Improving Dive Phase Definitions in Northern Resident Killer Whales

by

Ian Murphy

BMath, University of Waterloo, 2019

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES (Statistics)

The University of British Columbia (Vancouver)

August 2021

© Ian Murphy, 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Improving Dive Phase Definitions in Northern Resident Killer Whales

submitted by **Ian Murphy** in partial fulfillment of the requirements for the degree of **Master of Science** in **Statistics**.

Examining Committee:

Nancy Heckman, Professor, Statistics, UBC *Supervisor*

Marie Auger-Méthé, Assistant Professor, Statistics, Institute for Oceans & Fisheries, UBC *Co-supervisor*

Abstract

In contrast with the endangered southern resident killer whales (SRKWs), the northern resident killer whales (NRKWs) have been thriving in their habitats. The main hypotheses proposed to explain the differences in survival of these population are associated with differential reproductive output, body compositions, and feeding rates. Testing some of these hypotheses requires researchers to identify prey captures for these animals. As these events are difficult to directly observe through field operations, researchers equip whales with suction-cup attached biologgers and use kinematic variables during the bottom phase of a dive to predict prey captures. However, universal definitions of the bottom phase have not been established and often appear arbitrarily chosen, leading to potentially over or underestimating foraging events. Using the diving and kinematic data collected from three NRKWs, I show that modifying the bottom phase greatly impacts existing methods used to predict prey capture events. To investigate bottom phase definition variability, I then asked several whale researchers to identify the bottom phase of various dives via an interactive study. Linear mixed-effects model analyses showed that there exists substantial variation in bottom phase definitions across different researchers and across different dive types. I compared several statistical models of the start and end of the bottom phase of a dive, including modifications to existing methods, linear regression models, and functional linear regression models. Compared to the currently used bottom phase definitions, using the model based definitions resulted in significant improvements when predicting prey capture dives. Furthermore, these proposed models offer substantial increases in prediction accuracy of the bottom phase of a dive when comparing these model predictions and the currently used methods to the user-provided bottom phases. Finally, I formulated two methods to determine an adequate sample size for fitting these statistical models. The results of both methods show that an adequate sample size of approximately 50-100 dives can be used to obtain satisfactory model predictions for this data. This work shows that dive phase definitions may impact the results of many existing studies and should be emphasized as an important part of analyzing diving data.

Lay Summary

In contrast with the endangered southern resident killer whales (SRKWs), the northern resident killer whales (NRKWs) are comparatively healthier with a larger population size, higher reproductive success and generally better body composition. A leading hypothesis proposed to explain the differences in survival of these populations is based on differences in feeding success, and requires researchers to identify prey captures. As the events are difficult to directly observe, researchers often predict prey captures using statistical models. Prediction of prey capture dives depends on several key components of a dive, called dive phases (typically descent, bottom, and ascent phases), which are defined by the researcher. In modifying these dive phase definitions, I show that predicted prey capture dives change substantially. This thesis seeks to ameliorate these user-defined dive phases through a formal statistical approach. This work shows that dive phase definitions affect a broad range of important dive metrics, and thus, dive phase definitions may impact the results of many existing studies.

Preface

This thesis was prepared by the author, Ian Murphy, under the supervision of Professor Nancy Heckman and Assistant Professor Marie Auger-Méthé of the University of British Columbia. The data analyses in Chapters 2-5 are my original work. Both supervisors assisted in developing and refining the research topic and provided supervision during the analyses. The data used in Chapters 2-5 from the killer whales were collected by Dr. Sarah Fortune and Dr. Andrew Trites of the Marine Mammal Research Unit of the University of British Columbia under Animal Care Permits XMMS 6 2019 and A19-0053.

I was responsible for all major areas of concept formation, data collection, analysis, and reporting for the study conducted in Chapter 3 with Dr. Nancy Heckman as the primary investigator, and this project was approved by the University of British Columbia's Research Ethics Board [certificate #H20-02096-A002].

None of the text of the dissertation is taken directly from previously published or collaborative articles.

Table of Contents

y Sur	nmary
	-
eface	· · · · · · · · · · · · · · · · · · ·
ble of	Contents
st of]	Γables
st of I	igures
know	ledgments
Intro	oduction
Exp	loratory Analysis
2.1	Motivation & Background
2.2	Data Description
2.3	Longitudinal Variables Calculated from Raw Data
2.4	Dive Level Characteristics
2.5	Detection of Foraging Dives 11
2.6	Modifying the Bottom Phase Definition
2.7	Conclusions
Gett	ing to the Bottom of It: An Experimental Study
3.1	Background & Motivation
3.2	Experimental Design
	ble of st of T st of T know Intro 2.1 2.2 2.3 2.4 2.5 2.6 2.7 Gett 2.1

	3.4	Results
	3.5	Discussion
4	Stat	istical Models for the Bottom Phase
	4.1	Background
	4.2	Methods
		4.2.1 Percentage Threshold
		4.2.2 Linear Regression
		4.2.3 Functional Regression
		4.2.4 Model Comparison
	4.3	Results
		4.3.1 Percentage Threshold
		4.3.2 Linear Model
		4.3.3 Functional Model
		4.3.4 Model Comparison
	4.4	Conclusions
5	Hou	w Many Divag To Chaogo? Mothoda fan Dagoanahang
3	П0М 5 1	Deckground
	5.1	
	5.2	Methods
		5.2.1 Probabilistic Approach
	5.0	5.2.2 User-Provided Approach
	5.3	Results 60 Quarkariana 60
	5.4	Conclusions
6	The	Bottom Line
Bi	bliogi	raphy
٨	Sum	norting Materials 7
A		Matrix and Vactor Formulation of Models from Chapter 3
	л.1 л ว	Model Diagnostics for Descentage Threshold Model
	Λ.2	Model Diagnostics for the Linear Model
	A.3	Model Diagnostics for the Euleric Pagression Models
	A.4	Product Diagnostics for the Functional Regression Models $\dots \dots \dots$
	A.J	

List of Tables

Several summary statistics of important variables for each whale. The VeDBA	
for a single dive is calculated as the average VeDBA over the duration of the dive.	9
Model predictions versus actual prey capture status for dives from Whale #3	14
Dive numbers for the predicted successful prey capture dives for multiple differ-	
ent bottom phase definitions	22
Dive metrics from dive 67 from Whale #3	24
Values for the dive metrics used to identify successful prey captures using the	
criteria established by Tennessen et al. [1] stratified by different bottom phase	
definitions according to maximum depth thresholds. Rows highlighted in blue	
indicate predicted successful prey capture dives.	26
Values for the dive metrics used to identify successful prey captures using the	
criteria established by Tennessen et al. [1] stratified by different bottom phase	
definitions according to vertical velocity thresholds. Rows highlighted in blue	
indicate predicted successful prey capture dives	26
Summary of output of the mixed-effects model presented in Equation 3.3 for	
both the proportion of time into the dive at the start and end of the bottom phase.	35
Intraclass Correlation Coefficients (ICCs) for each of the estimated variances of	
the random intercepts in the two models	36
Results from fitting the percentage threshold models (from Equations 4.3 and	
4.4) with 95% boostrapped confidence intervals.	47
Results from fitting the basic linear regression model.	52
Estimates of the regression coefficients from fitting Equation 4.8 for the start and	
end of the bottom phase for each functional covariate. Adjusted R^2 is also given.	57
	Several summary statistics of important variables for each whale. The VeDBA for a single dive is calculated as the average VeDBA over the duration of the dive. Model predictions versus actual prey capture status for dives from Whale #3 Dive numbers for the predicted successful prey capture dives for multiple different bottom phase definitions

Table 4.4	Results from running 5- and 10-fold cross validation for each of the models men-	
	tioned in this chapter.	60
Table 4.5	Confusion matrices for predicting prey captures via Tennessen et al. [1]'s criteria,	
	but with the bottom phase determined by the different models. The last table	
	is the confusion matrix when predicting prey captures using the user provided	
	bottom phases.	61

List of Figures

Figure 1.1	A time-depth profile of a dive of a NRKW from the datasets used in this thesis, where the bottom phase (in red) has been defined using the 70% maximum				
	depth threshold	2			
Figure 2.1	The time-depth plot of Whale #1 during the tag attachment period	6			
Figure 2.2	The time-depth plot of Whale #2 during the tag attachment period	6			
Figure 2.3	The time-depth plot of Whale #3 during the tag attachment period	7			
Figure 2.4	A visualization of the three axes of rotation as well as the three directions of				
	movement for an animal freely moving in three dimensions.	8			
Figure 2.5	Dive 64 from Whale #1 which occurred from approximately 13:51:43 PM until				
	13:52:29 PM (about 46 seconds in duration).	13			
Figure 2.6	Dive 142 from Whale #1 which occurred from approximately 15:01:59 PM until				
e	15:02:31 PM (about 32 seconds in duration).	13			
Figure 2.7	Predicted prey capture dives for Whale #3 using the 70% depth threshold and				
e	applying the methods in Tennessen et al. [1]. Confirmed prev capture dives are				
	highlighted in red.	15			
Figure 2.8	Boxplots of the distributions of log ierk peaks for various bottom phase defini-				
8	tion thresholds involving maximum depth. Horizontal line indicates one of the				
	criteria a dive needs to meet to be classified as a successful prev capture, based				
	on Tennessen et al. [1]	17			
Figure 2.9	Boxplots of the distributions of rolls at jerk peak for various bottom phase def-	17			
115010 2.9	inition thresholds involving maximum denth. Horizontal line indicates one of				
	the criteria a dive peeds to meet to be classified as a successful providenture				
	the chieffa a dive needs to meet to be classified as a successful prey capture,	10			
	based on Tennessen et al. [1]	18			

Figure 2.10	Boxplots of the distributions of heading variance for various bottom phase def-	
	inition thresholds involving maximum depth. Horizontal line indicates one of	
	the criteria a dive needs to meet to be classified as a successful prey capture,	
	based on Tennessen et al. [1]	19
Figure 2.11	Boxplots of the distributions of log jerk peaks for various bottom phase defini-	
	tion thresholds involving velocity. Horizontal line indicates the criterion used	
	in Tennessen et al. [1]	20
Figure 2.12	Boxplots of the distributions of rolls at jerk peak for various bottom phase defi-	
	nition thresholds involving velocity. Horizontal line indicates the criterion used	
	in Tennessen et al. [1]	20
Figure 2.13	Boxplots of the distributions of heading variance for various bottom phase defi-	
	nition thresholds involving velocity. Horizontal line indicates the criterona used	
	in Tennessen et al. [1]	21
Figure 2.14	Time-Depth plot of dive 67 from Whale #3	24
Figure 2.15	Dive 126 from Whale #2. Plots of depth and jerk with two definitions of bottom	
	phase labeled via vertical lines. The red line indicates a 60% depth threshold	
	whereas the blue line indicates a 70% depth threshold. The 70% threshold ex-	
	cluded the sharp jerk measure as seen between the two vertical lines on the	
	right-hand side. This is the reason why the dive was not predicted as a suc-	
	cessful prey capture using this threshold, but it was predicted using the other	
	threshold which contained this sharp jerk measurement	27
Figure 3.1	Example of a dive from the Shiny App	31
Figure 3.2	Plots of the responses for the start of the bottom phase for dives that were	
	viewed twice by the researchers.	36
Figure 3.3	Plots of the responses for the end of the bottom phase for dives that were viewed	
C	twice by the researchers.	37
Figure 4.1	Plots of the depth curves for the 250 dives from Whale #3 to be used as func-	
	tional covariates in Equation 4.8. The times for each depth curve were scaled to	
	[0,1]	44
Figure 4.2	Plots of the speed curves for the 250 dives from Whale #3 to be used as func-	
	tional covariates in Equation 4.8. The times for each speed curve were scaled	
	to [0,1]	45

Figure 4.3	Plots of the acceleration curves for the 250 dives from Whale #3 to be used as functional covariates in Equation 4.8. The times for each depth curve were	
	scaled to [0,1].	46
Figure 4.4	Sum of squared errors versus theta for the start of the bottom phase for values	
	of theta over 0 to 1	48
Figure 4.5	Sum of squared errors versus theta for the end of the bottom phase for values of	
	theta over 0 to 1. \ldots	48
Figure 4.6	Identified outlier dives for the percent of max depth model for the start of the	
	bottom phase. The red line indicates the user-provided start of the bottom phase,	
	and the turquoise line indicates the predicted start of the bottom phase from the	
	model.	49
Figure 4.7	Identified outlier dives for the percent of max depth model for the end of the	
	bottom phase. The red line indicates the user-provided end of the bottom phase,	
	and the turquoise line indicates the predicted end of the bottom phase from the	
	model	50
Figure 4.8	Predicted responses for the start of the bottom phase as a function of duration	
	for several values of max depth	51
Figure 4.9	Predicted responses for the end of the bottom phase as a function of duration	
	for several values of max depth	51
Figure 4.10	Identified outlier dives for the linear regression model in Equation 4.5 for the	
	start of the bottom phase. The red line indicates the user-provided start of the	
	bottom phase, and the turquoise line indicates the predicted start of the bottom	
	phase from the model.	53
Figure 4.11	Identified outlier dives for the linear regression model in Equation 4.5 for the	
	end of the bottom phase. The red line indicates the user-provided end of the	
	bottom phase, and the turquoise line indicates the predicted end of the bottom	
	phase from the model.	54
Figure 4.12	Fitted functional coefficients for the model presented in Equation 4.8 for the	
	different functional covariates.	55
Figure 4.13	Example dive to interpret the functional regression coefficients.	56
Figure 4.14	The outlier dives for the models for the start of the bottom phase	58
Figure 4.15	The outlier dives for the models for the end of the bottom phase	59
Figure 5.1	Results of the simulation study for the basic linear regression model for the	
	probabilistic method.	69

Figure 5.2	Results of the simulation study for the functional regression model for the prob- abilistic approach.	70
Figure 5.3	Results of the simulation study for the user-provided approach. The plots on	
	the top row correspond to the RMSE metric while the plots on the bottom row	
	correspond to the MAE metric. The left column plots are for the start of the	
	bottom phase, and the right column plots are for the end of the bottom phase.	
	The x-axis is the sample size and the y-axis is the value of the metric. Each	
	model is represented by a different coloured line. Error bars are omitted for	
	clarity	71
Figure A.1	Residuals versus fitted values for the start of the bottom phase for the percent	
	of max depth model.	81
Figure A.2	Residuals versus fitted values for the end of the bottom phase for the percent of	
	max depth model	82
Figure A.3	Residual plots for the linear model in Equation 4.5 for the start of the bottom	
	phase	83
Figure A.4	Residual plots for the linear model in Equation 4.5 for the end of the bottom phase	84
Figure A.5	Standardized residual plots versus index for both the start and end of the bottom	
-	phase	85

Acknowledgments

First and foremost, I must thank my two wonderful supervisors—Dr. Nancy Heckman and Dr. Marie Auger-Méthé.

Nancy, you have provided me with some of the most insightful, intriguing, and invigorating conversations and discussions that only someone of your calibre could generate. I am entirely grateful to you and without your constant support and encouragement, I could not have accomplished what I have at UBC. You have showed me that particularities are not something to shy away from, but to embrace. You've allowed me to grow as a statistician and develop my mind in ways I never knew I could. You are always there to bounce ideas off, whether it's a Friday night at 8 PM or a bright and early Sunday morning. I can not only count on you to provide me with intellect and thought-provoking conversation, but I can always count on you to help me with life decisions, or to just have a fun conversation. Thank you.

Marie, thank you for being a great supervisor. I would write more, but I know how much you appreciate succinctness.

In all seriousness, thank you Marie for showing me a world of research that I never knew I would enjoy so much. Marine mammal research has become something I am fascinated by, not something I wish to leave in the dark. Your kindest, openness, and overall demeanor has made my time at UBC some of the best years of my life. Welcoming me into your lab with open arms—literally—and introducing me to everyone else in the lab was a kind and generous "thing to do" which not a lot of students get to experience. The late night dinner parties with all the lab members at your house gave me a sense of belonging for someone who was brand new to the West coast. It meant the world to me, and for that, I am truly grateful. Thank you.

To the wonderful people I have met through the Marine Mammal Research Unit. Sarah and Andrew-thank you for providing me with such amazing experiences as well as the abundance of fascinating and detailed data that statisticians dream about. I truly appreciate it. As well, thank you to all the funding partners and sponsors that made all of this work possible, including Dr. Nancy Heckman's NSERC discovery grant [RGPIN-2020-04629], Fisheries and Oceans Canada

(Whale Science for Tomorrow Initiatives), Hakai Institute, Pacific Salmon Foundation, and Vancouver Whale Watch.

To my UBC friends, thank you. As a newbie to the West coast, I knew nobody at UBC. I was alone. But not for long, as we all began to talk, hangout, laugh, argue, and enjoy each others' company. I would not have changed this cohort for the world. Evan, thank you for being my "Killer Whale Partner in Crime". It was such a relief that I was able to work with someone who is so compassionate, easy-going, and fun to be around. I would also like to thank all the administrators at UBC who were always so kind and knowledgeable about everything.

I would also like to thank all my friends that I have met throughout the years—in person and online—who have supported me and brought smiles to my face countless times. The late night chats, most random online interactions, and the ever-so-fun spontaneous FaceTime calls, all of you have brought something meaningful to my life. I am incredibly grateful to have met each and every one of you, and the support you've given me throughout this journey was immense. In particular, Vanessa, Jill, and Morné—my team, the dream team! You three have been such great friends to me, and I am so glad that I was able to find such a great group of friends that I can count on at all hours of the day. Demetri, Brett, Liam, and all of my other internet friends—thank you for putting up with me and my shenanigans. And of course, my childhood friends who have been with me since the beginning: Dalton, Marisa, Chloe, and Kristin, thank you for being there for me through it all.

Finally, I would like to thank my family for supporting me throughout my journey at UBC. Even though I was living halfway across Canada, I felt like you were only around the corner. To my siblings, Mac and Maggie, I appreciate and love every single moment I get to spend with you. Growing up with the best siblings in the entire world isn't an easy job, but someone has to do it. Buster, my furry friend who definitely cannot read this–love you buddy. And to my Mom and Dad, I love you to the moon and back. Without your support, I would not be the person I am today. You have guided me and showed me how to be the best possible person I could ever be. I am forever grateful and indebted to you, and I can only hope that one day I am half the person that you both are. Your kindness and generosity has been paramount to my success. Thank you and I love you.

Chapter 1

Introduction

The truth, however ugly in itself, is always curious and beautiful to seekers after it. — Agatha Christie, The Murder of Roger Ackroyd

The northern resident killer whales (NRKWs) are a population of killer whales (*Orcinus orca*) which lives along the west coast of the United States and Canada from northern Washington to southern Alaska. The NRKWs are four times more abundant than the endangered southern resident killer whales (SRKWs) [2] and have been thriving. Their population size has been relatively stable with some evidence to suggest it may be increasing [3]. The survival success of these whales is directly related to their ability to find and capture prey during foraging dives. As such, researchers are interested in comparing the key features of foraging dives across these two populations of whales to better understand the survival differences. This may help us understand why the SRKW's population has been relatively stagnant while the NRKW population has been increasing [4].

A major problem with identifying successful prey capture dives is the lack of visual or acoustic confirmation of prey capture events. Many researchers tag these whales with suction-cup attached biologgers equipped with an array of sensors such as 3-axis accelerometers, providing information on their kinematic movements underwater. Along with a hand-full of visually confirmed surface prey captures and acoustic signals that the whale is "crunching" a prey [3, 5, 6], these kinematic movements may be used to identify signature features of prey capture dives, which can then be applied to dives where there is no visual or acoustic data available [1]. To implement these procedures, key variables during the *bottom phase* of a dive have been identified as predictive of prey capture events [1] since the bottom phase of a dive is usually where killer whales capture their prey [3].

Despite the ecological significance of the bottom phase of a dive, the current scientific literature lacks a cohesive definition for the bottom phase of a dive. Some recent studies use definitions that involve the maximum depth of the dive [1, 7], and others involve changes in the vertical speed of

the whale [8, 9]. The definitions vary across researchers and as such, different definitions could result in different values of the summary statistics used in predicting prey captures. The goal of this thesis is threefold: 1) to motivate the potential issues with different bottom phase definitions, 2) identify the main sources of this variation, and 3) propose a reproducible statistical framework that provides researchers flexibility in defining the bottom phase of a dive without using *arbitrary* threshold values which may not accurately represent the bottom phase. For example, a common bottom phase definition for a dive is the first and last time the whale exceeds 70% of its maximum depth [1, 7]. This percentage threshold value may not be appropriate for all dives and may even look entirely incorrect for some dives (Figure 1.1).



Figure 1.1: A time-depth profile of a dive of a NRKW from the datasets used in this thesis, where the bottom phase (in red) has been defined using the 70% maximum depth threshold.

In Chapter 2, I introduce the data collected from the accelerometer tags of three NRKWs. Using these data, methods are applied from a recent paper (Tennessen et al. [1]) to identify successful prey capture dives using the 70% of maximum depth bottom phase definition. In modifying the bottom phase definition of the dives (e.g., increase or decrease the threshold value), I re-predict prey capture dives and find different results. Moreover, since we have confirmed prey captures from one whale, I compared the model predictions of prey capture dives using the different bottom phases to the true prey capture dives. These results demonstrate how bottom phase definitions can impact predicted

prey captures using the methods in Tennessen et al. [1].

Chapter 3 describes an experimental study wherein I asked several researchers to identify the location of the bottom phases of a set of dives. Using these data, I quantified the major sources of variability in bottom phase classifications, including across-researchers, across-dives, and within-researcher variations. This chapter provides further motivation for the need to have a consistent and reproducible method in identifying the bottom phase of a dive that can be tailored to each researcher's own datasets.

Chapter 4 considers several statistical models of the start and end of the bottom phase of a dive, including a modifying percentage threshold model, as well as linear and functional regression models, in order to improve the definition of bottom phase. A dataset of 250 dives is used where a whale researcher has defined the bottom phase for all dives. Moreover, I also construct confusion matrices of predicted prey captures versus actual prey captures as in Chapter 2. The chapter concludes by using cross-validation to compare the models' predictions of bottom phase to the bottom phases provided by the user.

Chapter 5 provides an answer to a commonly asked question when building statistical models: How large of a dataset is required to build the models and have them perform well? Since many whale researchers have thousands of dives in their datasets, it would be impractical to mark the bottom phase for all dives. As such, a researcher must determine an adequate sample size when building these statistical models. Chapter 5 considers two possible approaches to this question with simulation studies to see how each model performs as the sample size of the training data changes.

In Chapter 6, I conclude with an overview of the significance of this thesis. As well, I discuss several possible next steps that can be taken to improve upon the methods I developed.

Chapter 2

Exploratory Analysis

It is the brain, the little gray cells on which one must rely. One must seek the truth within-not without. — Agatha Christie, Poirot Investigates

2.1 Motivation & Background

In the study of marine diving animals, dive phase definitions are critical since several key statistics used to determine the animals' behavioural states are computed during different dive phases. For example, dive efficiency, which is the ratio of time during the bottom phase to the entire dive duration, helps to quantify cost-efficiency of foraging behaviour in several species including otters (*Lutra lutra*) [10], tiger sharks (*Galeocerdo cuvier*) [11], emperor penguins (*Aptenodytes forsteri*) [12], sperm whales (*Physeter macrocephalus*) [13], bottle-nose dolphins (*Tursiops truncatus*) [14], and killer whales (*Orcinus orca*) [15]. In changing the definition of the bottom phase of a dive, the resulting analyses can be skewed and as such, may not be entirely accurate of the animals' behaviour. Hence, defining the bottom phase of a dive correctly is of utmost importance for marine mammal researchers and needs to be approached in a methodical and thorough manner.

In this chapter, I begin by describing the data that have been collected from three NRKWs. Afterwards, I provide a brief introduction to the important variables that have been collected from these whales, as well as the variables which could be important for determining foraging success. Next, I discuss how each dive can be divided into its respective phases since it is believed that the majority of feeding occurs during the bottom phase of a dive [3]. I then provide a review of the rules derived in Tennessen et al. [1] to detect foraging dives and apply it to the datasets. Finally, I consider what would happen if the definition of the bottom phase changes. Specifically, I note how this affects which dives are determined to be foraging based on Tennessen et al. [1]'s rules.

I also look at the distributions of the kinematic variables used to detect foraging and how these distributions change as the bottom phase definition changes. I conclude this chapter by presenting the main concerns with these rules—they rely on very precise definitions of dive phases.

2.2 Data Description

Each of the three datasets used in this thesis contain information on a single, tagged whale. Each whale was tagged with a Customized Animal Tracking Solutions (CATs) tag [16]. The CATs tag was attached to the first whale (referred to as Whale #1) for approximately 6 hours, from 13:16 PM until 18:30 PM (Pacific Daylight Time) on September 2nd, 2019. The second whale (referred to as Whale #2) was tagged for about 5 hours, from 10:30 AM to 15:15 PM (Pacific Daylight Time) on August 31st, 2019. Finally, the third whale (referred to as Whale #3) was tagged for approximately 13 hours from 10:15 AM to 11:30 PM (Pacific Daylight Time) on August 25th, 2020. Notice that Whales #1 and #2 were tagged in 2019 and Whale #3 was tagged in 2020. The CATs tag collected information from the whales at 50 Hertz (Hz) meaning there were 50 measurements taken every second. As such, these datasets are large, each containing hundreds of thousands of rows of data. The CATs tag collected information on the following variables:

- Date and time (local)
- Depth (meters)
- Acceleration measured by an accelerometer (3 axes) (m/s^2)
- Angular Velocity measured by a gyroscope (3 axes) (mrad/s)
- Magnetic Field measured by a magnetometer (3 axes) (microTesla [μT])

The CATs tag also has a forward facing camera to allow for visually confirmed prey captures when adequate light was available. Moreover, these tags contained a passive acoustic recorder to acoustically identify when these whales were hunting (via echolocation) and when successful prey captures were made (via prey handling sounds like crunches). Figures 2.1, 2.2, and 2.3 show the complete time-depth plot for the whales during the entire time of tag deployment. We can see that there are periods of time where data were not collected on Whales #1 and #2, which could have been due to the CATs tag malfunctioning. It would be difficult to impute the missing data during those time periods as it is difficult to determine if only a single dive occurred during that time period or if it contains multiple dives.



Figure 2.1: The time-depth plot of Whale #1 during the tag attachment period.



Figure 2.2: The time-depth plot of Whale #2 during the tag attachment period.



Figure 2.3: The time-depth plot of Whale #3 during the tag attachment period.

2.3 Longitudinal Variables Calculated from Raw Data

The key question for these whales is if there are any potential foraging activities occurring during the tagged period. These foraging dives tend to have similar kinematic characteristics. By understanding the kinematic behaviour of the whales during foraging, then researchers can better understand how much energy is being expended and how this affects the survival of the whale.

The kinematic variables that are used to detect prey captures generally utilize dynamic acceleration. The total acceleration of any object is the sum of two distinct accelerations: static and dynamic. Static acceleration affects all objects on earth and is due to gravity. Dynamic acceleration is the acceleration directly due to the movements of that object, and is thus the most crucial value as it contains the information about the whales' kinematic behaviours without any influence of gravity. Any further mention of acceleration in this thesis will be referring to dynamic acceleration. The CATs tags do not automatically collect dynamic acceleration data from the whales; however, I used the CATs Visualizer (CV) to decompose total acceleration into static and dynamic components [16].

Following the methods described by Tennessen et al. [1], there are three variables, each measured as a function of time, that are important to identify a successful prey capture dive, and one variable which is used as a proxy for energy expenditure [17]. These variables are defined below.

1. Jerk: Jerk is defined as the rate of change of dynamic acceleration. Jerk can be calculated by taking the derivative along each axis of the dynamic acceleration vector. There is a measure-

ment for jerk at each time point in the dataset.

Roll: Roll is defined as the rotation about the *x*-axis (longitudinal axis) of the whale, measured in degrees, ranging from -180° to 180°. We define right-side up as 0°, so a value of 180° or -180° indicates that the whale is upside down. Below is a picture with arrows indicating the different rotations about the different axes.



Figure 2.4: A visualization of the three axes of rotation as well as the three directions of movement for an animal freely moving in three dimensions.

3. Heading: Heading is defined as the forward direction or bearing of the whale, relative to the magnetic North Pole. If the whale is travelling towards the North Pole, then the heading would be 0°. Heading ranges from -180° to 180°. This can be calculated mathematically using the magnetometer data as follows:

Heading =
$$\frac{180}{\pi} \arctan 2(m_x, m_y)$$

where $\arctan 2$ is defined as the 2-argument arctangent function, and m_x and m_y are the magnetometer readings for the x and y axis respectively. This formula will produce an angle determined by the x-axis and the point (m_x, m_y) .

4. Vectorized Dynamic Body Acceleration (VeDBA): VeDBA is defined as the Euclidean norm of the dynamic acceleration vector. VeDBA is a proxy for energy expenditure [17]. It can be calculated for each time point and can be written mathematically as

$$VeDBA = \sqrt{a_x^2 + a_y^2 + a_z^2}$$

where a_x , a_y , and a_z are the components of the dynamic acceleration vector.

2.4 Dive Level Characteristics

These datasets contain multiple, separate dives that need to be isolated since a dive is the unit of interest. A dive is defined as the period when the whale descends below the surface, surpasses the user-defined minimum depth requirement, and then returns to the surface. The surface could be defined as 0 meters, but since a dive begins and ends when a whale intakes oxygen, the surface is defined as any point between sea level and the length of the animal. Between these two points, the whale could potentially be breathing, hence, it has returned to the surface. For these datasets, the surface threshold was chosen to be 1 meter, even though killer whales can be upwards of 9 meters in length [18]. Moreover, the minimum depth requirement for a dive was chosen to be 1.5 meters. These criteria were established in collaboration with biologists who agreed that these values were appropriate for these datasets.

I have excluded from the analysis any dive that contained missing data. Upon applying the surface threshold and minimum depth criteria, we found that there were 300 dives during the tag deployment for Whale #1, 246 dives for Whale #2, and 250 dives for Whale #3. Table 2.1 provides summary statistics of several commonly reported variables regarding dives.

	Whale #1 (300 dives)		W (24	Whale #2 (246 dives)		Whale #3 (250 dives)	
	Mean	SD	Mea	un SD	Mean	SD	
Max Depth (m)	4.08	3.82	5.7	/8 9.66	25.39	41.75	
Dive Duration (s)	39.12	42.41	36.0	08 57.36	92.24	112.79	
Max Jerk – Whole Dive (m/s^3)	37.78	35.81	26.4	6 33.03	664.87	572.49	
VeDBA – Whole Dive (m/s^2)	0.72	0.16	1.5	0.63	1.04	0.85	

Table 2.1: Several summary statistics of important variables for each whale. The VeDBA for a single dive is calculated as the average VeDBA over the duration of the dive.

From Table 2.1, we can see that Whale #3 tends to have much deeper dives, on average, with a higher variation in the max depth than both Whale #1 and Whale #2. This is evident from Figures 2.1, 2.2, and 2.3 as Whale #3 has many deep dives exceeding the deepest dives from the other two whales. Moreover, we can also look at the mean VeDBA between these whales, which is higher in Whale #2. This could indicate that this whale was expending more energy than Whale #1 and Whale #3 as VeDBA is a proxy for energy expenditure. Furthermore, Whale #3 has much richer data than the other two whales. This tag was attached for over 12 hours and does not have any missing data. Moreover, this whale had significantly deeper and longer dives. From 2.1 and 2.2, the majority of missing data for these whales occurs during very deep dives. Because of this, the statistics presented in this table could be biased.

The next step is to divide each dive into its corresponding phases: descent, bottom, and ascent. The bottom phase is likely where the majority of foraging occurs, and indeed Tennessen et al. [1] uses summary statistics during the bottom phase of a dive to determine foraging activity.

Tennessen et al. [1], following Arranz et al. [7], defines the beginning of the bottom phase of a dive as the first time the whale achieves 70% of the maximum depth of that dive, and the end of the bottom phase as the last time the whale reaches 70% of the maximum depth of that dive. Arranz et al. [7] justify this definition by citing Hooker and Baird [19] where the authors set out rules for the bottom phase to be 85% of maximum depth. Tennessen et al. [1] modifies this percentage threshold for the bottom phase so that it fits better with their datasets. They chose this value by visually confirming the bottom phase of a random sample of 25% of their dives. To be consistent with Tennessen et al. [1], I decided on using the 70% of maximum depth threshold definition when applying their methods to detect foraging activity. This concept of defining the proper bottom phase is a central aspect to this thesis which will be discussed in more detail in Chapters 3 and 4, including the potential benefits and downfalls to having a rigid definition for the bottom phase of a dive.

Once the bottom phase has been identified, the next steps taken in Tennessen et al. [1] were to determine which kinematic variables were highly predictive of successful prey captures. The primary kinematic variables of interest, as described in Tennessen et al. [1], are as follows.

- 1. Jerk Peak: Jerk Peak is defined as the maximum jerk during the bottom phase of a dive, scaled by the median jerk of the bottom phase of that dive. This is measured in ms⁻³. To compute jerk peak, the magnitude of the jerk is calculated at each time point, then the maximum jerk during the bottom phase is found. Finally, the maximum jerk during the bottom phase is divided by the median jerk during the bottom phase so that jerk peaks are comparable across multiple dives. As such, there is a clear distinction between maximum jerk and jerk peak.
- 2. Bottom to Whole Ratio: Bottom to Whole Ratio (BTWR) is defined as the ratio of the duration

of the bottom phase of a dive to the duration of the entire dive. This is a value between 0 and 1, exclusively.

- 3. Rate of Ascent: Rate of Ascent is defined as the average speed, in m/s, that the whale achieves during its ascent phase.
- 4. Roll at Jerk Peak: Roll at Jerk Peak is defined as the absolute value of the instantaneous roll at the same time of the jerk peak. This is measured in degrees (°). Since we are using absolute values, the roll at jerk peak is strictly positive.
- 5. Mean VeDBA: Mean VeDBA is defined as the average VeDBA during the bottom phase of the dive.
- 6. Heading Variance: Heading Variance is defined as the circular variance of the sequence of heading values during the bottom phase of the dive. Circular variance must be used when calculating heading variance since this is a circular quantity. Circular variances can only take on values between 0 and 1, where 0 indicates that all the data points are concentrated at one point, whereas a value of 1 indicates that there is a large spread amongst all of the points. A large heading variance for a whale indicates that the whale has made several changes in its heading during the bottom phase of the dive. Such variance could imply that the whale is chasing prey, which may be moving erratically to evade being captured.

2.5 Detection of Foraging Dives

To detect foraging dives, Tennessen et al. [1] built a generalized linear mixed effects model (GLMM) with a binary response variable (successful prey capture or not), a logistic link function, and several fixed and random effects. The authors labelled each dive in their dataset as either having a successful prey capture or not. To determine which dives were successful prey capture dives, they identified 12 dives with a visually confirmed prey capture event. The fixed effects included in the model are the kinematic variables described in the previous section, other kinematic variables that are of less interest, as well as the sex of the whale. The year and the tag deployment were included in the model as random effects.

Using this GLMM, the authors determined that only three of these variables are significant predictors of successful prey capture events: Jerk Peak, Roll at Jerk Peak, and Heading Variance. They then derived a "prey detector" from this model. Given a single dive, if all of these three predictors met or exceeded a minimum threshold value, then this model would classify the dive as a successful prey capture dive. The three criteria which a dive must satisfy in order to be classified as a successful prey capture dive are:

- 1. Jerk Peak $\geq 14.38 \ ms^{-3}$,
- 2. Roll at Jerk Peak $\geq 22.93^{\circ}$,
- 3. Heading Variance ≥ 0.40 .

The authors also verified the model by using it to predict prey capture success for all dives in their dataset, then confirming the results with acoustic data. Acoustic data continuously collected sounds in the immediate vicinity of the whales during the entire deployment. In doing so, they were able to identify the true positive and false positive rates for their model which were 78.7% and 0.2% respectively.

The models used in Tennessen et al. [1] provide a basis for detecting prey capture dives based on those three kinematic variables. It would be useful to generate and build our own prey capture detection model for our data, however, we do not have any confirmed prey capture events for Whale #1 or #2, and we only have 7 confirmed prey capture events for Whale #3. Building a logistic regression model from this data will not provide meaningful results since there are too few successes and too many failures in our datasets. As such, we rely on Tennessen et al. [1]'s rules to detect successful prey captures.

Upon applying these three criteria to our datasets, their model predicts 2 successful prey capture dives from Whale #1, 0 from Whale #2, and 21 from Whale #3. From Whale #1, I labelled these dives as dive 64 and dive 142 due to their relative location in the tag deployment. These dives are both shallow, reaching approximately 6 meters in depth (Figs 2.5 and 2.6). Occasionally, killer whales are known to chase their prey towards the surface of the water, then consume the prey during the next, shallow dive [5]. As such, it is possible that both of these dives from Whale #1 were in fact successful prey captures, however, without visual or acoustic confirmation, it is difficult to confirm. Since these dives were shallow and prey captures tend to occur during deeper dives, [3], these dives could also be false positives, or the whales could be exhibiting some other associated behaviour. Furthermore, the dives immediately preceding both dive 64 and dive 142 do not exhibit signs of potential foraging activity based on the dive metrics used.

Since Whale #3 had many deep, long dives, we expected there to be more predicted prey captures from this model, and indeed there are 21. Of these 21 dives, 6 of them are in fact successful prey capture dives. From Whale #3, we have 250 dives in total, of which, 7 are confirmed prey capture dives, 179 are confirmed non-prey capture dives, and 64 are not known since there was no acoustic data for validation during these dives. We can tabulate the total dives classified by the prediction algorithm against the true state of these dives (Table 2.2).

From this table, we can see that this model correctly predicted 6 out of the 7 successful prey capture dives, but failed to identify 1 successful prey capture. This model predicted 15 additional



Figure 2.5: Dive 64 from Whale #1 which occurred from approximately 13:51:43 PM until 13:52:29 PM (about 46 seconds in duration).



Figure 2.6: Dive 142 from Whale #1 which occurred from approximately 15:01:59 PM until 15:02:31 PM (about 32 seconds in duration).

			Truth		
		Prey Capture	Non-Prey Capture	Unknown	Total
Dradiction	Prey Capture	6	15	0	21
Flediction	Non-Prey Capture	1	164	64	229
	Total	7	179	64	250

Table 2.2: Model predictions versus actual prey capture status for dives from Whale #3

dives as successful prey captures when they were not. There are also 64 dives which do not have a true label attached to them. However, this model failed to predict any of them as prey captures. As such, we can claim that the false positive rate, among the dives for which we have a correct classification of prey capture status, is about 8.38 % (15 / 179) since we incorrectly identified 15 dives as prey capture dives when they in fact were not.

Figure 2.7 shows the time-depth plots for all the predicted prey capture dives from Whale #3 using the 70% depth threshold. All the dives have been scaled on the *x*-axis to be between 0 and 1 for demonstration purposes. Dives 47, 54, 62, 124, 125, and 136 are actual successful prey captures and they are highlighted in red. A clear commonality between these dives is that they are all relatively deep, except for Dive 126 which is not a successful prey capture dive.

2.6 Modifying the Bottom Phase Definition

A critical aspect to the methods used in Tennessen et al. [1] is the bottom phase definition. They define the bottom phase using a 70% of max depth threshold. As mentioned earlier, this threshold value was derived based on past studies and visual assessments, which could be considered *ad hoc*. In this section, we will repeat the previous analysis for detecting successful prey capture dives as we did in Section 2.5 but now we consider different threshold values for the definition of the bottom phase of a dive. The thresholds are 50%, 60%, 70%, 80% and 90% of maximum depth.

Moreover, we use other bottom phase definitions which do not depend on depth. For example, we use vertical velocity, as calculated from the time-depth data, as a threshold for the start and end of the bottom phase. Viviant et al. [8] and Viviant et al. [9] define the start of the bottom phase for Antarctic fur seals (*Arctocephalus gazella*) as the first time the seal surpasses 0.4 m/s, and the end of the bottom phase as the last time the animal reaches 0.4 m/s. Le Bras et al. [20] also uses a similar approach except they use the value 0.75 m/s. We denote this as the vertical velocity threshold. However, for some dives, these velocities are never achieved, and thus, those dives would not have a bottom phase. To address this issue, we use a similar technique as we did for the max depth thresholds by using the percentage of minimum (or maximum) velocity in a dive as the threshold. For example, suppose the velocity threshold is 50%. The whale begins the bottom phase when it



Figure 2.7: Predicted prey capture dives for Whale #3 using the 70% depth threshold and applying the methods in Tennessen et al. [1]. Confirmed prey capture dives are highlighted in red.

reaches 50% of its minimum velocity for the first time, and it finishes the bottom phase when it reaches 50% of its maximum velocity for the last time.

There is a necessary distinction between minimum and maximum velocity because during the descent of a dive, the vertical velocity is negative, and during the ascent of a dive, the vertical velocity is positive. Moreover, the magnitude of the minimum and maximum velocities will not necessarily be equal. Biologically, whales usually do not ascend and descend at the same velocities [21]. For example, the minimum velocity could be -1 m/s whereas the maximum velocity could be 3 m/s. If the bottom phase begins and ends at 50% of the maximum velocity, then there would not be a well-defined start of the bottom phase for this dive since the whale never achieves -1.5 m/s during the descent. Hence, we could run into the same issue as before of either not having a start or end of the bottom phase since the whale never actually reaches the percentage threshold of the velocity.

To decide on threshold values for velocity, we considered several plots of time-depth data for individual dives. On these plots, we looked at how the bottom phase definitions change for different values of these thresholds. As the threshold values increase, the bottom phase becomes wider and wider. When the threshold values for velocity are small, the bottom phase is narrower. This threshold value for velocity is in contrast to the max depth thresholds, which result in a smaller, narrower bottom phases as the threshold value increases. As such, we decided on five thresholds which would be reasonable for this data: 10%, 20%, 30%, 40%, and 50%. Since this is for demonstration purposes, the exact values are not critical, rather, it is important to be using a wide range of values.

Figures 2.8, 2.9, and 2.10 show the distribution of the dive metrics for the whales using the varying bottom phase definitions that depend on the maximum depth thresholds. Notice that these dive metrics have been log-transformed since they are heavily skewed.



Figure 2.8: Boxplots of the distributions of log jerk peaks for various bottom phase definition thresholds involving maximum depth. Horizontal line indicates one of the criteria a dive needs to meet to be classified as a successful prey capture, based on Tennessen et al. [1].



Figure 2.9: Boxplots of the distributions of rolls at jerk peak for various bottom phase definition thresholds involving maximum depth. Horizontal line indicates one of the criteria a dive needs to meet to be classified as a successful prey capture, based on Tennessen et al. [1].



Figure 2.10: Boxplots of the distributions of heading variance for various bottom phase definition thresholds involving maximum depth. Horizontal line indicates one of the criteria a dive needs to meet to be classified as a successful prey capture, based on Tennessen et al. [1].

We see that the distributions of these variables tend to change, albeit minimally, for different definitions of bottom phase based on maximum depth. A fair number of dives surpass the successful prey capture criterion for jerk peak, as indicated by all the points lying above the horizontal line. As we increase the percentage of maximum depth threshold, we notice that the distribution begins to shift towards zero as the bottom phase is getting smaller and smaller, thus excluding more dives from satisfying this criteria. In contrast, we note that as the maximum depth threshold increases, the distributions for the log of the heading variance shift away from zero, becoming more negative (Fig 2.10). However, since the heading variance is on the log scale, it indicates that the untransformed heading variance is actually trending towards 0. Intuitively, we know that as the maximum depth threshold increases, the bottom phase becomes narrower and narrower. The heading variance is only calculated during the bottom phase of a dive. Since we only use a very narrow interval of data points to compute the heading variance, we may find an overall smaller value for heading variance unless there are sudden and rapid changes in this interval.

Similarly for the velocity thresholds, boxplots of the distributions of the dive metrics are given in Figures 2.11, 2.12, and 2.13.



Figure 2.11: Boxplots of the distributions of log jerk peaks for various bottom phase definition thresholds involving velocity. Horizontal line indicates the criterion used in Tennessen et al. [1].



Figure 2.12: Boxplots of the distributions of rolls at jerk peak for various bottom phase definition thresholds involving velocity. Horizontal line indicates the criterion used in Tennessen et al. [1].



Figure 2.13: Boxplots of the distributions of heading variance for various bottom phase definition thresholds involving velocity. Horizontal line indicates the criterona used in Tennessen et al. [1].

These plots provide a similar interpretation as the max depth threshold plots. As the speed threshold increases, we generally see a widening of the bottom phase since the vertical speed of the whale close to the bottom of the dive tends towards zero resulting in a low percentage threshold. On the other hand, as the percentage threshold increases, the bottom phase tends to get wider.

Next, for each of these bottom phase definitions, I applied the rules from Tennessen et al. [1] to detect successful prey capture dives.
	Whale #1	Whale #2	Whale #3
Depth Threshold	1		
50%	None	126, 210	14 44 46 47 53 54 62 63 77 88 117 119 121 122 124 125 126 129 136 148 169 171 172 180 184 185 186 187 194
60%	142	126, 210	14 16 44 46 47 54 62 63 77 88 117 119 121 122 124 125 126 129 136 148 169 171 172 180 184 186 187 188
70%	64, 142	None	14 16 44 46 <mark>47 54 62</mark> 63 77 119 121 122 <mark>124</mark> 125 126 129 136 148 169 180 184
80%	142	None	14 16 44 46 <mark>47 54 62</mark> 63 77 119 121 122 <mark>124</mark> 125 126 129 136 148 169 171 184
90%	None	None	14 16 46 47 54 62 77 119 121 122 124 125 126 129 136 148
Speed Threshold	1		
10%	None	None	14 44 46 <mark>47 54 62</mark> 77 119 121 122 124 125 126 129 136 148 169 172 184 185 186
20%	None	None	14 43 44 46 47 54 62 77 95 106 113 119 121 122 124 125 126 129 136 148 169 172 184 185 187
30%	None	126	14 44 46 47 54 62 63 77 88 95 106 113 119 121 122 124 125 126 129 136 148 169 172 184 185 186 187 188 194
40%	None	126	14 44 47 54 62 63 77 88 106 113 119 121 122 124 125 126 129 136 148 169 172 184 185 186 187 194
50%	None	126	14 44 47 54 62 63 77 88 106 113 119 121 122 124 125 126 129 136 148 169 171 172 184 185 186 187

Table 2.3: Dive numbers for the predicted successful prey capture dives as per the rules determined by Tennessen et al. [1] for Whale #1, #2, and #3 for multiple different bottom phase definitions. Dives in red are confirmed prey capture dives, and dives in blue are not confirmed to be either prey capture or non-prey capture dives.

The identification of predicted successful prey capture dives is not consistent across all bottom phase definitions (Table 2.3). We can see that successful prey captures predicted by the methods described in Tennessen et al. [1] vary widely depending on which definition of bottom phase is used, especially when varying the depth thresholds. However, using the velocity threshold results in more consistent predicted successful prey capture dives, although the results conflict with the predictions using the depth thresholds.

Without visual or acoustic confirmation for Whale #1 and #2, we do not know which of these dives are actually successful prey captures. For Whale #3, however, the rules were able to detect 6 of the 7 successful prey capture dives (47, 54, 62, 124, 125, and 136) for every bottom phase definition, as shown in red. On the other hand, the number of additional predicted prey capture dives does vary as the definitions of bottom phase change. Notice that the dives shown in blue are not confirmed to be either prey capture or non-prey capture dives, as there was no acoustic data for validation, as explained earlier.

The one prey capture dive that this model was not able to detect was dive 67. Since this dive is consistently misidentified as a non-prey capture dive, I looked at the values for the three important dive metrics – jerk peak, roll at jerk peak, and heading variance – for each bottom phase definition (Table 2.4). This dive is different from the other prey captures since the jerk peak and roll at jerk peak are quite low for all definitions. Furthermore, the heading variance is also low, although it is near the required threshold for several of the bottom phase definitions. The time-depth plot of this dive (Figure 2.14) is indicative of a prey capture dive, however, this dive did not exhibit extreme kinematic variables resulting in the failure to detect it based on these methods.

	Jerk Peak	Roll at Jerk Peak	Heading Variance
Depth Threshold			
50%	9.41	5.48	0.31
60%	9.41	5.48	0.29
70%	9.41	5.48	0.30
80%	9.41	5.48	0.29
90%	9.16	15.13	0.21
Speed Threshold			
10%	9.17	15.13	0.14
20%	9.17	15.13	0.14
30%	9.17	15.13	0.19
40%	9.41	5.48	0.36
50%	9.41	5.48	0.36

Table 2.4: Dive metrics from dive 67 from Whale #3.



Figure 2.14: Time-Depth plot of dive 67 from Whale #3

Of the three kinematic variables used to predict successful prey capture dives, jerk peak tends to be the least volatile measurement as the bottom definition changes. Since jerk peak is computed by finding the maximum jerk during the bottom phase, then dividing by the median jerk of the bottom phase, as long as the new definition of bottom phase does not exclude this maximum jerk value, the jerk peak will remain relatively stable since the median jerk is robust to these changes. Similarly, roll at jerk peak tends to remain relatively constant for each dive. The heading variance seems to be the variable which changes abruptly as the bottom phase changes because it is computed by finding the circular variance using the data points during the bottom phase. If the bottom phase includes more or less points, then this value may change accordingly.

The values for jerk peak, roll at jerk peak, and heading variance for each of the aforementioned potential prey capture dives, for Whale #1 and #2, stratified by the bottom phase definition for max depth thresholds are shown below (Table 2.5). The same variables for the single dive which was predicted to be a successful prey capture using the speed thresholds is also shown (Table 2.6). From these tables, it is apparent that dive 126 from Whale #2 could potentially be a successful prey capture dive since it constantly appears to satisfy the three dive metric criteria for several definitions of bottom phase. Notice however that it does not appear as a successful prey capture dive when using the 70% of max depth threshold which is why it was not detected previously. The time-depth plot of dive 126 is given in Figure 2.15 with a plot of the jerk as well. This dive is much more indicative of a successful prey capture dive than the previously identified dives since it is almost 20 meters deep and is approximately 2 minutes in duration. We can also see the reason why this dive was not predicted to be a successful prey capture dive at the 70% depth threshold, but it was for the 60% depth threshold. Since the 70% threshold barely excluded that large jump in jerk, the jerk peak metric was not large enough to predict the dive as a prey capture. However, upon widening the bottom phase using the 60% threshold, we include this sharp measure in jerk, allowing the model to detect this dive as a successful prey capture dive. This example only further demonstrates the importance of a well-defined bottom phase as well as well-defined metrics used for detecting prey capture dives.

	(W1) Dive 64			(W1) Dive 142		(W2) Dive 126			(W2) Dive 210					
	JP	RaJP	HV		JP	RaJP	HV	-	JP	RaJP	HV	JP	RaJP	HV
Depth Thresholds														
50	9.28	148.97	0.30	44.	01	170.22	0.32		22.98	29.75	0.56	20.41	39.18	0.52
60	14.13	148.97	0.59	40.	88	170.22	0.40		23.05	29.75	0.57	19.38	39.18	0.431
70	14.44	148.97	0.47	29.	55	170.22	0.55		8.02	7.59	0.57	17.26	42.39	0.25
80	13.43	148.97	0.37	17.	03	170.22	0.72		8.02	7.59	0.58	3.68	1.54	0.17
90	14.69	148.97	0.17	13.	58	170.22	0.73		6.45	14.17	0.75	4.11	1.54	0.01

Table 2.5: Values for the dive metrics used to identify successful prey captures using the criteria established by Tennessen et al. [1] stratified by different bottom phase definitions according to maximum depth thresholds. Rows highlighted in blue indicate predicted successful prey capture dives.

	(W2) Dive 126				
	JP RaJP HV				
Speed Thresholds					
10	8.22	7.59	0.62		
20	8.13	7.59	0.59		
30	23.96	29.75	0.46		
40	23.84	29.75	0.45		
50	23.45	29.75	0.44		

Table 2.6: Values for the dive metrics used to identify successful prey captures using the criteria established by Tennessen et al. [1] stratified by different bottom phase definitions according to vertical velocity thresholds. Rows highlighted in blue indicate predicted successful prey capture dives.



Figure 2.15: Dive 126 from Whale #2. Plots of depth and jerk with two definitions of bottom phase labeled via vertical lines. The red line indicates a 60% depth threshold whereas the blue line indicates a 70% depth threshold. The 70% threshold excluded the sharp jerk measure as seen between the two vertical lines on the right-hand side. This is the reason why the dive was not predicted as a successful prey capture using this threshold, but it was predicted using the other threshold which contained this sharp jerk measurement.

2.7 Conclusions

Through this exploratory analysis of the datasets, I uncovered several inherent limitations of using a static threshold for bottom phase definitions and that these definitions can affect the results of prey capture detection algorithms. In particular, by applying the thresholds in Tennessen et al. [1] to our dataset while varying the definition of the bottom phase of a dive, we saw that the predicted successful prey capture dives were dependent on which definition was used. There appears to be no consensus on the definition of a bottom phase in the animal movement literature (see definitions used in [7], [8], [9], [22], and [23], as well as the methods used in the R Package diveMove [24]). Thus, there is a need to develop a more quantitative and flexible definition of the bottom phase of a dive. I devote the remainder of the thesis to the bottom phase definition and present potential remedies using formal statistical models.

Chapter 3

Getting to the Bottom of It: An Experimental Study

No man can hope to find out the truth without investigation — George F. Richards

3.1 Background & Motivation

The definition of the bottom phase of a dive is dependent not only on the type of animal being studied but also on the authors' preferences for their specific datasets. For example, Tennessen et al. [1] define the bottom phase of a dive as the part between the first and last occurrences of the whale reaching 70% of the maximum depth for that dive. In the seminal methodological review, Hooker and Baird [19] use an 85% threshold rather than 70%. The authors also state that this percentage is subject to change depending on the dives that are being studied as well as the authors' preferences, but do not provide explicit guidance on how to modify this percentage. Halsey et al. [25] provides a thorough method for understanding and classifying diving behaviour of seabirds with methods that can be applied to dives in general, also indicating that bottom phases of dives can be quite complex and require detailed descriptions depending on several factors. Many authors visually assess a subset of their dives to determine a reasonable bottom phase definition. These definitions of bottom phase are variable across not only different researchers, but also across dives. In this chapter, I use an experimental approach to formally investigate these variations.

3.2 Experimental Design

To quantify the sources of variability in the classification of dive phases, I designed a study in which multiple researchers classify the bottom phase of several different dives. I recruited 18 volunteer marine mammal researchers via word-of-mouth (Ethics Approval Permit ID: H20-02096). We used diving data from Whale #1 and #2 for this experiment since we did not have data from Whale #3 at the time. The researchers were asked several preliminary questions such as how much experience they have working with diving data. Researchers in this study were experts in this research area, with 12 participants having over 10 years of experience. Each researcher was presented with a sequence of 75 dives. These dives were shown one at a time and in a random order. For each dive, the researcher was asked to identify the bottom phase. Once the researcher was confident with their decision, they submitted their response and proceeded to the next dive, repeating the same procedure until all 75 dives were classified. Once the researcher completed the study, they were asked how confident they were in their responses.

Of the 75 dives that each researcher saw, 25 dives were only classified by that researcher, and another 25 dives were classified by all researchers, twice. We refer to the former as the *unique* dives and the latter as the *common* dives. The 25 common dives allow us to assess two contributions of variability. By comparing how researchers classified the 25 common dives, we quantify the variability across researchers. By comparing how the same researcher classifies the same dive twice, we can quantify the consistency of researchers. The 25 unique dives allow us to quantify how the dive phase definitions vary dive-to-dive. We will use these data to make inferences about the variability across researchers, within researcher, and across dives.

To conduct this study, I created a web application that makes it easier for someone to classify the bottom phase of a dive, with the resulting data being stored remotely. Using the shiny package in R, I built the app which plots an individual dive along with 2 vertical lines that determine the bottom phase. The user is able to move these verticals lines, then submit their information once they are satisfied with their choice (Figure 3.1).

3.3 Notation & Definitions

In the model to determine the variability of bottom phase definitions, the response variable of interest, denoted Y, is the proportion of time into a dive when the bottom phase begins. The analysis is the same for when Y is the proportion of time into a dive when the bottom phase ends. Note that Yis untransformed even though it is a proportion since the following analysis is quite similar with and without a transformation, but the untransformed model is easy to interpret. There are r researchers in total indexed by $i = 1, \dots, r$.

Determine Beginning and End of the Bottom Phase



Figure 3.1: Example of a dive that a researcher would be presented with on the Shiny App. The user is free to move the vertical lines to determine where they believe the bottom phase of a dive to be.

The general forms of this model for the common and unique dives are defined separately. For the common dives, the model is

$$Y_{i\ell k}^{C} = \mu + R_{i} + D_{\ell}^{C} + \left(RD^{C}\right)_{i\ell} + \varepsilon_{i\ell k}^{C}$$

$$(3.1)$$

and for the unique dives, the model is

$$Y_{ij}^{U} = \mu + R_i + D_{ij}^{U} + (RD^{U})_{ij} + \varepsilon_{ij}^{U}$$
(3.2)

where $Y_{i\ell k}^{C}$ is the *i*th researcher's *k*th response for the ℓ^{th} common dive, where $i = 1, \dots, r$, $\ell = 1, \dots, d_C$, $k = 1, \dots, K_{i\ell}$, and the total number of common dives that each researcher sees is denoted d_C . For our study where all common dives are observed twice, $K_{i\ell} = 2$. Moreover, Y_{ij}^{U} is the response from the *i*th researcher's *j*th unique dive for $j = 1, \dots, m_i$, so that researcher *i* observes m_i unique dives. I acknowledge that the index *j* depends on *i*, however, we omit the dependency in the notation for simplicity. Finally, μ is the mean proportion of time into the dive for the start (or end) of the bottom phase.

I define these two models separately to emphasize the different indices which represent the complex structure of this dataset, specifically the occurrence of repeated common dives. In essence, these two models are equivalent, and their marginal distributions are technically the same.

The random effects from these models are defined as follows:

• R_i is the random effect for researcher i, i = 1, ..., r. We assume that

$$R_i \sim N(0, \sigma_R^2).$$

• D_{ℓ}^{C} is the random effect for the ℓ^{th} common dive, $\ell = 1, ..., d_{C}$. We assume that

$$D_{\ell}^C \sim N(0, \sigma_D^2).$$

• D_{ij}^U is the random effect for the i^{th} researcher's j^{th} unique dive, i = 1, ..., r and $j = 1, ..., m_i$. We assume that

$$D_{ii}^U \sim N(0, \sigma_D^2),$$

noting that the σ_D^2 here is the same as the one presented above since all dives are assumed to come from the same distribution, regardless of if they are unique or common to all researchers.

• $(RD^{C})_{i\ell}$ is the random effect for the interactions between researcher and common dive for

 $i = 1, ..., r, \ell = 1, ... d_C$. We assume that

$$(RD^C)_{i\ell} \sim N(0, \sigma_{RD}^2).$$

• $(RD^U)_{ij}$ is the random effect for the interactions between researcher and unique dive for $i = 1, ..., r, j = 1, ..., m_i$. We assume that

$$(RD^U)_{ii} \sim N(0, \sigma_{RD}^2).$$

Again, σ_{RD}^2 is the same as above for similar reasons as already stated.

• $\varepsilon_{i\ell k}^C$ is the random error of the response variable associated with the common dives respectively, for $i = 1, ..., r, k = 1, ..., K_{i\ell}, \ell = 1, ..., d_C$, and $k = 1, ..., K_{i\ell}$. We also assume that

$$\varepsilon_{i\ell k}^C \sim N(0, \sigma_E^2).$$

• ε_{ij}^U is the random error of the response variable associated with the unique dives for i = 1, ..., r, $j = 1, ..., m_i$. We also assume that

$$\varepsilon_{ij}^U \sim N(0, \sigma_E^2).$$

Again, σ_E^2 is the same as above for similar reasons as already stated.

We also assume all of the above random effects and random errors are independent of one another. The i^{th} researcher's response vector for the common dives is

$$\mathbf{Y}_{i}^{C} = \left(Y_{i11}^{C}, \cdots, Y_{i1K_{i1}}^{C}, Y_{i21}^{C}, \cdots, Y_{i2K_{i2}}^{C}, \cdots, Y_{id_{C}1}^{C}, \cdots, Y_{id_{C}K_{id_{C}}}^{C}\right)^{T} \in \mathbb{R}^{K_{i}'}$$

where $K'_i = \sum_{p=1}^{d_c} K_{ip}$. The vector of responses for the unique dives is defined as

$$\mathbf{Y}_{i}^{U} = \left(Y_{i1}^{U}, Y_{i2}^{U}, \cdots, Y_{im_{i}}^{U}\right)^{T}$$

Thus, we can write the models in Equations 3.1 and 3.2 in matrix notation as

$$\mathbf{Y}_{i}^{C} = \mu \mathbf{1}_{K_{i}^{\prime}} + R_{i} \mathbf{1}_{K_{i}^{\prime}} + \mathbb{I}_{i} \mathbf{D}^{C} + \mathbb{I}_{i} \left(\mathbf{R} \mathbf{D}^{C} \right)_{i} + \boldsymbol{\varepsilon}_{i}^{C}$$

and

$$\mathbf{Y}_{i}^{U} = \boldsymbol{\mu} \mathbf{1}_{m_{i}} + R_{i} \mathbf{1}_{m_{i}} + \mathbf{D}_{i}^{U} + \left(\mathbf{R}\mathbf{D}^{U}\right)_{i} + \boldsymbol{\varepsilon}_{i}^{U}$$

where $\mathbf{1}_k$ denotes a vector of ones of length k, and

$$\mathbb{I}_{i} = \begin{pmatrix} \mathbf{1}_{K_{i1}} & 0 & 0 & \dots & 0 \\ 0 & \mathbf{1}_{K_{i2}} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{1}_{K_{id_{C}}} \end{pmatrix}$$

is a K'_i by d_C matrix of ones and zeros. Finally, we stack both the common and unique responses into the complete response vector for the *i*th researcher as

$$\mathbf{Y}_i = \begin{pmatrix} \mathbf{Y}_i^C \\ \mathbf{Y}_i^U \end{pmatrix}$$

and thus, we define

$$\mathbf{Y} = \left(\mathbf{Y}_1^T, \mathbf{Y}_2^T, \cdots, \mathbf{Y}_r^T\right)^T$$

which leads to the complete model in matrix form as

$$\mathbf{Y} = \boldsymbol{\mu} \mathbf{1} + \mathbf{Z}\mathbf{R} + \mathbf{W}\mathbf{D} + \mathbf{X}\mathbf{T} + \boldsymbol{\varepsilon}$$
(3.3)

where \mathbf{Z} , \mathbf{W} , and \mathbf{X} are the design matrices of zeroes and ones for the random effects for the researcher, \mathbf{R} , for the dive, \mathbf{D} , and for the interaction between researcher and dive, \mathbf{T} , respectively. The exact structure and form of these vectors and matrices are given in Appendix A.1 for completeness.

I fit the model in Equation 3.3 using the data collected from the experiment. The only fixed effect in this model is the mean, μ , while the remaining parameters of interest are the variance components of the random effects and the errors. Since this is a linear mixed-effects model with only random intercepts, the computation of these estimates is straightforward and can be done using the lme4 package in R [26]. The confidence intervals for the estimated variance parameters were computed using likelihood profiles via the confirmt function in R.

3.4 Results

The estimated variance components of the random effects indicate clear variation across researchers, across dives, and within researchers as well (Table 3.1). An easier way to compare the relative magnitudes of these variance estimates is to consider the proportion of variance explained by each random intercept, also called the intraclass correlation coefficient (ICC) or the variance partition coefficient [27]. The ICC for the i^{th} variance component is defined as:

$$ho_i = rac{\sigma_i^2}{\sigma_{ t total}^2}$$

where σ_{total}^2 is the total variance, here $\sigma_{total}^2 = \sigma_D^2 + \sigma_R^2 + \sigma_{RD}^2 + \sigma_E^2$. The ICC will range between 0 and 1 since it is the ratio of an individual variance component to the total variance [28, p. 448].

The ICC values indicate that the dive-to-dive variability of the start of the bottom phase is much larger than the dive-to-dive variability of the end of the bottom phase ($\hat{\sigma}_D$ in Table 3.2). We can also see a substantial researcher-to-researcher variability as estimated by $\hat{\sigma}_R$ for both the start and end of the bottom phase. Finally, $\hat{\sigma}_E$, which quantifies the variation in the response when a researcher classifies the same dive twice, is the smallest estimate for both models.

Table 3.1: Summary of output of the mixed-effects model presented in Equation 3.3 for both the proportion of time into the dive at the start and end of the bottom phase. Notice that estimates are for the standard deviations rather than the variances. These two analyses are computed separately.

	Proportion of Time into Dive at							
	Start of th	e Bottom Phase	End of the Bottom Phase					
	Estimate	95% CI	Estimate	95% CI				
û	0.339	[0.312, 0.367]	0.696	[0.668, 0.724]				
$\hat{\sigma}_D$	0.089	[0.081, 0.099]	0.058	[0.048, 0.068]				
$\hat{\sigma}_R$	0.054	[0.039, 0.077]	0.056	[0.040, 0.079]				
$\hat{\sigma}_{RD}$	0.050	[0.046, 0.055]	0.061	[0.056, 0.066]				
$\hat{\sigma}_{F}$	0.039	[0.036, 0.042]	0.038	[0.036, 0.041]				

Figures 3.2 and 3.3 show the responses (proportion of time into the dive) from each researcher for the start and end of the bottom phase, respectively, for the dives that they saw twice. Each plot shows how consistent the researcher was; the closer the points are to the line, the more consistent the researcher.

3.5 Discussion

My analysis indicates that specifications of the bottom phase for dives tends to vary more for the start than for the end, as indicated by the estimated standard deviation for the random effect for dive. A

Table 3.2: Intraclass Correlation Coefficients (ICCs) for each of the estimated variances of the random intercepts in the two models.

	Model for Proportion of Time Into Dive for								
	Start of the Bottom Phase End of the Bottom Phase								
	ICC	ICC							
$\hat{\sigma}_D^2$	53%	29%							
$\hat{\sigma}_{R}^{2}$	20%	27%							
$\hat{\sigma}_{RD}^2$	17%	32%							
$\hat{\sigma}_{E}^{2^{-}}$	10%	12%							



Figure 3.2: Plots of the responses for the start of the bottom phase for dives that were viewed twice by the researchers. On the *x*-axis is the response for the first time a researcher viewed the dive. The *y*-axis is the response for the second time the researcher viewed the dive. Each plot is for a different researcher, and the line y = x is included.



Figure 3.3: Plots of the responses for the end of the bottom phase for dives that were viewed twice by the researchers. On the *x*-axis is the response for the first time a researcher viewed the dive. The *y*-axis is the response for the second time a researcher viewed the dive. Each plot is for a different researcher, and the line y = x is included.

possible reason for this is that as the whales descend, they may slowly transition to the bottom phase of the dive, making the distinction between descent and bottom phase difficult to identify. Moreover, in speculating the possible reasons for this difference in variation, the distinct, sharper change in the whale's depth, resulting in a "corner" or "kink" between the bottom phase and the ascent phase, could potentially be due to the whale needing to return to the surface to breathe. Hence, it may be easier to distinguish the end of the bottom phase. On the other hand, these differences in variation could actually be due to the bathymetry, or bottom depth, of the ocean floor, forcing the whale along a particular vertical path. Without bathymetry data, it is difficult to determine whether this is true or not. There could also be other biological reasons why the start of the bottom phase is more variable than the end, specifically due to buoyancy. For example, North American right whales (*Eubalaena* *glacialis*) have been shown to be positively buoyant during their descent, meaning they must fight against the buoyancy whereas they are negatively buoyant during their ascent [29]. Killer whales (*Orcinus orca*) are known to be buoyant and they accumulate fat to increase their buoyancy [30]. The buoyant force could result in a more distinct end of the bottom phase compared to the start, resulting in lower variation for the end. Both of these reasons are plausible, and moreover, are both potentially contributing to this variation.

This analysis has also revealed that a large researcher-to-researcher variability exists, and it accounts for approximately 20% and 27% of the variability of the identified starts and ends of the bottom phase respectively. A clear implication of this result is that a one-size-fits-all approach may not be appropriate for bottom phase definitions since there exists clear variability across researchers, indicating that researchers may have personal preferences in defining dive phase definitions, especially depending on the species that they are studying. Moreover, the majority of the researchers said that they were confident in their classifications, and yet substantial researcher-to-researcher variation exists.

Furthermore, the low values for $\hat{\sigma}_E$ imply that researchers tend to be consistent in their labeling of the start and end of the bottom phase when presented with the same dive multiple times. As such, there may be some "true" bottom phase for any given researcher. However, researchers varied in how consistent they were when presented with the same dives twice, with some researchers being far more variable than others (Figures 3.2 and 3.3). For example, for the start of the bottom phase, we can see that researchers 3 and 18 were somewhat inconsistent with their labelling, whereas researchers 6 and 13 were highly consistent. Using the additional information provided by the researchers, we found that both researchers 3 and 18 had less experience working with diving data than researchers 6 and 13, which may be a reason why their bottom phase definitions were less consistent.

All of this evidence suggests that a formal, reproducible method of modelling the bottom phase should be considered. These models, which can be used to predict the bottom phase of dives, should incorporate the researcher's own data allowing for more personalized predictions. This will be considered in the next chapter. Furthermore, the shiny app that I developed for this experiment is available for researchers to use to make the labelling of bottom phases easier and readily usable [31]. Future studies may wish to further quantify the researcher-to-researcher variability or the consistency within researchers on a larger scale since this study was preliminary and included only 18 participants.

Chapter 4

Statistical Models for the Bottom Phase

To every problem, there is a most simple solution. — Agatha Christie, The Clocks (1963)

4.1 Background

As was shown in Chapter 2, the definition of the dive phases is important when attempting to predict prey capture dives. Changing this definition can result in different predicted prey capture dives. The current definitions usually involve a percentage threshold of maximum depth or speed, and these definitions are not necessarily informed by the data [1, 7–9]. Moreover, Chapter 3 showed that researchers tend to classify the bottom phase of dives differently from one another. To address some of these issues, I developed several statistical models for the bottom phase of a dive, with the goal being to predict the bottom phase of dives in a systematic and reproducible way, from researchers' phase specifications.

In this chapter, I consider three models: percentage of maximum depth, linear regression, and functional regression. Tennessen et al. [1] define the start and end of the bottom phase of a dive as the first and last time the whale achieves 70% of its maximum depth for that dive, respectively. This percentage is referred to as the *percentage threshold*. The authors chose this percentage threshold and then validated it visually using a random subset of their dives. A simple improvement to this definition is to use a training dataset to estimate the percentage threshold to define the bottom phase. Furthermore, the percentage threshold need not be the same for the start and end of the bottom phase.

Since there exists plenty of data from each dive, and the bottom phase may depend on these additional covariates, I also consider a linear regression model. These linear regression models are simple to understand and use and, they can provide improvements in modeling the bottom phase over the 70% threshold model or the estimated percentage threshold model.

Functional linear regression extends the linear regression model by including a function or curve as a covariate in the model [32, p. 157]. Each functional covariate, which in this case is a depth, speed, or acceleration curve, is scaled to [0, 1] on the time axis. However, because the functional curve is measured at many discrete time points, including all of this data in the model as a vector of covariates would result in an overparametrized model. Furthermore, each curve has a different number of measurements which would make it difficult to include as a vector of covariates since the lengths differ. Finally, there is an inherent time ordering within this type of data which would be lost if the data was included as a covariate vector. Functional regression properly accounts for these issues by representing this discretely measured data as a single smooth function [33, 34].

For each model, I will specify the mathematical form and present the fitted results. I will also perform model diagnostics. I will apply the methods from Chapter 2 to predict prey capture dives for each model's fitted bottom phase classification. I will construct confusion matrices of these results to compare the predicted versus actual prey capture dives. Moreover, I will also perform k-fold cross validation of these models to see how they improve upon the current 70% max depth definition.

4.2 Methods

To fit statistical models for the bottom phase, Dr. Sarah Fortune, a marine mammal whale researcher, independently identified the bottom phases of all 250 non-randomized dives from Whale #3. I selected this whale for this chapter, because unlike for Whale #1 and Whale #2, the data of Whale #3 have no sequences that are missing, have several deep dives, and have confirmed prey captures. Thus, the data of Whale #3 could be used to assess the effect of my proposed dive phase models on prey capture predictions.

4.2.1 Percentage Threshold

Let Y_i^S and Y_i^E be the percentage of time into a dive, entered by a researcher, for the start and end of the bottom phase respectively for the i^{th} dive, where i = 1, ..., N. Let $X_i^S(\theta)$ and $X_i^E(\theta)$ be the percentage of time into the i^{th} dive when the whale reaches $(100 \times \theta)\%$ of its maximum depth for the first and last time, respectively for $\theta \in (0, 1)$. To avoid any extraneous distributional assumptions, we assume that

$$E(Y_i^S|D_i) = X_i^S(\theta) \tag{4.1}$$

where D_i is the time-depth trajectory of the i^{th} dive. Similarly, we assume that

$$E(Y_i^E|D_i) = X_i^E(\theta) \tag{4.2}$$

for the end of the bottom phase. We estimate θ by minimizing the sum of squared errors

$$\sum_{i=1}^{N} \left(Y_i^S - X_i^S(\boldsymbol{\theta}) \right)^2 \tag{4.3}$$

for the start of the bottom phase. In a similar manner, for the *end* of the bottom phase, we estimate θ by minimizing

$$\sum_{i=1}^{N} \left(Y_i^E - X_i^E(\theta) \right)^2.$$
(4.4)

These equations are appealing in that you obtain different percentage threshold values for the start and end of the bottom phase which are easily interpretable; however, these models are simplistic in that their definition of the bottom phase only focuses on the maximum depth of the dive. From now on, I refer to these models as M_{θ} .

The estimation of θ for the start and end of the bottom phase involves minimizing the nonlinear functions in Equations 4.3 and 4.4. Using a set of appropriate starting values, I use the optimize function in R [35] to minimize the function and yield values for $\hat{\theta}$. Confidence intervals are computed using standard bootstrap techniques with B = 500 samples and a confidence level of 95% [36, p. 13].

4.2.2 Linear Regression

Let Y_i be the researcher's response for the i^{th} dive, where Y_i is the percentage of time into the dive when the bottom phase begins (or ends). Let $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{ip})^T$ be a $p \times 1$ vector of covariates for the i^{th} dive, i = 1, ..., n. We can write a basic linear regression model as

$$Y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \tag{4.5}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^T$ is a vector of unknown coefficients and $\varepsilon_i \sim N(0, \sigma^2)$ are i.i.d. random errors.

Since the two most common measures available from diving data are maximum depth and total duration (e.g. [7], [8]), I consider a linear regression model with maximum depth and total duration as covariates, along with their product, which I refer to as their interaction term. We will refer to these models as M_{LM} .

4.2.3 Functional Regression

Assume for dive *i*, there is a single scalar response variable, Y_i , a set of fixed scalar covariates, $x_{i1}, x_{i2}, \ldots, x_{ip}$, as well as D_i , a functional covariate that yields densely sampled data measured at

times t_{ik} , for $k = 1, ..., K_i$ and i = 1, ..., N. Each D_i is scaled to [0, 1] on the time axis. In a linear model, the response vector, Y_i , is regressed on the covariates, $\mathbf{x}_i^T \boldsymbol{\beta}$. Note that $\mathbf{x}_i^T \boldsymbol{\beta}$ can be written as $\langle \mathbf{x}_i, \boldsymbol{\beta} \rangle$ where \langle , \rangle is the standard Euclidean inner product on \mathbb{R}^p . To incorporate D_i into a model with its own estimable functional coefficient, say γ , we add $\langle \gamma, D_i \rangle$ into the model, where \langle , \rangle is an inner product over the vector space of square-integrable functions defined on the closed interval [0, 1]. The usual functional regression model is written as

$$Y_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \int_0^1 \gamma(t) D_i(t) dt + \varepsilon_i, \qquad (4.6)$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is a vector of unknown coefficients, γ is an unknown, smooth functional coefficient, and ε_i are i.i.d. normal random variables with mean 0 and constant variance σ^2 . Since γ is a function in an infinite dimensional space, a finite number of basis functions must be used to represent γ . Let $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_M)$ be a vector of known basis functions defined on the interval [0, 1]. Commonly used bases are B-splines bases, polynomial bases, and Fourier bases. Assume $\gamma(t)$ is of the form

$$\gamma(t) = \sum_{j=1}^{M} g_j \phi_j(t).$$
 (4.7)

The g_j 's are unknown and will be estimated. The value of M determines the degree of smoothness for $\gamma(t)$. Using this assumption, we write the model in Equation 4.6 as

$$Y_i = \sum_{w=1}^p \beta_w x_{iw} + \sum_{j=1}^M g_j \int_0^1 \phi_j(t) D_i(t) dt + \varepsilon_i$$

Since we have extremely dense tag data (sampled at 50 Hz), we can approximate $\int_{0}^{1} \phi_j(t) D_i(t) dt$ using a Reimann sum, yielding the final model

$$Y_i \approx \sum_{w=1}^p \beta_w x_{iw} + \sum_{j=1}^M g_j x_{ij}^* + \varepsilon_i$$
(4.8)

where $x_{ij}^* = \sum_{k=1}^{K_i} \phi_j(t_{ik}) D_i(t_{ik}) (\Delta t)_i$ are known covariates for the i^{th} dive and the j^{th} basis function, K_i is the number of measurements taken on the i^{th} dive, and $(\Delta t)_i = 1/K_i$ is the time between two consecutive measurements for the i^{th} dive, i = 1, ..., N and j = 1, ..., M. For large enough values of K_i , this approximation will be accurate. To simplify notation, and since for our dataset $523 < K_i < 23,491$, I replace the approximation sign with an equality sign. For the two different response variables (percent of time into a dive for the start and end of the bottom phase), I fit the model in Equation 4.8 for three different functional covariates: depth, speed, and acceleration (Figures 4.1–4.3). From now on, these models will be referred to as $M_{\gamma D}$ (using depth), $M_{\gamma S}$ (using speed), and $M_{\gamma A}$ (using acceleration) (Table 4.3). I used cubic B-splines as basis functions for γ , with M = 20, since cubic B-splines are ubiquitous in the literature for functional data, and they have desirable properties such as being twice differentiable and non-oscillatory unlike Fourier series. Cubic B-splines also exhibit local behaviour and have good computational properties since their support is small. The choice of M is arbitrary, but M = 20 provides adequate detail and smoothness for the functional regression coefficient, γ . Each model also includes maximum depth, total duration, and their product as fixed covariates.

The response variable is a percent of time into the dive when the bottom phase begins or ends and this is bounded between 0 and 100. Thus, this response variable can never be normally distributed. However, for our dataset, the response variables actually resemble a normal distribution and this assumption fits well. We also tried several transformations of the response variable such as a *logit* transformation, however, these transformations did not provide any additional benefits to the fit of the models. As such, we maintained the untransformed response variable for ease of interpretation.



Figure 4.1: Plots of the depth curves for the 250 dives from Whale #3 to be used as functional covariates in Equation 4.8. The times for each depth curve were scaled to [0, 1].



Figure 4.2: Plots of the speed curves for the 250 dives from Whale #3 to be used as functional covariates in Equation 4.8. The times for each speed curve were scaled to [0, 1].



Figure 4.3: Plots of the acceleration curves for the 250 dives from Whale #3 to be used as functional covariates in Equation 4.8. The times for each depth curve were scaled to [0,1].

4.2.4 Model Comparison

To compare the different models, including the linear models, the percentage threshold models, and the functional regression models, I will consider k-fold cross validation, as well as comparing predicted prey capture dives. For the k-fold cross validation, I compute the mean squared error (MSE) and mean absolute error (MAE) for each of the k withheld groups, then I average the k

MSEs and the k MAEs resulting in one metric for each model. To make everything comparable and on the same scale, I take the square root of the averaged MSE and refer to this as the root mean squared error (RMSE). The MSE and MAE are defined as:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \text{ and } MAE = \frac{1}{n} \sum_{i=1}^{n} |Y_i - \hat{Y}_i|$$
(4.9)

where Y_i is the actual response for the i^{th} dive and \hat{Y}_i is the predicted response from a model and n = 250/k.

For the predicted prey capture comparison, audio and visual data confirmed 7 prey capture dives from Whale #3. For several of the remaining 243 dives, the whale could have been feeding, but there are no acoustic or visual data available to confirm or deny any prey capture events. Specifically, dives #1 through #186 have been correctly classified as either prey capture dives or non-prey capture dives, whereas Dives #187 to #250 are unclassified. I then create fitted values for the bottom phase for each of the aforementioned models and apply the prey capture detection algorithm described in Chapter 2.

4.3 Results

4.3.1 Percentage Threshold

A linear grid search over the interval (0,1), which is not computationally intensive, provides appropriate starting values for the minimization (Figures 4.4 and 4.5).

 Table 4.1: Results from fitting the percentage threshold models (from Equations 4.3 and 4.4) with 95% boostrapped confidence intervals.

Percentage Threshold Model for							
Start of	the Bottom Phase	End of the Bottom Phase					
$\hat{ heta}$	95% CI	$\hat{ heta}$	95% CI				
0.8051	[0.7611, 0.8543]	0.8242	[0.7707, 0.8750]				

The estimates for both the start and the end of the bottom phase are quite similar to one another (Table 4.1), however, the 70% threshold from Tennessen et al. [1] is not in either of these confidence intervals, indicating that the 70% threshold is unlikely to be appropriate for this dataset. Model diagnostics indicate a few outlier points with very large residuals (see Appendix A.2 for diagnostic



Figure 4.4: Sum of squared errors versus theta for the start of the bottom phase for values of theta over 0 to 1.



Figure 4.5: Sum of squared errors versus theta for the end of the bottom phase for values of theta over 0 to 1.

plots). The difference between the user-defined bottom phase and the predicted bottom phase from the model for these dives (Figures 4.6 and 4.7) is most likely due to the awkward shapes. Since this model uses a proportion of maximum depth to define the bottom phase, it tends to misrepresent the bottom phase of dives which have a slower ascent as well as dives with a second, albeit smaller, descent phase sometimes referred to as a "step in the ascent phase" [7] (see Dives 21 and 188 in Figure 4.7 as examples). As such, the percentage of max depth model may not adequately handle these specific types of dives.



Figure 4.6: Identified outlier dives for the percent of max depth model for the start of the bottom phase. The red line indicates the user-provided start of the bottom phase, and the turquoise line indicates the predicted start of the bottom phase from the model.

4.3.2 Linear Model

The estimated coefficients from fitting the linear model to the dataset are all highly significant (Table 4.2), except the maximum depth coefficient for the start of the bottom phase. The fitted models indicate that start (end) of the bottom phase tends to be closer to the beginning of the dive



Figure 4.7: Identified outlier dives for the percent of max depth model for the end of the bottom phase. The red line indicates the user-provided end of the bottom phase, and the turquoise line indicates the predicted end of the bottom phase from the model.

(closer to the end of the dive) for longer duration dives, and that as the maximum depth increases, this relationship with duration is more pronounced (Figures 4.8 and 4.9).

Diagnostic plots show the normality assumptions are appropriate, although the tails are slightly heavy, and that there are no obvious temporal trends in residuals or clear heteroskedasticity (see Appendix A.3 for diagnostic plots). While there are no influential points, there are a few outlier dives with large standardized residuals that have awkward shapes and are typically skewed, meaning the bottom phase occurs in either the first half or second half of the dive, rather than somewhere in the middle (Figures 4.10 and 4.11). To address some of these minor model assumption violations, I considered several transformations of both the response variable and the covariates, but I found negligible improvements in the residual plots. Since this model is meant to be a foundational, basic model with easier interpretations, I settled on using the original model in Equation 4.5 without any transformations.



Figure 4.8: Predicted responses for the start of the bottom phase as a function of duration for several values of max depth



Figure 4.9: Predicted responses for the end of the bottom phase as a function of duration for several values of max depth

	Response	Variable [†]
	Start of the	End of the
	Bottom Phase	Bottom Phase
Intercept	35.79*	66.58*
-	(0.91)	(0.88)
Max Depth	-0.01	-0.24^{*}
Duration	-0.08*	0.04*
	(0.01)	(0.02)
Max Depth × Duration	0.00042* (0.00011)	0.00053* (0.00014)
Number of Observations	250	250
\mathbb{R}^2	0.20	0.23
Adjusted R ²	0.19	0.22
Residual Std. Error	9.1	8.7
F Statistic (df = 3; 246)	20.75*	24.37*

 Table 4.2: Results from fitting the basic linear regression model.

*p<0.01

[†] Response variable is the percentage of time from the start of the dive.

4.3.3 Functional Model

The estimated fixed covariates show high significance for most of the covariates in each model (Table 4.3). The functional regression coefficients (Figure 4.12) can be interpreted by considering a simple example of predicting the start of the bottom phase of a dive (Figure 4.13) using the model with depth as a functional predictor (Equation 4.8). For dives that are very deep near the beginning of the dive, this model tends to predict an earlier start to the bottom phase since the integral will be very negative. On the other hand, for dives which are shallow near the beginning of the dive, this integral will be less negative, resulting in a predicted start of the bottom phase closer to the middle of the dive. Similarly, for the end of the bottom phase (Figure 4.12), the estimated functional regression coefficient is positive in the second half of the time interval, which implies that dives that



Figure 4.10: Identified outlier dives for the linear regression model in Equation 4.5 for the start of the bottom phase. The red line indicates the user-provided start of the bottom phase, and the turquoise line indicates the predicted start of the bottom phase from the model.

are very deep near the end of the time domain will have a late bottom phase. These interpretations can be extrapolated to the other functional regression models.

Visual inspection of diagnostics plots indicates that the functional regressions appear to fit the data well, except for a few non-influential outliers and some moderate heteroskedasticity (see Appendix A.4 for details). For the three models for the start of the bottom phase, the same six dives have fairly large residuals for each model, indicating that these dives are outliers (Figure 4.14). For the three models for the end of the bottom phase, six dives have large residuals for each model, although these outliers are not the same outliers for the start of the bottom phase (Figure 4.15). The outlier dives in Figure 4.14 are mostly "V-shaped" dives where the deepest portion of the dive occurs in the second half of the time domain, producing predictions which occur much earlier than what





the whale researcher had marked. Similarly, in Figure 4.15, the outlier dives are mainly "V-shaped" dives where the deepest portion occurs in the first half of the time domain, resulting in predictions that occur a lot later in the dive than where the researcher had marked.

4.3.4 Model Comparison

Overall, using any model results in improvements compared to using the 70% threshold method, $M_{70\%}$, based on the RMSE and MAE (Table 4.4). The percentage threshold model, M_{θ} , performs well in terms of MAE, but performs poorly in terms of RMSE, due to a few extreme outliers. The linear regression model, M_{LM} , tends to perform poorly compared to the other models based on both the RMSE and MAE. The functional regression models, $M_{\gamma D}$, $M_{\gamma S}$, and $M_{\gamma A}$, all tend to perform similarly for both the start and end of the bottom phase, and are better than the other models. As



Figure 4.12: Plots of the fitted regression coefficients, $\hat{\gamma}$, from Equation 4.8, with 95% pointwise confidence bands in red. The left and right hand columns are the results for the start and the end of the bottom phase, respectively. The rows correspond to the models for the different functional covariates, depth, speed, and acceleration.



Figure 4.13: Plots illustrating the interpretation of the impact of the functional predictor on the response variable. Top plot is of the estimated functional regression curve when depth is used as a functional covariate (Equation 4.8). Middle plot is of a single, deep dive. The bottom plot is the estimated functional regression curve multiplied (pointwise) by the depth curve. Integrating the bottom plot over the entire time domain gives the value of the functional predictor.

	Dependent Variable							
	Start	of Bottom	Phase	End	End of Bottom Phase			
Functional Covariate	Depth $(M_{\gamma D})$	Speed $(M_{\gamma S})$	$\begin{array}{c} \mathrm{Acc} \\ (M_{\gamma\mathrm{A}}) \end{array}$	$ \begin{array}{c} Depth \\ (M_{\gamma D}) \end{array} $	Speed $(M_{\gamma S})$	$\begin{array}{c} \text{Acc} \\ (M_{\gamma \text{A}}) \end{array}$		
Coefficient								
Intercept	35.81	35.84	35.60	66.00	66.06	66.10		
	(0.83)	(0.97)	(0.89)	(0.84)	(0.95)	(0.88)		
Max Depth	-0.29	-0.33	-0.21	0.37	0.32	0.25		
	(0.18)	(0.18)	(0.19)	(0.16)	(0.18)	(0.19)		
Duration	-0.064	-0.066	-0.064	0.037	0.037	0.033		
	(0.013)	(0.012)	(0.015)	(0.012)	(0.013)	(0.015)		
Max Depth \times Duration	-0.0009	-0.0008	-0.0007	0.0001	0.0001	0.0002		
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0001)		
Adj R ²	0.347	0.346	0.344	0.301	0.306	0.309		

Table 4.3: Estimates of the regression coefficients from fitting Equation 4.8 for the start and end of the bottom phase for each functional covariate. Adjusted R^2 is also given.

such, these functional regression models should be considered as the best alternative to the 70% threshold method.

When applying the prey capture detection algorithm from Chapter 2 using the different bottom phases predicted from these models, 6 out of the 7 confirmed successful prey capture dives are detected, except when using bottom phases from the linear model, which results in only 4 out of the 7 (Table 4.5). These are the same 6 dives that were previously mentioned in Chapter 2 (Dives 47, 54, 62, 124, 125, and 136). Dive 67 is never detected as a successful prey capture from any of the models. When considering only confirmed non-prey capture dives, the 70% threshold method results in a false positive rate of 8.37% (15/179), whereas the models containing either depth or acceleration as a functional covariate resulted in the lowest false positive rate (8/179 = 4.47%), which is a 46% reduction in false positives. Since all of these models have the same true positive rate of false positives and the same rate of true positives than other methods, defining the bottom phase using a functional regression model with depth or acceleration as a functional covariate improves


Figure 4.14: The outlier dives for all 3 models for the start of the bottom phase from Equation 4.8. The purple vertical lines indicate what the whale researcher said was the bottom phase, and the remaining 3 lines (which overlap) indicate the model predictions.

the capacity of Tennessen et al. [1]'s method to predict prey capture events for our data. Finally, in Table 4.5, I provide the results of the prey capture detection algorithm for the user provided bottom phase as well as for comparison purposes. While the user provided bottom phases perform similarly to the other models in terms of false positives, some models ($M_{\gamma}D$ and $M_{\gamma}A$) outperformed the user-provided data.

4.4 Conclusions

My results suggest that modeling the bottom phase using expert information and formal statistical models rather than an arbitrarily set threshold appears to improve the identification of the bottom phase, which in turn improves our capacity to predict key ecological features such as prey captures. Of the many models I explored, the functional models outperformed the linear models and the percentage threshold models in terms of RMSE and MAE, as shown by the cross-validation procedure, as well as reduced false positive rates for predicted prey capture dives.



Figure 4.15: The outlier dives for all 3 models for the end of the bottom phase from Equation 4.8. The purple vertical lines indicate what the whale researcher said was the bottom phase, and the remaining 3 lines (which overlap) indicate the model predictions.

models has its own set of advantages and disadvantages.

The percentage threshold model extends the basic 70% threshold method by incorporating a user's own data and estimating this percentage. This model was shown to have major improvements over the 70% method in terms of both reductions in RMSE and MAE, but also in terms of predicting prey capture dives. The percentage threshold model may however not be detailed enough since it only incorporates maximum depth and does not incorporate other important variables such as total dive duration. The linear regression model tends to be easy to understand and interpret, but may not be detailed enough to model the complexity of the bottom phase. The linear model provides minor improvements to the 70% threshold model, but when predicting prey capture dives, the predicted bottom phases from this model failed to capture enough information to detect all the prey capture dives. As such, these two simple models may provide slight benefits over the 70% threshold method, but the improvements are marginal.

The functional regression models are promising, alternative models that incorporate functional covariates. While they are more complicated models, they appear to provide better predictions

	Dependent Variable										
	Start of Bottom Phase					End of Bottom Phase					
	Avg RMSE		Avg MAE			Avg RMSE		Avg MAE			
	k = 5	k = 10	k = 5	k = 10		k = 5	k = 10	k = 5	k = 10		
Model											
M _{70%}	0.103		0.091			0.125		0.110			
$M_{ heta}$	0.089	0.089	0.060	0.057		0.102	0.104	0.062	0.063		
M_{LM}	0.093	0.093	0.072	0.072		0.089	0.089	0.066	0.066		
$M_{ m \gamma D}$	0.082	0.082	0.063	0.064		0.085	0.083	0.062	0.062		
$M_{\gamma \mathrm{S}}$	0.083	0.082	0.063	0.063		0.085	0.084	0.062	0.062		
ΜγΑ	0.083	0.083	0.063	0.062		0.085	0.084	0.061	0.061		

 Table 4.4: Results from running 5- and 10-fold cross validation for each of the models mentioned in this chapter.

for detecting prey capture dives. From the analysis in this chapter, the functional regression models, regardless of the functional covariate, performed better than all the other models during cross validation, except in minor cases. Moreover, the models provide flexibility to the standard and widely accepted bottom phase definitions. These models also outperform the raw inputted data directly from the user, according to our results using our dataset. Hence, in using these models, researchers can produce more accurate bottom phase definitions which can then be reproduced by other researchers. Thus, researchers should state the parameter estimates for these models when disseminating their results so that others can use their model estimates as well. There are also several limitations to these functional models. While these models can produce global standard errors for the functional coefficient, the resulting bands are often too wide to be useful. As such, we rely on the point-wise confidence bands which may not be representative of the bounds for the functional coefficient. Moreover, these models require a large dataset to estimate the large number of parameters. As such, we will discuss possible approaches to determine just how many dives are needed to fit these models in the next chapter. These functional regression models may not be appropriate for every scenario since they require a deeper understanding of the methodology in order to properly interpret the results.

For researchers to collect bottom phase data from their dives, I have developed a Shiny App [37] which allows a user to view a dive then mark the bottom phase and save the results. The relevant

		Truth							
Model	Prediction	Prey Capture	Non-Prey Capture	Unknown	Total				
M _{LM}	Prey Capture	4	9	1	14				
	Non-Prey Capture	3	170	63	236				
$M_{ heta\%}$	Prey Capture	6	13	0	19				
	Non-Prey Capture	1	166	64	231				
<i>M</i> _{70%}	Prey Capture	6	15	0	21				
	Non-Prey Capture	1	164	64	229				
$M_{\gamma D}$	Prey Capture	6	8	1	15				
	Non-Prey Capture	1	171	63	235				
$M_{\gamma S}$	Prey Capture	6	10	1	17				
	Non-Prey Capture	1	169	63	229				
ΜγΑ	Prey Capture	6	8	1	15				
	Non-Prey Capture	1	171	63	235				
Truth									
	Prediction	Prey Capture	Non-Prey Capture	Unknown	Total				
User	Prey Capture	6	10	0	16				
Provided	Non-Prey Capture	1	169	64	234				

Table 4.5: Confusion matrices for predicting prey captures via Tennessen et al. [1]'s criteria, but with the bottom phase determined by the different models. The last table is the confusion matrix when predicting prey captures using the user provided bottom phases.

code and instructions comprise my custom R package called bustR, which can be found on my personal GitHub page [31].

Chapter 5

How Many Dives To Choose? Methods for Researchers

Decision making is an art only until the person understands the science. — Pearl Zhu

5.1 Background

In Chapter 4, I demonstrated that several statistical models for the start and end of the bottom phase of a dive outperformed current methods. However, diving datasets typically contain thousands of dives (e.g. Tennessen et al. [1] analyzed over 6,000 dives and Viviant et al. [8] analyzed over 5,000 dives). These models require the researcher to define the bottom phase for each dive. Since it is impractical to classify the bottom phase for thousands of dives, researchers need to determine an adequate number of dives for these models. In this chapter, I address possible solutions to this problem with two potential approaches: one relying on matching two model predictions, one from a fit to the whole data set and the other from a fit to a subsample of the data set, while the other approach matches model predictions to the user-provided response.

Let Y_i be the proportion of time into a dive for the start (or end) bottom phase that the researcher *would mark* for the *i*th dive in their dataset, i = 1, ..., N. Let \hat{Y}_i^k be the predicted bottom phase for the *i*th dive, i = 1, ..., N, using a model that was fit using only $k \le N$ randomly selected dives. Note that when k = N, \hat{Y}_i^N is the predicted bottom phase for the *i*th dive using a model that was fit using all N dives in the dataset. The two possible approaches that I will discuss to determine an adequate sample size to fit these models are as follows.

1. A probabilistic approach based on the distribution of $\hat{Y}_i^N - \hat{Y}_i^k$, requiring covariates from all *N* dives and a value for the regression error variance.

2. A user-provided approach to directly compare Y_i to \hat{Y}_i^k , requiring users to prove $Y_1, ..., Y_N$ and possibly other covariates.

This chapter shows researchers how to assess the *trends* in prediction accuracy as the sample size, k, approaches the total number of dives in the dataset, N. For the two approaches discussed in this chapter, I use the 250 dives from Whale #3 which were used to fit the models in Chapter 4, and in particular, I use the user's responses of the bottom phase for the second approach.

5.2 Methods

For the probabilistic approach, I consider a linear model which assumes normality (see Section 4.2.2). Nonlinear models (such as the percentage threshold model) cannot be handled by the first method, so they are not considered here. The linear model includes maximum depth, total duration, and their product (referred to as the interaction term) as fixed covariates. I also consider a functional regression model which includes depth as a functional predictor. See Section 4.2.3 for complete details of functional regression models. Interestingly, the first approach does not require the user's responses, rather it only requires a reasonable estimate for the variability of the data.

In the second approach, using the user's responses, I consider both linear and non-linear models since the methods used are more flexible to model choice. For this so-called user-provided approach, I consider the 70% threshold method used in Tennessen et al. [1], the estimated percentage threshold model used in Section 4.2.1, a simple linear regression model used in Section 4.2.2 with the same covariates as described above, and a functional regression model with depth as a functional predictor from Section 4.2.3 again with the same covariates as described above. The other two predictors in the functional regression models (speed and acceleration) are omitted since they behave very similarly to the depth model.

5.2.1 Probabilistic Approach

The probabilistic approach is based on the distribution of the difference between \hat{Y}_i^N and \hat{Y}_i^k , using a linear model. All notation is defined as in Equation 4.5. Let \mathbf{X}_N be the $N \times p$ matrix of covariates for all data points, and \mathbf{X}_k be the matrix of covariates for the *k* randomly chosen dives. The associated response vectors are denoted \mathbf{Y}_N and \mathbf{Y}_k . For convenience, construct \mathbf{X}_N and \mathbf{Y}_N so that their first *k* rows contain the *k* randomly chosen dives. Now, we derive the conditional probability distribution of $\hat{\mathbf{Y}}^N - \hat{\mathbf{Y}}^k$ given \mathbf{X}_N and \mathbf{X}_k as a means to determine an adequate sample size, *k*.

Claim 5.2.1. Assume that both \mathbf{X}_N and \mathbf{X}_k are full rank. Then $\hat{\mathbf{Y}}^N - \hat{\mathbf{Y}}^k | \mathbf{X}_N, \mathbf{X}_k$ has a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$ where $\boldsymbol{\Sigma} = \sigma^2 \mathbf{X}_N \left[\left(\mathbf{X}_k^T \mathbf{X}_k \right)^{-1} - \left(\mathbf{X}_N^T \mathbf{X}_N \right)^{-1} \right] \mathbf{X}_N^T$ with $Var(\boldsymbol{\varepsilon}_i | \mathbf{X}_N, \mathbf{X}_k) = \sigma^2$.

Proof. Fitting the linear model in Equation 4.5 to k randomly selected dives results in

$$\hat{\boldsymbol{\beta}}^{k} = \left(\mathbf{X}_{k}^{T}\mathbf{X}_{k}\right)^{-1}\mathbf{X}_{k}^{T}\mathbf{Y}_{k}$$

and when k = N, then $\hat{\boldsymbol{\beta}}^{k} = \hat{\boldsymbol{\beta}}^{N} = (\mathbf{X}_{N}^{T}\mathbf{X}_{N})^{-1}\mathbf{X}_{N}^{T}\mathbf{Y}_{N}$. The predicted responses for all N dives in the dataset are given by:

$$\hat{\mathbf{Y}}^k = \mathbf{X}_N \hat{\boldsymbol{\beta}}^k.$$

From this, we can see that

$$E\left[\hat{\mathbf{Y}}^{N}-\hat{\mathbf{Y}}^{k} \mid \mathbf{X}_{N}, \mathbf{X}_{k}\right] = E\left[\hat{\mathbf{Y}}^{N} \mid \mathbf{X}_{N}, \mathbf{X}_{k}\right] - E\left[\hat{\mathbf{Y}}^{k} \mid \mathbf{X}_{N}, \mathbf{X}_{k}\right]$$
$$= \mathbf{X}_{N}\boldsymbol{\beta} - \mathbf{X}_{N}\boldsymbol{\beta} = \mathbf{0}_{N \times 1}.$$

To derive the variance-covariance matrix, write $\hat{\mathbf{Y}}^k$ as

$$\begin{aligned} \hat{\mathbf{Y}}^{k} &= \mathbf{X}_{N} \hat{\boldsymbol{\beta}}^{k} \\ &= \mathbf{X}_{N} \left(\mathbf{X}_{k}^{T} \mathbf{X}_{k} \right)^{-1} \mathbf{X}_{k}^{T} \mathbf{Y}_{k} \\ &= \mathbf{X}_{N} \left(\mathbf{X}_{k}^{T} \mathbf{X}_{k} \right)^{-1} \mathbf{X}_{k}^{T} \left[\mathbf{X}_{k} \boldsymbol{\beta} + \boldsymbol{\varepsilon}_{k} \right] \\ &= \mathbf{X}_{N} \boldsymbol{\beta} + \mathbf{X}_{N} \left(\mathbf{X}_{k}^{T} \mathbf{X}_{k} \right)^{-1} \mathbf{X}_{k}^{T} \left[\mathbf{I}_{k} \quad \mathbf{0}_{k \times (N-k)} \right] \boldsymbol{\varepsilon}_{N}. \end{aligned}$$

Similarly, write $\hat{\mathbf{Y}}^N$ as $\hat{\mathbf{Y}}^N = \mathbf{X}_N \boldsymbol{\beta} + \mathbf{X}_N \left(\mathbf{X}_N^T \mathbf{X}_N\right)^{-1} \mathbf{X}_N^T \boldsymbol{\varepsilon}_N$. Thus,

$$\begin{aligned} \hat{\mathbf{Y}}^{N} - \hat{\mathbf{Y}}^{k} &= \mathbf{X}_{N} \left\{ \left(\mathbf{X}_{N}^{T} \mathbf{X}_{N} \right)^{-1} \mathbf{X}_{N}^{T} - \left(\mathbf{X}_{k}^{T} \mathbf{X}_{k} \right)^{-1} \mathbf{X}_{k}^{T} \left[\mathbf{I}_{k} \quad \mathbf{0}_{k \times (N-k)} \right] \right\} \boldsymbol{\varepsilon}_{N} \\ &= \mathbf{X}_{N} \left[\mathbf{A}_{N} - \mathbf{A}_{k} \right] \boldsymbol{\varepsilon}_{N} \end{aligned}$$

where \mathbf{A}_N and \mathbf{A}_k are defined accordingly to simplify notation. Now,

$$Var\left[\hat{\mathbf{Y}}^{N}-\hat{\mathbf{Y}}^{k} \mid \mathbf{X}_{N}, \mathbf{X}_{k}\right] = \mathbf{X}_{N} \left(\mathbf{A}_{N}-\mathbf{A}_{k}\right) Var\left[\boldsymbol{\varepsilon}_{N} \mid \mathbf{X}_{N}, \mathbf{X}_{k}\right] \left[\mathbf{X}_{N} \left(\mathbf{A}_{N}-\mathbf{A}_{k}\right)\right]^{T}$$
$$= \boldsymbol{\sigma}^{2} \mathbf{X}_{N} \left(\mathbf{A}_{N}-\mathbf{A}_{k}\right) \left(\mathbf{A}_{N}-\mathbf{A}_{k}\right)^{T} \mathbf{X}_{N}^{T}$$

which, after calculations, simplifies to

$$Var\left[\mathbf{\hat{Y}}^{N}-\mathbf{\hat{Y}}^{k} \mid \mathbf{X}_{N}, \mathbf{X}_{k}\right] = \sigma^{2}\mathbf{X}_{N}\left[\left(\mathbf{X}_{k}^{T}\mathbf{X}_{k}\right)^{-1}-\left(\mathbf{X}_{N}^{T}\mathbf{X}_{N}\right)^{-1}\right]\mathbf{X}_{N}^{T} \equiv \mathbf{\Sigma}$$

which must be positive definite since it is a variance-covariance matrix. I show directly in Appendix

A.5 that $(\mathbf{X}_k^T \mathbf{X}_k)^{-1} - (\mathbf{X}_N^T \mathbf{X}_N)^{-1}$ is positive definite—a surprising result on its own. Finally, since $\hat{\mathbf{Y}}^N - \hat{\mathbf{Y}}^k = \mathbf{X}_N [\mathbf{A}_N - \mathbf{A}_k] \boldsymbol{\varepsilon}_N$ and $\boldsymbol{\varepsilon}_N \stackrel{iid}{\sim} MVN(0, \sigma^2 I)$ then

$$\left\| \mathbf{\hat{Y}}^N - \mathbf{\hat{Y}}^k \right\| \mathbf{X}_N, \mathbf{X}_k \sim MVN\left(\mathbf{0}, \mathbf{\Sigma}\right)$$

By utilizing the above result, then

$$Pr\left(-\mathbf{d} \le \hat{\mathbf{Y}}^N - \hat{\mathbf{Y}}^k \le \mathbf{d} \,\middle|\, \mathbf{X}_N, \mathbf{X}_k\right) \tag{5.1}$$

can be computed for a given $\mathbf{X}_N, \mathbf{X}_k, \sigma^2$, and **d**, where **d** is a vector of so-called tolerance levels. The inequality, $-\mathbf{d} \leq \hat{\mathbf{Y}}^N - \hat{\mathbf{Y}}^k \leq \mathbf{d}$, implies that every entry of the vector $\hat{\mathbf{Y}}^N - \hat{\mathbf{Y}}^k$ is between the corresponding entries in $-\mathbf{d}$ and \mathbf{d} . For this chapter, $\mathbf{d} = d\mathbf{1}$, so all entries of **d** are equal to *d*; however, this does not need to be the case.

To determine an adequate value for k, we compute

$$Pr\left(-\mathbf{d} \leq \hat{\mathbf{Y}}^{N} - \hat{\mathbf{Y}}^{k} \leq \mathbf{d} \mid \mathbf{X}_{N}, k \text{ dives chosen}\right) = \frac{1}{\binom{N}{k}} \sum_{\mathbf{X}_{k} \in \mathcal{X}} Pr\left(-\mathbf{d} \leq \hat{\mathbf{Y}}^{N} - \hat{\mathbf{Y}}^{k} \leq \mathbf{d} \mid \mathbf{X}_{N}, \mathbf{X}_{k}\right)$$
(5.2)

where X denotes the set of design matrices from all possible samples of size k generated from the N dives. Since generating all possible samples would be computationally impractical, instead, Equation 5.2 is estimated using 500 randomly chosen samples of size k. This simulation study proceeds as follows:

- 1. Choose a set \mathcal{K} of sample sizes, a set \mathcal{S} of σ values, and a set \mathcal{D} of d values.
- 2. For each $k \in \mathcal{K}$, $\sigma \in S$, and $d \in \mathcal{D}$, randomly select *k* rows from \mathbf{X}_N to create \mathbf{X}_k and compute the probabilities given in Equation 5.1.
- 3. Repeat Step 2 500 times and approximate Equation 5.2 by the average of these five hundred probabilities.

The sets chosen for this simulation are

- $S = \{0.05, 0.10, 0.15, 0.20\}$ for the linear model and $S = \{0.010, 0.025, 0.050, 0.100\}$ for the functional regression model.
- $\mathcal{D} = \{0.010, 0.025, 0.050, 0.100\}$

• $\mathcal{K} = \{5, 10, \dots, 95, 100, 110, \dots, 230, 240, 249\}$ for the linear model and $\mathcal{K} = \{30, 35, \dots, 95, 100, 110, \dots, 230, 240, 249\}$ for the functional model.

For reference, the estimate of σ for the linear model in Section 4.5 for our data is $\hat{\sigma} = 0.091$ for the start of the bottom phase and $\hat{\sigma} = 0.087$ for the end of the bottom phase when using all 250 dives. For the functional model with depth as a functional predictor from Section 4.8, $\hat{\sigma} = 0.078$ for the start of the bottom phase and $\hat{\sigma} = 0.075$ for the end of the bottom phase when using all 250 dives.

5.2.2 User-Provided Approach

For this approach, the RMSE and MAE is computed using \hat{Y}_i^k and Y_i in order to determine an adequate value of k. I perform simulations as follows:

- 1. For each $k \in \mathcal{K}$, where $\mathcal{K} = \{30, 35, \dots, 245, 250\}$, randomly select *k* dives from all *N* dives, and fit each model to this random sample.
- 2. Compute the fitted values, \hat{Y}_i^k , for every dive, and use \hat{Y}_i^k and Y_i to compute the RMSE and MAE using Equation 4.9.
- 3. Repeat steps 1 and 2 500 times.

5.3 Results

The first simulation for the probabilistic approach in Section 5.2.1 shows that as the sample size k increases, the probability also increases (Figures 5.1, 5.2). This positive relationship can be thought of as an increase in a user's confidence in that a certain level of precision will be ascertained as the sample size increases. Moreover, as the tolerance level increases (i.e. target precision decreases), and all else remaining constant, the confidence that a user has in the model will also increase. Furthermore, for highly variable data (large σ), more data are required to fit these models to achieve the same level of confidence in these predictions. Finally, comparing the linear regression model results to the functional regression model results, more data are needed for the functional regression model since there are more parameters to be estimated.

For the user-provided approach in Section 5.2.2, all models needed less than 100 user-marked dives to perform well (Figure 5.3), according to both the RMSE and MAE metric. The 70% threshold model, which does not depend on the training data and hence is constant for all sample sizes, underperforms against every other model when k > 50. The RMSE and MAE for the percentage threshold model also remain relatively constant for all sample sizes; however, this model performs better than the 70% threshold model for both the start and end of the bottom phase. The estimated

percentage threshold value, $\hat{\theta}$, stabilizes early, resulting in stable values for RMSE and MAE for this model. Hence, for this dataset, the percentage threshold model does not require a large sample size. The RMSE and MAE for the basic linear regression model are variable for low sample sizes, but eventually level out and tend to perform better than the 70% threshold model, but worse than the functional regression models. The RMSE and MAE for the functional regression model are high for small sample sizes, as expected; however, when the sample size is large (k > 50), the RMSE and MAE for the functional models are low compared to the other models. Finally, the RMSE and MAE are both lower for the end of the bottom phase than for the start of the bottom phase.

5.4 Conclusions

The probabilistic approach provides a significant advantage in determining an adequate sample size as the researchers can carry out these simulations using their own datasets, which will depend on the covariates used in the models as well as the total number of dives in the dataset. This approach is limited to a linear model and assumes normality. This method also requires a reasonable estimate for σ , the variability in the error, which can be estimated by fitting the models to a preliminary sample of dives. To account for error in the estimation of σ , I recommend conducting the simulations for several values of σ which are close to the estimate from the preliminary sample. In doing so, these simulations can allow a researcher to determine an adequate value of *k* for their own datasets.

The meaning of the tolerance level d, which is on the same scale as the prediction vectors (proportion of time into the dive), can be understood by considering a simple example. Suppose the researcher decides to use a tolerance level of d = 0.01. Then, the researcher chooses a sample size, k, and computes the probability in Equation 5.1 to get 0.85. As such, the user can be 85% confident that the difference between \hat{Y}_i^k and \hat{Y}_i^N equates to 1% of the dive duration. So, if a dive is 100 seconds in duration, then this would equate to a difference of 1 second, whereas for a dive that is 200 seconds in duration, this difference would be 2 seconds. Moreover, if 85% confidence is not high enough, then the sample size can be increased until the desired confidence is achieved. A researcher could also increase the tolerance level to achieve a higher confidence level.

The user-provided approach shows that, for this dataset, only about 100 dives needed to have the bottom phase marked before the RMSE and MAE both leveled off, after which point, the improvements are marginal regardless of the model chosen. This approach shows that the functional regression model tends to perform better than the other models. However, all models still perform better than the traditional 70% threshold method. Moreover, the RMSE and MAE are generally smaller for all models when comparing the start of the bottom phase to the end of the bottom phase. This result is consistent with the prior hypothesis that the end of the bottom phase is less variable than the start, as discussed in Section 3.4. While replicating these simulations would require the user to have a user-provided bottom phase, Y_i , for every dive, the results from this simulation can give an indication of the sample size needed for each model, before there are no further improvements in the RMSE or MAE (approximately 100 dives for each model). As such, I recommend using the *results* of the simulation presented here to guide researchers in the trends as *k* increases, rather than replicating the simulation.



Figure 5.1: Results of the simulation study for the basic linear regression model in Equation 4.5 for the probabilistic approach in Section 5.2.1. Each plot corresponds to a different value for σ . The *x*-axis is the sample size and the *y*-axis is the probability. The tolerance levels, *d*, are shown by the different coloured lines.



Figure 5.2: Results of the simulation study for the functional regression model in Equation 4.8 for the probabilistic approach in Section 5.2.1. Each plot corresponds to a different value for σ . The *x*-axis is the sample size and the *y*-axis is the probability. The tolerance levels, *d*, are shown by the different coloured lines.



Figure 5.3: Results of the simulation study for the user-provided approach. The plots on the top row correspond to the RMSE metric while the plots on the bottom row correspond to the MAE metric. The left column plots are for the start of the bottom phase, and the right column plots are for the end of the bottom phase. The *x*-axis is the sample size and the *y*-axis is the value of the metric. Each model is represented by a different coloured line. Error bars are omitted for clarity.

Chapter 6

The Bottom Line

The end of a melody is not its goal: but nonetheless, had the melody not reached its end it would not have reached its goal either. A parable. — Nietzsche

This work highlights proposed solutions for a reproducible and statistically sound definition for the bottom phase of a dive. Through using different bottom phase definitions, which are used when computing summary statistics to predict prey capture dives (see Section 2.5), I demonstrated the importance of having an accurate and well-defined bottom phase for a dive. I determined the major sources of variation in user-based dive phase definitions using an experimental study, discovering substantial researcher-to-researcher variation, within-researcher variation, and across-dives variation in bottom phase definitions, indicating a strong need for a statistical model for bottom phase definitions rather than relying on threshold values which may not represent the data appropriately. I then presented several models that could be used by individual researchers to model the bottom phase of a dive while incorporating their own personal datasets as well as bottom phase identifications that they deem appropriate. These statistical models proved to be more effective at matching the researcher's choices of the bottom phase compared to the previously used common threshold definitions. Finally, I established two approaches to determine an adequate sample size for these statistical models and implemented them via simulation studies. Of these sample size calculations, the probabilistic approach was designed in such a way so that a researcher can recreate the simulation and produce their own results for their dataset, allowing them to determine an adequate sample size.

The methods proposed in this thesis show several advantages over the currently used methods. Firstly, the statistical models I proposed provide a more theoretical, reproducible, and justifiable framework in defining the start and end of the bottom phase of a dive. These models incorporate the researcher's own tagging data, as well as their own researcher-specified starts and ends of bottom phases to train said models. These models can be easily presented and possibly used by other researchers, assuming that both datasets are similar and that both researchers have similar variation in user-defined bottom phases. As such, it is imperative for a researcher to report the model estimates in their findings to allow for reproducibility. Moreover, these models provide substantial improvements in matching user bottom phase markings of a dive compared to the commonly used threshold methods. Specifically, I showed that by implementing a functional regression model with the time-depth curve of a dive as a functional predictor and using the researcher's choices of bottom phase as the response variable, I reduced the number of false positive prey capture dives by nearly 50% using the aforementioned prey capture detection algorithm from Section 2.5. These findings make intuitive sense since the threshold models did not work well for complex dives with awkward shapes, whereas the functional models are able to handle these types of dives.

Several extensions to the methods presented in this thesis can be considered as a next step. I first note that the start and end of the bottom phase have an inherent dependence in that the start must occur before the end. As such, one could jointly model the start and end of the bottom phase simultaneously to guarantee that the start occurs before the end [38]. Even though the analyses presented in this thesis never encountered this issue, jointly modeling the response variables may be provide more advantages since the inherent dependence is accounted for. Moreover, other variables such as jerk, roll, or heading, as well as acoustic data, all of which are measured over the same time domain as the depth curve, could be incorporated into these models as functional predictors. One may also consider models where the functional predictors depend on other aspects of the dive itself, such as shape or behaviour exhibited [39]. Of course, this would require each dive to be labeled accordingly, either by the researcher or by other techniques such as a hidden Markov model [40–42]. Furthermore, if multiple whales are being analyzed, then random effects can be used to deal with the natural correlation of the dives clustered within a whale [39]. Finally, in this thesis, I only used the data from three NRKWs and more specifically, I only used data from one NRKW in the modeling section. The small sample size used here is a clear limitation, and future work should attempt to incorporate more data into the analyses to achieve more accurate and general results.

The models and methods presented in this thesis can be extended to other species beyond killer whales. Seals, sharks, otters, penguins, dolphins, and other species of whales are air breathing marine animals which exhibit diving behaviour, and predicting prey captures for these animals is also critical [8, 10–15]. Moreover, in contrast to solely predicting prey capture dives, other variables are used in the analysis of these diving data. For example, dive efficiency, which is the ratio of the time during the bottom phase of a dive to the total duration of the dive, is used when analyzing energy expenditures of diving animals [11–13]. Having a well-defined bottom phase can produce better estimates of dive efficiency.

Furthermore, some dives may exhibit more than one possible bottom phase. Sometimes, these are referred to as "steps in the ascent/descent" [8] and occur when an animal changes its vertical trajectory during either the ascent or descent phase, hence creating a pseudo-bottom phase. These steps typically occur due to changes in bathymetry (the ocean floor) [8], the animals engaging in social behaviours [3], or the animals searching for prey usually via echolocation [9]. Modeling and investigating the multiple bottom phases of these dives would be another extension of the methods presented in this thesis. Studying the implications of including these pseudo-bottom phases when attempting to detect prey capture dives could provide for interesting future work.

The parameter estimates and threshold values identified by Tennessen et al. [1] to predict prey capture dives depend on the bottom phase definition, species being studied, and potentially extraneous factors like time of year or researcher analyzing the data. As such, further work in this area could include creating a prey capture detection model akin to the ones presented in Tennessen et al. [1], except modifying the bottom phase definition using the methods shown in Chapter 4. Since these prey capture detection models depend on the bottom phase definition, one can compare the new estimated threshold values (similar to the values shown in Section 2.5) when using model-based bottom phase definitions to the currently existing thresholds values and examine how much they differ and the implications thereof. Moreover, one may find that these new threshold values, which are used to predict prey capture events, may include variables other than the ones determined by Tennessen et al. [1], such as ascent velocity, descent velocity, or dive efficiency. Overall, it would be pertinent to investigate new prey capture detection algorithms using the model-based bottom phase definitions.

Since the overarching issue here is to study the reasons for the differences in survival rates between the NRKWs and the SRKWs, studying the energy consumption of these two populations of whales is imperative. To do so requires knowledge of prey capture events, which can be identified in several ways such as by using a prey capture detection algorithm or by visually observing the events. When using a predictive model to determining prey capture events, the bottom phase definition has been shown to be highly important. The models used in this thesis may be able to provide more accurate estimates of the bottom phase of a dive, and thus, more accurate estimates of energy consumption allowing for accurate comparisons of survival rates between species.

Bibliography

- J.B. Tennessen, M.M. Holt, M.B. Hanson, C.K. Emmons, D.A. Giles, and J.T. Hogan. Kinematic signatures of prey capture from archival tags reveal sex differences in killer whale foraging activity. *J Exp Biol*, 222, 2019. doi:10.1242/jeb.191874. viii, ix, x, xi, 1, 2, 3, 4, 7, 10, 11, 12, 14, 15, 17, 18, 19, 20, 21, 22, 23, 26, 28, 29, 39, 47, 58, 61, 62, 63, 74
- [2] National Oceanic and Atmospheric Administration. The Endangered Species Act. Federal Register, 70(222), 2005. doi:041213348-5285-02. 1
- [3] B.M. Wright, J.K.B. Ford, G.M. Ellis, V.B. Deecke, A.D. Shapiro, B.C. Battaile, and A.W. Trites. Fine-scale foraging movements by fish-eating killer whales (*Orcinus orca*) relate to the vertical distributions and escape responses of salmonid prey (*Oncorhynchus spp.*). Mov Ecol, 5(3), 2017. doi:10.1186/s40462-017-0094-0. 1, 4, 12, 74
- [4] Fisheries and Oceans Canada. Recovery Strategy for the Northern and Southern Resident Killer Whales (Orcinus orca) in Canada. Species at Risk Act Recovery Strategy Series, ix + 81 pp, 2008. 1
- [5] B.M. Wright, V.B. Deecke, G.M. Ellis, A.W. Trites, and J.K.B. Ford. Behavioral context of echolocation and prey-handling sounds produced by killer whales (*Orcinus orca*) during pursuit and capture of Pacific salmon (*Oncorhynchus spp.*). *Mar Mammal Sci*, (1-26), 2021. doi:10.1111/mms.12836. 1, 12
- [6] M.M. Holt, D.P. Noren, and C.K. Emmons. An investigation of sound use and behavior in a killer whale (*Orcinus orca*) population to inform passive acoustic monitoring studies. *Mar Mammal Sci*, 29(193-202), 2013. doi:10.1111/j.1748-7692.2012.00599.x. 1
- [7] P. Arranz, S.L. DeRuiter, A.K. Stimpert, S. Neves, A.S. Friedlaender, J.A. Goldbogen,
 F. Visser, J. Calambokidis, B.L. Southall, and P.L. Tyack. Discrimination of fast click-series produced by tagged Risso's dolphins (*Grampus griseus*) for echolocation or communication. *J Exp Biol*, 219(2898-2907), 2016. doi:10.1242/jeb.144295. 1, 2, 10, 28, 39, 41, 49
- [8] M. Viviant, P. Monestiez, and C. Guinet. Can We Predict Foraging Success in a Marine Predator from Dive Patterns Only? Validation with Prey Capture Attempt Data. *PLoS ONE*, 9(3), 2014. doi:10.1371/journal.pone.0088503. 2, 14, 28, 41, 62, 73, 74

- [9] M. Viviant, T. Jeanniard-du Dot, P. Monestiez, M. Authier, and C. Guinet. Bottom time does not always predict prey encounter rate in Antarctic fur seals. *Functional Ecology*, 30 (1834-1844), 2016. doi:10.1111/1365-2435.12675. 2, 14, 28, 39, 74
- B.A. Nolet, D.E.H. Wansink, and H. Kruuk. Diving of Otters (*Lutra lutra*) in a Marine Habitat: Use of Depths by a Single-Prey Loader. *J Anim Ecol*, 62(1):22–32, 1993. doi:10.2307/5479. 4, 73
- [11] S. Andrzejaczek, A.C. Gleiss, K.O. Lear, C. Pattiaratchi, T.K. Chapple, and M.G. Meekan. Depth-dependent dive kinematics suggest cost-efficient foraging strategies by tiger sharks. *Royal Society Open Science*, 7(8):200789, 2020. doi:10.1098/rsos.200789. 4, 73
- [12] I. Zimmer, R.P. Wilson, M. Beaulieu, Y. Ropert-Coudert, A. Kato, A. Ancel, and J. Plotz. Dive efficiency versus depth in foraging emperor penguins. *Aquat Biol*, 8:269–277, 2010. doi:10.3354/ab00213. 4
- [13] S.L. Watwood, P.J.O. Miller, M. Johnson, P.T. Madsen, and P.L. Tyack. Deep-diving foraging behaviour of sperm whales (*Physeter macrocephalus*). J Anim Ecol, 75(3):814–825, 2006. doi:10.1111/j.1365-2656.2006.01101.x. 4, 73
- [14] R.C. Skrovan, T.M. Williams, P.S. Berry, and P.W. Moore and R.W. Davis. The diving physiology of bottlenose dolphins (*Tursiops truncatus*). II. Biomechanics and changes in buoyancy at depth. *J Exp Biol*, 202(20):2749–2761, 1999. doi:10.1242/jeb.202.20.2749. 4
- [15] R.R. Reisinger, M. Keith, R.D. Andrews, and P.J.N. de Bruyn. Movement and diving of killer whales (*Orcinus orca*) at a Southern Ocean archipelago. *Journal of Experimental Marine Biology and Ecology*, 473:90–102, 2015. doi:10.1016/j.jembe.2015.08.008. 4, 73
- [16] CATs Visualizer. http://www.cats.is/products/cats-visualizer/, 2019. Accessed: 2020-09-01. 5, 7
- [17] R.P. Wilson, L. Börger, M.D. Holton, D.M. Scantlebury, A. Gómez-Laich, F. Quintana, F. Rosell, P.M. Graf, H. Williams, R. Gunner, L. Hopkins, N. Marks, N.R. Geraldi, C.M. Duarte, R. Scott, M.S. Strano, H. Robotka, C. Eizaguirre, A. Fahlman, and E.L.C. Shepard. Estimates for energy expenditure in free-living animals using acceleration proxies: A reappraisal. *J Anim Ecol*, 89(161-172), 2020. doi:10.1111/1365-2656.13040. 7, 9
- [18] R.W. Baird. Killer Whales of the World: Natural History and Conservation. MBI Publishing Company LLC, Cambridge, Massachusetts, 1st edition, 2006. 9
- S.K. Hooker and R.W. Baird. Diving and ranging behaviour of odontocetes: a methodological review and critique. *Mammal Rev*, 31(81-105), 2001. doi:10.1046/J.1365-2907.2001.00080.X. 10, 29
- [20] Y. Le Bras, J. Jouma'a, and C. Guinet. Three-dimensional space use during the bottom phase of southern elephant seal dives. *Mov Ecol*, 5(18), 2017. doi:10.1186/s40462-017-0108-y. 14

- [21] S.H. Ridgeway, C.A. Bowers, D. Miller, M.L. Schultz, C.A. Jacobs, and C.A. Dooley. Diving and blood oxygen in the white whale. *Can J Zool*, 62(2349-2351), 1984. 16
- [22] A.K. Stimpert, S.L. DeRuiter, B.L. Southall, D.J. Moretti, E.A. Falcone, J.A. Goldbogen, A. Friedlaender, G.S. Schorr, and J. Calambokidis. Acoustic and foraging behavior of a Baird's beaked whale, *Berardius bairdii*, exposed to simulated sonar. *Scientific Reports*, 4 (7031), 2014. doi:10.1038/srep07031. 28
- [23] H. Bailey, G. Shillinger, D. Palacios, S. Bograd, J. Spotila, F. Paladino, and B. Block. Identifying and comparing phases of movement by leatherback turtles using state-space models. *J Exp Mar Biol Ecol*, 356:128–135, 2008. 28
- [24] S.P. Luque. Diving Behaviour Analysis in R. *R News*, 7(3):8–14, June 2007. ISSN -2022. URL https://cran.r-project.org/doc/Rnews/. Contributions from: John P.Y. Arnould, Laurent Dubroca, and Andy Liaw. 28
- [25] L. G. Halsey, C.A. Bost, and Y. Handrich. A thorough and quantified method for classifying seabird diving behaviour. *Polar Biol*, 30(991-1004), 2007. doi:10.1007/s00300-007-0257-3. 29
- [26] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting Linear Mixed-Effects Models Using Ime4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi:10.18637/jss.v067.i01. 34
- [27] J. Hox, M. Moerbeek, and R. van de Schoot. *Multilevel Analysis*. Routledge, New York, New York, 2nd edition, 2010. 34
- [28] A. Gelman and J. Hill. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, Cambridge, UK, 1st edition, 2007. 35
- [29] D.P. Nowacek, M.P. Johnson, P.L. Tyack, K.A. Shorter, W.A. McLellan, and D.A. Pabst. Buoyant Balaenids: The Ups and Downs of Buoyancy in Right Whales. *Biological Sciences*, 268(1478):1811–1816, September 2001. doi:10.1098/rspb.2001.1730. 38
- [30] G. Lauriano, E. Pirotta, T. Joyce, R.L. Pitman, A. Borrell, and S. Panigada. Movements, diving behaviour and diet of type-C killer whales (*Orcinus orca*) in the Ross Sea, Antarctica. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 30(12):2428–2440, 2020. doi:10.1002/aqc.3371. 38
- [31] I. Murphy. bustR: An R package for dive phase definitions. https://github.com/iamurphy/bustR, 2021. 38, 61
- [32] J.O. Ramsay and B.W. Silverman. Functional Data Analysis. Springer, 2010. 40
- [33] J.O. Ramsay and B.W. Silverman. *Applied Functional Data Analysis Methods and Case Studies*. Springer, 2002. 40

- [34] J. Goldsmith, J. Bobb, C.M. Crainiceanu, B. Caffo, and D. Reich. Penalized Functional Regression. *Journal of Computational and Graphical Statistics*, 20(4):830–851, 2011. doi:10.1198/jcgs.2010.10007. 40
- [35] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/. 41
- [36] B. Efron and R. Tibshirani. An Introduction to the Bootstrap. Chapman and Hall/CRC, 1993.41
- [37] W. Chang, J. Cheng, J.J. Allaire, Y. Xie, and J. McPherson. *shiny: Web Application Framework for R*, 2020. URL https://CRAN.R-project.org/package=shiny. R package version 1.5.0. 60
- [38] C. Tofallis. Model Building with Multiple Dependent Variables and Constraints. Journal of the Royal Statistical Society. Series D (The Statistician), 48(3):371–378, 1999. URL https://www.jstor.org/stable/2680999. 73
- [39] J. Goldsmith, C. M. Crainiceanu, B. Caffo, and D. Reich. Longitudinal Penalized Functional Regression for Cognitive Outcomes on Neuronal Tract Measurements. *Journal of the Royal Statistical Society. Series C, Applied statistics*, 61(3):453–469, 2012. doi:10.1111/j.1467-9876.2011.01031.x. 73
- [40] W. Zucchini, I.L. MacDonald, and R. Langrock. *Hidden Markov Models for Time Series: An Introduction Using R.* Chapman and Hall, Boca Raton, Florida, 2016. 73
- [41] V. Leos-Barajas, Gangloff E.J., T. Adam, R. Langrock, F.M. van Beest, J. Nabe-Nielsen, and J.M. Morales. Multi-scale Modeling of Animal Movement and General Behavior Data Using Hidden Markov Models with Hierarchical Structures. *Journal of Agricultural, Biological and Environmental Statistics*, 22(3):232–248, 2017. doi:10.1007/s13253-017-0282-9.
- [42] E. Sidrow, N. Heckman, S.M.E. Fortune, A.W. Trites, I. Murphy, and M. Auger-Méthé. Modelling multi-scale state-switching functional data with hidden Markov models, 2021. URL https://arxiv.org/abs/2101.03268. 73
- [43] R.D. Cook. Detection of Influential Observation in Linear Regression. *Technometrics*, 19(1): 15–18, February 1977. doi:10.2307/1268249. 81, 84

Appendix A

Supporting Materials

A.1 Matrix and Vector Formulation of Models from Chapter 3

Here, we present the exact structure and form of the vectors and matrices from Eq 3.3. For completeness, Eq 3.3 states

$$\mathbf{Y} = \boldsymbol{\mu} \mathbf{1} + \mathbf{Z}\mathbf{R} + \mathbf{W}\mathbf{D} + \mathbf{X}\mathbf{T} + \boldsymbol{\varepsilon}$$

From this model, we have the following vectors and matrices:

• Y is the complete vector of responses containing data from each researcher on their unique and common dives.

$$\mathbf{Y} = \left(\mathbf{Y}_1^T, \mathbf{Y}_2^T \cdots, \mathbf{Y}_r^T\right)^T \tag{A.1}$$

- μ is the overall mean and **1** is a vector of ones of length *n* where $n = \sum_{i=1}^{r} K'_{i}$.
- **R** is a vector of researcher effects given as $\mathbf{R} = (R_1, R_2, \cdots, R_r)^T$, and thus the matrix **Z** is defined as

$$\mathbf{Z} = \begin{pmatrix} \mathbf{1}_{K'_1+m_1} & 0 & 0 & \dots & 0 \\ 0 & \mathbf{1}_{K'_2+m_2} & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{1}_{K'_r+m_r} \end{pmatrix}$$
(A.2)

• **D** is the vector of dive effects given as $\mathbf{D} = (\mathbf{D}^C, \mathbf{D}_1^U, \cdots, \mathbf{D}^C, \mathbf{D}_r^U)^T$. We define \mathbf{W}_i as the matrix for the *i*th researcher as

$$\mathbf{W}_{i} = \begin{pmatrix} \mathbf{1}_{K_{i1}} & 0 & 0 & \dots & 0 & 0 \\ 0 & \mathbf{1}_{K_{i2}} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{1}_{K_{id_{C}}} & 0 \\ 0 & 0 & 0 & \dots & 0 & \mathbb{I}_{m_{i} \times m_{i}} \end{pmatrix}$$
(A.3)

Then, putting each one of these together, we get

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 & 0 & 0 & \dots & 0 \\ 0 & \mathbf{W}_2 & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{W}_r \end{pmatrix}$$
(A.4)

• **T** is the vector of interaction effects given as $\mathbf{T} = ((\mathbf{RD}^C)_1, (\mathbf{RD}^U)_1, \cdots, (\mathbf{RD}^C)_r, (\mathbf{RD}^U)_r)^T$ and thus we define the matrix \mathbf{X}_i as

$$\mathbf{X}_{i} = \begin{pmatrix} \mathbf{1}_{K_{i1}} & 0 & 0 & \dots & 0 & 0 \\ 0 & \mathbf{1}_{K_{i2}} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \\ 0 & 0 & 0 & \dots & \mathbf{1}_{K_{id_{C}}} & 0 \\ 0 & 0 & 0 & \dots & 0 & \mathbb{I}_{m_{i} \times m_{i}} \end{pmatrix}$$
(A.5)

Then, stacking these matrices together for each researcher, we get

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_{1} & 0 & 0 & \dots & 0 \\ 0 & \mathbf{X}_{2} & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{X}_{r} \end{pmatrix}$$
(A.6)

• Finally, we define $\boldsymbol{\varepsilon}$ as the vector of random errors as

$$\boldsymbol{\varepsilon} = \left(\boldsymbol{\varepsilon}_1^C, \boldsymbol{\varepsilon}_1^U, \cdots, \boldsymbol{\varepsilon}_r^C, \boldsymbol{\varepsilon}_r^U\right)^T \tag{A.7}$$

The notation for these models is quite complex, but when it is broken down to its base components, it is really just matrices of ones and zeros arranged in a specific way to accommodate the desired form of this model.

A.2 Model Diagnostics for Percentage Threshold Model



Residual plots for the percentage threshold model are given here.

Figure A.1: Residuals versus fitted values for the start of the bottom phase for the percent of max depth model.

A.3 Model Diagnostics for the Linear Model

To check the model assumptions for the linear model, consider the diagnostic plots (Figures A.3 and A.4). The standardized residuals versus fitted values tend to vary randomly about the blue smoothed line, albeit there are some outliers. For the start of the bottom phase (Figure A.3), there is minimal heteroskedasticity and most residuals are between -2.5 and 2.5, with a few exceptions. The assumptions of independent residuals seems to be satisfied for this model. For the end of the bottom phase (Figure A.4), there is variation in the residuals indicating potential violation of equal variance assumption. Both QQ-plots indicate that the Normality assumption is appropriate, since most sample quantiles follow along the line y = x, as expected, with few exceptions near the tails.

Cook [43] proposed a method to detect influential points, called Cook's D, where any point which is larger than $F_{(p,n-p)}(0.5)$, the median quantile of the *F*-distribution with *p* and *n* – *p* degrees of freedom, where *p* is the total number of covariates in the model (including an intercept) and *n* is the total number of data points in the dataset, is deemed influential. For our dataset, $F_{(p,n-p)}(0.5) = F_{(4,246)}(0.5) = 0.8415$, since p = 4 and n = 250 and no points are deemed to be influential (Figures



Figure A.2: Residuals versus fitted values for the end of the bottom phase for the percent of max depth model.

A.3 and A.4).



Figure A.3: Residual plots for the linear model in Equation 4.5 for the start of the bottom phase



Figure A.4: Residual plots for the linear model in Equation 4.5 for the end of the bottom phase

Residuals over time help to identify temporal patterns and trends as well as changes in the variance (Figure A.5). There is a slight change in the variance near the end of the residuals in both plots. One reason for this change in variance could be that the researcher became more consistent in defining the bottom phase of dives as time progressed. Also, there are several shallow dives near the end of the tag deployment which could result in more similar and consistent bottom phase definitions.

A.4 Model Diagnostics for the Functional Regression Models

Residual plots, QQ plots, and Cook's distance plots for the fitted models in Equation 4.8 are presented here for the functional regression models. The residual versus fitted plots are all similar for the start of the bottom phase, and there appears to be some distinct heteroskedasticity present. A similar, but less severe heteroskedasticity exists for the models for the end of the bottom phase as well. The QQ plots show that normality is a reasonable assumption for each model, although the tails are fairly heavy possibly due to a few outliers. Using the same rule discussed in Cook [43], the



Figure A.5: Standardized residual plots versus index for both the start and end of the bottom phase.

data points with a Cook's D larger than $F_{(24,226)}(0.5) = 0.9752$ are influential (since there are 20 basis function coefficients, an intercept, and three additional covariate coefficients to be estimated). However, no dives achieved a Cook's D higher than this threshold, indicating that there are unlikely to be highly influential points in this dataset.



Figure A.6: Residuals versus Fitted Values for the functional regression models in Chapter 4. Plots 1, 2, and 3 are for the depth, speed, and acceleration models for the start of the bottom phase respectively. Plots 4, 5, and 6 are for the depth, speed, and acceleration models for the end of the bottom phase respectively.



Figure A.7: QQ plots for the functional regression models in Chapter 4. Plots 1, 2, and 3 are for the depth, speed, and acceleration models for the start of the bottom phase respectively. Plots 4, 5, and 6 are for the depth, speed, and acceleration models for the end of the bottom phase respectively.



Figure A.8: Cook's distance plots for the functional regression models in Chapter 4. Plots 1, 2, and 3 are for the depth, speed, and acceleration models for the start of the bottom phase respectively. Plots 4, 5, and 6 are for the depth, speed, and acceleration models for the end of the bottom phase respectively. Each line corresponds to an individual data point.

A.5 Positive Semi-Definiteness of Σ

Lemma A.5.1. Let A, B be two positive definite symmetric matrices. Then, if B - A is positive definite, then $C^{T}(B-A)C$ is also positive definite for any matrix C with appropriate dimensions.

Proof. Consider the quantity $u^T C^T (B - A)Cu$ for any vector u of appropriate dimension. This can be written as $(Cu)^T (B - A)Cu$. If we let x = Cu, then we have $x^T (B - A)x > 0$ since B - A is positive definite. Thus, $u^T C^T (B - A)Cu > 0$ and hence $C^T (B - A)C$ is positive definite as required.

Lemma A.5.2. Let *B* be a positive definite symmetric matrix. If $B - \mathbb{I}$ is positive definite, then $\mathbb{I} - B^{-1}$ is positive definite.

Proof. Firstly, we note that every positive definite matrix is invertible, so B^{-1} exists. Next, if $B - \mathbb{I}$ is positive definite, then $(B^{-1/2})^T (B - \mathbb{I})B^{-1/2}$ is also positive definite by Lemma A.5.1. Simplifying everything yields

$$(B^{-1/2})^T (B - \mathbb{I}) B^{-1/2} = B^{-1/2} (B - \mathbb{I}) B^{-1/2}$$
$$= B^{-1/2} B B^{-1/2} - B^{-1/2} \mathbb{I} B^{-1/2}$$
$$= \mathbb{I} - B^{-1}$$

Thus $\mathbb{I} - B^{-1}$ is also positive definite.

Lemma A.5.3. Let A, B be symmetric and positive definite matrices. If B - A is positive definite, then $A^{-1} - B^{-1}$ is positive definite.

Proof. If B - A is positive definite, then by Lemma A.5.1, we have that $A^{-1/2}(B - A)A^{-1/2}$ must also be positive definite (using the fact that A is symmetric as well). Then,

$$A^{-1/2}(B-A)A^{-1/2} = A^{-1/2}BA^{-1/2} - A^{-1/2}AA^{-1/2}$$
$$= A^{-1/2}BA^{-1/2} - \mathbb{I}$$

is positive definite. Now, applying Lemma A.5.2, we must also have that $\mathbb{I} - [A^{-1/2}BA^{-1/2}]^{-1}$ is positive definite, which upon simplifying, we get $\mathbb{I} - A^{1/2}B^{-1}A^{1/2}$. Now, applying Lemma A.5.1 to this expression by multiplying on the left and right by $A^{-1/2}$, we get

$$A^{-1/2} \left(\mathbb{I} - A^{1/2} B^{-1} A^{1/2} \right) A^{-1/2} = A^{-1/2} \mathbb{I} A^{-1/2} - A^{-1/2} A^{1/2} B^{-1} A^{1/2} A^{-1/2}$$
$$= A^{-1} - B^{-1}$$

is positive definite.

Theorem A.5.4. Suppose X_N is an $N \times p$ matrix and X_k is a $k \times p$ matrix of k randomly selected rows from X_N . Suppose X_N is constructed such that $X_N = \begin{bmatrix} X_k^T & (X^*)^T \end{bmatrix}^T$ where X^* are the remaining rows not chosen from X_N . Then, $X_N \begin{bmatrix} (X_k^T X_k)^{-1} - (X_N^T X_N)^{-1} \end{bmatrix} X_N^T$ is positive definite.

Proof. It should be clear that both $X_k^T X_k$ and $X_N^T X_N$ are positive definite. We can see that

$$\begin{split} X_N^T X_N &= \left[\begin{array}{cc} X_k^T & X^{*T} \end{array} \right] \left[\begin{array}{c} X_k \\ X^* \end{array} \right] \\ &= X_k^T X_k + X^{*T} X^* \end{split}$$

Hence, $X_N^T X_N - X_k^T X_k = X^{*T} X^*$. Clearly, $X^{*T} X^*$ is positive definite, so $X_N^T X_N - X_k^T X_k$ is positive definite. Using Lemma A.5.3, we get that $(X_k^T X_k)^{-1} - (X_N^T X_N)^{-1}$ is positive definite too. Applying Lemma A.5.1, we get that $X_N \left[(X_k^T X_k)^{-1} - (X_N^T X_N)^{-1} \right] X_N^T$ is positive definite, as required. \Box