Habitat Associations and Reproduction of Fishes on the Northwestern Gulf of Mexico Shelf Edge

Elizabeth Marie Keller

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Applied Statistics Commons, Aquaculture and Fisheries Commons, Biodiversity Commons, Marine Biology Commons, Multivariate Analysis Commons, Oceanography Commons, Population Biology Commons, and the Statistical Models Commons

Recommended Citation

Keller, Elizabeth Marie, "Habitat Associations and Reproduction of Fishes on the Northwestern Gulf of Mexico Shelf Edge" (2019). LSU Doctoral Dissertations. 5092.
https://digitalcommons.lsu.edu/gradschool_dissertations/5092

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
11-6-2019

Habitat Associations and Reproduction of Fishes on the Northwestern Gulf of Mexico Shelf Edge

Elizabeth Marie Keller

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Applied Statistics Commons, Aquaculture and Fisheries Commons, Biodiversity Commons, Marine Biology Commons, Multivariate Analysis Commons, Oceanography Commons, Population Biology Commons, and the Statistical Models Commons
HABITAT ASSOCIATIONS AND REPRODUCTION OF FISHES ON THE NORTHWESTERN GULF OF MEXICO SHELF EDGE

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Oceanography and Coastal Sciences

by

Elizabeth Marie Keller
B.S., Cornell University, 2014
December 2019
“The truth. It is a beautiful and terrible thing and should therefore be treated with caution.”

—A.P.W.B.D.
ACKNOWLEDGMENTS

I would like to thank everyone who contributed to the success of this project and experience:

My advisor, Dr. Jim Cowan, without whom I would not have had the opportunity to do this research.

My committee—Drs. Bargu, Bentley, Galvez, Marx, and Midway—for their support, edits, and willingness to fill out numerous meeting scheduling polls.

All the Cowan lab members, Dave Nieland, captains and crew, and friends and helpers in the field: this work could not have been done without you.

Hilary Glenn, Tori Barker, Alayna Petre, and Toby Jarvis for equipment and software training.

Sarah Margolis and Kat Ellis, my project teammates.

All my friends and roommates, who helped me enjoy my time in Louisiana.

Kendall Valentine, for helping me maintain my sanity and survive the PhD later years.

My family, for supporting my crazy dream to pursue science and go to graduate school.

And thank you to JC Guerrero, my partner and my light at the end of the tunnel.

Funding for this research was provided by the National Oceanic and Atmospheric Administration Saltonstall-Kennedy Grant Program and Louisiana Department of Wildlife and Fisheries. Access to oil and gas platforms was provided by Apache Shelf, Inc., Fieldwood Energy LLC, and Arena Offshore, LP.
# CONTENTS

Acknowledgments.......................................................................................................................... iii

List of Tables .................................................................................................................................. vi

List of Figures ............................................................................................................................... vii

Abstract ........................................................................................................................................ ix

General Introduction ....................................................................................................................... 1
  Literature Cited ............................................................................................................................. 6

Chapter 1. Bayesian Estimation of Length at Sexual Transition for Six Species of Grouper
(Epinephelidae) in the Northwestern Gulf of Mexico ................................................................. 9
  Introduction ................................................................................................................................. 9
  Materials and Methods .............................................................................................................. 13
  Results ...................................................................................................................................... 15
  Discussion ................................................................................................................................. 19
  Literature Cited ......................................................................................................................... 26

Chapter 2. Assessing Optimal Number of Cameras for a Multi-Camera Stationary Baited
Remote Underwater Video System ............................................................................................... 34
  Introduction ............................................................................................................................... 34
  Materials and Methods .............................................................................................................. 36
  Results ....................................................................................................................................... 40
  Discussion ................................................................................................................................. 47
  Literature Cited ......................................................................................................................... 51

Chapter 3. Relationships and Prediction Between Fish Assemblage and Habitat on the
Northwestern Gulf of Mexico Shelf-Edge Banks ........................................................................ 57
  Introduction ............................................................................................................................... 57
  Methods ..................................................................................................................................... 59
  Results ....................................................................................................................................... 66
  Discussion ................................................................................................................................. 76
  Literature Cited ......................................................................................................................... 84

Chapter 4. Predicting Fish Backscatter and Density Based on Habitat Along the
Northwestern Gulf of Mexico Shelf-Edge .................................................................................... 89
  Introduction ............................................................................................................................... 89
  Methods ..................................................................................................................................... 91
  Results ....................................................................................................................................... 98
  Discussion ............................................................................................................................... 107
  Literature Cited ....................................................................................................................... 110

General Summary and Conclusions .......................................................................................... 114
  Literature Cited ....................................................................................................................... 116
LIST OF TABLES

Table 1.1. Selected life history parameter estimates for six Epinephelidae grouper species ....... 12
Table 1.2. Prior distributions used for the Bayesian logistic regression model....................... 15
Table 1.3. Sample sizes and total length ranges (mm TL) of grouper specimens by sex .......... 16
Table 1.4. Parameter estimates for Bayesian model.......................................................... 17
Table 1.5. Parameter estimates for frequentist GLM model............................................... 18
Table 2.1. Site demographics of oil and gas platforms sampled in the Eugene Island (EI) Oil and Gas Lease Block................................................................. 37
Table 2.2. Detection rate and MaxN for observed taxa......................................................... 41
Table 2.3. Metrics for taxonomic richness, diversity, and MaxN observed by the three BRUVS configurations................................................................. 46
Table 3.1. Habitat characteristics measured via video.......................................................... 63
Table 3.2. Score scale for percent cover estimates............................................................... 63
Table 3.3. Binary classification evaluation measures........................................................... 65
Table 3.4. Habitat characteristics....................................................................................... 66
Table 3.5. Metrics for identified taxa.................................................................................. 67
Table 3.6. Biological community metrics............................................................................ 69
Table 4.1. Habitat variable predictors included in boosted regression models.................... 96
Table 4.2. Binary classification evaluation measures.......................................................... 98
Table 4.3. Summary statistics for select habitat predictors recorded and derived from split-beam data................................................................. 99
Table 4.4. Summary statistics for select habitat predictors derived from multibeam sonar data............................................................................................................. 100
Table 4.5. Error rates and metrics for probability cutoff values of 0.5 and 0.13............... 106
LIST OF FIGURES

Figure 1.1. Specimen lengths for *Cephalopholis fulva*, *Dermatolepis inermis*, *Epinephelus adscensionis*, *Epinephelus guttatus*, *Hyporthodus flavolimbatus*, and *Mycteroperca phenax*. 16

Figure 1.2. Bayesian logistic regression curves of the relationship between total length (mm) and sex. 17

Figure 1.3. Frequentist logistic regression curves of the relationship between total length (mm) and sex. 18

Figure 2.1. Study sites in the Eugene Island Oil and Gas Lease Block off the coast of Louisiana. 37

Figure 2.2. Aerial view of the BRUVS configurations. 39

Figure 2.3. Violin plots of percentage detections, richness, and MaxN detected by 4-camera and 2-camera BRUVS relative to six-camera BRUVS configuration. 43

Figure 2.4. Comparison of metrics between BRUVS configurations. 44

Figure 3.1. Six banks surveyed along the northwestern Gulf of Mexico shelf edge. 59

Figure 3.2. Aerial view of the BRUVS configurations. 61

Figure 3.3. Plot of NMDS1 and NMDS2 for a three-dimensional nonmetric multidimensional scaling. 70

Figure 3.4. Class error rates for the three random forest models. 72

Figure 3.5. Classification metric for three random forest types. 73

Figure 3.6. Histograms for AUC values for three random forest models. 74

Figure 3.7. Partial dependency plots for *Serranus annularis*. 77

Figure 3.8. Partial dependency plots for *Mycteroperca interstitialis*. 77

Figure 3.9. Trade-offs between class error rates and their relationship with detection frequency and PR AUC. 83

Figure 4.1. Schematic of acoustics sampling cruise track. 92

Figure 4.2. Dataflow diagram representing acoustic data processing in Echoview. 93

Figure 4.3. Relative influence of habitat predictors for BRT model of MVBS. 101

Figure 4.4. Partial dependence plot for the effects of cell depth and bottom offset. 102
Figure 4.5. Histogram of residuals for values predicted by the MVBS BRT model .................. 103

Figure 4.6. Relationship between observed MVBS and MVBS predicted by the BRT model for a dataset independent of model training ................................................................. 103

Figure 4.7. Relative influence of habitat predictors for binary fish density BRT model ....... 104

Figure 4.8. Error rates and classification metrics for a range of probability cutoff value for predicting a positive value ................................................................. 105

Figure 4.9. Overall model accuracy as a function of probability at which values were predicted in the positive class ................................................................. 106
ABSTRACT

Several of the northwestern Gulf of Mexico (GOM) shelf-edge banks provide critical hard bottom habitat for coral and fish communities, supporting a wide diversity of ecologically and economically important species. These sites may be fish aggregation and spawning sites and provide important habitat for fish growth and reproduction. Already designated as habitat areas of particular concern, many of these banks are also under consideration for inclusion in the expansion of the Flower Garden Banks National Marine Sanctuary. This project aimed to gain a more comprehensive understanding of the communities and fish species on shelf-edge banks by way of gonad histology, baited remote underwater video, and hydroacoustics, as well as traditional statistical analyses, Bayesian estimation, and machine learning techniques. The study had several objectives: (1) estimate size at sexual transition for six GOM grouper species, (2) determine the optimal number of cameras on a baited remote underwater video system, (3) create a predictive model to provide presence of fish species based on habitat, and (4) grow a model to predict fish backscatter and density based on habitat parameters. Bayesian estimation allowed for size at sexual transition determinations for the six grouper species, outperforming the tradition frequentist models, especially for situations where tradition models failed to converge. Random forests based on video data had mixed results, but models for several species were able to predict fish presences with class and overall accuracies of greater than 80%. Boosted regression trees based on hydroacoustic data reinforced the importance of depth as a driving factor in fish distributions. The study provided greater understanding and predictive ability regarding fish on the bank habitats.
GENERAL INTRODUCTION

About 100 to 200 kilometers off the coast of Louisiana and Texas, the northwestern Gulf
of Mexico (GOM) is dotted with banks that in large part have formed on the top of salt diapirs
along the continental shelf edge. These 130 topographic features of the region provide most of
the 1 – 3% hard bottom occurring on the otherwise soft bottom of the northern GOM shelf
(Parker et al. 1983, Schmahl et al. 2008). This hard bottom provides critical habitat that supports
corals and associated communities on these banks.

Carbonate-capped bedrock along the shelf-edge was pushed by salt diapirs into a series of
hard substrate topographic features known as banks (Dennis and Bright 1988, Rezak et al. 1990).
The region of the continental shelf is composed of predominantly terrigenous sediment,
positioned away from coastal runoff and eutrophication, and subject to warm currents from the
Loop Current with temperatures ranging from about 18 – 30°C making it uniquely suitable for
coral communities (Schmahl et al. 2008). It is also important to note that these relatively high-
latitude coral communities are not simply a series of isolated favorable habitats, but are a system
of interconnected features (Gardner and Beaudoin 2005).

While fishermen have known of the fish diversity and value of the northwest GOM banks
since the 1890s, the bulk of the scientific research did not begin until the 1970s and ‘80s when
there was need to understand the ecosystems in the face of potential oil and gas development in
the area (Dennis and Bright 1988, Schmahl et al. 2008). These studies classified the benthic
communities at many of the topographic features along the shelf. Many studies have been
conducted since, but the vast majority focus within the Flower Garden Banks National Marine
Sanctuary or token other banks.
Shelf-edge banks support commercially and recreationally important fish species—fish species caught around these banks valued $7.5 million for the state of Texas and $14.5 million for Louisiana in 2005—as well as a diverse collection of both reef building and non-reef building coral species including deep corals (Schmahl et al. 2008). A recent census of global marine biodiversity ranked the GOM in the top five for species per area but also for threats to biodiversity (Costello et al. 2010). Deep-sea corals in particular have garnered attention and are thus in need of research and management. The National Oceanic and Atmospheric Administration (NOAA) has a strategic plan for deep-sea coral ecosystems. Goals for the GOM initiative include mapping the distribution of deep-sea coral ecosystems, characterizing habitat, and determining what drives differences between low-relief and high-relief habitats (National Oceanic and Atmospheric Administration Coral Reef Conservation Program 2010, Wagner et al. 2017). Preliminary studies also indicate that these features may serve as important areas for spawning and aggregation sites of notable species such as scamp (*Mycteroperca phenax*) and marbled grouper (*Dermatolepis inermis*) (Schmahl et al. 2008).

As an acknowledgment of the importance of these habitats, many banks (including all banks in this study) are designated as both Essential Fish Habitat (EFH) and Habitat Areas of Particular Concern (HAPC) by the National Oceanic and Atmospheric Association (NOAA) as well as No Activity Zones for oil and gas activities regulated through the Minerals Management Service. Despite these designations, most sites do not have special regulations or protections as HAPC zones. Furthermore, only the banks within the Flower Gardens Bank National Marine Sanctuary are subject to regular monitoring, leaving many of the other dozens of banks along the shelf edge unprotected, understudied, and unquantified in regards to their fish communities and habitat use (Kraus et al. 2006).
The lack of protection could change with the proposed expansion of the sanctuary. NOAA’s preferred alternative for expansion would add approximately 847 km² beyond the current sanctuary boundaries, including five of the six study sites (Office of National Marine Sanctuaries 2016). National Marine Sanctuary regulations would prohibit anchoring, subsurface salvage, dredging, explosives, drilling, trawling, and waste discharge (15 C.F.R. §922.61). An economic valuation study found that on average American households would be willing to pay from $35 to $107 for boundary expansion, underscoring the value of these ecosystems and their protection (Stefanski and Shimshack 2016). Understanding the true biological worth of additional protections of the ecosystems hinges on gathering additional data.

Since the push for research of these habitats in the 1970s, there has been repeated identification of distinct and consistent biotic habitat zonation. Rezak et al. (1990) described seven zones of biological communities which were later updated by Schmahl et al. (2008) and condensed into five biological zones. The biotic zones are primarily depth related, although the depth at which one zone transitioned to another depends on the bank (Rezak et al. 1990). These zones are home to distinct fish assemblages (Schmahl et al. 2008, Langland 2015). Habitat zones alone cannot adequately describe the groupings of fish. Interpretations of fish at these habitats need to consider multiple aspects of habitat and incorporate the interactions and complexities of the communities. Studies at an increasing breadth of the banks is needed to assess commonalities between the banks, especially in light of their interconnected nature (Schmahl et al. 2008). The goal of this project was to gain a more comprehensive understanding of the communities and fish species on the shelf-edge banks.

The first chapter focuses on several grouper species (Epinephelidae) found in the GOM that are protogynous hermaphrodites: coney (Cephalopholis fulva), red hind (Epinephelus
guttatus), rock hind (*Epinephelus adscensionis*), marbled grouper (*Dermatolepis inermis*), scamp (*Mycteroperca phenax*), and yellowedge grouper (*Hyporthodus flavolimbatus*). Information is lacking regarding grouper reproductive traits, especially in the Gulf of Mexico. For protogynous hermaphrodites, parameters such as length at sexual transition are critical to estimates of spawning stock biomass, and in turn stock assessments, which are an important part of maintaining a healthy population and meeting rebuilding targets (Trippel 1995, Collins et al. 1996, Woods 2003, Hood et al. 2007, Strelcheck and Hood 2007). Lack of information on groupers’ susceptibility to changes in sex ratio due to fishing pressure may result in incorrect management (Bates and Colvin 2016). Histology can also provide novel information on very poorly understood species, like marbled grouper (*Dermatolepis inermis*). The objective of the study was to establish size at sexual transition for the six species. Gonad histology was used to establish sex, while logistic models were used to model sex as a function of fish size. Frequentist and Bayesian estimation was used to determine the size at which half the individuals of a species had likely transitioned to become male. Five of these species’ populations have never been assessed in the northern Gulf of Mexico, and none have a reported stock status, making these data valuable as model inputs. To the authors’ knowledge, this is the first study to utilize multi-level Bayesian methods to estimate length at sexual transition.

Chapter 2 assesses an underwater video system to evaluate efficiency and richness of data collection. Baited remote underwater video systems (BRUVS) are an effective methodology to assess underwater populations. BRUVS have been utilized in many past studies which demonstrated the effectiveness in assessing the underwater environment due to less avoidance behavior, increased accuracy of measurements, lower variance in results, permanent data records, and their ability to sample areas that may otherwise be inaccessible (Gledhill 2001,
Langlois et al. 2010, Unsworth et al. 2014). However, they are limited by video processing time. This section considers the optimization of the number of cameras—and therefore necessary processing effort—for a 6-camera BRUVS. The objective was to determine the amount of data—measured by richness, diversity, and measured abundance—that is gained with use of additional cameras and where the optimal point of effort will be depending on survey goals.

Chapter 3 describes how the BRUVS system was used to record underwater surveys of the bank habitats to assess fish assemblages. Video was used to identify and count fish, as well as characterize surrounding habitat. Random forest models established the connections between habitat and fish species present. The goal was to (1) investigate the relationships between fish assemblages and habitat characteristics to determine what drives abundance of fish overall as well as of particular species of interest and (2) develop a predictive model to forecast presence of fish species based on habitat characteristics alone.

The fourth and final chapter examines the distribution of fish biomass and density across habitat features in the GOM. Hydroacoustics surveys were used to describe the biomass and density of fish via acoustic backscatter while habitat was measured via multibeam acoustic data. Hydroacoustic techniques allow for a quick and non-invasive survey approach (Langland 2015, Benoit-Bird and Lawson 2016). These data will allow analysis of how biomass of fish changes with habitat and environmental characteristics. Boosted regression trees (BRT) were used create models of fish backscatter and positive density. The objective was to create a model that could accurately predict fish backscatter and positive density based on habitat parameters.

The results of these studies will result in a better understanding of the species assemblages, habitat function, and fish reproductive capacity associated with shelf-edge banks that have not been comprehensively studied or fully understood. A variety of field and laboratory
methods were used with corresponding data analysis to extract information from data, even in the face of challenges such as limited data and imbalanced distributions. A combination of histology, BRUVS, and hydroacoustic surveys better characterized fish biology, habitats, and their associated assemblages, providing a better understanding of GOM fish that can inform managers. Comprehending the full scope of the value of these shelf-edge bank habitats is necessary because they may become even more important to these species’ growth and reproduction as the quantity of habitat declines due to the removal of oil and gas platforms, the progression of climate change, and the degradation of other reef communities.

**Literature Cited**


Bates, C., and R. Colvin. 2016. Comparison of scamp grouper (Mycteroperca phenax), growth off of the West Florida shelf and the coast of Louisiana. LaGrange College.


Langland, T. 2015. Fish assemblage structure, distribution, and trophic ecology at Northwestern Gulf of Mexico banks. Louisiana State University, Baton Rouge, LA.


CHAPTER 1. BAYESIAN ESTIMATION OF LENGTH AT SEXUAL TRANSITION FOR SIX SPECIES OF GROUPER (EPINEPHELIDAE) IN THE NORTHWESTERN GULF OF MEXICO

INTRODUCTION

Grouper species of the family Epinephelidae are predominantly hermaphroditic fish, making their reproduction interesting. A couple of dozen families of fish exhibit hermaphroditic species. Hermaphroditic species tend to be tropical and marine, the highest incidence of hermaphroditism being in coral reef fishes (Sadovy de Mitcheson and Liu 2008).

Hermaphroditism is a successful reproductive strategy when it is advantageous for there to be a skewed sex ratio or for one sex to be larger than the other. For example, female to male sex change can provide the opportunity for males to mate with multiple females, a situation which can balance the up to daily frequency of female spawning versus the much higher spawning frequency of males (Shapiro 1984). An individual changing sex is primarily influenced by behavioral and social controls (Sadovy de Mitcheson and Liu 2008).

Hermaphroditism is challenging to study. The mechanisms behind hermaphroditism are poorly understood. Sequential sex reversal has evolved independently in many families and is expressed variably, sometimes limited to single genera or population (Shapiro 1984, Sadovy de Mitcheson and Liu 2008). Within families where sex change has evolved, the strategy as to why some species exhibit sex reversal where others do not is not clear (Sadovy and Shapiro 1987, Shapiro et al. 1993).

Classifying a species as hermaphroditic can be difficult in and of itself. Some traits indicative of a hermaphroditic species, such as bimodal size frequency distribution between sexes, are not a conclusive determination of sex change. More unique factors like ovarian-like lumen in testis provides stronger support for protogynous hermaphroditism (Sadovy and Shapiro 1987). However, not all species or individuals will have a membrane-lined cavity in the testis.
Strong indicators of hermaphroditism all require histology, assessing tissue types and stages. The most conclusive indicator of hermaphroditism is the observation of individuals with proliferating tissue of one sex and degenerating tissues of the other (Sadovy and Shapiro 1987); however, such individuals are not commonly collected in the field. Investigation of sexual process is further muddled by the fact that there are both functional and non-functional hermaphrodites—individuals or species who have the necessary structure to change sex, but function gonochoristically (Sadovy de Mitcheson and Liu 2008). Even once identified, study of populations and life history can be difficult to assess for reasons like difficulty differentiating immature females from mature, resting females (Brulé and Colás-Marrufo 2013).

Hermaphroditism makes sexual transition an additional reproductive characteristic to study, alongside factors such as age and size at maturity, fecundity, and egg size that are important to both hermaphroditic and gonochoristic species (Smith 1965, Erisman et al. 2009). Understanding the process of sexual transition in hermaphroditic species is critical to fully grasp their life history. Length at sexual transition is related to the sex ratio and spawning stock biomass, which are crucial in maintaining a healthy population as well as assessing and managing fish stocks. Fishing pressure has been shown to alter sex ratios and influence length at sexual transition by preferentially removing large males and inducing relatively smaller females to transition to males (Coleman et al. 1996, Armsworth 2001, Heppell et al. 2006, Provost and Jensen 2015). This can result in reduced size and fecundity of mature individuals, negatively affecting reproduction and population growth (Reñones et al. 2010, Lowerre-Barbieri et al. 2011, Lombardi-Carlson et al. 2012, Provost and Jensen 2015). Furthermore, ignoring sex change in stock assessments and lack of information regarding groupers’ susceptibility to changes in sex ratio can lead to incorrect management (Alonzo et al. 2008, Bates and Colvin 2016).
Data on grouper spawning and reproduction are limited, especially in regions outside the Pacific and Caribbean. Studies in the GOM tend to focus on the eastern Gulf, leaving populations in the northwestern GOM understudied. Grouper populations from different regions have been shown to differ in life history parameters, such as length at maturity and sexual transition, so data from each different region are critical (Lombardi-Carlson et al. 2012, Nolan et al. 2017). Life history parameters such as length at maturity and sexual transition are not well understood for all grouper species (Table 1.1).

Six species are included in this study: coney Cephalopholis fulva (Linnaeus 1758), marbled grouper Dermatolepis inermis (Valenciennes 1833), red hind Epinephelus guttatus (Linnaeus 1758), rock hind Epinephelus adscensionis (Osbeck 1765), scamp Mycteroperca phenax (Jordan and Swain 1884), and yellowedge grouper Hyporthodus flavolimbatus (Poey 1865). Five of these species have been confirmed as protogynous hermaphroditic species, having individuals born as females and some transitioning to male during their lifespan (Sadovy et al. 1992, Cook 2007, Erisman et al. 2009, Freitas et al. 2011, Kline et al. 2011, Lombardi-Carlson et al. 2012, Brulé et al. 2016). The sexual pattern of D. inermis has not been confirmed but may very likely be a protogynous hermaphrodite as well considering the overwhelming protogyny among closely related species (Brulé et al. 2004, Erisman et al. 2009). Hyporthodus flavolimbatus is the only of the six species in this analysis currently studied via stock assessments in the region, but all six species had commercial landings in the GOM in 2016 (personal communication from the National Marine Fisheries Service, Fisheries Statistics Division [April 19 2018]).

The majority of data on these species is outside of the northwestern GOM region, while for rare species such as D. inermis nearly no reproductive information is known, regardless of
region. This study explores these grouper species in the northwestern GOM at banks along the continental shelf edge, with the primary intent to estimate each species’ length at sexual transition.

Table 1.1. Selected life history parameter estimates for six Epinephelidae grouper species, where TL indicates total length, FL indicates fork length, and SL indicates standard length. Length at maturity and length at transition are the length at which 50% of females are mature and 50% of individuals are male, respectively. All lengths are in millimeters.

<table>
<thead>
<tr>
<th>Species</th>
<th>Length at Maturity</th>
<th>Length at Transition</th>
<th>Maximum Length</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Cephalopholis fulva</em></td>
<td>130 FL</td>
<td>About 270 TL</td>
<td>491 TL (this study)</td>
</tr>
<tr>
<td></td>
<td>133 SL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>At or before 160 TL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>180 TL</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Dermatolepis inermis</em></td>
<td>Unknown</td>
<td>Unknown</td>
<td>910 TL</td>
</tr>
<tr>
<td><em>Epinephelus guttatus</em></td>
<td>201 FL</td>
<td>280 TL</td>
<td>760 TL</td>
</tr>
<tr>
<td></td>
<td>215 FL</td>
<td>378 FL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>240 TL</td>
<td>380 TL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>250 TL</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Epinephelus adscensionis</em></td>
<td>200 TL</td>
<td>418 TL</td>
<td>606 TL</td>
</tr>
<tr>
<td></td>
<td>250 SL</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>289 TL</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Mycteroperca phenax</em></td>
<td>332 FL</td>
<td>566 FL</td>
<td>1070 TL</td>
</tr>
<tr>
<td></td>
<td>353 TL (1990-97)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>374 TL (1979-89)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Hyporthodus flavolimbatus</em></td>
<td>512 TL whole GOM</td>
<td>815 TL</td>
<td>1170 TL</td>
</tr>
<tr>
<td></td>
<td>533 TL western GOM</td>
<td>817 TL</td>
<td></td>
</tr>
<tr>
<td></td>
<td>569 TL</td>
<td>811 TL eastern GOM</td>
<td></td>
</tr>
<tr>
<td></td>
<td>580 TL</td>
<td>840 TL whole GOM</td>
<td></td>
</tr>
</tbody>
</table>

MATERIALS AND METHODS

Specimen Collection

Grouper were sampled along the Texas-Louisiana shelf-edge in the northwestern GOM, approximately 215 km offshore in water depths between 50 and 150 meters. Samples were collected at locations between 89.82 and 93.45° W and 27.87 and 28.67° N. Fish were collected from 2011 to 2018 in all seasons, using single hook rods as well as longlines of five to ten hooks of alternating hook sizes. Total length (TL), fork length (FL), and standard length (SL) in millimeters (mm) were recorded for all specimens. Gonads were collected and refrigerated at sea until lab processing.

Tissue and Slide Preparation

Gonads were weighed to the nearest 0.01 gram and preserved in 10% formalin for at least two weeks. Species of grouper have been shown to have undelimited gonads which develop symmetrically and homogeneously (Smith 1965, Sadovy and Shapiro 1987, Collins et al. 1996, Collins et al. 1998, Cook 2007, Gaspare and Bryceson 2013). One sample from a gonad of each fish was therefore sufficient to represent the fish in histological slide preparation and characterization of reproductive characteristics. A second sample from another location within the gonad was taken from any gonad with both female and male tissue present in the initial sample (Sadovy and Shapiro 1987, Alonso-Fernández et al. 2011).

Histological slides were prepared from 2-mm gonadal cross-sections processed in a Leica ASP6025 tissues processor and embedded in paraffin using a Leica EG 1150H embedding station. Embedded tissues were cut to 4 μm with a Leica RM2125 RTS microtome and stained and counterstained with hematoxylin and eosin by a Leica ST5020 slide stainer. Slides were
assessed by microscopic examination and development classified according to terminology given by Brown-Peterson et al. (2011).

**Regression Analysis**

Logistic regression was used to assess sex as a function of length. A logistic model was a natural fit given the binary nature of sex. A multi-level Bayesian model was used to include both the individual level as well as species-level information. The nested structure allowed the fitting of species with relatively few samples for which estimates may have otherwise been impossible. At least one individual of each sex was collected for six species—*C. fulva, D. inermis, E. adscensionis, E. guttatus, H. flavolimbatus, M. phenax*—which were included in the models.

The response variable *sex* was Bernoulli distributed, where $y_{ij} = 0$ if individual $i$ of species $j$ was female, and $y_{ij} = 1$ if individual $i$ of species $j$ was male otherwise ($y_{ij} \sim Bernoulli(\pi_{ij})$). The form of the model was as follows:

$$
\text{logit}(\pi_{ij}) = \alpha_j + \beta_j x_i, \text{ for } i = 1, \ldots, n
$$

$$
\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2), \text{ for } j = 1, \ldots, J
$$

$$
\beta_j \sim N(\mu_\beta, \sigma_\beta^2), \text{ for } j = 1, \ldots, J
$$

where $\alpha_j$ is the species-specific intercept, $\beta_j$ is the species-specific slope, and $x_i$ is the total length of each fish. Prior distributions were all diffuse and followed conventional distributions (Table 1.2). The length at 50% sexual transition was determined by the relationship

$$
L_{50} = -\alpha_j / \beta_j,
$$

which gives the inflection point of the logistic curve. This relationship is comprised of the species-specific intercept and slope, resulting in an estimate of length at sexual transition for each of the six species in the model.
Table 1.2. Prior distributions used for the Bayesian logistic regression model.

<table>
<thead>
<tr>
<th>PARAMETER</th>
<th>PRIOR DISTRIBUTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species level</td>
<td></td>
</tr>
<tr>
<td>$\alpha_j$</td>
<td>Normal ($\mu_\alpha, \sigma_\alpha^2$)</td>
</tr>
<tr>
<td>$\beta_j$</td>
<td>Normal ($\mu_\beta, \sigma_\beta^2$)</td>
</tr>
<tr>
<td>Global level: hyperpriors</td>
<td></td>
</tr>
<tr>
<td>$\mu_\alpha$</td>
<td>Normal (0, 10000)</td>
</tr>
<tr>
<td>$\sigma_\alpha$</td>
<td>Uniform (0, 10)</td>
</tr>
<tr>
<td>$\mu_\beta$</td>
<td>Normal (0, 10000)</td>
</tr>
<tr>
<td>$\sigma_\beta$</td>
<td>Uniform (0, 10)</td>
</tr>
</tbody>
</table>

The Bayesian model was fit using JAGS (Plummer 2003) executed within R version 3.4.3 (R Core Team 2017) using the R2jags package (Su and Yajima 2015). Three parallel chains of 100,000 were run, each with the first 10,000 samples discarded and every third sample retained for a total of 90,000 samples. The scale reduction factor ($\hat{R}$) for each parameter was examined for convergence and found to be near one ($\hat{R} \leq 1.002$), indicating good convergence and low variation between chains (Gelman 1996).

The Bayesian model was compared to a similar generalized linear model using frequentist estimation. Standard error of frequentist estimates of length at transition were calculated using the MASS package (Venables and Ripley 2002). Data for all species were then analyzed in a means parameterized logistic regression with species as a factor. Length at maturity was not modeled due to insufficient numbers of immature individuals.

**RESULTS**

**Length Distributions**

Lengths varied by species, but range of lengths for males was generally larger than females for each species, consistent with protogyny (Figure 1.1). Lengths ranged from 333 mm TL of a female *C. fulva* to a 980 mm TL for a male *H. flavolimbatus* (Table 1.3). Most individuals were within the expected size limits for each species, the exception being two males
C. fulva caught at a total length of 433 and 491 mm, which is larger than the largest maximum length of 410 mm listed in grouper references (Craig et al. 2011). Species with greater sample sizes highlighted a larger range of potential lengths of individuals; however, expected differences in overall expected size between species was still evident for smaller samples (Figure 1.1).

Figure 1.1. Specimen lengths for Cephalopholis fulva, Dermatolepis inermis, Epinephelus adscensionis, Epinephelus guttatus, Hyporthodus flavolimbatus, and Mycteroperca phenax. Each species is separated by sex, circles on the left being female and triangles on the right being male. Points were jittered horizontally to minimize overlap.

Table 1.3. Sample sizes and total length ranges (mm TL) of grouper specimens by sex.

<table>
<thead>
<tr>
<th>SPECIES</th>
<th>FEMALE</th>
<th></th>
<th>MALE</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TOTAL</td>
<td>SIZE RANGE</td>
<td>TOTAL</td>
<td>SIZE RANGE</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>(mm)</td>
<td>N</td>
<td>(mm)</td>
</tr>
<tr>
<td>Cephalopholis fulva</td>
<td>9</td>
<td>333–395</td>
<td>3</td>
<td>406–491</td>
</tr>
<tr>
<td>Dermatolepis inermis</td>
<td>3</td>
<td>696</td>
<td>2</td>
<td>787–795</td>
</tr>
<tr>
<td>Epinephelus guttatus</td>
<td>4</td>
<td>455–504</td>
<td>2</td>
<td>487–512</td>
</tr>
<tr>
<td>Epinephelus adscensionis</td>
<td>3</td>
<td>365</td>
<td>2</td>
<td>442–465</td>
</tr>
<tr>
<td>Mycteroperca phenax</td>
<td>49</td>
<td>434–700</td>
<td>22</td>
<td>596–793</td>
</tr>
<tr>
<td>Hyporthodus flavolimbatus</td>
<td>15</td>
<td>388–746</td>
<td>3</td>
<td>803–980</td>
</tr>
</tbody>
</table>

Logistic Regressions

The Bayesian model converged with all six species producing estimates of species-specific relationships between length and sex and all $\hat{R} < 1.1$ (Figure 1.2). The mean slope for all
species, $\mu_\beta$, was 0.060 (95% credible interval [CI] = 0.019 – 0.105) and the mean intercept, $\mu_\alpha$, was -32.72 (95% CI = -52.205 – -17.211). Estimates of mean length at sexual transition ($L_{50}$) ranged from 405 mm for *E. adscensionis* to 781 mm for *H. flavolimbatus*. A list of coefficient estimates for each species can be found in Table 1.4.

![Bayesian logistic regression curves](image)

**Figure 1.2.** Bayesian logistic regression curves of the relationship between total length (mm) and sex, where female = 0 and male = 1. Each line represents a species. Species in legend are listed as lines appear left to right. The solid black line represents a population-average of all species.

<table>
<thead>
<tr>
<th>Species</th>
<th>$\alpha_j$</th>
<th>$\beta_j$</th>
<th>$L_{50}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Cephalopholis fulva</em></td>
<td>-30.47 (-52.60, -14.27)</td>
<td>0.075 (0.034, 0.131)</td>
<td>409 (378, 451)</td>
</tr>
<tr>
<td><em>Dermatolepis inermis</em></td>
<td>-35.34 (-57.70, -18.01)</td>
<td>0.049 (0.025, 0.079)</td>
<td>722 (632, 800)</td>
</tr>
<tr>
<td><em>Epinephelus guttatus</em></td>
<td>-31.56 (-53.51, -15.13)</td>
<td>0.064 (0.031, 0.109)</td>
<td>492 (448, 537)</td>
</tr>
<tr>
<td><em>Epinephelus adscensionis</em></td>
<td>-30.20 (-52.51, -13.44)</td>
<td>0.075 (0.034, 0.131)</td>
<td>405 (344, 464)</td>
</tr>
<tr>
<td><em>Mycteroperca phenax</em></td>
<td>-32.63 (-50.34, -18.48)</td>
<td>0.051 (0.029, 0.079)</td>
<td>636 (617, 656)</td>
</tr>
<tr>
<td><em>Hyporthodus flavolimbatus</em></td>
<td>-36.24 (-58.59, -19.15)</td>
<td>0.046 (0.024, 0.076)</td>
<td>781 (726, 847)</td>
</tr>
</tbody>
</table>

Table 1.4. Parameter estimates for Bayesian model (95% credible intervals in parentheses). A separate slope and intercept were estimated for each species. Length at 50% transitioned ($L_{50}$) represent total lengths in millimeters.

The frequentist model failed to converge for some species. Models converged only for species where lengths for female and male observations overlapped: *E. guttatus* and *M. phenax*. 
(Figure 1.3). The other four species modelled had complete separation of zeros and ones, resulting in probabilities numerically zero or one. Estimates of $L_{50}$ ranged from 400.5 mm for *C. fulva* to 774.3 mm for *H. flavolimbatus*. A list of coefficient estimates for each species can be found in Table 1.5. Error for the species that failed to converge was high.

![Frequentist logistic regression curves of the relationship between total length (mm) and sex, where female = 0 and male = 1. Each line represents a species. Species in legend are listed as lines appear left to right.](image)

**Table 1.5.** Parameter estimates for frequentist model. A separate slope ($\beta$) and intercept ($\alpha$) were estimated for each species. Length at 50% transitioned ($L_{50}$) represent total lengths in millimeters. SE is standard error.

<table>
<thead>
<tr>
<th>SPECIES</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$L_{50}$</th>
<th>SE ($L_{50}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Cephalopholis fulva</em></td>
<td>-1303.428</td>
<td>3.255</td>
<td>400.5</td>
<td>1017.8</td>
</tr>
<tr>
<td><em>Dermatolepis inermis</em></td>
<td>-361.161</td>
<td>0.487</td>
<td>742.2</td>
<td>57808.0</td>
</tr>
<tr>
<td><em>Epinephelus guttatus</em></td>
<td>-24.172</td>
<td>0.049</td>
<td>490.7</td>
<td>22.98</td>
</tr>
<tr>
<td><em>Epinephelus adscensionis</em></td>
<td>-229.692</td>
<td>0.568</td>
<td>404.1</td>
<td>43704.5</td>
</tr>
<tr>
<td><em>Mycteroperca phenax</em></td>
<td>-26.235</td>
<td>0.041</td>
<td>636.2</td>
<td>10.95</td>
</tr>
<tr>
<td><em>Hyporthodus flavolimbatus</em></td>
<td>-481.264</td>
<td>0.621</td>
<td>774.3</td>
<td>4851.71</td>
</tr>
</tbody>
</table>

The Bayesian and frequentist estimations of $L_{50}$ were fairly similar. Estimates for *E. guttatus*, *E. adscensionis*, and *M. phenax* were nearly identical between the two models. For these species, there was no shrinkage of estimates due to other species in the model. Estimates
for *C. fulva* and *H. flavolimbatus* were 9 and 7 mm different between models. Marbled grouper, *D. inermis*, had the largest difference between models: 20 mm difference between length at transition estimates. The most notable difference was the lack of convergence and the large error ranges associated with estimates of the frequentist GLM.

**DISCUSSION**

**Species-Specific Results**

The most precise estimation of length at transition was that for *M. phenax* (*n* = 49). Our estimated length at sexual transition of 636 mm TL (95% credible interval (CI) = 617 – 656 mm TL). The previous estimate by Lombardi-Carlson et al. (2012) was 566 mm FL, which is equivalent to approximately 620 mm TL as evidenced by a simple linear regression of fork length (FL) and total length (TL) for all *M. phenax* caught in this study offers. The previous length at transition estimate from Lombardi-Carlson et al. (2012) is at the lower end of the 95% credible interval calculated here. Lombardi-Carlson et al. (2012) sampled a much greater number of fish (*n* = 2,269), but from a slightly different region, the eastern GOM along the West Florida shelf. The slight difference in length at transition estimates could be an artifact of sample size or differences in regional sample collection.

Difference in mean length at transition could be due to regional population differences between the northwestern GOM *M. phenax* individuals analyzed in this study and the northeastern GOM fish studied by Lombardi-Carlson et al. (2012), similar to differences seen in *H. flavolimbatus* between Gulf regions (Cook 2007). Regional differences may be related to differential fishing pressure between the northeastern and northwestern GOM (Coleman et al. 1996). Commercial landings of *M. phenax* in 2016 for the Florida West Coast were 133 metric tons, versus 19.4 metric tons of landings from Texas, Louisiana, and Alabama combined (NOAA
Understanding trends and regional differences in life history parameters of *M. phenax* is critical to a proper evaluation for future stock assessments, especially considering the species’ “Data Deficient” IUCN RedList status (Afonso et al. 2018).

The estimate of length of sexual transition for *H. flavolimbatus* was 781 mm (95% credible interval [CI] = 726 – 847 mm TL). This is a smaller length at transition than the estimate by Cook (2007) of 865 mm from 69 individuals of the species in the western GOM. This difference illustrates the continuation of the long-term decrease in transitional size since the onset of commercial fishing noted by Cook (2007). The previously reported value is outside the 95% credible interval; based on 15 fish, these results estimate that there is a 95% probability that the length at sexual transition is between 726 and 847 mm TL. While catches may not be at historical highs, the commercial landings of *H. flavolimbatus* were still 382 metric tons in 2016, meaning fishing pressure could still be altering population demographics (personal communication from the National Marine Fisheries Service, Fisheries Statistics Division [April 19, 2018]). The indication that these lengths have shifted and potential for continued shift in lengths at sexual transition call attention to the importance of proper monitoring and assessment. This is especially true considering the relatively recent onset of management as of 2013 and the uncertain stock biomass at the time of the most recent stock assessment (SEDAR 2011). The IUCN RedList classifies *H. flavolimbatus* as “Vulnerable” globally and identifies a decrease in the GOM population (Padovani-Ferreira et al. 2018).

*Dermatolepis inermis* was estimated to sexually transition at about 722 mm (95% CI = (632 – 800 mm TL). While there are some caveats to this estimate—low sample size (*n* = 3) and reliance on pooling information from other species—it is the first value proposed for the species. Little is known about *D. inermis* overall, and almost nothing is known regarding its reproduction.
It is a rare species, but has been recorded numerous times at the Flower Garden Banks and nearby banks such as McGrail, suggesting that these shelf-edge banks provide critical habitat (Schmahl and Hickerson 2006, Schmahl et al. 2008, Clark et al. 2014). There is also indication that these banks may be the first known spawning aggregation site for *D. inermis* (Schmahl 2016).

Length at sexual transition for *E. adscensionis* was similar to the previous estimate from the South Atlantic. Nolan et al. (2017) estimated 50% transition at 418 mm TL which is close to our estimate of 405 mm TL and fits in the center of our 95% credible interval of 344 to 464 mm TL. Lengths from both Nolan et al. (2017) and this study are larger lengths for transition than indicated by data from other regions. *E. adscensionis* in the eastern GOM ranged from 252 to 355 mm TL for females and 287 to 375 mm TL for males (Bullock and Smith 1991). In Brazil, transitional size was 260 to 330 mm TL (Marques 2011). This is in contrast to the size ranges Nolan et al. (2017) observed for female and male fish—266 to 480 and 351 to 606 mm TL, respectively—which are similar to the sizes observed in this study (Table 1.3). It has been suggested that the sexual transition of the species varies between regions (Marques 2011). Based upon data in the present study and Nolan et al. (2017), *E. adscensionis* from the northwestern GOM are more similar to those at Ascension Island in the South Atlantic than those in the eastern GOM or Brazil.

*Cephalopholis fulva* and *E. guttatus* each have somewhat higher $L_{50}$ estimates than expected from previous species estimates and the 95% credible intervals do not include the values from previous literature. High estimates for both species may be an artifact of chance samples composed only of larger than average individuals, which drove the estimates given small sample sizes ($n < 10$). Estimates of length at sexual transition may differ from previous
information due to regional population differences. None of the values published in the literature for these two species were from northwestern GOM populations (Table 1.1). The western GOM may be a population with larger individuals and/or a larger length at sexual transition.

Many sources list the maximum length of *C. fulva* or largest observed specimen at 400 mm TL or less (Thompson and Munro 1978, Heemstra and Randall 1993, Figuerola et al. 2001, Trott 2006, Araújo and Martins 2009, Freitas et al. 2011). Trott (2006) observed transitional individuals ranging from 194 mm to 375 mm FL in Bermuda, while Figuerola et al. (2001) observed transitional individuals ranging from 186 to 258 mm FL in Puerto Rico. Given that three of our nine *C. fulva* specimens were above 400 mm TL, and the largest was 491 mm TL—larger than observed in other studies—it is not surprising that the size at transition estimated here would be larger than some previous literature. The larger observed lengths in this study may indicate overall lengths in the populations differ from previous data. Both de Araujo and Martins (2006) and Coelho et al. (2012) observed a small number of individuals greater than 400 mm in length, while Figuerola et al. (2001) indicates that fish upwards of 440 mm FL have been caught previously in the commercial fishery. However, no studies that include these larger individuals involved study of length at transition.

The *E. guttatus* individuals observed in this study were similarly larger than many other populations reported. All collected individuals were at the upper ranges of lengths reported in the Caribbean (Sadovy et al. 1992, Sadovy et al. 1994). All four were larger than any transitional individuals observed in Puerto Rico (Shapiro et al. 1993) and larger than the length at transition reported for the Caribbean (Thompson and Munro 1978) and southern GOM (Caballero-Arango 2013) (Table 1.1). However, the length ranges observed (Table 1.3) fit well inside those reported
by Brulé et al. (2016) in the southern GOM who found females from 205 to 510 mm FL, males from 263 to 575 mm FL, and transitional individuals from 320 to 500 mm FL.

Multiple factors could influence sex change. Fishing pressure has been documented to affect size at sexual transition through “fisheries-induced evolution” (Lowerre-Barbieri et al. 2011, Provost and Jensen 2015). Grouper species declines have been observed even in areas with low fishing pressure (Nolan et al. 2017). In the absence of fishing pressure, predation pressure can have similar effects (Reñones et al. 2010). Grouper populations are also impacted by habitat degradation—especially that of spawning habitats—through activities such as trawling and dredging (Koenig et al. 2000, Afonso et al. 2018). Gulf of Mexico grouper are dependent on shelf-edge habitats for spawning (Marancik et al. 2012). Length at transition is also related to overall size of individuals in a population and the growth rates of individuals, which will be related to factors such as prey availability and abiotic stressors. Grouper size at sexual transition is known to differ through time and space (Cook 2007, Lombardi-Carlson et al. 2012). Some of these factors would be prime for inclusion in an additional layer of the model. A future version of the model may explore the changes in regression parameters with changes in geographic location or through time.

Assessment of Models

Length at sexual transition is typically estimated using logistic regressions, similarly to length at maturity. Parameters are traditionally estimated using maximum likelihood; however, this requires relatively large sample sizes and can result in biased coefficients when sample sizes are low (Peduzzi et al. 1996). Bayesian inference results in more precise estimates of uncertainty, provides more information about the population of interest with posterior distributions, and makes a direct probability statement about parameters of interest (Jiao et al. 2011, Doll and
Lauer 2013). The frequentist model using maximum likelihood estimation in this study resulted in large error surrounding estimates.

Bayesian hierarchical models provide a solution for data-poor situations. These models allow groups with small sample sizes to borrow strength from groups with larger sample sizes and are able to converge and provide coefficient estimates in situations where traditional frequentist methods cannot (Gelman et al. 2004, Doll and Lauer 2013). A comparison of the two types of logistic models highlights the utility of using Bayesian estimation. When the values of zeros and ones do not overlap, the frequentist GLM does not converge, whereas the Bayesian model converged for all species. While the borrowing of information between the species in the Bayesian model can help regressions fit, it can also lead to sharing of information between species when it would not be appropriate. This change of an estimate towards the overall mean is known as “shrinkage.” For instance, the estimate for *D. inermis* showed considerable shrinkage: a difference of 20 mm between the two models.

Bayesian hierarchical models are being used in fisheries science for many objectives: estimations of fish growth (e.g. Froese et al. 2014), population growth (e.g. Jiao et al. 2011), stock-recruitment relationships (e.g. Forrest et al. 2010), stock assessments (e.g. Kuparinen et al. 2012), species occurrence, (e.g. Midway et al. 2014), abundance (e.g. DuFour et al. 2014), mortality (e.g. DuFour et al. 2014), recruitment (e.g. Hansen et al. 2018), and fecundity (e.g. Dick et al. 2017). Bayesian hierarchical models have also been implemented for length at maturity (Punt et al. 2006, Doll and Lauer 2013, Feiner et al. 2015, Rufener et al. 2017). It is a natural progression to extend these methods to modeling length at sexual transition, especially in a situation with data-poor species.
Estimates for these six grouper species are worth reporting, despite potential caveats, to supplement current incomplete knowledge. Five of these species have never been assessed in the northern Gulf of Mexico, and stock status is unknown for *H. flavolimbatus*, the only species consider here which is assessed. This makes the data and estimates valuable for understanding these populations in addition to use as population model inputs. However, there are limits to how such data-limited estimates should be used. The Bayesian models successfully converged for all species to supply information even for species with three or four data points, but population estimates based on so few points could still be very biased by chance observation that may not reflect the overall distribution. While these data are valuable for the expanding of scientific knowledge, they should be applied cautiously in any management setting.

To my knowledge, this is the first application of Bayesian multi-level modeling to estimate length at sexual transition. The Bayesian model successfully estimated parameters and credible intervals for all six species. For most species included, sample size would have been too low to achieve reliable estimates from a frequentist model as can be seen in the high error around estimates. Due to the ability of the model to borrow strength from species with relatively more data to supplement the data-poor species, it was able to construct estimates for species such as *D. inermis*. While there are limits to how much such a data-limited estimate could be utilized, it is a valuable starting point for a species where there is virtually no data available.

The true value of the underlying multilevel Bayesian structure lies in opportunities to expand this model with more data. The Bayesian model presented here could accommodate individuals from different regions to explicitly understand potential regional population differences in estimates of length at sexual transition. Further research should also attempt to expand sample sizes to develop more robust estimates and potentially broaden understanding to
additional species. Acquisition of smaller individuals of each species would make the
determination of length at maturity possible. Otolith aging of these samples could be an
additional tool to enhance our grouper life history understanding, including estimating age at
maturity and sexual transition.

**LITERATURE CITED**


Use of Histological Techniques to Study the Reproductive Biology of the Hermaphroditic

Alonzo, S. H., T. Ish, M. Key, A. D. MacCall, and M. Mangel. 2008. The importance of


Bates, C., and R. Colvin. 2016. Comparison of scamp grouper (Mycteroperca phenax), growth off of the West Florida shelf and the coast of Louisiana. LaGrange College.


Brulé, T., D. Caballero-Arango, X. Renan, and T. Colas-Marrufo. 2016. Confirmation of


Cook, M. 2007. Population dynamics, structure and per-recruit analyses of yellowedge grouper, Epinephelus flavolimbatus, from the northern Gulf of Mexico. The University of Southern Mississippi, Hattiesburg, MS.


for Pacific rockfishes (Sebastes spp.) under alternative assumptions about the stock–
recruit function. Canadian Journal of Fisheries and Aquatic Sciences 67:1611-1634.

patterns of commercially important reef fish (Lutjanidae and Serranidae) in the tropical
western South Atlantic. Scientia Marina 75:135-146.

Froese, R., J. T. Thorson, and R. B. Reyes. 2014. A Bayesian approach for estimating length-

Gaspare, L., and I. Bryceson. 2013. Reproductive Biology and Fishery-Related Characteristics of
the Malabar Grouper (Epinephelus malabaricus) Caught in the Coastal Waters of Mafia

Richardson, and D. J. Spiegelhalter, editors. Markov Chain Monte Carlo in Practice.
Chapman and Hall, Florida.

and Hill, Boca Raton, Florida.

resilient to warming water temperatures in lakes with abundant largemouth bass
populations. Canadian Journal of Fisheries and Aquatic Sciences 75:106-115.

of Marine Science 70:113-132.

world (family Serranidae, subfamily Epinephelinae). An annotated and illustrated
catalogue of the grouper, rockcod, hind, coral grouper and lyretail species known to date.
Food and Agriculture Organization of the United Nations, Rome.


Trott, T. M. 2006. Preliminary analysis of age, growth, and reproduction of coney (Cephalopholis fulva) at Bermuda. Pages 385-399 57th Gulf and Caribbean Fisheries Institute, Saint Petersburg, FL.

CHAPTER 2. ASSESSING OPTIMAL NUMBER OF CAMERAS FOR A MULTI-CAMERA STATIONARY BAITED REMOTE UNDERWATER VIDEO SYSTEM

INTRODUCTION

Baited remote underwater video systems (BRUVS) have proven to be an effective methodology to assess marine fish populations. They are non-extractive, non-destructive, relatively cheap and low-risk, and can be used at a range of depths (Cappo et al. 2003, Murphy and Jenkins 2010). BRUVS simultaneously sample multiple species, detect species that would avoid divers or traps, produce less variable estimates, and can be used on complex or sensitive topographies (Bortone et al. 1989, Cappo et al. 2004, Cappo et al. 2007, Stobart et al. 2007, Bacheler et al. 2013, Whitmarsh et al. 2017).

BRUVS have been used since at least the 1970s and have grown more common over time (Letessier et al. 2013). These systems have been used in most of the world’s oceans and seas and a wide variety of habitat types (Whitmarsh et al. 2017). Applications of BRUVS span evaluations of marine protected areas, behavioral studies, and investigations of habitat relationships. The National Marine Fisheries Service (NMFS) has used an underwater video system to assess fish populations in the Gulf of Mexico (GOM) since 1991, and video data are used increasingly for stock assessments (Gledhill et al. 1996, Campbell et al. 2015). Data collected by BRUVS can be especially important because these ecosystems may be inaccessible to other sampling techniques.

With the rise in usage of BRUVS, many different methodologies have been used. Studies vary in their camera configuration, soak time, and abundance indices. BRUVS may consist of a single camera (e.g. Ellis and Demartini 1995), single stereo pair of two cameras facing forward (e.g. Harvey et al. 2007), cameras facing downward to the seafloor (e.g. Willis and Babcock 2000), or multiple cameras facing different directions (e.g. Gledhill et al. 1996, Kilfoil et al. 2013).
Whitmarsh et al. (2017) found that 60% of BRUVS studies use a single camera, with 36% using stereo-video systems; few studies used more than two cameras. Using pairs of cameras in a stereo configuration on the BRUVS has the advantage of allowing the measurement of fish lengths.

A four-camera system was used by Gledhill (2001) and for the National Marine Fisheries Service (NMFS) Southeast Area Monitoring and Assessment Program (SEAMAP) reef fish video survey beginning in the 1990s. This system consisted of four cameras mounted orthogonally to achieve a near 360° view around the BRUVS. Recent work has further explored full panoramic and full spherical BRUVS (Kilfoil et al. 2017, Campbell et al. 2018). More traditional BRUVS with one or two cameras have a relatively restricted field of view, pointing in a single direction. This limited field of view may result in missing the areas of interest and limit detection, especially when habitats or fish occur in patchy, uneven distributions (Campbell et al. 2018). Traditional BRUVS’s restricted field of view results in saturation at high densities and inaccurate counts whereas systems with additional cameras providing a fully circular or spherical field of view have been shown to increase detection probabilities, increase accuracy of fish counts, and address issues of saturation at high fish density (Campbell et al. 2015, Kilfoil et al. 2017, Campbell et al. 2018). However, additional cameras incur cost both in terms of equipment as well as processing time of increased quantities of video.

The length of video analyzed can vary greatly, with some recommendations of a minimum of three-hour deployments to capture all species (Letessier et al. 2013). While long deployments certainly increase the likelihood of detecting all species in the area, they also dramatically increase the processing time and effort. There is a resulting optimization between number of species seen and amount of video viewed. Video processing time has been
acknowledged as a primary constraint of video analysis (Stobart et al. 2007, Shortis et al. 2013, Mallet and Pelletier 2014, Stobart et al. 2015, Misa et al. 2016, Shafait et al. 2016). Multiple studies have investigated and addressed the trade-offs between soak time, collecting adequate data, and the time to deploy and process (Gledhill et al. 1996, Gladstone et al. 2012, Harasti et al. 2015, Misa et al. 2016). Analysis of these trade-offs can result in adoption of new methodologies. For example, Gledhill (2001) found that approximately 68% of taxa in a 60-minute sample were recorded in a 20-minute sample. Twenty-minute read times were adopted thereafter as the standardized NMFS procedure (Campbell et al. 2015).

There is a similar opportunity to optimize the number of cameras used on a BRUVS, because each camera requires additional viewing time. This study considers the number of cameras viewed on a multi-camera 6-camera BRUVS. The objectives are to investigate the trade-off between using two cameras (single stereo pair facing one direction), four cameras (two stereo pairs facing two opposite directions), or six cameras (two stereo pairs and two single cameras for a four-directional view) and its effect on: 1) taxonomic richness, 2) diversity, and 3) abundance of fish detected. Understanding differences resulting from usage of multiple cameras can also elucidate data comparisons as new studies move to adoption of more multi-camera BRUVS for larger fields of view.

**MATERIALS AND METHODS**

**Field Methods**

The northern GOM contains about 1,800 oil and gas platforms that act as artificial reefs and home to thousands of reef-associated fishes (Stanley and Wilson 2000, BSEE 2018). Five such standing platforms were sampled (Table 2.1). All platforms were located in the Eugene
Island (EI) Oil and Gas Lease Block, approximately 130 km off the coast of Louisiana (Figure 2.1).

Table 2.1. Site demographics of oil and gas platforms sampled in the Eugene Island (EI) Oil and Gas Lease Block. Values obtained from Bureau of Ocean Energy Management (2018).

<table>
<thead>
<tr>
<th>Platform</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Depth (m)</th>
<th>Installed Piles</th>
<th>BRUVS deployments</th>
</tr>
</thead>
<tbody>
<tr>
<td>EI303A</td>
<td>28° 18’ 30.29” N</td>
<td>91° 26’ 37.91” W</td>
<td>69</td>
<td>2005</td>
<td>3</td>
</tr>
<tr>
<td>EI320B</td>
<td>28° 15’19.04” N</td>
<td>91°26' 1.29” W</td>
<td>77</td>
<td>2006</td>
<td>3</td>
</tr>
<tr>
<td>EI325A</td>
<td>28° 14’39.59” N</td>
<td>91°27’ 26.31” W</td>
<td>78</td>
<td>1989</td>
<td>4</td>
</tr>
<tr>
<td>EI342C</td>
<td>28° 11’3.99” N</td>
<td>91°30’ 23.95” W</td>
<td>87</td>
<td>1986</td>
<td>4</td>
</tr>
<tr>
<td>EI346A</td>
<td>28° 9’49.97” N</td>
<td>91°22’ 8.19” W</td>
<td>96</td>
<td>2000</td>
<td>4</td>
</tr>
</tbody>
</table>

Figure 2.1. Study sites in the Eugene Island Oil and Gas Lease Block off the coast of Louisiana: EI 303A, EI 320B, EI 325A, EI 342C, and EI 346A. Depth contours of the main map are at 50-meter intervals. Bolded depth contours on the inset map represent 500-meter contours. Map bathymetry data are from the ETOP01 database hosted by NOAA, accessed via the marmap package in R (Pante and Simon-Bouhet 2013). Coastline data were sourced from the oce and ocedata R packages (Kelley 2015, Kelley and Clark 2018).

Sampling was conducted two to four times per year from August 2013 to November 2015. Video data were collected during daylight hours using a horizontal BRUVS containing six Canon VIXIA HF G10 cameras lit with four 5000-lumen lights. The system consisted of two
stereo-pairs facing opposite directions, each 70 cm apart and angled inward at 7°, and two single cameras mounted orthogonally on the metal cage for a nearly 360° panoramic view (Figure 2.2 C). The system was baited with whole Gulf menhaden (*Brevoorita patronus*) attached to the system with zip ties, as well as ground chub mackerel (*Scomber japonicas*) contained in a metal mesh basket. The BRUVS was deployed suspended on the side of each platform near a piling at three depth strata, each representing approximately one third of the water column. Depth bins were 0 – 30 m, 30 – 60 m, and > 60 m. The cameras recorded at the middle of each stratum for 20 minutes, similar to video viewing procedures for the NMFS video sampling (Gledhill et al. 1996, Campbell et al. 2015, Campbell et al. 2018). The set of six simultaneously collected 20-minute videos (one from each camera on the BRUVS) from each depth stratum of each deployment was considered one sample.

**Video Processing**

Twenty minutes of video from each camera in each depth layer were reviewed using EventMeasure (SeaGIS Pty Ltd) software to identify fish to the lowest taxonomic level possible. MaxN, the maximum number of a given species in any single frame, was determined for each camera or stereo pair for each sample as a conservative measure of abundance (Ellis and Demartini 1995). Stereo pairs were viewed simultaneously as a set, while single cameras were each viewed separately. These data were initially described and analyzed by Reynolds (2015) and Barker (2016). Species of several genera (*Chaetodon, Epinephelus, Holacanthus, Kyphosus, Lipogramma, Pterois*, and *Stegastes*) and two families (Carcharhinidae, Labridae) were grouped, as individual species were difficult to identify.
Figure 2.2. Aerial view of the BRUVS configurations. Each white rectangle is a camera housing holding a camera facing outward from the metal cage structure. Shaded areas show the fields of view of each camera.
Analysis

MaxN counts were condensed across cameras for each sample, such that the largest MaxN from all cameras was used as the single MaxN for that taxon for that sample. Because the six cameras recorded simultaneously but not all six cameras were synced, the greatest MaxN of the cameras rather than sum of MaxN was used to avoid double-counting fish. This was done thrice: once using data from all six cameras and again for four cameras (the two stereo pairs) as well as two cameras (a single stereo pair) (Figure 2.2). Due to the arrangement of stereo cameras on the BRUVS, all three configurations compared in this study allow the measurement of fish lengths.

Taxonomic richness ($S$) and diversity ($H'$) were calculated for each depth layer at each site each time it was surveyed. Taxonomic richness is the total number of taxa detected. Diversity was calculated as the Shannon Diversity index ($H'$) using the vegan package in R (Oksanen et al. 2018) following the equation:

$$H' = - \sum_i p_i \ln p_i$$

where $p_i$ is the proportion of total fish of taxa $i$. Taxonomic richness, diversity, and abundance (MaxN) from the full six-camera BRUVS were compared to the same metrics from only four cameras and two cameras. Differences in metrics were calculated as the 4- or 2- camera metric subtracted from the 6-camera metric. Percentages were calculated as percentages of the 6-camera data. All data manipulation and visualizations were done in R v3.5.0 (R Core Team 2019).

RESULTS

A total of 26 BRUVS deployments were made, each sampled at three depths, resulting in 78 total samples. Five samples detected no taxa on any cameras. A total of 8,963 individual fish, comprising thirty-seven taxa, were observed and enumerated across all surveys and depths,
representing 19 families, 27 genera, and 28 species (Table 2.2). Counting each taxon detected in a sample as a single detection, there were 413 detections across the 78 samples. Jacks (carangids) were the most common, both in terms of estimated abundance as well as frequency of detection. The most commonly detected species in the 78 samples were greater amberjack (*Seriola dumerili, n = 49 samples*), red snapper (*Lutjanus campechanus, n = 47*), and almaco jack (*Seriola rivoliana, n = 41*). Most numerous in terms of cumulative MaxN as well as mean MaxN across samples were blue runner (*Caranx crysos, n = 3,848 fish in 35 samples*) and little tunny (*Euthynnus alletteratus, n = 1,479 in 8 samples*).

Table 2.2. Detection rate and MaxN for observed taxa. Detections refers to the number of samples in which the taxon was observed. Detection rate is the percentage of samples in which a taxon was observed, relative to the 6-camera BRUVS. Mean MaxN is the average MaxN value over all samples in which the taxon was observed. Mean percent MaxN is the percentage of 6-camera MaxN detected by the other configurations averaged over detection per taxon. Each metric is listed for the separate BRUVS configurations.

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Detections</th>
<th>Detection Rate</th>
<th>Mean MaxN</th>
<th>Mean % MaxN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. of Cameras:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abudefduf saxatilis</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Balistes capriscus</td>
<td>10</td>
<td>9</td>
<td>5</td>
<td>0.9</td>
</tr>
<tr>
<td>Canthidermis sufflamen</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>0.88</td>
</tr>
<tr>
<td>Carangoides bartholomae</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>Caranx crysos</td>
<td>35</td>
<td>32</td>
<td>28</td>
<td>0.91</td>
</tr>
<tr>
<td>Caranx hippos</td>
<td>26</td>
<td>20</td>
<td>14</td>
<td>0.77</td>
</tr>
<tr>
<td>Caranx latus</td>
<td>24</td>
<td>21</td>
<td>12</td>
<td>0.88</td>
</tr>
<tr>
<td>Caranx ruber</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>0.6</td>
</tr>
<tr>
<td>Carcharhinidae spp.</td>
<td>27</td>
<td>20</td>
<td>15</td>
<td>0.74</td>
</tr>
<tr>
<td>Chaetodon spp.</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0.5</td>
</tr>
<tr>
<td>Elagatis bipinnulata</td>
<td>16</td>
<td>13</td>
<td>11</td>
<td>0.81</td>
</tr>
<tr>
<td>Epinephelus spp.</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0.75</td>
</tr>
<tr>
<td>Euthynnus alletteratus</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>0.75</td>
</tr>
<tr>
<td>Holacanthus spp.</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>0.75</td>
</tr>
<tr>
<td>Hyporthodus flavolimbatus</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Kyphosus spp.</td>
<td>18</td>
<td>16</td>
<td>11</td>
<td>0.89</td>
</tr>
</tbody>
</table>

(table continued)
### Table: Species Detection Rates and Mean MaxN

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Detections</th>
<th>Detection Rate</th>
<th>Mean MaxN</th>
<th>Mean % MaxN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No. of Cameras:</strong></td>
<td>6 4 2</td>
<td>4 2</td>
<td>6 4 2</td>
<td>4 2</td>
</tr>
<tr>
<td>Labridae spp.</td>
<td>12 7 6</td>
<td>0.58 0.5</td>
<td>3.25 4.43</td>
<td>2 0.58 0.44</td>
</tr>
<tr>
<td>Lipogramma spp.</td>
<td>1 0 0</td>
<td>0 0</td>
<td>9 --</td>
<td>-- 0 0</td>
</tr>
<tr>
<td>Lutjanus campechanus</td>
<td>47 45 41</td>
<td>0.96 0.87</td>
<td>18.85 17.27</td>
<td>14.17 0.81 0.62</td>
</tr>
<tr>
<td>Lutjanus cyanopterus</td>
<td>1 0 0</td>
<td>0 0</td>
<td>1 --</td>
<td>-- 0 0</td>
</tr>
<tr>
<td>Lutjanus griseus</td>
<td>19 16 13</td>
<td>0.84 0.68</td>
<td>3.47 2.44</td>
<td>1.46 0.72</td>
</tr>
<tr>
<td>Lutjanus jocu</td>
<td>1 1 1</td>
<td>1 1</td>
<td>1 1</td>
<td>1 1</td>
</tr>
<tr>
<td>Mycteroperca interstitialis</td>
<td>1 1 0</td>
<td>1 0</td>
<td>1 1</td>
<td>-- 1 0</td>
</tr>
<tr>
<td>Mycteroperca phenax</td>
<td>4 3 1</td>
<td>0.75 0.25</td>
<td>1.25 1.33</td>
<td>1 0.75 0.12</td>
</tr>
<tr>
<td>Paranthias furcifer</td>
<td>4 3 2</td>
<td>0.75 0.5</td>
<td>4.75 2.33</td>
<td>3 0.75 0.5</td>
</tr>
<tr>
<td>Polyipnus laternatus</td>
<td>1 1 0</td>
<td>1 0</td>
<td>1 1</td>
<td>-- 1 0</td>
</tr>
<tr>
<td>Pomacanthus paru</td>
<td>1 0 0</td>
<td>0 0</td>
<td>2 --</td>
<td>-- 0 0</td>
</tr>
<tr>
<td>Pterois spp.</td>
<td>2 0 0</td>
<td>0 0</td>
<td>1 --</td>
<td>-- 0 0</td>
</tr>
<tr>
<td>Rachycentron canadum</td>
<td>3 2 2</td>
<td>0.67 0.67</td>
<td>2.33 2.5</td>
<td>1.5 0.67 0.5</td>
</tr>
<tr>
<td>Remora</td>
<td>2 2 1</td>
<td>1 0.5</td>
<td>2.5 2.5</td>
<td>1 1.12</td>
</tr>
<tr>
<td>Rhomboplites aurorubens</td>
<td>9 6 4</td>
<td>0.67 0.44</td>
<td>9.78 10.17</td>
<td>7 0.59 0.28</td>
</tr>
<tr>
<td>Selene vomer</td>
<td>1 0 0</td>
<td>0 0</td>
<td>13 --</td>
<td>-- 0 0</td>
</tr>
<tr>
<td>Seriola dumerili</td>
<td>49 42 39</td>
<td>0.86 0.8</td>
<td>5.41 5.38</td>
<td>4.26 0.75 0.56</td>
</tr>
<tr>
<td>Seriola rivoliana</td>
<td>41 40 32</td>
<td>0.98 0.78</td>
<td>2.66 2.4</td>
<td>2.56 0.89 0.67</td>
</tr>
<tr>
<td>Sphyraena barracuda</td>
<td>16 14 11</td>
<td>0.88 0.69</td>
<td>1.25 1.14</td>
<td>1.09 0.81 0.62</td>
</tr>
<tr>
<td>Sphyrrna mokarran</td>
<td>3 2 0</td>
<td>0.67 0</td>
<td>1.67 2</td>
<td>-- 0.67 0</td>
</tr>
<tr>
<td>Stegastes spp.</td>
<td>1 0 0</td>
<td>0 0</td>
<td>30 --</td>
<td>-- 0 0</td>
</tr>
</tbody>
</table>

Counting each taxon detected in each depth sample as a single “detection,” the 4-camera system detected 83.3% of 6-camera detections, while the 2-camera system captured 65.1% (Figure 2.3). The 6-camera system produced 413 detections across all deployments, whereas the 4-camera system produced 344 and two cameras produced 269. There were 69 instances where the 4-camera system did not detect a taxon found with six cameras and 144 missed by the two-camera system (Figure 2.4 A, B).
Figure 2.3. Violin plots of percentage detections, richness, and MaxN detected by 4-camera and 2-camera BRUVS relative to six-camera BRUVS configuration. Percentage of detections was calculated per taxon ($N = 37$). Percentage taxonomic richness was calculated for every sample where at least one taxon was detected by the six-camera configuration ($N = 73$), while percentage MaxN was calculated per taxon detection ($N = 413$).
Figure 2.4. Comparison of metrics between BRUVS configurations. Lefthand panels (A, C, E, G) illustrate comparisons between 6-camera and 4-camera BRUVS. Righthand panels (B, D, F, H) illustrate comparisons between 6-camera and 2-camera BRUVS. A-B) Detection frequency (caption continued)
(per taxon, \(n = 37\), C-D) taxonomic richness (per sample, \(n = 78\)), E-F) diversity (per sample, \(n = 78\)), and G-H) MaxN (per detection, \(n = 413\)). Size of points represents frequency of values. The grey line has a slope of 1 for reference. Note: Two MaxN points were excluded from the range with values of 770 and 1,341; MaxN was equivalent between 4- and 6-camera configurations for both these points, while fewer fish were detected on the 2-camera BRUVS.

Some taxa were never detected by the 4- or 2-camera BRUVS configuration. Other taxa experienced up to 100% detection, where the taxon was detected by the 4- or 2-camera BRUVS in every sample it was detected by the 6-camera BRUVS (Table 2.2). Overall, the 4-camera BRUVS had an average detection rate of 68.1% per taxon, while the 2-camera BRUVS averaged 45.3% detection per taxon (Figure 2.3). The taxa that were never detected by the 4- and 2-camera BRUVS were taxa detected by the 6-camera BRUVS in three or fewer of the samples (Table 2.2). Six taxa were detected on the 6-camera BRUVS, but never on either the 4- or 2-camera configurations: *Lipogramma* spp., *Lutjanus cyanopterus*, *Pomacanthus paru*, *Pterois* spp., *Selene vomer*, and *Stegastes* spp. Each of these was detected in only one sample except for *Pterois* spp. which were detected in two samples. Four additional taxa were detected on the 6- and 4-camera configurations but never on the 2-camera BRUVS: *Chaetodon* spp., *Sphyra mokarran*, *Mycteroperca interstitialis*, and *Polyipnus laternatus*. Except the six taxa missed by the 4-camera BRUVS, all other 31 taxa had 50% or greater detection rate across samples compared to the 6-camera BRUVS. The 2-camera BRUVS had detections rates lower than 50% for four of the 31 taxa: *Epinephelus* spp., *Mycteroperca phenax*, *Caranx ruber*, and *Rhombopeltes aurorubens*.

Taxonomic richness, diversity, and MaxN all declined with the reduction in number of cameras on the BRUVS (Table 2.3). Taxonomic richness peaked at 22 for the six-camera BRUVS, 17 for four-camera, and 16 for two-camera (Figure 2.4 C, D). Mean richness across samples was 5.29 for six cameras, 4.41 for four, and 3.45 for two. Difference in taxonomic
richness from the 6-camera BRUVS ranged from zero to five for the 4-camera BRUVS and zero to nine for the 2-camera BRUVS. Averaging across the 73 samples where the six-camera BRUVS detected at least one taxon, the four-camera and two-camera BRUVS detected on average 81.2% and 65.8% of the taxonomic richness detected by the six-camera configuration, respectively (Figure 2.3). Thirty-seven of the 78 (47.4%) samples had no difference in taxonomic richness between 4- and 6-camera configurations, while 26 (33.3%) differed by one taxon. Richness detected by the 2-camera configuration did not differ from the 6-camera BRUVS for 23 (29.5%) samples, while 21 (26.9%) samples differed by one species.

Table 2.3. Metrics for taxonomic richness, diversity, and MaxN observed by the three BRUVS configurations. Metrics for taxonomic richness, diversity, and MaxN observed by the three BRUVS configurations. Metrics for taxonomic richness and diversity are calculated for all samples \( (N = 78) \), while those for MaxN are calculated over taxon detections \( (N = 413) \).

<table>
<thead>
<tr>
<th>Taxonomic Richness</th>
<th>Cameras</th>
<th>Six</th>
<th>Four</th>
<th>Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>5.29</td>
<td>4.41</td>
<td>3.45</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>4.11</td>
<td>3.41</td>
<td>2.85</td>
<td></td>
</tr>
<tr>
<td>max</td>
<td>22</td>
<td>17</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diversity</th>
<th>Cameras</th>
<th>Six</th>
<th>Four</th>
<th>Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>0.83</td>
<td>0.73</td>
<td>0.66</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>0.56</td>
<td>0.54</td>
<td>0.53</td>
<td></td>
</tr>
<tr>
<td>max</td>
<td>2.10</td>
<td>1.87</td>
<td>1.81</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MaxN</th>
<th>Cameras</th>
<th>Six</th>
<th>Four</th>
<th>Two</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>21.70</td>
<td>17.04</td>
<td>8.34</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>84.15</td>
<td>80.27</td>
<td>24.39</td>
<td></td>
</tr>
<tr>
<td>max</td>
<td>1341</td>
<td>1341</td>
<td>267</td>
<td></td>
</tr>
</tbody>
</table>

Diversity \( (H') \) ranged from 0 to 2.10 with a mean of 0.83 for six cameras, 0 to 1.87 with a mean of 0.73 for four cameras, and 0 to 1.81 with a mean of 0.66 for two cameras (Table 2.3, Figure 2.4 E, F). Difference in diversity between 6- and 4-camera BRUVS for each sample
ranged from -0.34 to 0.77 with a mean of 0.10. Difference in diversity between 6- and 2-camera BRUVS for each sample ranged from -1.05 to 1.33 with a mean of 0.16.

MaxN varied among taxa and samples, with the largest MaxN of 1,341 *Euthynnus alletteratus*. The 4- and 2-camera systems accounted for 73.3% and 48.5%, respectively, of MaxN found using six cameras averaged across the 413 species detections (Figure 2.3). Both percent richness and percent MaxN detected by the four- and two- camera BRUVS ranged from zero—where the reduced camera configuration detected none of the taxa detected by the six-camera BRUVS—to one—where all the configurations detected or counted the same taxa. For the 4-camera BRUVS, 249 taxa detections (60.3%) had no difference in MaxN from the 6-camera BRUVS, while the 2-camera BRUVS had no difference for 135 detections (32.7%). Mean difference in MaxN between camera systems was 4.67 fish and 13.4 fish for the 4- and 2-camera BRUVS, respectively. Large differences in MaxN between camera configurations were dominated by carangid species. All instances of a difference in MaxN of 40 or more between 4-camera and 6-camera BRUVS were carangid species: *Caranx crysos*, *Caranx latus*, and *Elagatis bipinnulata*. Most large differences between 2-camera and 6-camera BRUVS were also carangids, but also one instance of missed detection of a very large school of *Euthynnus alletteratus*. The number of detections with zero percent MaxN was much greater for the 2-camera system than the 4-camera system (Figure 2.3).

**DISCUSSION**

The 4-camera and 2-camera systems demonstrated highly correlated taxonomic richness, diversity, and abundance estimates detected using the 6-camera system (Figure 2.4). Because the cameras used for the 2- and 4-camera configurations were stereo pairs, processing time of the pair is roughly equivalent to a single camera due to simultaneous viewing and high degree of
overlap and allows for length measurements. Greater number of cameras and field of view resulted in higher detection probability, similar to previous work (Kilfoil et al. 2017). Data derived from only four cameras captured 83.3% of total detections, an average of 81.2% of taxonomic richness per nonzero sample, and an average of 73.3% of MaxN per detection. This shows a large majority of information was still gained with half the processing effort. The 2-camera system captured 65.1% of detections, an average of 65.8% richness per nonzero sample, and an average of 48.5% MaxN per detection. The 2-camera BRUVS captures about half the information of the 6-camera BRUVS with one quarter of the processing effort.

This suggests that using fewer cameras optimizes data collection and processing if the goal is maximum number of species detected and counted per time spent processing. In fact, from a purely efficiency standpoint, the 2-camera BRUVS resulted in a higher ratio of metrics to processing time spent than either the 4- and 6-camera configurations. However, if research goals focus on capturing all possible species for a full census of the area, the additional data garnered from additional cameras and a more complete field of view may be critical. Overall, average statistics show a more efficient low processing effort sampling by the 2-camera system, but many fish were not detected and species not counted compared to 4- or 6-camera systems. While number of detections and measured richness were noticeably lower in the 2-camera scenario, there was a particularly marked increase in the number of low and zero MaxN counts (Figure 2.4). It is worth noting that for these comparisons, the deployment time in the field was the same, regardless of number of cameras on the system; only processing time was changed by reducing the number of cameras used.

The largest differences in MaxN between BRUVS configurations were for species of jacks and little tunny. This demonstrates that comparisons and trade-offs between BRUVS
methods depend on species of interest. Large discrepancies in MaxN are likely due to aggregating and schooling behavior. Campbell et al. (2018) found that jack data trends different from other species groups, attributing it to schooling behaviors; however, they also show that as camera field of view increases, aggregation has less impact on fish counts. For lower density species, these issues are less of a problem, where large combined field of view has less effect on counts (Kilfoil et al. 2017). The taxa that were never detected on the 4- and 2-camera systems were detected infrequently on the 6-camera BRUVS, so it is hard to know whether those taxa truly have differing detection probabilities between BRUVS configurations or if lack of detection was just an artifact of low sample size and those taxa.

The sites in this study were standing platforms which provide artificial habitat for fish, which differ in some ways from nearby natural habitats. Multiple studies have found higher species richness on natural habitats compared to platforms in the northern GOM as well as different species assemblages (Sonnier et al. 1976, Rooker et al. 1997, Wilson et al. 2003, Wilson et al. 2006, Langland 2015). Langland (2015) found that the fish assemblage at a platform differed significantly from three natural habitat types at the shelf-edge banks. Fish counts at platform sites tend to be dominated by midwater pelagic species (carangids and scombrids), while natural habitats have higher proportions of reef-associated taxa and lower numbers of midwater pelagics (Rooker et al. 1997, Wilson et al. 2003, Wilson et al. 2006). However, Rooker et al. (1997) asserts that although midwater pelagic species overwhelm platform data numerically, these species are transient, and the fish assemblage at platforms is similar to that of adjacent natural communities. It is difficult to say how a change in species composition would affect the percentage of detections between systems. Lower numbers of reef-dependent species were detected at the platform sites, but we cannot know whether this is related
to detectability with the BRUVS or adequate detectability but simply lower abundance at platforms.

Some studies show density of fish to be an order of magnitude higher at platforms than at natural reefs, while others show areas where densities are comparable (Wilson et al. 2003, Wilson et al. 2006). Habitats with lower fish density may have more discrepancy between 6-camera and 4-camera systems because patchy or sparse distributions would be more likely to be captured only by a single camera angle, whereas densely populated fish would be more likely to be seen on multiple cameras.

While MaxN is the most commonly used abundance metric, it has come under scrutiny. MaxN is robust over time and space and has greater precision than the alternative metric, MeanCount (Letessier et al. 2013, Campbell et al. 2015, Whitmarsh et al. 2017); however, it has been shown to be nonlinearly related to true abundance with issues of hyperstability (Schobernd et al. 2014). Stobart et al. (2015) demonstrated that MeanCount suffers from the same issue of saturation at high densities, which is supported by Kilfoil et al. (2017) who did not find MeanCount to reduce hyperstability. Despite a non-linear relationship with abundance, MaxN can accurately represent relative abundances, and average abundances are generally below the saturation point (Campbell et al. 2015, Stobart et al. 2015). We find MaxN to be an appropriate relative abundance measurement. Furthermore, larger field of view can decrease and all but eliminate those hyperstability issues. This has been shown theoretically as well as empirically (Campbell et al. 2015, Kilfoil et al. 2017, Campbell et al. 2018).

Unlike other spherical and 360° BRUVS technologies, the cameras on the BRUVS in this study were not all synced with each other or stitched together for viewing. While this system may not fully be able to address hyperstability of MaxN due to use of separate counts from each
camera set, the multiple cameras provide additional field of view and increased detection probability in a way similar to true spherical BRUVS. The comparison of number of cameras gives an indication of what one- or two-camera BRUVS may be missing in their limited field of view; however, neither the six-camera BRUVS nor even a truly spherical BRUVS will have 100% detection probability (Kilfoil et al. 2017).

Ongoing research into automated detection and identification of fish in video may change the dynamics of trade-offs between detail and processing time moving forward. As automation significantly reduces video processing time, the additional camera information will likely outweigh the cost of processing (Shafait et al. 2016, Shafait et al. 2017, Siddiqui et al. 2018). Until automated methods can be widely utilized, a reduced number of cameras can lessen the necessary processing time.

Overall, the analysis suggests that viewing four cameras produces more than half the information of six cameras with half of the viewing time, while two cameras produce about half the information in one quarter of the viewing time. In some cases, taxonomic richness and MaxN were reduced, but there was a greater amount of information gained per unit processing time which presents a trade-off depending on research priorities. If the goal is a detailed, comprehensive understanding of a community, especially with interest in rare species, maintaining more cameras may be worth the processing time; however, if sampling multiple areas is a research priority, using fewer cameras cuts processing time, potentially allowing for more locations or habitats to be assessed for the same processing effort. Further research is needed to evaluate comparisons in other habitats.

**Literature Cited**

for indexing reef fish presence and abundance in the southeast United States. Fisheries Research 143:81-88.

Barker, V. A. 2016. The effects of artificial light on the community structure and distribution of reef-associated fishes at oil and gas platforms in the northern Gulf of Mexico. Louisiana State University, Baton Rouge, LA.


Langland, T. 2015. Fish assemblage structure, distribution, and trophic ecology at Northwestern Gulf of Mexico banks. Louisiana State University, Baton Rouge, LA.

application of demersal video techniques to the mid-water environment. Methods in Oceanography 8:41-55.


Reynolds, E. M. 2015. Fish biomass and community structure around standing and toppled oil and gas platforms in the northern Gulf of Mexico using hydroacoustics and video surveys. Louisiana State University, Baton Rouge, LA.


Wilson, C. A., M. W. Miller, Y. C. Allen, and K. M. Boswell. 2006. Effects of Depth, Location, and Habitat Type on Relative Abundance and Species Composition of Fishes Associated with Petroleum Platforms and Sonnier Bank in the Northern Gulf of Mexico. OCS Study MMS 2006-037, New Orleans, LA.

CHAPTER 3. RELATIONSHIPS AND PREDICTION BETWEEN FISH ASSEMBLAGE AND HABITAT ON THE NORTHWESTERN GULF OF MEXICO SHELF-EDGE BANKS

INTRODUCTION

The Gulf of Mexico (GOM) is home to a diverse array of fish species. The GOM has been ranked top five globally in species per area, but also top five in threats to biodiversity (Costello et al. 2010). A chain of high-latitude coral reefs exists on a series of bathymetric highs along the edge of the continental shelf, formed by salt diapirism during the Jurassic (Hickerson et al. 2008, Schmahl et al. 2008). These domes are referred to as “banks” (Gardner and Beaudoin 2005). These banks provide most of the 1 – 3% hard bottom occurring on the otherwise soft substrate on the shelf, serving as critical habitat (Parker et al. 1983, Schmahl et al. 2008). Hard substrate combined with relatively stable temperatures provide conditions suitable for the establishment of reef-building and reef-associated species (Dennis and Bright 1988, Rezak et al. 1990, Studivan and Voss 2018). Dozens of species, including economically valuable snappers and groupers, form the fish communities around the hard bottom and vertical relief provided by the banks (Gledhill 2001, Schmahl and Hickerson 2006).

Studies over the past several decades have attempted to resolve divisions in fish assemblages and understand what factors best categorize fish into characteristic groupings. Rezak et al. (1985) first characterized “habitat zones,” home to distinct species assemblages and grouped by vegetation and habitat characteristics. These habitat zones have been described by many studies since. Dennis and Bright (1988) suggest that bank structure, depth, and location determined reef fish communities. Rezak et al. (1990) described the depth-related biotic zonation common on these bank habitats, dependent upon substrate, current regimes, temperature, river influence, and depth. In 2001, Gledhill displayed results that depth, substrate complexity, and
reef area were the most important factors in determining fish assemblage. Schmahl et al. (2008) refined and updated Rezak’s characterization into four habitat zones. Langland (2015) explored the distinctions between these habitat zones. He found that overall habitat characteristics were consistent within and distinct between habitat zones and that fish assemblage was strongly associated with habitat zone. Live cover discriminated most between natural habitats. Artificial habitat of oil platforms and the coral community zone both divided clearly; however, deep coral and coralline algal habitat displayed a less definite clustering. Understanding what drives the assemblage in the coralline algal zone is critical due to the large fraction of area this habitat occupies on the banks (Clark et al. 2014, National Ocean Service 2017). Clearly, more than habitat zone alone is needed to understand fish distributions. Several authors have highlighted lack of sufficient data and identified need for addition study of these mesophotic habitats and their species assemblages (Brooke and Schroeder 2007, Schmahl et al. 2008, Langland 2015).

This study investigates the relationships between fish assemblages and habitat to determine what drives abundance and assemblages using underwater video. The goal is to predict fish species occurrence based on a number of the habitat characteristics that delineate the range of habitat types on the shelf-edge banks. This study extends investigation to a broad range of depths from bank crests to base, as well as rarely studied banks. Successful prediction of fish species presence based on habitat will aid in management, expanding our understanding of species’ habitat use and distribution. The area is being considered for an over 800 km² expansion of the Flower Garden Banks National Marine Sanctuary (Office of National Marine Sanctuaries 2016). Understanding which species utilize habitats on these banks and what factors drive assemblages will directly inform the utility of sanctuary expansion and the habitats most in need
of additional protection. Species, such as those of grouper and snapper, are of particular interest due to their economic importance and management status. McGrail and Geyer banks in particular have been noted to be important habitat for marbled grouper (*Dermatolepis inermis*), a rare species, and their juveniles (Schmahl and Hickerson 2006).

**METHODS**

**Study Area**

Sampling was conducted at six shelf-edge banks 90 to 130 kilometers off the coast of Louisiana: Rankin, Bright, McGrail, Sidner, Parker, and Jakkula (Figure 3.1). These banks formed as the result of salt diapirism, which created a variety of shapes and depths (Rezak et al. 1990). The banks in this study range from approximately 3 to 20 km² in area and from 30 m at the highest crest down to 185 m in depth. Five of the banks are in regions declared Habitat Areas of Particular Concern (HAPC) under NOAA Fisheries Essential Fisheries Habitat legislation which identifies these areas as of special interest for species assessments (Fisheries Leadership & Sustainability Forum 2016). These banks are also designated as No Activity Zones for the oil and gas industry by the Bureau of Ocean Energy Management (BOEM 2014). All six banks are being
considered for an expansion of the Flower Garden Banks National Marine Sanctuary, with five banks in the preferred expansion plan (Office of National Marine Sanctuaries 2016).

Field Methods

The six banks were sampled in six surveys from July 2015 to June 2016. Two to four banks were sampled each survey. Each bank was sampled in three surveys, each in a different month of the year. Video data were collected during daylight hours using a horizontal BRUVS containing four Canon VIXIA HF G10 cameras lit with four 5000-lumen lights. The system consisted of two stereo-pairs facing opposite directions, each 70 cm apart and angled inward at 7° at a height of 0.5 meters off the seafloor (Figure 3.2). The BRUVS was similar to that used by the SEAMAP offshore reef survey as well as Langland (2015), but with the removal of the two single cameras to maximize area surveyed relative to video processing time (Gledhill 2001). The system was baited with whole Gulf menhaden (Brevoortia patronus) attached to the system with zip ties, as well as ground chub mackerel (Scomber japonicas) contained in a metal mesh basket. The BRUVS was deployed to rest on the seafloor with a rope-attached buoy. Cameras were deployed during daylight hours—defined as 30 minutes after sunrise and 30 minutes before sunset—for safety and to reduce diel variation in fish assemblages.

Four sites were chosen at each bank in each survey at varying depths in the attempt to capture the breadth of depth-related habitat zonation. The depth range of each bank was divided, and one site chosen at a random depth within each interval. Deployment sites at multiple depths as well as dual-facing cameras were used in an attempt to account for spatially heterogenous fish distributions.
Figure 3.2. Aerial view of the BRUVS configurations. Each white rectangle is a camera housing holding a camera facing outward from the metal cage structure. Shaded areas show the fields of view of each camera.

**Video Processing**

Videos were reviewed using EventMeasure (SeaGIS Pty Ltd) software for a duration of 20 minutes after settling to optimize viewing effort (Gledhill et al. 1996, Gledhill 2001, Campbell et al. 2015, Campbell et al. 2018). At least the first five minutes after the BRUVS landed on the seafloor were not examined to allow time for sediment to settle and fish to adjust to the stationary camera system. Fish were identified to the lowest taxonomic level possible.

Stereo pairs were viewed simultaneously as a set. MaxN, the maximum number of individuals of a given species seen at one time, was determined for each stereo pair as a conservative measure of abundance (Priede et al. 1994, Ellis and Demartini 1995). MaxN avoids double-counting fish that may repeatedly enter the field of view during the deployment. Because the two stereo pairs recorded simultaneously but were not synced with each other, the greatest MaxN of the cameras in each deployment rather than sum of cameras was used as the MaxN for the deployment to avoid potential double-counting fish. Species of several genera (*Carcharhinus*, *Halichoeres*, *Lactophrys*, *Opistognathus*, *Ptereleotris*, *Pterois*, *Scomberomorus*, and *Sparisoma*) and two
families (Echeneidae and Holocentridae) were grouped, as individual species were difficult to identify.

**Habitat Characterization**

Video data were used to assess habitat characteristics of each site. View, visibility, relief, and rugosity were recorded using the definition and scale define in Table 3.1. Percentages of hard substratum, live cover, and three substrate types were recorded on a scale of 1 to 5, where each increment represented a 20% increase in cover (Table 3.2). Substrate types were 1) sand/mud/clay, 2) gravel/rubble/shell, and 3) rock. Because sand and rubble were not included in the definition of hard substratum, the measure of rock substrate and hard substrate were equivalent. Live cover was assessed separately from hard substratum and substrate type because it was growing upon one the three substrates. Deployments with a view or visibility scores of zero were excluded from analysis. At each BRUVS deployment site, depth measurements from a BioSonics DT-X echosounder as well as longitude and latitude via GPS were recorded. Distance from the coast (depth of zero) and distance from the shelf edge (approximated by the 200 m isobath) were calculated using the dist2isobath function in the marmap package in R (Pante and Simon-Bouhet 2013).

**Statistical Analysis**

Taxonomic richness ($S$), diversity ($H'$), and evenness ($J'$) were calculated for each deployment. Taxonomic richness is the total number of taxa detected. Diversity was calculated as the Shannon Diversity index ($H'$) using the vegan package in R (Oksanen et al. 2018) following the equation:

$$H' = -\sum_i p_i \ln p_i$$
where \( p_i \) is the proportion of total fish of taxa \( i \) (Shannon 1948). Evenness was calculated as Pielou’s evenness index:

\[
J' = \frac{H'}{H'_{\text{max}}} = \frac{H'}{\log S}
\]

where \( H' \) is the observed value of Shannon diversity and \( H'_{\text{max}} \) is the maximum possible value of Shannon diversity for that site (Pielou 1966).

Table 3.1. Habitat characteristics measured via video.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Scale</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>View</td>
<td>0 to 2</td>
<td>0 Field of view completely obscured by habitat close to the camera, or camera facing up or down 1 Partially restricted view by habitat close to the camera 2 Unobstructed view</td>
<td>Hannah and Blume (2012); Easton et al. (2015)</td>
</tr>
<tr>
<td>Visibility</td>
<td>0 to 2</td>
<td>0 Poor view of surrounding substrate is completely obscured by turbidity or marine snow 1 Medium: view of surrounding substrate is not obscured, but viewing distance is limited 2 Good: view of surrounding substrate is clear to the limit of the lighted area</td>
<td>Easton et al. (2015)</td>
</tr>
<tr>
<td>Relief</td>
<td>1 to 3</td>
<td>1 Flat (sand, flat bedrock, gravel or pebble, hash) 2 Low (cobble, small boulder, bedrock) 3 High (large boulder, vertical wall, crevice)</td>
<td>Easton et al. (2015)</td>
</tr>
<tr>
<td>Rugosity</td>
<td>1 to 5</td>
<td>Visual topographic estimate of the substratum 1 essentially flat substrate (e.g. sand) 2 Light: topographically simple 3 Medium: topographic complexity moderate (e.g. coral reefs) 4 Complex: topographically complex (e.g. coral reefs) 5 Highly complex substrate (e.g. branching coral)</td>
<td>Gratwicke and Speight (2005)</td>
</tr>
<tr>
<td>Hard substratum</td>
<td>1 to 5</td>
<td>Percentage of substratum that was not mud, sand or rubble</td>
<td>Gratwicke and Speight (2005)</td>
</tr>
<tr>
<td>Live cover</td>
<td>1 to 5</td>
<td>Percentage cover of living organisms (e.g. corals, macroalgae, sponges)</td>
<td>Gratwicke and Speight (2005)</td>
</tr>
</tbody>
</table>

Table 3.2. Score scale for percent cover estimates.

<table>
<thead>
<tr>
<th>Score</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent cover</td>
<td>0 – 19%</td>
<td>20 – 39%</td>
<td>40 – 59%</td>
<td>60 – 79%</td>
<td>80 – 100%</td>
</tr>
</tbody>
</table>
Non-metric multidimensional scaling (NMDS) was used to elucidate patterns in habitat characteristics that result in more similar fish assemblages. NMDS was based on a Bray-Curtis similarities matrix of species MaxN to create ordinations of sites such that more similar sites are closer together (Shepard 1962, Kruskal 1964). Two- and three-dimensional ordinations were considered and plotted.

Random forest (Breiman 2001) methods were used to predict species’ presence or absence. Random forests combine bagging and random selection to build a large number of trees then the trees vote to assign the most popular class (Breiman 2001). There are many benefits to using random forests including high classification accuracy and ability to model complex interactions (Cutler et al. 2007).

Detection/non-detection data for the 74 species and 11 habitat characteristics across sites were used to grow three sets of classification random forests using the randomForestSRC package in R (Ishwaran and Kogalur 2019): 1) separate univariate random forests for each species, 2) a multivariate random forest model packaging all species’ forests together using the Multivar formula option, and 3) separate random forests for each species using the imbalanced function. These will be referred to as “univariate”, “multivariate”, and “imbalanced” forests. Forests were grown by trying four habitat variables (m = 4) at each tree split and grown to 5000 trees (ntree = 5000). The imbalanced forest was run using perf.type=“misclass” for error calculation to facilitate comparison with the other two forest methods by calculating error in the same manner. Variable importance and minimal depth were calculated within the randomForestSRC package for each species-habitat variable combination for each random forest.

Error rates and other model evaluation measures were calculated for each random forest methodology for each species as specified in Table 3.3. All error rates and predicted values used
were out-of-bag, meaning they were calculated from data not used to train each tree. Out-of-bag estimation has been shown to be just as accurate as using a test set as large as the training set, therefore avoiding the need to separate data into training and test sets (Breiman 1996). Receiver operating characteristic (ROC) and Precision-recall (PR) curves were plotted and area under the curve (AUC) calculated. ROC curves plot true positive rate as a function of false positive rate. PR curves plot precision as a function of recall. A combination of these metrics was used to assess model performance.

Table 3.3. Binary classification evaluation measures (partially adapted from Saito and Rehmsmeier [2015]). TP = true positive; TN = true negative; FN = false negative; FP = false positive.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>( \frac{(TP + TN)}{(TP + TN + FN + FP)} )</td>
</tr>
<tr>
<td>Error</td>
<td>( \frac{(FP + FN)}{(TP + TN + FN + FP)} )</td>
</tr>
<tr>
<td>Class 0 error, False positive rate</td>
<td>( \frac{FP}{(FP + TN)} )</td>
</tr>
<tr>
<td>Class 1 error</td>
<td>( \frac{FN}{(TP + FN)} )</td>
</tr>
<tr>
<td>Sensitivity, Recall, True positive rate</td>
<td>( \frac{TP}{(TP + FN)} )</td>
</tr>
<tr>
<td>Specificity</td>
<td>( \frac{TN}{(TN + FP)} )</td>
</tr>
<tr>
<td>Precision, Positive predictive value</td>
<td>( \frac{TP}{(TP + FP)} )</td>
</tr>
</tbody>
</table>

To be deemed “acceptable,” models must have class accuracy rates of at least 50% and positive values of precision and sensitivity. ROC AUC values ranging from 0.7 to 0.8 are considered “acceptable,” 0.8 to 0.9 are “excellent,” and 0.9 to 1 are considered “outstanding” in terms of model performance (Hosmer and Lemeshow 2000). Because the baseline of PR AUC shifts with the class distribution of the data, model performance bins are less consistently established (Davis and Goadrich 2006, Saito and Rehmsmeier 2015). PR AUC values were assessed as a difference between AUC and the expected AUC from a random model. Using ROC AUC as a guide (where the baseline is 0.5), the “acceptable” category begins 0.2 above the baseline. All PR AUC values greater than 0.2 above the baseline established for the species by its
class distribution will be regarded as “acceptable” or better model performance. Values from 0.3 – 0.4 are “excellent”, and 0.4 – 0.5 are “outstanding.”

All data manipulation and visualizations were done in R v3.5.3 (R Core Team 2019).

RESULTS

From July 2015 to June 2016 six shelf-edge banks were sampled in 68 BRUVS deployments across six months. Each bank was surveyed in three separate months, resulting in ten to twelve total successful deployments per bank. Habitat characteristics varied across sites (Table 3.4). Depths ranged from 50 to 157 meters. Samples spanned the possible range of relief, sand/mud, gravel/rubble, and live cover. Cover of rock substrate was observed from 1 to 4 but did not reach 5. Similarly, rugosity ratings were observed from 1 to 3, but neither 4 nor 5, due to lack of highly rugose corals found only in shallower habitat.

Table 3.4. Habitat characteristics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth (m)</td>
<td>50</td>
<td>157</td>
<td>81.96</td>
<td>22.15</td>
</tr>
<tr>
<td>Latitude</td>
<td>27.86862</td>
<td>28.00458</td>
<td>27.93942</td>
<td>0.03518</td>
</tr>
<tr>
<td>Longitude</td>
<td>-93.45252</td>
<td>-91.64773</td>
<td>-92.56748</td>
<td>0.65389</td>
</tr>
<tr>
<td>Relief (1 – 3)</td>
<td>1</td>
<td>3</td>
<td>1.46</td>
<td>0.63</td>
</tr>
<tr>
<td>Rugosity (1 – 5)</td>
<td>1</td>
<td>3</td>
<td>2.24</td>
<td>0.85</td>
</tr>
<tr>
<td>Sand/mud (1 – 5)</td>
<td>1</td>
<td>5</td>
<td>2.76</td>
<td>1.65</td>
</tr>
<tr>
<td>Gravel/rubble (1 – 5)</td>
<td>1</td>
<td>5</td>
<td>2.34</td>
<td>1.31</td>
</tr>
<tr>
<td>Rock (1 – 5)</td>
<td>1</td>
<td>4</td>
<td>1.51</td>
<td>0.82</td>
</tr>
<tr>
<td>Live cover (1 – 5)</td>
<td>1</td>
<td>5</td>
<td>2.49</td>
<td>1.34</td>
</tr>
<tr>
<td>Distance to coast (km)</td>
<td>142.06</td>
<td>203.60</td>
<td>176.37</td>
<td>20.76</td>
</tr>
<tr>
<td>Distance to shelf (km)</td>
<td>1.47</td>
<td>13.79</td>
<td>6.64</td>
<td>3.34</td>
</tr>
</tbody>
</table>

Seventy-four taxa were identified, representing 64 species across 52 genera and 30 families (Table 3.5). A total of 6,831 individual fish were counted across deployments. Twenty-two taxa were detected each in only one of the deployments. Only one deployment had no fish detections on any camera. Fifteen detections could not be made to a low enough taxonomic level to be included in this analysis. The largest MaxN count of a single species in a deployment was
654 for a school of *Choranthias tenuis*. The most abundant species across deployments were *Choranthias tenuis* (29.4% of total MaxN across deployments), followed by *Schultzea beta* (20.7%), and *Paranthias furcifer* (8.3%). The most commonly detected species, measured by number of deployments in which they were detected, were *Seriola rivoliana* (31 deployments), followed by *Seriola dumerili* (27), *Serranus annularis* (27), *Lutjanus campechanus* (25), and *Malacanthus plumieri* (25). Notable species such as *Dermatolepis inermis* (marbled grouper) were detected as well as juveniles of several species including *Holacanthus tricolor*, *Rhomboplites aurorubens*, and *Pomacanthus arcuatus*.

Table 3.5. Metrics for identified taxa. MaxN metrics include Max (the maximum MaxN for the taxa in a single deployment), Cum (the cumulative MaxN across all deployments), Mean (the mean MaxN across deployments where the taxon was detected), and SD (the standard deviation of MaxN across deployments where the taxon was detected). The Min and Max Depth are the minimum and maximum deployment depths where the taxon was detected. Freq is the detection frequency in total number deployments.

<table>
<thead>
<tr>
<th>Family</th>
<th>Taxa</th>
<th>MaxN</th>
<th>Depth</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Max</td>
<td>Cum</td>
<td>Mean</td>
</tr>
<tr>
<td>Acanthuridae</td>
<td><em>Acanthurus bahianus</em></td>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>Balistidae</td>
<td><em>Balistes capriscus</em></td>
<td>1</td>
<td>6</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td><em>Balistes vetula</em></td>
<td>2</td>
<td>10</td>
<td>1.25</td>
</tr>
<tr>
<td></td>
<td><em>Canthidermis sufflamen</em></td>
<td>5</td>
<td>5</td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td><em>Xanthichthys ringens</em></td>
<td>4</td>
<td>7</td>
<td>2.33</td>
</tr>
<tr>
<td>Carangidae</td>
<td><em>Caranx bartholomae</em></td>
<td>14</td>
<td>21</td>
<td>7.00</td>
</tr>
<tr>
<td></td>
<td><em>Caranx lugubris</em></td>
<td>3</td>
<td>3</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td><em>Seriola dumerili</em></td>
<td>15</td>
<td>53</td>
<td>1.96</td>
</tr>
<tr>
<td></td>
<td><em>Seriola rivoliana</em></td>
<td>112</td>
<td>235</td>
<td>7.58</td>
</tr>
<tr>
<td>Carcharhinidae</td>
<td><em>Carcharhinus spp</em></td>
<td>2</td>
<td>16</td>
<td>1.14</td>
</tr>
<tr>
<td>Chaetodontidae</td>
<td><em>Prognathodon sedentarius</em></td>
<td>7</td>
<td>38</td>
<td>2.38</td>
</tr>
<tr>
<td></td>
<td><em>Prognathodes aculeatus</em></td>
<td>2</td>
<td>8</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td><em>Prognathodes aya</em></td>
<td>1</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td>Congridae</td>
<td><em>Conger oceanicus</em></td>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td>Echeneidae</td>
<td><em>Echeneidae spp</em></td>
<td>1</td>
<td>2</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td><em>Dermatolepis inermis</em></td>
<td>2</td>
<td>3</td>
<td>1.50</td>
</tr>
<tr>
<td></td>
<td><em>Hyporthodus flavolimbatus</em></td>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td><em>Mycteroperca bonaci</em></td>
<td>1</td>
<td>1</td>
<td>1.00</td>
</tr>
</tbody>
</table>

(table continued)
<table>
<thead>
<tr>
<th>Family</th>
<th>Taxa</th>
<th>MaxN</th>
<th>Max</th>
<th>Cum</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mycteroperca interstitialis</td>
<td>2</td>
<td>17</td>
<td>1.31</td>
<td>0.48</td>
<td></td>
<td>58</td>
<td>96</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Mycteroperca microlepis</td>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>NA</td>
<td></td>
<td>62</td>
<td>62</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Mycteroperca phenax</td>
<td>4</td>
<td>29</td>
<td>2.07</td>
<td>1.07</td>
<td></td>
<td>61</td>
<td>96</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Mycteroperca venenosa</td>
<td>1</td>
<td>2</td>
<td>1.00</td>
<td>0.00</td>
<td></td>
<td>61</td>
<td>63</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Paranthias furcifer</td>
<td>99</td>
<td>559</td>
<td>39.93</td>
<td>31.97</td>
<td></td>
<td>55</td>
<td>90</td>
<td>14</td>
</tr>
<tr>
<td>Holocentridae</td>
<td>Holocentridae spp</td>
<td>3</td>
<td>9</td>
<td>2.25</td>
<td>0.96</td>
<td></td>
<td>57</td>
<td>90</td>
<td>4</td>
</tr>
<tr>
<td>Labridae</td>
<td>Bodianus pulchellus</td>
<td>5</td>
<td>13</td>
<td>2.60</td>
<td>1.52</td>
<td></td>
<td>57</td>
<td>86</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Halichoeres spp</td>
<td>3</td>
<td>28</td>
<td>1.65</td>
<td>0.79</td>
<td></td>
<td>55</td>
<td>93</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Apsilus dentatus</td>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>NA</td>
<td></td>
<td>89</td>
<td>89</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Lutjanus buccanella</td>
<td>2</td>
<td>8</td>
<td>1.33</td>
<td>0.52</td>
<td></td>
<td>62</td>
<td>90</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>Lutjanus campechanus</td>
<td>15</td>
<td>61</td>
<td>2.44</td>
<td>2.97</td>
<td></td>
<td>56</td>
<td>118</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Lutjanus jocu</td>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>NA</td>
<td></td>
<td>63</td>
<td>63</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Pristipomoides aquilonaris</td>
<td>14</td>
<td>42</td>
<td>3.23</td>
<td>3.77</td>
<td></td>
<td>93</td>
<td>157</td>
<td>13</td>
</tr>
<tr>
<td>Lutjanidae</td>
<td>Rhomboplites auro rubens</td>
<td>140</td>
<td>432</td>
<td>48.00</td>
<td>48.95</td>
<td></td>
<td>62</td>
<td>110</td>
<td>9</td>
</tr>
<tr>
<td>Malacanthida</td>
<td>Caulolatilus cyanops</td>
<td>3</td>
<td>3</td>
<td>3.00</td>
<td>NA</td>
<td></td>
<td>95</td>
<td>95</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Malacanthus plumieri</td>
<td>4</td>
<td>49</td>
<td>1.96</td>
<td>0.93</td>
<td></td>
<td>53</td>
<td>87</td>
<td>25</td>
</tr>
<tr>
<td>Monacanthida</td>
<td>Aluterus scriptus</td>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>NA</td>
<td></td>
<td>87</td>
<td>87</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Cantherhines pullus</td>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>NA</td>
<td></td>
<td>63</td>
<td>63</td>
<td>1</td>
</tr>
<tr>
<td>Muraenidae</td>
<td>Gymnothorax moringa</td>
<td>1</td>
<td>2</td>
<td>1.00</td>
<td>0.00</td>
<td></td>
<td>61</td>
<td>83</td>
<td>2</td>
</tr>
<tr>
<td>Myliobatidae</td>
<td>Aetobatus narinari</td>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>NA</td>
<td></td>
<td>55</td>
<td>55</td>
<td>1</td>
</tr>
<tr>
<td>Opistognathida</td>
<td>Opistognathus spp</td>
<td>37</td>
<td>65</td>
<td>5.00</td>
<td>9.71</td>
<td></td>
<td>55</td>
<td>87</td>
<td>13</td>
</tr>
<tr>
<td>Ostraciida</td>
<td>Acanthostracion quadricornis</td>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>NA</td>
<td></td>
<td>81</td>
<td>81</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Lactophrys spp</td>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>NA</td>
<td></td>
<td>65</td>
<td>65</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Centropyge argi</td>
<td>19</td>
<td>129</td>
<td>5.61</td>
<td>4.18</td>
<td></td>
<td>50</td>
<td>90</td>
<td>23</td>
</tr>
<tr>
<td>Pomacanthida</td>
<td>Holacanthus bermudensis</td>
<td>1</td>
<td>2</td>
<td>1.00</td>
<td>0.00</td>
<td></td>
<td>67</td>
<td>70</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Holacanthus ciliaris</td>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>NA</td>
<td></td>
<td>63</td>
<td>63</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Holacanthus tricolor</td>
<td>2</td>
<td>6</td>
<td>1.20</td>
<td>0.45</td>
<td></td>
<td>55</td>
<td>85</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Pomacanthus arcuatus</td>
<td>2</td>
<td>3</td>
<td>1.50</td>
<td>0.71</td>
<td></td>
<td>56</td>
<td>65</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Pomacanthus paru</td>
<td>2</td>
<td>7</td>
<td>1.75</td>
<td>0.50</td>
<td></td>
<td>58</td>
<td>66</td>
<td>4</td>
</tr>
<tr>
<td>Pomacentridae</td>
<td>Chromis enchrysurus</td>
<td>91</td>
<td>439</td>
<td>27.44</td>
<td>28.60</td>
<td></td>
<td>53</td>
<td>90</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Chromis insolata</td>
<td>19</td>
<td>59</td>
<td>8.43</td>
<td>7.02</td>
<td></td>
<td>53</td>
<td>67</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Stegastes partitus</td>
<td>23</td>
<td>28</td>
<td>9.33</td>
<td>11.85</td>
<td></td>
<td>53</td>
<td>58</td>
<td>3</td>
</tr>
<tr>
<td>Priacanthida</td>
<td>Pristigenys alta</td>
<td>1</td>
<td>1</td>
<td>1.00</td>
<td>NA</td>
<td></td>
<td>95</td>
<td>95</td>
<td>1</td>
</tr>
<tr>
<td>Ptereleotridae</td>
<td>Ptereleotris spp</td>
<td>8</td>
<td>22</td>
<td>5.50</td>
<td>2.38</td>
<td></td>
<td>65</td>
<td>81</td>
<td>4</td>
</tr>
<tr>
<td>Scaridae</td>
<td>Sparisoma spp</td>
<td>23</td>
<td>54</td>
<td>3.60</td>
<td>5.70</td>
<td></td>
<td>53</td>
<td>90</td>
<td>15</td>
</tr>
</tbody>
</table>

(table continued)
Species richness and diversity ranged widely between deployments (Table 3.6). The
highest taxonomic richness for a deployment was 27, whereas another deployment detected no
species. Shannon diversity ranged from 0 to 2.41.

Table 3.6. Biological community metrics.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Richness</td>
<td>0</td>
<td>27</td>
<td>7.60</td>
<td>6.075</td>
</tr>
<tr>
<td>Diversity</td>
<td>0</td>
<td>2.41</td>
<td>1.10</td>
<td>0.571</td>
</tr>
<tr>
<td>Evenness</td>
<td>0</td>
<td>1</td>
<td>0.71</td>
<td>0.256</td>
</tr>
</tbody>
</table>

Nonmetric multidimensional scaling with two dimensions resulted in stress of 0.155,
while three dimensions had stress of 0.122. Fish assemblage data showed clear patterns relative
Figure 3.3. Plot of NMDS1 and NMDS2 for a three-dimensional nonmetric multidimensional scaling. Each point is a single deployment, and its color represents the value of that habitat variable. Blue color points indicate highest values of each habitat variable; red indicates lowest values.
to depth, sand/mud, gravel/rubble, live cover and rugosity (Figure 3.3). All four were along a similar gradient. These variables were all correlated (0.7 or greater correlation) and are all related to a group of coinciding changes with depth through the water column. Two- and three-dimensional scaling had similar results, but three-dimensional NMDS isolated these gradients more cleanly to NMDS1. Relief and percent rock substrate showed patterns comparable to each other, exhibiting the gradient along NMDS1 similar to depth-related factors, but also some change along NMDS2.

**Random Forests**

Performance of the random forests differed by model type. The univariate and multivariate models had similar overall error rates as well as class error rates (Figure 3.4). The imbalanced model had higher overall error and higher class 0 error, but lower class 1 error than the other two models. The univariate and multivariate models showed a trend in error rate relative to number of positive detections a species had: species with lower numbers of positive detections across sites had lower class 0 error and higher class 1 error. Conversely, species with higher positive detections had lower class 1 error and higher class 0 error. The imbalanced forests did not exhibit a trend between error rates and number of positive detections.

Variable importance tracked roughly 1:1 for univariate and multivariate models with a much less clear relationship to the imbalanced model. The imbalanced model had a larger range in values: -0.059 to 0.38 compared to -0.013 to 0.082 of the univariate and multivariate. In contrast, the imbalanced model’s minimal depth, another indicator of variable significance, had a very tight 1:1 linear relationship with the minimal depth of the univariate model showing near identical values for each species-habitat combination. The habitat predictors with the highest variable importance were notably different between forest types, especially those in the
imbalanced model. For instance, latitude had the highest mean variable importance across species in the imbalanced forests, whereas it was ranked 8 in the univariate and 10 in the multivariate forests out of the 11 predictors.

Figure 3.4. Class error rates for the three random forest models. Point size corresponds to the number of overlapping points.

Sensitivity and specificity showed the divergence between the imbalanced model and the other two (Figure 3.5). The univariate and multivariate models were similar to one another but had higher specificity and lower sensitivity than the imbalanced model. The increased sensitivity
of the imbalanced forest was more pronounced the smaller the number of positive detections for a species; species with more numerous detections (and therefore more balanced data) had more similar sensitivity values between forest models. There were 30 species for which sensitivity was zero for all model types. These 30 species had one or two detections each, with the exception of one that was detected five times. Precision had a similar range of values for all models and comparisons were mixed depending on species. Comparisons of precision were limited by the inability to calculate precision for many species in the univariate and multivariate models. Calculation of precision depends on presence of at least one predicted 1. Random forests predicted only zeros for 37 species in the univariate and 40 species in the multivariate models making it impossible to calculate precision.

Figure 3.5. Classification metric for three random forest types: imbalanced (red), multivariate (green), and univariate (blue). Point size corresponds to cumulative frequency of positive detections for the species across sites. Precision values could not be calculated and therefore were not plotted for 37 species for univariate forests and 40 species for multivariate forests. Points were jittered to reduce overlap between points.
Area under the curve (AUC) for ROC curves and PR curves gave an additional indicator of model performance. The PR AUC was very similar between models. Values for ROC AUC were also very similar, although there was a subset of taxa with lower AUC values for the multivariate model indicating poorer performance (Figure 3.6). Distributions of AUC for the multivariate forests were bimodal for ROC. The univariate and imbalanced forests had similar

Figure 3.6. Histograms for AUC values for three random forest models: imbalanced (red), multivariate (green), and univariate (blue).
AUC values centered around 0.72 for ROC and 0.61 for PR. Using criteria in Hosmer and Lemeshow (2000) for ROC AUC values, univariate forests included 7 outstanding, 15 excellent, and 20 acceptable species forests. The imbalanced forests included 8 outstanding, 18 excellent, and 17 acceptable. That is 57% of univariate species forests and 58% of imbalanced species forests had at least acceptable model performance based on ROC AUC. Frequency of detection ranged from 1 to 31 in the 68 samples resulting in baseline values from 0.0147 to 0.4559 for PR AUC from a random model. Fourteen species had class accuracy of 50% or greater as well as acceptable PR AUC values.

There were 30 taxa for which all models performed very poorly. These 30 taxa had class 1 error equal to 100%, sensitivity of zero, and zero or incalculable precision for all forest types. Their PR AUC were very close to and often less than what would be expected from a random model. Frequency of detection for these poorly modelled taxa were mostly one or two, except for one taxon with five positive detections. Thirteen taxa had just as poor models for both univariate and multivariate but slightly better results from the imbalanced method; however, class 1 error even for the imbalanced model was still 50% or greater and PR AUC values indicated performance similar to that of a random model. Detection frequency ranged from 2 to 6 with the exception of Carcharhinus species which were detected 14 times and Canthigaster rostrata which was detected 16 times. Seventeen taxa had univariate and multivariate models with class 1 error upwards of 0.52, but their imbalanced models’ performance was promising, if still unacceptable: PR AUC was between 0.10 and 0.20 above what would be expected from a random model, and overall and class accuracies were 50% or higher.

Models for Choranthias tenuis, Schultzea beta, Prognathodes aculeatus, Mycteroperca phenax, and Paranthias furcifer had acceptable imbalanced model performance, but were
unacceptable when analyzed using univariate or multivariate. The *Calamus* model was similarly unacceptable for univariate and multivariate but had excellent imbalanced performance. *Pronotogrammus martinicensis* had excellent model fit for imbalanced and multivariate, but greater than 50% class 1 error for univariate.

Seven taxa’s models performed well under all random forest scenarios: *Serranus annularis*, *C. sedentarius*, *Centropyge argi*, *Pristipomoides aquilonaris*, *Seriola rivoliana*, *Malacanthus plumieri*, and *Sparisoma* spp. These seven taxa were detected with a frequency of 13 to 31 times, making their data 0.191 to 0.456 proportion positive. Models had PR AUC values 0.285 to 0.510 above what would be expected from a random model. The best performing forest was for *Serranus annularis* with overall accuracy above 85% for all models and class errors above 80%. Overall, using thresholds of 50% class accuracy and PR AUC at least 0.2 above the baseline, univariate forests had acceptable predictions for 7 taxa, multivariate 8, and imbalanced 14.

Partial dependence plots can expound on relationships between predictions and specific predictor variables. For example, *Serranus annularis* are more likely to be found on areas of high live cover, less sand/mud substrate, and predicted presence decreases as depth increases from about 50 to 85 m (Figure 3.7 and Figure 3.8). *Mycteroperca interstitialis*, a data-limited species, tends towards rocky rugose habitat with moderate to high live cover and relief.

**Discussion**

Seventy-four taxa were identified, capturing a large range of species from small reef-associated to large pelagic to benthic fish species. The trade-offs regarding BRUVS suggest these detections are a good overall but imperfect representation of taxa present. Studies have shown BRUVS tend to be better at quantifying larger and more mobile species but are poorer in
Figure 3.7. Partial dependency plots for *Serranus annularis*.

Figure 3.8. Partial dependency plots for *Mycteroperca interstitialis*.
terms of smaller or cryptic species (Watson et al. 2010, Lowry et al. 2012). Gledhill (2001) found that while video cameras did a poor job sampling the cryptic species close to the bottom, cameras detected fish that divers missed while looking the opposite direction. Furthermore, small cryptic fishes tend to exhibit less fear and aversion to BRUVS if given time to acclimate to the stationary system.

Videos showed instances where additional species were identified or additional individuals were counted after the 20-minute viewing time. This occurred several times for grouper species, even including detection of a Warsaw grouper (*Hyporthodus nigritus*). This calls into question the choice of viewing time, especially if slow-moving species like grouper are targets of interest. Fast-moving, active swimmers like jack species typically came in and out of the frame numerous times during the 20-minute time frame. Viewing beyond 20 minutes was not part of the study, so the extent to which this would change presence and abundance data is unknown. Twenty minutes has been shown to be efficient for overall community composition (Gledhill 2011), but likely does not capture nearly the true present abundance and potentially not even the presence of some species. Viewing time affects the taxa recorded and some sources recommend longer time intervals (Bortone et al. 1989, Gledhill 2001, Misa et al. 2016).

Insufficient viewing time could be a contributing factor in detecting fewer species along these banks than several other studies, which found upwards of 90 or 100 species (Dennis and Bright 1988, Weaver et al. 2006); however, these differences could also be an artifact of total cumulative sampling effort. Other studies, such as Schmahl and Hickerson (2006) and Streich et al. (2017) found 78 and 79 species, respectively—very similar to the findings of this study.

These models are limited somewhat in their scope due to the incomplete range of some habitat variables. The shallowest site was 50 m deep, leaving coral reef and coral community
habitat zones un-sampled. This is likely a limiting factor on the taxonomic richness observed in this study; coral reef fish assemblages consist of a distinct collection of species compared to deeper habitat zones (Schmahl et al. 2008, Langland 2015). While true coral reef only exists in this area on East and West Flower Garden Banks, coral community is present on the crests of other banks above 52 m, such as Bright and McGrail (Schmahl and Hickerson 2006, Schmahl et al. 2008). Sites sampled in this study did not cover the full breadth of rugosity (peaked at 3 on a scale of 1 to 5). To use these models to predict species occurrence across the shelf-edge banks, additional sampling would need to be incorporated that covers shallower and more rugose environments.

Investigation of habitat drivers through factor analysis and NMDS was consistent with previous findings. Habitat variable and fish assemblage groupings identified consistent depth gradients and important variables such as rugosity, live cover, substrate type, and relief, consistent with previous literature (Dennis and Bright 1988, Rezak et al. 1990, Langland 2015).

Results from the random forests are promising for the prospect of using random forests to predict fish presence for a number of important species. Data from about a dozen species were sufficient to create predictive models based on the suite of eleven habitat predictors. Such predictions would be useful in understanding and managing fish distribution along the shelf-edge. Predictions based on relatively static habitat variables would allow less frequent and less time-intensive repeated data collection to garner information about fish distributions. Training forests on fish-habitat relationships across multiple banks allows the model to apply to a range across the chain of northwest GOM banks. Understanding fish habitat utilization and presence across the banks is relevant to the proposed Flower Garden Banks National Marine Sanctuary expansion (Office of National Marine Sanctuaries 2016). Random forest predictions of species
presence across habitats can help accurately identify and quantify the fish resources at the proposed expansion sites, providing background knowledge for effective management.

Comparisons between random forest types presented the anticipated trade-offs. The imbalanced model increased sensitivity at the expense of specificity. The imbalanced model specifically accounted for the skewed class distribution of the data rather than assuming a probability breakpoint of 0.5 between the two binary classes. This resulted in reduced class 1 error rates compared to the other models, but increased class 0 error rate. The univariate and multivariate forests had higher overall accuracy; however, especially for species with very few positive detections, the univariate and multivariate forests tended to predict most or entirely zeros. Models which predict solely zeros in order to achieve high accuracy values are not particularly useful in the practice of understanding fish distributions, so it makes sense to balance priority to sensitivity over specificity and recommend the imbalanced model. In fish predictions, risk of a false positive is not dire, so it is reasonable to sacrifice some specificity or greater ability to predict positive outcomes. For species with higher frequency of detections, which result in more balanced data, metrics were closer between models. For these more commonly detected species, an imbalanced model is less critical.

Of the three random forest types, the imbalanced forest tended to fit the data best. Imbalanced forests had acceptable or better performance in about twice as many models as univariate or multivariate forests. However, even the imbalanced model could not predict well for all species included. There were 30 samples that had zero specificity for the imbalanced model. The higher sensitivity of the imbalanced model still cannot predict when there are extremely low positive counts in the data. For some species, detections were simply too few for meaningful patterns to be drawn from the existing data.
Some metrics proved inadequate for assessing model performance. Accuracy alone misrepresented the utility of many of the models due to artifacts of the imbalanced nature of the data; models that predicted absence in all cases for rarely detected taxa were rewarded with high accuracy despite the lack of utility of such models. Davis and Goadrich (2006) and Saito and Rehmsmeier (2015) recommend PR AUC over ROC AUC when handling imbalanced data. ROC AUC values were high for models that performed poorly as assessed by class error rate, sensitivity, precision, and PR AUC, reinforcing that ROC can be a misleading metric for assessing model performance of imbalanced data. PR AUC values suffered much less from this and proved a good consistent indicator of good model performance. The main challenge is how to best account for the shifting baseline for PR curves.

A major factor in determining well-performing forests was frequency of detection. Taxa with more frequent detections, and therefore a more balanced ratio between the positive and negative class, resulted in better predictive models (Figure 3.9). The univariate and multivariate models performed well for taxa with 20 to 46% positive data. Considering the total sample size of this study was 68, it cannot be determined whether this relationship stems more from imbalanced classes or simply an overall lack of data, or some combination therein. Given the improvement of the imbalanced model compared to the univariate and multivariate forests in some cases and the relatively successful predictions of Calamus calamus—a species with only two detections—it seems likely that lack of data per species is the greater issue. One or two detections does not offer much information from which to draw a pattern. The success of C. calamus habitat prediction seems to relate to the anomalously narrow range and variance in habitat predictors. Other taxa had greater variance in observed habitat, leaving detection frequency as the most related factor to model performance.
Taxa with as few as 13 separate site detections produced models with 85% predictive accuracy. Sample size needs to be higher relative to the limited data in this study, but sample sizes can be relatively small compared to thresholds such as those for parametric statistical assumptions (i.e. \( n = 30 \)). *Carcharhinus* spp. performed opposite to the detection frequency trend seen for *C. calamus*: with a moderate sample size of 14 detections, the random forest models performed very poorly. The poor performance may indicate that these species should not be grouped, and this grouping is obscuring trends in habitat use and occurrence.

Several taxa observed in this study have been highlighted as data-limited. Detection and modeling of habitat associations for wenchman (*Pristipomoides aquilonaris*), yellowmouth grouper (*Mycteroperca interstitialis*), and almaco jack (*Seriola rivoliana*) is therefore particularly important to fill in data gaps to inform management (SEDAR 2016). Two of these three species—*S. rivoliana* and *P. aquilonaris*—were among the best performing random forests, 74% and 85% accuracy, respectively, a promising result for management applications. The model for *M. interstitialis* was deemed unacceptable but had the most promising performance in the imbalanced forest: overall accuracy of 66% with 69% accuracy for class 0 and 54% accuracy for class 1. Even without the most accurate model predictions, partial dependency plots could elucidate trends in habitat utilization notes as an area of needed research by SEDAR 49 (Figure 3.8).

Results from the random forests are promising in predicting fish presence based on habitat for some species. Multiple models achieved greater than 80% accuracy. Additional fish detection data would better train the model, especially for species with very few positive detections. The more data that can be included, the less likely artifacts of imbalanced data will dampen model performance and more likely consistent habitat use patterns can be established. A
Figure 3.9. Trade-offs between class error rates and their relationship with detection frequency and PR AUC. Size of points indicated the frequency a species was detected; large indicates more detections. The color of points indicates PR AUC values; darker is higher area under the curve.

Sampling of more complete habitat breadth and depth range is needed to train the model to apply to all habitats present across the shelf-edge banks. Most data in this study were collected in coralline algal and soft bottom sand or gravel habitats. Exploration of coral community and coral reef habitat would include the full range of habitat predictors as well as capture additional species. Ultimately, future research is needed to construct habitat maps similar to that of Clark et al. (2014) to serve as model inputs to predict fish presence across the series of banks.
LITERATURE CITED


Hickerson, E., G. Schmahl, M. Robbart, W. Precht, and C. Caldow. 2008. The state of coral reef ecosystems of the Flower Garden Banks, Stetson Bank, and other banks in the


Langland, T. 2015. Fish assemblage structure, distribution, and trophic ecology at Northwestern Gulf of Mexico banks. Louisiana State University, Baton Rouge, LA.


Saito, T., and M. Rehmsmeier. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 10:e0118432.


SEDAR. 2016. SEDAR 49 Gulf of Mexico Data-limited Species. Southeast Data Assessment and Review, North Charleston, SC.


CHAPTER 4. PREDICTING FISH BACKSCATTER AND DENSITY BASED ON HABITAT ALONG THE NORTHWESTERN GULF OF MEXICO SHELF-EDGE

INTRODUCTION

The northwestern Gulf of Mexico (GOM) banks provide a series of unique habitats along the continental shelf edge. These 130 topographic features of the region provide most of the 1–3% hard bottom occurring on the otherwise soft bottom of the northern GOM shelf (Parker et al. 1983, Schmahl et al. 2008). This hard bottom provides critical habitat that supports corals and associated communities on these banks. The shelf-edge banks support a diverse collection of both reef building and non-reef building coral species including deep corals (Schmahl et al. 2008). A recent census of marine biodiversity ranked the GOM in the top five globally for species per area but also for threats to biodiversity (Costello et al. 2010).

As an acknowledgment of the importance of these habitats, many banks are designated as both Essential Fish Habitat (EFH) and Habitat Areas of Particular Concern (HAPC) as well as No Activity Zones for oil and gas activities. Despite these designations, most sites do not have special regulations or protections as HAPC zones. Furthermore, only the banks within the Flower Gardens Bank National Marine Sanctuary are subject to regular monitoring, leaving many of the other dozens of banks along the shelf edge unprotected, understudied, and unquantified in regards to their fish communities and habitat use (Kraus et al. 2006). The lack of protection could change with the proposed expansion of the sanctuary. Understanding the true biological worth of additional protections of the ecosystems hinges on gathering additional data.

The bulk of the scientific research regarding the northwestern GOM shelf edge did not begin until the 1970s and ‘80s when there was need to understand the ecosystems in the face of potential oil and gas development in the area (Schmahl et al. 2008). Many studies have been...
conducted in the area since, but the vast majority focus on area around standing and toppled oil and gas platforms or within the Flower Garden Banks National Marine Sanctuary.

Many of the studies regarding biomass and density of fishes in the region utilize fisheries hydroacoustics. Hydroacoustics is a useful tool to assess fish distributions and biomass quickly and nondestructively. The technology works well for large areas, at depth, on varying habitat types, and with high spatial resolution that many other methodologies would not be able to achieve and has been utilized repeatedly to explore fish spatial distribution in the GOM (Gledhill et al. 1996, Stanley and Wilson 1996, 1997, 1998, 2000, Wilson et al. 2003, Wilson et al. 2006, Boswell et al. 2010, Simonsen et al. 2014, Langland 2015). Biomass has been shown to be associated with both low and high relief, related to depth, and varied between bathymetrically-defined habitats (Wilson et al. 2003, Wilson et al. 2006). Aspects of structural complexity, bathymetry, and other abiotic factors may be important in determining fish distribution, abundance, and biomass of fish assemblages in the northwestern GOM across the varying substrates, depths, and habitats (Dennis and Bright 1988, Gledhill 2001).

Many aspects of physical habitat can be derived from multibeam acoustic data that have already been collected (Gardner and Beaudoin 2005). Variables describing structural complexity, substrate, slopes, and many more variables can be extracted from this multibeam data (Wilson et al. 2007). Determining relationships between these multibeam habitat variables and fish biomass and density would lessen the need for detailed and frequent habitat monitoring to understand fish distribution across habitats.

This study aims to create accurate predictive models to estimate fish backscatter and density values across habitat gradients, including across different shelf-edge banks. Predictive models based on habitat would be useful in relating fish distribution to the abiotic environment
and assessing which habitats are most important for fish. Predictive models are particularly helpful for decision-making in management, including marine protected areas (Miller et al. 2004). These models can provide both a deeper understanding of which habitat variables and what spatial scales are most important for predicting fish biomass and density as well as reliable predictions of fish at new sites where split-beam hydroacoustic data have not been collected.

METHODS

Field Methods

Active hydroacoustic sampling took place across five banks along the GOM shelf edge: Rankin, Bright, McGrail, Sidner, and Jakkula. Sites were sampled from July 2015 to June 2016 in six surveys, with up to four banks visited in each survey. During each survey, four sites at each bank were chosen to capture varying depths across the bank. Sampling was conducted during daylight hours at an average vessel speed of five knots.

The hydroacoustics data collection system included three downward-facing split beam transducers (at frequencies of 70, 123, and 206 kHz) mounted 2 m below the water surface, a Biosonics DT-X scientific echosounder, and personal computer. Transducers were calibrated using the standard sphere method (Foote et al. 1987). The survey pattern consisted of a line every 30º around a center point. Each line extended 100 m from the video array, resulting in six 200 m intersecting transects (Figure 4.1). This resulted in a circular survey area of approximately 31,416 m² for each site. Acoustic data were recorded using Visual Acquisition 6.0 (Biosonics, Inc.). Data were collected at a ping rate of 0.4 ms. Time-synchronized GPS coordinates were recorded along with acoustics data. Immediately following completion of transects, a SeaBird SBE 25 Sealogger CTD or YSI sonde was deployed to measure conductivity, temperature, depth,
and salinity throughout the water column. In the event of equipment failure, abiotic data from the nearest site were used.

Figure 4.1. Schematic of acoustics sampling cruise track. Solid lines represent the six 200-meter intersecting transects; dashed lines represent turns where active acoustic data were not collected.

Hydroacoustic data were collected simultaneously with corresponding underwater video, such that the center of the hydroacoustic cruise track corresponds to the location of the underwater video array. Video data were described and analyzed in Chapter 3.

Data Processing

Acoustic backscatter data from the 123 and 70 kHz transducers were post-processed using Echoview software (Myriax Pty Ltd, Hobart, Tasmania, Australia) (Figure 4.2). Harmonic means of temperature and salinity were used to correct for sound speed and absorption. Backscatter was processed to get mean volume backscatter strength (MVBS, $S_v$, dB) and measure fish density. Mean volume backscatter was used as a proxy for biomass. Temperature
Figure 4.2. Dataflow diagram representing acoustic data processing in Echoview. Sv: Volume backscatter; TS: target strength; MVBS: mean volume backscattering strength.
and salinity data from CTD casts were used to correct for changes in the speed of sound and absorption underwater. Bad data, such as signal loss, bubbles, interference, or excess noise, were excluded from the echogram. The “best bottom candidate” detection algorithm was used to exclude data within 0.5 m of the seafloor and structure with manual edits as necessary. The 10 m nearest the surface was excluded to avoid surface noise and near field effects. Background noise was removed using the methods of De Robertis and Higginbottom (2007), and intermittent noise was removed using the “impulsive noise (IN)” and “transient noise (TN)” algorithms described in Ryan et al. (2015).

Decibel differencing was employed to isolate fish backscatter from other unwanted scattering, such as zooplankton. Different classes of organisms have unique patterns of backscatter across different frequencies. At the range of frequencies typical of fisheries acoustics, swim-bladdered fish tend to result in geometric scattering, in which backscatter is relatively stable across frequency (Kang 2002, Korneliussen 2002, Benoit-Bird and Lawson 2016). At these same frequencies, many zooplankton are Rayleigh scatterers, such that backscatter increases with greater frequency. Difference between backscatter strength at different frequencies can describe these different trends and characterize scatterers.

Ping times were match between the 70 and 123 kHz transducers for synced data. Backscatter from both frequencies with noise removed was smoothed using a three sample by three-ping median in XxY Statistic in Echoview. The difference between 123 kHz and 70 kHz was produced by a Minus operator (Figure 4.2). Difference values were used to define fish backscatter: values from -15 to 1 were classified as fish, 3 to 25 were classified as non-fish. Masks of these values produced fish and non-fish echograms of 70 kHz backscatter. The fish echogram was gridded into 10 m depth and 20 m distance cells and used for analysis.
Mean volume backscattering strength was obtained by integrating the volume backscattering coefficient \( s_v \) over each cell, according to the relationship:

\[
S_v = 10 \log_{10}(s_v)
\]

and exported for analysis (MacLennan et al. 2002). Fish density per cubic meter was calculated by scaling MVBS values by mean single-target strengths following MacLennan et al. (2002):

\[
Density_{cell} = 10^{(MVBS/10)} / 10^{(TS/10)}
\]

for each cell. Density and MVBS values were analyzed by cell.

### Habitat Predictors

A series of habitat variables were measured as potential predictors of fish acoustics backscatter (Table 4.1). Bank, date, time, depth, and GPS location were recorded during active acoustic data collection. Date was converted to an integer number of days in the year using the yday function in lubridate, and time was converted to number of minutes of the day using the times function in the chron package (Grolemund and Wickham 2011, James and Hornik 2018).

Mean depth of each cell and depth offset (the difference between mean cell depth and mean bottom depth below the cell) were derived from Echoview output for each cell. Bottom characteristics of mean depth, maximum depth, minimum depth, and relief (the difference between maximum and minimum depth) were calculated for each 20-m distance interval and used for all cells above that bottom interval.

GPS coordinates representing the mean of 20-m distance cells were used for latitude and longitude as well as determining distances. Distance to the coast and shelf were calculated using the dist2isobath function in the marmap package with NOAA bathymetry data extracted with the getNOAA.bathy function (Pante and Simon-Bouhet 2013). The shelf edge was roughly approximated here by the 200-m isobath.
Table 4.1. Habitat variable predictors included in boosted regression models.

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Predictor</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study design</td>
<td>Mean cell depth</td>
<td>cell</td>
</tr>
<tr>
<td></td>
<td>Depth offset</td>
<td></td>
</tr>
<tr>
<td>Split-beam</td>
<td>Mean bottom depth</td>
<td>20-m distance interval below cell</td>
</tr>
<tr>
<td></td>
<td>Min bottom depth</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max bottom depth</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Relief</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Latitude</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Longitude</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distance from shelf edge</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Distance to shore</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Date (day)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time (minutes)</td>
<td></td>
</tr>
<tr>
<td>Multibeam: backscatter</td>
<td>Acoustic backscatter: mean</td>
<td>radii of 10, 20, 50, 100 m</td>
</tr>
<tr>
<td></td>
<td>Acoustic backscatter: variance</td>
<td></td>
</tr>
<tr>
<td>Multibeam: bathymetry</td>
<td>Bathymetry: mean</td>
<td>radii of 10, 20, 50, 100 m</td>
</tr>
<tr>
<td></td>
<td>Bathymetry: variance</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Slope</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aspect: northness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Aspect: eastness</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Topographic Position Index</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Terrain Ruggedness Index</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Roughness</td>
<td></td>
</tr>
</tbody>
</table>

Multibeam bathymetry and acoustic backscatter data from the USGS described in Gardner and Beaudoin (2005) were used (Gardner et al. 2016). Seafloor characteristics were derived from multibeam data using the raster package (Hijmans 2019). Multibeam parameters were extracted at four spatial scales: a circle of radius 10 m from the mean coordinates of the cell, such that the diameter of the circle is the 20 m length of the cell, as well as circles of 25, 50, and 100 m radii to explore larger scales. Areas with these radii were also used as scales for measuring acoustic backscatter, both its mean and variance for each scale.

96
Mean and variance of bathymetry values were calculated at all spatial scales to measure depth and changes in depth at difference scales. Slope and aspect were calculated following Horn (1981) using the terrain function in the raster package (Hijmans 2019). Aspect was then converted to northness and eastness to handle circular nature of the variable. Topographic Position Index (TPI), Terrain Ruggedness Index (TRI), and roughness were computed according to Wilson et al. (2007) using the terrain function.

**Data Analysis**

Boosted regression trees (BRT) were used to describe and predict the relationship between the habitat predictors and two responses: linearized MVBS and fish density. Fish density was split into zero and nonzero categories for modeling due to large quantities of zero values. Areas of zero fish density occur due to the variable and spatially heterogenous nature of fish distributions. A grid search was used to tune the boosted model and choose optimal parameter settings.

MVBS was modeled using a gaussian distribution, learning rate of 0.05, tree complexity of 6, and 4,000 trees. Fish density zero/nonzero data were modeled with a Bernoulli distribution, learning rate of 0.01, tree complexity of 6, and 5,000 trees. Both BRT were grown with a bag fraction of 0.5 and 10-fold cross-validation. Trees were grown using the gbm package in R (Greenwell et al. 2019). Both BRT models were trained using half the available data sets. For each analysis, the total number of observations was randomly split in half, resulting in a training set of 13,238 observations (cells) as well as a complimentary independent set of 13,238 observations for model testing. Each model was tested with the independent data set using the number of trees found to have the best cross-validation iteration.
The MVBS regression model was evaluated with variable importance, partial dependence plots, root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). RMSE is the square root of the squared prediction errors (observed minus predicted values). It is the standard deviation of the residuals. MAE measures the average of the magnitude of the residuals. MAPE is similar to MAE but normalizes the mean absolute residuals by the actual values.

The binary fish density classification model was evaluated with overall and class error rates, metrics of sensitivity, specificity, and precision, as well as calculation of area under the Precision-Recall curve (Table 4.2) (Saito and Rehmsmeier 2015).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>((\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FN} + \text{FP}))</td>
</tr>
<tr>
<td>Error</td>
<td>((\text{FP} + \text{FN}) / (\text{TP} + \text{TN} + \text{FN} + \text{FP}))</td>
</tr>
<tr>
<td>Class 0 error, False positive rate</td>
<td>(\text{FP} / (\text{FP} + \text{TN}))</td>
</tr>
<tr>
<td>Class 1 error</td>
<td>(\text{FN} / (\text{TP} + \text{FN}))</td>
</tr>
<tr>
<td>Sensitivity, Recall, True positive rate</td>
<td>(\text{TP} / (\text{TP} + \text{FN}))</td>
</tr>
<tr>
<td>Specificity</td>
<td>(\text{TN} / (\text{TN} + \text{FP}))</td>
</tr>
<tr>
<td>Precision, Positive predictive value</td>
<td>(\text{TP} / (\text{TP} + \text{FP}))</td>
</tr>
</tbody>
</table>

**RESULTS**

Sampling occurred between 27.87 to 28.00 latitude and -93.45 to -91.65 longitude in bottom water depths of 45 to 166 meters (Table 4.3). Average relief along the bottom below a 20-m cell was just over one meter. Some multibeam habitat variables varied significantly across spatial scales, while other remained consistent (Table 4.4).

Mean variable backscatter (MVBS) had a mean of -74.01, ranging from -111.01 to -40.10 decibels. The MVBS BRT model found all 53 predictors had nonzero influence. Mean cell depth was the most influential predictor (relative influence = 21.86), closely followed by bottom offset.
Day of the year was the third most important predictor while time of day was the fourth. Of all the bottom variables derived from multibeam bathymetry, those at the 100 m radius scale tended to have the greatest influence. Relationship between predicted MVBS and cell depth saw an overall dome-shaped curve, with the highest fish backscatter at around 60 m depth and declining precipitously at around 125 m (Figure 4.4). A partial dependence plot of bottom offset revealed that MVBS was predicted to be much higher close to the bottom.

Table 4.3. Summary statistics for select habitat predictors recorded and derived from split-beam data. Min, Max, and SD represent minimum, maximum, and standard deviation, respectively.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell mean depth (m)</td>
<td>49.89</td>
<td>45.00</td>
<td>14.87</td>
<td>154.11</td>
<td>26.18</td>
</tr>
<tr>
<td>Bottom depth (m)</td>
<td>88.38</td>
<td>86.41</td>
<td>46.06</td>
<td>166.19</td>
<td>26.22</td>
</tr>
<tr>
<td>Bottom depth min (m)</td>
<td>87.86</td>
<td>85.63</td>
<td>45.34</td>
<td>165.97</td>
<td>26.19</td>
</tr>
<tr>
<td>Bottom depth max (m)</td>
<td>88.91</td>
<td>87.15</td>
<td>46.51</td>
<td>166.45</td>
<td>26.27</td>
</tr>
<tr>
<td>Relief (m)</td>
<td>1.05</td>
<td>0.71</td>
<td>0.00</td>
<td>11.12</td>
<td>1.08</td>
</tr>
<tr>
<td>Bottom offset (m)</td>
<td>38.49</td>
<td>34.36</td>
<td>-3.73</td>
<td>151.19</td>
<td>28.34</td>
</tr>
<tr>
<td>Latitude</td>
<td>27.94</td>
<td>27.93</td>
<td>27.87</td>
<td>28.00</td>
<td>0.04</td>
</tr>
<tr>
<td>Longitude</td>
<td>-92.60</td>
<td>-92.59</td>
<td>-93.45</td>
<td>-91.65</td>
<td>0.69</td>
</tr>
<tr>
<td>Distance to shelf (km)</td>
<td>6.15</td>
<td>4.58</td>
<td>1.35</td>
<td>13.99</td>
<td>3.63</td>
</tr>
<tr>
<td>Distance to coast (km)</td>
<td>177.19</td>
<td>179.20</td>
<td>142.19</td>
<td>203.69</td>
<td>22.41</td>
</tr>
<tr>
<td>Time (minutes)</td>
<td>732.12</td>
<td>693.45</td>
<td>453.31</td>
<td>1218.29</td>
<td>201.35</td>
</tr>
<tr>
<td>Day of year (days)</td>
<td>168.61</td>
<td>177.00</td>
<td>57.00</td>
<td>309.00</td>
<td>75.88</td>
</tr>
</tbody>
</table>
Table 4.4. Summary statistics for select habitat predictors derived from multibeam sonar data. Numbers following variables derived from multibeam data indicate the spatial scale, as meters of the area radius. TPI and TRI are Topographic Position Index, Terrain Ruggedness Index, respectively. Min, Max, and SD represent minimum, maximum, and standard deviation, respectively.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multibeam</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>backscatter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>204.28</td>
<td>205.89</td>
<td>182.44</td>
<td>231.35</td>
<td>7.97</td>
</tr>
<tr>
<td>25</td>
<td>204.30</td>
<td>205.85</td>
<td>183.18</td>
<td>224.38</td>
<td>7.68</td>
</tr>
<tr>
<td>50</td>
<td>204.31</td>
<td>206.05</td>
<td>183.22</td>
<td>219.43</td>
<td>7.50</td>
</tr>
<tr>
<td>100</td>
<td>204.37</td>
<td>205.84</td>
<td>183.87</td>
<td>217.39</td>
<td>7.21</td>
</tr>
<tr>
<td><strong>Backscatter variance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>31.94</td>
<td>16.31</td>
<td>2.37</td>
<td>987.93</td>
<td>54.05</td>
</tr>
<tr>
<td>25</td>
<td>35.51</td>
<td>19.66</td>
<td>5.65</td>
<td>382.00</td>
<td>39.22</td>
</tr>
<tr>
<td>50</td>
<td>39.13</td>
<td>30.30</td>
<td>7.23</td>
<td>302.68</td>
<td>28.69</td>
</tr>
<tr>
<td>100</td>
<td>42.72</td>
<td>41.13</td>
<td>10.10</td>
<td>144.63</td>
<td>16.66</td>
</tr>
<tr>
<td><strong>Bathymetry</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>-88.74</td>
<td>-87.47</td>
<td>-166.50</td>
<td>-47.17</td>
<td>26.09</td>
</tr>
<tr>
<td>25</td>
<td>-88.74</td>
<td>-87.35</td>
<td>-166.34</td>
<td>-47.20</td>
<td>26.04</td>
</tr>
<tr>
<td>50</td>
<td>-88.75</td>
<td>-87.11</td>
<td>-165.94</td>
<td>-47.19</td>
<td>25.89</td>
</tr>
<tr>
<td>100</td>
<td>-88.80</td>
<td>-86.97</td>
<td>-165.40</td>
<td>-46.36</td>
<td>25.42</td>
</tr>
<tr>
<td><strong>Bathymetry variance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.32</td>
<td>0.06</td>
<td>0.00</td>
<td>14.09</td>
<td>0.85</td>
</tr>
<tr>
<td>25</td>
<td>1.75</td>
<td>0.42</td>
<td>0.00</td>
<td>34.80</td>
<td>3.89</td>
</tr>
<tr>
<td>50</td>
<td>6.37</td>
<td>1.79</td>
<td>0.01</td>
<td>99.49</td>
<td>12.20</td>
</tr>
<tr>
<td>100</td>
<td>22.40</td>
<td>8.60</td>
<td>0.04</td>
<td>284.26</td>
<td>36.64</td>
</tr>
<tr>
<td><strong>Roughness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.77</td>
<td>0.51</td>
<td>0.03</td>
<td>5.79</td>
<td>0.80</td>
</tr>
<tr>
<td>25</td>
<td>0.78</td>
<td>0.54</td>
<td>0.04</td>
<td>4.40</td>
<td>0.74</td>
</tr>
<tr>
<td>50</td>
<td>0.78</td>
<td>0.58</td>
<td>0.07</td>
<td>3.85</td>
<td>0.67</td>
</tr>
<tr>
<td>100</td>
<td>0.80</td>
<td>0.67</td>
<td>0.09</td>
<td>3.28</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>Slope</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>4.27</td>
<td>2.90</td>
<td>0.13</td>
<td>31.61</td>
<td>4.27</td>
</tr>
<tr>
<td>25</td>
<td>4.28</td>
<td>3.03</td>
<td>0.23</td>
<td>23.49</td>
<td>3.98</td>
</tr>
<tr>
<td>50</td>
<td>4.31</td>
<td>3.20</td>
<td>0.36</td>
<td>20.80</td>
<td>3.58</td>
</tr>
<tr>
<td>100</td>
<td>4.42</td>
<td>3.76</td>
<td>0.49</td>
<td>17.72</td>
<td>3.08</td>
</tr>
<tr>
<td><strong>TPI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.22</td>
<td>0.19</td>
<td>0.02</td>
</tr>
<tr>
<td>25</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.10</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>50</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.04</td>
<td>0.04</td>
<td>0.01</td>
</tr>
<tr>
<td>100</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>TRI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.23</td>
<td>0.16</td>
<td>0.01</td>
<td>1.96</td>
<td>0.24</td>
</tr>
<tr>
<td>25</td>
<td>0.23</td>
<td>0.16</td>
<td>0.01</td>
<td>1.37</td>
<td>0.22</td>
</tr>
<tr>
<td>50</td>
<td>0.24</td>
<td>0.17</td>
<td>0.02</td>
<td>1.20</td>
<td>0.20</td>
</tr>
<tr>
<td>100</td>
<td>0.24</td>
<td>0.21</td>
<td>0.03</td>
<td>1.00</td>
<td>0.17</td>
</tr>
</tbody>
</table>
Figure 4.3. Relative influence of habitat predictors for BRT model of MVBS.
Figure 4.4. Partial dependence plots for the effects of cell depth and bottom offset: the two most influential predictors of both MVBS (top) and fish density (bottom) BRT models.

MVBS values of independent testing data ranged from -111.008 to -41.439 dB. Residuals were evenly centered around zero, ranging from -44.363 to 40.261 with a mean of 0.0282 (Figure 4.5). Average difference between observed and predicted values had RMSE of 4.446, MAE of 2.640, and MAPE of 0.0370 (Figure 4.6).

Fish density ranged from 0 to 1.139. The vast majority (88.6%) of cells had fish densities of zero, having had no single fish targets detected. Nonzero density values ranged from 0.00021 to 1.1388 with a mean of 0.0211. The binary classification fish density model found all 53 predictors had nonzero influence. Mean cell depth and bottom offset were again the most influential variables relative to the other predictors (Figure 4.7). Compared to the MVBS model,
these two predictors were even more influential: mean cell depth had relative influence of 28.376 and bottom offset had 17.202. Relative influence drops off rather precipitously after these two variables. Similar to the MVBS model, cell depth had a domed curve relationship with fish
Figure 4.7. Relative influence of habitat predictors for binary fish density BRT model.
density and bottom offset indicated greater fish density near the seafloor (Figure 4.4). The cell depth at which there is the largest response was greater for presence of fish density than MVBS. Bank was important for fish density modelling, but not for MVBS.

Fish density data were imbalanced, with 1,517 positive (presence) observations and 11,721 negative observations. Many metrics depend on the probability cutoff at which predicted values are assigned to the zero or one categories. The traditional default of 0.5 stems from balanced data, for which the expected value of each class is 0.5; however, in these data, a positive outcome for a cell has a probability of only 0.1146 based on the number of zeros and ones. Classification metrics and class error rates change significantly depending on what probability triggers assignment to the positive class (Figure 4.8). At a probability of 0.13, the class error rates are approximately equal. If both types of these errors are of equal importance, a 0.13 cutoff for assigning a one for the predicted value may be more appropriate than 0.5. Classification metrics were calculated for a probability cutoff of both 0.5 and 0.13 (Table 4.5).

Figure 4.8. Error rates and classification metrics for a range of probability cutoff values for predicting a positive value.
Table 4.5. Error rates and metrics for probability cutoff values of 0.50 and 0.13.

<table>
<thead>
<tr>
<th>Metric</th>
<th>0.50</th>
<th>0.13</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Negatives</td>
<td>11,420</td>
<td>10,508</td>
</tr>
<tr>
<td>True Positives</td>
<td>930</td>
<td>1,354</td>
</tr>
<tr>
<td>False Negatives</td>
<td>587</td>
<td>163</td>
</tr>
<tr>
<td>False Positives</td>
<td>301</td>
<td>1,213</td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.6131</td>
<td>0.8926</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.9743</td>
<td>0.8965</td>
</tr>
<tr>
<td>Precision</td>
<td>0.7555</td>
<td>0.5275</td>
</tr>
<tr>
<td>Class 0 error</td>
<td>0.0257</td>
<td>0.1035</td>
</tr>
<tr>
<td>Class 1 error</td>
<td>0.3869</td>
<td>0.1075</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>0.9298</td>
<td>0.8930</td>
</tr>
</tbody>
</table>

Adjusting the probability assigned to the positive class decreased class 1 (positive fish density) error at the expense of class 0. Overall accuracy decreased from 92.98% to 89.93%. Accuracy of the model peaked at a probability cutoff of 0.49 (Figure 4.9). Sensitivity increased at the expense of specificity and precision. Based on predicted probabilities, area under the receiver operating curve was 0.9579, while area under the precision-recall curve was 0.7694, indicating a good model.

![Figure 4.9. Overall model accuracy as a function of probability at which values were predicted in the positive class.](image-url)
DISCUSSION

Depth in the water column was clearly the most influential factor in predicting both MVBS and positive fish density. This is not surprising considering what is known about the multitude of changes that occur with depth in the aquatic environment. Depth is related to light, temperature, pressure, and substrate in this area (Rezak et al. 1983, Wilson et al. 2006). Major habitat types in this area, each with characteristic fish assemblages, are known to be related to depth (Rezak et al. 1985, Schmahl et al. 2008). Wilson et al. (2006) found a very similar relationship between MVBS and depth; however, this study extends the depth range by 90 m.

Measures of depth were represented in both models’ top two most influential predictors: mean cell depth and bottom offset. Cell depth expresses depth in relation to the water surface, while offset describes it in relation to the bottom. The ability of these two predictors together to describe depth, may be a reason why bottom depth—the factor used to approximate habitat zones—was not of high importance: cell depth and offset can be combined to determine bottom depth.

The marginal relationship with bottom offset showed that MVBS is greatest within 10 m of the bottom and positive fish density is greatest within 40 m of the bottom. These habitats are home to reef-associated fish species, which would be expected to be relatively close to the bottom, near structure (Schmahl et al. 2008). Another contributing factor is the possibility that some backscatter included was from sources other than fish near the bottom, such as vegetation or sediment. Correctly classifying acoustic backscatter is a main challenge of hydroacoustics data processing and interpretation (Simmonds and MacLennan 2005b, Jech and Michaels 2006). Decibel differencing classification values were chosen to best describe a target scattering sound in the pattern expected from a swim-bladdered fish; however, such classification is not exact and
is more of an approximation (Kang 2002, Korneliussen 2002). Therefore, it is likely that some classifications of backscatter as fish were incorrect. Near-bottom analysis is further complicated by the acoustic “dead zone” near the seafloor. The curved shape of the transducer beam results in an area close to the seafloor that is not adequately sampled, which could result in under-sampling of fish near the seafloor (Simmonds and MacLennan 2005a).

Measures of time had varying levels of importance to modeling. Both time during the year and time of day were moderately influential predictors for MVBS. In fact, day of the year was the third most influential predictor of volume backscattering. This indicates seasonal differences throughout the year affect the backscattering from fish and therefore the quantity and/or size of the fish. Other studies have also shown changes in fish occurrence with season as well as time of day (Stanley and Wilson 1997, Langland 2015).

Comparisons between different scales at which multibeam data were calculated showed that the largest spatial scale resulted in the greatest importance. This would indicate that large-scale bathymetry is more indicative of the fish biomass and density than local scale. However, this result may not be reliable due to the multicollinearities among the predictors. The strong correlations between the bathymetric variables likely also contributed to the relatively low influence of each variable; since they are correlated, the unique information each was able to add to the model was diminished. These variables all added value to the model, as evidenced by their nonzero influence. Because the focus on machine learning is on predictive accuracy, additional variables—even if they have multicollinearity—are advantageous because they improve predictions.

Fish density value calculated in this study were higher than those observed previously on bank habitats in the northwestern GOM (Wilson et al. 2003, Wilson et al. 2006, Langland 2015).
Wilson et al. (2003) and Wilson et al. (2006) observed similar dome-shaped fish density relationships with depth. They found peak fish density was found at approximately 20 m at Sonnier Bank and 30 m at West Flower Garden Bank. Sonnier is significantly farther from the shelf edge, while West Flower Garden Bank crests shallower compared to the banks studied here; both factors could be expected to affect fish distributions. The effect of the cell depth and bottom offset on presence of positive fish density had very similar shapes and peaks as those found in Langland (2015), whose study covered three of the five banks studied here. The similarity shows that the marginal effect of these variables is consistent over time and across different banks.

Imbalanced data, such as the density data here, present an additional challenge. It is necessary to determine the goals of the model predictions and whether the accuracy of each class is valued equally. Overall model error may be minimized by splitting predicted probabilities evenly, such that a probability of greater than 0.5 leads to the prediction of 1; however, a better balance between errors for each class will be achieved if some class 0 error is compromised for the sake of class 1 error rate. In this case, the ability to predict positive values as well as negative is of importance. Therefore, a lower probability cutoff results in a more useful model. Lowering the cutoff between the predicted classes does decrease overall accuracy, but it only drops to 89.3%, which is still a good model.

These BRT models show the predictive power of machine learning techniques for predicting fish across habitats. The MVBS BRT can predict volume backscattering from fish with a standard deviation of 4.45 dB. The fish density BRT can predict correctly positive or negative fish density for 89.3% of observations. These are vastly better than a random model. The BRT model grown here can be used to predict on any new test sites where multibeam data
are available, which is true of many of the GOM shelf-edge banks. The longitudinal span of these sites, as well as the consistency of these results with results from Langland (2015) from only a subset of these sites, provides a promising indication that these patterns can be extended to nearby banks and across years. Such predictive power and understanding of the importance of habitat could be an integral tool in assessing the value of sites being considered for addition to the Flower Garden Banks National Marine Sanctuary (Office of National Marine Sanctuaries 2016). These models could be improved with more research and attention to discriminating between types of scatterers, including types of fish, based on their acoustic properties.

**LITERATURE CITED**


Langland, T. 2015. Fish assemblage structure, distribution, and trophic ecology at Northwestern Gulf of Mexico banks. Louisiana State University, Baton Rouge, LA.


Saito, T., and M. Rehmsmeier. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 10:e0118432.


GENERAL SUMMARY AND CONCLUSIONS

There were four objectives in this study: (1) estimate size at sexual transition for six GOM grouper species, (2) determine the optimal number of cameras on a baited remote underwater video system, (3) create a predictive model to provide presence of fish species based on habitat, and (4) grow a model to predict fish backscatter and density based on habitat parameters. BRUVS provided a non-extractive method to assess fish assemblages to the species level, while hydroacoustic sampling contributed complimentary information fish distribution on a larger spatial scale. Gonad histology identified information about grouper reproduction that would otherwise be unable to determine.

Chapter 1 employed logistic models to determine groupers’ size at sexual transition. Bayesian logistic modeling techniques estimated lengths at sexual transition for six grouper species. The study demonstrated the utility of applying Bayesian estimation in this context. It was particularly effective at producing models that would not converge when frequentist statistics are applied. Data-limitation for many species highlighted the difficulty in ascertaining reliable estimates for such life history parameters. Steps towards dependable parameter estimates are crucial to manage these species. Future work should explore application of these methods with additional species and richer datasets. Ideally, with enough data for a species, such modeling could also incorporate time to explore trends in life history parameters over time.

Chapter 2 explored the trade-off between number of cameras on a BRUVS and the quality of recorded data. Considering raw species and diversity metrics relative to time, fewer cameras are more efficient, although this decreases the resolution of available data. The optimal number of cameras to use on a BRUVS, and the corresponding viewing effort, depends on the survey goals. The forefront of underwater video is now advancing into full 360° and spherical
systems with the ability to simultaneously view an entire area, with no gaps (Kilfoil et al. 2017). A spherical view negates many issues regarding double-counting and the metrics that are used to avoid it. Methodologies are also improving for automated video viewing, an advancement that would wholly shift the discussion around allocation of video processing effort (Shafait et al. 2016, Dawkins et al. 2017, Siddiqui et al. 2018).

Chapter 3 employed random forest machine learning to deduce connections between habitat variables and presence of fish species. Models for several species were able to achieve high accuracy, meaning the models trained on collected data could predict well the occurrence of those species based on a site’s habitat. The prevalence of zeros in addition to the limited number of ones for many species presented a modelling challenge. Models need to account for the known imbalanced nature of the data. Some species simply did not have enough instances of detection for the machine learning algorithm to establish accurate predictions. Additional occurrence data, potentially more targeted to specific species of interest, could aid in the modeling process.

Chapter 4 utilized hydroacoustic data to grow boosted regression trees to predict fish backscatter, a proxy for biomass, and positive fish density based on habitat characteristics. Multibeam data available from the USGS were incorporated to supplement recorded aspects of habitat. While the multibeam data allowed for calculation of numerous bathymetric and terrain metrics, depth variables recorded by the split-beam echosounder were most influential in modeling. Some predictors had low relative importance, but all variables had non-negative importance and added to the model. Ongoing additions to the regression trees could expand the range of environmental variables, many of which are included in ongoing data collection and monitoring efforts, such as ocean temperatures, primary productivity, or current regimes. The more different types of predictors, the more likely it is to account for all the driving factors of
fish assemblages. Research is ongoing in advancements classifying scatterers during acoustic data processing to better isolate targets of interest (Korneliussen 2018).

Results of this study have improved the understanding of the community and population dynamics of fish living on the GOM shelf-edge banks. A variety of field methods as well as a range of data analysis techniques were used in an attempt to glean the most information from the data despite data and system complexities. Developed models help both understand as well as predict aspects of these fish. Understanding and predicting fish habitat use is beneficial for determining which habitats are most in need of protection and determining the overall value of the habitats to fish species and fisheries. Species distributions also indicate which areas should be targeted for directed sampling moving forward.

**LITERATURE CITED**


Barker, V. A. 2016. The effects of artificial light on the community structure and distribution of reef-associated fishes at oil and gas platforms in the northern Gulf of Mexico. Louisiana State University, Baton Rouge, LA.

Bates, C., and R. Colvin. 2016. Comparison of scamp grouper (Mycteroperca phenax), growth off of the West Florida shelf and the coast of Louisiana. LaGrange College.


ftp.stat.berkeley.edu/pub/users/breiman/OOBestimation.ps.


Cook, M. 2007. Population dynamics, structure and per-recruit analyses of yellowedge grouper, Epinephelus flavolimbatus, from the northern Gulf of Mexico. The University of Southern Mississippi, Hattiesburg, MS.


Langland, T. 2015. Fish assemblage structure, distribution, and trophic ecology at Northwestern Gulf of Mexico banks. Louisiana State University, Baton Rouge, LA.


Lowry, M., H. Folpp, M. Gregson, and I. Suthers. 2012. Comparison of baited remote underwater video (BRUV) and underwater visual census (UVC) for assessment of


National Oceanic and Atmospheric Administration Coral Reef Conservation Program. 2010. NOAA Strategic Plan for Deep-Sea Coral and Sponge Ecosystems: Research,


Reynolds, E. M. 2015. Fish biomass and community structure around standing and toppled oil and gas platforms in the northern Gulf of Mexico using hydroacoustics and video surveys. Louisiana State University, Baton Rouge, LA.


Saito, T., and M. Rehmsmeier. 2015. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 10:e0118432.


SEDAR. 2016. SEDAR 49 Gulf of Mexico Data-limited Species. Southeast Data Assessment and Review, North Charleston, SC.


Trott, T. M. 2006. Preliminary analysis of age, growth, and reproduction of coney (Cephalopholis fulva) at Bermuda. Pages 385-399 57th Gulf and Caribbean Fisheries Institute, Saint Petersburg, FL.


VITA

Elizabeth Keller was born in Poughkeepsie, NY in 1992. She grew up in Fishkill, NY. Despite not living near an ocean, she always loved the beach and water with a general passion for zoology and learning. Lizz graduated from John Jay High School in Hopewell Junction, NY in 2010 and pursued higher education at Cornell University in Ithaca, NY. She majored in Biological Sciences with minors in Environmental and Resource Economics and Natural Resources, earning a Bachelor of Science in May 2014. She developed her specific interest in fish ecology during two summers researching lake populations of alewife in upstate New York at the Cornell Biological Field Station at Shackelton Point. Lizz completed an honors thesis of her research under supervision of Lars Rudstam. Elizabeth entered the graduate program at Louisiana State University in August 2014 as a master’s student, but after one year transitioned to the doctoral program under the supervision of Dr. James H. Cowan, Jr.. While at LSU, she served a year as each Social Chair and President of the Coast and Environment Graduate Organization (CEGO). Lizz is currently a degree candidate in the Department of Oceanography and Coastal Sciences doctoral program, degree to be awarded in December 2019.