3-28-2016

**Acropora** Habitat Evaluation and Restoration Site Selection Using a Species Distribution Modeling Approach

Katherine Wirt Ames

Follow this and additional works at: [http://scholarcommons.usf.edu/etd](http://scholarcommons.usf.edu/etd)

Part of the [Ecology and Evolutionary Biology Commons](http://scholarcommons.usf.edu/etd/), [Other Animal Sciences Commons](http://scholarcommons.usf.edu/etd/), and the [Other Oceanography and Atmospheric Sciences and Meteorology Commons](http://scholarcommons.usf.edu/etd/)

Scholar Commons Citation

http://scholarcommons.usf.edu/etd/6158

This Dissertation is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.
Acropora Habitat Evaluation and Restoration Site Selection Using a Species Distribution Modeling Approach

by

Katherine Wirt Ames

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy
College of Marine Science
University of South Florida

Major Professor: Pamela Hallock Muller, Ph.D.
David Palandro, Ph.D.
Kathleen Semon Lunz, Ph.D.
Kendra L. Daly, Ph.D.
Chuanmin Hu, Ph.D.

Date of Approval:
February 9, 2016

Keywords: Acropora palmata, A. cervicornis, restoration, GIS

Copyright © 2016, Katherine Wirt Ames
Dedication

This dissertation is dedicated to my husband, John. Thank you for your unwavering support throughout this journey. I cannot wait to undergo the rest of life’s adventures with you by my side. Thanks also go to my parents for always being my cheerleaders and encouraging me to chase after this very big dream.
Acknowledgements

I would like to acknowledge the countless people who have advised, encouraged, and assisted me throughout my graduate school career. Without the support and confidence from my academic advisor, Pamela Hallock Muller, this dissertation would not have been possible. She used every last one of her resources to find a way to bring me to USF, and for that I am eternally grateful. A special thanks goes out to David Palandro, who took me under his wing and began the process of shaping me into the scientist I am today. Thank you to my committee members, Kathleen Semon Lunz, Kendra Daly and Chuanmin Hu for their exceptional advice and expertise over the years, this dissertation would not be what it is without your input. To my colleagues and friends at FWRI, especially Luke McEachron. Luke’s creativity and knowledge played a huge role in the development of this project, and for that I owe him every ounce of appreciation. I am also grateful for the generous support of the USF College of Marine Science Endowed Fellowship Program. The support of the Southern Kingfish Association Fellowship, The Linton Tibbetts Graduate Fellowship and the Gulf Oceanographic Charitable Trust Endowed Fellowship provided the assistance needed to continue my graduate career. And lastly, thank you to my amazing friends and family that have stuck with me through this journey. Without the support of my wonderful husband, John, and my amazing parents, my dream of completing a Ph.D. would have been just that, a dream.
# Table of Contents

List of Tables ........................................................................................................................ vii

List of Figures ........................................................................................................................ ix

Abstract .................................................................................................................................... xiii

1 Introduction ............................................................................................................................. 1
   1.1 Background on Acropora spp. in the Western Atlantic and Caribbean ...................... 5
   1.2 Recent History of Acropora Decline .......................................................................... 6
       1.2.1 Wider Caribbean Acropora .............................................................................. 6
       1.2.2 Southeast Florida Acropora ........................................................................... 8
       1.2.3 Dry Tortugas Acropora .................................................................................. 9
   1.3 Characteristics and Importance of Acropora ............................................................. 10
   1.4 Habitat Requirements of Atlantic and Caribbean Acropora spp. ............................ 11
   1.5 Light Availability and Energy Requirements ............................................................. 13
   1.6 Previous Work ............................................................................................................ 14
   1.7 Overview of Dissertation ......................................................................................... 15
   1.8 References ................................................................................................................. 17

2 Potential Habitat of Acropora spp. on Reefs of Florida, Puerto Rico, and the US Virgin Islands ................................................................................................................................. 22
   2.1 Abstract ....................................................................................................................... 22
   2.2 Introduction ................................................................................................................ 23
       2.2.1 Recent threats to Acropora populations ......................................................... 23
       2.2.2 Habitat requirements of Atlantic and Caribbean Acropora spp. .................... 25
       2.2.3 Previous Research .......................................................................................... 26
       2.2.4 Objectives and potential significance ............................................................... 26
   2.3 Methods ...................................................................................................................... 27
       2.3.1 Field area .......................................................................................................... 27
       2.3.2 Data .................................................................................................................. 28
       2.3.3 Benthic habitat maps ......................................................................................... 29
       2.3.4 Determination of Acropora habitat ................................................................ 29
       2.3.5 Buffer generation ............................................................................................. 30
       2.3.6 Potential-habitat map generation ..................................................................... 31
   2.4 Results ......................................................................................................................... 31
       2.4.1 Habitats of observed Acropora spp. ................................................................. 32
       2.4.2 Buffer generation ............................................................................................. 32
       2.4.3 Generation of potential-habitat maps ............................................................... 35
4.4 Discussion .................................................................................................................. 117
4.5 References ................................................................................................................. 122

5 Final Model Application ................................................................................................. 126
  5.1 Abstract ..................................................................................................................... 126
  5.2 Introduction ................................................................................................................ 127
  5.3 Methods and Results ................................................................................................. 131
    5.3.1 Study Area ........................................................................................................... 131
    5.3.2 Environmental Data Sets .................................................................................... 131
    5.3.3 AdaBoost.M1 Boosted Classification Tree Model Development ...................... 132
    5.3.4 Analysis of Model Predictions ............................................................................ 135
    5.3.5 Analysis of Environmental Parameters ............................................................. 136
    5.3.6 Analysis of Restoration Locations ....................................................................... 139
    5.3.7 Spatial Analysis of Restoration Sites .................................................................... 142
  5.4 Discussion .................................................................................................................. 146
  5.5 References ................................................................................................................. 152

6 Summary and Conclusions .............................................................................................. 155
  6.1 Research Overview .................................................................................................... 156
  6.2 Future Research and Recommendations .................................................................. 159
  6.3 References ................................................................................................................ 160

Appendix A: Copyright Clearance ...................................................................................... 162

About the Author ................................................................................................................ End Page
List of Tables

Table 2.1: Details of the multiple benthic habitat maps used for this study ........................................29

Table 2.2: Coral reef and hardbottom classifications used for the benthic habitat maps in this study .........................................................................................................................................30

Table 2.3: Number of observations of Acropora spp. by region of the Florida Reef Tract after erroneous points were removed from the database ..........................................................32

Table 2.4: Buffer distances within which exactly 95% and 99% of points are included ..........34

Table 2.5: Results of K-S goodness of fit test comparing distributions in each region to distributions of the full reef tract ........................................................................................................35

Table 2.6: Buffer distances used for potential-habitat maps .........................................................37

Table 3.1: Summary of abbreviations and acronyms .................................................................53

Table 3.2: Summary of data layers and brief descriptions ..........................................................55

Table 3.3: Number of cells categorized as each reef habitat type throughout the study area ..............................................................................................................................................57

Table 3.4: Categories assigned to repeat monitoring Acropora cervicornis data .................67

Table 3.5: Number of observations in each of the A. cervicornis data sets .........................68

Table 4.1: Description of Random Forest algorithm steps .........................................................95

Table 4.2: Description of Boosted Classification Tree algorithm steps ..................................98

Table 4.3: Description of all predictor layers used in this study .................................................101

Table 4.4: Combinations of variables tested in the RF model .................................................106

Table 4.5: Results of variable combinations tried with the RF model ..................................106

Table 4.6: Combinations tried for the AdaBoost.M1 model .................................................109

Table 4.7: Results of the variable combinations tried with the AdaBoost.M1 algorithm ....110
Table 4.8: Results of the variable combinations tried with the SAMME algorithm ...............111
Table 4.9: Confusion matrix from final RF model using the training data..........................112
Table 4.10: Confusion matrix from the final AdaBoost.M1 model using the training data.......114
Table 4.11: Confusion matrix from the final SAMME model using the training data.............116
Table 4.12: Error statistics for all final models......................................................................117
Table 4.13: Number of cells predicted for each category for each final model....................118
Table 5.1: Description of environmental predictor data layers used in the development of the SDM in this study ........................................................................................................133
Table 5.2: Categories assigned to repeat-monitoring sites representing Acropora cervicornis data........................................................................................................................................134
Table 5.3: Confusion matrix from the final AdaBoost.M1 model using the training data.......134
Table 5.4: Number of restoration sites within each predicted category..............................140
List of Figures

Figure 1.1: The three main regions of the Florida Reef Tract ................................................................. 2
Figure 1.2: NOAA designated Critical Habitat in Florida ................................................................. 4
Figure 1.3: NOAA designated Critical Habitat in Puerto Rico and the U.S. Virgin Islands ........ 4
Figure 1.4: Two species of Acropora found in the western Atlantic and Caribbean: A. cervicornis and A. palmata ............................................................ 5
Figure 1.5: Growth extension rates of selected zooxanthellate scleractinian corals from the Caribbean region [Recreated from Dullo (2005)] ......................................................... 12
Figure 1.6: Percent of Caribbean localities with A. palmata and A. cervicornis as the dominant coral in the Late Pleistocene, Holocene, before 1983 and after 1983 [Recreated from Pandolfi (2002)] .......................................................... 13
Figure 2.1: Regions of focus in this study, with locations of Acropora spp. observations ........ 28
Figure 2.2: Percentage of observations of Acropora spp. located on mapped coral reef or hardbottom .......................................................... 33
Figure 2.3: Associated substrate of Acropora spp. observations in Florida ............................... 34
Figure 2.4: Associated substrate of A. palmata observations in Puerto Rico ............................. 35
Figure 2.5: Associated substrate of Acropora spp. observation in St. Croix .............................. 36
Figure 2.6: Associated substrate of Acropora spp. observations in St. Thomas/St. John .......... 37
Figure 2.7: Percentages of points included for various buffer distances by region ................... 38
Figure 2.8: 95% Acropora spp. potential habitat compared to previously determined NOAA critical habitat around Florida ......................................................... 39
Figure 2.9: 95% Acropora spp. potential habitat compared to previously determined NOAA critical habitat around Puerto Rico ......................................................... 40
Figure 2.10: 95% Acropora spp. potential habitat compared to previously determined NOAA critical habitat around St. Croix ......................................................... 41
Figure 2.11: 95% *Acropora* spp. potential habitat compared to previously determined NOAA critical habitat around St. Thomas and St. John .........................................................42

Figure 3.1: Extent of study boundary ....................................................................................................................72

Figure 3.2: The seven regions of the Florida Reef Tract used in this study ..........................................................73

Figure 3.3: Reef type polygons located within a single grid cell used to determine dominant reef type ..........................................................................................................................74

Figure 3.4: Raster representation of the reef type data layer ..................................................................................75

Figure 3.5: Example of a non-reef dominated grid cell that was classified as patch reef .................................76

Figure 3.6: Final distance from shore raster layer ..................................................................................................77

Figure 3.7: Depth layer used in this study ...............................................................................................................78

Figure 3.8: Final slope raster layer .......................................................................................................................79

Figure 3.9: Final SBT high and low raster layers ..................................................................................................80

Figure 3.10: Final light380 and light488 raster layers ..........................................................................................80

Figure 3.11: Results of simple linear regression testing for multicollinearity of the predictor data ..........................................................81

Figure 3.12: Out-of-bag (OOB) error estimates on different subsets of training data .........................................82

Figure 3.13: Spatial distribution of the training and validation *A. cervicornis* data sets ....................................83

Figure 3.14: Coverage of final point file with all predictors assigned .................................................................84

Figure 4.1: Example of a single classification tree using *A. palmata* presence/absence data as response, and depth, wave exposure and light availability as predictors. ..............................92

Figure 4.2: Average out-of-bag (OOB) error stabilizes when more trees are added to the random forest ..........................................................102

Figure 4.3: Results of the rfcv function on various values of mtry .......................................................................104

Figure 4.4: Mean Decrease in Accuracy for each predictor variable in the RF model .......................................105

Figure 4.5: Results of the tune.rpart on maxdepth ............................................................................................108
Figure 4.6: AdaBoost.M1 algorithm error rate vs. number of trees .................................................109
Figure 4.7: Mean variable importance for the AdaBoost.M1 algorithm ........................................110
Figure 4.8: SAMME algorithm error rate vs. number of trees .......................................................111
Figure 4.9: Mean variable importance for the SAMME algorithm ..................................................112
Figure 4.10: RF variable importance .................................................................................................113
Figure 4.11: Margin for random forest model ....................................................................................114
Figure 4.12: Relative importance of variables for the final AdaBoost.M1 model ......................115
Figure 4.13: Margin cumulative distribution for the AdaBoost.M1 model ................................115
Figure 4.14: Variable importance for the SAMME model ..............................................................116
Figure 4.15: Margin for the SAMME model ......................................................................................117
Figure 4.16: Map results of full predictions .......................................................................................120
Figure 5.1: Relative importance (%) of variables for the final AdaBoost.M1 model .............134
Figure 5.2: Margin cumulative distribution for the AdaBoost.M1 model .......................................135
Figure 5.3: The final model was used to predict unsampled locations along the full reef tract ........................................................................................................................................136
Figure 5.4: Mapped probabilities for each prediction category .......................................................137
Figure 5.5: Depth means and ranges for each predicted category ..................................................138
Figure 5.6: Percent light availability means and ranges for each predicted category .............139
Figure 5.7: Percent of solar radiation at 380 nm reaching the seafloor at the restoration locations compared to the ranges observed in the cells predicted by the model ......141
Figure 5.8: Percent of solar radiation at 488 nm reaching the seafloor at the restoration locations compared to the ranges observed in the cells predicted by the model ......141
Figure 5.9: Depth at the restoration locations compared to the ranges observed in the cells predicted by the model ...........................................................................................................142
Figure 5.10: Acropora cervicornis trend prediction on the east coast of Florida by a boosted classification tree at a 1 km scale ......................................................................................143
Figure 5.11: *Acropora cervicornis* trend prediction on the east coast of Florida by a boosted classification tree at a 1 km scale .................................................................144

Figure 5.12: *Acropora cervicornis* trend prediction on the east coast of Florida by a boosted classification tree at a 1 km scale .................................................................145
Abstract

While populations of nearly all stony coral species along the Florida reef tract have exhibited decline, the most notable decline has occurred in the once-dominant acroporid species (Acropora cervicornis, A. palmata). Both species were listed in 2006 as threatened under the Endangered Species Act. This listing, combined with their continued decline, has resulted in large-scale restoration efforts throughout Florida and the Western Caribbean. Currently, there is little to no information regarding spatial prioritization of sites for these restoration efforts. The primary objective of this dissertation was to utilize species distribution modeling, informed by existing data from the Florida reef tract, to identify sites for restoration of acroporid corals that should have strong likelihood for success.

The initial focus was to use a database of reported field observations, in combination with benthic habitat maps, to model the extent of suitable habitat for Acropora spp. The mapped coral reef and hardbottom classifications throughout Florida, Puerto Rico, and the US Virgin Island reef tracts were used to generate potential-habitat polygons using buffers that incorporated 95% and 99% of reported observations of Acropora spp. Resulting maps demonstrated that A. palmata habitat is relatively well defined, while that of A. cervicornis is more variable and difficult to constrain.

Thus, as the major focus of this dissertation, available monitoring data from the Florida reef tract were used to construct and compare two types of statistical species distribution models, random forest and boosted classification trees, as an approach to inform siting of restoration efforts for A. cervicornis. Boosted classification trees were more accurate than the random forest
model at classifying *A. cervicornis* population trends. Further analyses of the two most important environmental parameters identified by the model, depth and light availability, revealed that reef areas predicted to not have had *A. cervicornis* present from 1996-2013 were deeper, on average, and had lower light availability and greater variance than areas predicted to have had continuous or transient *A. cervicornis* presence over this time frame.

This study represents a first step at deriving an ecologically-guided approach to spatial prioritization of restoration efforts. The overarching goal of this project has been to design a strategy for creating models to define and predict where conservation and restoration actions should be most effective, that can be utilized for a variety of coral species. With existing populations mapped, the results can also aid in protecting the limited areas in which these species still occur.
**1 Introduction**

The combination of a broad, shallow continental shelf and warm water of the Gulf Stream provide a unique suite of benthic habitats that support Florida’s coral reef ecosystems (Andrews et al. 2005). Previous research estimated that $31 \times 10^3 \text{ km}^2$ of shallow-water inshore habitat around Florida has the potential to support coral reef ecosystems (Rohmann et al. 2005). Unfortunately, this estimation far exceeds the reality of area encompassed by living coral reefs in Florida.

There are three main regions of the Florida Reef Tract: the Tortugas Bank, the Florida Keys, and the southeast coast (Figure 1.1). All regions are subject to unique environmental conditions and, subsequently, varying degrees of coral cover and species composition (Andrews et al. 2005). No single region has shown to be resistant to the documented decline of coral reef ecosystems throughout the western Atlantic and Caribbean. All areas are subject to numerous stressors including, but not limited to, bleaching, diseases, water pollution, physical impacts, tropical storms and winter cold fronts (Andrews et al. 2005). While most studied coral reefs in the state have shown a declining trend, certain areas have exhibited slower decline than others.

Reefs throughout Florida and the Caribbean have experienced losses in both species richness and overall coral cover since the 1970s (Gardner et al. 2003). While nearly all of the stony coral species have exhibited decline, the most notable decline has occurred in the once-dominant acroporid species, the Staghorn (*Acropora cervicornis*, Lamark 1816) and Elkhorn (*A. palmata*, Lamark 1816) corals. Decline of these species has been noted throughout their range
over the last 40 years (Jaap et al. 1988; Porter and Meier 1992; Bythell and Sheppard 1993). This documented decline has led to an increase of propagation and gardening approaches for restoration (Epstein et al. 2003) in hopes to enhance and restore wild coral populations (Miller et al. 2014). A majority of restoration projects in Florida and the Caribbean focus on Acropora spp., particularly *A. cervicornis*, not only because they are keystone species, but also because they are fast growing and propagate by asexual fragmentation, which is conducive to coral-gardening approaches (Johnson et al. 2011).

Figure 1.1: The three main regions of the Florida Reef Tract.
Elkhorn and Staghorn corals were listed as threatened species under the Endangered Species Act in 2006 (National Marine Fisheries Service 2006). In 2008 both species were listed as critically endangered on the International Union for Conservation of Nature (IUCN) Red List of Threatened Species. The decline of these species is unprecedented given the vital role they historically played as major builders of western Atlantic and Caribbean coral reefs (Shinn 1966; Shinn et al. 1989; Jackson 1992a; Bruckner 2002; Precht et al. 2002). The overarching goal of this research is to develop a model to define and predict where conservation and restoration actions would be most effective. To accomplish this goal, this study builds upon previously created critical habitat maps for *A. palmata* and *A. cervicornis* (Figures 1.2–1.3) to show areas where these species currently exist, as well as areas that would be suitable for their (re)establishment through restoration. In a previous study, potential habitat maps were produced based on benthic substrata and depth throughout the Florida reef tract using GIS software. The first objective of this dissertation was to update these maps using more recent benthic habitat maps as well as to expand the study area to include Puerto Rico and the U.S. Virgin Islands.

Habitat parameters conducive to coral reef growth are not limited to benthic substrata and depth. Parameters such as light availability, geomorphology, and temperature play key roles in determining suitable habitat for many corals, including *Acropora* spp. (e.g., Hubbard et al. 1997, and references therein). As such, to best represent habitat suitable for restoration efforts of *A. cervicornis*, the second objective of this study is to include these additional parameters when creating species distribution models for the purpose of identifying suitable restoration sites along the Florida Reef Tract. The results of this research may be used not only to inform restoration site selection, but also to refine the current NOAA *Acropora* spp. critical habitat map as well as to provide a methodology for the creation of future coral critical habitat maps.
Figure 1.2: NOAA designated Critical Habitat in Florida.

Figure 1.3: NOAA designated Critical Habitat in Puerto Rico and the U.S. Virgin Islands.
1.1 Background on Acropora spp. in the western Atlantic and Caribbean

Corals of the genus *Acropora* have commonly epitomized coral reefs worldwide because of their typically shallow distributions (i.e., accessibility) and characteristic branching or bushy morphologies (Figure 1.4). Two distinct species occur in the western Atlantic and Caribbean region, *Acropora cervicornis* and *A. palmata*. A third form, commonly known as the fused Staghorn coral (*A. prolifera*, Lamarck 1816), is now recognized to be a hybrid between *A. cervicornis* and *A. palmata* (Van Oppen et al. 2000; Vollmer and Palumbi 2002).

![Figure 1.4: Two species of *Acropora* found in the western Atlantic and Caribbean: *A. cervicornis* and *A. palmata.*](Photo Credit: National Marine Fisheries Service 2015)

According to Veron (2000), *A. cervicornis* is characterized as arborescent with cylindrical branches that subdivide infrequently, thus its common name, Staghorn coral. Historically this species was common on upper to mid-reef slopes and in lagoons with clear waters. According to Veron (2000), *A. palmata* is characterized by parallel, obliquely inclined, very thick, tapered branches, thus its common name, Elkhorn coral. This species was historically common and conspicuous on shallow outer reef margins exposed to wave action. The common names for these corals are based on the European stag (*Cervus cervus*), which is similar to North American elk (*Cervus canadensis*), and the European “elk” (*Alces alces*), which in North America is known commonly as the moose.
The habitat of *A. palmata* has been so well defined that its distribution has been used to interpret both modern environments and paleoenvironments. Hubbard (1989; 1997) used the occurrence and morphologies of *A. palmata* as an indicator of wave and storm prevalence on reefs. Because this species is adapted to high light intensities and, therefore, water depths typically less than 10 m, its fossil distribution has been widely used to interpret rates of sea level rise (Blanchon and Shaw 1995; Toscano and Lundberg 1998; Blanchon and Eisenhauer 2001; Toscano and Macintyre 2003; Brock et al. 2008; Gabriel et al. 2009; Blanchon 2010).

The precipitous decline of *Acropora* spp. in the western Atlantic and Caribbean is a major issue in discussions of coral reef conservation. Their designation as candidate species for listing as threatened under the US Endangered Species Act (ESA) in 1999 (Diaz-Soltero 1999) and finally the formal listing as threatened in 2005 (*Acropora* Biological Review Team 2005) highlighted the concern for these historically abundant major reef-building corals, bringing attention to the overall decline in reef accretion over the past several decades.

1.2 Recent History of Acropora Decline

1.2.1 Wider Caribbean Acropora

Disturbance events, particularly exceptionally strong winter cold fronts and hurricanes that caused extensive mortality in *Acropora* spp., were recognized by researchers in the Dry Tortugas in the late 19th and early 20th centuries (Wells 1932; Jaap et al. 2008). The shallow habitats and branching morphologies of *Acropora* spp. made them particularly vulnerable to disturbance events, while their rapid growth rates and branching structures enabled populations to recover from such disturbances in a few years to decades (Gladfelter et al. 1978; Jaap et al. 1988). Thus, the cold-water event in January 1978 that resulted in extensive mortality of
Acropora spp. in the Dry Tortugas and elsewhere along the Florida reef tract was notable, but not of major concern (Davis 1982; Porter et al. 1982; Roberts et al. 1982). Similarly, Jaap (1979) noted bleaching on Middle Sambo Reef in 1973 and concluded that bleaching events of short duration have limited long-term effect on reef communities.

Unfortunately for the Florida reef communities, the cold-water event of 1978 was followed by the spread of white-band disease through Acropora populations throughout the western Atlantic and Caribbean (Gladfelter 1982). In addition, increasingly frequent disturbances have limited the recovery of extensive Acropora thickets in most parts of the Florida reef tract. These disturbances included widespread mass-bleaching events in 1983 (Jaap 1985) and 1987 (Lang et al. 1992), an exceptionally severe bleaching event in 1998 (Hoegh-Guldberg 1999), and the region-wide Diadema die off, also in 1983 (Lessios 1988). Bleaching was also observed in the Florida Keys in 1989, 1990, and 1991. Porter and Meier (1992) reported declines in live coral cover of up to 44% between 1982 and 1991 at several locations along the Florida reef tract. At Carysfort Reef in the Florida Keys, Palandro et al. (2003) detected a loss of ‘coral dominated’ bottom types from 52% to 6% over the years 1981-2000 using high resolution satellite imagery and aerial photographs. Somerfield et al. (2008) also noted declines in number of species, as well as coral cover on shallow and deep offshore reefs, following the bleaching event of 1998. The Coral Reef Evaluation and Monitoring Program (CREMP), which began annual assessments at 40 sites Keys wide in 1996, documented subsequent decline, such that by 2006, live coral cover averaged 6-7% (Callahan et al. 2007).

Thus, the two western Atlantic and Caribbean species of Acropora have been declining in abundance for the past 40 years, throughout the Florida Keys (Jaap et al. 1988; Porter and Meier 1992) and Caribbean (Aronson and Precht 2001b). For example, a study by Miller et al. (2002)
estimated a 93% decline of *A. palmata* and a 98% decline of *A. cervicornis* between 1983 and 2000 at Looe Key National Marine Sanctuary in the Florida Keys. Miller et al. (2002) attributed decline to a wide range of factors, including, but not limited to, storms, disease, high-temperature events that caused mass bleaching, water-quality decline, and ship groundings. Physical damage to these corals by anchor deployment (Halas 1985), boat grounding, diver disturbance, fishing lines, hooks, lobster pots, and buoys (Jaap et al. 1984) also have been commonly observed throughout the Florida reef tract.

Single events can result in multiple stressors on a coral community. In 1980 Hurricane Allen caused considerable physical damage to both *A. palmata* and *A. cervicornis* populations in Jamaica. In addition to the physical damage, corallivores out-survived their prey, which reduced the ability of the corals to recover (Hughes and Connell 1999). Signs of recovery were not apparent in the Caribbean throughout the 1980s and for a major part of the 1990s. Recovery has been observed in some areas in the late 1990s and into the early 2000s, but has been slow or unobserved in other areas (Grober-Dunsmore et al. 2006).

1.2.2 Southeast Florida Acropora

Human population increases in southeast Florida pose a variety of threats to coral reef ecosystems, including nutrient enrichment, diminished water transparency, phosphate inhibition of calcification, biotic replacement, and increased bioerosion (Weiss and Goddard 1977; Smith et al. 1981; Hallock and Schlager 1986; Yentsch et al. 2002). In spite of these anthropogenic influences, thickets of *A. cervicornis* have been found off the highly populated southeast coast of Florida. Significant populations have been reported in shallow, nearshore water off Fort Lauderdale (Thomas et al. 2000), where they are thriving at or near the latitudinal limits for the
species. The size of these thickets were found to range between 1,000 and 8,000 m$^2$, with $A. cervicornis$ representing 87–97% of all scleractinians (Vargas-Angel et al. 2003). These patches of $A. cervicornis$ were found to be fertile and spawned each summer (Vargas-Angel et al. 2006). These populations are believed to be the largest and northernmost in the continental USA and are a potential source of propagules to repopulate or replenish threatened populations in south Florida habitats (Vargas-Angel and Thomas 2002).

In addition to anthropogenically induced disturbances, south Florida populations are also exposed to natural threats, such as white-band disease, predation, and thermal stress. White-band disease was found on many thickets off Broward County in 2002, as was predation by $Hermodice carunculata$, a corallivorous worm (Vargas-Angel et al. 2003). Surface-water temperatures range from 22–25 °C in the winter (Vargas-Angel et al. 2003), which fall below the optimal temperatures for $Acropora$ spp. of between 25°C and 29°C (Jaap et al. 1989). A series of hurricanes, including Floyd in 1987, Andrew in 1992, Irene in 1999, Frances in 2004, and Katrina and Wilma in 2005, also have affected the southeast coast of Florida.

Extensive populations of $A. palmata$ are notably absent from southeast Florida habitats, although isolated colonies have been found (Banks et al. 2008). Unfortunately for the southeast Florida populations, these northernmost reefs of the Florida reef tract receive considerably less management than reefs in the Florida Keys and Dry Tortugas (Causey et al. 2002).

1.2.3 Dry Tortugas Acropora

The isolated, atoll-like reef system at the terminus of the Florida Keys, the Dry Tortugas, has the longest history of scientific investigations. Research in the Dry Tortugas began in 1881, when Alexander Agassiz mapped the benthos (Davis 1982). Reef research continued with the
establishment of the Carnegie Institute Tortugas Laboratory on Loggerhead Key in 1905 (Davis 1982; Shinn and Jaap 2005). Although the original habitat map by Agassiz showed 44 hectares (440,000 m²) of *A. palmata*, a study by Davis (1982) found that by 1976 *A. palmata* colonies had been reduced to two small patches that occupied less than 600 m², as well as a swatch of algal-covered *A. palmata* rubble on the reef crest. This same study found extensive stands of *A. cervicornis* covering a total of $4.78 \times 10^6$ m², accounting for 55% of the total scleractinian coral cover. Unfortunately, in January 1977 a severe cold front with water temperatures of 14°C to 16°C wiped out many of these *A. cervicornis* colonies in the Dry Tortugas and impacted the few remaining *A. palmata* colonies (Davis 1982; Porter et al. 1982; Jaap et al. 1989).

Populations of these species in the Dry Tortugas have not recovered to pre-1970s abundances and continued decline was documented at White Shoal from 1999 to 2005, as a result of bleaching and disease (Wheaton et al. 2005). Jaap and Sargent (1994) speculated that populations of *A. palmata* in the Dry Tortugas have not recovered to original levels due to loss of environmental conditions favorable for recruitment and growth. In 2007, the U.S. National Park Service (NPS) designated a “no-take” Research Natural Area (RNA) around the Dry Tortugas in hopes of protecting this unique region of the Florida reef tract. Kuffner et al. (2008) concluded that it was too early to speculate whether this RNA will would contribute to the restoration of the benthic community in Dry Tortugas National Park.

1.3 Characteristics and Importance of Acropora

The fast growth and calcification rates and their branching morphologies are attributes that make *A. cervicornis* and *A. palmata* important to reef communities (Gladfelter et al. 1978). *Acropora palmata* and *A. cervicornis* have the fastest growth and calcification rates of any
species in the Caribbean (Dullo 2005; Figure 1.5). Historically, *A. palmata* was the major reef-builder in the shallow forereef zones of the Florida reef tract (Shinn et al. 1989; Shinn 2004) and the extensive three-dimensional structure of *Acropora* thickets provided habitat for many reef fish (Gladfelter et al. 1978; Lirman 1999). *Acropora cervicornis* also played a major role in the structure and ecology of many Caribbean reefs by significantly contributing to reef accretion, framework construction, and habitat formation (Aronson and Precht 2001a). The decline of these species has resulted in both decline in reef accretion and loss of habitat for many reef constituents (Jackson 1992; Hughes 1994; Bak and Nieuwland 1995; Jackson et al. 2001).

Historically, *Acropora* populations also dominated reefs elsewhere in the western Atlantic and Caribbean. Stands of *Acropora* have been dominant features of Caribbean reefs for at least the last 500,000 years according to Pandolfi (2002; Figure 1.6). Often, the loss of the major stands of *Acropora* has been interpreted to be the result of the combination of disease, siltation, decline in water quality, and hurricanes (Norstrom et al. 2009). Others have attributed the decline to the mass mortality of *Diadema antillarum* in the 1980s (Pandolfi 2002). Yet other studies suggest that regional decline of *A. palmata* and *A. cervicornis* is due to white-band disease breakouts (Aronson and Precht 2001b). Most likely, the combination of all of these influences has contributed to the continuing decline of not only *Acropora* spp., but general coral cover throughout the Atlantic and Caribbean.

### 1.4 Habitat Requirements of Atlantic and Caribbean Acropora spp.

The specific habitat requirements for these two species are relatively well known. *Acropora palmata* has fairly discrete environmental requirements including clear, normal marine salinity, well-circulated water; solid substrata; and moderate water temperatures (optimally 25°C
to 29°C, without extreme seasonal variation (Jaap et al. 1989). During times of high abundance, both *A. palmata* and *A. cervicornis* were common in forereef zones. Prior to the 1970s, *A. palmata* was the dominant coral in wave-exposed and high-surge reef zones, typically at depths less than 10 m, throughout much of the Caribbean (Adey and Burke 1976). *Acropora cervicornis* was found at shallow to medium depths, as deep as 30 m, in optically shallow waters (Fenner 1988). *Acropora cervicornis* thickets in shallow backreef flats and patch reefs were common prior to the 1980s (Dustan 1985; Shinn et al. 1989). However, the extent of present, historical, and potential habitat for these two species along the Florida reef tract is not well known.

![Growth extension rates of selected zooxanthellate scleractinian corals from the Caribbean region. (Recreated from Dullo (2005))](image)
Recent studies on habitat distributions of both species found *A. cervicornis* distribution to be wider than *A. palmata* (Wirt et al. 2013), with colonies of the former found in a variety of habitats, including mid-channel and offshore patch reefs, as well as inner reef-tract sites (Miller et al. 2008). Miller et al. (2008) estimated that there may be $14 \pm 12$ million *A. cervicornis* colonies and $1.6 \pm 1.4$ million *A. palmata* colonies throughout the Florida Keys. Unfortunately, a majority of these colonies are undocumented, therefore, the specific habitat type of a majority of the populations can be speculated, but not fully verified by observation.

### 1.5 Light availability and energy requirements

Hermatypic zooxanthellate corals are symbiotic organisms that obtain a majority of their energy and carbon requirements from photosynthesizing zooxanthellae present in their tissues (Muscatine et al. 1981). It is well understood that hermatypic corals exist in areas where suitable substrata are present, but also where there is sufficient light energy available to maintain photosynthesis (Wells 1957). The maximum depth of coral reefs depends on the attenuation of
light, which often varies from location to location (Mass et al. 2010). Underwater light decreases exponentially with depth, following the Beer-Lambert law (Gordon 1989). The reduction of light intensity in the water column with depth is due not only to the absorbing properties of the water itself (Smith and Baker 1981), but also due to dissolved and particulate organic material and suspended sediments (Kirk 1994; Mass et al. 2010). The lower bathymetric limits for individual species are determined by their photosynthetic efficiency at reduced levels of irradiance. For many species, photosynthesis becomes limiting where light availability approaches 10% of surface light levels (Done 1983).

Previous studies have found that, under reduced light levels, corals often produce fragile skeletons, increasing the threat of fragmentation when subjected to high energy waves breaking across the reef (Cook et al. 1997; Yentsch et al. 2002). Yentsch et al. (2002) concluded that some Florida coral populations are squeezed between two bathymetric limitations: they are unable to grow at deeper depths due to light limitations, but are also unable to grow in shallower regions due to the high wave energy which limits colonization. This is very likely the case for Acropora spp. populations on Florida reefs and some Caribbean reefs.

1.6 Previous Work

In previous studies, Wirt (2011) and Wirt et al. (2013) compared a database of Acropora spp. observations to benthic habitat maps throughout the Florida reef tract. The results of this research provided a series of potential habitat maps for Acropora spp. throughout the Florida reef tract. These maps were based on the mapped coral reef and hardbottom classifications and incorporated 95% and 99% of reported observations of colonies of Acropora spp. One of the main conclusions of this study was that additional studies of Acropora spp., both with respect to
occurrence and quality of maps, are needed for the southeast coast of Florida and especially in the Dry Tortugas region. This work also needs to be expanded to include the U.S. Virgin Islands and Puerto Rico, where the currently defined critical habitat needs refinement.

1.7 Overview of Dissertation

The work presented in this dissertation is composed of six chapters, one of which has been published (Wirt et al. 2015). Chapters 2 – 5 are intended to stand alone, with relevant figures, tables and references therein. The sixth and final chapter reviews the previous chapters, notes future work in progress, and makes recommendations for future research directions. Each chapter of the dissertation contributes to, or fully addresses one of two overarching goals. The first goal focuses on ways to best refine the current Acropora spp. critical habitat map originally generated by NOAA, at the time of listing of the species. The second goal of this study is to identify suitable A. cervicornis restoration sites using a species distribution modeling approach and existing monitoring data throughout Florida. Chapter 1 serves as a literature synthesis related to Acropora spp. history and environmental requirements.

The topic addressed in Chapter 2 examines whether the methods developed in Wirt (2011), for southeast Florida and the Florida reef tract, are applicable to the U.S. Virgin Islands and Puerto Rico. This work was published by Wirt et al. (2015) in Global Ecology and Conservation.

Chapter 3 describes the compilation and processing steps used to amass relevant reef ecosystem data to use for multiple spatial modeling and mapping purposes. The primary use in this dissertation for the data described is to identify suitable Acropora restoration sites using a species distribution modelling (SDM) approach (Chapters 4 and 5). There are a variety of other
uses for the compiled data, however the specific steps required to format the data into a suitable format for SDM development is described herein.

In Chapter 4, two statistical models used for species distribution modeling are compared and evaluated, random forest and boosted classification trees. The methods were chosen based on their ability to handle multinomial response data and their non-parametric nature. Both methods have also been shown to perform well in ecological prediction studies. The models are used with repeat *A. cervicornis* monitoring data to assess the regional stability of these populations. The results of this work can be used to predict appropriate restoration locations.

Chapter 5 further examines the best statistical model identified in Chapter 4. The purpose of this chapter is to analyze the environmental parameters identified as most important by the model, and to provide an ecologically-guided approach to spatial prioritization of restoration efforts. The results of the model are also compared to current restoration locations.

The resulting models from this research define and predict where conservation and restoration actions will be most effective. With existing populations mapped, the results will also aid in protecting the limited areas in which these species still occur. The resulting dataset on existing populations can also be used by researchers to compare characteristics of locations where these species are still thriving with the characteristics of areas from which they have disappeared. Such comparisons could, for example, inform choices for likely sites for successful restoration projects. By protecting and restoring populations of these species that provide habitat for many reef fish, the commercial and recreational fishing industries will also potentially be enhanced. Preservation of these species will have benefits for many other organisms that rely on *Acropora* thickets for shelter.
1.8 References


Hughes TP (1994) Catastrophes, phase shifts and large-scale degradation of Caribbean coral reef. Science 265:1547-1551


Toscano MA, Macintyre IG (2003) Corrected western Atlantic sea-level curve for the last 11,000 years based on calibrated C-14 dates from Acropora palmata framework and intertidal mangrove peat. Coral Reefs 22:257-270


2 Potential Habitat of *Acropora* spp. on Reefs of Florida, Puerto Rico, and the US Virgin Islands

**Note to Reader**

This chapter has been published in full (Wirt et al. 2015), and is included with the permission of the publisher.

2.1 Abstract

Elkhorn and staghorn corals (*Acropora palmata, Acropora cervicornis*) were listed in 2006 as threatened under the Endangered Species Act. The goal of this study was to create model potential-habitat maps for *A. palmata* and *A. cervicornis*, while identifying areas for possible re-establishment. These maps were created using a database of reported field observations in combination with existing benthic habitat maps. The mapped coral reef and hardbottom classifications throughout Florida, Puerto Rico, and the US Virgin Island reef tracts were used to generate potential-habitat polygons using buffers that incorporated 95% and 99% of reported observations of *Acropora* spp. Locations of 92% of *A. palmata* observations and 84% of *A. cervicornis* observations coincided with mapped coral reef or hard-bottom habitat throughout the study area. These results indicate that potential habitat for *A. palmata* is currently well defined throughout this region, but that potential habitat for *A. cervicornis* is more variable and has a wider range than that for *A. palmata*. This study provides a novel method of combining data sets
at various geographic spatial scales and may be used to inform and refine the current National Oceanic and Atmospheric Administration critical habitat map.

2.2 Introduction

Corals of the genus *Acropora* have epitomized coral reefs worldwide because of their typical distribution in shallow water (making them more accessible and therefore more familiar) and characteristic branching morphologies. Two species occur in the western Atlantic and Caribbean region, *Acropora cervicornis* (Lamarck 1816) and *Acropora palmata* (Lamarck 1816).

The precipitous decline of *Acropora* spp. in the western Atlantic and Caribbean is a primary issue of discussion in coral reef conservation. Their formal listing as threatened in 2006 (Federal Register, 2006) highlighted the concern for these historically important reef-building corals, bringing attention to the overall decline in reef-building corals over the past several decades. Fortunately for conservation and management purposes, their distribution in shallow water makes their habitats relatively accessible and amenable to mapping techniques based on satellite or aircraft-based remote sensing.

2.2.1 Recent threats to *Acropora* populations

During the past three decades many western Atlantic coral communities, including *Acropora* spp. populations, have been affected by a series of disturbances. The most notable of these include the spread of white-band disease (WBD) through Atlantic and Caribbean *Acropora* populations beginning around 1976 (Gladfelter, 1982; Porter et al., 2001), a cold-water event in 1976–1977 (Davis, 1982; Lessios, 1988; Lirman et al., 2011; Porter et al., 1982; Roberts et al., 1982), a region-wide Diadema (sea urchin) die-off in 1983 (Lessios, 1988), widespread mass-bleaching events in 1983 (Jaap, 1985; Lang et al., 1992; Hoegh-Guldberg, 1999), and numerous
hurricanes (Lirman and Fong, 1996). In addition, the increase in frequency of such disturbances has limited recovery of extensive Acropora thickets in most of the western Caribbean.

The spread of WBD is widely considered one of the most significant disturbances affecting A. palmata and A. cervicornis. White-band disease was first documented in the early 1970s at Buck Island National Reef Monument in St. Croix, US Virgin Islands (Gladfelter et al., 1977). Prior to the spread of the disease, areas of the reef in St. Croix were composed of more than 50% live A. palmata (Mayor et al., 2006). Unfortunately, and as is the case for many populations throughout the species range, the combination of WBD and other disturbances has reduced populations more than 90% (Mayor et al., 2006). Similarly, WBD drastically reduced A. cervicornis populations in both the shelf lagoon and spur-and-groove zones of the Belizean Barrier Reef, where in some locations coverage dropped from ~70% to close to zero in just seven years (Aronson and Precht, 1997). Additionally, gene flow between populations has been found to be restricted, negatively contributing to the influence of WBD on many populations (Baums et al., 2010; Hemond and Vollmer, 2010; Vollmer and Palumbi, 2007). As in many locations, the losses of these two reef-building species have had lasting and widespread effects on reefs throughout Florida, Puerto Rico, and the US Virgin Islands (Rogers et al., 2008a, b).

Human population growth poses a variety of local threats to coral reef ecosystems, including nutrient enrichment, diminished water clarity, inhibition of calcification by phosphates, biotic replacement, and increased bioerosion (Hallock and Schlager, 1986; Simkiss, 1964; Smith et al., 1981; Weiss and Goddard, 1977). Nevertheless, significant thickets of A. cervicornis have been reported in shallow nearshore waters off of populous Fort Lauderdale, Florida (Thomas et al., 2000; Vargas-Angel et al., 2003), at or near the northern latitudinal limits of the species. These populations are believed to be among the largest and northernmost in the continental USA
and are a potential source of propagules for replenishment of threatened populations in south Florida habitats (Vargas-Angel and Thomas, 2002). Extensive populations of *A. palmata* are notably absent from southeast Florida coastal habitats, although isolated colonies have been found (Banks et al., 2008). Unfortunately for these populations, the reefs of southeast Florida are located outside formal marine management zones, unlike the reefs in the Florida Keys and Dry Tortugas (Causey et al., 2002).

In the US Virgin Islands, many locations have vast stands of dead *A. palmata*, but small areas with dense live *A. palmata* do exist (Rogers et al., 2008a,b; Zitello et al., 2009). Unfortunately, no reefs in the US Virgin Islands exhibit densities of *A. palmata* as high as those recorded in the 1960s and 1970s (Rogers et al., 2002). In many areas of Puerto Rico, the fire coral *Millepora* spp. has replaced *A. palmata* as the dominant reef-crest coral species (Ballantine et al., 2008).

### 2.2.2 Habitat requirements of Atlantic and Caribbean Acropora spp.

Habitat requirements for *A. palmata* and *A. cervicornis* are well known. *Acropora palmata* has sensitive environmental requirements including clear, well-circulated water of normal marine salinity, a solid substrate, and moderate water temperatures (optimally 25–29 °C, without extreme seasonal variation) (Jaap et al., 1989). The presence of a solid substrate such as coral reef or hard bottom is vital for the attachment of coral recruits during settlement (Harrison and Wallace, 1990). Before the 1970s, in much of the Caribbean *A. palmata* was the dominant coral in wave-exposed and high-surge reef zones, typically at depths of less than 10 m (Adey and Burke, 1976). At the same time, *A. cervicornis* was found at shallow to medium depths, as deep as 30m in clear water (Fenner, 1988). *Acropora cervicornis* thickets in shallow reef flats and
patch reefs were common before the 1980s (Dustan, 1985; Shinn et al., 1989). However, the full extent of historical and current distributions, as well as potential habitat for these two species in the western Atlantic and Caribbean, has not been fully documented.

Recent studies of both species found *A. cervicornis* distribution to be wider than that of *A. palmata*, with colonies found on a variety of habitats, including mid-channel and offshore patch reefs, as well as inner reef-tract sites (Miller et al., 2008). Following spatially extensive surveys performed in 2007, Miller et al. (2008) estimated that there were 14 ± 12 million *A. cervicornis* colonies and 1.6 ± 1.4 million *A. palmata* colonies in the Florida Keys. Most of these colonies, however, have not been documented, nor has their location been verified by observation.

2.2.3 Previous Research

Wirt et al. (2013) used reported observations of *A. palmata* and *A. cervicornis* to generate potential-habitat maps for the species along the Florida Reef Tract. The study found that locations of 99% of *A. palmata* and 84% of *A. cervicornis* observations coincided with previously mapped coral reef or hard-bottom habitat. A main conclusion of that research was that potential habitat for *A. palmata* is currently well defined in Florida, while *A. cervicornis*, habitat is more variable and has a wider range (Wirt et al., 2013).

2.2.4 Objectives and potential significance

The objectives of this study were:

1. to evaluate the accuracy of reported *Acropora* spp. observations and current benthic habitat maps of Florida, Puerto Rico, and the US Virgin Islands;
(2) to use reports of existing colonies of *A. palmata* and *A. cervicornis* and current benthic

habitat maps to model the distribution of potential habitat in Florida, Puerto Rico, and the US
Virgin Islands; and,

(3) to create potential-habitat maps showing areas in which these species exist, as well as areas

that would be suitable for their re-establishment.

The results of this research will help define where conservation actions will be most
effective. With existing populations mapped, the results will also aid in preventing destruction of
the limited areas in which these species occur. This study also provides a model for developing
future critical habitat maps, which will be necessary if other coral species are designated as
threatened or endangered.

2.3 Methods

2.3.1 Field Area

This study focused on the reef tracts of Florida, Puerto Rico, and the US Virgin Islands
(St. Thomas, St. John, and St. Croix). The Florida Reef Tract extends from Martin County
through the Dry Tortugas, and both species are present throughout. In Puerto Rico, the reef tract
encircles the island, but *Acropora* spp. is most prevalent off the southwest and northeast coasts.
In St. Croix, *Acropora* spp. is concentrated along the north and northeast coasts of the island.
*Acropora* spp. locations are distributed relatively evenly along the reef tract on all sides of St.
Thomas and St. John (Figure 2.1).
Figure 2.1: Regions of focus in this study, with locations of *Acropora* spp. observations.

2.3.2 Data

The Florida Fish and Wildlife Conservation Commission’s (FWC) Fish and Wildlife Research Institute (FWRI) received *Acropora* spp. location data in Florida, Puerto Rico, and the US Virgin Islands, as described by Wirt et al. (2013). Due to errors discovered in the original Florida data set, the data here represent a reanalysis of the Florida region. Observations of *Acropora* spp. were reported from surveys conducted between 1996 and 2012 by a range of groups, agencies, and institutions.

While locations of surveys that did not detect *Acropora* spp. are important, they were not addressed in this study. All results pertain to locations where surveys detected one or both species of *Acropora.*
2.3.3 Benthic habitat maps

Recently, benthic habitat maps created by FWRI, the National Park Service (NPS), Nova Southeastern University, and the National Oceanic and Atmospheric Administration (NOAA) were combined to create the Unified Florida Reef Tract Map, which provides a seamless spatial representation of the Florida Reef Tract benthos. The classification system used accommodates and integrates various classification schemes while retaining the original information. The original maps used to build the Unified Florida Reef Tract Map were created using a combination of IKONOS imagery, aerial photography (digital and analog), and LiDAR (Baumstark, 2013).

Benthic habitat maps used for Puerto Rico and the US Virgin Islands were digitized from aerial photography acquired in 1999 by the NOAA Biogeography Branch. This project mapped the insular shelf between the shoreline and the shelf edge throughout Puerto Rico and the US Virgin Islands (Kendall et al., 2001). While these two mapping projects (Florida and the Caribbean) used different classification schemes, the goal of this study was to extract only mapped coral reef and hardbottom areas, which in this study allowed for the use of multiple classification schemes. Information on benthic habitat maps used for this study, including the year and minimum mapping unit for each map, is provided in Table 2.1.

<table>
<thead>
<tr>
<th>Location</th>
<th>Year</th>
<th>Year of source imagery</th>
<th>Agency</th>
<th>Minimum mapping unit</th>
<th>Area covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>St. Thomas/St John/St. Croix</td>
<td>2001</td>
<td>1999</td>
<td>NOAA Biogeography</td>
<td>1 acre</td>
<td>490 km²</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>2001</td>
<td>1999</td>
<td>NOAA Biogeography</td>
<td>1 acre</td>
<td>1,600 km²</td>
</tr>
<tr>
<td>Florida</td>
<td>2013</td>
<td>2001-2012</td>
<td>FWRI</td>
<td>0.5 – 1 acre</td>
<td>9,102 km²</td>
</tr>
</tbody>
</table>

2.3.4 Determination of Acropora habitat

ArcGIS software was used to overlay the Acropora spp. observation database on the benthic habitat maps. Observations of Acropora spp. located entirely within mapped coral reef or
The various classifications, which were considered coral reef and hardbottom in each of the benthic habitat maps, are shown in Table 2.2. In situ observation data points that fell outside of mapped reef or hardbottom were extracted and further examined. For each point not on a coral reef or hardbottom classification, the type of substrate was identified from the underlying benthic habitat maps.

Table 2.2: Coral reef and hardbottom classifications used for the benthic habitat maps in this study

<table>
<thead>
<tr>
<th>Classification</th>
<th>Florida</th>
<th>Puerto Rico</th>
<th>St. Croix</th>
<th>St. Thomas/St. John</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate reef</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Colonized bedrock</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Colonized pavement</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Colonized pavement with sand channels</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Linear reef</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Patch reef (aggregated)</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Patch reef (individual)</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Pavement</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reef rubble</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ridge</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scattered coral/rock in unconsolidated</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Spur and groove reef</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

2.3.5 Buffer generation

Buffers of various widths were created around mapped coral reef and hardbottom to compensate for errors associated with lower-resolution habitat maps (Tveite and Langaas, 1999). The frequency of points included within each buffer distance was determined. Buffers were created at 1-m increments until 95% and 99% of all points were included. This process was automated by creating a program using ArcGIS’s ModelBuilder, which created a set of 10 buffers at a time at 1-m increments for each benthic habitat map.

Observations were then separated into four regions, Florida, Puerto Rico, St. Thomas/St. John, and St. Croix, based on the benthic habitat map to which they corresponded. The percentage of points within each buffer size was calculated to determine the cumulative distribution of the four regions. A Kolmogorov–Smirnov (K–S) test for goodness of fit (Sokal
and Rohlf, 1981) was performed to determine whether any of the four regions differed significantly from the overall cumulative percentage distribution ($\alpha = 0.05$). The K–S test was performed for *A. palmata* and *A. cervicornis* individually and for both species combined.

### 2.3.6 Potential-habitat map generation

Based on the frequencies recorded in the various buffer sizes and the results of the K–S tests, potential habitat was delineated for each of the four regions by merging the appropriate buffered coral reef and hardbottom polygons. These polygons were then clipped to appropriate depth limits, using USGS FLaSH bathymetry in Florida and isobaths from National Geospatial-Intelligence Agency Digital Nautical Charts in Puerto Rico and the USVI. These potential habitat maps are models representing habitats in which *Acropora* spp. would be expected to be found, based on mapped habitats of previously observed colonies.

Potential habitat was determined and mapped for each species and for both species together, and at two confidence levels, 95% and 99%. A total of six potential-habitat maps were created. Potential habitat was determined based on the buffer distance where exactly 95% and 99% of observations were included, as well as the results of the K–S test. The *Acropora* spp. and *A. cervicornis* potential habitat maps were clipped to a 30 m depth limit to represent the maximum depth range of *A. cervicornis*, while the *A. palmata* potential habitat maps were clipped to a 10 m depth limit.

### 2.4 Results

A total of 18,732 locations of *Acropora* spp. presence in Florida, Puerto Rico, and the US Virgin Islands were reported to the database. Of those records, 380 (2%) were clearly erroneous and removed based on unreasonable locations in relation to bathymetry and coastline. For example, observations located on land or at depths greater than 50 m were assumed to be
erroneous and were removed. A total of 18,352 recorded *in situ* observations of *Acropora* spp. presence were retained for use in this study (Table 2.3).

Table 2.3: Number of observations of *Acropora* spp. by region of the Florida Reef Tract after erroneous points were removed from the database

<table>
<thead>
<tr>
<th>Observation</th>
<th>Florida</th>
<th>Puerto Rico</th>
<th>St. Croix</th>
<th>St. Thomas/ St. John</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Acropora palmata</em> presence</td>
<td>3,885</td>
<td>1,901</td>
<td>2,900</td>
<td>5,594</td>
<td>14,280</td>
</tr>
<tr>
<td><em>A. cervicornis</em> presence</td>
<td>3,832</td>
<td>54</td>
<td>40</td>
<td>146</td>
<td>4,072</td>
</tr>
<tr>
<td>Total presence</td>
<td>7,717</td>
<td>1,955</td>
<td>2,940</td>
<td>5,740</td>
<td>18,352</td>
</tr>
</tbody>
</table>

2.4.1 Habitats of observed *Acropora* spp.

A total of 17,069 observations coincided with previously mapped coral reef or hardbottom, accounting for 93% of the observations. *Acropora palmata* had the highest percentage of occurrence on previously mapped coral reef or hardbottom in the St. Croix region and the lowest in Florida. *Acropora cervicornis* had the highest percentage of occurrence on mapped reef or hardbottom in Puerto Rico and the lowest in Florida. Overall, observations of both species combined were most often located on mapped reef or hardbottom in St. Croix. In that region, 99% of all observations were located on mapped reef or hardbottom (Figure 2.2). In all regions except Puerto Rico, both *A. palmata* and *A. cervicornis* were observed outside mapped reef or hardbottom. While 100% of *A. cervicornis* observations in Puerto Rico occurred on mapped reef or hardbottom, the limited number of observations (54) was not enough to increase the percentage of both species combined over that at St. Croix. The substrate associated with observations not on mapped reef or hardbottom is detailed by region in Figures 2.3–2.6.

2.4.2 Buffer generation

The buffer sizes at which 95% and 99% of points were included are detailed for each species by region in Table 2.4. The benthic habitat maps for St. Croix required no buffer to include 95%–99% of *Acropora* spp. observations, while those for Florida required a buffer as wide as 490 m to include 99% of *A. cervicornis* observations. Figure 2.7 compares the
percentage of points included within each buffer distance by region. Most observations of *A. cervicornis* were recorded in Florida, and very few occurred in other regions. Therefore, the percentage of points across all regions is mostly influenced by the observations in Florida, as seen in Figure 2.7(b).

![Bar chart showing percentage of observations of Acropora spp. located on mapped coral reef or hardbottom.](image)

Figure 2.2: Percentage of observations of *Acropora* spp. located on mapped coral reef or hardbottom.

Results of the K–S test (Table 2.5) suggest that, for both species combined, the distribution in each region is significantly different from that across all regions combined. Distributions of *A. palmata* observations in St. Thomas, St. John, and Puerto Rico are not significantly different from those of the entire study area combined, but the distributions of *A. palmata* in St. Croix and Florida are significantly different. No distribution of *A. cervicornis* in any region is significantly different from distributions in the entire study area.
Figure 2.3: Associated substrate of Acropora spp. observations in Florida.

Table 2.4: Buffer distances within which exactly 95% and 99% of points are included

<table>
<thead>
<tr>
<th></th>
<th>St. Thomas/St. John</th>
<th>Puerto Rico</th>
<th>St. Croix</th>
<th>Florida</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95%</td>
<td>99%</td>
<td>95%</td>
<td>99%</td>
<td>95%</td>
</tr>
<tr>
<td>A. palmata</td>
<td>2 m</td>
<td>15 m</td>
<td>0 m</td>
<td>10 m</td>
<td>0 m</td>
</tr>
<tr>
<td>A. cervicornis</td>
<td>2 m</td>
<td>22 m</td>
<td>0 m</td>
<td>0 m</td>
<td>86 m</td>
</tr>
<tr>
<td>Both species</td>
<td>2 m</td>
<td>15 m</td>
<td>0 m</td>
<td>10 m</td>
<td>0 m</td>
</tr>
</tbody>
</table>
Figure 2.4: Associated substrate of *A. palmata* observations in Puerto Rico.

<table>
<thead>
<tr>
<th>Region</th>
<th><em>A. cervicornis</em></th>
<th><em>A. palmata</em></th>
<th>Acropora spp.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>D_{max}/D_{0.05}</td>
<td>n</td>
</tr>
<tr>
<td>Florida</td>
<td>3,816</td>
<td>0.01/0.02</td>
<td>3,384</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>54</td>
<td>0.15/0.19</td>
<td>1,901</td>
</tr>
<tr>
<td>St. Croix</td>
<td>40</td>
<td>0.07/0.22</td>
<td>2,900</td>
</tr>
<tr>
<td>St. Thomas/St. John</td>
<td>146</td>
<td>0.11/0.11</td>
<td>5,594</td>
</tr>
</tbody>
</table>

2.4.3 Generation of potential-habitat maps

Table 2.6 presents the buffer distances used for the six potential-habitat maps. The final maps were created by merging the buffered benthic-habitat maps for each region to create a single representation of *Acropora* spp. potential habitat. While these maps are best suited for use with GIS, Figures 2.8–2.11 nevertheless provide useful visual comparisons between the 95% potential habitat maps created in this study and the previously designated NOAA critical-habitat maps. Merging maps from multiple data sources, as is done here, has been useful in other studies (Giglio et al., 2010). While regions in the backcountry of the Florida Keys (Florida Bay and
Biscayne Bay) have been mapped as hardbottom, they were excluded from all potential habitat maps because no observations were recorded there. Moreover, this area is not suitable habitat for *Acropora* spp. due to high turbidity and variability in salinity and water temperature (Roberts et al., 1982; Shinn et al., 1989).

![Figure 2.5: Associated substrate of *Acropora* spp. observations in St. Croix.](image-url)
Fig 2.6: Associated substrate of *Acropora* spp. observations in St. Thomas/St. John.

### Table 2.6: Buffer distances used for potential-habitat maps

<table>
<thead>
<tr>
<th></th>
<th>A. cervicorns</th>
<th>A. palmata</th>
<th>Acropora spp.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>95%</td>
<td>99%</td>
<td>95%</td>
</tr>
<tr>
<td>Florida</td>
<td>97 m</td>
<td>468 m</td>
<td>11 m</td>
</tr>
<tr>
<td>St. Croix</td>
<td>97 m</td>
<td>468 m</td>
<td>0 m</td>
</tr>
<tr>
<td>Puerto Rico</td>
<td>97 m</td>
<td>468 m</td>
<td>0 m</td>
</tr>
<tr>
<td>St. Thomas/St. John</td>
<td>97 m</td>
<td>468 m</td>
<td>0 m</td>
</tr>
</tbody>
</table>
Figure 2.7: Percentages of points included for various buffer distances by region a) percentages of *A. palmata* b) percentages of *A. cervicornis* c) percentages of both species combined.
2.5 Discussion

The results of this study provide the methodology for mapping the extent of habitats known to support or likely capable of supporting *A. palmata* and *A. cervicornis* throughout US jurisdictions. At the time of listing of the species, NOAA designated critical habitat maps throughout the species range within US territory (Federal Register, 2008). The current critical habitat maps cover large areas of reef tract in Florida, Puerto Rico and the US Virgin Islands (Figures 2.8–2.11). These results will aid in refining these critical-habitat maps for the species and in accurately representing habitat necessary for their re-establishment.

![Figure 2.8: 95% Acropora spp. potential habitat compared to previously determined NOAA critical habitat around Florida](image)

Ideally, all observations of *Acropora* spp. would be located on previously mapped coral reef or hardbottom. In this study approximately 7% of observations were located outside of these classified bottom types. Most of these points (673) were *A. palmata*, but that accounted for only...
5% of the *A. palmata* observations. There were fewer *A. cervicornis* locations in the database, so the 610 observations located outside of mapped coral reef or hardbottom accounted for a higher percentage (15%). This reflects more extensive potential habitat for *A. cervicornis*, because habitat for *A. palmata* is less variable and more often limited to reef margins. These results are comparable to those of Wirt et al. (2013), who found 1% of *A. palmata* and 7% of *A. cervicornis* observations outside of mapped reef or hardbottom. A key question can be addressed using this data set: Do the data points that do not fall on mapped coral reef or hardbottom reflect mapping error, the potential of *Acropora* spp. to occupy habitats other than coral reef or hardbottom, or still other phenomena?

Figure 2.9: 95% *Acropora* spp. potential habitat compared to previously determined NOAA critical habitat around Puerto Rico.

Low spatial resolution of benthic habitat maps could result in observations outside of mapped coral reef or hardbottom; for example, a small patch reef encompassing an area smaller
than the minimum mapping unit would not be mapped differently than the surrounding habitat. Temporal variability of the habitats could have the same result. Areas of thin sediment-cover where hardbottom occurs intermittently both spatially and temporally within seagrass and sand could be mapped as potential habitat or completely overlooked.

Figure 2.10: 95% Acropora spp. potential habitat compared to previously determined NOAA critical habitat around St. Croix

Habitats have changed since the benthic habitat maps were created, which also could have resulted in outlying points. Recent hurricanes affecting south Florida, Puerto Rico, and the US Virgin Islands, such as Katrina, Wilma, Otto, and Kyle, were strong enough to alter habitats. Strong hurricanes dislodge and move hard substrate (Geister, 1980) and expose hardbottom, creating Acropora spp. habitat in previously uninhabitable regions. This problem can be resolved only by more frequent mapping, especially after significant storms.
The present database is a compilation of reported observations from many groups, agencies, and individuals. A number of errors could occur in data acquisition, including errors in recording latitude and longitude and those in simple data handling. The most likely error would be in data recording, but the fact that 95% of the *A. palmata* records fell on previously mapped reef or hardbottom indicates that the data set is highly reliable. Any point-pattern data may be spatially biased towards accessible areas, resulting in errors of omission (false negatives) or commission (false positives, Rondinini et al., 2006). This data set is no exception and is subject to errors of omission because a complete census of *Acropora* spp. does not exist (Pressey, 2004). While the potential habitat in this study contains inherent commission errors, these errors are generally seen as less ‘costly’ for conservation and management actions than errors of omission (Guisan and Zimmermann, 2000).
The number of reports of *A. cervicornis* outside of mapped reef or hardbottom reflects the greater range of habitats available to this species. Larvae can recruit to habitats dominated by sand or seagrass so long as there is hard substrate upon which it can settle, such as a shell or a small outcropping of hardbottom. *Acropora cervicornis* is also common on inshore and offshore patch reefs, which are more difficult to map than are the reef crests, where *A. palmata* is more commonly found. Additionally, *A. cervicornis* has been found to be spatially and temporally dynamic, with patches capable of moving tens of meters in a few years time (Walker et al., 2012). This dynamic nature of *A. cervicornis* supports the use of a buffer around coral reef and hardbottom habitat, as used in this study, in order to account for possible movement into adjacent areas.

Field surveys are needed to examine sites at which discrepancies occur between mapped habitat and observations. The results of such surveys can provide guidance in determining whether it is feasible to better define the potential habitat for *A. cervicornis*. But since 95% of the reported observations of *A. cervicornis* occur within 99 m of previously mapped coral reef or hardbottom, the current maps should be suitable for delineating critical habitat. Therefore, future mapping efforts should focus on reef areas for which habitat maps do not yet exist, such as those outside the Dry Tortugas.

This study confirms that *A. cervicornis* has a wider habitat range than *A. palmata* throughout Florida, Puerto Rico, and the US Virgin Islands. This difference is especially apparent off southeast Florida, where *A. cervicornis* appears to be thriving outside mapped coral reef areas and at latitudes considered marginal for hermatypic corals. The documentation of a wider range of habitat for *A. cervicornis* than for *A. palmata* indicates that different management strategies may be needed for the two species.
When the species were listed as threatened, NOAA designated critical-habitat maps throughout the species range within US territory. The newly generated potential-habitat maps were compared with NOAA’s mapped critical habitat. Overall, the new potential-habitat maps are more refined in spatial coverage than the current critical-habitat maps. An important aspect of the new potential-habitat maps concerns the southeast coast of Florida. The northern extent of the NOAA-designated critical habitat is in the southern portion of Palm Beach County, whereas the northern extent of all versions of the new potential-habitat maps is farther north, in the southern portion of Martin County. While no observations were located in Martin County, 10 observations of *A. cervicornis* were located north of the NOAA critical habitat in Palm Beach County. The caveat associated with this extended region is that it is defined only by the presence of mapped hardbottom. The northernmost extent of *Acropora* spp. is still not precisely known, but all *A. cervicornis* observations in this study were located south of 26.682°N. The northern extension of potential habitat within Palm Beach County exhibits additional areas of reef and hardbottom available for *Acropora* spp. settlement. However, settlement of *Acropora* spp. in southern Martin County is unlikely due to unique hydrodynamics present in this region. A recent study by Walker and Gilliam (2013) suggests the unlikelihood of northern expansion of tropical coral communities into the Martin County region due to frequent and intense upwelling events that occur north of the Bahamas Fault Zone (Klitgord et al., 1984; Pitts, 1999; Pitts and Smith, 1997; Smith, 1983). Coral communities north of this fault zone are restricted to cold tolerant species such as *Siderastrea* and *Oculina* (Colella et al., 2012; Lirman et al., 2011; Walker and Gilliam, 2013).

Defined potential habitat around the Dry Tortugas is confined to the extent of the currently mapped reef and hardbottom in the region. NOAA’s critical-habitat map extends
farther outside of Dry Tortugas National Park and, therefore, encompasses regions that are believed to be *Acropora* spp. habitat. But the current potential-habitat map does not include these regions due to the limited extent of the mapped reef and hardbottom in the region. Throughout the Caribbean regions (Puerto Rico and the US Virgin Islands), the new potential-habitat maps do not encompass additional *Acropora* spp. habitat. Instead, they reduce the coverage to more manageable areas, areas in which *Acropora* spp. is likely to be found. But the number of observations of *A. palmata* greatly exceeds that of *A. cervicornis*. This is likely a result of the concentration of surveys on linear reef zones rather than on patch reefs, where *A. cervicornis* is commonly found. Therefore, the unevenness of these observations throughout this region may have resulted in the exclusion of some potential *A. cervicornis* habitat. Clearly, more observations of *A. cervicornis* are needed to improve the accuracy of the potential-habitat maps.

Multiple potential-habitat maps were produced from the results of this study that will have value in a number of management efforts. In general, the 95% *Acropora* spp. potential habitat maps may be best suited for general purposes, because their boundaries were determined based on the presence of both species. However, the 99% *A. cervicornis* potential-habitat map could be used when the maximum area of potential habitat is desired, given the wider range of *A. cervicornis*. Alternatively, the species maps can be used individually, recognizing the different environmental requirements of the two species relative to light availability and water movement. These maps show suitable substrate that could be available for the settlement of *A. palmata* and *A. cervicornis*, and do not assign any type of threat potential or probability of survival facing potential transplants.
2.5.1 Possibilities for future research

The potential for future work related to this study is extensive. More extensive in-water surveys are needed to clarify any anomalous observations in relation to mapped habitat. This study also revealed the need for additional mapping efforts in the Dry Tortugas, where a large portion of habitat is currently unmapped.

Determining the benthic habitat required for these species is only the first step in determining true critical habitat. Other factors, such as water transparency, water quality, and temperature, are also important determinants of suitable habitat for the re-establishment of the species. The next step should be to incorporate these parameters into the potential-habitat maps to more specifically identify critical habitat for Atlantic and Caribbean Acropora spp. The results of this and future studies can also provide the framework for determining critical habitat for other species of coral.

Future work could also examine habitats where Acropora spp. was once found but from which it has since disappeared. The detection of these locations could provide the opportunity to identify which environmental parameters are now different from those at locations where Acropora spp. is still present. This may improve understanding of the most crucial environmental parameters necessary for these species’ survival.

2.6 Acknowledgements

This project was funded by award NAIONMF4720029 from the National Oceanic and Atmospheric Administration’s National Marine Fisheries Service (NMFS), US Department of Commerce. The statements, findings, conclusions, and recommendations are those of the authors and do not necessarily reflect the views of NMFS or the US Department of Commerce. Special thanks go to FWRI’s Center for Spatial Analysis for resources and support necessary for this
project. The first author also received financial support from the Linton Tibbetts Graduate Fellowship, the Gulf Oceanographic Charitable Trust Endowed Fellowship, and the Southern Kingfish Association Endowed Fellowship, all from the University of South Florida’s College of Marine Science.

2.7 References


3 Species Distribution Model Data Compilation and Processing

3.1 Abstract

The compilation and processing steps used to amass relevant reef ecosystem data for spatial modeling and mapping purposes is often time consuming and tedious. The purpose of this chapter is to describe the spatial data compiled and the processing steps taken to format the data. The primary use of the data described here is for a study that identifies suitable restoration sites for the staghorn coral, *Acropora cervicornis*, using a species distribution modelling (SDM) approach. At the beginning of the project, a series of stakeholder workshops were held to identify the most important ecological and environmental data relevant to *A. cervicornis* distributions. The final data layers created in this study include 1 km² raster representations of: region, reef type, distance from shore, depth, slope, sea bottom temperature, and light availability. The *A. cervicornis* monitoring data used as the response variable was also analyzed and formatted for use in a SDM. The data layers described represent the currently best available environmental data to use for *A. cervicornis* SDMs. This study confirms that there is no shortage of spatial data available for the creation of an *A. cervicornis* SDM, and the data sets described can also be used for prediction of distributions of other shallow water hermatypic coral species throughout the Florida Reef Tract.

3.2 Introduction

The task of gathering and formatting data for the purpose of geographic analysis can often be the most time consuming and tedious aspect of a project. While laborious, carefully
formatting data is also one of the most important tasks, and sets a project up for success. The purpose of this chapter is to describe the compilation and processing steps used to compile relevant reef ecosystem data to use for multiple spatial modeling and mapping purposes. The primary use of the data described in the following sections is for a study that identifies suitable restoration sites using a species distribution modeling (SDM) approach (See Table 3.1 for all abbreviations and acronyms). There are a variety of other uses for the compiled data, however the specific steps required to format the data for SDM development are described herein.

Two types of data are essential to a SDM, responses and predictors. Response data are generally observations of species occurrence, and predictor data are the environmental or ecological data that are believed to contribute to the likelihood of a species existing in a certain location (Guisan and Zimmermann 2000; Guisan and Thuiller 2005; Elith and Leathwick 2009). Prediction performance of a SDM is enhanced when knowledge of ecological processes is incorporated in model development, as opposed to simply integrating any and all available data (Austin 2002). Therefore, by taking the time to develop species-specific predictor data to be used in model development, the resulting SDM will be more robust and will provide more informative ecological information.

Model predictors can be grouped into three categories: resource, direct and indirect gradients. Resource gradients are matter and energy consumed by plants or animals. Examples of resource gradients include nutrients, water, light for photosynthesis, and food. Direct gradients are environmental parameters with physiological importance, but are not consumed, such as temperature and pH. Indirect gradients have no direct physiological relevance for a species’ performance, but may replace a combination of different resources and direct gradients in a simpler way (Austin and Smith 1989; Guisan et al. 1999; Guisan and Zimmermann 2000).
Examples of indirect gradients include slope, depth, and habitat type. In the marine system, depth is an indirect proxy for several direct and resource gradients: temperature and its variability, water motion, salinity, and light (Elith and Leathwick 2009). The use of strictly indirect predictors limits a model’s transferability, because in different regions, the same indirect parameters may have very different direct and resource gradient correlations (Guisan and Zimmermann 2000). While models based on resource and direct gradients have been found to be the most robust and widely applicable (Austin 2002), indirect gradient data are essential when direct and resource gradient data are unavailable.

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDM</td>
<td>Species Distribution Model(ing)</td>
</tr>
<tr>
<td>PCS</td>
<td>Projected Coordinate System</td>
</tr>
<tr>
<td>UFRTM</td>
<td>Unified Florida Reef Tract Map</td>
</tr>
<tr>
<td>LiDAR</td>
<td>Light Detection and Ranging</td>
</tr>
<tr>
<td>SST</td>
<td>Sea Surface Temperature</td>
</tr>
<tr>
<td>SBT</td>
<td>Sea Bottom Temperature</td>
</tr>
<tr>
<td>PAR</td>
<td>Photosynthetically Available Radiation</td>
</tr>
<tr>
<td>CREMP</td>
<td>Coral Reef Evaluation and Monitoring Project</td>
</tr>
<tr>
<td>OOB</td>
<td>Out-of-bag</td>
</tr>
<tr>
<td>DIN</td>
<td>Dissolved Inorganic Nitrogen</td>
</tr>
<tr>
<td>TP</td>
<td>Total Phosphorus</td>
</tr>
<tr>
<td>GAM</td>
<td>Generalized Additive Model</td>
</tr>
<tr>
<td>FRT</td>
<td>Florida Reef Tract</td>
</tr>
</tbody>
</table>

3.3 Methods

3.3.1 Stakeholder Workshops

At the beginning of the project, two stakeholder workshops were held to inform the data analysis and modeling approach. A main discussion point was identification of the most important ecological and environmental data relevant to Acropora spp. distributions, as well as the availability of such data. Each workshop hosted 20-30 participants including researchers, nursery supervisors, managers, and other relevant coral-restoration experts, thus providing a wide range of expert opinion on data most important to predict restoration success. One of the
most important considerations when selecting data used in the modeling process is that the data are ecologically relevant to the species of interest (Elith and Leathwick 2009). The expert opinion provided in these stakeholder workshops guided the effort to select such data used to predict *A. cervicornis* distributions. The results of these discussions dictated the data layers created and further discussed in this document.

Restoration efforts were a core focus during the stakeholder engagements. Given the choice of focusing the restoration modeling efforts on all stony coral species combined, or to limit the focus to a few target species, it was recommended that focus should be concentrated on *A. cervicornis*. Currently, a majority of restoration efforts within the Florida Reef Tract (FRT) have targeted *A. cervicornis* and *A. palmata* (Johnson et al. 2011; Young et al. 2012). While the following data layers were developed with *A. cervicornis* modeling in mind, they are all appropriate and applicable to other shallow water stony coral species on the Florida Reef Tract.

### 3.3.2 Study Area

The combination of a broad, shallow continental shelf and warm waters of the Gulf Stream uniquely suit Florida’s benthic habitat for the support of coral reef ecosystems (Jaap and Hallock 1990; Andrews et al. 2005). The study area included the Florida Reef Tract, from Martin County through the Dry Tortugas, with the exception of Biscayne Bay due to limited *in situ* observations in Biscayne Bay (Figure 3.5). For each data layer described below, a boundary layer was used to clip the data to the appropriate geospatial extent.

### 3.3.3 Spatial Resolution and Data Formatting
One of the most important tasks involved in building a SDM is the determination of an appropriate spatial scale (Wiens 2002). For sessile organisms, such as coral species, finer spatial resolution data often provide better predictions in SDMs (Guisan and Thuiller 2005). There is no single natural scale at which ecological patterns should be studied (Levin 1992). Instead, appropriate scales are determined by the study goals, the system and data availability (Elith and Leathwick 2009).

The spatial resolution of all layers in this study was set to 1 km\(^2\), as designated by the coarsest resolution of the data sets used. All data layers were projected to the Albers projected coordinate system (PCS Albers) and converted to raster format, with the exception of the \textit{A. cervicornis} observation point data, which remained in vector point format, also projected in PCS Albers. Data layers used in this study are summarized in Table 3.2

<table>
<thead>
<tr>
<th>Data Layer</th>
<th>Format</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>Raster</td>
<td>Seven regions of the Florida Reef Tract</td>
</tr>
<tr>
<td>Reef type</td>
<td>Raster</td>
<td>Dominant reef type within 1 km(^2) cell</td>
</tr>
<tr>
<td>Distance from Shore</td>
<td>Raster</td>
<td>Distance from shore at 1 km increments</td>
</tr>
<tr>
<td>Depth</td>
<td>Raster</td>
<td>Depth interpolated from LiDAR and point bathymetric data</td>
</tr>
<tr>
<td>Slope</td>
<td>Raster</td>
<td>Slope derived from depth layer</td>
</tr>
<tr>
<td>SBT High</td>
<td>Raster</td>
<td>Number of weeks above 30(^\circ) C from 2008 – 2011</td>
</tr>
<tr>
<td>SBT Low</td>
<td>Raster</td>
<td>Number of weeks below 20(^\circ) C from 2008 – 2011</td>
</tr>
<tr>
<td>Light380</td>
<td>Raster</td>
<td>Percent total light at 380 nm available at depth</td>
</tr>
<tr>
<td>Light488</td>
<td>Raster</td>
<td>Percent total light at 488 nm available at depth</td>
</tr>
<tr>
<td>\textit{A. cervicornis} survey data</td>
<td>Point</td>
<td>Long term trends from repeat \textit{A. cervicornis} survey data, categorized by presence/absence trends</td>
</tr>
</tbody>
</table>

3.4 Data Layer Methods and Results

3.4.1 Region

There are three main regions of the Florida Reef Tract: the Tortugas Bank, Florida Keys, and the southeastern coast. All regions are subject to unique environmental conditions and, consequently, varying degrees of health and species composition (Andrews et al. 2005; Walker
2012). For this study, the Florida Reef Tract was divided into seven regions: Upper East Coast, Lower East Coast, Upper Keys, Middle Keys, Lower Keys, Marquesas and Tortugas (Figure 3.6).

The large latitudinal gradient of the reefs along the east coast of Florida designates the need for two regions in this area. Additionally, previous studies by Walker (2012) and Walker and Gilliam (2013) identified spatial barriers that coincided with distinct habitat differences. Based upon these findings, the boundary between the upper and lower east coast regions was positioned at Lake Worth inlet, and the boundary between the lower east coast and the upper Keys was positioned at Government Cut (B. Walker personal communication). All regional boundaries were created in polygon format and then converted to 1 km$^2$ raster format.

3.4.2 Reef Type

Several types of coral reef habitats are found along the Florida Reef Tract. Until recently, many different groups, agencies and institutions mapped the habitats of the reef tract, each using a unique classification system. To remedy this, the Florida Fish and Wildlife Conservation commission introduced the Unified Florida Reef Tract Map (UFRTM). The UFRTM integrates existing benthic habitat maps to provide a consistent classification of the Florida Reef Tract from Martin County to the Dry Tortugas (Baumstark 2013). According to the UFRTM, there are five main types of coral reef habitats: Aggregate Reef, Patch Reef, Pavement, Reef Rubble and Ridge.

The UFRTM was used to create a raster representation of the dominant reef types at 1 km$^2$ resolution. In order to only represent reef types, all coral reef and hardbottom polygons were selected out of the UFRTM. A 1 km$^2$ grid was overlaid on the extracted polygons, and the
dominant reef type by area was calculated per grid cell. Figure 3.3 shows a single grid cell with the individual reef-type polygons. In this example, the 1 km² grid cell was categorized as Aggregate Reef. The final product is a 1 km² resolution raster that represents the dominant reef type within that cell (Figure 3.4). This information was used to describe the reef habitat, while removing other features such as sand or seagrass that would often dominate grid cells. The reef type that most often dominated was Pavement, followed by Individual or Aggregated Patch Reef (Table 3.3).

<table>
<thead>
<tr>
<th>Habitat Type</th>
<th>Number of cells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate Reef</td>
<td>637</td>
</tr>
<tr>
<td>Individual or Aggregated patch</td>
<td>751</td>
</tr>
<tr>
<td>Pavement</td>
<td>1485</td>
</tr>
<tr>
<td>Reef Rubble</td>
<td>27</td>
</tr>
<tr>
<td>Ridge</td>
<td>309</td>
</tr>
<tr>
<td>Not reef</td>
<td>3216</td>
</tr>
</tbody>
</table>

While this layer provides the necessary information regarding reef type in raster format, there are a few known issues. Specifically, since only reef polygons were extracted, there are cases where cells that are predominantly not-reef (sand or seagrass) were still categorized as reef in some way. For example, Figure 3.5 represents a 1 km² grid cell that is predominantly not-reef. However the process of creating this layer resulted in that cell being categorized as patch reef. This issue is of little concern to the overall use of this layer, especially at the 1 km² resolution. The purpose of this layer was to depict the dominant reef type, and not quantify amount of reef per cell or represent other non-reef habitat types. Additionally, there are likely unmapped features in this grid cell, such as very small patch reefs that were not depicted in the UFRTM, justifying the classification of this and similar cells as patch reef.
3.4.3 Distance from shore

Inshore coral reef habitats are subject to higher nutrient levels (DIN and TP), turbidity, and light attenuation. As the distance from shore of a reef increases, these parameters decrease (Lirman and Fong 2007). Counterintuitive to this concept, studies have found that values of coral cover on the Florida Reef Tract decrease significantly with increased distance from shore (Lirman and Fong 2007). Regardless of the driving forces, proximity to shore has been found to correlate with coral cover throughout the Florida Keys. This concept was represented by creating a distance-from-shore layer.

To uniformly represent distance from shore throughout the Florida Reef Tract, a straight line along the coast of Florida was drawn for the entire study extent. Special attention was paid to the Biscayne Bay region, where a straight line was drawn across the inlet to the bay, rather than indenting into the bay. Doing so kept a strange feature from extrapolating out from this region. The multiple ring buffer tool in ESRI ArcGIS was used to create buffers at 1 km increments along this line. All buffer rings were then split in half and the landward sides of the rings were removed, leaving only concentric polygons at 1 km increments seaward from shore. These polygons were then converted to raster format (Figure 3.6).

This layer provides a proxy for multiple influences on reef health, such as anthropogenic stress, influence of elevated nutrients, as well as habitat type and quality. Additionally, this layer aims to represent the spatial distribution of ocean acidification threats. A study by Manzello (2012) hypothesized that inshore reefs closer to seagrass beds act as a potential refugia for ocean acidification. Quantifying the distance from shore attempts to quantify multiple impacts on coral reef development, both positive and negative.
3.4.4 Depth

Depth is one of the most important habitat requirements for many species of coral, with *A. cervicornis* restricted to depths less than 30 m (Fenner 1988). Prior to this study, the spatial coverage of detailed depth information was lacking throughout the Florida Reef Tract. Regional and small-scale acoustic and LiDAR data are available for some regions, but large gaps throughout the reef tract still exist. Lower resolution depth data, from sources such as nautical charts and bathymetric contours, exist throughout the reef tract; however, these sources are not in a data format usable for the geographic analysis required for this study.

To create the appropriate depth layer, LiDAR and acoustic data were combined with data points from nautical charts to create an interpolated raster coverage. First, all available LiDAR and acoustic data available were aggregated to 1 km$^2$ resolution based on cell averaging. Data sources included LiDAR from the Dry Tortugas (USGS), the Marquesas (NOAA), Martin County (NOVA Southeastern University), Miami-Dade County (NOVA Southeastern University), Palm Beach County (NOVA Southeastern University), Broward County (FWRI), Biscayne National Park (FWRI), and the Middle Keys (FWRI), as well as acoustic data from the Dry Tortugas, the Marquesas and the Lower Keys (FWRI and USF). Depth values were initially converted to feet, and centroids of the 1 km$^2$ aggregated raster cells were converted to points. The 1 km$^2$ centroid point file was appended to fill in spatial coverage gaps by adding 1,031 random depth points (converted to feet) taken from NOAA Nautical Coastal and Approach charts, resulting in a total of 4,700 depth points for the entire Florida Reef Tract. Any point above 0 feet (0 m) elevation, and beyond 200 feet (60 m) depth were excluded, resulting in a final total of 4,554 points (Figure 3.7a). Depth values were log-transformed to approximate normal residuals. An Empirical Bayesian Kriging (EBK) model interpolated depth values in
ArcGIS geostatistical analyst, with 1,000 simulations. The final model resulted in a Root Mean Square prediction error of 11.83 feet (3.6 m). The final EBK model was exported to a 1 km² grid and depth values were converted to meters (Figure 3.7).

3.4.5 Slope

Benthic geomorphology has been found to have an influence on coral species distributions (Goldberg 1973; Franklin et al. 2013). Given the scale of the study and the relatively simple geomorphology of the Florida Reef Tract at that scale, the only measure of benthic geomorphology considered was slope. Slope was derived from the previously described depth layer using the slope tool in ArcGIS spatial analyst (Figure 3.8). This tool calculates the steepest angle, measured in degrees, of a plane defined for a depth grid cell and its eight neighbors. There is no directionality associated with the slope, it is only a representation of the coarse relief of the area.

During the previously mentioned stakeholder workshops, it was noted that this layer may not be particularly useful at the 1 km² scale. However, it was included in model development, and later dropped if a model test identified it as a non-constructive predictor.

3.4.6 Sea Bottom Temperature

The distributions of many marine species are strongly controlled by sea temperature, and mortality of nursery corals has often been attributed to temperature disturbance (Quinn and Kojis 2006; Schopmeyer et al. 2012). Sea-surface temperature (SST) is the most common indicator of temperature used in the marine environment. Generally, SST is an appropriate indicator of thermal stress experienced by coral. However, coastal waters in the Florida Keys have been shown to exhibit inverted thermocline events over the summer (Porter et al. 1999). These events
are driven largely by the discharge of hypersaline water from Florida Bay and contribute to elevated bottom temperatures experienced along portions of the Florida Reef Tract. Historically, water in Florida Bay has been hyposaline, allowing it to safely pass over the reefs of the Florida Keys (McIvor et al. 1994; Pitts 1994; Porter et al. 1999). In the 1950’s, synonymous with the construction of a series of levees and canals in the Everglades by the Army Corps of Engineers (Light and Dindeen 1994), researchers began recording hypersaline conditions year round (McIvor et al. 1994). McIvor et al. (1994) suggests that the disruption of freshwater flow from the Everglades, combined with lower than average rainfall, has contributed to the hypersaline conditions in Florida Bay. In the summer, the hypersaline waters of Florida Bay rise in temperature. This hot, hypersaline water then moves from the shallow areas of Florida Bay and enters the Florida Keys reef tract. Due to its increased density, this warm water sinks and creates a stable inverted thermocline (Porter et al. 1999).

Traditional coral bleaching models using only SST may overestimate or underestimate bleaching risk in systems where sea bottom temperature (SBT, the temperature that corals experience) differs considerably from SST. Cold-water-induced stress and mortality are also important factors in coral reef communities (Glynn and Stewart 1973; Glynn 1976; Coles and Fadlallah 1991; Hoegh-Guldberg et al. 2005). The Florida Keys have experienced periodic cold-water events that have been found to have pronounced effects on coral mortality and benthic cover (Colella et al. 2012).

The use of modeled environmental parameters in SDMs has been increasing and including such parameters contributes to more ecologically meaningful SDMs (Guisan and Zimmermann 2000). Given the sparse spatial coverage of in situ bottom temperature data throughout the Florida Reef Tract, a model of bottom temperature was required. A Generalized
Additive Model (GAM) was created to predict SBT from satellite-derived SST, using \textit{in situ} SBT observations collected from areas throughout the reef tract. The satellite and \textit{in situ} observations were collected from 2008 to 2011. The resulting GAM was used to generate 1 km$^2$ raster coverages of daily SBT conditions. These daily SBT coverages were averaged by week and the number of weeks above 30˚ C and below 20˚ C per 1 km$^2$ grid cell were tabulated. The resulting raster coverages depict areas of warm and cold thermal stress, as predicted from SBT (Figure 3.9). Full details on the creation of the GAM can be found in McEachron et al. (in review). Low temperature stress is much more difficult to predict, given the episodic and widespread nature of such events. Patterns of low-stress are much less distinct, with low temperature events impacting the full reef tract relatively evenly. Discerning any spatial patterns of low temperature stress is therefore problematic.

\subsection*{3.4.7 Light Availability}

Zooxanthellate corals are symbiotic organisms that obtain a majority of their energy and carbon requirements from photosynthesizing zooxanthellae present in their tissues (Muscatine et al. 1981). It is well understood that hermatypic corals exist in areas where suitable substrate is present, but also where there is sufficient light available to maintain photosynthesis (Wells 1957). The lower bathymetric limits for individual species are determined by their photosynthetic efficiency at reduced levels of irradiance. For many species, photosynthesis becomes limiting where light availability approaches 10\% of surface light levels (Done 1983). Previous studies have found that under reduced light levels, corals often produce fragile skeletons, increasing the threat of fragmentation when subject to high energy waves breaking across the reef (Cook et al. 1997; Yentsch et al. 2002). Yentsch et al. (2002) concluded that some Florida coral populations are squeezed between two bathymetric limitations: they are unable to
grow at deeper depths due to light limitations but are also unable to grow in shallower regions due to high wave energy, which limits colonization. This is very likely the case for *A. cervicornis* populations on Florida reefs. Therefore, light availability combined with depth becomes one of the most vital predictors of coral reef cover status and trends along the Florida Reef Tract.

MODIS/Aqua satellite-derived values of the diffuse attenuation coefficient (*K*<sub>d</sub>) were obtained from University of South Florida’s Optical Oceanography Laboratory. The Optical Oceanography Laboratory used improved algorithms to obtain *K*<sub>d</sub> values over the optically shallow water of the Florida Keys (Barnes et al. 2013; Barnes et al. 2014). Two yearly climatological coverages were provided, *K*<sub>d</sub>(380) and *K*<sub>d</sub>(488) for the MODIS data period of 2002 - 2015. These wavelengths were used as a proxy of two important measurements of light: UV light (λ = 380 nm) and photosynthetically available radiation (PAR, λ = 488 nm) reaching the benthos. While PAR refers to a large spectral band (400 – 700 nm), 488 nm serves as a proxy for PAR reaching the benthos. Studies by Kratzer et al. (2003), Pierson et al. (2008), and Zhao et al. (2013) have found a direct relationship between *K*<sub>d</sub>(PAR) and either *K*<sub>d</sub>(490) or euphotic zone depth [directly proportional to *K*<sub>d</sub>(490)], therefore justifying the use of *K*<sub>d</sub>(488) to approximate PAR reaching the benthos in this study, as done in a study by Barnes et al. (2015).

The previously described depth layer was converted to points, and the Extract Values to Points tool in ArcGIS Spatial Analyst was used to append the *K*<sub>d</sub> values coincident with the depth points. As the improved algorithm was designed for optically shallow waters (limited to 5 – 30 m), *K*<sub>d</sub> values for depths less than 5 m were not available. Much of the study area, especially along the east coast of Florida, was located in depths less than 5 m, leaving large gaps in the light availability data coverage. To remedy these coverage issues, a nearest neighbor resampling approach was used to fill in missing data values. The nearest neighbor resampling approach was
deemed appropriate, as $K_d$ values do not vary significantly inshore to offshore within the current study area (B. Barnes, personal communication). Additionally, zooxanthellate coral living in depths less than 5 m are not limited by light reaching the benthos. Nearest neighbor resampling was conducted for any depth point of the study area where MODIS-derived $K_d$ values were not available, a total of 188/5,629 points were resampled for the $K_d(488)$ layer and 197/5,637 points were resampled for the $K_d(380)$ layer.

Once all depth points were assigned $K_d$ values, the Beer-Lambert Law (Equation 1) was used to calculate percent light availability at depth (Gordon 1989):

$$E_d(z) = E_d(0)e^{-K_d(z)}$$  (1)

Where $z$ is depth (m), $K_d$ is the diffuse attenuation coefficient for downwelling irradiance and $E_d(0)$ is the downwelling irradiance at the surface (Kirk 1994). A value of 100 was used for $E_d(0)$ to obtain percent total light available at depth.

3.4.8 Multicollinearity analysis

An important step in model development is checking data layers for multicollinearity. Multicollinearity occurs when high correlation exists between several predictors (Guisan and Zimmermann 2000; Guisan and Thuiller 2005). Multicollinearity is of greatest concern when using linear or generalized linear models for the SDM approach, and of little concern when using classification trees. However, diagnosing multicollinearity early in the study analysis assists with decisions to remove variables during the model fitting process.

A stratified random sample of 10% of all variables by region was taken across the entire study area. Stratifying the random sample by region ensured equal distribution across the entire
reef tract. The random sample resulted in 450 records of each predictor variable covering the entire study area. A simple linear regression was run on all combinations of variables reasonably assumed to be correlated.

No strong linear relationships were found between any of the predictor variables. While many of the p-values were significant, this is likely an artifact of the large sample size used for the regression. The $R^2$ values and the corresponding linear regression plots in Figure 3.11 indicate that there were no strong correlations between variables, and all predictor variables can be included in the model fitting process. Regardless of the significant p-values, the assumptions of the non-parametric classification models used in the following study were not violated. Classifiers look at each predictor variable separately, rather than using relationships between variables to fit the model, therefore correlations between predictor variables do not impact the resulting model fit.

3.4.9 Acropora cervicornis survey data

The use of coral survey data in SDM development is rare, but the amount of high quality data available makes coral species ideal targets for SDM analysis. The Florida Fish and Wildlife Conservation Commission collects annual data on the status of coral habitats along the Florida Reef Tract through the Coral Reef Evaluation and Monitoring Project (CREMP). The CREMP team began collecting data at 40 sites in the Florida Keys in 1996, expanded with three additional sites in the Dry Tortugas in 1999 and 10 more sites along Florida’s southeast coast in 2003. In addition to stony coral benthic habitat cover data, CREMP also records stony coral richness at the species level (Callahan et al. 2007). CREMP is the largest, most comprehensive source of repeat monitoring data in Florida, and therefore, lends itself to a multitude of data analysis opportunities.
All repeat monitoring data on *A. cervicornis* presence/absence within the study area were extracted from the CREMP database. CREMP surveys were comprised of four stations per site. These four stations were lumped into a single point, and the data were analyzed on a per-site scale. The four stations per CREMP site were not sufficiently spatially distinct at the resolution of this study to allow for individual analysis. By lumping individual stations into a single site, spatial autocorrelation errors were avoided, and confusing the model with multiple points per 1 km² grid cell was avoided.

In addition to the CREMP survey data, the *Acropora* spp. location database described in Wirt et al. (2013) was used to identify any non-CREMP repeat *A. cervicornis* monitoring in Florida using the ‘find identical’ tool in ArcGIS. Any survey location with four or more years of *A. cervicornis* presence/absence monitoring was included for analysis. These survey points were merged with the CREMP survey data. Any CREMP locations with single sample years were removed. Each survey location was analyzed individually, and one of three categories was assigned to the location based on the trend of the multi-year data (Table 3.4). The original five categories described in Table 3.3 were pooled into three, more general, categories. The original “Disappeared” and “Appeared” categories had limited observations, thus were pooled into the “Transient Presence” category.

Once each observation was assigned a category, the spatial distribution of the survey data was examined for spatial autocorrelation. Any points within 100 m of one another were combined and categories were assigned based on the most logical value. Nearly all combined observations were categorized uniformly from the start, so the combination did not result in any major data changes.
Table 3.4: Categories assigned to repeat monitoring *Acropora cervicornis* data.

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Pooled Category</th>
<th>Number of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never Present</td>
<td>Surveys never recorded presence of the species at any year</td>
<td>1</td>
<td>56</td>
</tr>
<tr>
<td>Transient Presence</td>
<td>The species was detected some survey years, but not others</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>Continuously Present</td>
<td>Surveys detected at least one observation of the species every year</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>Disappeared</td>
<td>The species was detected the first several survey years, but then was not detected the most recent survey years</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Appeared</td>
<td>The species was not detected the first several survey years, but then was continuously detected the most recent survey years</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

For SDM development, it is common to separate data into training and validation data sets (Pearson et al. 2007; Elith and Leathwick 2009). Training data are species records used to fit the model and validation data are those observations held out for evaluation of predictive performance once the model has been parameterized (Elith and Leathwick 2009). Given the small size of the *A. cervicornis* data set, the sensitivity of the data in relation to the number of records held out for validation purposes was evaluated. A basic Random Forest model (Breiman 2001) was run multiple times, using different percentages of data set aside for training and validation. For each run, a stratified random sample was taken from the data set and used to fit a random forest with 500 trees, and all other parameters set to the default values. Samples of 90%, 80%, 70%, 60% and 50% of the data were used to train the model. To account for the random factor in the appropriately named random forest, each subset option was run 10 times, and the out-of-bag (OOB) error estimates were averaged (Figure 3.12).

When using a Random Forest model, there is no need for cross-validation or a test set to get an unbiased estimate of the training set error. This value is estimated internally and called the OOB error estimate (Svetnik et al. 2003; Peters et al. 2007). Each tree in the random forest is
constructed using a different bootstrap sample from the original data, then compared to the portion of the data left out from the bootstrap (Svetnik et al. 2003; Cutler et al. 2007). The results of this study suggest that leaving out 10-30% of the data results in essentially the same error estimates, with higher error estimates when using 60% or less of the data for training. The runs using 90% of the data have a low standard deviation, as well as the lowest error rate, suggesting that the data are not particularly sensitive and that modifying the model with 90% of the data will result in higher confidence in the results.

Based on the results of the above error study, the A. cervicornis survey data was separated into 90% training and 10% validation data sets. The same data sets were used for all models for consistency purposes. The stratified random sample of the data ensured that both the training and the validation data sets had all three categories represented (Table 3.5). Additionally, visualization of the spatial distribution of both data sets ensured that points were distributed throughout the study area and not clumped in a single location (Figure 3.13).

Table 3.5: Number of observations in each of the A. cervicornis data sets.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Category 1</th>
<th>Category 2</th>
<th>Category 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>51</td>
<td>24</td>
<td>13</td>
</tr>
<tr>
<td>Validation</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

3.4.10 Data compilation

Once all predictor and response data were properly formatted, the data were compiled into files to be used for the species distribution modeling analysis. Using ArcGIS’s Spatial Analyst extension, the extract-multi-values-to-points tool was used to append the Acropora trend data with all predictor variable values. This resulted in a table with all Acropora trend data points and their coincident values for all nine predictor variables. All A. cervicornis trend points were analyzed for NA values, and points with true NA values for any of the predictor data layers were
removed. Any point with a NA value that was located within 100 m of a predictor layer was assigned the nearest neighbor value.

To complete the species distribution modeling analysis and predict to unsampled locations, another point file was created that covered the entire extent of the study area. A layer with points located 1 km apart was created. To speed up processing, a script was written in R version 3.2.1, that assigned the values of the raster predictor layers to the point file covering the full extent. Once all values were assigned to the point value, all values of NA were removed. This resulted in a predictor point layer of points where all nine predictors had values (Figure 3.14). The resulting point file contained 4,469 points covering the full study extent. This layer was then used after model development to predict the *A. cervicornis* trend in all areas of the reef tract.

3.5 Discussion

The previously described data layers represent the best available environmental data to use for *A. cervicornis* species distribution models. As with any study, there are areas where the data could be improved. The most obvious improvement will be in the spatial resolution. The current study was restricted by the highest resolution available from the analyzed satellite data (1 km$^2$). The importance of the layers created from the 1 km$^2$ spatial resolution MODIS data dictated the need for the remaining layers to be represented at the same spatial scale. Instead of a site-specific product, the results of the SDMs should be used as a regional guide for restoration efforts. To provide additional guidance, the results of the SDM could possibly be clipped to reef area using existing benthic habitat maps. This would provide slightly more detailed recommendations than the current 1 km$^2$-resolution raster coverage. Until light availability and temperature data (and corresponding algorithms) are available at resolutions higher than 1 km$^2$,
future studies will also be restricted to this resolution. However, even at the 1 km² resolution, the resulting SDM product provides valuable information for restoration projects. Mortality of nursery corals has been described as highly site-specific, with threats ranging from temperature anomalies and predation to poor water quality (Quinn and Kojis 2006; Schopmeyer et al. 2012; Young et al. 2012). The site specificity of mortality enhances the need for a spatial tool useful for restoration activities.

Data were separated into categories instead of using the more typical presence/absence survey data to allow for examination of the long-term trends of A. cervicornis throughout the reef tract. Predicting the status and trends over longer time frames will provide useful information to researchers focusing on restoration efforts. This project was guided by restoration needs, as restoration projects often need more than simple presence/absence information. Resources for restoration projects are often limited, making effective and efficient planning essential (Thomson et al. 2009; McBride et al. 2010). The results of this study will be used to assist restoration planning, and the category format will be useful no matter the goals of the restoration effort. For example, a restoration study could focus on the “Transient Presence” category if the goals of the restoration effort are to enhance struggling reefs, or focus on reefs where A. cervicornis has disappeared or is declining. Currently, focusing restoration efforts on areas where A. cervicornis was once present is considered a ‘best practice’ for Caribbean Acropora restoration (Johnson et al. 2011).

While this study used all available Florida Reef Tract data, there are some coral reef ecosystem metrics that are noticeably missing, namely predation and disease layers. Acropora cervicornis populations are susceptible to predation by the fireworm, Hermodice carunculata. Predation is prevalent on both wild and restocked populations of A. cervicornis, posing
significant threats to colonies in the Florida Keys (Miller et al. 2014a). Unfortunately, the amount of quantitative data on fireworm impacts is limited and spatial data are nearly nonexistent. For future studies, it would be advantageous to analyze spatial patterns of fireworm predation, in hopes of determining the location of any regional hotspots. While representing predator dynamics continues to be problematic, distributions of benthic invertebrates are strongly controlled by ambient environmental conditions, rather than predator-prey dynamics (Robinson et al. 2011), therefore, modelling the preferred habitat without such predator data is more realistic.

Just as problematic as predation, if not more, is the impact disease has on current A. cervicornis populations. The spread of syndromes known as white-band disease (WBD) is widely considered one of the most significant disturbances affecting Acropora spp. populations (Aronson and Precht 2001). While most literature describes any disease impacting A. cervicornis populations as WBD, there are multiple varieties of disease, all of which are difficult to diagnose in the field (Miller et al. 2014b). Confounded by the struggles to accurately and uniformly identify various diseases that impact A. cervicornis, the availability of accurate spatial data for disease is sparse. Additionally, devastating disease outbreaks have been characterized as intermittent in wild and restored populations, making it even more difficult to spatially model and predict areas of increased disease occurrence. Previous studies on the prevalence and occurrence of various Acropora spp. diseases have drawn attention to the ongoing difficulty of spatially predicting disease occurrence (Miller et al. 2014b).

In conclusion, this analysis confirms that there is no shortage of spatial data available for the creation of an A. cervicornis SDM. The aforementioned data can also be used for prediction of distributions of other shallow water hermatypic coral species throughout the Florida Reef
Tract. To ensure data quality, all previously mentioned processing steps should be followed for any new data layers considered.

3.6 Figures

Figure 3.5 Extent of study boundary. All survey and field data were clipped to this extent.
Figure 3.6: The seven regions of the Florida Reef Tract used in this study.
Figure 3.7 Reef type polygons located within a single grid cell used to determine dominant reef type.
Figure 3.8 Raster representation of the reef type data layer.
Figure 3.9 Example of non-reef dominated grid cell that was classified as patch reef.
Figure 3.10 Final distance from shore raster layer
Figure 3.11 Depth layer used in this study a) Point file used as input in EBK depth model b) resulting 1 km² raster grid after interpolation.
Figure 3.12: Final slope raster layer.
Figure 3.13 Final SBT high and low raster layers. Values represent number of weeks above 30°C or below 20°C from 2008 – 2011.

Figure 3.14: Final light380 and light488 raster layers. Values represent percent light reaching the bottom.
Figure 3.15: Results of simple linear regressions testing for multicollinearity of the predictor data.
Figure 3.16 Out-of-bag (OOB) error estimates on different subsets of training data. Each group was run 10x, with a separate stratified random sample taken before each run.
Figure 3.17 Spatial distribution of the training and validation *A. cervicornis* data sets.
Figure 3.18: Coverage of final point file with all predictors assigned.

3.7 References


Glynn PW, Stewart RH (1973) Distribution of coral reefs in Pearl Islands (Gulf of Panama) in relation to thermal conditions. Limnol Oceanogr 18:367-379

85


87

4 Model Comparison

4.1 Introduction

Spatial patterns of species distributions are of basic interest to ecologists, and understanding the environmental factors that influence these patterns is important for management of species facing threats such as climate change, habitat destruction and anthropogenic stress (Robinson et al. 2011). Species Distribution Models (SDMs) are tools that relate field observations to environmental predictor variables using either statistical or theoretical relationships (Guisan and Zimmermann 2000). Species distribution models are especially helpful when field survey data are sparse and information across the entire landscape is desired. The modeled relationship created by the SDM can be used to predict across unsampled space, where \textit{in situ} data are unavailable.

The use of SDMs has been growing in recent years, with applications in fields from conservation biology to climate change research (Guisan and Zimmermann 2000). Species distribution modeling has been widely used in terrestrial studies for decades, but applications of the SDM approach to marine species and environments are rare by comparison (Robinson et al. 2011). Even rarer is species distribution modeling applied to shallow-water stony coral species, with very limited studies just starting to explore the possibility of using this technique for predictive mapping (Franklin et al. 2013). The majority of marine SDM research has focused on deep-water coral (Bryan and Metaxas 2007; Davies et al. 2008; Tittensor et al. 2009; Woodby et al. 2009; Howell et al. 2011), fish (Leathwick et al. 2008; Valavanis et al. 2008; Maxwell et al. 2008;
2009), gastropods (Mieszkowska et al. 2013), or coarse-scale global prediction of shallow-reef locations (Couce et al. 2012). However, the abundance of shallow-water stony coral monitoring data, combined with an ever-increasing amount of environmental data in coral reef environments, perfectly suits shallow-water coral taxa to be assessed using SDMs.

The staghorn (*Acropora cervicornis*) and elkhorn (*A. palmata*) corals were the first Caribbean corals listed as threatened on the Endangered Species List in 2006 (NMFS 2006). Given their threatened status, restoration efforts in Florida have focused almost exclusively on *A. cervicornis* outplanting projects. Most of these efforts have relied solely on local knowledge to determine appropriate outplanting locations. The need for this study arose from the realization that no spatial prioritization tool existed to identify suitable restoration locations. In fact, both the NOAA *Acropora* Biological Review Team (*Acropora* Biological Review Team 2005) and the National Marine Fisheries Service (NMFS 2015) identified the need for quantitative predictions based on modeling analysis for conservation. Therefore, the purpose of this study is to provide a tool to help restoration experts and managers determine areas to focus outplanting efforts.

Species distribution models usually have one of two objectives: to develop the best single predictive model or to achieve the best explanations of the causal relationship between the response and the proposed predictors. These two objectives require very different approaches to model development and, therefore, the goal of the model must be determined early on in the modeling process (Mac Nally 2000). The objective of this study is to select a SDM modeling approach best suited to predicting sites with potential for successful restoration of *A. cervicornis* populations along the Florida Reef Tract.
4.1.1 The SDM approach

There are many types of SDMs and they all differ in how they handle data, the assumptions they make about the data, and their ease of use and applicability to certain types of data. Classification and Regression Trees (CART) are one family of SDMs that utilize a “machine-learning” method. Classification and Regression Tree models operate by recursively partitioning data through decision trees based on predictor variables (Breiman et al. 1984). These models attempt to partition the data into the most homogeneous nodes based on a series of decisions. The result is a decision tree that can be used to predict the response of new cases (Breiman et al. 1984; Vayssieres et al. 2000).

The most basic type of CART is a single decision tree. The single decision tree shown in Figure 4.1 uses three environmental conditions—depth, wave exposure and light availability—to partition Acropora palmata presence/absence data into homogeneous groups. As data travel down the tree, they are separated into smaller and smaller groups, until a homogeneous group is created and no more partitions are necessary. In this example, the decision tree has created a rule stating that A. palmata is present in depths between 5-10 m with high light availability or in depths less than 5 m with high wave exposure. This decision tree can then be applied to unsampled areas using the same environmental conditions to predict whether A. palmata is likely to be present or absent.
The previous example is very simple and ecological data are rarely partitioned so smoothly. Often, observations fall into multiple categories or the splits are much less distinct. For example, the light availability predictor was split at either high or low in Figure 4.1. A key question that may be asked is: what is the cutoff value to separate high versus low? Likely, there were several values where light availability levels could have been split. Due to intricacies such as this, single classification trees have been found to be weak predictors (Svetnik et al. 2003). These models often work well at describing the training data used to build the model, but often fall short at predicting to new locations.

While single decision trees have shortcomings, they remain a common tool for species distribution modeling. Decision trees are intuitive and conceptually easy to visualize. Additionally, decision trees are capable of handling many types of predictor variables (numeric, binary, categorical, etc.) with little to no requirement for removing irrelevant predictors from the model (Elith et al. 2008). Guided by the popularity of CART models, several versions of ensemble machine-learning methods have been developed that have been found to improve the
predictive performance of single decision trees in ecological modeling (Breiman 2001; Svetnik et al. 2003; Hamza and Larocque 2005). A main goal of ecological and environmental analysis is commonly prediction, and the ability to accurately predict complex responses has improved with the advancement of machine-learning methods and high-order computation analyses. Examples of learning algorithms include neural networks, support-vector machines, bagging, Random Forests, and boosting (De’ath 2007). The similar goal in all of these classification methods is to create a classifier that predicts with high accuracy on test data (sometimes referred to as validation data, as in this study). This is achieved by confirming that the classifier first fits the training data well without being too complex (Mukherjee and Schapire 2013).

For this study, ensemble-tree methods were selected to model the spatial categorical trends of A. cervicornis. Ensemble tree methods are flexible and can be used for classification-type analysis; they are non-parametric, able to handle and ignore irrelevant predictors, and are suited for multinomial response data. The ensemble methods used in this study are Random Forest (RF) and boosted classification trees. All trees used in this study are stochastic, meaning they contain a random or probabilistic component. This stochasticity improves the predictive performance, as a random subset of the data is used to fit each new tree. As a result, the final model will be subtly different each time it is run (Elith et al. 2008).

4.1.2 Statistical Model Description

Two statistical models were evaluated in this study, RF and boosted classification trees. The methods were chosen based on their ability to handle multinomial response data and their non-parametric nature. Both methods have also been shown to perform well in ecological prediction studies.
4.1.2.1 Random Forest Models

Random Forest is an ensemble-learning technique that generates many classification trees that are aggregated to compute a final classification (Breiman et al. 1984; Breiman 2001). Random Forests have been found to be among the best predictors (Svetnik et al. 2003). Unlike standard decision trees, which consider all predictor variables at each split, a random forest only considers a subset of predictors randomly chosen at each node (Breiman 2001; Liaw and Wiener 2002). Each individual classification tree built in the forest is unique as the training data are resampled without replacement (called bootstrap sampling), and the predictor datasets to consider at each split of the tree are randomly changed. One desirable feature of the RF is its built-in estimation of predictor accuracy due to the bootstrap sampling method, called the out-of-bag (OOB) error estimation. Additionally, RFs only have two parameters to set when tuning the model: the number of variables to consider at each node \( m_{\text{try}} \) and the number of trees in the forest \( n_{\text{tree}} \) (Liaw and Wiener 2002). Table 4.1 outlines the steps for the RF algorithm as well as for the error estimation that takes place within the RF.

During model development, two parameters need to be optimized to decrease the error, \( m_{\text{try}} \) and \( n_{\text{tree}} \). The \( n_{\text{tree}} \) parameter needs to be set sufficiently high to allow for convergence of the OOB error rate. The \( m_{\text{try}} \) parameter has an influence on both the strength of the individual trees in the forest, as well as the correlation between trees in the forest. Reducing \( m_{\text{try}} \) reduces both strength and correlation. Reduction of strength results in increased error, while reduction of correlation results in decreased error. Therefore, \( m_{\text{try}} \) must be optimized to minimize the error (Peters et al. 2007).
Table 4.4: Description of Random Forest algorithm steps

<table>
<thead>
<tr>
<th>Random Forest Algorithm Steps:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Draw ( n_{\text{tree}} ) bootstrap samples from the training data.</td>
</tr>
<tr>
<td>2. For each bootstrap sample, grow a classification or regression tree to its full size. Instead of considering all possible predictors at each node of the tree, choose the best split among a random sample of size ( m_{\text{try}} ) of the predictors and choose the best split from only those predictors.</td>
</tr>
<tr>
<td>3. Use the fully grown trees to predict to new data by aggregating the predictions of all trees combined (majority vote becomes the prediction).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Forest Error Estimation Steps:</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. For each tree (and subsequently each bootstrap sample), predict the data not in the bootstrap sample using the tree grown with the bootstrap sample. The data not in the bootstrap sample are called the “out-of-bag” or OOB data (Breiman 2001).</td>
</tr>
<tr>
<td>2. Aggregate the OOB predictions to calculate the error rate. This is called the OOB estimate of error rate and has been shown to be both an accurate estimate of the generalization error and similar to the more traditional model error estimate procedure of using cross-validation (Bylander 2002; Liaw and Wiener 2002).</td>
</tr>
</tbody>
</table>

In addition to the above mentioned parameters, the number of predictor variables to include must also be set during model development. In general, the most simplified model is desired, with as few predictor variables as possible to appropriately model the response. Simplified models are desired to avoid mathematical artifacts that result from including an excess number of predictors. In fact, it is always possible to model the data better by including more predictor terms. When the number of predictor terms approaches the number of response observations, a near perfect fit is possible and represents a mathematical artifact rather than a modeling success (Mac Nally 2000). Additionally, the confidence of the model is lessened when more predictor terms are included in the model as prediction error increases (Breiman 1995). While these suggestions were originally formulated based on the more traditional linear regression models, they also apply to CART models. While CART models are less sensitive to including irrelevant predictors than other methods [linear regression, Generalized Additive Models (GAMs), etc.], reducing the number of predictors in CART models is still desirable for predictive purposes.
Once the model development is complete, RF models provide additional useful products for ecological studies. The most widely used of these products are variable-importance plots. Random forest models estimate the importance of predictor variables by analyzing how much the OOB error increases when OOB data for that variable are permuted, while all other variables are left unchanged. This is a complicated task, as the importance of any single variable may be due to its interaction with one or multiple other variables (Breiman 2001; Liaw and Wiener 2002).

The RF model provides two measures of variable importance: the mean decrease in accuracy and mean decrease in Gini coefficient. The mean decrease in accuracy (referred to as just mean decrease accuracy in the model) is computed at the same time as the OOB error. The RF calculates the decrease in the accuracy of the model when a single variable is excluded. The greater the decrease in accuracy of the model with the exclusion of a particular variable, the greater the importance of that variable. The Gini coefficient is a measure of homogeneity of the final RF. The mean decrease Gini coefficient calculated by the RF model represents how each of the predictor variables contributes to this homogeneity. This is essentially a measure of how well each predictor classifies the data. When the goal of the model is prediction, it is recommended to use mean-decrease accuracy measures to determine which variables to include (Breiman 2001).

4.1.2.2 Boosted Classification Tree Models

Boosted classification trees are another group of ensemble learning techniques that work by creating a single classifier that is extremely accurate in classifying the training data. Predictions from a set of weak classifiers (such as single trees) are combined to create a strong classifier with low prediction error (De'ath 2007). Boosting uses the performance of previous trees to inform trees being built later in the process.
The boosting process works by growing a sequence of trees, with successive trees grown based on re-weighted versions of the training data. At each step when a new tree is grown, the training data are classified based on the previously grown trees. The results of the classification of the training data at this step are used when fitting the next tree in the sequence. Incorrectly classified cases receive more weight than the correctly classified ones. As a result of this weighting process, the data points that are most difficult to classify receive the greatest attention, thereby increasing their chance of being classified correctly. The final classification results from the weighted majority of classifications from all trees in the sequence (De'ath 2007).

There are several versions of boosting algorithms and they vary in how they calculate the lack of fit and how they weight the data for the next iteration (Elith et al. 2008). The common goal of these different algorithms is to improve the accuracy of an ensemble-tree method by combining individual classifiers that are as precise and different as possible (Alfaro et al. 2013). The known weaknesses of classification trees can be overcome by the use of boosted trees, which are widely known to be excellent predictors and depict complex relationships in a simple manner (De'ath 2007).

Two boosting algorithms were evaluated in this study, AdaBoost.M1 (Freund and Schapire 1996) and SAMME (Zhu et al. 2009). These were chosen because of their ability to be used for multiclass classification problems. AdaBoost.M1 measures the performance of the weak classifiers using ordinary error (Freund and Schapire 1996; Mukherjee and Schapire 2013). The SAMME algorithm requires less error than random guessing on any distribution of the examples (Zhu et al. 2009; Mukherjee and Schapire 2013). Table 4.2 contains the steps for each of the boosting algorithms. The only difference between the two algorithms (AdaBoost.M1 and
SAMME) is how the alpha constant is calculated. The number of classes \((k)\) is taken into account for the SAMME algorithm and not for the AdaBoost.M1 algorithm (Alfaro et al. 2013).

Table 4.5: Description of Boosted Classification Tree algorithm steps

**AdaBoost.M1 Algorithm Steps:**

1. Given a set of training data \(T_n = [(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_n, y_n)]\), assign the default weight \(w_{0(i)} = 1/n, i = 1, 2, \ldots, n\) to each observation.
2. Repeat the following process for \(b = 1, 2, \ldots, B\) where \(B\) is the number of iterations set by the user
   a. Fit the classifier \(C_b(x_i) = (1, 2, \ldots, k)\) using weights \(w_b(i)\) on \(T_b\) which is a subset of \(T_n\).
   b. Compute the error \((e_b)\) as: \(e_b = \sum_{i=1}^{n} w_b(i) I(C_b(x_i) \neq y_i)\) then the alpha coefficient: \(\alpha_b = \frac{1}{2} \ln \left( \frac{1-e_b}{e_b} \right)\)
   c. Update the weights using the formula: \(w_{b+1}(i) = w_b(i) \exp(\alpha_b I(C_b(x_i) \neq y_i))\) then normalize the weights
3. Output the final classifier: \(C_f(x_i) = \arg\max_{j \in Y} \sum_{b=1}^{B} \alpha_b I(C_b(x_i) = j)\)

**SAMME Algorithm Steps:**

1. Start with \(w_b(i) = \frac{1}{n}, i = 1, 2, \ldots, n\)
2. Repeat for \(b = 1, 2, \ldots, B\)
   a. Fit the classifier \(C_b(x_i) = (1, 2, \ldots, k)\) using weights \(w_b(i)\) on \(T_b\) which is a subset of \(T_n\).
   b. Compute the error \((e_b)\) as: \(e_b = \sum_{i=1}^{n} w_b(i) I(C_b(x_i) \neq y_i)\) then the alpha coefficient: \(\alpha_b = \ln \left( \frac{1-e_b}{e_b} \right) + \ln(k - 1)\)
   c. Update the weights using the formula: \(w_{b+1}(i) = w_b(i) \exp(\alpha_b I(C_b(x_i) \neq y_i))\) then normalize the weights
3. Output the final classifier: \(C_f(x_i) = \arg\max_{j \in Y} \sum_{b=1}^{B} \alpha_b I(C_b(x_i) = j)\)

Like the RF algorithm, the boosted classification tree algorithms also provide information on variable importance, but in a slightly different fashion. Instead of mean decrease accuracy and mean decrease Gini coefficient, the boosting package simply exports an importance measure, which is the relative importance of each variable in the classification. This measure takes into account the Gini index of a variable in a tree, as well as the weight of that particular tree. The statistical package used to implement these algorithms combines these measures to produce a final measure of importance for each variable.
4.1.3 Model Accuracy Measures

Random Forest models are unique in that they provide a built-in measure of predictor accuracy from the OOB error estimate. The boosted classification trees do not provide this estimation, however other methods are used to estimate the accuracy of the final predictor. These measures of accuracy are based on a confusion or error matrix. Confusion matrices cross-tabulate the observed and predicted responses and can be used to calculate a variety of accuracy measures (Fielding and Bell 1997). One of the most commonly used measures of accuracy from the confusion matrix is percent correctly classified (PCC), which is simply the percentage of points where the observation and prediction agree. Another commonly used diagnostic measure of prediction accuracy is the kappa statistic. Kappa measures model accuracy by correcting for accuracy expected to occur by chance (Cohen 1960). A combination of these measures of model accuracy were used for model selection in this study.

In addition to the accuracy statistics summarized above, the margins are also calculated for each of the models. The margin is related to the certainty of classification. It is calculated as the difference between the support of the correct class and the maximum support of an incorrect class (Alfaro et al. 2013). Margins closest to one signify a highly certain classification. Negative margins represent an incorrect classification.

This study describes the approaches used to select a SDM best suited for predicting sites with potential for successful restoration of A. cervicornis populations along the Florida Reef Tract. The results of this study provide recommendations on model development, model selection and application in the coral reef ecosystem. The recommendations made in this study provide guidance for future coral reef modeling projects along the Florida Reef Tract. The final
models developed in this study are also useful for spatially prioritizing *A. cervicornis* restoration locations throughout the Florida Reef Tract.

4.2 Methods

4.2.1 Study Area

This study focused on the Florida Reef Tract (FRT), which spans from Martin County through the Dry Tortugas. The FRT is a subtropical barrier reef system that contains a combination of patch, linear and aggregate reefs (Shinn et al. 1989; Precht and Miller 2007). The reef tract occurs near the latitudinal limits of subtropical waters and, thus, experiences an abundance of natural stressors. Conditions along the reef tract include variable temperatures reaching extreme highs and lows, which are not generally favorable for reef development. Nevertheless, corals continue to occur in this region (Kruczynski and McManus 2002).

4.2.2 Data Sets

This study used the *A. cervicornis* trend dataset described in Chapter 3 as the response data, and the nine environmental data sets described in Chapter 3 as the predictor data sets. The nine data sets included: light at 380 nm, light at 488 nm, depth (m), region, sea bottom temperature (SBT) high, SBT low, reef type, distance to shore, and slope (Table 4.3). Full details regarding the details each of these data sets are available in Chapter 3.

The lack of an independent data set for model evaluation required splitting the full *A. cervicornis* data set into two sets, one for model training and the other for model validation. The training data set comprised 90% of all data, and the validation set comprised 10% of all data, using a stratified random sample approach to ensure each set contained an appropriate number of
observations from each category. Full details regarding the training and validation data separation are described in Chapter 3. The same training and validation data sets were used for each model.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>region</td>
<td>Seven regions of the Florida Reef Tract</td>
</tr>
<tr>
<td>reeftype</td>
<td>Dominant reef type within 1 km² cell</td>
</tr>
<tr>
<td>shoredist</td>
<td>Distance from shore at 1 km increments</td>
</tr>
<tr>
<td>depth</td>
<td>Depth interpolated from LiDAR and point bathymetric data</td>
</tr>
<tr>
<td>slope</td>
<td>Slope derived from depth layer</td>
</tr>
<tr>
<td>sbhigh</td>
<td>Number of weeks above 30°C from 2008 – 2011</td>
</tr>
<tr>
<td>sblow</td>
<td>Number of weeks above 30°C from 2008 – 2011</td>
</tr>
<tr>
<td>light380</td>
<td>Percent total light at 380 nm available at depth</td>
</tr>
<tr>
<td>light488</td>
<td>Percent total light at 488 nm available at depth</td>
</tr>
</tbody>
</table>

4.2.3 Model Construction, Calibration and Resulting Parameters

4.2.3.1 Random Forest Model

For RF model development, the randomForest package (Liaw and Wiener 2002) within the statistical software R v. 3.2.1 was used (R Core Team 2015). The RF model was calibrated to optimize the \( n_{tree} \) and \( m_{try} \) parameters in order to minimize the generalization error. To account for the uneven sample size, all RF models were run with the sample size parameter set to sample 10 points from each category of the A. cervicornis response data. This ensured that the training data were not biased towards the most dominant category in the data set (category 1). As suggested by Peters et al. (2007), the \( n_{tree} \) parameter was optimized first, followed by the \( m_{try} \) parameter.

One benefit of using RF models is that they do not overfit the data (Breiman 2001). Adding more trees to the forest (increasing \( n_{tree} \)) will limit the generalization error. A series of random forest models were constructed on the training data, varying the \( n_{tree} \) parameter from 10–
10,000 on a logarithmic scale. The model was run ten times for each \( n_{\text{tree}} \) value, and the average and standard deviation of the OOB error were computed (Figure 4.2). The default value for \( m_{\text{try}} \) was used for all runs.

![Figure 4.2: Average out-of-bag (OOB) error stabilizes when more trees are added to the random forest. The OOB error \% here represents the average over 10 runs for each value, with the standard deviation.](image)

Figure 4.2 presents the OOB error as a function of the number of trees. At 1,000 trees, the OOB error reaches 35.7\% ±1.4\%. Adding more trees does not considerably decrease or increase the OOB error estimate, suggesting that using 1,000 trees is appropriate for this study. The mean OOB error reaches a minimum at 300 trees. Svetnik et al. (2003), however, suggests that the only limitation to including more trees is increased computing requirements; thus, 1,000 trees are appropriate for this study. Also, the error rate at 1,000 trees is within the standard deviation of the error rate at 300 trees. Since the only limitation is that of computing time, and computing time was not an issue with this dataset, 1000 trees were used. Additionally, including more trees in the model reduces the variance and ensures that each predictor variable has a strong chance to be included in the forest prediction process.

Based on the above findings, a random forest with 1,000 trees was constructed to determine the optimal value for the number of predictor variables to consider at each node, or
At each split down the tree, a random subset of \( m_{try} \) predictor variables is considered, and the best splitting variable among those is used to split the node. The value of \( m_{try} \) remains constant throughout the forest-growing process. This value is dependent upon the number of predictors initially included in the model and, therefore, must be chosen at the same time as when determining the number of predictor variables to include.

Based on suggestions by Zhao and Chen (2014), 10-fold cross validation (CV) using the training data was performed to determine the optimal number of predictor variables to include. The \texttt{rfcv} function in the ‘randomForest’ package in R creates a series of cross-validated random forest models and provides the error rates of classification at each step in the function (Svetnik et al. 2003). The \texttt{rfvc} function removes the least important variable(s) at each step, generates a new model and, thus, performs automated variable selections. The package validates the accuracy of the new model until a desired number of variables is reached. The \texttt{rfcv} function was set to remove one variable at a time and replicate the process 100 times to determine the optimal number of variables to include in the model. The \texttt{rfcv} function was run with a variety of \( m_{try} \) values: \( p \), \( p/2 \), \( p/4 \) and \( \sqrt{p} \), where \( p \) is the total number of variables.

The results of the \texttt{rfcv} function at various \( m_{try} \) values are shown in Figure 4.3. The most accurate model was developed when five variables were used, as five variables achieved the lowest CV error rate, for all \( m_{try} \) values. This suggests that the five most important predictors should be used in the final predictive RF model. However, models that included 1 – 8 variables performed similarly and could also be considered when constructing the final model. In addition, Figure 4.3 shows that all values of \( m_{try} \) performed similarly, therefore the default value of \( m_{try} = \sqrt{p} \) was used for the remainder of model development.
The rfcv function informs the appropriate number of variables to use, but does not reveal which predictor variables were used. The stochastic nature of the RF model needs to be considered when constructing the final model, and is most apparent when determining predictor importance, due to the random selection of predictors at each node. For this reason, the least important predictor variables may change with each iteration of the model (Zhao and Cen 2014). Two factors contribute to the randomness of the variable importance in the rfcv function. First, each time the 10-fold cross validation is run, a slightly different data set is used to train the model. Additionally, each time a RF model is run, a random selection of variables \([\sqrt{p}]\) is considered at each split. Therefore, the predictors included in the model must be manually changed to identify the best model.

A RF model was run using the parameters previously determined \([n_{\text{tree}} = 1000, m_{\text{try}} = \sqrt{p}]\) 20 times and the most important predictors, as well as their values corresponding to variable importance (mean decrease accuracy), were recorded. Figure 4.4 shows the mean decrease in accuracy of all 20 runs for each variable.
As the model is to be used for prediction, the mean decrease accuracy was used to select the variables to include in the model. Over all 20 runs, depth and SBT high were identified as the two most important predictors. These two variables also had the greatest mean decrease accuracy when averaged over the 20 runs.

The model was then simplified using backward variable selection to remove unimportant predictors to attempt to decrease the OOB error rate of the model. Guided by the results of Figures 4.3 and 4.4, two sets of variables were tested for best performance. As recommended by the rfcv results, the top five variables were tested. Initial results indicated that using five variables resulted in poor accuracy measures, so the top eight variables were also tested (Table 4.3). Each model was run 20 times with the each set of predictor variables and the accuracy measures on both the test data and the validation data were averaged (Table 4.4).

The results indicate that including eight variables produces better results than including five. The final RF model chosen included all predictor variables except for SBT low. The final RF model used the following parameters: $n_{tree} = 1000$, $m_{try} = \sqrt{p}$, and eight variables. When using eight variables, $m_{try}$ is set to two, meaning at each node of the tree, the best split among
two randomly selected predictor variables is chosen. All statistics related to the model prediction accuracy on the test data, as well as to the validation data, were recorded, including the confusion matrices on both the test and validation data sets.

<table>
<thead>
<tr>
<th>Combination 1</th>
<th>Combination 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td>Depth</td>
</tr>
<tr>
<td>Sbthigh</td>
<td>Sbthigh</td>
</tr>
<tr>
<td>Region</td>
<td>Region</td>
</tr>
<tr>
<td>Slope</td>
<td>Slope</td>
</tr>
<tr>
<td>Shoredistance</td>
<td>Shoredistance</td>
</tr>
<tr>
<td>Light380</td>
<td>Light380</td>
</tr>
<tr>
<td>Reeftype</td>
<td>Reeftype</td>
</tr>
<tr>
<td>Light488</td>
<td>Light488</td>
</tr>
</tbody>
</table>

Table 4.7: Combinations of variables tested in the RF model

<table>
<thead>
<tr>
<th></th>
<th>5 Variables</th>
<th>8 Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>OOB Error</td>
<td>38.92</td>
<td>33.07</td>
</tr>
<tr>
<td>Kappa (test)</td>
<td>0.36</td>
<td>0.46</td>
</tr>
<tr>
<td>PCC (test)</td>
<td>61.08</td>
<td>66.93</td>
</tr>
<tr>
<td>Kappa (validation)</td>
<td>0.003</td>
<td>0.14</td>
</tr>
<tr>
<td>PCC (Validation)</td>
<td>22.5</td>
<td>40.0</td>
</tr>
</tbody>
</table>

**4.2.3.2 Boosted Classification Tree Models**

For the boosted classification tree model development, the adabag package within the statistical software R 3.2.1 was used (R Core Team 2015). The two boosting algorithms, AdaBoost.M1 and SAMME, were tuned separately, using the same procedure. The boosting models have two important user-defined parameters that must be tuned for maximum performance. Similar to the RF, the number of trees must be optimized. In addition, a parameter called maxdepth needs to be set to control the size of the trees and prevent overfitting the data.

The adabag package was built to utilize Freund and Schapire’s (1996) Adaboost.M1 algorithm and, more recently, Zhu’s (2009) SAMME algorithm for boosted classification trees. The adabag package calls another package in R, named rpart, which is the basic package for
recursive partitioning for classification trees (Therneau et al. 2015). To optimize models created using the adabag package, certain parameters for both the adabag package and the rpart package must be optimized.

The primary rpart parameter that must be optimized is maxdepth. The optimal maxdepth of the trees is dependent upon the data used, not the specific model algorithm and, therefore, should be tuned prior to changing the model algorithms. Maxdepth determines the size of the final classification tree and controls the maximum depth of any node of the final tree, with the root node counted as depth 0. The data should be analyzed using a variety of maxdepth values to create sufficiently large trees without over fitting the data.

To optimize maxdepth, the tune.rpart function in the e1071 package in R (Meyer et al. 2015) was used. Tune.rpart was set to sequentially try a maxdepth of values 1-10 at each run and record the error associated with each value of maxdepth. Tune.rpart was repeated 20 times and the results were averaged (Figure 4.5). Tune.rpart uses cross-validation to randomize the data sets before constructing the classification trees. The results of Figure 4.5 suggest that a max depth of five results in the lowest mean classification error for the data. A value of five for maxdepth was used for all future steps of the boosted classification tree development, for both algorithms tested.

4.2.3.2.1 AdaBoost.M1 Model

The last parameters that must be optimized for boosted classification trees are dependent upon the boosting algorithm used. Therefore, the AdaBoost.M1 algorithm was tuned separately from the SAMME algorithm. First, the mfinal parameter, which corresponds to the total number of trees built during the boosting process, was optimized. Using the adabag package in R, a
boosted classification tree was constructed using the AdaBoost.M1 algorithm and mfinal set to 1,000. The evolution of the error rate was then plotted against the number of trees (Figure 4.6). As the number of trees mostly influences the computing time, instead of accuracy of a model, mfinal was set to 200 trees.

![Graph showing error rate against maxdepth](image)

Figure 4.23: Results of tune.rpart on max depth. Average of all 20 runs is in bold black line.

Using the same procedure as in the RF model development, the number of predictor variables to include was also tuned. However, the adabag package does not have a function similar to rfcv, so the process was completed manually. The above parameters were set (mfinal = 200 and maxdepth = 5) and a series of 20 boosted classification trees were constructed, noting the variable importance at each run. The average variable importance was computed over all 20 runs (Figure 4.7). Depth was consistently the most important parameter.

To investigate the importance of variables further, 20 runs of several combinations of variables were tested. Combinations tested were based on the observed groupings from Figure 4.7 (Table 4.5). For each combination, the kappa values and the PCC were averaged for both the
training data sets and the validation data sets (Table 4.6). Due to the similarity of statistics on the training data, the best model was selected based on the statistics on the validation data. The combination of variables selected as the best was Combination 3: depth, light 380, light 488, SBT high, slope and region. This combination maximized the PCC on the validation data, without sacrificing much accuracy on the training data.

![Ensemble error vs number of trees](image)

**Figure 4.24:** AdaBoost.M1 algorithm error rate vs number of trees

<table>
<thead>
<tr>
<th>Combination 1</th>
<th>Combination 2</th>
<th>Combination 3</th>
<th>Combination 4</th>
<th>Combination 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td>Depth</td>
<td>Depth</td>
<td>Depth</td>
<td>Depth</td>
</tr>
<tr>
<td>Light 488</td>
<td>Light 488</td>
<td>Light 488</td>
<td>Light 488</td>
<td>Light 488</td>
</tr>
<tr>
<td>Light 380</td>
<td>Light 380</td>
<td>Light 380</td>
<td>Light 380</td>
<td>Light 380</td>
</tr>
<tr>
<td>SBT High</td>
<td>SBT High</td>
<td>SBT High</td>
<td>SBT High</td>
<td>SBT High</td>
</tr>
<tr>
<td>Slope</td>
<td>Slope</td>
<td>Slope</td>
<td>Slope</td>
<td>Slope</td>
</tr>
<tr>
<td>Region</td>
<td>Region</td>
<td>Region</td>
<td>Region</td>
<td>Region</td>
</tr>
<tr>
<td>Shore Distance</td>
<td>Shore Distance</td>
<td>Shore Distance</td>
<td>Shore Distance</td>
<td>Shore Distance</td>
</tr>
<tr>
<td>Reef Type</td>
<td>Reef Type</td>
<td>Reef Type</td>
<td>Reef Type</td>
<td>Reef Type</td>
</tr>
<tr>
<td>SBT Low</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.10: Results of the variable combinations tried with the AdaBoost.M1 algorithm. “Combo” = combination, as shown in Table 4.5

<table>
<thead>
<tr>
<th></th>
<th>Combo 1</th>
<th>Combo 2</th>
<th>Combo 3</th>
<th>Combo 4</th>
<th>Combo 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kappa (test)</strong></td>
<td>0.978</td>
<td>0.975</td>
<td>0.967</td>
<td>0.971</td>
<td>0.862</td>
</tr>
<tr>
<td><strong>PCC (test)</strong></td>
<td>98.75</td>
<td>98.58</td>
<td>98.12</td>
<td>98.35</td>
<td>92.22</td>
</tr>
<tr>
<td><strong>Kappa (validation)</strong></td>
<td>0.273</td>
<td>0.283</td>
<td><strong>0.463</strong></td>
<td>0.238</td>
<td>0.117</td>
</tr>
<tr>
<td><strong>PCC (validation)</strong></td>
<td>62.5</td>
<td>63.13</td>
<td><strong>68.13</strong></td>
<td>50</td>
<td>42.5</td>
</tr>
</tbody>
</table>

4.2.3.2 SAMME Model

Using the adabag package in R, a boosted classification tree was constructed using the SAMME algorithm and mfinal set to 1,000. The evolution of the error rate was then plotted against the number of trees (Figure 4.8). Based on the results, mfinal was again set to 200 trees.

The same procedure was used as with the AdaBoost.M1 algorithm to optimize the number of parameters to include in the SAMME model. The model was run 20 times and all variable importance statistics were recorded and averaged (Figure 4.9). Depth and light 488 were the most important predictor variables, each being the most important variable 50% of the time (10 runs each). The combinations of variables tested were again based on the results from Figure 4.9 and are the same as from the AdaBoost.M1 algorithm (Table 4.5). For each combination, the
kappa values and the PCC were averaged for both the training data sets and the validation data sets across all 20 runs (Table 4.7). Again, because of the similarity of the statistics on the training data, the best model was selected based on the statistics on the validation data. The combination of variables selected as best was Combination 1, which included all nine predictor variables. This combination maximized the kappa and PCC on both the training and validation data.

![Ensemble error vs number of trees](image)

Figure 4.26: SAMME algorithm error rate vs number of trees

Table 4.11: Results of the variable combinations tried with the SAMME algorithm. “Combo” = combination, as shown in Table 4.5

<table>
<thead>
<tr>
<th></th>
<th>Combo 1</th>
<th>Combo 2</th>
<th>Combo 3</th>
<th>Combo 4</th>
<th>Combo 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kappa (test)</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.980</td>
<td>0.939</td>
</tr>
<tr>
<td>PCC (test)</td>
<td>98.86</td>
<td>98.86</td>
<td>98.86</td>
<td>98.86</td>
<td>96.53</td>
</tr>
<tr>
<td>Kappa (validation)</td>
<td>0.266</td>
<td>0.219</td>
<td>0.146</td>
<td>0.149</td>
<td>-0.195</td>
</tr>
<tr>
<td>PCC (validation)</td>
<td>60.6</td>
<td>57.5</td>
<td>52.5</td>
<td>42.5</td>
<td>36.3</td>
</tr>
</tbody>
</table>
4.3 Results

4.3.1 Random Forest Model

The final RF model chosen consisted of 1,000 trees, considered 2 variables at each split of the trees and included eight predictor variables. The only predictor variable excluded from the model was SBT low. Of the predictor variables used, depth and SBT high were the most important (Figure 4.10). The final OOB error estimate of this model was 33%, with class errors summarized in Table 4.9.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Class error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36</td>
<td>9</td>
<td>6</td>
<td>29.4%</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>16</td>
<td>5</td>
<td>33.3%</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>46.2%</td>
</tr>
</tbody>
</table>

Table 4.12: Confusion matrix from the final RF model using the training data
A unique margin was calculated for each observation, and all observations were compiled to represent the cumulative distribution of the margin on the training data (Figure 4.11). The final RF model had an average margin of 0.19, and a maximum margin of 0.93. About 34% of the training data falls below the zero mark, meaning that 34% of the data were classified incorrectly.

4.3.2 AdaBoost.M1 Model

The final boosted classification tree using the AdaBoost.M1 algorithm consisted of 200 trees, had a maxdepth of five for each tree and included six predictor variables. The final model did not use shore distance, reef type or SBT low as predictor variables. Of the six variables used
in the model, depth was again the most important predictor (Figure 4.12). The final model correctly classified 98% of the training data, with class errors summarized in Table 4.9.

The final AdaBoost.M1 model had an average margin of 0.40 and a maximum margin of 0.91. Only about 1% of the training data falls below the zero mark, meaning that 1% or less of the training data were incorrectly classified (Figure 4.13).

![Margin cumulative distribution](image)

Figure 4.29: Margin for random forest model

<table>
<thead>
<tr>
<th>Class Error</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>51</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>12</td>
</tr>
</tbody>
</table>

### 4.3.3 SAMME Model

The final boosted classification tree using the SAMME algorithm consisted of 200 trees, had a maxdepth of five for each tree and included all nine predictor variables. The light 488 layer was considered the most important predictor variable in the final model (Figure 4.14). The final model correctly classified 99% of the training data, with class errors summarized in Table 4.10.
The final SAMME model had an average margin of 0.19 and a maximum margin of 0.45. Only about 1% of the training data falls below the zero mark, meaning that 1% or less of the training data were incorrectly classified (Figure 4.15).
Figure 4.32: Variable importance for the SAMME model. light488: percent of surface light at 488 nm reaching the benthos, depthm: bottom depth in meters, light380: percent of surface light at 380 nm reaching the benthos, sbthigh: number of weeks above 30º C from 2008 – 2011, region: region of the FRT, slope: slope of the seafloor at 1km² scale, shoredist: distance from shore at 1 km increments, reeftype: dominant reef type within 1 km² grid cell, sbtlow: number of weeks below 20º C from 2008 – 2011.

Table 4.1: Confusion matrix from the final SAMME model using the training data

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Class Error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong></td>
<td>51</td>
<td>0</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>2</strong></td>
<td>0</td>
<td>24</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>3</strong></td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>7.7%</td>
</tr>
</tbody>
</table>

Accuracy statistics for the final versions of all three models are summarized in Table 4.11. The AdaBoost.M1 and the SAMME algorithms performed similarly in regards to percent correctly classified (PCC) on both the training and the validation data. The AdaBoost.M1 algorithm out-performed the SAMME algorithm in terms of both the kappa statistic and the average margin. The final models were used to predict to the full reef tract, and these results are shown in Figure 4.16. The resulting maps are meant to be used in digital format and are therefore difficult to interpret in print. The summary of cells predicted in each category is represented in Table 4.12 for comparative purposes.
Figure 4.33: Margin for the SAMME model

Table 4.15: Error statistics for all final models

<table>
<thead>
<tr>
<th>Model</th>
<th>UW kappa (t)</th>
<th>PCC (t)</th>
<th>UW kappa (v)</th>
<th>PCC (v)</th>
<th>Average Margins (t)</th>
<th>Average Margins (v)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>0.46</td>
<td>67.05</td>
<td>0.11</td>
<td>37.5</td>
<td>0.19</td>
<td>-0.01</td>
</tr>
<tr>
<td>AdaBoost.M1</td>
<td><strong>0.98</strong></td>
<td><strong>98.86</strong></td>
<td>0.38</td>
<td>62.5</td>
<td><strong>0.40</strong></td>
<td>0.06</td>
</tr>
<tr>
<td>SAMME</td>
<td><strong>0.98</strong></td>
<td><strong>98.86</strong></td>
<td>0.27</td>
<td>62.5</td>
<td>0.19</td>
<td>0.04</td>
</tr>
</tbody>
</table>

4.4 Discussion

The most common algorithms used in species distribution modeling are parametric, linear regression-type models. These models make a variety of assumptions about the data, and are only suited for a specific type of response data. For example, many algorithms assume a species’ response follows some functional form, such as the typical bell-shaped curve. Species’ responses to environmental gradients rarely follow such functional forms therefore, the assumptions of these models are often violated (Austin and Smith 1989). Non-parametric models, such as the models tested in this study, make no such assumptions, and are therefore more suited to model responses typically observed in the environment (Vayssieres et al. 2000).
Table 4.16: Number of cells predicted for each category for each final model

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Category 1</th>
<th>Category 2</th>
<th>Category 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Forest</td>
<td>1219</td>
<td>783</td>
<td>217</td>
</tr>
<tr>
<td>AdaBoost.M1</td>
<td>1208</td>
<td>727</td>
<td>284</td>
</tr>
<tr>
<td>SAMME</td>
<td>1346</td>
<td>573</td>
<td>300</td>
</tr>
</tbody>
</table>

Of the three model algorithms tested, the two boosted classification trees performed better than the RF at classifying the *A. cervicornis* response data on both training and validation data sets. The two boosted classification tree algorithms performed similarly, but the AdaBoost.M1 algorithm slightly out-performed the SAMME algorithm. A previous study by Pittman et al. (2009) found boosted classification trees to be an appropriate statistical technique to model the complex nonlinear relationships often found in coral reef ecosystems. These results were further confirmed in this study.

It is important to note that the small sample size in this study likely had an impact on the error estimates, as when sample sizes are small (n < 1,000), the small size of the training and validation data sets can increase the prediction error (De'ath 2007). The small size of the validation data set in this study likely played a role in the relatively high prediction errors observed. In addition, the error rates are also likely influenced by the categorical nature of the data. Attempting to predict to three categories, rather than the typical presence/absence categories, is much more difficult, and therefore is subject to higher error rates.

The two boosted classification trees performed exceptionally well on the training data, with 99% of data correctly classified. The 1% classification errors of the training data compare well to the classification errors recorded on data sets that were used to initially develop the boosted classification script used for this study. One of the most well-known datasets in R is referred to as the iris data set. This data set is often used to illustrate the use of various models in R and serves as a guideline for model performance. The authors of the adabag package in R
constructed a variety of boosted classification trees on the iris data set, with error rates between 4 – 5% (Alfaro et al. 2013). This comparison confirms the strong performance of the boosted classification trees on the A. cervicornis trend data used in this study.

A variety of other algorithms not discussed were also tested during the model building process. Using the same methods as discussed above, bagged classification trees were calibrated. Due to an unknown error, the bagged classification trees resulted in predictions of only category 1 of the response data. This error was not investigated further, as random forest models with the \( m_{tr} \) parameter set to \( p \) (the total number of variables) is the same as a bagged classification tree. The random forest models with \( m_{tr} \) set to \( p \) did not perform well in comparison, therefore the analysis of bagged classification trees was dropped from this study. In addition, it has been shown empirically that both random forest and boosted classification trees outperform bagged classification trees, which verifies the results of the bagged classification trees explored during model development in this study (Svetnik et al. 2003).

One of the most important steps in model development and calibration is identification of appropriate predictor variables. These variables should reasonably be assumed to influence the response variable (Flack and Chang 1987), thus ecological knowledge should be used when identifying appropriate predictor variables. Intelligent selection of predictor variables from the start of model development should never be substituted with statistical tampering to find the best model. Some researchers include all available data layers and hope that the unimportant variables fall out during the model selection process. This method is not recommended and it has been found that the variable selection process is improved if it is based upon existing knowledge and theory regarding the response and predictor variable relationship (Mac Nally 2000).
Figure 4.34: Map results of full predictions. a) Random forest, b) AdaBoost.M1, c) SAMME
As described in Chapter 3, the environmental predictor layers used in this study were selected based on expert opinion in a series of stakeholder workshops. This initiative helped to assist in the final variable selection process during model calibration. The ecological relevance of the predictor variables used are confirmed by the inclusion of all nine predictor variables in the final SAMME model, which performed well in comparison to the other models evaluated in this study.

A major accomplishment of the current study is the representation of sub-surface environmental variables as model predictors and the confirmation that these layers are usable for species distribution modeling in the coral reef environment. In the past, researchers attempting statistical predictive modeling have struggled with the quality and timeliness of bathymetric environmental predictors relevant to benthic species such as coral and seagrass (Kelly et al. 2001). The recent improvements in availability of sub-surface data for light availability and temperature made this work possible (e.g., Barnes et al. 2013). In fact, the high temperature layer used in this study was identified as among the most important predictor variables in all three models.

The results of the variable selection process from this study identified several environmental predictor variables that were consistently important in the prediction process. These layers, specifically depth and SBT high, provide valuable information for future studies aiming to predict coral species-response data. Prior to this study, high spatial resolution raster depth information was not available across the entire study area. The methods discussed in Chapter 3, which resulted in a continuous raster coverage of depth, albeit at a coarse spatial resolution, provided valuable information to be used in the modeling process. Future studies
should aim to continue to improve data layers such as depth to optimize modeling results and to assist in improving the resolution of the model results.

While the majority of predictive modeling history is rooted in terrestrial ecosystems, this analysis shows that these techniques are useful tools in marine ecosystems as well. Accurate maps predicting coral trends provide valuable baseline data, especially for unsurveyed regions (Pittman et al. 2009). These models can be a valuable tool for the management of many marine ecosystems, not just coral reefs (Kelly et al. 2001). The use of these model results in combination with the visualizations capabilities of GIS software make it possible for managers to analyze environments on a landscape-scale (Remillard and Welch 1993). One of the greatest research needs along the Florida reef tract is for studies that take a holistic approach at analyzing the reef tract, rather than on a reef-by-reef or region-by-region basis. This study attempts to provide a reef-tract-wide prediction of *A. cervicornis* trend data to provide a baseline for future reef-tract-wide studies.

4.5 References


Howell KL, Holt R, Endrino IP, Stewart H (2011) When the species is also a habitat: Comparing the predictively modelled distributions of Lophelia pertusa and the reef habitat it forms. Biol Conserv 144:2656-2665


5 Final Model Application

5.1 Abstract

In the face of the current ecosystem-wide coral reef decline, large-scale restoration efforts have been increasing throughout the Florida Reef Tract. Currently, there is little to no information regarding the spatial prioritization for restoration efforts. This study provides the first step at statistically informing site selection for large scale restoration efforts that are currently taking place in Florida. Nine environmental data layers were used to model the staghorn coral, *Acropora cervicornis*, presence/absence trend data using a non-parametric boosted classification tree. The most important environmental parameters identified by the model, depth and light availability, were further analyzed. It was found that reef areas predicted to not have had *A. cervicornis* present from 1996-2013 were deeper, and had lower light availability and greater variance than areas predicted to have had continuous or transient *A. cervicornis* presence in this time frame. Model predictions were also compared to current restoration efforts using a database of *A. cervicornis* restoration locations provided by The Nature Conservancy. Many restoration locations occurred in areas predicted to not have had *A. cervicornis* since at least 1996, but were located in areas of similar depth and light availability as areas predicted to have had continuous *A. cervicornis* presence since 1996. The results of this study provide the first step at deriving an ecologically-guided approach to spatial prioritization of restoration efforts.
5.2 Introduction

Previous research has estimated that $31 \times 10^3 \text{ km}^2$ of shallow-water inshore habitat around Florida has the potential to support coral reef ecosystems (Rohmann et al. 2005). Unfortunately, this estimate far exceeds the reality of area encompassed by living coral reefs in Florida. Over the past several decades, the coral reefs of the Florida Reef Tract (FRT) have experienced unprecedented declines in both species richness and coral cover. As best described by the Florida Fish and Wildlife Research Institute’s Coral Reef Evaluation and Monitoring Project (CREMP), there has been an overall net loss in species richness in the Florida Keys since the program’s inception in 1996 (Callahan et al. 2007).

Not only has there been a loss in species richness, but overall coral cover has declined throughout the Caribbean since the 1970s (Gardner et al. 2003). While nearly all of the stony coral species along the Florida Reef Tract have exhibited decline, the most notable has occurred in the once-dominant acroporid species, staghorn (*Acropora cervicornis*) and elkhorn (*A. palmata*) corals. Decline of these species has been reported throughout the western Atlantic and Caribbean over the last 30+ years (Jaap et al. 1988; Porter and Meier 1992; Bythell and Sheppard 1993; Aronson and Precht 2001).

Conservation and sustainable-use measures are insufficient for many ecosystems, so that restoration efforts are necessary to maintain biodiversity and ecosystem services (Aronson and Alexander 2013). The documented decline of coral cover and species richness along the FRT has led to an increase of propagation and ‘gardening’ approaches for restoration (Epstein et al. 2003), with goals to enhance and restore wild coral populations (Miller et al. 2014). The majority of restoration projects in Florida and the Caribbean focus on *Acropora* spp. not only because they
are keystone species, but also because they are fast growing and propagate by asexual fragmentation, which is conducive to coral-gardening approaches (Johnson et al. 2011).

In the past, restoration efforts have been criticized for being ad hoc, site or situation specific (Hobbs and Norton 1996), with a prime example being the site-specific restoration effort that took place at Molasses Reef in the Florida Keys after the M/V Wellwood ran aground and destroyed thousands of square meters of living corals (Hudson and Diaz 1988). While localized restoration efforts can be valuable, restoration efforts must be expanded in a systematic manner to larger scales in the face of ecosystem-wide coral decline. Large-scale restoration efforts have achieved success in other regions, despite unpredictable disturbance regimes and severe baseline degradation (Shaish et al. 2010; Gomez et al. 2014). This study provides the first step at statistically informing site selection for larger scale restoration efforts that are currently taking place in Florida.

While the best practices for coral restoration methods have been previously summarized (Johnson et al. 2011; Schopmeyer et al. 2012), there is little to no information regarding the spatial prioritization for restoration. The first step in determining appropriate restoration sites is to identify the ongoing threats contributing to coral reef decline, both anthropogenic and environmental in nature. Human population increases in southeast Florida pose a variety of threats to coral reef ecosystems, including nutrient enrichment, diminished water transparency, phosphate inhibition of calcification, biotic replacement, and increased bioerosion (Weiss and Goddard 1977; Smith et al. 1981; Hallock et al. 1993; Yentsch et al. 2002). In addition to the anthropogenic threats facing coral reefs, environmental threats include temperature anomalies, disease, predation and physical damage from tropical storms. Many restoration efforts have
focused on resilient species of coral that can be transplanted to address local and acute stressors, while chronic widespread stressors are mitigated via management (Gomez et al. 2014).

Stony coral restoration often involves the introduction of live coral branches from healthy colonies, in an attempt to create a competitive advantage over faster growing macroalgae, hydroids and sponges (Vermeij et al. 2009). These coral fragments are transplanted at a sufficiently large size that they are able to establish in the extremely competitive environment. After successful attachment to the substrate, the restored fragments ideally should be able to grow and compete with nearby sessile species, as these coral transplants bypass the vulnerable larval-settlement stage (Gomez et al. 2014).

The selection of an appropriate site for coral restoration is a critical first step (Yap 2003). It is essential to understand the biology and ecology of the coral species being used, and the ecological conditions of the location where the species will be transplanted (Yap 2003). The models developed in this study aim to identify sites that will be appropriate for *A. cervicornis* restoration, guided by information from population trends of *A. cervicornis* throughout the FRT.

The task of identifying the threats to coral reefs is one challenge, the mission of amassing spatial data to represent these threats is even more challenging. Until satellite imagery is sufficiently sophisticated to map individual coral communities, techniques such as species distribution modeling (SDM), which statistically relate measureable environmental variables to the occurrence of these species, are essential tools. The SDMs are especially important in environments too extensive for detailed surveys.

Statistical methods that model species distributions using the relationship between the species occurrence data and environmental variables are quite common in terrestrial studies (see
Araujo and Guisan 2006), but much less commonly used in marine environments (Bryan and Metaxas 2007; Davies et al. 2008; Tittensor et al. 2009). This study uses models in the classification and regression tree (CART) family, which are nonparametric, probabilistic, machine-learning methods. They induce a set of rules to classify the response data based on the values of the predictor data (Breiman et al. 1984). This method was chosen given the models’ ability to utilize multinomial (categorical) data and because many of the predictor data used are non-linear, are often very closely related, and interactive. The SDM approach used in this study is one way to incorporate disturbances and environmental parameters into restoration site selection.

The goal of this study was to use predictive modeling to understand the environmental controls responsible for *A. cervicornis* presence trends (continuously present, transient presence, never present) and to predict areas that are likely to exhibit these trends. To predict trend distribution, a boosted classification tree model was used with *A. cervicornis* repeat-monitoring data and nine environmental data layers. The boosted classification tree was chosen given its ability to use multinomial, categorical data as the response, and because of the limited number of assumptions of the model. Many environmental variables in marine ecosystems strongly interact in ways that are often not understood and do not typically conform to standard distributions. Therefore, a non-parametric approach is required to avoid violating assumptions made by the more traditional parametric modeling methods. The model output was then used to construct maps of predicted population trends of *A. cervicornis*.
5.3 Methods and Results

5.3.1 Study Area

This study focused on the Florida Reef Tract (FRT), which spans from Martin County in the northeast through to the Dry Tortugas in the southwest. The FRT is a subtropical barrier reef system that contains a combination of patch, linear and aggregate reefs (Shinn et al. 1989; Precht and Miller 2007). The reef tract occurs near the latitudinal limits of subtropical waters and, thus, experiences an abundance of natural stressors. Conditions along the reef tract include variable temperatures reaching extreme highs and lows, which are not generally favorable for reef development. Nevertheless, corals continue to occur in this region (Kruczynski and McManus 2002).

5.3.2 Environmental Data Sets

At the beginning of the project, a series of stakeholder workshops were held to inform the data analysis and modeling approach. A main discussion point was identification of the most important ecological and environmental data relevant to Acropora spp. distributions, as well as the availability of such data. One of the most important considerations when selecting data used in the modeling process is that the data are ecologically relevant to the species of interest (Elith and Leathwick 2009). The expert opinion provided in these stakeholder workshops guided the effort to select the data used to predict A. cervicornis distributions.

All environmental data layers, as described in Chapter 3, were formatted to consistent resolution and projection. The spatial resolution of all layers in this study was determined by the coarsest resolution of a data set used. Moderate-resolution imaging spectroradiometer (MODIS) imagery, which was a source for multiple environmental data layers, has a 1 km² resolution.
Therefore, all other environmental data layers were mapped to this resolution. All data layers were projected to the Albers projected coordinate system (PCS Albers) and converted to raster format, with the exception of the *A. cervicornis* observation-point data, which remained in vector-point format, also projected in PCS Albers.

Nine environmental data layers were used as model predictors (Table 5.1). Full details regarding the formatting and processing of the data layers are provided in Chapter 3. Data sets that reported multi-year monitoring of *A. cervicornis* were used to compile the response layer for the model. Monitoring data of *A. cervicornis* spanned the years 1996 – 2013. Each observation was converted to a categorical trend value. Each survey location was analyzed individually, and one of three categories was assigned to the location based on the trend of the multi-year data (Table 5.2). The original five categories described in Table 5.2 were pooled into three, more general, categories. The original “Disappeared” and “Appeared” categories had limited observations, thus were pooled into the “Transient Presence” category. The *A. cervicornis* data were then separated into training and validation sets, 90% for training and 10% for validation purposes. All model calibration used the training data set, and final model statistics were compared using the validation data set.

5.3.3 *AdaBoost.M1* Boosted Classification Tree Model Development

The response data in this study are multinomial categorical, thus traditional linear regression techniques cannot be applied. Instead, a non-parametric boosted classification tree was used to predict the probability of *A. cervicornis* presence/absence trend as a function of the nine environmental parameters. These models are robust to correlations among environmental variables and do not make assumptions about data layer distributions.
The adabag package within the statistical software R 3.2.1 was used (R Core Team 2015) to calibrate and run the boosted classification tree using the AdaBoost.M1 algorithm (Freund and Schapire 1996). As described in Chapter 4, all model parameters were tuned for optimal performance, and the environmental predictor layers used for the final model were reduced to the appropriate number.

Table 5.17: Description of environmental predictor data layers used in the development of the SDM in this study.

<table>
<thead>
<tr>
<th>Layer Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>Raster representation of the seven regions of the Florida Reef Tract</td>
</tr>
<tr>
<td>Reef Type</td>
<td>Representation of the dominant reef type present in each 1 km$^2$ grid cell. Reef types include: Aggregate Reef, Patch Reef, Pavement, Reef Rubble and Ridge</td>
</tr>
<tr>
<td>Distance from Shore</td>
<td>Distance in km from nearest shoreline</td>
</tr>
<tr>
<td>Depth</td>
<td>Raster representation of depth (m) taken from interpolated LiDAR/Acoustic data supplemented with data from NOAA Nautical Coastal and Approach Charts.</td>
</tr>
<tr>
<td>Slope</td>
<td>Slope (degrees) derived from raster depth</td>
</tr>
<tr>
<td>Sea Bottom Temperature (SBT)- High</td>
<td>Results of statistical model, which relates in situ SBT temperature data to satellite-derived SST data. The results of the model were used to sum total number of weeks where temperatures exceeded 30˚ C per 1 km$^2$ grid cell</td>
</tr>
<tr>
<td>SBT – Low</td>
<td>Results from the same statistical model used to derive the SBT high layer. The results of the model were used to sum total number of weeks where temperatures fell below 20˚ C per 1 km$^2$ grid cell</td>
</tr>
<tr>
<td>Light Availability at 380 NM</td>
<td>Percent of 380 nm light available at the bottom derived from MODIS/Aqua satellite measured diffuse attenuation coefficient ($K_d$) and the raster depth layer used in this study. Light at 380 nm represents light in the UV portion of the spectrum, which can often be damaging to corals</td>
</tr>
<tr>
<td>Light Availability at 488 NM</td>
<td>Similar to the Light Availability at 380 nm layer, percent light availability was derived for the 488 nm portion of the spectrum. The light in this portion of the spectrum represents the Photosynthetically Available Radiation (PAR) necessary for photosynthesis in zooxanthellate corals</td>
</tr>
</tbody>
</table>

The final boosted classification tree using the AdaBoost.M1 algorithm consisted of 200 trees, had a maxdepth of five for each tree and included six predictor variables (depth, light380, light488, region, sbthigh, slope). The final model did not use shore distance, reef type or SBT low as predictor variables. Of the six variables used in the model, depth was the most important predictor (Figure 5.1). The final model correctly classified 98% of the training data, with class errors summarized in Table 5.3. The final model had an average margin of 0.40 and a maximum
margin of 0.91. Only about 1% of the training data falls below the zero mark, meaning that 1% or less of the training data were incorrectly classified (Figure 5.2).

Table 5.2: Categories assigned to repeat-monitoring sites representing *Acropora cervicornis* data. Each repeat monitoring site consisted of a minimum of four years of *A. cervicornis* presence/absence observations. Each location was analyzed individually and one of the three pooled categories was assigned to the location based on the trend of the multi-year data.

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Pooled Category</th>
<th>Number of Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Never Present</td>
<td>Presence never recorded</td>
<td>1</td>
<td>56</td>
</tr>
<tr>
<td>Transient Presence</td>
<td>Presence was recorded some survey years, but not others.</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td>Continuously Present</td>
<td>Presence recorded by at least one survey each year</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>Disappeared</td>
<td>Presence recorded the first several survey years, but not in most recent surveys</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Appeared</td>
<td>Presence not recorded the first several survey years, but was continuously recorded in more recent surveys</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.3: Confusion matrix from the final AdaBoost.M1 model using the training data. Class error represents the percentage of times the training data points were misclassified.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>Class Error</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong></td>
<td>51</td>
<td>0</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>2</strong></td>
<td>0</td>
<td>24</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td><strong>3</strong></td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>7.7%</td>
</tr>
</tbody>
</table>

**Figure 5.1:** Relative importance (%) of variables for the final AdaBoost.M1 model
Figure 5.2: Margin cumulative distribution for the AdaBoost.M1 model. The margin is related to the certainty of classification. Margins closest to one signify a highly certain classification. Negative margins represent an incorrect classification. The cumulative distribution represents the percentage of points which equal or fall below a particular margin. For example, in this study, approximately 70% of the training data had a margin of 0.5 or less. Less than 1% of the training data had a margin of less than 1%.

5.3.4 Analysis of Model Predictions

Once the model was fully calibrated, the trained classification tree was used to predict the full reef tract (Figure 5.3). Each cell of the study area has a probability of each category associated with it. The category with the highest probability in each cell becomes the final prediction. High probabilities indicate areas that are likely to exhibit the particular modeled trend. Highly confident predictions will have a high probability for one category, and very small probabilities for the other prediction, which is also measured by the margin of the data.

Probabilities can be thought of as the confidence of the model in the prediction. For each observation, there is a set of optimum site conditions where the probabilities of each category are highest, the highest probability for each cell results in the final predicted category. The probabilities of each category were mapped across the full reef tract (Figure 5.4). Both the maps of the final predicted categories, as well as the probability of each category, can be useful for restoration managers.
Figure 5.3: The final model was used to predict unsampled locations along the full reef tract. The model used the observed environmental parameters at each location to predict which category that region would be assigned

5.3.5 Analysis of Environmental Parameters

One benefit of using boosted classification trees is their ability to accurately predict and minimize classification error on training data. Unfortunately, the nature of the boosting process results in the loss of the interpretability of simple trees, thereby reducing the amount of ecological information discernable from the final models. The final classifier built in this study is a weighted sum of trees, which can often be difficult to interpret. However, some interpretation of the ecological data can be achieved by analyzing the values of the environmental parameters from the cells of each predicted category. The values of environmental predictors in each category provide important ecological information regarding the species being modeled.
Figure 5.4: Mapped probabilities for each prediction category for *A. cervicornis*. Category 1 – Never Present, Category 2 – Transient Presence, Category 3 – Continuously Present. Each cell in the study area has a probability of each category associated with it. The final prediction is the category with the highest probability in each cell.

The values of the environmental parameters were extracted for each cell of the final model, and these values were then analyzed by the final predicted category of the cells. The most
important parameters, depth and light availability, were analyzed for each category. The cells predicted as Category 1 (Never Present) exhibited the deepest depth on average, 16.5 m, with values ranging from 1.1 – 49.8 m. The cells predicted as Category 2 (Transiently Present) and Category 3 (Continuously Present) were shallower on average, 7.2 m and 8.7 m, respectively. The cells predicted to be Category 3 had the least variation, with depths ranging from 4.6 – 18.3 m (Figure 5.5). This information provides an indication of the depth ranges appropriate for A. cervicornis habitat in south Florida today.

The percent of the light available in both the 380 nm and 488 nm wavelengths were analyzed together. For both wavelengths, the cells predicted as Category 3 exhibited higher average light availability than either of the other two categories. Cells predicted to be Category 3 had an average of 32% benthic light availability for the 380 nm wavelength, and an average of 60% benthic light availability for the 488 nm wavelength. In addition, the variability of light availability values for both wavelengths is less for cells predicted as Category 3 than cells
predicted to be either of the other two categories (Figure 5.6). This limited variability represents a more defined light availability range for observations in Category 3.

![Figure 5.6: Percent light availability means and ranges for each predicted category.](image)

5.3.6 Analysis of Restoration Locations

A database of restoration locations was acquired from The Nature Conservancy (TNC). This database contained the spatial information for *A. cervicornis* outplant locations throughout south Florida. These data were used to compare to the final model prediction to understand how the model predictions compare to restoration actions currently taking place.

Duplicate and auto-correlated points in the TNC database were removed. The outplant locations were analyzed, and any location where more than one point occurred within a single pixel of the final model was consolidated to a single point per pixel. The final filtered dataset
consisted of 59 unique restoration sites in all regions of the Florida Reef Tract except for the lower east coast and the Marquesas regions.

The outplant locations were compared to the model predictions and the number of restoration sites within each predicted category was recorded (Table 5.4). Six of the restoration sites fell outside of the extent of the model, leaving a total of 53 points within the study extent. A majority of current restoration sites (51%) occurred within Category 1 – Never Present. It is important to note that the Category 1 Prediction of the model does not necessarily represent unsuitable habitat for *A. cervicornis*. Rather, these are locations that are predicted to have never had *A. cervicornis* present within the years used to train the model, 1996 – 2013.

### Table 5.4: Number of restoration sites within each predicted category

<table>
<thead>
<tr>
<th>Category</th>
<th>Restoration Sites</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – Never Present</td>
<td>27</td>
<td>50</td>
</tr>
<tr>
<td>2 – Transient Presence</td>
<td>13</td>
<td>25</td>
</tr>
<tr>
<td>3 – Continuously Present</td>
<td>13</td>
<td>25</td>
</tr>
</tbody>
</table>

The values of the most important environmental parameters were analyzed at each restoration location, by the model category in which the observation coincided (Figure 5.7 – 5.9). These figures provide an idea of the conditions of the restoration locations in terms of the important environmental parameters. As shown in Figures 5.7 and 5.8, the majority of restoration sites located in Category 1 of the model were still located in areas of high light availability. Figure 5.9 illustrates that the restoration locations observed in Category 1 were also located in shallower depths, and falling within the range of depths predicted to Category 3. All restoration locations, no matter the category, were observed between 3.5 – 19 m depth.
Figure 5.7: Percent of solar radiation at 380 nm reaching the seafloor at the restoration locations compared to the ranges observed in the cells predicted by the model. The box represents the average for all pixels in that category, and the error bars denote the maximum and minimum observed values.

Figure 5.8: Percent of solar radiation at 488 nm reaching the seafloor at the restoration locations compared to the ranges observed in the cells predicted by the model. The box represents the average for all pixels in that category, and the error bars denote the maximum and minimum observed values.
5.3.7 Spatial Analysis of Restoration Sites

To visualize the locations of restoration sites in comparison to the model predictions, as well as the most important environmental parameters, the restoration locations were mapped alongside the model predictions and values for those parameters. The example area shown in Figures 5.10–5.12 illustrate that, while many restoration locations are located in areas predicted by the model as Category 1, many are on the edge of that category predictions, and still within areas with similar light availability and depth as areas observed in Categories 2 and 3.
Figure 5.10: *Acropora cervicornis* trend prediction on the east coast of Florida by a boosted classification tree at a 1 km scale. Current restoration locations, as well as observed light availability at 380 nm at depth, are displayed. Current restoration locations within cells predicted to Category 1 are still located within areas of similar light availability at depth as the areas predicted as Category 2 and 3.
Figure 5.11: *Acropora cervicornis* trend prediction on the east coast of Florida by a boosted classification tree at a 1 km scale. Current restoration locations, as well as observed light availability at 488 nm at depth, are displayed. Current restoration locations within cells predicted to Category 1 are still located within areas of similar light availability at depth as the areas predicted as Category 2 and 3.
Figure 5.12: Acropora cervicornis trend prediction on the east coast of Florida by a boosted classification tree at a 1 km scale. Current restoration locations, as well as modeled depth, are displayed. Current restoration locations within cells predicted to Category 1 are located within areas of similar depth as the areas predicted as Category 2 and 3.
5.4 Discussion

This study is the first attempt at spatially prioritizing restoration activities in the Florida Keys. In the past, restoration activities were guided by local knowledge and intuition. The development of this product allows for an ecologically guided approach at spatial prioritization of restoration. When using the results of this study, the goals of a restoration project must first be identified. If the primary goal is to restore A. cervicornis, then this product can be used in its current form to guide the placement of coral outplants. If the goals of the restoration project are more geared toward full reef ecosystem restoration, then the results of this study should be combined with other products to address the objective of the restoration project. Regardless of the goal, the results of this study provide the invaluable first step at spatially prioritizing restoration efforts throughout the FRT.

The results of this study provide flexibility for application based on unique restoration goals. For example, if the purpose of a restoration project is to place outplants in regions where success is most likely, in an attempt to enhance existing relatively healthy reefs, then the project could focus on regions predicted to be Category 3 (continuously present). Restoration efforts could just as easily focus on areas predicted to Category 1 (Never Present) in this study. Category 1 areas do not necessarily represent unsuitable habitat, rather they represent areas predicted to not have had A. cervicornis present between the years 1996 – 2013. These could be areas that lost A. cervicornis populations prior to 1996 and have yet to recover, reefs that are dominated by other species, or where A. cervicornis has never successfully established. Restoration approaches such as this are considered to be more dramatic and utilize the “assisted colonization” approach, by moving target species to sites where they do not currently exist or have not existed in recent history (Hoegh-Guldberg et al. 2008).
If the restoration project is more experimental in nature, the results of this modeling effort can provide testable hypotheses on potential success of location for restoration efforts. One question, which could be addressed using this model, is whether sites predicted to be Category 1 can be suitable for restoration efforts. This would require separation of the Category 1 sites into multiple sub-categories, ranging from clearly not suitable sites to potentially suitable sites. Further refining the Category 1 sites would require additional information from other coral reef ecosystem models, such as ones that model coral cover or species richness. Alternatively, environmental attributes such as light availability and depth could be used to determine which Category 1 sites are most similar to the Category 3 sites.

The primary goal of this study was to create a regional product for coral restoration activities in Florida. The precise placement of coral outplants should be determined in the field, using expert opinion and local knowledge. This product is not a substitute for coral reef ecosystem knowledge during the restoration process, and the best practices for local placement should still be followed (Johnson et al. 2011; Schopmeyer et al. 2012). Recommendations in terms of spacing, specific substrate available, and small scale zoning on the reef should be considered when placing outplants in the regionally identified areas. Small spatial scale interactions must be considered when using this tool as local interactions could determine the difference between success and failure of restored colonies. Interactions such as competition between transplants themselves, or competition with algal species, are of primary concern (Yap 2003). Spatial statistical models such as these do not inform the user about these biological interactions that influence the distribution of stony coral species. However, no previous studies have attempted to predict A. cervicornis trends for restoration purposes, thus this study fulfills that need with room for continued improvement.
As with all modeled representations of ecosystems, these models are not able to represent the full complexity of the system, and the results should be viewed with a cautionary lens. To quote the late George Box, “All models are wrong, some are useful” (Box and Draper 1987). Every aspect of the enormously complex coral reef ecosystem is impossible to consider in the model, but simplifying the ecosystem in such a model allows for the simplification of reality, moving one step closer to explaining the system as a whole.

There are a wide variety of statistical models, which could be applied to the coral reef environment. The use of boosted classification trees was identified as the most appropriate tool in this particular study. The advantages of using boosted classification trees far outweighed the downfalls, but with any study, there are still disadvantages that must be considered. The nature of the boosted classification trees result in the loss of the simple interpretability of single classification trees. Instead of being able to pull ecological information out of one tree, a series of trees must be examined. Variable importance plots must be relied on more heavily, with information about the environmental data layers as a whole, rather than the optimal levels of that environmental condition.

The analysis of the environmental data layers presented in this study provides one method to gather ecological information from a boosted classification tree model. The final model identified depth as the most important parameter, which is logical given the well-known depth limits of *A. cervicornis* (e.g., Huston 1985 a, b, and references therein). Analysis of cells predicted to be Category 3 indicate that the depth ranges of *A. cervicornis* in Florida is 4.6–18.3 m. This information is reinforced by the depth of the current restoration activities, which were found to take place within very similar depth ranges. Previous studies have suggested that *A. cervicornis* can exist in depths up to 30 m in areas of consistently very clear waters (Fenner
My results suggest that the depth limit of *A. cervicornis* is <20 m in Florida waters, consistent with estimates of Yentsch et al. (2002).

In addition to depth, light availability is another important environmental parameter, which must be considered. This study found that areas predicted to be Category 3 (Continuously Present) exhibited on average more light reaching the seafloor than in areas predicted to the other two categories. In addition to these higher levels, the areas predicted to be Category 3 also had less variability in light availability. Reduced variability, therefore more stability in light availability, has previously been found to be advantageous to coral growth and survival (Fisher et al. 2007; Ayoub et al. 2008; Ayoub et al. 2012).

One environmental parameter not analyzed in detail in this study, but very important to reef-coral growth and survival, is temperature. The sea-bottom temperature (SBT) product used in this study provided less of an opportunity to examine yearly averages and variability at locations, and instead served as a proxy for thermal stress. Previous research has identified variability in sea surface temperature (SST), rather than average values, as strong indicators of coral reef species composition and abundance (Vega-Rodriguez et al. 2015). Future modeling efforts would benefit from utilizing environmental parameters focusing on SST or SBT variability at locations, rather than climatological averages. Additionally, given the unique thermodynamic regimes experienced along portions of the Florida Reef Tract, future modeling efforts should continue to use SBT temperature rather than SST products when possible, to best represent the temperature anomalies experienced at depth.

The model presented may be improved in multiple ways. First, improving the spatial resolution of the predictor variables will improve the usefulness and applicability of the results.
The current model was restricted by the resolution of the coarsest environmental data layer used. Improved resolution of environmental layers such as light availability and temperature will greatly improve the applicability of the model to restoration efforts. There are several options to address the spatial resolution concerns of the model in its current state. One possibility is to clip the results of this study to smaller spatial scales, for example, by using benthic habitat maps and clipping the prediction to only available reef and hardbottom polygons.

In addition to improving the accuracy of the models in this study, the validation of the models would be greatly improved with an independent data set, as well as additional repeat coral monitoring data. The limited amount of training and validation data available to this study prevented the use of a purely independent data set, and validation was restricted to a relatively small data set. Validating with a previously collected data set, or by using the results from this study to design a restoration project, would assess the accuracy of the developed model. Comparing the model predictions to a database of restoration efforts, which document success and failure in the field, is the ultimate validation for this study and would increase the confidence of using the model to plan future restoration work.

This model was developed specifically for the Florida Reef Tract, and the environmental parameters used to model the A. cervicornis response were also specific. However, the statistical methods used in this study are transferable to other reef environments, such as the other U.S. jurisdictions where Acropora spp. are listed as threatened: Puerto Rico and the U.S. Virgin Islands. Once appropriate environmental predictor variables are identified for a geographic area, non-parametric classification trees are appropriate to use for response data, as done in this study. The results of this research confirm the applicability of the species distribution modeling approach to stony coral species trends, with applications in restoration science, conservation
science, and management. This work also confirms that SDM research is not restricted to simply presence/absence models, and that more ecological information can be predicted using these techniques.

The possibilities for future research in this field are extensive. First and foremost, this study focused on a single species. This approach has unique benefits as it allows for the exploration of their realized niche (Austin and Smith 1989). However, there is also value in modeling at the community level. A community modeling approach would be able to identify healthy reefs, rather than locations where single species are present. The models developed in this study for a single species generally require more sophisticated and refined statistical fitting than models that attempt to predict community structure (Zimmermann and Kienast 1999). Therefore, a natural extension of this study would be to attempt to model the community, with less refinement of the statistical model required. A previous study by Zimmerman and Kineast (1999) proposed that a model of a single species is more ‘biological’ in nature and therefore more realistic at predicting large-scale patterns compared to community-based models. The latter must also consider competition that occurs in addition to environmental requirements that result in a specific community response. Modeling the community structure could, therefore, result in more refined products, at higher resolution.

In conclusion, this tool is the first of its kind to provide spatial information regarding the placement of restoration projects along the Florida Reef Tract. When implementing any restoration effort, it is important to consider the ultimate goal, restoring the health of Florida’s coral reef ecosystem. Unfortunately, there is still debate regarding resilience trajectories of Caribbean coral reefs. Are we attempting to restore something that is doomed, or is there still hope (Mumby et al. 2013; Bozec and Mumby 2015)? The results of this research provide yet
another step towards determining whether these important ecosystems have the capability to
bounce back.

5.5 References

Biogeogr 33:1677-1688

Aronson J, Alexander S (2013) Ecosystem restoration is now a global priority: Time to roll up
our sleeves. Restor Ecol 21:293-296

reefs. Hydrobiologia 460:25-38


coral reefs by controlling exposure to UVR. Proc 11th Int Coral Reef Symp.
http://nsuworks.nova.edu/cgi/viewcontent.cgi?filename=117&article=1000&context=occ
_i CRS&type=additional

Keys phytoplankton vary with distance from shore and CDOM: Implications for coral

Sons, New York, NY

Bozec YM, Mumby PJ (2015) Synergistic impacts of global warming on the resilience of coral

Chapman & Hall/CRC, Boca Raton

Bryan TL, Metaxas A (2007) Predicting suitable habitat for deep-water gorgonian corals on the

26:296-297

Callahan M, Wheaton J, Beaver CR, Brooke S, Johnson D, Kidney J, Kupfner S, Porter JW,
Project 2006 Executive Summary, 25

Davies AJ, Wisshak M, Orr JC, Roberts JM (2008) Predicting suitable habitat for the cold-water

Elith J, Leathwick JR (2009) Species distribution models: Ecological explanation and prediction

Epstein N, Bak RPM, Rinkevich B (2003) Applying forest restoration principles to coral reef


corals Montastraea spp. as indicators of colony condition. Mar Ecol Prog Ser 339:61-71

Proceedings 13th International Conference pp 148-156
Huston M (1985a) Variation in coral growth rates with depth at Discovery Bay, Jamaica. Coral Reefs 4:19-25


6 Summary and Conclusions

Elkhorn and Staghorn corals (*Acropora palmata, A. cervicornis*) were listed as threatened species under the Endangered Species Act in 2006 (National Marine Fisheries Service 2006). In 2008 both species were listed as critically endangered on the International Union for Conservation of Nature (IUCN) Red List of Threatened Species. The decline of these species is unprecedented given the vital role they historically played as major builders of western Atlantic and Caribbean coral reefs. The primary goal of this study was to build upon previously created critical habitat maps for *A. palmata* and *A. cervicornis* to show areas where these species currently exist, as well as areas that would be suitable for their (re)establishment through restoration using a database of reported *in situ* observations and existing mapped data. In previous work, potential habitat maps were produced based on benthic substrata throughout the Florida Reef Tract. The first objective of this study was to update these maps using more recently updated benthic habitat maps, as well as to expand the study area to include Puerto Rico and the U.S. Virgin Islands (Wirt et al. 2015).

Habitat parameters conducive to coral reef growth include not only benthic substrate and depth. Parameters such as light availability, geomorphology, and temperature play key roles in determining suitable habitat for many corals, including *Acropora* spp. As such, to best represent habitat suitable for restoration efforts of *A. cervicornis*, the second objective of this study was to include these additional parameters when creating species distribution models for the purpose of identifying suitable restoration sites. The results of this research may be used to (1) inform
restoration site selection(2) inform the current NOAA *Acropora* spp. critical habitat map, and (3) provide a basis for future coral critical habitat maps. These objectives are presented in a series of four chapters in this dissertation; each chapter describes a subset of the research necessary to fulfill the overall objectives. Chapter 1 provided the background and literature review related to *Acropora* spp. history and environmental requirements.

6.1 Research Overview

Chapter 2 examined whether the methods developed in my Master’s Thesis (Wirt 2011) for the Florida Reef Tract were applicable to the U.S. Virgin Islands and Puerto Rico. This work was published by Wirt et al. (2015) in *Global Ecology and Conservation*. The results of the study indicate that potential habitat for *A. palmata* is currently well defined throughout the study region, but that potential habitat for *A. cervicornis* is more variable and has a wider range than that for *A. palmata*. The results of this study prompted the further examination of *A. cervicornis* habitat suitability using other environmental parameters, in the subsequent chapters of this dissertation.

Chapter 3 is primarily a methods paper that describes the compilation and processing steps used to amass relevant coral reef ecosystem data to use for multiple spatial modeling and mapping purposes. The data layers described represent the best currently available environmental data to use for *A. cervicornis* species distribution models. The results of Chapter 3 provide a model for compilation of spatial data available for the creation of coral reef species distribution models in Florida. Improvements can be made in terms of spatial resolution, however the data represented are appropriate for the creation of a regional model.
Chapter 4 compared and evaluated two statistical models, random forest and boosted classification trees, for their performance at predicting *A. cervicornis* trend data along the Florida Reef Tract. The models were used with repeat *A. cervicornis* monitoring data, to predict the regional stability of these populations. The results of this chapter show that species distribution modeling techniques are useful tools in marine ecosystems. The boosted classification trees outperformed the random forest models at classifying the *A. cervicornis* response data. The two boosted classification tree algorithms performed similarly, but the AdaBoost.M1 algorithm slightly out-performed the SAMME algorithm. Maps, such as the ones produced in Chapter 4, provide valuable baseline data, especially for unsurveyed regions (Pittman et al. 2009). The model results are valuable tools for managers and restoration program designers that allow them to analyze environments on a landscape-scale (Remillard and Welch 1993; Kelly et al. 2001).

Chapter 5 used predictive modeling to understand the environmental controls responsible for *A. cervicornis* presence trends (continuously present, transient presence, never present) and to predict areas that are likely to exhibit these trends. To predict trend distribution, a boosted classification tree model was used with *A. cervicornis* repeat-monitoring data and nine environmental data layers. Depth, light availability, slope, sea bottom temperature and region were identified as the most important predictor layers. Depth and light availability were further analyzed by comparing the means and ranges for each predicted category. Cells predicted as Categories 2 (Transient Presence) and 3 (Continuously Present) were shallower on average and had smaller depth ranges than the cells predicted to be Category 1 (Never Present). This analysis provides an indication of the depth ranges appropriate for *A. cervicornis* habitat in South Florida today. Similar analysis for light availability was performed, revealing that cells predicted as Category 3 had less variability in light reaching the seafloor for both wavelengths assessed (380...
nm and 488 nm) than cells predicted to be either of the other two categories. These findings further supported results from previous studies that found stability in light availability to be advantageous to coral growth and survival (Fisher et al. 2007; Ayoub et al. 2008; Ayoub et al. 2012). The reduced variability of the light reaching the benthos in areas predicted as Category 3 likely contributes to the stable presence of the species in these areas. The model output was used to construct maps of predicted population trends of *A. cervicornis* along the Florida Reef Tract. The results of the model were compared to a database of restoration efforts currently taking place in South Florida. While many restoration locations are located in areas predicted by the model as Category 1 (Never Present), many are on the edge of the cells, and still within areas with similar light availability and depth as cells predicted as Categories 2 and 3. This research is the first of its kind to attempt to spatially prioritize restoration activities along the Florida Reef Tract. The development of this product allows for an ecologically-guided approach at spatial prioritization of restoration.

A major conclusion of this research is that environmental parameters, when combined with proper models, can be used to predict presence/absence trends of *A. cervicornis* observations along the Florida Reef Tract. Species Distribution Modeling is a useful tool for marine ecologists to use when monitoring data are sparse. The results of this research confirm the applicability of the species distribution modeling approach to stony coral species trends, with applications in restoration science, conservation science, and management. This work also confirms the fact that SDM research is not restricted to simply presence/absence models, and that more ecological information can be predicted using these techniques. There are countless varieties of SDMs available to ecologists, and there is no single method that will work in all situations. The conclusions of this research confirm that when modeling the coral reef
environment, multiple approaches should be taken and a variety of models should be tested before identifying the best model. Continued usage of statistical models, such as those explored in this research, will be vital for the future of coral reef conservation.

6.2 Future Research and Recommendations

The opportunity for future research in this field is extensive. This dissertation represents only the beginning of the potential to use species distribution modeling for prioritization of restoration locations. As the use of species distribution models in the marine ecosystem continues to demonstrate applicability, improvements can be made in all areas related to model development. First and foremost, improvements in spatial resolution of the models can be made. The results of this research provide regional recommendations for restoration projects, rather than recommendations for specific reefs. Within each 1 km$^2$ cell assessed by the model, restoration practitioners must still use expert opinion to determine precise placement of coral outplant sites. Recommendations in terms of spacing, specific substrate available, and small scale zoning on the reef must be considered when placing outplantings in the regionally identified areas. Continuous improvement of environmental data layers resulting from improved spatial resolution of satellite data may eventually allow for finer-scale prediction of suitable restoration sites.

In addition to improving the spatial resolution of the models presented in this study, additional models that predict species richness and coral cover will be extremely useful to combine with the results of this research. Combining predicted species richness and coral cover data with the $A.\ cervicornis$ trend predicted in this study will help to identify reefs that have lost $A.\ cervicornis$, but may support restoration efforts due to the presence of other species and/or
substantial coral cover. The ultimate goal of research in this field is to create a customizable and interactive tool to that can be adjusted to support the unique goals of individual restoration projects.

Future research in this field should also explore the idea of modeling the reefs along the southeast coast of Florida separately from the Florida Keys and the Dry Tortugas. The models presented in this dissertation could be improved in this area. Reefs along the east coast exist in a much narrower shelf than the reefs of the Keys and the Dry Tortugas. The environmental data layers used may not be of appropriate resolution to distinguish the fine-scale differences in the habitats along the narrow reef tract of this region. Higher spatial resolution data should substantially improve model performance along the southeast coast. The reefs of the east coast of Florida experience different hydrodynamics and ultimately ecological drivers than reefs further to the south along the Florida Reef Tract. The ecological differences of these regions must be further explored to best represent what may be two very different systems in terms of coral reef condition and survival.

6.3 References

Appendix A: Copyright Clearance

Dear Katie:

As an Elsevier journal author, you retain the right to Include the article in a thesis or dissertation (provided that this is not to be published commercially) whether in part or in toto, subject to proper acknowledgment; see http://www.elsevier.com/about/company-information/policies/copyright/personal-use for more information. As this is a retained right, no written permission from Elsevier is necessary.

If I may be of further assistance, please let me know.

Best of luck with your dissertation and best regards,

Hop

Hop Wechsler
Permissions Helpdesk Manager

Elsevier
1600 John F. Kennedy Boulevard
Suite 1800
Philadelphia, PA 19103-2899
Tel: +1-215-239-3520
Mobile: +1-215-900-5674
Fax: +1-215-239-3805
E-mail: h.wechsler@elsevier.com

Contact the Permissions Helpdesk:
+1-800-523-4069 x 3808 permissionshelpdesk@elsevier.com
Personal use

Authors can use their articles, in full or in part, for a wide range of scholarly, non-commercial purposes as outlined below:

- Use by an author in the author’s classroom teaching (including distribution of copies, paper or electronic)
- Distribution of copies (including through e-mail) to known research colleagues for their personal use (but not for Commercial Use)
- Inclusion in a thesis or dissertation (provided that this is not to be published commercially)
- Use in a subsequent compilation of the author’s works
- Extending the Article to book-length form
- Preparation of other derivative works (but not for Commercial Use)
- Otherwise using or re-using portions or excerpts in other works

These rights apply for all Elsevier authors who publish their article as either a subscription article or an open access article. In all cases we require that all Elsevier authors always include a full acknowledgement and, if appropriate, a link to the final published version hosted on Science Direct.
About the Author

Katherine (Katie) Wirt Ames was born in Lafayette, Indiana and was raised in Brownsburg, Indiana. She attended DePauw University in Greencastle, Indiana, where she began her collegiate career as a Physics major. During her sophomore year, she traveled to Cozumel for an off-campus Winter Term, where she discovered her passion for diving and interest in the coral reef ecosystem. Upon her return from Cozumel, she immediately began investigating marine research study abroad options. Katie spent the Fall semester of 2007 studying Marine Ecology on South Caicos, Turks and Caicos Islands, where she solidified her notion that she wanted to spend her life as a marine ecologist. Upon her return to the states, she immediately changed her major and graduated with a B.A. in Biology in 2009.

Katie entered the College of Marine Science at the University of South Florida in the Fall of 2009 with Dr. Pamela Hallock Muller as her academic advisor and began working at the Florida Fish and Wildlife Research Institute (FWRI) in the Center for Spatial Analysis with Dr. David Palandro. While at FWRI, Katie refined her GIS and mapping skills, and began using statistical models in the coral reef environment. Today, Katie is a Research Associate at FWRI working on multiple marine and coral reef mapping and modeling projects.

In her free time, Katie enjoys spending time with her husband and her three dogs. Katie is an avid runner and cyclist, and loves baking and taking on projects in her family’s 1925 bungalow.