage position data were directly extracted from each steady-state region of each run. The carriage position data were differentiated to calculate an estimate of speed, and smoothed with a 1-second span moving average to remove high-frequency noise. The turbine spin data were left unfiltered. The mean spin rates of the turbine and the speeds of the carriage were then extracted from the steady-state of each run and compared via a linear regression. In addition to the linear fit coefficients and offsets, the R^2, P-values, and the 95%

CIs of the linear fits were computed. To numerically represent a CI of a linear fit, the differences between the best-fit line and 95% CI bound coefficients and offsets were computed. For example: for Turbine 1, the 95% CI of the linear fit is bounded below by the line $u_{T1_-} = 0.026f - 0.131$, and above by $u_{T1_+} = 0.028f + 0.168$. Thus for Turbine 1, this yields a $\Delta_{coef} = \pm 1.48 \times 10^{-3}$ and $\Delta_{offset} = \pm 0.150$ for a 95% CI.

The inter-and-intra-sensor variation analysis was performed by computing the Coefficients of Variation (CV) of the turbine spin rates at the four separate speeds between sensors and between runs of a particular sensor, respectively. A one-way analysis of covariance of the regressions was computed with an $\alpha = 0.05$, and used in a multiple comparison of coefficients to determine if the turbines' individual linear regressions were significantly different. The linear regression results also illustrate the differences in responses each sensor has to similar flow profiles.

3.2.4.2 Results

During the trials to examine the uncertainty of measurements made using sensor Configuration B, the carriage speed during steady-state was extremely stable, with an overall standard deviation from the target speeds of only 4.28×10^{-4} $\mathrm{m\,s^{-1}}$. As a result, carriage speed variation was not considered in the analysis. The results presented in Table 3.6 indicate that the response of the four tags with sensor Configuration B were all linear, with R^2 values over 0.98, and excepting Turbine 1, very small 95% CI bounds (Fig. 3.13, top). Figure 3.13, bottom, presents data from experimental trials measured using Turbine 3, and illustrates a representative example of the ability of a sensor to repeatedly track the speed of the carriage. In general, there was variability between the responses of the four tags, particularly at the higher speeds. As speeds increase, each data set remains clustered closely to itself, but separates from the other data sets (Fig. 3.13, top). This result is expected after comparing coefficients of the individual regressions in Table 3.6, where in an extreme case, Turbine 4 is 22% more sensitive (i.e. higher rotations per $\mathrm{m\,s^{-1}}$) than Turbine 1.

The CV results also show that the performance discrepancy between individual tags with sensor Configuration B increases at higher speeds. Furthermore, the mean CVs for each tag, all in the range of $\sim 1\%$, are a fraction of the mean for the combined dataset, at 5.31%. Comparing the linear regression values to those of the calibration in Configuration A (coefficient=0.019, offset=0.17), we see the individual and combined regressions all have similarly low line offsets, but also all use larger coefficients in their linear regressions. All individual linear regressions of sensors in Configuration B were determined to be statistically different from each other, as the multiple comparison of coefficients yielded p-values $< 10^{-2}$ for all comparisons.

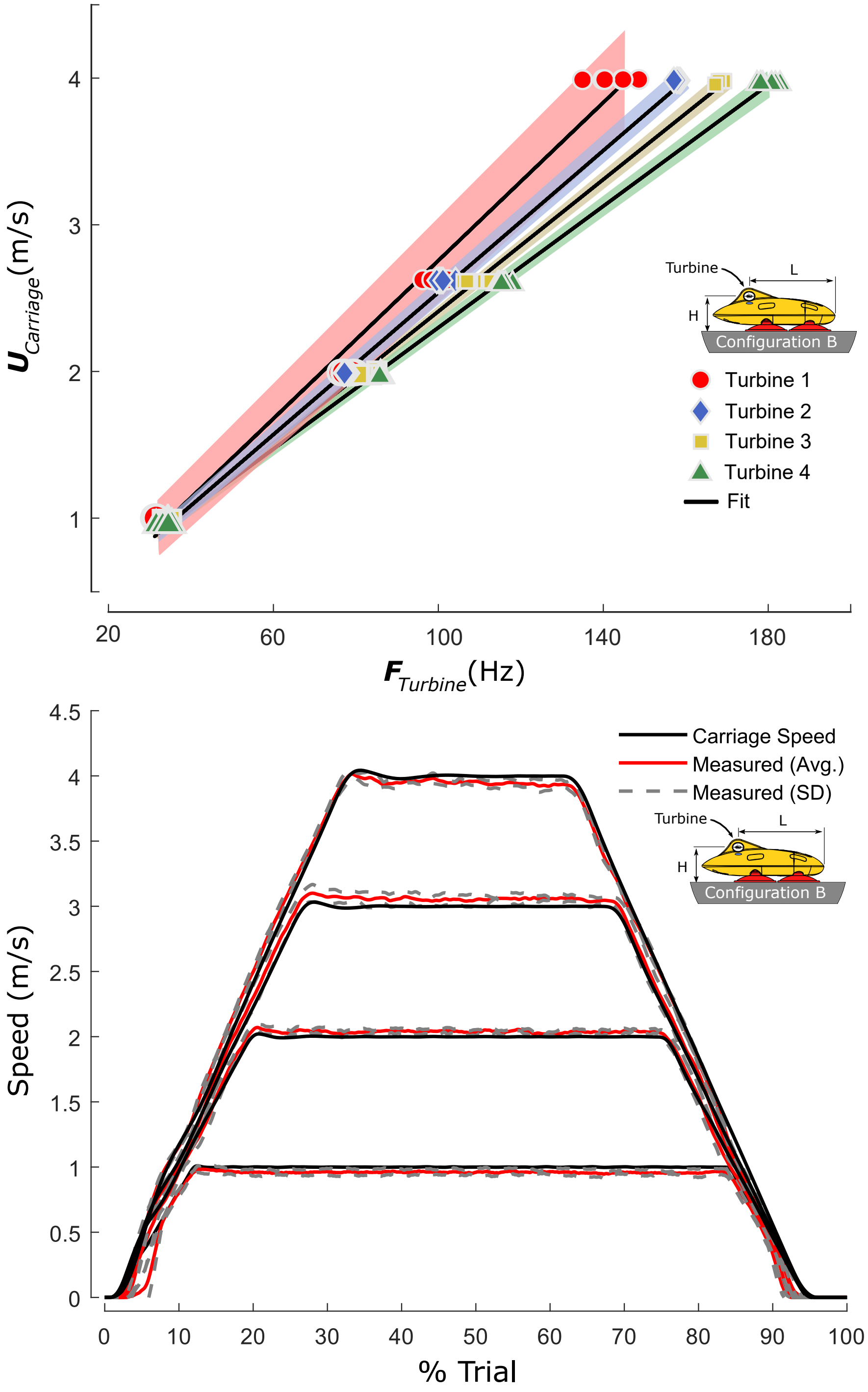


Figure 3.13: Results from the uncertainty analysis experiment. TOP: Linear regression results for all four turbines, with best-fit lines (black) of the raw data (shape markers) flanked by 95% confidence interval bounds (shaded regions). The data markers and confidence interval shadings have colors corresponding to their respective turbines. BOTTOM: Comparison of carriage speed (black) versus speed measured by Turbine 3 (red), flanked by 1 standard deviation bounds of the measured values (gray dashed), visualized according to % of trial. The measured speed results were collated and averaged by speed section (1 – 4 $\mathrm{m\,s^{-1}}$), and smoothed with a 2%-span moving average for visibility. Turbine 3's specific calibration was used to compute the measured values.

Table 3.6: Uncertainty analysis results

Parameter	T1	T2	T3	T4	Comb.
Line Coef.	0.027	0.022	0.024	0.021	0.023
Line Offset	0.018	0.210	0.158	0.259	0.228
R^2	0.985	0.999	0.998	0.999	0.971
P-value	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$	$< 10^{-4}$
CI Δ_{coef}	1.48E-3	2.70E-4	4.83E-4	3.17E-4	8.07E-4
CI Δ_{offset}	0.150	0.031	0.052	0.038	0.089
CV% - 1 m/s	0.306	1.306	2.619	2.011	2.317
CV% - 2 m/s	1.085	1.275	0.540	0.391	4.719
CV% - 3 m/s	1.253	1.344	1.000	0.492	5.368
CV% - 4 m/s	2.475	0.318	0.255	0.905	8.819
CV% - Mean	1.280	1.061	1.103	0.950	5.306

3.2.5 Sensor Calibration General Discussion

In the steady (unobstructed) flow, the relationship between sensor Configuration A measurements and the fluid speeds from PIV were linear and highly correlated ($R^2 = 0.99$, P$< 10^{-4}$). The flow speed near the sensor was affected by the tag geometry, but this effect can be corrected by calibrating the sensor against the free-stream PIV data. This enables “corrected” speed measurements of the animal as long as the tag is placed in a location where the flow conditions are similar to this experiment (i.e. where boundary layer thickness on the animal is less than the sensor height).

The oscillatory flow trials demonstrated that sensor Configuration A was able to measure mean speed and frequency of oscillation of the local flow, but it did not track the amplitude of the disturbed flow precisely. This was particularly true when the speed of the oscillating flow dropped below the turbine stall speed ($U \approx 0.25$ m s^{-1}). While dolphin swimming speed can vary widely, the oscillating speed created by animal fluking should be well above the stall velocity for the sensor during routine animal swimming speeds. Nominal dolphin swimming speeds have been reported to range from 1.6 – 5.6 m s^{-1}, with minimum transport costs observed at swimming speeds of 2.1 – 2.5 m s^{-1} [37, 105, 117, 118]. These results from the literature, along with our own experimental results, indicate that the expected oscillatory swimming speeds will be at magnitudes measurable by the proposed sensor system.

Importantly, we were able to verify the performance of sensor Configuration A with free swimming animals. While the work in the flow chamber was conducted at speeds of around 1 m s^{-1}, the experimental verification of the sensor with free-swimming dolphins successfully demonstrated that the our PIV-based calibration can be extended for use in the field. As an example, Figure 3.12

indicates swimming speeds agree well with the measurements calculated from the camera data, and measured speeds 5 times higher than those we were able to create in the flume. In addition to the qualitative comparisons, the accuracy of the sensor was demonstrated using the turbine-camera distance estimate comparisons, with a mean percentage error of less than 7% for the 41 trials.

Complementing the results from the first two experiments, the uncertainty analysis performed using sensor Configuration B indicates that the response of sensor Configuration B to flow speeds in our range of interest is linear, but that there is variability between the individual tags. This could be due to variability in the manufacturing of the turbine and tag body, or variability between tags at the housing/impeller interface.

3.3 Conclusion

The contributions in this chapter represent the tools built for marine animal monitoring that enable the development of further work in this area. Neural network object detection provides new opportunities in managed animal tracking by simultaneously reducing localization error, increasing repeatability, and simplifying data processing for extended monitoring sessions. The capabilities of the method are further explored through the analysis of animal intra-day kinematics trends and habitat use, demonstrating the flexibility of the approach in monitoring animal behaviors throughout large datasets. As such, the localization method detailed in Section 3.1 provides a persistent, automated approach for estimating animal location in managed settings with accuracy and scalability not feasible with established tracking methods.

The speed sensor implemented, calibrated, and validated in Section 3.2 has large ramifications for biologging-tag animal monitoring. It displayed low noise, a highly linear response to a wide range of fluid speeds, and demonstrated its viability through real-world on-animal measurement tasks. The accuracy, reliability, and size limitations of previously developed speed estimation methods (dedicated hardware or otherwise) has thus far limited their performance or inclusion on animal-borne tags. This sensor can further the capabilities of these tags by providing robust and predictable speed measurements, and its straightforward physical integration and miniaturized profile simplify its addition to a device platform.

CHAPTER 4

Persistent Marine Mammal Energetics: A Physics-Based Approach

This contribution investigates bottlenose dolphin swimming biomechanics using tag-measured swimming kinematics, through the extension of a physics-based hydrodynamic model to estimate propulsive power, work, and COT during locomotion. Tag data enabled the use of information from thousands of fluke strokes to estimate mechanical work and power during both steady-state and transient swimming behaviors. Prescribed straight-line swimming was used to characterize steady-state performance of six animals over a range of swimming speeds, and a 24-hour period of data during an animal's daily life was used to quantify preferred swimming biomechanics. The contributions of this chapter include: 1) the experimental design used to collect data for the framework from animals in human care (Section 4.1); 2) experimental investigation of mechanical work and power during high-effort steady-state swimming (Section 4.3); and 3) the quantification of the day-scale biomechanics of a free-swimming animal in a managed environment (Section 4.4). Importantly, this chapter demonstrates the viability of the approach for analyzing swimming mechanics during an animal's daily life by estimating its energy budget during a 24-hour period, an important step for the use of this approach with animals in the wild.

4.1 Energetics Monitoring Framework

4.1.1 Tag Hardware

MTags (Movement Tags) are persistently-monitoring biologging tags with internal electronics built on the OpenTag3 platform with the addition of integrated speed sensing. These are effectively

---

Content of this chapter also to be submitted as:

J. Gabaldon, D. Zhang, J. Rocho-Levine, M. Moore, J. van der Hoop, K. Barton, and K. A. Shorter, "Tag-based estimates of bottlenose dolphin swimming biomechanics and energetics," in *The Journal of Experimental Biology*. Copyright may be transferred without notice, after which this version may no longer be accessible.

a refined iteration of Tag Configuration B in Section 3.2.1. Kinematic sensors include: 3-axis accelerometer, 3-axis magnetometer, 3-axis gyroscope, temperature sensor, and ambient pressure sensor. Speed through water was measured using a secondary circuit board with a 1-axis Hall-effect sensor, and a free-spinning uniaxial magnetic micro-turbine mounted in line with the tag fin (Fig. 4.1, popout). Rotations created by water moving past the turbine were recorded by the Hall-effect sensor. Forward velocity (v_{tag}) was obtained using methods outlined Section 3.2.4. The accelerometer, magnetometer, and gyroscope were sampled at 50Hz, and the remaining sensors at 5Hz. The MTag electronics were powered by a 1100 mAh lithium-ion battery, enough to record continuously for ~3 days. Four silicone suction cups were used to secure each tag to the animal (Fig. 4.1). Three MTag units were used for the 2018 experimental sessions, and three different units were used for the 2019 sessions.

4.1.2 Power Estimation

In this section it is assumed that the tagged animal can control buoyancy to balance the gravitational force acting on the body, so the remaining forces acting on the animal are combined into net thrust and net drag (Fig. 4.1) that are related to the animal's motion by:

$$m^{\dagger} a_{com} = F_{thrust} + F_{drag} \tag{4.1}$$

where $m^{\dagger}$ is the total effective mass and a_{com} is the Center of Mass (COM) acceleration of the animal. Power (P) is related to force (F) and velocity (v) through $P = F \cdot v$:

$$P_{com} = m^{\dagger} a_{com} \cdot v_{com} = F_{thrust} \cdot v_{com} + F_{drag} \cdot v_{com} \tag{4.2}$$

The power generated by the animal during swimming locomotion was then defined as follows:

$$P_{thrust} = \underbrace{m^{\dagger} a_{com} \cdot v_{com}}_{\text{Inertial}} - \underbrace{F_{drag} \cdot v_{com}}_{\text{Drag}} \tag{4.3}$$

where $m^{\dagger}$ is the mass of the animal (m) plus the induced (added) mass of the fluid displaced by the animal during movement (m_{add}): $m^{\dagger} = m + m_{add}$. For a swimming animal, $m_{add} = 0.4\rho V$, where ρ is the density of the fluid medium and V is animal volume [119]:

$$m^{\dagger} = m + 0.4\rho V \tag{4.4}$$

Here, ρ is the density of seawater ($\rho = 1030$ kg/m^3), and V was obtained from a 3D model of an animal. MTag-estimated speed (v_{tag}) was used to approximate v_{com}, and a_{com} was approximated

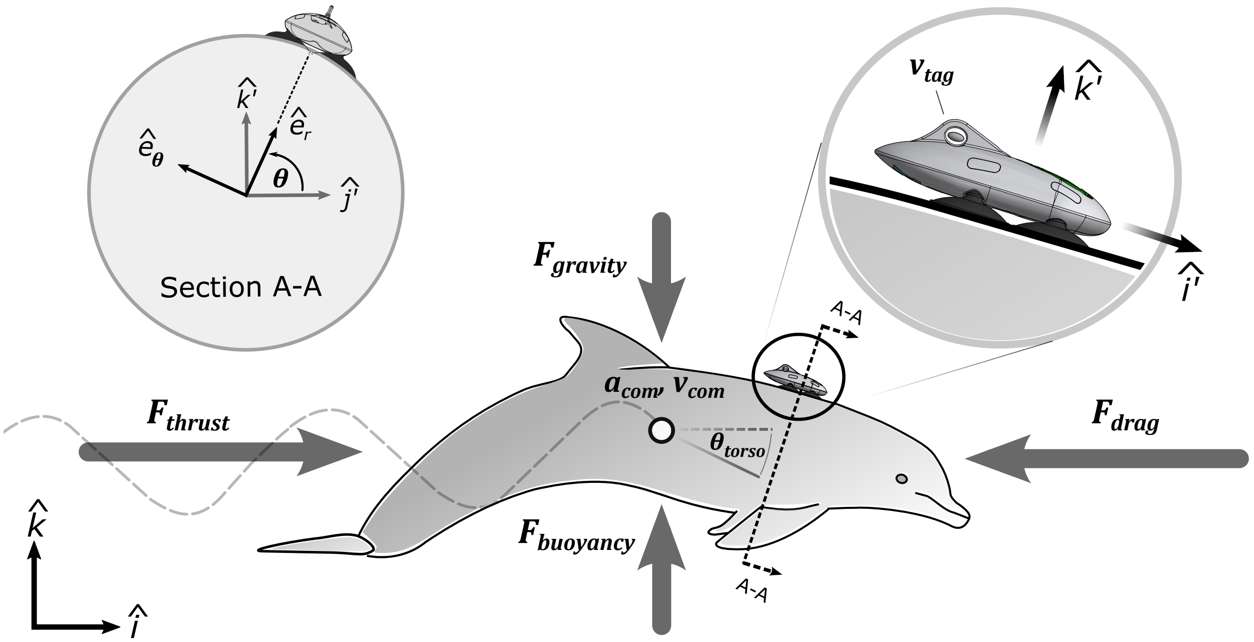


Figure 4.1: Diagram of forces on a swimming dolphin. This contribution focuses on the thrust and drag forces that act in the animal's direction of travel, and assumes that the buoyancy and gravitational forces cancel. The approximate MTag placement on the animal is displayed between the animal's blowhole and dorsal fin, with the fin of the tag parallel to the dorsal fin. POPOUT: The location of the micro-turbine is indicated (v_{tag}), along with the x and z tag accelerometer axes. SECTION A-A: View of the tag placement with animal-frame coordinate axes.

by numerically differentiating v_{tag} and smoothing the result using a 2-second moving average to reduce noise.

Drag acting on the animals was modeled according to work by F. Fish [25]:

$$F_{drag} = -0.5\rho A_s C_D v_{com}^2 \tag{4.5}$$

where A_s is the wetted surface area of the animal and C_D is the drag coefficient. However, Equation 4.5 does not account for increased drag due to surface drag effects [120]. A multiplier, γ, based on on depth was applied to the drag coefficient to account for surface effects: $C_{Dd} = C_D\gamma$. Applying the modified drag coefficient to Equation 4.5 yields:

$$F_{drag} = -0.5\rho A_s C_D \gamma v_{com}^2 \tag{4.6}$$

Animal surface area was obtained using $A_s = 0.08m^{0.65}$, where m is animal mass [25]. Non-depth normalized drag coefficients were computed using the relation $C_D = 16.99\,\mathrm{Re}^{-0.47}$ [41], with

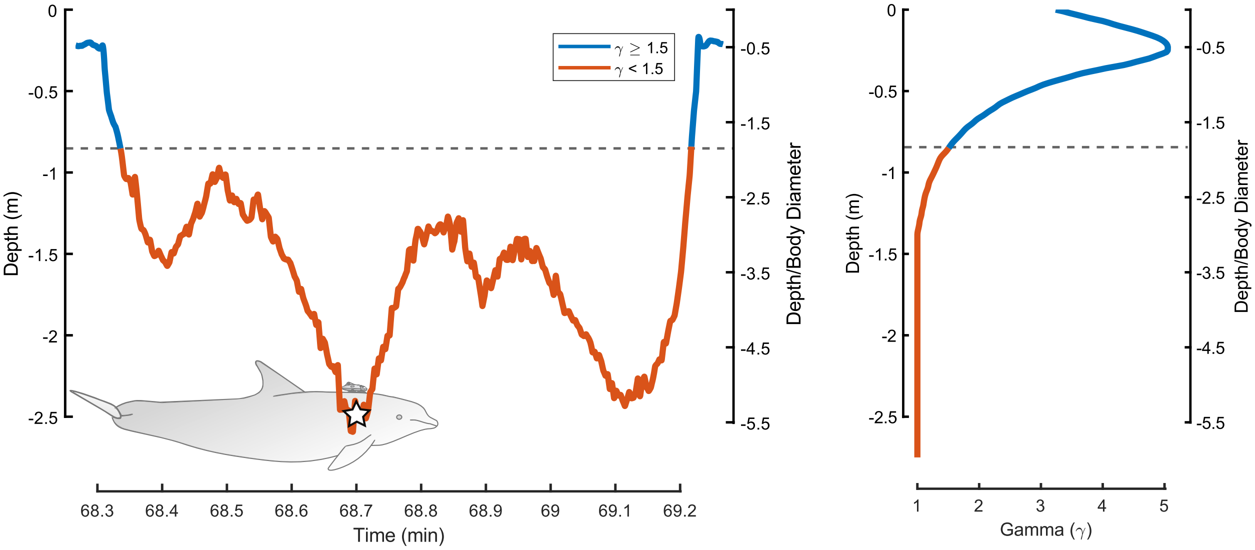


Figure 4.2: Illustration of animal drag multiplier γ and its relation to depth. LEFT: Example dive profile for dolphin T2, with high γ regions ($\gamma \geq 1.5$) in blue and low γ regions ($\gamma < 1.5$) in red. Depth is estimated for the COM of the animal, so the depth will not read 0 m during a surfacing event (i.e. only the blowhole is at the surface with the rest of the body underwater). RIGHT: Plot of the drag multiplier γ due to an animal's proximity to the surface, as a function of body diameters below the surface. As before, high γ is plotted in blue and low γ in red. The multiplier maximizes at -0.5 depth/body diameter, at $\gamma = 5.05$. Note: The true depth vs. γ relation presented here is specific to dolphin T2; only the depth/body diameter vs. γ relation applies in the general case.

$\mathrm{Re} = Lv_{tag}/\nu$. L is animal length and $\nu = 1.044\times10^{-6}$ m^2/s is the kinematic viscosity of seawater. Values for γ range from 1 to 5.05 for underwater swimming, and are dependent on the number of body diameters the animal is below the surface (Fig. 4.2, right). The depth/body diameter to γ relation was obtained from Hertel [121]. As before, v_{tag} is used as the best approximation of v_{com}.

The final expression for thrust power is then:

$$P_{thrust} = \underbrace{(m + 0.4\rho V)a_{fwd}v_{tag}}_{\text{Inertial}} + \underbrace{0.5\rho A_s C_D \gamma v_{tag}^3}_{\text{Drag}} \tag{4.7}$$

A nondimensional form of thrust power was also computed to better compare animals of varying lengths and masses:

$$P_{t,nd} = P_{thrust}/(mg^{1.5}L^{0.5}) \tag{4.8}$$

4.1.3 Work and COT Estimation

Work estimates (W_{thrust}) were obtained by numerically integrating sequences of P_{thrust} data using a trapezoidal sum. Only positive power components were included to ensure that periods where there was no active fluking did not artificially lower the work estimates (e.g. negative power/work during braking). Total Distance Traveled (TDT) estimates for discrete time intervals were calculated by numerically integrating v_{tag}, also using a trapezoidal sum. Instantaneous mass-specific cost of transport ($\text{COT}_{MS,I}$) was obtained with the expression:

$$\text{COT}_{MS,I} = P_{thrust}/(mv_{tag}) \tag{4.9}$$

Time-interval mass-specific COT ($\text{COT}_{MS,T}$) was obtained using:

$$\text{COT}_{MS,T} = W_{thrust}/(m \cdot \text{TDT}) \tag{4.10}$$

To account for energy loss as chemical energy is converted into mechanical energy, the mammalian metabolic-to-muscle power efficiency (chemical) was taken to be $\eta_{ms} = 0.25$ [122]. Muscle-to-propulsion power efficiency (mechanical) was taken to be $\eta_{sp} = 0.85$ for *Tursiops truncatus* [26]. Using these efficiencies, and an animals' Resting Metabolic Rate (RMR), total metabolic power requirements for propulsion were computed as:

$$P_{met} = P_{thrust}/(\eta_{ms}\eta_{sp}) + P_{rmr} \tag{4.11}$$

where P_{rmr} is the RMR power in Watts. Numerically integrating P_{met} (again only using $P_{thrust} > 0$) yielded the metabolic work W_{met}. Similar to $\text{COT}_{MS,I}$, the mass-specific metabolic COT for a given animal can be calculated as:

$$\text{COT}_{met} = (P_{thrust}/(\eta_{ms}\eta_{sp}) + P_{rmr})/(mv_{tag}) \tag{4.12}$$

4.1.3.1 Resting Metabolic Rate

A previous study at Dolphin Quest Oahu had assessed the mass-specific RMR of dolphin T2 (the animal used in the 24-hour study) to be 6.33 ml O_2 kg^{-1} min^{-1} ([38], Table 1, non-fasted). A non-fasted/fasted ratio of 1.53 was computed from the multi-animal RMR results presented in [38]. Dividing T2's non-fasted RMR by this ratio yielded a fasted mass-specific RMR of 4.13 ml O_2 kg^{-1} min^{-1}. RMR in ml O_2 min^{-1} was computed using T2's mass of $m_{T2} = 209$ kg, for 1,322 (non-fasted) and 863 (fasted) ml O_2 min^{-1} (Table 4.1). Dolphin T2's RMR was converted to units of power using the relation of 20.1 kJ (or 4.8 kcal) per liter O_2, to obtain $P_{rmr_{nf},T2} = 442.9$ W (non-fasted) and $P_{rmr_f,T2} = 289.0$ W (fasted) [107].

4.1.3.2 Metabolic Power and COT versus Speed

In order to find direct expressions from animal speed to P_{met} and COT_{met} for dolphin T2, the best-fit relation between T2's 24-hour steady-state low-γ ($\gamma < 1.5$) Day time (08:00-18:00) P_{thrust} and v_{tag} data was computed and then applied to Equations 4.11 and 4.12. A zero-intercept third-order polynomial was used for the initial fit (corresponding to the $P_{thrust} \propto v_{tag}^3$ relation), and fit using the least squares method, yielding the expression:

$$\bar{P}_{thrust}(v_{tag}) = av_{tag}^3 + bv_{tag}^2 + cv_{tag} \tag{4.13}$$

where a, b, and c are fitting coefficients. This was applied to Equation 4.11 to produce the metabolic power/speed relation:

$$\bar{P}_{met,T2}(v_{tag}) = (av_{tag}^3 + bv_{tag}^2 + cv_{tag})/(\eta_{ms}\eta_{sp}) + P_{rmr_{nf},T2} \tag{4.14}$$

$\bar{P}_{thrust}$ was then applied to Equation 4.12 to produce the metabolic COT/speed relation:

$$\overline{\text{COT}}_{met,T2}(v_{tag}) = \frac{1}{m_{T2}v_{tag}}\left(\frac{av_{tag}^3 + bv_{tag}^2 + cv_{tag}}{\eta_{ms}\eta_{sp}} + P_{rmr_{nf},T2}\right) \tag{4.15}$$

4.1.4 MTag Data Post-Processing

An estimate of the dolphin orientation was used to compute dolphin center-of-mass depth to determine the corresponding surface drag coefficient. Accelerometer, magnetometer, and gyroscope data were used with the Madgwick orientation filter to estimate animal orientation [10]. Before the orientation filter was applied, accelerations resulting from the the dynamic motion of the animal were removed from the accelerometry measurements using the following method. Note: axes with apostrophes (e.g. $\hat{i}'$) belong to the tag reference frame, and those without belong to the world reference frame.

The measured acceleration on the tag ($\mathbf{A}_m$) consists of the acceleration in the direction of the animal's motion ($\mathbf{A}$) plus the tangential and radial components of the acceleration due to animal rolling (cylindrical representation, Fig. 4.1, Section A-A):

$$\mathbf{A}_m = \mathbf{A} + [r_d\omega_{\hat{i}'}^2]\hat{e}_r + [r_d\alpha_{\hat{i}'}]\hat{e}_\theta \tag{4.16}$$

Here, $[r_d\omega_{\hat{i}'}^2]\hat{e}_r$ is the radial (centripetal) acceleration effect on $A_{m,\hat{k}'}$, $[r_d\alpha_{\hat{i}'}]\hat{e}_\theta$ is the tangential acceleration effect on $A_{m,\hat{j}'}$, and $\omega_{\hat{i}'}$ and $\alpha_{\hat{i}'}$ are the tag-frame angular velocity and acceleration, respectively, about $\hat{i}'$. These two acceleration effects must be removed for a more accurate estimate of $\mathbf{A}$. The magnitude of the radial acceleration measured at the accelerometer during a roll was computed directly using the dolphin radius r_d and the gyroscope measurement of $\omega_{\hat{i}'}$ (Fig. 4.1, popout). To obtain $\alpha_{\hat{i}'}$, the gyroscope reading of $\omega_{\hat{i}'}$ was smoothed using a Savitsky-Golay filter with a 0.25 second window and numerically differentiated. This produced the estimate of angular acceleration about $\hat{i}'$: $\bar{\alpha}_{\hat{i}'}$. The tangential acceleration effect was estimated as $[r_d \bar{\alpha}_{\hat{i}'}]\hat{e}_\theta$. The error-compensated tag-frame acceleration estimate was then:

$$\bar{\mathbf{A}} = \mathbf{A}_m - [r_d \omega_{\hat{i}'}^2]\hat{e}_r - [r_d \bar{\alpha}_{\hat{i}'}]\hat{e}_\theta \tag{4.17}$$

$\bar{\mathbf{A}}$ was used in the orientation filter alongside the gyroscope and magnetometer raw data. The filter was implemented as recommended in [10], with $\beta = 0.041$, using the Matlab toolbox distributed by the authors of the filter.

4.1.4.1 Depth Correction

The orientation from the filter was used to compute the MTag $\hat{k}$-location with respect to the animal centerline one animal radius beneath the tag. This location was selected instead of true animal COM as the COM location with respect to the MTag changed according to animal morphology during locomotion (spine flexing and curling through fluking), and it was a similar depth to COM during horizontal swimming. The orientation filter output the tag orientation in quaternion form ($\mathbf{q}_t$ at time step t), and taking its conjugate ($\mathbf{q}_t^*$) yielded the tag orientation with respect to the world frame. Converting $\mathbf{q}_t^*$ to rotation-matrix form ($\mathbf{R}_t^* \in \mathbb{R}^{3\times 3}$) made it possible to convert the vector distance from the tag to the animal centerline from the tag frame (δ) into the world frame ($\mathbf{\Delta}$):

$$\mathbf{\Delta}_t = \mathbf{R}_t^* \delta; \quad \delta = [\delta_{\hat{i}'}, \delta_{\hat{j}'}, \delta_{\hat{k}'}]^\top = [0, 0, r_d]^\top \tag{4.18}$$

where $\mathbf{\Delta}_t = [\Delta_{\hat{i},t}, \Delta_{\hat{j},t}, \Delta_{\hat{k},t}]^\top$. $\Delta_{\hat{k},t}$ was the t-th time step world-frame vertical distance from the MTag to the animal centerline, and this was used to more closely estimate animal depth:

$$d_t^\dagger = d_t - \Delta_{\hat{k},t} \tag{4.19}$$

where d_t was the original animal depth estimated from the pressure sensor at time step t.

To correct for measurement error created by fluid moving across the cylindrical cavity that housed the pressure sensor, Bernoulli's equation for incompressible flow along a streamline was used (seawater is insignificantly compressible for this analysis):

$$p + 0.5\rho v^2 + \rho g h = \text{constant} \tag{4.20}$$

where p is the fluid pressure, ρ is the density, v is the fluid velocity, g is the acceleration due to gravity, and h is the height in the water column (positive upwards). When comparing two different points on a streamline, Equation 4.20 becomes:

$$p_1 + 0.5\rho v_1^2 + \rho g h_1 = p_2 + 0.5\rho v_2^2 + \rho g h_2 \tag{4.21}$$

For this case, it was assumed Point 1 was static ($v_1 = 0$) directly in front of the tag, and Point 2 was moving across the pressure sensor. The depth difference between points was assumed to be minimal ($h_1 = h_2$). Applying these assumptions to Equation 4.21 yielded the pressure difference due to relative fluid velocity change:

$$p_1 - p_2 = 0.5\rho v_2^2 \tag{4.22}$$

The relation between a fluid pressure differential and the corresponding depth difference was also computed with Equation 4.21, this time neglecting the v-terms (pressure change due to *only* a change in depth):

$$p_1 - p_2 = \rho g(h_2 - h_1) \tag{4.23}$$

Combining Equations 4.22 and 4.23 yielded the depth estimation error due to fluid flow across the pressure sensor:

$$\Delta_h = h_2 - h_1 = v_2^2/(2g) \tag{4.24}$$

with $v_2 = v_{tag}$. The final estimate for animal depth was obtained by combining the results from Equations 4.19 and 4.24:

$$d_t^{\ddagger} = d_t - \Delta_{\hat{k},t} - \Delta_{h,t} \tag{4.25}$$

where $\Delta_{h,t}$ is the velocity-induced depth error Δ_h at time step t. The corrected depth ($d^{\ddagger}$) values were then used for estimating γ in the thrust power computation.

4.1.4.2 Surface Speed Estimation

The micro-turbine on the MTag was designed for underwater-only speed estimation. As a result, each time the tag is out of the water, the speed measurement is interrupted, causing periodic errors in the thrust power estimates. Speed gap data indices were identified by extracting regions of zero speed bounded by low-then-high acceleration jumps outside of heuristically-defined limits ($\mp 3.4\,\mathrm{m\,s^{-1}}^2$). To capture only transient animal surfacings, gaps that were longer than 3 seconds or bounded by speeds lower than 0.25 $\mathrm{m\,s^{-1}}$ were ignored (both thresholds also heuristically determined). This ensured that true sustained zero-speed data (e.g. when a dolphin was observing something outside of its habitat, above the water's surface) remained unaltered. Linear interpolations between values immediately before to immediately after each gap were used to estimate the missing speed data.

4.1.5 Steady-State Segmentation

Steady-state portions of animal swimming data (regions of minimal speed fluctuation) were extracted using a heuristically-tuned automated method. First, speed data (v_{tag}) were smoothed using a 2-second Savitzky-Golay filter to produce $\bar{v}_{tag}$. This filtering method was chosen for its ability to perform smoothing while preserving overall signal shape, which was useful when identifying specific time indices. A 2-second moving-window standard deviation was computed for $\bar{v}_{tag}$, to produce $\sigma_{\bar{v}}$. Regions with minimal variation in speed would then result in low $\sigma_{\bar{v}}$, and the smoothing ensured that only overall speed trend changes (rather than noise) were represented in the standard deviation estimate. Low $\sigma_{\bar{v}}$ alone was not sufficient to extract steady-state portions of swimming data, as small $\sigma_{\bar{v}}$ values also occurred when an animal was not moving. During steady-state swimming, a dolphin would be fluking to overcome drag effects, which would result in positive P_{thrust} rather than zero during no motion, hence minimal-movement low $\sigma_{\bar{v}}$ segments were filtered out by removing segments with low thrust power. Thrust power data were also smoothed with a 2-second Savitzky-Golay filter to produce $\bar{P}_{thrust}$, for noise reduction. A dolphin was then considered to be in a steady-state swimming pattern when $\sigma_{\bar{v}} \leq 0.045\,\mathrm{m\,s^{-1}}$ and $\bar{P}_{thrust} \geq 50\,\mathrm{W}$, both thresholds heuristically determined. Close-up examples of steady-state swimming are shown in Figure 4.3 (left side, yellow highlight).

4.2 Experimental Deployment

Experiments were conducted at Dolphin Quest Oahu, Oahu, HI, with six bottlenose dolphins participating in the study (morphometric measurements are reported in Table 4.1). Data were collected during prescribed swimming (lap trials), and from a single 24-hour session where the animal swam freely. For both cases, dolphins were trained by the ACSs to wear biologging MTags placed between the blowhole and dorsal fin, in the orientation indicated in Fig. 4.1. For the lap trials, the dolphins were asked to start at a floating dock (denoted “at station”), swim around an ACS located in the water typically 35 meters from the dock, and then return to station to complete the lap, all underwater. During the lap trials the animals completed up to 16 consecutive laps per session with one to three breaths taken between each lap. An example lap trajectory is displayed in Fig. 4.3, overlaid on a diagram of the lagoon. During the experiment, thirty trials were conducted, for a total of 282 laps between all six dolphins. The number of completed laps per dolphin is reported in Table 4.1. For the 24-hour trial, one dolphin (T2) was asked to wear an MTag for approximately 24 hours, from ~09:00 to ~09:00 the next day. Dolphin T2 was exempt from the lap trials for that

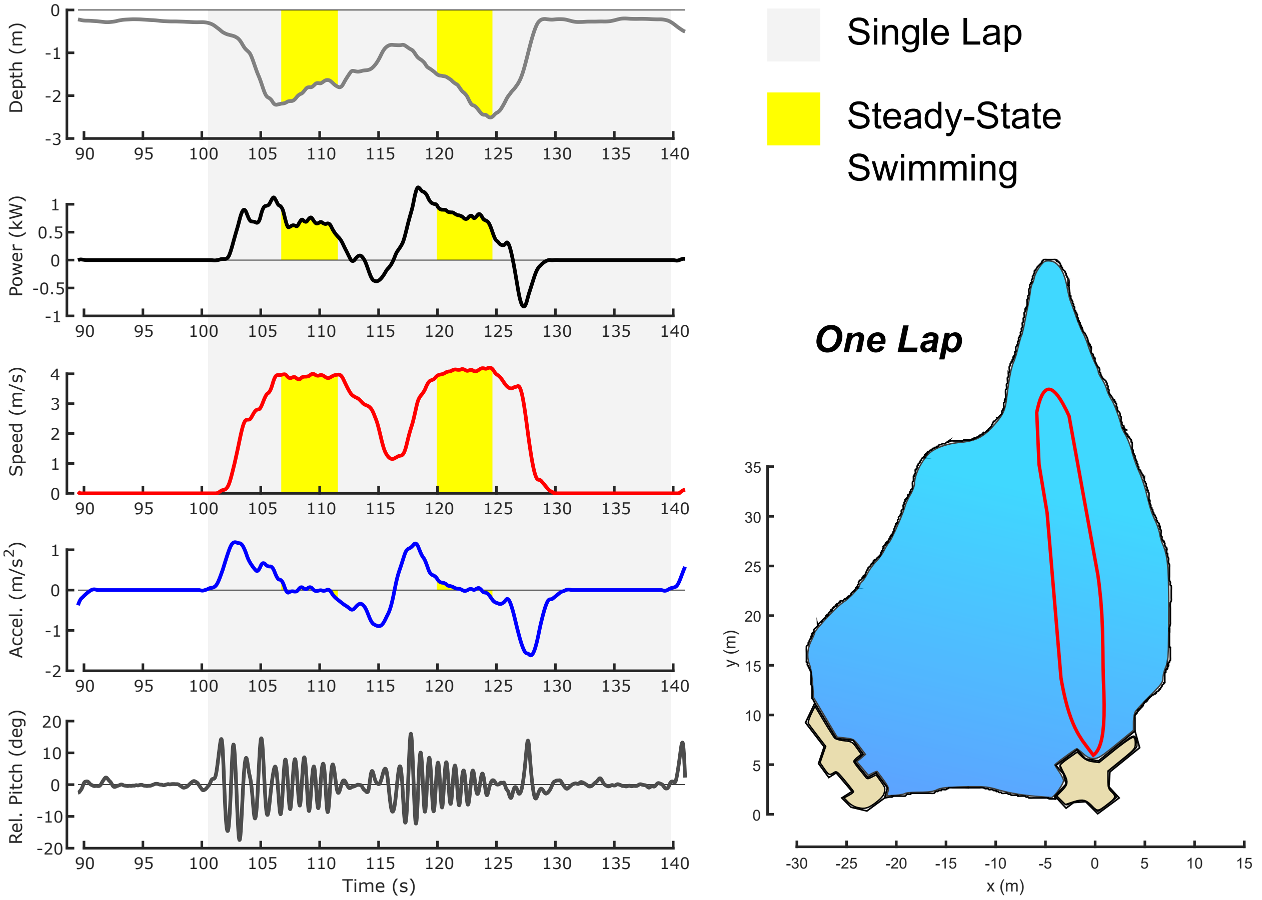


Figure 4.3: Dolphins were asked to perform laps in the Dolphin Quest Oahu Lagoon 2 (bottom-right). Nominal lap trajectory is shown by the red loop. Laps began at the dock (beige), looped around an ACS in the water (hairpin turn), and ended at the same dock. The depth (light gray), power (black), speed (red), forward acceleration (blue), and relative pitch (dark gray) of a sample lap (gray highlight) are shown on the left, with steady-state swimming regions highlighted in yellow.

day, so as to observe the animal's activity levels as it went through a standard day consisting of: free-swimming, standard ACS interactions, and public interaction sessions. Experimental sessions occurred between May 13th-28th and October 10th-13th in 2018, and May 5th-9th in 2019.

4.3 Lap Trials

4.3.1 Results

Mean lap trial swimming metrics were extracted for each dolphin with respect to full lap and steady-state time intervals. Swimming metrics from low-γ ($\gamma < 1.5$, threshold heuristically determined) steady-state segments were further isolated from the complete set to allow for comparisons to results from existing literature, and provide a view of animal swimming profiles during higher

Table 4.1: Lap trial metrics: animal measurements and summary parameters for the lap swimming trials. Parameters were extracted with respect to full lap, steady-state, and low-γ (surface drag multiplier < 1.5) steady-state time intervals. Dolphins ranged from 2.31-2.72 m in length and 143-245 kg in mass. Most animals (T1-T5) spent $> 33\%$ of swimming time in steady-state, with a separate majority (T2-6) maintaining mean γ penalties of $\leq 5\%$ while in steady-state. Mean lap distances ranged from 62.2-86.3 m, and mean steady-state speeds ranged from 2.8-5.0 $\mathrm{m\,s^{-1}}$.

	Dolphin	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**
Phys.	Mass (kg)	245	209	186	156	143	186
	Length (m)	2.72	2.57	2.46	2.37	2.31	2.52
	Radius (cm)	22.7	23.0	21.2	19.4	20.1	20.2
SS	Segment Count	86	95	85	74	119	10
	Time Fraction of Lap (%)	36.7	34.7	33.3	35.0	33.4	7.6
	Mean γ	1.32	1.00	1.00	1.00	1.02	1.05
	Mean Speed ($\mathrm{m\,s^{-1}}$)	5.0	3.7	4.2	2.8	3.9	3.4
	Mean SS Work/Lap (kJ)	12.2	5.1	6.0	3.1	3.9	0.7
	Mean SS TDT/Lap (m)	31.6	29.4	31.6	29.3	27.0	4.6
	Mean COT_{MS} ($\mathrm{J\,[kg\,m]^{-1}}$)	1.59	0.79	1.02	0.61	0.98	0.81
SS $\gamma < 1.5$	Segment Count	65	95	85	74	118	10
	Mean Duration (s)	4.7	4.1	4.2	4.0	4.2	3.5
	Mean Speed ($\mathrm{m\,s^{-1}}$)	5.2	3.7	4.2	2.8	3.9	3.4
	Mean Power (kW)	1.71	0.67	0.81	0.28	0.56	0.56
	Mean COT_{MS} ($\mathrm{J\,[kg\,m]^{-1}}$)	1.32	0.79	1.02	0.61	0.98	0.81
Full Laps	Lap Counts	65	48	47	28	70	24
	Mean γ	1.59	1.15	1.12	1.08	1.14	1.26
	Mean Speed/Lap ($\mathrm{m\,s^{-1}}$)	4.3	3.3	3.6	2.6	3.4	2.9
	Mean Power/Lap (kW)	1.71	0.67	0.75	0.27	0.61	0.54
	Mean Work/Lap (kJ)	29.2	14.2	15.9	8.0	11.8	10.1
	Mean TDT/Lap (m)	79.9	80.7	86.3	83.6	78.1	62.2
	Mean COT_{MS}/Lap ($\mathrm{J\,[kg\,m]^{-1}}$)	1.50	0.83	0.99	0.61	1.07	0.87

efficiency propulsion. The extracted metrics corresponded to animal speed, work/power, distance traveled, COT, and achieved γ values. Full lap metrics were computed first with respect to each lap, and these were averaged with respect to each animal to produce the final metrics estimates. Steady-state metrics were computed first for each steady-state interval, and these were then averaged with respect to each animal.

General metrics for "Full Laps", differentiated by animal, are reported in Table 4.1. Each dolphin completed between 28 (T4) and of 70 (T5) laps across all trials. Mean speed per lap varied between animals, from T4 at 2.6 $\mathrm{m\,s^{-1}}$ to T1 at 4.3 $\mathrm{m\,s^{-1}}$. The mean work per lap varied greatly between animals, from T4 at 8.0 kJ to T1 at 29.2 kJ. The mean total distance traveled per lap was relatively consistent across dolphins T1-5, from 78.1 m to 86.3 m, with T6 as the notable outlier at 62.2 m. These work and distance estimates resulted in differences in COT_{MS} per lap that ranged from T4 at the lowest value of 0.61 J/[kg m] (the lowest work per lap but the second-highest distance), up to T1 with 1.50 J/[kg m] (highest work per lap and middling distance).

Steady-state segments were extracted from each lap, for a total of 469 segments over the 282 laps. Individual animal steady-state segment counts and percent time per lap spent in steady-state were reported in Table 4.1, heading "SS". The majority of the dolphins (T1-5) spent one-third or more of their lap time in steady-state, with T6 as the outlier at 7.6%. All animals except T1 maintained mean steady-state γ close to 1, indicating the effects of surface drag were typically low during the lap trials.

Isolating the low-γ steady-state segments left 447 intervals across all animals (Table 4.1, heading "SS $\gamma < 1.5$"). Nearly all high-γ ($\gamma \geq 1.5$) segments were from T1 (T5 had one segment), with approximately 24% of steady-state segments for T1 having mean $\gamma \geq 1.5$. Dolphin low-γ steady-state segment durations ranged from 3.5 s (T6) to 4.7 s (T1), with most (T2-5) maintaining a duration closer to 4 s. Mirroring the full lap data, mean low-γ steady-state speeds ranged from 2.8 $\mathrm{m\,s^{-1}}$ (T4) to 5.2 $\mathrm{m\,s^{-1}}$ (T1). Mean powers ranged from 0.28 kW (T4) to 1.71 kW (T1), a factor of ~ 6. When compared to the full lap values, mean low-γ steady-state COT_{MS} for T2-4 remained similar ($< |4|$% change), with a change of -12% for T1, -8.0% for T5, and -7.5% for T6.

The per-segment means of length-normalized dolphin speed ($v_{nd} = v_{tag}/L$) and nondimensional thrust power ($P_{t,nd}$) data from the low-γ segments were correlated using a zero-intercept power fit (Fig. 4.4, top). The nondimensionalized power fit relation was found to be $P_{t,nd} = 0.023 v_{nd}^{2.54}$, with an adjusted R^2 of 0.917. The per-segment means of dolphin speed (v_{tag}) and thrust power (P_{thrust}) data from the low-γ segments were also correlated using a zero-intercept power fit for comparison to existing literature (Fig. 4.4, bottom). The power fit relation was found to be $P_{thrust} = 10.08 v_{tag}^{3.04}$, with an adjusted R^2 of 0.916.

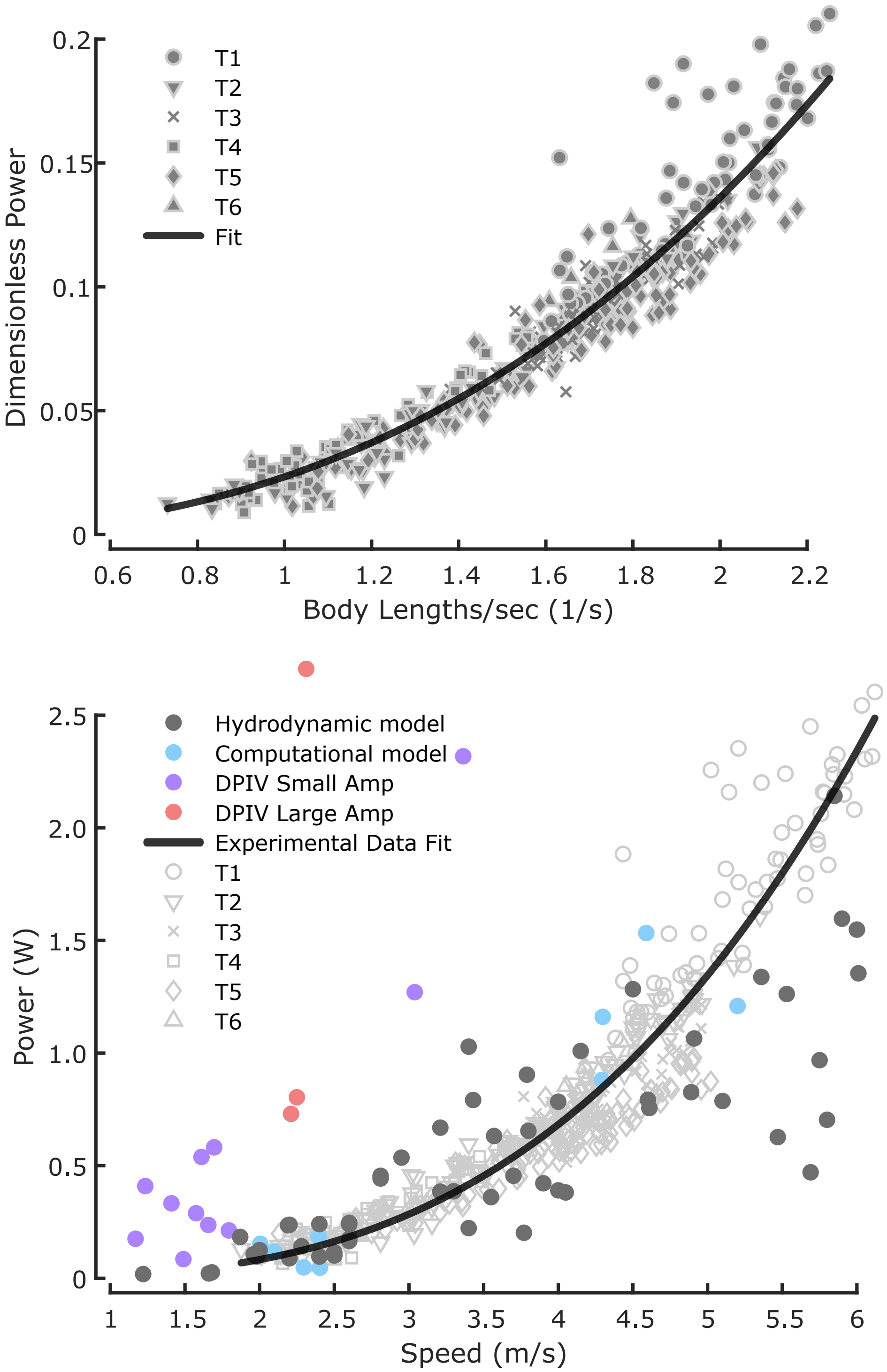


Figure 4.4: Comparisons of animal speed vs. power for steady-state time segments. TOP: Dimensionless animal swim trial data (gray shapes) were fit to a zero-intercept power curve (black). Each data point corresponds to the average body length-normalized speed and dimensionless power of a single steady-state segment, and only segments with $\gamma < 1.5$ were used (average duration 4.19 s). BOTTOM: The unmodified steady-state experimental data (light gray) and power fit (black) are compared to results from existing literature (solid circles: [26, 41]).

4.3.2 Discussion

During the prescribed swimming trials, data from 282 trials were collected and analyzed from 6 animals to investigate steady-state work and power during a straight-line swimming task. Each trial consisted of a lap from station to an ACS and back (range of 36.7 – 98.6 m), resulting in 22.4 km of swimming data for the analysis. During the trials, the animals swam over a range of speeds (lap means of 1.9 – 6.1 $\mathrm{m\,s^{-1}}$), spanning preferred speeds that have been reported in the literature [105]. Lap trial data were separated into transient and steady-state components to facilitate the comparison with results from the literature. Additionally, only the steady-state portions of the trials that were deep enough to avoid high surface drag effects were used for the analysis (Fig. 4.4).

During the lap trials, the dolphins tended to swim at depths that reduced the effects of surface drag on swimming performance. All six dolphins generally maintained relatively efficient depths (such that $\gamma < 1.5$) while in steady-state motion over a wide range of speeds and steady-state segment durations, with dolphins T1–5 in the steady-state swimming mode for more than 30% of the laps. Additionally, dolphins T2–6 *maintained hydrodynamically efficient depths* (minimal deviation from $\gamma = 1$) during high-effort steady-state swimming. Only T1 swam at steady-state depths with non-trivial penalties (mean $\gamma = 1.32$, 24% of laps $\gamma \geq 1.5$). Despite the $\sim$ 30% increase in drag at this average swimming depth, T1 had the highest average swimming speed of $\sim 5\ \mathrm{m\,s^{-1}}$, along with the highest per-lap swimming power. The resulting average cost of transport for T1 was more than twice as large as the most efficient swimmer, T4 (1.5 vs. 0.61 $\mathrm{J\,[kg\,m]^{-1}}$). The thrust power estimated from the low-γ steady-state swimming compared well to the existing literature (Fig. 4.4, bottom, in SI units). This was true for both Chopra's hydrodynamic model [26], and Fish's experimental methods [41], verifying the results from the proposed approach.

4.4 24-Hour Session

4.4.1 Results

Transient, steady-state, and general (transient $\bigcup$ steady-state) swimming metrics were extracted for dolphin T2 from the 24-hour session. Low-γ steady-state segments were also isolated from this complete set to provide comparisons between swimming profiles for T2's lap trial and 24-hour results. Extracted values for the 24-hour data concerned the same metric types as those in the lap trials. General and transient metrics were computed with respect to their entire data set/subset: general metrics were computed directly from the entire data set, and transient data intervals were first concatenated and then processed appropriately according to metric type (e.g. averaged, numerically integrated, etc.). 24-hour steady-state metrics were computed in the same format as the lap trials to allow for comparisons to existing literature and for consistency when performing statistical comparisons between lap trial and 24-hour session data. All metrics for the 24-hour session were computed for the Day, Night, and Overall (full session) time intervals.

Animal metrics for dolphin T2's 24-hour recording session were reported in Table 4.2, in terms of the following swimming modes: Transient, Steady-State, Steady-State $\gamma < 1.5$ (as a subset of the full Steady-State dataset), and General. Examples of 24-hour steady-state swimming data were highlighted (yellow) in Figure 4.5, top. Within these modes, the data were further split into "Day" (08:00-18:00) and "Night" (18:00-08:00) columns, with values in the "Overall" column using data from the entire 24-hour session. While in transient motion, T2 was less active (half the mean speed, one third the work done, and 77% distance traveled) at Night versus during the Day. Overall COT_{MS} values for T2 during the 24-hour trial were lower than those of the lap trials, and transient Night COT_{MS} values were less than half of those during the day. Mean γ values remained relatively high (≥ 1.5) during transient motion for T2.

To visualize dolphin T2's activity and work done over time, the mean thrust power (averaged within 1-minute bins) and thrust work per hour (neglecting the efficiency losses) were visualized in Figure 4.6 (top and bottom, respectively). Both plots have Day and Night time intervals indicated, and the work done per hour was further decomposed into steady-state and transient. T2 exhibited the highest sustained thrust power at three intervals in the Day: 09:00-11:25, 11:50-14:05, and 14:30-16:20. The highest observed 1-minute thrust power was 571 W at 09:28. T2 had sporadic peaks of high activity at Night at 19:25, 03:52, and 07:33, with corresponding powers of 297 W, 327 W, and 355 W. General mean thrust power was 137 W during the Day and 29 W at Night. The maximum thrust work done in a single hour interval was 916 kJ, from 10:00-11:00. This interval also featured the highest single-hour steady-state and transient work values, at 220 kJ and 697 kJ, respectively. The minimum thrust work done in an hour interval was 30 kJ, from 21:00-22:00.

The highest percentage of hourly work done while in steady-state was 32.6%, from 14:00-15:00, and the lowest was 0.2% from 00:00-01:00. T2 exerted a total thrust work of 6.64 MJ for the 24-hour period, with 5.07 MJ during the Day (1.07 MJ steady-state, 4.00 MJ transient), and 1.57 MJ at Night (0.10 MJ steady-state, 1.47 MJ transient) (Table 4.2). T2 was fed throughout the Day time period, and was considered to be under the non-fasted RMR condition from hours 08:00 to 18:00, transitioning from non-fasted to fasted from 18:00 to 22:00 (the 4 hours following the Day interval: food passage time through a bottlenose dolphin's digestive tract was estimated to be $\sim$ 4 hours [123]), and under the fasted condition otherwise. The transition from non-fasted to fasted RMR was modeled as a linear trajectory from $P_{rmr_{nf},T2}$ (442.9 W) at hour 18:00 to $P_{rmr_{f},T2}$ (289.0 W) at hour 22:00. Accounting for efficiency losses and adding the varying RMR across the 24-hour interval, T2 achieved a metabolic work of $W_{met} = 62.9$ MJ for that day, or 15,030 kCal.

During the 24-hour session, the majority of T2's time in steady-state occurred during the Day (80%), with 12.6% of its Day time swimming in steady-state as compared to 1.7% at Night. Similar

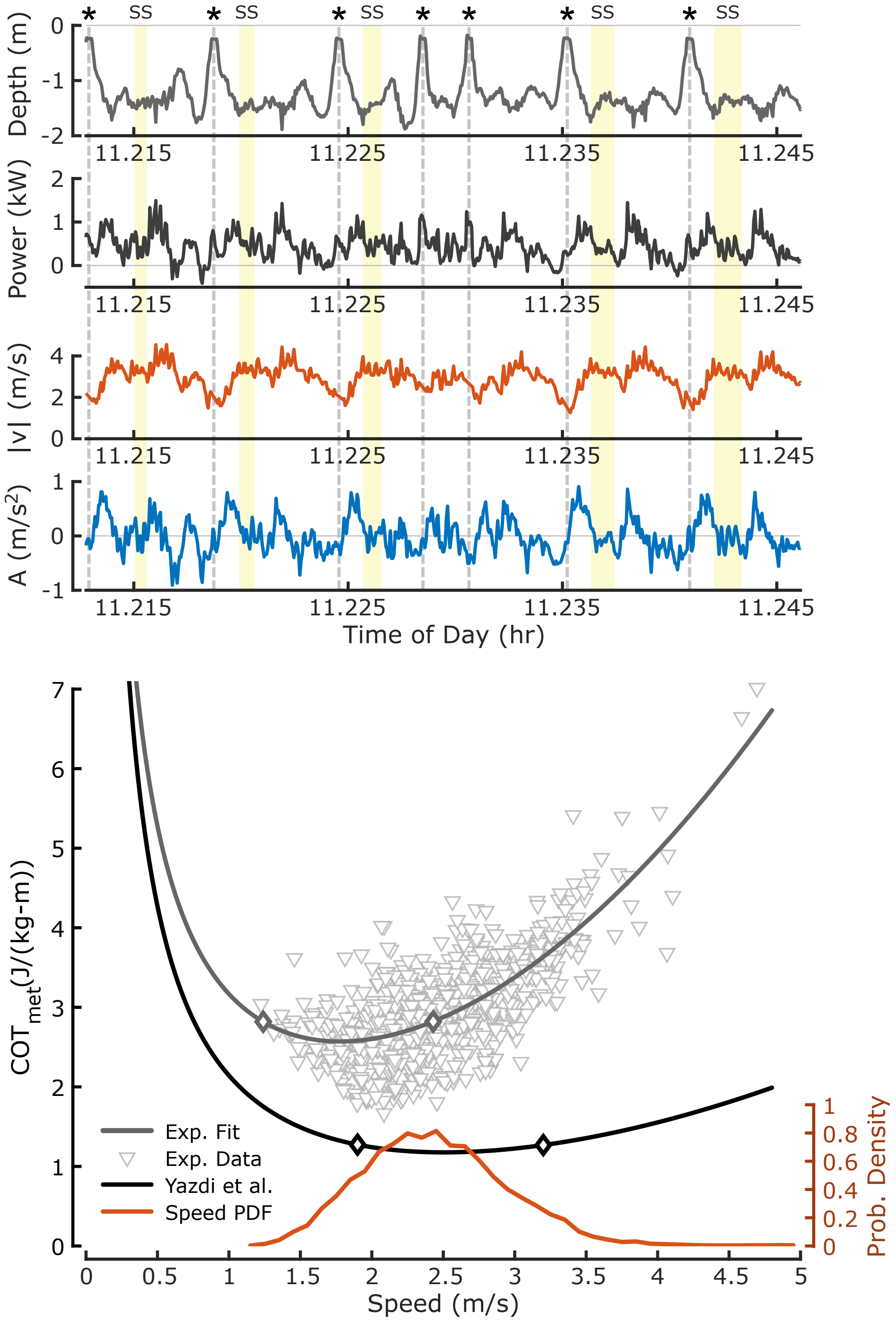


Figure 4.5: TOP: High activity time window displaying a 2-minute sample of T2's depth (gray), thrust power (black), forward speed (red), and forward acceleration (blue). Surfacing events are indicated (light gray, dashed/starred). Steady-state segments are highlighted (yellow). BOTTOM: Steady-state low-γ Day COT$_{met}$ and speed data points (gray triangles) from the T2 24-hour session, with the corresponding best-fit curve (gray). The COT$_{met}$ curve from Yazdi et al. (black) is shown for comparison. Diamonds on the curves show the optimal swimming speed region limits (Exp. Fit: 1.2–2.4 m s^{-1}, Yazdi et al.: 1.9–3.2 m s^{-1}). T2's speed PDF during steady-state low-γ Day swimming (red) corresponds to the right-sided y-axis.

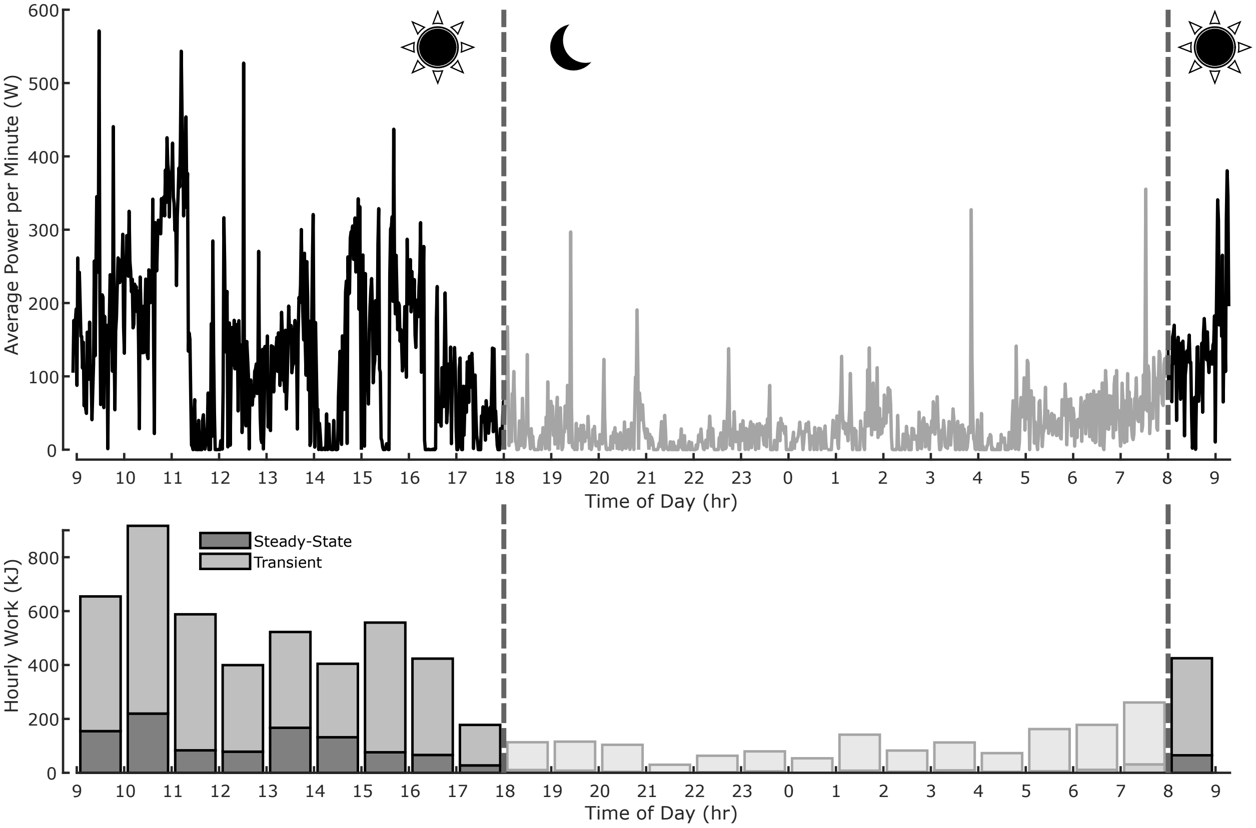


Figure 4.6: TOP: Example power values for dolphin T2 over a 24hr period, beginning at 08:55 and ending at 09:18 the next day. Day (08:00-18:00, when the ACSs were present) and night (18:00-08:00, when the ACSs were not present) time intervals are indicated. For visibility, values were averaged within bins of 1-minute durations. BOTTOM: Total propulsive work performed by T2 for each hour interval between 09:00 and 09:00 of the next day, with day and night intervals in line with the top figure. The work per interval is separated into steady-state (dark) and transient (light) components.

Table 4.2: T2 24-hour session metrics: summary parameters for T2's 24-hour free-swimming session. Parameters were extracted with respect to transient, steady-state, low-γ (surface drag multiplier < 1.5) steady-state, and general (transient & steady-state) time intervals. Parameters were calculated for Day (08:00-18:00), Night (18:00-08:00), and Overall (Day & Night) for each time interval type. Over the 24-hour period, T2 swam 78.1 km and produced 6.64 MJ of propulsive work (before accounting for efficiencies). Note: values in the Overall column were averaged from the Day and Night values using the durations of the Day and Night time intervals as weights.

	Metric	**Day**	**Night**	**Overall**
Transient	Mean Speed ($\mathrm{m\,s^{-1}}$)	1.20	0.59	0.83
	Work Done (MJ)	4.00	1.47	5.47
	TDT (km)	37.8	29.0	66.8
	Mean COT_{MS} ($\mathrm{J\,[kg\,m]^{-1}}$)	0.51	0.24	0.39
	Mean γ	2.66	2.84	2.77
Steady-State	Segment Count	1243	268	1511
	Mean Duration (s)	3.66	3.17	3.57
	Time Fraction in SS (%)	12.6	1.7	6.2
	Mean Speed ($\mathrm{m\,s^{-1}}$)	2.22	1.44	2.10
	Mean Power (W)	237	119	216
	Work Done (MJ)	1.07	0.10	1.17
	TDT (km)	10.1	1.2	11.3
	Mean COT_{MS} ($\mathrm{J\,[kg\,m]^{-1}}$)	0.51	0.41	0.50
	Mean γ	1.59	2.33	1.72
	Time Frac. $\gamma < 1.5$ (%)	71.0	37.7	65.8
SS $\gamma < 1.5$	Segment Count	864	104	968
	Mean Duration (s)	3.74	3.08	3.67
	Mean Speed ($\mathrm{m\,s^{-1}}$)	2.43	1.65	2.35
	Mean Power (W)	233	107	219
	Work Done (MJ)	0.76	0.04	0.79
	TDT (km)	7.8	0.5	8.4
	Mean COT_{MS} ($\mathrm{J\,[kg\,m]^{-1}}$)	0.43	0.27	0.41
General	Mean Speed ($\mathrm{m\,s^{-1}}$)	1.33	0.60	0.90
	Work Done (MJ)	5.07	1.57	6.64
	TDT (km)	47.9	30.2	78.1
	Mean COT_{MS} ($\mathrm{J\,[kg\,m]^{-1}}$)	0.51	0.25	0.41
	Mean γ	2.53	2.83	2.70

to the transient results, activity levels for T2 were much lower at Night (65% of the mean speed and half the mean power). Compared to the transient values, COT_{MS} during steady-state was equivalent during the Day and higher at Night (171%), for an overall higher mean COT_{MS} during steady-state (128%). Again compared to transient, the steady-state γ values were lower during both Day (60%) and Night (82%). Dolphin T2 spent a majority of its Day steady-state swimming in regions of low-γ, and a majority of its Night steady-state swimming in regions of high-γ. Combining distances traveled for transient and steady-state, T2 swam a total of 78.1 km over the entire 24-hour session.

Equation 4.12 was used to combine $P_{rmr_{nf},T2}$, η_{ms}, and η_{sp} with mean P_{thrust} and v_{tag} values from T2's 24-hour steady-state low-γ Day segments to generate the experimental speed versus COT_{met} data points (Fig. 4.5, bottom). The best-fit curve of the experimental data (gray) was contrasted to the COT curve computed by Yazdi et al. (black) [37]. The best-fit speed versus $\bar{P}_{met}$ and $\overline{\mathrm{COT}}_{met}$ relations for T2 (Equations 4.14 and 4.15) were found to have fit coefficients of $a = 14.0$, $b = -20.1$, and $c = 52.5$. The optimal (minimum) COT_{met} for T2 was found to be 2.57 $\mathrm{J/(kg\,m)}$ at 1.78 $\mathrm{m\,s^{-1}}$, with a maximum range speed interval (speeds yielding COT_{met} within 10% of the minimum) of 1.24–2.43 $\mathrm{m\,s^{-1}}$. The steady-state Day time low-γ speed PDF for T2 was plotted on a secondary axis (Fig. 4.5, bottom, red). The speed PDF peaked at 2.40 $\mathrm{m\,s^{-1}}$ (std. dev. of 0.50 $\mathrm{m\,s^{-1}}$) with values ranging from 1.13–4.98 $\mathrm{m\,s^{-1}}$.

4.4.2 Discussion

The 24-hour data collection session was used to create a continuous measure of animal activity and estimated energy expenditure over an extended duration of free-swimming. These data yielded a detailed picture of T2's activity, with variable levels during the day and reduced activity at night. During the 24-hour session, T2's activity peaked in the morning (~10:00 local), decreased during the rest of the day, and dropped at night (Fig. 4.6). This behavioral pattern has been reported in the literature for managed dolphins [51], where activity levels of the animals are closely associated with the schedule put in place by the ACSs. The Hourly Work plot (Fig. 4.6, bottom, per-hour transient vs steady-state work) indicates that T2 engaged in more sustained (i.e. steady-state) bouts of swimming during the Day time not just overall, but also as a larger fraction of its active time when compared to Night. This pattern is present qualitatively in the Hourly Work plot, and quantitatively in Table 4.2, where the steady-state time fraction was an order of magnitude larger for the Day versus Night time.

The 24-hour free-swimming session also provided an opportunity to collect baseline data about T2's movement in the lagoon environment. T2 swam 78.1 km (47.9 km during the Day and 30.2 km during at Night) while wearing the tag, with a mean speed of 1.3 $\mathrm{m\,s^{-1}}$ during the Day and 0.6 $\mathrm{m\,s^{-1}}$ at Night. The 78 km traveled was higher than expected, and represents *the first persistently-*

estimated 24-hour range measurement for a bottlenose dolphin. Interestingly, the animal still swam $\sim$ 30 km at night when activity levels were significantly reduced. The longest distance traveled in a day for wild dolphins reported in the literature is 47 km/day between two separate sightings [124]. This was recorded as a minimum distance as the estimate only considered the regional travel distance and not local transient travel (foraging, diving, etc.). Bottlenose dolphins can have extended habitat ranges [124, 125], and the total distance traveled observed in this lagoon environment sheds light on the per-day travel capabilities of these animals.

Steady-state power estimates and the corresponding distances traveled were used to calculate T2's cost of transport over the 24-hour tag deployment. The steady state COT_{met} curve during the day is presented in Figure 4.5, and indicates a minimum COT_{met} of 2.57 $\mathrm{J/(kg\,m)}$ at a swimming speed of 1.78 $\mathrm{m\,s^{-1}}$, with an optimal speed range (where COT_{met} is within +10% of the minimum) of 1.24 – 2.43 $\mathrm{m\,s^{-1}}$. The mean recorded steady-state swimming speed of 2.22 $\mathrm{m\,s^{-1}}$ is at the high end of what would be considered optimal, and may indicate that T2's preferred swimming speed was influenced by factors beyond cost minimization alone. A total COT_{met} of 0.51 $\mathrm{J/(kg\,m)}$ during the Day and 0.25 $\mathrm{J/(kg\,m)}$ at Night was calculated from the data. In addition to average cost of transport, T2's 24-hour metabolic energy usage was estimated to be $W_{met} = 15,030$ kCal, which was compared to its reported daily caloric intake. T2 was fed 13,337 kCal of a mixture of mackerel and squid each day for the month of May 2018, assuming an assimilation efficiency of 87%, T2 had $\sim 11,600$ kCal of available metabolic energy per day [126]. By this measure, W_{met} overestimated T2's metabolic energy use by 30% for that day.

The overestimate of T2's daily energy consumption may be attributed to several factors. The research presented here uses a drag coefficient (C_D) formula estimated from animals engaged in low amplitude swimming [41], and this C_D was employed in our calculations for all identified periods of locomotion. Future work could explore the use of separate expressions for C_D for fluking and gliding swimming modes, as these may differ in drag characteristics. Next, the assumed resting metabolic rate for the animals accounts for a significant portion of the estimated daily metabolic energy requirements (50.3% for T2's 24-hour session). Our work used values measured by van der Hoop et al. [38], and additional data collection could improve animal-specific measurements to raise the accuracy of the approach. Finally, as thermoregulatory costs constitute a portion of a dolphin's RMR, there is reason to explore a variable RMR dependent on the dolphin's propulsive effort [37]. As propulsive effort increases, it is expected that the waste heat from muscular inefficiency (η_{ms}) provides some portion of the animal's thermoregulatory requirements. This would lower the overall RMR requirements during motion and lower the estimated energy consumption.

When comparing T2's best-fit COT_{met} curve (Eq. 4.15) to the COT curve from Yazdi et al. [37], our model was higher over all speed ranges, penalized higher speeds more heavily, and had slower and smaller-range optimal speed regions (this experiment: 1.24–2.43 $\mathrm{m\,s^{-1}}$, Yazdi et al.:

1.9–3.2 $\mathrm{m\,s^{-1}}$). T2's steady-state low-γ Day time swimming speed PDF peak (2.40 $\mathrm{m\,s^{-1}}$) fell within the optimal speed region proposed by Yazdi et al., and just inside the optimal region found in this research. For reference, Yazdi et al. employed respirometry to measure the metabolic rates of the dolphins in a controlled environment (exempting their data from efficiency conversions), and estimated animal swimming speeds through direct observation (speed extrapolation from duration and distance traveled). They observed speeds in the range of 0.8–2.6 $\mathrm{m\,s^{-1}}$, with speeds above this range (denoted "high-speed") assigned a general value of 3 $\mathrm{m\,s^{-1}}$ (it was also noted that there was uncertainty on the accuracy of the high-speed values as they could not be reliably measured). In contrast, the speeds used in our COT estimation ranged from 1.13–4.98 $\mathrm{m\,s^{-1}}$, all of which were directly measured by the MTag's onboard sensor.

A separate study performed by Williams et al. used electrocardiography to estimate bottlenose dolphin metabolic rates (using a data-driven relation of heart rate to metabolic rate) as they kept pace with a boat traveling in open water [36]. These speed and power data were used to produce COT estimates for their animals, and for higher speed (non wave-riding) travel, they found that their dolphins had a mean COT of 2.85 J $[\mathrm{kg\,m}]^{-1}$ at 2.9 $\mathrm{m\,s^{-1}}$. At the same speed, the model in Yazdi et al. predicted 1.21 J $[\mathrm{kg\,m}]^{-1}$ (58% lower), and the model from this research predicted 3.26 J $[\mathrm{kg\,m}]^{-1}$ (14% higher). As each study used widely different methods of estimating metabolic rate (respirometry, electrocardiography, and fluid dynamics + RMR), a further investigation where all three data streams are used to estimate cost is warranted.

4.5 Experimental Condition Comparison

Dolphin T2's low-γ lap trial versus 24-hour session steady-state swimming speed, power, and COT were compared via averages, with statistical significance evaluated using the two-sample *t*-test. The Day interval values were used for this comparison, as this was when the animal was assumed to be "active" and not in a resting state. Further comparisons were made between T2's achieved γ values during active swimming: full lap and Day time general γ means were compared, and lap trial steady-state and Day time steady-state γ means were compared. The statistical significance was again evaluated using the two-sample *t*-test.

T2's free-swimming results differed from the lap trials in several key aspects: general activity level, depth selection, and percent of time spent in steady-state. To make a direct comparison between similar swimming modes, steady-state low-γ segments from the 24-hour session (Table 4.2, heading "SS $\gamma < 1.5$") were compared with the swimming metrics for T2 from the lap trials (Table 4.1, heading "SS $\gamma < 1.5$"). Assuming T2's active portion of the 24-hour session to be during Day time, activity levels for this period were lower during free-swimming than in the lap trials. Day time steady-state low-γ COT was significantly lower for the free-swimming session versus the lap trials (0.43 vs. 0.79 J $[\mathrm{kg\,m}]^{-1}$ or 54%, $p = 6.7 \times 10^{-68}$). This directly follows from the corresponding lower relative swimming speed (2.4 vs. 3.7 $\mathrm{m\,s^{-1}}$ or 66%, $p = 5.2 \times 10^{-87}$) and power (233 vs. 672 W or 35%, $p = 5.3 \times 10^{-95}$) results.

Although T2 achieved a lower COT during the 24-hour session than during the lap trials, this was not due to minimizing surface drag. T2's γ-values seen during Day time free-swimming were significantly higher than those achieved during the lap trials (General: 2.53 vs. 1.15 or 220%, $p < 10^{-307}$; Steady-State: 1.59 vs. 1.00 or 159%, $p = 9.1{\times}10^{-9}$). This indicated that while T2 used efficient swimming depths at high activity levels when completing a task (lap trials), *minimizing surface drag was not heavily prioritized at lower travel speeds* (24-hour session). The difference in T2's surface drag results indicates that it adjusted to the higher drag experienced during the high-intensity lap trials by prioritizing efficient travel depths. Further research is necessary to determine whether this behavior is observed in other dolphins in similar swimming conditions.

Dolphin T2 also did not engage in steady-state swimming nearly as frequently in the 24-hour session as in the lap trials (12.6% [Day] vs. 34.7%). Similarly, the percent distance traveled while in steady-state was lower in the 24-hour session than in the lap trials (21.1% [Day] vs. 36.4% [TDT/Lap]), as is the case with the total work done (21.1% [Day] vs. 35.9% [Work/Lap]). This could be due to increased use of a mixed fluke-and-glide gait during self-selected swimming that dolphins are known to use for extended-duration motion [67], as the acceleration-to-deceleration profile of this gait is not consistent with the uniform-speed requirements of steady-state. The dolphin's fluctuating use of transient versus steady-state fluking dependent on swimming condition (mixed-speed free-swimming versus high-speed lap trials) demonstrates the *necessity for metabolics estimation methods to be robust to both steady-state and transient swimming*. The method presented in this chapter is intended to satisfy this necessity through its ability to persistently monitor animal speed and power, allowing for fine-scale metabolics analyses.

4.6 Future Work and Conclusions

While the lap trial results offered insights into the similarities and differences between individual dolphin swimming behaviors, more work can be done to relate mechanical work and power to metabolic cost. Combining tag-based power estimates with respirometry-based metabolic measurements has the potential to enhance our understanding of the relation between animal propulsive power and metabolic rate [127]. The residual of the measured metabolic cost after accounting for thrust power can be used to provide insight into the non-propulsive components of a dolphin's energy expenditures (thermoregulation, digestion, etc.) during controlled experiments. Quantifying these additional components would then help build a more accurate model to estimate an animal's energetics using this combination of tag-based kinematic estimates and hydrodynamic models.

The 24-hour trial offered quantitative insight into day-scale biomechanics, though more data collection is required to build a baseline data set that can be used to examine generalizable trends observed for animals in this environment. To extend this research, additional dolphins from this population can be tagged for extended durations in repeated sessions during different times of the year. These data could be used to investigate topics concerning this population such as: detailed circadian rhythm, activity level dependency on proximity to conspecifics, generalized cost of transport estimation, and seasonal energy expenditure changes. This framework can also be extended to wild animals with small modifications to the biologging tags used in this work. The tags currently in use have not been outfitted for open-water retrieval, which will be necessary due to their archival (versus transmitting) design. Additionally, the speed sensor requires the tag to be aligned with the animal's body for optimal performance, so it may be necessary to update the suction cup design to limit the chance of tag slippage (and misalignment) after initial placement.

This contribution paired measured kinematics with a physics-based thrust model which yielded insights into the swimming biomechanics of managed bottlenose dolphins and how they modulate their propulsive effort during high-intensity directed tasks and varying-intensity free-swimming. The ability to measure speed from the animal using a biologging tag was a key technical component that enabled the framework presented in this chapter. Additionally, this research produced new knowledge about the day-scale range capability of a bottlenose dolphin in a lagoon environment, including a 78 km total distance traveled estimate for the animal. Finally, the estimated propulsive power combined with a resting metabolic rate from the literature were used to estimate an animal's total metabolic rate and steady-state cost of transport throughout a free-swimming recording session. While the total metabolic rate model can be refined, the results in this manuscript offer a representative example of the opportunities for investigation afforded by the proposed approach.

CHAPTER 5

High-Fidelity 3D Monitoring: Spatially Contextualized Animal Metrics

The ability to accurately and precisely track animals in their habitat can enable an entirely new vein of research on managed marine animals. While bottlenose dolphins are presented in this chapter as a case study, the monitoring framework presented in this chapter can be modified and extended to apply to other commonly tagged underwater animals. At its heart, the framework is based on fusing high-frequency dead-reckoning using the approach from Section 3.1 with low-frequency external localizations through pose-graph optimization [44], and this property is what allows it to be generalized to other localization applications. As a biologging tag is core to the operation of this method, tag-based metrics on animal behavior can now be examined with respect to where they occurred within the environment. To demonstrate this capability, animal energetics are estimated using the method from Chapter 4 and contextualized via location. Since the majority of the method is automated, the framework presented here also makes processing long-duration datasets more feasible, simplifying and standardizing large-scale raw data analysis. This chapter details the hardware and analysis methods used as part of the monitoring framework (Section 5.1), demonstrates their use by tracking a bottlenose dolphin in a managed setting (Section 5.2), and

---

Content of this chapter also published as:

J. Gabaldon, D. Zhang, K. Barton, M. Johnson-Roberson and K. A. Shorter, "A framework for enhanced localization of marine mammals using auto-detected video and wearable sensor data fusion," *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2017, pp. 2505-2510, doi: 10.1109/IROS.2017.8206069. ©IEEE 2017. Reprinted with permission.

D. Zhang, J. Gabaldon, L. Lauderdale, M. Johnson-Roberson, L. J. Miller, K. Barton, and K. A. Shorter, "Localization and Tracking of Uncontrollable Underwater Agents: Particle Filter Based Fusion of On-Body IMUs and Stationary Cameras," *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 6575-6581, doi: 10.1109/ICRA.2019.8794141. ©IEEE 2019. Reprinted with permission.

Content of this chapter also to be submitted as:

J. Gabaldon, D. Zhang, L. Lauderdale, L. Miller, M. Johnson-Roberson, K. Barton, K. A. Shorter, "High-fidelity bottlenose dolphin localization and monitoring in a managed setting," *Science Robotics*. Copyright may be transferred without notice, after which this version may no longer be accessible.

explores the opportunities of the resulting data in evaluating swimming behavior and energetics as they pertain to location (Section 5.3).

5.1 Framework Structure

This section describes the general structure of the framework, with the following subsections describing each process in more detail. Data for this research was collected through a combination of a biologging tag attached to the animal, and a pair of environmentally-mounted video cameras (Fig. 5.1, A). This is the same experimental environment used in Section 3.1.2, and uses the MTag hardware described in Section 4.1.1. First, data from the tag was used to produce an odometry (dead-reckoning) estimate of animal movement. Next, the locations of all animals in the environment were tracked in recorded video through a CNN object detector. These detections were error-corrected dependent on depth information from the tag, and were combined with the odometry estimate through a particle filter to produce a drift-corrected track of the dolphin. Finally, animal localization was further improved by: 1) exploiting temporal associations between CNN localizations to identify detections that corresponded to the tagged animal, and 2) fusing the odometry and identified detections through pose-graph optimization.

5.1.1 Odometry Generation

Before animal odometry can be produced, the raw tag data must be processed. Odometry requires two principal components at each time step: 1) body orientation, and 2) distance traveled. Body orientation was obtained by passing IMU data through the Madgwick orientation filter [10]. To do this, acceleration data ($\mathbf{A}$) were first pre-processed to remove effects on the tag from animal rolling actions (tangential and centripetal), using the method described in Section 4.1.4. The modified acceleration data ($\bar{\mathbf{A}}$) were then combined with the gyroscope and magnetometer data in the Madgwick filter to produce the full animal orientation as a set of 3D rotation matrices from animal to world-frame ($\mathbf{R} \in \mathbb{R}^{3\times3}$). This research uses right-handed axis conventions of x-right, y-front-to-back, and z-up with respect to the local habitat for the world frame (Fig. 5.1, top), and x-forward, y-left, and z-up with respect to the orientation of the MTag for the animal frame. World-frame velocity ($\mathbf{v}_w$) was generated by rotating the animal-frame velocity ($\mathbf{v}_a$) using $\mathbf{v}_w = \mathbf{R}\mathbf{v}_a$, where $\mathbf{v}_a = [v_{tag}, 0, 0]^\top$, at each time step. Specific acceleration ($\mathbf{A}_s$) was computed by rotating $\bar{\mathbf{A}}$ into the world frame and subtracting gravitational effects at each time step:

$$\mathbf{A}_{s,k} = \mathbf{R}_k\bar{\mathbf{A}}_k - [0, 0, g]^\top, \ \bar{\mathbf{A}}_k \in \mathbb{R}^{3\times1} \tag{5.1}$$

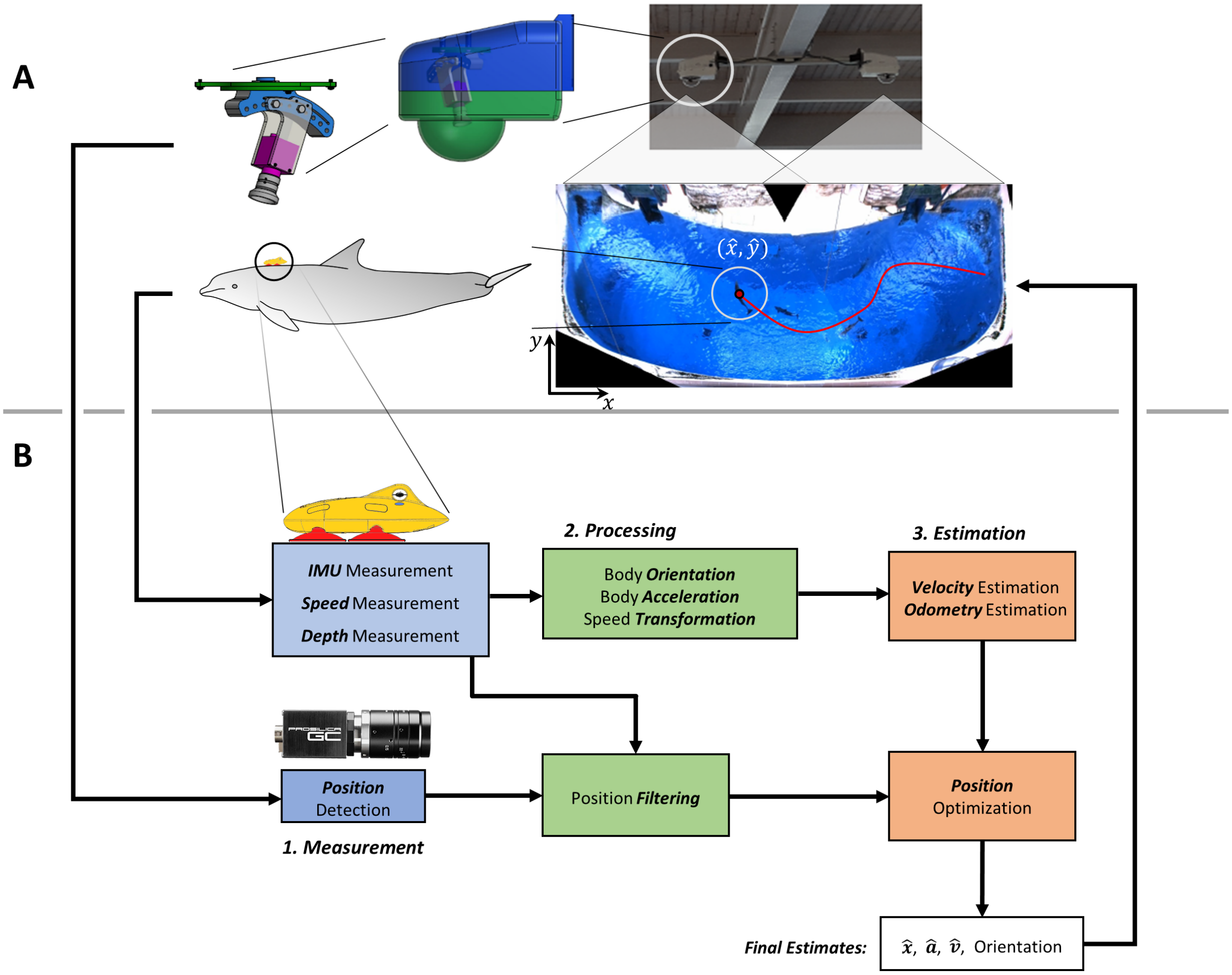


Figure 5.1: Experimental setting/sensing hardware relation to the localization framework methods. A: Diagram illustrating the experimental habitat, on-animal tag placement, and environmental camera setup used in this research. The main habitat of the Seven Seas Dolphinarium (bottom-right) is overlooked by two cameras in weatherproof housing (top). These are used to visually record a tagged bottlenose dolphin (bottom-left). B: Block diagram of the localization framework. Sensor data from the cameras and biologging tag are fed into two interconnected computation streams, which are combined to provide a high-fidelity estimate of animal locations and kinematics.

Procedure 5.1 Drift-Correcting Kalman Filter

```
procedure MODIFIED KALMAN FILTER (3D)($\mathbf{u}$, $\mathbf{z}$, $\mathbf{n}_{u,a}$, $\mathbf{n}_{u,d}$, $\mathbf{n}_z$, $T$)
    $\mathbf{F} = \begin{bmatrix} 1 & -T & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -T & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -T \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \ \mathbf{B} = \begin{bmatrix} T & -T^2/2 & 0 & 0 & 0 & 0 \\ 0 & T & 0 & 0 & 0 & 0 \\ 0 & 0 & T & -T^2/2 & 0 & 0 \\ 0 & 0 & 0 & T & 0 & 0 \\ 0 & 0 & 0 & 0 & T & -T^2/2 \\ 0 & 0 & 0 & 0 & 0 & T \end{bmatrix}$
    $\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$
    $\mathbf{Q} = \mathbf{B}\,\mathrm{diag}([n^2_{u,a_x}, n^2_{u,d_x}, n^2_{u,a_y}, n^2_{u,d_y}, n^2_{u,a_z}, n^2_{u,d_z}])\mathbf{B}^\top$
    $\mu_1 = \mathbf{0}_{6\times1}, \ \mathbf{\Sigma}_1 = \mathbf{0}_{6\times6}$
    for $k \in [2, N]$ do
        $\bar{\mu}_k = \mathbf{F}\mu_{k-1} + \mathbf{B}\mathbf{u}_k$
        $\bar{\mathbf{\Sigma}}_k = \mathbf{F}\mathbf{\Sigma}_{k-1}\mathbf{F}^\top + \mathbf{Q}$
        $\mathbf{R} = \mathrm{diag}([n^2_{z_{x,k}}, n^2_{z_{y,k}}, n^2_{z_{z,k}}])$
        $\mathbf{K}_k = \bar{\mathbf{\Sigma}}_k\mathbf{H}^\top(\mathbf{H}\bar{\mathbf{\Sigma}}_k\mathbf{H}^\top + \mathbf{R})^{-1}$
        $\mu_k = \bar{\mu}_k + \mathbf{K}_k(\mathbf{z}_k - \mathbf{H}\bar{\mu}_k)$
        $\mathbf{\Sigma}_k = (\mathbf{I}_{6\times6} - \mathbf{K}_k\mathbf{H})\bar{\mathbf{\Sigma}}_k$
    end for
return $\mu$
end procedure
```

where k is the time step and g is gravitational acceleration (magnitude). World-frame velocity and specific acceleration were fused to create a filtered velocity signal ($\hat{\mathbf{v}}_w$) using a drift-correcting Kalman filter (Alg. 5.1). Odometry position ($\hat{\mathbf{p}}_{od}$) is computed by numerically integrating ($\hat{\mathbf{v}}_w$) via trapezoidal sum. Full odometry pose is then represented by the combination of odometry position and animal body orientation at each time step k:

$$\mathbf{P}_{od,k} = \begin{bmatrix} \mathbf{R}_k^{(3\times3)} & \hat{\mathbf{p}}_{od,k}^{(3\times1)} \\ \mathbf{0}^{1\times3} & 1 \end{bmatrix} \tag{5.2}$$

Notes on algorithm inputs and component variables:

- $\mathbf{u}$: Primary input array of acceleration data $\mathbf{A}_s$. Format: $\mathbf{u} = [\mathbf{u}_x, 0, \mathbf{u}_y, 0, \mathbf{u}_z, 0]^\top$.
- $\mathbf{z}$: Sensor update array of velocity data $\mathbf{v}_w$. Format: $\mathbf{z} = [\mathbf{z}_x, \mathbf{z}_y, \mathbf{z}_z]$.
- $\mathbf{n}_{u,a}$: Noise standard deviation for input array. Format: $\mathbf{n}_{u,a} = [\mathbf{n}_{u,a_x}, \mathbf{n}_{u,a_y}, \mathbf{n}_{u,a_z}]$.
- $\mathbf{n}_{u,d}$: Drift standard deviation for input array. Format: $\mathbf{n}_{u,d} = [\mathbf{n}_{u,d_x}, \mathbf{n}_{u,d_y}, \mathbf{n}_{u,d_z}]$.

- $\mathbf{n}_z$: Noise standard deviation for sensor update. Format: $\mathbf{n}_z = [\mathbf{n}_{z_x}, \mathbf{n}_{z_y}, \mathbf{n}_{z_z}]$.
- T: Iteration period (seconds).
- μ: Output state vector. Format: $\mu = [\mathbf{v}_x, \mathbf{v}_{x,\delta}, \mathbf{v}_y, \mathbf{v}_{y,\delta}, \mathbf{v}_z, \mathbf{v}_{z,\delta}]$. $[\mathbf{v}_x, \mathbf{v}_y, \mathbf{v}_z]$ represent the velocity components and $[\mathbf{v}_{x,\delta}, \mathbf{v}_{y,\delta}, \mathbf{v}_{z,\delta}]$ represent the velocity drift components.

5.1.2 Animal Tracking

The same validation and training data as in Section 3.1.2.3 were used to train updated Faster R-CNN object detectors (one for each camera) using the second classifier structure from Section 3.1.1. Bounding-box conflicts between overlapping detections were resolved and false-positives were removed using the methods found in Section 3.1.2.4. This portion of the framework did not change significantly from the approach in Section 3.1, with the most meaningful modification being made to the CNN classifier structure for more robust animal detection.

5.1.3 Depth-Based Position Correction and Particle Filtering

Just as in Section 3.1, animal detected locations were subject to systematic errors due to depth-based camera perspective and optical refraction effects. The method for correcting these errors follows the same concept as in Section 3.1.2.6. However, in this case tag-based depth readings are present to enable exact position corrections. These corrections were applied to all detections regardless of tagging condition: since only the tagged animal is being tracked by the framework, erroneous position corrections applied to other animals in the area are irrelevant. As MTags are used in this framework, two sources of error in tag-estimated depth must be addressed: 1) errors due to animal in-place rotations, and 2) errors due to the fluid pressure drop during dolphin swimming. These were corrected using the method derived in Section 4.1.4.1.

The following method was devised and executed by D. Zhang[1]*, and is briefly described here for completeness.* Depth-corrected animal detections were fused with 2D (world-frame x-y) odometry data in a non-causal particle filter [128]. The odometry was sub-sampled to match the sampling frequency of the detection data. Randomly instantiated particles were weighted ("rewarded") based on their proximity to a detection within a general radius of attraction (4.5 meters). Particles were then propagated forward in time according to the yaw and speed of the x-y components of the odometry. This reward-propagate process would continue with particles being resampled if differences in weights surpassed a heuristically-defined threshold. The estimated position of the animal

---

[1] In: D. Zhang, J. Gabaldon, L. Lauderdale, M. Johnson-Roberson, L. J. Miller, K. Barton, and K. A. Shorter, "Localization and Tracking of Uncontrollable Underwater Agents: Particle Filter Based Fusion of On-Body IMUs and Stationary Cameras," *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 6575-6581, doi: 10.1109/ICRA.2019.8794141.

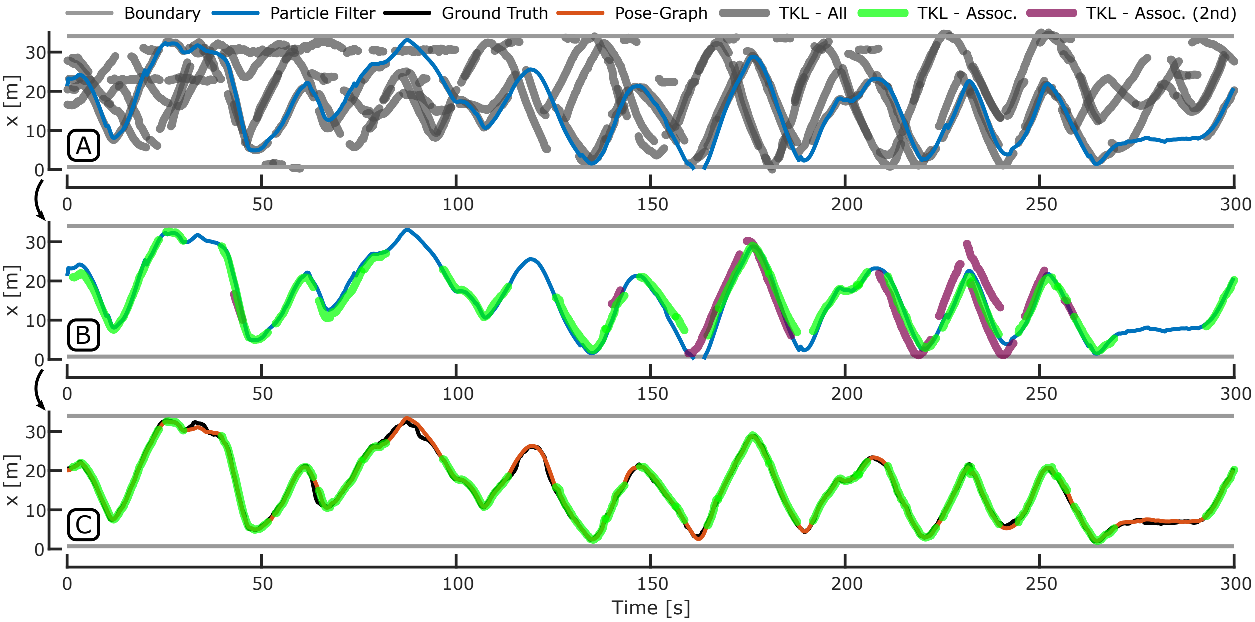


Figure 5.2: Demonstration of the tracklet association process. A: Particle filter *x*-position versus time (blue), overlaid with all existing tracklets (gray) during the five-minute subsection of a processed track. As there can be up to seven animals in the main habitat at one time, the association tool must be able to contend with all of these at once. B: Particle filter *x*-position versus time, overlaid with all potentially associated tracklets. Confirmed associations (post tie-breaking) are shown in green and discarded (secondary) associations are shown in purple. C: Final pose-graph *x*-position (red) and ground-truth (black) versus time, overlaid with the confirmed associated tracklets. This demonstrates how the associated tracklets, when present, enable the localization framework to closely follow the true animal position.

at each time step was defined according to the weighted mean of particle locations. As this computation was performed entirely in-post, particles could receive additional priority dependent on the magnitude of rewards they would receive in subsequent time steps. Any particles that were prioritized this way were defined as “non-causal” and were used in a separate weighted mean for a secondary, more accurate track estimate. The work in this chapter uses the non-causal filtering approach for enhanced tracking performance.

5.1.4 Detection Association

This section represents one of the primary method contributions of the chapter: building the localization framework mainly involved the integration of methods previously detailed in Chapters 3 and 4, and what follows is a novel method required to link animal tag data with the animal’s corresponding localizations. To associate specific detections with the tagged animal, they were first temporally linked through the process of tracklet generation as described in Section 3.1.2.5.

Tracklet generation provides the benefit of linking localizations that correspond to a single animal; however, they cannot identify the animal. A further step was required to associate individual tracklets to the tagged animal, which is not a trivial endeavor when multiple animals are present in the monitored environment (Fig. 5.2, A). Applying the particle filter reduced localization drift while preserving the spatial shape of the tagged animal's movements. This made it possible to associate tracklets to the tagged animal by comparing tracklet shape to the shape of the particle filter output track. Segments of the particle filter track were extracted at time intervals corresponding to each tracklet's time interval. As the particle filter was set to match the detection data sampling frequency, each extracted track and corresponding tracklet had an equal number of points. For each comparison, x-y particle filter and tracklet location data were first conditioned to reduce scale discrepancies caused by short-term drift. Conditioning was performed as follows, for a 2D position sequence $\mathbf{p} \in \mathbb{R}^{2\times n}$, where n is the number of points in the sequence:

$$\mathbf{o} = [\text{mean}(\mathbf{p}_x), \text{mean}(\mathbf{p}_y)] \tag{5.3}$$

$$s = \sqrt{2}/\text{mean}\left(\sqrt{(\mathbf{p}_x - o_x \cdot \mathbf{1}^{1\times n})^2 + (\mathbf{p}_y - o_y \cdot \mathbf{1}^{1\times n})^2}\right) \tag{5.4}$$

$$\mathbf{T} = \begin{bmatrix} s & 0 & -o_x s \\ 0 & s & -o_y s \\ 0 & 0 & 1 \end{bmatrix} \tag{5.5}$$

$$\bar{\mathbf{p}} = \mathbf{T}[\mathbf{p}^\top, \mathbf{1}^{n\times 1}]^\top \tag{5.6}$$

where the first two rows of $\bar{\mathbf{p}}$ are the conditioned x and y location values, respectively. This conditioning method was sourced from the normalized Direct Linear Transform [109], and was found to aid in tracklet association robustness in this research.

Conditioned track and tracklet data were then aligned using the Iterative Closest Point (ICP) scan-matching method [129]. Figure 5.3 displays four examples of conditioned particle filter track segments and tracklets both before and after scan-matching. The track segments were used as the point models to which the tracklets were matched. Tracklets were discarded if they failed a set of threshold requirements that evaluated the differences from their corresponding track segments. Each tracklet was evaluated according to four metrics, the thresholds heuristically determined:

1. Distance: The Root Mean Square (RMS) distance between a tracklet and its track segment (before conditioning) was computed, with an allowable maximum of 10 meters. This ensured the overall distance between point sets was not extreme (e.g. opposite ends of the habitat).

2. Root Mean Square Error (RMSE): The x-y position RMSE of the conditioned tracklet as compared to its corresponding track segment was computed, both before and after scan-matching, with an allowable maximum of 0.7 meters. The minimum of the two was used, as

in rare cases circular movement patterns could cause the ICP process to arbitrarily rotate the tracklet, yielding excessive errors. This ensured the overall shapes matched.

3. Rotation: The 2×2 rotation matrix output from the ICP scan-matching was converted to its corresponding angular rotation, with an allowable maximum magnitude of 15°. The rotation metric was only checked if the post-ICP RMSE was used, as otherwise circular tracklets could experience a false rejection. This ensured the tracklet could not be a similar shape but a completely different orientation compared to the track segment.

4. Travel Direction: The conditioned track segment and tracklet position data were numerically differentiated and passed through the `atan2` function to obtain their estimated directions of travel between each time step. The same tracklet position data that produced the minimum for the RMSE metric were used in this computation for consistency. Tracklet data were smoothed pre-differentiation using a 1 second window Savitsky-Golay filter to reduce noise. The RMS difference between these was then computed, with an allowable maximum of 60°. This ensured that the tracklet and particle filter segment were not traveling in opposing directions, as ICP does not explicitly account for point ordering.

Tracklets satisfying these requirements were placed into a set of potential associations (Fig. 5.2, B). Dolphins do engage in paired swimming, and as a result there can exist tracklets of similar shape and position occurring simultaneously. It is also possible for similarly-shaped simultaneous movement patterns to occur purely by chance when there are multiple dolphins in the same habitat. Potential associations without temporal overlaps were placed into a set of confirmed associations.

When determining overlaps, special cases were handled where a potentially associated tracklet had a temporal overlap with another tracklet less than 1 second long and under 25% of its total duration. The overlapping section was removed, and the two tracklets were re-labeled as not overlapping, preserving any other overlaps with other tracklets. This could occur when another tracklet was instantiated in close proximity to the tracklet in question, and erroneously hijacked what would be the rest of the tracklet. This would end the first tracklet prematurely with both tracklets corresponding to the same animal, and a short overlap between the two.

After handling the special cases, tie-breakers were performed to identify the most-likely associated tracklets among the overlapping subset. Potential associations were scored according to how close they came to surpassing each threshold for the distance, RMSE, and travel direction metrics. The total score for a potential association was computed as the mean of the metric fractions (computed metric value divided by its maximum allowable threshold). Overlapping tracklets with the lowest scores among their overlap group were placed into the confirmed set.

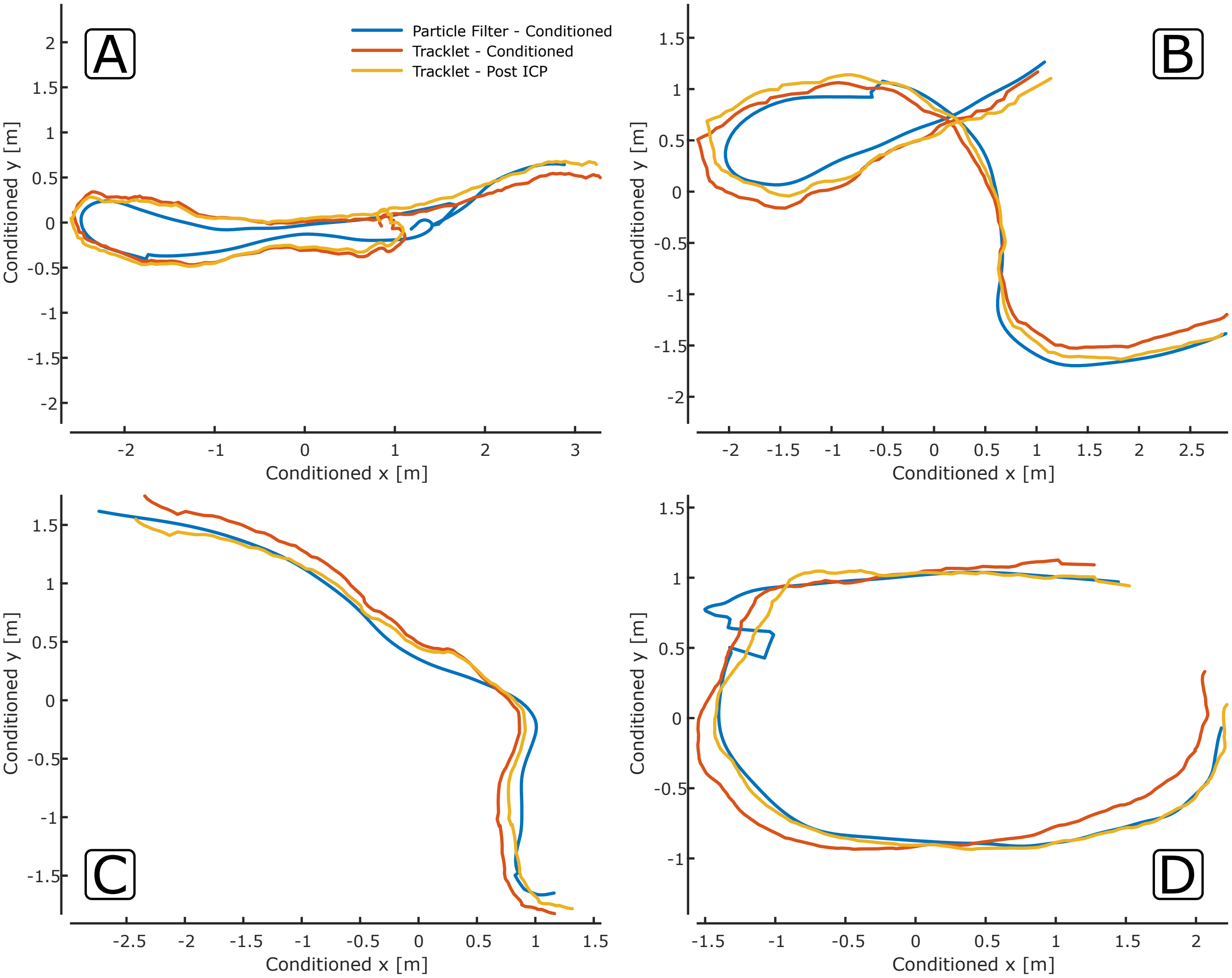


Figure 5.3: Four examples of conditioned particle filter, tracklet, and post-ICP scan-matched tracklet position data. All four tracklets were identified as associated with the tagged dolphin. These tracklets were selected to demonstrate the flexibility of the method, which is able to associate track shapes with high eccentricity (A), fully enclosed narrow loops (B), simple linear progression (C), and circular wide loops (D).

5.1.5 Pose-Graph Optimization

The full 3D odometry ($\mathbf{P}_{od}$) was fused with the associated tracklet position data through the iSAM pose-graph optimization method [44], as implemented in the GTSAM MATLAB toolbox [130]. As the environmentally-mounted cameras did not offer a full stereo view of the habitat, the camera data could not provide depth information. The MTag-based depth data were used as a substitute, and when paired with the associated tracklet data provided absolute position estimates used in the pose-graph localization method. To evaluate the performance of the individual tracking steps (odometry, particle filtering, and pose-graph optimization), the x-y positions of each were compared to manually-tracked x-y position data by computing their 2-norm error values.

5.2 Experimental Deployment

5.2.1 Experimental Setup

One bottlenose dolphin from a group of seven individuals was outfitted with an MTag biologging device (Section 4.1.1) and recorded in video in the Seven Seas dolphinarium at the Brookfield Zoo, Brookfield, IL using the same camera setup as described in Section 3.1.2. Recordings were captured at 20 Hz in the RGB-24 color format and object detection was performed at 10 Hz to reduce computation time. The dataset used to validate the framework performance was recorded during a tagging session in September 2018, which contained two segments (18 and 19 minutes each) where the tagged dolphin was continuously present in the portion of the habitat monitored by the cameras. Ground-truth data were generated by hand-tracking the tagged dolphin in a panoramic stitched video that combined the two camera views (e.g. Fig. 5.1, A, bottom-right), with a tracking rate of 10 Hz. For the work performed in this chapter, an updated computer system was used for detection inference, and was outfitted with an AMD 5900X processor and a Nvidia RTX 3090 graphics processing unit to expedite the inference process. The MTag housing and internal hardware remained unchanged from Chapter 4.

5.2.2 Localization Results and Performance

For both segments, the pose-graph (PG) optimized track was more accurate than the particle filter (PF) result for all metrics (Table 5.1). Note: the z-axis was not considered when computing the errors as the ground-truth obtained from manual video annotation was only capable of providing x and y localizations, so in this case, the tag provides the sole depth estimate. Mean PG errors were less than half those of the PF, and median PG errors were roughly one-third of the PF results. PG error standard deviations were relatively high compared to the means, which is understandable

Table 5.1: Localization framework performance

Parameter		Segment 1	Segment 2
Duration [min]		18	19
Err_{PF}	Mean [m]	1.80	2.66
	Median [m]	1.55	2.24
	Std. Dev. [m]	1.21	1.97
Err_{PG}	Mean [m]	0.85	1.03
	Median [m]	0.58	0.63
	Std. Dev. [m]	0.79	1.11
Mean Err_{PG} – With TKL [m]		0.50	0.61
Mean Err_{PG} – No TKL [m]		1.33	1.52

Notation: *PF* corresponds to the particle filter track and *PG* to the pose-graph track.

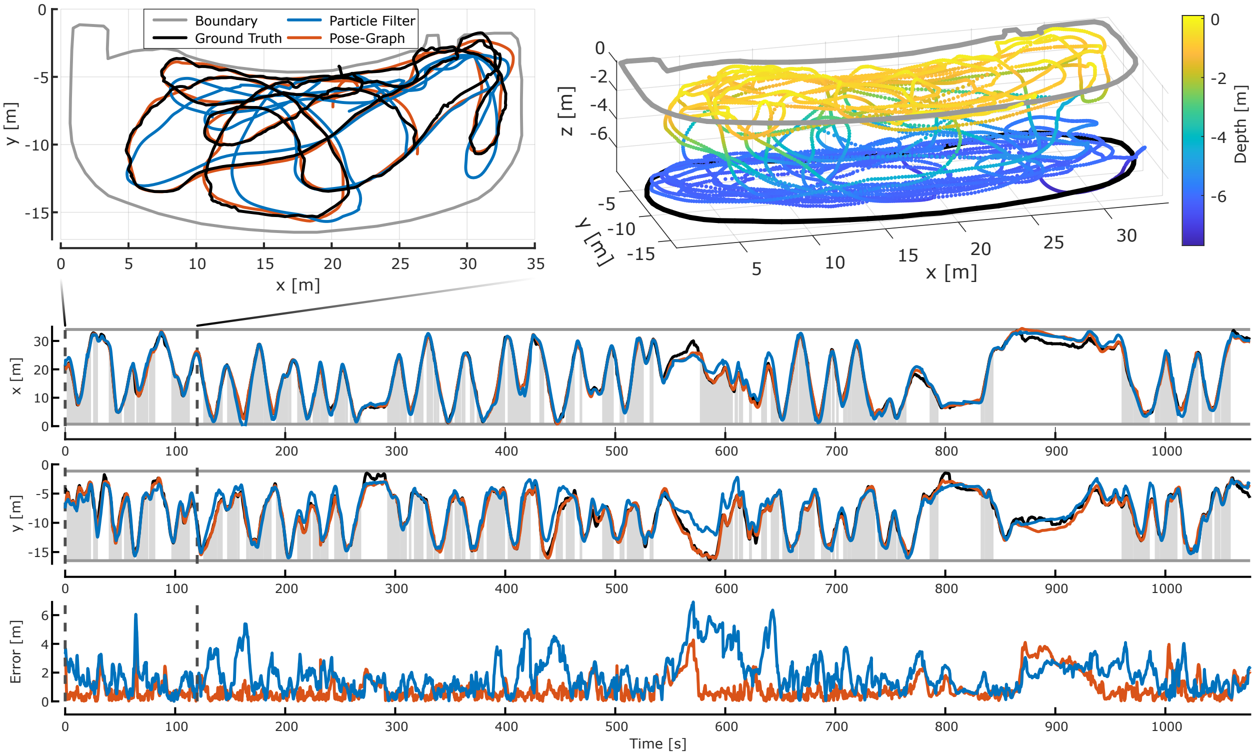


Figure 5.4: Comparison of particle filter, pose-graph, and ground-truth tracks for Segment 1. TOP-RIGHT: 3D view of the dolphin's trajectory for this segment, color-coded according to depth for visibility. The upper edge of the habitat is lined in gray, and the bottom edge in black. BOTTOM: x and y position versus time results for the particle filter (blue) and pose-graph (red) tracks as they compare to the ground-truth positions (black), with the habitat bounds marked in gray. Shaded regions under the x and y plots represent times where animal-associated tracklets were present. The tracking error for each position estimate is shown in the final plot, demonstrating the overall performance advantage for the pose-graph result over the particle filter. TOP-LEFT (POPOUT): Spatial x-y comparison between the particle filter and pose-graph tracks as they compare to the ground-truth for a 2-minute subset of the data at the beginning of the segment.

when observing Figure 5.4: tracking error only significantly increased when no tracklets were detected for the animal (absence of shaded regions in the plot) for tens of seconds at a time, and these spikes in error contributed to the higher standard deviation values. This is to be expected, as the odometry is guaranteed to drift over time, and is the driving component in the localization process when no external observations are present to anchor the track. This indicates that improvements to the framework's tracking performance would primarily require further advancements in odometry estimation, as mean errors when the tracklets were present are well below 1 meter, versus more than double when no animal-associated tracklets appear (Table 5.1).

The dolphin traversed the majority of the habitat throughout Segment 1, primarily focusing on swimming in the upper third (42% of the time) and at the bottom third (51%), with the middle third (7%) of the environment primarily reserved for transient actions such as diving and surfacing (Fig. 5.4, top-right). Despite the segment only being 18 minutes in length, animal exploration for this period was relatively thorough. However, to produce generalized statistics on habitat use and related animal behavior metrics similar to those in Section 3.1, significantly more recording sessions are required.

5.3 Contextualized Monitoring

To provide an example of the potential provided by contextualizing animal swimming behavior via localization, animal COT (non-metabolic) was calculated using the approach in Chapter 4 and analyzed with respect to its track, using Segment 1 as a case study. Cost of Transport is an important metric as it explicitly tracks an animal's locomotive efficiency. Monitoring COT then provides an understanding of the propulsive effort ramifications of an animals' chosen gait at a given time. This can inform researchers on correlations between different animal behavioral states and power modulation, indicating which situations result in an animal prioritizing efficiency, and which do not. An extreme case of varying efficiency prioritization was explored in Chapter 4, where an animal avoided parasitic drag during high-intensity directed swimming tasks, and made no such effort when in low-intensity free-swimming. COT analysis can also provide a more nuanced view of how animals approach propulsion according to behavioral sub-states. For example, paired versus unpaired swimming (i.e. when an animal is alongside/tailing another; occurs during free-swimming), may have differing effort profiles due to the effects of the social interactions between that particular pair of dolphins. It is then important to verify the significance of such interactions on animal behavior, and quantify how these effects manifest themselves.

Figure 5.5 displays a temporal view of animal position throughout Segment 1, and has been color-coded according to the animal's mass-specific COT (Eq. 4.9). Note: while the dolphin achieved a peak of $3.75\ \mathrm{J/[kg\,m]}$, this was a transient shallow-water high-speed activity spike

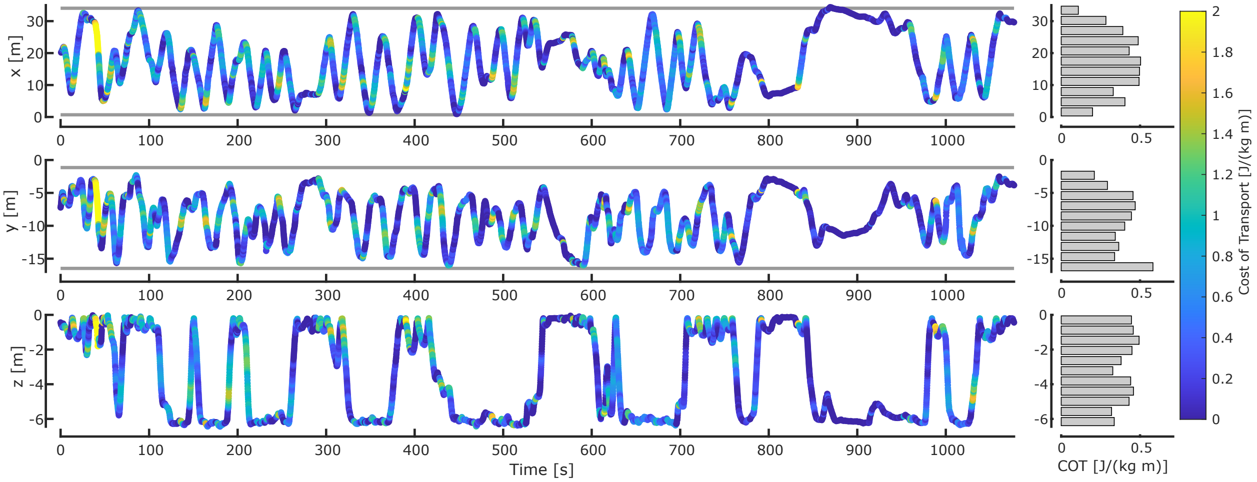


Figure 5.5: Segment 1 pose-graph animal track, color-coded according to swimming COT. This provides a temporal view of animal swimming effort with respect to 3D position, allowing for qualitative analyses of swimming behavior trends. Additionally, mean animal COT is discretized according to each axis (gray bars). The animal engaged in looping swimming patterns throughout the segment, with COT generally rising during the central portions of the x and y axes. Swimming effort increased in shallow water during the diving portions of its short swimming bouts. During transitions between shallow and deep regions, if a COT spike was observed it tended to occur at the beginning of each long dive/ascent. COT peaks in deep water were short but sporadic, and tended to occur 1 meter above the bottom of the environment.

within the first minute of the track. All other values in the segment fell below 2.0 J/[kg m], so the colormap was capped at this value to better differentiate COT peaks and troughs throughout the entire segment. The dolphin engaged in paired swimming with another animal for the majority of Segment 1, interacting with floating enrichment for the first 100 seconds and briefly (~ 5 s) right before the 10-minute mark. These behaviors were manually observed in the video recording.

Despite the tagged dolphin exhibiting consistent looping patterns in the x-y plane (varying amplitude sinusoidal motion in the x and y axes) both at the surface and at the bottom of the habitat (Fig. 5.5), temporal patterns in COT primarily correlated to depth, rather than any particular phase in an x-y loop. The most predictable pattern occurred in the top third of the habitat, where the animal would increase swimming effort immediately after respiration during a surfacing event as it dove back into the water. This aligns with the "respiration → dive → fluke → glide → repeat" gait pattern seen in literature [67]. Overall COT was higher in the top third of the habitat (0.65 ± 0.50 std. dev. J/[kg m]) than the bottom third (0.46 ± 0.37 std. dev. J/[kg m]). Figure 5.6 aids in this visualization, presenting average animal propulsive COT according to spatial location for both the top and bottom thirds of the environment (middle is not shown due its low use rate). The environment was subsampled into 1×1 meter squares similar to Section 3.1.3, however in this case it is now possible to differentiate animal swimming behavior according to depth condition.

The most likely explanation for the discrepancy in COT between the top and bottom thirds of

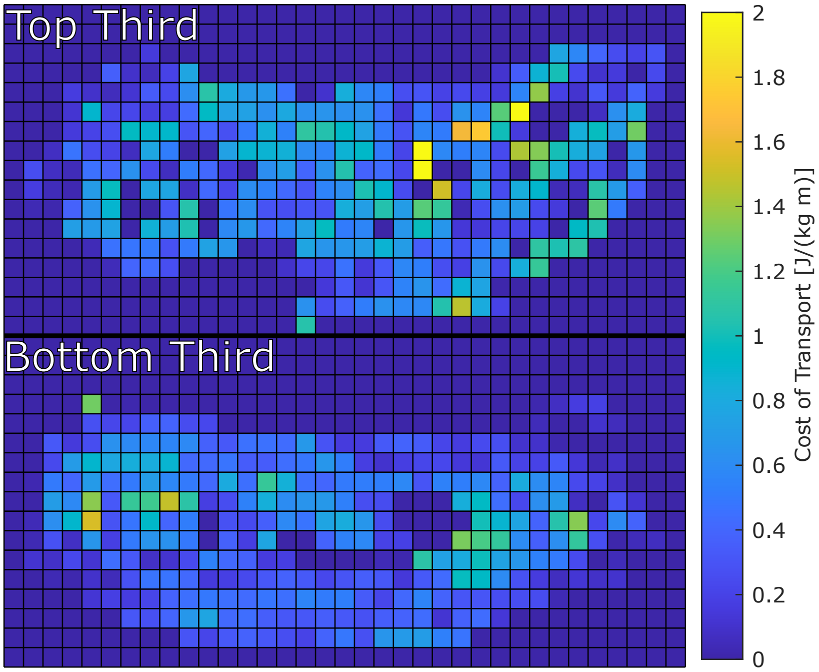


Figure 5.6: Animal mean COT with respect to position for top and bottom thirds of the habitat. Each square represents a 1 × 1 meter area. Top third COT values are higher overall, though the bottom third sees more even habitat coverage. The middle third was neglected due to its low (7%) usage rate, only employed by the dolphin for transitions between the other two regions.

the habitat can be given through a combination of two factors: wave drag and oxygen availability. As shown in Chapter 4, wave drag drastically increases an animal's required propulsive effort dependent on the animal's proximity to the surface, consequently decreasing its efficiency. Further, an animal can afford to sacrifice efficiency at the surface as it has quick access to a fresh supply of oxygen. However, these factors may not fully explain the spatial patterns observed in Figure 5.6: the tracked animal engaged in paired swimming (with the same conspecific) regardless of depth condition, yet the spatial spread of propulsive efficiencies varied drastically dependent on where in the water column the swimming took place. The top third displayed a near-random COT distribution with large scattered spikes, while the bottom third displayed more even coverage along the edges of the environment. Despite the animal being in the same social state for both depth regions, there was a fundamental shift in swimming behavior as the animal switched between the two. This is supported through manual inspection of the video data: shallow swimming entailed more varied movement patterns, and the opposite was true for deep water swimming. Further recording sessions are required to determine the significance of this observation.

5.4 Future Work and Conclusions

The localization framework itself represents the primary contribution of this chapter, and thus far has demonstrated its capability in providing across-the-board improvements in tracking precision and accuracy, in some cases achieving legitimately sub-1-meter errors. This is not possible even when using state-of-the-art localization techniques originally developed for the wild [24]. Further, it achieves continuous localization while preserving high-frequency animal kinematics information through the combination of tag and camera sensor fusion, which has not been feasible with existing camera-only indoor tracking solutions [28, 29].

The method can still be improved, however, with the introduction of additional hardware. The ability to achieve full stereo coverage of the monitored habitat would provide a secondary depth estimate, allowing for enhanced vertical accuracy during high-speed swimming: as the depth correction process relies on fluid speed information, any noise or lag in the signal has the potential to reduce the efficacy of the corrections. A secondary depth estimate could then provide additional robustness to edge cases, including out-of-water leaping height estimates, which are currently not modeled at all. The odometry performance also indicates that continuously improved speed and orientation estimation may also be required. While the speed sensor represents an advancement in accuracy and robustness compared to existing methods, it is limited by its uniaxial structure. As dolphins can drift slightly during banking, the assumption that velocity is primarily in the direction of the animal's orientation begins to break down as turns sharpen or speeds increase. The development of an omnidirectional speed sensor of similar or better performance versus the micro-turbine used in this dissertation would then enable more accurate animal odometry.

The contextualized monitoring example presented in this chapter is only intended to serve as a case study, and for the full potential of this framework to be realized a longer, more in-depth monitoring experiment is required. Further recording sessions with the same and additional dolphins are necessary to achieve statistically significant results on animal kinematics and energetics and how these metrics relate to location-specific stimuli. Consistent recording sessions can then yield insights into how behaviors vary throughout a typical day and across seasons, and the use of the framework presented in this chapter will enable higher precision than in Section 3.1, which attempted such a feat. Further, as this framework is equipped to perform animal identification, specific individual monitoring is possible. This has the potential to inform both researchers and managing institutions on animal engagement, biomechanics, and welfare, offering the opportunity to gain new insights into behavioral patterns and continuously refine institution management practices to safeguard animal mental and physical health.

CHAPTER 6

Conclusion

6.1 Contribution Summary

The limitations in marine animal monitoring inspired the contributions of this dissertation in pushing advancements to both hardware and computation methods. These in turn led to new insights on animal behaviors concerning kinematics, energetics, and habitat exploration. As existing methods have been primarily developed for use in wild animal monitoring, they do not abide by the constraints set by the managed environment. The contributions presented here have been designed to overcome these particular hurdles by adopting and blending methods from the field of robotics with those of marine mammal science.

Before any novel insights on animal behavior could be made, additional tools were necessary to enable this research. Chapter 3 presents two contributions of this nature: CNN-based visual localization, and a robust on-tag speed sensor. While Section 3.1 did not advance the field of CNN object detection itself, the unique method in this case concerned the structure surrounding the detector. By accounting for the limitations set by the environment and the complications of simultaneously tracking multiple targets, the object detection method was able to provide generalized statistics describing the behaviors of a population of bottlenose dolphins, gaining insights into environmental attractors of their attention and intra-day kinematics trends. This research provides a direct example of the integration of a method novel to the field of marine mammal science furthering our understanding of animal behavior.

In contrast, the speed sensor represents the introduction and validation of a tool that has not existed thus far. Section 3.2 proposes, calibrates, and verifies the performance of a miniaturized speed measurement device, intended to provide direct animal kinematic monitoring capabilities to biologging tags. The validation of the sensor's performance then lent confidence to the research relying upon it further into the dissertation.

Chapter 4 extends state-of-the-art physics-based animal energetics models [25], to account for the newfound kinematics monitoring capabilities of a speed-instrumented biologging tag. The result was a refined method to persistently monitor animal work and power without the use of external observers. This enabled a new look at dolphin swimming behaviors and how they relate to conditions requiring varying propulsive intensities, prompting an enhanced understanding of efficiency-driven behavior (or lack thereof). Further, the method presents new opportunities in long-term swimming effort analysis, which was demonstrated through the first persistently-monitored 24-hour range estimate for a bottlenose dolphin, in addition to a detailed quantitative view of its energetics fluctuations throughout this time.

The culmination of this dissertation is realized in Chapter 5 by combining the contributions in Chapters 3 and 4 to produce a high-fidelity 3D monitoring framework. By fusing biologging tag and recorded video data, fine-scale localization became possible for cetaceans in a highly dynamic multi-animal environment. This heavily relied on techniques originally developed in the field of robotics, in total combining robust IMU orientation estimation [10], Faster R-CNN object detection [15], particle filtering [128], ICP scan-matching [129], and pose-graph optimization [44]. The completed framework represents a contribution to the field of robotics, not just marine mammal science: while it was developed with the intent of tracking animals, it operates on the principle of combining on-agent and environmental sensor streams through the identification and localization of a target agent in a crowded environment. To demonstrate an example of the contextualized monitoring capability of the framework, animal COT was computed with methods from Chapter 4, and was analyzed with respect to the animal's trajectory.

6.2 Research Impacts

Both the individual tools (Chapter 3) and the enhanced monitoring methods (Chapters 4 and 5) have potential in implementations beyond the scope of this dissertation. While they are generally suited for use in managed settings, some extensions can be made for wild animal monitoring.

The CNN-based animal tracking method has already demonstrated its capabilities in not only localization, but general animal kinematics monitoring as well. This opens up a host of additional opportunities such as tracking trends in activity levels across years and seasons, or observing animal responses to changing environmental conditions (e.g. toy presence, human crowd densities, etc.). Further, such an approach can be deployed in any habitat where cameras can be placed with clear views of regions of interest. In general, camera-based monitoring is flexible as it does not explicitly require permanent hardware installations [29].

The ramifications of the speed sensor and the ensuing kinematics/energetics monitoring framework are applicable regardless of environmental condition (wild or managed), although there may be some limitations with compatible species. Since the MTag design functions as a standalone device, it is possible to be deployed on any animal that supports suction-cup adhesion. Similar to the DTAG line of devices [6, 23], the internal MTag hardware (e.g. IMU, temperature, pressure) will function regardless of orientation on the animal. However, as the speed sensor is uniaxial, its performance will suffer when not aligned with the animal's primary direction of motion. The methods defined in Chapter 4 *can* function for any animal the MTag can be stuck to, though this will require careful placement and orientation, which is not possible when tagging larger cetaceans (i.e. whales in the wild). Therefore, while the energetics monitoring approach can apply to any setting, it is best suited for wild tagging studies where animals can be captured and released, or managed studies where animal cooperation is possible.

The benefits of the energetics monitoring framework are not restricted to answering purely biological questions. As marine mammal science can be inspired by robotics, so too can robotics rely on bio-inspired design. Cetaceans are known to be highly efficient swimmers [26], and understanding how they modulate their fluking frequency, amplitude, and duty cycle (for fluke-and-glide propulsion) can inform physical designs on potential opportunities in advancing underwater robot efficiency and mobility. By emulating highly adapted marine animals, there is potential to further the field of robotics in turn.

In its current state, the 3D monitoring framework is primarily suited for managed settings due to the limitations of camera performance in the wild. Despite it requiring the most setup, involving both tagging and camera recordings, this framework does offer the richest possible options in animal kinematics and localization analysis. Deploying the framework in a managed habitat can enable researchers and ACSs to understand how swimming animals use their environment with exacting detail, while allowing them to process data on a scale otherwise untenable with manual annotation methods. This has the potential to inform highly in-depth studies on specific behavior patterns, habitat region use, location-specific kinematics trends, and in the case where multiple animals are tagged at once, behavioral effects of social interactions with conspecifics.

Beyond the scope of marine mammal science, the 3D monitoring framework also has potential in localizing sensor-limited mobile robots. Miniaturized robot platforms may not support a full suite of self-localizing sensors (e.g. Light Detection and Ranging (LIDAR), multi-camera arrays) due to size or power constraints. To address this, a customized version of the framework could allow odometry-enabled robot localization by combining their dead-reckoning data with drift-correcting absolute position updates from environmentally-mounted sensors. This would require that individual identification still be possible through associating robot odometry with partially-continuous external sensor-based tracks (in line with Section 5.1.4). Sensor-limited robots could then have higher accuracy trajectory estimates without the introduction of additional onboard hardware beyond the minimum requirements to perform odometry.

6.3 Future Extensions

While the contributions in this dissertation offer general performance improvements in localization and new opportunities overall in kinematics and energetics monitoring, there are still improvements that can be made and additional paths to explore. Modifications range from general advancements in the capabilities of the monitoring frameworks to the full decoupling of tag and external camera sensor streams for animal localization.

With respect to the CNN object detection method, one clear revision can be the introduction of full stereo vision. As the system currently in place at the Brookfield Zoo uses two cameras with minimally-overlapping FOVs, the use of additional sensors or wide-angle lenses offers the potential for complete localization [29], rather than the current method of x-y position with uncertain depth. Further, the modification of the CNN classifier to handle additional dynamic environmental elements such as humans or enrichment will offer the potential to observe behaviors as they relate to the evolving habitat. These changes would provide localization performance benefits and additional behavior monitoring options without drastically changing the structure of the method.

The energetics monitoring framework has the potential for improvements in both hardware and its mathematical model. The most important hardware modification would come in the form of a new speed sensor. While the current sensor's performance is robust, its uniaxial design limits its capabilities in monitoring larger animals in the wild where device placement cannot be guaranteed. An omnidirectional sensor would then offer the potential for kinematics and energetics tracking regardless of orientation, which additionally guards against device shifting that may occur over long recording sessions, where the tag's suction cups can slide along the animal during highly dynamic actions or in the event of cup partial pressure loss. This cascades into improvements for the localization framework as well, which can experience disruptions in the event of a tag slide.

In terms of mathematical model, modifications can be made to the computation of animal drag coefficients. As the current model relies on steady-state swimming kinematics [25, 41], this disregards the fact that cetaceans commonly engage in fluke-and-glide swimming for propulsive efficiency benefits [67], and in this case the gliding component would rely on a separate drag model. Employing a multi-state drag model would then enhance the power estimate accuracy. Separately, an additional complication appears when attempting to estimate full animal metabolic expenditures, as these rely on RMR estimates to provide a significant portion of the total energy rate (Section 4.1.3.1). However, these values vary according to feeding patterns [123] and activity levels as a result of thermoregulatory requirements [37]. Accurately estimating true metabolic rate then requires a more sophisticated RMR model.

Overall, the hardware changes proposed to aid the CNN object detection technique and the energetics monitoring framework can improve the performance of the localization framework as well. The inclusion of depth estimates from stereo camera data provides more information for the pose-graph optimizer to access, and an omnidirectional speed sensor can offer improved odometry for lower drift when no camera detections are available. However, this still requires that the framework rely on environmentally-mounted cameras to provide absolute position estimates, which has several drawbacks:

1. Occlusions in the environment, such as water-surface glare or other animals, can directly impede animal localization resulting in incomplete camera information.
2. Water opacity and surface ripples prevent individual animal identification due to image obfuscation, requiring additional methods to associate portions of camera track data with the tagged animal.
3. Environmentally-mounted cameras are ineffectual in the wild due to limited sight lines. Instrumenting a wild habitat with cameras would provide anecdotal information as animals would only appear in-frame when in close proximity. Otherwise, a prohibitively large number of cameras would be required for more complete environmental coverage.

One potential solution would be to forgo environmentally-mounted cameras in their entirety, and instead adopt absolute position estimation into the tag itself. This is already attempted through GPS-based technologies [8], although such localizations are only possible when an animal has surfaced, yielding no potential for drift correction underwater. This problem has been studied for autonomous underwater vehicle localization, and involves performing Simultaneous Localization and Mapping (SLAM) on camera-generated maps of the environment [9]. Theoretically, it is possible to localize a marine animal in this way through the inclusion of one or more cameras into a biologging tag. However, this will not work in open water due to habitat visibility constraints, and as such would be relegated to regions with exceptionally clear water or shallow depths (e.g. Sarasota, Florida, USA). This method would invariably perform more optimally in managed environments due to their relatively shallow habitats and limited spatial scales. Regardless, such an endeavor would render environmentally-mounted camera localization unnecessary, simplifying the animal monitoring process while providing additional information as it explores its habitat.

While it is clear there is much work to be done in this field, much of it just to *enable* animal observations, with a wide-reaching multidisciplinary approach it is possible to overcome the challenges put forth by the environment itself.

BIBLIOGRAPHY

[1] M. R. Payne, "Growth of a fur seal population," *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, vol. 279, no. 963, pp. 67–79, 1977.
(Back reference pages: iv, 2, and 8)

[2] R. S. Wells and M. D. Scott, "Estimating bottlenose dolphin population parameters from individual identification and capture-release techniques," *Report - International Whaling Commission, Special Issue*, vol. 12, pp. 407–415, 1990.
(Back reference pages: iv, 2, and 8)

[3] B. Wilson, P. S. Hammond, and P. M. Thompson, "Estimating size and assessing trends in a coastal bottlenose dolphin population," *Ecological Applications*, vol. 9, no. 1, pp. 288–300, 1999.
(Back reference pages: iv, 2, and 7)

[4] P. B. Best, "TRENDS IN THE INSHORE RIGHT WHALE POPULATION OFF SOUTH AFRICA, 1969–1987," *Marine Mammal Science*, vol. 6, no. 2, pp. 93–108, apr 1990. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1748-7692.1990.tb00232.xhttps://onlinelibrary.wiley.com/doi/abs/10.1111/j.1748-7692.1990.tb00232.xhttps://onlinelibrary.wiley.com/doi/10.1111/j.1748-7692.1990.tb00232.x
(Back reference pages: iv, 2, and 7)

[5] M. A. McDonald and C. G. Fox, "Passive acoustic methods applied to fin whale population density estimation," *The Journal of the Acoustical Society of America*, vol. 105, no. 5, pp. 2643–2651, may 1999.
(Back reference pages: iv, 2, 8, and 14)

[6] M. P. Johnson and P. L. Tyack, "A digital acoustic recording tag for measuring the response of wild marine mammals to sound," *IEEE Journal of Oceanic Engineering*, vol. 28, no. 1, pp. 3–12, 2003.
(Back reference pages: iv, 2, 10, 16, and 107)

[7] C. Ware, R. Arsenault, M. Plumlee, and D. Wiley, "Visualizing the underwater behavior of humpback whales," *IEEE Computer Graphics and Applications*, vol. 26, no. 4, pp. 14–18, jul 2006.
(Back reference pages: iv, 4, and 15)

[8] P. J. Wensveen, L. Thomas, and P. J. Miller, "A path reconstruction method integrating dead-reckoning and position fixes applied to humpback whales," *Movement Ecology*, vol. 3, no. 1, pp. 1–16, 2015. http://dx.doi.org/10.1186/s40462-015-0061-6
(Back reference pages: iv, 2, 4, 10, 17, and 109)

[9] A. Kim and R. M. Eustice, "Real-time visual SLAM for autonomous underwater hull inspection using visual saliency," *IEEE Transactions on Robotics*, vol. 29, no. 3, pp. 719–733, 2013.
(Back reference pages: v, 5, 17, and 109)

[10] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," *IEEE International Conference on Rehabilitation Robotics*, 2011.
(Back reference pages: v, 5, 17, 72, 73, 91, and 106)

[11] A. Cavallo, A. Cirillo, P. Cirillo, G. De Maria, P. Falco, C. Natale, and S. Pirozzi, "Experimental comparison of sensor fusion algorithms for attitude estimation," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, vol. 19, no. 3. IFAC Secretariat, jan 2014, pp. 7585–7591.
(Back reference pages: v, 17)

[12] R. G. Valenti, I. Dryanovski, and J. Xiao, "Keeping a good attitude: A quaternion-based orientation filter for IMUs and MARGs," *Sensors (Switzerland)*, vol. 15, no. 8, pp. 19 302–19 330, aug 2015. www.mdpi.com/journal/sensorsArticle
(Back reference pages: v, 17)

[13] Á. Odry, R. Fullér, I. J. Rudas, and P. Odry, "Kalman filter for mobile-robot attitude estimation: Novel optimized and adaptive solutions," *Mechanical Systems and Signal Processing*, vol. 110, pp. 569–589, sep 2018.
(Back reference pages: v, 17)

[14] S. O. Madgwick, S. Wilson, R. Turk, J. Burridge, C. Kapatos, and R. Vaidyanathan, "An Extended Complementary Filter for Full-Body MARG Orientation Estimation," *IEEE/ASME Transactions on Mechatronics*, vol. 25, no. 4, pp. 2054–2064, aug 2020.
(Back reference pages: v, 17)

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, jun 2017.
(Back reference pages: v, 22, 27, and 106)

[16] K. A. Shorter, A. Arbor, T. Hurst, W. Hole, and M. Johnson, "Finalizing the DTAG : Implementation and Testing of Design Improvements for Reliability and Availability," Office of Naval Research, Tech. Rep., sep 2014. http://www.dtic.mil/docs/citations/ADA615982
(Back reference pages: vi, x)

[17] P. B. Best, *Exploitation and recovery of right whales, Eubalaena australis, off the Cape Province*. Division of Sea Fisheries, 1970.
(Back reference pages: 2, 7)

[18] P. S. Hammond, P. Berggren, H. Benke, D. L. Borchers, A. Collet, M. P. Heide-Jørgensen, S. Heimlich, A. R. Hiby, M. F. Leopold, and N. Øien, "Abundance of harbour porpoise and other cetaceans in the North Sea and adjacent waters," *Journal of Applied Ecology*, vol. 39, no. 2, pp. 361–376, apr 2002. http://www.ascobans.org
(Back reference pages: 2, 7)

[19] W. J. Richardson, C. R. Greene, C. I. Malme, and D. H. Thomson, "CHAPTER 3 - MEASUREMENT PROCEDURES," in *Marine Mammals and Noise*, W. J. Richardson, C. R. Greene, C. I. Malme, and D. H. Thomson, Eds. San Diego: Academic Press, 1995, ch. 3, pp. 33–58. https://www.sciencedirect.com/science/article/pii/B9780080573038500069
(Back reference pages: 2, 8)

[20] S. M. Wiggins, M. A. McDonald, and J. A. Hildebrand, "Beaked whale and dolphin tracking using a multichannel autonomous acoustic recorder," *The Journal of the Acoustical Society of America*, vol. 131, no. 1, pp. 156–163, jan 2012.
(Back reference pages: 2, 8, and 14)

[21] M. Poupard, M. Ferrari, J. Schluter, R. Marxer, P. Giraudet, V. Barchasz, V. Gies, G. Pavan, and H. Glotin, "Real-time Passive Acoustic 3D Tracking of Deep Diving Cetacean by Small Non-uniform Mobile Surface Antenna," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2019-May. Institute of Electrical and Electronics Engineers Inc., may 2019, pp. 8251–8255.
(Back reference pages: 2, 8, and 14)

[22] G. L. Kooyman, "Techniques used in measuring diving capacities of Weddell Seals," *Polar Record*, vol. 12, no. 79, pp. 391–394, 1965. https://www.cambridge.org/core/journals/polar-record/article/techniques-used-in-measuring-diving-capacities-of-weddell-seals/ABD5262D6ABA9A69C02B7FAEBE4877C8
(Back reference pages: 2, 10, and 12)

[23] M. Johnson, N. A. De Soto, and P. T. Madsen, "Studying the behaviour and sensory ecology of marine mammals using acoustic recording tags: A review," *Marine Ecology Progress Series*, vol. 395, pp. 55–73, 2009.
(Back reference pages: 2, 10, 16, and 107)

[24] A. M. Dujon, R. T. Lindstrom, and G. C. Hays, "The accuracy of Fastloc-GPS locations and implications for animal tracking," *Methods in Ecology and Evolution*, vol. 5, no. 11, pp. 1162–1169, nov 2014. http://cddis.
(Back reference pages: 2, 4, 13, and 104)

[25] F. E. Fish, "POWER OUTPUT AND PROPULSIVE EFFICIENCY OF SWIMMING BOTTLENOSE DOLPHINS (TURSIOPS TRUNCATUS)," Tech. Rep., 1993.
(Back reference pages: 3, 4, 9, 25, 69, 105, and 108)

[26] ——, “Comparative kinematics and hydrodynamics of odontocete cetaceans: Morphological and ecological correlates with swimming performance,” *Journal of Experimental Biology*, vol. 201, no. 20, pp. 2867–2877, oct 1998.
(Back reference pages: xiii, 3, 9, 25, 71, 79, 80, and 107)

[27] H. Tanaka, G. Li, Y. Uchida, M. Nakamura, T. Ikeda, and H. Liu, “Measurement of time-varying kinematics of a dolphin in burst accelerating swimming,” *PLoS ONE*, vol. 14, no. 1, p. e0210860, jan 2019. https://doi.org/10.1371/journal.pone.0210860
(Back reference pages: 3, 4, 9, and 25)

[28] J. Karnowski, E. Hutchins, and C. Johnson, “Dolphin detection and tracking,” in *Proceedings - 2015 IEEE Winter Conference on Applications of Computer Vision Workshops, WACVW 2015*, 2015, pp. 51–56.
(Back reference pages: 3, 15, and 104)

[29] P. Rachinas-Lopes, R. Ribeiro, M. E. Dos Santos, and R. M. Costa, “D-Track—A semi-automatic 3D video-tracking technique to analyse movements and routines of aquatic animals with application to captive dolphins,” *PLoS ONE*, vol. 13, no. 8, p. e0201614, aug 2018. https://doi.org/10.1371/journal.pone.0201614.g001
(Back reference pages: 3, 15, 104, 106, and 108)

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
(Back reference pages: 3, 21)

[31] R. W. Davis, L. A. Fuiman, T. M. Williams, and B. J. Le Boeuf, “Three-dimensional movements and swimming activity of a northern elephant seal,” in *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, vol. 129, no. 4. Elsevier Inc., jul 2001, pp. 759–770.
(Back reference pages: 3, 4, 16, and 17)

[32] R. P. Wilson, N. Liebsch, I. M. Davies, F. Quintana, H. Weimerskirch, S. Storch, K. Lucke, U. Siebert, S. Zankl, G. Müller, I. Zimmer, A. Scolaro, C. Campagna, J. Plötz, H. Bornemann, J. Teilmann, and C. R. McMahon, “All at sea with animal tracks; methodological and analytical solutions for the resolution of movement,” *Deep-Sea Research Part II: Topical Studies in Oceanography*, vol. 54, no. 3-4, pp. 193–210, feb 2007.
(Back reference pages: 3, 4, 16, and 17)

[33] J. A. Goldbogen, N. D. Pyenson, and R. E. Shadwick, “Big gulps require high drag for fin whale lunge feeding,” *Marine Ecology Progress Series*, vol. 349, pp. 289–301, nov 2007. www.int-res.com
(Back reference pages: 3, 15)

[34] E. L. Shepard, R. P. Wilson, N. Liebsch, F. Quintana, A. Gómez Laich, and K. Lucke, "Flexible paddle sheds new light on speed: A novel method for the remote measurement of swim speed in aquatic animals," *Endangered Species Research*, vol. 4, no. 1-2, pp. 157–164, dec 2008. www.int-res.com
(Back reference pages: 3, 16)

[35] N. Aguilar Soto, M. P. Johnson, P. T. Madsen, F. Díaz, I. Domínguez, A. Brito, and P. Tyack, "Cheetahs of the deep sea: Deep foraging sprints in short-finned pilot whales off Tenerife (Canary Islands)," *Journal of Animal Ecology*, vol. 77, no. 5, pp. 936–947, sep 2008. https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2656.2008.01393.xhttps://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2656.2008.01393.xhttps://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.1365-2656.2008.01393.x
(Back reference pages: 3, 10, 12, and 15)

[36] T. M. Williams, W. A. Friedl, M. L. Fong, R. M. Yamada, P. Sedivy, and J. E. Haun, "Travel at low energetic cost by swimming and wave-riding bottlenose dolphins," *Nature*, vol. 355, no. 6363, pp. 821–823, 1992. https://www.nature.com/articles/355821a0
(Back reference pages: 3, 24, and 87)

[37] P. Yazdi, A. Kilian, and B. M. Culik, "Energy expenditure of swimming bottlenose dolphins (Tursiops truncatus)," *Marine Biology*, vol. 134, no. 4, pp. 601–607, sep 1999. https://link.springer.com/article/10.1007/s002270050575
(Back reference pages: 3, 9, 24, 65, 85, 86, and 108)

[38] J. M. Van Der Hoop, A. Fahlman, T. Hurst, J. Rocho-Levine, K. A. Shorter, V. Petrov, and M. J. Moore, "Bottlenose dolphins modify behavior to reduce metabolic effect of tag attachment," *Journal of Experimental Biology*, vol. 217, no. 23, pp. 4229–4236, dec 2014.
(Back reference pages: 3, 11, 24, 71, and 86)

[39] A. Fahlman, R. Wilson, C. Svärd, D. A. Rosen, and A. W. Trites, "Activity and diving metabolism correlate in Steller sea lion Eumetopias jubatus," *Aquatic Biology*, vol. 2, no. 1, pp. 75–84, mar 2008. www.int-res.com
(Back reference pages: 3, 24, and 25)

[40] R. P. Wilson, C. R. White, F. Quintana, L. G. Halsey, N. Liebsch, G. R. Martin, and P. J. Butler, "Moving towards acceleration for estimates of activity-specific metabolic rate in free-living animals: The case of the cormorant," *Journal of Animal Ecology*, vol. 75, no. 5, pp. 1081–1090, sep 2006. https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2656.2006.01127.xhttps://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2656.2006.01127.xhttps://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.1365-2656.2006.01127.x
(Back reference pages: 3, 24)

[41] F. E. Fish, P. Legac, T. M. Williams, and T. Wei, "Measurement of hydrodynamic force generation by swimming dolphins using bubble DPIV," *Journal of Experimental Biology*, vol. 217, no. 2, pp. 252–260, jan 2014.
(Back reference pages: xiii, 4, 25, 69, 79, 80, 86, and 108)

[42] Y. Le Bras, J. Jouma'a, and C. Guinet, "Three-dimensional space use during the bottom phase of southern elephant seal dives," *Movement Ecology*, vol. 5, no. 1, p. 18, aug 2017. https://movementecologyjournal.biomedcentral.com/articles/10.1186/s40462-017-0108-y
(Back reference pages: 4, 17)

[43] ARGOS, "Argos User's Manual," 2016. https://www.argos-system.org/wp-content/uploads/2016/08/r363_9_argos_users_manual-v1.6.6.pdf
(Back reference pages: 4, 11, and 13)

[44] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 216–235, 2012. http://journals.sagepub.com/doi/10.1177/0278364911430419
(Back reference pages: 5, 17, 90, 99, and 106)

[45] P. L. Tyack and C. W. Clark, *Communication and Acoustic Behavior of Dolphins and Whales*. New York, NY: Springer New York, 2000, pp. 156–224. https://doi.org/10.1007/978-1-4612-1150-1_4
(Back reference page: 8)

[46] N. Sobel, A. Y. Supin, and M. S. Myslobodsky, "Rotational swimming tendencies in the dolphin (Tursiops truncatus)," *Behavioural Brain Research*, vol. 65, no. 1, pp. 41–45, nov 1994.
(Back reference page: 8)

[47] C. Ugaz, R. A. Valdez, M. C. Romano, and F. Galindo, "Behavior and salivary cortisol of captive dolphins (Tursiops truncatus) kept in open and closed facilities," *Journal of Veterinary Behavior: Clinical Applications and Research*, vol. 8, no. 4, pp. 285–290, jul 2013.
(Back reference pages: 8, 9)

[48] K. A. Waples and N. J. Gales, "Evaluating and minimising social stress in the care of captive bottlenose dolphins (Tursiops aduncus)," *Zoo Biology*, vol. 21, no. 1, pp. 5–26, jan 2002. https://onlinelibrary.wiley.com/doi/full/10.1002/zoo.10004https://onlinelibrary.wiley.com/doi/abs/10.1002/zoo.10004https://onlinelibrary.wiley.com/doi/10.1002/zoo.10004
(Back reference pages: 8, 9)

[49] I. L. Clegg, H. G. Rödel, M. Cellier, D. Vink, I. Michaud, B. Mercera, M. Böye, M. Hausberger, A. Lemasson, and F. Delfour, "Schedule of human-controlled periods structures bottlenose dolphin (tursiops truncatus) behavior in their free-time," *Journal of Comparative Psychology*, vol. 131, no. 3, pp. 214–224, aug 2017. /record/2017-14080-001
(Back reference pages: 8, 11)

[50] I. L. Clegg, C. E. Van Elk, and F. Delfour, "Applying welfare science to bottlenose dolphins (Tursiops truncatus)," *Animal Welfare*, vol. 26, no. 2, pp. 165–176, 2017. www.ufaw.org.uk
(Back reference pages: 8, 9)

[51] Y. Sekiguchi and S. Kohshima, “Resting behaviors of captive bottlenose dolphins (Tursiops truncatus),” *Physiology and Behavior*, vol. 79, no. 4-5, pp. 643–653, 2003.
(Back reference pages: 9, 47, and 85)

[52] B. S. Harvey, K. M. Dudzinski, and S. A. Kuczaj, “Associations and the role of affiliative, agonistic, and socio-sexual behaviors among common bottlenose dolphins (Tursiops truncatus),” *Behavioural Processes*, vol. 135, pp. 145–156, feb 2017.
(Back reference page: 9)

[53] M. C. Caldwell and D. K. Caldwell, “Vocalization of naive captive dolphins in small groups,” *Science*, vol. 159, no. 3819, pp. 1121–1123, mar 1968. http://science.sciencemag.org/
(Back reference page: 9)

[54] B. McCowan and D. Reiss, “Maternal aggressive contact vocalizations in captive bottlenose dolphins (Tursiops truncatus): Wide-band, low-frequency signals during mother/aunt-infant interactions,” *Zoo Biology*, vol. 14, no. 4, pp. 293–309, 1995.
(Back reference page: 9)

[55] A. E. Ames, R. P. Macgregor, S. J. Wielandt, D. M. Cameron, S. A. Kuczaj, and H. M. Hill, “Pre-and post-partum whistle production of a bottlenose dolphin (Tursiops truncatus) social group,” *International Journal of Comparative Psychology*, vol. 32, pp. 1–17, 2019.
(Back reference page: 9)

[56] L. E. Freitag and P. L. Tyack, “Passive acoustic localization of the Atlantic bottlenose dolphin using whistles and echolocation clicks,” *The Journal of the Acoustical Society of America*, vol. 93, no. 4, pp. 2197–2205, 1993. https://doi.org/10.1121/1.406681
(Back reference pages: 9, 14)

[57] S. K. Hooker and R. W. Baird, “Deep-diving behaviour of the northern bottlenose whale, Hyperoodon ampullatus (Cetacea: Ziphiidae),” *Proceedings of the Royal Society B: Biological Sciences*, vol. 266, no. 1420, pp. 671–676, apr 1999. https://royalsocietypublishing.org/
(Back reference pages: 9, 12)

[58] C. Pomilla and H. C. Rosenbaum, “Against the current: An inter-oceanic whale migration event,” *Biology Letters*, vol. 1, no. 4, pp. 476–479, dec 2005. https://royalsocietypublishing.org/
(Back reference page: 10)

[59] A. J. Read, “Telemetry,” in *Encyclopedia of Marine Mammals*. Elsevier Ltd, jan 2009, pp. 1153–1156.
(Back reference pages: 10, 11, and 13)

[60] M. D. Scott and S. J. Chivers, "Movements and diving behavior of pelagic spotted dolphins," *Marine Mammal Science*, vol. 25, no. 1, pp. 137–160, jan 2009. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1748-7692.2008.00241.xhttps://onlinelibrary.wiley.com/doi/abs/10.1111/j.1748-7692.2008.00241.xhttps://onlinelibrary.wiley.com/doi/10.1111/j.1748-7692.2008.00241.x
(Back reference pages: 10, 11, and 12)

[61] D. Costa, "The Secret Life of Marine Mammals: Novel Tools for Studying Their Behavior and Biology at Sea," *Oceanography*, vol. 6, no. 3, pp. 120–128, 1993. https://www.jstor.org/stable/43924653
(Back reference pages: 10, 12)

[62] P. T. Madsen, M. Johnson, N. Aguilar De Soto, W. M. Zimmer, and P. Tyack, "Biosonar performance of foraging beaked whales (Mesoplodon densirostris)," *Journal of Experimental Biology*, vol. 208, no. 2, pp. 181–194, jan 2005.
(Back reference pages: 10, 12)

[63] S. Isojunno, C. Curé, P. H. Kvadsheim, F. P. A. Lam, P. L. Tyack, P. J. Wensveen, and P. J. O. Miller, "Sperm whales reduce foraging effort during exposure to 1-2 kH z sonar and killer whale sounds," *Ecological Applications*, vol. 26, no. 1, pp. 77–93, jan 2016. https://esajournals.onlinelibrary.wiley.com/doi/full/10.1890/15-0040https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/15-0040https://esajournals.onlinelibrary.wiley.com/doi/10.1890/15-0040
(Back reference pages: 10, 12)

[64] B. S. Stewart, S. Leatherwood, P. K. Yochem, and M. Heide-Jørgensen, "HARBOR SEAL TRACKING AND TELEMETRY BY SATELLITE," *Marine Mammal Science*, vol. 5, no. 4, pp. 361–375, oct 1989. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1748-7692.1989.tb00348.xhttps://onlinelibrary.wiley.com/doi/abs/10.1111/j.1748-7692.1989.tb00348.xhttps://onlinelibrary.wiley.com/doi/10.1111/j.1748-7692.1989.tb00348.x
(Back reference pages: 11, 12, and 13)

[65] R. D. Andrews, R. L. Pitman, and L. T. Ballance, "Satellite tracking reveals distinct movement patterns for Type B and Type C killer whales in the southern Ross Sea, Antarctica," *Polar Biology*, vol. 31, no. 12, pp. 1461–1468, nov 2008. https://link.springer.com/article/10.1007/s00300-008-0487-z
(Back reference pages: 11, 13)

[66] Y. Akiyama, Y. Matsuda, N. Sakurai, and K. Sato, "Evaluation of wave drag on bottlenose dolphin Tursiops truncatus from swimming effort," *Coastal marine science*, vol. 38, pp. 42–46, 2015.
(Back reference page: 11)

[67] K. Alex Shorter, Y. Shao, L. Ojeda, K. Barton, J. Rocho-Levine, J. van der Hoop, and M. Moore, “A day in the life of a dolphin: Using bio-logging tags for improved animal health and well-being,” *Marine Mammal Science*, vol. 33, no. 3, pp. 785–802, jul 2017. https://onlinelibrary.wiley.com/doi/full/10.1111/mms.12408https://onlinelibrary.wiley.com/doi/abs/10.1111/mms.12408https://onlinelibrary.wiley.com/doi/10.1111/mms.12408
(Back reference pages: 11, 88, 102, and 108)

[68] O. S. Eyobu and D. S. Han, “Feature representation and data augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network,” *Sensors (Switzerland)*, vol. 18, no. 9, p. 2892, sep 2018. www.mdpi.com/journal/sensors
(Back reference page: 12)

[69] P. L. Tyack, M. Johnson, N. Aguilar Soto, A. Sturlese, and P. T. Madsen, “Extreme diving of beaked whales,” *Journal of Experimental Biology*, vol. 209, no. 21, pp. 4238–4253, nov 2006.
(Back reference page: 12)

[70] S. M. Tomkiewicz, M. R. Fuller, J. G. Kie, and K. K. Bates, “Global positioning system and associated technologies in animal behaviour and ecological research,” pp. 2163–2176, jul 2010.
(Back reference page: 13)

[71] GPS.gov, “GPS.gov: GPS Accuracy.” https://www.gps.gov/systems/gps/performance/accuracy/
(Back reference page: 13)

[72] L. N. Frazer and P. I. Pecholcs, “Single-hydrophone localization,” *The Journal of the Acoustical Society of America*, vol. 88, no. 2, pp. 995–1002, aug 1990. https://doi.org/10.1121/1.399750
(Back reference page: 14)

[73] E. K. Westwood and D. P. Knobles, “Source track localization via multipath correlation matching,” *The Journal of the Acoustical Society of America*, vol. 102, no. 5, pp. 2645–2654, nov 1997. https://doi.org/10.1121/1.420318
(Back reference page: 14)

[74] P. Giraudet and H. Glotin, “Real-time 3D tracking of whales by echo-robust precise TDOA estimates with a widely-spaced hydrophone array,” *Applied Acoustics*, vol. 67, no. 11-12, pp. 1106–1117, nov 2006.
(Back reference page: 14)

[75] C. O. Tiemann, M. B. Porter, and L. N. Frazer, “Localization of marine mammals near Hawaii using an acoustic propagation model,” *The Journal of the Acoustical Society of America*, vol. 115, no. 6, pp. 2834–2843, jun 2004. https://doi.org/10.1121/1.1643368
(Back reference page: 14)

[76] W. Selby, P. Corke, and D. Rus, “Autonomous aerial navigation and tracking of marine animals,” in *Proceedings of the 2011 Australasian Conference on Robotics and Automation*, 2011.
(Back reference page: 14)

[77] R. P. Wilson and M.-P. Wilson, “Dead reckoning: a new technique for determining penguim movements at sea,” *Meeresforschung (Hamburg)*, vol. 32, no. 2, pp. 155–158, 1988.
(Back reference page: 15)

[78] R. A. Hyde, L. P. Ketteringham, S. A. Neild, and R. J. Jones, “Estimation of upper-limb orientation based on accelerometer and gyroscope measurements,” *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 2, pp. 746–754, feb 2008.
(Back reference page: 16)

[79] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, “Activation Functions: Comparison of trends in Practice and Research for Deep Learning,” nov 2018. http://arxiv.org/abs/1811.03378
(Back reference pages: 18, 19)

[80] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation Applied to Handwritten Zip Code Recognition,” *Neural Computation*, vol. 1, no. 4, pp. 541–551, dec 1989. http://direct.mit.edu/neco/article-pdf/1/4/541/811941/neco.1989.1.4.541.pdf
(Back reference pages: 18, 21)

[81] Y. LeCun, K. Kavukcuoglu, and C. Farabet, “Convolutional networks and applications in vision,” in *ISCAS 2010 - 2010 IEEE International Symposium on Circuits and Systems: Nano-Bio Circuit Fabrics and Systems*, 2010, pp. 253–256.
(Back reference page: 18)

[82] X. Glorot and Y. Bengio, “Understanding the difficulty of training deep feedforward neural networks,” in *Journal of Machine Learning Research*, vol. 9. JMLR Workshop and Conference Proceedings, mar 2010, pp. 249–256. http://www.iro.umontreal.
(Back reference page: 19)

[83] K. Fukushima, “Neocognitron: A hierarchical neural network capable of visual pattern recognition,” *Neural Networks*, vol. 1, no. 2, pp. 119–130, jan 1988.
(Back reference page: 19)

[84] V. Nair and G. E. Hinton, “Rectified linear units improve Restricted Boltzmann machines,” in *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, 2010, pp. 807–814.
(Back reference page: 19)

[85] A. L. Maas, A. Y. Hannun, and A. Y. Ng, “Rectifier nonlinearities improve neural network acoustic models,” in *in ICML Workshop on Deep Learning for Audio, Speech and Language Processing*, vol. 30, no. 1. Citeseer, 2013, p. 3.
(Back reference page: 19)

[86] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2015 Inter, 2015, pp. 1026–1034.
(Back reference pages: 19, 27)

[87] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical Evaluation of Rectified Activations in Convolutional Network," may 2015. http://arxiv.org/abs/1505.00853
(Back reference page: 20)

[88] H. Gholamalinezhad and H. Khosravi, "Pooling Methods in Deep Neural Networks, a Review," sep 2020. http://arxiv.org/abs/2009.07485
(Back reference page: 20)

[89] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, and L. Jackel, "Handwritten digit recognition with a back-propagation network," *Advances in neural information processing systems*, vol. 2, 1989.
(Back reference page: 20)

[90] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5455–5516, dec 2020. https://doi.org/10.1007/s10462-020-09825-6
(Back reference page: 21)

[91] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, sep 2015. http://www.robots.ox.ac.uk/
(Back reference pages: 21, 28)

[92] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07-12-June, 2015, pp. 1–9.
(Back reference pages: 21, 28)

[93] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, 2016, pp. 2818–2826.
(Back reference page: 22)

[94] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, 2016, pp. 770–778. http://image-net.org/challenges/LSVRC/2015/
(Back reference pages: 22, 28)

[95] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-ResNet and the impact of residual connections on learning,” in *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, vol. 31, no. 1, feb 2017, pp. 4278–4284. www.aaai.org
(Back reference page: 22)

[96] R. Girshick, “Fast R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, dec 2015.
(Back reference page: 22)

[97] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, oct 2017.
(Back reference pages: 23, 27)

[98] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, 2016, pp. 779–788. http://pjreddie.com/yolo/
(Back reference pages: 23, 27)

[99] J. Redmon and A. Farhadi, “YOLO9000: Better, Faster, Stronger,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, jul 2017.
(Back reference page: 23)

[100] ——, “YOLOv3: An Incremental Improvement,” *CoRR*, vol. abs/1804.0, 2018. http://arxiv.org/abs/1804.02767
(Back reference page: 23)

[101] R. J. Full and M. S. Tu, “Mechanics of a rapid running insect: Two-, four- and six-legged locomotion,” *Journal of Experimental Biology*, vol. 156, no. 1, pp. 215–231, mar 1991.
(Back reference page: 23)

[102] H. Pontzer, “A unified theory for the energy cost of legged locomotion,” *Biology Letters*, vol. 12, no. 2, feb 2016. http://dx.doi.org/10.1098/rsbl.2015.0935orviahttp://rsbl.royalsocietypublishing.org.Biomechanics
(Back reference page: 23)

[103] D. P. Costa and R. L. Gentry, “Chapter 5. Free-Ranging Energetics of Northern Fur Seals,” in *Fur Seals*. Princeton University Press, dec 2014, pp. 79–101.
(Back reference page: 24)

[104] J. R. Speakman and E. Król, “Comparison of different approaches for the calculation of energy expenditure using doubly labeled water in a small mammal,” pp. 650–667, jul 2005. https://www.journals.uchicago.edu/doi/abs/10.1086/430234
(Back reference page: 24)

[105] T. M. Williams, W. A. Friedl, and J. E. Haun, “The physiology of bottlenose dolphins (Tursiops truncatus): heart rate, metabolic rate and plasma lactate concentration during exercise.” *The Journal of experimental biology*, vol. 179, no. 1, pp. 31–46, jun 1993.
(Back reference pages: 24, 65, and 80)

[106] J. M. van der Hoop, A. Fahlman, K. A. Shorter, J. Gabaldon, J. Rocho-Levine, V. Petrov, and M. J. Moore, “Swimming energy economy in bottlenose dolphins under variable drag loading,” *Frontiers in Marine Science*, vol. 5, no. DEC, dec 2018.
(Back reference pages: 24, 25)

[107] K. Schmidt-Nielsen, *Animal physiology: adaptation and environment*. Cambridge university press, 1997.
(Back reference pages: 24, 71)

[108] Mathworks, “Define a Custom Deep Learning Layer with Learnable Parameters,” 2019. https://www.mathworks.com
(Back reference page: 27)

[109] R. I. Hartley, “In Defense of the Eight-Point Algorithm,” *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 19, no. 6, 1997.
(Back reference pages: 33, 96)

[110] F. C. Porter, “Testing Consistency of Two Histograms,” apr 2008. http://arxiv.org/abs/0804.0380
(Back reference page: 39)

[111] T. Y. Moore, K. L. Cooper, A. A. Biewener, and R. Vasudevan, “Unpredictability of escape trajectory explains predator evasion ability and microhabitat preference of desert rodents,” *Nature Communications*, vol. 8, no. 1, pp. 1–9, dec 2017. www.nature.com/naturecommunications
(Back reference page: 47)

[112] C. E. Willert and M. Gharib, “Digital particle image velocimetry,” *Experiments in Fluids*, vol. 10, no. 4, pp. 181–193, 1991. https://link.springer.com/article/10.1007/BF00190388
(Back reference page: 49)

[113] R. J. Adrian, “Particle-imaging techniques for experimental fluid mechanics,” *Annual Review of Fluid Mechanics*, vol. 23, no. 1, pp. 261–304, 1991. www.annualreviews.org
(Back reference page: 49)

[114] R. W. Fox, A. T. McDonald, and J. W. Mitchell, *Fox and McDonald’s introduction to fluid mechanics*. John Wiley & Sons, 2020.
(Back reference page: 53)

[115] B. Wieneke, “PIV uncertainty quantification from correlation statistics,” *Measurement Science and Technology*, vol. 26, no. 7, p. 074002, jun 2015. https://iopscience.iop.org/article/10.1088/0957-0233/26/7/074002https://iopscience.iop.org/article/10.1088/0957-0233/26/7/074002/meta
(Back reference page: 55)

[116] D. Scaramuzza, A. Martinelli, and R. Siegwart, “A flexible technique for accurate omnidirectional camera calibration and structure from motion,” in *Proceedings of the Fourth IEEE International Conference on Computer Vision Systems, ICVS’06*, vol. 2006, 2006, p. 45.
(Back reference page: 60)

[117] S. H. Shane, “Behavior and Ecology of the Bottlenose Dolphin at Sanibel Island, Florida,” *The Bottlenose Dolphin*, pp. 245–265, 1990.
(Back reference page: 65)

[118] C. Lockyer and R. Morris, “Observations on diving behaviour and swimming speeds in a wild juvenile bottlenose dolphins,” *Aquatic Mammals*, vol. 13, no. 1, pp. 31–35, 1987.
(Back reference page: 65)

[119] S. Vogel, *Life in Moving Fluids: The Physical Biology of Flow-Revised and Expanded Second Edition*. Princeton University Press, 2020.
(Back reference page: 68)

[120] D. Au and D. Weihs, “At high speeds dolphins save energy by leaping,” *Nature*, vol. 284, no. 5756, pp. 548–550, 1980.
(Back reference page: 69)

[121] H. Hertel, “Structure, form, movement,” 1966.
(Back reference page: 70)

[122] S. Faraji, A. R. Wu, and A. J. Ijspeert, “A simple model of mechanical effects to estimate metabolic cost of human walking,” *Scientific Reports*, vol. 8, no. 1, pp. 1–12, 2018.
(Back reference page: 71)

[123] R. A. Kastelein, C. Staal, and P. R. Wiepkema, “Food consumption, food passage time, and body measurements of captive Atlantic bottlenose dolphins (¡I¿Tursiops truncatus¡/I¿),” *Aquatic Mammals*, vol. 29, no. 1, pp. 53–66, 2005.
(Back reference pages: 81, 108)

[124] R. H. Defran, D. W. Weller, D. L. Kelly, and M. A. Espinosa, “Range characteristics of Pacific coast bottlenose dolphins (Tursiops truncatus) in the Southern California Bight,” *Marine Mammal Science*, vol. 15, no. 2, pp. 381–393, 1999.
(Back reference page: 86)

[125] A. Dinis, C. Molina, M. Tobeña, A. Sambolino, K. Hartman, M. Fernandez, S. Magalhães, R. P. dos Santos, F. Ritter, V. Martín, N. A. de Soto, and F. Alves, “Large-scale movements of common bottlenose dolphins in the Atlantic: Dolphins with an international courtyard,” *PeerJ*, vol. 9, p. e11069, 2021.
(Back reference page: 86)

[126] A. C. Bejarano, R. S. Wells, and D. P. Costa, “Development of a bioenergetic model for estimating energy requirements and prey biomass consumption of the bottlenose dolphin Tursiops truncatus,” *Ecological Modelling*, vol. 356, pp. 162–172, 2017.
(Back reference page: 86)

[127] A. S. Allen, A. J. Read, K. A. Shorter, J. Gabaldon, A. M. Blawas, J. Rocho-Levine, and A. Fahlman, “Dynamic body acceleration as a proxy to predict the cost of locomotion in bottlenose dolphins,” *Journal of Experimental Biology*, 2021.
(Back reference page: 88)

[128] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking,” *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, feb 2002.
(Back reference pages: 94, 106)

[129] P. Bergström and O. Edlund, “Robust registration of point sets using iteratively reweighted least squares,” *Computational Optimization and Applications*, vol. 58, no. 3, pp. 543–561, 2014.
(Back reference pages: 96, 106)

[130] Georgia Tech BORG Lab, “GTSAM 3.2.1.” http://www.borg.cc.gatech.edu/download.html
(Back reference page: 99)