From AI to IA: Towards Intelligent Analysis of Cooperative Behavior in Bottlenose Dolphins

by

Ding Zhang

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

Doctoral Committee:

Assistant Professor K. Alex Shorter, Co-Chair
Associate Professor Kira Barton, Co-Chair
Assistant Professor Talia Moore
Associate Professor Jenna Wiens

Ding Zhang

zhding@umich.edu

ORCID iD: 0000-0002-3501-5486

© Ding Zhang 2021

ACKNOWLEDGMENTS

It has been a challenging but exciting journey in accomplishing this dissertation. There are so many people I am grateful to.

Starting with my doctoral research mentors, Kira Barton and Alex Shorter. They made sure that the research vessel was always heading in novel and exciting directions. Kira and Alex also showed me how to be an inspiring and supportive collaborator and advisor at the same time. I will try to be a good collaborator and advisor myself in the future to pass the merits along.

As the doctoral committee members, Talia Moore and Jenna Wiens have contributed many valuable insights in shaping this dissertation. In particular, Talia's expertise in biology and biomechanics has enhanced this dissertation in terms of its biological reasoning. And Jenna's knowledge in AI and machine learning has strengthened the analysis methods of this dissertation.

Lauro Ojeda had mentored my Master's research project, which led to the work in this dissertation. Lauro's enthusiasm for technology had motivated me to bring more state-of-the-art techniques to applicable interdisciplinary investigations.

My labmates from the Barton Research Group and ESTAR Lab have accompanied and supported me over the years. To name some of them: Joaquin Gabaldon, Isaac Spiegel, Yeonjoon Cheong, Ilya Kovalenko, Lai-Yu Leo Tse, Efe Balta, Delaramm Afkhami, Nazanin Farjam, Yassine Qamsane, Dory Yang, Yaqing Xu, Loubna Baroudi, Deema Totah, Mike Quann, Anne Gu, Meghna Menon, Mingjie Bi, Bo Fu, Max Wu, Tyler Toner, Max Toothman, Miguel Saez, Berk Altın, Christopher Pannier.

All the friends I have made in Michigan have inspired me to be a rigorous researcher, a hard-working individual, and a thoughtful team player. Here is an incomplete list of them: Xin Dong, Da Li, Hao Li, Ziwei Zeng, Bo Lin, Molong Duan, Zhiyuan Zuo, Shaowu Pan, Xingyu Li, Nan Li, Haotian Feng, Yan Zhao, Daning Huang, Zhixiong Zhang, Xiuyuan Yang, Jiacheng Huang, Yiding Liang, Zihang Wei, Yang Wang, Chengyuan Wu, Xudong Hao.

My fiancee has accompanied me through this journey with tireless support and understanding. My strength has been augmented and will always be.

Finally, I am grateful to my parents for the continuous support and love, together with the philosophical guidance and everything else.

TABLE OF CONTENTS

Acknowledgments . . . . . **ii**

List of Figures . . . . . **vi**

List of Tables . . . . . **xiii**

List of Abbreviations . . . . . **xiv**

Abstract . . . . . **xvi**

Chapter

1 Introduction **1**

1.1 Intelligence, Coordinated Behaviors, and Biomimicry 1
1.1.1 Scope of This Dissertation 2
1.2 The Common Bottlenose Dolphins (*Tursiops truncatus*) 4
1.3 Biologging Tags for Cetaceans 5
1.4 Computer Vision for Dolphins: Detection and Tracking 7
1.5 Dolphin Locomotion and Dynamics 9
1.5.1 Environment, Coordination, and Behavior 10
1.6 Dissertation Structure 10

2 Estimation with Biologging Tags: Robust Pose-Gait Analysis and Characterization **12**

2.1 Background 12
2.2 The Biologging Tag: MTag 14
2.3 Pose-Gait Patterns in Tag Data and Assumptions 15
2.4 Nonparametric Tag Shift Detection 16
2.5 Pattern Based Orientation Correction 17
2.6 Animal Pose Calculation and Gait Characterization 20
2.7 Validation and Results 22
2.8 Discussion 26
2.9 Summary 29

3 Estimation with Biologging Tags: Sequential Motion State Classification **31**

3.1 Background 31
3.2 Problem Formulation and Methods Outline 33

3.3 Feature Engineering . . . 36
3.3.1 Window Size Selection . . . 36
3.3.2 Feature Library and Feature Selection . . . 37
3.3.3 Adaptive Weighted Sampling Scheme . . . 38
3.4 Sequential Frame Classification with HMM . . . 39
3.4.1 Model Training and Data Fitting/Testing . . . 40
3.5 Working without Feature Engineering:
An LSTM-HMM Framework for Sequential State Classification . . . 44
3.6 A Case Study on Dolphin Motion Classification . . . 45
3.6.1 Data Preparation . . . 46
3.6.2 LSTM RNN Configurations . . . 47
3.7 Validation Results and Discussions . . . 48
3.7.1 Feature Engineering - HMM . . . 48
3.7.2 LSTM - HMM . . . 52
3.8 Summary . . . 56

4 Identity Preserving Multi-Target Tracking with Cameras and Biologging Tags . . . 57

4.1 Background . . . 57
4.2 Problem Formulation . . . 60
4.2.1 Experimental Environment . . . 60
4.2.2 Objective and Assumptions . . . 60
4.2.3 Framework Outline . . . 62
4.3 External Measurement – Cameras . . . 62
4.3.1 Camera Detection . . . 62
4.3.2 2D-3D Projection . . . 62
4.3.3 General Multi-Target Tracking using Above Habitat Video . . . 63
4.4 Internal Measurement – Tags . . . 66
4.4.1 Heading Estimation . . . 66
4.4.2 Motion Model and Odometry . . . 66
4.4.3 Noise Model . . . 67
4.5 Particle Filter (PF) based Fusion of Cameras and Biologging Tags . . . 67
4.5.1 Preliminaries . . . 67
4.5.2 Fuzzy Observations . . . 69
4.5.3 Non-causal State Estimation . . . 70
4.6 Results and Discussions . . . 71
4.7 Summary . . . 74

**5 Data-Driven Modeling of Dolphin Dynamics:
Hydrodynamic Drag, Transport Cost, and Gait . . . 75**

5.1 Background . . . 75
5.1.1 Gait and Swimming Efficiency . . . 75
5.1.2 Tag Induced Drag . . . 76
5.1.3 Research Contributions . . . 77
5.2 Swimming Biomechanics . . . 77
5.2.1 The Fluke-and-Glide (FG) Gait and its Dynamics . . . 77

5.2.2 Kinetics of the Fluke-and-Glide (FG) Gait . . . 80
5.2.3 Mechanical Energy Cost of Transport (MECOT) . . . 81
5.3 Data-Driven Gait Parametrization using Biologging Tags . . . 81
5.3.1 Data collection . . . 81
5.3.2 Glide Extraction and Fluke-and-Glide (FG) Bout Identification . . . 82
5.3.3 Drag Estimation with Parameter Optimization . . . 83
5.3.4 Speed Distribution and Normalization . . . 83
5.4 Hydrodynamic Impact of Biologging Tags: Experiments and Simulations . . . 85
5.4.1 Data Collection . . . 85
5.4.2 Video Processing and Data Extraction . . . 87
5.4.3 Parameter Synthesis and Compensation . . . 89
5.4.4 Computational Fluid Dynamics (CFD) . . . 90
5.4.5 Statistical Analysis . . . 91
5.5 Results . . . 92
5.5.1 Estimated Drag of the Tag . . . 92
5.5.2 Fluke and Glide (FG) Gait Characteristics . . . 92
5.5.3 Energetics as A Function of Gait . . . 94
5.5.4 Assessing the Hydrodynamic Impact of a Tag . . . 95
5.6 Discussion . . . 101
5.6.1 Energetic Benefits of Fluke and glide gait . . . 101
5.6.2 On the Hydrodynamic Impact of a Tag . . . 102
5.7 Summary . . . 106

6 Fine-Scale Behavior Analysis using Cameras and Biologging Tags . . . 107

6.1 Background . . . 107
6.2 Materials and Methods . . . 109
6.2.1 Environment Setups . . . 109
6.2.2 Target Detection . . . 110
6.2.3 Tracking of Targeted Group and Individuals . . . 111
6.2.4 Movement Synthesis in the Environment . . . 112
6.2.5 Motion States and Swimming Statistics of Tagged Individuals . . . 113
6.2.6 Engaging with Enrichment . . . 114
6.3 Results . . . 114
6.3.1 General Group Kinematic . . . 114
6.3.2 Group Enrichment Engagement . . . 117
6.3.3 Individual Behavior During Interactions with Conspecifics and Enrichment 118
6.4 Discussion . . . 126
6.5 Summary . . . 130

7 Conclusions . . . 131

7.1 Synthesis of Research Contributions and Broader Impacts . . . 131
7.2 Future Directions . . . 134

Bibliography . . . 137

LIST OF FIGURES

1.1 The scope (**left**) and structure (**right**) of this dissertation. This dissertation is at the intersection of sensor fusion, machine learning, computer vision, data-driven modeling, biologging, animal behavior, and biomechanics. The work leverages and advances frameworks from Artificial Intelligence (AI) to enable the Intelligent Analysis (IA) methods used to extract the dynamic state of the animal. The framework was used to investigate the energetic cost of swimming gait and the coordinated behavior of bottlenose dolphins. This work brings us one step closer to the further goal of studying bio-inspired artificial multi-agent systems. Dashed arrows indicate a connection that has not been explicitly made in this dissertation, but is important future work. 3

1.2 A bottlenose dolphin (*Tursiops truncatus*, Section 1.2) equipped with a suction cup based biologging tag (animal-borne wearable sensor package, Section 1.3) in Hawaii (*photo by D. Zhang*). 4

2.1 Biologging tag attached to a bottlenose dolphin (MTag, **left**) and a humpback whale (DTag, **right**) with coordinate definitions. Sometimes the tag's coordinates are aligned with the animal (**left**) but oftentimes they are not (**right**). It is also possible that a tag's position on the animal changes during a deployment. For applications like animal pose estimation, tag data needs to be represented in the animal's body coordinates during post processing, which implicitly requires the knowledge of relative orientation between tag and animal. Note that a positive animal pitch is a negative rotation around the body's y-axis under this coordinate system. 14

2.2 **Orientation sphere** for a bottlenose dolphin with the biologging tag aligned with the animal's body, (i.e. data in the animal's body coordinates without the need for a correction). Each data point is clustered into 'flat', 'ascend', or 'descend' depending on its depth speed. The dominant directions for the 3 clusters are marked by $\boldsymbol{d}_{flat}$, $\boldsymbol{d}_{ascend}$, and $\boldsymbol{d}_{descend}$, to be discussed in Section 2.5. 15

2.3 **Orientation sphere** for a bottlenose dolphin with the data in the tag coordinates before any correction. The presented dataset contains 4 simulated tag shifts. 16

2.4 Tag shift detection algorithm. Each branch is marked by a circled numbers (1 to 12). . 18

2.5 **Orientation sphere** for a data segment of a bottlenose dolphin (**left**) and a humpback whale (**right**). The plots visualize the orientation correction method being applied to a uncorrected data segment to find the dominant directions in the tag's coordinates. The dominant directions can then be mapped to their assumed directions in the body's coordinates, Figure 2.2. 19

2.6 **Orientation spheres** for a bottlenose dolphin (left column), a humpback whale (center column) and a beluga whale (right column). The plot of accelerometer data clustered by depth speed is referred to as **orientation sphere**. The top row of subplots presents the orientation spheres in the animal's body coordinates, after detecting tag shifts and correcting tag orientation misalignment. The middle row presents the data in the tag coordinates before any correction. The bottom row visualizes the orientation correction method being applied to a uncorrected data segment. The dolphin dataset (left column) contains 4 simulated tag shifts. The humpback whale dataset (center column) has 5 detected tag shifts. While the beluga dataset (right column) has no detected tag shifts (i.e. no relative motion between tag and animal was detected). . 24
2.7 Average precision and recall of tag shift detection (**left**) and the average absolute error of the calculated animal pose after tag orientation correction (**right**) over simulated tag shifts with random direction and fixed degrees (x-axis). SD denotes the average of the standard deviation values for roll, pitch, and yaw at each offset increment. 25
2.8 Example data segments from a bottlenose dolphin (**left**) and a beluga whale (**right**) with animal's gait parameterized via **dynamic pitch** ($pitch_{dp}$). When the animal has a big roll angle (e.g. during the time around 240 seconds in the bottlenose dolphin dataset), the fluking 'signature' (i.e. the sinusoidal fluctuations in each signal channel) transfers from pitch to yaw in the orientation estimations. $Pitch_{dp}$ is used to have a pose invariant descriptor of the gait of the animal, that is with respect to the animal itself rather than the environment. Inactive swimming periods (e.g. gliding) are automatically identified in $pitch_{dp}$ while fluking frequency and amplitude are calculated from $pitch_{dp}$, which can be used for further gait analysis. 27

3.1 Architecture of 'single frame' classification. The example signal consists of 11 channels. Window size is denoted as W, overlap between windows is r. The signal is segmented into T windows with each window considered as a discrete state. For each state (window), M features are extracted and taken as observations of each state (e.g. $\boldsymbol{F}_t = \{F_{1t}, F_{2t}, \cdots, F_{Mt}\}$ for the t-th state). The motion class of each state is denoted by random variable X, with lower case x denotes the instance (i.e. actual value), e.g. $X_t = x_t$ means state X_t takes value x_t, where x_t belongs to one of the n motion classes ($x_t \in \{c_1, \cdots, c_n\}$). 35
3.2 Block diagram of the proposed framework. 36
3.3 Example Hidden Markov Model with $n = 3$ motion classes $\{c_1,\ c_2,\ c_3\}$. The probability of state transitions from class c_i to class c_j is denoted by a_{ij}. Multivariate random variable $\boldsymbol{F}$ represents the observation emitted/obtained from each class. $\boldsymbol{b}_i$ models the probability distribution of $\boldsymbol{F}$ that emitted from a specific class c_i. 41
3.4 Architecture of 'sequential frame' classification using Hidden Markov Model (HMM). Extended on Figure 3.1, each feature vector $\boldsymbol{F}_t = \{F_{1t}, \cdots, F_{Mt}\}$ is now inputted into a Gaussian Mixture Model (GMM) to generate the vector of **affiliation probabilities** $\boldsymbol{P}_t = \{p_{1t}, \cdots, p_{it}, \cdots, p_{nt}\}$ with $p_{it} = P(X_t = c_i|\boldsymbol{F}_t)$, Eq. 3.17. The vectors of affiliation probabilities are connected by a Markov Structure, with π_i denotes the prior distribution of class c_i and a_{ij} the transition probability from class c_i to c_j. The sequence of most likely states $\{x_t, \cdots, x_T\}$ are estimated by the *Viterbi* algorithm [143]. 42

3.5 The Long Short-Term Memory (LSTM)-HMM architecture. Advanced on Figure 3.4, each signal segment (e.g. $\boldsymbol{S}_t$ for the t-th state) is now directly inputted into a LSTM structure to generate the vector of affiliation probabilities $\boldsymbol{P}_t = \{p_{1t}, \cdots, p_{it}, \cdots, p_{nt}\}$ with $p_{it} = P(X_t = c_i|\boldsymbol{S}_t)$. The vectors of affiliation probabilities $\{\boldsymbol{P}_1, \cdots, \boldsymbol{P}_T\}$ are then synthesised by the Markov model to produce the sequence of most likely states $\{x_t, \cdots, x_T\}$. 45
3.6 A bottlenose dolphin (*Tursiops truncatus*) equipped with a biologging sensor package on its back. 46
3.7 Classification performance of different classifiers over the number of features included. Top plot shows overall accuracy vs. number of features used, middle plot shows the average precisions of each classifier while the bottom plot shows average recall. Window size $W = 0.6$ seconds, window overlap $r = 0.12$ seconds (20% of W), animal data set TT-A, data train-test ratio is 70% - 30%. 49
3.8 Adaptive weighted sampling scheme applied on the relationship curve between *accuracy* and *number of features*. The overall number of samples took was 30% as the size of the search space (i.e. all possible choices of *number of feature*). Window size $W = 0.6$ seconds, window overlap $r = 0.12$ seconds (20% of W), using animal data set TT-A, data train-test ratio is 70% - 30%. 50
3.9 Classification performance of different classifiers over window size selections. Top plot shows best accuracy vs. window size used, while bottom plot gives the corresponding number of features used for the best accuracy. Best accuracy is found according to Eq. 3.3. Window overlap r is 20% of window size W, animal data set TT-A, data train-test ratio is 70% - 30%. 51
3.10 Classification performance of different classifiers over different train/test ratios. Top plot uses animal data set TT-A, bottom plot uses TT-B. Best accuracy is found according to Eq. 3.3. Window size $W = 0.6$ seconds, window overlap $r = 0.12$ seconds (20% of W). 52
3.11 Comparison between classification methods. Average classification accuracy (over 10 random runs) plotted against varying training data volume. Total data volume is 35 minutes. 53
3.12 Comparison between Bidirectional and Forward LSTM. Average classification accuracy (over 10 random runs) plotted against varying training data volume. Total data volume is 35 minutes. 54
3.13 State transition probabilities obtained over all data (35 minutes). The entry at i-th row j-th column denotes the state transition probability (%) from class i to j, which is also color coded. Note that the transition probabilities from each row sum to 1 (100%). . . 55
3.14 State transition probabilities obtained over one instance of training data set. The volume of this training set is 5.1 minutes. 55

4.1 System block diagram of the proposed multiagent localization and tracking system. . . 58

4.2 (**A**) Transformed top-down view of the experimental environment. The outline of the water at the bottom is shown in black, and at the surface in yellow. Despite the size difference in the image, their projections to the world horizontal plane overlap each other. (**B**) Diagram of projection and refraction effects on estimated dolphin location. The yellow region represents the camera FOV. This cross section is shown in (**A**) as the white dash-dot line. (**C**) Diagram of tag (with state notations) and tagged agent. . . 61
4.3 Diagram of the Kalman filter based tracklet generation. Shown is the x-y view of example tracklets (red and green on gray lines) of two dolphins (highlighted light orange). 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 next-frame position ($\hat{\mathbf{p}}^{(j,t+1)}$), next-frame agent detection ($\mathbf{u}^{(j,t+1,i')}$), and Kalman-filtered next-frame position ($\mathbf{p}^{(j,t+1)}$). 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 ($\dot{\theta}^{(j,t')}$). 64
4.4 (A) Visualization of the tracking system with 30-second track plotted, including annotations of track components, where 'NC' stands for Non-Causal, 'PF' for Particle Filter. (B) Non-causally estimated 3D track of Agent 2 in second deployment, color-coded by depth. 72

5.1 (**A**): A bottlenose dolphin with a biologging tag. (**B**): Simplified free-body diagrams of an animal fluking and then gliding. (**C**): Simplified speed model of Fluke-and-Glide (FG) gait and continuous fluking gait over two cycles. 79
5.2 Representative data from a glide when the animal is wearing a tag. (**Top Left**) top view of the optimization surface used to select the model parameters (v_0 and C_d). Parameters that minimize fit residuals are indicated by the red circle. (**Top Middle**) RMS (root mean square) value of fitting residual versus initial speed v_0. (**Top Right**) RMS value of fitting residual versus drag coefficient C_d. (**Middle**) Measured (solid black) and the corresponding dynamic model prediction of the speed using the parameters selected using optimization. (**Bottom**) A representative iteration of the resampling and refitting to illustrate how bootstrapping (Section 5.4.5) was used to create a confidence interval for the drag coefficient. 84
5.3 Illustration of the experiment (**A**) conducted to collect kinematic data during different drag loading conditions (**B** and **C**): control, tag, tag + 4 and tag + 8. (**A**): Zinc oxide applied to the dolphin's skin was used to create the kinematic markers. Planar video data of the animals during gliding were collected using an underwater camera. (**B**): The modular system including the tag and drag elements, built to increase drag loading on the animals. (**C**): Expected net drag in the different drag loading conditions. 86

5.4 (**A**) Example video frame of a gliding dolphin with kinematic markers (zinc oxide) under the dorsal fin. (**B**) Pre-processing was used to enhance the visibility of the markers for automatic tracking of the data. (**C**) Tracked kinematic markers with the identified features are visualized with red circles. The average feature position is shown as a black triangle. The distance (Δx) between the two fixed kinematic markers as the dolphin glides through the frame was calculated for calibration and undistortion. (**D**) The horizontal position of animal in the video frame is shown on the x-axis. Pincushion distortion created by the lens results in as much as an 8-mm change in distance between the markers at the edges of the frame. 88
5.5 Illustration of the dolphin-tag model in the computational domain used for the simulations of all tag conditions. Velocity results from a representative simulation in 3 m/s steady state flow are presented in the figure. 91
5.6 Signal features of an example section of data (**A**-**D**) and three extracted gliding events' speed profiles (G1, G2, G3) with fitted dynamic models (**E**) are shown in the first 5 plots. Plot (**F**) presents the resulting 532 pairs of $\{Re, C_d\}$ values from all animals in grey markers (up triangle, down triangle, and square for the 3 animals respectively). The pairs have been clustered into 7 groups based on the Re values. The mean values (red circle) as well as the ± Standard Deviation (SD) range are shown for each group and a linear fit to all the mean values is represented by the black solid line. Results from low amplitude dolphin fluke (blue diamond, [101]) and low speed dolphin glide (black pentagram, Section 5.5.4) are also shown. 93
5.7 The speed distribution of all locomotion, sampled FG bouts as well as normalized FG bouts (Sections 5.3.4 & 5.5.2). 94
5.8 (**Top**) Average and individual initial velocities calculated during the parameter optimization for both animals. Error bars indicate one SD around the mean value. (**Bottom**) Mean and individual glide duration of each successful trial. Error bars indicate one SD around the mean value. 97
5.9 Computational Fluid Dynamics simulation results from the dolphin and combined dolphin-tag models that illustrate how the tag and added drag elements affect the velocity of the flow (**left**) and the resulting pressure differentials (**right**) around the model of the animal. 98
5.10 (**Top**) Net drag forces calculated from the Computational Fluid Dynamics simulation results. The control and tag conditions have similar net forces over the range of simulated speeds, but the additional of the drag elements greatly increases the force acting on the combined body. The yellow circle highlights the results near 1 m/s, a comparable speed to the gliding trials. (**Bottom**) A comparison of the estimated drag force acting on TT02 at speed of 1 m/s during the four experimental conditions over simulation (Sim) and experiment results that have been corrected for surface drag effect (Exp Cor) as well as the uncorrected raw drag force (Exp Uncor). The percent increase over the respective control trial is provided for all of the tagged conditions. The 90% confidence interval is also plotted with the experimental data. 100

5.11 Drag coefficients of dolphins from historical data compared to this study. Data from this study follow the same convention as our other figures (black circles = no tag; blue triangles = tag; green diamonds = tag+4; red squares = tag+8). The uncertainty ellipse around each of the experimentally derived drag coefficients was formed using the 90% confidence bound calculated using bootstrapping and the SD of the initial velocity data. Unfilled symbols represent results from Computational Fluid Dynamics. Results from this work are shown alongside data from the literature as presented in [101]. Experiments on rigid models, towed bodies and gliding animals (maroon circles), from hydrodynamic models based on swimming kinematics (orange circles), from a rigid 'dolphin' model with the shape of a solid of revolution of the NACA 66 series (yellow circles), and from particle image velocimetry (teal circles). As in [101], the black line represents the frictional drag coefficient for a flat plate with turbulent boundary layer flow and the dashed line is for a flat plate with laminar boundary flow. 105

6.1 Monitoring camera's view of the experimental environment (Section 6.2.1). 110
6.2 Example detections of a dolphin, floating enrichment (a buoy), and people in the lagoon environment. Each detected object is shown with its corresponding mask and the likelihood of detection. 111
6.3 Detections and tracklets are shown for four animals in the lagoon. 112
6.4 An illustration of how waiting time was defined for the enrichment. Eight time steps are shown for a dolphin (blue) and a piece of enrichment (orange). A point of interaction occurs at t_5 when the dolphin detection is within 2 m of the enrichment at the same time. The waiting time is then defined as the time from the initial detection of the animal in its current tracklet (t_1) to the point of interaction (t_5). 115
6.5 Results comparing the three environmental conditions: baseline, enrichment, and people. (**A**) Quiver plots for all three conditions show counter-clockwise movement in the lagoon, with the most concentric pattern occurring during the baseline condition. (**B**) Position heatmaps for the conditions show how the presence of enrichment or people modified where the animals spent time in the environment. (**C**, **D**) Probability density functions for the speed and heading rate of the animals derived from the tracklets. The animals tended to swim faster during the baseline condition, and spent more time stationary when people where present in the environment. Mean heading rates for all three conditions were positive, indicating counter-clockwise swimming trend. 116
6.6 Animal movement when enrichment was present in the environment. (**A**) The quiver plots show a more varied movement pattern when enrichment is initially introduced. (**B**) The heatmaps also indicate that the use of the environment during the first 20% of the time with enrichment was more variable. For all three heatmaps, 1 in color bar represents the maximum occupancy density based on the corresponding conditions. (**C**, **D**) Probability density of tracklet speed and heading rate show that the animals tended to move faster during the middle and last 20%, and preference for counter-clockwise swimming was not as strong when the enrichment was initially introduced. 119

6.7 Average waiting time and SD (**top**) and number of enrichment interactions (**bottom**) are shown for each 200-second-wide bins for two types of enrichment (Buoys and Jolly Balls). The waiting time for buoys is lower than that for Jolly Balls in general (note that the y-axes for waiting time are on different scales). The numbers of interactions for both enrichments show decreasing trends as the trials progress. 120
6.8 Continuous 3D tracks of 3 animals during 3 trials. Each trial lasted 30 minutes. The depth information of each animal is color coded, with light yellow indicates surfacing and dark blue for underwater locations. 122
6.9 State distributions of three tagged animals during three focal trials. **Steer** distributions are shown on the left while **Pose** distributions are shown on the right. 123
6.10 State transition matrices of animal TT01 during the three trials (Trial A, B, and C). **Steer** transitions are shown on the left while **Pose** transitions are shown on the right. The number/element at the i-th row and j-th column characterize the transition probability (%) from state i to state j, i.e. Eq. 3.5: $a_{ij} = P(X_{t+1} = c_j \mid X_t = c_i)$. The transitions of animals TT02 and TT03 are shown in Figure 6.11. 124
6.11 State transition matrices of animal TT02 and TT03 during Trial A and B. The transitions of animal TT01 are shown in Figure 6.10. 125

7.1 The scope (**left**) and structure (**right**) of this dissertation, modified from Figure 1.1 from Chapter 1. This work brings us one step closer to the further goal of studying bio-inspired artificial multi-agent systems by achieving the solid black arrows in the chart. Yet dashed red arrows represent a connection that has not been explicitly made in this dissertation, but is important to be made in the future. Particularly, the energetic reasoning (right side) needs to be considered while analyzing the coordinated behavior. And making the connection between dolphin coordination and artificial multi-agent systems (left side) can be very insightful to both fields. 136

LIST OF TABLES

2.1 Resulting pose angle differences with different conditions and methods. The direct difference (e.g. $A - B$) is used to detect a bias in the difference, while the absolute difference (e.g. $|A-B|$) returns the magnitude of the difference. Pose angle differences are in degree, each cell gives mean ± standard deviation. 25

3.1 Top 30 features in ranked feature list. Window size $w = 0.6$ seconds, window overlap $r = 0.12$ seconds (20% of W), using animal data set TT-A. Feature type plus signal channel forms a feature. 47

4.1 Summary of detection and tracking performance. 71

5.1 Characteristics of FG Gait vs. Continuous Fluke Gait under varying Duty Factor and Gliding Duration 94

5.2 Weight (kg) and length (m) of the two bottlenose dolphins in the study, their estimated wetted surface area (calculated from [171]) and the number of glides they each successfully performed for each experimental condition. Girth was measured approximately two fingers in front of the dorsal fin after animals exhaled. 95

5.3 Drag coefficients (C_d) calculated from the individual fitting to speed data, the number of successful trials, and the resulting fit residuals (R). The results from fitting to all of the data for a given trial (C_{d_all}) as well as the corresponding drag coefficient that has compensated for surface drag ($C_{d_all_correct}$), the 90% confidence interval are presented below. 96

5.4 Drag coefficients (C_d) calculated from the individual fitting to speed data, the number of successful trials, and the resulting fit residuals (R). The results from fitting to all of the data for a given trial (C_{d_all}) as well as the corresponding drag coefficient that has compensated for surface drag ($C_{d_all_correct}$), the 90% confidence intervals are presented below. 99

6.1 Summary measures calculated from the tracklets for the three conditions (baseline, enrichment, and people). Mean, SD, and the results of the K-S tests are reported. The enrichment and people conditions are compared with Baseline in the K-S tests (Section 6.2.4.2). 117

6.2 Statistic of the tracklets' speed and heading rate for the first 20%, middle 20%, and last 20% time sessions under the enrichment condition. All three sessions were compared with the Baseline condition in the K-S tests (Section 6.2.4.2). 117

6.3 The gait statistics of three tagged animals in three focal trials. 126

LIST OF ABBREVIATIONS

AI Artificial Intelligence

IA Intelligent Analysis

LSTM Long Short-Term Memory

RNN Recurrent Neural Network

HMM Hidden Markov Model

wrt with respect to

GPS Global Positioning System

IMU Inertial Measurement Unit

MEMS Micro-Electro-Mechanical Systems

DOF Degrees of Freedom

GMM Gaussian Mixture Model

SVM Support Vector Machine

KNN K-Nearest Neighbors

DT Decision Tree

NB Naive Bayes

PDF Probability Density Function

EM Expectation-Maximization

PF Particle Filter

CNN Convolutional Neural Network

CFD Computational Fluid Dynamics

MECOT Mechanical Energy Cost of Transport

FG Fluke-and-Glide

SLAM Simultaneous Localization and Mapping

ODBA Overall Dynamic Body Acceleration

SD Standard Deviation

ABSTRACT

As an intelligent species, bottlenose dolphins (*Tursiops truncatus*) demonstrate sophisticated coordinated behavior. This behavior depends on each individual's temporal behavior profile and the spatial locations of the individuals over time. An individual animal's decision-making in the environment is also influenced by its own dynamics and energy cost. As a result, the study of such a coordinated system requires a framework to synthesize multiple streams of information with appropriate models for extended periods. However, tools to investigate the fine-scale dynamical behavior of multiple animals during coordinated tasks are lacking. To address this gap, we present Intelligent Analysis (IA) methods that leverage experimental data to study the coordinated behavior of bottlenose dolphins and investigate how energetic cost may influence behavior.

Specifically, this dissertation addresses the following research questions: **Q1**: *Can we develop a data-driven analysis framework that enhances dynamical behavior characterization and classification while minimizing the need for human input?* **Q2**: *Can existing perception and localization tools be extended to a framework for tracking multiple uncontrollable agents within a dynamic underwater environment with quantified uncertainties?* **Q3**: *What are the energetic trade-offs of different swimming gaits measured using a biologging tag (wearable sensors for animals)?* **Q4**: *How do varying environmental conditions or social coordination result in changes in the behavior and swimming gait of an animal?*

This work is at the intersection of machine learning, computer vision, sensor fusion, data-driven modeling, animal behavior, and biomechanics. The AI-driven analysis framework developed in this work was used with data from animal-borne biologging tags and cameras placed in the environment to derive interpretable information about an animal's dynamical state (e.g. gait,

pose, position in the environment, and energetics). Knowledge of these dynamics facilitated the investigation of animal behavior during interactions with conspecifics and environmental features. The resulting identity preserving hour-scale tracks for each individual, together with their dynamic profiles, demonstrate that individuals respond to the same environmental features very differently and that coordinated events significantly modify an individual's behavior. Research outcomes from this work contribute to an improved understanding of bottlenose dolphin behavior and biomechanics and can be applied to biologging tag data from other cetaceans both in managed settings and the wild. Further, the algorithms and framework developed here can be extended for applications ranging from state estimation in vehicles to multi-target tracking in robotics.

CHAPTER 1

Introduction

1.1 Intelligence, Coordinated Behaviors, and Biomimicry

As humans, we are all scientifically classified as *Homo sapiens*, that is 'man' (*Homo*) the 'wise' (*sapiens*) in Latin. Intelligence is so important to us that we use its level to distinguish us (*Homo sapiens*) from earlier species of man, e.g. *Homo erectus* ('Upright Man'), *Homo rudolfensis* ('Man from Lake Rudolf'), and *Homo ergaster* ('Working Man') [1]. Trying to understand how a person or an animal can perceive, process, and operate in a world that is far more complicated than the subject itself is one of our fundamental curiosities. As a result, the study of intelligence has a long history [2, 3]. What is more, scientists and researchers are not only working on understanding intelligence but also building intelligent entities. The emerging field of Artificial Intelligence (AI), focuses mainly on how an artificial system could perceive the environment and act rationally to achieve the best expected outcome [4, 5]. As one of the newest fields in science and engineering, AI rose to prominence in the late twentieth century and is continuously drawing more attention as the next 'big thing' following the internet revolution. While the studies of human intelligence and artificial intelligence have been historically influenced by and benefited from each other [5], research on animal intelligence can also benefit from and contribute to AI.

Animal intelligence, often considered to be under the umbrella of general animal cognition, takes many forms from skill learning and problem solving to language utilization and coordination [6]. Some of the most widely recognized intelligent animals include ravens, crows, parrots, chimpanzees, gorillas, elephants, pigs, dogs, octopuses, and dolphins [7]. Coordinated animal behavior has attracted interest from both researchers and the general public because of the intricacy of an individual's behavior and the system level interactions between animals [8]. Examples of coordinated groups in nature include ant trails, fish schools, bird flight formations, honeybee swarms [8], large chimpanzee hunting groups with structured levels of cooperation [9], male bottlenose dolphin alliances to coordinate the herding of females for mating [10], and wolves hunting in packs [11, 12]. Coordinated behavior often enables more complex and dynamic objectives, which could

inform the coordination of engineered systems.

Researchers have long been interested in using nature as inspiration for technology. The domain of biologically inspired work covers a wide range of engineering and science, from structural materials [13] and optimization algorithms [14, 15, 16] to the control of engineered flapping flight [17], various forms of robotics [18, 19] and multi-agent systems for reconfigurable manufacturing systems [20]. Multi-agent systems, in particular, can be arguably seen as the counterpart to animal coordination in AI, which are composed of multiple computing elements, called *agents*, that are capable of taking certain autonomous actions and interacting with other agents [21]. The research field of multi-agent systems has grown rapidly over the decades since being established in the 1990s, with the belief that the agent-based paradigm can effectively represent the variety of distributed systems around us, and addressing challenges associated with these systems [21]. It is therefore worthwhile for us to seek inspiration from natural coordination to further our design in artificial systems.

1.1.1 Scope of This Dissertation

The work presented in this dissertation lies at the intersection of coordinated animal behavior and AI and seeks to leverage and advance techniques from AI and robotics towards the Intelligent Analysis (IA) of bottlenose dolphins. The long-term goal is to use our understanding of coordinated dolphin behavior to inform the exploration of bio-inspired multi-agent control strategies, Figure 1.1-Left. As an intelligent species, bottlenose dolphins demonstrate sophisticated coordinated behavior. But an understanding of the fine-scale dynamical behaviors during coordination is currently lacking due to technical challenges associated with persistent monitoring of the highly dynamic marine animals. Automated tools are needed to advance data processing and analysis abilities to the next level to enable the investigation of cooperative multi-animal systems from a new perspective that is persistent, detailed, and featurizable[1]. The work presented in this dissertation addresses this gap by creating an analysis framework that draws from machine learning, computer vision, sensor fusion, and data-driving modeling to intelligently parse data from biologging tags[2] and cameras into interpretable information about an animal's dynamical status (i.e. gait, motion states, positions in the environment, and energetic). This information enables the investigation of animal behavior during coordination. Four research questions are addressed in this work:

> **Q1**: *Can we develop a data-driven analysis framework that enhances dynamical behavior characterization and classification while minimizing the need for human input?*

[1] Being able to turn abstract information into numerical feature vectors for further statistical, data-driven, or machine learning based processing.

[2] The animal-borne 'wearable sensors', to be introduced in Section 1.3.

Q2: *Can existing perception and localization tools be extended into a framework for tracking multiple uncontrollable agents within a dynamic underwater environment with quantified uncertainties?*

Q3: *What are the energetic trade-offs of different swimming gaits measured using a biologging tag?*

Q4: *How do varying environmental conditions or social coordination result in changes in the behavior and swimming gait of an animal?*

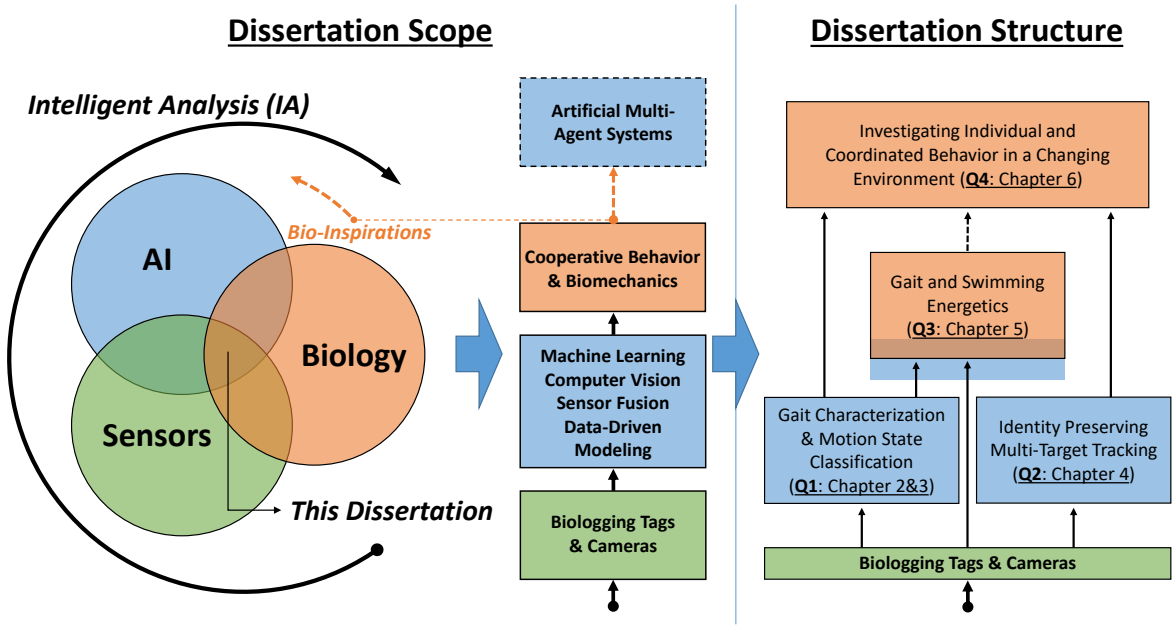


Figure 1.1: The scope (**left**) and structure (**right**) of this dissertation. This dissertation is at the intersection of sensor fusion, machine learning, computer vision, data-driven modeling, biologging, animal behavior, and biomechanics. The work leverages and advances frameworks from AI to enable the IA methods used to extract the dynamic state of the animal. The framework was used to investigate the energetic cost of swimming gait and the coordinated behavior of bottlenose dolphins. This work brings us one step closer to the further goal of studying bio-inspired artificial multi-agent systems. Dashed arrows indicate a connection that has not been explicitly made in this dissertation, but is important future work.

1.2 The Common Bottlenose Dolphins (*Tursiops truncatus*)

Bottlenose dolphins (*Tursiops truncatus*[3], Figure 1.2) are a well-known cetacean, thriving in temperate or tropical marine waters around the globe [24]. While the observations and descriptions of these animals can be traced back to Aristotle, Oppian, and Pliny the Elder, modern studies of these animals have taken place in both the wild and managed marine environments [23, 24]. During our long history of interacting with dolphins, they have impressed us as highly intelligent animals that live and hunt in groups while being excellent mammal swimmers at the same time. They can hold a social network of more than 400 individuals, with males forming alliances at different (nested) levels, including stable alliances in small groups and labile alliances among bigger groups, which, to some degree, is reminiscent of human political organizations [25, 26]. Research has shown that the alliances between males are formed based on a cooperation-based social concept, to defend reproductive females against other alliances or to attack other alliances to improve access to the females [10]. Dolphins are able to classify alliance membership according to cooperative payoffs and fulfill their role accordingly.

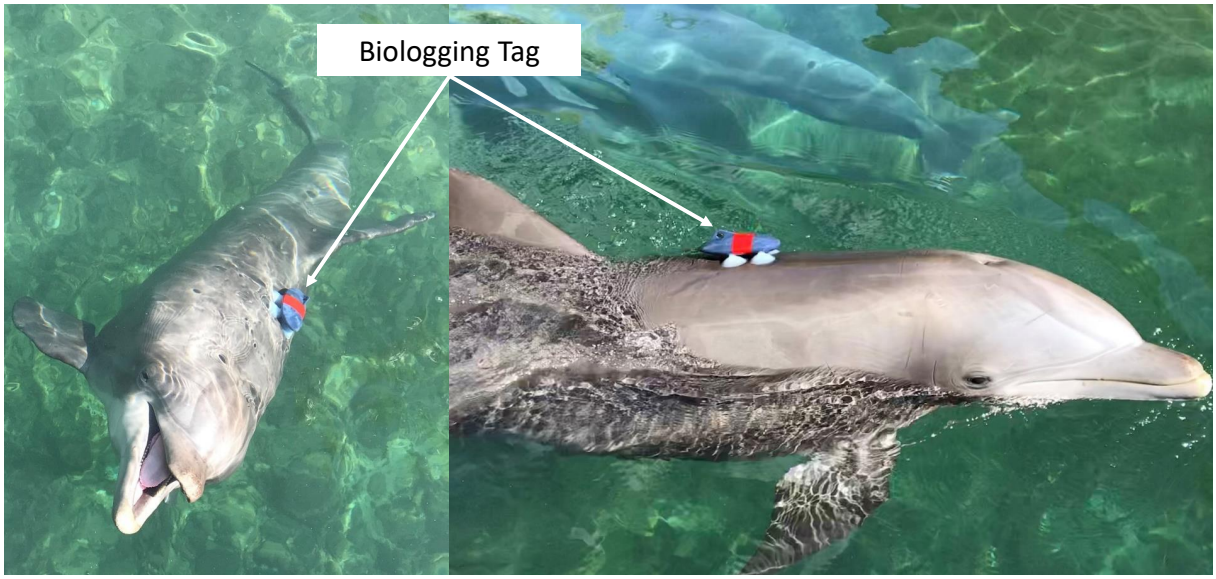


Figure 1.2: A bottlenose dolphin (*Tursiops truncatus*, Section 1.2) equipped with a suction cup based biologging tag (animal-borne wearable sensor package, Section 1.3) in Hawaii (*photo by D. Zhang*).

The high level of intelligence required for such sophisticated coordination is related to the large brain these animals have evolved [27]. Bigger brains require more energy, which requires more

[3]Scientific two-part (genus-species) Latin name of the common bottlenose dolphins, to be distinguished from the Indo-Pacific bottlenose dolphins (*Tursiops aduncus*). Due to certain taxonomic uncertainties within the genus *Tursiops*, *Tursiops aduncus* were not distinguished from *Tursiops truncatus* in some studies [22]. For those who are interested, *Tursiops* stands for 'dolphin-like' and *truncatus* means 'shortened', so *Tursiops truncatus* is 'a dolphin-like (animal) with shortened (teeth)' [23, 24]. *Tursiops aduncus* stands for 'a dolphin-like (animal) with a hooked (jaw)'.

food. Dolphins use echolocation, vision, and coordinated behavior when foraging. In addition to a standard pursuit and catch hunting method, multiple dolphins have been observed working cooperatively to herd fish into a ball to improve the foraging opportunity for the group [24, 28]. Animals have also been observed using features in the environment during foraging. Bottlenose dolphins in Florida Bay were recorded working together to form a mud plume ring to trap fish inside, and then catching fish trying to jump out of the trap [29]. Beyond the cooperation within the species, they were also observed feeding cooperatively with false killer whales (*Pseudorca crassidens*) in New Zealand [30] and even managed to 'team up' with local artisanal fishermen in Brazil [31]. In managed settings, studies have shown that bottlenose dolphins can communicate with vocal signals to solve a cooperative task designed by researchers [32]. These observational studies demonstrate that bottlenose dolphins are capable of learning and executing cooperative tasks. However, current work in the literature lacks quantitative analysis of dolphin coordination, and little is known about the detailed motion dynamics (position, speed, orientation) of each individual in the group during a coordinated event.

Measurements and estimates of animal dynamics would help answer questions related to animal energetics and behavior during coordination. For example, how much energy does an individual save when foraging in a group versus individually? How do animals move efficiently during a coordinated foraging event? How does each individual within the group position itself and regulate its motion so the attack from another alliance group can be defended? Or vice versa, what does each individual do during the entire process of an effective offense? The answers to these questions might directly connect to existing theoretical frameworks for artificial multi-agent systems and robotics, and even the frameworks yet to be created. Investigating these questions, however, requires the development of advanced sensing, modeling, and analysis methods to parse the fine-scale dynamics and spatial relationships of multiple individuals during a behavioral event.

1.3 Biologging Tags for Cetaceans

A common way to study animal behavior is through human observation, either directly in the field or by scoring recorded videos ([25, 33, 34, 35, 36]). This approach can be very detailed and precise when the animals are visible to the observer but does not scale with the demands for larger datasets to provide stronger statistical power. Relying on direct observation alone for studies of fine-scale cooperative behavior of marine animals is particularly challenging because underwater animals are not visible at all times and there are multiple individuals that must be tracked simultaneously. Fortunately, the rapid development of MEMS[4] technology makes wearable sensors[5] affordable, low

[4]Micro-Electro-Mechanical Systems, e.g. accelerometer, gyroscope, and pressure sensor.

[5]For example, the 'Apple Watch', 'Fitbit', 'Amazon Halo' etc.

power, accurate, and reliable. The development of technology for mobile computing and communication facilitates persistent monitoring of animals (including humans). The subject of interest can be monitored for days or weeks in conditions where direct observation is not feasible (e.g. in the dark, underwater, or in the wild). The term 'biologging' is often used to refer to animal-borne wearable sensors that are used to record and/or transmit data about animal movement, physiology, or the environment [37]. Biologging tags (e.g. Figure 1.2) have become an increasingly popular tool for the investigation of animal behavior and locomotion in their natural habitat over the past decades [38]. A wide variety of research topics on cetaceans have also been facilitated by biologging tags in recent years. For example, acoustic tags were used to study the bioacoustics and foraging of marine mammals [39, 40, 41]. The locomotion biomechanics, time budget, and diving behaviors of cetaceans have been studied using biologging tags as well [42, 43, 44, 45, 46, 47, 48]. These tag systems have also enabled position tracking of marine animals [49, 50, 51] to provide information about animal behavior in terms of habitat usage.

While the development and use of biologging tags have grown rapidly in the past two decades, the data processing capability has not yet caught up. Biologging tag datasets were often complemented with the direct expert observation of the tagged animals or expert interpretation of the data to identify specific behaviors that can then be described and quantified in detail. As datasets become larger, the inefficiency of these manual approaches limits the benefits from the available information. Custom software scripts have been developed for specific analysis, as will be discussed in detail in Chapters 2 and 3, but are limited in functionality and accessibility. As the biologging community grows, systematic development and public presentation/dissemination of the data processing tools are necessary to maximize the information extracted from these datasets. In particular, intelligent analysis of cooperative behavior of bottlenose dolphins requires automated analysis tools that can reliably parse the multi-channel tag data into 'interpretable' information about the dynamics and behavior of the animal, while minimizing human effort. These tools will not only facilitate our own studies but can also be used by other researchers, including the field experts, who would benefit from more readily available analysis tools.

With these objectives in mind, the first research task addresses the following question: (**Q1**) *What data-driven methods can be developed or leveraged for enhanced dynamical behavior characterization and classification while minimizing the need for human input?* Chapter 2 first introduces the custom/specific biologging tag used in this dissertation, followed by the formulation of a data-driven analysis framework to estimate animal pose and extract key gait parameters from the data. This framework characterizes the pose and gait of the animal by exploiting features in the tag data that result from the animal's motion. And it is robust to relative orientation uncertainty between the tag and animal body. Chapter 3 presents a systematic machine learning framework for signal feature generation and the selection and sequential motion state classification, which re-

duces demands on user knowledge for feature engineering and machine learning yet maintains the high accuracy of the system. Chapter 3 subsequently demonstrates the feasibility of replacing all of the operations related to feature engineering with a Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN), which is coupled with a Hidden Markov Model (HMM) structure at an upper level. The HMM structure helps with achieving higher accuracy by considering the temporal relationships of states at the upper level, yet, keeping the upper level a 'white box'.

1.4 Computer Vision for Dolphins: Detection and Tracking

The nature of cooperative behavior results in a problem that depends on individual temporal behavior profiles and the spatial relationships between individuals that evolve over time. Building on the work presented here on parsing temporal behavior profiles of an individual animal (Chapter 2 and Chapter 3), this dissertation addresses how to characterize and quantify spatial relationships between individuals at the sub-meter-scale during behavioral events. This task addresses the question: (**Q2**) *Can existing perception and localization tools be extended into a framework for tracking multiple uncontrollable agents within a dynamic underwater environment with quantified uncertainties?*

While many methods can be used to locate a particular individual in its environment over time, the accuracy and required sensor data vary by approach. Data from biologging sensors, like acceleration and heading, can be integrated to form continuous dead-reckoning tracks. However, track drift created by accumulated sensor errors severely limits estimated positional accuracy [52]. Additional information is required to enhance the accuracy of these tracks. Global Positioning System (GPS) is a straight forward technique for obtaining absolute positional information and can be combined with dead-reckoning information to obtain finer tracks [51]; however, it is not available underwater or indoors, nor is it accurate enough for meter-scale applications. Setting up landmarks/beacons with known positions for observers to perceive is often very effective (e.g. [53, 54, 55]), but this approach only applies to limited types of systems and can require additional infrastructure and sensors. Using an observer as a vehicle to sense and/or reconstruct the environment itself, such as the kidnapped robot [56] and Simultaneous Localization and Mapping (SLAM) [57, 58, 59] problems, can also improve the localization[6] of an observer. However, the required online computing power for data processing, as well as the power and size requirements of the sensors (e.g. camera, lidar, etc.), make these localization systems currently impractical for on animal tagging systems.

Alternatively, cameras placed in the environment have good potential for tracking multiple dol-

[6]The problem regarding finding the position of an observer with respect to its environment. In this work, the biologging tag is effectively an observer.

phins in habitat during an event. The manually parsed video has provided insightful observations about the behavior and dynamics of these animals [10, 60]. But automated tools are needed to investigate the dynamics of a cooperative multi-animal system in a featurizable manner.

Advancements in computer vision and deep learning [61] in the past two decades provide the foundation for the development of an automated tracking system for marine animals. With appropriate training data and configurations, many of the well-established object detection frameworks [62, 63, 64, 65, 66, 67, 68] can be used to reliably detect, locate, and classify objects of interests (e.g. dolphins, people, buoys or a bicycle if needed) that are visible within each image frame. However, the challenge in this work requires that we link detections in consecutive image frames into dynamic tracks across the video while maintaining the identity of each track through the entire event. This is not an easy task, particularly in a marine environment.

Significant environmental occlusions (e.g. surface glare) make it nearly impossible for a camera-based multi-target tracking system to continuously track an object of interest, especially in the case where multiple similarly-looking highly mobile objects (e.g. dolphins) are moving independently in the environment. Popular tracking techniques are limited in these situations. For example, the Joint Probabilistic Data-Association Filter (JPDAF) [69] tracks a fixed and known number of objects without any merged measurements in its basic implementation. Additionally, Multiple Hypothesis Tracking (MHT) [70, 71, 72] can be used to achieve state-of-the-art performance when augmented with higher-order appearance information about each object [73]. However, long-term occlusions from other objects or the environment are not accounted for in this approach, and the objects need to appear visually distinct to assist the tracking. The Explicit Occlusion Model (EOM) is constructed under a network flow framework to track long-term, inter-object occlusions in [74], though occlusions from the environment are not addressed. In [75], appearance cues of objects are used to prevent identity switching, which is not applicable to indistinguishable objects (e.g. dolphins). The main limitation of these approaches is that there is no information about the object during occlusion, and broken associations remain broken.

To address this technical challenge, Chapter 4 introduces a particle filter based approach to perform persistent identity preserving tracking of multiple bottlenose dolphins in their habitat using the above-habitat camera and on-body biologging tags. The chapter includes the implementation of a Kalman filter based tracking method that uses only camera data to generate multiple unassociated partial tracks of the animals. These partial tracks, together with the estimated pose and speed of the animals from tags (Chapter 2), are used to generate continuous estimates of the tagged animals in the environment.

1.5 Dolphin Locomotion and Dynamics

Animal behavior is related to and constrained by its dynamics, environment, and energetic cost. The ability to model and characterize individual dynamics provides insight and improved understanding of the observed behavior. In nature, efficient locomotion is essential for animals that travel long distances, and various morphological or behavioral adaptations have been observed to this end [76, 77, 78]. Dolphins, in particular, are well equipped to be excellent swimmers. The muscular body with a streamlined shape reduces hydrodynamic drag and resembles engineered 'laminar' profiles [79]. Dolphin skin includes certain features to further increase the animal's hydrodynamic performance underwater [80]. Dolphins can regulate their behavior to improve locomotion efficiencies. For example, they have been observed leaping out of the water to reduce drag and save energy at high speed [81], or riding on the waves generated by boats [82]. Young dolphin calves have been observed 'drafting' next to their mothers to keep up with a rapidly moving group of animals [83].

Different gaits can also be used by an animal to conserve energy, without the need for enabling external conditions (e.g. being near to the surface, having a ship passing by, or being close to their mothers). Intermittent locomotion [84], where an animal employs a gait composed of periods of active flapping/stroking/bursting and inactive gliding/coasting, is arguably a convergent evolutionary trend in both aerial and marine animals [78] that has been observed with birds [85, 86] and fish [87, 88, 89] during their daily activities. While the energetic benefits of intermittent locomotion have been studied for certain fish from various perspectives [87, 88, 89, 90, 91, 92, 93, 94], little work has been done on dolphins. Because dolphins are excellent swimmers, and burst-and-coast swimming is not always energetically beneficial in fish [94], it is important to investigate the energetic trade-offs of intermittent locomotion for these animals and identify when they select such gaits. For cetaceans, published work has tended to focus on how animals extend dive duration and conserve energy by taking advantage of buoyancy changes while gliding during descent or ascent in a dive [42, 95, 96]. However, a model-based investigation of the energetic benefits of intermittent horizontal locomotion for transit, specifically a fluke-and-glide (FG) gait, is lacking for these animals. Further, how animals regulate their gait in the presence of other animals or features in the environment is also an open question.

The evaluation of energetic cost requires an accurate estimate of the hydrodynamic drag acting on the animal during swimming. Various methods have been used to obtain such estimations for marine animals, including theoretical computations [79, 90, 91], computational fluid dynamic simulations [97, 98], towing experiments [99], camera assisted experiments [76, 42, 60, 100, 101], and field biologging tag deployments where speed can be derived from the data to estimate drag [43]. However, estimates of drag during gliding over a range of speeds for bottlenose dolphins are

currently lacking. Additionally, the hydrodynamic impact of biologging tags on these streamlined animals remains to be assessed.

These gaps are addressed in Chapter 5 of this dissertation, with the following research question: (**Q3**) *What are the energetic trade-offs of different swimming gaits measured using a biologging tag?* In Chapter 5, we first propose a model of the fluke-and-glide gait of the animal and review how the dynamic characteristics could be used to estimate passive drag. Then, a data-driven approach is introduced to estimate passive drag using measurements from a large biologging tag dataset, which further links the energetic cost of the animal with the gait parameters extracted from the biologging tag data collected from our first set of experiments. Subsequently, the second set of experiments and Computational Fluid Dynamics (CFD) simulations are presented to assess the hydrodynamic impact of biologging tags on bottlenose dolphins.

1.5.1 Environment, Coordination, and Behavior

Leveraging the developed IA framework, a pilot investigation of the coordinated behavior of bottlenose dolphins (discussed in Section 1.2) in a managed environment is presented in Chapter 6 to investigate animal behavior signatures during interaction with other animals and environmental features using information from cameras and the tag. The associated research question is: (**Q4**) *How do varying environmental conditions or social coordination result in changes in the behavior and swimming gait of an animal?*

1.6 Dissertation Structure

This dissertation aims at addressing the following research questions towards the IA of cooperative behaviors of bottlenose dolphins:

> **Q1**: *Can we develop a data-driven analysis framework that enhances dynamical behavior characterization and classification while minimizing the need for human input?*
>
> **Q2**: *Can existing perception and localization tools be extended into a framework for tracking multiple uncontrollable agents within a dynamic underwater environment with quantified uncertainties?*
>
> **Q3**:*What are the energetic trade-offs of different swimming gaits measured using a biologging tag?*
>
> **Q4**: *How do varying environmental conditions or social coordination result in changes in the behavior and swimming gait of an animal?*

The research outcomes from this work will not only contribute to bottlenose dolphin studies but also benefit the research of other cetaceans even broader species range, e.g. the gait characterization of a humpback whale or the tracking of multiple squids. The intelligent analysis of coordinated dolphin behaviors also brings us one step closer to the overarching goal of studying bio-inspired multi-agent cooperation strategies. Further, the algorithms and framework developed here can be extended for applications ranging from state estimation in vehicles to multi-agent tracking in robotics.

The rest of this dissertation is organized as follows (Figure 1.1-Right): Chapters 2 and 3 present the enabling data-driven tools to automatically parse biologging tag data into 'interpretable' information about the pose, gait, and motion states of the animal. Then, in Chapter 4, the video stream from overhead cameras is used together with tag data to achieve prolonged identity-preserving object tracking. Next, experiments and dynamic models are introduced in Chapter 5 to quantify the hydrodynamic impact of the tag as well as the relationship between gait and energetic cost of locomotion using tag data. Chapter 6 investigates animal behavior signatures during interaction with other animals and environmental features using information from cameras and the tag. Chapter 7 concludes this dissertation and discusses future directions.

CHAPTER 2

Estimation with Biologging Tags: Robust Pose-Gait Analysis and Characterization

2.1 Background

Biologging tags use a combination of sensors (e.g. accelerometer, magnetometer, gyroscope, speed, pressure sensor, camera, micro/hydrophone) to record data about animal movement, behavior, and the environment. For studies of cetaceans in particular, a wide variety of research topics utilize data from biologging tags. For example, tags are commonly used to study the bioacoustics of marine mammals [39, 40, 41]. The biomechanics, time budget, and diving behaviors of cetaceans are also often studied using biologging tags [43, 44, 42, 45, 46, 47, 48]. The kinematic data from these tag systems are also used to estimate the position of the animal in the environment and provide information about animal behavior in the context of habitat usage [49, 50, 51]. What is more, machine learning based approaches have been applied to tag data to infer the behavior of the animal when direct human observation/interpretation is expensive, time consuming, or not possible [103, 104]. Key to many of these studies is the ability to accurately estimate animal pose (roll, pitch and yaw, Figure 2.1).

Pose is particularly important for studies that require an estimated spatial trajectory of the animal or are investigating locomotion and gait [43, 46, 44, 49, 50, 51]. In the literature, pose is typically estimated using accelerometer and magnetometer data [39, 46]. When gyroscope data is available, methods like [105] can be used to improve the estimated animal pose. A swimming gait is essentially composed of a repetitive sequence of poses, but how to efficiently parametrize

Content of this chapter is mainly associated with the following publication:

[102] D. Zhang*, K. Goodbar, N. West, V. Lesage, S. Parks, D. Wiley, K. Barton, and K. A. Shorter, "Pose-gait analysis for cetaceans with biologging tag," *in Preparation for Submission*, 2021.

* Corresponding Author

the gait of the animal from a pose sequence remains an open question. Studies in the literature typically use pitch to identify and quantify gait using parameters like frequency, amplitude, and duty-factor [43, 46, 44]. But these pitch based gait descriptors are sensitive to the roll of the animal. For example, the pitch of an animal derived from accelerometer data captures little to no information about the animal's gait if the animal is fluking sideways (a 90 degree roll angle). To characterize an animal's gait, we need a measure with respect to (wrt) the animal's body, rather than the world. To address this, the measured angular rate from a gyroscope can be integrated to estimate the rotation angle wrt the animal itself. But, this estimation is subject to accumulated sensor error and gyroscopes are not available in some tag platforms. As such, it is important to develop approaches that can address this in accelerometer based estimates of orientation.

Because many cetacean tags use a suction cup based attachment method, the location and orientation of the tag wrt the animal can change during a field deployment [39, 106]. As a result, the relative orientation of the tag wrt the animal must be determined before data can be used for animal pose calculation [39, 46]. Further, even with the best practices for deploying tags [106], the relative orientation of the tag can shift during a deployment (i.e. tag slide on the animal, we refer to this event as a tag **shift**). Identifying these shifts in orientation is essential for determining the correct relative orientation. Currently, relative orientation tends to be determined manually or heuristically using portions of data where the animal's orientation can be inferred. For example, when an animal breathes at the surface (surfacing), it is assumed that the pose should not have significant roll. A surfacing event can be inferred from pressure data and the roll estimate can then be corrected accordingly [39, 46]. Tag orientation shifts tend to be identified by human inspection using sensor streams, like accelerometer data, to identify features, such as an impact to the tag by another animal or object [39]. However, the tag may be impacted by hydrodynamic forces, which can result in slowly changing orientation (slides) that are not captured by an accelerometer. Simulation and experimental studies have been used to estimate hydrodynamics forces that are acting on tags or imparted to the animal, but it is difficult to predict when the combined hydrodynamic and inertial forces resulting from animal motion will result in a slide [98, 97]. These heuristic methods require a level of knowledge and experience that can affect the reliability and repeatability of the results. Currently, an automated approach to identify these relative changes in orientation is lacking.

To address these gaps, this chapter presents an automated data processing pipeline (and software) that takes advantage of common characteristics of pose and gait of cetaceans to: (1) Identify time instances associated with the occurrences of relative orientation change between the tag and animal; (2) Identify the relative orientation of tag wrt the animal's body for a given data segment assuming no relative motion between the two; and (3) Extract gait parameters that are invariant to animal pose. Biologging tag data from a bottlenose dolphin, humpback whale, and beluga whale were used to validate and demonstrate the methods. The proposed analysis approach will facilitate

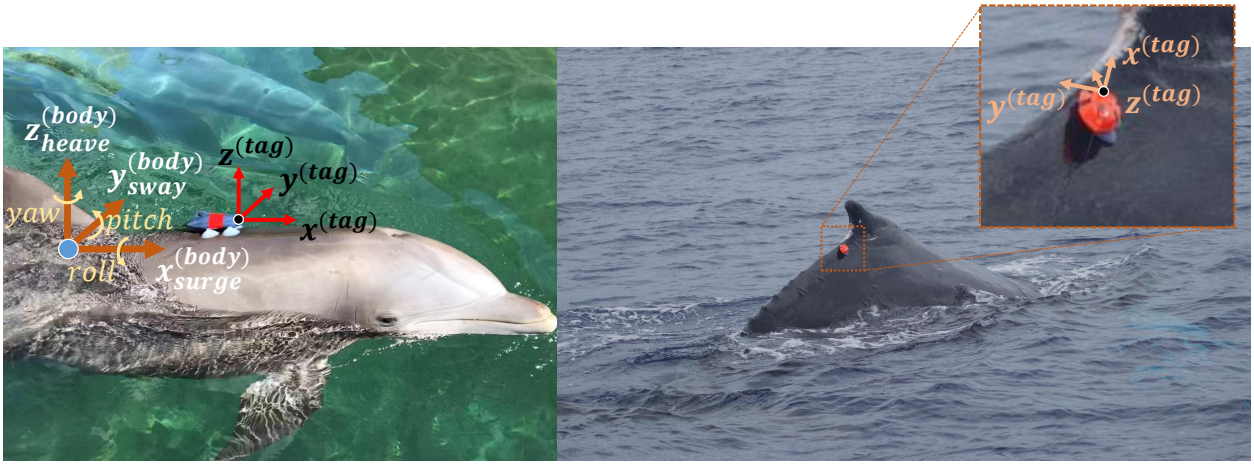


Figure 2.1: Biologging tag attached to a bottlenose dolphin (MTag, **left**) and a humpback whale (DTag, **right**) with coordinate definitions. Sometimes the tag's coordinates are aligned with the animal (**left**) but oftentimes they are not (**right**). It is also possible that a tag's position on the animal changes during a deployment. For applications like animal pose estimation, tag data needs to be represented in the animal's body coordinates during post processing, which implicitly requires the knowledge of relative orientation between tag and animal. Note that a positive animal pitch is a negative rotation around the body's y-axis under this coordinate system.

the use of biologging tags for the study of cetacean locomotion and behavior, and the method can be used directly with cetacean data from any tag platform equipped with an accelerometer, magnetometer, and pressure sensor. Discussion and suggestions related to data processing best practices are also provided.

2.2 The Biologging Tag: MTag

The main biologging tag considered in this dissertation is the MTag platform (Figure 2.1-Left) designed and built by the ESTAR Lab lead by Dr. K. Alex Shorter at the University of Michigan. The suction cup based tag has electronics encapsulated in 3D printed housings. Sensors on board include a 9 Degrees of Freedom (DOF) Inertial Measurement Unit (IMU) with an accelerometer, gyroscope, and magnetometer, as well as additional sensors on-board to record temperature and pressure. The board also supports the addition of an external 1-DOF Hall effect sensor, which functions as a flow speed sensor when used in conjunction with a magnetic micro-turbine mounted outside the tag housing [107]. The tags were configured to record the IMU at 50 Hz, and all other sensors at 10 Hz. To be minimally invasive and to minimize deployment complexity, the tags are attached to the animals with a set of four silicone rubber suction cups. Deployments can be done within a few seconds.

Methods presented in this chapter also support other cetacean tag platforms that are equipped

with an accelerometer, magnetometer, and pressure sensor. For example, the DTag platform [39] has been commonly used for wild cetacean deployments (Figure 2.1-Right). The DTag also attaches to an animal via suction cups and contains an accelerometer, magnetometer, pressure sensor, high performance hydrophone, but no gyroscope.

2.3 Pose-Gait Patterns in Tag Data and Assumptions

The proposed work uses biologging tag data (from a 3-axis accelerometer, 3-axis magnetometer, and pressure/depth sensor) and the general data patterns associated with cetacean gait to estimate animal pose. While an accelerometer measures acceleration due to both gravity and motion, we apply a simplified approach in this work and use the low-pass filtered 3-axis accelerometer data as the measurement associated with gravity in the tag coordinate frame ($\boldsymbol{A}^{(tag)}$). The magnitude and direction of the gravity force (vertically downwards) are known, and the resulting associated tag measurement always points vertically upwards in the world coordinates.

As the animal changes pose, the amount of the gravitational force measured by each of the component axes of the accelerometer changes in the tag coordinate frame ($\boldsymbol{A}^{(tag)}$), but the magnitude of the total signal is constant. We use $\boldsymbol{A}^{(tag)}$, together with rate of the depth change, to reveal the pattern associated with the animal's gait, and we refer to the plot as the **orientation sphere** (Figure 2.2, with tag aligned with animal). Even though the animal could be in any pose at a particular time instance, the authors assume that there are common characteristics about the general gait patterns.

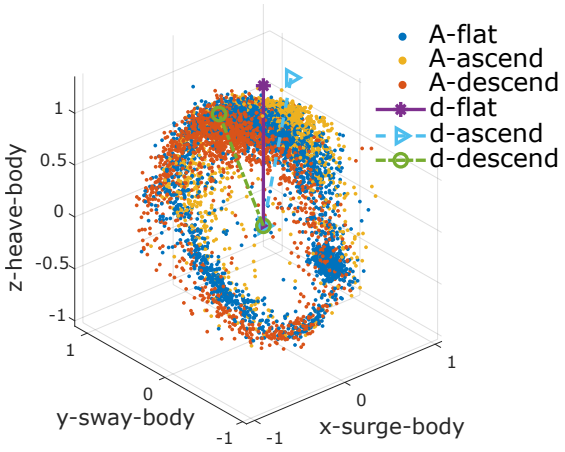


Figure 2.2: **Orientation sphere** for a bottlenose dolphin with the biologging tag aligned with the animal's body, (i.e. data in the animal's body coordinates without the need for a correction). Each data point is clustered into 'flat', 'ascend', or 'descend' depending on its depth speed. The dominant directions for the 3 clusters are marked by $\boldsymbol{d}_{flat}$, $\boldsymbol{d}_{ascend}$, and $\boldsymbol{d}_{descend}$, to be discussed in Section 2.5.

In particular, three assumptions are made about the gait of the animal and the associated data pattern under *general* situations. The first and most important assumption is that cetaceans swim with a neutral roll angle more often than either a leftward or a rightward roll. Second, it is assumed that the animal has positive pitch when ascending and a negative pitch when descending. Third, the pattern of one signal segment is assumed to match another segment for the same animal. Tag shift detection and orientation correction are based on these assumptions. More discussions shall be made about the solidness and influence of these assumptions through the rest of the chapter.

2.4 Nonparametric Tag Shift Detection

The third assumption stated above, that the pattern of one signal segment matches another segment for the same animal, will be violated when the relative orientation of the tag to the animal changes, i.e. the tag **shifts**, Figure 2.3. Even though significant animal gait changes could also lead to violation of the assumption, tag shift detection is achieved by identifying the time instances when signal patterns change. At a higher level, the method focuses on a comparison of data distributions from different data segments such that 'abnormal' segments in the dataset can be located. The comparisons are performed on individual data points, without the need for prior knowledge of the specific data distribution.

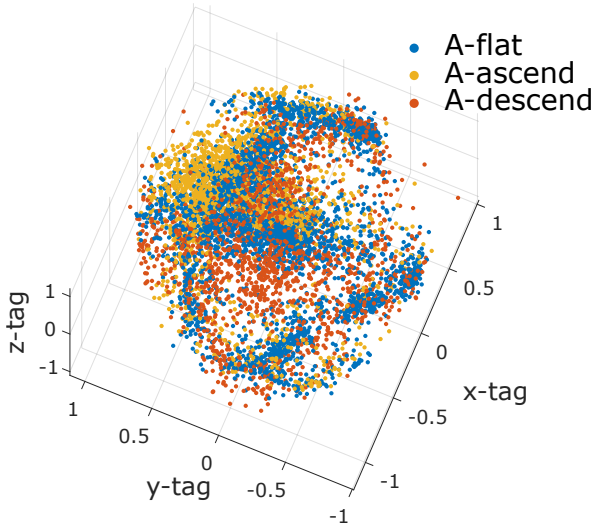


Figure 2.3: **Orientation sphere** for a bottlenose dolphin with the data in the tag coordinates before any correction. The presented dataset contains 4 simulated tag shifts.

Before solving the entire shift detection problem, we first describe our approach to a more constrained subproblem: how to identify a tag shift, if it exist, within a given signal segment S

that contains at most one shift. For this subproblem, we equally divide S into temporally adjacent segments S_1 and S_2, knowing that a shift could lie in either S_1 or S_2, but not both. Without loss of generality, we assume the shift lies in S_2. S_1 then serves as the template for identifying the time instance t_2 when the shift happens in S_2. With t_2, S_2 can be further divided into $S_2^{(t<t_2)}$ and $S_2^{(t\geq t_2)}$. Then, we know $S_2^{(t<t_2)}$ shares the same distribution as S_1 and $S_2^{(t\geq t_2)}$ diverges from S_1.

The question now is how to locate such t_2 in S_2. For each data point in S_2, we find its K[1] nearest spacial neighbors in S_1 and compute the average distance from the point to its neighbors. If this distance is within an empirically defined threshold, the data point in S_2 is considered an **inlier** of S_1, otherwise it is classified as an **outlier**. After this process, all of the points in S_2 are assigned with a value of 1 (inlier) or 0 (outlier). A temporal moving average filter is then applied to these 1s and 0s to obtain a local inlier percentage ($InPct$) value. The time instance t_2 is found by identifying the first time $InPct$ drops below a defined threshold. If $InPct$ never drops below, this indicates that a shift does not lie in S_2. This procedure can then be applied to check S_1 by using S_2 as a template, aiming at finding time instance t_1 in S_1 that corresponds to a shift.

With a strategy defined above to find one (or zero) shifting instance within two segments, S_1 and S_2, an algorithm for finding all tag shifts is presented in Figure 2.4. Built on the procedures described above, the algorithm starts with adjacent data segments S_1 and S_2. If neither segment contains a shift (branch 10), then S_1 remains unchanged and S_2 moves forward to the next data segment to check for a shift. Once a shift is detected, both segments will be redefined such that S_1 starts from the identified shift and S_2 adjacently follows (branches 6 to 9).

Every time S_1 is redefined (i.e. S_1 and S_2 are adjacent), the algorithm searches for a shift to exist in either S_1 or S_2, starting with S_2 (branch 1). If S_2 does not contain a shift (branch 3), then the algorithm checks S_1 for a shift (branch 5, followed by branches 9 & 10). If S_2 contains a shift instance t_2 (branch 2), and t_2 is not near[2] S_1 (branch 8), then t_2 is recorded as a shift. However, if t_2 is near S_1 (branch 4), the algorithm will also check S_1 to decide whether the shift is lying in S_1 but appears to be t_2 initially. Then either t_1 or t_2 will be recorded as the shift, depending the search result of S_1 (branches 6 & 7). When only S_2 is redefined (i.e. S_1 and S_2 are not adjacent), the algorithm only searches for a shift in S_2 (branches 8 & 11). These steps are repeated until the entire dataset has been searched (branch 12).

2.5 Pattern Based Orientation Correction

For a given segment of data that contains no tag shift but has an unknown tag-animal configuration, the tag data from accelerometer and magnetometer (and gyroscope, if equipped) can be rotated (i.e.

[1] An empirically determined constant value for the method. In this work, $K = 30$.

[2] The time interval is within an empirically defined threshold. In this work, the threshold interval is 3 minutes.

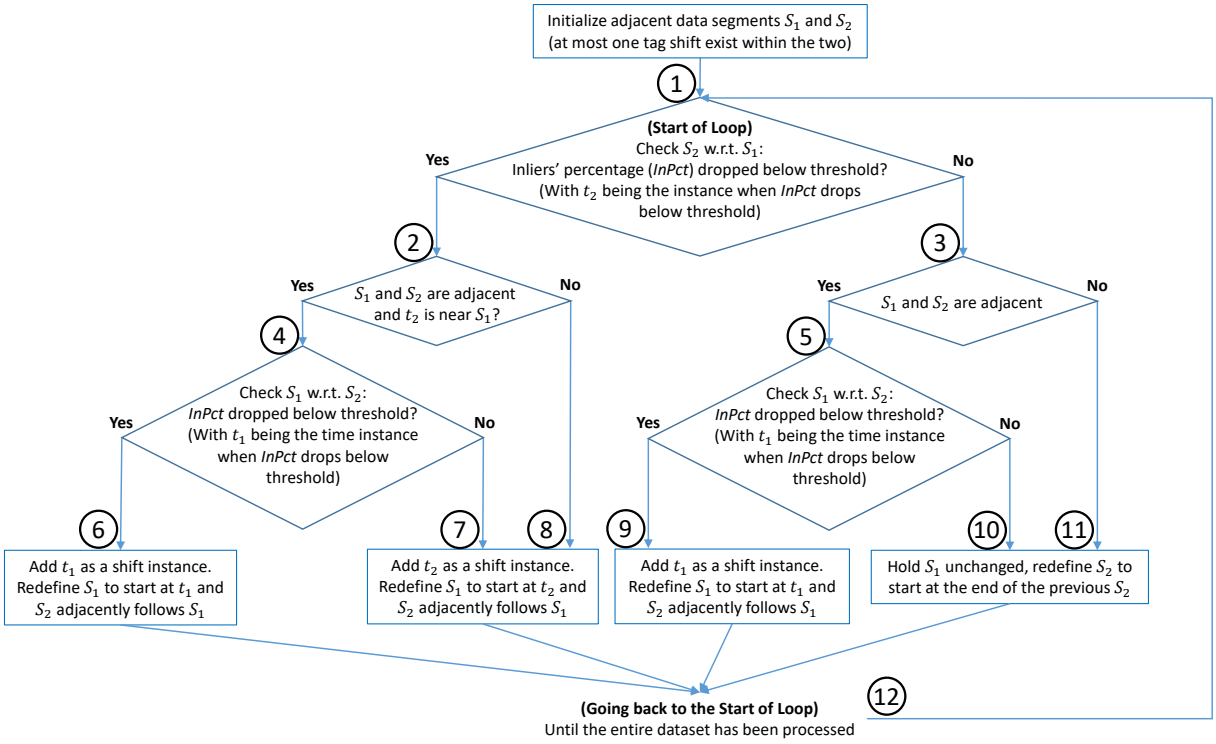


Figure 2.4: Tag shift detection algorithm. Each branch is marked by a circled numbers (1 to 12).

change of coordinates) to transform the data from tag coordinates to the animal's body coordinates. In this work, such a rotation is found by matching the measured general motion pattern, i.e. orientation sphere, to the assumed one, as discussed in this section. Depth measurement is used to group the tag coordinates' gravity measurements $\boldsymbol{A}^{(tag)}$ into three clusters: $\boldsymbol{A}^{(tag)}_{flat}$, $\boldsymbol{A}^{(tag)}_{ascend}$ and $\boldsymbol{A}^{(tag)}_{descend}$, in accordance with the animal swimming horizontally, ascending, and descending.

We further define the **dominant direction** as the gravity measurement direction associated with the most common pose of the animal for each of the three swimming conditions: $\boldsymbol{d}^{(tag)}_{flat}$, $\boldsymbol{d}^{(tag)}_{ascend}$, and $\boldsymbol{d}^{(tag)}_{descend}$. By the assumptions about the general motion of these animals (Section 2.3), the most common pose of a cetacean under each of the three swimming conditions is assumed to be known to us. In particular, $\boldsymbol{d}^{(tag)}_{flat}$ would correspond to the gravity measurement when the animal has zero pitch and zero roll, while $\boldsymbol{d}^{(tag)}_{ascend}$ corresponds to positive pitch and zero roll, and $\boldsymbol{d}^{(tag)}_{descend}$ corresponds to negative pitch and zero roll. In other words, we know how the dominant directions are orientated in the animal's body coordinates. Specifically, $\boldsymbol{d}^{(body)}_{flat}$ is aligned with the body's z-axis ($z^{(body)}$), and $\boldsymbol{d}^{(body)}_{ascend}$ lies in the first quadrant of the plane formed by the $x^{(body)}$ and $z^{(body)}$ axes, while $\boldsymbol{d}^{(body)}_{descend}$ lies in the second quadrant (Figure 2.2). So if $\boldsymbol{d}^{(tag)}_{flat}$, $\boldsymbol{d}^{(tag)}_{ascend}$ and $\boldsymbol{d}^{(tag)}_{descend}$ can be found in the tag coordinates (Figure 2.5), the rotations that map them to $\boldsymbol{d}^{(body)}_{flat}$, $\boldsymbol{d}^{(body)}_{ascend}$ and $\boldsymbol{d}^{(body)}_{descend}$ can be found and applied to map other data from tag coordinates to animal body

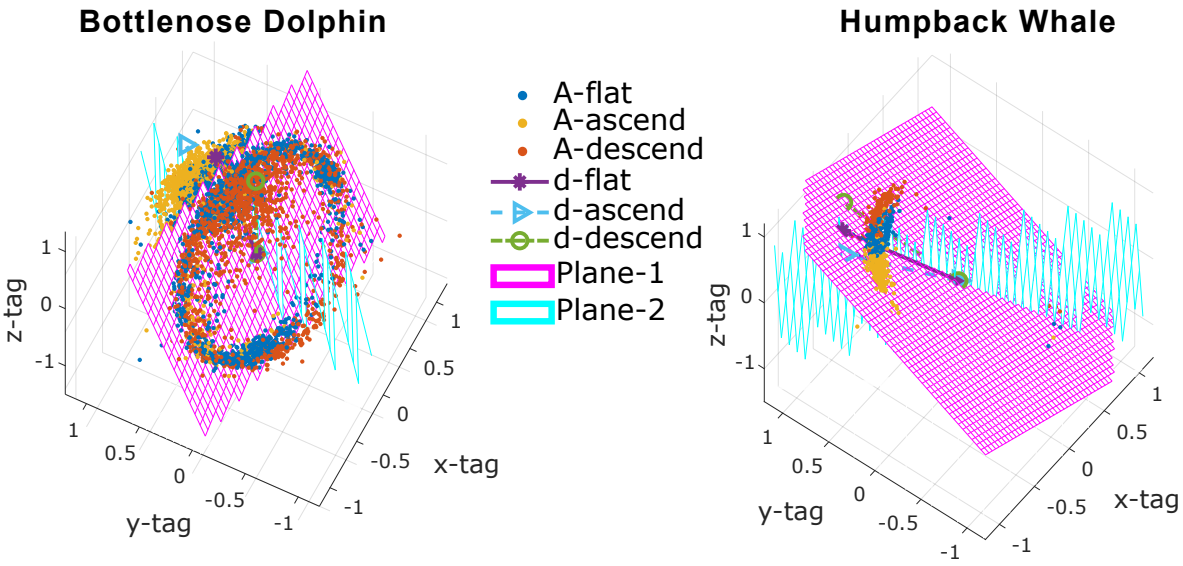


Figure 2.5: **Orientation sphere** for a data segment of a bottlenose dolphin (**left**) and a humpback whale (**right**). The plots visualize the orientation correction method being applied to a uncorrected data segment to find the dominant directions in the tag's coordinates. The dominant directions can then be mapped to their assumed directions in the body's coordinates, Figure 2.2.

coordinates (Figure 2.2).

To determine the dominant directions within uncorrected data, two perpendicular planes are fitted to $\boldsymbol{A}^{(tag)}$ using the RANSAC algorithm [108] (Figure 2.5), which is particularly robust to outliers as compared to least square based approaches. The two planes essentially correspond to two physical motions: rolling with zero pitch for the **rolling plane** and pitching with zero roll for the **pitching plane**.

Firstly, plane-1 (through the origin) is fitted to the most significant distribution in the data, which could be either $\boldsymbol{A}^{(tag)}_{flat}$ (e.g. Figure 2.5-Left, plane-1 is a **rolling plane**) or $\boldsymbol{A}^{(tag)}_{ascend}$ and $\boldsymbol{A}^{(tag)}_{descend}$ (e.g. Figure 2.5-Right, plane-1 is a **pitching plane**), depending on the animal and environment. Secondly, plane-2 (also through the origin) is fitted to the rest of the data under the constraint that it is perpendicular to plane-1. The toolbox can assign the roles of **rolling plane** and **pitching plane** to the two planes automatically.

The intersection line of the two planes is colinear with $\boldsymbol{d}^{(tag)}_{flat}$ while both $\boldsymbol{d}^{(tag)}_{ascend}$ and $\boldsymbol{d}^{(tag)}_{descend}$ lie in the **pitching plane**. If we further define $\boldsymbol{a}^{(tag)}_{flat}$, $\boldsymbol{a}^{(tag)}_{ascend}$, and $\boldsymbol{a}^{(tag)}_{descend}$ to be the average values of $\boldsymbol{A}^{(tag)}_{flat}$, $\boldsymbol{A}^{(tag)}_{ascend}$, and $\boldsymbol{A}^{(tag)}_{descend}$, then projecting $\boldsymbol{a}^{(tag)}_{flat}$ onto the intersection line produces $\boldsymbol{d}^{(tag)}_{flat}$, while projecting $\boldsymbol{a}^{(tag)}_{ascend}$ and $\boldsymbol{a}^{(tag)}_{descend}$ onto the **pitching plane** returns $\boldsymbol{d}^{(tag)}_{ascend}$ and $\boldsymbol{d}^{(tag)}_{descend}$ (Figure 2.5), respectively.

With $\boldsymbol{d}^{(tag)}_{flat}$, $\boldsymbol{d}^{(tag)}_{ascend}$, and $\boldsymbol{d}^{(tag)}_{descend}$ identified in the tag coordinates, the data can be put into the animal's body coordinates (i.e. $\boldsymbol{A}^{(tag)} \rightarrow \boldsymbol{A}^{(body)}$) with two rotations: first, a rotation ($\boldsymbol{R}_1$) to

align $\boldsymbol{d}_{flat}^{(tag)}$ with the body's z-axis ($z^{(body)}$, heave) and second, a rotation ($\boldsymbol{R}_2$) around the z-axis so $\boldsymbol{d}_{ascend}$ and $\boldsymbol{d}_{descend}$ are on the x-z plane with $\boldsymbol{d}_{ascend}$ on the positive x-axis side and $\boldsymbol{d}_{descend}$ on the negative x-axis side (Figure 2.2). Other data are mapped from tag coordinates to body coordinates using $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$:

$$\boldsymbol{A}^{(body)} = \boldsymbol{A}^{(tag)}\boldsymbol{R}_1\boldsymbol{R}_2, \tag{2.1}$$

$$\boldsymbol{M}^{(body)} = \boldsymbol{M}^{(tag)}\boldsymbol{R}_1\boldsymbol{R}_2, \tag{2.2}$$

with $\boldsymbol{A}^{(body)}$ and $\boldsymbol{M}^{(body)}$ being the transformed (animal's body coordinates) data from the accelerometer and magnetometer.

2.6 Animal Pose Calculation and Gait Characterization

Roll, pitch, and yaw are widely used to describe the pose of animals [39, 46]. In this work we are computing them in a conventional [39] way using the transformed (animal's body coordinates) data from accelerometer $\boldsymbol{A}^{(body)}$ and magnetometer $\boldsymbol{M}^{(body)}$. Define $\boldsymbol{R}_x(\alpha)$ to be the rotation matrix for rotating the data about the coordinates' x-axis with angle α. And similarly $\boldsymbol{R}_y(\beta)$ about the y-axis with angle β and $\boldsymbol{R}_z(\gamma)$ about the z-axis with angle γ. For time instance t, the 3-axes body coordinates' accelerometer and magnetometer data are represented as $\boldsymbol{A}_t^{(body)} = [A_{xt}^{(b)}, A_{yt}^{(b)}, A_{zt}^{(b)}]$ and $\boldsymbol{M}_t^{(body)} = [M_{xt}^{(b)}, M_{yt}^{(b)}, M_{zt}^{(b)}]$. Roll, pitch, and yaw are calculated as:

$$roll_t = \arctan(A_{zt}^{(b)}/A_{yt}^{(b)}), \tag{2.3}$$

$$pitch_t = \arcsin(A_{xt}^{(b)}/norm([A_{xt}^{(b)}, A_{yt}^{(b)}, A_{zt}^{(b)}]), \tag{2.4}$$

$$[M_{xt}^{(2)}, M_{yt}^{(2)}, M_{zt}^{(2)}] = [M_{xt}^{(b)}, M_{yt}^{(b)}, M_{zt}^{(b)}]\boldsymbol{R}_x(roll_t)\boldsymbol{R}_y(-pitch_t), \tag{2.5}$$

$$yaw_t = -\arctan(M_{yt}^{(2)}/M_{xt}^{(2)}), \tag{2.6}$$

where we note that the body coordinates' $x^{(body)}$, $y^{(body)}$, and $z^{(body)}$ axes are defined to be pointing forward, leftward, and upward wrt the animal's body, respectively (Figure 2.1).

Let a set of moving coordinates $\{x^{(move)}, y^{(move)}, z^{(move)}\}$ be initially aligned with the earth's inertial coordinates[3] $\{x^{(inertial)}, y^{(inertial)}, z^{(inertial)}\}$. Then yaw, pitch, and roll rotates the moving coordinates from the inertial coordinates to the animal's body coordinates in 3 steps:

1. Yaw corresponds to a positive rotation around the moving coordinates' current $z^{(move)}$ axis (same as the inertial $z^{(inertial)}$ axis at the moment), rotates the moving coordinates so $x^{(move)}$ is aligned with the projection of the animal's $x^{(body)}$ axis on the earth horizontal plane.

[3]For the inertial coordinates, we define $x^{(inertial)}$ points to magnetic north, $z^{(inertial)}$ points vertically up, and $y^{(inertial)}$ follows the right hand rule points to the magnetic west.

2. Pitch measures the angle between the body coordinates' $x^{(body)}$ axis and the earth horizontal plane, rotates around the moving coordinates' current $y^{(move)}$ axis and makes $x^{(move)}$ tilted from horizontal to be aligned with the animal's $x^{(body)}$ axis. Note that a positive pitch (i.e. animal head up) corresponds to a negative rotation around $y^{(move)}$, with $y^{(body)}$ defined pointing to the left of the animal.

3. Roll represents a positive rotation around moving coordinates' current $x^{(move)}$ axis (which is now aligned with the animal's $x^{(body)}$ axis) that brings moving coordinates' $y^{(move)}$ axis from horizontal to the animal's $y^{(body)}$ axis. After this step, $x^{(move)}$, $y^{(move)}$, and $z^{(move)}$ are aligned with $x^{(body)}$, $y^{(body)}$, and $z^{(body)}$, respectively.

This pitch estimation is commonly used to describe the fluking gait of the animal [43, 46, 44]. But because pitch only measures the angle between the animal's $x^{(body)}$ axis and the earth horizontal plane, it does not capture the gait characteristics well when the animal is rolling. The situation degrades more when the animal is fluking with a roll angle of 90° where the fluking motion barely shows up in the pitch measurement. Ideally, we want an angle measurement wrt the animal's own $y^{(body)}$ axis to describe its gait. Even though this can be done by integrating the gyroscope measured angular rate about the $y^{(body)}$ axis of the animal, a gyroscope is not available in all tag platforms and the integration is subject to measurement drift due to accumulated errors.

$\{x^{(body)}, y^{(body)}, z^{(body)}\}$ are moving wrt the earth inertial coordinates at all times following the animal's motion (both **high** frequency, e.g. fluking; and **low** frequency, e.g. turning from ascending to descending). We now define a new set of coordinates $\{\bar{x}^{(body)}, \bar{y}^{(body)}, \bar{z}^{(body)}\}$ that only follow the **low** frequency motion of the animal. Then $\{\bar{x}^{(body)}, \bar{y}^{(body)}, \bar{z}^{(body)}\}$ can be used to characterize animal's **high** frequency motion (e.g. fluking) wrt the animal itself, rather than wrt the earth's inertial coordinates.

Specifically, representing $x^{(body)}$ in $\{\bar{x}^{(body)}, \bar{y}^{(body)}, \bar{z}^{(body)}\}$ returns the dynamic heading of the animal wrt its own neutral body pose (e.g. head tilting up and down wrt its neutral body pose results in the vector $x^{(body)}$ swinging in the $\bar{x}^{(body)}$-$\bar{z}^{(body)}$ plane). We now refer to the vector $x^{(body)}$, represented in $\{\bar{x}^{(body)}, \bar{y}^{(body)}, \bar{z}^{(body)}\}$, as the **vector of dynamic pose** ($\boldsymbol{V}_{dp}$):

$$\boldsymbol{V}_{dp(t)} = [1,0,0]\boldsymbol{R}_y(-pitch_t)\boldsymbol{R}_z(yaw_t)\boldsymbol{R}_z^{-1}(\overline{yaw}_t)\boldsymbol{R}_y^{-1}(-\overline{pitch}_t)\boldsymbol{R}_x^{-1}(\overline{roll}_t), \tag{2.7}$$

where $[1,0,0]$ is unit vector pointing forward (surge) while $\overline{roll}_t$, $\overline{pitch}_t$ and $\overline{yaw}_t$ are the low-pass filtered roll, pitch and yaw values at time t. We note that a positive animal pitch is a negative rotation around the y-axis under the current axes definition. Within the equation, '$[1,0,0]\boldsymbol{R}_y(-pitch_t)$ $\boldsymbol{R}_z(yaw_t)$' finds the animal's current 3D heading in the inertial coordinates (i.e. $x^{(body)}$ in the earth inertial coordinates). '$\boldsymbol{R}_z^{-1}(\overline{yaw}_t)\boldsymbol{R}_y^{-1}(-\overline{pitch}_t)\boldsymbol{R}_x^{-1}(\overline{roll}_t)$' then maps $x^{(body)}$ from the inertial coordinates to the low frequency body coordinates $\{\bar{x}^{(body)}, \bar{y}^{(body)}, \bar{z}^{(body)}\}$.

The resulting $\boldsymbol{V}_{dp}$ gives the dynamic 3D heading of the animal wrt the animal's current neutral body pose (i.e. the low frequency body pose). Since the fluking motion of cetaceans is happening in the $\bar{x}^{(body)}$-$\bar{z}^{(body)}$ plane (i.e. **pitching plane**), the resulting $\boldsymbol{V}_{dp}$ would also be swinging in the $\bar{x}^{(body)}$-$\bar{z}^{(body)}$ plane, regardless of the actual orientation of the animal[4]. We further define **dynamic pitch** ($pitch_{dp}$) and **dynamic yaw** (yaw_{dp}) to be the angles between $\boldsymbol{V}_{dp}$ and the $\bar{x}^{(body)}$-$\bar{y}^{(body)}$ plane as well as the $\bar{x}^{(body)}$-$\bar{z}^{(body)}$ plane, respectively:

$$pitch_{dp(t)} = \arcsin(V_{dpz(t)}/norm(\boldsymbol{V}_{dp(t)})), \tag{2.8}$$

$$yaw_{dp(t)} = \arcsin(V_{dpy(t)}/norm(\boldsymbol{V}_{dp(t)})), \tag{2.9}$$

with $\boldsymbol{V}_{dp(t)} = [V_{dpx(t)}, V_{dpy(t)}, V_{dpz(t)}]$ at time t. The **dynamic pitch** ($pitch_{dp}$) and **dynamic yaw** (yaw_{dp}) approximate the animal's dynamic angle changes around its $\bar{y}^{(body)}$ and $\bar{z}^{(body)}$ axes, respectively. The $pitch_{dp}$ measure for a cetacean is what we desired earlier: 'an angle measurement wrt the animal's own body $y^{(body)}$ axis to describe its gait.'

An animal's per-stroke fluking period and amplitude are calculated by automatically locating and processing the positive and negative peaks in $pitch_{dp}$, which enables a wide variety of further analysis (e.g. investigating the animal's fluking frequency and amplitude when diving speed is greater than 2 m/s). Meanwhile, data segments that do not contain any peak are marked as passive behaviors (e.g. gliding).

2.7 Validation and Results

Biologging tag (MTag) data collected from bottlenose dolphins (*Tursiops truncatus*) under human care in Dolphin Quest Oahu Hawaii were used for quantitative validation of the methods. The biologging tag was aligned with the animal and attached 20 cm behind the blowhole in a non-invasive manner via four silicone suction cups (Figure 2.1-Left). As a correctly aligned tag measurement differs from a misaligned tag primarily by a rotation to the data, we injected artificial rotations to the originally aligned data to simulate the effects of tag shifts. For validation purposes, randomly designed rotations were applied to each randomly defined data segment to simulate the effects of a tag shift. More specifically, 18 datasets, with an average duration of 87 (±23) minutes, were included for the validation and 100 random simulations were conducted for each dataset. For each run, the dataset was randomly broken into $k + 1$ segments, with the random integer k ranging between 1 and 6, representing the number of injected tag shift instances. A random rotation was then applied to each data segment to simulate the effect of a tag shift. Figure 2.6-Left Column presents

[4]If $\boldsymbol{V}_{dp}$ is applied to a shark, for example, the resulting vector would be mainly swinging in the $\bar{x}^{(body)}$-$\bar{y}^{(body)}$ plane of the shark's low frequency body coordinates.

the orientation sphere of an example dolphin dataset with 4 simulated tag shifts.

The validation objectives were to assess: (1) how precise can the proposed method identify the injected shift instances; (2) how well can the proposed method correct for the applied random rotations; and (3) how accurate is the calculated animal pose compared to ground truth.

For tag shift detection we define **error** of a detection as the absolute time difference between the detected shift instance and the nearest injected shift instance. Further, the detection is a **positive detection** if the error is within 300 seconds (5 minutes), otherwise a **negative detection**. Then **precision** and **recall** are calculated as:

$$precision = \frac{\text{\# positive detections}}{\text{\# all detections}}, \tag{2.10}$$

$$recall = \frac{\text{\# positive detections}}{\text{\# injected shifts}}. \tag{2.11}$$

Among all 100 random runs over the 18 datasets, the average precision is 0.87, average recall is 0.89 while the average error is 37.5 (±18.4) seconds.

For tag orientation correction, to assess whether the random rotations have been corrected well, poses (roll, pitch and yaw) calculated from **corrected** simulated data segments (condition **A**) were compared with poses calculated from the original data segments (no random rotation injected, condition **B**), with results shown in Table 2.1.

Meanwhile, to evaluate the stated pose calculation method, the poses calculated from the original data segments using the stated method (involved accelerometer and magnetometer, condition **B**) were further compared with the poses calculated using a gradient descent based filtering approach [105] (referred to as the *Madgwick*'s approach after the author's name), which used accelerometer, magnetometer and **gyroscope** data (condition **C**, Table 2.1).

To assess the sensitivity of the proposed methods, we repeated the computation experiments by injecting simulated tag shifts, with the simulated tag rotations having random directions (same as before) but fixed degree for each run. Specifically, 50 random runs were made to each dataset for each fixed degree. For each run, k rotations were injected into the dataset (as before), and each rotation had a random direction with the fixed degree. The average precision and recall of the method in detecting the shifting instances were calculated for each fixed degree over all datasets and random runs, Figure 2.7-Left. The absolute angle differences between poses calculated from corrected simulated data segments (condition **A**) and poses calculated from the original data segments (condition **B**) were calculated for each fixed angle, Figure 2.7-Right.

DTag [39] data from wild humpback whales and belugas were also used to evaluate the method. Figure 2.6-Center Column presents the orientation sphere of an example dataset from a wild humpback whale (22.03 hours) with multiple tag shifts detected. The top row shows the orientation sphere after shift corrections, the middle row provides the raw data, while the bottom row shows

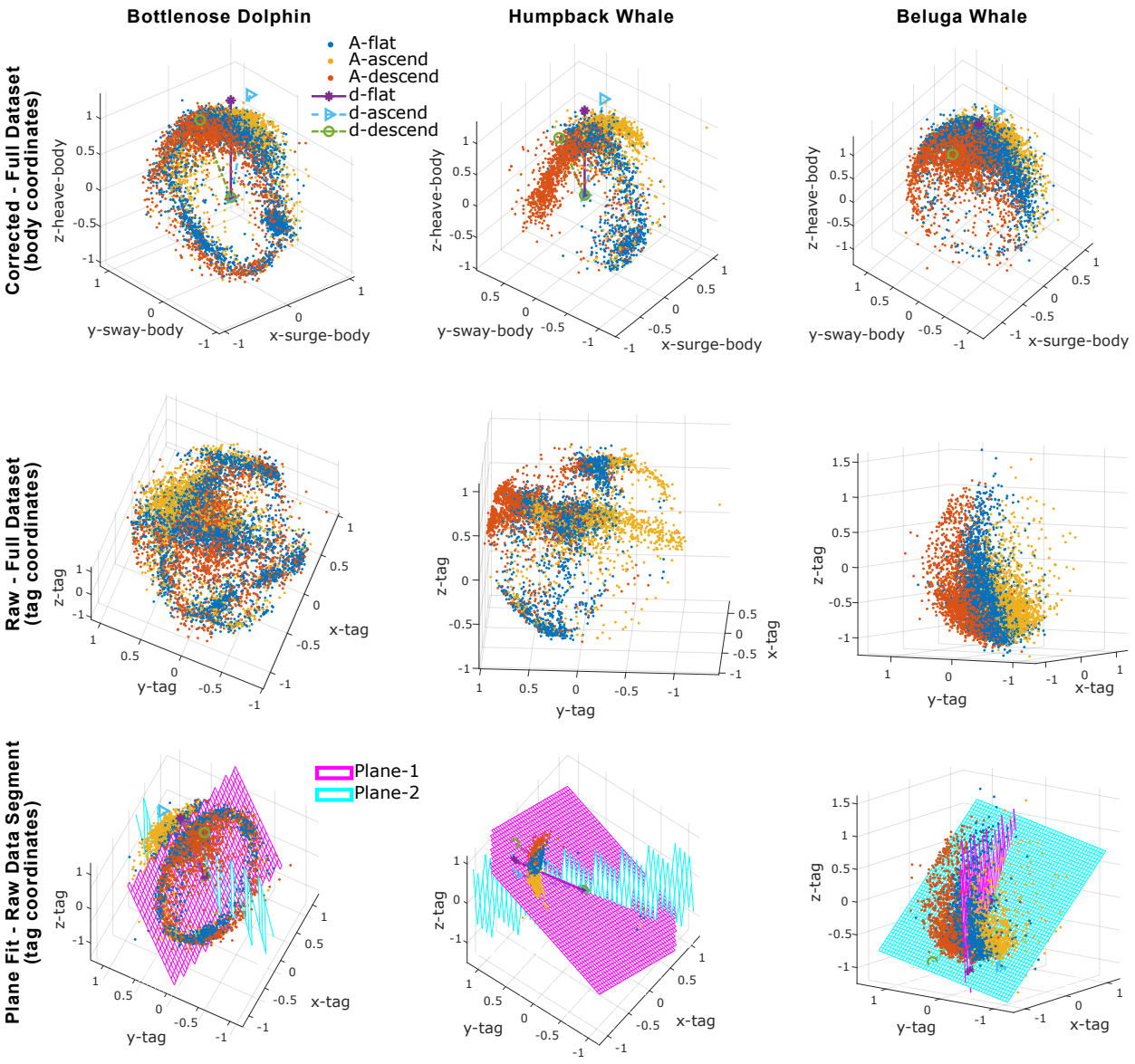


Figure 2.6: **Orientation spheres** for a bottlenose dolphin (left column), a humpback whale (center column) and a beluga whale (right column). The plot of accelerometer data clustered by depth speed is referred to as **orientation sphere**. The top row of subplots presents the orientation spheres in the animal's body coordinates, after detecting tag shifts and correcting tag orientation misalignment. The middle row presents the data in the tag coordinates before any correction. The bottom row visualizes the orientation correction method being applied to a uncorrected data segment. The dolphin dataset (left column) contains 4 simulated tag shifts. The humpback whale dataset (center column) has 5 detected tag shifts. While the beluga dataset (right column) has no detected tag shifts (i.e. no relative motion between tag and animal was detected).

Table 2.1: Resulting pose angle differences with different conditions and methods. The direct difference (e.g. $A - B$) is used to detect a bias in the difference, while the absolute difference (e.g. $|A - B|$) returns the magnitude of the difference. Pose angle differences are in degree, each cell gives mean ± standard deviation.

	A vs. B		**B vs. C**	
Pose	$A - B$	$\|A - B\|$	$B - C$	$\|B - C\|$
Roll	2.4 ± 10.5	6.6 ± 9.2	0.2 ± 3.0	1.3 ± 2.7
Pitch	6.8 ± 4.9	8.1 ± 3.1	-0.4 ± 2.3	1.3 ± 1.9
Yaw	0.8 ± 7.6	5.8 ± 5.8	-0.3 ± 5.7	2.1 ± 5.3

A: Pose calculated from simulated data after the simulated tag shifts have been corrected using the proposed method.
B: Pose calculated from the original data (without tag shift).
C: Pose calculated from the original data (without tag shift) using the Madgwick's filter [105], which involves the additional use of gyroscope.

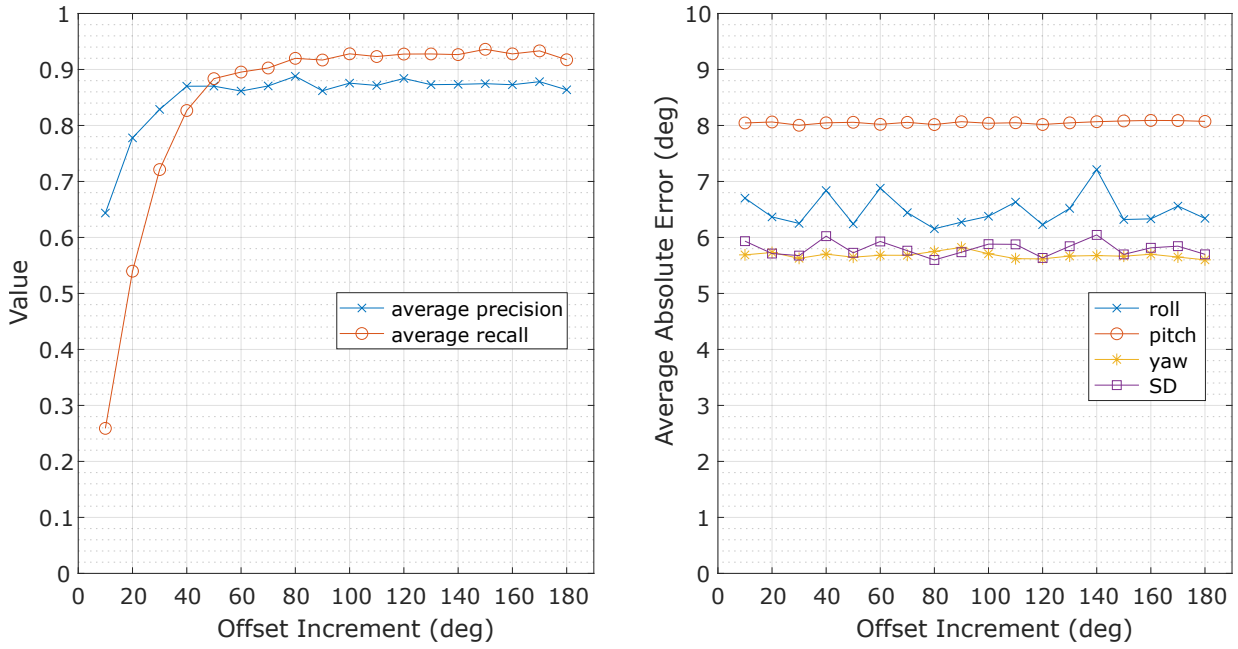


Figure 2.7: Average precision and recall of tag shift detection (**left**) and the average absolute error of the calculated animal pose after tag orientation correction (**right**) over simulated tag shifts with random direction and fixed degrees (x-axis). SD denotes the average of the standard deviation values for roll, pitch, and yaw at each offset increment.

the orientation method being applied to one of the data segments. Similarly, Figure 2.6-Right Column presents the orientation sphere from a wild beluga whale dataset (2.25 hours) where the tag was not aligned with the animal but contains no detected tag shift (i.e. no relative motion between tag and animal was detected).

Figure 2.8 presents example data segments from one of the bottlenose dolphins and the beluga whale datasets, where the pose of the animal was calculated from the corrected body coordinates' data, and the gait of the animal was characterized using the $pitch_{dp}$ (**dynamic pitch**) of the animal. The mean fluking period of this bottlenose dolphin is 1.0 s with an average amplitude of 14.1 deg. The mean fluking period of the beluga whale is 1.6 s with an average amplitude of 16.8 deg. For the humpback whale dataset, the mean fluking period is 6.7 s with amplitude of 12.2 deg.

2.8 Discussion

The simulated datasets from bottlenose dolphins in a managed setting provided an efficient validation environment for the proposed methods. A large validation dataset that covered a wide variety of tag shifts was obtained. And, the ground truth measurements for the validation dataset were immediately available. The automatic shift detection algorithm is the first of its kind, and had a detection precision of 0.87 and recall of 0.89 with average error of 37.5 seconds. The missed detections that were not 'recalled' by the method could be due to multiple injected shifts being temporally close to each other. As the method was designed to identify *occasional* tag shifts, the method may not detect all shifts if there are more than two shifts within two consecutive data segments. The length of a signal segment was determined empirically to be 10 to 15 minutes. If the length of a data segment is too short, there is not enough data to form a meaningful orientation sphere. While if the length of a data segment is too long, nearby shifts may not be separately detected.

Another reason for missed detections could be that the tag shift was too small to be detected. The method's sensitivity to shifting angles was investigated accordingly and presented in Figure 2.7. With a 10 deg offset, less than 30% of the shifts were detected. While with a 20 deg offset, around 54% of the shifts were detected with a detection precision of around 79%. When the offset reaches to 30~40 deg, the method's performance starts being more reliable with a recall in the 70~80 percentile and precision being ~85%. When the offset reaches to 50 deg and above, the method's performance converges to ~90% precision and recall. False positives (i.e. detections that did not correspond to a tag shift) were generated when the animals switched from one gait to another (very different) gait. For example, the method could be tricked when the animal switched from a gait without any roll to a sideways swimming gait (i.e. swimming with ~90 deg roll). In further developments, the method's performance could be further improved by taking into account

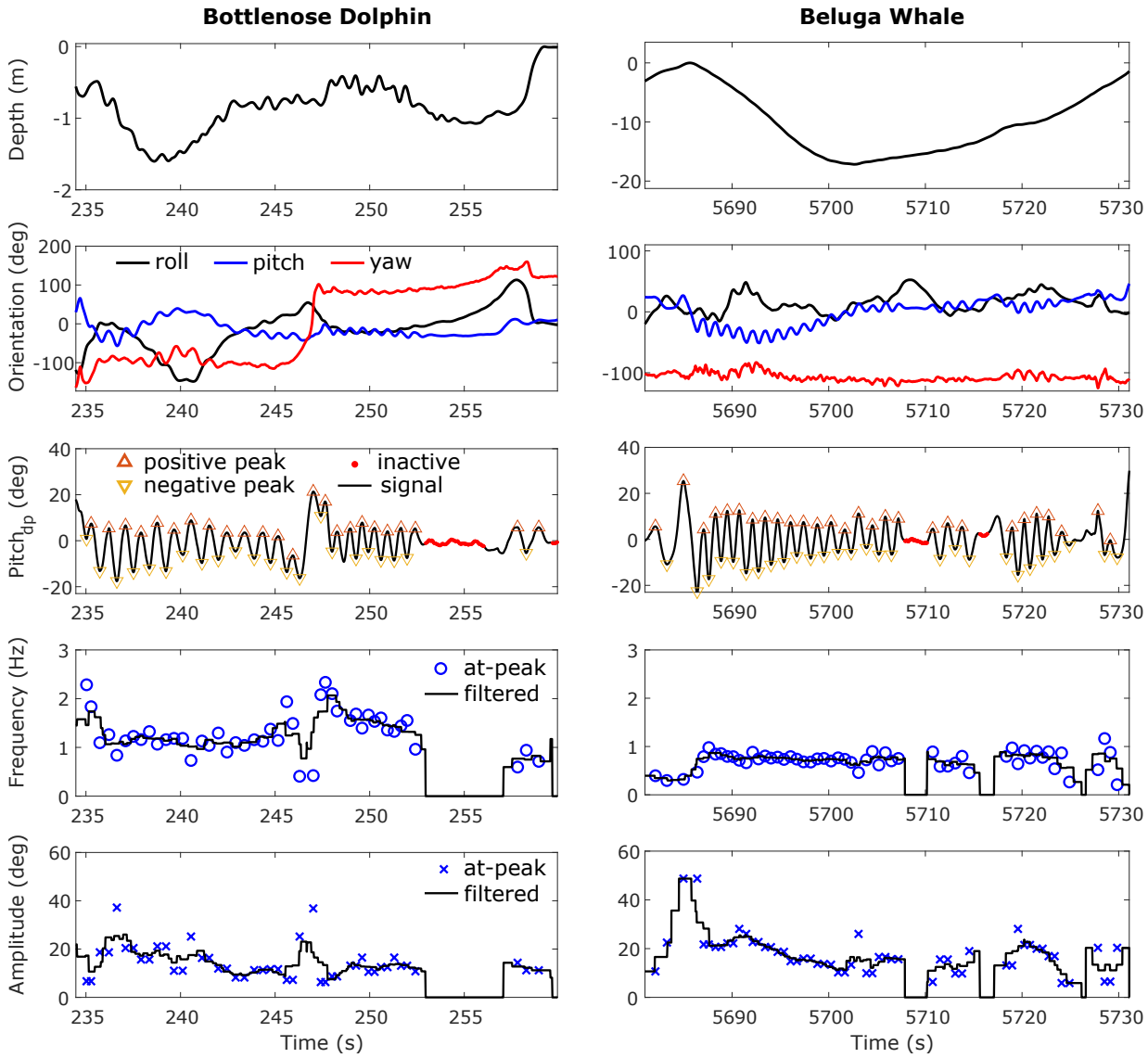


Figure 2.8: Example data segments from a bottlenose dolphin (**left**) and a beluga whale (**right**) with animal's gait parameterized via **dynamic pitch** ($pitch_{dp}$). When the animal has a big roll angle (e.g. during the time around 240 seconds in the bottlenose dolphin dataset), the fluking 'signature' (i.e. the sinusoidal fluctuations in each signal channel) transfers from pitch to yaw in the orientation estimations. $Pitch_{dp}$ is used to have a pose invariant descriptor of the gait of the animal, that is with respect to the animal itself rather than the environment. Inactive swimming periods (e.g. gliding) are automatically identified in $pitch_{dp}$ while fluking frequency and amplitude are calculated from $pitch_{dp}$, which can be used for further gait analysis.

events associated with high acceleration impacts to the tag, as tag shifts were oftentimes caused by physical impact to the tag [39], such as an impact from a conspecific animal in the group.

To decouple the evaluations of tag shift detection and tag orientation correction, the tag orientation correction method was applied to each of the data segments that had been transformed by a rotation. Poses (roll, pitch and yaw) of the animal were calculated after orientation correction and then compared with poses calculated using the original (un-rotated) data. The average differences between the two were within 10 deg in all cases (Table 2.1-**A** vs. **B** and Figure 2.7-Right). This indicates that the tag orientation correction method is able to work reliably no matter where the tag was located as long as the given data segment did not contain a tag shift. Meanwhile, the poses calculated using the stated method (involved accelerometer and magnetometer) in this work were compared with the established Madgwick's filtering approach (involved accelerometer, magnetometer and gyroscope [105], Table 2.1-**B** vs. **C**) and resulted in average differences that were within 5 deg. The main error for the proposed pose estimation method came from the estimation of gravity direction using the accelerometer measurements.

In this work, the low-pass filtered 3-axis accelerometer data was used as the measurement associated with gravity ($\boldsymbol{A}^{(tag)}$) for the pose calculation. However, this is an inaccurate assumption in practice, since the accelerometer measurements result from both the animal's linear acceleration and gravity. The animal's motion affects accelerometer measurements in 3 aspects:

(a1) Measurements associated with animal's linear acceleration. Since normally the animal would not maintain a constant acceleration for more than a few seconds, all linear accelerations are considered high frequency.

(a2) Measurements associated with gravity, driven by **high** frequency body orientation changes.

(a3) Measurements associated with gravity, driven by **low** frequency body orientation changes.

Because measurements associated with gravity (a2 & a3) are needed for estimating animal pose, a1 needs to be decoupled from a2 and a3. An accurate decoupling would not be possible using an accelerometer alone. But because these animals' linear accelerations are generally much smaller than gravity, a low-pass filter can be applied to attenuate a1 as well as a2 to approximately decouple linear acceleration measurement (a1) from gravity (a2 & a3).

To better decouple linear acceleration (a1) from gravity (a2 & a3), [109] proposed a *magnetometer* approach that takes advantage of the fact that a magnetometer is only sensitive to orientation changes, not acceleration. The method assumes that high frequency body orientation changes are due to the 'pitching' motion of the animal, and relies on this assumption to use the magnetometer to assist finding a2, with a3 obtained via low-pass filtering, and a1 the remaining acceleration after subtracting a2 and a3. The method works well in decoupling a1 from a2 and a3 for most cases. However, we note that because the method assumes high frequency body orientation changes are happening in the pitching plane, the method tends to be unstable when the animal does quick back-

and-forth rolling or turning. An implementation of the method is included in the toolbox proposed in this work and we recommend using it for animals with less frequent back-and-forth rolling or turning. For tags that have a gyroscope, together with an accelerometer and magnetometer, a gradient descent based filtering approach [105] (*Madgwick*'s approach) can be used to better estimate the tag's orientation wrt the world. The method uses the gyroscope to provide a direct measure for a2, thus reducing uncertainty in the measurement. A wrapper of the Madgwick's approach is also included in the toolbox and we recommend using it when gyroscope measurements are available.

Even though the relative orientation of the tag was corrected to be aligned with the animal, the actual location of the tag (e.g. back vs. peduncle) could still affect tag measurements and results, particularly the estimated fluking amplitude of the animal. For example, a tag located closer to the fluke will have higher estimated fluking amplitude than a tag located near the dorsal fin for the same gait. So, fluking amplitude would be a good indicator of animal gait for fine-scale studies (e.g. comparing an animal's fluking amplitude between descent vs. ascent), but not for global comparisons (e.g. comparing one animal to another, with different tag locations). On the other hand, the fluking frequency and period are not affected by the location of the tag, and can be used for both fine-scale studies and global investigations.

The orientation spheres (Figure 2.4) can also serve as a visualization tool for animal behavior studies. For example, a clear sign of lateralized behavior [110] appears in the orientation sphere of the humpback whale (Figure 2.4-Center Column), with an unbalanced right 'wing' corresponding to a dominating left roll of the animal during foraging at the ocean bottom. Also, this humpback whale did not roll much during ascending/descending while the bottlenose dolphin (Figure 2.4-Left Column) and beluga whale (Figure 2.4-Right Column) rolled more frequently during ascending/descending. Even though the beluga whale rolled frequently, it hardly ever rolled completely upside down, which can be observed from the empty bottom of the orientation sphere. Unlike the beluga, the dolphin data shown in Figure 2.4 indicate that the animal rolled all the way around during the studied period.

2.9 Summary

This chapter systematically presents an automated data processing framework (and software) that takes advantage of the common characteristics of pose and gait of the animal to process tag data and analyze the animals' gait. Three aspects were investigated in particular: (1) The identification of time instances associated with occurrences of relative motion between tag and animal. (2) The identification of the relative orientation of a tag wrt the animal's body for a given data segment. (3) The extraction of gait parameters that are invariant to pose and tag orientation. Biologging tag data from bottlenose dolphins, a humpback whale, and a beluga whale were used to validate and

demonstrate the approach. Results show that the average relative orientation error of the tag wrt the dolphin's body after processing were within 10 degrees in roll, pitch and yaw directions. The average precision and recall for identifying relative tag motion were 0.87 and 0.89, respectively. Examples of the resulting pose and gait analysis demonstrate the potential of this approach in not only gait analysis but also animal behavioral studies. The proposed analysis approach and software will facilitate the use of biologging tags for the study of cetacean locomotion and behavior. The method and software can be directly used with cetacean data from any tag platform that uses an accelerometer, magnetometer, and pressure sensor.

CHAPTER 3

Estimation with Biologging Tags: Sequential Motion State Classification

3.1 Background

Biologging tags are a key tool for the study of animal dynamics and behavior, enabling the investigation of fundamental questions in biology and ecology [113, 114]. Many of the advancements in bio-inspired robotics [115] and multi-agent control [116, 117] are built on an understanding of animal movements, decision making strategies, and interactions. For many animals it is possible to use direct observation to study behavior in the wild [33]. However, continuous observation of an animal for days or weeks is difficult even for a team of human observers, and behavioral observation does not directly capture dynamical movement information. Further, for some animals, especially in the marine environment, direct observation is often not viable. For these cases, biologging tags are the only way to monitor and study these animals in their natural environment [39]. The ability to measure dynamical (e.g. acceleration, angular velocity) and environmental (e.g. magnetic field, temperature, pressure) information makes these tags ideal tools to study the dynamics and behaviors of animals. With the rapid growth of Micro-Electro-Mechanical Systems (MEMS) technology, cost effective and low power IMUs (Inertial Measurement Unit) have become available and are increasingly used in biologging tags. Sensors can be placed on animals

Content of this chapter is mainly associated with the following publications:

[111] D. Zhang*, K. A. Shorter, J. Rocho-Levine, J. van der Hoop, M. Moore, and K. Barton, "Behavior Inference From Bio-Logging Sensors: A Systematic Approach for Feature Generation, Selection and State Classification," in *Proceedings of the ASME 2018 Dynamic Systems and Control Conference (DSCC)*. September 30 – October 3, 2018, Atlanta, Georgia, USA: ASME, 2018, pp. 1–10.

[112] R. Sibal, D. Zhang*, J. Rocho-Levine, K. A. Shorter, and K. Barton, "Bidirectional LSTM Recurrent Neural Network Plus Hidden Markov Model for Wearable Sensor-Based Dynamic State Estimation," *ASME Letters in Dynamic Systems and Control*, vol. 1, no. 2, pp. (021 002) 1–5, 2021.

* Corresponding Author

for days or even weeks at a time, creating large sets of data about the animals. These datasets are often complemented with direct observation of the tagged animals by a trained observer to identify specific behaviors that can then be described and quantified in detail using the tag data. Having well-labeled high-quality data provides the foundational information for automatic motion state classification.

In this work, biologging is synonymous with *wearable*, a term commonly used to refer to devices that measure human activities. Despite the similarities in sensor suites, data analysis tools for biologging devices have tended to lag behind those used in the field of wearable sensing. Animal-focused studies [118, 119] generally rely more on manual data processing and expert knowledge to interpret information. Human studies [120, 121], on the other hand, more often leverage machine learning techniques to classify behaviors.

In the biologging field, many machine learning classification schemes are 'single frame', where each data point is considered to be independent of all neighboring data points and classified accordingly. For example, a multi-class Support Vector Machine (SVM) has been used with three-dimensional accelerometer data [122] to classify behavioral patterns of dairy cows. In [123], multiple classifiers were applied with GPS and accelerometer data to classify behavioral modes of griffon vultures. The classifiers are all single frame and include: artificial neural networks (ANN), classification and regression trees (CART), linear discriminant analysis (LDA), random forest (RF) and SVM. Aggregated continuous accelerometer and GPS data were also used to classify the behavior of six cheetahs in Botswana at an 80-90 percent accuracy rate [124]. In [125], a captive surrogate (domestic dog) is used to train a SVM that is then used to classify behaviors of similar species that are difficult to observe. Contrary to 'single frame' classifiers, a 'sequential frame' classifier considers the probability of a sequence of data points (states), while classifying a single data point. HMM, in particular, uses a set of hidden states (stochastic processes) to derive an understanding of the temporal information between sequences [126]. This type of technique have been widely used in applications such as speech recognition [127], but has only recently been considered for biologging applications [128].

Other important considerations when implementing machine learning algorithms for behavioral classification are the selection of the feature sets and window sizes used to generate the feature sets. As will be formally introduced in the following sections, **feature** is one of the key components in the classification tasks. However, current applications in biologging tend to rely on empirical results or domain expertise to determine these algorithm settings [122, 123, 125], which may not extend to the study of new species or behavioral sets with performance guarantees.

Deep neural networks are gaining popularity due to their ability to learn representations of data with multiple levels of abstraction [61]. In [129], researchers used deep learning to detect irregularities in the behaviors of elderly individuals suffering from dementia to prevent mishaps

from occurring. Though neural networks are becoming state-of-the-art in a wide variety of fields [130], their use to monitor animals is still fairly limited [130]. However, the use of machine learning methods can provide valuable insights not only into fundamental biological theories, but also animal locomotion, dynamics, and decision making [113]. The Artificial Neural Networks (ANN), on the other hand, have been shown to reduce the amount of expert intervention as the input data does not have to be heavily pre-processed [131]. RNN in particular has been found to work well for applications using time series data, e.g. activity recognition through "sample by sample prediction of physical activities" [132]. Within RNN's, the LSTM RNN is able to exploit both short-term and long-term connections within data, which outperforms traditional RNN that only maintains short-term memories [133]. While there are several studies that use LSTM to classify human behavior during daily activities [132, 134], there are very few examples where LSTM has been used for animal behavior classification. The strong ability of LSTM model to process time series data is very beneficial for animal behavior classification. However, in practice, long duration data is often broken into many sequences to be processed individually, and information between sequences is commonly ignored.

Building on the studies in the literature, this work first presents a systematic framework for feature generation, selection and sequential state classification for use with biologging devices, with an HMM structure used at the top level to link and improve results from single frame classifications at the bottom level. Then, this chapter demonstrates how an LSTM framework can be used at the bottom level to replace the single frame feature generation, selection and state classification, and forms an LSTM-HMM architecture that exploits temporal information within both the bottom and the top level. The overall objective of this chapter is to develop a systematic framework for sequential motion state classification, that has an automated 'black box' at the bottom level and an interpretable 'white box' at the top level. So the methods would not only have the advantages of being more accurate (due to the extra multi-frame temporal information) and more adaptive to new tasks (due to the 'white box' at the top level), but also being user friendly by shading away the requirement for user to have expertise in machine learning to deal with feature engineering.

3.2 Problem Formulation and Methods Outline

The objective of this work is to develop a framework for classifying animal motion states using biologging sensor data. This section presents the high-level mathematical description of the problems addressed in this work. Figure 3.1 illustrates a standard classification approach using an 11-signal-channel dataset as an example. The complete signal sequence is segmented into T segments ($\{\boldsymbol{S}_1, \cdots, \boldsymbol{S}_T\}$) using windows each of size W (e.g. $w = 0.6$ second) with overlaps of size r. Each segment (e.g. $\boldsymbol{S}_t$ for the t-th segment) is considered as a **state**, with the entire time sequence

separated into T states. A segment of raw signal sequence $\boldsymbol{S}_t = [\boldsymbol{q}_1^{(t)}; \cdots; \boldsymbol{q}_l^{(t)}; \cdots; \boldsymbol{q}_L^{(t)}]$ needs to be **feature engineered** into a M-by-1 feature vector $\boldsymbol{F}_t = [F_{1t}, \cdots, F_{mt}, \cdots, F_{Mt}]^T$ before the classifier can make any prediction. Here $\boldsymbol{q}_l^{(t)} = [q_{l1}^{(t)}, \cdots, q_{lz}^{(t)}, \cdots, q_{lZ}^{(t)}]$ denotes the l-th data point in $\boldsymbol{S}_t$, and $q_{lz}^{(t)}$ is the z-th dimension (i.e. the z-th signal channel) of this data point.

M descriptive signal features, also known as summary statistics, are extracted from each state (e.g. $[F_{1t}, F_{2t}, \cdots, F_{Mt}]^T$ for t-th state). Some commonly used features are, for example, the mean value of x-axis acceleration, or the entropy of y-axis gyroscope reading, or the covariance of y and z-axis angular speed, etc. Mathematically, if the m-th feature is the mean value of the k-th signal channel, then it is computed as $F_{mt} = mean([q_{1z}^{(t)}, \cdots, q_{lz}^{(t)}, \cdots, q_{Lz}^{(t)}])$ for the t-th state. Further details of the features will be provided in the next section.

The features are considered as the (sensor) observations of the states. And each state needs to be classified as a specific motion class, which is denoted by the random variable X. Letting there be n motion classes total, then for the t-th state, its motion state X_t takes value from the set $\{c_1, c_2, \cdots, c_n\}$. An example set of motion classes could be $\{c_1 =$ 'turn left', $c_2 =$ 'ascend', $\cdots, c_n =$ 'leaping'$\}$. The state X_t can then be classified by feeding $\boldsymbol{F}_t$ into a classifier (e.g. SVM, K-Nearest Neighbors (KNN), Decision Tree (DT)).

For notation purposes, upper case X_t denotes random variable, while lower case x_t is the instantiation of the random variable. For example, $P(X_t = x_t \mid x_t = c_i, \boldsymbol{F}_t)$, or simply $P(X_t = c_i \mid \boldsymbol{F}_t)$, denotes the probability of having state X_t being of class c_i given the feature vector $\boldsymbol{F}_t$. To address the classification process in a systematic manner, we now define the problems as follow:

1. Does the selection of window size W affect the classification result? How should W be selected?

2. What is the optimal feature set, $\{F_{1t},\ F_{2t},\ \cdots,\ F_{Mt}\}$, to use for a given set of motion classes $\{c_1, \cdots, c_n\}$ in a specific classification task?

3. Given the contiguous observations of states, $\{\boldsymbol{F}_1,\ \boldsymbol{F}_2,\ \cdots,\ \boldsymbol{F}_T\}$, how to estimate the corresponding motion classes $\{x_1,\ x_2,\ \cdots,\ x_n\}$?

4. Are features $(\{\boldsymbol{F}_1, \cdots, \boldsymbol{F}_T\})$ necessary in representing the signal segments $(\{\boldsymbol{S}_1, \cdots, \boldsymbol{S}_T\})$?

A block diagram of the proposed framework is presented in Figure 3.2. As will be presented in Section 3.3, the selection of the window size is handled by the outer loop, while the optimal set of features is determined via the inner loop, both are conducted using an efficient sampling strategy. The system further boosts classification performance by utilizing the sequential temporal information present in the data via an HMM structure (Section 3.4). Then in Section 3.5, it is shown how feature engineering can be avoided by employing an LSTM RNN at the bottom level.

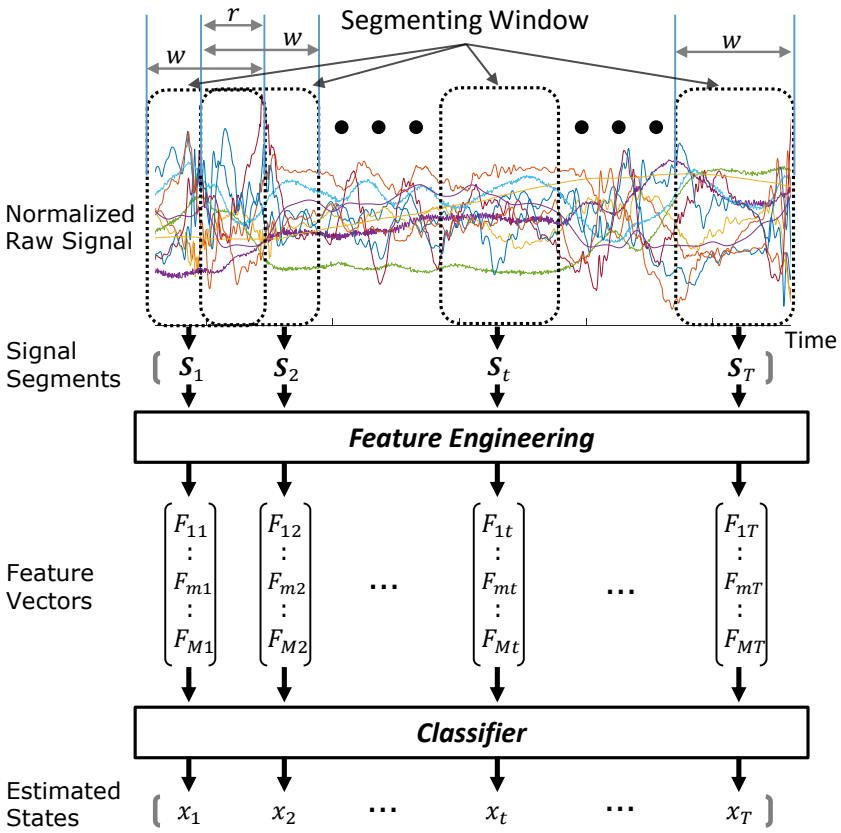


Figure 3.1: Architecture of 'single frame' classification. The example signal consists of 11 channels. Window size is denoted as W, overlap between windows is r. The signal is segmented into T windows with each window considered as a discrete state. For each state (window), M features are extracted and taken as observations of each state (e.g. $\boldsymbol{F}_t = \{F_{1t}, F_{2t}, \cdots, F_{Mt}\}$ for the t-th state). The motion class of each state is denoted by random variable X, with lower case x denotes the instance (i.e. actual value), e.g. $X_t = x_t$ means state X_t takes value x_t, where x_t belongs to one of the n motion classes ($x_t \in \{c_1, \cdots, c_n\}$).

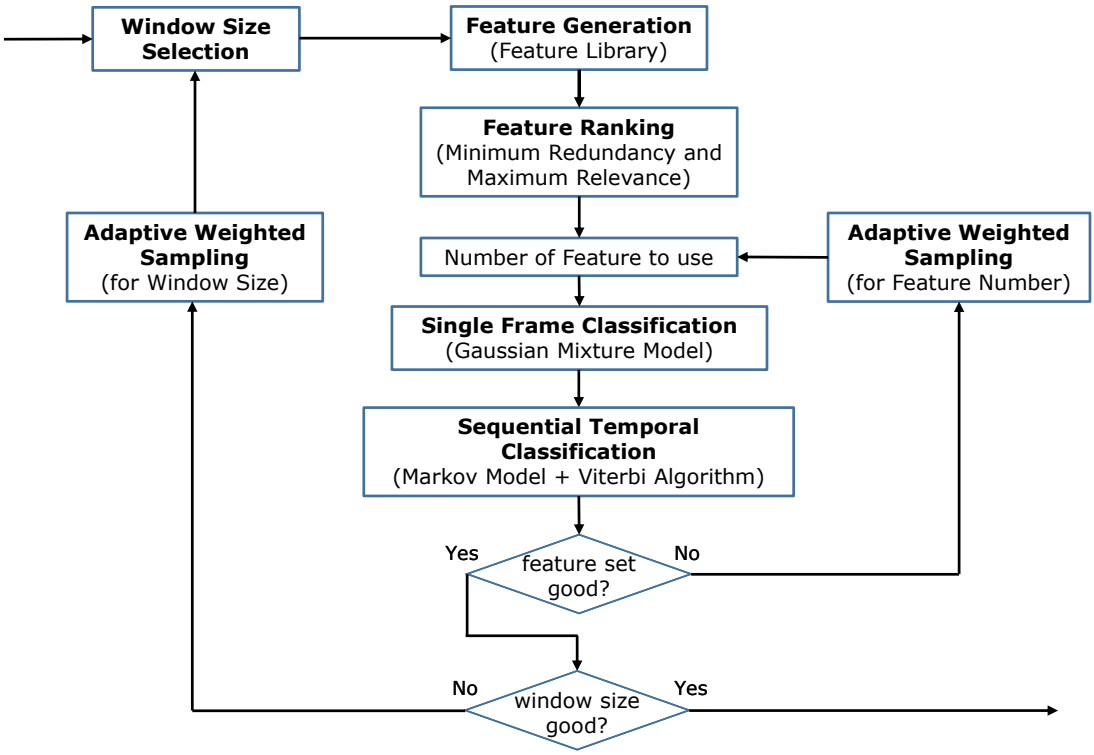


Figure 3.2: Block diagram of the proposed framework.

A case study on bottlenose dolphin motion state classification is presented in Section 3.6, followed by the results and discussions in Section 3.7.

3.3 Feature Engineering

3.3.1 Window Size Selection

The selection of window size W during signal segmentation has been shown to influence animal behavior classification performance [135, 136]. However, there is no prevalent method available in the field of animal behavior classification to select the window size. Meanwhile, the activity-based and event-based windowing method [137, 138, 139] in the field of human activity classification/recognition are more domain specific and hard to be generalized toward different behavior sets and different animal species. A conceptually simple yet powerful window searching method is used here instead.

If there is an optimal window size w^* for generating features to classify a certain set of behaviors of a certain animal species, what needs to be done is to 'simply' search through all the possible

window size selections and find the optimal one. A fine-scale grid can be defined to replace the continuous search space. Combined with an efficient searching method, as will be introduced in *3.3.3 Adaptive Weighted Sampling Scheme*, the near optimal window size can be found.

3.3.2 Feature Library and Feature Selection

Features are the fundamental elements used for all state classification and identification tasks that are to be performed in this work. Insufficient feature sets cannot provide enough information to adequately train the system or classify unlabeled observations correctly. Large feature sets are computationally expensive to work with and can make it hard to understand the underlying process that generated the data [140]. Besides, as the number of feature increases, the volume of the data space grows dramatically, which makes available data sparse and not as sufficient for training purpose [141]. Thus, extra attention is required to select the appropriate feature set for each specific task (e.g. certain animal species and/or certain behavior set). Instead of fully relying on expert knowledge to decide what features to use for a particular application, a combination of a feature library and a feature selection scheme is proposed to handle the selection of the appropriate feature sets. In this approach, a feature library includes as many features as possible in the beginning stage of the process, and then feature selection is applied to identify the optimal subset for a specific task.

Features typically vary from task to task and a library is used to store all available features. Features are extracted from the signal channel for each window (state) as illustrated in Figure 3.1. The feature library in this work includes the following feature types: mean, Standard Deviation (SD), mean of absolute value, SD of absolute value, SD of absolute value of diff value (where diff value is the approximate derivative of the signal), root mean square (rms), maximum value (max), minimum value (min), max minus min (range), skewness, kurtosis, entropy, mean frequency, band power, waveform-length (wfl) (Eq. 3.1), max fractal length (mfl) (Eq. 3.2), and the correlation coefficients over different signal channels. For the k-th signal channel of a data segment $\boldsymbol{S} = [\boldsymbol{q}_1; \cdots; \boldsymbol{q}_l; \cdots; \boldsymbol{q}_L]$, with $\boldsymbol{q}_l = [q_{l1}, \cdots, q_{lz}, \cdots, q_{lZ}]$ denotes the l-th data point in $\boldsymbol{S}$, and q_{lz} being the z-th signal channel of this data point, wfl and mfl are calculated as:

$$F_{(wfl)} = \sum_{l=1}^{L-1} |q_{(l+1)z} - q_{lz}|, \tag{3.1}$$

$$F_{(mfl)} = \log\left(\sum_{l=1}^{L-1} (q_{(l+1)z} - q_{lz})^2\right). \tag{3.2}$$

A mutual information based greedy feature selection method, *Minimal Redundancy Maximal*

Relevance (mRMR) [142], is used to rank all features in the feature library. The feature selection method, as its name suggests, adds features to the rank list one-by-one such that the newly added feature has minimal redundancy with respect to the features already in the list and maximum relevancy to the class label.

After all features are ranked based on their 'usefulness', the proposed framework determines how many features to include in the final set based on overall classification performance.

3.3.3 Adaptive Weighted Sampling Scheme

Instead of an exhaustive search, an adaptive weighted sampling scheme is formulated to ease the computational burden in determining the feature number and window size selections. The objectives for these two tasks are very similar and given as follows. For optimal feature number selection, it is desired that high classification performance is achieved with the minimum number of features. For window size selection, the objective is to use a relatively small window size to generate better classification results so that behaviors of animals are captured with higher resolution. In the adaptive weighted sampling scheme, regions of high interest are sampled with higher probability, while the low interest regions are sampled with lower probability. During the sampling process, the estimation of 'interest' over the sampling domain is continuously updated as new samples are collected. This iterative process alternates between the **update** step and the **sample** step.

The interest or reward of a position is defined as $J(e)$, where e is a discrete variable denoting either the number of features or the window size, depending on its application. The classification performance (e.g. accuracy) associated with e is denoted as $H(e)$, which is unknown and to be sampled. $J(e)$ is estimated over all e based on the sampled $H(e)$ using linear interpolation in the **update** step. The estimation formula for $J(e)$ in the **update** step depends on 5 circumstances:

1. $H(e)$ exists in the database, i.e. e has been sampled already. Then $J(e)$ is calculated by Eq. 3.3, with a small positive coefficient λ posing cost (negative reward) on an increasing e.

$$J(e) = H(e) - \lambda e \tag{3.3}$$

2. $H(e)$ does not exist in the database, but e *tightly* lies between two positions e_α and e_β that have been sampled before. That is: (1) $\exists e_\alpha, e_\beta$ s.t. $e_\alpha < e < e_\beta$ and $H(e_\alpha)$ and $H(t_\beta)$ exist in the database (i.e. have been sampled); and (2) $\nexists e'$ s.t. $e_\alpha < e' < e_\beta$ and $H(e')$ exists in the database. The second condition makes sure that there does not exist a sampled e' that is closer to e than either e_α or e_β, which ensures that e_α and e_β are the closest (tightest) sampled positions on either side of e. In this case, $J(e)$ is calculated by linearly interpolating between $J(e_\alpha)$ and $J(e_\beta)$ (Eq. 3.4).

$$J(e) = \frac{e - e_\alpha}{e_\beta - e_\alpha}\big(J(e_\beta) - J(e_\alpha)\big) + J(e_\alpha) \tag{3.4}$$

3. $H(e)$ does not exist in the database, yet e is tightly on the left of a previously sampled position t_α with no lower bound. That is: $\exists e_\alpha$ s.t. $e < e_\alpha$ and $H(e_\alpha)$ exists in the database; and (2) $\nexists e'$ s.t. $e' < e_\alpha$ and $H(e')$ exists in the database. In this case, $J(e)$ takes the value of $J(e_\alpha)$, i.e. $J(e) = J(e_\alpha)$.

4. Similar to case 3, $H(e)$ does not exist in the database, but e is tightly on the right of previously sampled position e_β with no upper bound. That is: (1) $\exists e_\beta$ s.t. $e_\beta < e$ and $H(e_\beta)$ exists in database; and (2) $\nexists e'$ s.t. $e_\beta < e'$ and $H(e')$ exists in the database. Then $J(e)$ takes the value of $J(e_\beta)$, i.e. $J(e) = J(e_\beta)$.

5. Lastly, no e has been sampled, which is: $\nexists t$ s.t. $H(e)$ exists in the database. In this case, $J(e)$ assumes a constant value C, i.e. $J(e) = C$. Which indicates that all of the positions are sampled uniformly at random.

In **sample** step, a new sample is taken at e_{new} based on the probability weighting, which depends on $J(e)$ for all e. The adaptive weighted sampling scheme then switches between an **update** (of $J(e)$) step and a **sample** (of $H(e)$) step until the global density of sampled points reaches a desired level. Note that this process automatically provides higher resolution at the high 'interest' region. Meanwhile, the probability weighting can be normalized differently to make the sampling scheme more or less aggressive.

3.4 Sequential Frame Classification with HMM

In this work, animal motion states are modeled as a discrete time Markov chain to utilize temporal information in the process. Since the motion classes of the states are not observable (directly from sensor data), this Markov model is, more specifically, an HMM. Figure 3.3 illustrates an example HMM with only three motion classes. As defined in the problem formulation (Section 3.2) and Figure 3.1, for the t-th state, motion state X_t is assigned a value from the set of motion classes $\{c_1, \cdots, c_n\}$, while the corresponding state observation is denoted as $\boldsymbol{F}_t = [F_{1t}, F_{2t}, \cdots, F_{Mt}]^T \in \mathbb{R}^M$. A state transition probability matrix $\boldsymbol{A} = \{a_{ij}\} \in \mathbb{R}^{n \times n}$ characterizes the probability of the state transitioning from one class to another, or to itself (Eq. 3.5 and Eq. 3.6).

$$a_{ij} = P(X_{t+1} = c_j \mid X_t = c_i) \tag{3.5}$$

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}, \; a_{ij} \geq 0, \; \sum_{j=1}^{n} a_{ij} = 1 \tag{3.6}$$

The initial distribution (often called a *prior* probability distribution) of motion classes, $\boldsymbol{\pi} = [\pi_1, \cdots, \pi_i, \cdots, \pi_n]$, characterizes the probabilities of a state sequence to start at each particular class. E.g. π_i denotes the probability of the initial motion state being c_i (Eq. 3.7).

$$\pi_i = P(X_1 = c_i) \tag{3.7}$$

The Probability Density Function (PDF) of state observation $\boldsymbol{F}$ emitted from motion class c_i (Eq. 3.8) is represented by $\boldsymbol{b}_i$. $\boldsymbol{b}_i$ is characterized by a multivariate Gaussian Mixture Model (GMM) which is a weighted sum of K mixtures with each mixture representing a multivariate Gaussian distribution (Eq. 3.9). In Eq. 3.9, ω_{ik} is the weight of the k-th mixture, while $\boldsymbol{\mu}_{ik}$ and $\boldsymbol{\Sigma}_{ik}$ are the mean vector and covariance matrix of the k-th mixture.

$$\boldsymbol{b}_i = \text{PDF}(\boldsymbol{F} \mid X = c_i) \tag{3.8}$$

$$\boldsymbol{b}_i = \sum_{k=1}^{K} \omega_{ik} \, \mathcal{N}(\boldsymbol{\mu}_{ik}, \boldsymbol{\Sigma}_{ik}), \; \omega_{ik} \geq 0, \; \sum_{k=1}^{K} \omega_{ik} = 1 \tag{3.9}$$

3.4.1 Model Training and Data Fitting/Testing

During the model training stage, full system information, both state observations and motion classes of the states from training data, is required. Given $\{\boldsymbol{F}_1, \cdots, \boldsymbol{F}_T\}$ and $\{x_1, \cdots, x_T\}$, the state transition matrix $\boldsymbol{A}$ is obtained by determining the frequency of all class transitions in $\{x_1, \cdots, x_T\}$. Note that in practice, a small perturbation is added to the transition matrix to capture the probability of transitioning between any two motion classes, unless the transitions are infeasible in the real world. The initial state distribution $\boldsymbol{\pi}$ is estimated by computing the frequency of the appearance of each class in $\{x_1, \cdots, x_T\}$.

To obtain a specific $\boldsymbol{b}_i$ using GMM, first identify the set of all individual observations $\boldsymbol{F}_t$ that are emitted from the motion class c_i, i.e. $X_t = c_i$. Then, the model parameters, $\{\omega_{i1}, \cdots, c_{iK}\}$, $\{\boldsymbol{\mu}_{i1}, \cdots, \boldsymbol{\mu}_{iK}\}$ and $\{\boldsymbol{\Sigma}_{i1}, \cdots, \boldsymbol{\Sigma}_{iK}\}$ are estimated from the training data via the Expectation Maxi-

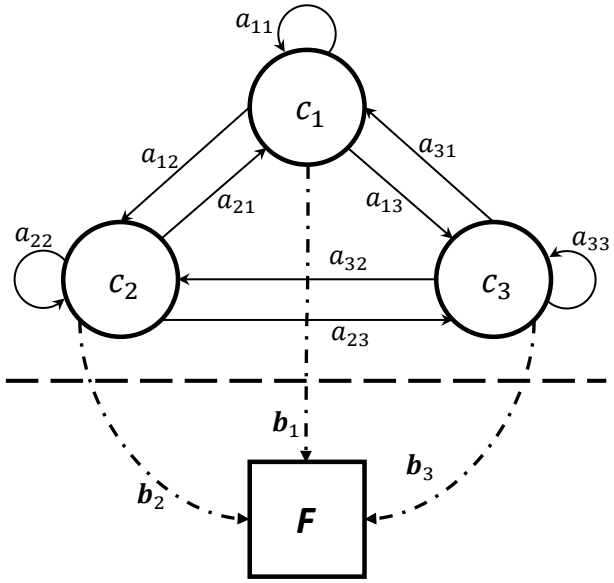


Figure 3.3: Example Hidden Markov Model with $n = 3$ motion classes $\{c_1,\ c_2,\ c_3\}$. The probability of state transitions from class c_i to class c_j is denoted by a_{ij}. Multivariate random variable $\boldsymbol{F}$ represents the observation emitted/obtained from each class. $\boldsymbol{b}_i$ models the probability distribution of $\boldsymbol{F}$ that emitted from a specific class c_i.

mization (EM) algorithm [144]. Model training is completed after $\{\boldsymbol{b}_1,\ \cdots,\ \boldsymbol{b}_n\}$ are all obtained.

When applying the train framework to classification tasks, as demonstrated in Figure 3.4, the aim is to find the sequence of estimated motion classes $\{x_1,\ \cdots,\ x_T\}$ for a sequence of observations $\{\boldsymbol{F}_1,\ \cdots,\ \boldsymbol{F}_T\}$, such that $P\big(X_1 = x_1,\ \cdots,\ X_T = x_T \ \big|\ \boldsymbol{F}_1,\ \cdots,\ \boldsymbol{F}_T\big)$ is maximized.

The first step approaching this is to estimate the *confidence* that a particular observation $\boldsymbol{F}_t$ has emitted from a particular class c_i via the function $b_i(\cdot)$ (Eq. 3.10). $b_i(\boldsymbol{F}_t)$ gives an estimated (Mahalanobis based) distance from the observation $\boldsymbol{F}_i$ to the distribution $\boldsymbol{b}_i$ (Eq. 3.11, Eq. 3.12, with M being the dimension of $\boldsymbol{F}_t$). The closer they are, the higher confidence we have that the observation $\boldsymbol{F}_t$ is emitted from distribution $\boldsymbol{b}_i$. To derive the *probability* rather than the confidence level, $b_i(\boldsymbol{F}_t)$ for all n classes must to be evaluated.

$$C\big(\boldsymbol{F}_t \ \big|\ X_t = c_i\big) = \boldsymbol{b}_i(\boldsymbol{F}_t), \tag{3.10}$$

$$b_i(\boldsymbol{F}_t) = \sum_{k=1}^{K} \omega_{ik}\ b_{ik}(\boldsymbol{F}_t)\ \ , \quad \text{where} \tag{3.11}$$

$$b_{ik}(\boldsymbol{F}_t) = \frac{\exp\big(-0.5 \cdot (\boldsymbol{F}_t - \boldsymbol{\mu}_{ik})\boldsymbol{\Sigma}_{ik}^{-1}(\boldsymbol{F}_t - \boldsymbol{\mu}_{ik})^T\big)}{(2\pi)^{(M/2)}\sqrt{|\det(\boldsymbol{\Sigma}_{ik})|}}. \tag{3.12}$$

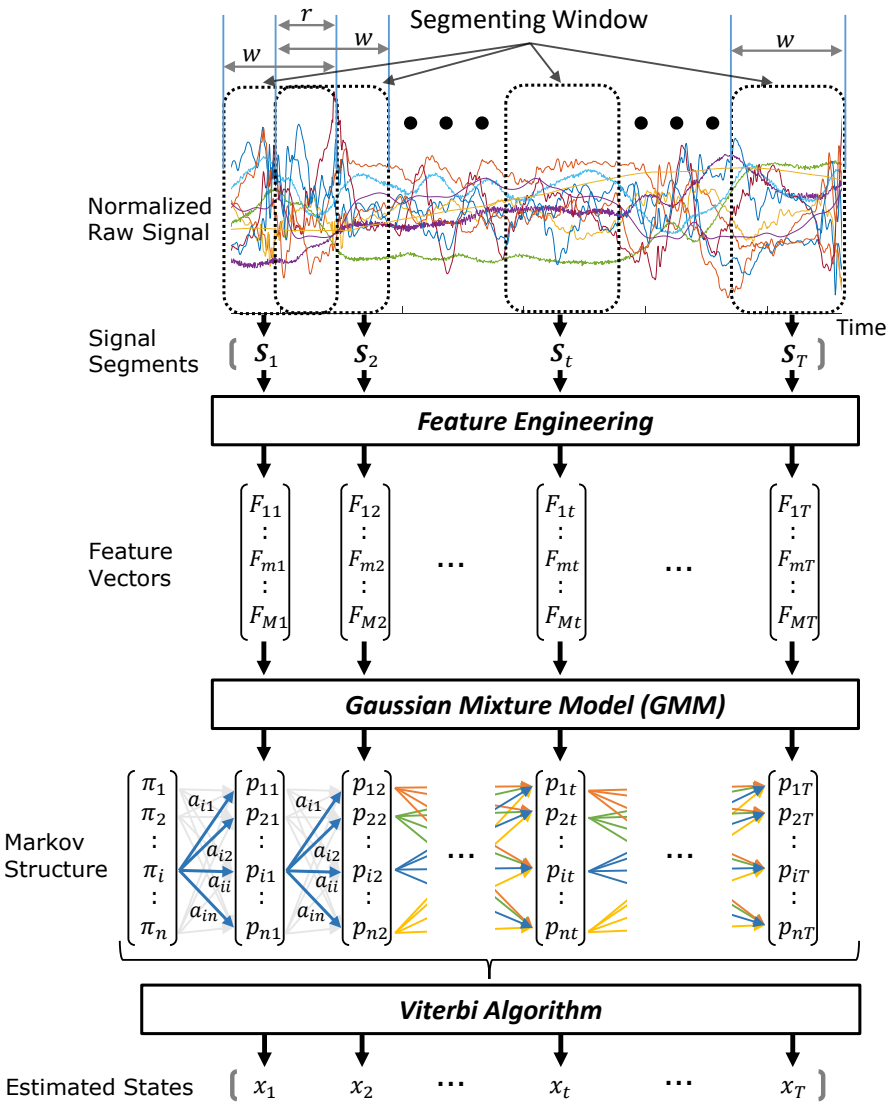


Figure 3.4: Architecture of 'sequential frame' classification using HMM. Extended on Figure 3.1, each feature vector $\boldsymbol{F}_t = \{F_{1t}, \cdots, F_{Mt}\}$ is now inputted into a GMM to generate the vector of **affiliation probabilities** $\boldsymbol{P}_t = \{p_{1t}, \cdots, p_{it}, \cdots, p_{nt}\}$ with $p_{it} = P(X_t = c_i|\boldsymbol{F}_t)$, Eq. 3.17. The vectors of affiliation probabilities are connected by a Markov Structure, with π_i denotes the prior distribution of class c_i and a_{ij} the transition probability from class c_i to c_j. The sequence of most likely states $\{x_t, \cdots, x_T\}$ are estimated by the *Viterbi* algorithm [143].

Since $\{\boldsymbol{b}_1, \cdots, \boldsymbol{b}_n\}$ have been identified in the training stage, the confidence value for each observation emitted from every class is captured in the confidence matrix $\boldsymbol{C}$ in Eq. 3.14. Normalizing each column of $\boldsymbol{C}$ gives the **affiliation probability** matrix $\boldsymbol{P}$ for all observations and motion classes combinations (Eq. 3.17). Following normalization, each column of $\boldsymbol{P}$ sums to 1, indicating that each observation must has been emitted from one of the motion classes.

$$\boldsymbol{C} = \begin{bmatrix} b_1(\boldsymbol{F}_1) & \dots & b_1(\boldsymbol{F}_T) \\ \vdots & \ddots & \vdots \\ b_n(\boldsymbol{F}_1) & \dots & b_n(\boldsymbol{F}_T) \end{bmatrix} \tag{3.13}$$

$$= \begin{bmatrix} C\big(X_1 = c_1 \,\big|\, \boldsymbol{F}_1\big) & \dots & C\big(X_T = c_1 \,\big|\, \boldsymbol{F}_T\big) \\ \vdots & \ddots & \vdots \\ C\big(X_1 = c_n \,\big|\, \boldsymbol{F}_1\big) & \dots & C\big(X_T = c_n \,\big|\, \boldsymbol{F}_T\big) \end{bmatrix} \tag{3.14}$$

$$\boldsymbol{P} = Normalize(\boldsymbol{C}) \tag{3.15}$$

$$= \begin{bmatrix} p_{11} & \dots & p_{1T} \\ \vdots & \ddots & \vdots \\ p_{n1} & \dots & p_{nT} \end{bmatrix} \tag{3.16}$$

$$= \begin{bmatrix} P\big(X_1 = c_1 \,\big|\, \boldsymbol{F}_1\big) & \dots & P\big(X_T = c_1 \,\big|\, \boldsymbol{F}_T\big) \\ \vdots & \ddots & \vdots \\ P\big(X_1 = c_n \,\big|\, \boldsymbol{F}_1\big) & \dots & P\big(X_T = c_n \,\big|\, \boldsymbol{F}_T\big) \end{bmatrix} \tag{3.17}$$

Note that a motion state x_t for observation $\boldsymbol{F}_t$ can be found at this point by looking for the class that gives the highest emission probability in $\boldsymbol{P}$ (Eq. 3.18). This is what has been called a **single frame** classifier. While this identifies the most likely class at each time instant, the probability of a sequence of states is not considered. A **sequential frame** classifier must be used to leverage sequential information carried by the data.

$$x_t = \underset{c_i \in \{c_1, \cdots, c_n\}}{\arg\max} \; P\big(X_t = c_i \,\big|\, \boldsymbol{F}_t\big) \tag{3.18}$$

The well known dynamic programming algorithm, *Viterbi Algorithm* [143], is used to find the most likely sequence of motion classes $\{x_1, \cdots, x_T\}$ given the sequence of state observations $\{\boldsymbol{F}_1, \cdots, \boldsymbol{F}_T\}$. Provided with the dynamical information of the system (i.e. transition matrix $\boldsymbol{A}$, initial distribution $\boldsymbol{\pi}$ and emission (affiliation probability) matrix $\boldsymbol{P}$), the *Viterbi Algorithm* conducts a forward pass to estimate the probability of sub-paths that are associated with the motion

class selection for each observation. The algorithm then executes a backward pass along the highest probability (sub) path and determines the motion class for each observation, i.e. $\{x_1, \cdots, x_T\}$. This completes the sequential frame classification.

3.5 Working without Feature Engineering: An LSTM-HMM Framework for Sequential State Classification

An LSTM RNN [133] consists of cells feeding inputs to one another in a chain-like fashion. These cells act as memory, with both long-term and short-term dependencies. Thus, the prediction of an output is based on the history of inputs and not just the current timestep information. The integration of this memory loop allows for the flow of information between timesteps, making the framework ideal for the processing of time series data. A Bidirectional LSTM [145] can process a segment of data both forward and backward to provide more insights about the signal segment.

For the signal sequence associated with each state, (e.g. $\boldsymbol{S}_t = [\boldsymbol{q}_1^{(t)}; \cdots; \boldsymbol{q}_l^{(t)}; \cdots; \boldsymbol{q}_L^{(t)}]$ for t-th state with $\boldsymbol{q}_l^{(t)}$ being the l-th data point of the state), the trained LSTM network will parse the signal sequence $\boldsymbol{S}_t$ and provide a vector of **affiliation probabilities** ($\boldsymbol{P_t}$, Eq. 3.19) through its *softmax* layer [145]. With the i-th entry of $\boldsymbol{P_t}$ characterizing the affiliation probability of the motion state being c_i, i.e. $p_{it} = P(X_t = c_i|\boldsymbol{S}_t)$, the collection of $\{\boldsymbol{P_1}, \cdots, \boldsymbol{P_T}\}$ serves exactly the same function as the $\boldsymbol{P}$ matrix (Eq. 3.17) required by the *Viterbi* algorithm in Section 3.4.1.

$$\boldsymbol{p}_t = \begin{bmatrix} p_{1t} \\ p_{2t} \\ \vdots \\ p_{it} \\ \vdots \\ p_{nt} \end{bmatrix} = \begin{bmatrix} P\big(X_t = c_1 \mid \boldsymbol{S}_t\big) \\ P\big(X_t = c_2 \mid \boldsymbol{S}_t\big) \\ \vdots \\ P\big(X_t = c_i \mid \boldsymbol{S}_t\big) \\ \vdots \\ P\big(X_t = c_n \mid \boldsymbol{S}_t\big) \end{bmatrix} \tag{3.19}$$

As a result, an LSTM framework can be employed to take place of the feature engineering and GMM operations shown in Figure 3.4, and being combined with a Markov structure to form the LSTM-HMM framework. Figure 3.5 gives a global view of the LSTM-HMM framework. The raw sensor signal first is normalized so that each signal channel has mean zero and SD one, then is segmented into multiple (overlapped) windows. Consecutive windows are overlapped to better capture sudden 'events' that might be cut in half if there is no overlap. And we define each

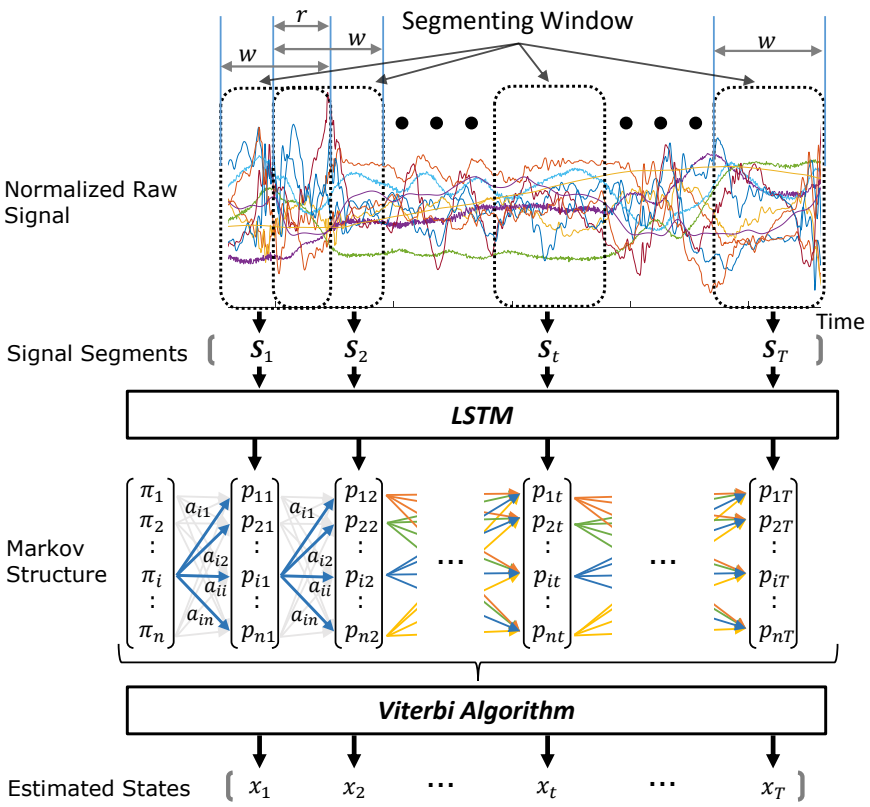


Figure 3.5: The LSTM-HMM architecture. Advanced on Figure 3.4, each signal segment (e.g. $\boldsymbol{S}_t$ for the t-th state) is now directly inputted into a LSTM structure to generate the vector of affiliation probabilities $\boldsymbol{P}_t = \{p_{1t}, \cdots, p_{it}, \cdots, p_{nt}\}$ with $p_{it} = P(X_t = c_i|\boldsymbol{S}_t)$. The vectors of affiliation probabilities $\{\boldsymbol{P}_1, \cdots, \boldsymbol{P}_T\}$ are then synthesised by the Markov model to produce the sequence of most likely states $\{x_t, \cdots, x_T\}$.

window to correspond to a state as before (Section 3.4). The combined LSTM-HMM method leverages the LSTM recurrent neural network [133] to process each signal segment individually and an HMM framework to link outputs from LSTM across multiple windows to make the final decision about state assignments. This two-layer architecture allows the framework to take into account the surrounding dynamic states to provide a better prediction of the current state. Yet with the HMM synthesising LSTM outputs, the framework remains a 'white box' at the top level.

3.6 A Case Study on Dolphin Motion Classification

The proposed frameworks were used to classify swimming states of bottlenose dolphins (*Tursiops truncatus*) using data from biologging tag. Figure 3.6 shows a dolphin with a biologging sensor package (MTag, Section 2.2) on its back. The IMU's orientation with respect to the dolphin is: x-forward, y-left, z-up. A total of 80 minutes and 11 seconds of data was collected from two

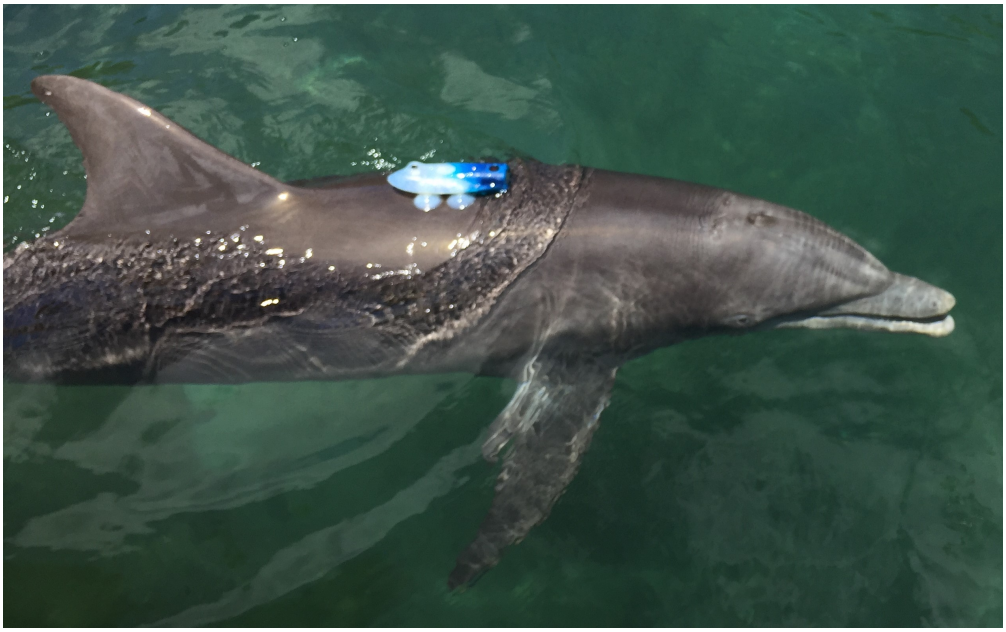


Figure 3.6: A bottlenose dolphin (*Tursiops truncatus*) equipped with a biologging sensor package on its back.

bottlenose dolphins (TT-A and TT-B) located at Dolphin Quest Oahu, Honolulu, HI. The study protocol was approved by the Woods Hole Oceanographic Institution and University of Michigan Institutional Animal Care and Use Committees.

3.6.1 Data Preparation

There were two sources of data in this case study: 1) movement data from a biologging sensor package on the animal, and 2) an overhead video recording.

For biologging sensor data, the *Madgwick* orientation filter [105] was first applied to accelerometer, gyroscope and magnetometer signals to generate an additional 3 data channels that describe the orientation (i.e. roll, pitch and yaw) of the animal. The filter addresses magnetic field distortion and gyroscope drift in quaternion space using a gradient-descent based method, and has been shown to have high accuracy with low computation cost. Features were then extracted from the following 7 data channels: 3-axis accelerometers (Ax, Ay, Az), 3-axis heading (roll, pitch, yaw) and depth. A total of $M_{total} = 79$ features were generated in this case study and an example of the 30 most useful features are provided in Table 3.1.

Concurrent overhead video recordings were scored every second by a human expert to create the 'true' labels of the motion classes (or swimming classes) of the animal to train and evaluate the frameworks. In case the window of a state covered two or more different motion classes

Table 3.1: Top 30 features in ranked feature list. Window size $w = 0.6$ seconds, window overlap $r = 0.12$ seconds (20% of W), using animal data set TT-A. Feature type plus signal channel forms a feature.

Rank	Channel	Feature Type	Rank	Channel	Feature Type
1	Ax	min	16	Depth	waveform length
2	Ay	entropy	17	Ax	max
3	Depth	min	18	Az	max fractal length
4	Ax	mean frequency	19	Depth	rms
5	Yaw	range	20	Ax	std
6	Pitch	max fractal length	21	Ay	mean of abs
7	Ay	mean frequency	22	Az	min
8	Ax	mean	23	Yaw	std
9	Pitch	std-abs-diff	24	Pitch	waveform length
10	Depth	max	25	Depth	mean
11	Ax	waveform length	26	Yaw	max fractal length
12	Ax	entropy	27	Ax	range
13	Ay	max	28	Ay	min
14	Yaw	waveform length	29	Depth	range
15	Ax&Az	correlation	30	Roll	waveform length

according to video labels, the dominant (greater than 75%) motion class would be considered as the 'true' motion class of this state. A total of 7 motion classes were examined in this case study: (c_1) swimming sideways; (c_2) swimming upside down; (c_3) sharp turning; (c_4) respiration; (c_5) targeted straight line swimming; (c_6) Cruising, i.e. swimming in a large circular path; (c_7) interacting, i.e. interacting with human trainer.

3.6.2 LSTM RNN Configurations

The Forward and Bidirectional LSTM in this work were implemented in MATLAB and ran in GPU mode. The mini-batch size was chosen heuristically to be 16. Both the Forward and Bidirectional networks contained two hidden layers with 100 hidden units for each layer. The network was configured to use a sequence of data input (i.e. $\boldsymbol{S}_t$) to estimate one class label as well as the corresponding affiliation probability $\boldsymbol{P}_t$ (Eq. 3.19) for all classes, which was later used in HMM to make improved state estimation. The number of training epochs varied from 55 to 170 depending on the size of the training data and networks.

3.7 Validation Results and Discussions

Two general frameworks are introduced in this chapter: (1) Systematic feature engineering combined with HMM for sequential state classification and (2) LSTM combined with HMM for sequential state classification. This section presents the results for the two frameworks one at a time.

3.7.1 Feature Engineering - HMM

In this work, training and testing were conducted for each dataset (TT-A or TT-B) separately. Each dataset was first segmented into T states using window size W as discussed in Section 3.2 and $M_{total} = 79$ features were generated for each state to form the feature library as stated in Section 3.3.2. Then a portion of randomly selected states were used as the training set, while the rest of the states were reserved for validation and testing. The randomly selected training set is also a good analogy to the real-world case where behaviors of animals are recorded by human observers sporadically. Note that 'true' motion classes of states $\{x_1, \cdots, x_T\}$ for both the training set and testing set were fully obtained via video processing.

The proposed sequential-frame classification strategy performed more accurately, especially when the number of features used was small, than the other single-frame classifiers: GMM (work in a single-frame manner as discussed in Section 3.4), Naive Bayes (NB), SVM, KNN, DT. After feature generation and ranking for a certain window size ($w = 0.6$ seconds in this example), the effect of the number of features used on performance of the classifiers is demonstrated via Figure 3.7. The specified number of features were selected from the ranked feature list (e.g. Table 3.1), with the top ranked features chosen first. For example, if the number of features equals eight, then the top 8 features from the ranked list would be selected for classification and analysis. Accuracy was defined as the ratio of the number of correctly classified states over the number of all states. Classification precision and recall were calculated for each motion class, with the average precision and recall for all classes reported in Figure 3.7. From the results it can be seen that classification performance was poor when fewer than 10 features were selected. However, performance was not seriously affected by redundant features in this case study either. In general, all classification methods examined here performed well once more than 20 features were used, but computation efficiency was reduced as more features were used for the classification.

The adaptive weighted sampling scheme presented in this work provided a systematic method for determining the best number of features for the classification (inner loop in Figure 3.2), while easing the computation burden of the algorithm. In Figure 3.8, the *truth accuracy curve* and the *truth interest curve* were considered unknown and waited to be sampled. Recall that the *truth interest curve* is estimated from the *truth accuracy curve* via Eq. 3.3, where a cost (negative reward) was assessed on the number of selected features. This resulted in an increased probability of

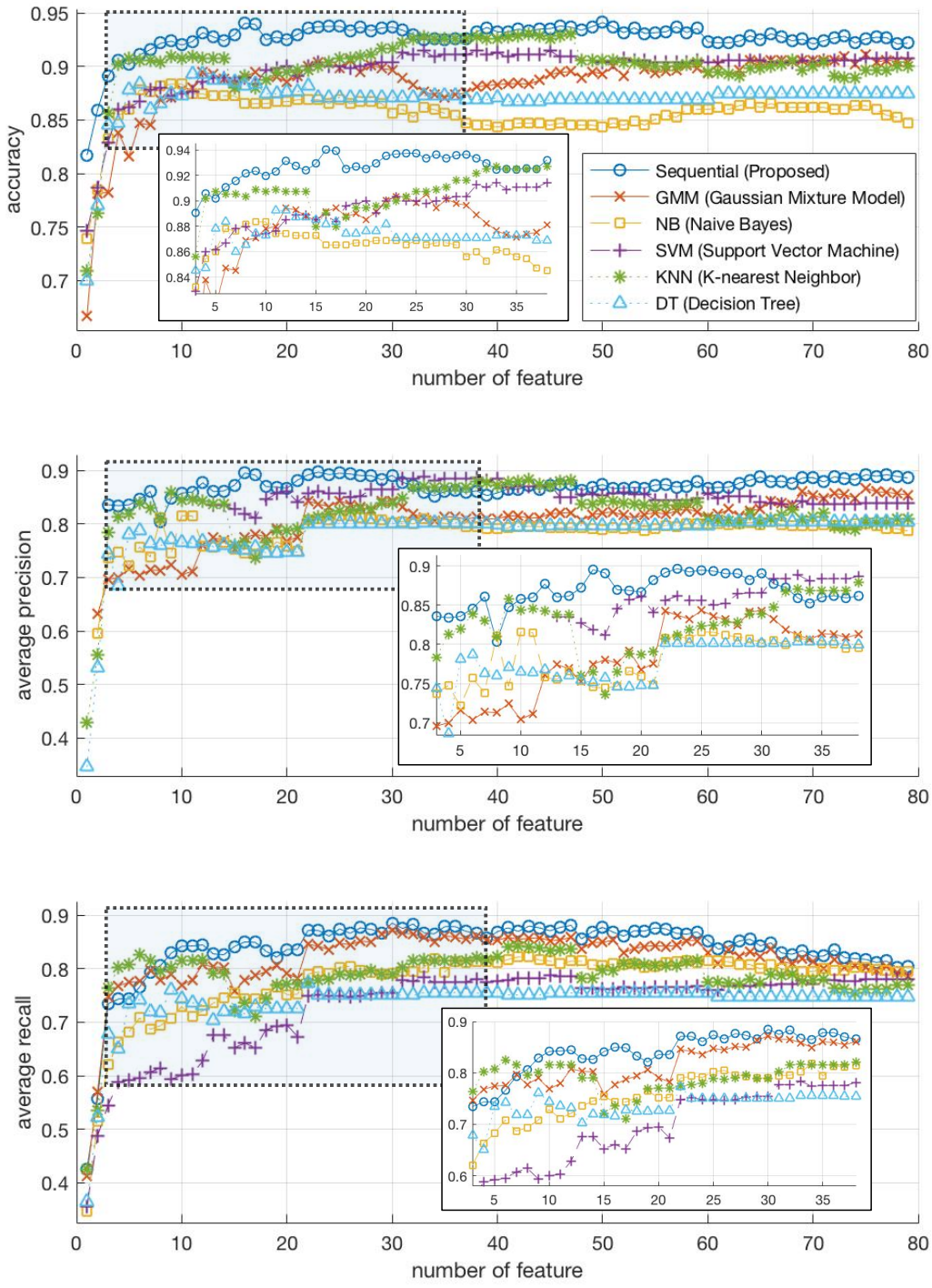


Figure 3.7: Classification performance of different classifiers over the number of features included. Top plot shows overall accuracy vs. number of features used, middle plot shows the average precisions of each classifier while the bottom plot shows average recall. Window size $W = 0.6$ seconds, window overlap $r = 0.12$ seconds (20% of W), animal data set TT-A, data train-test ratio is 70% - 30%.

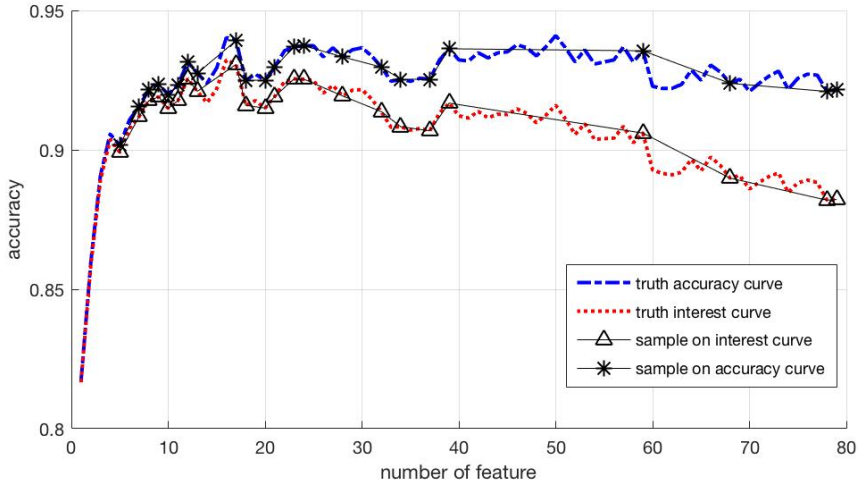


Figure 3.8: Adaptive weighted sampling scheme applied on the relationship curve between *accuracy* and *number of features*. The overall number of samples took was 30% as the size of the search space (i.e. all possible choices of *number of feature*). Window size $W = 0.6$ seconds, window overlap $r = 0.12$ seconds (20% of W), using animal data set TT-A, data train-test ratio is 70% - 30%.

sampling regions that had both good classification accuracy and a smaller number of features. As the number of sampled points increased, the regions of high interest were characterized with much higher resolution than low interest regions. Using this approach, 16 features were selected for the example setting.

Performance of the classifiers over different window sizes was determined after the best feature set was selected for each window size (outer loop in Figure 3.2). Figure 3.9 presents classification accuracy for different algorithms over different window sizes (top plot) and the corresponding number of features used to reach to that accuracy (bottom plot). Accuracy was higher for window sizes $W = \{0.3,\ 0.6,\ 1.0,\ 1.5\}$ (seconds) while it was lower for $W = \{0.4,\ 0.8,\ 1.3\}$ (seconds). Both animal datasets (TT-A, TT-B) follow this trend while only TT-A is presented here. The window size with the highest interest (according to Eq. 3.3) was selected for the analysis: $W = 0.6$ (seconds). Using the 0.6 seconds window size and the classifiers' corresponding best number of features, the proposed sequential-frame classification strategy performed better (on average over datasets TT-A and TT-B) than the single-frame classifiers by: 4.8% (GMM), 8.3% (NB), 4.0% (SVM), 2.4% (KNN), 8.1% (DT).

To investigate how the classifiers perform with a limited set of training data, different training-testing ratios were tested over the two datasets (TT-A, TT-B). As shown in Figure 3.10, the performance of all the classifiers was fairly stable over most ratios, while the sequential-frame classification strategy continues to achieve the highest accuracy value. The proposed sequential-frame

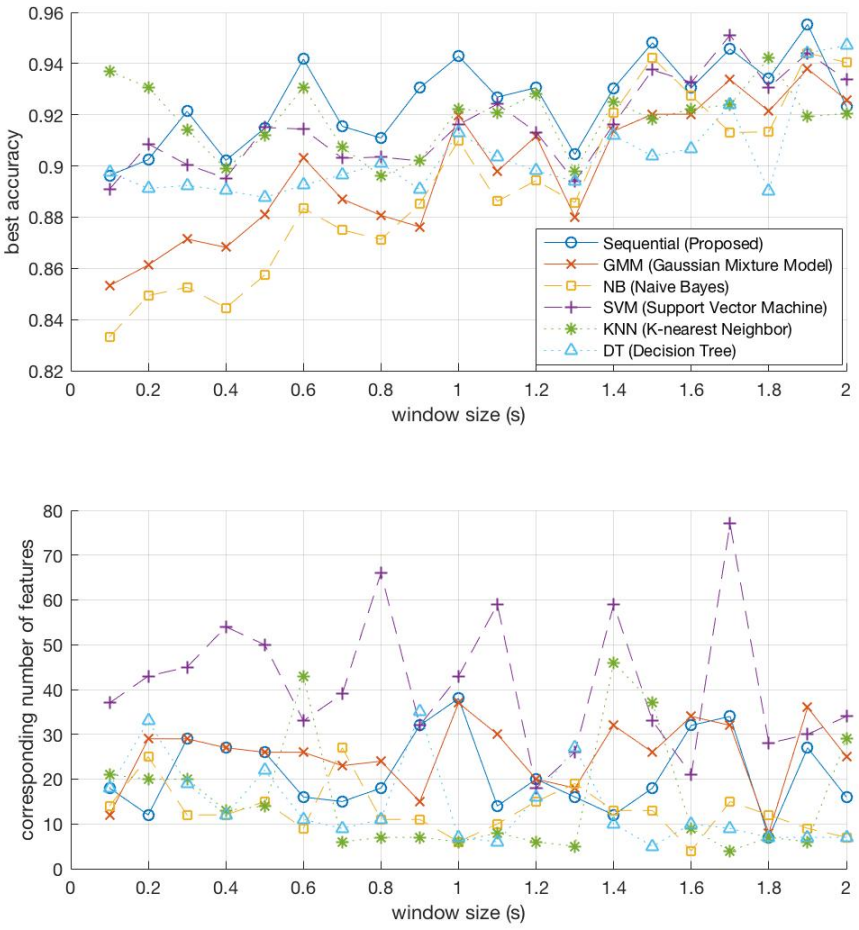


Figure 3.9: Classification performance of different classifiers over window size selections. Top plot shows best accuracy vs. window size used, while bottom plot gives the corresponding number of features used for the best accuracy. Best accuracy is found according to Eq. 3.3. Window overlap r is 20% of window size W, animal data set TT-A, data train-test ratio is 70% - 30%.

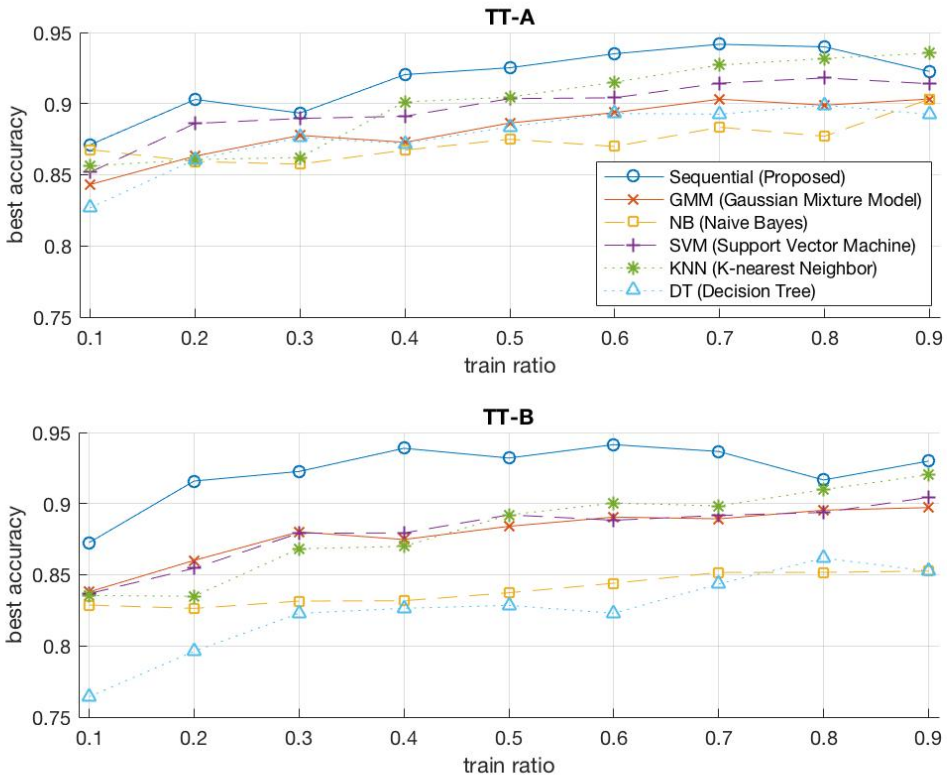


Figure 3.10: Classification performance of different classifiers over different train/test ratios. Top plot uses animal data set TT-A, bottom plot uses TT-B. Best accuracy is found according to Eq. 3.3. Window size $W = 0.6$ seconds, window overlap $r = 0.12$ seconds (20% of W).

classification strategy showed accuracy values of 87% or greater starting with a training ratio of 10%. These results indicate that the proposed framework should accurately classify states in long datasets where a human observer can only make sparse observations about the motion class of the animal.

3.7.2 LSTM - HMM

The proposed LSTM-HMM framework was compared to several well known classification methods including KNN, SVM, NB and DT classifiers. We adapted the choice of window size of 0.6 seconds and the top 30 most useful features identified in Section 3.7.1 for the other classifiers to compare the LSTM-HMM framework with. In practice, the distribution of motion classes is likely to be unbalanced, where certain classes appear more frequently than others. In Section 3.7.1, a fixed **portion** (e.g. 60%) of data corresponding to each class was used for training, resulted in an unbalanced training set given that the more popular classes in the data set showed up more in the training set. In the present section, a fixed **volume** (e.g. 10 minutes) of data corresponding

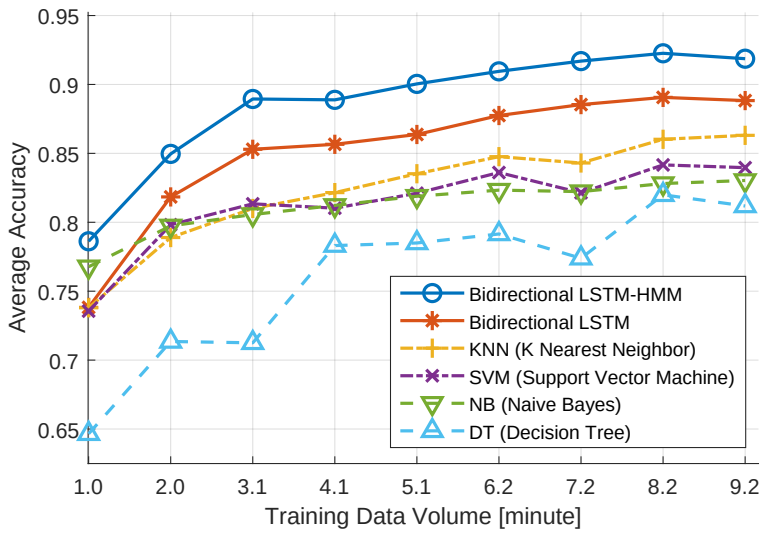


Figure 3.11: Comparison between classification methods. Average classification accuracy (over 10 random runs) plotted against varying training data volume. Total data volume is 35 minutes.

to each class was used to train the framework. In this case, the training set was balanced and appeared closer to the application where people collect training data for each class and then apply the trained framework to a larger dataset with unknown class distributions. Note that special compensations (e.g. dedicated data collections, elimination of a class) are needed in the presence of rare classes that occur every few hours or even every few days. The performance of different volumes of training data were investigated in this section. Ten random runs were performed for each training-testing cycle and the average classification **accuracy** was reported to characterize the preliminary performance. The metric **accuracy** was defined as the ratio of the number of correctly classified states over the number of total states as before (Section 3.7.1), which is one of the most important aspects people care about when dealing with real world data.

Figure 3.11 compares the average classification accuracy (over 10 random runs) of the proposed LSTM-HMM framework and alternative methods over varying amounts of training data volume. The LSTM-HMM framework is denoted as "Bidirectional LSTM-HMM", which demonstrates the highest accuracy, achieving greater than 90% (a minimum of 6% more accurate than other methods) when trained using over 5 minutes of balanced data. In comparison, the 'Bidirectional LSTM' (label predicted by LSTM directly without HMM) results in an accuracy level that is only 3% higher than the most accurate of the alternative methods, and 3% lower than the LSTM-HMM method. The intuition behind the accuracy improvement of LSTM-HMM framework is that HMM is 'biasing' the estimation results from LSTM towards the more 'correct' direction by utilizing the state transition structure.

The comparison of accuracy of Bidirectional LSTM and Forward LSTM is shown in Fig-

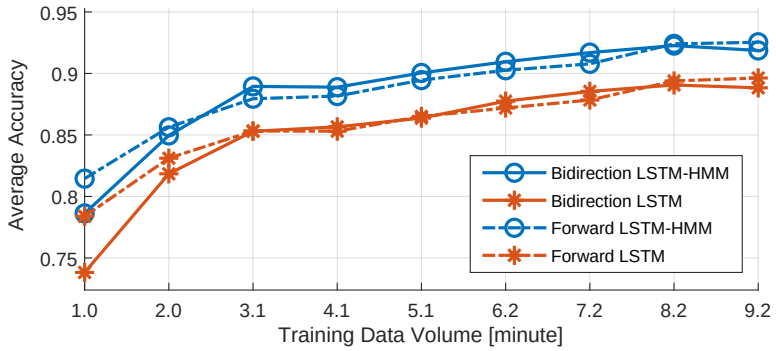


Figure 3.12: Comparison between Bidirectional and Forward LSTM. Average classification accuracy (over 10 random runs) plotted against varying training data volume. Total data volume is 35 minutes.

ure 3.12. For the "LSTM-HMM" and "LSTM" approaches, the Bidirectional and Forward LSTM methods are comparable with Bidirectional demonstrating slightly higher accuracy values. But we do note that the Bidirectional method learns faster (in terms of training epochs) than the forward LSTM method and is generally able to handle more complex system dynamics.

In addition to outperforming many traditional methods, the proposed LSTM-HMM framework strongly focuses on transitions between dynamical states, which not only helps with classifying dynamic states more accurately, but also provides insights into the animal's motion pattern that may be of interest to biologists or software specialists aiming to mimic an animal's behavior. Figure 3.13 visualizes the state transition probabilities over the entire data set. The strongest state transitions are of the states transitioning to themselves, with probabilities averaging in the low nineties.

To evaluate the robustness of the state transition probabilities as a function of training data, Figure 3.14 shows the state transition probabilities over one instance of a 5.1 minute training set. The state transition probabilities in Figure 3.14 are quite similar to those in Figure 3.13. This indicates that the state transition matrix captures the essence of the transitions regardless of training data. Yet, this speaks to the LSTM-HMM framework's potential in adapting to different animals. As the motion states of the animals under different conditions would vary in terms of the transition matrices, the LSTM-HMM framework could be further coupled with an Expectation-Maximization (EM) algorithm at the top level to update and learn the transition matrix with data from a new scenario automatically without human label. This is part of the benefit brought to us by the 'white box' nature of HMM. A framework like LSTM-LSTM, would not be able to achieve that easily since both levels of LSTM need to be trained by labeled data.

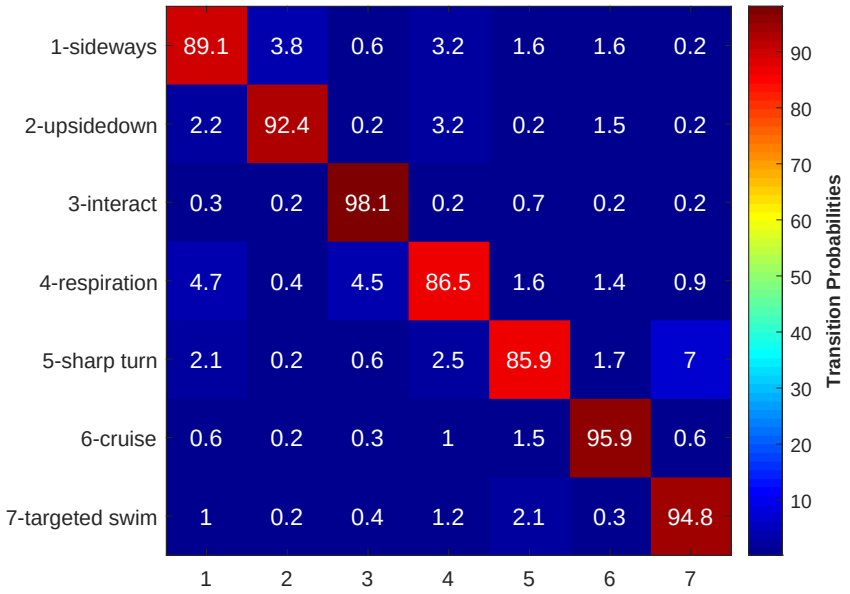


Figure 3.13: State transition probabilities obtained over all data (35 minutes). The entry at i-th row j-th column denotes the state transition probability (%) from class i to j, which is also color coded. Note that the transition probabilities from each row sum to 1 (100%).

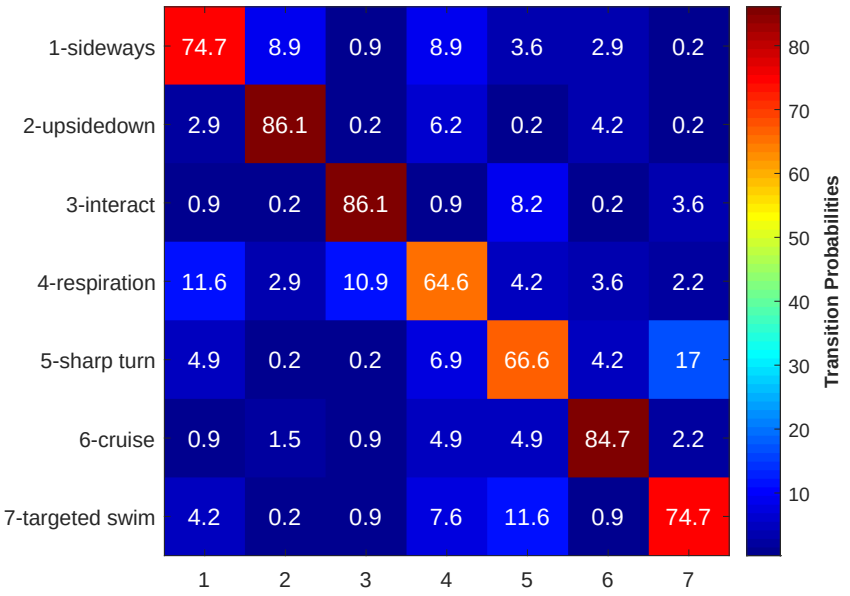


Figure 3.14: State transition probabilities obtained over one instance of training data set. The volume of this training set is 5.1 minutes.

3.8 Summary

This work first presents a systematic approach for feature generation, feature selection and state classification from biologging sensor data, and demonstrates how the proposed framework can be used to accurately classify bottlenose dolphin swimming states when only sparse motion observations are available for training. The proposed framework addresses questions about how to generate and select features from data, and provides a systematic approach for efficient window size selection. Sequential temporal information was also used to improve classification accuracy via a HMM during state classification, resulting in 4.8% (GMM), 8.3% (NB), 4.0% (SVM), 2.4% (KNN), 8.1% (DT) improvements when comparing with single-frame classification schemes.

This work then proposes an LSTM-HMM framework that achieves above 90% classification accuracy by heavily utilizing sequential temporal information that is embedded in the time series signal at both a point-to-point bottom level (LSTM) and state-to-state top level (HMM). Importantly, the LSTM-HMM framework does not need statistical feature engineering and only requires a small amount of training data to achieve reasonable classification accuracy, making this approach easily adaptable to new tasks. Also, as compared with LSTM-LSTM, the 'white box' at the top level of LSTM-HMM framework makes it possible to apply model trained on one dataset to another dataset with different state distributions while also making the state inferences interpretable.

Overall, the heavy focus posed on the transitions between motion states is capable of providing valuable intuition to more biologically driven questions, since it is often of interest to understand why a certain motion state arises. Moreover, the hybrid frameworks can be of use in cases where there are complex high level behaviors that encompass several low level motion states. The HMM step can identify loosely defined sequential information to draw conclusions about higher level behaviors.

Depending on the application scenario, feature engineering and LSTM can both be of great help in practice. Further study needs to be conducted to understand why window size selection affected classification performance. It could be similar to the idea of 'resonance frequency', where a particular window size fits better with the motion signal. Future work shall include: (1) Investigating how to identify the best window size and feature set for individual motion classes, as opposed to using one window and feature set for the entire set of classes. (2) Further testing and validation of classification performance when the algorithm has been trained using data sets with labeled information from a particular animal, but applied to new data that were not in the original training sets (i.e. data from different trials of the same animal or from different animals). That is, validating the framework with a larger dataset to explore the limits of the frameworks under extreme conditions. Finally, we will apply the framework towards the classification and analysis of higher level behaviors (e.g. coordinated swimming, playing etc.) in addition to motion states.

CHAPTER 4

Identity Preserving Multi-Target Tracking with Cameras and Biologging Tags

4.1 Background

As cameras and biologging sensors are becoming ever more ubiquitous, enabling techniques that leverage these distributed sensor networks have the potential to benefit a myriad of research areas, including but not limited to: performance analysis of the coordination of controllable and uncontrollable agents, development of bio-inspired multi-agent control strategies, and the study of biological systems themselves. The combined camera-tag sensor system presented in this chapter enables the analysis of both positions and detailed dynamics of the agents, which enables the investigation of challenging research questions about how best to combine the sensor streams to calculate or estimate parameters about the agents of interest. This work aims at tracking and localizing underwater agents in particular and the agents of interest are bottlenose dolphins (*Tursiops truncatus*). A framework (Figure 4.1) is developed to address the multi-target tracking problem from a localization perspective, using per-frame detections from video collected by stationary cameras, and dynamic information from the agents of interest measured with biologging sensors. The glare created by light reflecting off the water, the refraction caused by the air/water interface, and the 3D motion of the agents all complicate this tracking problem. By posing a tracking problem from a localization perspective, this approach has advantages over both pure-camera-based tracking and pure-biologging-based localization.

Content of this chapter is mainly associated with the following publication:

[146] 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," in *IEEE International Conference on Robotics and Automation (ICRA)*. May 20-24, 2019, Montreal, QC, Canada: IEEE, 2019, pp. 6575–6581

* Corresponding Author

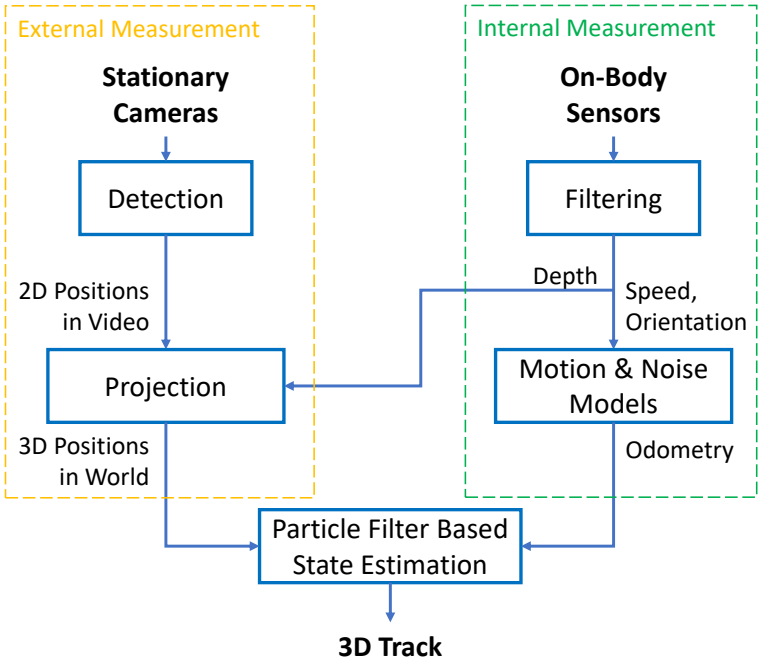


Figure 4.1: System block diagram of the proposed multiagent localization and tracking system.

Advancements in computer vision and deep learning [61] in the past two decades provide the foundation for the development of the automated tracking system for marine animals. With appropriate training data and configurations, many of the well established object detection frameworks [62, 63, 64, 65, 66, 67, 68] can be used to reliably detect, locate and classify objects of interests (e.g. dolphins, people, buoys or a bicycle if needed) that are visible within each image frame. However, the challenge in this work is how can we link the detections in consecutive image frames into dynamic tracks across the video while maintaining the identity of each track through the entire event. This is not an easy task, particularly in the marine environment.

Significant environmental occlusions (e.g. surface glare) make it nearly impossible for a camera-based multi-target tracking system to continuously track an object of interest, especially in the case where multiple similarly-looking highly mobile objects (e.g. dolphins) are moving independently in the environment. Popular tracking techniques are limited in these situations. For example, the Joint Probabilistic Data-Association Filter (JPDAF [69]) tracks a fixed and known number of objects without any merged measurements in its basic implementation. Additionally, Multiple Hypothesis Tracking (MHT [70, 71, 72]) can be used to achieve state-of-the-art performance when augmented with higher-order appearance information about each object [73]. However, long-term occlusions from other objects or the environment are not accounted for in this approach, and the objects need to demonstrate appearance differences to assist the tracking. The Explicit Occlusion Model (EOM) is constructed under a network flow framework to track through long-term inter-object occlusions in [74], though occlusions from the environment are still not

addressed. In [75], appearance cues of objects are used to prevent identity switching, which is not applicable to indistinguishable objects (e.g. dolphins). For all of these approaches, the main challenge is that there is no information about the object during occlusion, and broken associations remain broken.

Alternatively, data from biologging sensors, like acceleration and heading, can be integrated to form continuous dead-reckoning tracks, but track drift created by accumulated sensor errors are well understood [52]. Additional information is required to enhance the accuracy of the tracks generated from tag inertial data. GPS is a straight forward technique for obtaining absolute positional information and can be combined with dead-reckoning information to obtain finer tracks [51], but it is not available underwater or indoors, and is not accurate enough for meter-scale applications. Setting up landmarks/beacons with known positions for agents to observe is often very effective (e.g. [53, 54, 55]), but only applies to limited types of systems and can require additional infrastructure and sensors. Using the agent as a vehicle to sense and/or reconstruct the environment itself, such as in the kidnapped robot (e.g. [56]) and SLAM (Simultaneous Localization and Mapping) problems (e.g. [57, 58, 59]), can also be used to improve agent localization. However, the required online computing power for data processing, as well as the power and size requirements of the sensors (e.g. camera, lidar, etc.), make these localization systems (currently) hard to fit in a reasonably sized biologging package.

In common localization practices (e.g. [55, 56, 59]), an agent's observations about the environment, or the observations about the relationship between the agent and the environment, are made either from the agent or uniquely associated with the agent. This is natural, as localization concerns finding 'where I am' and the observations are made by or about 'me'. In contrast, the question tracking systems attempt to answer is 'where is everybody' and the observations are about 'everyone'. The challenge of approaching a tracking problem from a localization perspective is how to find 'where I am' using information about 'everybody', given the observer cannot discriminate 'me' from 'everyone else'. As will be discussed, this is addressed using a particle filter embedded with a concept of **fuzzy observation**.

This work presents a platform to track multiple uncontrollable underwater agents using the above-water cameras and on-body sensors. The platform is straightforward in implementation for many similar environments, and can handle nonlinear agent dynamics, non-Gaussian noise, and scattered observations. The system is particularly designed for post-hoc analysis and validations, and is able to maintain a robust track in a noisy environment.

4.2 Problem Formulation

4.2.1 Experimental Environment

The experimental environment in this work was located in the Seven Seas Dolphinarium at the Brookfield Zoo, in Brookfield, IL. Two Allied Vision Prosilica GC1380C cameras, outfitted with ThorLabs MVL5M23 lenses, were mounted to the ceiling to cover the entire environment (Figure 4.2-A). Each camera was enclosed in a Dotworkz D2 security housing to protect it from the saltwater environment, which were in turn mounted to a ceiling support beam through an 80/20 aluminum frame. The cameras were connected to a desktop PC through an Intel PRO/1000 Pt PCIe Ethernet card, for separate direct connections to each camera. Recordings were controlled through the MATLAB Image Acquisition Toolbox, and were captured in the full RGB 24-bit color format, at 20 Hz.

Bottlenose dolphins (*Tursiops truncatus*) served as the uncontrollable agents for this study, and were outfitted with an in-house assembled tag (MTag, Section 2.2), based on the Loggerhead Instruments OpenTag3 board (Figure 4.2-C). The OpenTag3 contains a 9-DOF IMU with an accelerometer, gyroscope, and magnetometer, and has additional sensors on-board to record temperature and pressure. The board also supports the addition of an external 1-DOF Hall effect sensor, which functions as a flow speed sensor when used in conjunction with a magnetic micro-turbine mounted outside the tag housing [107]. The tags were configured to record the IMU at 50 Hz, and all other sensors at 5 Hz. To be minimally invasive and to minimize deployment complexity, the tags are attached to the agents with a set of four silicone rubber suction cups. Deployments can be done within a few seconds.

4.2.2 Objective and Assumptions

The objective of this work is to maintain a track of each tagged agent throughout the entire experiment via the fusion of overhead camera and biologging sensor data. The proposed framework bears the following assumptions about the environment and agents:

- The experimental environment is bounded and the agents that are being tracked stay within the boundary. However, there is no assumption about the agents that are not being tracked; they are free to join or leave.
- There are regions in the environment that have low camera observability, i.e. environmental occlusions.
- Agents can be close to each other to a degree that two or more agents are perceived by

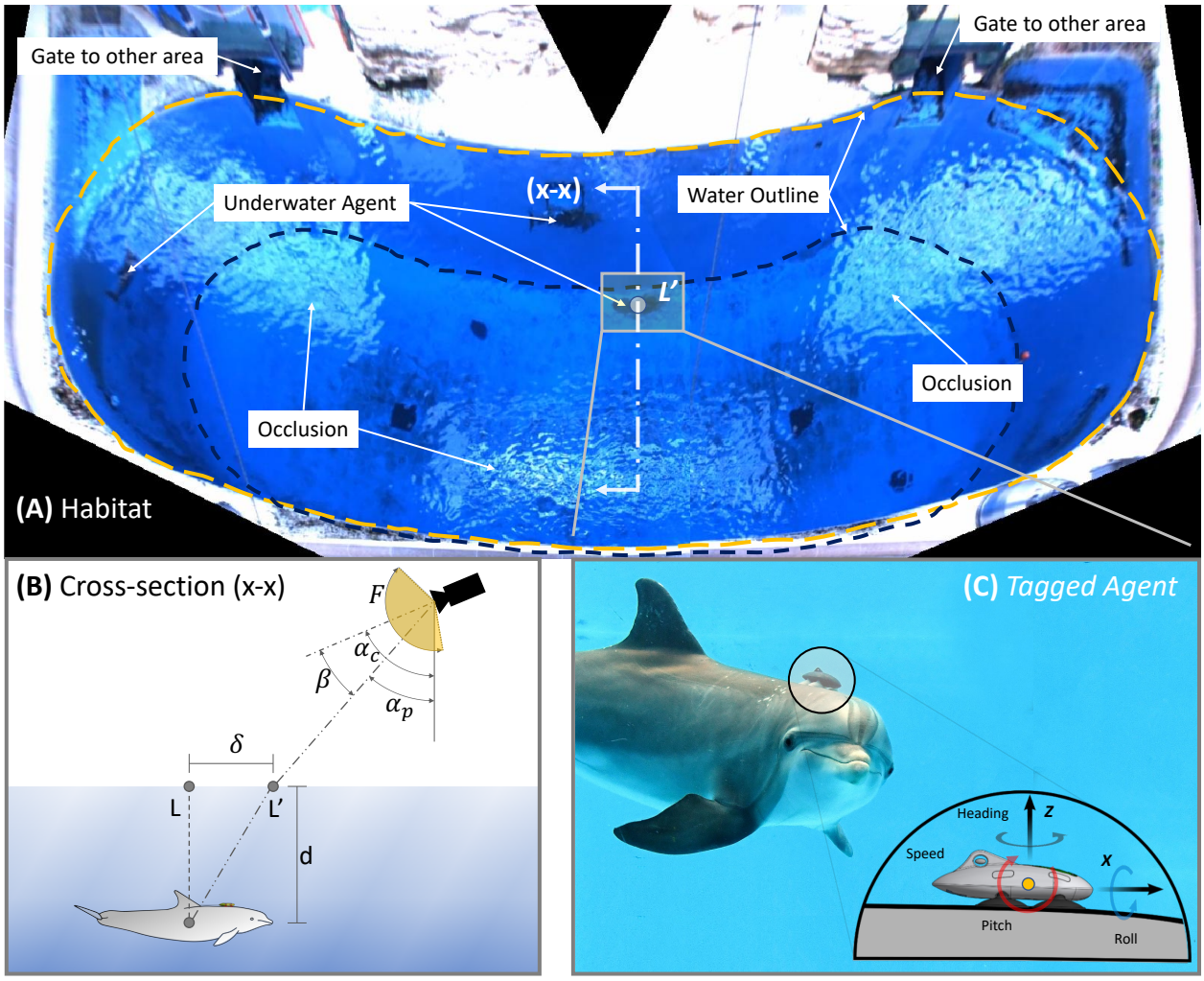


Figure 4.2: (**A**) Transformed top-down view of the experimental environment. The outline of the water at the bottom is shown in black, and at the surface in yellow. Despite the size difference in the image, their projections to the world horizontal plane overlap each other. (**B**) Diagram of projection and refraction effects on estimated dolphin location. The yellow region represents the camera FOV. This cross section is shown in (**A**) as the white dash-dot line. (**C**) Diagram of tag (with state notations) and tagged agent.

the camera as only one, for (potentially) long periods of time, i.e. long-term inter-agent occlusions.

- The initial positions of tagged agents are NOT known.

4.2.3 Framework Outline

The proposed system contains three major components (Figure 4.1): (A) Convolutional Neural Network (CNN) based per-frame detections of agents and the associated handling of position mapping from image to real world, as well as a Kalman filter based camera-only tracking method; (B) Motion & noise model of the tagged agents and the resulting odometry; (C) Particle Filter (PF) based position estimation, which combines information from both cameras and tags. These three components are introduced in the following sections.

4.3 External Measurement – Cameras

4.3.1 Camera Detection

A Faster R-CNN object detector [62] was trained and used to detect both tagged and untagged agents indiscriminately, for each camera. Each detector was trained with separate sets of over 900 labeled images. Locations of all detected agents from both cameras were projected into the top-down view (Figure 4.2-A) using two corresponding camera-to-world homographies, one for each camera.

4.3.2 2D-3D Projection

Since the homographies 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. The edge of the water is one example to visualize this effect: in the transformed top-down view (Figure 4.2-A), the edge of the water at the bottom appears smaller than the edge at the top, when in reality they cover the same general 2D area. This is due to refraction and perspective effects (Figure 4.2-B), where an agent's observed 2D position is shifted from L to L'.

For accurate 3D localization/tracking, it is important to find the 'shifts' for the tagged agents when they are below the surface, i.e. correctly project the 'shifted' 2D image position to a 3D world position. For a known depth provided by the tag, a position in top-down view 2D (Figure 4.2-A) is projected into 3D via the following steps:

- 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.

- Assuming the camera orientation is known, α_c, the angle of the camera image center with respect to the world frame z-axis, 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:

$$\beta = \tan^{-1}\left[(2p/r - 1)\tan(F/2)\right] \tag{4.1}$$

  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:

$$\alpha_p = \beta + \alpha_c \tag{4.2}$$

- 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:

$$\delta = d\tan\left[\sin^{-1}\left(\frac{n_a}{n_w}\sin(\alpha_p)\right)\right] \tag{4.3}$$

  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.

- 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.

- 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:

$$L = L' + \delta_w \tag{4.4}$$

  where L' is the original location estimate.

4.3.3 General Multi-Target Tracking using Above Habitat Video

Each experimental session involved the detection of multiple agents throughout their environment. However, agent detections were done independently for each frame of the video. To extract kinematic information from the agents in the video, the detection associations needed to be preserved

across frames. In this section[1], we demonstrate how short continuous track (i.e. **tracklet**, without individual's identity) was generated for a detected agent by identifying the most likely detection of that agent 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, also as in Figure 4.3.

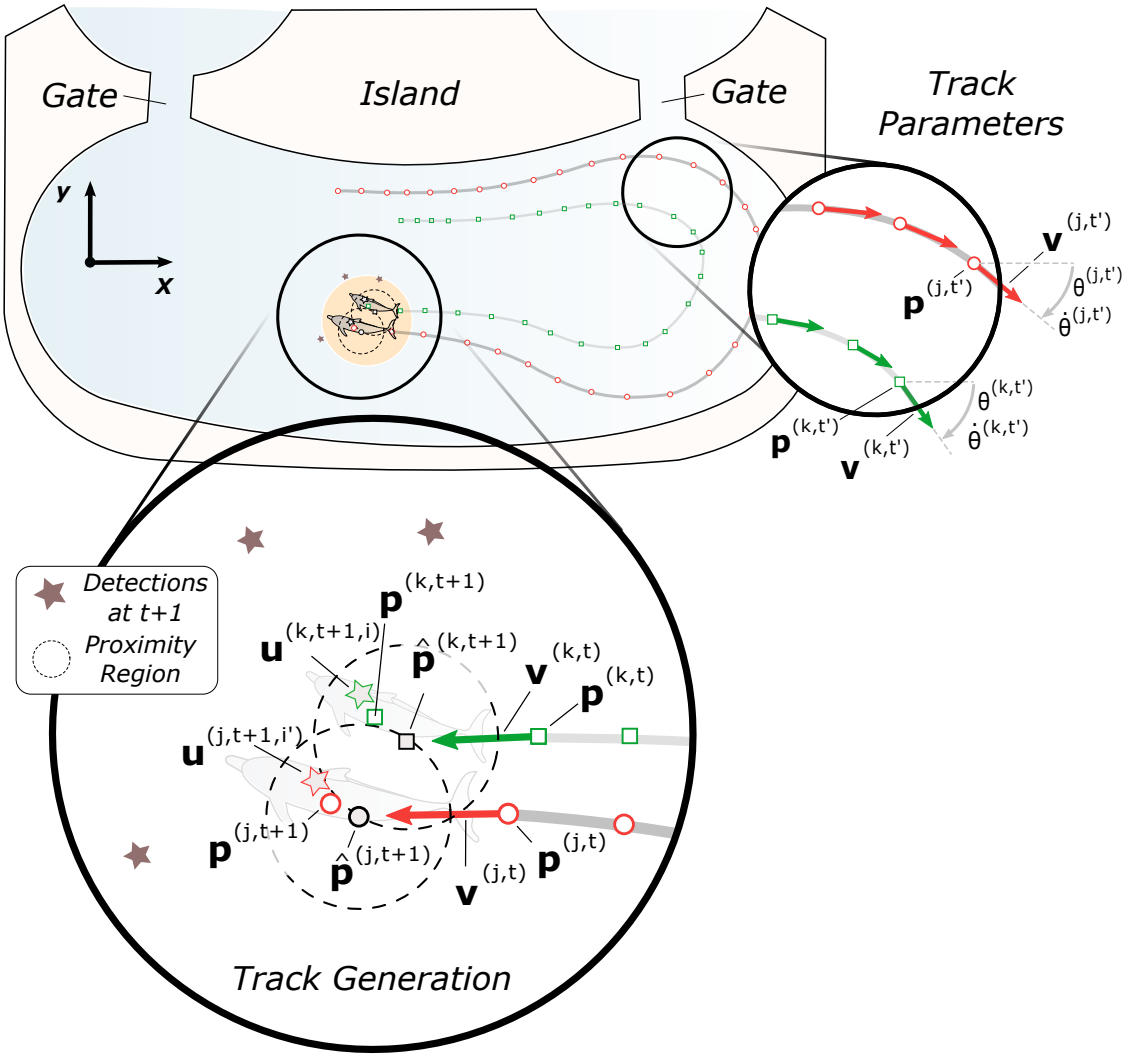


Figure 4.3: Diagram of the Kalman filter based tracklet generation. Shown is the x-y view of example tracklets (red and green on gray lines) of two dolphins (highlighted light orange). ZOOM (LEFT): Illustration of tracklet generation, with detections (stars) and tracklet proximity regions (dashed). Example notation for tracklet j (red): position ($\mathbf{p}^{(j,t)}$), velocity ($\mathbf{v}^{(j,t)}$), Kalman-predicted next-frame position ($\hat{\mathbf{p}}^{(j,t+1)}$), next-frame agent detection ($\mathbf{u}^{(j,t+1,i')}$), and Kalman-filtered next-frame position ($\mathbf{p}^{(j,t+1)}$). 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 ($\dot{\theta}^{(j,t')}$).

[1]This part of the content is associated with our previous publication:

[147] J. Gabaldon, D. Zhang, L. Lauderdale, L. Miller, K. Barton, and K. A. Shorter, "Vision-based monitoring and measurement of bottlenose dolphins' daily habitat use and kinematics," *bioRxiv*, 2021.

The position of the i-th detected agent 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 agent in the next frame ($\widehat{\mathbf{T}}^{(k,t+1)}$) based on the current state information of the agent $\mathbf{T}^{(k,t)}$.

$$\mathbf{T}^{(k,t)} = [\mathbf{p}^{(k,t)}, \mathbf{v}^{(k,t)}] \tag{4.5}$$
$$= [p_x^{(k,t)}, p_y^{(k,t)}, v_x^{(k,t)}, v_y^{(k,t)}] \tag{4.6}$$
$$\widehat{\mathbf{T}}^{(k,t+1)} = [\hat{\mathbf{p}}^{(k,t+1)}, \hat{\mathbf{v}}^{(k,t+1)}] \tag{4.7}$$
$$= [\hat{p}_x^{(k,t+1)}, \hat{p}_y^{(k,t+1)}, \hat{v}_x^{(k,t+1)}, \hat{v}_y^{(k,t+1)}] \tag{4.8}$$

where $\mathbf{p}^{(k,t)} = [p_x^{(k,t)}, p_y^{(k,t)}]$ denotes the filtered position of the agent 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.

$$\hat{p}_x^{(k,t+1)} = p_x^{(k,t)} + v_x^{(k,t)}\Delta t \tag{4.9}$$
$$\hat{p}_y^{(k,t+1)} = p_y^{(k,t)} + v_y^{(k,t)}\Delta t \tag{4.10}$$
$$\hat{v}_x^{(k,t+1)} = v_x^{(k,t)} \tag{4.11}$$
$$\hat{v}_y^{(k,t+1)} = v_y^{(k,t)} \tag{4.12}$$

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 agent 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:

$$\mathbf{T}^{(k)} = [\mathbf{T}^{(k,t_{start})}, \ldots, \mathbf{T}^{(k,t-1)}, \mathbf{T}^{(k,t)}, \mathbf{T}^{(k,t+1)}, \ldots, \mathbf{T}^{(k,t_{end})}]^T \tag{4.13}$$

As illustrated in Figure 4.3, the tracklet formation operation linked each agent's individual

detections (**u**) over consecutive frames and returned not only the positions (**p**) of the agents, 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 agents. We note again that a tracklet would be terminated once the agent became invisible for a few consecutive frames, and it did not maintain the identity of whichever agent was just being tracked. But tracklet formation is still an effective step in 'filtering' out false detections that might have been generated due to water wave glare etc. Also, as will be discussed in Section 4.6, the tracklets facilitate the precise tracking of the agents once their identities are resolved.

4.4 Internal Measurement – Tags

4.4.1 Heading Estimation

Estimated heading of a tagged agent is obtained from the tag IMU (accelerometer, gyroscope and magnetometer) using the Madgwick orientation filter [105]. The filter processes data from the 9-DOF IMU with a gradient-descent based method in a quaternion representation, and has been shown to have high accuracy as well as the ability to handle gyroscope drift and static magnetic field distortion. The estimated orientation can then be directly converted to a rotation matrix, which is used to project the measured speed from tag to horizontal world frame. Special attention should be given if the orientation is represented in the form of Euler angles, as the heading estimation can diverge at certain configurations due to gimbal-locking, a singularity problem inherent in Euler angles.

4.4.2 Motion Model and Odometry

From this point until the end of this chapter, the motion of the agent is considered 2D (x-y plane, the horizontal plane) instead of 3D (x-y-depth). This reduction is possible because the position of the agent in the third dimension (depth) is directly measurable by the tag's pressure sensor, which is not subject to drift, bias or any other significant error. Meanwhile, both the measured speed and camera detections are projected into the x-y plane.

The odometry of the agent (Eq. 4.14) is straightforward after speed (v) and heading (or yaw, θ) are measured. The (higher order) mismatch between the model and the actual motion of the agent is represented by Gaussian noises η_x and η_y. Also, when an agent performs a turn, the centripetal force required to maintain its circular motion is provided by the water, but the fluidity of water causes it to yield to the agent during turning. This provides insufficient centripetal force for an exact turn, which causes a drift. The sideways drift of the agent during turning is modeled as

$v^{\perp}$, where the direction is perpendicular to the heading of the agent, and the value is a function of measured forward speed (v) and turning rate ($\dot{\theta}$) (Eq. 4.15). Δt is time duration between the current state and the next state, i.e. ($Time_{t+1}$ - $Time_t$). In Eq. 4.15, an empirical constant k_{water} characterizes the inability of water to provide sufficient reactive centripetal force during turns, and ΔT is another empirical constant that defines the discretized time period corresponding to the drifting instance. Note that $f(v, \dot{\theta})$ is a rough approximation of $v^{\perp}$; the precise model is beyond the scope of this work.

$$\begin{bmatrix} Px_{t+1} \\ Py_{t+1} \end{bmatrix} = \begin{bmatrix} Px_t \\ Py_t \end{bmatrix} + \begin{bmatrix} \eta_x \\ \eta_y \end{bmatrix} + \left(\begin{bmatrix} \cos(\theta) \\ \sin(\theta) \end{bmatrix} v + \begin{bmatrix} \cos(\theta - 90 \\ \sin(\theta - 90 \end{bmatrix} v^{\perp} \right) \cdot \Delta t \tag{4.14}$$

$$v^{\perp} = f(v, \dot{\theta}) = k_{water} \cdot v \cdot \dot{\theta} \cdot \Delta T \tag{4.15}$$

4.4.3 Noise Model

There are two critical measurements within the system: forward speed (v) and heading (θ). The noise model of these measurements directly affects the performance of the filter.

Heading error can primarily be characterized by Gaussian noise, denoted as η_θ. However, when the tagged agent is close to the edge of the water, the heading error tends to be dominated by some unknown bias, say ϵ_θ, which is likely caused by a locally distorted magnetic field near the concrete wall. Note that a **random error** tends to continuously change value, while a **bias** maintains its value, once selected.

Similarly, the error in a speed measurement is also modeled with a combination of Gaussian noise (η_v) and bias (ϵ_v). The bias is incorporated to capture the influence of local flow (current or turbulence) on speed measurements. As will be discussed in the following subsection, biases will be assigned to each particle once the cluster is resampled.

4.5 Particle Filter (PF) based Fusion of Cameras and Biologging Tags

4.5.1 Preliminaries

One of the major challenges this work addresses is the association between a tracked agent and its corresponding detection among all detections (of all agents). From a localization perspective, the situation differs from a traditional filtering problem in that the observations about the environment (i.e. all detections of all agents, denoted as z_t for time t) do not uniquely correspond to the agent

Procedure 4.1 $(\boldsymbol{S}_{all}, \boldsymbol{X}_{ft}, \boldsymbol{t}_{rs}) \leftarrow \text{Propagate}(\boldsymbol{Y}_{in}, \boldsymbol{Y}_{ex}, w_{ts})$

input: $\boldsymbol{Y}_{in}$, internal measurements
$\boldsymbol{Y}_{ex}$, external measurements
w_{ts}, weight difference threshold
output: $\boldsymbol{S}_{all}$, states, weights and biases of all particles over all times
$\boldsymbol{X}_{ft}$, filtered states of tracked agent
$\boldsymbol{t}_{rs}$, times of resampling instances
$\boldsymbol{S}_{all}^{(t=0)} \leftarrow$ Initialize-Particles(); $\boldsymbol{t}_{rs} \leftarrow \emptyset$
for $t = 1, \cdots, T$ **do**
  $\boldsymbol{S}_{all}^{(t)} \leftarrow$ Predict-Particle-States$(\boldsymbol{S}_{all}^{(t-1)}, \boldsymbol{Y}_{in}^{(t-1)})$
  $\boldsymbol{S}_{all}^{(t)} \leftarrow$ Update-Particle-Weights$(\boldsymbol{S}_{all}^{(t-1)}, \boldsymbol{Y}_{ex}^{(t-1)})$
  $\boldsymbol{X}_{ft}^{(t)} \leftarrow$ Compute-Weighted-Sum$(\boldsymbol{S}_{all}^{(t)})$
  $w_{diff} \leftarrow$ Compute-Weight-Difference$(\boldsymbol{S}_{all}^{(t)})$
  if $w_{diff} > w_{ts}$ **then**
    $\boldsymbol{S}_{all}^{(t)} \leftarrow$ Resample$(\boldsymbol{S}_{all}^{(t)})$
    $\boldsymbol{S}_{all}^{(t)} \leftarrow$ Reset-Bias-and-Weight$(\boldsymbol{S}_{all}^{(t)})$
    $\boldsymbol{t}_{rs} \leftarrow$ Add-New(t)
  end if
end for
return $(\boldsymbol{S}_{all}, \boldsymbol{X}_{ft}, \boldsymbol{t}_{rs})$

of interest. Because the association between an agent and its own detection is unknown, it is not certain whether an observation ($\boldsymbol{z}_t$) contains the detection for the agent of interest. If the observation ($\boldsymbol{z}_t$) does not contain the detection for the agent of interest, then it is independent of the state (denoted as $\boldsymbol{x}_t$ for time t) of the agent of interest. Mathematically, it is difficult to have a closed form representation for the observation model $p(\boldsymbol{z}_t|\boldsymbol{x}_t)$, which is essential in many filtering schemes. This subsection introduces how a particle filter is embedded with a fuzzy observation concept and a non-causal state estimation method to handle the problem mentioned above.

A particle filter [148] is conceptually simple yet powerful. It is capable of handling nonlinear agent dynamics and non-Gaussian noise, at the cost of higher computation overhead. The probability distribution of the states of the system are approximated by a cluster of weighted particles. As the particles evolve forward, based on the provided dynamics model, the weights of the particles are updated according to external measurements (observations). When the weighting differences between particles surpasses a threshold, the entire cluster of particles will be resampled such that low weight particles are eliminated and high weight particles are duplicated. The states of the system are then estimated by taking the weighted sum of the states of all the particles. In this work, each tracked agent is considered a dynamical system with their horizontal positions serving as their

Procedure 4.2 $\boldsymbol{X}_{nc} \leftarrow$ Non-Causal-Estimate$(\boldsymbol{S}_{all}, \boldsymbol{t}_{rs}, \tau)$

input: $\boldsymbol{S}_{all}$, states, weights and biases of all particles
over all times
$\boldsymbol{t}_{rs}$, times of when resampling instances
τ, percentage of particles to use
output: $\boldsymbol{X}_{nc}$, non-causal estimated states
for $t = 0, \cdots, T$ **do**
  find n s.t. $\boldsymbol{t}_{rs}^{(n-1)} \leq t < \boldsymbol{t}_{rs}^{(n)}$
  if n exist **then**
    $\boldsymbol{I} \leftarrow$ Find-Top-Weighted-Particles$(\boldsymbol{S}_{all}, \boldsymbol{t}_{rs}^{(n)}, \tau)$
  else
    $\boldsymbol{I} \leftarrow$ Find-Top-Weighted-Particles$(\boldsymbol{S}_{all}, t, \tau)$
  end if
  $\boldsymbol{X}_{nc}^{(t)} \leftarrow$ Compute-Weighted-Sum$(\boldsymbol{S}_{all}^{(t)}, \boldsymbol{I})$
end for
return $\boldsymbol{X}_{nc}$

states (i.e. Px_t and Py_t for time t).

Algorithm 4.1 presents the main component of the employed particle filter, with tag data and camera detections of all agents serving as inputs, with the estimated track of the tagged agent as the output. Function $Predict\text{-}Particle\text{-}States$ propagates the particles forward by incorporating the motion and noise models introduced previously. $Update\text{-}Particle\text{-}Weights$ is performed according to the proposed **Fuzzy Observations** concept, to be introduced later. The times when particles are resampled are also recorded (Algorithm 4.1 line 17), and output for use in Algorithm 4.2, which performs the non-causal state estimation process that improves estimation accuracy, also to be introduced later in the chapter.

4.5.2 Fuzzy Observations

To handle the detection uncertainties mentioned above, each detection is considered as a possible detection for the tagged agent, yet none of them are fully trusted (in terms of the association to the tagged agent). In other words, we are avoiding the decision of whether a detection association is completely true (1) or completely false (0). Instead, the trusts in the detection associations are treated as unknown values between true (1) and false (0), which limits the influence of detections on the particles' forward propagations. As the particle states evolve, employing tag measurements for propagation, each particle collects small rewards depending on its proximity to a detection. These reward values serve as the defining factors for selecting particles to resample: over time, the particles with a higher reward status are more likely to be resampled. In this way, the detections are affecting particles in a conservative and gradual manner. The main concept is that when the

number of particles is sufficiently large, there will always exist particles that closely represent the true state of the tagged agent. As a result, these particles will (statistically) collect more rewards, given the detectability of a tagged agent is comparable to any other agent. This assumption is supported by the experimental results presented in the next section. For mathematical preciseness, the weights of the particles are not to be associated with **probabilities** in this work. Instead, they are to be considered **confidence scores**, which are indicators of how many rewards the particles have collected.

Now, let $D_{(i)}^{(t)}$ denote the shortest (Euclidean) distance from particle i to any detection at time t. $W_{(i)}^{(t)}$ and $W_{(i)}^{(t+1)}$ are the confidence scores (weights) of the i-th particle at time t and $t+1$. Then the confidence score is updated as follows:

$$W_{(i)}^{(t+1)} = \begin{cases} W_{(i)}^{(t)}\left(1 + R^{+} \cdot \frac{D_{px} - D_{(i)}^{(t)}}{D_{px} + D_{(i)}^{(t)}}\right), & \text{if } D_{(i)}^{(t)} \leq D_{px} \\ W_{(i)}^{(t)}(1 - R^{-}), & \text{otherwise} \end{cases} \tag{4.16}$$

Where D_{px} is a predefined proximity radius, R^{+} is the positive reward constant for a particle if it is in proximity to a detection, while R^{-} is the negative reward constant. Note that both R^{+} and R^{-} are small in magnitude, such that the confidence scores of the particles are not significantly influenced by any single step observations.

4.5.3 Non-causal State Estimation

The core idea for the non-causal implementation is to determine the states of the agent by not only using information from the past and present, but also from the future. This is somewhat similar to the idea of smoothing schemes ([149, 150, 151]) in common particle filter practices. For an individual particle, it may not be clear at present whether it is carrying the 'more correct' state information, or whether its sensor measurement bias is correctly assigned, due to the low confidence placed in the external measurements. However, if this particle is receiving more rewards than other particles over time in the future, it is more likely that this particle is carrying the 'more correct' state information at present. The major difference between smoothing techniques and the proposed non-causal state estimation is that in smoothing, the operation is performed on the entire population of particles, and results in a smoother trajectory than one obtained using only a forward filtering process. In the proposed non-causal state estimation method, only a fraction of the particles are identified and used for state estimation, as it is understood that a good portion of particles are not carrying the 'correct' information.

Each time the particle cluster resamples itself, the time instance is recorded (Algorithm 4.1)

Table 4.1: Summary of detection and tracking performance.

Parameters	Deployment-1 A1-1	Deployment-1 A1-2	Deployment-2 A1	Deployment-2 A2	Overall
Duration [minute]	22.6	30.9	48.9	49.2	151.6
Converging Time [s]	32.5	10.5	14.1	22.9	19.0
Detectability – Man	70.4%	86.0%	100%	85.6%	88.1%
Detectability – CNN	44.6%	54.7%	50.2%	59.2%	53.2%
True Detect Ratio	17.5%	24.4%	18.1%	21.0%	20.2%
Reward Ratio – All	62.6%	58.7%	67.7%	67.6%	65.1%
Reward Ratio – NC	70.2%	63.5%	70.0%	71.5%	69.2%
Track Error – All Mean / (SD) [meter]	2.33 (1.49)	1.58 (1.15)	1.63 (1.15)	1.87 (1.27)	1.80 (1.23)
Track Error – NC Mean (SD) [meter]	1.67 (1.27)	1.20 (1.00)	1.40 (1.10)	1.41 (1.09)	1.40 (1.10)

and the top τ percent of highly weighted particles at that time are found (Algorithm 4.2). Those highly weighted particles are then used to estimate the states between the previous and current resampling instances (Algorithm 4.2). Specifically, only the particles that become highly weighted in the future will be used to estimate the current state, even though all particles contribute to propagating the state forward. The state estimation is more accurate using this method, while the particle cluster still maintains high diversity.

4.6 Results and Discussions

Two deployments were conducted for validation of the proposed framework. In the first deployment, one agent (named A1) was tagged and tracked while the total number of agents in the monitored area varied from 2 to 6. This deployment was separated into two sections that were evaluated individually. The second deployment involved 4 agents in the monitored area while two of them, denoted as A1 and A2, were tagged and tracked. Ground truth tracks of all tagged agents were made by experienced human trackers. With ground truth positions, the **true detections** of the tagged agents were found among all detections. Figure 4.4 visualizes 2D and 3D tracks of one tagged agent at different time horizons. Table 4.1 summarizes the performance of the system over all deployments from the following perspectives.

Converging Time is the time duration for particles to converge from random initial positions to the proximities of the tagged agent. Because the initial positions of tagged agents were assumed unknown, the states of the particles were initialized randomly across the entire monitored environment. Convergence speed depends on how distinct the motion of the tagged agent was compared

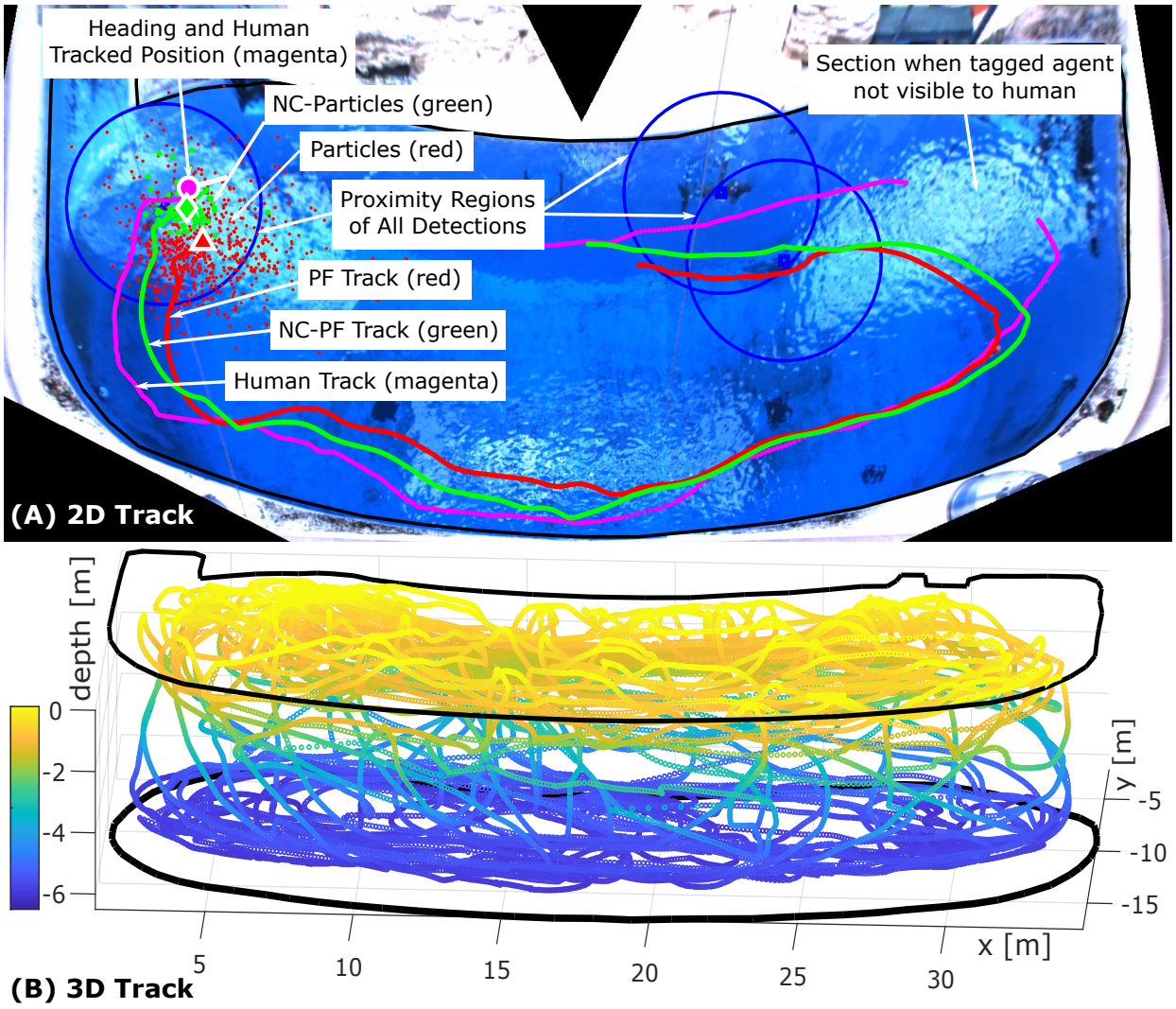


Figure 4.4: (A) Visualization of the tracking system with 30-second track plotted, including annotations of track components, where 'NC' stands for Non-Causal, 'PF' for Particle Filter. (B) Non-causally estimated 3D track of Agent 2 in second deployment, color-coded by depth.

to the rest of the agents, i.e. more distinct motion results in a faster convergence.

Detectability is the percent of total time the tagged agent was able to be detected by the human trackers or the CNN detectors over the corresponding deployment duration. Note that in these deployments, the tags were colored so that human trackers can distinguish them and generate ground truths for each tagged agent. In normal deployments, tag colors were similar to the agent, which makes the tags difficult to observe/distinguish for the human trackers. As in Table 4.1, the tagged agent was detected by the CNN detector ∼60% of the time when the tagged agent was detectable to the human tracker. The precision of the CNN detector can be increased at the cost of a lower recall, which is not always desired.

True Detect Ratio is the ratio between the total number of **true detections**, i.e. the detections of tagged agents, over all detections. This number is roughly equal to or slightly smaller than the ratio of the number of tagged agents over all agents. The smaller this number is, the harder it is for particles to collect rewards from **true detections**.

Reward Ratio records the ratio of particles' reward collected from **true detections** over all detections. This is an indicator of the influence of **true detections** on the particles. For example, if all particles are moving randomly, then the **Reward Ratio** would approach the **True Detect Ratio**. However, the **Reward Ratio** is significantly bigger than **True Detect Ratio**, which is an empirical proof that the particle filter with fuzzy observations automatically relies heavily on **true detections**, even though it does not know which detections are 'true'. In Table 4.1, *All* indicates all particles, and *NC* denotes the subset of particles identified using the non-causal approach. It is observed that the non-causal implementation relies even more heavily on **true detections**.

Track Error gives the mean and SD of tagged agents' position estimation errors w.r.t. to human tracks. The non-causal implementation improves tracking accuracy by 22% w.r.t. *All*, and the estimated position of the tagged agent is very close to the true value, given that the agent itself has a length of around 2.5 meters. Even though the mean error for the estimated track is relatively low, the estimated track does not converge to the exact track of the agent. This is expected, as the filter is designed to not trust any detection well enough to fully converge to it. Hence, the quality of the track can be further improved with two more steps. First, a heuristic tracklet selection method can be included to select the most likely tracklets that are associated with the estimated track of a tagged agent. (A tracklet is a sequence of detections that are associated to one agent with high probability, Section 4.3.3.) Then, the entire track, with unconnected sections of tracklets and continuous tag data, can be smoothed with an efficient optimization technique as discussed in our previous work [152]. Tracks with both high frequency and precision can be obtained as a result of this implementation. These two extra steps are not included in this dissertation to keep the presented framework concise but are under active development and testing. The tracking error of a tag-only dead-reckoning track is an order of magnitude higher than the methods discussed here

and thus not compared.

4.7 Summary

This chapter presents a framework that approaches the tracking problem from a localization perspective, utilizing a combination of biologging sensors and stationary cameras. Specifically, this framework deals with the situation where the agents are outfitted with biologging sensors and primarily subsist underwater, while the cameras are mounted above water. Our results demonstrate that the platform is readily implementable in managed environments, and is able to handle nonlinear agent dynamics, non-Gaussian noise, and scattered camera observations while maintaining robust localization and tracking ability. The method maintains the identity of the tracked agents comparing to camera-only tracking techniques, while it is significantly more accurate than biologging-only dead-reckoning tracks. In conclusion, this method can be applied to cover the general tracking scenario where sensor-augmented agents cannot be continuously monitored. As will be presented in Chapter 6, the identity preserving multi-target tracking ability enabled by this work shall be used to study the coordinated behavior of bottlenose dolphins.

CHAPTER 5

Data-Driven Modeling of Dolphin Dynamics: Hydrodynamic Drag, Transport Cost, and Gait

5.1 Background

5.1.1 Gait and Swimming Efficiency

Animal behavior is related to and constrained by its dynamics, environment, and energetic cost. Being able to model and characterize individual dynamics provides insight and improved understanding of observed behavior. In nature, efficient locomotion is essential for animals that travel long distances, and various strategies, both morphological or behavioral, have been observed to this end [76, 77, 78]. Dolphins, in particular, are well equipped to be excellent swimmers. A muscular streamlined body shape reduces hydrodynamic drag and resembles engineered 'laminar' profiles [79]. Dolphin skin includes various features to further increase the animal's hydrodynamic performance underwater [80]. Dolphins also regulated their behavior to improve locomotion efficiencies. For example, they have been observed leaping out of the water to reduce drag and save energy at high speed [81], or riding on the waves generated by boats [82]. Young dolphin calves have been observed 'drafting' next to their mothers to keep up with a rapidly moving group of animals [83].

Different gaits can also be used by an animal to conserve energy, without the need for enabling external conditions (e.g. being near to the surface, having a ship passing by, or being close to

Content of this chapter is mainly associated with the following publications:

[153] D. Zhang, J. M. van der Hoop, V. Petrov, J. Rocho-Levine, M. J. Moore, and K. A. Shorter*, "Simulated and experimental estimates of hydrodynamic drag from bio-logging tags," *Marine Mammal Science*, vol. 36, no. 1, pp. 136–157, 2020.

[154] D. Zhang*, Y. Wang, L. K. Lauderdale, J. Gabaldon, L. J. Miller, K. Barton, and K. A. Shorter, "Energetic benefits of fluke and glide gait for Bottlenose dolphins," in *Preparation for Submission*, 2021.

* Corresponding Author

their mothers). Intermittent locomotion [84], where an animal employs a gait composed of periods of active flapping/stroking/bursting and inactive gliding/coasting, is a convergent evolutionary trend in both aerial and marine animals [78] that has been observed with birds [85, 86] and fishes [87, 88, 89] during their daily activities. While the energetic benefits of intermittent locomotion have been studied for certain fishes from various perspectives [87, 88, 89, 90, 91, 92, 93, 94], little work has been done on dolphins. Because dolphins are excellent swimmers, and burst-and-coast swimming is not always energetically beneficial in fish [94], it is important to investigate the energetic trade-offs of intermittent locomotion for these animals and identify when they choose to use such gaits. For cetaceans, published work has tended to focus on how animals extend dive duration and conserve energy by taking advantage of buoyancy changes while gliding during descent or ascent in a dive [95, 96, 42]. However, a model based investigation of the energetic benefits of intermittent gait during horizontal locomotion for transit, specifically a Fluke-and-Glide (FG) gait, is lacking for these animals. Further, how animals regulate their gait in the presence of other animals or features in the environment is also an open question.

The evaluation of energetic cost requires an accurate estimate of the hydrodynamic drag experienced by the animal during all stages of swimming. Various methods have been used to obtain such estimations for marine animals, including theoretical computations [79, 90, 91], computational fluid dynamic simulations [98, 97], towing experiments [99], camera assisted experiments [76, 42, 60, 100, 101], and field biologging tag deployments where speed can be derived from the data to estimate drag [43]. Glide deceleration [155, 156] has been widely used to estimate passive drag coefficients from free-swimming animals. During a glide of a neutrally buoyant animal, a net drag force opposes the movement of the body, slowing it down. A measured rate of deceleration can then be used to calculate drag force [155, 156]. Along with glide deceleration, wind tunnels with physical models [157, 158] and tow experiments with live animals in a rigid body position [76] have been used to measure static/passive drag coefficients. However, estimates of gliding drag over a range of speeds for bottlenose dolphins are currently lacking.

5.1.2 Tag Induced Drag

Data from bio-logging tags [43, 159, 160], along with estimated body parameters, have been used to estimate drag forces on free-swimming animals. While methods using glide data are only useful for models that describe or quantify gliding behaviors, they still provide a baseline for relative comparison of different drag conditions, e.g. with instrumentation [42] or during and after pregnancy [100]. However, measurements made from animal-borne sensors are often required for these estimates, leaving the additional contribution created by the tag difficult to discern. The hydrodynamic impact of biologging tags on the the streamlined bottlenose dolphins remains to be assessed.

In the marine environment, placing tags on a highly streamlined body has the potential to affect swimming mechanics by increasing drag on the body. During swimming, the thrust output required to overcome drag is central to energetic expenditure; quantifying forces acting on the body when swimming with tags is essential to determine the potential for and magnitude of tag effects on energy consumption or swimming behavior, e.g. [97, 98]. Direct measurement of propulsive forces during swimming is difficult, but estimates of drag loading and direct measurement of the animal's specific acceleration can be used to estimate propulsive force [160, 161].

To complement measurements of static/passive drag coefficients, researchers have performed controlled experiments with managed animals trained to execute prescribed swimming tasks to estimate drag forces and coefficients [100]. The use of a controlled experimental environment allows data collection during un-instrumented trials, enabling the quantification of a tag's contribution to the net drag acting on the combined system. In contrast with the experimental methods described above, CFD models are becoming an increasingly popular method to simulate the expected effects of instrumentation and to refine tag design [97, 98, 162, 163, 164, 165]. The same technique has also been used to aid and optimize the design of underwater vehicles [166, 167, 168]. Working in a simulated environment enables a rapid evaluation of a range of tag designs in multiple fluid-flow conditions, but these simulations can be difficult to compare with experimental results.

5.1.3 Research Contributions

This chapter address the following research question: (**Q3**) *What are the energetic trade-offs of different swimming gaits measured using a biologging tag?*

We first propose a model of the fluke-and-glide gait of the animal and investigate how the dynamic characteristics could be used to estimate passive drag. Then, a data-driven method is introduced to estimate passive drag using measurements from 84 hours of biologging tag data from 3 animals. These results are used to investigate the energetic cost of the animal using gait parameters extracted from the biologging. Finally, the second set of experiments and CFD simulations are presented to assess the hydrodynamic impact of biologging tags on bottlenose dolphins.

5.2 Swimming Biomechanics

5.2.1 The Fluke-and-Glide (FG) Gait and its Dynamics

Similar to the flapping and gliding model for birds [169], dolphin FG gait is modeled as a two-phase process in this work (Figure 5.1-B): (1) an accelerating fluking phase with duration T_f and (2) a decelerating gliding phase with duration T_g. The maximum and minimum speeds in this process

are denoted by v_{max} and v_{min}, with an average speed during a bout of $v_{mid} = (v_{max} + v_{min})/2$. We further simplify the analysis by assuming that the animal is neutrally buoyant and swimming straight (Figure 5.1-B). With these assumptions, the thrust force during fluking at time instance t can be related to the hydrodynamic drag force $F_{drag}(t)$ and animal's acceleration $a(t)$ by:

$$F_{thrust}(t) = F_{drag}(t) + m_b a(t), \tag{5.1}$$

where m_b is the induced virtual mass of the animal [170]. The drag force $F_{drag}(t)$ opposes the direction of motion and is modeled in a conventional way [101] by:

$$F_{drag}(t) = \frac{1}{2}\rho v_{(t)}^2 A_w C_d, \tag{5.2}$$

where ρ is the density of sea water, $v(t)$ is forward speed with respect to water at t, A_w is the wetted surface area of the animal [171], and C_d is the dimensionless drag coefficient for the animal.

The drag coefficient during fluking is denoted as $C_d^{(f)}$ and during gliding as $C_d^{(g)}$. These two parameters are critical for the estimation of drag during FG gait. For this work, we use the $C_d^{(f)}$ estimated for bottlenose dolphins by [101], and estimate $C_d^{(g)}$ using biologging tag data.

During a glide, there is no thrust force and Eq. 5.1 reduces to:

$$m_b a(t) = m_b \frac{dv_{(t)}}{dt} = -F_{drag}(t). \tag{5.3}$$

Combining Equations 5.2 and 5.3, we obtain the differential equation that describes how speed $v_{(t)}$ changes during a glide as:

$$\frac{dv_{(t)}}{dt} = -\frac{\rho v_{(t)}^2 A_w C_d^{(g)}}{2m_b}, \tag{5.4}$$

which can be used to solve for speed $v_{(t)}$ as a function of time t and initial speed $v_{(0)}$:

$$v_{(t)} = \frac{1}{\frac{\rho A_w C_d^{(g)}}{2m_b} t + \frac{1}{v_{(0)}}}. \tag{5.5}$$

Equation 5.5 was then used with *measured* speed data during periods of gliding to identify the drag coefficient $C_d^{(g)}$, as in Section 5.3.3. Drag coefficients were related to Reynolds number (Re) by $C_d = bRe^c$ [101], which is a linear relationship in the log scale. The Reynolds number for each

gliding event was calculated as:

$$Re = \frac{Lv_{(0)}}{\nu}, \tag{5.6}$$

where L is the body length of the animal, $v_{(0)}$ is the identified initial speed of the glide, and ν is the kinematic viscosity of seawater (1.05×10^{-6} m^2/s).

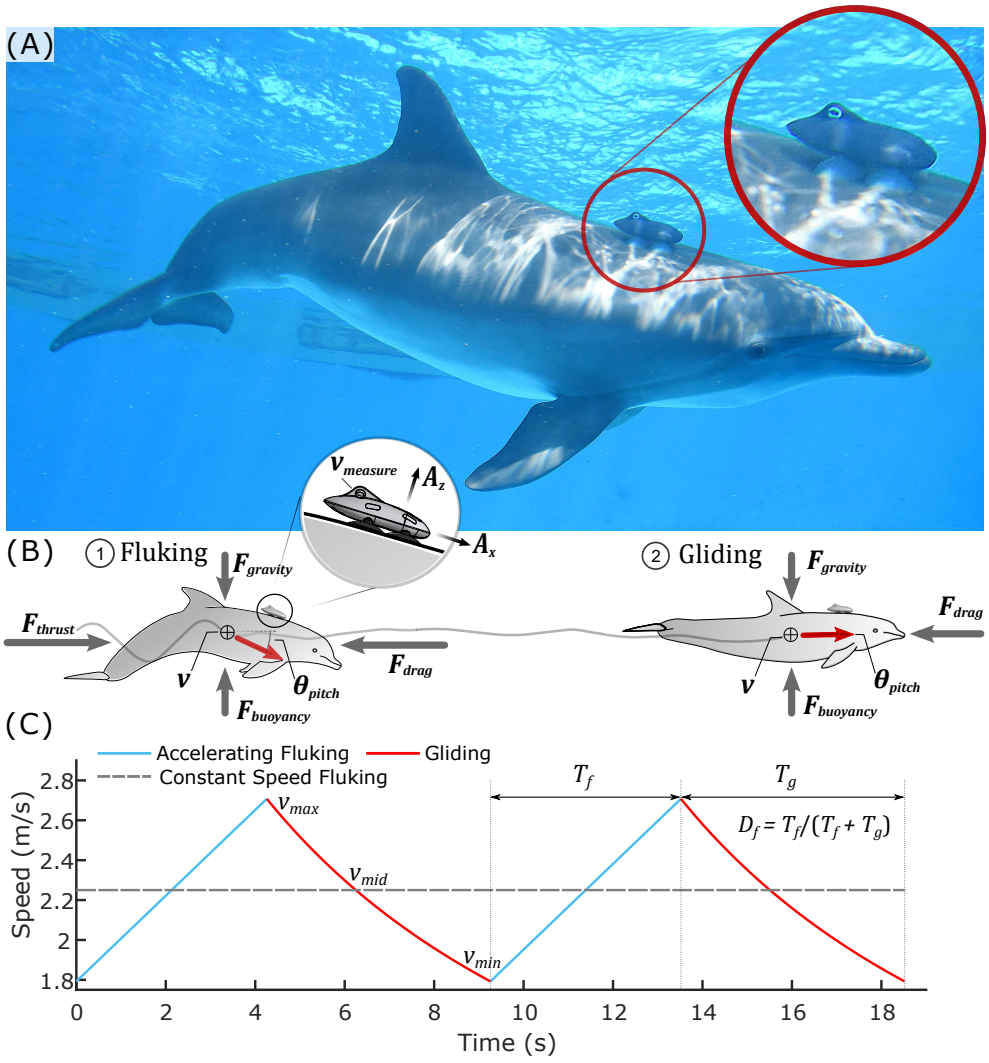


Figure 5.1: (**A**): A bottlenose dolphin with a biologging tag. (**B**): Simplified free-body diagrams of an animal fluking and then gliding. (**C**): Simplified speed model of FG gait and continuous fluking gait over two cycles.

5.2.2 Kinetics of the Fluke-and-Glide (FG) Gait

The speed profile of the proposed FG gait model (Figure 5.1-C) can be uniquely determined given an average bout speed v_{mid} and a gliding duration T_g. The starting and ending speeds of the glide (v_{max} and v_{min}) can be derived from Eq. 5.5 by plugging in:

$$v_{(0)} = v_{max}, \tag{5.7}$$

$$v_{(t)} = v_{min}, \tag{5.8}$$

$$v_{mid} = (v_{max} + v_{min})/2, \tag{5.9}$$

$$t = T_g. \tag{5.10}$$

Resulting with:

$$v_{max} = \frac{\Psi v_{mid} T_g - 1 + \sqrt{\Psi^2 v_{mid}^2 T_g^2 + 1}}{\Psi T_g}, \tag{5.11}$$

$$v_{min} = 2v_{mid} - v_{max}, \tag{5.12}$$

where

$$\Psi = \frac{\rho A_w C_d^{(g)}}{2m_b} \tag{5.13}$$

is the coefficient of t in Eq. 5.5. With the dynamics for gliding phase resolved, the fluking duration T_f and acceleration $a(t)$ can be obtained as:

$$T_f = \frac{D_f}{1 - D_f} T_g, \tag{5.14}$$

$$a(t) = (v_{max} - v_{min})/T_f. \tag{5.15}$$

The entire speed profile of a FG gait cycle can be thus represented as:

$$v_{(t)} = \begin{cases} v_{min} + a(t)t & 0 < t \leq T_f \\ \dfrac{v_{max}}{\Psi v_{max}(t - T_f) + 1} & T_f < t \leq T_f + T_g \end{cases}$$

with the two conditions corresponding to speed of fluking and gliding respectively.

5.2.3 Mechanical Energy Cost of Transport (MECOT)

In this work we use Mechanical Energy Cost of Transport (MECOT) to quantify the swimming effort of an animal. MECOT, to be distinguished from metabolic energy cost of transport [172, 173], for a FG gait is the amount of mechanical energy (in joules) required to transport one unit of body weight (1 kilogram) over one unit of distance (1 meter):

$$MECOT = \frac{\int_0^{T_f} F_{thrust}(t)v(t)dt}{m\int_0^{T_f+T_g} v(t)dt}, \tag{5.16}$$

where m is the body mass of the animal and $F_{thrust}(t)$ can be obtained from Eq. 5.1 together with the speed profile $v(t)$ of the proposed FG model (Figure 5.1-C, Section 5.2.2). The FG model and MECOT were used to explore how changes in swimming parameters affected the cost of transport. For this, the reference swimming speed v_{mid} was held constant, while gliding duration T_g and duty factor D_f were varied and the corresponding MECOT were computed. In addition to the FG gait, Eq. 5.16 characterizes a continuous constant speed fluking gait when $T_g = 0$ and $v_{(t)} = v_{mid}$. For the analysis, values for T_g and D_f were identified using biologging tag data collected from the animal (Section 5.3.2).

5.3 Data-Driven Gait Parametrization using Biologging Tags

5.3.1 Data collection

Three bottlenose dolphins (*Tursiops truncatus*) from the Brookfield Zoo, Chicago Zoological Society, participated in this study: age (38, 16 and 5 years), mass (202, 153 and 150 kg), body length (266, 242 and 226 cm) and girth (139, 128 and 127 cm). A biologging tag was attached to the animal 20 cm behind the blowhole via four silicone suction cups (Figure 5.1-A) by an animal care specialist. The MTag (Section 2.2) was specifically used for this study. Sensors on board include 3-axis accelerometer, gyroscope, magnetometer (50 Hz), as well as a pressure sensor, a temperature sensor and a speed sensor (10 Hz). The speed sensor measures relative fluid speed by counting the revolving frequency of a micro turbine on top of the tag, which is then mapped to the animal's forward speed in water [107]. The animal swam freely in the main aquarium habitat (33.5 m (length) by 12.2 m (width) by 6.7 m (depth)) during the deployments. The tags were deployed for an average of 2 hours at a time (84 hrs total).

5.3.2 Glide Extraction and Fluke-and-Glide (FG) Bout Identification

Calibrated accelerometer, gyroscope, and magnetometer data were used to estimate orientation (pitch, roll, and yaw) of the animal using a gradient-descent based optimization method in the quaternion space [105]. When the animals were moving (forward speed > 1.0 m/s), three key criteria were used to identify gliding events for analysis. First, the glide had to occur at a depth of greater than 2.5 body diameters to avoid surface drag [79], for all animals a depth greater than 3.5 m was used to satisfy this condition. Second, the animal was not actively swimming or changing direction during the gliding event. The postural configuration and the use of control surfaces (fins and flukes) for active maneuvering would affect drag acting on the body. We required the 1-norm of jerk (i.e. the derivative of the 3-axis measurements from the accelerometer) $||\text{jerk}||_1$ to be less than 0.5 g/s to exclude active swimming, and |pitch rate| < 20 deg/s, |yaw rate| < 22 deg/s to remove active turning. Third, to mitigate the effects of gravity and buoyancy, gliding events for analysis were restricted to constant depth (|depth rate| < 0.2 m/s and |pitch| < 17 deg). When the animal was moving at a constant depth it was assumed that buoyancy and gravitational force were equal and opposite, and did not affect the animal's forward motion. A decision tree was employed to detect *gliding points* in the data that satisfied all of the above conditions:

1) forward speed > 1.0 m/s

2) depth > 3.5 m

3) |pitch rate| < 20 deg/s

4) |yaw rate| < 22 deg/s

5) $||\text{jerk}||_1 < 0.5$ g/s

6) |depth rate| < 0.2 m/s

7) |pitch angle| < 17 deg

Temporally adjacent *gliding points* were grouped together to form a *gliding event*. Speed measurements during a gliding event were then used with Eq. 5.5 to estimate the gliding drag coefficient $C_d^{(g)}$, using a dual parameter sweep fitting approach presented in the next section (Section 5.3.3).

To characterize FG gait during a range of behaviors, we relaxed the restrictions on depth, pitch and yaw rate to include more FG bouts under different conditions. Manually selected data segments composed of accelerating fluking followed by a period of gliding were included for the analysis. The start of an event was identified using the fluking signature in the data, and the end point during the glide was selected to equal the segment's starting speed (with difference < 0.1 m/s).

5.3.3 Drag Estimation with Parameter Optimization

A dual parameter sweep was performed to identify the initial speed (v_0) and the coefficient of drag (C_d) for the model (Eq. 5.5) that best matched the measured speed profile. v_0 was considered unknown because it was a critical parameter for the curve-fitting process, and the noise in the extracted speed data made it difficult to accurately estimate v_0 directly. For the sweep, the initial speed (v_0) took values in the range $[\max(\hat{v}_0 - 1, 0), \hat{v}_0 + 1]$ with an interval step of 0.01 (m/s), with $\hat{v}_0$ being the **measured** initial speed. The candidates for the coefficient of drag (C_d) ranged from 0 to 0.3 with an interval step of 0.0005.

These parameter ranges were based on published values in the literature [174, 100, 101]. Each combination of v_0 and C_d candidates yields a predicted speed curve according to Eq. 5.5; the fit quality of the curve was quantified using the RMS (Root Mean Square) value of the residual, R, between measured speed and predicted speed. A surface, called the R-surface (Figure 5.2-Top), in 3D space was then formed by combining all pairs of $\{v_0, C_d\}$, and the resulting R. Given that a smaller R value indicates a better fit, the values of v_0 and C_d located at the global minima of the R-surface were selected for the final models (Figure 5.2-Middle).

5.3.4 Speed Distribution and Normalization

To assess the statistics of the FG gait in the data, the speed distributions under different conditions were calculated using steps described in this section. Let $P_{all}(v)$ be the probability of the animal's general locomotion speed to lie within the left-closed interval [v - 0.25, v + 0.25) (m/s) during all locomotion, then:

$$P_{all}(v) = \frac{C_{all}(v)}{\sum_v C_{all}(v)}, \tag{5.17}$$

where $C_{all}(v)$ is the counted number of occurrences within the interval associated with v during all locomotion. Similarly, we define the probability of an animal's speed during FG gait:

$$P_{fg}(v) = \frac{C_{fg}(v)}{\sum_v C_{fg}(v)}, \tag{5.18}$$

where $C_{fg}(v)$ is the counted number of occurrences for speed within the interval of v during FG gait. The normalized count for FG gait speed is then defined as:

$$C_{fg/all}(v) = \frac{C_{fg}(v)}{C_{all}(v)}, \tag{5.19}$$

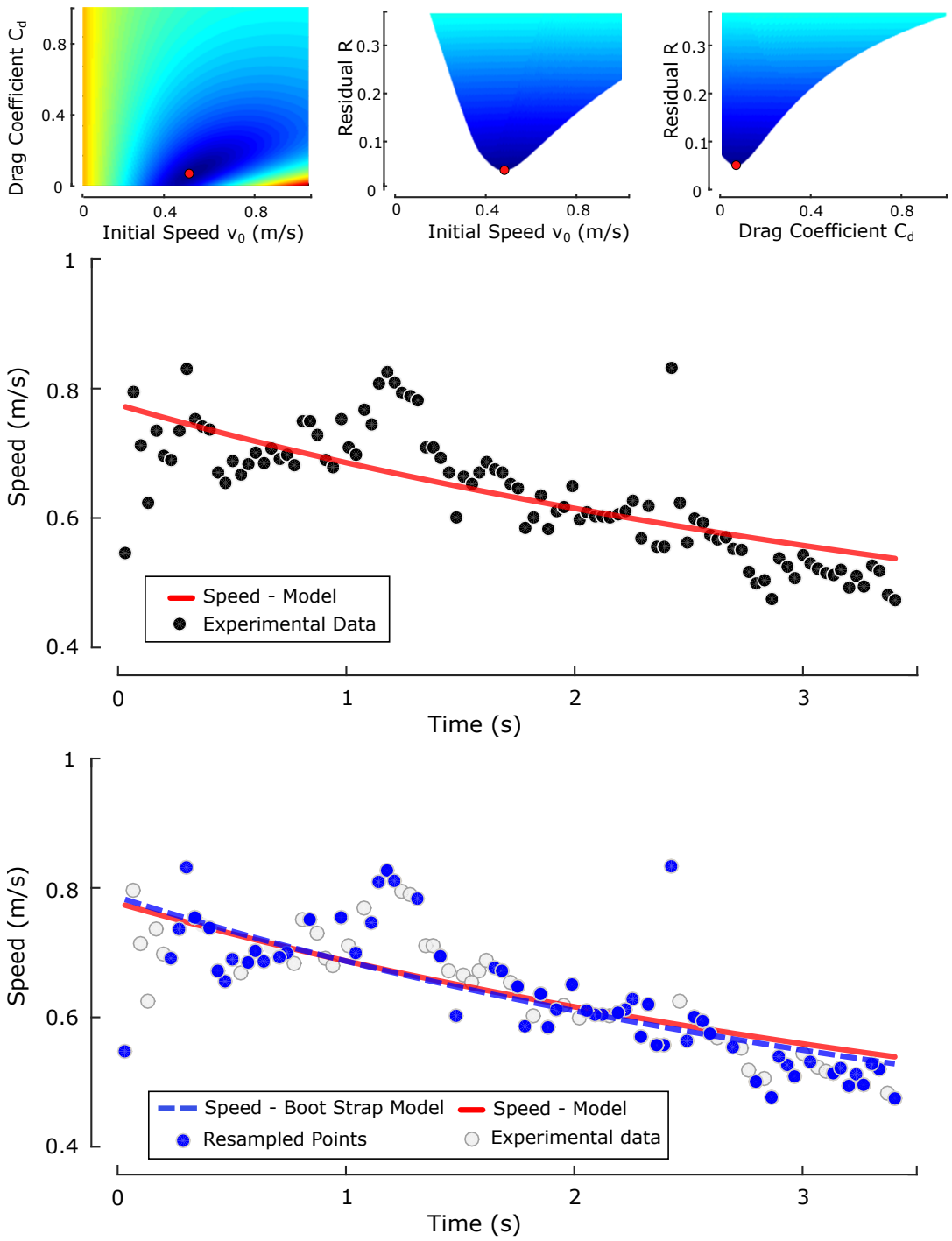


Figure 5.2: Representative data from a glide when the animal is wearing a tag. (**Top Left**) top view of the optimization surface used to select the model parameters (v_0 and C_d). Parameters that minimize fit residuals are indicated by the red circle. (**Top Middle**) RMS (root mean square) value of fitting residual versus initial speed v_0. (**Top Right**) RMS value of fitting residual versus drag coefficient C_d. (**Middle**) Measured (solid black) and the corresponding dynamic model prediction of the speed using the parameters selected using optimization. (**Bottom**) A representative iteration of the resampling and refitting to illustrate how bootstrapping (Section 5.4.5) was used to create a confidence interval for the drag coefficient.

which represents the animal's relative preference for using FG gait at speed v. And the associated probability for the normalized FG speed is:

$$P_{fg/all}(v) = \frac{C_{fg/all}(v)}{\sum_v C_{fg/all}(v)}. \tag{5.20}$$

$P_{fg/all}(v)$ represents the probability of a FG gait to have speed v *if* the animal has spent a uniform amount of time in its overall locomotion over all v choices. $P_{fg/all}(v)$ is thus a normalized indicator of the animal's preference for using a FG gait at speed v. The resulting $P_{all}(v)$, $P_{fg}(v)$ and $P_{fg/all}(v)$ over all choices of v from the data are later presented in Figure 5.7 (Section 5.5.2).

5.4 Hydrodynamic Impact of Biologging Tags: Experiments and Simulations

5.4.1 Data Collection

To experimentally investigate the effects of tag drag, we fabricated a modular set of elements to increase frontal area by up to eight times the tag alone (Figure 5.3). Two male bottlenose dolphins at Dolphin Quest Oahu ((Honolulu, HI, USA) were trained to glide with no locomotory movement on cue between animal care specialists standing or floating in water 8-12 m apart (Figure 5.3). One specialist provided the initial propulsive force before releasing the animal to glide to the second specialist. Dots of zinc oxide on the dolphin's skin were used to create kinematic markers visible in underwater video footage, and were used to calculate the position and speed of the animal during the trial. Glides were filmed perpendicular to the path of the animal at 30 fps with a Canon PowerShot G12 digital camera in a waterproof housing (Canon WP-DC34). The distance between the markers was measured following the application to each subject. The camera and housing were mounted on a monopod, which was rested on the bottom of the lagoon to reduce the camera's motion from water movement. The experimental lagoon was 1.5-2.4 m deep. The dolphins were asked to perform glides without the tag (*control*), wearing a tag (*tag*) and wearing the tag with either four (*tag+4*) or eight (*tag+8*) drag elements (Figure 5.3). Each drag element had the same cross-sectional area as the tag itself. The *tag+4* configuration increased the cross-sectional area by four while the *tag+8* configuration increased it by eight.

5.4.1.1 Identification of Successful Trials for Analysis

We used the following criteria to select glides for analysis: (1) no visible body movement, (2) no visible changes in the horizontal plane (i.e., the dolphin did not noticeably move closer to or away

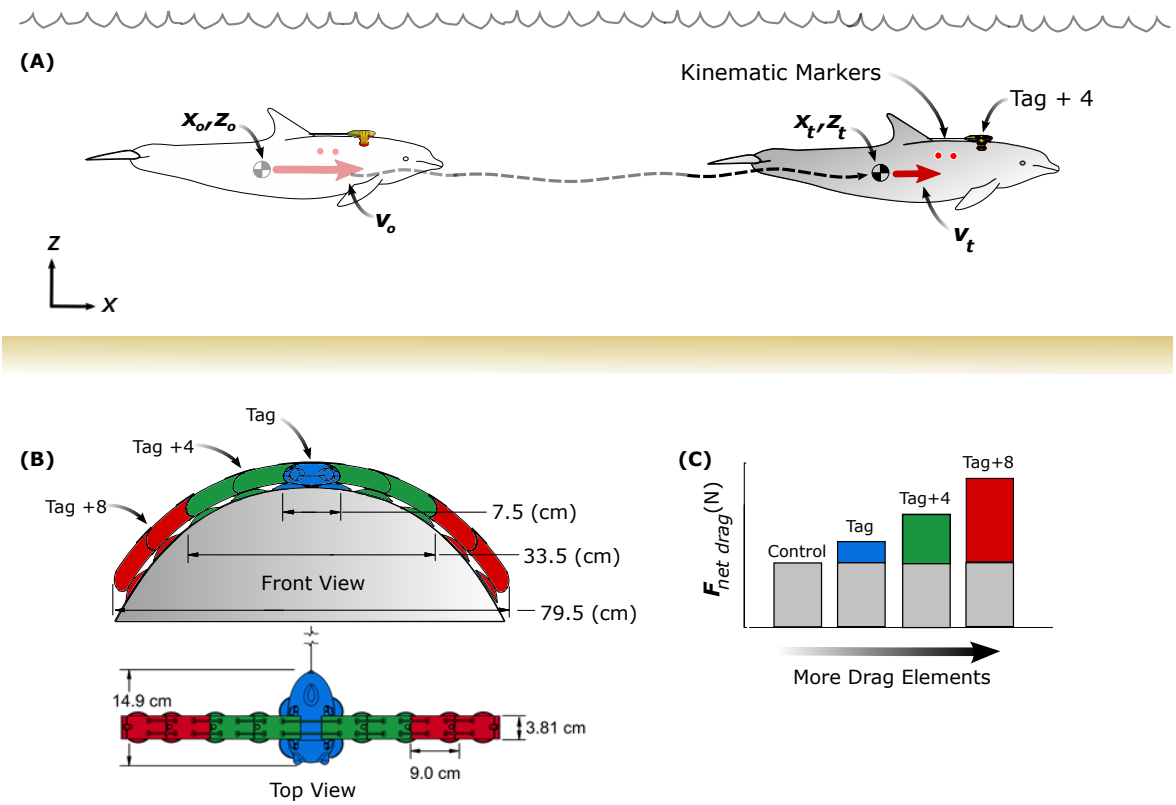


Figure 5.3: Illustration of the experiment (**A**) conducted to collect kinematic data during different drag loading conditions (**B** and **C**): control, tag, tag + 4 and tag + 8. (**A**): Zinc oxide applied to the dolphin's skin was used to create the kinematic markers. Planar video data of the animals during gliding were collected using an underwater camera. (**B**): The modular system including the tag and drag elements, built to increase drag loading on the animals. (**C**): Expected net drag in the different drag loading conditions.

from the camera), (3) minimal changes in depth, (4) 1 s minimum glide duration, and (5) a change in acceleration < 0 (i.e., there was a detectable decrease in speed) [175, 100]. Glide duration was calculated from the time the kinematic markers entered the frame to when the dolphins showed detectable body movement. We calculated success rates based on the number of glides the dolphins were asked to perform and the number of glides that were successful for each individual and each tag condition.

5.4.2 Video Processing and Data Extraction

All video processing was conducted in MATLAB and made use of the 'PointTracker' object from the 'Computer Vision System Toolbox'. An example video frame of a gliding dolphin with kinematic markers (zinc oxide) is shown in Figure 5.4. Videos were first preprocessed using the MATLAB 'Image Processing Toolbox' to enhance marker visibility. Each frame was converted into gray scale and then object edges were identified using the 'edge' function. Invalid small objects/edges were removed using the 'bwareaopen' function, and the 'imclose' function was used to close edges in each frame morphologically. Finally, the original frames were overplayed with the resulting binary frames to obtain the preprocessed video.

A semi-automatic tracking program was developed to extract kinematic marker positions from the enhanced video. Features around the kinematic markers were first identified in each frame (Figure 5.4-C). Marker position during the trial was then estimated by averaging tracked feature positions. In most cases, feature identification was automatic; however, when the number of tracked features per marker dropped below two, a human analyst was queried to identify marker position in the frame.

In this work, the known distance between the two kinematic markers was used to correct for lens distortion and to map dolphin speed from video to world frame. Figure 5.4-D shows an example of the pincushion distortion created by the camera lens/housing that was as large as five pixels at the frame edge. To correct the error, a third-order polynomial was fit to the distorted distance data (Figure 5.4-D, red line). The functional fit was used to undistort position data around the center of the frame. Speed was then calculated by numerically differentiating position data and mapped to the real world using the following relationship:

$$v = \frac{d_w}{d_v} v_v \tag{5.21}$$

Where v and v_v are speeds of dolphin in real world and in the video, d_w and d_v are distance between kinematic markers in real world and in video respectively.

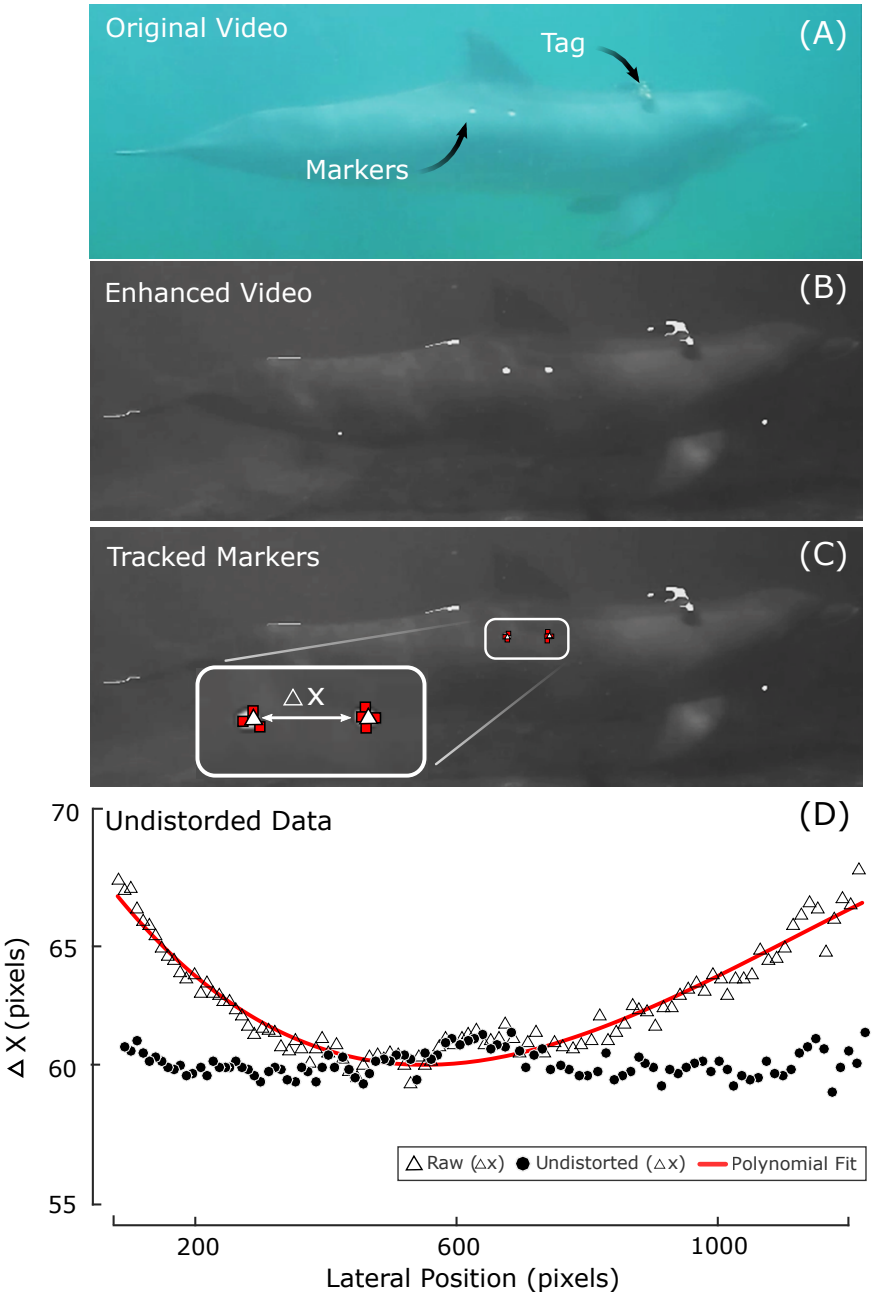


Figure 5.4: (**A**) Example video frame of a gliding dolphin with kinematic markers (zinc oxide) under the dorsal fin. (**B**) Pre-processing was used to enhance the visibility of the markers for automatic tracking of the data. (**C**) Tracked kinematic markers with the identified features are visualized with red circles. The average feature position is shown as a black triangle. The distance (Δx) between the two fixed kinematic markers as the dolphin glides through the frame was calculated for calibration and undistortion. (**D**) The horizontal position of animal in the video frame is shown on the x-axis. Pincushion distortion created by the lens results in as much as an 8-mm change in distance between the markers at the edges of the frame.

5.4.3 Parameter Synthesis and Compensation

In addition to the parameter optimization of individual gliding events (Section 5.3.3), speed profiles were pooled to estimate the collective drag coefficient when multiple qualified trials (each indexed by k) were obtained for an experimental condition. A similar optimization approach was used to identify the best fit from N coefficient candidates. The goodness of each drag coefficient candidate $C_d^{(n)}$ was evaluated individually. The constant $C_d^{(n)}$ was applied for the k-th trial, together with a varying initial speed (v_0), to find the corresponding lowest cost $R_{(k)}^{(n)}$ for the k-th trial. In other words, $R_{(k)}^{(n)}$ is the smallest residual that can be obtained by varying only initial speed (v_0) for the k-th trial and n-th drag coefficient. The overall cost of all K trials corresponding to the n-th drag coefficient $C_d^{(n)}$ is defined as:

$$R_{all}^{(n)} = \sum_{k=1}^{K} R_{(k)}^{(n)}, \tag{5.22}$$

where small values of $R_{all}^{(n)}$ indicate values of $C_d^{(n)}$ that result in a good fit. The $C_d^{(n)}$ with the lowest residual, among all N drag coefficient candidates (i.e. $C_d^{(1)} \cdots C_d^{(N)}$), was then selected for the model. The resulting drag coefficient is denoted as C_{d_all}:

$$C_{d_all} = C_d^{(n)}, \quad with \quad n = \underset{n\in[1,N]}{\arg\min}\, R_{all}^{(n)} \tag{5.23}$$

The average drag force acting on the animal could then be calculated using Eq. 5.2 by plugging the average speed during a particular trial in as $v(t)$. The percent increase in drag coefficient between the tagged conditions and the nontagged animal was calculated for both the experimental and simulation results:

$$\frac{C_d^{(tagged)} - C_d^{(no_tag)}}{C_d^{(no_tag)}} \times 100. \tag{5.24}$$

This provides a reference of the increase in a nondimensional quantity for comparison of tag effects [158]. We also calculated the Reynolds number (Re, Eq. 5.6) to compare drag coefficients within this study and to other values in the literature.

For studying the hydrodynamic impact of the tag, all experimental data were collected within 2 to 2.6 body diameters of the surface, and are subject to additional drag force created by dynamic interactions with the surface [79]. The average depth of each animal over all the experimental trials was used to determine the value of γ based on [79]. We use $C_{d_correct}$ to denote the drag coefficient that has been corrected for additional surface drag, and used it to compare with results from CFD and the literature:

$$C_{d_correct} = C_d/\gamma. \tag{5.25}$$

5.4.4 Computational Fluid Dynamics (CFD)

Building on previous work [165], CFD simulations were performed using CAD (computer-aided design) models of the tag and drag elements with representative animal geometry. Autodesk Inventor 2013 (Autodesk Inc., San Rafael, CA, USA) was used to model the tag geometry, and dolphin geometry was scaled to match the average length of the animals in the experimental trials (2.48 m) at Dolphin Quest Oahu (Honolulu, HI, USA).

Simulations were performed using STAR-CCM+ (version 9.04; CD-adapco, Melville, NY, USA/London, England). This commercial code (STAR-CCM+ 2014) solves the transport equations for continuity and 3D momentum on a very fine 3D mesh. We modeled turbulence with the two-layer Reynolds-Averaged Navier-Stokes (RANS) approach for the solution of the k-epsilon transport equations (STAR-CCM+ 2014; [176]). Mesh sensitivity tests were performed in 2D with a dolphin middle cross-section profile to estimate required cell size. We tested three different mesh sizes (coarse = 0.05 m, medium = 0.025 m, and fine = 0.01 m) with 0.365 million, 0.998 million, and 6.50 million cells, respectively. Variation in drag and lift forces from medium to fine mesh was 1.85%. The 3D simulations were performed with a trimmed cell mesh (0.25 m cell base size) with 50-57 million cells; the number of cells depends on the tag configuration with more drag elements requiring more cells. The 3D computation domain has a cylindrical shape with a length of 10.4 m and diameter of 6 m (Figure 5.5). The modeled dolphin length is 2.6 m with a body diameter approximately 0.2 m at the dorsal fin. During all simulations of the combined dolphin-tag model, the tag was located 0.74 m from the rostrum midway between the blow hole and dorsal fin. Simulations were constructed with a uniform velocity profile over a range of speeds (1-6 m/s). Maximum collection blockage area is below 0.6% with the object to the domain volume ratio of 0.4%. Boundary conditions were applied as: a) uniform velocity at the inlet; b) outflow boundary (zero normal gradients) at the outlet; c) slip wall around the domain.

For all simulations, a polynomial function was used to interpolate forces at flow speeds between simulated points. Drag coefficients (C_{d_sim}) were calculated for each condition using:

$$C_{d_sim} = \frac{\sum_f (F_f^{pressure} + F_f^{shear}) \cdot u_D}{\frac{1}{2}\rho_{ref} v_{ref}^2 A_{ref}}, \tag{5.26}$$

where $F_f^{pressure}$ and F_f^{shear} are the pressure and shear force vectors on the surface face f and ρ_{ref}, v_{ref}, and A_{ref} are the reference density, velocity, and area, respectively. u_D is a user-specified direction vector. The pressure force vector on the surface was computed as:

$$F_f^{pressure} = (p_f - p_{ref}) A_f, \tag{5.27}$$

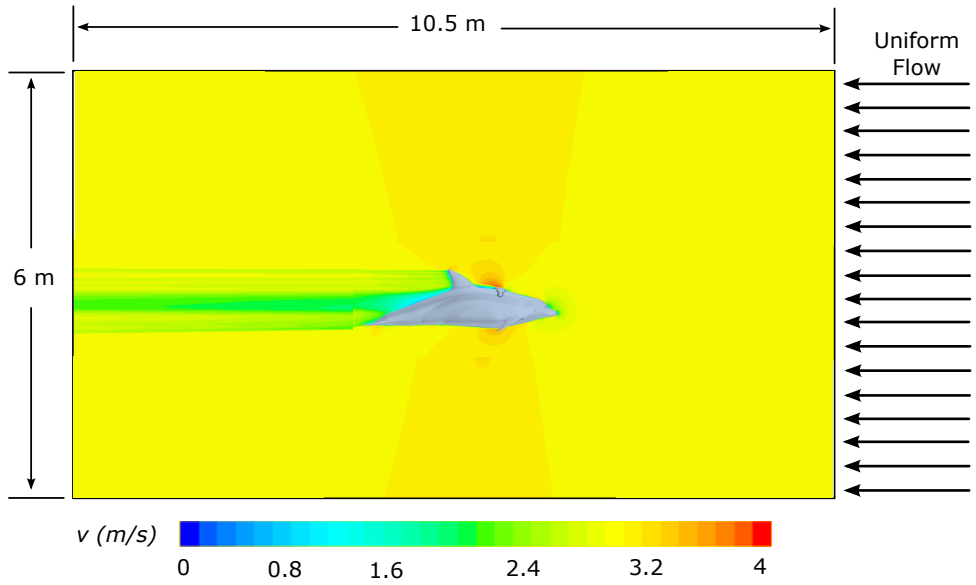


Figure 5.5: Illustration of the dolphin-tag model in the computational domain used for the simulations of all tag conditions. Velocity results from a representative simulation in 3 m/s steady state flow are presented in the figure.

where p_f is the face static pressure, A_f is the face area vector, and p_{ref} is the reference pressure. Shear force is also exerted on the surface face f by the moving fluid and was computed as:

$$F_f^{shear} = E_f \cdot A_f, \tag{5.28}$$

where E_f is the stress tensor at face f. The achieved level of residual convergence for all equations was below 1×10^{-6}. Additional simulations were performed to confirm that simulation results were independent of upstream and downstream domain lengths.

5.4.5 Statistical Analysis

Bootstrapping was used to create a confidence interval for the drag coefficients derived from the experimental data [177]. All experimental data for a given trial/animal was treated as a complete population. A simulated data set, with each condition trial containing the same number of data points as the original data set, was created by sampling with replacement. Regression and parameter estimation from the simulated data were performed in the same manner as described above to obtain an estimated drag coefficient, $\hat{C}_d^{(b)}$. This was repeated B times (B = 500) to obtain a set of estimates $\{\hat{C}_d^{(1)}, \cdots, \hat{C}_d^{(b)}, \cdots, \hat{C}_d^{(B)}\}$, which were used to obtain a 90% confidence interval (CI) for the experimentally derived drag coefficient.

A two-way ANOVA with post hoc Tukey's HSD was used to test whether drag coefficients from CFD simulations were significantly different between tag conditions (*no_tag/control*, *tag*, *tag+4* and *tag+8*), while considering the effect of velocity. We used ANOVA with post hoc Tukey's HSD to test for the effects of drag loading and individual on (1) drag coefficients estimated from experimental data and (2) glide duration.

5.5 Results

5.5.1 Estimated Drag of the Tag

Biologging tag data from a representative dive, along with the identified gliding events, are shown in Figure 5.6 (A-D). Equation 5.5 was used with the gliding speed profiles (Figure 5.6-E) to calculate the gliding drag coefficient $C_d^{(g)}$ [1]. A total of 532 gliding segments with an average duration of 1.97 ± 1.09 sec (mean ± SD) were identified from the animals for the analysis. The 532 pairs of $\{Re, C_d^{(g)}\}$ values are shown in Figure 5.6-F, which were further clustered into 7 balanced groups (76 pairs per group) to better visualize the distributions. A linear fit to the mean values from each group was obtained as:

$$\log_{10}(C_d^{(g)}) = -0.597 \log_{10}(Re) + 1.749 \tag{5.29}$$

which can be rearranged to give $C_d^{(g)} = 56.13 Re^{-0.60}$ and was used to estimate the gliding drag coefficient $C_d^{(g)}$ for the modeling analysis. For comparison, $C_d^{(f)}$ from [101] with the relationship $C_d^{(f)} = 16.99 Re^{-0.47}$, together with a $C_d^{(g)}$ estimate under *tag* condition from Section 5.5.4, are plotted along with the gliding coefficients found here in Figure 5.6-F.

5.5.2 Fluke and Glide (FG) Gait Characteristics

For the modeling analysis, 478 FG bouts, with an average fluking duration T_f of 4.3 ± 2.3 sec and gliding duration T_g of 5.0 ± 2.1 sec, were sampled from the data of one animal. Duty factor, $D_f = \frac{T_f}{T_f + T_g}$, of the FG gait had a value of 0.46 ± 0.12. The speed distributions of the animal while swimming (speed greater than 0.3 m/s) are shown in Figure 5.7. An overall swimming speed of around 1 m/s was most likely during all swimming gaits for this animal, while speed during FG gait was more likely to occur at a slightly higher speed (1.25 m/s). However, if we normalize the FG speed distribution over the general speed distribution (Section 5.3.4), to obtain a FG speed distribution associated with a *uniformly distributed* general speed, a FG speed of around 2.25 m/s

[1] To be distinguished with $C_d^{(f)}$, the fluking drag coefficient, in this part of the work.

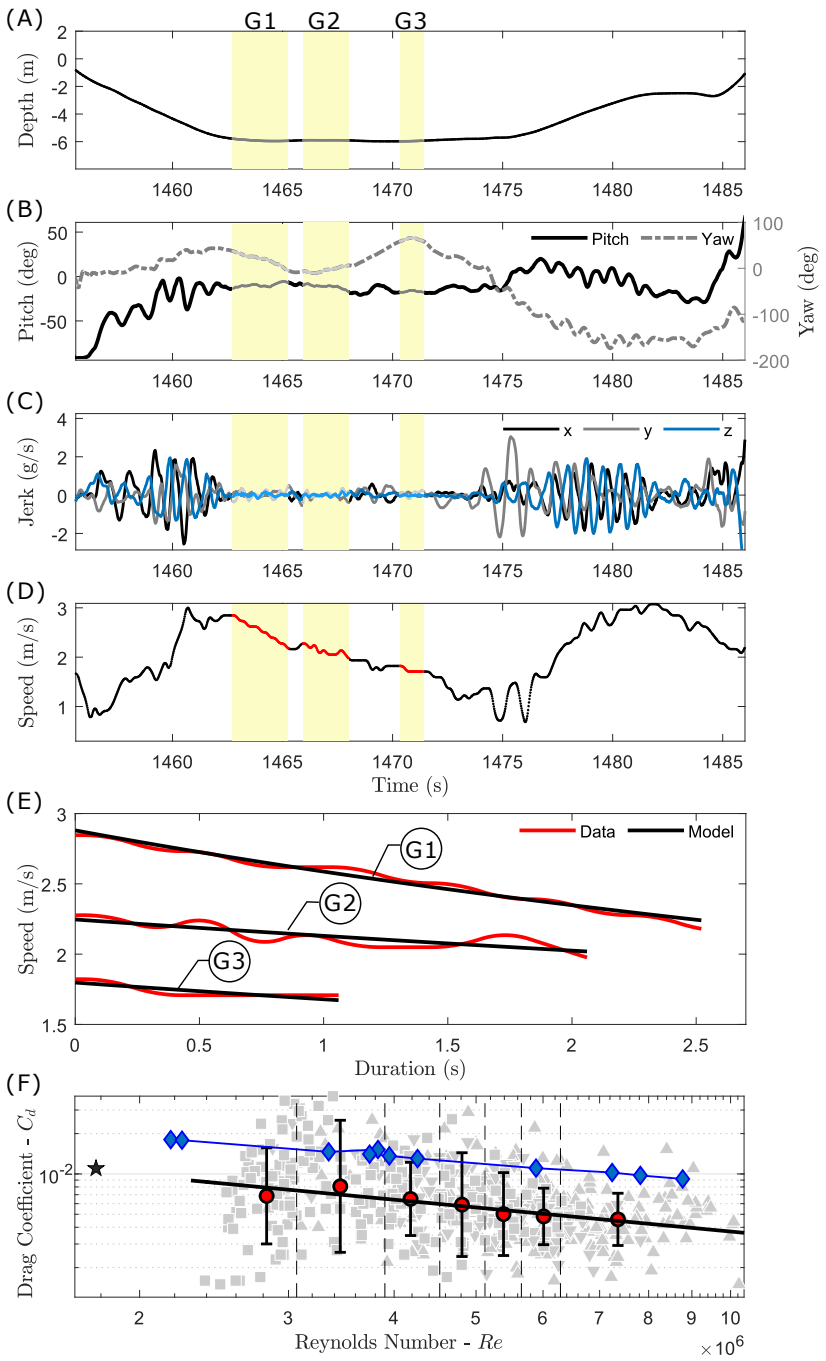


Figure 5.6: Signal features of an example section of data (**A**-**D**) and three extracted gliding events' speed profiles (G1, G2, G3) with fitted dynamic models (**E**) are shown in the first 5 plots. Plot (**F**) presents the resulting 532 pairs of $\{Re, C_d\}$ values from all animals in grey markers (up triangle, down triangle, and square for the 3 animals respectively). The pairs have been clustered into 7 groups based on the Re values. The mean values (red circle) as well as the ± SD range are shown for each group and a linear fit to all the mean values is represented by the black solid line. Results from low amplitude dolphin fluke (blue diamond, [101]) and low speed dolphin glide (black pentagram, Section 5.5.4) are also shown.

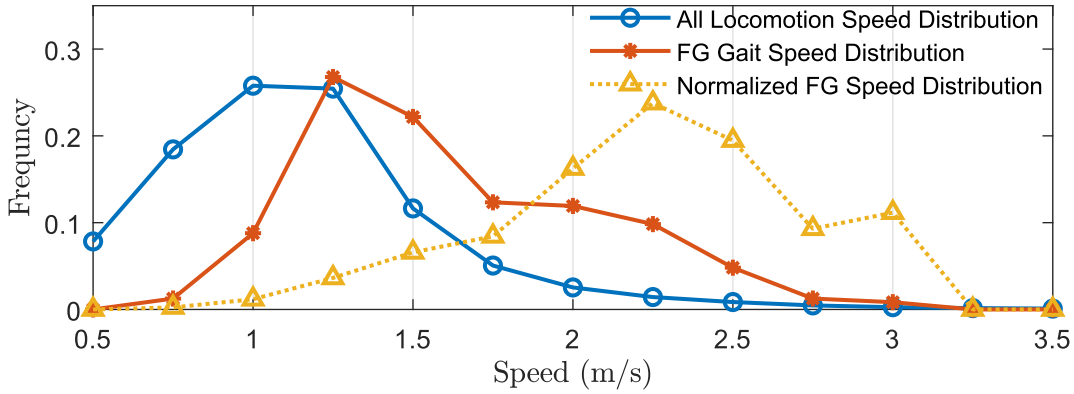


Figure 5.7: The speed distribution of all locomotion, sampled FG bouts as well as normalized FG bouts (Sections 5.3.4 & 5.5.2).

Table 5.1: Characteristics of FG Gait vs. Continuous Fluke Gait under varying Duty Factor and Gliding Duration

Gliding Duration T_g (s)	**Duty Factor*** $D_f = T_f/(T_f + T_g)$						
	0.34	***0.38***	***0.42***	***0.46***	***0.5***	***0.54***	***0.58***
3	0.284[a] / 34.9[b]	0.294 / 32.7	0.304 / 30.5	0.314 / 28.2	0.323 / 26.0	0.333 / 23.8	0.343 / 21.5
4	0.286 / 34.5	0.296 / 32.3	0.306 / 30.0	0.316 / 27.7	0.326 / 25.4	0.336 / 23.1	0.346 / 20.9
5	0.288 / 34.1	0.298 / 31.7	0.309 / 29.4	**0.319 / 27.0**	0.329 / 24.7	0.339 / 22.3	0.349 / 20.0
6	0.290 / 33.5	0.301 / 31.1	0.312 / 28.6	0.322 / 26.2	0.333 / 23.8	0.343 / 21.4	0.354 / 19.0
7	0.293 / 32.9	0.304 / 30.4	0.315 / 27.8	0.326 / 25.3	0.337 / 22.8	0.348 / 20.4	0.359 / 17.9

*Evaluated at the reference constant fluking speed of 2.25 m/s

*Associated mechanical energy cost of transport (MECOT) of constant speed fluking is 0.437 J $\cdot$ m^{-1} $\cdot$ kg^{-1}

[a]Left: MECOT of fluke-and-glide (FG) gait (J $\cdot$ m^{-1} $\cdot$ kg^{-1})

[b]Right: Percentage of MECOT saved as compared to constant speed fluking (%)

Bold cell is associated with the average duty factor and gliding duration obtained from the experimental data

would be most likely. This is a speed value near the previously reported optimal bottlenose dolphin swimming speed of 2.1 m/s [82].

5.5.3 Energetics as A Function of Gait

A representative reference speed of 2.25 m/s was used to compare the MECOT of the FG gait with the continuous constant speed fluking gait using the proposed model. While holding the reference swimming speed constant at 2.25 m/s, the model (Section 5.2.1, Figure 5.1-C) was used to explore how varying the gliding duration T_g and duty factor D_f affected the MECOT (Section 5.2.3). For example, the estimated MECOT of a continuous fluking gait at 2.25 m/s was 0.437 J $\cdot$ m^{-1} $\cdot$ kg^{-1}, while a FG gait with $T_g = 5.0$ sec and $D_f = 0.46$, resulted in a MECOT of 0.319 J $\cdot$ m^{-1} $\cdot$ kg^{-1}, a 27% reduction in cost. More results from different combinations of duty factor and gliding duration are shown in Table 5.1.

Table 5.2: Weight (kg) and length (m) of the two bottlenose dolphins in the study, their estimated wetted surface area (calculated from [171]) and the number of glides they each successfully performed for each experimental condition. Girth was measured approximately two fingers in front of the dorsal fin after animals exhaled.

Subject	Weight (kg)	Length (m)	Girth (m)	Frontal Area (m^2)	Wetted surface area (m^2)	No Tag	Tag	Tag +4	Tag +8
TT01 (63H4)	176	2.52	1.23	0.12	2.39	5	1	7	0
TT02 (01L5)	154	2.37	1.21	0.12	2.19	4	3	4	3

5.5.4 Assessing the Hydrodynamic Impact of a Tag

A total of 27 glides fit the necessary criteria for analysis (Table 5.2). Speed data for the model optimization were calculated from undistorted displacement data (Figure 5.4-D). Representative speed data from a glide, results from the parametric optimization and an example of the resampled model fit used in the bootstrapping analysis are presented in Figure 5.2. In the example, the parameters that minimized the fit residual to the speed data were $C_d = 0.0265$ and $v_0 = 0.78$ (Figure 5.2-Middle). The parameters that fit the resampled data were $\hat{C}_d = 0.0290$ and $\hat{v}_0 = 0.79$ (Figure 5.2-Bottom).

The uncorrected drag coefficients for the full data optimizations (C_{d_all}) ranged from 0.011 to 0.028, with the identified initial velocities ranging from 0.5 to 1.07 m/s (Table 5.3, Figure 5.8). Fit residuals for the full data optimization ranged from 0.12 for the *tag* to 0.21 for the *tag+8* condition for the two animals. There was a significant effect of *tag* condition on the estimated drag coefficients (ANOVA; $F_{3,22} = 7.5318$, $p = 0.0012$); we detected no significant effect of individual ($F_{1,22} = 0.6934$, $p = 0.4139$). There was no detectable increase between animals when not instrumented and when wearing the tag alone (post hoc Tukey's; $p = 0.7505$). The *tag+4* significantly increased C_d, by 126.4% ($p = 0.0019$) and the *tag+8* by 192.5% ($p = 0.0013$) compared to when the dolphins were not instrumented. From the experimental data, the *tag+4* and *tag+8* did not lead to significantly greater C_d compared to the *tag*-only ($p = 0.1557$ and 0.1664, respectively). The surface-induced drag corrected coefficients ($C_{d_all_correct}$) are also reported in Table 5.3, and are comparable to the simulated results.

The glide durations lasted on average 2.1 s (±0.75 SD) and, with the exception of TT01's single trial during the *tag* condition, were relatively consistent. We detected no significant effect of tag condition on glide duration ($F_{3,22} = 1.2501$, $p = 0.3157$). Success rates varied between the different conditions and individuals, with dolphin TT01 not able to successfully complete a glide

Table 5.3: Drag coefficients (C_d) calculated from the individual fitting to speed data, the number of successful trials, and the resulting fit residuals (R). The results from fitting to all of the data for a given trial (C_{d_all}) as well as the corresponding drag coefficient that has compensated for surface drag ($C_{d_all_correct}$), the 90% confidence interval are presented below.

Conditions & Animals	Parameters	Trial # 1	Trial # 2	Trial # 3	Trial # 4	Trial # 5	Trial # 6	Trial # 7	Mean	SD	C_{d_all}	$C_{d_all_correct}$
No Tag (TT01)	C_d	0.017	0.010	0.013	0.017	0.013	-	-	0.014	0.003	**0.0137 ± 0.006**	**0.0125 ± 0.0055**
	R	0.077	0.107	0.040	0.123	0.128	-	-	0.095	0.037	-	-
	$R\ (C_{d\ all})$	0.077	0.107	0.040	0.123	0.128	-	-	0.095	0.037	-	-
	U_0	0.780	0.790	0.790	0.990	1.050	-	-	0.874	0.136	-	-
	Duration	1.933	1.567	0.933	0.967	1.533	-	-	1.387	0.429	-	-
Tag (TT01)	C_d	0.011	-	-	-	-	-	-	0.011	0.000	**0.011 ± 0.01**	**0.010 ± 0.0091**
	R	0.209	-	-	-	-	-	-	0.209	0.000	-	-
	$R\ (C_{d\ all})$	0.209	-	-	-	-	-	-	0.209	0.000	-	-
	U_0	0.740	-	-	-	-	-	-	0.740	0.000	-	-
	Duration	3.567	-	-	-	-	-	-	3.567	0.000	-	-
Tag+4 (TT01)	C_d	0.027	0.018	0.015	0.023	0.038	0.034	0.039	0.027	0.010	**0.0225 ± 0.002**	**0.0205 ± 0.0018**
	R	0.062	0.042	0.054	0.062	0.081	0.066	0.053	0.060	0.012	-	-
	$R\ (C_{d\ all})$	0.063	0.045	0.060	0.062	0.090	0.069	0.056	0.064	0.014	-	-
	U_0	0.780	1.050	1.070	1.040	1.060	0.970	0.930	0.986	0.104	-	-
	Duration	3.367	1.967	1.933	2.100	2.000	1.333	0.900	1.943	0.765	-	-
No Tag (TT02)	C_d	0.023	0.020	0.011	0.004	-	-	-	0.014	0.009	**0.0158 ± 0.005**	**0.0121 ± 0.0038**
	R	0.050	0.151	0.203	0.186	-	-	-	0.147	0.069	-	-
	$R\ (C_{d\ all})$	0.052	0.151	0.204	0.190	-	-	-	0.149	0.069	-	-
	U_0	0.920	0.950	1.020	0.890	-	-	-	0.945	0.056	-	-
	Duration	1.800	2.233	2.633	1.967	-	-	-	2.158	0.364	-	-
Tag (TT02)	C_d	0.027	0.013	0.024	-	-	-	-	0.021	0.007	**0.021 ± 0.005**	**0.0161 ± 0.0038**
	R	0.055	0.025	0.200	-	-	-	-	0.093	0.094	-	-
	$R\ (C_{d\ all})$	0.056	0.027	0.201	-	-	-	-	0.094	0.093	-	-
	U_0	0.720	0.650	0.980	-	-	-	-	0.783	0.174	-	-
	Duration	1.667	1.733	3.100	-	-	-	-	2.167	0.809	-	-
Tag+4 (TT02)	C_d	0.026	0.031	0.023	0.039	-	-	-	0.030	0.007	**0.0265 ± 0.009**	**0.0203 ± 0.0069**
	R	0.171	0.150	0.142	0.274	-	-	-	0.184	0.061	-	-
	$R\ (C_{d\ all})$	0.171	0.151	0.142	0.274	-	-	-	0.184	0.061	-	-
	U_0	0.670	0.720	0.800	0.700	-	-	-	0.723	0.056	-	-
	Duration	2.667	2.667	2.300	1.033	-	-	-	2.167	0.775	-	-
Tag+8 (TT02)	C_d	0.033	0.023	0.037	-	-	-	-	0.031	0.007	**0.028 ± 0.01**	**0.0214 ± 0.0077**
	R	0.100	0.188	0.078	-	-	-	-	0.122	0.059	-	-
	$R\ (C_{d\ all})$	0.100	0.189	0.079	-	-	-	-	0.123	0.059	-	-
	U_0	0.500	0.800	0.750	-	-	-	-	0.683	0.161	-	-
	Duration	2.033	3.467	1.600	-	-	-	-	2.367	0.977	-	-

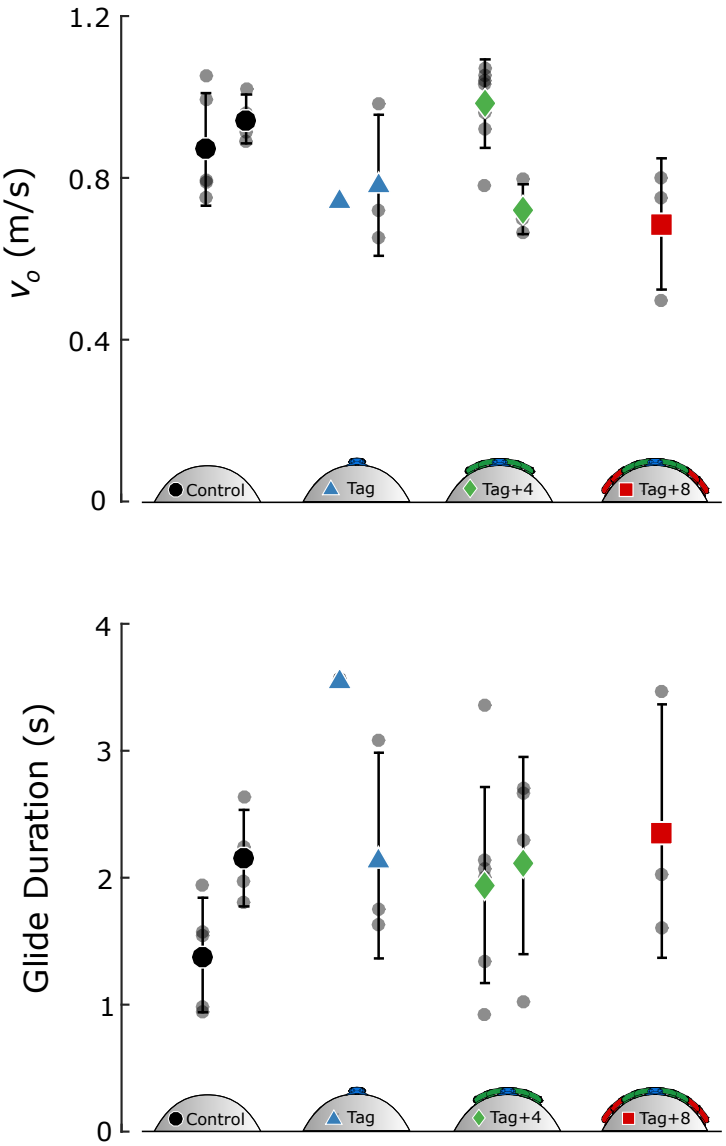


Figure 5.8: (**Top**) Average and individual initial velocities calculated during the parameter optimization for both animals. Error bars indicate one SD around the mean value. (**Bottom**) Mean and individual glide duration of each successful trial. Error bars indicate one SD around the mean value.

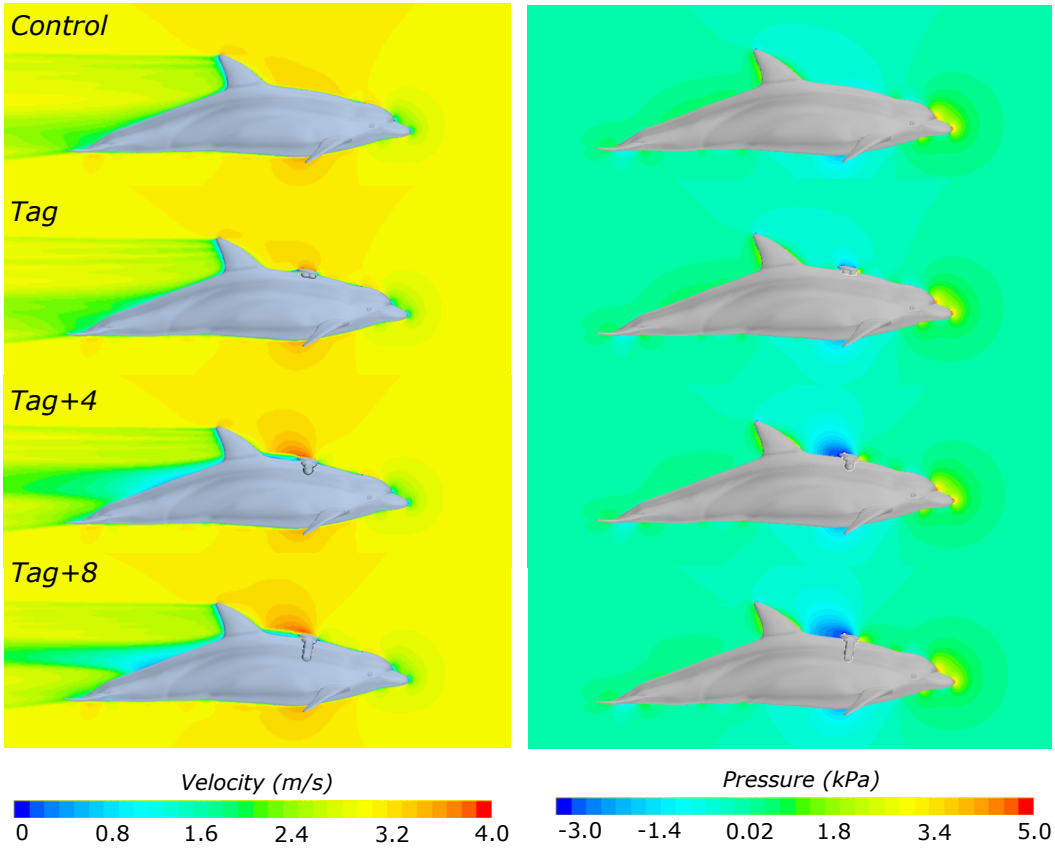


Figure 5.9: Computational Fluid Dynamics simulation results from the dolphin and combined dolphin-tag models that illustrate how the tag and added drag elements affect the velocity of the flow (**left**) and the resulting pressure differentials (**right**) around the model of the animal.

with all the drag elements (*tag+8*; Table 5.2).

Qualitatively, CFD simulations show the flow disturbance during the loading conditions (Figure 5.9). The *tag* condition creates a minor disturbance to the flow, forming a small region of recirculation upstream of the dorsal fin. Additional drag elements (*tag+4* and *tag+8*) significantly increased the recirculation region, which eventually occupies the entire space between the tag and the dorsal fin. This effect is also visible downstream from the dolphin body. The flow redistribution was reflected in the increased magnitude of the force acting on the coupled system.

Drag coefficients calculated from the CFD simulations increased as more elements were added (*tag+4* and *tag+8*), and ranged from 0.0071 to 0.0207. These results were comparable to the corrected experimental drag coefficients, with all but the CFD *tag* condition within a calculated 90% confidence interval of the experimental results. In the region near the tag, the velocity of the fluid increased as it moved over the top of the body creating an area of low-pressure behind the tag. These flow disturbances increased drag forces on the dolphin body (Figure 5.10-Top). As speed increased from 1-6 m/s, drag forces on the animal model increased from 15 N to 387 N (Table 5.4).

Table 5.4: Drag coefficients (C_d) calculated from the individual fitting to speed data, the number of successful trials, and the resulting fit residuals (R). The results from fitting to all of the data for a given trial (C_{d_all}) as well as the corresponding drag coefficient that has compensated for surface drag ($C_{d_all_correct}$), the 90% confidence intervals are presented below.

Condintion	A_w (m^2)	F_d (N) 1 m/s	C_d	F_d (N) 2 m/s	C_d	F_d (N) 3 m/s	C_d	F_d (N) 4 m/s	C_d	F_d (N) 5 m/s	C_d	F_d (N) 6 m/s	C_d
Control	2.915	14.9	0.010	49.8	0.008	102.4	0.008	177.2	0.007	272.0	0.007	387.0	0.007
Tag	2.939	15.6	0.010	52.6	0.009	109.0	0.008	188.6	0.008	290.2	0.008	412.5	0.008
Tag+4	2.983	22.5	0.015	81.6	0.013	175.8	0.013	308.6	0.013	475.2	0.012	686.1	0.013
Tag+8	3.029	32.3	0.021	122.3	0.020	267.8	0.019	472.2	0.019	734.5	0.019	1053.8	0.019

The tag added 1-25 N across these speeds, increasing the drag coefficient by only 4%, from 0.0076 to 0.0080. There were significant effects of flow speed (two-way ANOVA; $F_{1,24} = 70.3$, $p < 0.0001$) and *tag* condition ($F_{4,24} = 541.3$, $p < 0.0001$) on CFD-derived drag coefficients. The addition of the tag did not yield a significant increase in drag coefficient (post hoc Tukey's; $p = 0.6119$). In contrast, the *tag+4* and *tag+8* conditions increased the drag coefficient by 67% and 152% respectively, compared to the noninstrumented condition (i.e. *no_tag*, Table 5.4; $p < 0.0001$ for both comparisons). Additionally, drag coefficients for the *tag+4* and *tag+8* were significantly higher than the *tag* alone ($p < 0.0001$ for both comparisons).

To compare the experimental and simulation results directly, drag force acting on TT02 in 1 m/s flow were calculated for all tag conditions using (1) the experimentally derived drag coefficient (C_{d_all}); (2) the corrected experimental coefficients ($C_{d_all_correct}$); and (3) simulated coefficients (C_{d_sim}), Figure 5.10-Bottom. The 90% confidence intervals for each of the experimental force estimates are presented to capture uncertainty in the estimates. The magnitude of the simulated force estimates was smaller than the forces predicted using the experimental results. The relative increase in drag force created by the tag conditions varied between the two estimates, with the relative increase predicted by the experiment larger for the *tag* condition (4% vs. 33%), comparable for the *tag+4* (47% vs. 68%), and smaller (108% vs. 77%) for the *tag+8*.

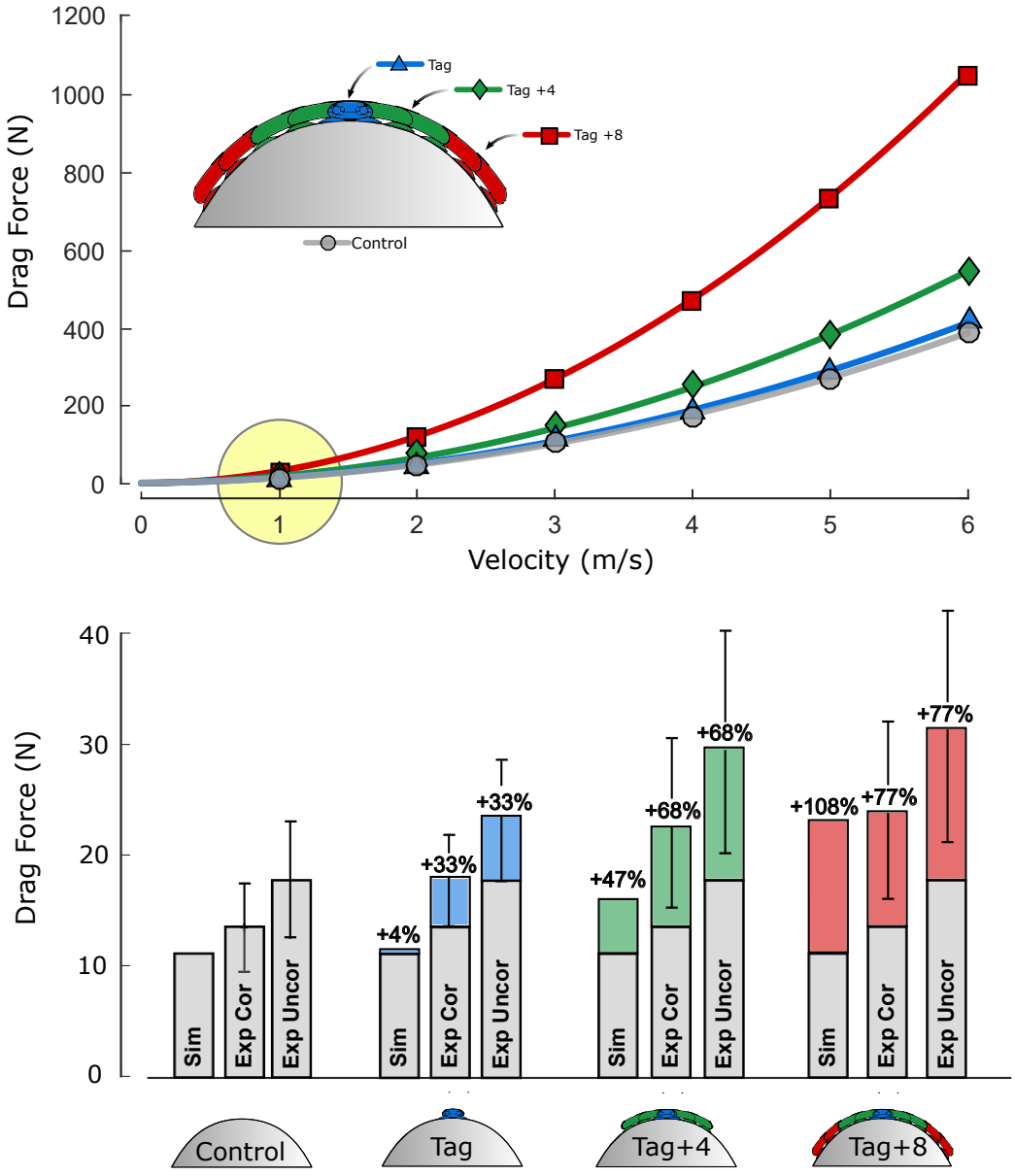


Figure 5.10: (**Top**) Net drag forces calculated from the Computational Fluid Dynamics simulation results. The control and tag conditions have similar net forces over the range of simulated speeds, but the additional of the drag elements greatly increases the force acting on the combined body. The yellow circle highlights the results near 1 m/s, a comparable speed to the gliding trials. (**Bottom**) A comparison of the estimated drag force acting on TT02 at speed of 1 m/s during the four experimental conditions over simulation (Sim) and experiment results that have been corrected for surface drag effect (Exp Cor) as well as the uncorrected raw drag force (Exp Uncor). The percent increase over the respective control trial is provided for all of the tagged conditions. The 90% confidence interval is also plotted with the experimental data.

5.6 Discussion

5.6.1 Energetic Benefits of Fluke and glide gait

The results from the investigation of the FG gait of the animal provide insight into the energetic benefits of a FG gait using new estimates of drag on gliding bottlenose dolphins over a range of swimming speeds. The data-driven approach used in this work combined a physics based model and experimental data to investigate dolphin swimming efficiency. The benefits of the approach were threefold: first, there was no need for dedicated experiments that involve divers or underwater cameras for drag estimation. Second, data were collected during the animal's daily routine, linking the theoretical model with the animals' natural behavior. Third, the many hours of tag data resulted in a large number of events for the analysis, helping to reduce the noise and uncertainty in the estimated parameters.

As expected, the range of drag coefficients estimated for gliding (Figure 5.6-F) are lower than those calculated during low amplitude fluking [101]. The changing body posture of the animal during swimming contributes to the increased drag. The dynamic dorsal-ventral bending of the body used to generate propulsive force for locomotion results in body configurations that generate more drag than the static pose used during a glide. The Bone-Lighthill boundary layer thinning hypothesis also suggests that undulating motion increases skin friction drag [90, 91]. The reduction in drag is significant across the range of speeds. For example, an animal experiences 2.2 times more drag during low amplitude fluking than during a glide at a Reynolds number of 5.216×10^6. This significant reduction contributes to the energetic benefits of the FG gait observed in the swimming patterns of whales and dolphins [44].

Energetic benefits of the FG gait were then further investigated using the proposed model and parameters identified from observed bouts of FG gait. Results from the model at a 2.25 m/s swimming speed show a 27% reduction in the MECOT compared to constant speed fluking. A 27% savings for bottlenose dolphins is smaller than the over 50% estimated savings for fish [92], but for dolphins that may travel more than 50 km per day [45], a 27% cost savings would result in a significant reduction in the foraging effort required to maintain the animal's energy budget. Additional results in Table 5.1 demonstrate that MECOT of a FG gait is lower when the duty factor is lower, that is, swimming efficiency improves when the animal spends more time gliding and less time fluking. While a small duty factor with as much gliding as possible leads to improved MECOT, a low duty factor also requires higher acceleration during the limited amount of fluking to build up speed for the glide. Higher acceleration leads to two potential drawbacks: first, the metabolic cost to support the burst could be higher. And second, the animal may have to increase its fluking amplitude to satisfy the kinematic demand of higher acceleration. The drag coefficient for fluking ($C_d^{(f)}$) in this work comes from estimates made by [101] during low amplitude dolphin

fluking. We speculate that the drag coefficient associated with large amplitude fluking would be higher. Another trend present in Table 5.1 is that longer gliding duration leads to higher cost at the same duty factor. This is because longer glides require higher initial gliding speeds, which are penalized by the fact that drag force increases quadratically with speed (Eq. 5.2).

While expert knowledge was used to select features from the sensor data to create a decision tree classifier, and the qualified FG bouts were identified by experts, other automated machine learning algorithms could be used to identify events of interest in the data ([178], Chapter 3). These automated approaches could further reduce the amount of expert knowledge required for signal feature selection and event detection for potentially more complicated movement patterns, but would still require a set of labeled data to train the algorithms. As an extension of this work, the on-body tag measurements could be combined with a dedicated model of the animal's full body kinematics to provide thrust and drag estimates under different scenarios, which could provide more detailed insights about the locomotion of these animals. Future work will also use tag data from dolphins in the wild to investigate the cost of transport associated with swimming gaits used in a different environmental context.

5.6.2 On the Hydrodynamic Impact of a Tag

The ability to estimate drag force from a swimming animal is key to an improved understanding of how biologging tags may impact animal behavior and energetic cost. To this end, we modified drag on gliding bottlenose dolphins with a modular tag system, collected experimental data that were used to estimate drag, and compared those experimental estimates to results from CFD. The simulations and experiments complement each other in this part of the work, with the physical experiments providing data to verify the simulations, and the virtual model extending the study to conditions that are hard to achieve experimentally.

The free-stream speeds used in simulation (1-6 m/s; 0.4-2.5 l/s, Reynolds numbers between 2×10^6-1×10^7) reflect the natural range of swimming speeds in dolphin species. Although dolphins are known to have a high capacity for maximum swimming speeds when trained (6.5-8.2 m/s; 2.5-3.8 l/s) [179] or when performing burst behaviors (15 m/s; 5.9 l/s) [180], routine speeds are much lower (1.6-5.6 m/s; 0.6-2.2 l/s) [45, 181, 182, 183, 184]. Minimum transport costs (COT_{min}) occur at swimming speeds of 2.1-2.5 m/s (~1 l/s) [185, 186]. CFD simulations enabled a virtual representation of the dolphin-tag system, and the resulting flow visualization provided a qualitative means of assessing the impact of tag geometry on the flow field around the animal (Figure 5.9) [165].

Transition from laminar to turbulent flow occurs somewhere between 5×105 to 1×107 (Reynolds number), depending on the shape and smoothness of the body [170]. During a glide,

velocity increases at the rostrum and along the front of the body in areas of high pressure, maintaining a thin boundary layer with low viscous forces (Figure 5.9) [187]. Velocity and pressure decrease as flow passes the widest point of the body, 0.34-0.45 l [171]; the transition to turbulent flow is most likely to occur at this point [90]. Pressure-drag increases due to the imbalance of pressure in front of and behind this widest point [170, 187]. For non-instrumented animals at low gliding speeds, other work suggests that the boundary layer remains attached up to the base of the flukes [80, 188]. Larger tags and higher speeds may lead to short-term separation (Figure 5.9), though reattachment is likely, especially with active movement [189].

Our simulations show that tag geometry and placement create a greater imbalance in these pressure fields and increase drag (Figure 5.9 & 5.10). Over the range of the simulated free-stream fluid speeds, the tag increased the force acting on the modeled dolphin geometry by between 4% and 7%. The bluff elements used to increase the frontal area of the tag created large areas of fluid damming (high pressure in front of the tag), stagnant flow (low pressure) behind the tag, and flow separation. These elements also create large areas of recirculating flow behind the tag, perturbing the fluid flow around the animal. The additional elements increased the drag forces over the range of free-stream speeds by between 47% and 77% for the *tag+4* condition and between 108% and 172% with the *tag+8* condition. These simulation results provide important insight into the effects of the tag on the coupled system, but the mechanics of the fluid in the simulations are complex, necessitating the experimental verification of the simulated results.

Glide speeds during the experimental trials (0.45-1.1 m/s) were slower than routine animal swimming speeds, translating to Reynolds numbers between 1-4$\times 10^6$. The animals self-selected their gliding posture, making the experimental trials more dynamic than the simulated results [190]. Data were only used from glides selected with the protocol described in the methods, reducing the effect of body and fin movement on the experimental estimates of drag coefficients. As drag was added to the animals, variability in estimated drag coefficient increased and glide success was reduced (Figure 5.8). In general, animal performance of the gliding task deteriorated beyond the *tag+4* condition, with one animal unable to complete the *tag+8* condition successfully. These observations during the experiment agree with the concerns from previous work showing behavioral changes with instrumentation [97, 98, 191], and especially with large tags [192, 193, 194, 195, 196]. Despite the challenges associated with data collection from live animals, the experimental results show the trend predicted by simulations (Figure 5.9). During the experimental trials the drag elements created a large increase in drag acting on the animal, with the *tag+8* condition increasing the drag coefficient of the coupled system by 77%.

To provide context for the results presented here for the assessment of the hydrodynamic impact of tag, both the experimental and the simulation coefficients are plotted together with other measurements and estimates found in the literature (Figure 5.11). Drag coefficients in our study

were calculated at lower Reynolds numbers than other experimentally derived coefficients, but the control condition is comparable to drag coefficients calculated for the slow swimming speeds presented by Fish et al. ([101]; teal circles). Our results are lower than [101] but are for gliding (i.e. $C_d^{(g)}$), where as those in [101] are for low amplitude fluking animals (i.e. $C_d^{(f)}$). The experimental results are plotted with ellipses that represent the uncertainty in our estimate of the coefficient and the variability of the animal's speed during the trial. The results derived from the CFD are lower than the experimental coefficients, but the *control*, *tag+4* and *tag+8* values fall within the uncertainty of our experimental results. We also compared the relative increase in drag created by the tag configurations from both the simulated and experimental results (Figure 5.10). These results are mixed, with a comparable relative increase for the *tag+4* (+47% simulation vs. +66% experiment) and *tag+8* (+108% simulation vs. +77% experiment), but not as well for the *tag* condition (+4% simulation vs. +33% experiment). The mixed results could be explained by differences between the modeled and actual dolphin geometry, self-selected gliding posture or in the orientation of the control surfaces (pectoral fins).

For both the simulation and experimental results, the bluff elements created a significant amount of added drag to the animal, but only increased the wetted surface area of the animal by ~3% (*tag+4*) and ~5% (*tag+8*) and the frontal area of the combined system by ~9% (*tag+4*) and ~17% (*tag+8*). These results speak to benefits of tags design with hydrodynamic fairings, even if those tag designs have slightly larger surface areas or cross-sections [165, 197]. The experimental results have variability and measurement noise, but we expect our ability to estimate parameters from these data to improve as the number of experimental trials grows. As presented in Table 5.3, when a higher number of trials were used for the experimental estimate (e.g. *tag+4* (TT01), *no-tag* (TT01)), error bounds on C_{d_all} were smaller. Along these lines, we had comparable agreement between the relative increase in drag at the *tag+4* condition, where we had the most data to estimate the drag coefficient (11 trials).

More kinematic data from additional animals would improve the parameter estimates and reduce the variability seen in the experimental results. However, there are a limited number of facilities with animals capable of successfully performing controlled swimming tasks, resulting in a constraint on the amount of data that can be readily collected for studies like these. This makes methods that combine experimental and computation approaches appealing when investigating these types of questions. For example, the experimentally estimated drag coefficients presented here were calculated from relatively slow glides compared to average animal swimming speeds reported in the literature (0.8 m/s vs. 2 m/s), but there was no restriction on the speeds that could be investigated in simulation. Using the available experimental data to verify and improve the computational estimates will be key to improving the viability of future computational work. In this part of the work, differences were observed between the simulated and experimental results at

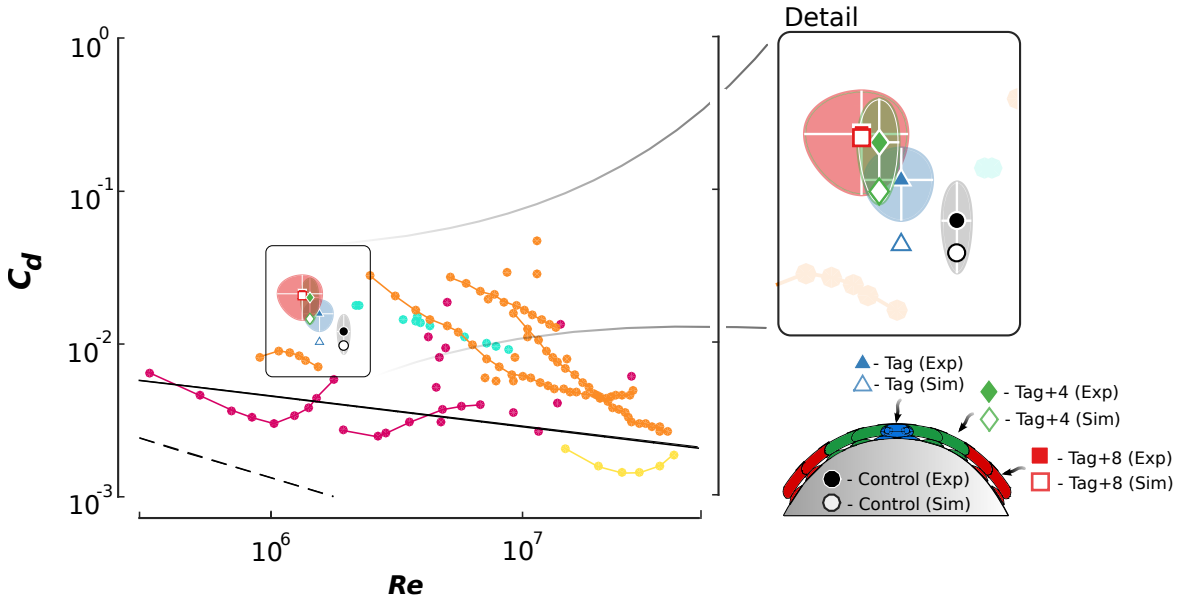


Figure 5.11: Drag coefficients of dolphins from historical data compared to this study. Data from this study follow the same convention as our other figures (black circles = no tag; blue triangles = tag; green diamonds = tag+4; red squares = tag+8). The uncertainty ellipse around each of the experimentally derived drag coefficients was formed using the 90% confidence bound calculated using bootstrapping and the SD of the initial velocity data. Unfilled symbols represent results from Computational Fluid Dynamics. Results from this work are shown alongside data from the literature as presented in [101]. Experiments on rigid models, towed bodies and gliding animals (maroon circles), from hydrodynamic models based on swimming kinematics (orange circles), from a rigid 'dolphin' model with the shape of a solid of revolution of the NACA 66 series (yellow circles), and from particle image velocimetry (teal circles). As in [101], the black line represents the frictional drag coefficient for a flat plate with turbulent boundary layer flow and the dashed line is for a flat plate with laminar boundary flow.

the slow gliding speeds (~1 m/s). The simulated and experimental drag coefficients were within the 90% confidence interval calculated for the experimental estimates for all but the tag condition, but the results could be improved by working to refine the animal geometry used in the modeling. Animal-specific scans could be used in both the CFD simulations and the force models used with the experimental data to produce individualized models for better comparison between methods.

While the drag estimates presented here only required planar kinematic measurements from one marker, the calibration of the camera data could be refined to improve the motion measurements. We used a simplified approach based on a known distance between two markers on the body to calibrate camera data; however, a full underwater camera calibration using methods presented in [198, 199] could reduce noise in the experimental data.

5.7 Summary

The results presented here indicate that the horizontal FG gait, in addition to porpoising/leaping [80, 81, 200], wave riding [82], drafting [83] and gliding during diving [95, 96, 42], is a behavioral mechanism that can enhance the locomotion efficiency of swimming dolphins. In this work we presented a data-driven physics based approach to estimate gliding drag coefficients and the mechanical energy cost of transport (MECOT) of bottlenose dolphins during horizontal intermittent locomotion. Kinematic measurements of animal motion were collected with a customized non-invasive biologging tag, and FG gait was modeled as a two-phase process. The MECOT of an average bout was estimated to be 0.319 $J \cdot m^{-1} \cdot kg^{-1}$ at the cruising speed of 2.25 m/s, which is 27% lower than the corresponding continuous fluking gait.

Quantifying the forces created by transmitting and archival tags used for animal science is an essential part of reducing the impact of these instruments. Our results suggest that computational and experimental methods can be used in a complementary manner to determine the increase in drag associated with tags, and the hydrodynamic impact of a single tag is quantifiably insignificant. The body of work seeking to assess the impacts of tags across taxa is growing [97, 98, 158, 191, 201, 202], and this work provides a framework that can be used to support these studies. Additionally, the computational approach used to estimate the drag and quantify the uncertainty of the estimate from the experimental data can be applied to other data sets collected from animals in both controlled and free-swimming environments.

CHAPTER 6

Fine-Scale Behavior Analysis using Cameras and Biologging Tags

6.1 Background

Features in an animal's environment have the potential to influence behavior, biomechanics, and movement patterns during daily life. For animals in managed settings, monitoring movement patterns to inform welfare practices is essential [205, 206, 207]. Behavior and movement patterns in response to common or long-term events in the surrounding environment can be predictable and identifiable [208, 209, 210]. For example, terrestrial animal studies [211] have shown that fixed feeding times during the day can result in stable and predictable behavior in stump-tailed macaques under human care. Previous animal studies have also investigated how enrichment and human interaction affect dolphin behavior in managed settings with trained observers monitoring and recording behavior [205, 212]. Key to the identification of how these environmental conditions influence the animals is the ability to measure and quantify movement without influencing behavior [213]. Traditionally, trained researchers directly observe and manually record animal behavior, but this approach can be time consuming and is limited to when the animals are visible to the observer [205, 212]. Observational approaches also tend to be qualitative and are not capable of capturing the kinematics (position, speed, acceleration) of the animals as they move in the environment.

Content of this chapter is mainly associated with the following publications:

[203] D. Zhang*, J. Gabaldon, K. Goodbar, N. West, K. Barton, and K. A. Shorter, "Fine-scale dynamic profiles of Bottlenose dolphin during conspecific and enrichment interactions," in *Preparation for Submission*, 2021.

[204] Z. Zhang, D. Zhang*, J. Gabaldon, K. Goodbar, N. West, K. Barton, and K. A. Shorter*, "Investigation of environmentally dependent movement of bottlenose dolphins (*Tursiops truncatus*)," *Journal of Zoological and Botanical Gardens*, vol. 2, no. 3, pp. 335–348, 2021.

* Corresponding Author

Biologging tags, which consist of sensors that record animal movement, sound, video, and other environmental measures, are used to provide high resolution measurements from animals ([39], Chapters 2&3). Tags provide direct measurements of kinematics, but do not capture much information about the environmental context that may be driving behavior. Camera data have been used in conjunction with direct observation to quantify swimming kinematics and body posture ([39], Chapters 3&4). Camera systems installed in the environment enable continuous records of movement, environmental use, social interaction, and animal kinematics [101, 112, 153, 152, 146, 214, 215]. For shorter duration sessions, manual tracking is possible [98]; however, this becomes intractable as video lengths grow. Ideally, cameras could be used to persistently monitor animals and their environment for days or weeks at a time, but this requires robust automated object tracking methods with sufficient precision to characterize animal movement.

Here we implement an automated computer vision framework using methods from robotics to track bottlenose dolphins, people, and enrichment in a lagoon environment (Section 4.3.3), and investigate how the presence of enrichment influences group movement. Further, biologging tag data in conjunction with overhead camera data is used to achieve prolonged identity preserving individual tracking (Chapter 4) and motion analysis (Chapters 2&3) during environmental interaction and conspecific interaction. Computer-automated animal tracking in a managed environment has only been explored recently, with initial research employing hand-crafted detection methods [152, 210], later followed by a convolutional neural-network approach [146] (Chapter 4). The detector used in this chapter, Mask R-CNN [63], is a state-of-the-art object detection technique that is an enhancement of the Faster R-CNN detector [62] employed in [146] (Chapter 4). Mask R-CNN was chosen for its additional precision over Faster R-CNN, and the ability to extract object profiles. Using this algorithm to simultaneously mask and track the animals as they engage and respond to people and enrichment provides important kinematic information to quantify movement during these conditions.

For the experiment, biologging tag data and video data were recorded at midday from animals in the main lagoon at Dolphin Quest Oahu ((Honolulu, HI, USA) during an extended period of unstructured swimming, both with and without enrichment. Positions and contours of dolphins, people, and enrichment were extracted from video frames and used to form dynamic tracks (tracklets, Section 4.3.3). Animal kinematics (position, speed, heading) from these tracks were used to investigate how the animals move in the environment both with and without enrichment. These data were also used to score dynamic interactions between animals and enrichment, a first step towards quantification of animal preference for different enrichment types. Gait and motion state information were further extracted from the tagged individuals to characterize their behavior during the interactions with conspecifics and enrichment. Importantly, the computer-vision-based analysis approach presented has broad applicability, and can be implemented to quantify movement and be-

havior as long as cameras and tags can be deployed in the environment, e.g., [152, 146, 214, 215].

6.2 Materials and Methods

6.2.1 Environment Setups

The proposed research was conducted in the main lagoon environment at Dolphin Quest Oahu, Hawaii. The lagoon is ~36 m wide and ~54 m long, with two floating docks separating the main environment from two adjacent lagoons, Figure 6.1. The docks are used by the animal care specialists to move within the lagoon system and interact with the animals. During the trials, the movement of the animals in the lagoon was recorded using a stationary overhead camera (Allied Vision Manta G-319) with a wide angle lens (Kowa LM4NCL ½" 3.5 mm F/1.4 C-Mount Lock Screw Industrial Lens) equipped with a linear glass polarizing filter (Edmund Optics, 70 mm unmounted linear polarizer, wavelength range 400 to 700 nm) mounted using custom 3D-printed hardware. The wide angle lens (with polarizing filter) was calibrated using the Fisheye Camera Calibration tools in MATLAB and was able to capture the entire lagoon, along with the animal care specialists when present on the docks. Data from four bottlenose dolphins were collected over 24 days during periods of free swimming at midday, with hour long sessions taking place from 12:00 pm to 1:00 pm or from 1:00 pm to 2:00 pm. During these lunchtime sessions the animals moved freely in the environment, and the natural lighting conditions improved the ability to observe the dolphins when swimming underwater. Three or four animals stayed in the main environment during each trial, with one to three of the animals wearing a MTag (Section 2.2) depending on the day.

Animal care specialists were free to move within the environment during the trials. For half of the sessions, one of two types of floating enrichment was introduced into the environment: red soft rounded jolly balls or black cylindrical buoys. Jolly balls are smaller and dolphins can grasp them with their mouths when swimming, while buoys are larger and harder for the dolphins to bring underwater. Both enrichment types have been routinely presented to the animals by the animal care specialists on a pseudo randomized schedule prior to this study as a means of promoting welfare.

For the analysis of an individual dolphin's behavior during enrichment interactions and conspecific interactions, camera and biologging tag data from 3 animals (TT01, TT02, and TT03) were selected from 3 experimental trials (Trial A, B, and C, 30 minutes each). Enrichment floats were available in Trials A & B. Animal TT01 was observed performing prolonged paired swimming[1] (with TT04, not tagged) in Trial B, as well as paired swimming and chasing in Trial C. Animals

[1]Two animals swimming coordinately side by side, with one animal swimming half a body length behind the other animal.

TT02 and TT03 did not engage in any prolonged conspecific interactions during the trials.

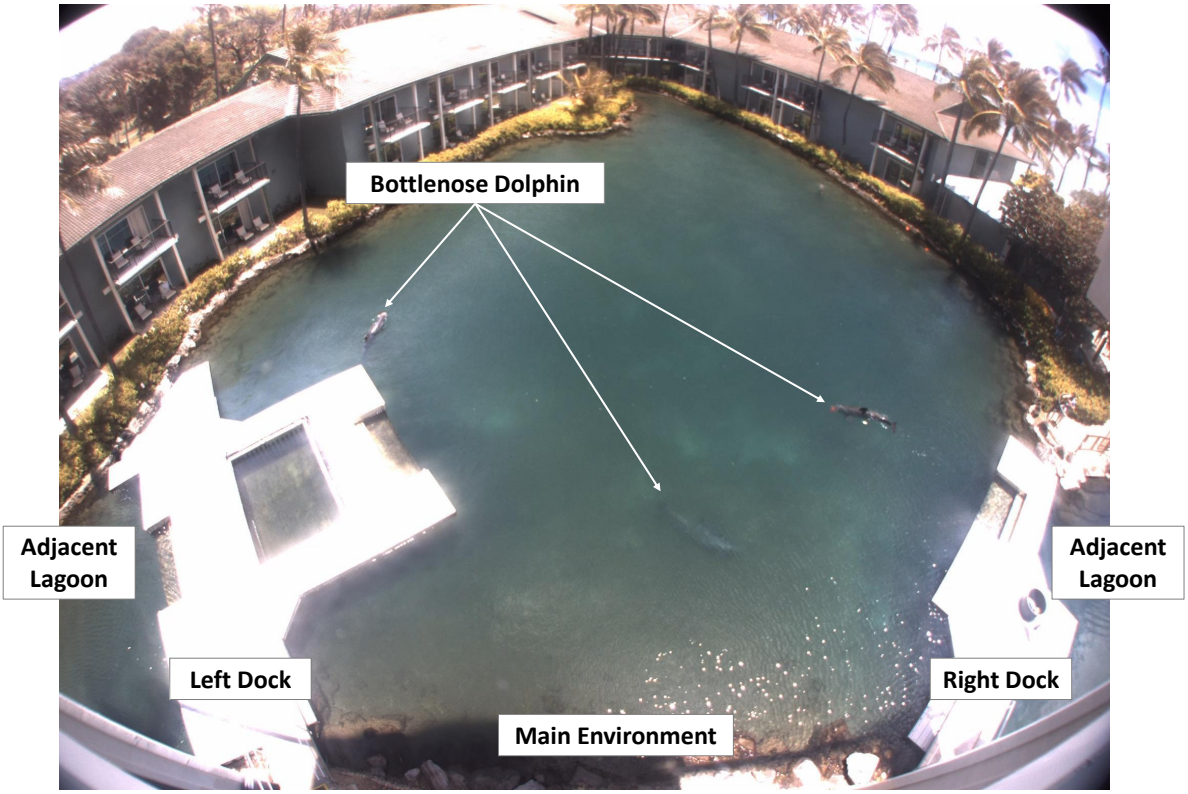


Figure 6.1: Monitoring camera's view of the experimental environment (Section 6.2.1).

6.2.2 Target Detection

Three objects were detected and tracked in this study to explore how environmental stimuli affect animal movement: dolphins, people (animal care specialists), and floating enrichment. A convolutional neural network, Mask R-CNN [63], was implemented and trained to detect the object types from multi-hour video recordings. Mask R-CNN outputs include instance detection, classification, and pixel-level object segmentation. Objects in the overhead video data are small and lack detail. Mask R-CNN was selected for this work because it is a well developed and validated object segmentation method, making it possible to reliably detect and mask objects using a more limited set of image features.

The model was trained using a dataset composed of 1200 frames selected from two 1-h sessions in August 2020. An additional 50 labeled frames for each type of enrichment were selected from 2 days to supplement the original training dataset. Objects (dolphins, people, enrichment) were manually annotated using the VGG Image Annotator [216], a stand-alone software package for manually annotating images. Each annotated object was also manually traced for the masking.

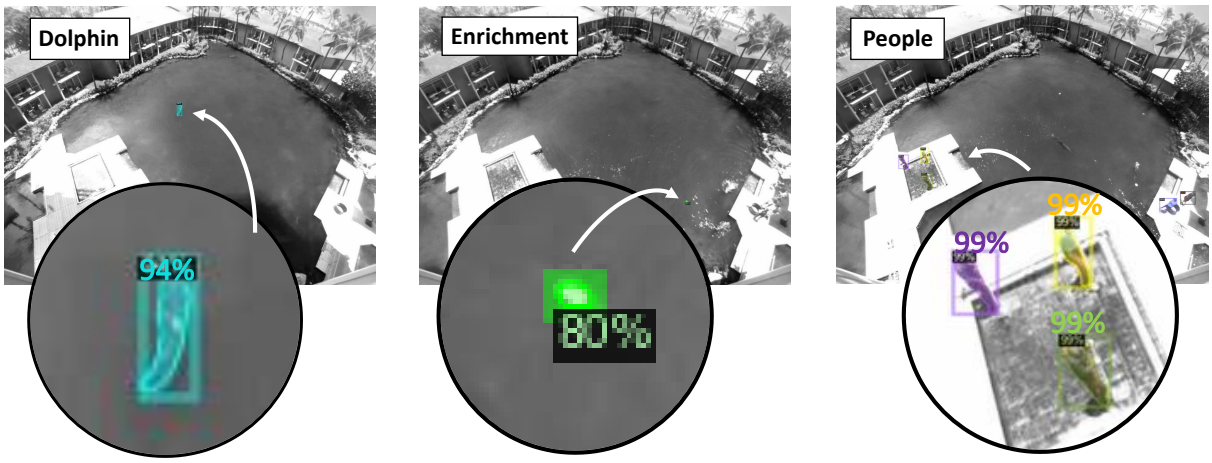


Figure 6.2: Example detections of a dolphin, floating enrichment (a buoy), and people in the lagoon environment. Each detected object is shown with its corresponding mask and the likelihood of detection.

Half of the labeled data were used to train the Mask R-CNN model, and the other half were used to test the classification results. The assessment results of the prediction confidence coefficient for the three detectors were greater than 80% (Figure 6.2). The camera frame rate was down sampled to 5 Hz, and the detectors were used to process 24 data sets collected in August, September, and October 2020.

The position of each detected object was estimated as the geometric center point of its mask and was transformed to correct for fisheye lens distortions using the Fisheye Camera Calibration tools in MATLAB. The corrected detections were mapped from the camera reference frame to the world reference frame by implementing a perspective transformation homography [147], and then used for the kinematic analysis.

6.2.3 Tracking of Targeted Group and Individuals

The world reference frame detections were initially independent for each image frame. To extract dynamic information from the targeted group (i.e. dolphins, floats or people), continuous tracklets were generated by applying a prediction-association constant velocity Kalman filter framework to the detections across video frames (Figure 6.3), as previously described in Section 4.3.3 from Chapter 4. This prediction-association process forms a tracklet by identifying a detection in the next frame that could be associated with the current tracklet while updating its belief about the target's current status. Kinetic information like speed, heading, and heading rate were extracted from each tracklet, and further used in the analysis of the targeted group's behavior. Note that individual identities were not distinguished during the group level analysis.

The continuous 3D track of each targeted (tagged) individual was further generated using data

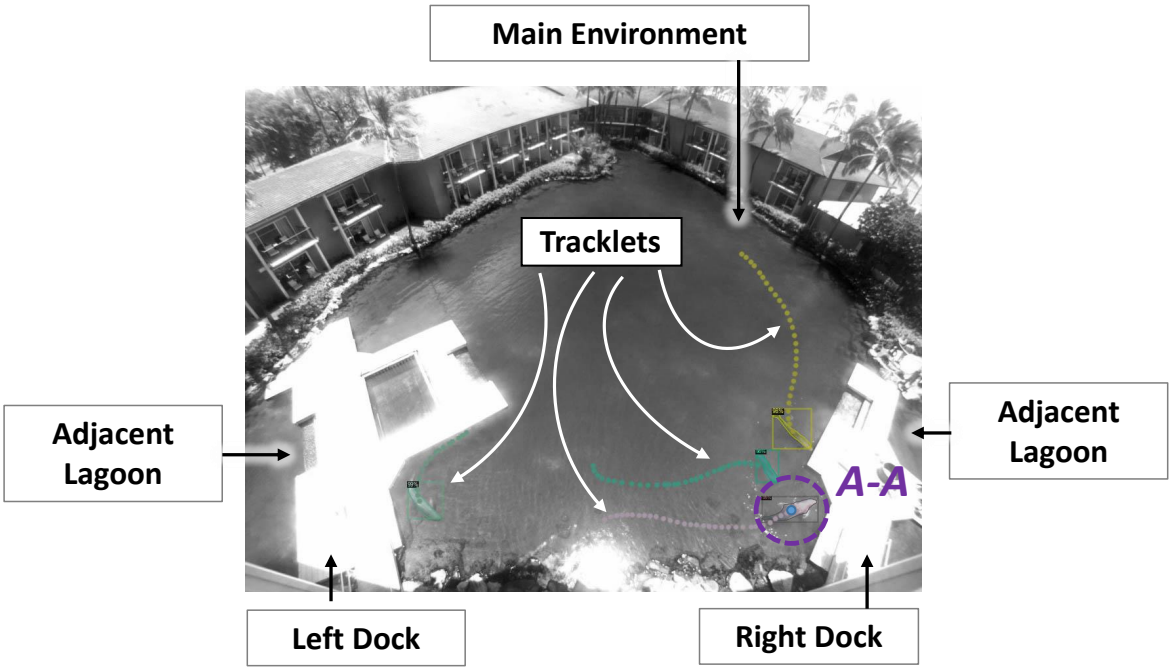


Figure 6.3: Detections and tracklets are shown for four animals in the lagoon.

from both a biologging tag and a camera via the particle filter based fusion framework presented in Chapter 4. This track provides complete kinetic information of the targeted individual for each trial. As will be discussed, the 3D tracks provide important information about the animal's motion, such as how the animal was regulating its motion to surface at similar locations in every lap.

6.2.4 Movement Synthesis in the Environment

6.2.4.1 Occupancy Heatmaps and Flow in the Environment

Occupancy heatmaps were use to characterize where the animals tended to spend time during the experiments [209]. The heatmaps were generated by integrating the occupancy probability distribution for all detections onto a blank map. The maximum occupancy density was then scaled to 1 for each heatmap, resulting in a relative habitat use for each condition. Percentage Use of the habitat was then calculated as:

$$\text{Percentage Use} = \frac{A_{(0,1]}}{A}, \tag{6.1}$$

where, A denotes the whole lagoon area and $A_{(0,1]}$ is the area where detections occurred (*Probability* greater than 0). Quiver plots were used to visualize the flow of the moving animals [147]. The environment was first subdivided into individual 2 m $\times$ 2 m cells. Vector headings for all tracklets that fell within an individual cell were then summed. If the animals tended to travel in the same

direction as they passed through a cell, the summation increased and the resulting flow vector exhibited a greater magnitude. Similarly, if the animals rarely moved through a cell or if they move through a cell in many different directions, the resulting magnitude of the flow was small.

6.2.4.2 Probability Density

Probability density functions for speed and heading rate were numerically determined for each experimental condition. A two-sample Kolmogorov-Smirnov (K-S) test [217] was used to compare the numerically derived distributions for movement in the presence of people and enrichment with the baseline condition. Comparing the density functions provides insight into trends present in the animals' movements during different environmental conditions. The K-S test statistic returns the maximum absolute difference between the empirical cumulative distributions of the two compared datasets. **Critical values** were determined using a 0.05 significance level, and the numbers of data points used to generate the density functions were: 416,257 for the baseline condition, 213,631 for the enrichment, and 196,185 when people were present. If a K-S statistic is greater than its critical value, the null hypothesis that the two datasets share the same distribution is rejected.

6.2.5 Motion States and Swimming Statistics of Tagged Individuals

The tagged animals' motion states (Chapter 3) were divided into two categories: the **Steer** category and the **Pose** category. The **Steer** category includes 7 empirical classes describing the steering status of the animal:

1. sharp turn left
2. medium turn left
3. cruising left
4. straight line swim
5. cruising right
6. medium turn right
7. sharp turn right

And the **Pose** category includes 5 empirical classes describing the posing status of the animal:

1. rolled upside down, lean towards the left side
2. rolled sideways towards the left side
3. neutral without significant rolling
4. rolled sideways towards the right side
5. rolled upside down, lean towards the right side

The transition matrices and distribution frequencies of these classes were calculated for each tagged animal for each trial. Further, we define **Steer Consistency** as the probability of the animal's current state and previous state sharing the same **Steer** class. And **Pose Consistency** is the probability of the animal's current state and previous state sharing the same **Pose** class.

In addition to the state transitions, the fluking frequency and amplitude of the tagged animals were obtained using methods presented in Chapter 2. Overall Dynamic Body Acceleration (ODBA) [218] was calculated as an indicator for activity level of each animal during the trials. The traveling speed and surface time duration were also measured and reported.

6.2.6 Engaging with Enrichment

A scoring system was used to evaluate how attractive an enrichment introduced into the environment was for the dolphins. Data from four trails, two with buoys and two with jolly balls, were examined for this analysis. The number of observed interactions and the time taken by the animals to engage with the enrichment (i.e. **waiting time**), were used to parameterize enrichment attractiveness. Using this parameter we would expect animals to engage with an attractive piece of enrichment quickly and frequently, while a less attractive object will induce a longer **waiting time** before an interaction. Waiting time was calculated as the difference between the start time of an enrichment tracklet and the time when an animal tracklet passes within 2 m of the enrichment. The example in Figure 6.4 illustrates a single interaction where the waiting time is the difference between t_1 and t_5 for the orange enrichment tracklet. To summarize the data, the hour long sessions were partitioned into eighteen 200 s bins. Average waiting time was calculated for all animal–enrichment interactions within that bin. Only enrichment tracklets with recorded interactions were included in the average.

6.3 Results

6.3.1 General Group Kinematic

During the experiment, ~9.5 h of baseline data and ~10 h of data with enrichment were collected. Over the course of all the trials, people were present in the environment for ~4.5 h, ~1.5 h of which involved no enrichment. The remaining ~3 h of data had both enrichment and people present, and was excluded from the analysis. During this time, there were around 825,000 detections of the three objects in the environment. Figure 6.5-A presents the flow of the animals and indicates a general counter-clockwise movement during all three conditions. The pattern of movement does vary between conditions, with a less defined center of rotation when enrichment or people were

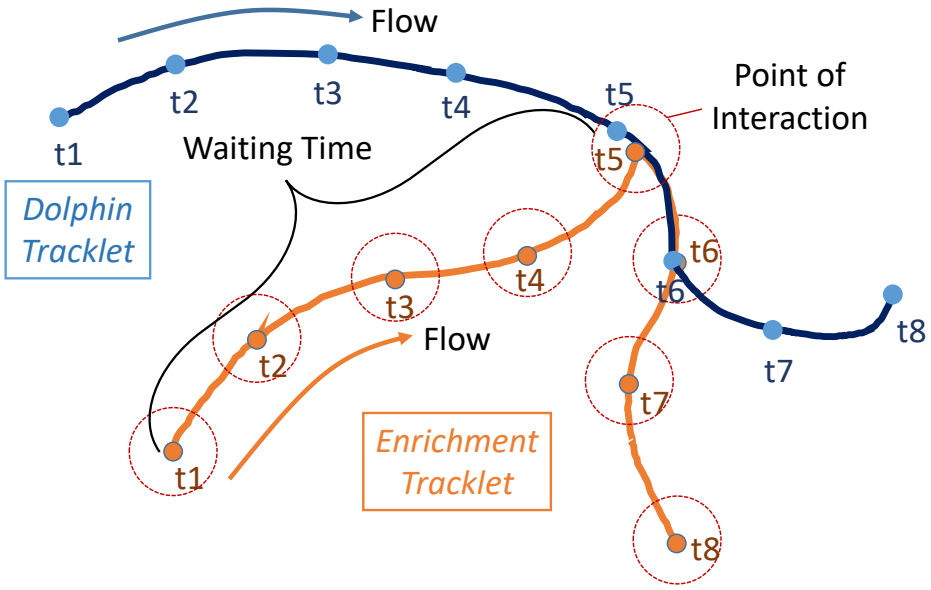


Figure 6.4: An illustration of how waiting time was defined for the enrichment. Eight time steps are shown for a dolphin (blue) and a piece of enrichment (orange). A point of interaction occurs at t_5 when the dolphin detection is within 2 m of the enrichment at the same time. The waiting time is then defined as the time from the initial detection of the animal in its current tracklet (t_1) to the point of interaction (t_5).

present. During the experiment, the animals had a preference for areas in the environment near the docks for all three conditions, Figure 6.5-B. The normalized probability of an animal visiting a location was highest (indicated by the yellow areas) near both docks during the baseline and people conditions, and near the right dock when enrichment was present. During the baseline condition, the animals had a larger and more defined circular area of use in the center of the environment, indicating more time spent in the circular counter-clockwise motion present in the flow. The percentage use of the environment was also the largest for the baseline condition, with an 89.3%, 72.6%, and 79.7% use for the baseline, enrichment, and people conditions, respectively.

The probability density functions calculated for speed and heading rate are presented in Figures 6.5-C&D. All three conditions had mean heading rates of 5 deg/s, confirming the counter-clockwise motion observed in the flow. Mean speed during the baseline condition (1.2 m/s) was higher than when either enrichment or people were present in the environment. The peaks in the density functions at the lower speeds (>0.5 m/s) indicate periods with little forward motion. This first peak is highest when people were present and lowest during the baseline condition. The density for baseline speed was used for comparison with the other two conditions (people and enrichment) in the two-sample Kolmogorov-Smirnov test. The resulting difference between the cumulative distribution of the baseline condition and the enrichment condition was 0.07, while the difference with people present was 0.08. The differences with the baseline distribution were

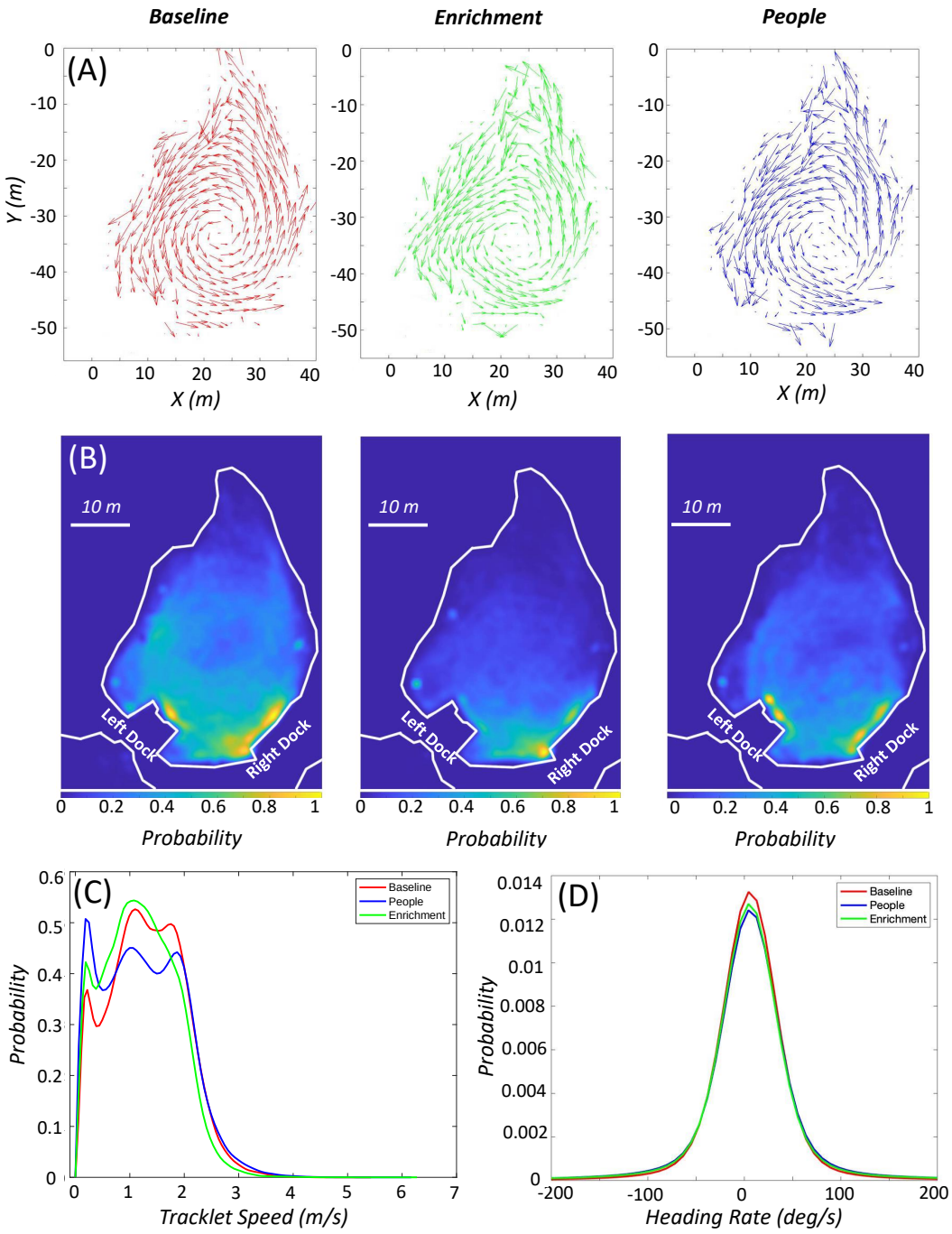


Figure 6.5: Results comparing the three environmental conditions: baseline, enrichment, and people. (**A**) Quiver plots for all three conditions show counter-clockwise movement in the lagoon, with the most concentric pattern occurring during the baseline condition. (**B**) Position heatmaps for the conditions show how the presence of enrichment or people modified where the animals spent time in the environment. (**C**, **D**) Probability density functions for the speed and heading rate of the animals derived from the tracklets. The animals tended to swim faster during the baseline condition, and spent more time stationary when people where present in the environment. Mean heading rates for all three conditions were positive, indicating counter-clockwise swimming trend.

Table 6.1: Summary measures calculated from the tracklets for the three conditions (baseline, enrichment, and people). Mean, SD, and the results of the K-S tests are reported. The enrichment and people conditions are compared with Baseline in the K-S tests (Section 6.2.4.2).

Parameter (Condition)	Mean	Standard Deviation	K-S Test statistic	Critical Values
Speed (Baseline)	1.20 m/s	0.74	0	0
Speed (Enrichment)	1.01 m/s	0.73	0.07	0.0036
Speed (People)	1.09 m/s	0.81	0.08	0.0037
Heading Rate (Baseline)	5.63 deg/s	43.01	0	0
Heading Rate (Enrichment)	5.33 deg/s	54.65	0.03	0.0036
Heading Rate (People)	5.94 deg/s	55.18	0.04	0.0037

Table 6.2: Statistic of the tracklets' speed and heading rate for the first 20%, middle 20%, and last 20% time sessions under the enrichment condition. All three sessions were compared with the Baseline condition in the K-S tests (Section 6.2.4.2).

Parameter (Condition)	Mean	Standard Deviation	K-S Test statistic	Critical Values
Speed (First 20%)	1.00 m/s	0.57	0.21	0.0064
Speed (Middle 20%)	1.20 m/s	0.69	0.06	0.0058
Speed (Last 20%)	1.33 m/s	0.72	0.03	0.0062
Heading Rate (First 20%)	4.26 deg/s	54.89	0.07	0.0064
Heading Rate (Middle 20%)	5.56 deg/s	74.04	0.05	0.0058
Heading Rate (Last 20%)	5.95 deg/s	52.77	0.01	0.0062

small for both conditions, but both test statistics are greater than the corresponding critical values (Table 6.1, Section 6.2.4.2).

6.3.2 Group Enrichment Engagement

The first 20%, middle 20%, and final 20% of the enrichment trials were examined to investigate how movement patterns change over time with enrichment present in the environment. The ~2 h of data for each period were used to generate quiver plots and heatmaps, Figure 6.6. The motion of the animals was more variable when enrichment was first introduced, with a more defined counter-clockwise movement pattern returning by the last 20 % of the trials. The heatmaps indicate that the animals were more likely to be present in the bottom half of the lagoon between the docks during the first 20%, with hotspots near the right dock as the trials progressed. By the last 20% of the trials the circulatory motion of the animals was again visible in the heatmap. The percentage use of the environment for the three periods were 75.33%, 61.79%, 67.06%. The mean speed of the animals increased from 1 m/s to 1.33 m/s over the course of the trials. The mean heading rate was around 5 deg/s for all three periods, Table 6.2.

The density functions for the three periods are compared to the baseline distribution in Fig-

ure 6.6-C&D. The probability distribution for the speed during the first 20% of the enrichment trials is visibly different from the other three, with slower overall speeds and a higher probability of slow forward motion (>0.5 m/s). The density for baseline speed was again used for comparison with the other two conditions in the two-sample Kolmogorov-Smirnov test. The K-S tests indicate that as the enrichment trials progressed, the probability distributions for speed and heading rate both became more similar to the baseline distributions, Table 6.2. All K-S test statistics for speed exceeded the critical values, but the difference between the cumulative distribution of the baseline condition and the first 20% was the largest at 0.21, and decreased to 0.03 for the last 20% of the trials. The K-S test statistics for the heading rate were all small and ranged from 0.07 to 0.01, but all exceeded their corresponding critical values.

Waiting time was used to further investigate interactions between the animals and enrichment. A 4-trial subset of the data, two hours with buoys and two hours with jolly balls, was used to evaluated engagement with the enrichment. Figure 6.7 presents waiting time and interaction counts for both types of enrichment. The average waiting time for the buoys was 2.0 s (4.7 SD), with the average waiting time per bin never exceeding 10 s. However, the number of interactions decreased throughout the trial, beginning with interaction counts as high as 20, but ending with counts between 1 and 5 for the 200-second-wide bins. The waiting time for the jolly balls had a higher average, 22.1 s (41.7 SD), and was more variable, with times ranging from 0.2 to 193.6 s. The number of interactions with the jolly balls was lower than the buoys (42 vs. 144), but the same decreasing trend is present in the data.

6.3.3 Individual Behavior During Interactions with Conspecifics and Enrichment

For individual behavior analysis, data from 3 focal animals (TT01, TT02, and TT03) over 3 trials (Trial A, B, and C, 30 minutes each) were synthesised. Enrichment floats were available in the environment for Trials A and B. The behavior of animal TT01 was dominated by 3 types of conspecific interaction conditions for the 3 trials: (1) solo for Trial A; (2) paired swim (with TT04) for Trial B; (3) paired swim and chase for Trial C. Animal TT01 stayed in the main environment with animals TT02, TT03, and TT04 (not tagged) for Trials A and B, and 3 other animals (not tagged) for Trial C. Neither TT02 nor TT03 engaged in prolonged conspecific interactions during the two trials in which they participated (Trials A and B).

The continuous 3D tracks of the targeted animals (TT01, TT02, and TT03) are shown in Figure 6.8 with light yellow indicating surfacings and dark blue indicating fully submerged/under water activities. The general track pattern of TT02 from Trial A matches that from Trial B, which consists of mainly circular motions with surfacing locations concentrating near the bottom right

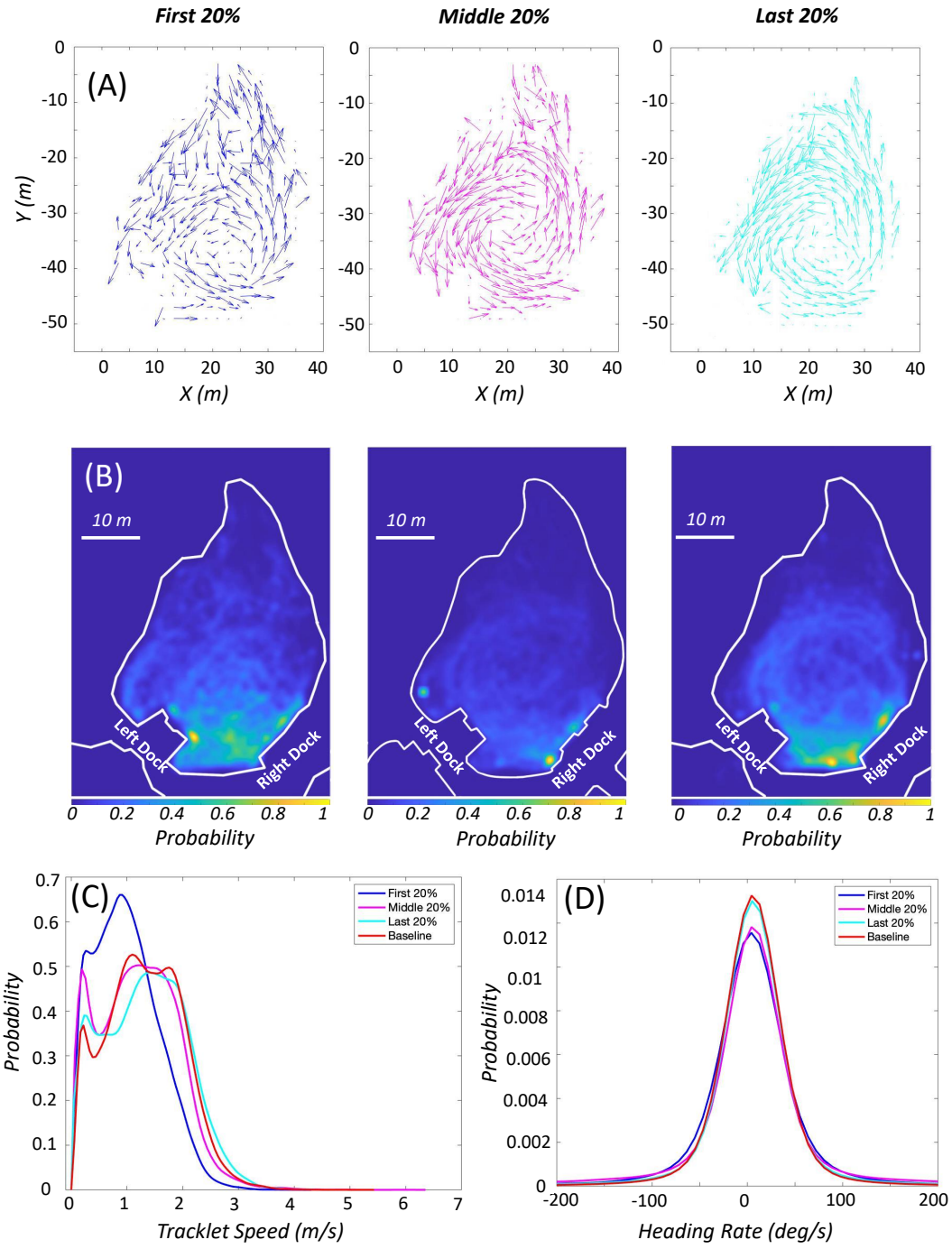


Figure 6.6: Animal movement when enrichment was present in the environment. (**A**) The quiver plots show a more varied movement pattern when enrichment is initially introduced. (**B**) The heatmaps also indicate that the use of the environment during the first 20% of the time with enrichment was more variable. For all three heatmaps, 1 in color bar represents the maximum occupancy density based on the corresponding conditions. (**C**, **D**) Probability density of tracklet speed and heading rate show that the animals tended to move faster during the middle and last 20%, and preference for counter-clockwise swimming was not as strong when the enrichment was initially introduced.

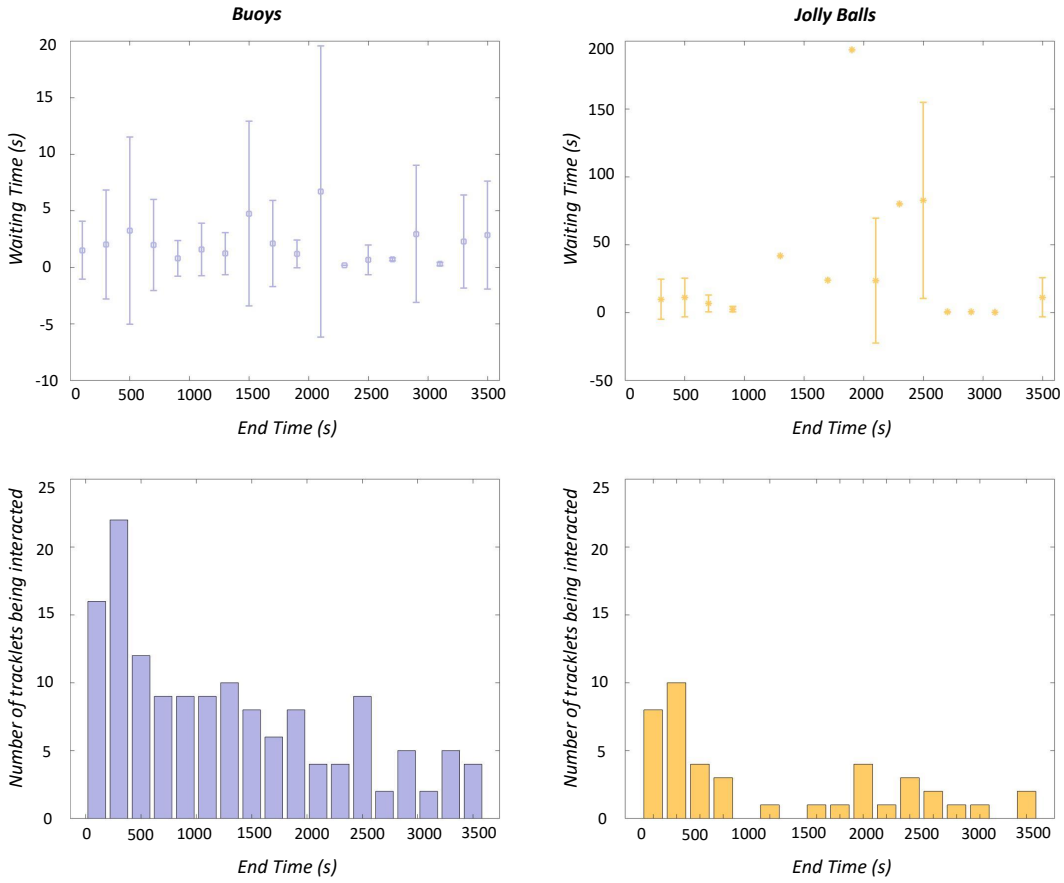


Figure 6.7: Average waiting time and SD (**top**) and number of enrichment interactions (**bottom**) are shown for each 200-second-wide bins for two types of enrichment (Buoys and Jolly Balls). The waiting time for buoys is lower than that for Jolly Balls in general (note that the y-axes for waiting time are on different scales). The numbers of interactions for both enrichments show decreasing trends as the trials progress.

edge of the environment. The pattern of TT03 from Trial A matches that from Trial B as well, but differs from the pattern significantly, and is more intertwined with surfacing locations spreading in the middle of the environment. For TT01, the pattern from Trial A resembles TT03, while patterns from Trials B and C resemble TT02.

The state distributions of the 3 animals over the trials are shown in Figure 6.9. The **Steer** of all 3 animals heavily favors left turns/cruises over right turns/cruises, indicating a counter-clockwise motion pattern in the environment. The **Steer** distributions of TT02 from Trial A quantitatively match those from Trial B, while the distributions of TT03 are consistent across Trial A and B as well. However, TT02's distribution is different from TT03's, with TT02's distribution favoring cruising left significantly more than TT03's. Similar to the 3D track patterns, TT01's state distribution matches TT03 for Trial A and TT02 for Trial B. In Trial C, TT01 demonstrates more medium left turns as compared to Trial B, which significantly changes the distribution. The **Pose** distributions from all 3 animals are similar in that all animals employ neutral poses more often than rolling to one side, yet differences still exist. TT02, when not having a neutral pose, would roll towards the left but not towards the right, while TT03 rolled both sides and even all the way to a upside down pose to have its belly pointing up. TT01 employed noticeably less un-neutral pose during the trials (Trials B and C) when it was interacting with other animals. Meanwhile, Figures 6.10 and 6.11 present the state transition matrices for the targeted animals over the trials. The most common transitions happen between a class and itself (i.e. the diagonal elements of the matrices). The bigger the diagonal elements are, the higher consistency the animal has demonstrated in its motion. The transition matrices together with the state distributions encode the motion pattern of the animal.

Other statistics of the targeted animals over each trial are presented in Table 6.3, including **Steer Consistency**, **Pose Consistency**, **Surface Time**, together with mean and SD of **Fluking Frequency**, **Fluking Amplitude**, **Speed**, and **ODBA**. TT02 demonstrates highest **Steer Consistency** (~89%) with TT03 being the lowest in the range 81% ~ 83%. The **Steer Consistency** of TT01 increases from 85% to ~89% over the trials. Meanwhile, the **Pose Consistency** for all 3 animals over all trials remains high with TT02 being the highest (~99%) and TT03 the lowest (~96%), and TT01 in the middle, showing an increasing trend over the trials. The **Fluking Frequency** of TT01 again shows an increasing trend over the trials while TT02 and TT03 share a similar frequency (0.62 ~ 0.68 Hz). The SD of **Fluking Amplitude** of TT03 during Trial B is noticeably higher than the other animals over all trials, while the **ODBA** of TT03 in Trial B is also significantly higher than the others. TT03 performs the lowest average swimming speed (~1.1 m/s) while TT02 maintains the highest average speed (~1.7 m/s) and TT01 demonstrates average speeds between TT02 and TT03. The **Surface Time** of TT01 in Trial A (21.6%) is similar to TT03 during Trial A and B (19~23.7%), but it dropped to 7.4% ~ 10.1% in Trial B and C for TT01. The

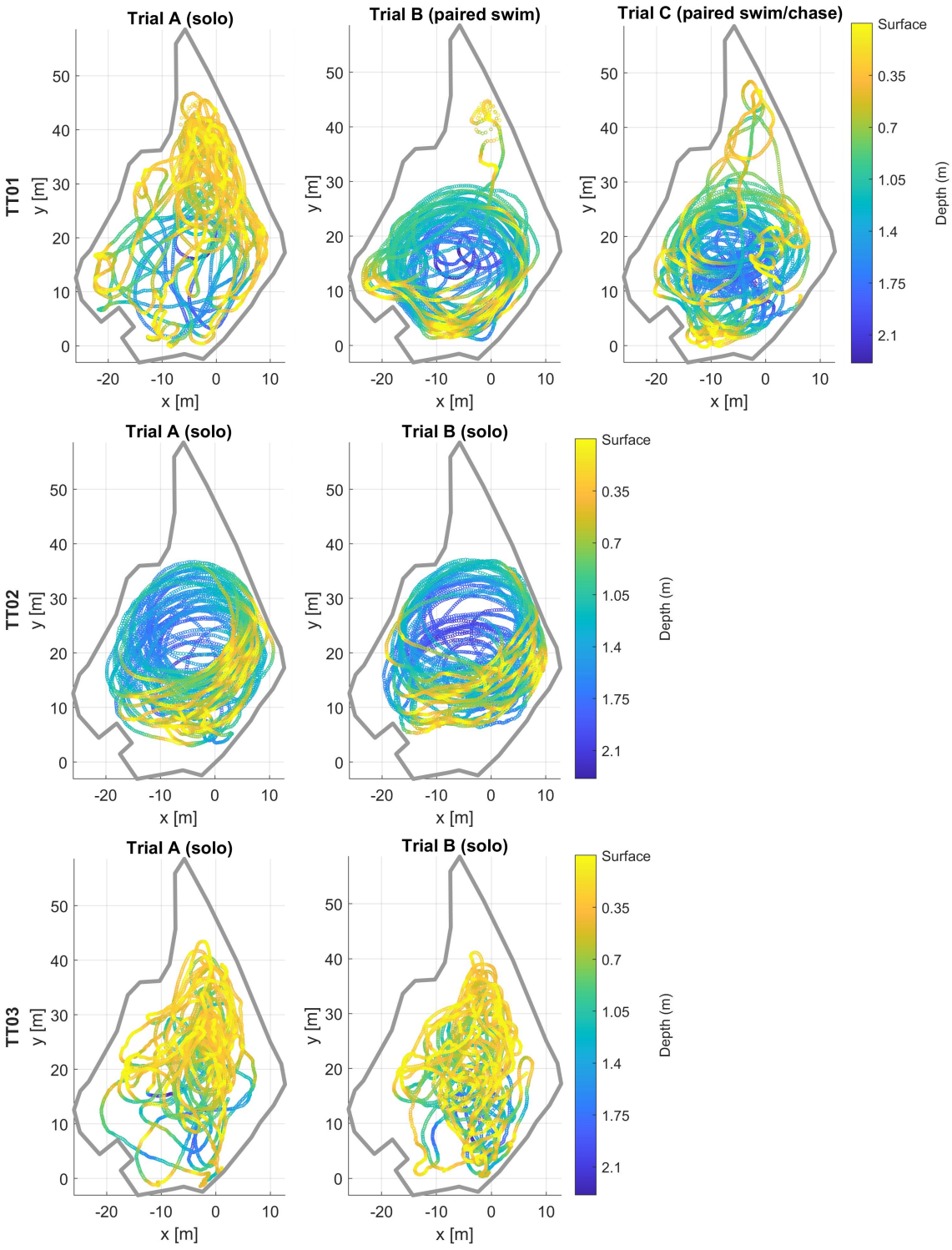


Figure 6.8: Continuous 3D tracks of 3 animals during 3 trials. Each trial lasted 30 minutes. The depth information of each animal is color coded, with light yellow indicates surfacing and dark blue for underwater locations.

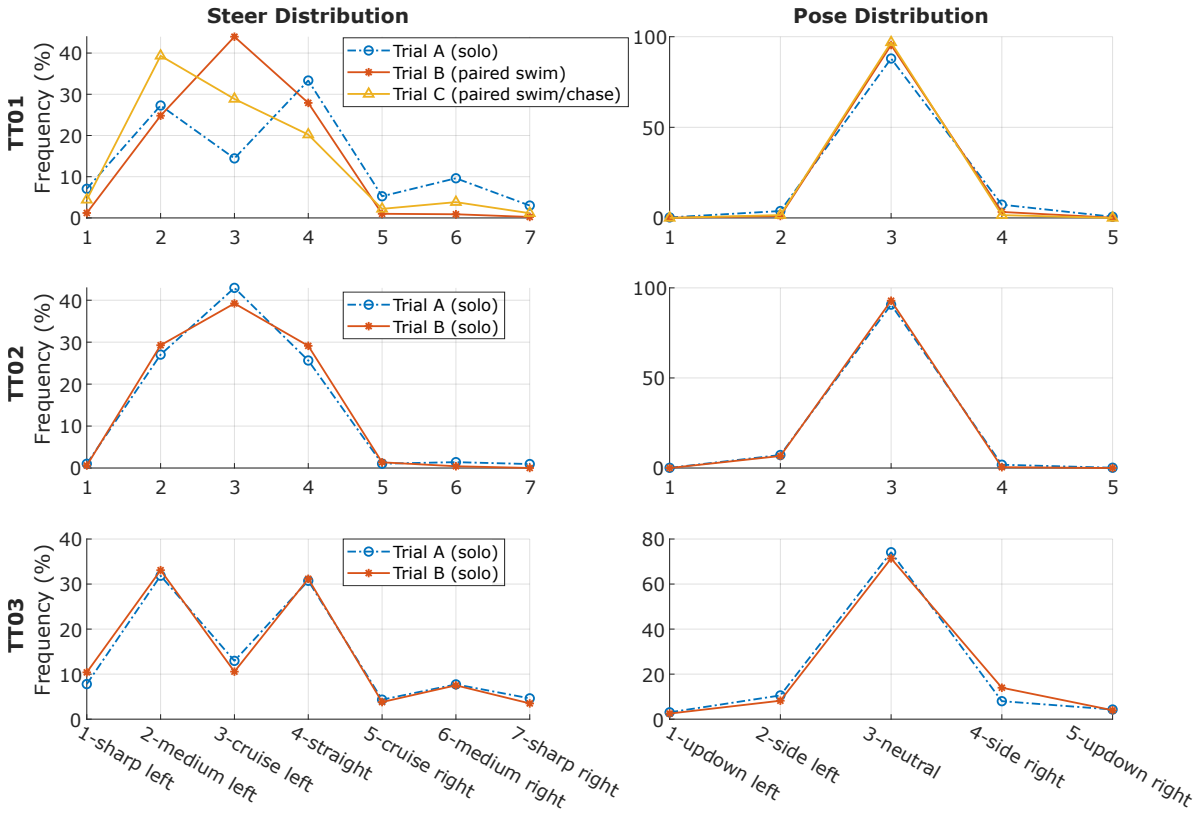


Figure 6.9: State distributions of three tagged animals during three focal trials. **Steer** distributions are shown on the left while **Pose** distributions are shown on the right.

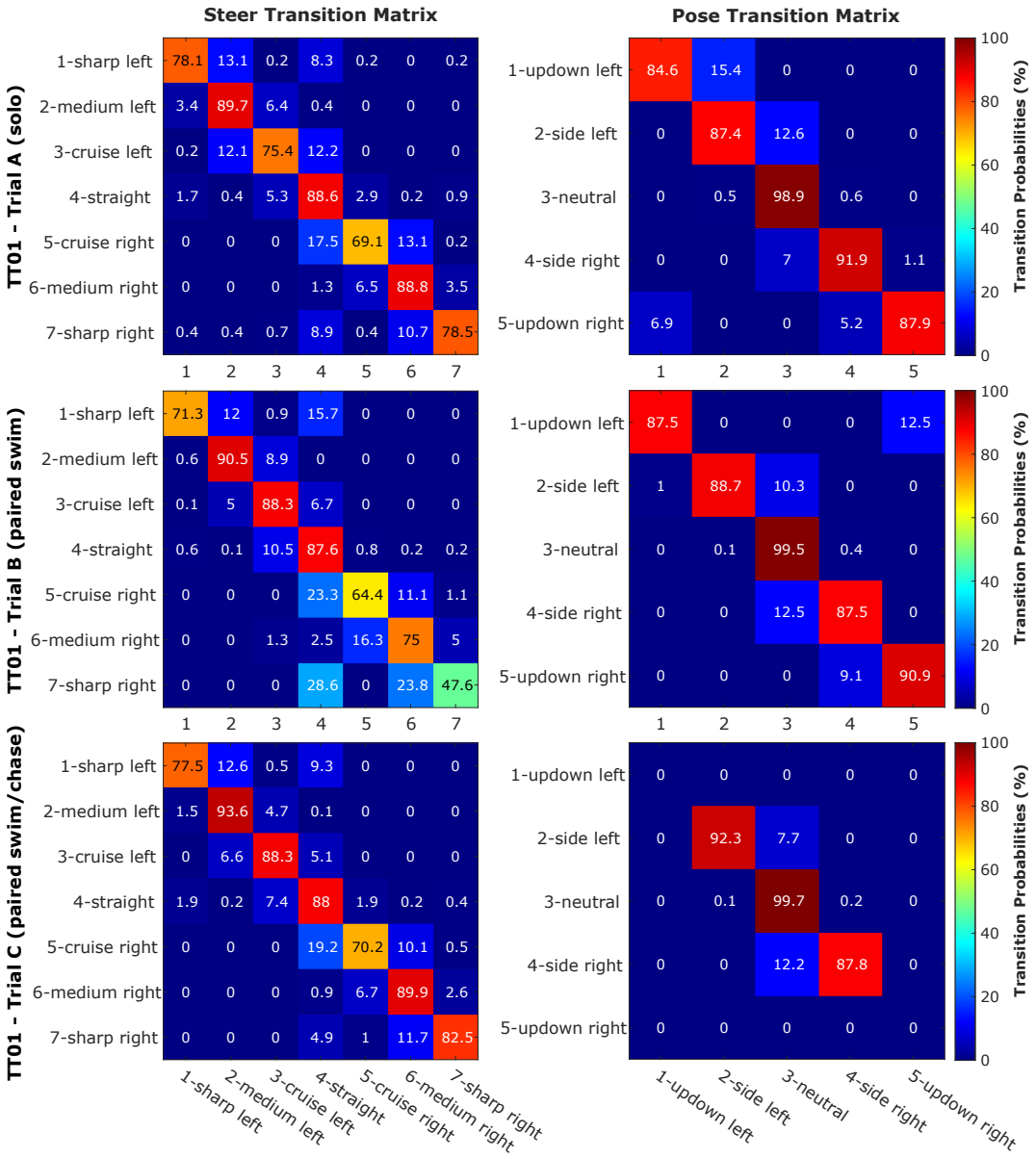


Figure 6.10: State transition matrices of animal TT01 during the three trials (Trial A, B, and C). **Steer** transitions are shown on the left while **Pose** transitions are shown on the right. The number/element at the i-th row and j-th column characterize the transition probability (%) from state i to state j, i.e. Eq. 3.5: $a_{ij} = P(X_{t+1} = c_j \mid X_t = c_i)$. The transitions of animals TT02 and TT03 are shown in Figure 6.11.

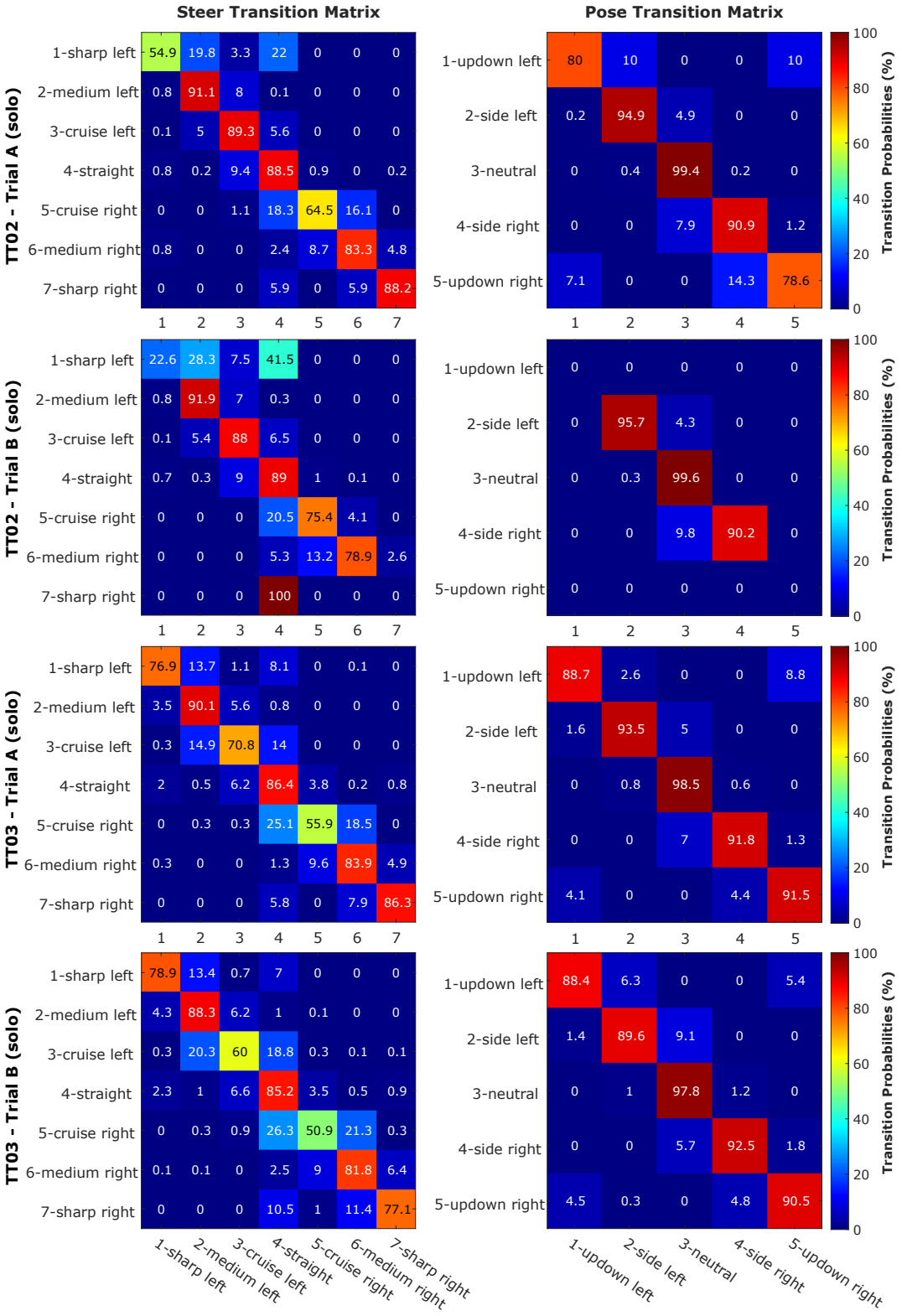


Figure 6.11: State transition matrices of animal TT02 and TT03 during Trial A and B. The transitions of animal TT01 are shown in Figure 6.10.

Table 6.3: The gait statistics of three tagged animals in three focal trials.

Statistics \ Animal ID	TT01			TT02		TT03	
	Trial A	Trial B	Trial C	Trial A	Trial B	Trial A	Trial B
Steer Consistency (%)	**85**	**88**	**89.5**	**88.9**	**89**	**83.3**	**81.1**
Pose Consistency (%)	**97.8**	**99**	**99.4**	**98.9**	**99.3**	**96.8**	**95.9**
Fluking Frequency	**0.58**	**0.64**	**0.73**	**0.68**	**0.63**	**0.62**	**0.66**
(Mean ± SD) (Hz)	(± 0.19)	(± 0.15)	(± 0.21)	(± 0.15)	(± 0.14)	(± 0.19)	(± 0.36)
Fluking Amplitude	**18.2**	**19.5**	**17.5**	**18.4**	**19.2**	**19.2**	**21.1**
(Mean ± SD) (Deg)	(± 7.9)	(± 7.1)	(± 7.3)	(± 7.2)	(± 7.4)	(± 10)	(± 14.3)
Speed	**1.3**	**1.6**	**1.5**	**1.7**	**1.6**	**1.1**	**1.1**
(Mean ± SD) (m/s)	(± 0.4)	(± 0.3)	(± 0.5)	(± 0.4)	(± 0.4)	(± 0.3)	(± 0.3)
ODBA	**2.3**	**2.3**	**2.3**	**2**	**1.8**	**2.6**	**3.2**
(Mean ± SD) (g/s)	(± 1.7)	(± 1.4)	(± 1.4)	(± 1.4)	(± 1.3)	(± 2.1)	(± 3.4)
Surface Time (%)	**21.6**	**7.4**	**10.1**	**5.8**	**5.3**	**19**	**23.7**

Surface Time for TT02 remains low (~5%) for both Trial A and Trial B.

6.4 Discussion

The ability to monitor how animals move in their environment while also identifying contextual factors that may influence movement is a challenging problem, especially when the animals are also engaged in conspecific interactions. In this work, we presented a computer vision and biologging based persistent monitoring and analysis framework to quantify bottlenose dolphin movement. Direct monitoring by trained behaviorists and observers provides important information about the animals in their environments. These observations are qualitative, and can be supplemented with information from cameras and other sensors placed in or around the environment. As sensing becomes integrated into more environments, the ability to efficiently and effectively extract information from sensor data will be key to maximizing the benefits of these systems. In this work, more than 21 h of data were examined, resulting in over 850,000 detected objects for the analysis. This detection density enabled the extraction of important animal kinematics (speed, heading, heading rate) that were used to characterize motion. If a trained observer was able to manually score one detection every second, it would take more than nine days of continuous work to label the entire dataset presented here. Further, the trained network can be used to classify objects from additional datasets, creating the opportunity to investigate behavior over the course of months or even different seasons.

To demonstrate the approach, we analyzed video data from four animals observed during hour

long unstructured swimming trials in a lagoon environment at midday. Enrichment was present for half the sessions, and animal care specialists were free to move in the environment during the recordings. Biologging tag data from 3 animals during 3 of the trials (Trial A and B had enrichment while C had not) were further included for the analysis of individual behavior. Qualitative differences from baseline are visible in the directional quiver plots and occupational heatmaps when enrichment or people were present. A general counter-clockwise flow was observed in the animal movement during all environmental conditions (i.e., baseline, enrichment, and people), Figure 6.5. This flow pattern is consistent with the previous observations that dolphins tend to swim in a counter-clockwise direction in the Northern Hemisphere when staying in a defined environment [219]. Although clockwise preferential swimming biases have been observed for dolphins in the Southern Hemisphere, it is speculated that the dominating flow direction is the result of global forces rather than individual dolphin anatomy [220].

Proximity to the docks occurred more frequently in the baseline condition than when enrichment was present. During baseline, the animals were observed swimming, interacting with another animal (e.g., paired swimming), or inspecting the adjacent lagoons separated by the floating docks. In the heatmaps, there were no people present on the docks, but the increased time spent near these areas could be driven by multiple reasons. It could be the result of their inspection of animals in neighboring lagoons, their interaction with the structure of the dock, an association of the docks with the trainers, or an anticipatory behavior. When enrichment was present the average speed decreased from 1.2 m/s (baseline) to 1.0 m/s, and the animals spent more time in the bottom half of the lagoon. Overall area of use also decreased from 89.3% to 72.6%. These changes in spacial distribution could be due to enrichment distribution in the lagoon. Enrichment was introduced from the docks, and wind would push the floating enrichment to the edges of the lagoon biasing distribution and the resulting animal movement.

The K-S statistics indicate that the speed and heading rate probability distributions when enrichment or people were present are different from the baseline condition, Table 6.1. Speed distributions have two sets of peaks: a first peak between 0 and 0.5 m/s and a second set of peaks between 0.5 and 4 m/s, Figure 6.5-C. The first peak indicates a distinct period of time when the animals were swimming slowly or were stationary. The highest first peak probability occurred when people were present in the environment. A higher probability might be expected because this condition includes periods of time at the start and end of the trials when the animals were at stations with the animal care specialists. Additional peaks in the probability functions around 1 m/s and 2 m/s are present during the baseline condition and when people are present. The distribution for the enrichment condition has a single peak around 1 m/s, which is in accordance with the overall drop in speed when enrichment is present. The probability distributions for the heading rate all indicate that the animals are turning counter-clockwise, confirming the direction of the flow in the quiver

plots, Figure 6.5-D.

Engagement with the enrichment was further investigated by calculating how often the animals interacted with the enrichment floats and how this changed over the course of the trial, Figures 6.6 and 6.7. When enrichment was introduced, the first 20% of the trial, there is less circularity in the movement flow and the animals tended to remain between the two docks in the bottom part of the lagoon. The K-S test statistic for speed distribution during the first 20% of the enrichment condition (0.21) shows the largest difference with baseline movement of and condition, Tables 6.1 and 6.2. However, as the trial progressed distributions for speed and heading rate reverted towards baseline behavior with K-S differences of 0.3 and 0.1, respectively. These temporal changes were also seen in the analysis of enrichment engagement.

Interactions between the animals and two types of enrichment floats (Buoys and Jolly Balls) were examined to demonstrate how this approach can be used to quantify animal movement and environmental features simultaneously. Engagement with the enrichment was quantified using interaction counts, and time duration between an interaction and the initial detection (waiting time). For the buoys, the average waiting time (2 s vs. 22 s) was shorter than for the Jolly Balls and the total number of interactions (144 vs. 42) was higher. The interaction counts for the first 20% of the trials (0–720 s) were the highest for both enrichment types, Buoys = ~56 and Jolly Balls = ~21, and dropped off considerably for the last 20% of the trials (2880–3600 s), Buoys = ~11 and Jolly Balls = ~4. These temporal changes could again be due to the distribution of the enrichment in the environment as wind tended to gradually blow the enrichment to the edges of the lagoon. The animals spend more time in the center of the lagoon environment, which reduces the likelihood of encountering enrichment at the edge of the environment. Additionally, accessing the enrichment at the shallow rocky edges of the lagoon could be more difficult for the animal, further reducing the opportunity for engagement. On the other hand, the animals 'allow' the enrichment to be slowly blown to the edges of the lagoon, which could indicate habituation to the specific enrichment. The two types of floating enrichment were physically similar when compared to other enrichment categories (e.g., food dispensing enrichment, trainer involved enrichment, etc., [221, 222, 223, 224]), yet the animals could still prefer one over the other. Jolly Balls are small and round with light colors that can be carried by mouth when swimming, while Buoys are larger and more difficult to bring underwater. The animals' variable response to the two types of floating enrichment could result from the different interaction mechanisms or simply the physical appearance of the objects. It is of great interest to extend this work to investigate and compare animals' response to other types of enrichment (e.g., food dispensing enrichment, trainer involved enrichment, etc., [221, 222, 223, 224]) as a means of detecting improvements in welfare.

The analysis of individual motion during enrichment or conspecific interaction (Section 6.3.3) provides significant further insight about animal behavior, in addition to the group level analysis

discussed. It is shown both qualitatively (Figure 6.8) and quantitatively (Figure 6.9) that individual behavior can be vastly different (e.g. TT02 vs. TT03) during the same trial where enrichment was available in the environment. Specifically, TT03 was visually observed as a 'playful' individual who demonstrated frequent enrichment interactions, that can be echoed by its significant longer **Surface Time** and higher activity level (**ODBA**) (Table 6.3) as well as its intertwined trajectory (Figure 6.8). In contrast, TT02 was known to its human animal care specialist as a dominant individual in the group and did not heavily engage with the enrichment. TT02 demonstrated mainly a cruising motion in combination with straight line swimming and medium turning (Figures 6.9 and 6.8). TT02's activity level (**ODBA**) was the lowest among the 3 tagged animals, yet its average swimming speed was the highest (Table 6.3). Hypothetically, swimming **and** playing with enrichment floats can result in higher measured activity levels. Meanwhile, TT02 also maintained the lowest SD of **Fluking Frequency** as well as the highest **Steer Consistency** and **Pose Consistency**, affirming that TT02 was more focused on doing routine swimming than other interactions. Noticeably, TT02 regulated its motion to have surfacing events near the bottom right edge of the environment (Figure 6.8), which could be due to an association of the docks with the trainers, or an anticipatory behavior.

Interestingly, TT01's behavior was similar to TT03's when TT01 was not paired with other animals (i.e. Trial A), and its behavior was similar to TT02's when it was engaged in paired swimming and chasing events (Trial B and C) with other animals. We thus speculate that coordinated behaviors, like paired swimming, require an individual to be more consistent in its motion and more predicable in its behavior. Further on this line, we speculate that consistency is a sign of social maturity, like what TT02 demonstrated, so other individuals can better coordinate with this specific individual when needed. In other words, if an intelligent entity (dolphin) derives internal models of other intelligent entities (other dolphins) during a coordinated event, then the more reliable the models are, the easier it is for this entity to make good decisions during the coordination. On the other hand, being more consistent and predicable as an intelligent entity makes it easier for other individuals in the group to build a reliable model and work efficiently with the specific entity.

The results of the computer vision and biologging based monitoring and analysis framework are promising, but there are limitations to consider. If different types of enrichment are introduced, new detectors will have to be trained using labeled data. Additionally, the detectors for this work were trained on data collected from the same time of day. To use these detectors with images from other periods of time, it may be necessary to retrain the network with images that cover a wide range of lighting conditions with the objects of interest. A single overhead camera mounted towards the bottom of the lagoon was used to collect the data for this work. A second camera placed to better observe the top half of the lagoon would improve the coverage of the environment.

6.5 Summary

A computer vision and biologging based persistent monitoring and analysis framework was presented to efficiently and effectively extract dolphin kinematics and track the movement of people and enrichment in the environment. This ability to simultaneously extract animal kinematics and the factors that may influence movement is an important complement to direct monitoring by trained behaviorists and observers. The framework was used to investigate bottlenose dolphin movement during three conditions: baseline, with enrichment and with people. Animal kinematics (position, speed, heading, heading rate) were automatically extracted from over 21 h of video footage, at a resolution that would be infeasible to label manually. Further, individual specific information (e.g. gait statistics, motion states, activity level, surfacing locations, and trajectory pattern) was extracted from the fusion of biologging tag and camera data. A general counterclockwise swimming flow was observed during the three conditions, and the presence of both enrichment and people in the environment did modify the observed kinematics. These changes were largest when enrichment was first introduced, with animal movement largely returned to baseline by the final 20% of the trials. Meanwhile, individual animal dynamical behavior can be altered significantly when interactions (e.g. paired swimming or chasing) with conspecifics are enabled. From the analysis, it appears that an individual dolphin's consistency in its motion increased during coordination. The proposed framework is flexible and can be implemented in other facilities where overhead video data can be collected or biologging tags can be deployed, enabling new data streams for the investigation of animal movement in these environments.

CHAPTER 7

Conclusions

7.1 Synthesis of Research Contributions and Broader Impacts

As an intelligent species, bottlenose dolphins demonstrate sophisticated coordinated behavior. But an understanding of the fine-scale dynamical behaviors during coordination was lacking, a result of technical challenges associated with persistent monitoring of the highly dynamic marine animals. The work presented in this dissertation addressed this gap by leveraging and advancing frameworks from machine learning, computer vision, sensor fusion, and data-driving modeling to intelligently parse data from biologging tags and cameras into interpretable information about an animal's dynamical status (i.e. gait, motion states, positions, and energetics), which then led to an investigation of animal behavior during interactions with conspecific and environmental features. Various aspects of the work done during the process of accomplishing this dissertation have been disseminated through a series of journal publications [102, 112, 153, 154, 203, 204, 147, 225, 226, 227, 228, 229, 230], peer-reviewed conference papers [111, 146, 152], and conference abstract presentations [231, 232, 233, 234, 235].

In Chapter 2, we systematically presented an automated data processing framework (and software) that took advantage of the common characteristics of pose and gait of the animal to process tag data and analyze the animals' gait. Three aspects were investigated in particular: (1) The identification of time instances associated with occurrences of relative motion between tag and animal. (2) The identification of the relative orientation of tag wrt the animal's body for a given data segment. (3) The extraction of gait parameters that are invariant to pose and tag orientation. Biologging tag data from bottlenose dolphins, a humpback whale, and a beluga whale were used to validate and demonstrate the approach. Results indicated that the average relative orientation error of the tag wrt the dolphin's body after processing were within 10 degrees in roll, pitch and yaw directions. The average precision and recall for identifying relative tag motion were 0.87 and 0.89, respectively. Examples of the resulting pose and gait analysis demonstrated the potential of this approach in not only gait analysis but also animal behavioral studies. The proposed analysis

approach and software will facilitate the use of biologging tags for the study of cetacean locomotion and behavior. The method and software can be directly used with cetacean data from any tag platform that uses accelerometer, magnetometer, and pressure sensor.

In Chapter 3, we first introduced a systematic approach for feature generation, feature selection and state classification from biologging sensor data, and demonstrated how the proposed framework can be used to accurately classify bottlenose dolphin swimming states when only sparse motion observations were available for training. The proposed framework addressed questions about how to generate and select features from biologging tag data, and provided a systematic approach for efficient window size selection. Sequential temporal information was also used to improve classification accuracy via a HMM during state classification, resulting in 4.8% (GMM), 8.3% (NB), 4.0% (SVM), 2.4% (KNN), 8.1% (DT) improvements when compared with single-frame classification schemes.

We then proposed an LSTM-HMM framework that achieved above 90% classification accuracy by heavily utilizing sequential temporal information that was embedded in the time series signal at both a point-to-point bottom level (LSTM) and a state-to-state top level (HMM). Importantly, the LSTM-HMM framework did not need statistical feature engineering and only requires a small amount of training data to achieve reasonable classification accuracy, making this approach easily adaptable to new tasks. Also, as compared with an LSTM-LSTM structure, the 'white box' at the top level of the LSTM-HMM framework made it possible to apply the model trained on one dataset to another dataset with different state distributions while leaving the state inferences interpretable. Depending on the application scenario, feature engineering and LSTM can both be of great help in practice. The heavy focus posed on the transitions between motion states is capable of providing valuable intuition to more biologically driven questions, since it is often of interest to understand why a certain motion state arises. Moreover, the hybrid frameworks can be of use in cases where there are complex high level behaviors that encompass several low level motion states. The HMM step can identify loosely defined sequential information to draw conclusions about higher level behaviors.

In Chapter 4, we showcased a framework that approached the tracking problem from a localization perspective, utilizing a combination of biologging tags and stationary cameras. Specifically, this framework dealt with the situation where the dynamic entities/agents of interest were outfitted with biologging tags and primarily subsist underwater, while the cameras were mounted above water. Our results demonstrated that the platform is readily implementable in managed environments, and is able to handle nonlinear agent dynamics, non-Gaussian noise, and scattered camera observations while maintaining robust localization and tracking ability. The method maintains the identity of the tracked agents comparing to camera-only tracking techniques, while it is significantly more accurate than biologging-only dead-reckoning tracks. This method can be applied to

cover the general tracking scenario where sensor-augmented entities cannot be continuously monitored. With appropriate adjustment, the framework can be used to track other marine animals in an aquarium environment or in the wild if a drone is available above the environment [10]. In addition, the framework also has good potential in the land based applications where GPS signal is unavailable or not sufficiently accurate.

In Chapter 5, we presented a data-driven physics based approach to estimate gliding drag coefficients and the mechanical energy cost of transport (MECOT) of bottlenose dolphins during horizontal intermittent locomotion. Kinematic measurements of animal motion were collected with a customized non-invasive biologging tag, and Fluke-and-Glide (FG) gait was modeled as a two-phase process. Additional experiments and CFD simulations were introduced to asses the hydrodynamic impact of biologging tags. Our results suggested that computational and experimental methods can be used in a complementary manner to determine the increase in drag associated with biologging tags, and the hydrodynamic impact of a single tag is quantifiably insignificant. While quantifying the forces created by transmitting and archival tags used for animal science is an essential part of reducing the impact of these instruments, the body of work seeking to assess the impacts of tags across taxa is growing [97, 98, 158, 191, 201, 202]. Our work provides a framework that can be used to support these studies. Additionally, the computational approach used to estimate the drag and quantify the uncertainty of the estimate from the experimental data can be applied to other data sets collected from animals in both controlled and free-swimming environments. Meanwhile, the results presented in Chapter 5 indicate that horizontal FG gait, in addition to porpoising/leaping [80, 81, 200], wave riding [82], drafting [83] and gliding during diving [95, 96, 42], is a behavioral mechanism that can enhance the locomotion efficiency of swimming dolphins.

In Chapter 6, we explored how a computer vision and biologging based persistent monitoring and analysis framework can be used to efficiently and effectively extract dolphin kinematics, and track the movement of people and enrichment in the environment. This ability to simultaneously extract animal kinematics and the factors that may influence movement is an important complement to direct monitoring by trained behaviorists and observers. The framework was used to investigate bottlenose dolphin movement during three conditions: baseline, with enrichment and with people. Animal kinematics (position, speed, heading, heading rate) were automatically extracted from over 21 h of video footage, at a resolution that would be infeasible to label manually. Further, individual specific information (e.g. gait statistics, motion states, activity level, surfacing locations, and trajectory pattern) was extracted from the fusion of biologging tag and camera data. A general counter-clockwise swimming flow was observed during the three conditions, and the presence of both enrichment and people in the environment did modify the observed kinematics. These changes were largest when enrichment was first introduced, with animal movement largely returned to baseline by the final 20% of the trials. Meanwhile, individual animal dynamical behavior

can be altered significantly when interactions (e.g. paired swimming or chasing) with conspecifics are enabled. From the analysis, it appears that an individual dolphin's consistency in its motion increased during coordination. The proposed framework is flexible and can be implemented in other facilities where overhead video data can be collected or biologging tags can be deployed, enabling new data streams for the investigation of animal movement in these environments.

7.2 Future Directions

While this dissertation is reaching to its end, the exploration and study on the associated subjects do not stop. Here we discuss the future directions of the work presented in this dissertation.

After the data processing introduced in Chapter 2, the relative orientation of the tag was correctly aligned with the animal. But the actual location (e.g. back vs. tail) of the tag can still affect tag measurements and results, particularly the estimated fluking amplitude of the animal. For example, a tag located near the tail is almost surely going to have higher estimated fluking amplitude than a tag located near the dorsal fin for the same gait of an animal. Even though this would not affect our ability to draw good conclusions about the animals, as discussed in Chapter 2, it is still important to systematically explore the sensing capability of the tag vs. tag location on the animal. Via both analytical derivations and experimental validations, a more complete picture can be provided for the gait analysis of cetaceans with biologging tags.

Based on Chapter 3, further study needs to be conducted to understand why window size selection affected classification performance. It could be similar to the idea of 'resonance frequency', where a particular window size fits better with the motion signal. Future work shall include: (1) Investigating how to identify the best window size and feature set for individual motion classes, as opposed to using one window and feature set for the entire set of classes. (2) Further testing and validation of classification performance when the algorithm has been trained using data sets with labeled information from a particular animal, but applied to new data that were not in the original training sets (i.e. data from different trials of the same animal or from different animals). That is, validating the framework with a larger dataset to explore the limits of the frameworks under extreme conditions.

Even though Chapter 4 demonstrated promising results in the prolonged identity preserving tracking of tagged animal, the quality of the track can be further improved with two more steps. First, a heuristic tracklet selection method can be included to select the most likely tracklets that are associated with the estimated track of a tagged agent. Then, the entire track, with unconnected sections of tracklets and continuous tag data, can be smoothed with an efficient optimization technique as discussed in our previous work [152]. Tracks with both high frequency and precision can be obtained as a result of this implementation.

As an extension of the data-driven modeling and investigation of the Fluke-and-Glide gait presented in Chapter 5, the on-body tag measurements could be combined with a dedicated model of the animal's full body kinematics to provide thrust and drag estimates under different scenarios, which could provide more detailed insights about the locomotion of these animals. Future work shall also use tag data from dolphins in the wild to investigate the cost of transport associated with swimming gaits used in a different environmental context.

Chapter 6 merely started peeking into the fine-scale dynamics of bottlenose dolphin behavior during interactions with conspecific animals and environmental features. Many exciting work can be done following this line. For example, the coordinated behavior investigated here (i.e. paired swimming) is a rather 'simple' one, it is of great interest to investigate more sophisticated coordinated behavior. To do so, we can design tasks that require coordination among the animals to achieve and observe how they form the coordination using the IA frameworks presented in this dissertation, Figure 7.1. Further, energetic reasoning needs to be considered while understanding the behavior of the animals, Figure 7.1-Right. Incorporating audio signals during coordination [32] into the analysis can provide a more complete picture of the behavior as well. Finally, making the connection between dolphin coordinations and artificial multi-agent systems will most likely provide valuable insights to both sides, Figure 7.1-Left.

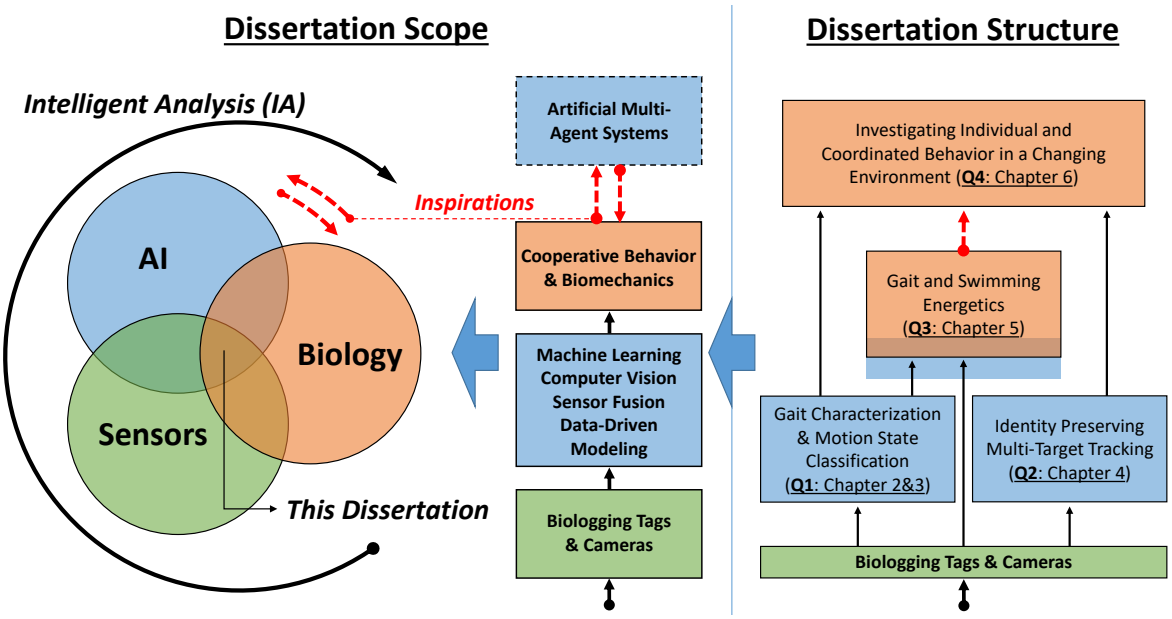


Figure 7.1: The scope (**left**) and structure (**right**) of this dissertation, modified from Figure 1.1 from Chapter 1. This work brings us one step closer to the further goal of studying bio-inspired artificial multi-agent systems by achieving the solid black arrows in the chart. Yet dashed red arrows represent a connection that has not been explicitly made in this dissertation, but is important to be made in the future. Particularly, the energetic reasoning (right side) needs to be considered while analyzing the coordinated behavior. And making the connection between dolphin coordination and artificial multi-agent systems (left side) can be very insightful to both fields.

BIBLIOGRAPHY

[1] Y. N. Harari, *Sapiens: A Brief History of Humankind*. HarperCollins, 2015.
(Back reference page: 1)

[2] J. P. Guilford, *The nature of human intelligence*. McGraw-Hill, 1967.
(Back reference page: 1)

[3] E. L. Thorndike and D. Bruce, *Animal Intelligence: Experimental Studies*. Routledge, 2017. https://doi.org/10.4324/9781351321044
(Back reference page: 1)

[4] D. Poole, A. Mackworth, and R. Goebel, *Computational Intelligence: A Logical Approach*. New York: Oxford University Press, 1998.
(Back reference page: 1)

[5] S. J. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Pearson Education, Inc., 2010.
(Back reference page: 1)

[6] S. J. Shettleworth, *Cognition, Evolution and Behavior*, 2nd ed. New York: Oxford University Press, 2010.
(Back reference page: 1)

[7] A. M. Helmenstine, "10 Most Intelligent Animals," 2020. https://www.thoughtco.com/most-intelligent-animals-4157712
(Back reference page: 1)

[8] D. J. T. Sumpter, *Collective Animal Behavior*. Princeton University Press, 2010.
(Back reference page: 1)

[9] C. Boesch and H. Boesch, "Hunting Behavior of Wild Chimpanzees in the Tai National Park," *American journal of physical anthropology*, vol. 78, no. 4, pp. 547–573, 1989. https://doi.org/10.1002/ajpa.1330780410
(Back reference page: 1)

[10] S. L. King, R. C. Connor, M. Krützen, and S. J. Allen, "Cooperation-based concept formation in male bottlenose dolphins," *Nature Communications*, vol. 12, p. 2373, 2021. http://dx.doi.org/10.1038/s41467-021-22668-1
(Back reference pages: 1, 4, 8, and 133)

[11] D. R. Stahler, D. W. Smith, and D. S. Guernsey, "Foraging and Feeding Ecology of the Gray Wolf (Canis lupus): Lessons from Yellowstone National Park, Wyoming, USA," *The Journal of Nutrition*, vol. 136, no. 7, pp. 1923–1926, 2006. https://doi.org/10.1093/jn/136.7.1923S
(Back reference page: 1)

[12] D. R. MacNulty, D. W. Smith, L. D. Mech, J. A. Vucetich, and C. Packer, "Nonlinear effects of group size on the success of wolves hunting elk," *Behavioral Ecology*, vol. 23, no. 1, pp. 75–82, 2011. https://doi.org/10.1093/beheco/arr159
(Back reference page: 1)

[13] U. G. K. Wegst, H. Bai, E. Saiz, A. P. Tomsia, and R. O. Ritchie, "Bioinspired structural materials," *Nature Materials*, vol. 14, pp. 23–36, 2015. https://doi.org/10.1126/science.1245842
(Back reference page: 2)

[14] S. Mirjalili, S. M. Mohammad, and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014. http://dx.doi.org/10.1016/j.advengsoft.2013.12.007
(Back reference page: 2)

[15] S. Mirjalili and A. Lewis, "The Whale Optimization Algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016. https://doi.org/10.1016/j.advengsoft.2016.01.008
(Back reference page: 2)

[16] X. Fan, W. Sayers, S. Zhang, Z. Han, L. Ren, and H. Chizari, "Review and Classification of Bio-inspired Algorithms and Their Applications," *Journal of Bionic Engineering*, vol. 17, pp. 611–631, 2020. https://doi.org/10.1007/s42235-020-0049-9
(Back reference page: 2)

[17] S.-J. Chung and M. Dorothy, "Neurobiologically Inspired Control of Engineered Flapping Flight," *Journal of Guidance, Control, and Dynamics*, vol. 33, no. 2, pp. 440–453, 2010. https://doi.org/10.2514/1.45311
(Back reference page: 2)

[18] R. Pfeifer, M. Lungarella, F. Iida, and B.-i. E. Systems, "Self-Organization, Embodiment, and Biologically Inspired Robotics," *Science*, vol. 318, no. 5853, pp. 1088–1093, 2007. https://doi.org/10.1126/science.1145803
(Back reference page: 2)

[19] J. Werfel, K. Petersen, and R. Nagpal, "Designing Collective Behavior in a Termite-Inspired Robot Construction Team," *Science*, vol. 343, no. 6172, pp. 754–758, 2014. https://doi.org/10.1126/science.1245842
(Back reference page: 2)

[20] P. Leitao, J. Barbosa, and D. Trentesaux, "Bio-inspired multi-agent systems for reconfigurable manufacturing systems," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 5, pp. 934–944, 2012. https://doi.org/10.1016/j.engappai.2011.09.025
(Back reference page: 2)

[21] M. Wooldridge, *An introduction to multiagent systems*, 2nd ed. John Wiley & Sons, 2009.
(Back reference page: 2)

[22] J. Y. Wang, "Bottlenose Dolphin, Tursiops Aduncus, Indo-Pacific Bottlenose Dolphin," in *Encyclopedia of Marine Mammals*, 3rd ed., B. Würsig, J. Thewissen, and K. M. Kovacs, Eds. Elsevier Inc., 2018, pp. 125–130. https://doi.org/10.1016/B978-0-12-804327-1.00073-X
(Back reference page: 4)

[23] J. E. Reynolds III, R. S. Wells, and S. D. Eide, *The Bottlenose Dolphin: Biology and Conservation*. University Press of Florida, 2000.
(Back reference page: 4)

[24] R. S. Wells and M. D. Scott, "Bottlenose Dolphin, Tursiops Truncatus, Common Bottlenose Dolphin," in *Encyclopedia of Marine Mammals*, 3rd ed., B. Würsig, J. Thewissen, and K. M. Kovacs, Eds. Elsevier Inc., 2018, pp. 118–125. https://doi.org/10.1016/B978-0-12-804327-1.00072-8
(Back reference pages: 4, 5)

[25] R. C. Connor, M. R. Heithaus, and L. M. Barre, "Superalliance of bottlenose dolphins," *Nature*, vol. 397, pp. 571–572, 1999. https://doi.org/10.1038/17501
(Back reference pages: 4, 5)

[26] R. C. Connor, W. R. Cioffi, S. Randić, S. J. Allen, J. Watson-capps, and M. Krützen, "Male alliance behaviour and mating access varies with habitat in a dolphin social network," *Scientific Reports*, vol. 7, p. 46354, 2017. https://doi.org/10.1038/srep46354
(Back reference page: 4)

[27] R. C. Connor, "Dolphin social intelligence: Complex alliance relationships in bottlenose dolphins and a consideration of selective environments for extreme brain size evolution in mammals," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 362, no. 1480, pp. 587–602, 2007. https://doi.org/10.1098/rstb.2006.1997
(Back reference page: 4)

[28] R. Vaughn, B. Würsig, and J. Packard, "Dolphin prey herding : Prey ball mobility relative to dolphin group and prey ball sizes , multispecies associates , and feeding duration," *Marine Mammal Science*, vol. 26, no. 1, pp. 213–225, 2010. https://doi.org/10.1111/j.1748-7692.2009.00317.x
(Back reference page: 5)

[29] L. G. Torres and A. J. Read, "Where to catch a fish? the influence of foraging tactics on the ecology of bottlenose dolphins (Tursiops truncatus) in Florida Bay, Florida," *Marine Mammal Science*, vol. 25, no. 4, pp. 797–815, 2009. https://doi.org/10.1111/j.1748-7692.2009.00297.x
(Back reference page: 5)

[30] J. R. Zaeschmar, S. L. Dwyer, and K. A. Stockin, "Rare observations of false killer whales (Pseudorca crassidens) cooperatively feeding with common bottlenose dolphins (Tursiops truncatus) in the Hauraki Gulf, New Zealand," *Marine Mammal Science*, vol. 29, no. 3, pp. 555–562, 2013. https://doi.org/10.1111/j.1748-7692.2012.00582.x
(Back reference page: 5)

[31] F. G. Daura-Jorge, M. Cantor, S. N. Ingram, D. Lusseau, and P. C. Simões-Lopes, "The structure of a bottlenose dolphin society is coupled to a unique foraging cooperation with artisanal fishermen," *Biology Letters*, vol. 8, no. 5, pp. 702–705, 2012. https://doi.org/10.1098/rsbl.2012.0174
(Back reference page: 5)

[32] S. L. King, E. Guarino, K. Donegan, C. McMullen, and K. Jaakkola, "Evidence that bottlenose dolphins can communicate with vocal signals to solve a cooperative task," *Royal Society Open Science*, vol. 8, no. 3, p. 202073, 2021. https://doi.org/10.1098/rsos.202073
(Back reference pages: 5, 135)

[33] P. Martin and P. Bateson, *Measuring Behaviour: An Introductory Guide*, 3rd ed. Cambridge University Press, 2007.
(Back reference pages: 5, 31)

[34] S. H. Shane, R. S. Wells, and B. Würsig, "Ecology, behaviour and organization of the bottlenose dolphin: A review," *Marine Mammal Science*, vol. 2, no. 1, pp. 34–63, 1986. https://doi.org/10.1111/j.1748-7692.1986.tb00026.x
(Back reference page: 5)

[35] B. Würsig and M. Würsig, "Behavior and ecology of the dusky dolphin, Lagenorhynchus obscurus, in the South Atlantic," *Fishery Bulletin*, vol. 77, no. 4, pp. 871–890, 1980.
(Back reference page: 5)

[36] M. J. Cremer, A. C. Holz, C. M. Sartori, B. Schulze, R. L. Paitach, and P. C. Simões-Lopes, "Behavior and Ecology of Endangered Species Living Together: Long-Term Monitoring of Resident Sympatric Dolphin Populations," in *Advances in Marine Vertebrate Research in Latin America*, M. R. Rossi-Santos and C. W. Finkl, Eds. Springer, 2018, vol. 22, pp. 477–508. https://doi.org/10.1007/978-3-319-56985-7{_}17
(Back reference page: 5)

[37] C. Rutz and G. C. Hays, "New frontiers in biologging science," *Biology Letters*, vol. 5, no. 3, pp. 289–292, 2009. https://doi.org/10.1098/rsbl.2009.0089
(Back reference page: 6)

[38] H. J. Williams, L. A. Taylor, S. Benhamou, A. I. Bijleveld, T. A. Clay, S. de Grissac, U. Demšar, H. M. English, N. Franconi, A. Gómez-Laich, R. C. Griffiths, W. P. Kay, J. M. Morales, J. R. Potts, K. F. Rogerson, C. Rutz, A. Spelt, A. M. Trevail, R. P. Wilson, and L. Börger, "Optimizing the use of biologgers for movement ecology research," *Journal of Animal Ecology*, vol. 89, no. 1, pp. 186–206, 2020. https://doi.org/10.1111/1365-2656.13094
(Back reference page: 6)

[39] 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. https://doi.org/10.1109/JOE.2002.808212
(Back reference pages: 6, 12, 13, 15, 20, 23, 28, 31, and 108)

[40] 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, 2005. https://doi.org/10.1242/jeb.01327
(Back reference pages: 6, 12)

[41] P. J. Miller, M. P. Johnson, and P. L. Tyack, "Sperm whale behaviour indicates the use of echolocation click buzzes 'creaks' in prey capture," *Proceedings of the Royal Society B: Biological Sciences*, vol. 271, no. 1554, pp. 2239–2247, 2004. https://doi.org/10.1098/rspb.2004.2863
(Back reference pages: 6, 12)

[42] R. Skrovan, T. M. Williams, P. S. Berry, P. W. Moore, and R. W. Davis, "The Diving Physiology of Bottlenose Dolphins," *Journal of Experimental Biology*, vol. 202, no. 20, pp. 2749–2761, 1999. https://doi.org/10.1242/jeb.202.20.2749
(Back reference pages: 6, 9, 12, 76, 106, and 133)

[43] P. J. Miller, M. P. Johnson, P. L. Tyack, and E. A. Terray, "Swimming gaits, passive drag and buoyancy of diving sperm whales Physeter macrocephalus," *Journal of Experimental Biology*, vol. 207, no. 11, pp. 1953–1967, 2004. https://doi.org/10.1242/jeb.00993
(Back reference pages: 6, 9, 12, 13, 21, and 76)

[44] K. A. 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, 2017. https://doi.org/10.1111/mms.12408
(Back reference pages: 6, 12, 13, 21, and 101)

[45] B. R. Mate, Kelly A. Rossbach, Sharon L. Nieukirk, R. S. Wells, A. B. Irvine, M. D. Scott, and A. J. Read, "Satellite-monitored movements and dive behavior of a bottlenose dolphin (Tursiops truncatus) in Tampa Bay, Florida," *Marine Mammal Science*, vol. 11, no. 4, pp. 452–463, 1995. https://doi.org/10.1111/j.1748-7692.1995.tb00669.x
(Back reference pages: 6, 12, 101, and 102)

[46] P. L. Tyack, M. Johnson, N. A. Soto, A. Sturlese, and P. T. Madsen, "Extreme diving of beaked whales," *Journal of Experimental Biology*, vol. 209, no. 21, pp. 4238–4253, 2006. https://doi.org/10.1242/jeb.02505
(Back reference pages: 6, 12, 13, 20, and 21)

[47] M. Johnson, P. T. Madsen, W. M. Zimmer, N. Aguilar De Soto, and P. L. Tyack, "Beaked whales echolocate on prey," *Proceedings of the Royal Society B: Biological Sciences*, vol. 271, no. SUPPL. 6, pp. 383–386, 2004. https://doi.org/10.1098/rsbl.2004.0208
(Back reference pages: 6, 12)

[48] S. L. Watwood, P. J. Miller, M. Johnson, P. T. Madsen, and P. L. Tyack, "Deep-diving foraging behaviour of sperm whales (Physeter macrocephalus)," *Journal of Animal Ecology*, vol. 75, no. 3, pp. 814–825, 2006. https://doi.org/10.1111/j.1365-2656.2006.01101.x
(Back reference pages: 6, 12)

[49] T. A. Patterson, L. Thomas, C. Wilcox, O. Ovaskainen, and J. Matthiopoulos, "State-space models of individual animal movement," *Trends in Ecology and Evolution*, vol. 23, no. 2, pp. 87–94, 2008. https://doi.org/10.1016/j.tree.2007.10.009
(Back reference pages: 6, 12)

[50] C. Laplanche, T. A. Marques, and L. Thomas, "Tracking marine mammals in 3D using electronic tag data," *Methods in Ecology and Evolution*, vol. 6, no. 9, pp. 987–996, 2015. https://doi.org/10.1111/2041-210X.12373
(Back reference pages: 6, 12)

[51] 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, p. 31, 2015. http://dx.doi.org/10.1186/s40462-015-0061-6
(Back reference pages: 6, 7, 12, and 59)

[52] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. The MIT Press, 2005. https://mitpress.mit.edu/books/probabilistic-robotics
(Back reference pages: 7, 59)

[53] H. Hu and D. Gu, "Landmark-based navigation of industrial mobile robots," *Industrial Robot*, vol. 27, no. 6, pp. 458–467, 2000. https://doi.org/10.1108/01439910010378879
(Back reference pages: 7, 59)

[54] S. Subedi, G. R. Kwon, S. Shin, S. S. Hwang, and J. Y. Pyun, "Beacon based indoor positioning system using weighted centroid localization approach," in *International Conference on Ubiquitous and Future Networks (ICUFN)*. IEEE, 2016, pp. 1016–1019. https://doi.org/10.1109/ICUFN.2016.7536951
(Back reference pages: 7, 59)

[55] S. Wang, D. Gu, L. Chen, and H. Hu, "Single Beacon-Based Localization with Constraints and Unknown Initial Poses," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 4, pp. 2229–2241, 2016. https://doi.org/10.1109/TIE.2015.2499253
(Back reference pages: 7, 59)

[56] C. N. Taylor and D. Mohler, "A study of particle filtering approaches for the kidnapped robot problem," in *Signal Processing, Sensor/Information Fusion, and Target Recognition XXVII*, Orlando, Florida, United States, 2018, p. 106460A. https://doi.org/10.1117/12.2305181
(Back reference pages: 7, 59)

[57] M. W. Gamini Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, and M. Csorba, "A solution to the simultaneous localization and map building (SLAM) problem," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 3, pp. 229–241, 2001. https://doi.org/10.1109/70.938381
(Back reference pages: 7, 59)

[58] R. Mur-Artal, J. M. Montiel, and J. D. Tardos, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," *IEEE Transactions on Robotics*, vol. 31, no. 5, pp. 1147–1163, 2015. https://doi.org/10.1109/TRO.2015.2463671
(Back reference pages: 7, 59)

[59] R. Mur-Artal and J. D. Tardos, "ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017. https://doi.org/10.1109/TRO.2017.2705103
(Back reference pages: 7, 59)

[60] T. G. Lang, "Speed, Power, And Drag Measurements of Dolphins and Porpoises," in *Swimming and Flying in Nature*, T. Y.-T. Wu, C. J. Brokaw, and C. Brennen, Eds. Boston: Springer, 1975, pp. 553–572. https://doi.org/10.1007/978-1-4757-1326-8{_}5
(Back reference pages: 8, 9, and 76)

[61] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015. https://doi.org/10.1038/nature14539
(Back reference pages: 8, 32, and 58)

[62] 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, 2017. https://doi.org/10.1109/TPAMI.2016.2577031
(Back reference pages: 8, 58, 62, and 108)

[63] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 386–397, 2020. https://doi.org/10.1109/TPAMI.2018.2844175
(Back reference pages: 8, 58, 108, and 110)

[64] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *IEEE conference on computer vision and pattern recognition (CVPR)*, 2016, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91
(Back reference pages: 8, 58)

[65] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *arXiv*, pp. 1–6, 2018. https://arxiv.org/abs/1804.02767
(Back reference pages: 8, 58)

[66] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *European Conference on Computer Vision (ECCV)*, vol. 9905 LNCS. Springer, 2016, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0{_}2
(Back reference pages: 8, 58)

[67] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 936–944. https://doi.org/10.1109/CVPR.2017.106
(Back reference pages: 8, 58)

[68] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, 2020. https://doi.org/10.1109/TPAMI.2018.2858826
(Back reference pages: 8, 58)

[69] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, 2001. https://doi.org/10.1002/0471221279
(Back reference pages: 8, 58)

[70] D. B. Reid, "An algorithm for tracking multiple targets," *IEEE Transactions on Automatic Control*, vol. 24, no. 6, pp. 843–854, 1979. https://doi.org/10.1109/TAC.1979.1102177
(Back reference pages: 8, 58)

[71] B.-S. Yaakov, Ed., *Multitarget-multisensor tracking: advanced applications*. Norwood: Artech House, Inc, 1990.
(Back reference pages: 8, 58)

[72] S. S. Blackman, "Multiple hypothesis tracking for multiple target tracking," *IEEE Aerospace and Electronic Systems Magazine*, vol. 19, no. 1, pp. 5–18, 2004. https://doi.org/10.1109/MAES.2004.1263228
(Back reference pages: 8, 58)

[73] C. Kim, F. Li, A. Ciptadi, and J. M. Rehg, "Multiple hypothesis tracking revisited," in *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2015, pp. 4696–4704. https://doi.org/10.1109/ICCV.2015.533
(Back reference pages: 8, 58)

[74] L. Zhang, Y. Li, and R. Nevatia, "Global data association for multi-object tracking using network flows," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008, pp. 1–8. https://doi.org/10.1109/CVPR.2008.4587584
(Back reference pages: 8, 58)

[75] H. B. Shitrit, J. Berclaz, F. Fleuret, and P. Fua, "Multi-commodity network flow for tracking multiple people," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1614–1627, 2014. https://doi.org/10.1109/TPAMI.2013.210
(Back reference pages: 8, 59)

[76] S. D. Feldkamp, "Swimming in the California sea lion: morphometrics, drag and energetics." *The Journal of Experimental Biology*, vol. 131, pp. 117–135, 1987. https://doi.org/10.1242/jeb.131.1.117
(Back reference pages: 9, 75, and 76)

[77] H. Weimerskirch, J. Martin, Y. Clerquin, P. Alexandre, and S. Jiraskova, "Energy saving in flight formation," *Nature*, vol. 413, pp. 697–698, 2001. https://doi.org/10.1038/35099670
(Back reference pages: 9, 75)

[78] A. C. Gleiss, S. J. Jorgensen, N. Liebsch, J. E. Sala, B. Norman, G. C. Hays, F. Quintana, E. Grundy, C. Campagna, A. W. Trites, B. A. Block, and R. P. Wilson, "Convergent evolution in locomotory patterns of flying and swimming animals," *Nature Communications*, vol. 2, no. 1, 2011. https://doi.org/10.1038/ncomms1350
(Back reference pages: 9, 75, and 76)

[79] H. Hertel, *Structure, form, movement*. New York: Reinhold Publishing Co., 1966.
(Back reference pages: 9, 75, 76, 82, and 89)

[80] F. E. Fish and C. A. Hui, "Dolphin swimming-a review," *Mammal Review*, vol. 21, no. 4, pp. 181–195, 1991. https://doi.org/10.1111/j.1365-2907.1991.tb00292.x
(Back reference pages: 9, 75, 103, 106, and 133)

[81] D. Au and D. Weihs, "At high speeds dolphins save energy by leaping," *Nature*, vol. 284, pp. 548–550, 1980. https://doi.org/10.1038/284548a0
(Back reference pages: 9, 75, 106, and 133)

[82] 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, pp. 821–823, 1992. https://doi.org/10.1038/355821a0
(Back reference pages: 9, 75, 94, 106, and 133)

[83] D. Weihs, "The hydrodynamics of dolphin drafting," *Journal of Biology*, vol. 3, no. 8, pp. 1–16, 2004. https://doi.org/10.1186/jbiol2
(Back reference pages: 9, 75, 106, and 133)

[84] D. L. Kramer and R. L. McLaughlin, "The behavioral ecology of intermittent locomotion," *American Zoologist*, vol. 41, no. 2, pp. 137–153, 2001. https://doi.org/10.1093/icb/41.2.137
(Back reference pages: 9, 76)

[85] B. W. Tobalske, "Morphology, velocity, and intermittent flight in birds," *American Zoologist*, vol. 41, no. 2, pp. 177–187, 2001. https://doi.org/10.1093/icb/41.2.177
(Back reference pages: 9, 76)

[86] J. M. Rayner, P. W. Viscardi, S. Ward, and J. R. Speakman, "Aerodynamics and energetics of intermittent flight in birds," *American Zoologist*, vol. 41, no. 2, pp. 188–204, 2001. https://doi.org/10.1093/icb/41.2.188
(Back reference pages: 9, 76)

[87] D. Weihs, "Mechanically efficient swimming techniques for fish with negative buoyancy," *Journal of Marine Research*, vol. 31, no. 3, pp. 194–209, 1973. https://www.researchgate.net/publication/284548959{%}0AMechanically
(Back reference pages: 9, 76)

[88] F. E. Fish, J. F. Fegely, and C. J. Xanthopoulos, "Burst-and-coast swimming in schooling fish (Notemigonus crysoleucas) with implications for energy economy," *Comparative Biochemistry and Physiology – Part A: Physiology*, vol. 100, no. 3, pp. 633–637, 1991. https://doi.org/10.1016/0300-9629(91)90382-M
(Back reference pages: 9, 76)

[89] U. K. Müller, E. J. Stamhuis, and J. J. Videler, "Hydrodynamics of unsteady fish swimming and the effects of body size: Comparing the flow fields of fish larvae and adults," *Journal of Experimental Biology*, vol. 203, no. 2, pp. 193–206, 2000. https://doi.org/10.1242/jeb.203.2.193
(Back reference pages: 9, 76)

[90] M. J. Lighthill, "Large-amplitude elongated-body theory of fish locomotion," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 179, pp. 125–138, 1971. https://doi.org/10.1098/rspb.1971.0085
(Back reference pages: 9, 76, 101, and 103)

[91] E. Akoz and K. W. Moored, "Unsteady propulsion by an intermittent swimming gait," *Journal of Fluid Mechanics*, vol. 834, pp. 149–172, 2018. https://doi.org/doi:10.1017/jfm.2017.731
(Back reference pages: 9, 76, and 101)

[92] D. Weihs, "Energetic advantages of burst swimming of fish," *Journal of Theoretical Biology*, vol. 48, no. 1, pp. 215–229, 1974. https://doi.org/10.1016/0022-5193(74)90192-1
(Back reference pages: 9, 76, and 101)

[93] J. J. Videler and D. Weihs, "Energetic advantages of burst-and-coast swimming of fish at high speeds." *Journal of Experimental Biology*, vol. 97, no. 1, pp. 169–178, 1982. https://doi.org/10.1242/jeb.97.1.169
(Back reference pages: 9, 76)

[94] I. Ashraf, S. van Wassenbergh, and S. Verma, "Burst-and-coast swimming is not always energetically beneficial in fish (Hemigrammus bleheri)," *Bioinspiration and Biomimetics*, vol. 16, no. 1, p. 016002, 2020. https://doi.org/10.1088/1748-3190/abb521
(Back reference pages: 9, 76)

[95] T. M. Williams, "Intermittent swimming by mammals: A strategy for increasing energetic efficiency during diving," *American Zoologist*, vol. 41, no. 2, pp. 166–176, 2001. https://doi.org/10.1093/icb/41.2.166
(Back reference pages: 9, 76, 106, and 133)

[96] T. M. Williams, R. W. Davis, L. A. Fuiman, J. Francis, B. J. Le Boeuf, M. Horning, J. Calambokidis, and D. A. Croll, "Sink or swim: Strategies for cost-efficient diving by marine mammals," *Science*, vol. 288, no. 5463, pp. 133–136, 2000. https://doi.org/10.1126/science.288.5463.133
(Back reference pages: 9, 76, 106, and 133)

[97] 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, 2014. https://doi.org/10.1242/jeb.108225
(Back reference pages: 9, 13, 76, 77, 103, 106, and 133)

[98] 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, p. 465, 2018. https://doi.org/10.3389/fmars.2018.00465
(Back reference pages: 9, 13, 76, 77, 103, 106, 108, and 133)

[99] P. E. Purves, W. H. Dudok van Heel, and A. Jonk, "Locomotion in dolphins Part I: Hydrodynamic experiments on a model of the bottle-nosed dolphin, Tursiops truncatus, (Mont.)," pp. 5–31, 1975. https://www.aquaticmammalsjournal.org/index.php?option=com{_}content{&}view=article{&}id=54{&}Itemid=34
(Back reference pages: 9, 76)

[100] S. R. Noren, J. V. Redfern, and E. F. Edwards, "Pregnancy is a drag: Hydrodynamics, kinematics and performance in pre-and postparturition bottlenose dolphins (Tursiops truncatus)," *Journal of Experimental Biology*, vol. 214, no. 24, pp. 4151–4159, 2011. https://doi.org/10.1242/jeb.059121
(Back reference pages: 9, 76, 77, 83, and 87)

[101] 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, 2014. https://doi.org/10.1242/jeb.087924
(Back reference pages: x, xi, 9, 76, 78, 83, 92, 93, 101, 104, 105, and 108)

[102] D. Zhang, K. Goodbar, N. West, V. Lesage, S. Parks, D. Wiley, K. Barton, and K. A. Shorter, "Pose-gait analysis for cetaceans with biologging tag," *in Preparation for Submission*, 2021.
(Back reference pages: 12, 131)

[103] L. Shamir, C. Yerby, R. Simpson, A. M. von Benda-Beckmann, P. Tyack, F. Samarra, P. Miller, and J. Wallin, "Classification of large acoustic datasets using machine learning and crowdsourcing: Application to whale calls," *The Journal of the Acoustical Society of America*, vol. 135, no. 2, pp. 953–962, 2014. https://doi.org/10.1121/1.4861348
(Back reference page: 12)

[104] V. Leos-Barajas, T. Photopoulou, R. Langrock, T. A. Patterson, Y. Y. Watanabe, M. Murgatroyd, and Y. P. Papastamatiou, "Analysis of animal accelerometer data using hidden Markov models," *Methods in Ecology and Evolution*, vol. 8, no. 2, pp. 161–173, 2017. https://doi.org/10.1111/2041-210X.12657
(Back reference page: 12)

[105] S. O. Madgwick, A. J. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *IEEE International Conference on Rehabilitation Robotics*. IEEE, 2011. https://doi.org/10.1109/ICORR.2011.5975346
(Back reference pages: 12, 23, 25, 28, 29, 46, 66, and 82)

[106] R. D. ANDREWS, R. W. BAIRD, J. CALAMBOKIDIS, C. E. GOERTZ, F. M. GULLAND, S. K. HEIDE-JØRGENSEN , MADS PETER HOOKER, M. JOHNSON, B. Mate, Y. MITANI, D. P. NOWACEK, K. OWEN, L. T. QUAKENBUSH, S. RAVERTY, J. ROBBINS, G. S. SCHORR, O. V. SHPAK, F. TOWNSEND JR I., M. UHART, R. S. WELLS, and A. N. ZERBINI, "Best practice guidelines for cetacean tagging," *Cetacean Research and Management*, vol. 20, no. 1, pp. 27–66, 2019. https://doi.org/10.47536/jcrm.v20i1.237
(Back reference page: 13)

[107] J. Gabaldon, E. J. Anderson, K. Barton, K. A. Shorter, M. Johnson-Roberson, E. L. Turner, and M. Johnson, "Integration, calibration, and experimental verification of a speed sensor for swimming animals," *IEEE Sensors Journal*, vol. PP, no. c, pp. 1–1, 2019.
(Back reference pages: 14, 60, and 81)

[108] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: a Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981. https://doi.org/10.1145/358669.358692
(Back reference page: 19)

[109] L. M. M. López, P. J. Miller, N. A. De Soto, and M. Johnson, "Gait switches in deep-diving beaked whales: Biomechanical strategies for long-duration dives," *Journal of Experimental Biology*, vol. 218, no. 9, pp. 1325–1338, 2015. https://doi.org/10.1242/jeb.106013
(Back reference page: 28)

[110] C. Canning, D. Crain, T. S. Eaton, K. Nuessly, A. Friedlaender, T. Hurst, S. Parks, C. Ware, D. Wiley, and M. Weinrich, "Population-level lateralized feeding behaviour in North Atlantic humpback whales, Megaptera novaeangliae," *Animal Behaviour*, vol. 82, no. 4, pp. 901–909, 2011. http://dx.doi.org/10.1016/j.anbehav.2011.07.031
(Back reference page: 29)

[111] D. Zhang, K. A. Shorter, J. Rocho-Levine, J. van der Hoop, M. Moore, and K. Barton, "Behavior Inference From Bio-Logging Sensors: A Systematic Approach for Feature Generation, Selection and State Classification," in *Proceedings of the ASME 2018 Dynamic Systems and Control Conference*. September 30–October 3, 2018, Atlanta, Georgia, USA: ASME, 2018, pp. 1–10. https://doi.org/10.1115/DSCC2018-9213
(Back reference pages: 31, 131)

[112] R. Sibal, D. Zhang, J. Rocho-Levine, K. A. Shorter, and K. Barton, "Bidirectional LSTM Recurrent Neural Network Plus Hidden Markov Model for Wearable Sensor-Based Dynamic State Estimation," *ASME Letters in Dynamic Systems and Control*, vol. 1, no. 2, pp. (021 002)1–5, 2021. https://doi.org/10.1115/1.4046685
(Back reference pages: 31, 108, and 131)

[113] A. Manning and M. S. Dawkins, *An Introduction to Animal Behaviour*, 6th ed. Cambridge University Press, 2012.
(Back reference pages: 31, 33)

[114] H. K. Preisler, A. A. Ager, B. K. Johnson, and J. G. Kie, "Modeling animal movements using stochastic differential equations," *Environmetrics*, vol. 15, no. 7, pp. 643–657, 2004. https://doi.org/10.1002/env.636
(Back reference page: 31)

[115] M. Raibert, K. Blankespoor, G. Nelson, and R. Playter, "BigDog , the Rough-Terrain Quadruped Robot," in *IFAC Proceedings Volumes*, vol. 41, no. 2. IFAC, 2008, pp. 10 822–10 825. https://doi.org/10.3182/20080706-5-KR-1001.01833
(Back reference page: 31)

[116] C.-h. Yu and R. Nagpal, "Biologically-Inspired Control for Multi-Agent Self-Adaptive Tasks," in *Twenty-Fourth AAAI Conference on Artificial Intelligence*, Atlanta, Georgia, USA, 2010, pp. 1702–1707. http://www.aaai.org/ocs/index.php/AAAI/AAAI10/paper/view/1755
(Back reference page: 31)

[117] Y. S. Ryuh and J. I. Moon, "Multi-agent control and implementation of Bio-inspired underwater robots for mariculture monitoring and control," in *2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. Guangzhou, China: IEEE, 2012, pp. 777–783. https://doi.org/10.1109/ROBIO.2012.6491062
(Back reference page: 31)

[118] R. W. Davis, L. A. Fuiman, T. M. Williams, S. O. Collier, W. P. Hagey, S. B. Kanatous, S. Kohin, and M. Horning, "Hunting behavior of a marine mammal beneath the antarctic fast ice," *Science*, vol. 283, no. 5404, pp. 993–996, 1999. https://doi.org/10.1126/science.283.5404.993
(Back reference page: 32)

[119] K. Yoda, K. Sato, Y. Niizuma, M. Kurita, C. A. Bost, Y. Le Maho, and Y. Naito, "Precise monitoring of porpoising behaviour of Adelie penguins determined using acceleration data loggers," *Journal of Experimental Biology*, vol. 202, no. 22, pp. 3121–3126, 1999. https://doi.org/10.1242/jeb.202.22.3121
(Back reference page: 32)

[120] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," *Lecture Notes in Computer Science*, vol. 3001, pp. 1–17, 2004. https://doi.org/10.1007/978-3-540-24646-6{_}1
(Back reference page: 32)

[121] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity Recognition from Accelerometer Data," in *The Twentieth National Conference on Artificial Intelligence and the Seventeenth Innovative Applications of Artificial Intelligence Conference*, Pittsburgh, Pennsylvania, USA, 2005, pp. 1541–1546. https://www.researchgate.net/publication/221604452{_}Activity{_}Recognition{_}from{_}Accelerometer{_}Data
(Back reference page: 32)

[122] P. Martiskainen, M. Järvinen, J. P. Skön, J. Tiirikainen, M. Kolehmainen, and J. Mononen, "Cow behaviour pattern recognition using a three-dimensional accelerometer and support vector machines," *Applied Animal Behaviour Science*, vol. 119, no. 1-2, pp. 32–38, 2009. https://doi.org/10.1016/j.applanim.2009.03.005
(Back reference page: 32)

[123] R. Nathan, O. Spiegel, S. Fortmann-Roe, R. Harel, M. Wikelski, and W. M. Getz, "Using tri-axial acceleration data to identify behavioral modes of free-ranging animals: General concepts and tools illustrated for griffon vultures," *Journal of Experimental Biology*, vol. 215, no. 6, pp. 986–996, 2012. https://doi.org/10.1242/jeb.058602
(Back reference page: 32)

[124] S. Grünewälder, F. Broekhuis, D. W. Macdonald, A. M. Wilson, J. W. McNutt, J. Shawe-Taylor, and S. Hailes, "Movement Activity Based Classification of Animal Behaviour with an Application to Data from Cheetah (Acinonyx jubatus)," *PLoS ONE*, vol. 7, no. 11, pp. 1–11, 2012.
(Back reference page: 32)

[125] H. A. Campbell, L. Gao, O. R. Bidder, J. Hunter, and C. E. Franklin, "Creating a behavioural classification module for acceleration data: Using a captive surrogate for difficult to observe species," *Journal of Experimental Biology*, vol. 216, no. 24, pp. 4501–4506, 2013. https://doi.org/10.1242/jeb.089805
(Back reference page: 32)

[126] Z. Ghahramani, "An introduction to Hidden Markov Models and Bayesian Networks," *Pattern Recognition and Artificial Intelligence*, vol. 15, no. 2, pp. 9–42, 2001. https://doi.org/10.1142/S0218001401000836
(Back reference page: 32)

[127] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989. https://doi.org/10.1109/5.18626
(Back reference page: 32)

[128] T. T. Hammond, D. Springthorpe, R. E. Walsh, and T. Berg-Kirkpatrick, "Using accelerometers to remotely and automatically characterize behavior in small animals," *Journal of Experimental Biology*, vol. 219, no. 11, pp. 1618–1624, 2016. https://doi.org/10.1242/jeb.136135
(Back reference page: 32)

[129] D. Arifoglu and A. Bouchachia, "Activity Recognition and Abnormal Behaviour Detection with Recurrent Neural Networks," *Procedia Computer Science*, vol. 110, pp. 86–93, 2017. https://doi.org/10.1016/j.procs.2017.06.121
(Back reference page: 32)

[130] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P. M. Agapow, M. Zietz, M. M. Hoffman, W. Xie, G. L. Rosen, B. J. Lengerich, J. Israeli, J. Lanchantin, S. Woloszynek, A. E. Carpenter, A. Shrikumar, J. Xu, E. M. Cofer, C. A. Lavender, S. C. Turaga, A. M. Alexandari, Z. Lu, D. J. Harris, D. Decaprio, Y. Qi, A. Kundaje, Y. Peng, L. K. Wiley, M. H. Segler, S. M. Boca, S. J. Swamidass, A. Huang, A. Gitter, and C. S. Greene, "Opportunities and obstacles for deep learning in biology and medicine," *Journal of the Royal Society Interface*, vol. 15, no. 141, p. 20170387, 2018. https://doi.org/10.1098/rsif.2017.0387
(Back reference page: 33)

[131] A. Vieira, "Predicting online user behaviour using deep learning algorithms," *ArXiv*, pp. 1–21, 2015. http://arxiv.org/abs/1511.06247
(Back reference page: 33)

[132] N. Y. Hammerla, S. Halloran, and T. Plötz, "Deep, convolutional, and recurrent models for human activity recognition using wearables," in *IJCAI International Joint Conference on Artificial Intelligence (IJCAI)*, 2016, pp. 1533–1540. https://dl.acm.org/doi/10.5555/3060832.3060835
(Back reference page: 33)

[133] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Computation*, vol. 12, no. 10, pp. 2451–2471, 2000. https://doi.org/10.1162/089976600300015015
(Back reference pages: 33, 44, and 45)

[134] K. Saleh, M. Hossny, and S. Nahavandi, "Driving behavior classification based on sensor data fusion using LSTM recurrent neural networks," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*. Yokohama, Japan: IEEE, 2018, pp. 1–6. https://doi.org/10.1109/itsc.2017.8317835
(Back reference page: 33)

[135] E. Walton, C. Casey, J. Mitsch, J. A. Vázquez-Diosdado, J. Yan, T. Dottorini, K. A. Ellis, A. Winterlich, and J. Kaler, "Evaluation of sampling frequency, window size and sensor position for classification of sheep behaviour," *Royal Society Open Science*, vol. 5, no. 2, p. 171442, 2018. https://doi.org/10.1098/rsos.171442
(Back reference page: 36)

[136] J. A. Vázquez Diosdado, Z. E. Barker, H. R. Hodges, J. R. Amory, D. P. Croft, N. J. Bell, and E. A. Codling, "Classification of behaviour in housed dairy cows using an accelerometer-based activity monitoring system," *Animal Biotelemetry*, vol. 3, no. 1, p. 15, 2015. https://doi.org/10.1186/s40317-015-0045-8
(Back reference page: 36)

[137] M. H. M. Noor, Z. Salcic, and K. I.-k. Wang, "Adaptive sliding window segmentation for physical activity recognition using a single tri-axial accelerometer," *Pervasive and Mobile Computing*, vol. 38, pp. 41–59, 2017. http://dx.doi.org/10.1016/j.pmcj.2016.09.009
(Back reference page: 36)

[138] M. Espinilla, J. Medina, J. Hallberg, and C. Nugent, "A new approach based on temporal sub-windows for online sensor-based activity recognition," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–13, 2018. http://dx.doi.org/10.1007/s12652-018-0746-y
(Back reference page: 36)

[139] Q. Ni, T. Patterson, I. Cleland, and C. Nugent, "Dynamic detection of window starting positions and its implementation within an activity recognition framework," *Journal of Biomedical Informatics*, vol. 62, pp. 171–180, 2016. http://dx.doi.org/10.1016/j.jbi.2016.07.005
(Back reference page: 36)

[140] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003. https://dl.acm.org/doi/10.5555/944919.944968
(Back reference page: 37)

[141] K. Koutroumbas and S. Theodoridis, *Pattern Recognition*, 4th ed. Academic Press, 2008. https://www.elsevier.com/books/pattern-recognition/koutroumbas/978-1-59749-272-0
(Back reference page: 37)

[142] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1226 – 1238, 2005. http://doi.org/10.1109/TPAMI.2005.159
(Back reference page: 38)

[143] A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967. https://doi.org/10.1109/TIT.1967.1054010
(Back reference pages: vii, 42, and 43)

[144] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 39, no. 1, pp. 1–22, 1977. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x
(Back reference page: 41)

[145] A. Graves and J. Schmidhuber, "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5-6, pp. 602–610, 2005. https://doi.org/10.1016/j.neunet.2005.06.042
(Back reference page: 44)

[146] 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," in *IEEE International Conference on Robotics and Automation (ICRA)*. May 20-24, 2019, Montreal, QC, Canada: IEEE, 2019, pp. 6575–6581. https://doi.org/10.1109/ICRA.2019.8794141
(Back reference pages: 57, 108, 109, and 131)

[147] J. Gabaldon, D. Zhang, L. Lauderdale, L. Miller, K. Barton, and K. A. Shorter, "Vision-based monitoring and measurement of bottlenose dolphins' daily habitat use and kinematics," *bioRxiv*, 2021. https://doi.org/10.1101/2021.07.23.453543
(Back reference pages: 64, 111, 112, and 131)

[148] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/nongaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002. https://doi.org/10.1109/78.978374
(Back reference page: 68)

[149] M. Klaas, M. Briers, N. De Freitas, A. Doucet, S. Maskell, and D. Lang, "Fast particle smoothing: If i had a million particles," in *ACM International Conference Proceeding Series - Proceedings of the 23rd international conference on Machine learning (ICML)*, vol. 148, 2006, pp. 481–488. https://doi.org/10.1145/1143844.1143905
(Back reference page: 70)

[150] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," *Handbook of nonlinear filtering*, vol. 12, no. 656-704, 2009. https://www.stats.ox.ac.uk/{~}doucet/doucet{_}johansen{_}tutorialPF2011.pdf
(Back reference page: 70)

[151] C. M. Carvalho, M. S. Johannes, H. F. Lopes, and N. G. Polson, "Particle learning and smoothing," *Statistical Science*, vol. 25, no. 1, pp. 88–106, 2010. https://doi.org/10.1214/10-STS325
(Back reference page: 70)

[152] 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," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. September 24–28, 2017, Vancouver, BC, Canada: IEEE, 2017, pp. 2505–2510. https://doi.org/10.1109/IROS.2017.8206069
(Back reference pages: 73, 108, 109, 131, and 134)

[153] D. Zhang, J. M. van der Hoop, V. Petrov, J. Rocho-Levine, M. J. Moore, and K. A. Shorter, "Simulated and experimental estimates of hydrodynamic drag from bio-logging tags," *Marine Mammal Science*, vol. 36, no. 1, pp. 136–157, 2020. https://doi.org/10.1111/mms.12627
(Back reference pages: 75, 108, and 131)

[154] D. Zhang, Y. Wang, L. K. Lauderdale, J. Gabaldon, L. J. Miller, K. Barton, and K. A. Shorter, "Energetic benefits of fluke and glide gait for Bottlenose dolphins," *in Preparation for Submission*, 2021.
(Back reference pages: 75, 131)

[155] T. G. Lang and D. A. Daybell, "Porpoise performance tests in a sea-water tank," *NAVAL ORDNANCE TEST STATION CHINA LAKE CA*, no. January, 1963. https://apps.dtic.mil/sti/pdfs/AD0298742.pdf
(Back reference page: 76)

[156] D. Bilo and W. Nachtigall, "A Simple Method to Determine Drag Coefficients in Aquatic Animals," *Journal of Experimental Biology*, vol. 87, no. 1, pp. 357–359, 1980. https://doi.org/10.1242/jeb.87.1.357
(Back reference page: 76)

[157] M. B. Hanson, "An evaluation of the relationship between small cetacean tag design and attachment durations: a bioengineering approach," Ph.D. dissertation, University of Washington, 2001.
(Back reference page: 76)

[158] T. T. Jones, K. S. Van Houtan, B. L. Bostrom, P. Ostafichuk, J. Mikkelsen, E. Tezcan, M. Carey, B. Imlach, and J. A. Seminoff, "Calculating the ecological impacts of animal-borne instruments on aquatic organisms," *Methods in Ecology and Evolution*, vol. 4, no. 12, pp. 1178–1186, 2013. https://doi.org/10.1111/2041-210X.12109
(Back reference pages: 76, 89, 106, and 133)

[159] K. Aoki, Y. Y. Watanabe, D. E. Crocker, P. W. Robinson, M. Biuw, D. P. Costa, N. Miyazaki, M. A. Fedak, and P. J. Miller, "Northern elephant seals adjust gliding and stroking patterns with changes in buoyancy: Validation of at-sea metrics of body density," *Journal of Experimental Biology*, vol. 214, no. 17, pp. 2973–2987, 2011. https://doi.org/10.1242/jeb.055137
(Back reference page: 76)

[160] C. Ware, A. W. Trites, D. A. Rosen, and J. Potvin, "Averaged propulsive body acceleration (APBA) can be calculated from biologging tags that incorporate gyroscopes and accelerometers to estimate swimming speed, hydrodynamic drag and energy expenditure for Steller sea lions," *PLoS ONE*, vol. 11, no. 6, pp. 1–27, 2016.
(Back reference pages: 76, 77)

[161] L. M. M. López, N. A. De Soto, P. Miller, and M. Johnson, "Tracking the kinematics of caudal-oscillatory swimming: A comparison of two on-animal sensing methods," *Journal of Experimental Biology*, vol. 219, no. 14, pp. 2103–2109, 2016. https://doi.org/10.1242/jeb.136242
(Back reference page: 77)

[162] A. A. Hazekamp, R. Mayer, and N. Osinga, "Flow simulation along a seal: The impact of an external device," *European Journal of Wildlife Research*, vol. 56, no. 2, pp. 131–140, 2010. https://doi.org/10.1007/s10344-009-0293-0
(Back reference page: 77)

[163] V. V. Pavlov and A. M. Rashad, "A non-invasive dolphin telemetry tag: Computer design and numerical flow simulation," *Marine Mammal Science*, vol. 28, no. 1, pp. 16–27, 2012. https://doi.org/10.1111/j.1748-7692.2011.00476.x
(Back reference page: 77)

[164] B. C. Balmer, R. S. Wells, L. E. Howle, A. A. Barleycorn, W. A. McLellan, D. Ann Pabst, T. K. Rowles, L. H. Schwacke, F. I. Townsend, A. J. Westgate, and E. S. Zolman, "Advances in cetacean telemetry: A review of single-pin transmitter attachment techniques on small cetaceans and development of a new satellite-linked transmitter design," *Marine Mammal Science*, vol. 30, no. 2, pp. 656–673, 2014. https://doi.org/10.1111/mms.12072
(Back reference page: 77)

[165] K. A. Shorter, M. M. Murray, M. Johnson, M. Moore, and L. E. Howle, "Drag of suction cup tags on swimming animals: Modeling and measurement," *Marine Mammal Science*, vol. 30, no. 2, pp. 726–746, 2014. https://doi.org/10.1111/mms.12083
(Back reference pages: 77, 90, 102, and 104)

[166] A. B. Phillips, S. R. Turnock, and M. Furlong, "The use of computational fluid dynamics to aid cost-effective hydrodynamic design of autonomous underwater vehicles," *Proceedings of the Institution of Mechanical Engineers Part M: Journal of Engineering for the Maritime Environment*, vol. 224, no. 4, pp. 239–254, 2010. https://doi.org/10.1243/14750902JEME199
(Back reference page: 77)

[167] T. H. Joung, K. Sammut, F. He, and S. K. Lee, "Shape optimization of an autonomous underwater vehicle with a ducted propeller using computational fluid dynamics analysis," *International Journal of Naval Architecture and Ocean Engineering*, vol. 4, no. 1, pp. 44–56, 2012. https://doi.org/10.2478/IJNAOE-2013-0077
(Back reference page: 77)

[168] S. G. Shereena, S. Vengadesan, V. G. Idichandyxs, and S. K. Bhattacharyya, "CFD study of drag reduction in axisymmetric underwater vehicles using air jets," *Engineering Applications of Computational Fluid Mechanics*, vol. 7, no. 2, pp. 193–209, 2013. https://doi.org/10.1080/19942060.2013.11015464
(Back reference page: 77)

[169] G. Sachs, "New model of flap-gliding flight," *Journal of Theoretical Biology*, vol. 377, pp. 110–116, 2015. https://doi.org/10.1016/j.jtbi.2015.03.022
(Back reference page: 77)

[170] S. Vogel, *Life in Moving Fluids: The Physical Biology of Flow*, 2nd ed. Princeton University Press, 1996. https://doi.org/10.1515/9780691212975
(Back reference pages: 78, 102, and 103)

[171] F. E. Fish, "Power Output and Propulsive Efficiency of Swimming Bottlenose Dolphins (Tursiops Truncatus)," *Journal of Experimental Biology*, vol. 185, no. 1, pp. 179–193, 1993.
(Back reference pages: xiii, 78, 95, and 103)

[172] K. Schmidt-Nielsen, "Locomotion: Energy cost of swimming, flying, and running," *Science*, vol. 177, no. 4045, pp. 222–228, 1972. https://doi.org/10.1126/science.177.4045.222
(Back reference page: 81)

[173] V. A. Tucker, "The energetic cost of moving about," *American Scientist*, vol. 63, no. 4, pp. 413–419, 1975. https://www.jstor.org/stable/27845576
(Back reference page: 81)

[174] F. E. Fish, "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, 1998. https://doi.org/10.1242/jeb.201.20.2867
(Back reference page: 83)

[175] L. L. Stelle, R. W. Blake, and A. W. Trites, "Hydrodynamic drag in Steller sea lions (Eumetopias jubatus)," *Journal of Experimental Biology*, vol. 203, no. 12, pp. 1915–1923, 2000. https://doi.org/10.1242/jeb.203.12.1915
(Back reference page: 87)

[176] W. Rodi, "Experience with two-layer models combining the k-epsilon model with a one-equation model near the wall," in *29th Aerospace Sciences Meeting*, 1991. https://doi.org/10.2514/6.1991-216
(Back reference page: 90)

[177] T. J. DiCiccio and B. Efron, "Bootstrap confidence intervals," *Statistical Science*, vol. 11, no. 3, pp. 189–228, 1996. https://doi.org/10.1214/ss/1032280214
(Back reference page: 91)

[178] M. Kabra, A. A. Robie, M. Rivera-Alba, S. Branson, and K. Branson, "JAABA: Interactive machine learning for automatic annotation of animal behavior," *Nature Methods*, vol. 10, no. 1, pp. 64–67, 2013. https://doi.org/10.1038/nmeth.2281
(Back reference page: 102)

[179] J. J. Rohr, F. E. Fish, and J. W. J. Gilpatrick, "Maximum swim speeds of captive and free-ranging Delphinids: Critical analysis of extraordinary performance," *Marine Mammal Science*, vol. 18, no. 1, pp. 1–19, 2002. https://doi.org/10.1111/j.1748-7692.2002.tb01014.x
(Back reference page: 102)

[180] C. Lockyer and R. Morris, "Observations on diving behaviour and swimming speeds in a wild juvenile Tursiops truncatus," pp. 31–35, 1987. https://aquaticmammalsjournal.org/index.php?option=com{_}content{&}view=article{&}id=44:volume-13{&}catid=3{&}Itemid=101
(Back reference page: 102)

[181] B. Würsig and M. Würsig, "Behavior and ecology of the bottlenose dolphin, Tursiops truncatus, in the south Atlantic," *Fishery Bulletin*, vol. 77, no. 2, pp. 399–412, 1979. https://www.google.com/books/edition/Fishery{_}Bulletin/1YiqSInWOaQC?hl=en{&}gbpv=0
(Back reference page: 102)

[182] A. B. Irvine, R. S. Wells, and M. D. Scott, "An evaluation of techniques for tagging small odontocete cetaceans [Tursiops truncatus, Atlantic bottlenose dolphins]," *Fishery Bulletin*, vol. 80, no. 1, pp. 135–143, 1982. https://agris.fao.org/agris-search/search.do?recordID=US19830951678
(Back reference page: 102)

[183] S. H. Shane, "Behavior and Ecology of the Bottlenose Dolphin at Sanibel Island, Florida," in *The Bottlenose Dolphin*, S. Leatherwood and R. R. Reeves, Eds. Academic Press, Inc., 1990, pp. 245–265. http://dx.doi.org/10.1016/B978-0-12-440280-5.50016-0
(Back reference page: 102)

[184] V. Ridoux, C. Guinet, C. L. P. Creton, R. Steenstrup, and G. Beauplet, "A video sonar as a new tool to study marine mammals in the wild: Measurements of dolphin swimming speed," *Marine Mammal Science*, vol. 13, no. 2, pp. 196–206, 1997. https://doi.org/10.1111/j.1748-7692.1997.tb00627.x
(Back reference page: 102)

[185] 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," *Journal of experimental biology*, vol. 179, no. 1, pp. 31–46, 1993. https://doi.org/10.1242/jeb.179.1.31
(Back reference page: 102)

[186] 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, 1999. https://doi.org/10.1007/s002270050575
(Back reference page: 102)

[187] B. E. Ungerechts, D. Daly, and J. P. Zhu, "What dolphins tell us about hydrodynamics," *Journal of Swimming Research*, vol. 13, pp. 1–7, 1998.
(Back reference page: 103)

[188] J. Rohr, M. I. Latz, S. Fallon, J. C. Nauen, and E. Hendricks, "Experimental approaches towards interpreting dolphin-stimulated bioluminescence," *Journal of Experimental Biology*, vol. 201, no. 9, pp. 1447–1460, 1998. https://doi.org/10.1242/jeb.201.9.1447
(Back reference page: 103)

[189] E. J. Anderson, W. R. McGillis, and M. A. Grosenbaugh, "The boundary layer of swimming fish," *Journal of Experimental Biology*, vol. 204, no. 1, pp. 81–102, 2001. https://doi.org/10.1242/jeb.204.1.81
(Back reference page: 103)

[190] P. W. Webb, "Hydrodynamics and Energetics of Fish Propulsion," *Bulletin Fisheries Research Boad of Canada*, no. 190, pp. 1–158, 1975. http://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail{&}idt=PASCAL7536509928
(Back reference page: 103)

[191] F. Broell, C. Burnell, and C. T. Taggart, "Measuring abnormal movements in free-swimming fish with accelerometers: Implications for quantifying tag and parasite load," *Journal of Experimental Biology*, vol. 219, no. 5, pp. 695–705, 2016. https://doi.org/10.1242/jeb.133033
(Back reference pages: 103, 106, and 133)

[192] R. P. Wilson, W. S. Grant, and D. C. Duffy, "Recording Devices on Free-Ranging Marine Animals: Does Measurement Affect Foraging Performance?" *Ecology*, vol. 67, no. 4, pp. 1091–1093, 1986. https://doi.org/10.2307/1939832
(Back reference page: 103)

[193] C. L. Littnan, J. D. Baker, F. A. Parrish, and G. J. Marshall, "Effects of video camera attachment on the foraging behavior of immature Hawaiian monk seals," *Marine Mammal Science*, vol. 20, no. 2, pp. 345–352, 2004. https://doi.org/10.1111/j.1748-7692.2004.tb01164.x
(Back reference page: 103)

[194] C. R. McMahon, I. C. Field, C. J. Bradshaw, G. C. White, and M. A. Hindell, "Tracking and data-logging devices attached to elephant seals do not affect individual mass gain or survival," *Journal of Experimental Marine Biology and Ecology*, vol. 360, no. 2, pp. 71–77, 2008. https://doi.org/10.1016/j.jembe.2008.03.012
(Back reference page: 103)

[195] J. L. Maresh, S. E. Simmons, D. E. Crocker, B. I. McDonald, T. M. Williams, and D. P. Costa, "Free-swimming northern elephant seals have low field metabolic rates that are sensitive to an increased cost of transport," *Journal of Experimental Biology*, vol. 217, no. 9, pp. 1485–1495, 2014. https://doi.org/10.1242/jeb.094201
(Back reference page: 103)

[196] A. Solsona Berga, A. J. Wright, A. Galatius, S. Sveegaard, and J. Teilmann, "Do larger tag packages alter diving behavior in harbor porpoises?" *Marine Mammal Science*, vol. 31, no. 2, pp. 756–763, 2015. https://doi.org/10.1111/mms.12179
(Back reference page: 103)

[197] G. Fiore, E. Anderson, C. S. Garborg, M. Murray, M. Johnson, M. J. Moore, L. Howle, and K. A. Shorter, "From the track to the ocean: Using flow control to improve marine bio-logging tags for cetaceans," *PLoS ONE*, vol. 12, no. 2, pp. 1–19, 2017.
(Back reference page: 104)

[198] J. Lavest, G. Rives, and J. Lapresté, "Underwater Camera Calibration," in *Vernon D. (eds) Computer Vision - ECCV 2000. ECCV 2000. Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 2000, pp. 654–668. https://doi.org/10.1007/3-540-45053-X{_}42
(Back reference page: 106)

[199] M. Shortis, "Calibration techniques for accurate measurements by underwater camera systems," *Sensors (MDPI)*, vol. 15, no. 12, pp. 30 810–30 826, 2015. https://doi.org/10.3390/s151229831
(Back reference page: 106)

[200] C. A. Hui, "Surfacing Behavior and Ventilation in Free-Ranging Dolphins," *Journal of Mammalogy*, vol. 70, no. 4, pp. 833–835, 1989. https://doi.org/10.2307/1381722
(Back reference pages: 106, 133)

[201] D. G. Barron, J. D. Brawn, and P. J. Weatherhead, "Meta-analysis of transmitter effects on avian behaviour and ecology," *Methods in Ecology and Evolution*, vol. 1, no. 2, pp. 180–187, 2010. https://doi.org/10.1111/j.2041-210X.2010.00013.x
(Back reference pages: 106, 133)

[202] B. M. Culik, R. Bannasch, and R. P. Wilson, "External devices on penguins: how important is shape?" *Marine Biology*, vol. 118, pp. 353–357, 1994. ttps://doi.org/10.1007/BF00350291
(Back reference pages: 106, 133)

[203] D. Zhang, J. Gabaldon, K. Goodbar, N. West, K. Barton, and K. A. Shorter, "Fine-scale dynamic profiles of Bottlenose dolphin during conspecific and enrichment interactions," *in Preparation for Submission*, 2021.
(Back reference pages: 107, 131)

[204] Z. Zhang, D. Zhang, J. Gabaldon, K. Goodbar, N. West, K. Barton, and K. A. Shorter, "Investigation of environmentally dependent movement of bottlenose dolphins (Tursiops truncatus)," *Journal of Zoological and Botanical Gardens*, vol. 2, no. 3, pp. 335–348, 2021. https://doi.org/10.3390/jzbg2030023
(Back reference pages: 107, 131)

[205] R. Kagan, S. Carter, and S. Allard, "A Universal Animal Welfare Framework for Zoos," *Journal of Applied Animal Welfare Science*, vol. 18, pp. 1–10, 2015. https://doi.org/10.1080/10888705.2015.1075830
(Back reference page: 107)

[206] J. C. Whitham and N. Wielebnowski, "New directions for zoo animal welfare science," *Applied Animal Behaviour Science*, vol. 147, no. 3-4, pp. 247–260, 2013. https://doi.org/10.1016/j.applanim.2013.02.004
(Back reference page: 107)

[207] L. J. Miller, G. A. Vicino, J. Sheftel, and L. K. Lauderdale, "Behavioral diversity as a potential indicator of positive animal welfare," *Animals*, vol. 10, no. 7, p. 1211, 2020. https://doi.org/10.3390/ani10071211
(Back reference page: 107)

[208] S. Daan, "Adaptive Daily Strategies in Behavior," in *Biological Rhythms*, J. Aschoff, Ed. Boston, MA: Springer, 1981, pp. 275–298. https://doi.org/10.1007/978-1-4615-6552-9{_}15
(Back reference page: 107)

[209] J. A. Stamps, "Individual differences in behavioural plasticities," *Biological Reviews*, vol. 91, no. 2, pp. 534–567, 2016. https://doi.org/10.1111/brv.12186
(Back reference pages: 107, 112)

[210] 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*. Waikoloa, HI, USA: IEEE, 2015, pp. 51–56. https://doi.org/10.1109/WACVW.2015.10
(Back reference pages: 107, 108)

[211] C. Waitt and H. M. Buchanan-Smith, "What time is feeding?: How delays and anticipation of feeding schedules affect stump-tailed macaque behavior," *Applied Animal Behaviour Science*, vol. 75, no. 1, pp. 75–85, 2001. https://doi.org/10.1016/S0168-1591(01)00174-5
(Back reference page: 107)

[212] G. J. Mason, "Species differences in responses to captivity: Stress, welfare and the comparative method," *Trends in Ecology and Evolution*, vol. 25, no. 12, pp. 713–721, 2010. https://doi.org/10.1016/j.tree.2010.08.011
(Back reference page: 107)

[213] A. F. McBride and D. O. Hebb, "Behavior of the captive bottle-nose dolphin, Tursiops truncatus," *Journal of Comparative and Physiological Psychology*, vol. 41, no. 2, pp. 111–123, 1948. https://doi.org/10.1037/h0057927
(Back reference page: 107)

[214] 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, pp. 1–12, 2018.
(Back reference pages: 108, 109)

[215] 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, pp. 1–25, 2019.
(Back reference pages: 108, 109)

[216] A. Dutta and A. Zisserman, "The VIA annotation software for images, audio and video," in *MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 2276–2279. https://doi.org/10.1145/3343031.3350535
(Back reference page: 110)

[217] V. W. Berger and Y. Zhou, "Kolmogorov – Smirnov Test: Overview," *Wiley StatsRef: Statistics Reference Online*, pp. 1–5, 2014. https://doi.org/10.1002/9781118445112.stat06558
(Back reference page: 113)

[218] 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, 2006. https://doi.org/10.1111/j.1365-2656.2006.01127.x
(Back reference page: 114)

[219] 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, 1994. https://doi.org/10.1016/0166-4328(94)90071-X
(Back reference page: 127)

[220] G. M. Stafne and P. R. Manger, "Predominance of clockwise swimming during rest in Southern Hemisphere dolphins," *Physiology and Behavior*, vol. 82, no. 5, pp. 919–926, 2004. https://doi.org/10.1016/j.physbeh.2004.08.001
(Back reference page: 127)

[221] S. Kuczaj, T. Lacinak, O. Fad, M. Trone, M. Solangi, and J. Ramos, "Keeping Environmental Enrichment Enriching," *International Journal of Comparative Psychology*, vol. 15, no. 2, pp. 127–137, 2002. https://escholarship.org/uc/item/1bj376tj
(Back reference page: 128)

[222] H. E. Harley, W. Fellner, and M. A. Stamper, "Cognitive Research with Dolphins (Tursiops truncatus) at Disney's The Seas : A Program for Enrichment , Science , Education , and Conservation," *International Journal of Comparative Psychology*, vol. 23, no. 3, pp. 331–343, 2010. https://escholarship.org/uc/item/92z8z9wq
(Back reference page: 128)

[223] L. K. Lauderdale and L. J. Miller, "Common bottlenose dolphin (Tursiops truncatus) problem solving strategies in response to a novel interactive apparatus," *Behavioural Processes*, vol. 169, no. October, p. 103990, 2019. https://doi.org/10.1016/j.beproc.2019.103990
(Back reference page: 128)

[224] L. Lauderdale and L. J. Miller, "Efficacy of an interactive apparatus as environmental enrichment for common bottlenose dolphins (Tursiops truncatus)," *Animal Welfare*, vol. 29, no. 4, pp. 379–386, 2020. https://doi.org/10.7120/09627286.29.4.379
(Back reference page: 128)

[225] S. Cones, D. Zhang, K. A. Shorter, K. Katija, D. Mann, F. Jensen, P. Afonso, J. Fontes, and T. A. Mooney, "Swimming behaviors during vertical migration in veined squid (Loligo forbesii)," *Submitted for Publication*, 2021.
(Back reference page: 131)

[226] 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," *Submitted for Publication*, 2021.
(Back reference page: 131)

[227] L. Lauderdale, K. A. Shorter, D. Zhang, J. Gabaldon, J. D. Mellen, M. T. Walsh, D. A. Granger, and L. J. Miller, "Bottlenose dolphin habitat and management factors related to activity and distance traveled in zoos and aquariums," *PLOS ONE*, vol. 16, no. 8, p. e0250687, 2021. https://doi.org/10.1371/journal.pone.0250687
(Back reference page: 131)

[228] L. K. Lauderdale, K. A. Shorter, D. Zhang, J. Gabaldon, J. D. Mellen, M. T. Walsh, D. A. Granger, and L. J. Miller, "Habitat characteristics and animal management factors associated with habitat use by bottlenose dolphins in zoological environments," *PLOS ONE*, vol. 16, no. 8, p. e0252010, 2021. https://doi.org/10.1371/journal.pone.0252010
(Back reference page: 131)

[229] L. K. Lauderdale, K. A. Shorter, D. Zhang, J. Gabaldon, J. D. Mellen, D. A. Granger, M. T. Walsh, and L. J. Miller, "Environmental Enrichment Factors Associated with the Energy Expenditure of Bottlenose Dolphins Under Professional Care," *Submitted for Publication*, 2021.
(Back reference page: 131)

[230] J. M. Kendall-Bar, D. A. Lozano, J. K. Pitman, R. Holser, R. Beltran, M. Schalles, C. L. Field, S. P. Johnson, D. Zhang, A. L. Vyssotski, D. P. Costa, and T. M. Williams, "Sleeping while Diving: Validating a non-invasive method to record sleep in wild marine mammals," *in Preparation for Submission*, 2021.
(Back reference page: 131)

[231] D. Zhang, Y. Wang, L. K. Lauderdale, J. Gabaldon, L. J. Miller, K. Barton, and K. A. Shorter, "A data driven approach for estimating hydrodynamic drag of bottlenose dolphins," *Integrative and Comparative Biology*, vol. 61, p. E1022, 2021.
(Back reference page: 131)

[232] D. Zhang, J. Gabaldon, J. Rocho-Levine, J. Van der Hoop, M. Moore, and K. Shorter, "Investigating bottlenose dolphin swimming biomechanics using biologging tags, tracking data, sensor fusion and estimation," *Integrative and Comparative Biology*, vol. 59, p. E259, 2019.
(Back reference page: 131)

[233] D. Zhang, K. Barton, F. Jensen, K. McHugh, R. T. Moore, L. Sayigh, P. Tyack, R. Wells, and K. A. Shorter, "Automatic locomotion: A machine learning approach for bottlenose dolphin locomotion analysis," *8th World Congress of Biomechanics*, p. O1673, 2018.
(Back reference page: 131)

[234] Z. Zhang, D. Zhang, J. Gabaldon, N. West, K. Barton, and K. A. Shorter, "Analysis of environment dependent locomotion of bottlenose dolphins using Mask R-CNN," *Integrative and Comparative Biology*, vol. 61, p. E1023, 2021.
(Back reference page: 131)

[235] K. A. Shorter, D. Zhang, J. Gabaldon, J. Rocho-Levine, J. van der Hoop, and M. Moore, "Estimating mechanical work during bottlenose dolphin swimming using biologging tags, tracking data, sensor fusion, and estimation," *Ocean Sciences Meeting*, 2020.
(Back reference page: 131)