November 2018

Integrating Towed Underwater Video with Multibeam Acoustics for Mapping Benthic Habitat and Assessing Reef Fish Communities on the West Florida Shelf

Alexander Ross Ilich
University of South Florida, ailich@mail.usf.edu

Integrating Towed Underwater Video with Multibeam Acoustics for Mapping Benthic Habitat and Assessing Reef Fish Communities on the West Florida Shelf

by

Alexander Ross Ilich

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
Department of Marine Science
with a concentration in Marine Resource Assessment
College of Marine Science
University of South Florida

Major Professor: Steven A. Murawski, PhD
Stanley D. Locker, PhD
Theodore S. Switzer, PhD

Date of Approval:
November 1, 2018

Keywords: Benthic Habitat Mapping, Multibeam Echosounder, Visual Surveys, Fish-Habitat Relationships

Acknowledgments

The research presented in this thesis was conducted as part of the Continental Shelf Characterization Assessment and Mapping Project (C-SCAMP), which is funded by the National Fish and Wildlife Foundation (NFWF).

I would like to thank my advisor Steve Murawski, and my committee members Stan Locker and Ted Switzer for their valuable help and feedback on this research. Additionally I would like to thank the C-SCAMP team and the staff of the Center for Ocean Technology at the University of South Florida College of Marine Science (USF CMS), particularly Chad Lembke, Sarah Grasty, Alex Silverman, Steven Butcher, Matt Hommeyer, Heather Broadbent, Ed Hughes, Gerardo Toro-Farmer, Abigail Vivlamore, Jennifer Brizzolara, John Gray, and Katie Davis, for their help on collecting, processing, and integrating various data products. Additionally, I would like to thank Dave Jones for helping with several statistical analyses. I would like to thank my fellow Murawski Minions (lab mates), my cohort at USF CMS, and the USF CMS community in general for being supportive throughout my graduate career. I would also like to thank the various people and organizations that have funded me throughout my graduate career including Anne and Werner Von Rosenstiel, Peter Betzer, Joni James, Beverly Young, Beverly Knight, Larry and Diana Foster, and Florida Sea Grant. I would like to thank the National Center for Ecological Analysis and Synthesis (NCEAS) for accepting me into their Open Science for Synthesis program and providing me with a lot of the technical skills that were necessary for completing this thesis. Last, but definitely not least I would like to thank my family for always being there for me.

Table of Contents

List of Tables

List of Figures

Abstract

Using a towed underwater video camera system, benthic habitats were classified along transects in a popular offshore fishing area on the West Florida Shelf (WFS) known as "The Elbow." Additionally, high resolution multibeam bathymetry and co-registered backscatter data were collected for the entire study area. Using these data, full coverage geologic and biotic habitat maps were developed using both unsupervised and supervised statistical classification methodologies. The unsupervised methodology used was k-means clustering, and the supervised methodology used a random forest algorithm. The two methods produced broadly similar results; however, the supervised methodology outperformed the unsupervised methodology. The results of the supervised classification demonstrated "substantial agreement" ($\kappa>0.6$) between observations and predictions for both geologic and biotic habitat, while the results of the unsupervised classification demonstrated "moderate agreement" ($\kappa>0.4$) between observations and predictions for both geologic and biotic habitat. Comparisons were made with the previously existing map for this area created by Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute (FWC-FWRI). Some features are distinguishable in both maps, but the FWC-FWRI map shows a greater extent of low relief hard bottom features than was predicted in our habitat maps. The areas predicted as low relief hard-bottom by FWC-FWRI often coincide with areas of higher uncertainty in the supervised map of geologic habitat from this study, but even when compared with ground-truth points from the towed video rather than predictions, the low relief hard bottom in FWC-FWRI's map still corresponds to what was identified as sand in the video 73% of the time. The higher uncertainty might be a result of the presence of mixed habitats, differing morphology of hard-bottom, or the presence of sand

intermixed with gravel or debris. More ground-truth samples should be taken in these areas to increase the confidence of these classifications and resolve discrepancies between the two maps.

Data from the towed video system were also used to assess differences in fish communities among habitat types and to calculate habitat-specific densities for each taxa. Fish communities were found to significantly differ between soft and hard bottom habitats as well as among the hard-bottom habitats with different vertical relief (flat hard-bottom vs more steeply sloping areas). Additionally, significant differences were found between the fish communities in habitats with attached fauna such as sponges and gorgonians, and areas without attached fauna; however, attached fauna require rock to attach to and the rock habitats rarely lacked attached fauna, so this difference may just reflect the difference between fish communities in sand and rock habitats without the consideration of vertical relief. Moreover, the species driving the differences in the fish communities were identified. Fish were more likely to be present and assemblages were more species rich in more complex habitats (rockier, higher relief, presence of attached fauna). Habitat specific densities were calculated for each species, and general trends are discussed.

Lastly the habitat-specific densities were extrapolated to the total area of habitat type (sand vs rock) as predicted by the supervised geologic habitat map. There is predicted to be approximately 111,000 fish (95% CI [67015, 169405]) within the study area based on this method, with ~47,000 (~43%) predicted to be within the sand habitat and ~64,000 (~57%) in the rock habitat. This demonstrates the potential of offshore rocky reefs as “critical habitats” for demersal fish in the offshore environment as rock accounts for just 4% of the study area but is expected to contain over half of the total abundance. The value of sand habitats is also shown, as

due to their large area they are able to contribute substantially to the total number of fish despite sustaining comparatively low densities.

Chapter 1: Project Overview

Importance and Objectives

The West Florida Shelf (WFS) sustains commercially and recreationally important fisheries for a variety of species, especially demersal fish which includes reef fishes such as snappers and groupers. Commercial fisheries in Florida contribute approximately $6 billion to the region's GDP and support almost 80 thousand jobs in Florida. Additionally, the recreational fishing sector on the WFS contributes another $4 billion to the region's GDP and supports over 61 thousand jobs (NMFS, 2017). This is a crucial economic sector for the region, and large demersal reef fishes such as groupers, snappers, jacks, and porgies are key resources for these industries. Many of these reef fishes have life history characteristics that make them particularly susceptible to overfishing such as slow growth and late maturity (Musick, 1999, Coleman et al., 2000). Traditional fisheries management is based on single-species population dynamics; however, fish populations can be affected by a number of external ecological, economic, and social dynamics. For example, there may be interactions between multiple fisheries if a fishery exists for both a predator as well as its prey, as increased fishing pressure on the prey species may reduce the sustainable yield level for the predator (Sinclair et al., 2002). To better account for these complexities, the National Oceanic and Atmospheric Administration (NOAA) is transitioning to a more comprehensive and holistic management scheme known as Ecosystem-Based Fisheries Management (EBFM), a move which is supported by the governmental and academic communities (McLeod et al., 2005).

One of the steps in implementing EBFM is understanding relationships between species distributions and identification of critical habitats and essential fish habitat. Habitat maps

combined with an understanding of the functional significance of each habitat type are critical to implementing effective and scientifically sound EBFM (Kendall, 2005, Shumchenia and King, 2010). Despite the tremendous importance of fisheries on the WFS, as of 2014, high resolution bathymetry existed for less than 5% of the WFS, and even less area had been "ground-truthed" using technologies such as underwater video (C-SCAMP, n.d.). Thus the relative importance, quantity, and distribution of benthic habitats along the WFS remains highly uncertain. It is thus impossible to understand how various habitat protections will affect reef fish populations. There are several known high value habitat areas along the WFS that support abundant and diverse communities of demersal reef fish as well as endangered sea turtles (Allee et al., 2011, Grasty, 2014, Hardy et al., 2014). These areas include the Madison-Swanson Marine Protected Area (MPA) which has large limestone ridges and is a confirmed site of Gag Grouper (*Mycteroperca microlepis*) spawning aggregations, the Steamboat Lumps MPA which contains a large number of grouper holes created by the Red Grouper (*Epinephelus morio*), and the Pulley Ridge and the Florida Middle Grounds Habitat Areas of Particular Concern (HAPC's) which both contain offshore coral reefs (Hine et al., 2008, Coleman et al., 2011, Wall et al., 2011). However, these areas likely represent just a small fraction of the totality of high value habitat areas that exist on the WFS. The Continental Shelf Characterization Assessment and Mapping Project (C-SCAMP) aims to approximately double the area of the WFS mapped with high resolution bathymetry (C-SCAMP, n.d.). From 2015 - 2018, approximately 1,850 Km^2 of WFS habitat has been mapped using a multibeam echosounder with over 330 hours of associated towed underwater video to ground-truth habitat and assess reef fish populations (Figure 1).

The research conducted for this thesis is part of the C-SCAMP project and will focus on integrating data from the towed underwater video system with data collected with a multibeam

echosounder in a popular offshore fishing area known as “The Elbow.” The objectives of my research are to:

1. Develop an objective and semi-automated methodology for creating full coverage habitat maps.
2. Develop quantitative relationships between fish abundance and community composition, with habitat characteristics.
3. Use the results to estimate the abundance of various demersal reef fish by habitat type.

This research aids in understanding the biology of several demersal reef fish, providing critical baseline data on the fish communities and identifying critical habitats of the often overlooked marine offshore environment. The lack of baseline data in the marine offshore environment was extremely evident in the wake of the *Deepwater Horizon* oil spill, when it was difficult to assess impacts related to the blowout and track recovery of ecosystems, as there was no known reference state for many ecosystems (Love et al., 2015). Additionally, although this thesis focuses on a single area, the methodology presented here is applicable to other areas on the WFS for which the C-SCAMP group has collected data. This research also demonstrates the utility of combining technologies of towed camera systems and multibeam echosounders for fisheries management, and can aid in operationalizing the use of these new and innovative technologies for assessing fish populations and simultaneously supporting fisheries and habitat management. The results of this research quantitatively link habitat and environmental characteristics to fish community composition and abundance to facilitate more accurate assessments of fish stocks. Specifically, habitat maps and fish-habitat relationships can aid in developing more accurate predictive models of fish communities, and can provide information useful for stratifying survey design of fisheries-independent surveys by habitat which could

improve sampling efficiency. Locating the habitats of interest was identified as one of the major challenges for fisheries-independent surveys for monitoring reef fish as the location of many habitats is unknown making optimum allocation of sampling effort difficult (Switzer et al., 2014). This research can aid in optimizing the sampling effort allocation and in increasing the statistical power of fisheries independent surveys by providing the location of these habitats of interest (Cogan et al., 2009, Switzer et al., 2014). The methodology employed here may also prove useful for determining which areas should be considered "essential fish habitat" as defined by the Magnuson-Stevens Act Fishery Conservation and Management Act of 1976, "critical habitat" as defined under the Endangered Species Act, or for designating habitat areas of particular concern (HAPC's) and marine protected areas (MPA's). Moreover, the results of my research will help further our understanding of the drivers behind what creates suitable habitat for different fish species and therefore can aid in locating of more of these high value habitats in the future.

The Camera-Based Assessment Survey System (C-BASS)

The C-BASS is a towed underwater camera system built for reef fish stock assessment by engineers from the Center for Ocean Technology at the University of South Florida College of Marine Science (Lembke et al., 2013, Lembke et al., 2017). The C-BASS is towed behind a research vessel at speeds of 1.5 - 2 ms^{-1} and between 2 - 4 meters above the seafloor (Lembke et al., 2013). The system consists of four LED lights, six underwater video cameras as well as various sensors (Figure 2; Lembke et al., 2013, Lembke et al., 2017). There are two forward-facing high definition (HD) cameras, one monochrome and one color. There are also four additional color standard definition (SD) cameras, two of which are front facing and two of which are angled to the sides. The forward-facing monochrome HD camera is the primary

camera used to identify fish and habitat types in this study as it consistently provided the clearest imagery. Other cameras are supplementary and can be used to aid in fish identification (e.g. if a fish swims out of frame before it could can be identified in the front camera but can be seen in the side camera) or to provide more perspective on the habitat of a given area if it is unclear in the primary camera (Grasty, 2014). The cameras are oriented at an angle below the main horizontal chassis plane, rather than directly downwards, as an oblique orientation increases the area observed, increases fish detection probability, and provides a perspective that aids in the identification of fish species and habitat characteristics (Bowden and Jones, 2016, Lembke et al., 2017). The sensors on the C-BASS include a compass to record the pitch, roll, and heading of the towbody, an altimeter to record height above the seafloor, and a CTD and fluorometer to record depth and ambient water properties (Lembke et al., 2017). All sensor data are recorded at a frequency of 1Hz or greater and exported to a single 1Hz table for ease of use.

Study Area

My research tests the utility of combining multibeam echosounders and towed underwater video for mapping benthic habitats and assessing fish communities in a popular offshore fishing area known as "The Elbow." The Elbow is hypothesized to be an ancient limestone coastline that was shaped by wave action approximately 12,000 years ago (Moe, 1963, Switzer et al., 2014). The area lies about 145 km northwest from the Sunshine Skyway Bridge at the mouth of Tampa Bay (Figure 1). The Elbow contains both hard-bottom and soft-bottom habitats, and benthic biological assemblages including sponges, gorgonians, and sea urchins, as well as a diverse community of reef fishes. In December of 2015, the C-SCAMP project mapped approximately 100 km^2 of The Elbow using a multibeam echosounder (Figure 3; C-SCAMP,

n.d.).This portion of The Elbow typically ranges in depth from 45 to 65 meters, and contains a long linear ridge that runs north to south for at least 16 km (Figure 3).

Data Collection

Multibeam Echosounder

Multibeam bathymetry of the Elbow was collected in December 2015 (Figure 3) using a Teledyne Reson SeaBat 7125, a dual-frequency multibeam swath sonar with 512 overlapping beams that can be operated at 400 or 200 kHz (C-SCAMP, n.d.). For this study the SeaBat7125 was operated at 400 kHz with a 140° swath which provides an across-track receive beam width of 0.5° and an along-track transmit beam width of 1°. The SeaBat 7125 was pole-mounted on the port side of the R/V Bellows. Navigation and motion compensation data were collected with the Applanix POS MV OceanMaster system. The POS MV system consists of an inertial motion unit (IMU) and a global positioning system (GPS) azimuth measurement system (GAMS) including two GPS receivers and has a position accuracy of 0.5 - 2 m^2 (Applanix, 2017). An AML Oceanographic Micro•X was used to correct for sound launch velocity at the sonar head and an AML Oceanographic Minos•X with an SV•Xchange sound velocity sensor was used for sound velocity profile correction.

C-BASS Transects

Video transects were planned by visual inspection of the multibeam bathymetry to maximize the likelihood of encountering all habitat types. Four video transects from the February 2016 C-SCAMP cruise on the R/V Weatherbird II were used for the analysis (Figure 4). These transects were collected over five days from February 17th to 21st, 2016. Transect six followed the main North-South ridge. Transect five “zigzags” across the entire study area crossing over the main ridge multiple times to sample a broad range of habitats. Transect three bisects the

study area from North to South. These three transects were collected during the day and were used as the training data set for creating habitat maps and were used in fish analyses. Transect one was collected at night follows a smaller ridge to the west of the main ridge. This transect was reserved as an independent validation transect for the habitat maps and was excluded from fish analyses.

Overview of Methods

This thesis will first relate ground-truth habitat observations from the towed video to the bathymetry and backscatter data collected by the multibeam system to create predicted habitat maps of geologic and biotic aspects of habitat. Terrain variables such as slope, rugosity, and curvature can be derived from the bathymetry surface, and texture metrics can be derived from the backscatter mosaic. These derived data sets as well as the bathymetry surface and backscatter mosaic themselves can be used for statistical classification of habitat (Lamarche et al., 2016). Predicted habitat maps were creating using both a supervised and unsupervised methodology. Supervised classification uses a set of training data to determine a statistical relationship between the observed habitat (e.g. from underwater video) and the available full coverage datasets (e.g. bathymetry backscatter, and their derivative features). These relationships are then used to predict habitat to the full extent of the study area. Unsupervised classification typically relies on clustering algorithms to segment the full coverage data sets within the study area into unique clusters without any consideration of the observed habitat from the ground-truth data. After segmentation, the ground-truth habitat observations are then used to interpret the clusters. The supervised methodology used in this thesis is a random forest algorithm, and the unsupervised methodology uses k-means clustering. The resulting maps are then assessed for accuracy using various metrics, and are compared to the existing geoform habitat map created by the Florida

Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute (FWC-FWRI).

In addition to habitat maps, the fish communities in this area are examined. Geometric calculations are used to calculate the area viewed by the towed camera system allowing for fish counts are converted to densities (Grasty, 2014). Principal Coordinates Analysis (PCoA) is used to identify the species driving the overall variation in fish community composition and abundance. This is followed up by a non-parametric Analysis-of-Variance (PERMANOVA) to test for significant differences in fish community composition and abundance among geologic and biotic habitats. A Canonical Analysis of Principal Coordinates (CAP) analysis is then used to determine the species driving those differences among habitat types. Additionally, these multivariate analyses are followed up with univariate analyses. The species richness among habitat types is explored, and habitat-specific densities are calculated for each taxa. These habitat-specific densities are then combined with the areas calculated from one of the habitat maps to provide estimates of total abundance for all observed taxa in the study area.

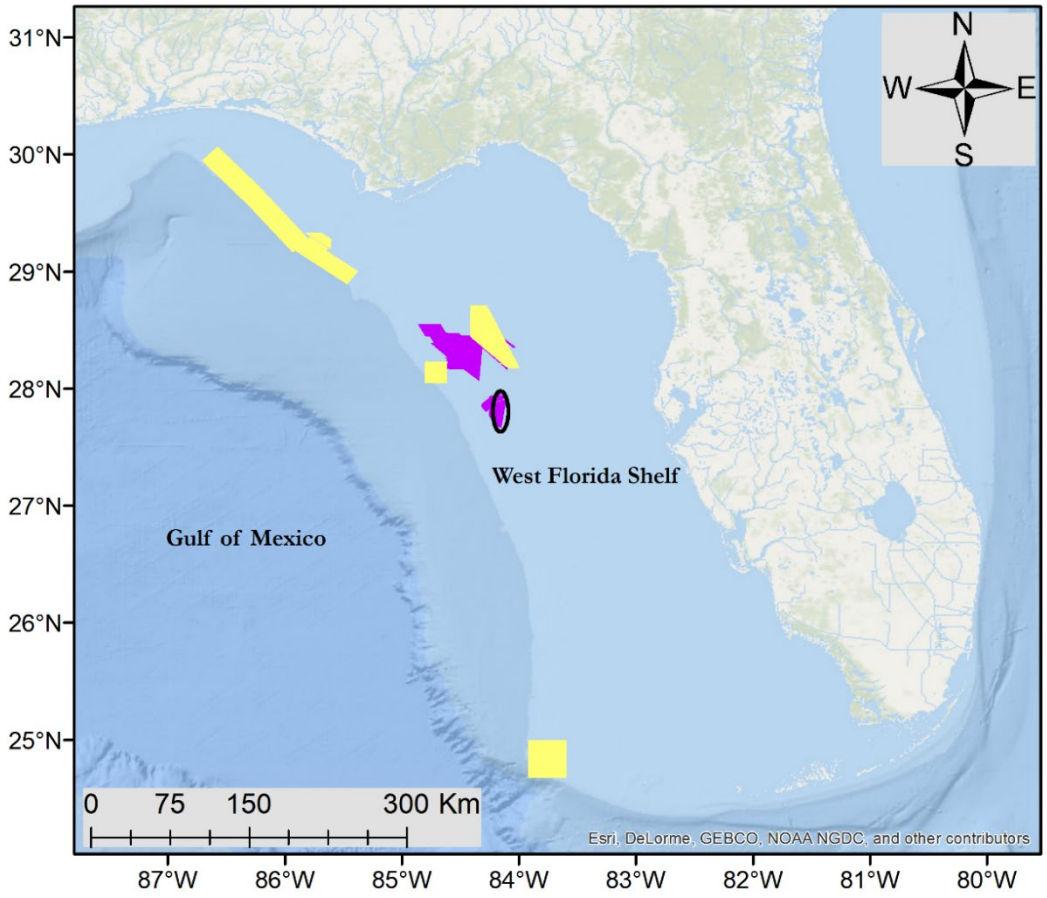


Figure 1: Map of areas previously mapped on the WFS at 10 m x 10 m resolution or finer (yellow) and those mapped by the C-SCAMP project (purple). The area circled in black is called “The Elbow” and is it the study area of this thesis.

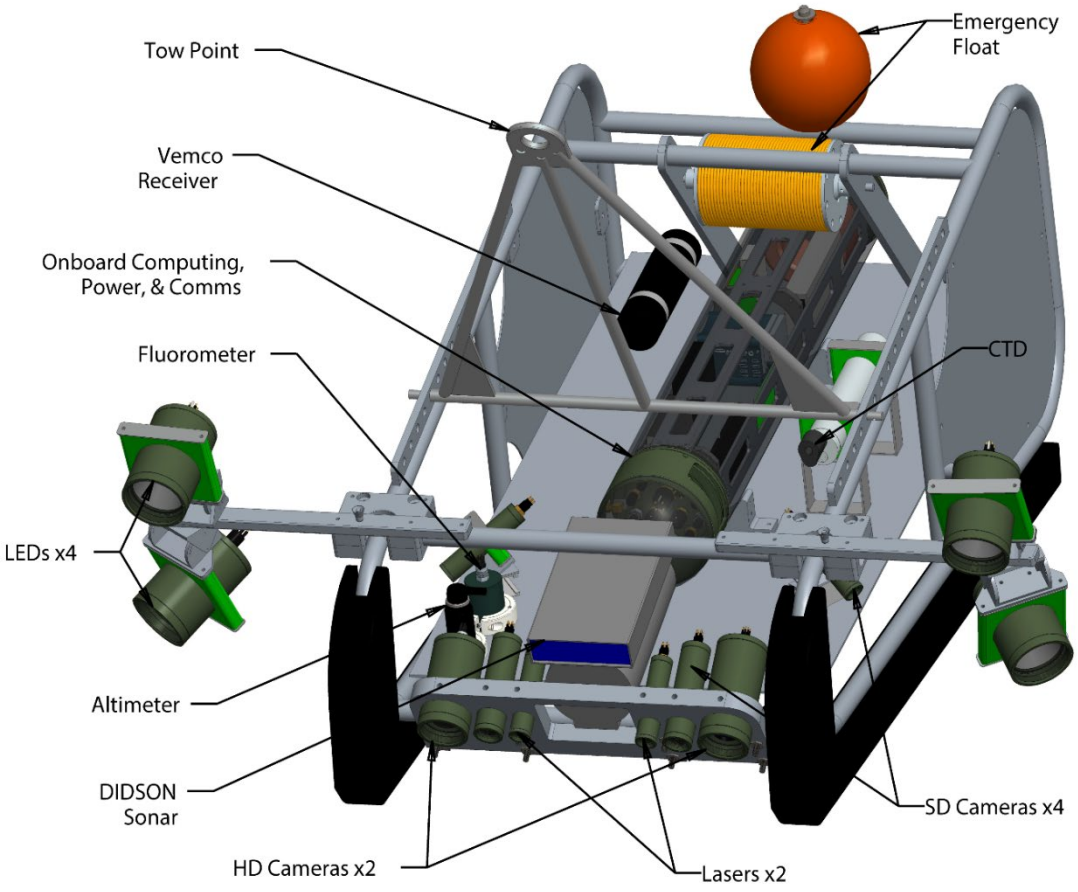


Figure 2: Schematic of the C-BASS towed video system

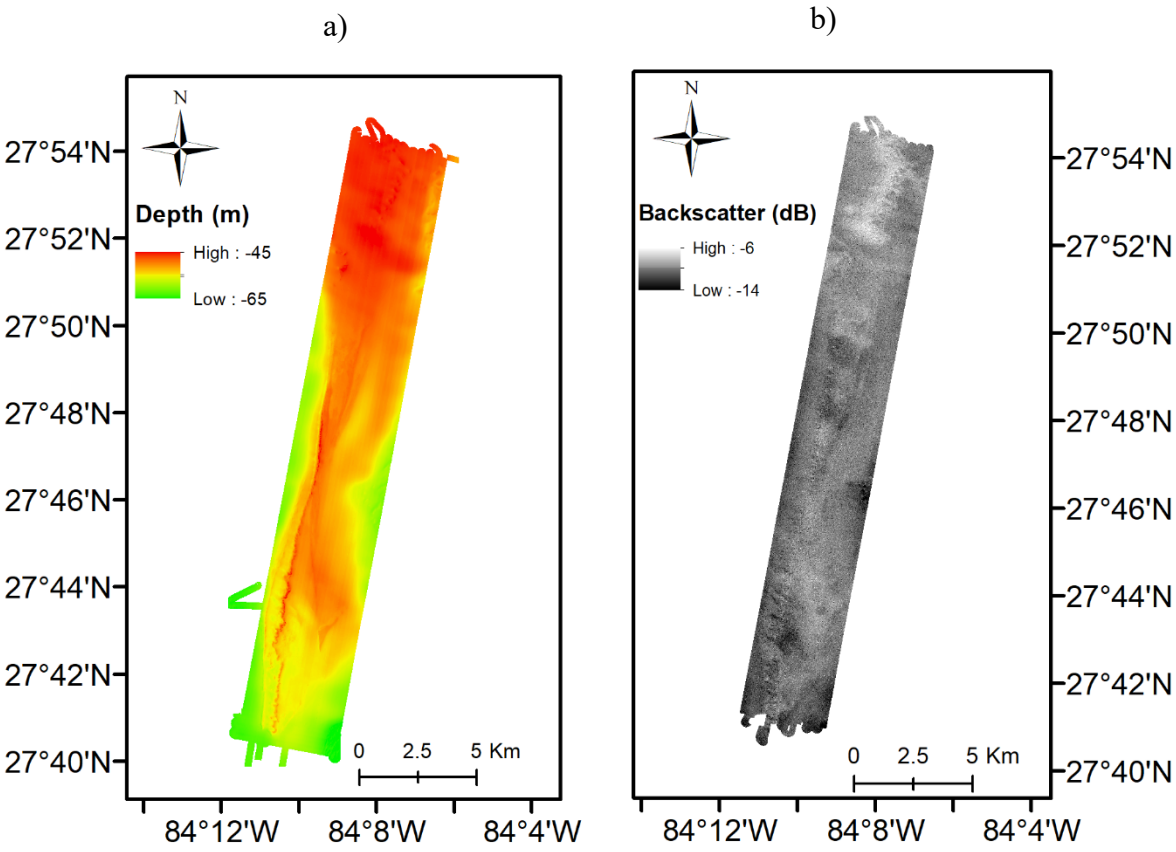


Figure 3: Full multibeam 2 m x 2 m bathymetry (a) and 1 m x 1 m backscatter (b) raster surfaces of The Elbow collected by the C-SCAMP project

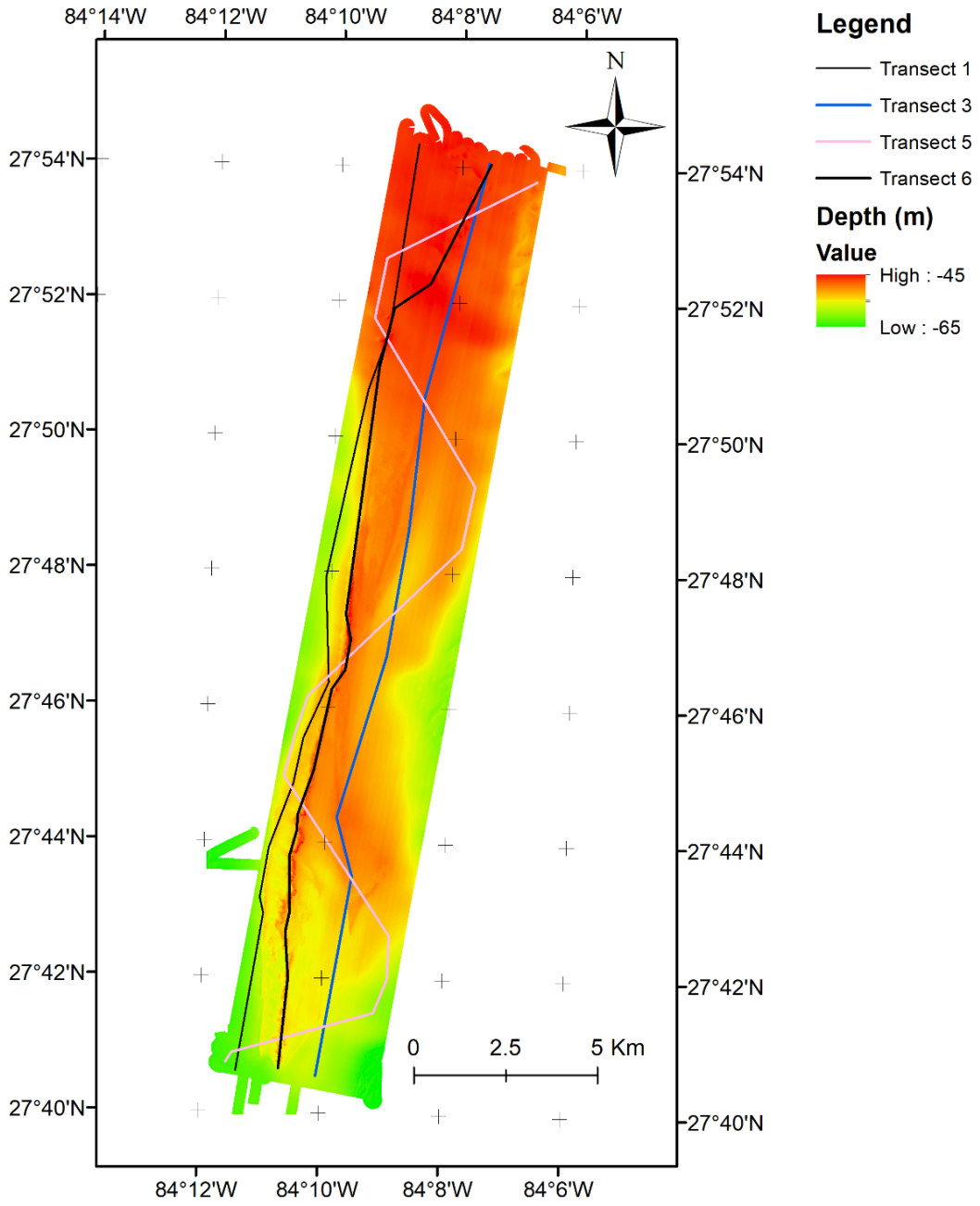


Figure 4: 2 m x 2 m bathymetry (m) of the Elbow with overlain towed video transects. Transects three, five, and six were daytime transects, and were used for training habitat models and for fish analyses. Transect one was a night transect and was used for validation of habitat models, and was not used in fish analyses.

Chapter 2: Benthic Habitat Mapping

Introduction

With increasing stress on the marine environment, maps of the seafloor are in high demand in order to better manage coastal and offshore resources as evidenced by several mapping initiatives at the regional, national, international, and even global scale (C-SCAMP, n.d., Andersen et al., 2018, Mayer et al., 2018). Maps of the seafloor can be useful for many sectors. For example, benthic habitat maps in the Gulf of Maine have been effectively used in Canada for siting of offshore facilities, MPA creation, and improving fisheries management (Pickrill and Todd, 2003).

Although there are several technologies available for mapping the seafloor, multibeam echosounders have rapidly become the most popular tool for surveying and mapping large portions of the seafloor as the large swath of beams can accurately and rapidly map the seafloor (Brown et al., 2011, Lamarche et al., 2016). Additionally, multibeam echosounders have the advantage over sidescan sonars of being able to simultaneously collect co-registered bathymetry and backscatter information (Brown et al., 2011, Lamarche et al., 2016). Collection of bathymetry provides a topographic map of the seafloor, and it has proved to be a very important predictor of habitat attributes (Hasan et al., 2014). This is likely because bathymetry and its derivatives (e.g. slope, curvature, aspect, rugosity) relate to complexity of the seafloor which may in turn relate to ecological processes such as providing shelter from predation for fish and mobile invertebrates, and providing areas to settle for benthic colonizers (Wilson et al., 2007). Backscatter, on the other hand is a related to how strong the echo returns, which can be a good

predictor of sediment grain-size, composition, and substrate type (Goff et al., 2000, Collier and Brown, 2005, Brown et al., 2011, McGonigle and Collier, 2014, Lamarche et al., 2016, Brizzolara, 2017). Therefore bathymetry and backscatter both provide different but complementary information describing the potential habitat of an area (Brown et al., 2011, Hasan et al., 2014). Bathymetry and backscatter both can be used to delineate habitat types on the seafloor, and including both bathymetry and backscatter as well as their derivatives increase the accuracy of habitat maps over using either one of them alone (Ierodiaconou et al., 2007). In addition to the collection of bathymetry and backscatter, it is critical to collect some form of ground-truth information (e.g. underwater video or sediment grabs) in order to inform and/or validate habitat classification products (Brown et al., 2011, Lamarche et al., 2016).

Traditionally, habitat maps have been created through manual delineation of boundaries by expert interpretation of acoustic data sets (Brown et al., 2011). This method, although effective in some scenarios, is subjective and can be time consuming, and is less reliable when contrast is more subtle which can occur for example when trying to identify flat hard bottom areas (Riggs et al., 1996, Cochrane, 2008). With the increasing volume of data and the desire to use these maps for management, there has been increased interest in developing semi-automated statistical classifiers that can create habitat maps in a more objective and repeatable manner (Cochrane, 2008, Brown et al., 2011, Brown et al., 2012, Diesing et al., 2014, Lecours, 2017). These statistical classifiers extrapolate habitat to the entire study area from a set of ground-truth observations. These classifiers typically fall into one of two categories: supervised or unsupervised classification (Brown et al., 2011). Supervised classification uses a set of training data to determine a statistical relationship between the observed habitat (e.g. from underwater video) and the available full coverage datasets (e.g. bathymetry, backscatter, and their derivative

features). These relationships are then used to predict the habitat to the full extent of the study area. In contrast, unsupervised classification typically relies on clustering algorithms to segment the full coverage data sets within the study area into unique clusters without any consideration of the observed habitat from the ground-truth data. After segmentation, the ground-truth habitat observations are then used to interpret the clusters. Many of these clustering algorithms require the number of clusters to be specified *a priori*; however, this can be difficult given that there is rarely a 1:1 correspondence between clusters and habitat types. Therefore, interpretation often requires several clusters to be merged into one habitat type (Brown et al., 2012, Stephens and Diesing, 2014). Results from unsupervised classifications thus can be sensitive to the number of clusters that is specified.

In addition to the call for increased objectivity in the delineation of marine habitats, it is widely recognized that there is a need to have a standard nomenclature in marine habitat classification (Greene et al., 1999, Greene et al., 2007, Costello, 2009, Federal Geographic Data Committee, 2012). The word "habitat" has been used in many different contexts in the scientific literature, as different studies may focus on different aspects of habitat (e.g. biotic vs geologic, or benthic vs pelagic), may use differing classification schemes, and may study habitat in regards to different organisms and at different spatial scales. This lack of consistency can make comparisons across studies difficult and reduces the ability to merge results from several studies into maps that can be used at the regional or national scale for resource management (Federal Geographic Data Committee, 2012). In order to address this issue The Coastal and Marine Ecological Classification Standard (CMECS) was developed to provide a standard framework and nomenclature for classifying coastal and marine environments in the United States and in 2012 it was adopted as the national standard for describing these habitats (Federal Geographic

Data Committee, 2012). The CMECS scheme is composed of six main elements (Figure 5). The biogeographic setting and the aquatic setting are hierarchical elements with three levels. The biogeographic setting represents ecoregions defined on the basis of climate, geology, and evolutionary history and the aquatic setting represents zones defined by salinity, coastal proximity, and tidal regime (Federal Geographic Data Committee, 2012). The biogeographic setting and aquatic setting both represent broad characteristics that are generally applicable to an entire study area. In addition to the biogeographic and aquatic settings, the CMECS scheme also consists of four components describing a different aspect of habitat. (1) The water column component which describes properties of the water column. (2) The geoform component which describes the geomorphological and structural characteristics of the seafloor. (3) The substrate component which describes what the seafloor is composed of. (4) The biotic component which describes both the planktonic and benthic biotic communities (Federal Geographic Data Committee, 2012). These components can be applied to individual sampling sites or sub-areas within the overall study area, and a given study can classify one or more of these components depending on their sampling methodology and research goals.

Methods

Data Processing

Multibeam Data

Bathymetry and backscatter data were processed by the C-SCAMP group. The bathymetry data were post processed according to IHO standards using Caris HIPS and SIPS 10.2 and meet or exceed IHO order 1A standards (IHO, 2008). The backscatter mosaic was created using the Caris SIPS Time-Series algorithm. The bathymetry surface and backscatter

mosaic were exported to 2 m x 2 m resolution and 1 m x 1 m resolution raster grids respectively (Figure 3b)

To align the raster grids and match the cell size to the scale of observations seen in the video, the bathymetry surface was aggregated to a 10m cell size, and the backscatter mosaic was first aggregated to 10 m resolution and then resampled to a matching raster grid using bilinear interpolation which calculates the values of each cell in the new resampled raster grid as a weighted average (by distance) of the four closest cells in the input raster. Due to changes in sonar settings, backscatter data was unavailable for part of the survey area, so both surfaces were trimmed to a common area where bathymetry and backscatter data were both available (Figure 6).

The matching 10 m bathymetry and backscatter grids were used to calculate various derivative features. These features include terrain attributes derived from the bathymetry surface as well as texture measures derived from the backscatter mosaic using gray level co-occurrence matrices (GLCM's; Haralick and Shanmugam, 1973, Wilson et al., 2007). A GLCM examines pairs of cells and is essentially a table of the relative frequencies at which different values occur next to each other (Hall-Beyer, 2017). The first step in constructing a GLCM is to scale the original data to a set number of discrete levels called grey levels. Then a window (e.g. 3 x 3 cell window) is created around a central cell. Within this window, frequencies at which different grey levels neighbor one another in a set direction (e.g. horizontally) are tallied in a matrix with rows and columns representing the corresponding grey levels. These frequencies are then converted to probabilities by dividing by the sum of all the frequencies in the matrix. The resulting GLCM is a matrix with row and column indices representing grey levels, and the values in each cell representing the probability of those two values neighboring one another. The resulting GLCM

can then be used to calculate texture metrics for that central cell describing an aspect of texture in the area at and around that central cell.

All bathymetric terrain attributes and backscatter texture metrics were computed using a 3 cell x 3 cell moving window. For texture measures, 32 gray levels were used, and the value of metrics was averaged over all directions (horizontally, vertically, and diagonally). Terrain features from the bathymetry were derived using the Raster package in R as well as the Benthic Terrain Modeler add-in for ArcGIS (Table 1, Appendix 1; Hijmans, 2016, Walbridge et al., 2018). Texture measures derived from the backscatter mosaic were calculated using the glcm R package (Table 2, Appendix 2; Zvoleff, 2015).

Video Data

Habitat Annotation

Habitat was classified from still images extracted from the video approximately every 15 seconds; however, scrolling a few seconds in each direction was allowed to provide context and ensure that the classification given adequately characterized the area. The primary camera used is the monochrome HD Point Grey Blackfly® camera as it has consistently provided the best imagery.

Habitat frames were classified according to a customized version of the CMECS Biotic, Substrate, and Geoform Components including modifiers for percent cover of primary induration (Appendix 3 and Appendix 4). These classifications informed the CMECS summary classification for the overall area; however, meaningful statistical analyses required generalizing these categories into broader classes due to issues of positional uncertainty and the need for sufficient sample sizes. For statistical analyses, these categories were reclassified into a simpler habitat scheme based on the substrate and biotic components of CMECS as well as visual relief (Figure 7 and Figure 8). The main distinction in The Elbow for substrate was between rock and

sand, although many areas tended to exhibit mixed classes. The CMECS documentation advocates using a 50% threshold to designate which substrate is dominant; however, a forward facing camera can make assessing percent cover difficult, and many areas exhibited rock substrate overlain by a thin sand veneer making it unclear exactly how to assess percent cover. Rather than using a 50% threshold, areas where a thin sand veneer was overlain on rock or where large high-relief features were exposed, were considered to be rock substrate. This is similar to the procedure used by Kingon (2013). Conversely, areas characterized by a few small isolated rocky features or rubble piles within larger expanses of sand were considered to be sand. In addition to these CMECS substrate categories, a modification was added to characterize the relief of rocky substrates according to three relief levels: low (covered to relatively flat exposed rock), moderate (small step like change in elevation or large but gradual change in elevation), and high (large steep change in elevation; Figure 9). The main distinction for the biotic component observable from the C-BASS video that was relevant to this study area was the presence or absence of attached fauna such as sponges and gorgonians (Figure 10). Sea urchin beds were also observed in this study area but could not be reliably identified unless C-BASS was very close to the seafloor. Additionally, benthic macroalgae was occasionally observed, but was not present at a large number of sites, or was simply difficult to detect. As such, the biotic component for this study area was collapsed to simply denoting whether or not attached fauna was present or absent (bare) at a given area .

C-BASS Position

Linking the C-BASS video data with the coinciding multibeam data requires position data for the towed camera system. To accomplish this, the position of the ship was logged using GPS and the distance of the C-BASS system behind the boat (layback) was calculated using the

cable-out from the winch which was logged manually, and the depth of the C-BASS which was collected by the C-BASS' CTD. The layback of the C-BASS system was calculated in Hypack® using the Towfish.DLL program using the "standard" method and a catenary factor of ~.89 (Equation 1; HYPACK, 2017). Using this information as well as the ships' position, heading, and motion, the Towfish.DLL program also estimates the speed and position of C-BASS at the same frequency as ship position (~2Hz). To match this position to the towed video and sensor data, the estimated Easting and Northing position of the C-BASS was then linearly interpolated from a 2Hz to a 1Hz frequency.

Eq 1: $$Layback = \sqrt{(\mathrm{k} * \mathrm{L})^2 - (I + z)^2}$$

L= cable out (m)
k= catenary factor
I= C-BASS Depth (m)
z= A-Frame Offset (m)

Data Analysis

Habitat Maps

Background on Classification Algorithms

Habitat maps were created using both supervised and unsupervised methodologies (Figure 11). The supervised habitat maps were created using a random forest algorithm (Breiman, 2001, Liaw and Wiener, 2002). Random Forests are a machine learning algorithm that is used in many applications including seafloor habitat mapping (Cutler et al., 2007, Stephens and Diesing, 2014, Hasan et al., 2014, Lucieer et al., 2013, Porskamp et al., 2018). The random forest algorithm works much like a traditional decision tree which at each node determines the optimal split in the predictor variables to best separate groups (De'ath and Fabricius, 2000, Breiman, 2001). Rather than simply fitting one decision tree, a "forest" of many decision trees (e.g. hundreds or thousands) are fit to the data with each tree differing in a "random" way as each

tree is fit using a bootstrap sample of the data (rather than the original data), and is only given access to a random subset of predictors at each node rather than all predictors. This creates many decision trees that are all different from one another. Classification of new data is then achieved by running a new data point through each decision tree and then aggregating results of the forest (e.g. by majority vote) to determine group membership of an observation. Like decision trees, random forests are efficient in dealing with many variables and complex non-linear relationships; however, random forests have been shown to be more accurate than traditional decision trees, and are less prone to overfitting of the data, which makes them more generalizable and robust for prediction (Breiman, 2001, Cutler et al., 2007). Moreover, random forests have been found to perform comparably well with other machine learning classifiers such as artificial neural networks, but are more user-friendly in that they only require two main parameters: the number of trees in the forest, and the number of predictors available at each node (Liaw and Wiener, 2002). Two additional benefits of the random forest algorithm include its ability to calculate a variable importance metric and being able to determine the probability of group membership for each new observation which allows for assessment of uncertainty.

The unsupervised model used was k-means clustering (MacQueen, 1967). Given a dataset and an *a priori* number of clusters, the algorithm will assign each point in the dataset to the cluster whose centroid is closest. The location of the centroids is determined by minimizing within cluster heterogeneity relative to other identified clusters in the data set, over several iterations.

Application of Classification Models

The RSToolbox package in R was used to run both the supervised and unsupervised models (Leutner and Horning, 2016, Wegmann et al., 2016). The same predictor variables,

training data, and validation data were used for both methodologies; however separate models were developed for classifying the geologic and biotic aspects of habitat. The predictor variables consist of the various acoustic raster data layers (Table 1, Appendix 1, Table 2, Appendix 2), and the training and validation data sets consist of the ground-truth habitat observations determined from the C-BASS video. Prior to running these models, geologic and biotic habitat were collapsed into binary categories: rock vs sand and attached fauna vs no attached fauna, respectively. Additionally, ground-truth habitat data from C-BASS video transects were split into training and validation sets. To reduce the influence of spatial autocorrelation on the accuracy assessment, one transect was reserved solely for validation (Figure 4). In order to reduce the effect of positional uncertainty of the C-BASS and confusion due to mixed habitats or habitat boundaries, only observations that were the same as their subsequent and previous class were retained in the training and validation set. This filtering of observations was done separately for the geologic and biotic aspects of habitat. Models were assessed using overall accuracy, as well as using Cohen's Kappa (κ), which adjusts the overall accuracy for what could occur by chance (Equation 2 and Equation 3; Cohen, 1960) . κ is equal to one if there is complete agreement between predictions and observations, is zero if the agreement is no greater than what could occur by chance, and is negative if it is less than what could occur by random chance. The performance of models were further assessed using confusion matrices as well as user's and producer's accuracy to see how well the model could predict each habitat type (particularly the rarer class). User's accuracy and producer's accuracy are complimentary assessments of accuracy that help portray a more detailed picture than overall accuracy alone. User's and producer's accuracy were calculated for each habitat class. User's accuracy describes accuracy from the perspective of the map user (e.g. if the map says an area is rock, how likely is that to be

correct?). Producer's accuracy describes accuracy from the perspective of the map producer (e.g. if an area truly is rock, how likely is it that my map correctly predicted that?).

The supervised model (random forest) requires two parameters, the number of trees, and the number of predictors available for the algorithm to search through at each node. To determine the optimal number of predictors available at each node, separate models were run with between two and 16 (total number of predictors - one) predictors per node. This was plotted against the κ based on five-fold cross-validation on the training data. Five-fold cross-validation splits the training data into five random partitions. The model is fit five times, each time leaving out a different partition. Each run, the model is tested on the partition not used to fit the model to calculate κ. The κ values for all five runs are then averaged. The optimal number of predictors available at each node was selected such as the one that maximized the κ, or the value at which the κ began to plateau. The number of trees was selected by plotting the "out of bag" (OOB) error rate against the number of trees. The OOB observations are the observations outside a tree's bootstrap sample. The number of trees vs OOB error rate plot typically resembles an exponential decline, and the value for this parameter was chosen as one that was sufficiently far into the plateau as to minimize error. This fitting was done separately for the geologic and biotic habitat data. These optimal models were then used to predict habitat for the full extent of the study area for both geologic and biotic habitat. Additionally, entropy maps which display the uncertainty of the classification were generated for geologic and biotic habitat using the proportion of trees in the model that voted for each class within a given cell. Entropy was calculated using the Shannon entropy formula with log base e (Equation 4; Shannon, 1948, Wegmann et al., 2016). Moreover, the variable importance of each predictor was determined by randomly permuting the values of that variable in the OOB observations for each tree and

calculating the mean decrease in accuracy that would occur for each variable (Breiman and Cutler, 2008, Strobl and Zeileis, 2008).

For the unsupervised classification, all predictors were z-score normalized to minimize the effects of differing ranges and units of among the various predictors. Then, a Principal Components Analysis (PCA) was conducted on the z-score normalized predictor variables to remove the effect of multi-collinearity by extracting the independent components of the data. To remove noise in these data, only a subset of the original Principal Components (PC's) were retained as each subsequent PC explains a smaller proportion of the variance than the preceding PC. There are several different methods used to determine the "correct" number of PC's to retain (Jackson, 1993). The method used here was to retain only the PC's that explain more than could be expected if the total variance was divided randomly amongst all the PC's as modelled by a broken-stick distribution (Frontier, 1976, Jackson, 1993). The retained PC's were then run through a k-means clustering procedure using the MacQueen algorithm (MacQueen, 1967). The ground-truth habitat points from the training data set were then added in *post-hoc* and used to interpret the statistical clusters. Each cluster was interpreted as a habitat type by assigning it a habitat class based on majority vote of all ground-truth habitat points from the training set contained within that cluster. This was done separately for the geologic and biotic habitat. Different numbers of clusters were tested to find the optimal number of clusters using 5 fold cross-validation. The number of clusters was plotted against the κ, and the optimal number of clusters was chosen as the number of clusters at which the κ was maximized or began to plateau.

After creating biotic and geologic habitat maps using both the supervised and unsupervised methodology, an accuracy assessment was conducted on the validation data. For

each map, a confusion matrix was created, and the overall accuracy, user's accuracy, producer's accuracy, and κ were calculated using the caret package in R (Kuhn, 2008).

Eq 2: $$c = \frac{1}{N^2}\sum_k n_{k\_pred} * n_{k\_obs}$$

c= chance agreement
n= number of observations
k= class number
n_{k_pred}= Number of class k predicted
n_{k_obs}= Number of class k observed

Eq 3: $$\kappa = 1 - \frac{1-a}{1-c}$$

a= overall accuracy
c= accuracy expected by random chance

Eq 4: $$Entropy = -\sum_{i=1}^{M}(p_i * ln(p_i))$$

p_i= probability that a cell is of class i
i= class number
M = number of classes

FWC-FWRI Elbow Map Comparison

The Elbow was previously mapped using a sidescan sonar and drop cameras by the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute (FWC-FWRI; Switzer et al., 2014). The habitat scheme they used is a modified version of the geoform component of CMECS, and delineations between habitats were made manually by visual inspection of the sidescan data, resulting in a vector map (polygons) of geoform habitat. As their map is focused on geologic habitat, I chose to compare it to the best performing of the two geologic habitat maps, which was the one created using the supervised classification. As there is overlap between our two study areas, qualitative and quantitative comparisons between the maps were made to assess the correspondence of habitat types in the two maps. Qualitative comparisons were made by visually assessing general trends between the two maps, and quantitative comparisons were made by extracting the raster values from the supervised geologic

habitat map contained within each one of their polygons to determine the percent of the time each of their habitat classes corresponds to rock and sand (as predicted by the supervised geologic habitat map).

Results

CMECS Summary

Using a combination of video observations, CTD data, and knowledge of the area, the following CMECS summary was compiled to describe the diversity of habitats encountered in the study area. The individual components and settings are bolded, and hierarchical levels are represented by indentation.

- **Biogeographic Setting**
    - Realm - Temperate Northern Atlantic
        - Province - Warm Temperate Northwest Atlantic
            - Ecoregion - Northern Gulf of Mexico
- **Aquatic Setting**
    - System – Marine
        - Subsystem – Marine Offshore
            - Tidal Zone – Marine Offshore Subtidal
- **Water Column Component**
    - Water Column Layer - Marine Offshore Lower Water Column
        - Salinity Regime – Euhaline Water (30 - 40 on Practical Salinity Scale)
        - Temperature Regime – Moderate Water (15ºC - 20ºC)
        - Temperature Regime – Warm Water (20ºC – 25ºC)
- **Geoform Component**
    - Tectonic Setting – Passive Continental Margin
    - Physiographic Setting – Continental/Island Shelf
    - Geoform Origin – Geologic
        - Geoform – Flat
        - Geoform - Ledge
        - Geoform - Ridge
        - Geoform - Ripples
        - Geoform - Rock Outcrop

                - Geoform Type - Authigenic Carbonate Outcrop
        - Geoform Origin – Biogenic
            - Geoform - Burrows/Bioturbation
- **Substrate Component**
    - Substrate Origin – Geologic Substrate
        - Substrate Class - Rock Substrate
        - Substrate Class – Fine Unconsolidated Substrate
            - Substrate Group – Sand
            - Substrate Group – Gravelly
                - Substrate Subgroup- Gravelly Sand
            - Substrate Group – Slightly Gravelly
                - Substrate Subgroup – Slightly Gravelly Sand
    - Substrate Origin – Biogenic Substrate
        - Substrate Class – Organic Substrate
            - Substrate Subclass – Organic Debris
- **Biotic Component**
    - Biotic Setting – Benthic/Attached Biota
        - Biotic Class – Faunal Bed
            - Biotic Subclass – Attached Fauna
                - Biotic Group – Attached Corals
                    - Biotic Community – Attached Gorgonians
                - Biotic Group – Attached Sponges
                - Biotic Group – Diverse Colonizers
                    - Biotic Community – Sponge/Gorgonian Colonizers
            - Biotic Subclass – Soft Sediment Fauna
                - Biotic Group – Sea Urchin Bed
        - Biotic Class – Aquatic Vegetation Bed
            - Biotic Subclass – Benthic Macroalgae

Habitat Maps

Ground-truth Data

Geologic Habitat

Although the substrate component consists of a variety of attributes, for map classification, geologic habitat was collapsed into a binary categorization of rock or sand. At this level of detail, the entire ground-truth dataset consisted of 473 observations of rock substrate, 3024 observations of sand, and 12 observations where habitat was not visible. After removing observations that were not the same as their previous and subsequent observation, observations

where habitat was not visible, and observations beyond the bounds of the trimmed bathymetry and backscatter layers, there were 238 observations of rock and 2531 observations of sand. These data were then split into training and validation sets by keeping transects three, five, and six for model training, and setting aside transect one for model validation. This resulted in a training data set consisting of 210 observations of rock and 1947 observations of sand, and a validation transect consisting of 28 observations of rock and 584 observations of sand (Figure 12).

Biotic Habitat

Biotic habitat attributes were collapsed into the binary categorization of the presence or absence of attached fauna. At this level of detail, the entire ground-truth dataset consisted of 435 observations of attached fauna, 3062 observations that were bare, and 12 observations where habitat was not visible. After removing observations that were not the same as their previous and subsequent observation, observations where habitat was not visible, and observations beyond the bounds of the trimmed bathymetry and backscatter layers, there were 206 observations of attached fauna and 2560 observations that were bare. This data set was then split into a training and validation set by keeping transects three, five, and six for model training, and setting aside transect one for model validation. This resulted in a training data set consisting of 183 observations of attached fauna and 1961 observations that were bare, and a validation transect consisting of 23 observations of attached fauna and 599 observations that were bare (Figure 13).

Unsupervised Classification

Principal Component Analysis

First, each raster data layer was z-score normalized. Then, a PCA was then run on these data to extract the independent components (Table 3, Appendix 5, Appendix 6). PC's were retained only if they explained more than could be expected if the total variance was divided randomly

amongst all the PC's as modelled by a broken-stick distribution (Frontier, 1976, Jackson, 1993). This led to the first four PC's being retained (Figure 14).

Geologic Habitat

A k-means clustering was run on the four retained PC layers. The model was run with between two and 12 clusters. More than 12 clusters led to some clusters not being able to be interpreted as some of the cross-validation sets did not have enough ground-truth points to assign a class to every cluster. The κ from five-fold cross validation was plotted against the number of clusters (Figure 15). Based on this plot it appears that beyond 10 clusters there is no improvement in κ, so 10 was chosen as the optimal number of clusters. The model was then run with 10 clusters (Figure 16), and trained using the entire training set of ground-truth points, with clusters being assigned a class by majority vote of the points within that cluster. The resulting map can be seen in Figure 17. This habitat map predicts that the study area consists of approximately 2.8 km^2 of rock and 84.2 km^2 of sand. The performance of this model was then assessed using the validation data set. The overall accuracy of this model on the validation data was 96.4% and the κ was 0.48. The confusion matrix can be seen in Table 4. The user's accuracy for rock was 68.8% and the producer's accuracy for rock is 39.3%. The user's accuracy for sand is 97.2% and the producer's accuracy for sand is 99.1%.

Biotic Habitat

A k-means clustering was run on the four retained PC layers. The model was run with two and 12 clusters. More than 12 clusters led to some clusters not being able to be interpreted as some of the cross-validation sets did not have enough ground-truth points to assign a class to every cluster. The κ from five-fold cross-validation was plotted against the number of clusters (Figure 18). Based on this plots it appears that beyond 10 clusters there is no improvement in κ,

so 10 was chosen as the optimal number of clusters. The model was then run with 10 clusters (Figure 16), and trained using the entire training set. The resulting maps can be seen in Figure 19. This habitat map predicts that the study area consists of approximately 2.8 km^2 of areas with attached fauna and 84.2 km^2 of bare habitat. The performance of this model was then assessed using the validation data set. The overall accuracy of this model on the validation data was 97% and the κ was 0.5. The confusion matrix can be seen in Table 5. The user's accuracy for attached fauna is 62.5% and the producer's accuracy for attached fauna is 43.5%. The user's accuracy for bare habitats is 97.9% and the producer's accuracy for bare habitats is 99%.

Supervised Classification

Geologic Habitat

To fit the random forest model, the classification was run using two – 16 predictors available at each node, and 5000 trees. The κ based on 5 fold cross-validation was plotted against the number of predictors available at each node (Figure 20). A value of two was chosen for the number of predictors available at each node as this maximized the κ. The number of trees in the forest was plotted against the OOB error (Figure 21). A value of 2500 trees was chosen as this was well into the plateau where error is minimized. The optimal model was then run using these values, and then used to create habitat predictions for the entire study area. The resulting habitat map can be seen in Figure 22. This habitat map predicts that the study area consists of approximately 3.5 km^2 of rock and 83.6 km^2 of sand. Performance was assessed using the validation data. On the validation score the model had a 96.9% overall accuracy and a κ of 0.66. The confusion matrix for the validation data is shown in Table 6. The user's accuracy for rock is 64.5% and the producer's accuracy for rock is 71.4%. The user's accuracy for sand is 98.6% and the producer's accuracy for sand is 98.1%. The entropy map can be seen in Figure 23. The

variable importance plot can be seen in Figure 24. The three most important predictors were the terrain ruggedness index, slope, and the surface area to planar area. The GLCM mean, backscatter, GLCM variance, topographic position index and profile curvature are the next most important variables. GLCM dissimilarity, GLCM contrast, and eastness have values that are just slightly positive showing that they provided a small benefit to the model. Planform curvature, northness, GLCM homogeneity, bathymetry, GLCM Entropy, and GLCM Angular Second Moment had negative values indicating that their inclusion did not provide benefits to the model.

Biotic Habitat

To fit the model, the classification was run using two – 16 predictors available at each node, and 5000 trees. The κ based on 5 fold cross-validation was plotted against the number of predictors available at each node (Figure 25). A value of three was chosen for the number of predictors available at each node as this maximized the κ. The number of trees in the forest was plotted against the OOB error (Figure 26). A value of 2000 trees was chosen as this was well into the plateau where error is minimized. The optimal model was then run using these values, and used to create habitat predictions for the entire study area. The resulting habitat map can be seen in Figure 27. This habitat map predicts that the study area consists of approximately 3.2 km^2 of areas with attached fauna and 83.9 km^2 of bare habitat. Performance was assessed using the validation data. On the validation score the model had a 97.3% overall accuracy and a κ of 0.67. The confusion matrix for the validation data is shown in Table 7. The user's accuracy for attached fauna is 60% and the producer's accuracy for attached fauna is 78.3%. The user's accuracy for bare habitats is 99.2% and the producer's accuracy for bare habitats is 98%. The entropy map can be seen in Figure 28. The variable importance plot can be seen in Figure 29. The three most important predictors were the terrain ruggedness index, slope, and the surface

area to planar area. GLCM mean, topographic position index, GLCM variance, profile curvature, backscatter, eastness, and GLCM contrast are the next most important variables. The rest of the variables have values that are just slightly positive showing that they provided a small benefit to the model.

FWC-FWRI Map Comparison

The geoform habitat map created by FWC-FWRI can be seen in Figure 30 (Switzer et al., 2014). The map has been trimmed to the extent of the study area, and areas not delineated as any habitat type are labelled as "sand" for clarity as the FWC-FWRI map only mapped hard-bottom features. The main hard-bottom features in their map are low relief hard-bottom, ledge, and mixed hard bottom. As FWC-FWRI mapped geologic habitat, I compared it to the geologic habitat map created using the supervised methodology as that had higher performance than the unsupervised map.

Qualitatively, both maps have labelled the main north-south ridge as a hard-bottom feature. In their map this is labelled as a ledge feature. The secondary ridge to the west is also picked up in both maps as hard-bottom, with their map calling most of it low relief hard-bottom, and smaller sections being considered mixed hard-bottom or ledge. Although the ridge features are both labeled as hard-bottom features in both maps, their map predicts more extensive areas of low relief hard bottom, while our map predicts many of those areas to be sand.

Quantitatively, the correspondence of the FWC-FWRI map with the supervised geologic habitat map was assessed by calculating the percent of rock vs sand (as predicted by the supervised geologic habitat map) contained within their polygons of a given class (Figure 31).

When FWC-FWRI labelled a habitat as ledges or boulder fields, that largely corresponded to rock substrate in the map from this study. Low relief hard-bottom however often

corresponded to sand habitat in the map from this study. Mixed hard-bottom and fragmented hard-bottom both corresponded to sand habitat approximately half the time and rock habitat the other half of the time. Potholes corresponded to sand habitats 100% of the time. Dredge deposits and unknown habitats also corresponded to sand habitats 100% of the time but were very rarely observed.

Discussion

Through visual examination of the geologic and biotic habitat maps, it is clear that the supervised and unsupervised procedures produce maps with the same general trends. Overall, both maps identified the main long rocky ridge that runs north to south, as well as a smaller ridge to the west, with both ridges appearing to have attached fauna across most of their extent. Both maps also reveal several areas where rock seems to be scattered throughout sandy areas, as well as some small isolated outcrops. Although the results are broadly similar, there are more subtle differences in the predicted maps from the two methodologies. For example, the unsupervised classification predicted a lower total area for the rarer habitats (rock/attached fauna) as compared to the supervised classification. This relates to the low producer's accuracy of the unsupervised classification which means that there were high errors of omission for rock and attached fauna indicating that the unsupervised classification is likely underestimating the true area of those habitats. Another difference is that in the supervised classification maps there appear to be thin stripes of rock/attached fauna while these do not appear in the unsupervised map. These stripes are likely not real features, and indicate that the supervised classification was more sensitive than the unsupervised classification to along-track artifacts in the multibeam data. This is likely because the artifacts have similar properties to a rocky ridge (linear pattern and vertical offset) which may be confusing the supervised classification. The unsupervised classification was less

sensitive to the artifacts and this is likely related to that it was based on a subset of the principal components which removes a lot of the noise in the data and/or that the merging of clusters smoothed out the effects of the artifacts in the final map.

In addition to the two methodologies producing similar maps, the biotic and geologic habitat maps show similar habitat boundaries, and for the unsupervised methodology, the geologic and biotic maps are actually identical. This occurs because attached fauna rely on a hard substrate to attach to so these two habitats tend to co-occur. In fact, within the 3 transects used for the training data set, 88% of observed rocky habitats also had attached fauna.

All of the maps showed very high accuracy (>96%); however, overall accuracy can be a misleading metric when classes are unbalanced as preference for simply guessing the majority class can lead to high accuracies. For example if 90% of observations are of class one, and 10% of observations are of class two, simply always guessing class one would lead to an overall accuracy 90%. This is why all maps were additionally assessed in terms of the user's and producer's accuracy as well as in terms of the κ which accounts for the agreement that could occur by random chance. All maps had very high user's and producer's accuracies for the majority class. However, for the rarer class, the unsupervised classification for both maps showed a slightly higher user's accuracy than the supervised maps, but had a much lower producer's accuracy. This means that the unsupervised classification had high errors of omission for predictions about the rarer class. The supervised classification maps showed intermediate levels of both user's and producer's accuracy for the rarer class, which indicates less of a preference for simply guessing the majority class. There are tradeoffs between overall accuracy, and the user's and producer's accuracy for each class; however, by examination of the κ we can see that overall, the supervised methodology performed better in both cases than the

unsupervised methodology for both geologic and biotic habitat. In both the supervised classification maps $\kappa > 0.6$ indicating "substantial agreement" between predictions and observations, while in both the unsupervised classification maps $\kappa > 0.4$ indicating "moderate agreement" between predictions and observations (Landis and Koch, 1977).

In addition to higher performance, the supervised map also had the added benefit of providing measures of variable importance and a measure of the uncertainty in the classification over space since the random forest algorithm was used. Uncertainty can be assessed with the entropy maps. These maps allow for an assessment of the confidence of the associated classification in a given area based on the number of trees that voted for each class in a given cell (e.g. 90% of trees voted that a cell is sand and 10% voted it is rock). In the maps produced in this study, there is more uncertainty in areas of the minority class (rock and attached fauna), as well as in a few other areas. These other areas may represent differing morphologies of hard-bottom than the one the model was trained on, mixed classes, areas with gravel or debris, or entirely new habitats that were not observed in the video transects. Future sampling efforts can be dedicated towards collecting more ground-truth observations in areas of greater uncertainty in order to improve the map over time. For both supervised maps the most important variables were slope and two different measures of terrain variability, all of which are derived from bathymetry. Moreover, the fourth most important variable for both supervised maps was the GLCM mean, and removing backscatter or one of its derivatives usually resulted in a decrease in accuracy, demonstrating that including features derived from both bathymetry and backscatter can improve classification accuracy. This finding is consistent with other studies (Ierodiaconou et al., 2007, Lucieer et al., 2013, Hasan et al., 2014, Ierodiaconou et al., 2018). Surprisingly, neither bathymetry or backscatter themselves were among the most important variables, and variable

importance for bathymetry was actually negative for the geologic habitat indicating it proved no benefit to the model, which contrasts with several other studies have found (Lucieer et al., 2013, Diesing et al., 2014, Hasan et al., 2014, Porskamp et al., 2018). This may be related to the scale (cell size and window size) at which the metrics were calculated, the thematic resolution of the habitat maps, or may be related to the characteristics of this specific environment. For example, in a more complex environment that had several different morphologies of hard-bottom, several sediment types, and various biotic habitats, the importance of each variable may be different.

In addition to comparing the different maps produced in this study, I compared the supervised geologic habitat map made in this study to the one previously created by FWC-FWRI through manual delineation of a sidescan mosaic. There is some correspondence between the two maps with both identifying two linear hard-bottom areas (the main and smaller ridge). Additionally, what they labeled as ledge or boulder field largely corresponded to rock in my map. Potholes are a micro-habitat which are too small to be captured in my map, and corresponded to sand. Mixed habitats such as mixed hard-bottom and fragmented hard-bottom corresponded to rock about half the time and sand the other half of the time which makes sense as these habitats are composed of both sand and rock. One major difference was low relief hard-bottom which largely corresponded to sand, leading them to predict a much greater area of hard-bottom. Although these areas were generally predicted as sand in my map, the entropy map shows that there is higher uncertainty in the classification in these areas. More ground-truth samples should be taken in this area to improve the classification in order to increase the confidence of these classifications. The higher uncertainty might be a result of the presence of mixed habitats, differing morphology of hard-bottom, or the presence sand intermixed with gravel or debris. There are several advantages and disadvantages associated with each map. The

FWC-FWRI map has more classes, and the use of manual delineation of high resolution sidescan data allowed for mapping of small scale features such as potholes. This approach however is subjective and can be time consuming. The supervised map was done in a more automated and objective manner, and estimates of uncertainty over space can be analyzed to improve the map over time. Additionally, with standardization of data collection and analysis protocols, approaches such as this can be extended to automatically classify other areas where multibeam data is collected, making it more scalable to regional level mapping initiatives and the associated large volumes of data. This method however is limited by the positional accuracy of the camera system which can make it difficult to predict small scale features, and the need for sufficient sample sizes can hinder the delineation of rarer classes.

Lastly, although the distribution of rock vs sand, and attached fauna vs bare habitats were mapped, there were also other habitats present and can be seen in the CMECS summary. For example, sea urchin beds and macro-algae were among the other biotic habitats, and several types of attached fauna were observed in video imagery. Additionally, burrows and small mounds of debris created by Sand Tilefish (*Malacanthus plumieri*) were often found. These mounds may be an important micro-habitat as they provide a unique habitat within an otherwise sandy area that can be utilized by other benthic organisms including fish and invertebrates (Büttner, 1996). Other habitats present in the summary include larger scale geoforms such as "ridge." The camera system proved effective for identifying substrate and smaller scale geoforms such as burrows. Large scale geoforms however can be difficult to identify as the video may be too "zoomed in" to see the broader context. Inspection of the habitat maps and the bathymetry shows a long linear rocky feature making the identification of this feature as a ridge much more apparent. This underscores the significance of combining information from both data sources as

neither the video or the multibeam provide the full picture, but when the two are combined there is a lot we can learn about the habitat of an area.

Conclusion

Both supervised and unsupervised methodologies provided broadly similar habitat maps; however, the supervised classification methodology outperformed the unsupervised classification methodology, as the results of the supervised classification demonstrated "substantial agreement" ($\kappa>0.6$) between observations and predictions for both geologic and biotic habitat, while the results of the unsupervised classification demonstrated "moderate agreement" ($\kappa>0.4$) between observations and predictions. These statistical classifiers were able to distinguish between areas of rock and sand, and between areas with attached fauna, and areas without attached fauna. In addition to higher performance, the random forest algorithm which was used for the supervised classification provides additional advantages of being able to measure variable importance as well as uncertainty in the classification over space. However, the unsupervised classification maps appeared to be less affected by artifacts in the multibeam data than the supervised classification maps. Comparisons with the map produced by FWC-FWRI demonstrate some correspondence with their ledge and boulder field habitats corresponding well with rock habitat identified in this study; however, they predicted much more extensive areas of low relief hard-bottom which in this study were predicted to be sand in most cases.

Future Work

Future work should focus on applying this methodology to mapping other areas along the WFS that the C-SCAMP project has collected data for, as well as improving upon this methodology. Moreover, sediment grabs, subsurface data, as well as more video transects have

been collected for this study area and could be integrated to improve the habitat maps for this area.

The field of automated habitat mapping is relatively new and there is little agreement in optimal protocols for determining inputs or statistical methods (Lecours, 2017, Lecours et al., 2017). That being said, explicit consideration of multiple scales is increasingly being recognized as important, and the methodologies here can be expanded to produce maps over a range of spatial, analytical, and thematic scales (Wilson et al., 2007, Lecours et al., 2015, Porskamp et al., 2018). Additionally, many of the classifiers used in habitat mapping are not spatially explicit; however improvements to tools and methods are being made to better account for the spatial nature of the data (Hengl et al., 2007, Hengl et al., 2018). Moreover, object based approaches have recently been applied in the seafloor mapping field (Lucieer, 2008, Bas, 2016, Diesing, 2016, Ierodiaconou et al., 2018). This is because with increasing resolution of new sonars, the cells are now often smaller than the objects of interest (Blaschke, 2010, Diesing, 2016). Object based approaches group cells that are similar to one another into objects which can provide benefits by allowing for the use of information of high resolution data, while taking into account the surrounding context, reducing the impact of noise and positional uncertainty of ground-truth observations, and facilitating multi-scale mapping (Burnett and Blaschke, 2003, Blaschke, 2010). Potential improvements to C-SCAMP surveys protocols can also increase the quality of maps. Improvements have already been made by adjusting sonar settings and implementing automatic recording of cable out when towing the C-BASS. Future improvements however could include using acoustic tracking of the C-BASS and adding a downward facing camera. These enhancements would improve the positional accuracy of ground-truth observations, allow for quantitative measures of percent cover of various habitat features, the creation of georeferenced

photomosaics, and would aid in utilizing finer resolution multibeam data (rather than resampling to 10 m resolution). This would improve the ability to map micro and mixed habitats, and aid in the consideration of habitat heterogeneity at multiple scales, which is increasingly being realized as relevant for relating these maps to our understanding of ecology (Pittman, 2013).

Lastly, although binary classification maps were used in this study the methods used here (e.g. random forest, k-means, confusion matrices, the various accuracy and performance metrics, and entropy maps) are all directly transferable to classifications with more than two categories. The binary classification was used here in order to have sufficient sample size; however, with greater sample size and positional accuracy it would be possible to map rarer and smaller habitat areas in order to provide full coverage maps with more detailed thematic and spatial resolution.

Table 1: Terrain attributes derived from the 10 m x 10 m bathymetry surface. All terrain attributes were calculated using a 3 cell x 3 cell moving window, and the resulting surfaces have 10 m x 10 m resolution.

Feature	Description	Software
Planform Curvature	Curvature perpendicular to the direction of maximum slope	ArcGIS Benthic Terrain Modeler
Profile Curvature	Curvature parallel to the direction of maximum slope	ArcGIS Benthic Terrain Modeler
Eastness	$\sin(aspect)$	Raster R Package
Northness	$\cos(aspect)$	Raster R Package
Slope	Measure of the rate of change in bathymetry. The Horn 1981 algorithm is used (Horn, 1981)	Raster R Package
Topographic Position Index	Indicates whether a location is a local high or low	Raster R Package
Ratio of Surface to Planar Area (planar area corrected for slope)	Measure of terrain variability using the surface area to planar area, also known as rugosity. This implementation decouples the metric from slope by correcting the planar area for local slope.	ArcGIS Benthic Terrain Modeler
Terrain Ruggedness Index	Measure of terrain variability that examines variation in bathymetry around a central cell	Raster R Package

Table 2: GLCM texture attributes derived from the 10 m x 10 m backscatter mosaic. All texture metrics were calculated using a 3 cell x 3 cell moving window and 32 gray levels. The resulting surfaces have 10 m x 10 m resolution. Formulas for texture metrics are from Hall-Beyer (2017).

N = Number of rows or columns in GLCM (Equal to the number of gray levels)
i = row indices of the GLCM matrix (equal to grey level of reference cell)
j = column indices of the GLCM matrix (equal to gray level of neighboring cell)
$P_{i,j}$ = Probability (relative frequency) of neighboring cells having gray levels i & j
μ_i = GLCM Mean

Feature	Description	Software
GLCM Mean	$\sum_{i,j=0}^{N-1} i(P_{i,j})$	glcm R package
GLCM Variance	$\sum_{i,j=0}^{N-1} P_{i,j}(1-\mu_i)^2$	glcm R package
GLCM Homogeneity	$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2}$	glcm R package
GLCM Contrast	$\sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2$	glcm R package
GLCM Dissimilarity	$\sum_{i,j=0}^{N-1} P_{i,j}\lvert i-j\rvert$	glcm R package
GLCM Entropy	$\sum_{i,j=0}^{N-1} P_{i,j}(-\ln(P_{i,j}))$	glcm R package
GLCM Angular Second Moment	$\sum_{i,j=0}^{N-1} P_{i,j}^2$	glcm R package

Table 3: Table of principle components of the raster layers for statistical habitat classification models, with the variation explained by each component, and the cumulative variation explained by that component and all previous components

Principal Component	Percent Variation Explained	Cumulative Percent Variation Explained
1	28.98	28.98
2	20.67	49.65
3	15.68	65.33
4	10.76	76.09
5	6.52	82.61
6	5.28	87.89
7	3.42	91.31
8	2.96	94.27
9	2.46	96.73
10	2.06	98.79
11	0.57	99.36
12	0.18	99.54
13	0.16	99.7
14	0.13	99.84
15	0.12	99.96
16	0.03	99.99
17	0.01	100

Table 4: Confusion matrix along with user's (row-wise) accuracy, producer's (column-wise) accuracy, overall accuracy, and κ for the unsupervised geologic habitat map

	Rock	Sand	User's Accuracy
Rock	11	5	68.8%
Sand	17	579	97.2%
Producer's Accuracy	39.3%	99.1%	**Overall Accuracy = 96.4%** **κ = 0.48**

Table 5: Confusion matrix along with user's (row-wise) accuracy, producer's (column-wise) accuracy, overall accuracy, and κ for the unsupervised biotic habitat map

	Attached Fauna	Bare	User's Accuracy
Attached Fauna	10	6	62.5%
Bare	13	593	97.9%
Producer's Accuracy	43.5%	99.0%	**Overall Accuracy = 97.0%** **κ = 0.50**

Table 6: Confusion matrix along with user's (row-wise) accuracy, producer's (column-wise) accuracy, overall accuracy, and κ for the supervised geologic habitat map

	Rock	Sand	User's Accuracy
Rock	20	11	64.5%
Sand	8	573	98.6%
Producer's Accuracy	71.4%	98.1%	**Overall Accuracy = 96.9%** **κ = 0.66**

Table 7: Confusion matrix along with user's (row-wise) accuracy, producer's (column-wise) accuracy, overall accuracy, and κ for the supervised biotic habitat map

	Attached Fauna	Bare	User's Accuracy
Attached Fauna	18	12	60.0%
Bare	5	587	99.2%
Producer's Accuracy	78.3%	98.0%	**Overall Accuracy = 97.3%** **κ = 0.67**

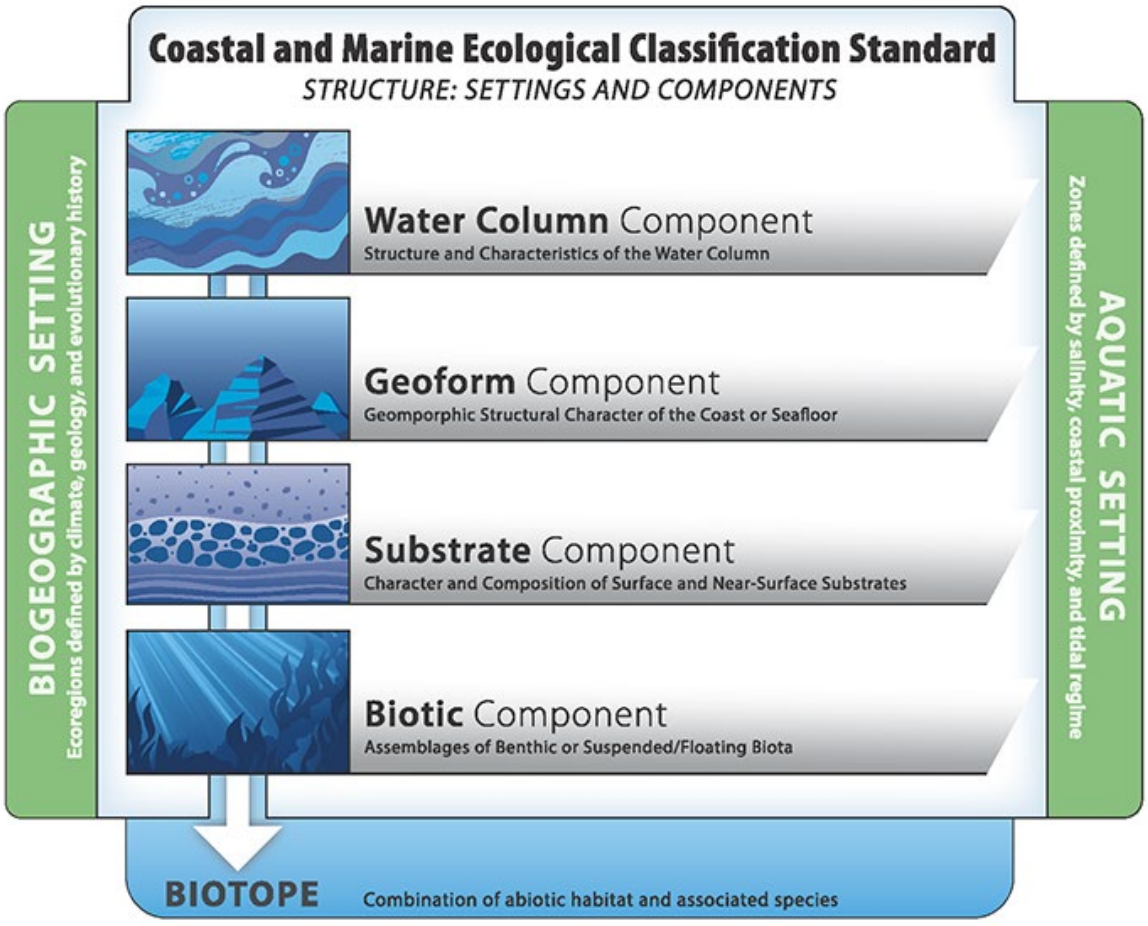


Figure 5: Overall structure of the Coastal and Marine Ecological Classification Standard (Retrieved from: https://iocm.noaa.gov/cmecs/)

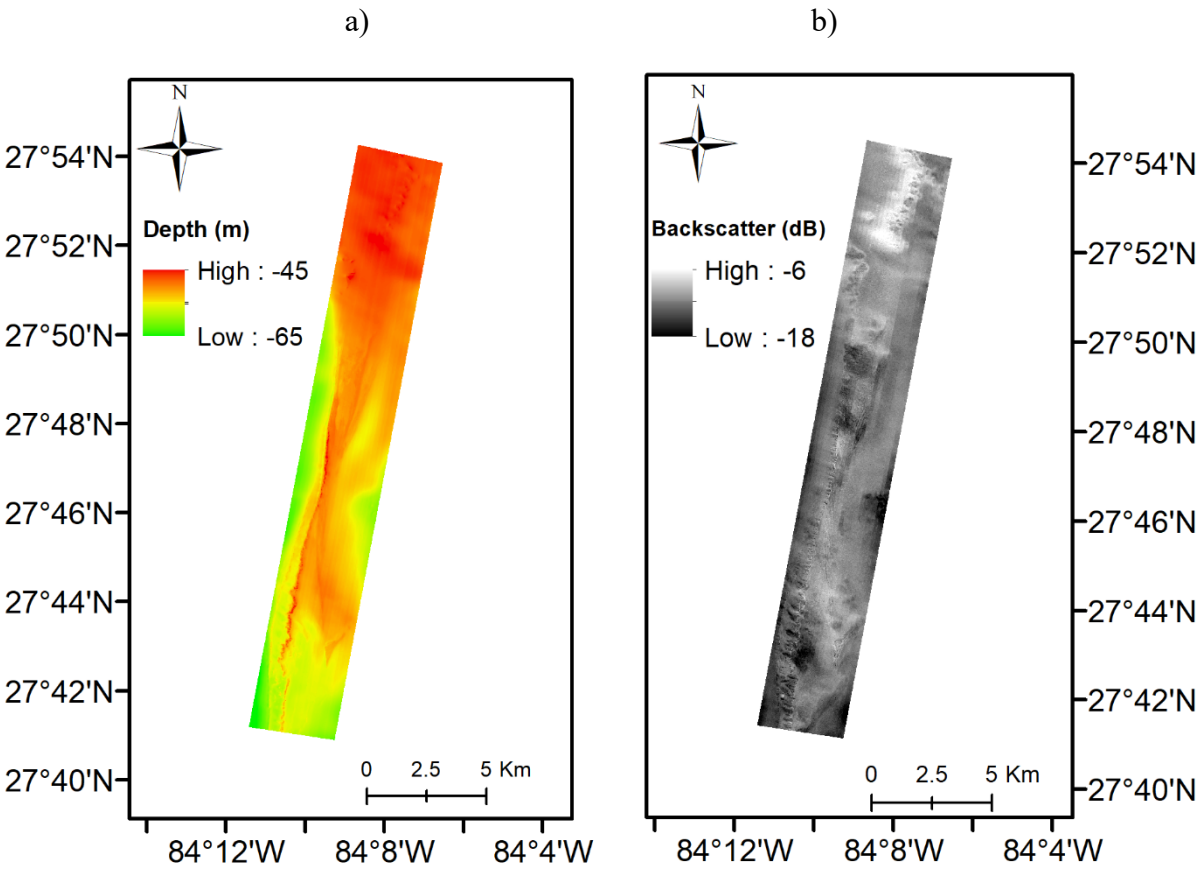


Figure 6: Spatially aligned 10 m x 10 m bathymetry (a) and backscatter (b) surfaces of The Elbow

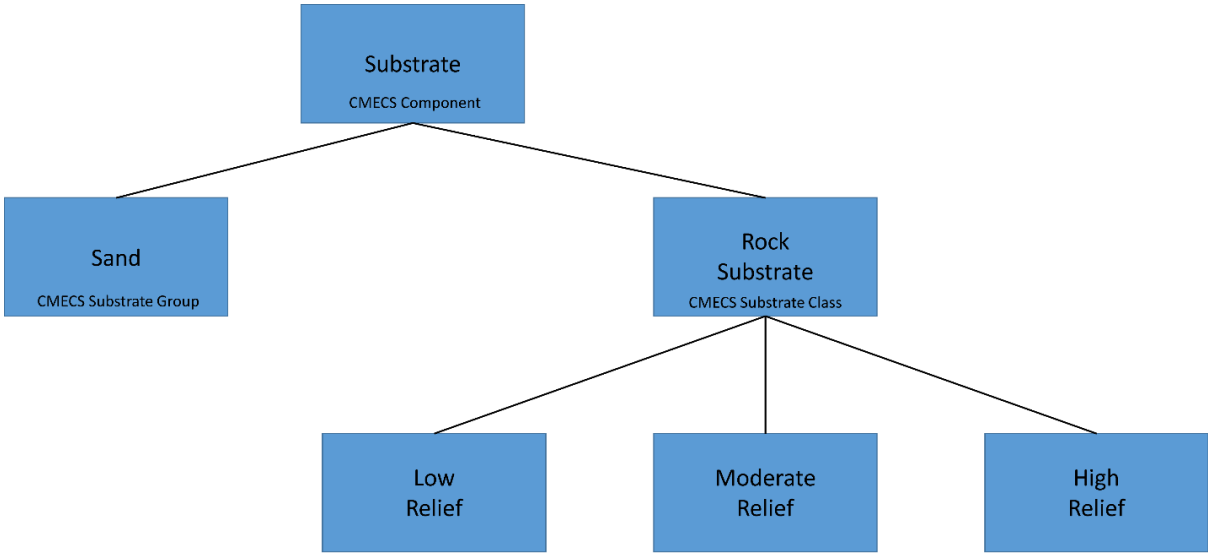


Figure 7: Flowchart representing the simplified substrate classification scheme for the main distinctions found in The Elbow. For classifications within the CMECS scheme, the level within the hierarchy is shown.

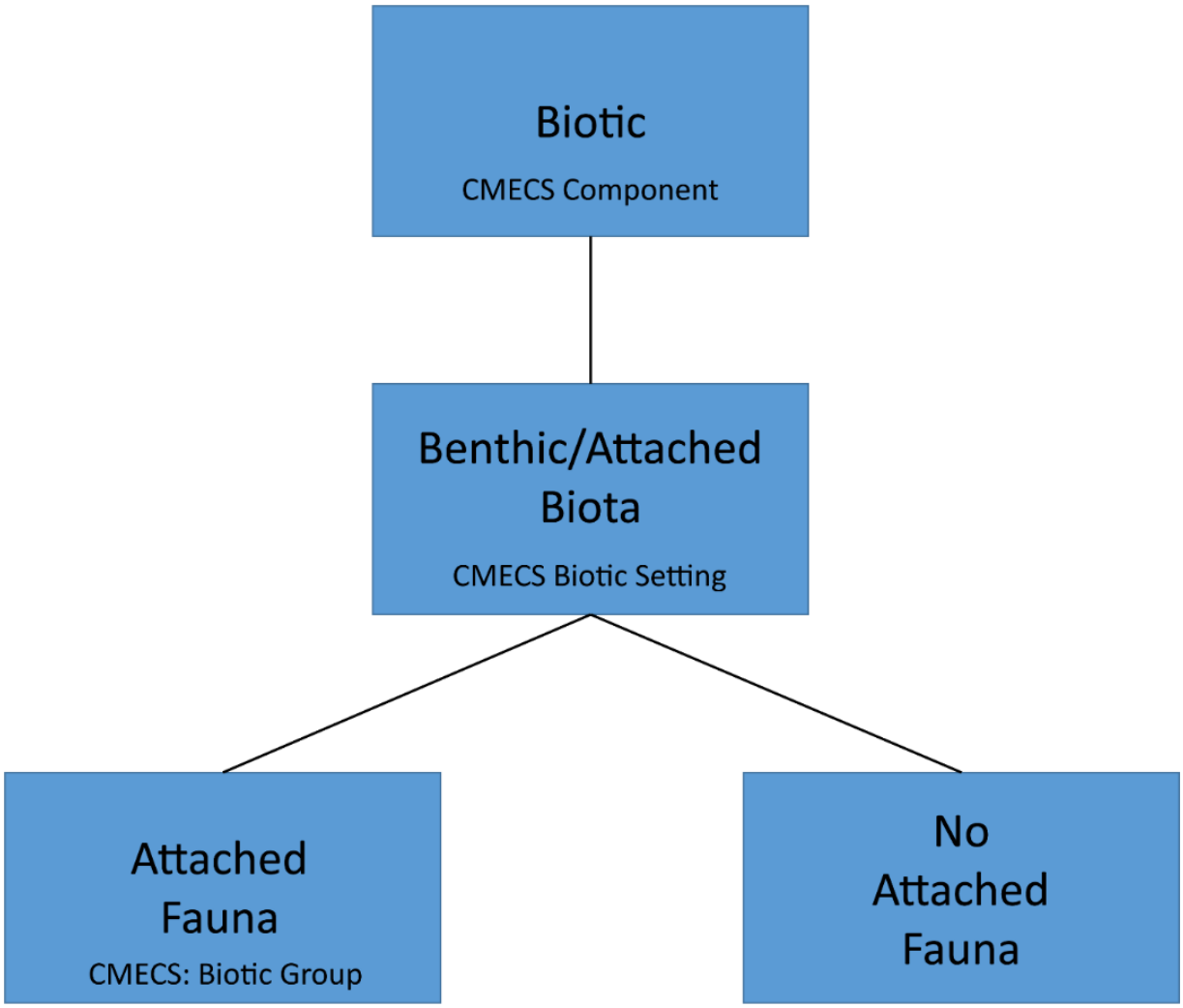


Figure 8: Flowchart representing the simplified biotic classification scheme for the main distinctions found in The Elbow. For classifications within the CMECS scheme, the level within the hierarchy is shown.

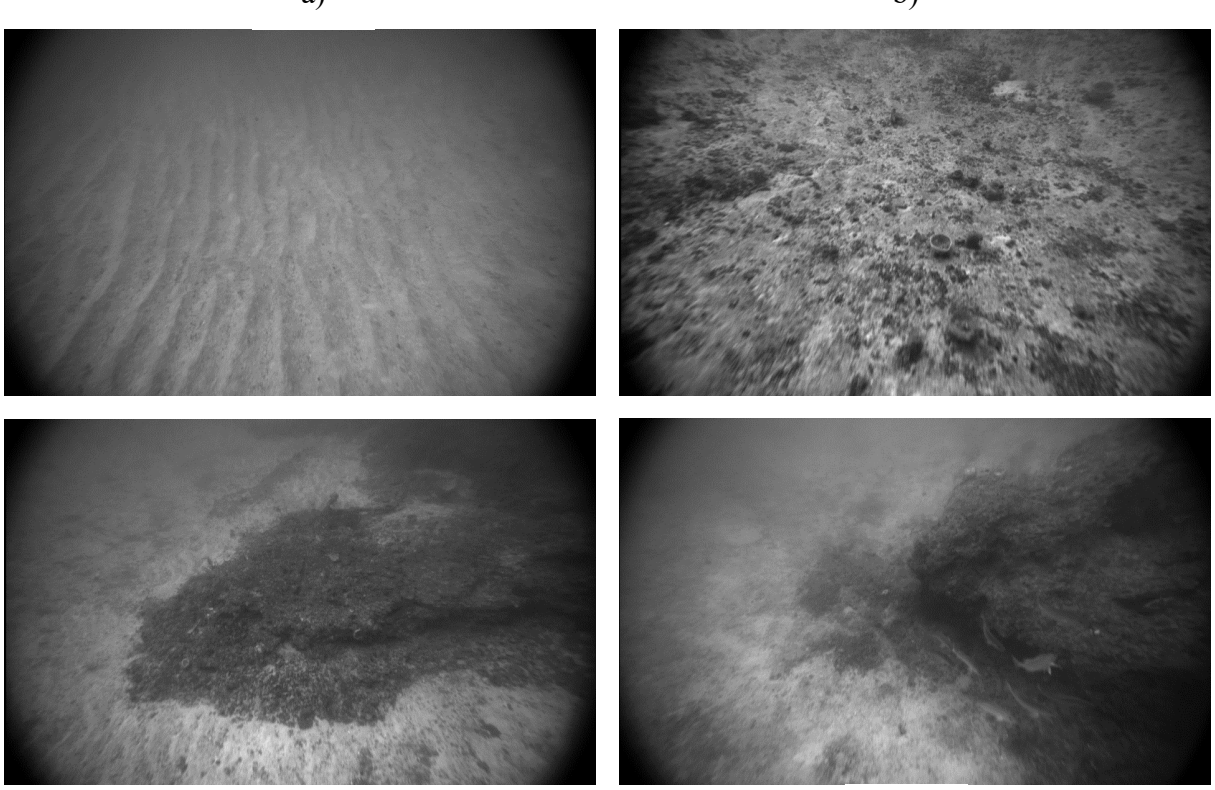


Figure 9: Examples of the main substrate types observed in The Elbow: sand (a), low relief rock (b), moderate relief rock (c), high relief rock (d)

Figure 10: Example of attached fauna. The presence or absence of attached fauna was the main distinction for biotic habitat observed in The Elbow

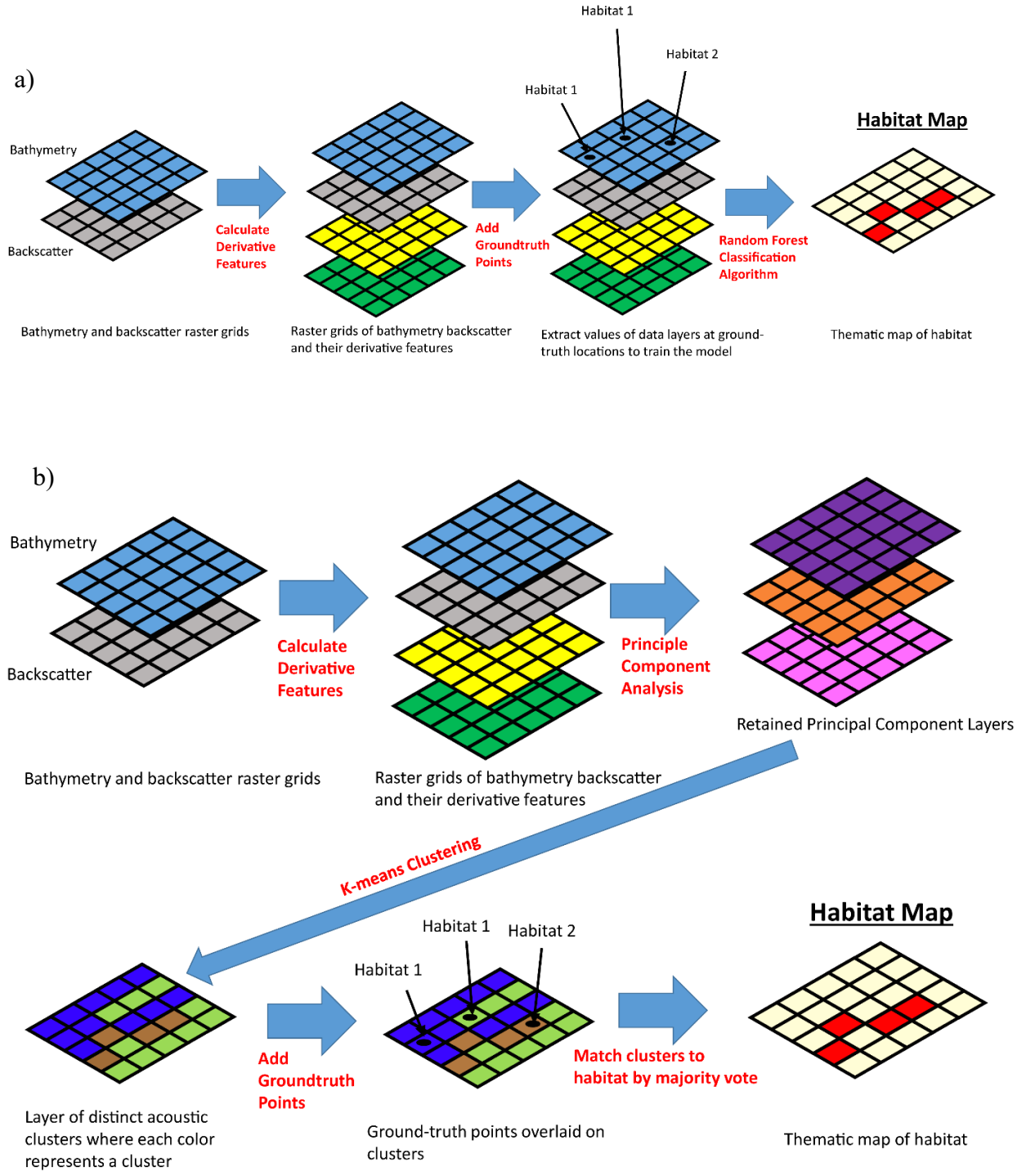


Figure 11: Graphical representation of the supervised (a) and unsupervised (b) classification models for creating predicted habitat maps

a)

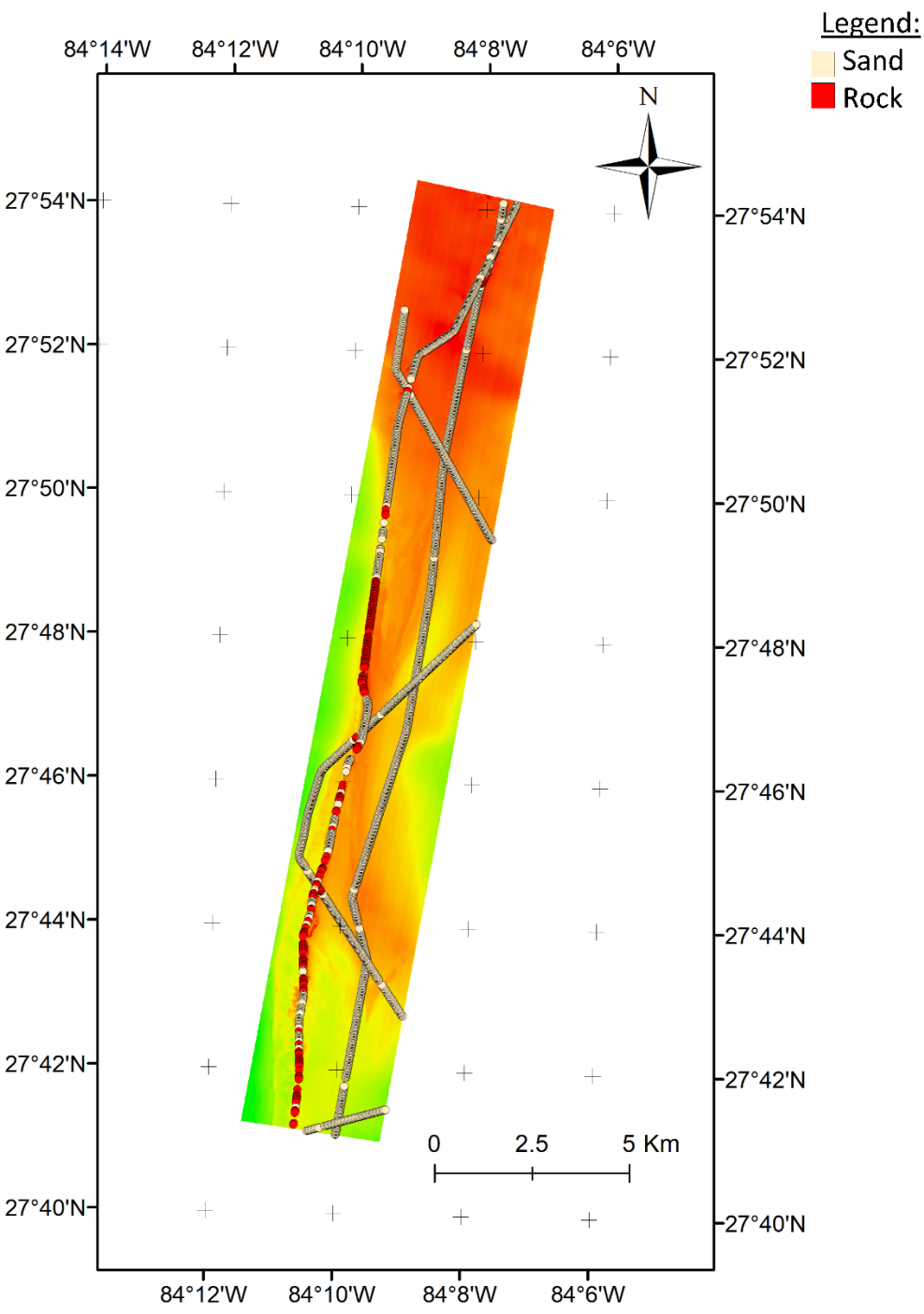

Legend:
Sand
Rock
N
84°14'W
84°12'W
84°10'W
84°8'W
84°6'W
27°54'N
27°52'N
27°50'N
27°48'N
27°46'N
27°44'N
27°42'N
27°40'N
0
2.5
5 Km

Figure 12a

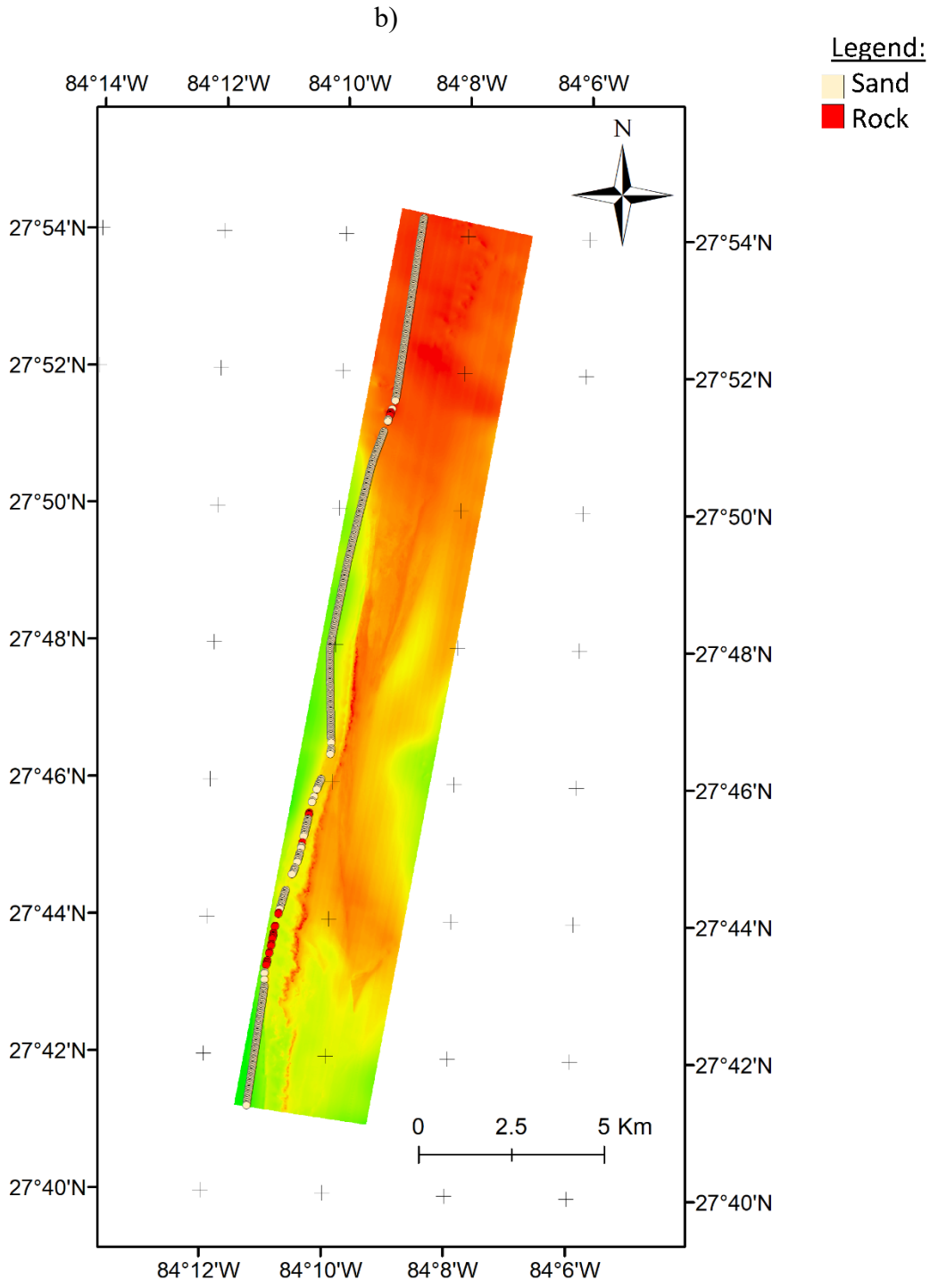


Figure 12: Ground-truth observations of geologic habitat from the C-BASS towed video for training (a) and validation (b) transects overlaid on the 10 m x 10 m bathymetry surface

a)

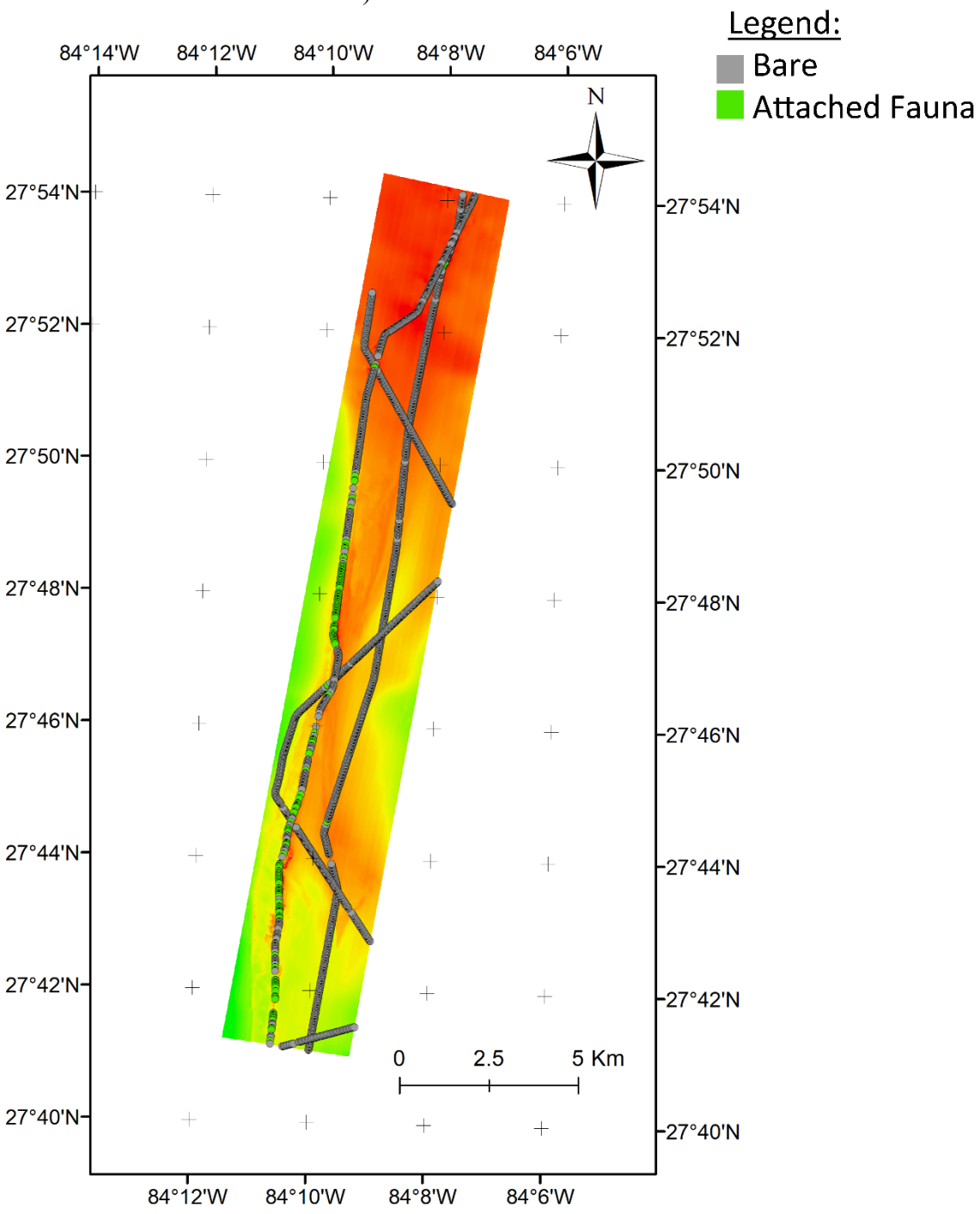

Legend:
Bare
Attached Fauna
84°14'W
84°12'W
84°10'W
84°8'W
84°6'W
N
27°54'N
27°52'N
27°50'N
27°48'N
27°46'N
27°44'N
27°42'N
27°40'N
0
2.5
5 Km

Figure 13a

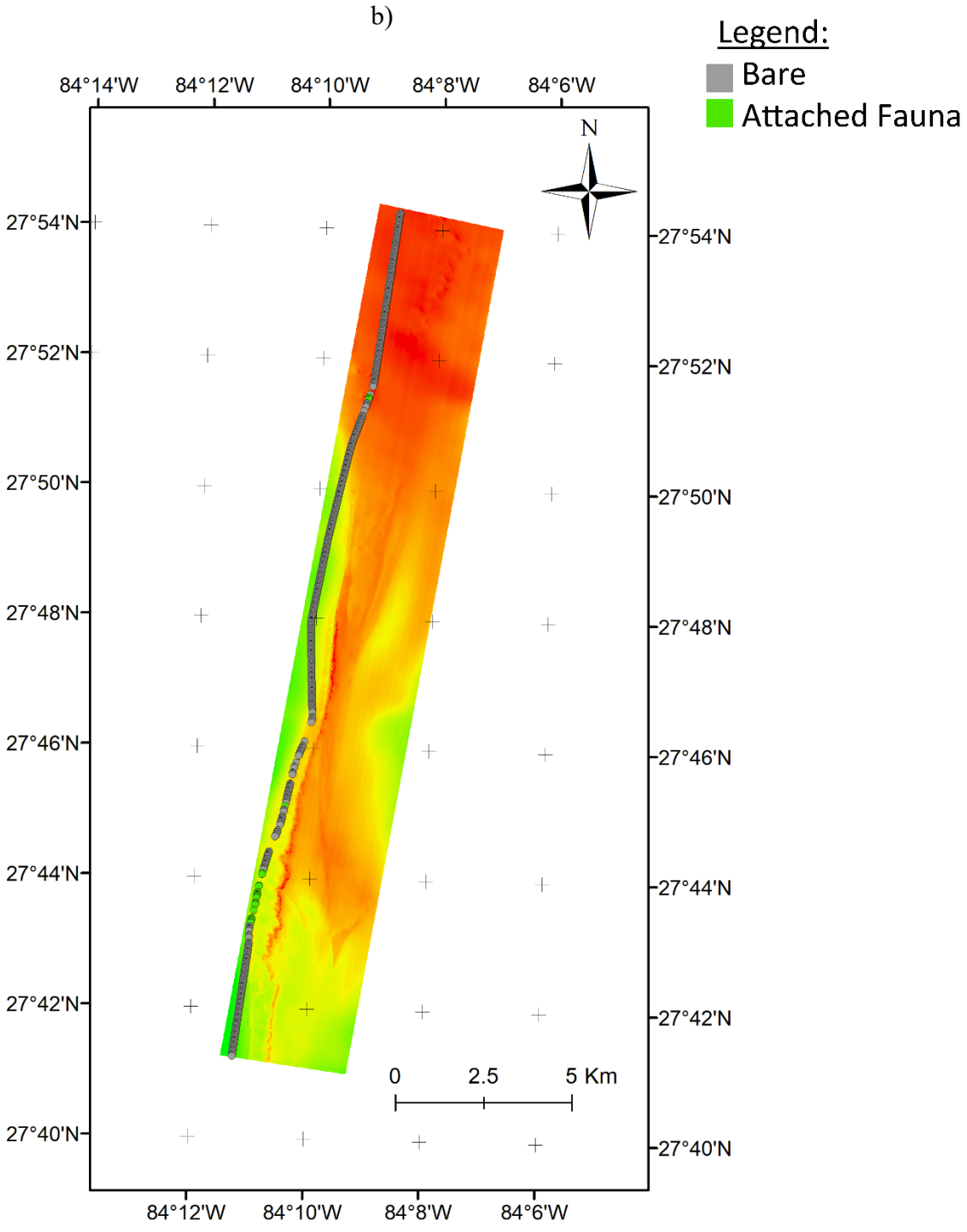


Figure 13: Groundtruth observations of biotic habitat from the C-BASS towed video for training (a) and validation (b) transects overlaid on the 10 m x 10 m bathymetry surface

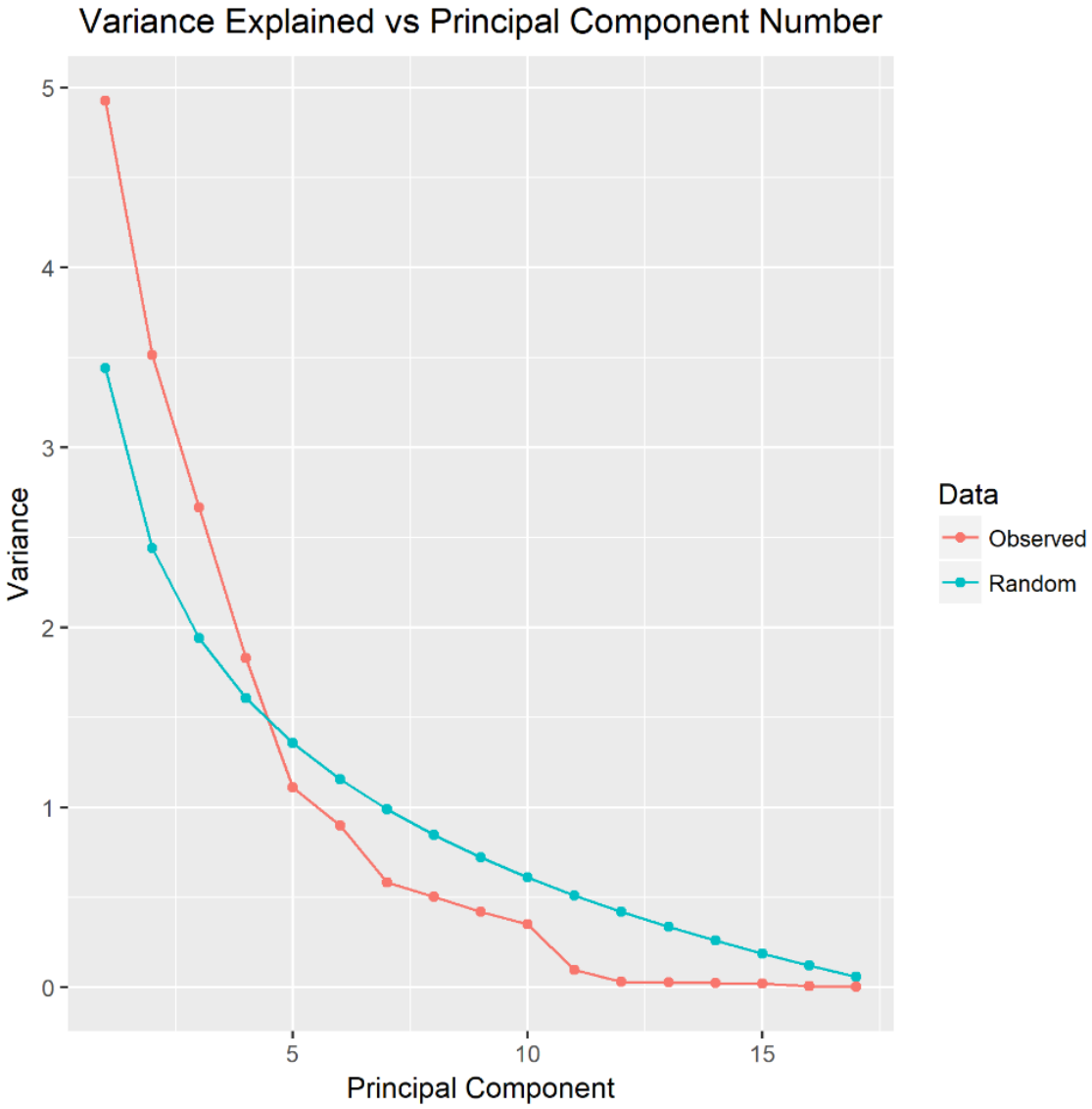


Figure 14: Plot of variance vs principal component for the observed and random data as modelled by a broken-stick distribution. This plot demonstrated that the first four Principal Components should be retained.

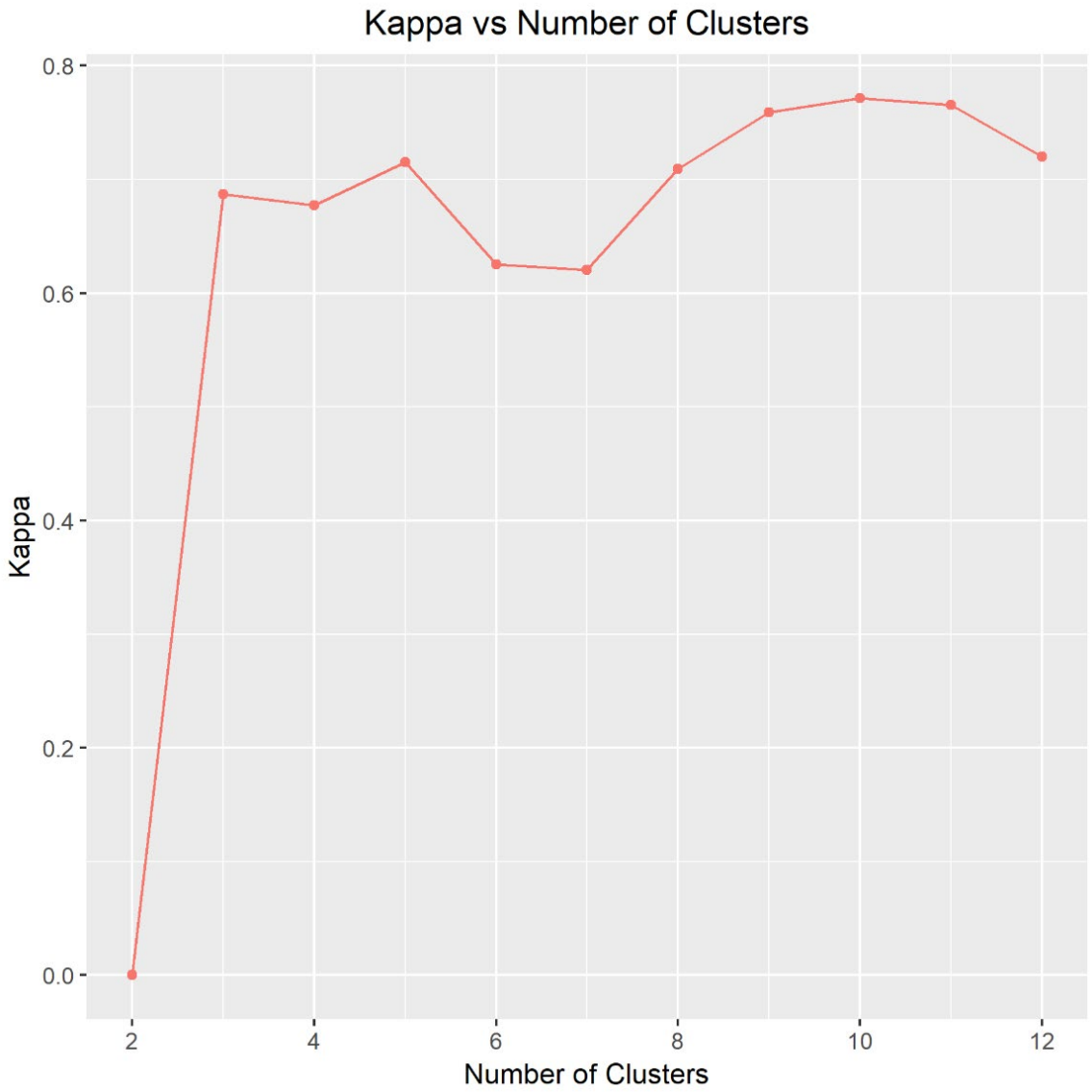


Figure 15: Plot of κ vs number of clusters for geologic habitat based on five-fold cross-validation of the training data plot to determine the optimal number of clusters

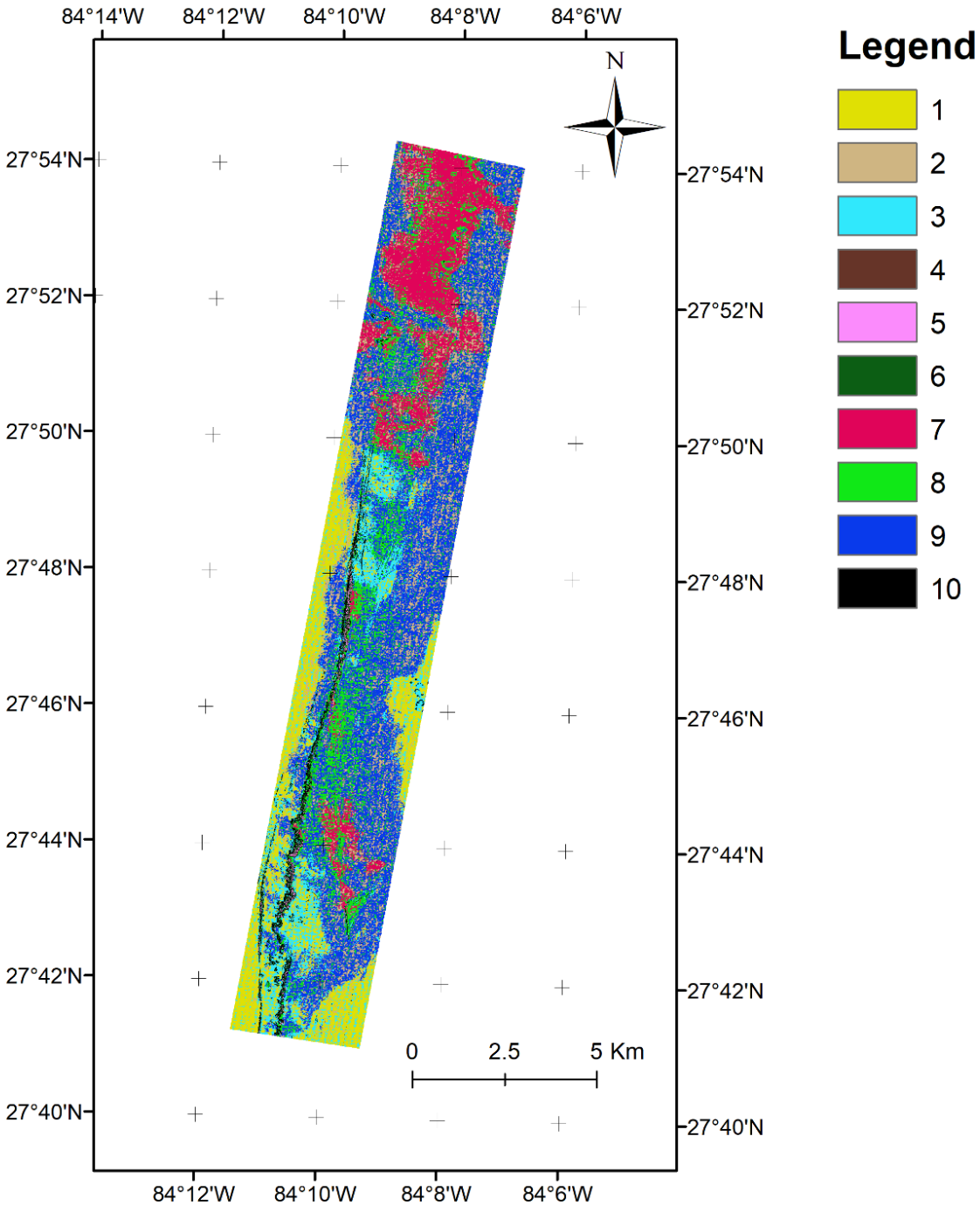


Figure 16: Map of the 10 acoustic clusters with 10 m x10 m resolution determined through k-means clustering of selected principal component layers.

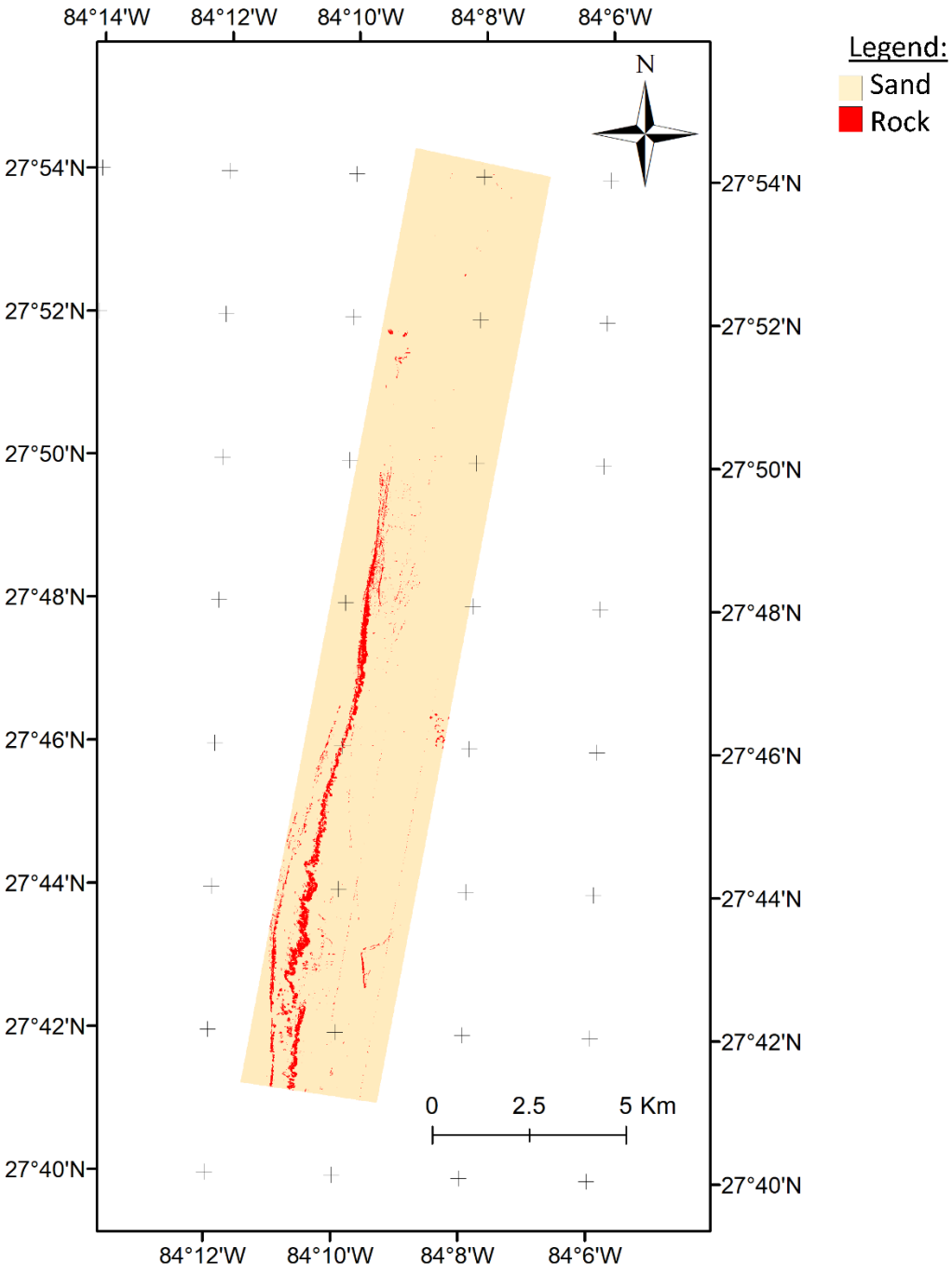


Figure 17: Map of geologic habitat with 10 m x 10 m resolution determined through unsupervised (k-means) classification

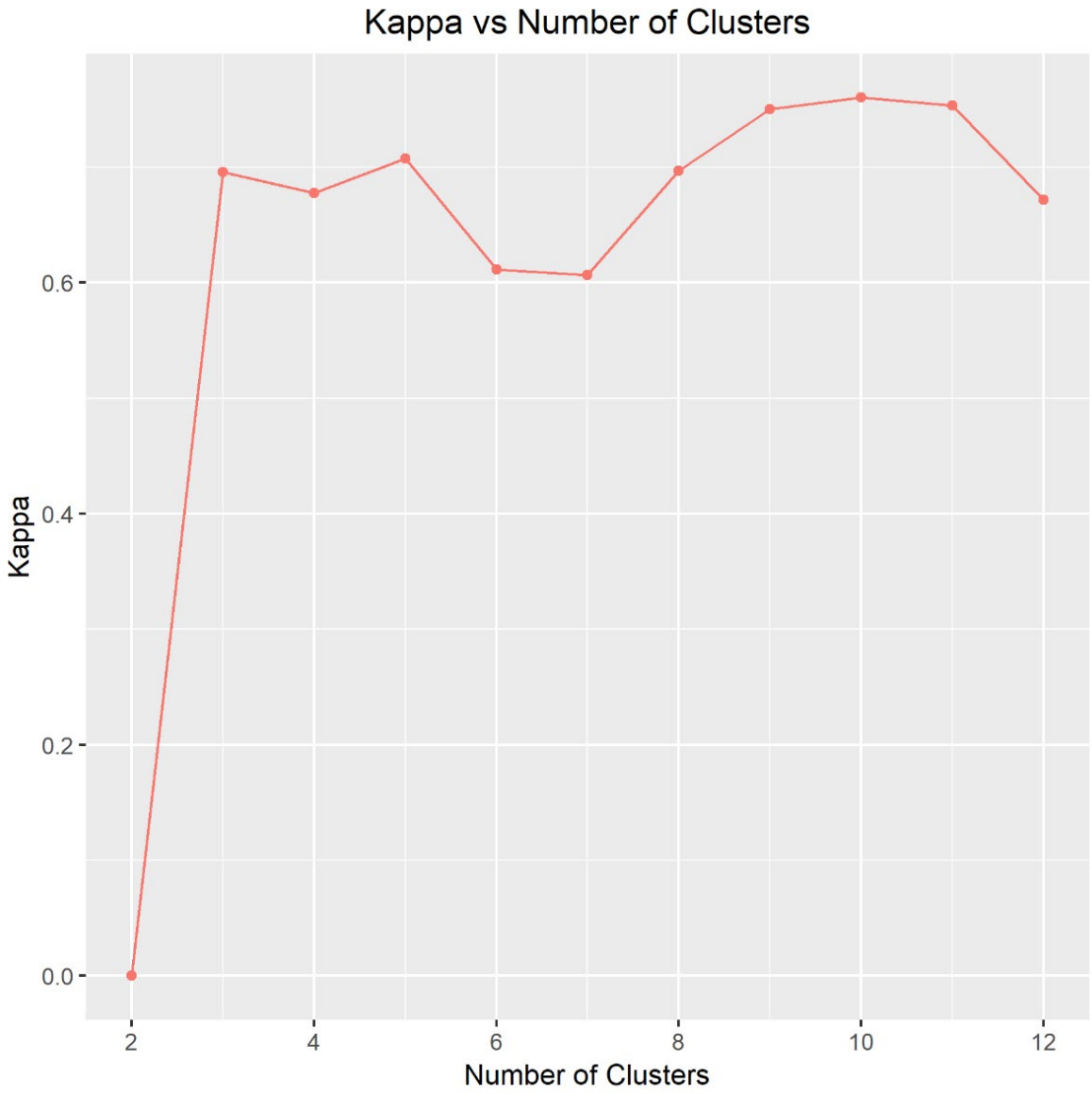


Figure 18: Plot of κ vs number of clusters for biotic habitat based on five-fold cross-validation of the training data plot to determine the optimal number of clusters.

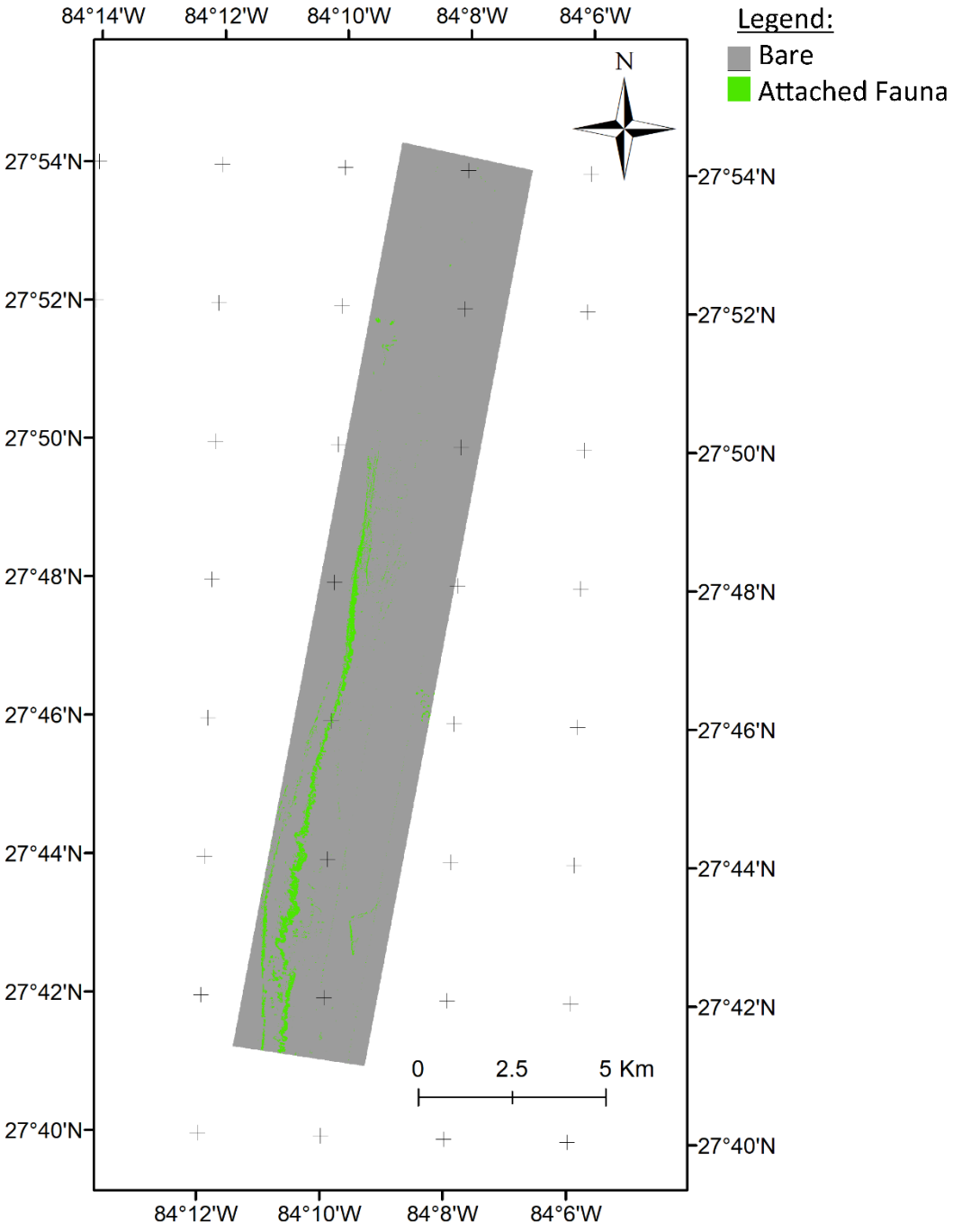


Figure 19: Map of biotic habitat with 10 m x 10 m resolution determined through unsupervised (k-means) classification

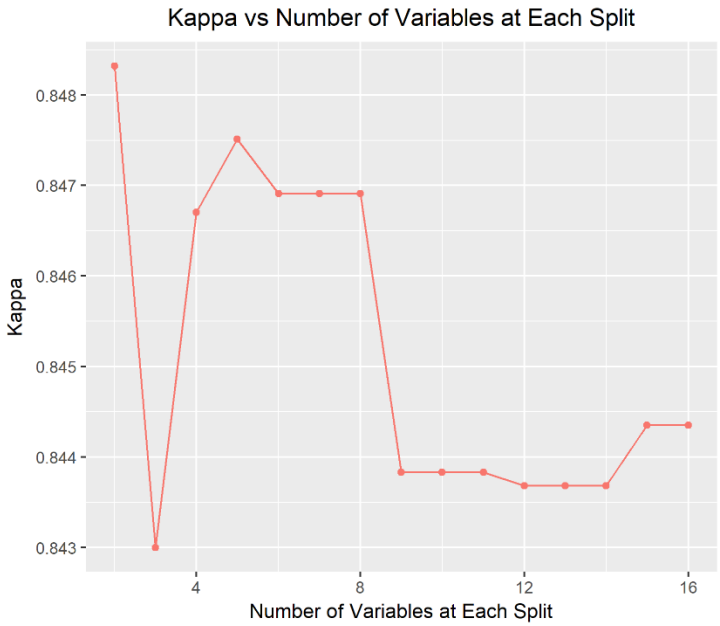


Figure 20: Plot of κ vs number of variables at each split for training random forest algorithm on geologic habitat

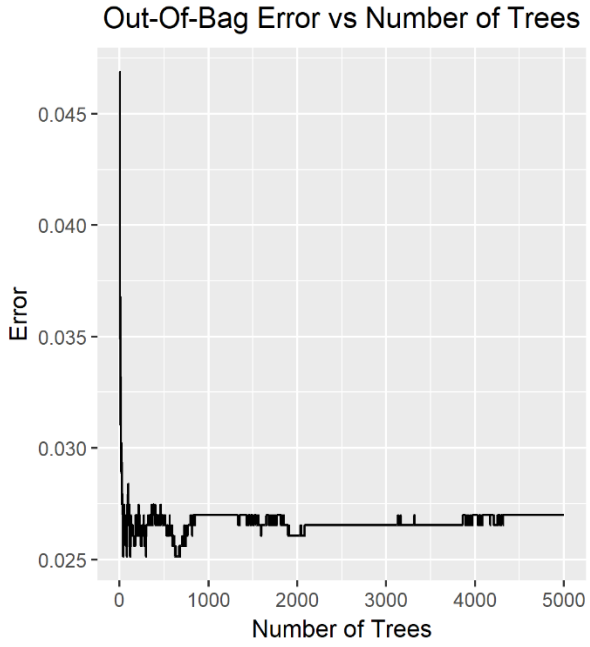


Figure 21: Plot of out-of-bag error vs number of trees in the random forest model for geologic habitat

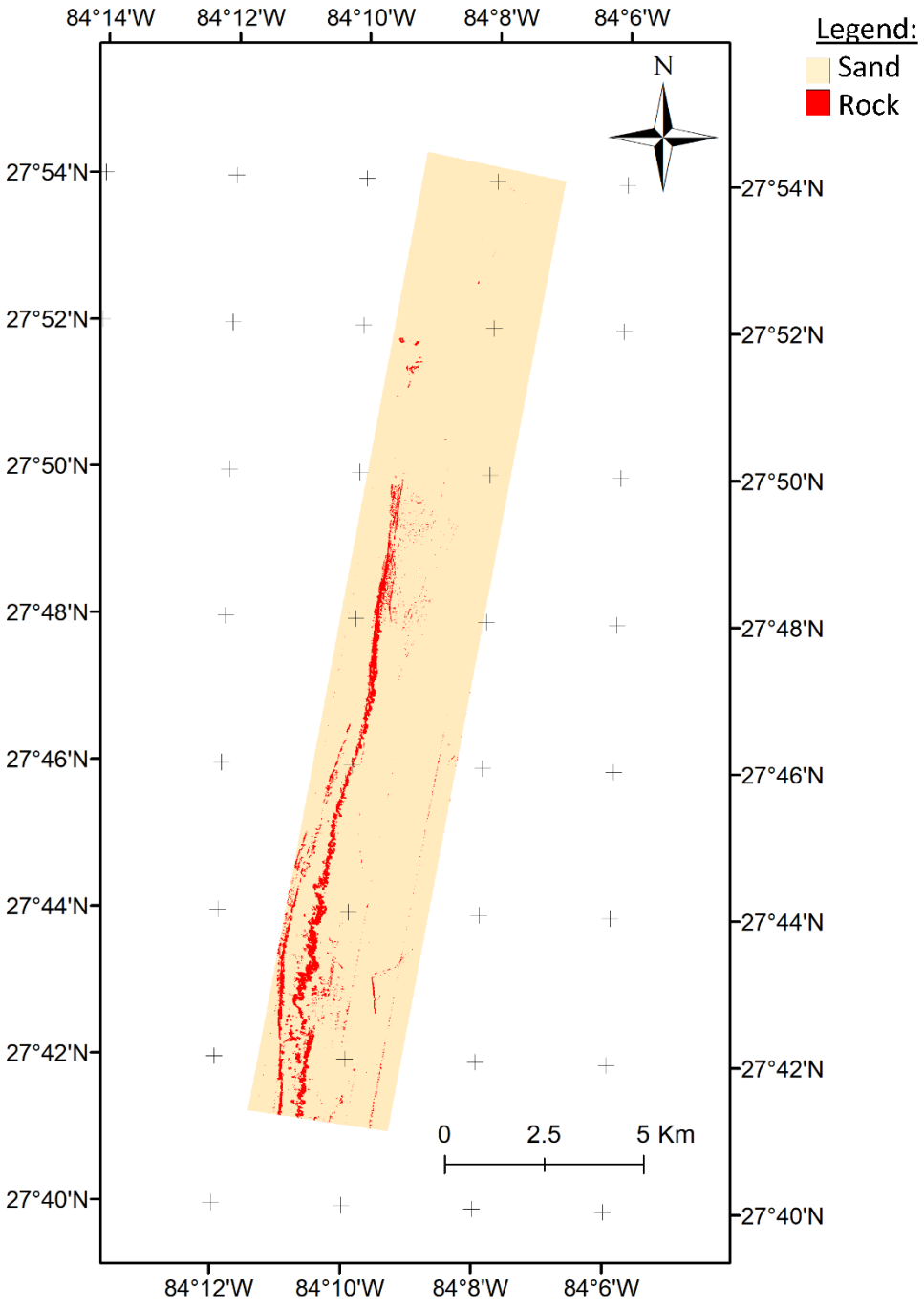


Figure 22: Map of geologic habitat with 10 m x 10 m resolution determined through supervised (random forest) classification

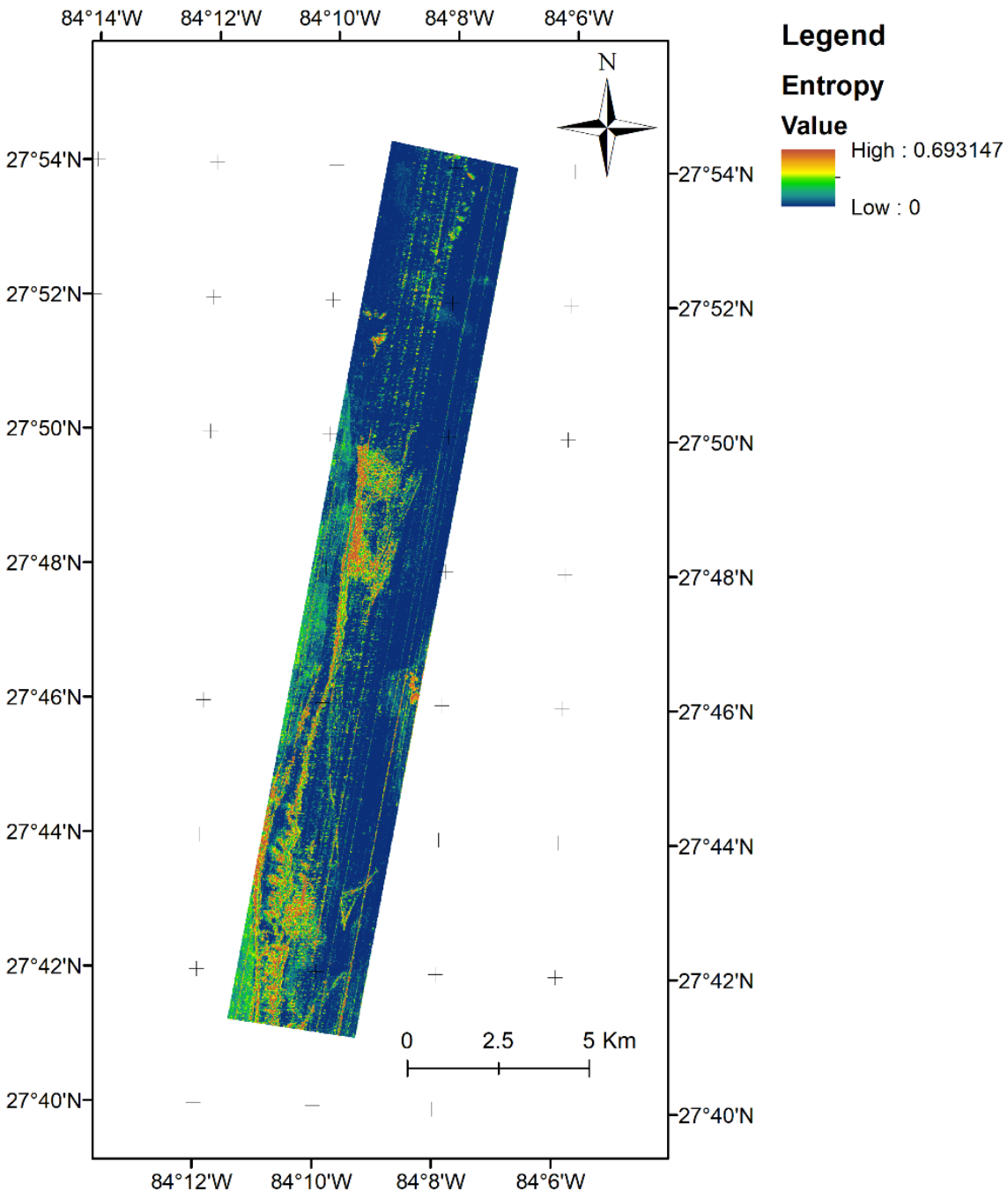


Figure 23: Entropy map with 10 m x 10 m resolution of geologic habitat classification from the supervised (random forest) classification

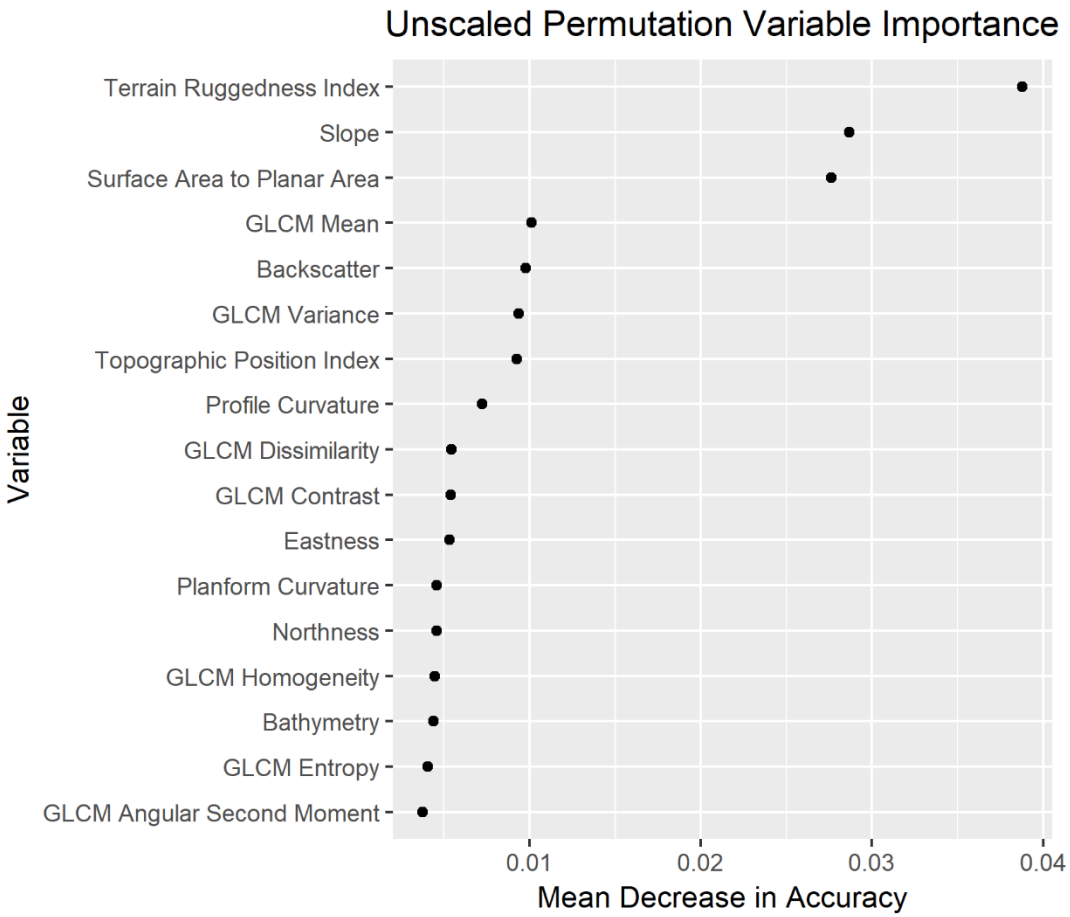


Figure 24: Variable importance for predicting geologic habitat (rock vs sand) according to mean decrease in accuracy as determined by permuting the OOB observations. Variable importance values reported are unscaled (not divided by standard deviation).

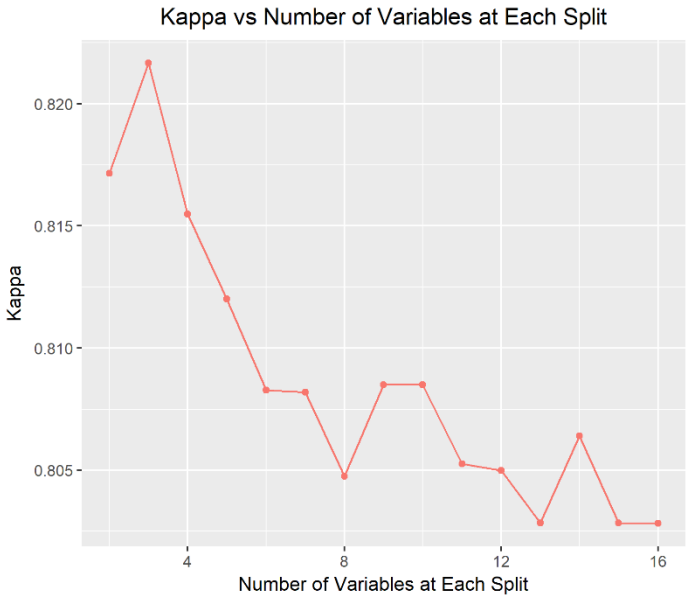


Figure 25: Plot of κ vs number of variables at each split for training random forest algorithm on biotic habitat

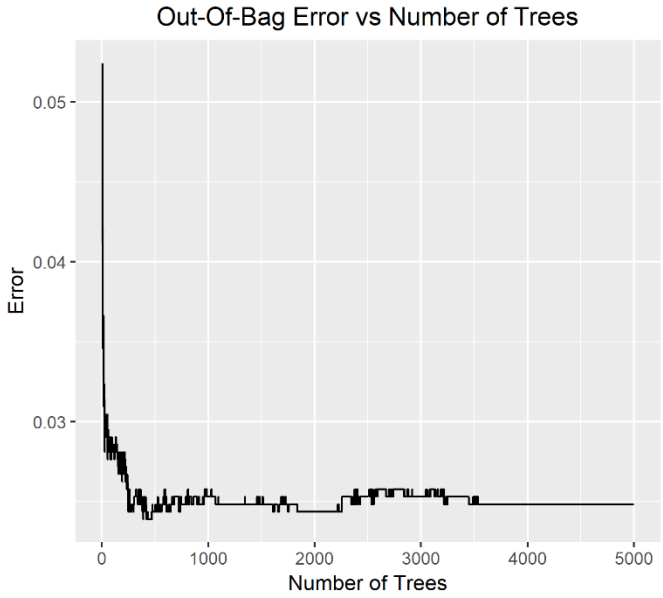


Figure 26: Plot of out-of-bag error vs number of trees in the random forest model for biotic habitat

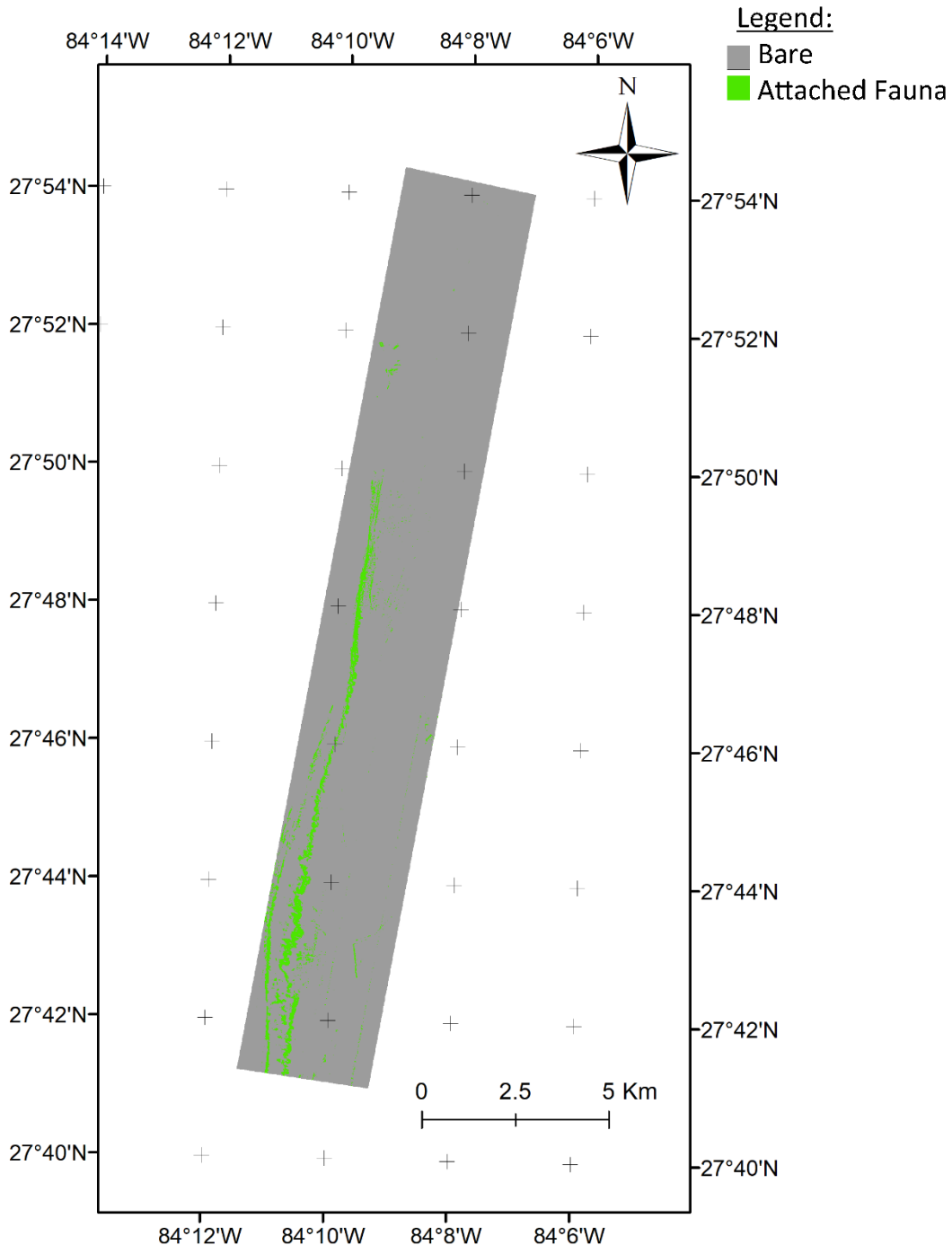


Figure 27: Map of biotic habitat with 10 m x 10 m resolution determined through supervised (random forest) classification

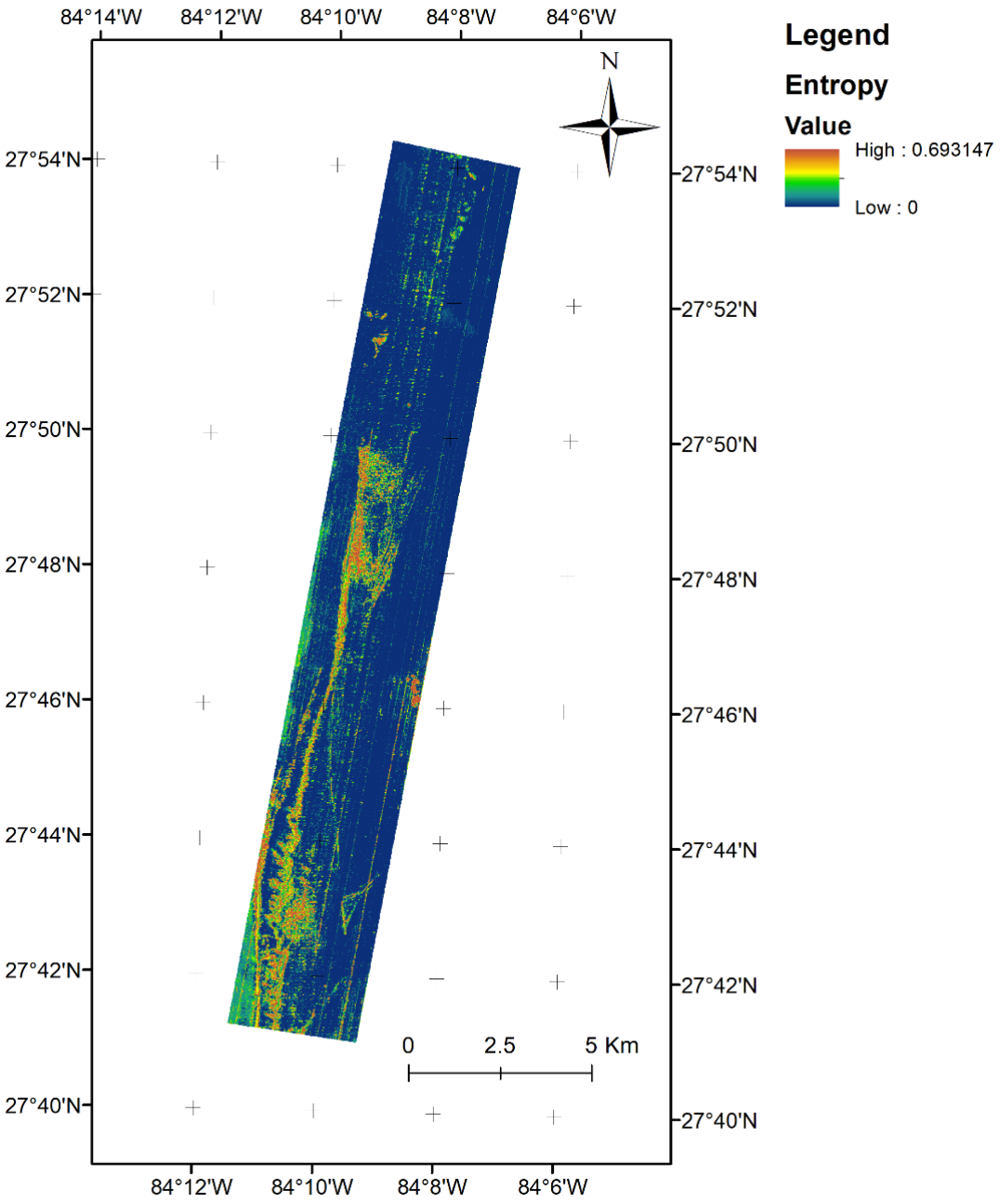


Figure 28: Entropy map with 10 m x 10 m resolution of biotic habitat classification from the supervised (random forest) classification

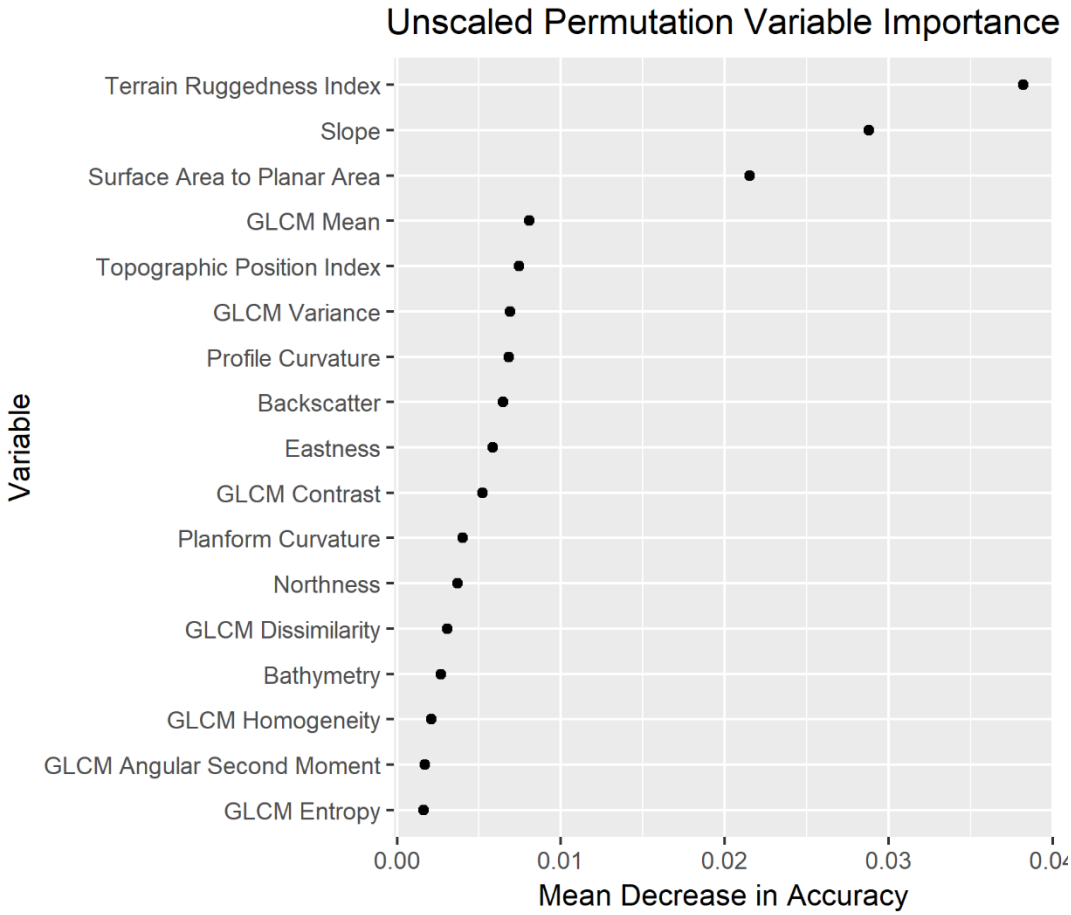


Figure 29: Variable importance for predicting geologic habitat according to mean decrease in accuracy as determined by permuting the OOB observations. Variable importance values reported are unscaled (not divided by standard deviation).

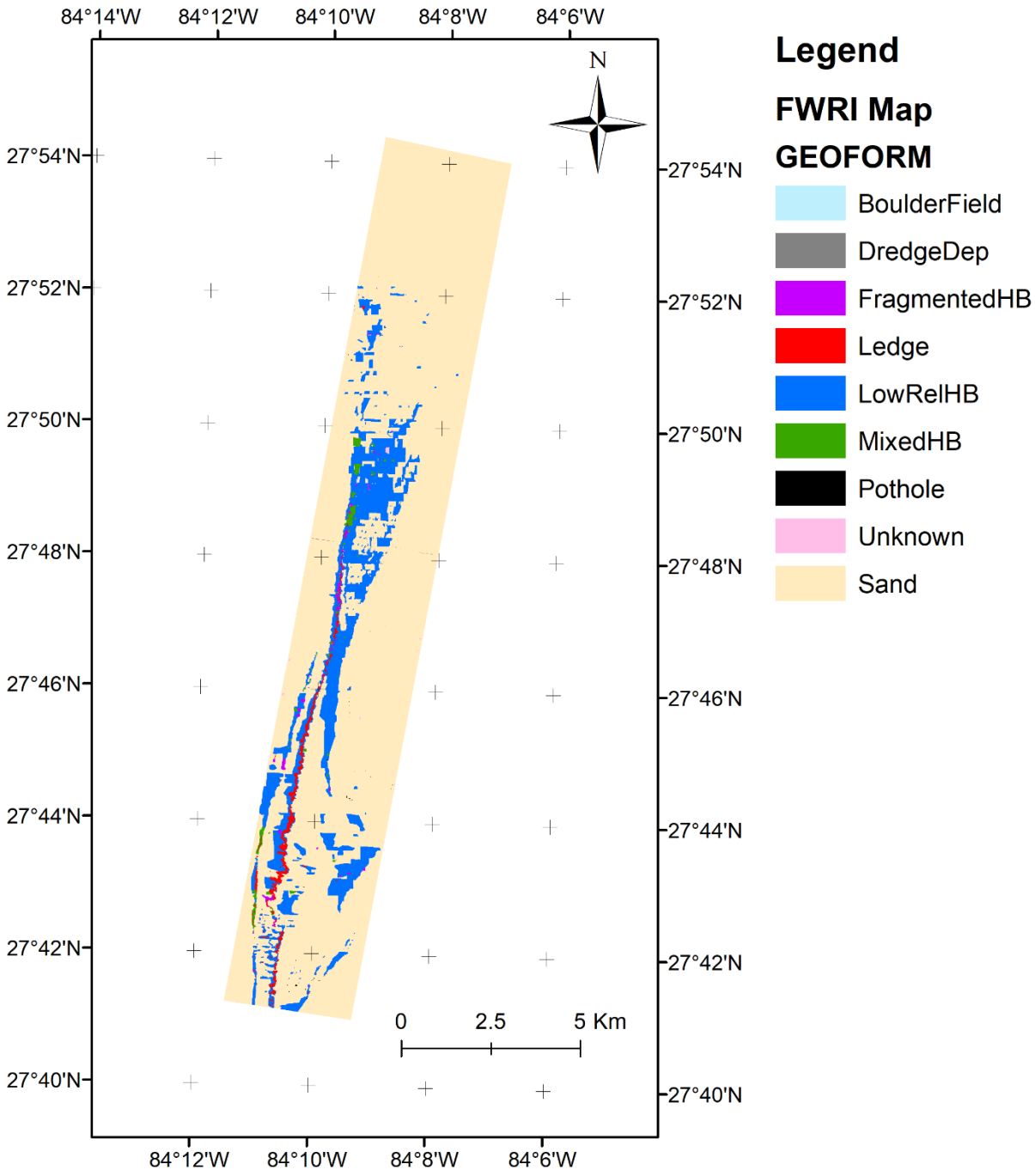


Figure 30: Vector map of geoform habitat of The Elbow created by FWC-FWRI. The map has been trimmed to the extent of the study area, and areas that FWC-FWRI did not classify as any type of hard-bottom are labelled as sand.

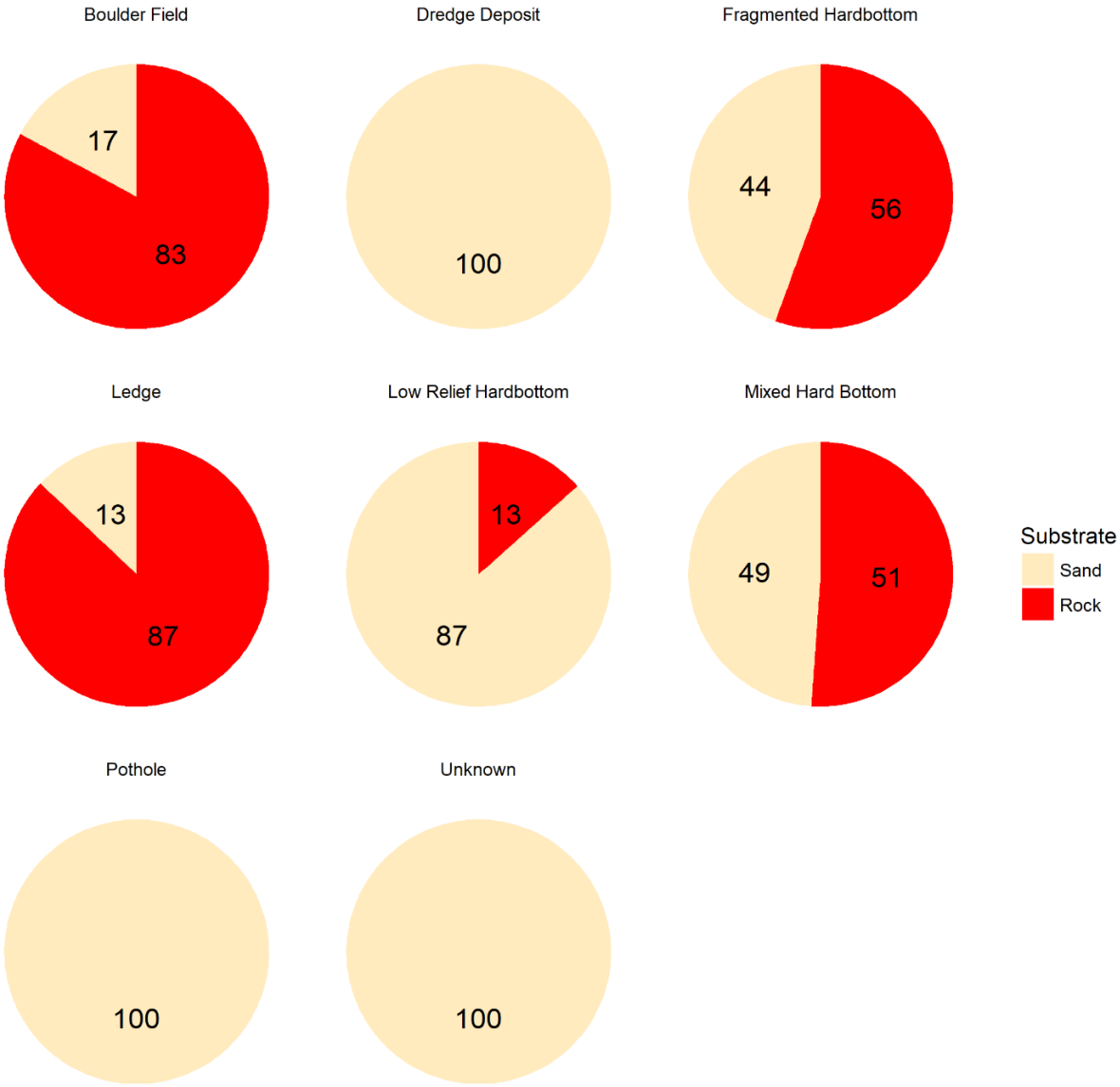


Figure 31: Correspondence between geoform categories determined by FWC-FWRI and the geologic habitat determined from the supervised classification in this study

Chapter 3: Fish Community Analysis

Introduction

The status of fisheries are commonly assessed using fisheries-dependent and/or fisheries-independent data (Shepherd, 1988). Fisheries-dependent monitoring employ catch data from the fishery to track changes in catch per unit effort (CPUE) over time for a fishery of a given species in order to assess whether the population is increasing or decreasing. Fisherman however often know where to go to target fish, and this may change over time due to changes in fish distribution, abundance or regulations, so the sampling may not be reflective of the entire population (Switzer et al., 2014). Fisheries-independent data on the other hand are collected from scientific surveys for the purpose of stock assessment, using more statistically robust sampling schemes in order to better reflect the status of the overall population (Switzer et al., 2014). The largest fisheries independent monitoring program in the Gulf of Mexico is the SEAMAP groundfish survey which uses a bottom trawl to catch fish and invertebrates near the seafloor. Bottom-trawls can cover a large area, but provide very course spatial resolution on fish abundance (at the scales of kilometers), and they can damage sensitive habitats as they are dragged along the seafloor (Board and Council, 2002, Kloser et al., 2007, Anderson et al., 2009). Additionally, bottom trawls can only be used to sample soft-bottom habitats as gear can hang on rockier and higher relief habitats, making them a poor candidate for assessing adult reef fish (Auster et al., 2001, Kingon, 2013, Lembke et al., 2013). On the West Florida Shelf (WFS), the National Marine Fisheries Service (NMFS) and the Florida Fish and Wildlife Conservation Commission's Fish and Wildlife Research Institute (FWC-FWRI) use alternative gear types such as fish traps, hooks, and stationary baited underwater cameras to assess reef fish stocks (Switzer

et al., 2014). Unlike bottom trawls, traps and hooks can be deployed in rockier and high relief areas, provide fine-scale spatial resolution on abundance, and they do not cause substantial damage to the seafloor habitats. They however are still inherently extractive as they remove individuals from the population, which can be undesirable if the population being assessed is endangered, or when monitoring sensitive or protected habitats. Baited stationary cameras provide a way to use non-extractive techniques to assess reef fish and have been found to more representatively sample the fish community compared to traps while concurrently collecting information about the surrounding habitat (Switzer et al., 2014).

With rapidly improving technology, visual sampling methods such as the baited stationary camera have become more common over the last decade (Mallet and Pelletier, 2014). There are many types of visual sampling methods, but most visual sampling methods targeted at assessing fish populations have the added benefit of providing information about the surrounding habitat, and allow for fine-scale (meters to 10's of meters) assessment of fish-habitat relationships (Cappo et al., 2003). Additionally, these methods are non-extractive, and have the potential to provide important archival data that can be reviewed and analyzed in the future to extract more information and assess trends over time (Bowden and Jones, 2016). Visual sampling methods however cannot provide certain life-history information that require the collection of a physical specimen, and data quality varies as a function of several factors (e.g. water clarity; Parker Jr et al., 1994, Switzer et al., 2014).

There are many different methods for visual sampling of fish. In shallow water SCUBA is a popular method; however this method becomes ineffective or impossible in deeper waters (Mallet and Pelletier, 2014). Three common methods of visual sampling for waters beyond diveable depths include the baited stationary cameras, remote operated vehicles (ROV's), and

towed cameras. In this study we use a towed video camera system flown near the seafloor. Towed camera systems have the advantage of being able to cover a larger area than both ROV's and stationary cameras (Lembke et al., 2013, Lembke et al., 2017, Logan et al., 2017). Additionally, as towed camera systems provide information across long continuous transects, this allows them to better characterize transitions between habitats as compared to either of these other technologies (Logan et al., 2017). Towed systems are also cheaper and able to handle rougher seas as compared to ROV's (Lembke et al., 2013, Bowden and Jones, 2016). Additionally, the use of towed video allows for a calculation of the "area swept" which in turn can be used to provide direct estimates of fish densities (abundance per unit area; McIntyre et al., 2013, Grasty, 2014). This can also be accomplished with an ROV, but cannot be accomplished with baited stationary cameras due to the unknown size of the "ring of attraction" created by the dispersal plume of the bait scent, and due to the angle of the camera which is typically oriented horizontally (Mallet and Pelletier, 2014). As a result, baited stationary cameras typically provide a relative index of abundance such as MaxN, which is the maximum number of individuals of a given species seen in the frame at the same time over a given time period (Logan et al., 2017). Towed camera systems are not without their drawbacks though as there are tradeoffs for using any type of sampling gear. The act of rapidly towing a large camera system through the water may cause attractance or more commonly avoidance behaviors by certain species, causing some species to be over or underrepresented respectively (Stoner et al., 2008, Grasty, 2014). Fish may react just as the camera system is approaching (near-field reactions), as well as when the camera is at considerable distance (far-field reactions; Stoner et al., 2008). Far-field reactions are much harder to evaluate as the fish are reacting before they come into view of the camera system, while near-field reactions can more easily be assessed as you can see how fish are reacting to the

system since they are in the view of the cameras (Stoner et al., 2008). Previous results have shown generally neutral to mild avoidance near-field behaviors by target reef fish species to the C-BASS system used in the current study (Grasty, 2014). Efforts to assess far-field reactions to the C-BASS system are ongoing through the paired use of towed and stationary cameras, and preliminary results showed no statistical difference in communities before and after the towed system had passed through an area, though this may be due to lack of statistical power to detect those differences (Grasty, 2014). In addition, towed camera systems are limited to deploying a single system at a time, while stationary cameras allow for multiple units to be deployed at different locations concurrently allowing for efficient sampling of target areas (Logan et al., 2017). Moreover, as a tradeoff for covering more area, towed systems also collect less detail about the environment than an ROV which can finely maneuver and closely zoom in on areas of interest (Bowden and Jones, 2016, Lembke et al., 2017). Stationary baited cameras have also been found to have higher statistical power than towed systems, and have been found to observe more abundant and diverse fish assemblages in some environments (Logan et al., 2017). Despite these drawbacks, the numerous advantages offered by a towed video systems make it an effective technology for assessing reef fish (Stoner et al., 2008, Grasty, 2014, Bowden and Jones, 2016, Lembke et al., 2017, Logan et al., 2017).

In this study a towed camera system is used to assess differences in fish community composition and abundance based on differences in substrate type, vertical relief, and the presence of attached fauna (e.g. sponges and corals). On the WFS, previous studies have demonstrated that these factors can be important in shaping fish communities (Allee et al., 2011, Grasty, 2014, Switzer et al., 2014). For example, hard bottom habitats on the WFS have different communities than soft bottom habitats (Allee et al., 2011, Grasty, 2014). Additionally, offshore

rocky reefs are known to be important for the spawning of groupers and snappers, with several areas on the WFS confirmed to be spawning sites for these taxa (Allee et al., 2011, Coleman et al., 2011). In general, most studies both on the WFS and around the world, have found habitat complexity to be positively related to the abundance and diversity of fish communities (Parker Jr et al., 1994, Gratwicke and Speight, 2005, Pittman et al., 2007, Kendall et al., 2009, Allee et al., 2011, Grasty, 2014, Switzer et al., 2014, Logan et al., 2017). Increasing habitat complexity may result from high rugosity, high relief, and the presence of attached fauna such as sponges and soft corals. This increased complexity is believed to lead to increased diversity and abundance by providing more opportunities for feeding, seeking refuge from predators, and through allowing for greater niche partitioning (MacArthur and Levins, 1964, Friedlander and Parrish, 1998, Almany, 2004). Although this general trend seems to hold true in many ecosystems, many fish-habitat relationships may be system specific or scale-dependent (Wiens, 1989, Levin, 1992, Kendall et al., 2009). Additionally, although habitat complexity is positively related to diversity and abundance of fish, different fish exhibit different habitat preferences, with some species preferring more complex habitats, some preferring less complex habitats, and some exhibiting more generalist behaviors by inhabiting a variety of habitats types (Allee et al., 2011, Grasty, 2014, Switzer et al., 2014). For example, although hard-bottom areas may be more diverse, sandy habitats are considered important for the Red Grouper (*Epinephelus morio)*, which is an ecologically and commercially important species, that creates large sand burrows called "Grouper holes" (Wall et al., 2011). Understanding which species show repeatable associations with certain habitat characteristics, and understanding the range and scale at which those relationships are applicable is key to developing accurate habitat stratified population estimates for fish stocks.

Methods

Data Processing

Fish Counts

All videos are watched in the CVision fish counting software (Woodward and Takahashi, 2017). All fish are identified and the exact frame-number is logged and exported to a csv file. If the habitat is complex, the video is viewed multiple times at different speeds in order to increase detection ability. Fish are only counted if they are observable in the primary camera, though other cameras are used to aid in identification (e.g. if the fish swims out to the side quickly and can be seen more clearly in a side camera). The primary camera used was the monochrome HD Point Grey Blackfly® camera as it has consistently provided the highest quality imagery. Fish are identified to the lowest taxonomic level possible given the visibility and quality of the data. Inability to identify fish down to the species level can occur for several reasons, including turbid water, poor viewing angle, or simply that several species in a taxonomic group look very similar and are difficult to differentiate from video sampling alone. If fish cannot be identified to the species level, they then are identified to a higher taxonomic level such as at the genus or family level, or if that is not possible as Large (> 15 cm) No ID. Fish under 15 cm that cannot be identified are not included in this analysis as counting them is difficult and unreliable as C-BASS's primary purpose is to survey large-bodied fish (Grasty, 2014). As such, fish under 15 cm that could not be identified were excluded from this analysis.

Linking Fish Counts to Habitat Observations

Using the recorded frame number for each fish and each annotated habitat observation, fish were considered to be associated with the habitat observation nearest in time in order to create a species-by-site matrix containing fish abundance per habitat observation (Figure 32).

Calculation of Area Viewed

The width of the frame was calculated using equations 5 – 9 (Grasty, 2014). The area viewed was then calculated by multiplying the width of the camera's field of view at the center of the frame by the distance traveled (Equation 10 and Equation 11). This was done for every 15 s interval using the median value of C-BASS' speed, altitude, and pitch over that time period. The median value was chosen in order to provide values of those parameters that are representative of that interval while protecting against the influence of outliers and faulty readings.

Eq 5: $$\boldsymbol{h_A = \cos(P) * h_R}$$

h_A = Adjusted altitude
P= Pitch
h_R = Raw altitude

Eq 6: $$\boldsymbol{\theta_G = \theta_C - P}$$

Θ_G = Camera angle to ground
Θ_C = Downwards angle of camera relative to horizontal axis of the chassis (32.8º)
P= Pitch

Eq 7: $$\boldsymbol{C = \frac{h_A}{\sin \theta_G}}$$

C = Center-line distance
h_A = Adjusted altitude
Θ_G = Camera angle to ground

Eq 8: $$\boldsymbol{HFOV_{sea} = 2 * \arcsin\left(\sin\left(\frac{HFOV_{air}}{2}\right) * \frac{n_{air}}{n_{sea}}\right)}$$

$HFOV_{sea}$ = Camera-specific field of view in seawater
$HFOV_{air}$ = Camera-specific field of view in air as specified by manufacturer (82.4º)
n_{air} = Index of refraction of air (1.000277)
n_{sea} = Index of refraction of seawater (4/3)

Eq 9: $$\boldsymbol{W = 2 * C * \tan(\frac{HFOV_{sea}}{2})}$$

W = Width of center of frame
C = Center-line distance
$HFOV_{sea}$ = Camera-specific field of view in seawater

Eq 10: $$\boldsymbol{L = S * t}$$

L = Distance covered
S = Speed
t = time (15 seconds)

Eq 11: $$\boldsymbol{A = W * L}$$

A = Area Covered
W = Width of center of frame
L = Distance covered
D= Density

Conversion from Fish Counts to Densities

Prior to analysis, all fish counts were converted to densities to account for changes in the area viewed that occur due to changes in speed and altitude. This was done by taking the fish counts for each species associated with a given habitat observation and dividing the counts by the area viewed by the C-BASS system over that 15 second window.

Data Analysis

Habitat Groupings

Both geologic and biotic habitats were examined in this analysis based on the simplified scheme presented in the previous chapter (Figure 7 and Figure 8). For these analyses, moderate and high relief rocky habitats were merged into one class in order to increase the sample size while still providing the ability to analyze the influence of vertical relief on fish communities, as previous studies have demonstrated this to be a potentially important driver. This led to three groups for geologic habitat (sand, low relief rock, and moderate/high relief rock), and two groups

for biotic habitat (presence or absence of attached fauna). For geologic habitat, there were 2,344 observations for sand, 369 observations for low relief rock, and 23 observations of moderate/high relief rock. For biotic habitat there were 2,369 observations of bare habitat, and 367 observations of attached fauna habitat.

Species Richness

The number of species within each 15 second bin was counted. When fish were identified at a courser taxonomic level than species (e.g. Family), then that was only counted towards the species richness if there were no other fish within that taxonomic group. The relative frequency of species richness was then plotted by habitat.

Habitat-Specific Densities

The average densities of each species over each habitat class was calculated for both the geologic and biotic habitat, and confidence intervals (95%) were determined using bootstrap resampling with 999 iterations after subsetting the data to the habitat type. Additionally, densities and confidence intervals were calculated for a two-class version of geologic habitat (rock vs sand) consistent with the habitat maps presented in the previous chapter.

Multivariate Community Analyses

For multivariate analyses, all densities were square root transformed to reduce the influence of occasional large aggregations. Additionally, the Bray-Curtis dissimilarity metric was used, which calculates the dissimilarity between samples in a more ecologically appropriate way than traditional Euclidean distance (Clarke and Warwick, 2001). It does this by calculating the dissimilarity between each pair of samples only in terms of species that are present in at least one of the samples, therefore preventing areas to be considered similar on the basis of joint species absences. All observations where no fish were observed were removed, as this is a requirement

for calculating Bray-Curtis dissimilarity since it cannot calculate dissimilarity between pairs of samples solely on the basis of joint absences. This resulted in 95 observations of sand, 173 observations of low relief rock, and 18 observations of moderate/high relief rock for geologic habitat, and 105 observations of bare habitat and 181 observations of attached fauna for the biotic habitat. Significance was assessed at the level of $\alpha = 0.05$. The following multivariate analyses of the fish community were conducted using the Fathom Toolbox in MATLAB (Jones, 2014).The overall variation in the fish community composition and abundance was explored using Principal Coordinates Analysis (PCoA), and the variance explained by each principal coordinate was adjusted to account for variance inflation that occurs due to a mathematical artefact related to negative eigenvalues which explain negative percentage of the variance (Legendre and Legendre, 1998). A two way non-parametric permutation based ANOVA (PERMANOVA) was then run to test the following three null hypotheses (Anderson, 2001):

1. There is no significant difference in fish community composition and abundance among geologic habitat classes.
2. There is no significant difference in fish community composition and abundance among biotic habitat classes.
3. There is no significant interaction effect between geologic and biotic habitat on fish community composition and abundance.

Prior to running the 2-way PERMANOVA, the assumption of homogenous multivariate dispersion among habitat classes was verified for both geologic and biotic habitat using a multivariate analogue of the Levene's Test (Anderson, 2006). As PERMANOVA's are an omnibus test it does not tell you which of the groups are significantly different from one another, only that there may be at least one significant difference. Therefore, if significant differences

were detected and more than two groups were being compared, that analysis was followed up by pairwise tests to test which groups were significantly different from one another using Holm's adjusted p-values for multiple comparisons (Holm, 1979). A plot was then generated using Canonical Analysis of Principal Coordinates (CAP) in order to determine which species were most responsible for driving those differences (Anderson and Willis, 2003).

Results

Multivariate Community Analyses

Over the three transects a total of 2,032 different individual fish and one sea turtle were observed spanning at least 33 different species and 20 different families (Appendix 7).

To assess overall trends in the fish community, a PCoA was conducted. The first two axes of the PCoA explained 8.4% percent of the total variation after correcting for negative eigenvalues, with the first component describing 5.01% and the second axis explaining 3.39% of the total variation (Figure 33a). Examination of the PCoA weighted biplot vectors shows that there are three main species driving the overall variation in species composition and abundance in the study are (Figure 33b). These species are the Lionfish (*Pterois* spp.), Squirrelfish (Holocentridae), and Sand Tilefish (*Malacanthus plumieri*).

To assess differences in fish communities among habitat types, a two-way PERMANOVA was conducted to examine if there are significant differences in fish community composition and abundance among differing geologic and biotic habitat types. Prior to conducting the PERMANOVA, the assumption of homogeneity of multivariate dispersion was validated for both biotic and geologic habitat through the use of a multivariate analogue to the Levene's test. The results of this showed no significant differences in multivariate dispersion among groups for either geologic ($F=0.28$, $p=0.811$) or biotic habitats ($F=0.39$, $p=0.555$)

indicating that this assumption is met. The results of the two-way PERMANOVA are shown in Table 8. As the interaction term is not significant, the effects of geologic and biotic habitat can be interpreted separately without controlling for the other. The PERMANOVA shows that the fish community composition differed significantly among both geologic and biotic habitats. Since there were more than two groups for the geologic habitat, the PERMANOVA was followed up with pairwise comparisons to identify which groups significantly differed from each other (Table 9). The pairwise comparisons found significant differences in fish species composition and abundance for all pairwise comparisons (sand vs low relief, sand vs moderate/high relief, and low relief vs moderate/high relief).

In order to determine what species are driving the differences among habit groups, two CAP analyses were run: one for geologic habitat and one for biotic habitat. The number of PCoA axes retained for each CAP was determined by finding the number of axes that maximized accuracy according to Leave-One-Out-Cross-Validation (LOO-CV). A total of 13 PCoA axes were retained in the CAP for the geologic habitat, and five PCoA axes were retained in the CAP for the biotic habitat (Figure 34). The results of the CAP analyses with the corresponding species correlation vectors overlaid are visualized in Figure 35. The CAP of geologic habitat (Trace statistic = 0.5227, p = 0.0001, m = 13, variability of Y_{dis} expl. = 96.13%) has substantial overlap, especially among low relief rock and sandy habitats, while moderate/high relief habitats seem to be a bit more clearly differentiated. General trends however show that fish communities in sand habitats tend to be characterized by more Sand Tilefish than the fish communities in the other geologic habitats. Fish communities in low relief rocky habitats on the other hand generally are characterized by having more Squirrelfish and Surgeonfish (Acanthuridae) than the fish communities in the other geologic habitats. Fish communities in moderate/high relief rocky

habitats are characterized by more Creolefish (*Paranthias furcifer*), Spanish Hogfish (*Bodianus rufus)*, Gray Snapper (*Lutjanus griseus)*, Goliath Grouper (*Epinephelus itajara)*, and Spotted Goatfish (*Pseudupeneus maculatus)* than the fish communities in the other geologic habitats. Blue Angelfish (*Holacanthus bermudensis)* and Lionfish appear to be associated with both low relief and moderate/high relief rocky habitats indicating that fish communities in these two habitats have more of Blue Angelfish and Lionfish than fish communities in sand habitats. The CAP of biotic habitat (Trace statistic = 0.1594, p = 0.0001, m = 5, variability of Y_{dis} expl. = 71.98%) displays substantial overlap between the two classes; however on average it appears that fish communities in bare habitats are characterized by having more Sand Tilefish, and fish communities in habitats with attached fauna are typically characterized by having more Squirrelfish, Blue Angelfish, and Lionfish.

Species Richness

The plot of relative frequency of species richness by geologic habitat type for each 15s bin can be seen in Figure 36a. Sand habitats were occupied only 5% of the time, and when occupied were generally limited to one species. Low relief rock were occupied 50% of the time and when fish were present there were generally between one and three species present. Moderate/high relief rock were occupied 83% of the time, and when occupied generally had between one and four species present with a maximum value of nine species within a 15s bin.

The plot of relative frequency of species richness by biotic habitat type for each 15s bin can be seen in Figure 36b. Bare habitats were occupied only 6% of the time, and when occupied were generally limited to one species. Attached Fauna habitats were occupied 52% of the time and when fish were present there were generally between one and three species present.

Habitat-Specific Densities

The average densities for each species by habitat can be seen in Figures 37-39.

Discussion

My results show that as a whole the variation in species abundance and composition in this area is driven mostly by three species: Lionfish, Squirrelfish, and Sand Tilefish. I found that fish communities significantly differed among all three geologic habitat groups as well as between both biotic habitat groups. For the geologic habitat, fish communities in flat, sandy habitats were generally differentiated by more Sand Tilefish, while fish communities in low relief rocky habitats were differentiated by more Squirrelfish and Surgeonfish, and fish communities in moderate to high-relief rocky habitats were differentiated by more Creolefish, Spanish Hogfish, Gray Snapper, Goliath Grouper, and Spotted Goatfish. Blue Angelfish and Lionfish were not characteristic of fish communities in one specific habitat, but rather were characteristic of fish communities in both low and moderate/high relief rock habitats thus differentiating communities in rock vs sand habitats. For the biotic habitat, examination of what species were driving those differences revealed that fish communities in bare habitats were characterized by Sand Tilefish, while fish communities in areas with attached fauna were characterized by having more Squirrelfish, Blue Angelfish, and Lionfish. Similarly to previous results on the WFS, I also found that fish communities differed among hard and soft-bottom habitats (Allee et al., 2011, Grasty, 2014, Switzer et al., 2014). Moreover, much like Switzer et al. (2014) which also studied a portion of The Elbow, I found that fish communities differed among different types of hard-bottom habitats. Although fish communities significantly differed between attached fauna and bare habitats, rock habitats were very rarely bare so the comparisons generally reflected the differences between rock and sand without the consideration of vertical relief.

The results of these multivariate analyses however only examine sites where fish were present. Much of the sand habitat however was often barren, while rockier habitats were much more frequently occupied by fish. Therefore it is important to follow up these multivariate analysis with univariate analyses by species in order to get a more complete picture of fish-habitat densities relationships. I found that fish densities and species richness were typically higher on more complex habitats (rockier, higher relief, presence of attached fauna; Figs 36-39). These findings are in line with previous results that have also found this trend (Parker Jr et al., 1994, Gratwicke and Speight, 2005, Pittman et al., 2007, Kendall et al., 2009, Allee et al., 2011, Grasty, 2014, Switzer et al., 2014, Logan et al., 2017). Additionally the likelihood of finding fish within a 15 second bin greatly increases with habitat complexity, being just 5% for sand habitats, 50% for low relief rock, and 83% for moderate/high relief rock. When examining the differences between various hard-bottom habitats, it is clear that on average higher relief features have greater fish densities when looking at all species together, but when analyzing at species level, the relationships are more variable (Figure 37). The Angelfishes (Blue and Gray) show a positive relationship with vertical relief showing increased density having the lowest densities over flat sand, intermediate densities over low relief habitats, and the highest densities over high relief habitats. This same trend is observed for Lionfish and Porgies (Sparidae). Creolefish and Gray Snappers have the highest densities over high relief habitats and very low densities over other habitats. Big Eyes (Priacanthidae), Sand Tilefish, and Triggerfish (Balistidae) had the highest densities over low relief rocky habitats, and low to intermediate densities over sand habitats. Lastly, Scamp (*Mycteroperca phenax*) and Squirrelfish showed roughly uniform densities over low relief and higher relief rocky habitats with much lower densities over sand habitats. When comparing these relationships to those determined in Switzer et al 2014, some of the results show

similar conclusions, while others do not. For example, both studies showed Blue Angelfish and some Porgies to be associated with higher relief features. Red Grouper and Lionfish however, were found to be uniformly distributed over different hard-bottom habitats in that study, while Red Grouper were found to be most abundant over low relief rocky habitats, and Lionfish showed a preference for more complex habitats with the greatest densities over high relief rocky habitats. Moreover, Switzer found more Sand Tilefish over higher relief features, while this study found them to be most abundant over low relief rocky features (Switzer et al., 2014). Additionally, although not within the Elbow Allee et al. (2011) was conducted on the WFS in the Madison Swanson area and found Creolefish to be associated with higher relief features which aligns with what was found in both this study and Switzer et al. (2014).

Conclusion

This chapter demonstrated the utility of using a towed camera system to find trends in fish communities and for determining habitat-specific fish densities. Although overlap occurred significant differences were found among fish communities in differing geologic and biotic habitats, and the species driving those differences were determined. Fish densities and species richness were generally higher in more complex habitats (rockier, higher relief, presence of attached fauna), and the likelihood of observing fish greatly increased with increasing habitat complexity. The towed camera system also proved effective for getting precise estimates of habitat-specific densities for some species. Species such as Blue Angelfishes, Bigeyes, Lionfish, Sand Tilefish, and Squirrelfish that were frequently observed and tended to swim alone or in small groups could be estimated most precisely, while species that were rarely observed or exhibited schooling behavior such as the Creolefish and Gray Snapper had less precise estimates of habitat-specific densities. These large schools often occurred over high relief areas, so

dedicating greater sampling effort to these areas would likely aid in more precise estimates (Cochran, 1977).

Future Work

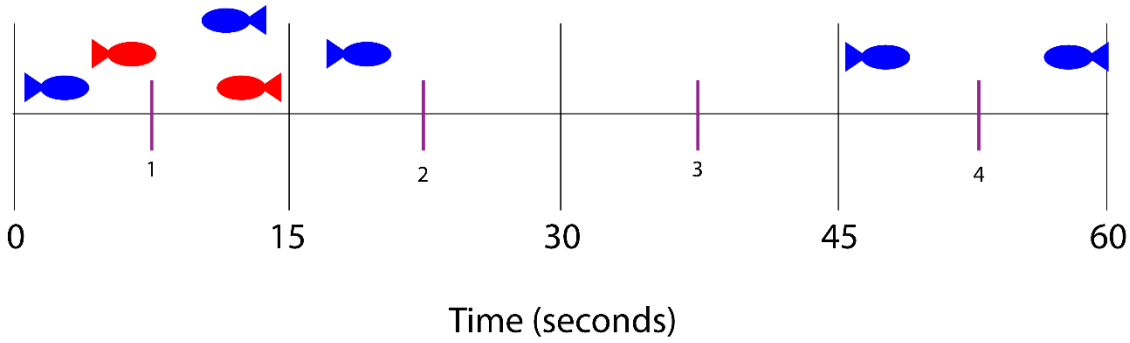
Over the last several years the C-SCAMP project has collected a large volume of video data on the WFS. Although the scope of this analysis was restricted to one area at one time period, future analyses could compare the fish communities among different areas on the WFS, and analyze whether or not the same relationships between fish and habitat hold true in these different areas. Moreover, the effect of season and time of day could be analyzed as fish are known to have both diel patterns in habitat use, as well as seasonal migrations. For example, Gag are known to inhabit the mid-shelf during most of the year and then migrate to the outer shelf for spawning in the winter, and fish communities have been found to differ on WFS depending on whether sampling was conducted during the day or at night (Allee et al., 2011, Kilborn, 2017). Moreover, fish counts can be combined with multibeam data and oceanographic data to predict the distribution of species. Lastly, future research could investigate bias associated with observing fish communities using a towed system. Our research group has been coordinating with FWC-FWRI, which assesses reef fish communities using stationary cameras. We have conducted joint cruises that analyzed the same areas over similar times so that comparisons can be made between the observed fish communities.

Table 8: Results of the 2-way PERMAOVA assessing differences in fish communities among geologic and biotic habitats

	df	SS	MS	F	p
Geologic	2	4.6673	2.3336	6.2399	0.0001
Biotic	1	3.0857	3.0857	8.2509	0.0001
Geologic x Biotic	2	0	0	0	1
Residual	280	104.72	0.37399		
Total	285	112.47			

Table 9: Results from pairwise comparisons assessing differences in fish community composition and abundance among geologic habitat types

	t	p	p (Holm's Corrected)
Sand vs Low Relief	3.0449	0.0001	0.0003
Sand vs Moderate - High Relief	1.9699	0.0019	0.0038
Low Relief vs Moderate - High Relief	1.7387	0.0049	0.0049



Habitat Observation	Red Fish	Blue Fish	Total Fish
1	2	2	4
2	0	1	1
3	0	0	0
4	0	2	2

Figure 32: Graphical representation of how fish counted continuously along a transect are linked to the periodic (every 15 s) habitat observations represented by the numbered purple lines. The resulting species-by-site matrix is also shown.

a) b)

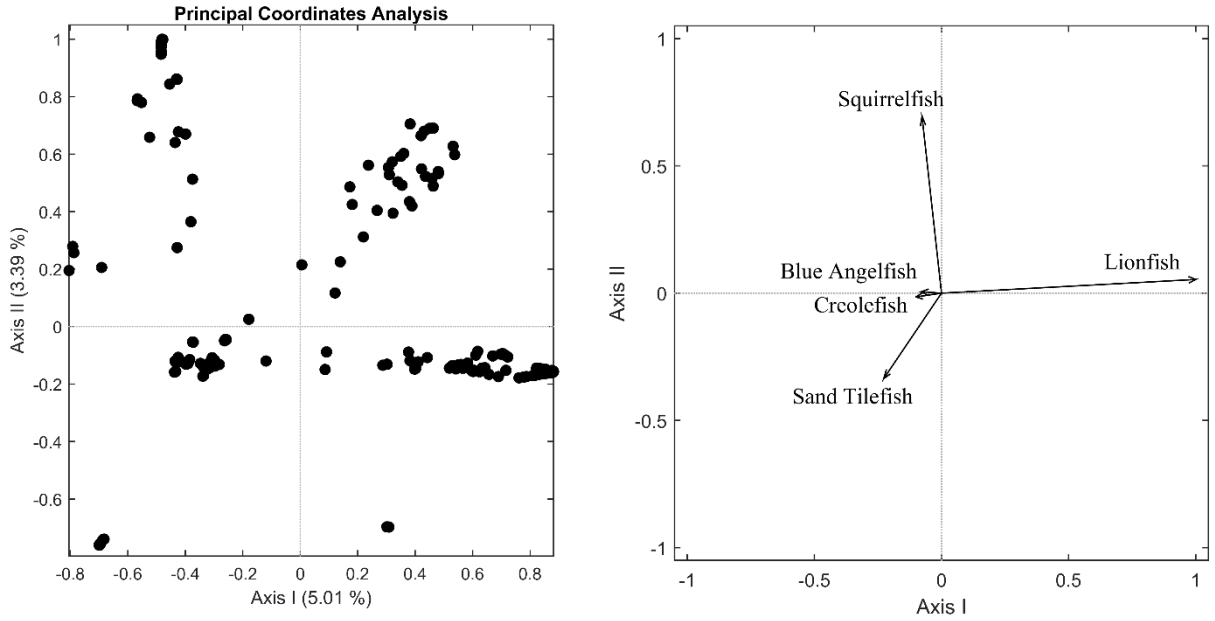
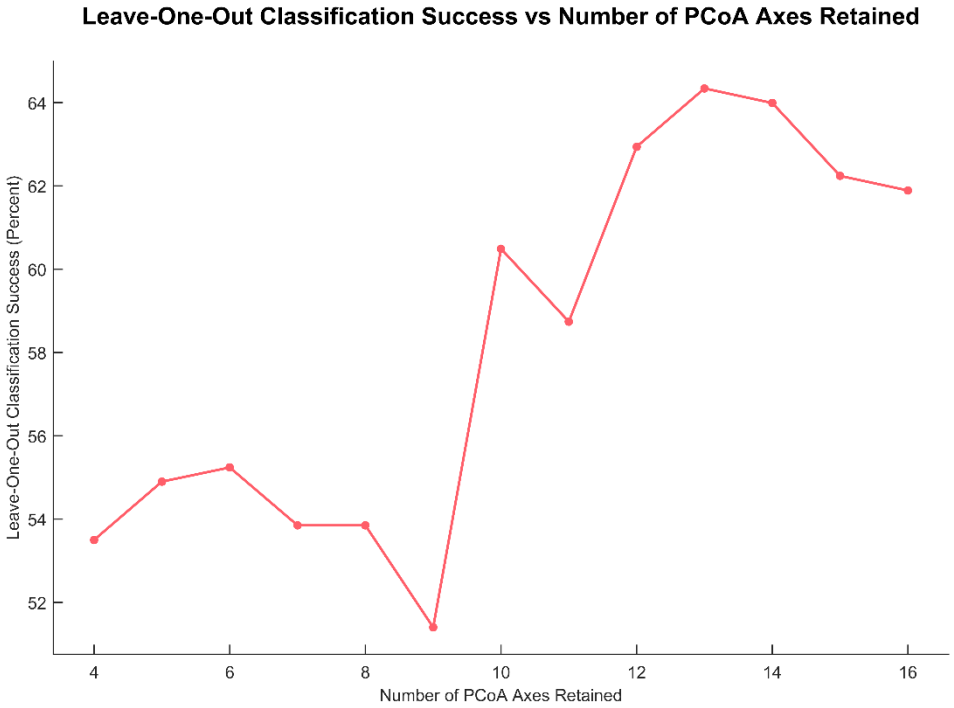


Figure 33: Plot of Principle Coordinates Analysis (a) and the corresponding species weighted species biplot vectors (b). For clarity only the 5 longest species biplot vectors are displayed.

a)



b)

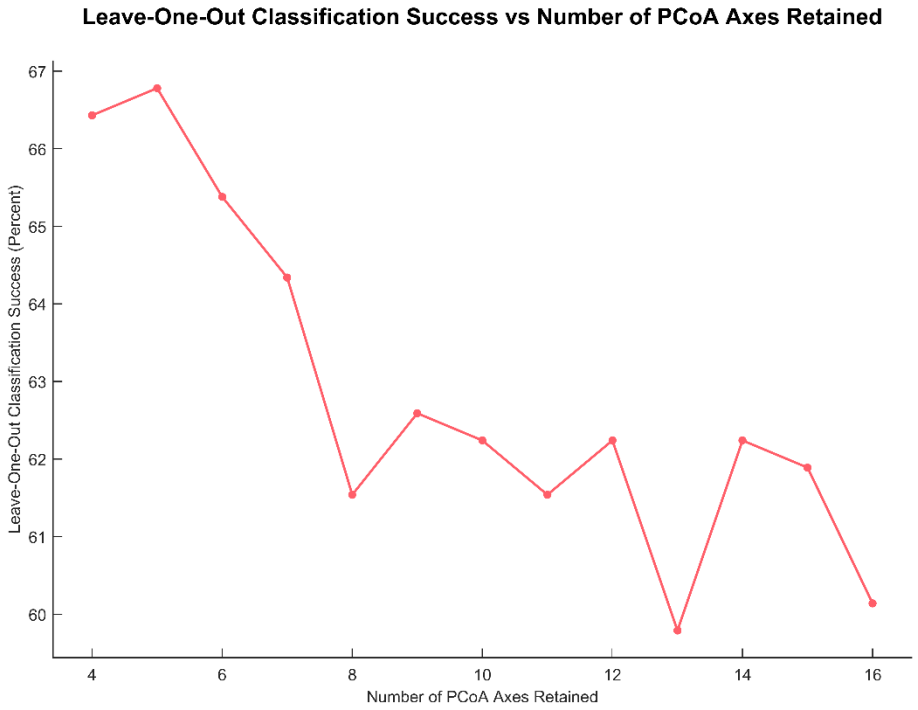


Figure 34: Plot of number of Principle Coordinates Axes retained vs LOO-CV classification accuracy to determine the optimal number of PCoA axes for the CAP analysis for geologic (a) and biotic (b) habitat

a)

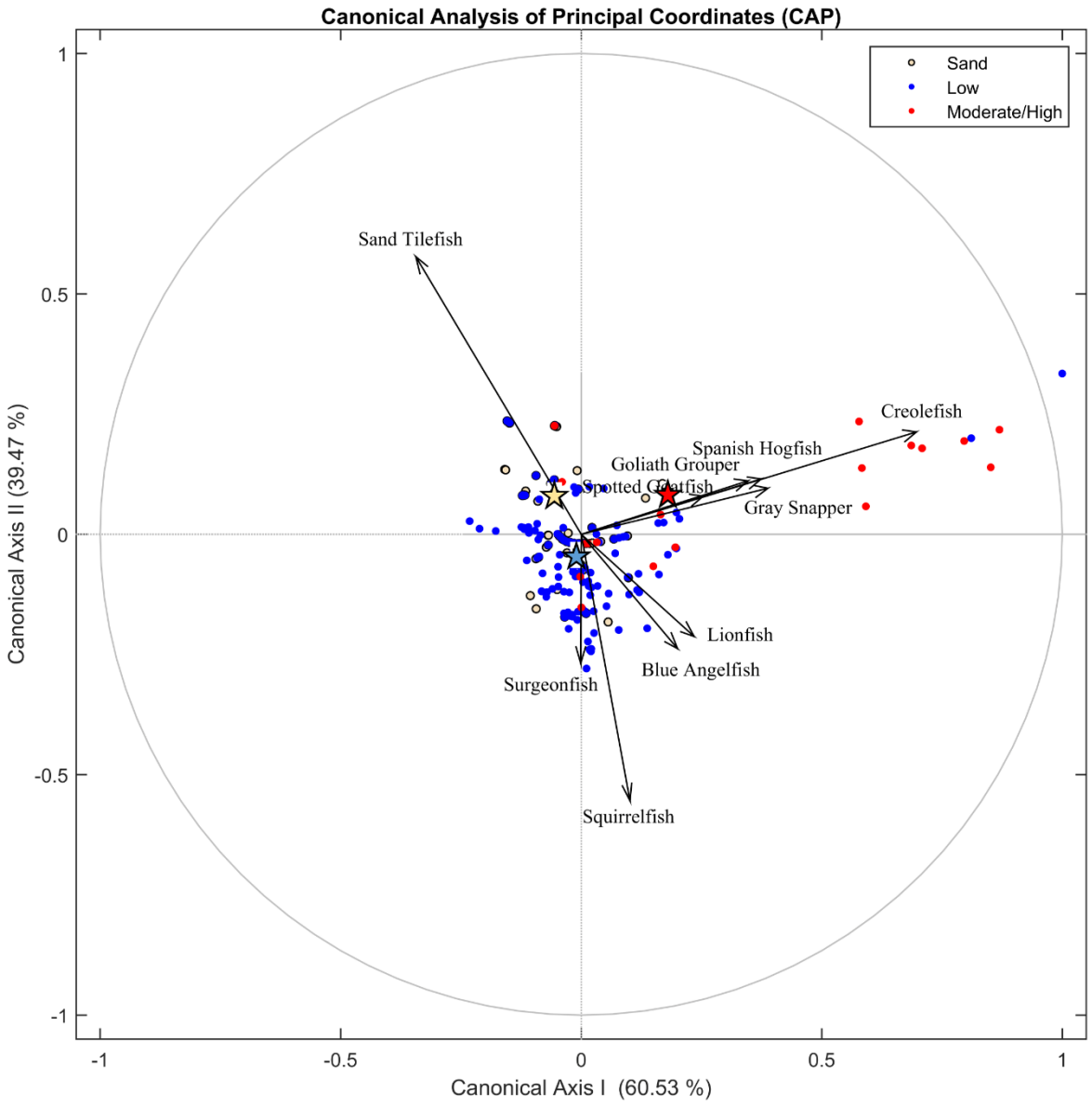

Canonical Analysis of Principal Coordinates (CAP)
Sand
Low
Moderate/High
Sand Tilefish
Creolefish
Spanish Hogfish
Goliath Grouper
Spotted Goatfish
Gray Snapper
Lionfish
Blue Angelfish
Surgeonfish
Squirrelfish
Canonical Axis II (39.47 %)
Canonical Axis I (60.53 %)
1
0.5
0
-0.5
-1
-1
-0.5
0
0.5
1

Figure 35a

b)

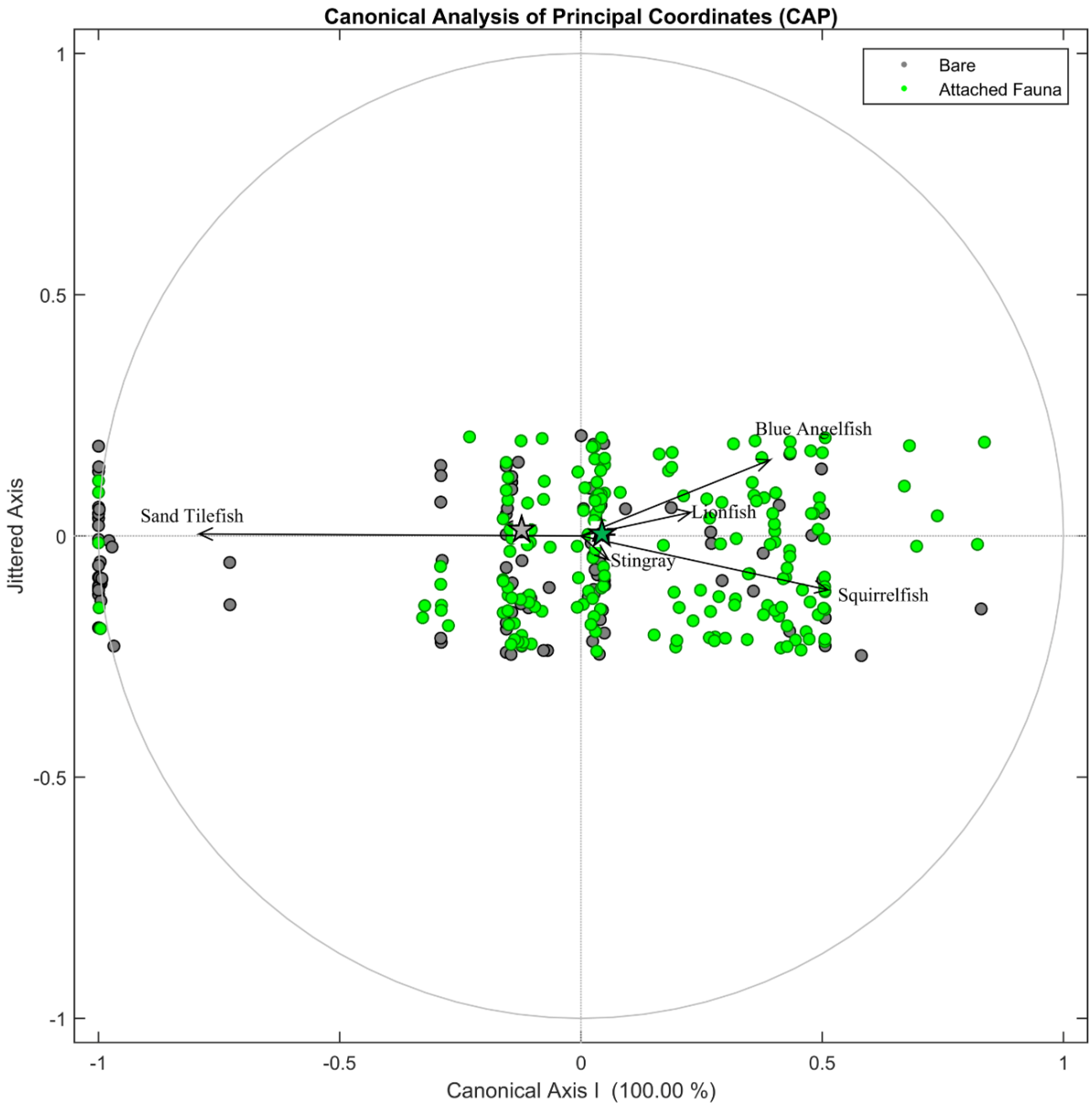


Figure 35: Plot of the CAP Analyses for geologic and biotic habitat with species correlation vectors overlaid on the plot. For clarity the 10 longest species correlation vectors are displayed for the CAP plot based on geologic habitat and the 5 longest species correlation vectors are displayed for the CAP plot based on biotic habitat. The centroids for each habitat group are represented by a star of the corresponding color for that group.

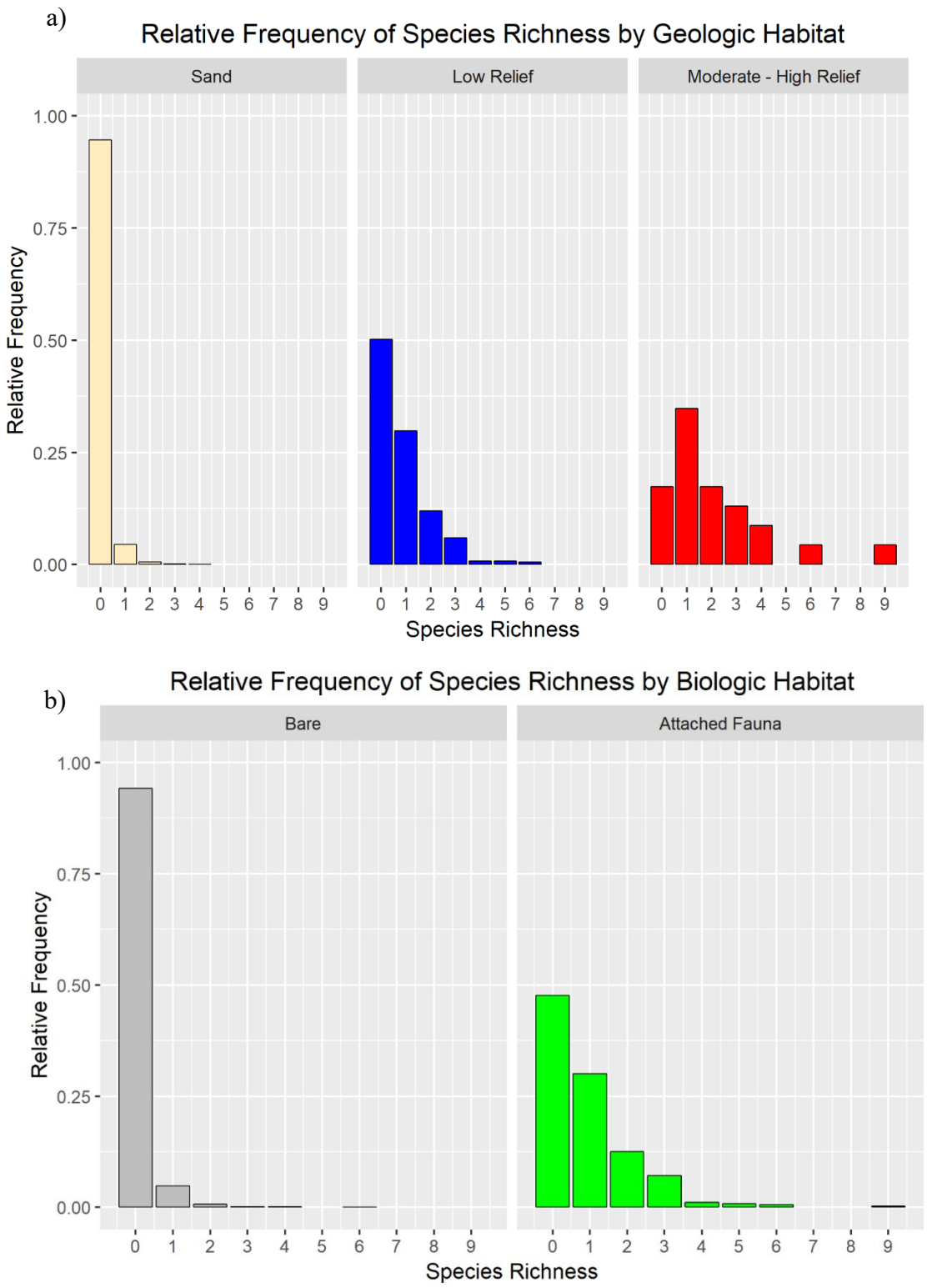


Figure 36: Plot of the relative frequency of species richness within each 15s bin by geologic (a) and biotic (b) habitat.

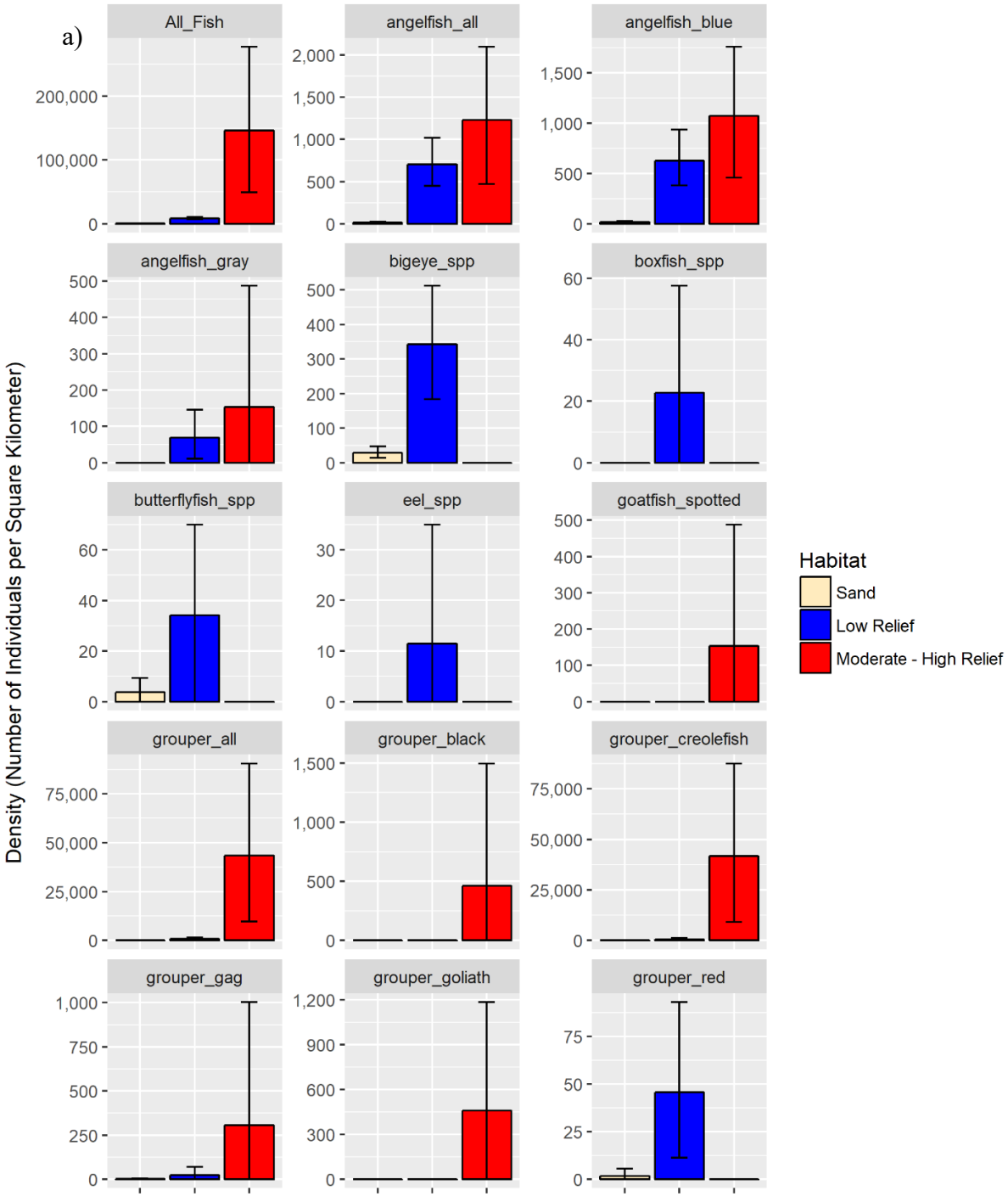

a)
All_Fish
angelfish_all
angelfish_blue
angelfish_gray
bigeye_spp
boxfish_spp
butterflyfish_spp
eel_spp
goatfish_spotted
grouper_all
grouper_black
grouper_creolefish
grouper_gag
grouper_goliath
grouper_red
Density (Number of Individuals per Square Kilometer)
Habitat
Sand
Low Relief
Moderate - High Relief

Figure 37a

b)

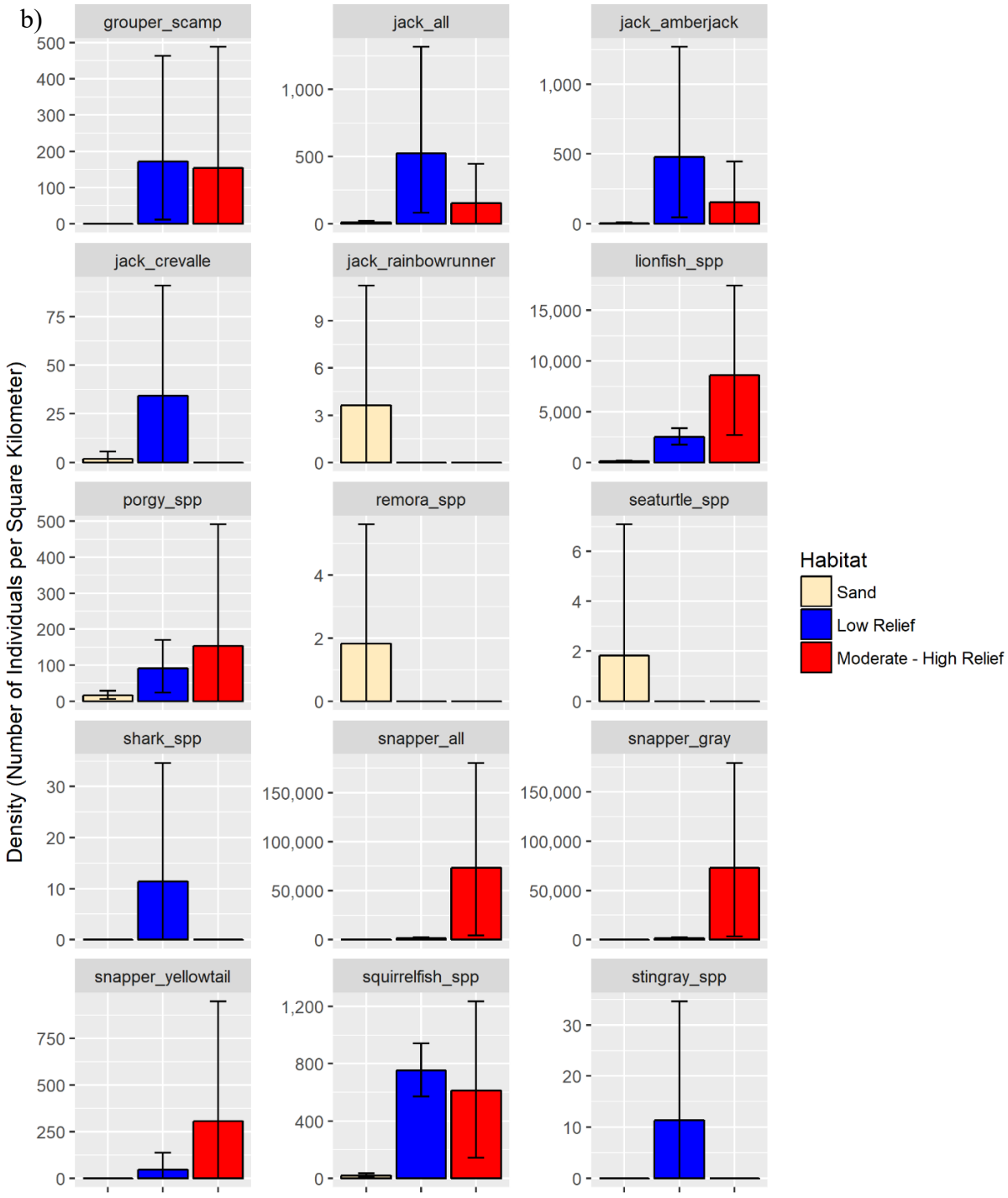

grouper_scamp
jack_all
jack_amberjack
jack_crevalle
jack_rainbowrunner
lionfish_spp
porgy_spp
remora_spp
seaturtle_spp
shark_spp
snapper_all
snapper_gray
snapper_yellowtail
squirrelfish_spp
stingray_spp
Density (Number of Individuals per Square Kilometer)
Habitat
Sand
Low Relief
Moderate - High Relief

Figure 37b

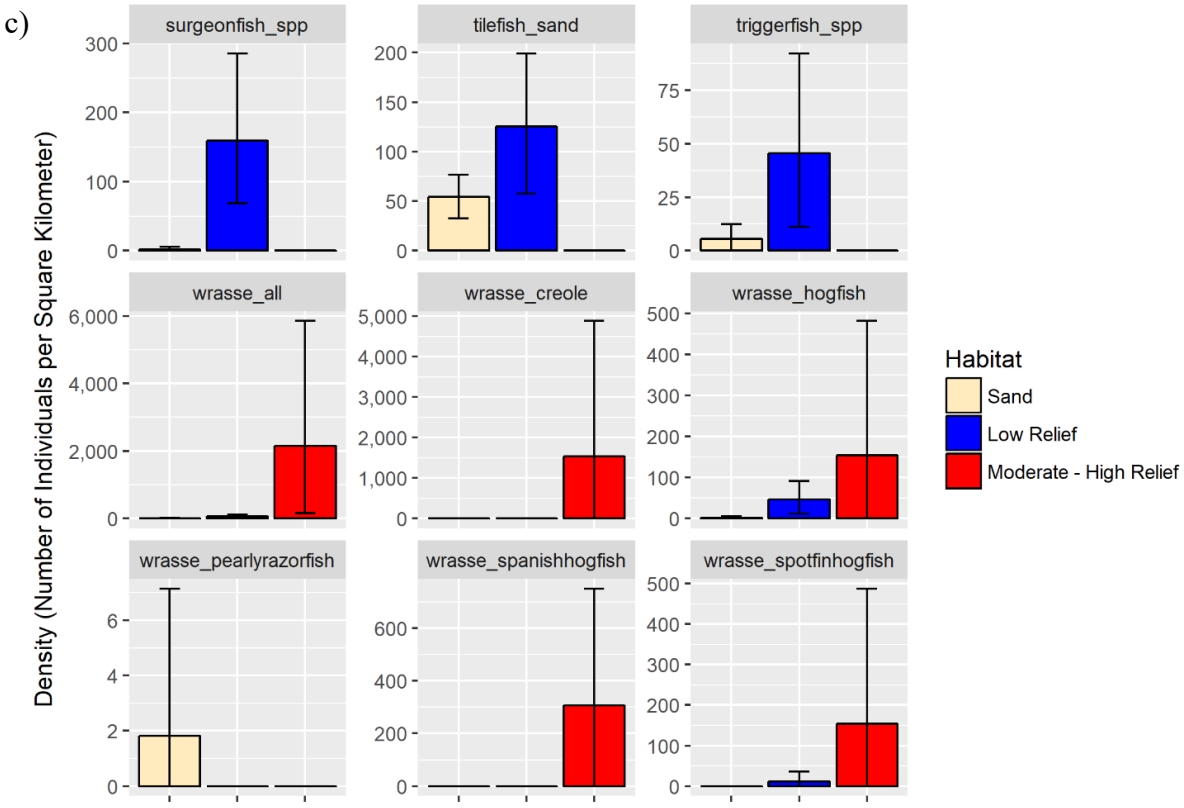


Figure 37: The habitat-specific densities for sand, low relief rock, and moderate/high relief rock determined from the C-BASS towed video transects. Error bars represent the 95% bootstrap confidence intervals. Taxa are sorted alphabetically with All_Fish-grouper_red shown in part a, grouper_scamp- stingray_spp in b, and surgeonfish-wrasse_spotfinhogfish in c.

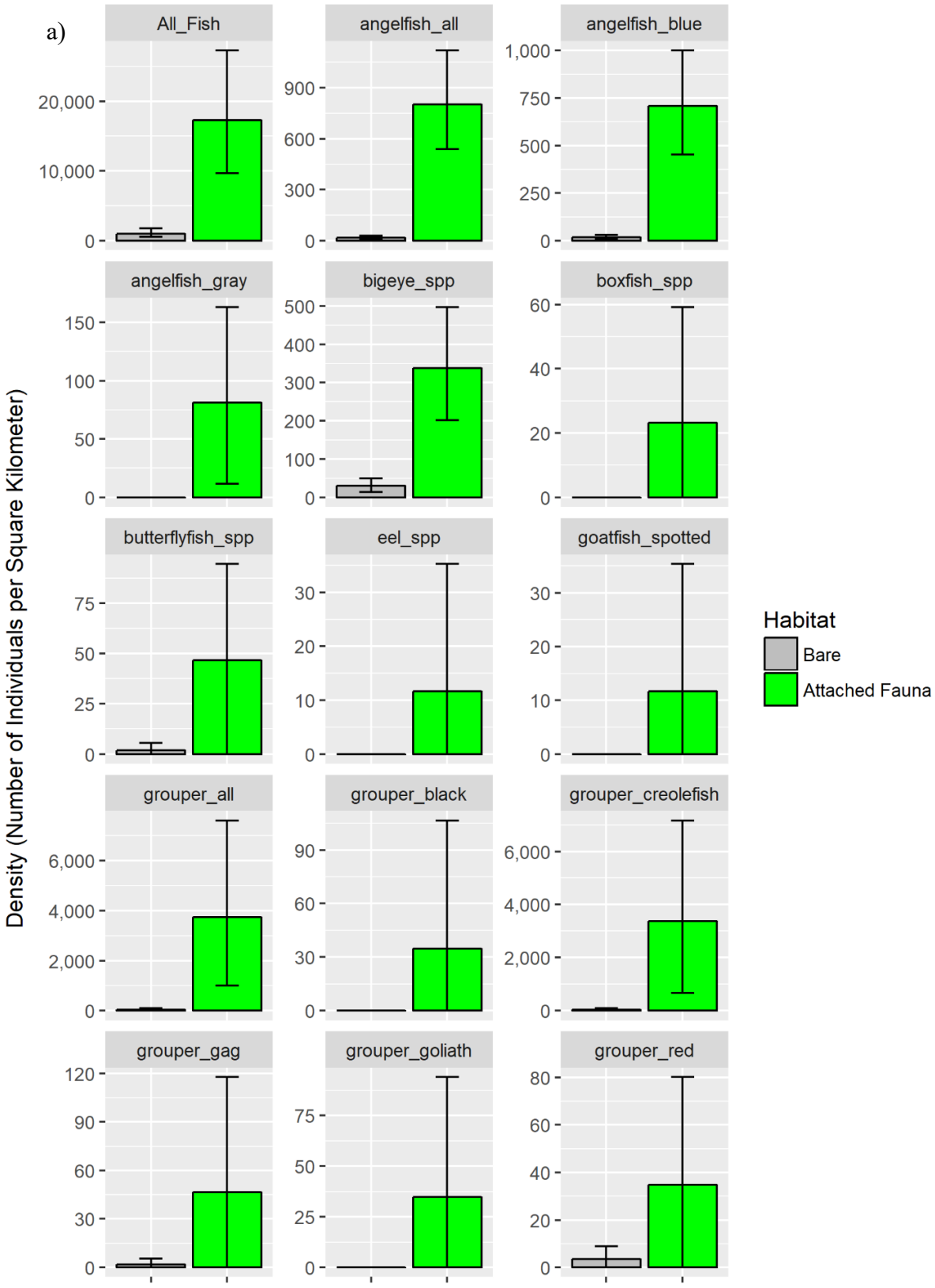

a)
All_Fish
angelfish_all
angelfish_blue
angelfish_gray
bigeye_spp
boxfish_spp
butterflyfish_spp
eel_spp
goatfish_spotted
grouper_all
grouper_black
grouper_creolefish
grouper_gag
grouper_goliath
grouper_red
Density (Number of Individuals per Square Kilometer)
Habitat
Bare
Attached Fauna

Figure 38a

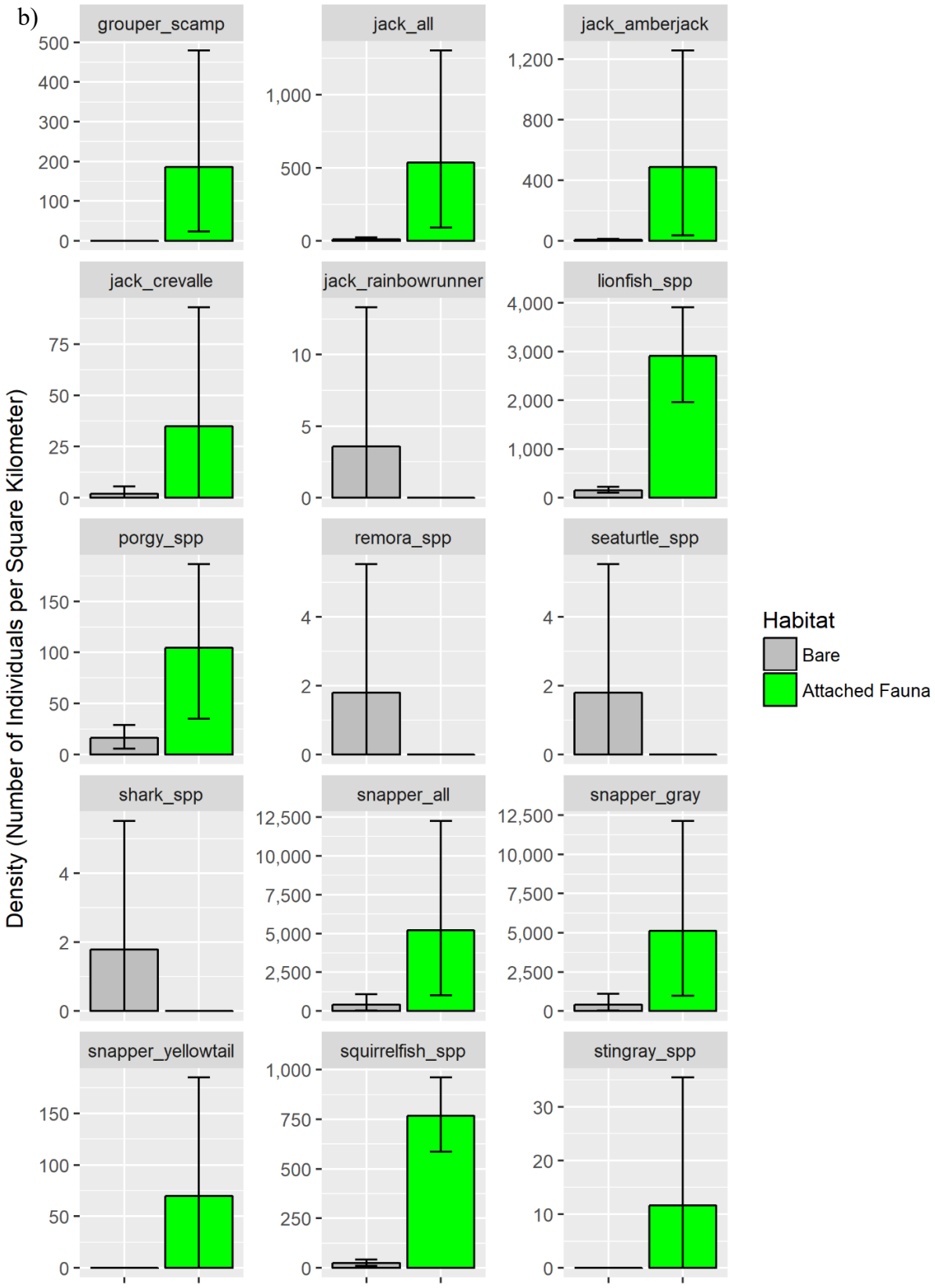

b)
grouper_scamp
jack_all
jack_amberjack
jack_crevalle
jack_rainbowrunner
lionfish_spp
porgy_spp
remora_spp
seaturtle_spp
shark_spp
snapper_all
snapper_gray
snapper_yellowtail
squirrelfish_spp
stingray_spp
Density (Number of Individuals per Square Kilometer)
Habitat
Bare
Attached Fauna

Figure 38b

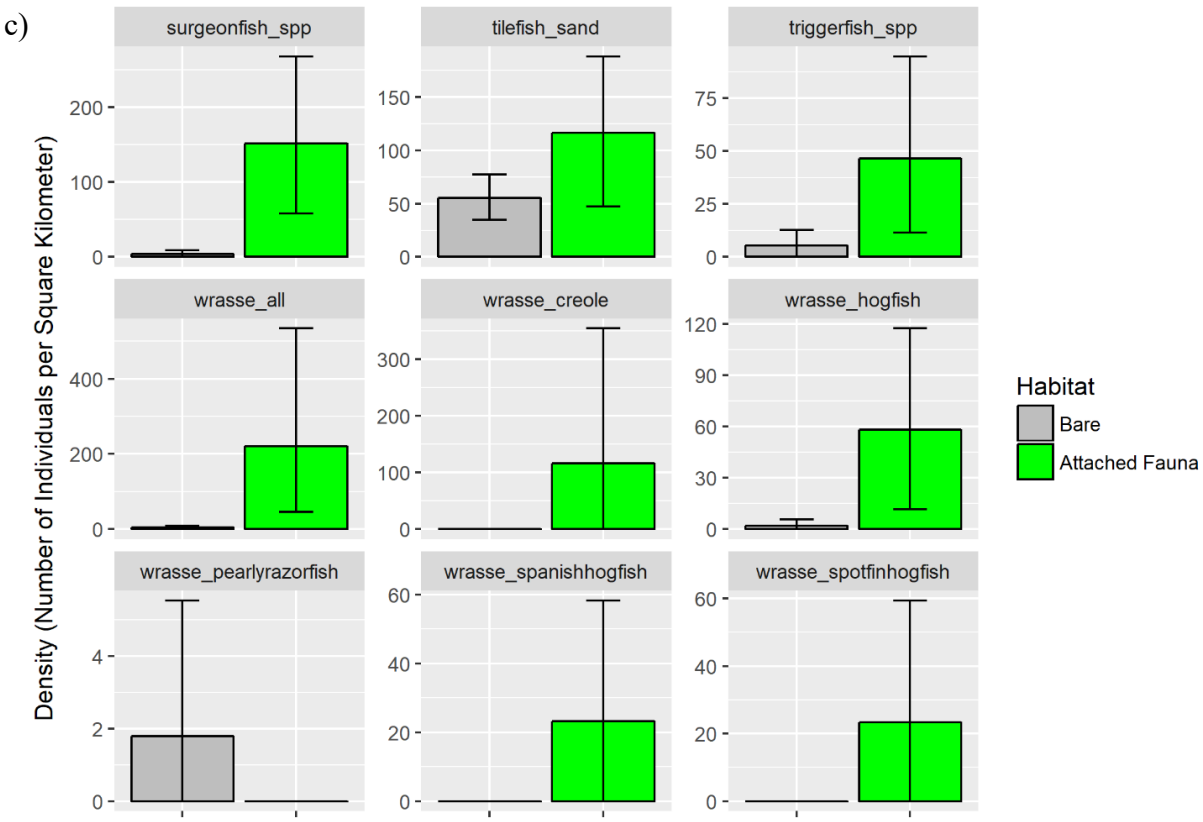


Figure 38: The habitat-specific densities for bare and attached fauna as determined from the C-BASS towed video transects. Error bars represent the 95% bootstrap confidence intervals. Taxa are sorted alphabetically with All_Fish-grouper_red shown in part a, grouper_scamp- stingray_spp in b, and surgeonfish-wrasse_spotfinhogfish in c.

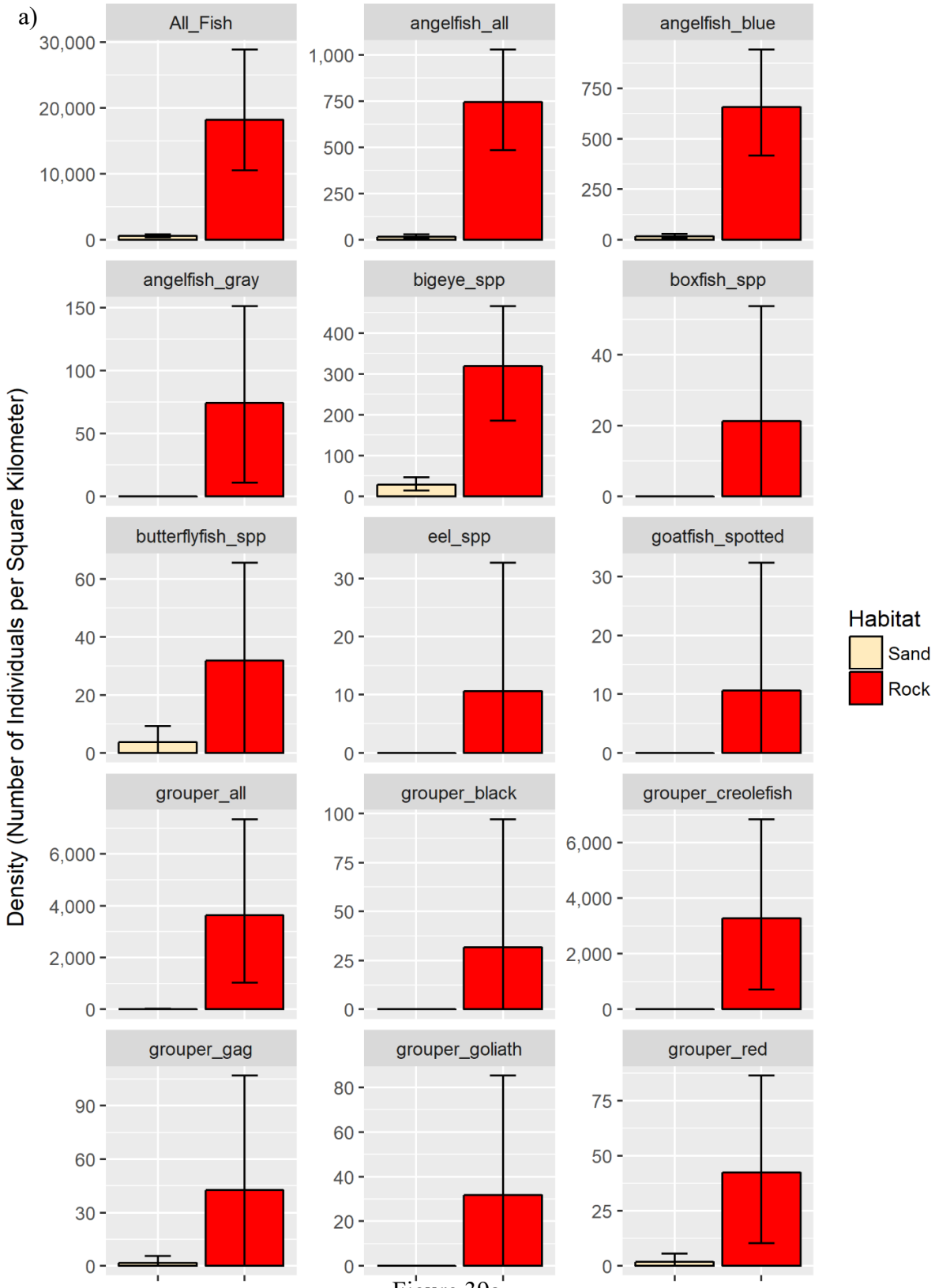

a)
All_Fish
angelfish_all
angelfish_blue
angelfish_gray
bigeye_spp
boxfish_spp
butterflyfish_spp
eel_spp
goatfish_spotted
grouper_all
grouper_black
grouper_creolefish
grouper_gag
grouper_goliath
grouper_red
Density (Number of Individuals per Square Kilometer)
Habitat
Sand
Rock

Figure 39a

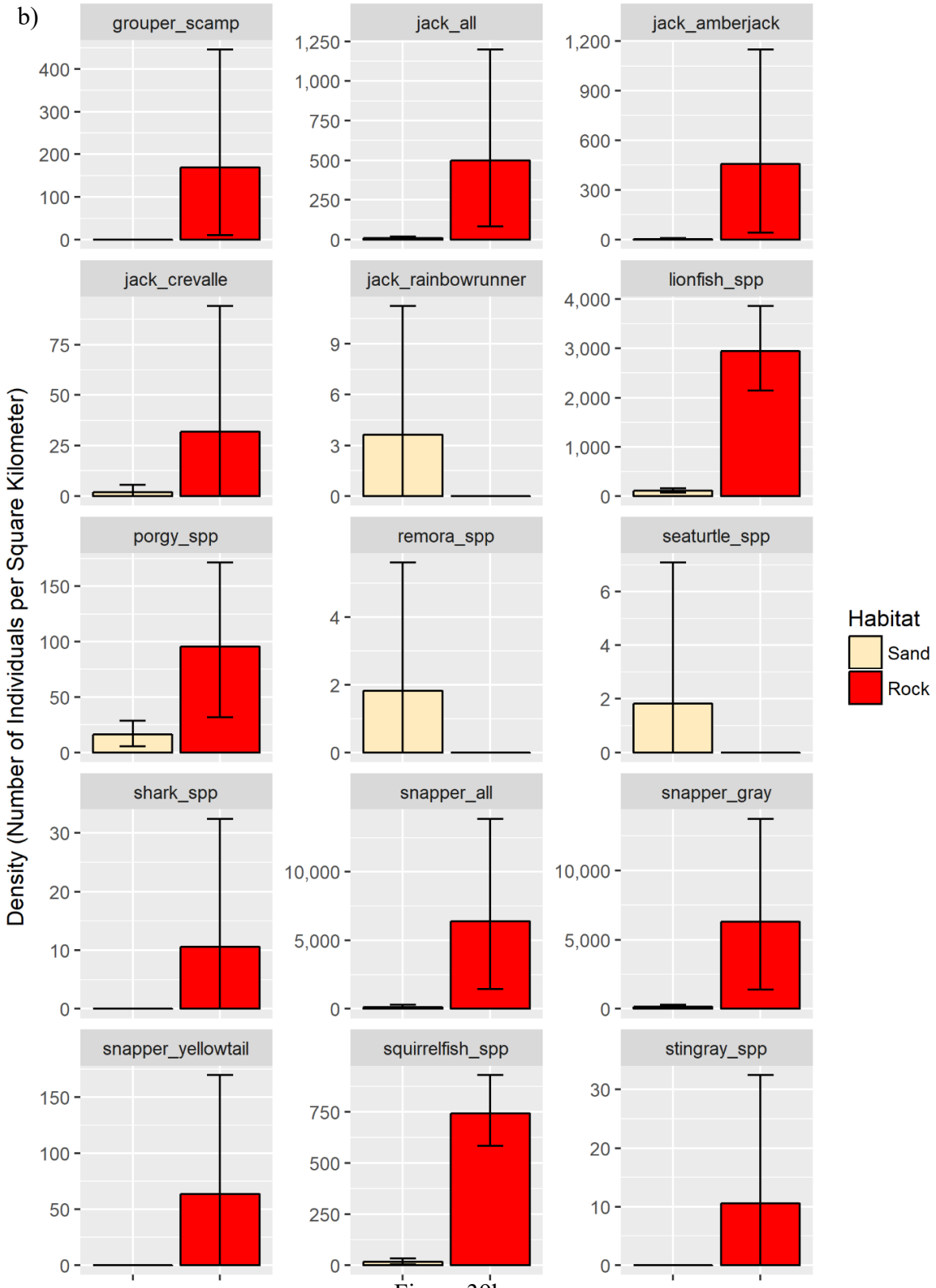

b)
grouper_scamp
jack_all
jack_amberjack
jack_crevalle
jack_rainbowrunner
lionfish_spp
porgy_spp
remora_spp
seaturtle_spp
shark_spp
snapper_all
snapper_gray
snapper_yellowtail
squirrelfish_spp
stingray_spp
Density (Number of Individuals per Square Kilometer)
Habitat
Sand
Rock

Figure 39b

c)

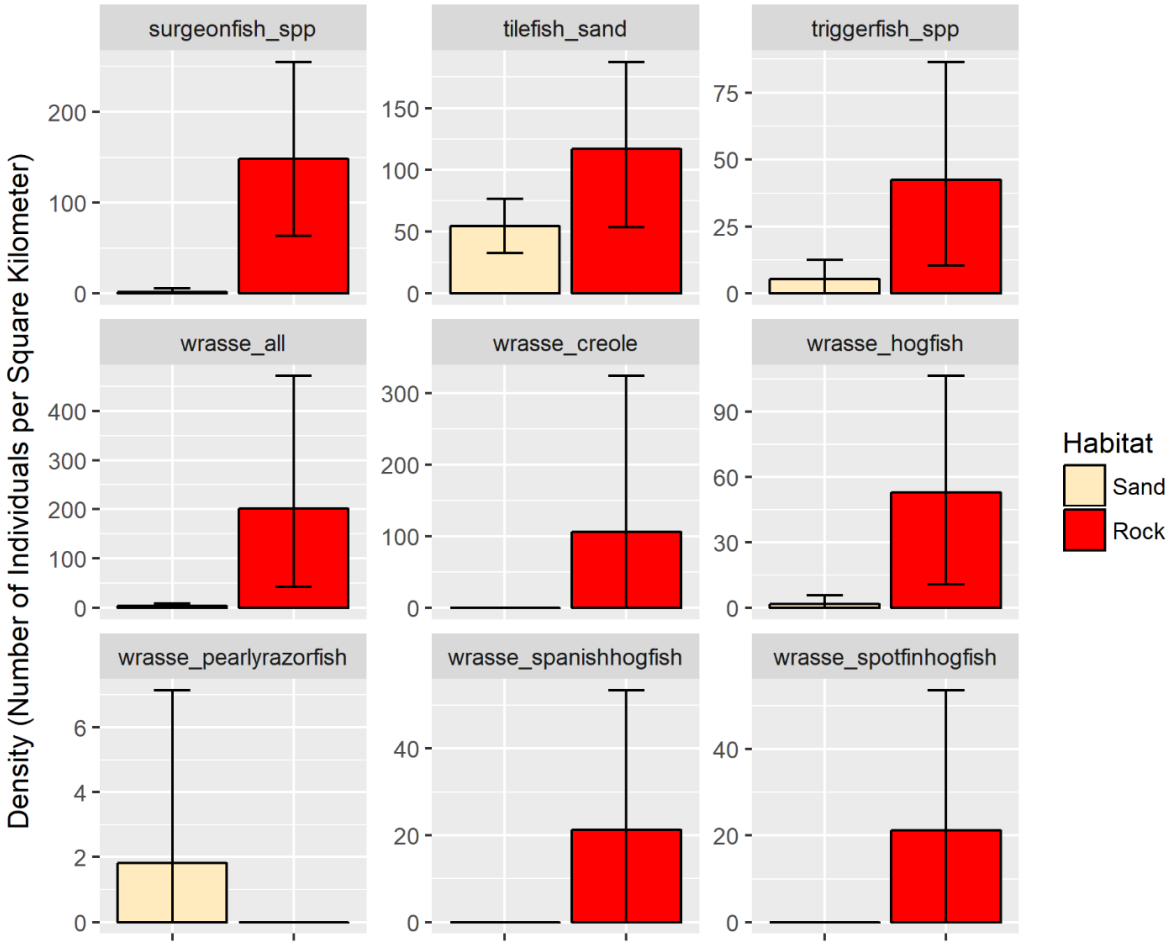


Figure 39: The habitat-specific densities for sand and rock as determined from the C-BASS towed video transects. Error bars represent the 95% bootstrap confidence intervals. Taxa are sorted alphabetically with All_Fish-grouper_red shown in part a, grouper_scamp-stingray_spp in b, and surgeonfish-wrasse_spotfinhogfish in c.

Chapter 4: Synthesis

Total Abundance Estimates

As stated earlier the three objectives of this thesis were:

1. Develop an objective and semi-automated methodology for creating full coverage habitat maps.
2. Develop quantitative relationships between fish abundance and community composition with habitat characteristics.
3. Use the results of the previous two objectives in order to estimate the abundance of various demersal reef fish.

The previous two chapters worked to answer objectives one and two. This chapter will focus on the third objective which integrates the results from the previous two objectives. In the previous two chapters we were able to calculate habitat-specific densities for fish species, as well as create habitat maps which give us the area of the different habitats. These two sources of information provide the necessary information to provide habitat stratified total abundance estimates of fish species within the study area. This can be accomplished by simply multiplying the habitat-specific densities by the area of that habitat. The total abundance can be calculated by summing the total abundances calculated for each habitat. Additionally, confidence intervals can be created by using the lower and upper bounds of the density confidence intervals in the calculations; however, it is important to note this does not take into account uncertainty in the habitat map itself. As an example, I will use the supervised habitat map of geologic habitat (Figure 22) to extrapolate abundances. The area of rock and sand from the supervised geologic

habitat map can be seen in Table 10, and the habitat-specific densities for rock vs sand can be seen in Figure 39. Combining these results via multiplication results in estimates of total abundance for observed taxa (Figure 40). There are an estimated ~111,000 fish (95% CI [67015, 169405]) within the study area that are large enough to be observed by the C-BASS. Of the ~111,000 fish, ~47,000 (~43%) are predicted to be within the sand habitat and ~64,000 (~57%) are predicted to be in the rock habitat. This demonstrates the potential of offshore rocky reefs as "critical habitats" for demersal fish in the offshore environment as just 4% of the study area is expected to contain over half of the total abundance. Additionally, sand habitats despite sustaining lower densities of fish contribute substantially to the total number of individuals due to its much larger area.

This method represents a simple way of combining these two results; however, to get truly representative estimates of density and abundance the catchability (proportion of fish observed by the system) of the C-BASS for each species must be determined. Ongoing work with FWRI through the use of paired experiments using stationary and towed cameras is being done to address this. Additionally, more complicated analyses can be done. For example, other factors such as vertical relief and biotic habitat could be considered. Also, more sophisticated techniques that take into account spatial relationships and multiple scales could be utilized as the relationships between fish and habitats can differ depending on the scale of inquiry, and there may be interactions among the different scales (Wiens, 1989, Levin, 1992). For example, in continental shelf environments off the coast of California it has been found that many relationships between fish and habitats at small-scales (one - 10's of meters) depend on the broader scale context (10-100's of meters), and that many fish species within a broad-scale habitat had different small-scale habitat associations (Anderson et al., 2009). Appreciating the

interactions of different scales and accounting for the configuration and heterogeneity of habitats is well established in studies of the terrestrial environment, but has been less studied in the marine environment. This new and evolving field known as "seascape ecology" and is largely based on its terrestrial counterpart landscape ecology (Pittman, 2013). Incorporating seascape ecology techniques as well as geostatistical methods will help make future analyses more robust and allow for the examination of more exciting and spatially explicit questions. Improved habitat maps that incorporate analyses at multiple scales will aid in conducting these types of analyses. Lastly, as habitat maps become more commonly used in conservation and management, addressing the full propagation of uncertainty throughout all analyses represents a difficult but important challenge to be addressed (Lecours, 2017).

Table 10: Area of rock vs sand habitat in km^2 and percentage of total area within the study area based on the supervised geologic habitat map

	Area (km^2)	Area (% of Study Area)
Rock	3.49	4.01
Sand	83.57	95.99

a)

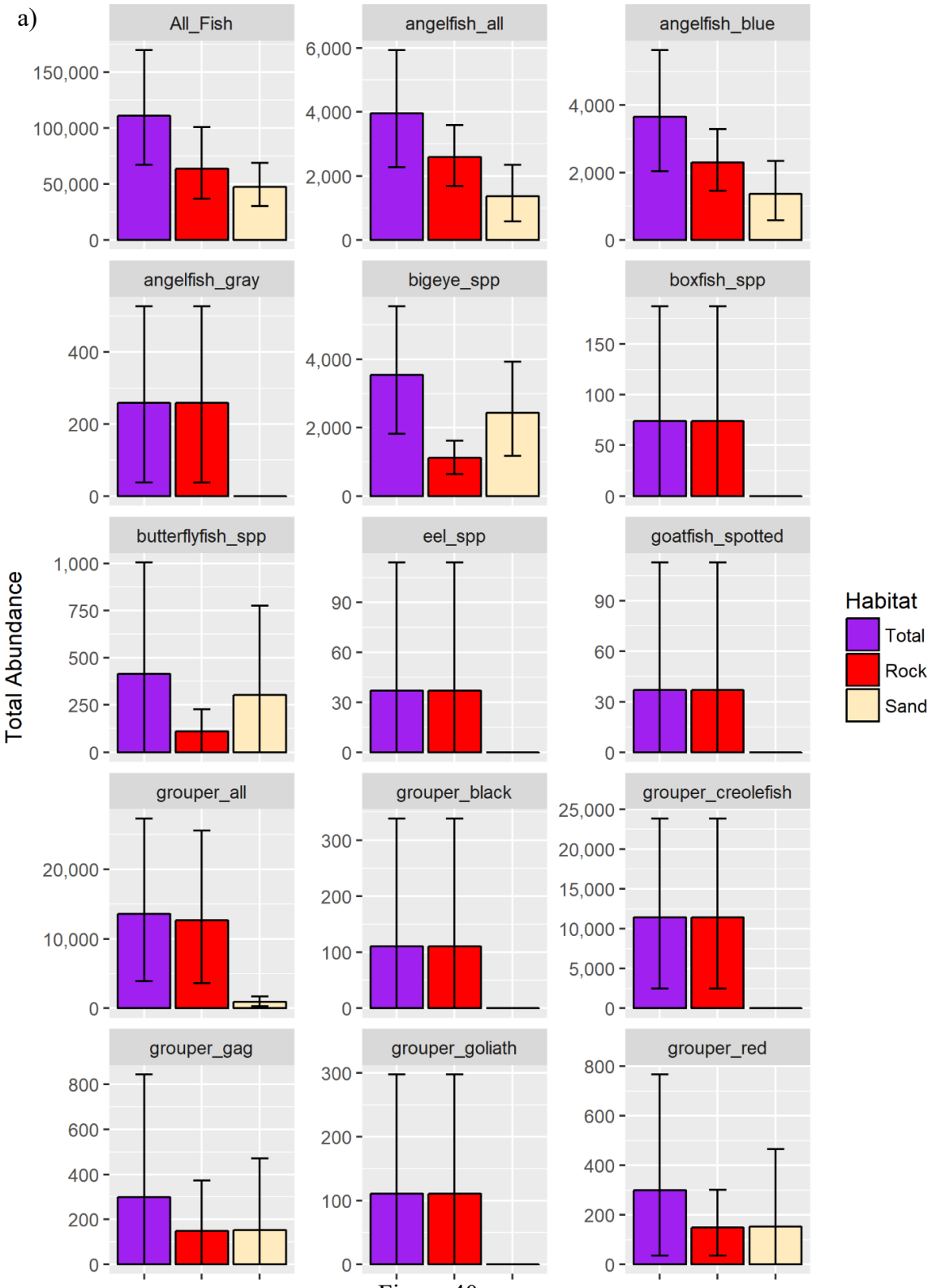

All_Fish
angelfish_all
angelfish_blue
angelfish_gray
bigeye_spp
boxfish_spp
butterflyfish_spp
eel_spp
goatfish_spotted
grouper_all
grouper_black
grouper_creolefish
grouper_gag
grouper_goliath
grouper_red
Total Abundance
Habitat
Total
Rock
Sand

Figure 40a

b)

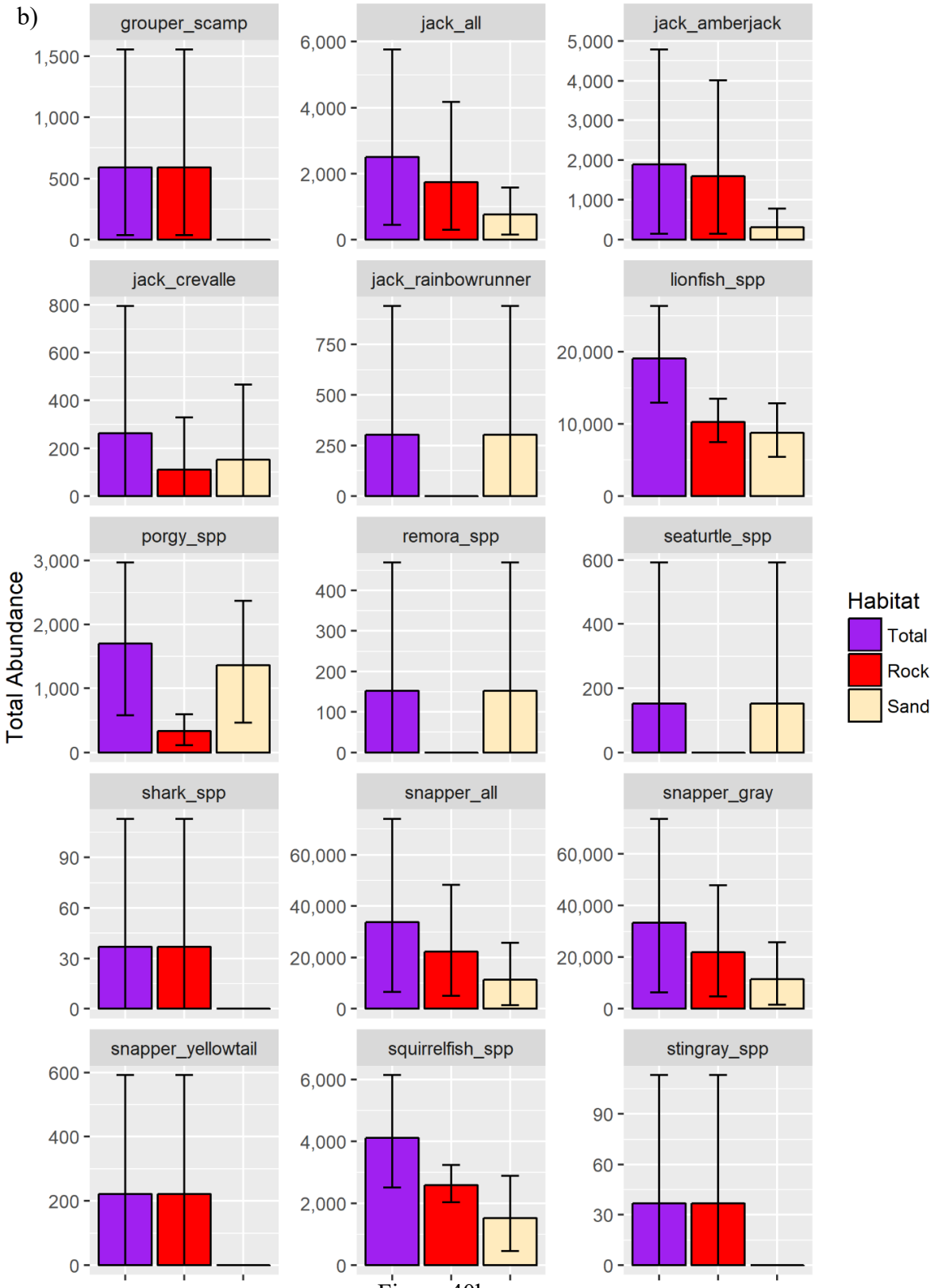

grouper_scamp
jack_all
jack_amberjack
jack_crevalle
jack_rainbowrunner
lionfish_spp
porgy_spp
remora_spp
seaturtle_spp
shark_spp
snapper_all
snapper_gray
snapper_yellowtail
squirrelfish_spp
stingray_spp
Total Abundance
Habitat
Total
Rock
Sand

Figure 40b

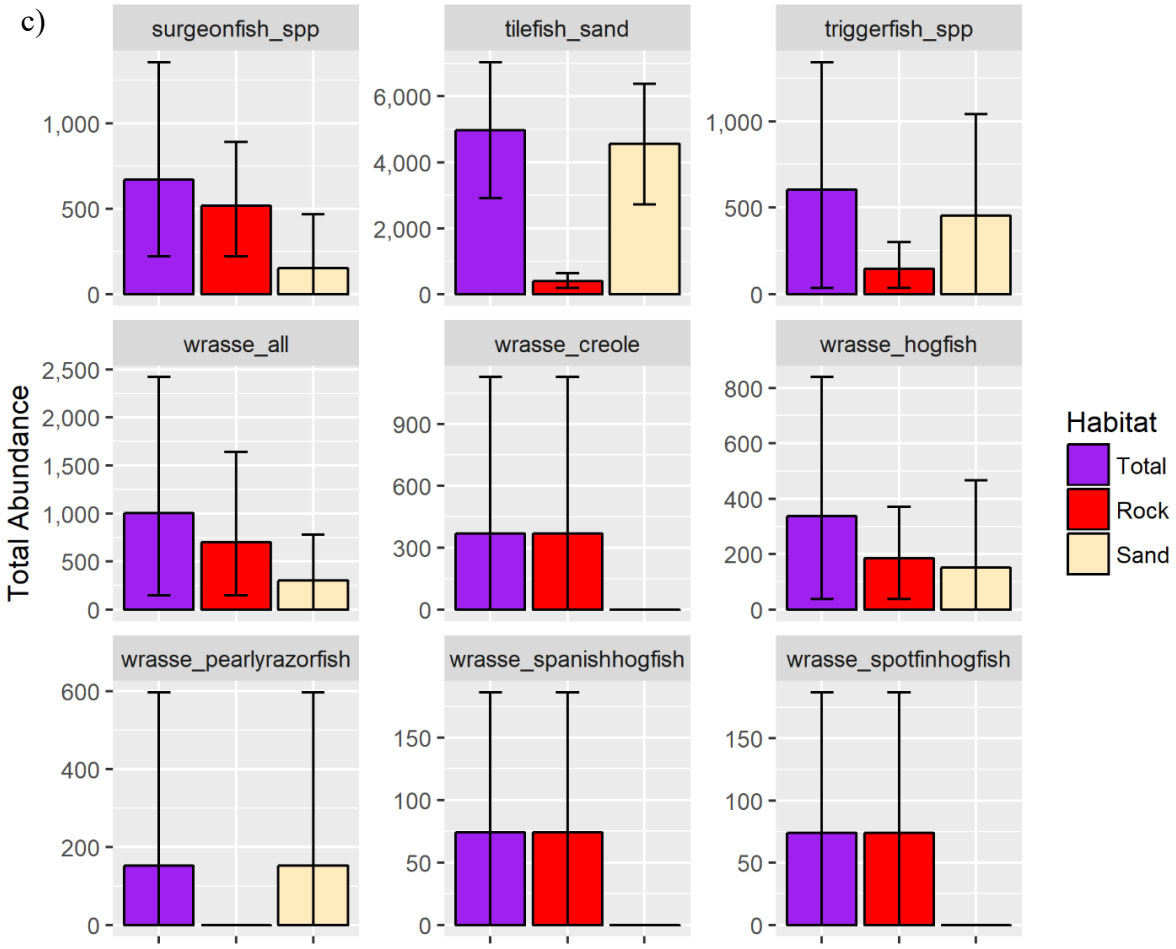


Figure 40: Estimates of total abundance within the study area for each taxa observed by the C-BASS. The predicted contribution of sand and rock to the estimated total abundance estimates is also shown. Extrapolations are based on the area of sand vs rock determined in the geologic habitat map created using the supervised methodology in chapter 1, and the habitat-specific densities over sand and rock determined in chapter 2. Error bars represent the 95% bootstrap confidence intervals. Taxa are sorted alphabetically with All_Fish-grouper_red shown in part a, grouper_scamp-stingray_spp in b, and surgeonfish-wrasse_spotfinhogfish in c.

References

ALLEE, R. J., DAVID, A. W. & NAAR, D. F. 2011. Two Shelf-Edge Marine Protected Areas in the Eastern Gulf of Mexico. Seafloor Geomorphology as Benthic Habitat: GeoHAB Atlas of Seafloor Geomorphic Features and Benthic Habitats**,** 435.

ALMANY, G. R. 2004. Does increased habitat complexity reduce predation and competition in coral reef fish assemblages? Oikos, 106**,** 275-284.

ANDERSEN, J. H., MANCA, E., AGNESI, S., AL-HAMDANI, Z., LILLIS, H., MO, G., POPULUS, J., REKER, J., TUNESI, L. & VASQUEZ, M. 2018. European Broad-Scale Seabed Habitat Maps Support Implementation of Ecosystem-Based Management. Open Journal of Ecology, 8**,** 86-103.

ANDERSON, M. J. 2001. A new method for non-parametric multivariate analysis of variance. Austral ecology, 26**,** 32-46.

ANDERSON, M. J. 2006. Distance-based tests for homogeneity of multivariate dispersions. Biometrics, 62**,** 245-253.

ANDERSON, M. J. & WILLIS, T. J. 2003. Canonical analysis of principal coordinates: a useful method of constrained ordination for ecology. Ecology, 84**,** 511-525.

ANDERSON, T. J., SYMS, C., ROBERTS, D. A. & HOWARD, D. F. 2009. Multi-scale fish–habitat associations and the use of habitat surrogates to predict the organisation and abundance of deep-water fish assemblages. Journal of Experimental Marine Biology and Ecology, 379**,** 34-42.

APPLANIX. 2017. POS MV Oceanmaster Specifications. Available: https://www.applanix.com/downloads/products/specs/posmv/POSMV_OCEANMASTER.pdf.

AUSTER, P. J., JOY, K. & VALENTINE, P. C. 2001. Fish species and community distributions as proxies for seafloor habitat distributions: the Stellwagen Bank National Marine Sanctuary example (Northwest Atlantic, Gulf of Maine). Environmental Biology of Fishes, 60**,** 331-346.

BAS, T. 2016. RSOBIA-A new OBIA Toolbar and Toolbox in ArcMap 10. x for Segmentation and Classification.

BLASCHKE, T. 2010. Object based image analysis for remote sensing. ISPRS journal of photogrammetry and remote sensing, 65**,** 2-16.

BOARD, O. S. & COUNCIL, N. R. 2002. Effects of trawling and dredging on seafloor habitat, National Academies Press.

BOWDEN, D. A. & JONES, D. O. B. 2016. Towed Camera Systems. In: CLARK, M. R., CONSALVEY, M. & ROWDEN, A. A. (eds.) Biological Sampling in the Deep Sea.

BREIMAN, L. 2001. Random forests. Machine learning, 45**,** 5-32.

BREIMAN, L. & CUTLER, A. 2008. Random forests—Classification manual. URL http://www. math. usu. edu/~ adele/forests.

BRIZZOLARA, J. 2017. Characterizing Benthic Habitat Using Multibeam Sonar and Towed Underwater Video in Two Marine Protected Areas on the West Florida Shelf, US. M.S., University of South Florida.

BROWN, C. J., SAMEOTO, J. A. & SMITH, S. J. 2012. Multiple methods, maps, and management applications: purpose made seafloor maps in support of ocean management. Journal of Sea Research, 72**,** 1-13.
BROWN, C. J., SMITH, S. J., LAWTON, P. & ANDERSON, J. T. 2011. Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. Estuarine, Coastal and Shelf Science, 92**,** 502-520.
BURNETT, C. & BLASCHKE, T. 2003. A multi-scale segmentation/object relationship modelling methodology for landscape analysis. Ecological modelling, 168**,** 233-249.
BÜTTNER, H. 1996. Rubble Mounds of Sand Tilefish Mala Canthus Plumieri (Bloch, 1787) and Associated Fishes in Colombia. Bulletin of Marine Science, 58**,** 248-260.
C-SCAMP. n.d.: The Continental Shelf Characterization Assessment and Mapping Project. Available: https://www.marine.usf.edu/scamp/ [Accessed].
CAPPO, M., HARVEY, E., MALCOLM, H. & SPEARE, P. 2003. Potential of video techniques to monitor diversity, abundance and size of fish in studies of marine protected areas. Aquatic Protected Areas-what works best and how do we know**,** 455-464.
CLARKE, K. & WARWICK, R. 2001. CHANGE IN MARINE COMMUNITIES.
COCHRAN, W. G. 1977. Sampling Techniques: 3d Ed, Wiley New York.
COCHRANE, G. R. 2008. Video-supervised classification of sonar data for mapping seafloor habitat. Marine habitat mapping technology for Alaska: Fairbanks, University of Alaska, Alaska Sea Grant College Program**,** 185-194.
COGAN, C. B., TODD, B. J., LAWTON, P. & NOJI, T. T. 2009. The role of marine habitat mapping in ecosystem-based management. ICES Journal of Marine Science: Journal du Conseil, 66**,** 2033-2042.
COHEN, J. 1960. A coefficient of agreement for nominal scales. Educational and psychological measurement, 20**,** 37-46.
COLEMAN, F., KOENIG, C., HUNTSMAN, G., MUSICK, J., EKLUND, A., MCGOVERN, J., SEDBERRY, G., CHAPMAN, R. & GRIMES, C. 2000. Long-lived reef fishes: the grouper-snapper complex. Fisheries, 25**,** 14-21.
COLEMAN, F. C., SCANLON, K. M. & KOENIG, C. C. 2011. Groupers on the edge: shelf edge spawning habitat in and around marine reserves of the northeastern Gulf of Mexico. The Professional Geographer, 63**,** 456-474.
COLLIER, J. & BROWN, C. 2005. Correlation of sidescan backscatter with grain size distribution of surficial seabed sediments. Marine Geology, 214**,** 431-449.
COSTELLO, M. J. 2009. Distinguishing marine habitat classification concepts for ecological data management. Marine Ecology Progress Series, 397**,** 253-268.
CUTLER, D. R., EDWARDS, T. C., BEARD, K. H., CUTLER, A., HESS, K. T., GIBSON, J. & LAWLER, J. J. 2007. Random forests for classification in ecology. Ecology, 88**,** 2783-2792.
DE'ATH, G. & FABRICIUS, K. E. 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology, 81**,** 3178-3192.
DIESING, M. 2016. Application of geobia to map the seafloor.
DIESING, M., GREEN, S. L., STEPHENS, D., LARK, R. M., STEWART, H. A. & DOVE, D. 2014. Mapping seabed sediments: comparison of manual, geostatistical, object-based image analysis and machine learning approaches. Continental Shelf Research, 84**,** 107-119.

FEDERAL GEOGRAPHIC DATA COMMITTEE 2012. Coastal and Marine Ecological Classification Standard. Publication# FGDC-STD-018-2012.
FRIEDLANDER, A. M. & PARRISH, J. D. 1998. Habitat characteristics affecting fish assemblages on a Hawaiian coral reef. Journal of Experimental Marine Biology and Ecology, 224**,** 1-30.
FRONTIER, S. 1976. Étude de la décroissance des valeurs propres dans une analyse en composantes principales: Comparaison avec le moddle du bâton brisé. Journal of Experimental Marine Biology and Ecology, 25**,** 67-75.
GOFF, J., OLSON, H. & DUNCAN, C. 2000. Correlation of side-scan backscatter intensity with grain-size distribution of shelf sediments, New Jersey margin. Geo-Marine Letters, 20**,** 43-49.
GRASTY, S. 2014. Use of a towed camera system for estimating reef fish population dynamics on the West Florida Shelf. Master's, University of South Florida.
GRATWICKE, B. & SPEIGHT, M. 2005. The relationship between fish species richness, abundance and habitat complexity in a range of shallow tropical marine habitats. Journal of fish biology, 66**,** 650-667.
GREENE, H. G., BIZZARRO, J. J., O'CONNELL, V. M. & BRYLINSKY, C. K. 2007. Construction of digital potential marine benthic habitat maps using a coded classification scheme and its application. Mapping the seafloor for habitat characterization: Geological Association of Canada Special Paper, 47**,** 141-155.
GREENE, H. G., YOKLAVICH, M. M., STARR, R. M., O'CONNELL, V. M., WAKEFIELD, W. W., SULLIVAN, D. E., MCREA, J. E. & CAILLIET, G. M. 1999. A classification scheme for deep seafloor habitats. Oceanologica acta, 22**,** 663-678.
HALL-BEYER, M. 2017. GLCM Textrure Tutorial v3.0.
HARALICK, R. M. & SHANMUGAM, K. 1973. Textural features for image classification. IEEE Transactions on systems, man, and cybernetics**,** 610-621.
HARDY, R. F., TUCKER, A. D., FOLEY, A. M., SCHROEDER, B. A., GIOVE, R. J. & MEYLAN, A. B. 2014. Spatiotemporal occurrence of loggerhead turtles (Caretta caretta) on the West Florida Shelf and apparent overlap with a commercial fishery. Canadian Journal of Fisheries and Aquatic Sciences, 71**,** 1924-1933.
HASAN, R. C., IERODIACONOU, D., LAURENSON, L. & SCHIMEL, A. 2014. Integrating multibeam backscatter angular response, mosaic and bathymetry data for benthic habitat mapping. Plos one, 9**,** e97339.
HENGL, T., HEUVELINK, G. B. & ROSSITER, D. G. 2007. About regression-kriging: from equations to case studies. Computers & geosciences, 33**,** 1301-1315.
HENGL, T., NUSSBAUM, M., WRIGHT, M. N., HEUVELINK, G. B. & GRÄLER, B. 2018. Random Forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ, 6**,** e5518.
HIJMANS, R. 2016. raster: geographic data analysis and modeling. R package ver. 2.5-8.
HINE, A. C., HALLEY, R. B., LOCKER, S. D., JARRETT, B. D., JAAP, W. C., MALLINSON, D. J., CIEMBRONOWICZ, K. T., OGDEN, N. B., DONAHUE, B. T. & NAAR, D. F. 2008. Coral reefs, present and past, on the west Florida shelf and platform margin. Coral Reefs of the USA. Springer.
HOLM, S. 1979. A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics**,** 65-70.
HORN, B. K. 1981. Hill shading and the reflectance map. Proceedings of the IEEE, 69**,** 14-47.

HYPACK 2017. Common HYPACK® Drivers: Interfacing Notes.
IERODIACONOU, D., LAURENSON, L., BURQ, S. & RESTON, M. 2007. Marine benthic habitat mapping using Multibeam data, georeferencedvideo and image classification techniques in Victoria, Australia. Journal of Spatial Science, 52**,** 93-104.
IERODIACONOU, D., SCHIMEL, A. C., KENNEDY, D., MONK, J., GAYLARD, G., YOUNG, M., DIESING, M. & RATTRAY, A. 2018. Combining pixel and object based image analysis of ultra-high resolution multibeam bathymetry and backscatter for habitat mapping in shallow marine waters. Marine Geophysical Research, 39**,** 271-288.
IHO 2008. IHO standards for hydrographic surveys. International Hydrographic Bureau Monaco.
JACKSON, D. A. 1993. Stopping rules in principal components analysis: a comparison of heuristical and statistical approaches. Ecology, 74**,** 2204-2214.
JONES, D. 2014. The Fathom Toolbox for MATLAB: software for multivariate ecological and oceanographic data analysis. College of Marine Science, University of South Florida, Tampa, FL, USA. Petersburg, FL, USA.
KENDALL, M. S. A method for investigating seascape ecology of reef fish. Proc Gulf Carib Fish Inst, 2005. 1-11.
KENDALL, M. S., BAUER, L. J. & JEFFREY, C. F. 2009. Influence of hard bottom morphology on fish assemblages of the continental shelf off Georgia, southeastern USA. Bulletin of marine science, 84**,** 265-286.
KILBORN, J. P. 2017. Investigating Marine Resources in the Gulf of Mexico at Multiple Spatial and Temporal Scales of Inquiry.
KINGON, K. 2013. Mapping, classification, and spatial variation of hardbottom habitats in the northeastern Gulf of Mexico.
KLOSER, R., WILLIAMS, A. & BUTLER, A. 2007. Exploratory surveys of seabed habitats in Australia's deep ocean using remote sensing–needs and realities. Mapping the seafloor for habitat characterization: Geological Association of Canada, Special Paper. Geological Association of Canada, Special Paper, 472007**,** 93-110.
KUHN, M. 2008. Caret package. Journal of statistical software, 28**,** 1-26.
LAMARCHE, G., ORPIN, A. R., MITCHELL, J. S. & PALLENTIN, A. 2016. Benthic Habitat Mapping. In: CLARK, M. R., CONSALVEY, M. & ROWDEN, A. A. (eds.) Biological Sampling in the Deep Sea.
LANDIS, J. R. & KOCH, G. G. 1977. The measurement of observer agreement for categorical data. biometrics**,** 159-174.
LECOURS, V. 2017. On the Use of Maps and Models in Conservation and Resource Management (Warning: Results May Vary). Frontiers in Marine Science, 4**,** 288.
LECOURS, V., DEVILLERS, R., SCHNEIDER, D. C., LUCIEER, V. L., BROWN, C. J. & EDINGER, E. N. 2015. Spatial scale and geographic context in benthic habitat mapping: review and future directions. Marine Ecology Progress Series, 535**,** 259-284.
LECOURS, V., DEVILLERS, R., SIMMS, A. E., LUCIEER, V. L. & BROWN, C. J. 2017. Towards a framework for terrain attribute selection in environmental studies. Environmental Modelling & Software, 89**,** 19-30.
LEGENDRE, P. & LEGENDRE, L. 1998. Numerical Ecology, Volume 24, (Developments in Environmental Modelling).
LEMBKE, C., GRASTY, S., SILVERMAN, A., BROADBENT, H., BUTCHER, S. & MURAWSKI, S. 2017. The Camera-Based Assessment Survey System (C-BASS): A

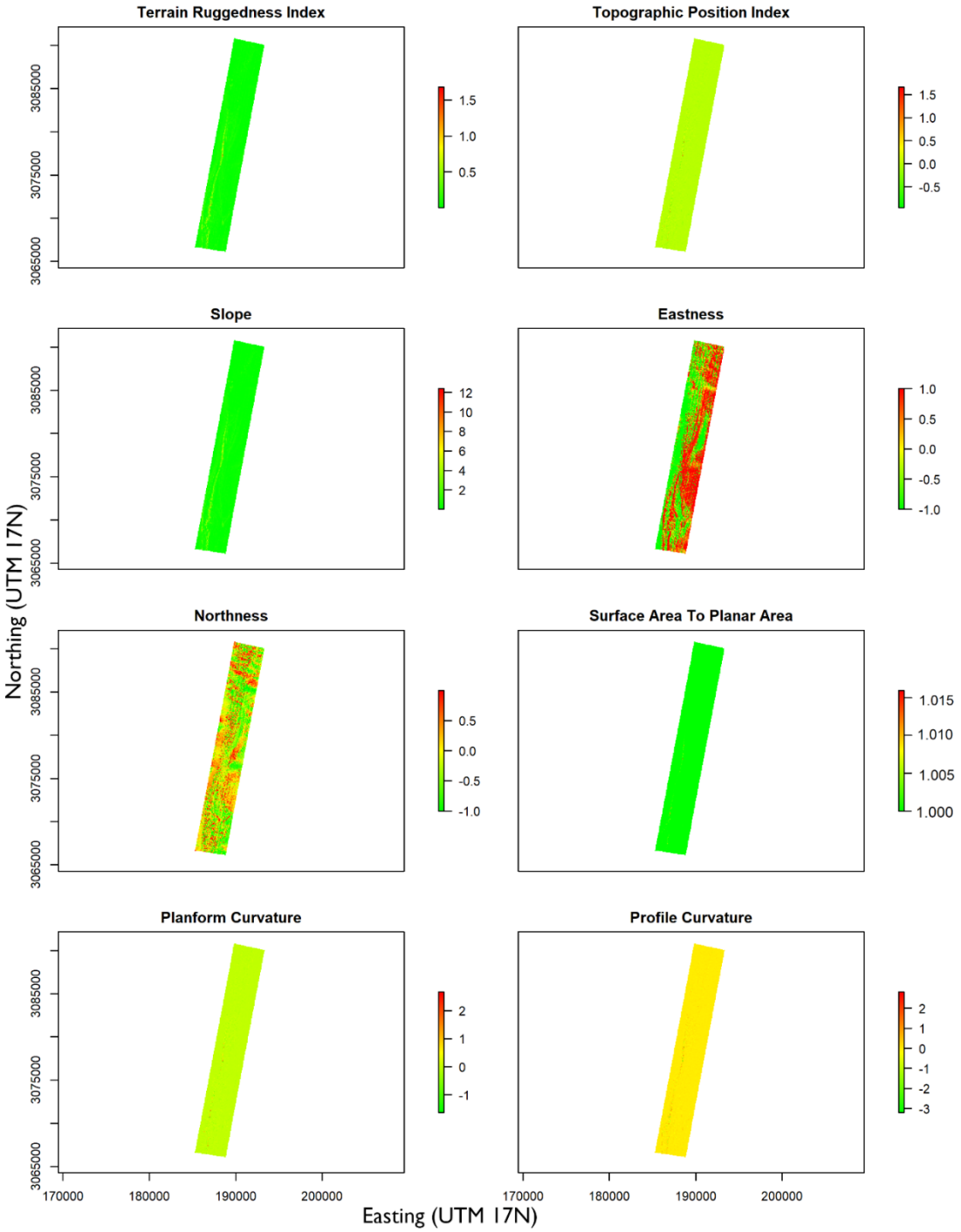
Towed Camera Platform for Reef Fish Abundance Surveys and Benthic Habitat Characterization in the Gulf of Mexico. Continental Shelf Research.
LEMBKE, C., SILVERMAN, A., BUTCHER, S., MURAWSKI, S., GRASTY, S., SHI, X. & IEEE 2013. Development and Sea Trials of a New Camera-Based Assessment Survey System for Reef Fish Stocks Assessment. 2013 Oceans - San Diego.
LEUTNER, B. & HORNING, N. 2016. RStoolbox: tools for remote sensing data analysis. R package version 0.1, 4.
LEVIN, S. A. 1992. The problem of pattern and scale in ecology: the Robert H. MacArthur award lecture. Ecology, 73**,** 1943-1967.
LIAW, A. & WIENER, M. 2002. Classification and regression by randomForest. R news, 2**,** 18-22.
LOGAN, J. M., YOUNG, M. A., HARVEY, E. S., SCHIMEL, A. C. & IERODIACONOU, D. 2017. Combining underwater video methods improves effectiveness of demersal fish assemblage surveys across habitats. Marine Ecology Progress Series, 582**,** 181-200.
LOVE, M., BALDERA, A., ROBBINS, C., SPIES, R. & ALLEN, J. 2015. Charting the Gulf: Analyzing the Gaps in Long-Term Monitoring of the Gulf of Mexico. New Orleans, LA: Ocean Conservancy.
LUCIEER, V. 2008. Object-oriented classification of sidescan sonar data for mapping benthic marine habitats. International Journal of Remote Sensing, 29**,** 905-921.
LUCIEER, V., HILL, N. A., BARRETT, N. S. & NICHOL, S. 2013. Do marine substrates 'look'and 'sound'the same? Supervised classification of multibeam acoustic data using autonomous underwater vehicle images. Estuarine, Coastal and Shelf Science, 117**,** 94-106.
MACARTHUR, R. & LEVINS, R. 1964. Competition, habitat selection, and character displacement in a patchy environment. Proceedings of the National Academy of Sciences, 51**,** 1207-1210.
MACQUEEN, J. Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, 1967. Oakland, CA, USA, 281-297.
MALLET, D. & PELLETIER, D. 2014. Underwater video techniques for observing coastal marine biodiversity: a review of sixty years of publications (1952–2012). Fisheries Research, 154**,** 44-62.
MAYER, L., JAKOBSSON, M., ALLEN, G., DORSCHEL, B., FALCONER, R., FERRINI, V., LAMARCHE, G., SNAITH, H. & WEATHERALL, P. 2018. The Nippon Foundation—GEBCO Seabed 2030 Project: The Quest to See the World's Oceans Completely Mapped by 2030. Geosciences, 8**,** 63.
MCGONIGLE, C. & COLLIER, J. S. 2014. Interlinking backscatter, grain size and benthic community structure. Estuarine, Coastal and Shelf Science, 147**,** 123-136.
MCINTYRE, F., COLLIE, N., STEWART, M., SCALA, L. & FERNANDES, P. 2013. A visual survey technique for deep-water fishes: estimating anglerfish Lophius spp. abundance in closed areasa. Journal of fish biology, 83**,** 739-753.
MCLEOD, K., LUBCHENCO, J., PALUMBI, S. & ROSENBERG, A. 2005. Scientific Consensus Statement on Marine Ecosystem-Based Management. Available: https://www.compassscicomm.org/ebm-consensus-statement-download.
MOE, M. A. 1963. A survey of offshore fishing in Florida, Florida Department of Natural Resources Marine Research Laboratory.

MUSICK, J. A. 1999. Life in the Slow Lane: Ecology and Conservation of Long-Lived Marine Animals. American Fisheries Society Symposium 23, American Fisheries Society, Bethesda, Maryland, USA.
NMFS 2017. Fisheries Economics of the United States 2015. In: COMMERCE, U. S. D. O. (ed.).
PARKER JR, R. O., CHESTER, A. J. & NELSON, R. S. 1994. A video transect method for estimating reef fish abundance, composition, and habitat utilization at Gray's Reef National Marine Sanctuary, Georgia. Fishery Bulletin, 92**,** 787-799.
PICKRILL, R. A. & TODD, B. J. 2003. The multiple roles of acoustic mapping in integrated ocean management, Canadian Atlantic continental margin. Ocean & Coastal Management, 46**,** 601-614.
PITTMAN, S., CHRISTENSEN, J., CALDOW, C., MENZA, C. & MONACO, M. 2007. Predictive mapping of fish species richness across shallow-water seascapes in the Caribbean. ecological modelling, 204**,** 9-21.
PITTMAN, S. J. 2013. Seascape ecology: A new science for the spatial information age. Marine Scientist**,** 20-23.
PORSKAMP, P., RATTRAY, A., YOUNG, M. & IERODIACONOU, D. 2018. Multiscale and Hierarchical Classification for Benthic Habitat Mapping. Geosciences, 8**,** 119.
RIGGS, S. R., SNYDER, S. W., HINE, A. C. & MEARNS, D. L. 1996. Hardbottom morphology and relationship to the geologic framework; Mid-Atlantic continental shelf. Journal of Sedimentary Research, 66**,** 830-846.
SHANNON, C. 1948. (1948)," A Mathematical Theory of Communication", Bell System Technical Journal, vol. 27, pp. 379-423 & 623-656, July & October.
SHEPHERD, J. 1988. Fish stock assessments and their data requirements. Fish population dynamics, 35**,** 50.
SHUMCHENIA, E. J. & KING, J. W. 2010. Comparison of methods for integrating biological and physical data for marine habitat mapping and classification. Continental Shelf Research, 30**,** 1717-1729.
SINCLAIR, M., ARNASON, R., CSIRKE, J., KARNICKI, Z., SIGURJONSSON, J., SKJOLDAL, H. R. & VALDIMARSSON, G. 2002. Responsible fisheries in the marine ecosystem. Fisheries Research, 58**,** 255-265.
STEPHENS, D. & DIESING, M. 2014. A comparison of supervised classification methods for the prediction of substrate type using multibeam acoustic and legacy grain-size data. PloS one, 9**,** e93950.
STONER, A. W., RYER, C. H., PARKER, S. J., AUSTER, P. J. & WAKEFIELD, W. W. 2008. Evaluating the role of fish behavior in surveys conducted with underwater vehicles. Canadian Journal of Fisheries and Aquatic Sciences, 65**,** 1230-1243.
STROBL, C. & ZEILEIS, A. 2008. Danger: High power!–exploring the statistical properties of a test for random forest variable importance.
SWITZER, T., KEENAN, S. & PURTLEBAUGH, C. 2014. Exploring the utility of side-scan sonar and experimental Z-traps in improving the efficiency of fisheries-independent surveys of reef fishes on the West Florida Shelf. In: FWC-FWRI (ed.). NOAA/NMFS Marine Fisheries Initiative (MARFIN) Final Report.
WALBRIDGE, S., SLOCUM, N., POBUDA, M. & WRIGHT, D. J. 2018. Unified geomorphological analysis workflows with benthic terrain modeler. Geosciences, 8**,** 94.

WALL, C. C., DONAHUE, B. T., NAAR, D. F. & MANN, D. A. 2011. Spatial and temporal variability of red grouper holes within Steamboat Lumps Marine Reserve, Gulf of Mexico. Marine Ecology Progress Series, 431**,** 243-254.
WEGMANN, M., LEUTNER, B. & DECH, S. 2016. Remote sensing and GIS for ecologists: using open source software, Pelagic Publishing Ltd.
WIENS, J. A. 1989. Spatial scaling in ecology. Functional ecology, 3**,** 385-397.
WILSON, M. F., O'CONNELL, B., BROWN, C., GUINAN, J. C. & GREHAN, A. J. 2007. Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. Marine Geodesy, 30**,** 3-35.
WOODWARD, B. & TAKAHASHI, J. 2017. CVision Video and Image Annotator.
ZVOLEFF, A. 2015. glcm: Calculate textures from grey-level co-occurrence matrices GLCMs in R. R package version 1.0.

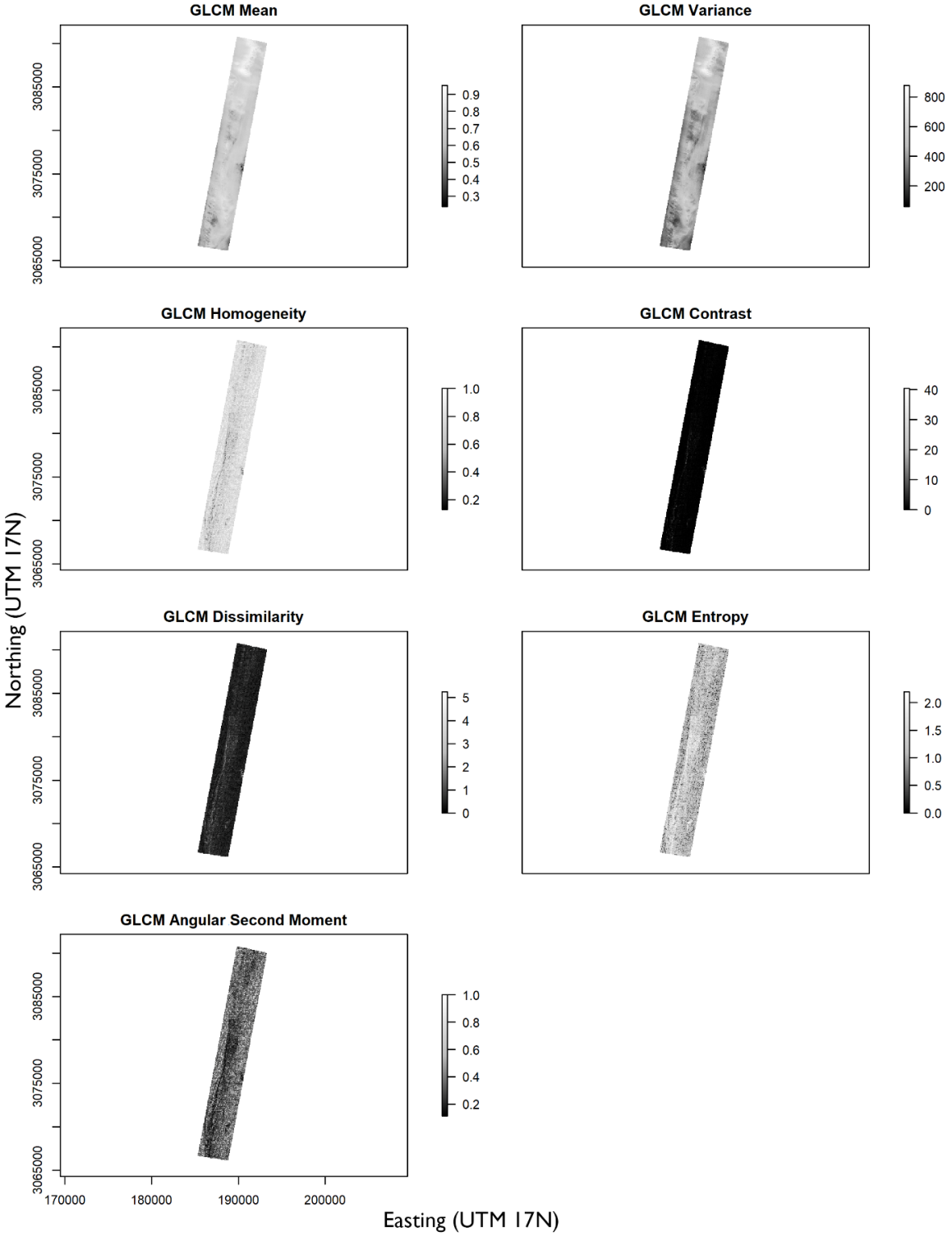
Appendices

Appendix 1: Bathymetric Derivative Features



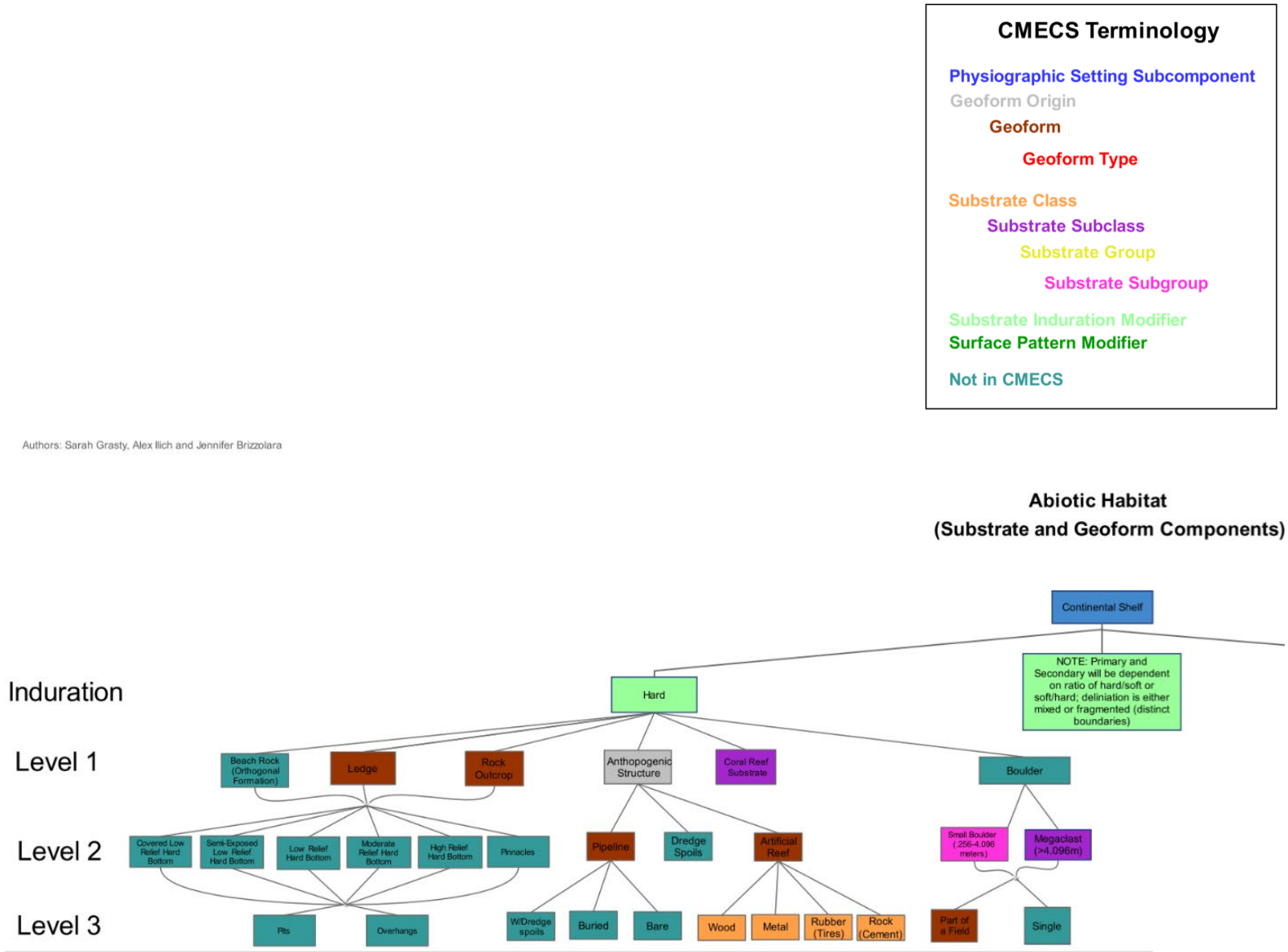
Terrain attributes derived from the bathymetry surface (10 m x 10 m resolution)

Appendix 2: Backscatter Derivative Features

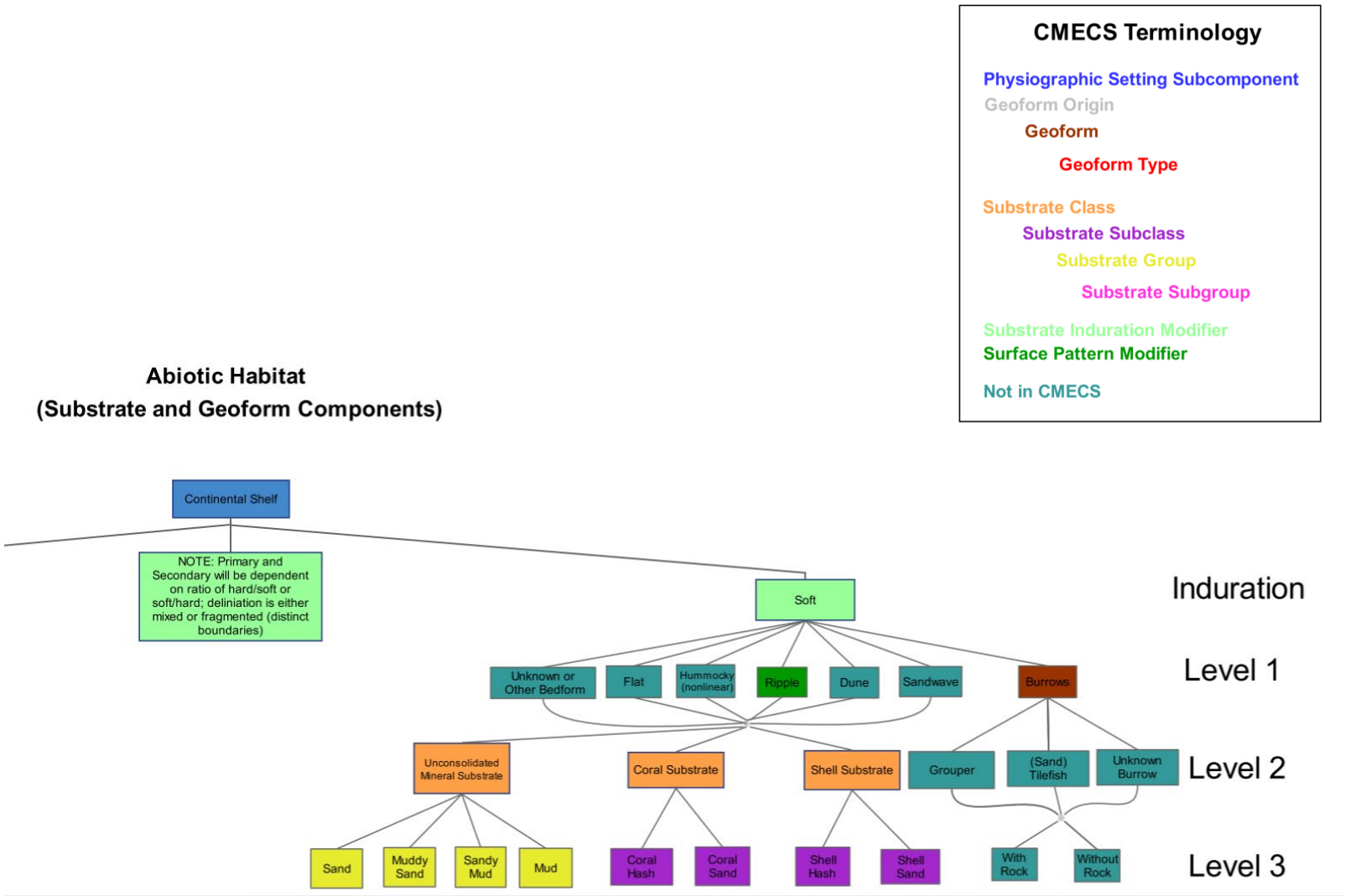


Texture metrics derived from the backscatter mosaic (10 m x 10 m resolution)

Appendix 3: Full Geologic Habitat Scheme

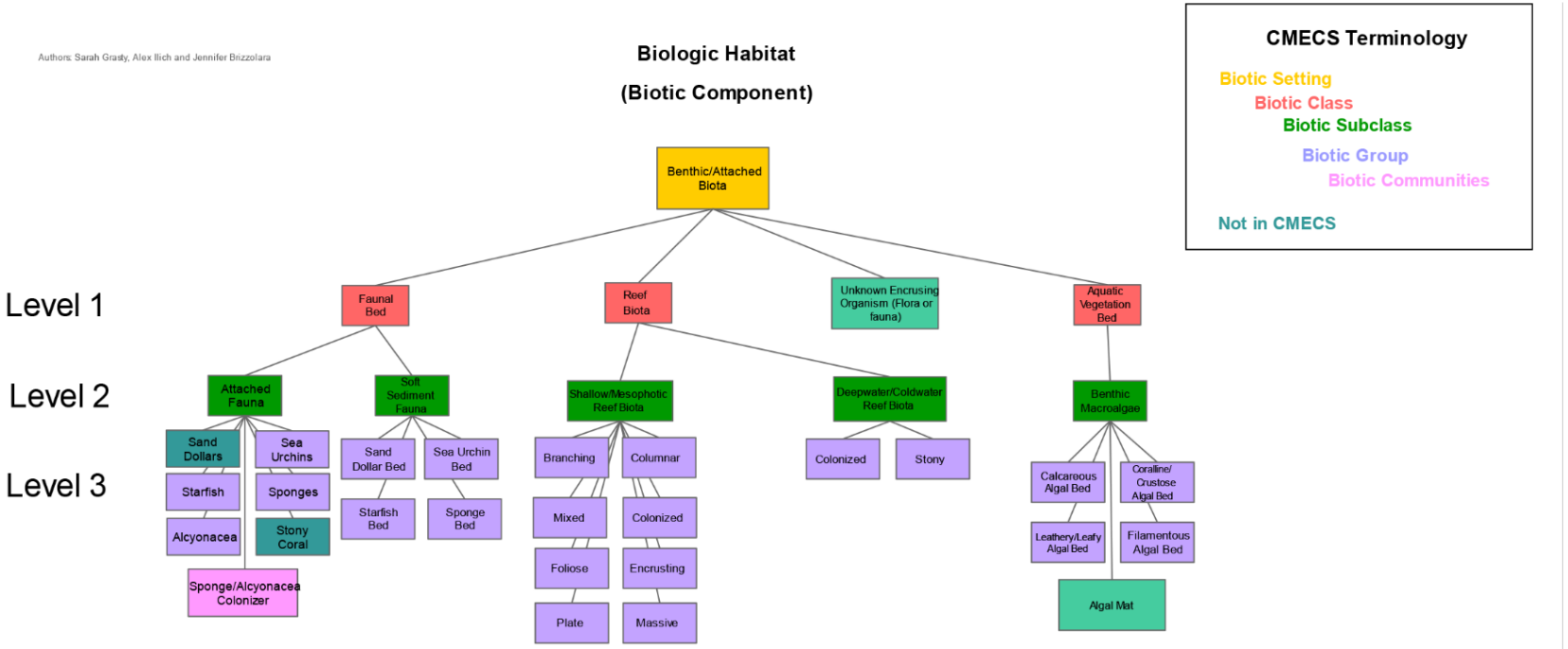


Full geologic habitat scheme (hard bottom)



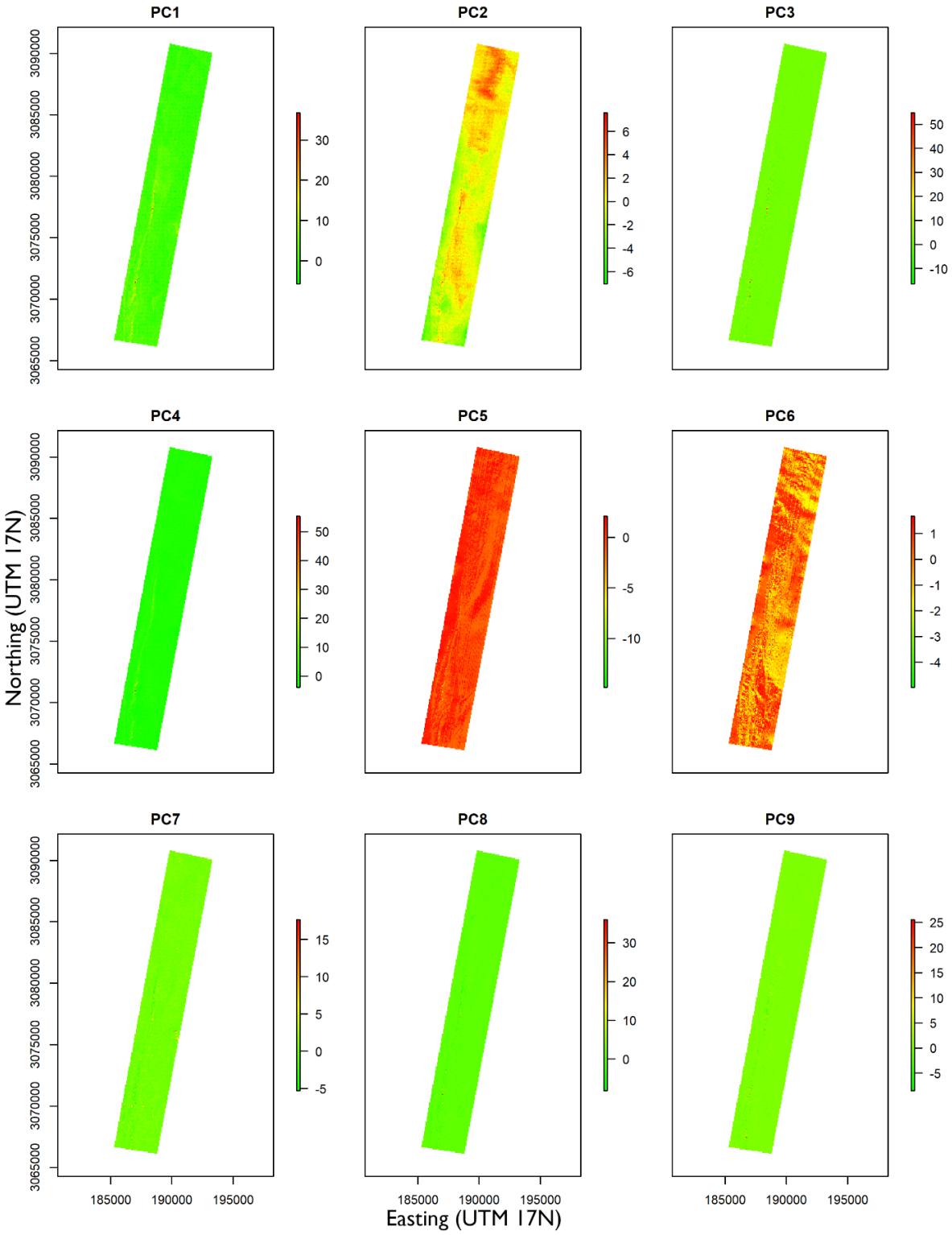
Full geologic habitat scheme (soft bottom)

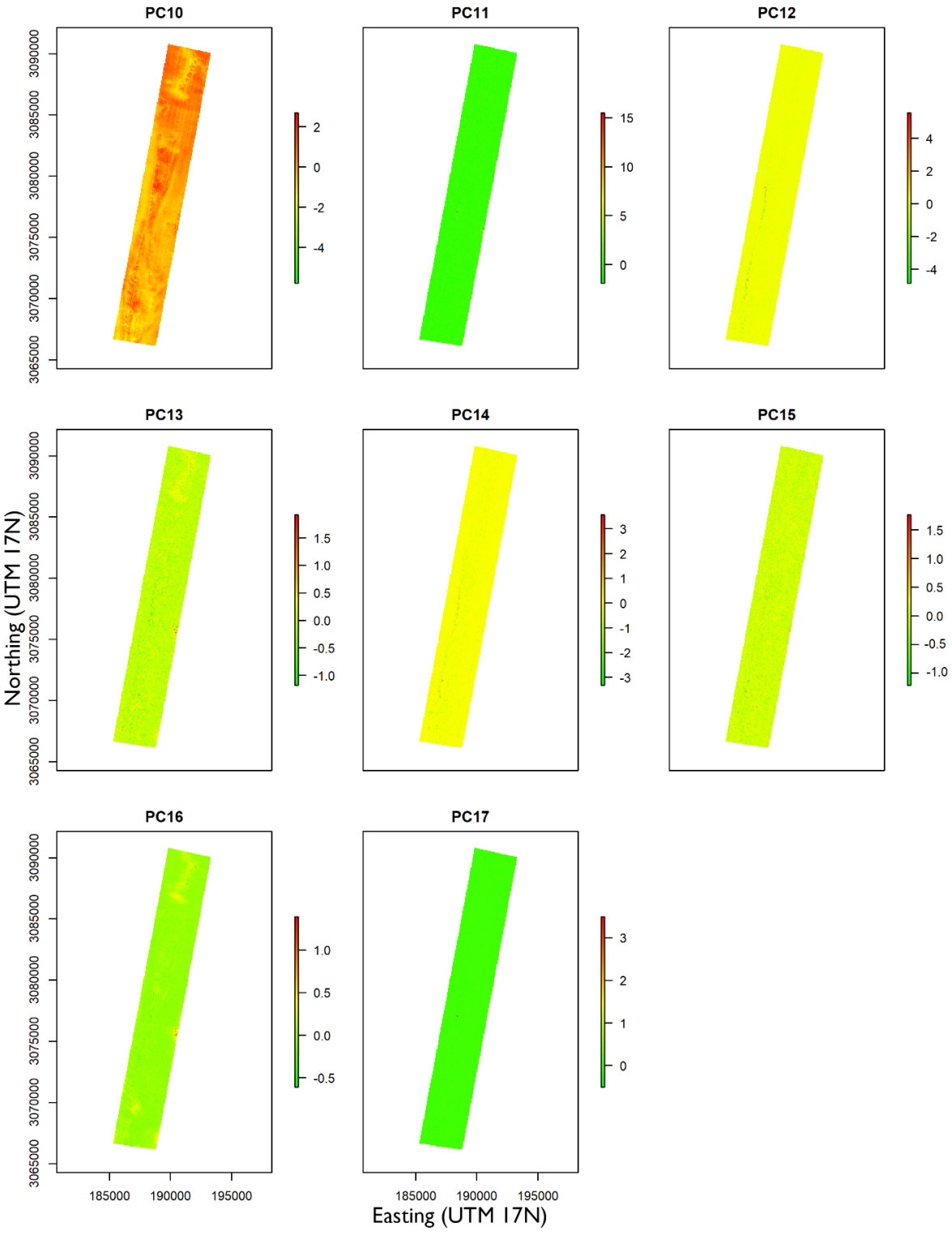
Appendix 4: Full Biotic Habitat Scheme



Full biotic habitat scheme

Appendix 5: Principal Components Raster Layers





Principal Components of bathymetry, backscatter and their derivative features (10 m x 10 m resolution)

Appendix 6: Principal Components Raster Layer Variable Loadings

Variable Loadings for each Principal Component

	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Bathymetry	-0.08617	0.421277	0.044653	0.083212	-0.12308	-0.11374
Terrain Ruggedness Index	0.293895	-0.04454	0.173827	0.480935	-0.04488	-0.00179
Topographic Position Index	0.076465	0.037148	0.554279	-0.26131	0.045161	0.014477
Slope	0.277316	-0.06182	0.101807	0.499139	-0.03516	0.009884
Eastness	-0.02642	0.040411	-0.01855	-0.14796	-0.68868	-0.66526
Northness	0.028424	-0.02213	0.004229	0.07651	0.675751	-0.73088
Ratio of Surface to Planar Area	0.20561	0.024071	0.301556	0.332264	-0.15027	-0.05629
Planform Curvature	0.106018	0.031187	0.500239	-0.11969	-0.01186	-0.00838
Profile Curvature	-0.02522	-0.0309	-0.46469	0.335739	-0.09104	-0.03422
Backscatter	-0.16761	0.471202	0.027796	0.109729	0.050224	0.033743
GLCM Mean	-0.16785	0.47894	0.032878	0.1144	0.050576	0.034138
GLCM Variance	-0.16139	0.480562	0.029037	0.110809	0.053318	0.034978
GLCM Homogeneity	-0.38921	-0.17594	0.143193	0.16153	-0.02813	-0.0163
GLCM Contrast	0.352369	0.109558	-0.03512	0.032877	-0.05768	-0.02044
GLCM Dissimilarity	0.40292	0.162735	-0.11145	-0.09594	-0.00142	0.00371
GLCM Entropy	0.366662	0.17143	-0.16351	-0.21267	0.047422	0.021976
GLCM Angular Second Moment	-0.33622	-0.17183	0.170691	0.245246	-0.05774	-0.02976

	PC 7	PC 8	PC 9	PC 10	PC 11	PC 12
Bathymetry	0.073338	-0.0003	0.093618	0.873497	-0.00265	0.000391
Terrain Ruggedness Index	-0.20183	-0.23157	-0.05496	0.022505	-0.01271	-0.24754
Topographic Position Index	-0.01926	-0.08062	-0.04242	0.00413	0.004159	0.73946
Slope	-0.25758	-0.34423	-0.20116	0.034192	0.025345	0.222819
Eastness	-0.10796	-0.11395	-0.0549	-0.17557	-0.0048	-0.00181
Northness	0.026375	0.033511	0.011951	0.002427	0.002534	0.00188
Ratio of Surface to Planar Area	0.090502	0.547525	0.618756	-0.14263	-0.03503	0.020103
Planform Curvature	-0.03713	0.439143	-0.60805	0.044809	-0.00023	-0.37321
Profile Curvature	0.015394	0.515212	-0.3955	0.030908	-0.003	0.448256
Backscatter	-0.0401	-0.03558	-0.03973	-0.25527	0.000125	-0.00676
GLCM Mean	-0.03665	-0.02798	-0.04121	-0.21989	-0.00162	0.007101
GLCM Variance	-0.02167	-0.03411	-0.04805	-0.24251	0.019659	-0.00154
GLCM Homogeneity	-0.01053	0.002216	0.014289	0.022428	0.641433	-0.01237
GLCM Contrast	0.713232	-0.13556	-0.11102	-0.07767	0.495824	-0.00409
GLCM Dissimilarity	0.265782	-0.05619	-0.05384	-0.04006	-0.34706	0.004407
GLCM Entropy	-0.30636	0.080952	0.05737	0.023077	0.225153	-0.03051
GLCM Angular Second Moment	0.433914	-0.13377	-0.09737	-0.01325	-0.41121	-0.02063

	PC 13	PC 14	PC 15	PC 16	PC 17
Bathymetry	-0.01828	0.016948	0.006024	0.017038	0.001616
Terrain Ruggedness Index	0.01469	-0.62843	0.305848	-0.00437	0.009405
Topographic Position Index	-0.00253	-0.23231	0.074265	0.003052	0.002717
Slope	-0.01189	0.5647	-0.25017	0.006788	-0.00247
Eastness	0.006185	-0.00518	0.002425	-0.00181	4.18E-05
Northness	-0.00112	0.005624	-0.00037	-0.00063	-0.0001
Ratio of Surface to Planar Area	-0.00145	0.114309	-0.06453	0.004361	-0.00031
Planform Curvature	-0.00542	0.12654	-0.04157	7.18E-06	-0.00116
Profile Curvature	-0.00337	-0.18	0.062208	0.002505	0.002013
Backscatter	-0.80841	-0.01361	0.013982	0.09919	-0.00049
GLCM Mean	0.316736	0.009387	0.002565	-0.75624	-0.00357
GLCM Variance	0.494886	-0.01041	-0.02503	0.645776	0.003008
GLCM Homogeneity	-0.00583	-0.07891	-0.12307	-0.00647	-0.57969
GLCM Contrast	-0.01172	0.017351	0.036275	-0.01794	0.253779
GLCM Dissimilarity	-0.00023	0.007854	0.005626	0.002556	-0.76854
GLCM Entropy	-0.01406	-0.32315	-0.70568	-0.01922	0.051341
GLCM Angular Second Moment	-0.00959	-0.2429	-0.55969	-0.01041	0.078258

Appendix 7: Observed Taxa

List of all taxa observed along with the number of sites (15 second bins) they were present in as the total number of individuals observed

Common Name	Scientific Name	Family	Order	Number of Individuals Observed	Number of Sites Present
Amberjack spp.	*Seriola spp.*	Carangidae (Jacks)	Perciformes	46	10
Angelfish spp.		Pomacanthidae (Angelfishes)	Perciformes	1	1
Blue Angelfish	*Holacanthus bermudensis*	Pomacanthidae (Angelfishes)	Perciformes	71	51
Gray Angelfish	*Pomacanthus arcuatus*	Pomacanthidae (Angelfishes)	Perciformes	7	5
Bigeye spp.		Priacanthidae (Bigeyes)	Perciformes	46	33
Boxfish spp.		Ostraciidae (Boxfishes)	Tetraodontiformes	2	2
butterflyfish_spp		Chaetodontidae (Butterflyfishes)	Perciformes	5	5
Eel spp.			Anguilliformes	1	1
Spotted Goatfish	*Pseudupeneus maculatus*	Mullidae (Goatfishes)	Perciformes	1	1
Grouper spp.		Serranidae (Groupers and Sea Basses)	Perciformes	8	7
Black Grouper	*Mycteroperca bonaci*	Serranidae (Groupers and Sea Basses)	Perciformes	3	1
Atlantic Creolefish	*Paranthias furcifer*	Serranidae (Groupers and Sea Basses)	Perciformes	309	10
Gag Grouper	*Mycteroperca microlepis*	Serranidae (Groupers and Sea Basses)	Perciformes	5	3
Goliath Grouper	*Epinephelus itajara*	Serranidae (Groupers and Sea Basses)	Perciformes	3	2
Red Grouper	*Epinephelus morio*	Serranidae (Groupers and Sea Basses)	Perciformes	5	5
Scamp	*Mycteroperca phenax*	Serranidae (Groupers and Sea Basses)	Perciformes	16	5

Jack spp.		Carangidae (Jacks)	Perciformes	1	1
Crevalle Jack	*Caranx hippos*	Carangidae (Jacks)	Perciformes	4	3
Rainbowrunner	*Elagatis bipinnulata*	Carangidae (Jacks)	Perciformes	2	1
Lionfish spp.	*Pterois spp.*	Scorpaenidae (Scorpionfishes)	Scorpaeniformes	335	121
Porgy spp.		Sparidae (Porgies)	Perciformes	18	16
Remora spp.		Echeneidae (Remoras)	Perciformes	1	1
Sea Turtle spp.			Testudines	1	1
Shark spp.			Superorder: Selachimorpha	1	1
Snapper spp.		Lutjanidae (Snappers)	Perciformes	3	2
Gray Snapper	*Lutjanus griseus*	Lutjanidae (Snappers)	Perciformes	670	30
Yellowtail Snapper	*Ocyurus chrysurus*	Lutjanidae (Snappers)	Perciformes	6	2
Squirrelfish spp.		Holocentridae (squirrelfishes)	Beryciformes	80	73
Whiptail Stingray spp.		Dasyatidae (Whiptail Stingrays)	Rajiformes	1	1
surgeonfish_spp		Acanthuridae (Surgeonfishes)	Perciformes	15	12
Sand Tilefish	*Malacanthus plumieri*	Malacanthidae (tilefishes)	Perciformes	41	38
triggerfish_spp		Balistidae (Triggerfishes)	Tetraodontiformes	7	7
Creole Wrasse	*Clepticus parrae*	Labridae (Wrasses)	Perciformes	10	1
Hogfish	*Lachnolaimus maximus*	Labridae (Wrasses)	Perciformes	6	6
Pearly Razorfish	*Xyrichtys novacula*	Labridae (Wrasses)	Perciformes	1	1
Spanish Hogfish	*Bodianus rufus*	Labridae (Wrasses)	Perciformes	2	2
Spotfinhogfish	*Bodianus pulchellus*	Labridae (Wrasses)	Perciformes	2	2
Large No ID				297	108