DISTRIBUTION MODELING OF MESOPHOTIC CORALS AND ALGAE TO FURTHER ECOLOGICAL UNDERSTANDING AND RESOURCE MANAGEMENT OF HAWAIIAN CORAL REEFS

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

ZOOLOGY

MAY 2018

By

Lindsay Marion Veazey

Dissertation Committee:

Robert Toonen, Chairperson

Christopher Kelley

Erik Franklin

Peter Marko

Tamara Ticktin

Keywords: species distribution modeling, spatial ecology, mesophotic coral reef ecosystems

ACKNOWLEDGMENTS

First, I dedicate this manuscript to my parents, whose encouragement, love, hiking companionship, and appreciation of all non-human animal friends strengthened my dedication to the discipline of biological sciences and emboldened my passion for environmental stewardship and animal welfare.

The completion of the work presented in this document would not have been possible without the support, conversation, commiseration, patience, and ample good humor of my ‘ohana near and far. The attainment of this degree is the culmination of many, many years of curiosity and wonder shared with my friends and family who have resolutely encouraged my pursuit of scientific truth. I dedicate this manuscript to all of you.

Finally, to all of the women- particularly women scientists- who confronted a world shaped by and for men, and decided they would do it anyway: my path was clearer because of your resilience. I count the submission of this dissertation as one more seed of progress that may facilitate the realization of a world shaped by strong, tough, educated, and independent women.

ABSTRACT

Hawaiian photic reef ecosystems, found from the surface to 30 m depth, are imperiled by anthropogenic stressors such as climate change, overfishing, introduction of invasive species, and sedimentation from coastal construction. Mesophotic coral ecosystems (MCEs) found from 30 to 180 m may be buffered from these stressors by their deeper depth range. However, their vulnerability is not well known because minimal data exist on their location and community structure around the main islands. Consequently, I developed multiple, spatially explicit models to predict the distribution of important MCE members, which included scleractinians of the genera *Montipora* and *Leptoseris*, and the calcified alga *Halimeda kanaloana.* Similar models were also created for the invasive alga *Avrainvillea amadelpha* to better understand its potential impact on Hawaiian MCEs. Using the results of my models, I created maps across the main Hawaiian Islands, identifying the most suitable habitat for (i.e., the most probable location for colonization by) these species of interest. These are the first pan-Hawai‘i predictive distribution maps for corals and algae within the mesophotic zone, which will hopefully serve as a valuable source of information to resource managers, scientists, and stakeholders in future marine spatial planning efforts.

LIST OF TABLES

LIST OF FIGURES

CONTENTS

Chapter 1: Introduction

1.1 Painting a picture of the unseen

Approximately one-quarter of the estimated 8.8 million eukaryotic species on earth are marine (Mora et al. 2011), and given that less than 5% of the ocean has been mapped in detail, this number may markedly underestimate the true biological wealth hidden in the depths of the sea. Technological advancements in survey methods, including satellite imagery, submersible vehicles, and closed-circuit diving, combined with the powerful processing capabilities of modern computers, facilitate the investigation of underexplored ocean regions like never before (Whitcomb et al. 2000; Sieber & Pyle 2010; Morrow & LeTraon 2012; Blondeau-Patissier et al. 2014; Pyle et al. 2016). The creativity and dedication of scientists across disciplines promoted the discovery of tens of thousands of terrestrial and marine plant and animal species in the past decade alone (Appeltans et al. 2012; Christenhusz & Byng 2016).

Despite the apparent abundance of information we have yet to uncover about the natural world, evidence suggests we are in the midst of the sixth mass extinction event- and fault lies entirely with us (Wake & Vredenburg 2008; Steffen et al. 2011). Industrialized animal agriculture and overconsumption of animal products (e.g., Carlsson-Kanyama et al. 2009; Tilman et al. 2002; Gill et al. 2010), overpopulation (Pimentel 2012; Butler 1994), changes in land use (e.g., Gauthier et al. 2005), and the broad, unstoppable impacts of human-caused climate change (Intergovernmental Panel on Climate Change 2014) are all playing a role in the decline of natural habitat that is contributing to the current extinction event. Two-thirds of monitored invertebrate populations show a global decline in mean abundance of almost 50% in the past several hundred years (Dirzo et al. 2014). With extinction rates at an estimated $10^2 – 10^3$ times higher than pre-Anthropocene levels, scientists estimate an annual loss of 11,000 – 58,000 species worldwide (Pimm et al. 1995; Scheffers et al. 2012; Mora et al. 2013).

A report by the National Research Council explains the irreversibility of crossing a climate threshold: “Technically, an abrupt climate change occurs when the climate system is forced to cross some threshold, triggering a transition to a new state at a rate determined by the climate system itself and faster than the cause” (National Research Council 2013). We are on the cusp of such a threshold. Now, more than ever before, we need to assess the world around us to define some baseline for species biodiversity, abundance, and distribution.

1.2 Niche theory

When introducing a new term, one would customarily start with the definition. For the term "niche", it's not so simple; for almost a century, ecologists have wrestled with the complexity of such a keystone concept. The term itself is likely derived from the Middle French word *nicher*, which translates as *to nest.* I will briefly introduce the three predominant perspectives on the concept of an ecological niche, starting with the Grinnellian niche, followed by the Eltonian niche, and finally, the most widely accepted Hutchinsonian niche.

Grinnell (1917) is the first researcher to apply this term in a paper about adaptations of the California thrasher (*Toxostoma redivivum*)- i.e., strong legs, long tail, and short wings, all perfectly conducive to hopping slyly through brush- to its woody chaparral habitat. The Grinnellian niche is environment-centric in that the habitat determines the various physiological and behavioral adaptations of a species. This definition permits the existence of both overlapping and empty niche spaces.

A more food-specific definition comes from Elton (1927). This definition is dynamic: the feeding habits of an animal may change over its lifespan or due to external forcings, such as changes in climate or prey populations. Elton grouped species into niches based on their foraging habits (e.g., "birds of prey"). This perspective introduces the idea of the effects of a species on both the biotic and abiotic components of their surrounding environment.

The Hutchinsonian niche is defined as an *n*-dimensional volume (i.e., "hypervolume") configured by the resources and environmental state which collectively determine the requirements of a species to persist. Theoretically, each species has a "fundamental niche" of usable resources that it could exploit if freed from interference by other species. Of course, species never exist in such a vacuum; thus, the "realized niche" is the true, real-world niche space of a species. Defined this way, each species has its own unique niche; in other words, this is very much a species-centric definition, and it is the one I have subscribed to while writing this dissertation.

In seeking to understand the role of a species in a given ecosystem, we are actively looking to define its niche space. Scientists can procure this understanding through long-term *in situ* observation, but this type of study can become cost-prohibitive, time-intensive, and logistically challenging. This is when ecological modeling can help.

1.3 Ecological modeling: species distributions and habitat suitability

Ecological modeling, alternatively defined as “environmental niche modeling” or, more commonly, “species distribution modeling”, combines mathematical formulae and computing power to uncover drivers that determine where and why organisms persist in particular areas. Models can provide a mathematically-sound basis for assessing the probability of the presence of a study organism in a given location at a given time; this may also be interpreted as the suitability of a given area for habitation by that organism. Factors affecting the performance of models include the quality and availability of input data used to train the model (Wisz et al. 2008; Guisan et al. 2007; Kramer-Schadt et al. 2013), the complexity of the model (Syfert et al. 2013), and unincorporated factors affecting the realized niche space of a species, such as life history traits, interspecific competition, or dispersal barriers (Araujo & Guisan 2006; Václavík & Meentemeyer 2009).

Despite the aforementioned challenges modelers must address, ecological models may offer an otherwise infeasible snapshot of the spatial ecology of unsurveyed regions or project how a community may change as the result of some perturbation (e.g., Gottfried et al. 1999; Van der Putten et al. 2010; Barve et al. 2011). Ecological models are particularly useful in areas that are hard to reach. The marine region of study explored in this dissertation, the mesophotic zone, is an example of an underexplored, hard-to-reach habitat.

1.4 Mesophotic coral ecosystems (MCEs): globally and in Hawai‘i

The name of the mesophotic (“meso” + “photic” = “middle light”) zone conveys its very nature: a light-starved transition zone extending from 30 – 180 m in depth and found in tropical and subtropical oceans (across the equator to about ~ 40° north and south latitude). Dr. Rich Pyle applied the “twilight zone” label to reefs in this depth range to convey its uniquely light-limited, relatively understudied qualities (Pyle 1996). The transitional nature of this zone plays a part in the relative dearth of data: the upper depth boundary (30 m) is the lower limit for standard open circuit diving, while the lower depth boundary (~165 – 200 m) is too shallow to justify costly, extensive exploration via submersibles or robots.

The potential biological and economic importance of MCEs cannot be underestimated. The sessile inhabitants of this zone, including corals, algae, and sponges, provide structurally complex habitat that is buffered by depth from thermal stress events, which may provide refuge for juvenile fishes, stressed shallow-water populations, or migratory organisms from other regions (Glynn 1996). Mesophotic depths are part of essential bottomfish habitat (Blyth-Skyrme

et al. 2013) and host commercially valuable black corals (Wagner et al. 2011). Researchers have observed up to 100% endemism within mesophotic reef fish assemblages in the Northwestern Hawaiian Islands (Kosaki et al. 2017). As the number of processed samples of mesophotic flora and fauna steadily increases, so do the records of newly identified algae (Spalding 2012; Spalding et al. 2016) and coral (e.g., Rooney et al. 2010) species, many of which may be endemic to the Hawaiian Islands.

The mesophotic zone may be our only hope for reef persistence as we face the looming climate change crisis. Coral bleaching (Hoegh-Guldberg 1999; Baker et al. 2008), overfishing (Mumby 2016), and land-based nutrient and sediment influxes (Pastorok & Bilyard 1985; Iar & Thurber 2015; Zaneveld et al. 2016) plague shallow (< 30 m) reefs, which have experienced a decline of up to 80% across certain regions (Gardner et al. 2003). In a survey of Caribbean reefs, Bak et al. (2005) found that the percent coral cover of deep coral communities remained fairly stable over several decades, and that these deep reefs experienced losses only in response to periodic, catastrophic events, such as storms or cold water upwelling episodes (it should be noted that thermal stress and storm events are also periodic stressors across shallow reef systems). However, MCEs are not invincible in the face of climate change; they face some of the same stressors as shallow reefs - including disease (Smith et al. 2016), storm damage (White et al. 2017), pollution (Appeldoorn et al. 2016; Etnoyer et al. 2016), invasive species (Andradi-Brown et al. 2017), and overfishing (Loya et al. 2016). Their depth makes them uniquely susceptible to the potential effects of a steadily shoaling aragonite saturation horizon (Sarma et al. 2002; Thresher et al. 2015) and increased light attenuation, and drowning due to sea level rise (Sanborn et al. 2017).

Given that we now know of at least some of the ecological and economic importance of MCEs, and given that we are in the midst of a human-driven mass extinction event, it serves scientists, coastal communities, policy makers, and resource managers to gather as much information as possible about our marine resources and use this knowledge to establish a baseline for management and conservation of these resources.

1.5 Dissertation objectives

This thesis uses biological observational data of hard corals (*Leptoseris* sp. and *Montipora* sp.) and green algae (*Halimeda kanaloana* and *Avrainvillea amadelpha*) and explanatory environmental data to create species distribution models at different spatial scales

and across different study areas within the main Hawaiian Islands (MHI). The work composing this dissertation achieves the following objectives:

1. Generate the first models of hard coral distribution across the mesophotic zone (30 – 180 m) of the entire span of the MHI;
2. Explore the influence of different environmental predictors at varying spatial scales on the distribution of mesophotic corals and algae;
3. Generate the first model of predicted native invasive *Avrainvillea amadelpha* distribution and identify "hotspot" regions of concern across nearshore to mesophotic (0 – 90 m) O‘ahu and surrounding MHI.

My hope in this endeavor is to figuratively "shed light" on the mysteries of the underexplored mesophotic reef system: What does it look like? What drivers shape certain mesophotic coral and algae communities? This work serves to answer those questions and provide scientists and managers with a greater understanding of the ecology of these communities so that we may all continue to work together to conserve endemic marine fauna of Hawai‘i, especially those just beyond reach.

1.6 References

Appeldoorn, R., Ballantine, D., Bejarano, I., Carlo, M., Nemeth, M., Otero, E., Pagan, F., Ruiz, H., Schizas, N., Sherman, C., and Weil, E. (2016). Mesophotic coral ecosystems under anthropogenic stress: a case study at Ponce, Puerto Rico. *Coral Reefs* 35(1): 63 – 75.

Appeltans, W., Ahyong, S. T., Anderson, G., Angel, M. V., Artois, T., Bailly, N., Bamber, R., Barber, A., Bartsch, I., Berta, A., and Błażewicz-Paszkowycz, M. (2012). The magnitude of global marine species diversity. *Current Biology* 22(23): 2189 – 2202.

Andradi-Brown, D. A., Grey, R., Hendrix, A., Hitchner, D., Hunt, C. L., Gress, E., Madej, K., Parry, R. L., Régnier-McKellar, C., Jones, O. P., and Arteaga, M. (2017). Depth-dependent effects of culling—do mesophotic lionfish populations undermine current management? *Royal Society Open Science* 4(5): 170027.

Araujo, M. B., and Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33(10): 1677 – 1688.

Bak, R. P., Nieuwland, G., and Meesters, E. H. (2005). Coral reef crisis in deep and shallow reefs: 30 years of constancy and change in reefs of Curacao and Bonaire. *Coral Reefs* 24(3): 475 – 479.

Baker, A. C., Glynn, P. W., and Riegl, B. (2008). Climate change and coral reef bleaching: An ecological assessment of long-term impacts, recovery trends and future outlook. *Estuarine, Coastal and Shelf Science* 80(4): 435 – 471.

Barve, N., Barve, V., Jiménez-Valverde, A., Lira-Noriega, A., Maher, S. P., Peterson, A. T., Soberon, J., and Villalobos, F. (2011). The crucial role of the accessible area in ecological niche modeling and species distribution modeling. *Ecological Modelling* 222(11): 1810 – 1819.

Blondeau-Patissier, D., Gower, J. F., Dekker, A. G., Phinn, S. R., and Brando, V. E. (2014). A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. *Progress in Oceanography* 123: 123 – 144.

Blyth-Skyrme, V. J., Rooney, J. J., Parrish, F. A., and Boland, R. C. (2013). Mesophotic coral ecosystems—Potential candidates as essential fish habitat and habitat areas of particular concern. US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Pacific Islands Fisheries Science Center.

Butler, C. (1994). Overpopulation, overconsumption, and economics. *The Lancet* 343(8897): 582 – 584.

Carlsson-Kanyama, A., and González, A. D. (2009). Potential contributions of food consumption patterns to climate change. *The American Journal of Clinical Nutrition* 89(5): 1704S – 1709S.

Christenhusz, M. J., and Byng, J. W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa*, 261(3): 201 – 217.

Dirzo, R., Young, H. S., Galetti, M., Ceballos, G., Isaac, N. J., and Collen, B. (2014). Defaunation in the Anthropocene. *Science* 345(6195): 401 – 406.

Elton, C. S. (1927). *Animal Ecology*. University of Chicago Press.

Etnoyer, P. J., Wickes, L. N., Silva, M., Dubick, J. D., Balthis, L., Salgado, E., and MacDonald, I. R. (2016). Decline in condition of gorgonian octocorals on mesophotic reefs in the northern Gulf of Mexico: before and after the Deepwater Horizon oil spill. *Coral Reefs* 35(1): 77 – 90.

Gardner, T. A., Cote, I. M., Gill, J., Gant, A., and Watkinson, A. R. (2003). Long-term region-wide declines in Caribbean corals. *Science* 301: 958 – 960

Gauthier, G., Giroux, J. F., Reed, A., Bechet, A., and Belanger, L. (2005). Interactions between land use, habitat use, and population increase in greater snow geese: what are the consequences for natural wetlands? *Global Change Biology* 11(6): 856 – 868.

Gill, M., Smith, P., and Wilkinson, J. M. (2010). Mitigating climate change: the role of domestic livestock. *Animal* 4(3): 323 – 333.

Glynn, P.W. (1996). Coral reef bleaching: facts, hypotheses and implications. *Global Change Biology* 2: 495 – 509.

Gottfried, M., Pauli, H., Reiter, K., and Grabherr, G. (1999). A fine-scaled predictive model for changes in species distribution patterns of high mountain plants induced by climate warming. *Diversity and Distributions* 5(6): 241 – 251.

Grinnell, J. (1917). The niche-relationships of the California Thrasher. *The Auk* 34(4): 427 – 433.

Guisan, A., Graham, C. H., Elith, J., and Huettmann, F. (2007). Sensitivity of predictive species distribution models to change in grain size. *Diversity and Distributions* 13(3): 332 – 340.

Hoegh-Guldberg, O. (1999). Climate change, coral bleaching and the future of the world's coral reefs. *Marine and Freshwater Research* 50(8): 839 – 866.

Intergovernmental Panel on Climate Change (2014). *Climate Change 2014–Impacts, Adaptation and Vulnerability: Regional Aspects*. Cambridge University Press.

Kosaki, R. K., Pyle, R. L., Leonard, J. C., Hauk, B. B., Whitton, R. K., and Wagner, D. (2017). 100% endemism in mesophotic reef fish assemblages at Kure Atoll, Hawaiian Islands. *Marine Biodiversity* 47(3): 783 – 784.

Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schröder, B., Lindenborn, J., Reinfelder, V., Stillfried, M., Heckmann, I., Scharf, A. K., Augeri, D. M., and Cheyne, S. M. (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions* 19(11): 1366 – 1379.

Loya, Y., Eyal, G., Treibitz, T., Lesser, M. P., and Appeldoorn, R. (2016). Theme section on mesophotic coral ecosystems: advances in knowledge and future perspectives. *Coral Reefs* 35(1): 1 – 9.

Mora, C., Rollo, A., and Tittensor, D. P. (2013). Comment on “Can I name Earth’s species before they go extinct?” *Science 341*: 237.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., and Worm, B. (2011). How many species are there on Earth and in the ocean? *PLoS Biology* 9(8): e1001127.

Morrow, R., and Le Traon, P. Y. (2012). Recent advances in observing mesoscale ocean dynamics with satellite altimetry. *Advances in Space Research* 50(8): 1062 – 1076.

Mumby, P. J. (2016). Stratifying herbivore fisheries by habitat to avoid ecosystem overfishing of coral reefs. *Fish and Fisheries* 17(1): 266 – 278.

National Research Council (2013). *Abrupt impacts of climate change: Anticipating surprises*. National Academies Press.

Pastorok, R. A., and Bilyard, G. R. (1985). Effects of sewage pollution on coral-reef communities. *Marine Ecology Progress Series* 21: 175 – 189.

Pimentel, D. (2012). World overpopulation. *Environment, Development and Sustainability* 14(2): 151 – 152.

Pimm, S. L., Russell, G. J., Gittleman, J. L., and Brooks, T. M. (1995). The future of biodiversity. *Science* 269(5222): 347.

Pyle, R. L. (1996). The twilight zone. *Nat. Hist. Mag.* 105: 59 – 62.

Pyle, R. L., Boland, R., Bolick, H., Bowen, B. W., Bradley C. J., Kane, C., Kosaki, R. K., Langston, R., Longenecker, K., Montgomery, A., Parrish, F. A., Popp, B. N., Rooney, J., Smith, C. M., Wagner, D., and Spalding, H. L. (2016) A comprehensive investigation of mesophotic coral ecosystems in the Hawaiian Archipelago. *PeerJ* 4:e2475 DOI:10.7717/peerj.2475.

Rooney, J., Donham, E., Montgomery, A., Spalding, H., Parrish, F., Boland, R., Fenner, D., Gove, J., and Vetter, O. (2010). Mesophotic coral ecosystems in the Hawaiian Archipelago. Coral Reefs 29(2): 361 – 367.

Sanborn, K.L., Webster, J.M., Yokoyama, Y., Dutton, A., Braga, J.C., Clague, D.A., Paduan, J.B., Wagner, D., Rooney, J.J. and Hansen, J.R. (2017). New evidence of Hawaiian coral reef drowning in response to meltwater pulse-1A. *Quaternary Science Reviews* 175: 60 – 72.

Sarma, V. V. S. S., Ono, T., and Saino, T. (2002). Increase of total alkalinity due to shoaling of aragonite saturation horizon in the Pacific and Indian Oceans: Influence of anthropogenic carbon inputs. *Geophysical Research Letters* 29(20).

Scheffers, B. R., Joppa, L. N., Pimm, S. L., and Laurance, W. F. (2012). What I know and don't know about Earth's missing biodiversity. *Trends in Ecology & Evolution* 27(9): 501 – 510.

Sieber, A., and Pyle, R. (2010). A review of the use of closed-circuit rebreathers for scientific diving. *Underwater Technology* 29(2): 73 – 78.

Smith, T. B., Gyory, J., Brandt, M. E., Miller, W. J., Jossart, J., and Nemeth, R. S. (2016). Caribbean mesophotic coral ecosystems are unlikely climate change refugia. *Global Change Biology* 22(8): 2756 – 2765.

Spalding, H. L. (2012). *Ecology of mesophotic macroalgae and* Halimeda kanaloana *meadows in the main Hawaiian Islands*. Doctoral dissertation. University of Hawai‘i at Mānoa.

Spalding, H. L., Conklin, K. Y., Smith, C. M., O'Kelly, C. J., and Sherwood, A. R. (2016). New Ulvaceae (Ulvophyceae, Chlorophyta) from mesophotic ecosystems across the Hawaiian Archipelago. *Journal of Phycology* 52(1): 40 – 53.

Steffen, W., Persson, Å., Deutsch, L., Zalasiewicz, J., Williams, M., Richardson, K., Crumley, C., Crutzen, P., Folke, C., Gordon, L., and Molina, M. (2011). The Anthropocene: From global change to planetary stewardship. *AMBIO: A Journal of the Human Environment* 40(7): 739 – 761.

Syfert, M. M., Smith, M. J., and Coomes, D. A. (2013). The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS One* 8(2): e55158.

Thresher, R. E., Guinotte, J. M., Matear, R. J., and Hobday, A. J. (2015). Options for managing impacts of climate change on a deep-sea community. *Nature Climate Change* 5(7): 635 – 639.

Tilman, D., Cassman, K. G., Matson, P. A., Naylor, R., and Polasky, S. (2002). Agricultural sustainability and intensive production practices. *Nature* 418(6898): 671.

Václavík, T., and Meentemeyer, R. K. (2009). Invasive species distribution modeling (iSDM): Are absence data and dispersal constraints needed to predict actual distributions? *Ecological Modelling* 220(23): 3248 – 3258.

Van der Putten, W. H., Macel, M., and Visser, M. E. (2010). Predicting species distribution and abundance responses to climate change: why it is essential to include biotic interactions across trophic levels. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 365(1549): 2025 – 2034.

Wagner, D., Papastamatiou, Y., Kosaki, R., Gleason, K., McFall, G., Boland, R., Pyle, R., and Toonen, R. (2011). New records of commercially valuable black corals (Cnidaria: Antipatharia) from the Northwestern Hawaiian Islands at mesophotic depths. *Pacific Science* 65(2): 249 – 255.

Wake, D. B., and Vredenburg, V. T. (2008). Are we in the midst of the sixth mass extinction? A view from the world of amphibians. *Proceedings of the National Academy of Sciences* 105(Supplement 1): 11466 – 11473.

Wear, S. L., and Thurber, R. V. (2015). Sewage pollution: mitigation is key for coral reef stewardship. *Annals of the New York Academy of Sciences* 1355(1): 15 – 30.

Whitcomb, L., Yoerger, D. R., Singh, H., and Howland, J. (2000). Advances in underwater robot vehicles for deep ocean exploration: Navigation, control, and survey operations. *Robotics Research*: 439 – 448.

White, K. N., Weinstein, D. K., Ohara, T., Denis, V., Montenegro, J., and Reimer, J. D. (2017). Shifting communities after typhoon damage on an upper mesophotic reef in Okinawa, Japan. *PeerJ* 5:e3573.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., and Guisan, A. (2008). Effects of sample size on the performance of species distribution models. *Diversity and Distributions* 14(5): 763 – 773.

Zaneveld, J. R., Burkepile, D. E., Shantz, A. A., Pritchard, C. E., McMinds, R., Payet, J. P., Welsh, R., Correa, A. M., Lemoine, N. P., Rosales, S., Fuchs, C., Maynard, J.A., and Thurber, R.V. (2016). Overfishing and nutrient pollution interact with temperature to disrupt coral reefs down to microbial scales. *Nature Communications* 7: Article ID 11833.

Chapter 2: The implementation of rare events logistic regression to predict the distribution of mesophotic hard corals across the main Hawaiian Islands

Published as Veazey, L. M., Franklin, E. C. Kelley, C., Rooney, J., Frazer, L. N. and Toonen, R. J. (2016) The implementation of rare events logistic regression to predict the distribution of mesophotic hard corals across the main Hawaiian Islands. *PeerJ* 4:e2189.

2.1 Abstract

Predictive habitat suitability models are powerful tools for cost–effective, statistically robust assessment of the environmental drivers of species distributions. The aim of this study was to develop predictive habitat suitability models for species in two genera of scleractinian corals (*Leptoseris* and *Montipora*) found within the mesophotic zone across the main Hawaiian Islands (MHI). The mesophotic zone (30 – 180 m) is challenging to reach, and therefore historically understudied, because it falls between the maximum limit of SCUBA divers and the minimum typical working depth of submersible vehicles. Here, I implement a logistic regression with rare events corrections to account for the scarcity of presence observations within the dataset. These corrections reduced the coefficient error and improved overall prediction success (73.6% and 74.3%) for both original regression models. The final models included depth, rugosity, slope, mean current velocity, and wave height as the best environmental covariates for predicting the occurrence of the two genera in the mesophotic zone. Using an objectively

selected theta ("presence") threshold, the predicted presence probability values (average of 0.051 for *Leptoseris* and 0.040 for *Montipora*) were translated to spatially–explicit habitat suitability maps of the main Hawaiian Islands at 25 m grid cell resolution. My maps are the first of their kind to use extant presence and absence data to examine the habitat preferences of these two dominant mesophotic coral genera across Hawai‘i.

2.2 Introduction

Consistent and pervasive deterioration of marine ecosystems worldwide highlights significant gaps in current management of ocean resources (Foley et al. 2010; Douvere 2008; Crowder & Norse 2008). One such gap is acquiring the data required for informed marine spatial planning, a management approach that synthesizes information about the location, anthropogenic use, and value of ocean resources to achieve better management practices such as defining marine protected areas and implementing harvesting restrictions (Jackson et al. 2000, Larsen et al. 2004). The creation of spatial predictive models for improved marine planning is a relatively low–cost and non–invasive technique for projecting the effects of present–day human activities on the health and geographic distribution of marine ecosystems.

Defining and managing the biological and physical boundaries of ecosystems is a complicated but essential component of marine spatial planning (McLeod et al. 2005). The heterogeneous nature of ecological datasets can require the time–intensive development of problem–specific ecosystem models (Cramer et al. 2001; Tyedmers et al. 2005). Scientists frequently use straightforward, easy–to–implement regression methods to analyze complex datasets. The development of software accessible to relative novices has contributed to the growing popularity of regression methods (e.g., Lambert et al. 2005; Tomz et al. 2003).

Here, I employ a logistic regression with rare events corrections (King & Zeng 2001) to analyze the presence and absence data of two coral genera (*Leptoseris* and *Montipora*) and, thus, develop a predictive framework for the geographic mapping of mesophotic coral reef ecosystems (MCEs) across the main Hawaiian Islands (MHI). MCEs, located at depths of 30 – 180 meters, are considered to be extensions of shallow reefs because they harbor many of the same reef species present at shallower depths, and are also oases of endemism in their own right (Grigg 2006; Lesser et al. 2010; Kane et al. 2014; Hurley et al. 2016). MCE habitats are formed primarily by macroalgae, sponges, and hard corals tolerant of low light levels (Lesser et al. 2009). Corals of genus *Montipora* colonize primarily the shallow reef zone (< 30 m), but some

species, particularly *Montipora capitata* (Rooney et al. 2010), are able to extend their settlement into mesophotic depths. Corals of genus *Leptoseris* construct extremely efficient, light–capturing skeletons that facilitate their habitation of the lower mesophotic zone (Kahng et al. 2012) and are considered to be exclusively mesophotic dwellers (Kahng & Kelley 2007).

Ecological studies in the mesophotic zone are sharply limited in contrast to the shallower photic zone more accessible by open circuit SCUBA, but steady advances in diving, computing, and remotely operated vehicle technologies continue to facilitate interdisciplinary mesophotic research (Pyle 1996; Puglise et al. 2009). Mesophotic research in Hawai‘i has been conducted primarily in the ‘Au‘au Channel, Maui, a relatively shallow, semi–enclosed waterway between the islands of Maui and Lāna‘i that is among the most geographically sheltered and accessible areas in the Hawaiian Archipelago, and, as a result, much of the existing video and photo records of MCEs are from this area. This concentration of historic surveys highlights the importance of creating a pan–Hawai‘i predictive habitat model to identify likely areas of MCEs across unexplored areas of Hawai‘i's mesophotic zone. Increasing my knowledge about the habitat preferences of the deep extensions of shallow coral species is critical given that approximately 40% of shallow (< 20 m) reef–building corals face a heightened extinction risk from the effects of climate change (Carpenter et al. 2008). Here, I model the habitat associations of mesophotic scleractinian corals because of both their intrinsic biological value as well as their potential to recolonize globally threatened shallow reef areas and serve as a refuge to mobile reef organisms (Bongaerts et al. 2010; Kahng et al. 2014).

Previous research about the environmental variables driving mesophotic scleractinian colonization in Hawai‘i suggests that distinct variation in community structure exists between the upper (30 – 50 m) and mid to lower mesophotic (50 – 180 m) depths (Rooney et al. 2010; Kahng et al. 2010; Kahng et al. 2014). Potentially influential environmental variables include photosynthetically active radiation (PAR) levels (Goreau & Goreau 1973; Fricke et al. 1987; Kahng and Kelley 2007; Kahng et al. 2010), isotherms (Grigg 1981; Kahng & Kelley 2007; Rooney et al. 2010), and hard substrate availability (Kahng & Kelley 2007; Costa et al. 2012). Rooney et al. (2010) noted that hard coral abundance declined dramatically below 100 m despite high (> 25%) availability of colonizable substrate; this sudden reduction in coral cover occurs at increasingly shallower depths across the northwestern Hawaiian Ridge and may be driven by the synchronously shallower occurrence of isotherms.

Light and temperature intensity (Jokiel & Coles 1977; Rogers 1990), physical stress (e.g., wave energy or uncontrolled tourism) (Dollar 1982; Nyström et al. 2000; Franklin et al. 2013), and availability of colonizable substrate (Jokiel et al. 2004; Franklin et al. 2013) are known drivers of shallow (< 30 m) reef coral distributions across the world. I expect that my model will capture the influence of these abiotic variables on the distribution of mesophotic corals, especially in the shallower mesophotic zone. I speculate that my model may detect unexpected drivers of *Leptoseris* distribution, particularly because *Leptoseris* is known to colonize deeper depths that bear little resemblance to shallow reefs (Lesser et al. 2009; Rooney et al. 2010). Finally, previous predictive modeling research about the drivers of Hawaiian mesophotic coral colonization identified depth, distance from shore, euphotic depth, and sea surface temperature as potentially influential environmental variables (Costa et al. 2012; Costa et al. 2015). My novel modeling approach utilizes all observational data (corals present and absent), which I believe will offer more insight into the dynamics that facilitate and inhibit coral colonization across the mesophotic zone.

2.3 Materials and methods

2.3.1 Organismal and environmental data

The Hawai‘i Undersea Research Laboratory (HURL) and the Pacific Islands Fisheries Science Center (PIFSC) provided video and photo records from MCEs in the Hawaiian Islands for my analyses. This imagery came from 19 dives conducted using submersibles, remotely operated vehicles (ROVs), autonomous underwater vehicles (AUVs), and tethered optical assessment devices (TOADs) in the ‘Au‘au Channel, Maui (13 dives) and two other geographically distinct regions: south O‘ahu (5 dives) and southeast Kaua‘i (1 dive). These dives were conducted between 2001 and 2013. I analyzed dive video using the Coral Point Count with Excel extensions (CPCe) tool (Kohler & Gill 2006) in combination with a modified PIFSC 2011 mapping protocol (PIBHMC 2015). PIFSC has used this type of combined analysis, referred to as the random five point overlay method (RFPOM), to process coral reef ecosystem benthic imagery throughout the U.S. Pacific Islands Region since August 2011, and my use of it ensures database consistency with regions processed prior to this study. The CPCe software placed five points randomly on each snapshot, which I then assessed for coral presence. If any of the five points was on coral, that observation was recorded as a “presence”. In an effort to evaluate the accuracy of RFPOM, I counted all corals in 200 randomly selected screengrabs and found that

this method misses 2.4% of coral observations recorded in these images. I categorized corals by genus, because both *Montipora* (Forsman et al. 2010) and *Leptoseris* (Luck et al. 2013) contain species complexes that remain the subject of taxonomic uncertainty which prevent us from being able to reliably identify corals to the species level from photographs.

I recorded snapshots every 30 seconds for the duration of each dive video. In addition to an existing database of 40,193 records from dives in the ʻAuʻau Channel, 3517 new snapshots were collected from the additional dives across south Oʻahu and Kauaʻi (Fig. 1). Of these 43,710 total images, 20,980 were discarded because either: 1) crucial observational data were absent, 2) they were redundant due to an extended stationary period, or 3) they fell outside the study depth range of 30 – 180 m. Of the remaining 22,714 records, I analyzed 2757 unprocessed images using the RFPOM (Table 1).

I selected my environmental covariates, listed in Table 2, based on the sufficiency of the data and the potential significance of each variable as an indicator of hard coral habitat suitability (e.g., Dolan et al. 2008; Rooney et al. 2010; Costa et al. 2012). I defined summer and winter seasons as May – September and October – April, respectively (Kay 1994; Rooney et al. 2010).I delineated significant wave height estimates and mean current velocities by season and direction. I extracted and averaged significant wave height data from 144 days per season of twenty–four hourly PacIOOS Simulating WAves Nearshore (SWAN) regional wave models estimated values for 2011 – 2015 (see website: http://oos.soest.hawaii.edu/las/). Mean current velocity values were available from 0:00 – 21:00 every three hours for all months from 2013 – 2015; for each season and direction, 48 mean current velocity values were extracted and averaged from the PacIOOS Regional Ocean Modeling System (see website: http://oos.soest.hawaii.edu/las/). This model has a 4 km horizontal resolution with 30 vertical levels across seafloor terrain. I sourced monthly MODIS Aqua chlorophyll *a* averages for the year 2012 from the NOAA PIFSC OceanWatch Live Access Server (see website: http://oceanwatch.pifsc.noaa.gov/). Using the Morel (2007) method, we applied the following cubic polynomial equation to obtain logged euphotic depth:

$$\log_{10} Z_{eu} = 1.524 - 0.436x - 0.0145\,x^2 + 0.0186x^3, \qquad (1)$$

where x represents the measured Chlorophyll *a* concentrations (mg/m^3) at sea surface. Euphotic depth is the depth at which the level of photosynthetically active radiation (PAR), a limiting factor for many heterotrophic mesophotic corals, is at 1% of surface PAR. In total, we used 14

environmental predictor variables to shape our model (Table 2) (Supplementary material, Figs. A1 – A5).

The spatial resolution of the bathymetry data was 50 m x 50 m for all islands. We resampled the bathymetry raster to a cell size of 25 x 25 m consistent with a conservatively estimated $\pm$ 25 m positioning error margin observed at a depth of ~800 m. We estimated an average camera swath value of 3.24 m (range 2.45 – 4.54 m) based on previous measurements from 19 still image screenshots taken when the submersible was located at different heights above the seafloor. Our geopositional error for the images is $\pm$ 5 m and we can expect that the location data are within a circle with a 10 m diameter. Our observation sampling area is projected out from the location area a distance of $\leq$ 5m. Addition of a conservative 5 m observation area buffer to the location error area produces an observational data position of $\pm$ 20 m from the given coordinates of a data point.

We removed all subsampling within cells due to slight variations in camera angles or vessel speed through a point–to–raster conversion. We categorized all cells with $\geq$ 1 presence observation as "present" cells and all cells with only absence observations as "absent" cells. This removal of multiple observations within the same 25 x 25 m pixel effectively eliminates pseudoreplication within the data. We used ArcToolbox and the Benthic Terrain Modeler Toolbox to calculate slope, curvature, rugosity, and aspect (compass direction) values (Wright et al. 2012). We performed a spatial join based on proximity to observation point data to assign values for surface Chlorophyll *a* concentration, mean current velocities, distance from shore, and significant wave heights.

Regression methods

In describing the relationship between a response variable and one or more predictor variables, we use a logistic regression model because the response variable is dichotomous (Hosmer and Lemeshow 2004). The ordinary logistic regression (OLR) model is defined as:

$$\theta = \text{expit}(\mu) = \frac{1}{1+\exp(-\mu)} \ , \qquad (2)$$

where θ is the probability that the species of interest is present $(y = 1)$, and $1-\theta$ is the probability it is absent $(y = 0)$.The logit function is the inverse of the expit function, and

$$\text{logit}(\theta) = \mu = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n \qquad (3)$$

is the linear sum of predictor variables, $x_1, x_2, ..., x_n$, with intercept β_0 and regression coefficients $\beta_1, \beta_2, ..., \beta_n$. In the language of generalized linear models (GLM), OLR is said to have the logit function as its link function and the expit function as its inverse link function. Logistic regression provides a straightforward, meaningful interpretation of the relationship between a dichotomous dependent variable y and a set of predictor variables (Allison 2001).

Despite the popularity of OLR, it may yield extremely biased results when an imbalance exists in the proportion of the response variable data (e.g., such as in our case, when $y = 0 >> y = 1$) (Van Den Eeckhaut et al. 2006). King and Zeng (2001) coined the term "rare events logistic regression" to describe their corrective methodology in dealing with unbalanced binary event data:

1. The first step requires the selection of a representative sample. Though researchers generally prefer to work with more uniform response data (e.g., Liu et al.2005), selection of an unusually high proportion of the rare event (in this case, $y = 1$) to "balance" the dataset and increase θ estimates will yield nonsensical results. We divided the data in half to create our training and testing datasets and checked that each set of observations had an approximately equal proportion ($\bar{y}$) of presence observations to better reflect the "true state" of the full dataset.
2. The second step rectifies any bias that might be introduced when dividing the dataset. This prior correction on the intercept (β_0) can be calculated as:

$$\hat{\beta}_0 = \tilde{\beta}_0 - \ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{\bar{y}}{1-\bar{y}}\right)\right]; \qquad (4)$$

   here, $\hat{\beta}_0$ is the corrected intercept, $\tilde{\beta}_0$ is the uncorrected intercept, τ is the true proportion of 1s in the population; and $\bar{y}$ is the observed proportion of 1s in the training sample.
3. The third step rectifies any underestimation of the probabilities of the independent variables $\beta_{1...n}$ from the substitution of the intercept value, obtained as:

$$P(y_i = 1) = \tilde{\theta}_i + C_i \ , \qquad (5)$$

   where the correction factor C_i is given by:

$$C_i = (0.5 - \tilde{\theta}_i)\tilde{\theta}_i(1 - \tilde{\theta}_i)XV(\tilde{\beta}_i)X' \,, \qquad (6)$$

where X is a $1 \times (n+1)$ vector of values for each independent variable β_i, X' is the transpose of X, and $V(\tilde{\beta}_i)$ is the variance covariance matrix. We obtained the improved probability estimates through estimation of β_i via $\tilde{\beta}_i$, thereby considered "mostly" Bayesian (King and Zeng 2001). Our priors in this case would be uninformative, which means that we lack sufficient knowledge to estimate the probability distributions of our data and our parameter of interest, θ. This is often the case when working with sparse ecological datasets. As the uninformative prior for a regression coefficient with domain $(\infty, -\infty)$ is uniform, a full Bayesian estimation with uninformative priors is equivalent to a traditional logistic regression. Therefore, this correction is effectively a correction to the approximate Bayesian estimator, and its addition improves our regression by lowering the mean squared error of our estimates. We implemented this rare events logistic regression using the 'Zelig' package run in R (Imai et al. 2008, Choirat et al. 2015).

We constructed a correlation scatterplot matrix per coral genus to observe correlation levels between all variables. In choosing which highly correlated variables to exclude from the analyses, we followed the criteria outlined by Dancey and Reidy (2004) and Tabachnick and Fidell (1996), who suggest a cutoff correlation value of 0.7. Only mean significant wave height parsed by season consistently overreached this threshold; the covariate that was least correlated with the response variable was removed. We excluded predictors that lacked a clear distribution pattern or correlated minimally (< 0.05) with the response variable.

One of the more studied habitat preferences of *Leptoseris* and *Montipora* is the influence of depth on their distribution (Rooney et al. 2010, Costa et al. 2012, Kahng et al. 2010). Increasing depths often correlate with greater distance from shore. The inclusion of squared terms (e.g., $x_2 = x_1^2$) in our regression equation $\text{expit}(\theta) = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n$ permits the logistic curve to reflect the bell curve shape expected in plotting the distribution of these animals across a range of depths or distance from shore. In order to account for these trends, we added Depth Squared and Distance Squared as potential variables for consideration in our final model. As depth or distance increases, its square increases even more rapidly, allowing the squared term to eventually dominate and "pull down" the probability curve.

We withheld 50% of our information per genus as testing (i.e., validation) data. Using the remaining 50% (our training data), we performed the rare events corrected logistic regression described above. Using an exhaustive iterative algorithm (Calcagno and Mazancourt 2010), we modeled all possible combinations of included covariates. We ranked models using the corrected Akaike information criterion (AICc) (Hurvich and Tsai, 1989), which is considered an excellent comparative measurement of model strength, especially for sparse datasets. For both genera, the models with the lowest (lowest = best) AICc scores were lower than the "second best" AICc scores by at least 2 (i.e., Δ AICc $\geq$ 2), indicating strong preference for the best model (e.g., Hayward et al. 2007).

In an ideal and unrealistic study, all biotic and abiotic components of a model would be homogenous and evenly distributed across a sampling space. Our sampling design includes overlapping submarine dive tracks and the inherent heterogeneity of the marine environment, which could problematically violate our model's underlying assumption regarding the independence of our biological and environmental data. We removed all instances of pseudoreplication (multiple observations in one grid cell) when we assigned each grid cell to a category of "corals present" or "corals absent". After we removed subsampling within our observational data, we checked for the presence of clustering, or spatial autocorrelation, within these data. Uncorrected spatial autocorrelation between observational data points confounds and undermines any biological inferences drawn from model predictions.

We checked small–scale, local spatial autocorrelation using Geary's C statistic (Geary 1954), based on the deviations in the responses of observation points with one another:

$$C = \frac{n-1}{2S_0} \frac{\sum_i \sum_j w_{ij}(x_i - x_j)^2}{\sum_i (x_i - \bar{x})^2}. \qquad (7)$$

Here, x is the variable of interest, i and j are locations (where $i \neq j$), w_{ij} represents the components of the weight matrix, and S_0 is the sum of the components of the weight matrix. Geary's C ranges from 0 (maximal positive autocorrelation) to 2 for high negative autocorrelation. In the absence of autocorrelation, its expectation is 1 (Sokal and Oden 1978).

We also examined global spatial autocorrelation using Moran's I statistic, which measures cross–products of deviations from the mean (Moran 1950):

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}. \qquad (8)$$

Moran's I values generally range from –1 to 1, with 0 as the expectation when no spatial autocorrelation is present.

We also verified the spatial independence of our observational point data using a semivariogram, which is a graphical method of quantifying spatial correlation in a set of points (Figs. 2 – 3). We selected our theoretical semivariogram to fit the empirical semivariance using the ordinary least squares (OLS) method (Jian et al. 1996, Kendall et al. 2005). The spherical model had the best quantitative fit based on OLS estimates (Table 3). For each dataset, the low thresholds at which semivariance stopped increasing indicated the almost complete absence of spatial autocorrelation for each genus.

2.3.3 Model assessment

Evaluation of the rare events logistic regression model output is more complicated than for the typical linear model. For example, R^2 values, although calculated, have little applicability to logistic regressions and are therefore ignored (Menard 2000, Peng et al. 2002). Sample size and selected threshold largely influence the results of the Hosmer and Lemeshow goodness–of–fit test (Hosmer et al. 1997). Accordingly, we use model classification accuracy as a second measure of goodness–of–fit (in addition to Δ AICc). We want to maximize true positives (TP) and true negatives (TN) while minimizing false positives (FP) and false negatives (FN). The sensitivity–specificity sum maximization approach (Cantor et al. 1999) therefore maximizes

$$SS_{\max} = \frac{TP}{TP + FN} + \frac{TN}{TN + FP}, \qquad (9)$$

which is equivalent to finding the point on the ROC (receiver operating characteristics) curve at which the tangent slope is 1, indicating the optimal cutoff point at which "cost" (here, the number of FN and FP) and "benefit" (the number of TN and TP) is balanced. We chose this technique because we aim to identify regions devoid of hard corals as well as regions deemed potentially suitable for habitation.

ROC curves plot the true positive test rate against the false positive test rate across different theta cutoff points (Hadley and McNeil 1982). We calculated values for sensitivity and

specificity for threshold increments of 0.005 $\pm$ 1 standard deviation of the rounded mean for each model. Because each theta threshold value varied based on the genus and model, the threshold–independent area under the curve (AUC) test statistic best reflects the predictive accuracy of the model.

In addition to creating ROC curves, we also took into account the overall prediction success of each model, given as:

$$OPS = \frac{TP + TN}{TP + TN + FP + FN}. \qquad (10)$$

Overall prediction success is a measure of total correct classification of both present and absent observations. While this is a good final assessment of model classification error, consideration of the prediction success alone is not a viable evaluation method when binary data is highly imbalanced, as a value given by this method may primarily represent model success in identifying the most common observation type (Fielding and Bell 1997). We plotted our sensitivity and specificity values on a ROC curve to show how each model performed relative to chance (Fig. 4). All models fall in the range $0.7 \leq \text{AUC} < 0.9$, which indicates good discrimination and reliability of model predictions (Hosmer and Lemeshow 2004).

We also created maps of individual and summed predicted occurrence probabilities of both coral genera across the main Hawaiian Islands and ran a hotspot analysis using the ArcGIS Getis–Ord G_i* Hotspot Analysis tool. We constructed a polygon fishnet composed of 1 x 1 km cells which encompassed all islands. We summed each 25 x 25 m raster cell value for probability of *Leptoseris* occurrence and probability of *Montipora* occurrence. We performed a spatial join of raster cell values within each polygon for an average value of summed probabilities. The Getis–Ord G_i* statistic identifies clusters within these polygons that display values higher in magnitude than random chance would permit. The Getis–Ord local statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \overline{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{1}{n-1}\left[n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j} \right)^2 \right]}}. \qquad (11)$$

Here, $w_{i,j}$ represents the spatial weights between features i and j; n represents the total number of features; x_j is the attribute value for feature j; $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} x_j$; and $S = \sqrt{\frac{1}{n}\sum_{j=1}^{n} x_j^2 - (\overline{X})^2}$.

2.4 Results

Geary's C test statistic is a measure of local (small–scale) spatial autocorrelation; in the absence of correlation, 1 is the expected value of Geary's C. Moran's I is a measure of global (large–scale) spatial autocorrelation; in the absence of correlation, a value of 0 is expected for the Moran's I test statistic. For my *Leptoseris* dataset, Geary's C = 0.990; for my *Montipora* dataset, Geary's C = 0.996. For my *Leptoseris* dataset, Moran's I = 0.006; for my *Montipora* dataset, Moran's I = 0.003. These values do not indicate any local clustering or global spatial autocorrelation within either dataset. I observed negligible levels of autocorrelation up to ~100 m for *Montipora* (Fig. 3).By ensuring that spatial autocorrelation is not present in my data, I do not violate the assumption that my response data are independently observed, which enables me to draw robust conclusions about the ecological factors influencing the distribution of these coral genera within the mesophotic zone across the MHI.

The OLR covariate coefficients were modified using the rare events corrections proposed by King and Zeng (2001), resulting in a change in predictive power (Table 4). Rare events corrected models usually performed better than the uncorrected models, in terms of improved specificity and prediction success. My sensitivity values for both corrected models were slightly lower than the corresponding OLR sensitivities, but in each case, specificity and prediction success values were improved. Additionally, standard errors of the coefficient estimates were lower for corrected models than for uncorrected models (Tables A1– A4).

Leptoseris corals inhabit mesophotic regions with high slope and rugosity values, high to moderate perennial current flow, and their occurrence peaks around 100 m (Table A3, Figs. A6 – A10). *Montipora* corals peak in occurrence around 60 m and colonize regions less exposed to high energy winter swells (Table A4; Figs. A11 – A12). Predicted presence probability values (θ) averaged 0.051 for *Leptoseris* and 0.040 for *Montipora* models in the validation data (Figs. 5 – 6). These values agree well with the actual presence frequencies in that data (0.052, 0.042). To better interpret these realistically low theta values, I chose a theta threshold to transform the probability estimates to presence/absence values. This is standard practice when examining the

results of a rare events logistic regression, but less common when performing OLR (Liu et al. 2005; Bai et al. 2011). Objective selection of a theta threshold on a per–model basis is more scientifically sound than, for example, an arbitrary assignment of 0.5 (Cramer 2003). The transformed model is valid if a threshold value yields a high percentage of correctly classified observations and a low number of FP and FN observations (Gobin et al. 2001). I selected an appropriate threshold for each model (Table 4) in order to maximize SS_{max} (Liu et al. 2005).

My final hotspot maps show the results of my analysis for *Leptoseris, Montipora,* and both genera combined across all islands (Figs. 7 – 9). I show hotspots of habitat suitability for both coral genera in red for areas of highest suitability and blue for areas of lowest suitability. I identify a cell as a hotspot when the sum of its value and the values of its nearest neighbors is much higher or lower than the mean over all cells. When the local sum of a cluster is very different from the expected value, a statistically significant hotspot is identified (G_i* statistic $\geq$ 1.96 or G_i* statistic $\leq$ –1.96). Neither genus clearly dominated the summed probabilities hotspot identification across any of the islands. Large *Leptoseris* hotspots were identified in southwest Moloka‘i, northeast O‘ahu, west Hawai‘i, and the central ‘Au‘au Channel. *Montipora* hotspots were identified in east Ni‘ihau, southwest Kaua‘i, west and south O‘ahu, west Hawai‘i, and the central ‘Au‘au Channel.

2.5 Discussion

In this study, I used logistic regression with rare events corrections to predict the habitat preferences of two dominant scleractinian coral genera across the entire mesophotic zone surrounding the main Hawaiian Islands. The habitat preferences of *Montipora* in the mesophotic zone appear distinct from those of *Leptoseris*. *Montipora* prefers the middle mesophotic zone (50 – 80 m) of reefs less exposed to high–energy winter swells. *Leptoseris* prefers steep, rugose slopes and the lower mesophotic zone (> 80 m) in regions of high year–round current flow.

2.5.1 Important environmental covariates

Predicted *Montipora* presence peaks at about 60 meters (median occurrence probability = 7.5%); *Leptoseris* presence peaks at about 100 meters (median occurrence probability = 7.5%). These predictions are consistent with the inferences of Rooney et al. (2010), which separates mesophotic reefs into three distinct depth sections: upper (30 – 50 m), branching/plate dominated (50 – 80 m), and *Leptoseris* dominated ($\geq$ 80 m). The depth at which suitability peaks for

Leptoseris occurs at a range where steep ridges and drop–offs are plentiful in my study region, and therefore the mean preferred depth may be prone to slight overestimation.

In addition to depth, four environmental covariates appeared to influence the distribution of *Leptoseris*: rugosity, slope, summer mean current velocity (northward), and winter mean current velocity (eastward). Scleractinians easily colonize environments that are relatively calm and rugose due to the larger amount of available surface area, and this positive correlation was reflected in my model. *Leptoseris* habitat preference was also positively associated with slope, which was not observed for *Montipora.* Corals that inhabit the upper mesophotic zone may be more susceptible to damage from debris displaced by high wave energy, and are therefore less likely to colonize steep slopes (e.g., Harmelin–Vivien & Laboute 1986; Bridge & Guinotte 2013). The deeper distribution of *Leptoseris* may protect it from damage related to wave intensity, allowing it to colonize slopes (e.g., White et al. 2013). Another possibility is that the model is picking up drop–offs from masses accreted during the last glacial maximum. These steep drop–offs are present between 90 – 120 m in the *Leptoseris*–dominated lower mesophotic zone (Yokoyama et al. 2001; Webster et al. 2004).

Leptoseris also favors well–flushed areas exposed to year–round moderate current flow (i.e., up to 0.3 m/s). The plate–like morphology of *Leptoseris* corals effectively boosts sunlight capture by its symbiotic zooxanthellae and zooplankton capture by the corals themselves, but it also makes the coral vulnerable to smothering by sediment accumulation (Bak et al. 2005, Bongaerts et al. 2010; Marcellino et al. 2013). The success of *Leptoseris* corals in areas of moderate current flow may be related to the improbability of sediment settlement and accumulation. While the model did not capture the same effect of current flow on *Montipora* distribution, I recognize that the morphology of some *Montipora* species is extremely similar to that of *Leptoseris.* I do not expect either genus to readily colonize highly turbid regions, especially given that certain species of heterotrophic *Montipora* are thought to exploit strong currents to meet their energy requirements (Grottoli et al. 2006; Rooney et al. 2010).

Substrate hardness, a variable known to influence coral colonization, was notably absent from each model. Substrate hardness values were derived from acoustic backscatter imagery readings. The base resolution of these readings (50 x 50 m) was not sufficiently detailed for purposes of this analysis. I noted plentiful coral colonization along larger surfaces like lava fingers, the hardness of which would be detectable by backscatter surveys, as well as across

small rock fragments strewn across a sand flat, which would be obscured by the softness of the surrounding benthos. I can conclude that measurements of benthic hardness are not detailed enough for predictive modeling purposes at a 25 x 25 m resolution.

I emphasize that the purpose of this study was to build a pan–Hawai‘i predictive habitat map for two dominant coral genera within the mesophotic zone. Because the scope of this study included all MHI, my work was constrained by the coarseness of available full–coverage environmental data. As I build on this analysis, I plan to use my maps to identify areas of interest for further study at higher resolution and to include additional variables currently only available in certain regions, such as light intensity and temperature at depth. For example, my predictive and hotspot maps identify Penguin Bank (southwest Moloka‘i) as particularly suitable for *Leptoseris* colonization, which has not been verified by video or photo records. High resolution backscatter data (1 x 1 m) exist for this region, and incorporation of these data into new analyses of subsets of my study area may refine my conclusions.

2.5.2 Error sources and model reliability

I examined two types of error (false negatives and false positives) and analyzed my models without giving preference to either one. This approach is widely accepted as the best method of overall error minimization (e.g., Liu et al. 2005; Fielding & Bell 1997). Rare events corrected models for both *Leptoseris* and *Montipora* achieved levels of specificity and sensitivity well above the null, indicating good predictive power. Additionally, both models attained about 74% overall prediction success. I assumed coral detectability was constant across the study region and that I can therefore consider the true absence observations to be reliable indicators of a potentially unsuitable habitat for corals. For each genus, the model tended to slightly over–predict presence observations; large numbers of false positives lowered sensitivity values. This is inevitable in the analysis of severely imbalanced or sparse binary data; the ongoing addition of presence observations to the dataset will improve overall model classification accuracy.

While the consistent identification of southern coastal areas as suitable is reliable, the comparatively infrequent selection of northern coasts is likely due to the source of the model–building observations. The vast majority of mesophotic exploration has been along southern coastlines, which is often where waters are calmest in Hawai‘i. It is speculated that because mesophotic corals are more shielded from winter long–period wave energy than their shallow water counterparts, they are able to flourish at depth along northern coastlines (Grigg 1998;

Rooney et al. 2010). The addition of data sourced from northern expeditions would likely improve predictive power of the model across north–facing coastlines (Alin 2010).

I acknowledge that the original data were not collected in a standardized fashion (e.g., variation in vessel traveling speed or differences in data collection vessel and/or quality). My careful exclusion of overlapping observation points within each 25 x 25 m rectified this sampling design flaw as much as possible and eliminated pseudoreplication.

2.5.3 Distinctions between coral genera

My *Montipora* model was simpler than the *Leptoseris* model in that the only variable included other than depth was winter significant wave height. Though uncertainty was highest at lower values of significant wave height, *Montipora* demonstrated a preference in colonizing habitats that experience lower significant wave height during winter. This preference contrasts with *Montipora* species in shallow waters that were more likely to be observed in higher wave height environments (Franklin et al. 2013). This likely influenced the inability of the model to identify any suitable habitat around Niʻihau, where the average winter significant wave height equaled 1.78 meters, almost double the mean significant wave height of my model training data (0.91 m). Though mesophotic corals are generally thought to be exempt from the growth limitations faced by shallow water corals in regions of high wave energy, prolonged wave intensity has been shown to negatively affect the colonization of upper mesophotic scleractinians, especially in sloping areas prone to debris avalanches (Bridge & Guinotte 2013; Kahng et al. 2014). Continuation of this work might include a more in–depth examination of the relationship of this coral genus with the combined effects of slope of available substrate and exposure to wave energy.

I found no records of *Montipora* presence when processing my Oʻahu dataset, which probably contributed to the very low predicted mean probability of *Montipora* occurrence there (0.1%). I believe this is due in part to the sampling pattern across south Oʻahu; I recorded 62.3% of all observations processed for this region at a depth of 75 m or greater. *Montipora* prevalence is greater in the upper–to–middle mesophotic zone, and the relative deepness of the Oʻahu dives likely influenced their nonappearance in this portion of the dataset. I emphasize that the dearth of *Montipora* observations around Oʻahu is an artifact of the dataset I used to construct my model; *Montipora* corals have been observed in mesophotic depths across Oʻahu (e.g., Fig. 4b; Rooney et al. 2010).The mean significant wave height across the mesophotic zone was lower across the

southern and western coasts (1.50 m) than that observed across the northern and eastern coasts (2.37 m) of the island. As at Niʻihau, I assume that this high northern and eastern average height, coupled with the absence of *Montipora* presences in Oʻahu in the training dataset, greatly impacted my model's ability to detect areas of suitable habitat around the island. The results of my Getis–Ord G_i* Hotspot Analysis corroborate the findings of Costa et al. (2015), who used Maximum Entropy software to predict the highest occurrence probability of *Leptoseris* and *Montipora* in the middle and mid–coastal ʻAuʻau Channel, respectively (Costa et al. 2015).

The factors influencing the distribution of coral species in shallow and mesophotic habitats differ. One of the fundamental drivers of the occurrence and abundance of coral species on shallow reefs in Hawaiian waters is wave stress (Dollar 1982; Grigg 1983; Franklin et al. 2013). Given the depth range of MCEs, wave stress is unlikely to serve as a direct influence on coral occurrence but may provide secondary effects as wave events lead to debris reaching MCEs (Kahng 2014). Furthermore, the decoupled effects of environmental drivers on shallow and mesophotic zones extend between the islands. In shallow reef communities *Montipora* species become relatively more dominant from Hawaii Island to Niʻihau (Franklin et al. 2013), but appear to peak in occurrence in the mesophotic zone of Maui Nui. While strong environmental drivers influence the distributions of shallow corals, the occurrence patterns of mesophotic corals may reflect a more stable environment with an increased influence of biotic factors such as interspecific competition in a habitat zone with limited light and space resources available.

2.6 Conclusions

I implemented a rare events corrected logistic regression to determine the most influential environmental predictors of *Montipora* and *Leptoseris* colonization in the mesophotic zone. Habitat preference differences between these genera appear distinct and multi–faceted. *Montipora* thrives in the middle mesophotic zone in areas sheltered from high intensity winter swells, while *Leptoseris* tends to colonize steep, rugose, well–flushed areas in the lower mesophotic zone. Improved understanding of the distribution of mesophotic corals will enable resource managers to propose the construction of seafloor power cables and other offshore infrastructure in areas less likely to contain coral communities. Results will likewise facilitate efforts to protect these communities by supplementing scientific dive planning and strategies for conservation, such as marine spatial planning.

2.7 Acknowledgments

Funding for this research was provided by the NOAA Coral Reef Conservation Program grant no. NA14NOS4820092. I thank the many staff members of HURL, PIFSC, and SOEST who helped acquire and process my data. I thank Dr. David Hondula (Researcher at Arizona State University) for his dedicated GIS instruction. I dedicate this chapter to the memory of my *PeerJ* manuscript coauthor, Dr. John Rooney. *A hui hou*, John.

2.8 Figures

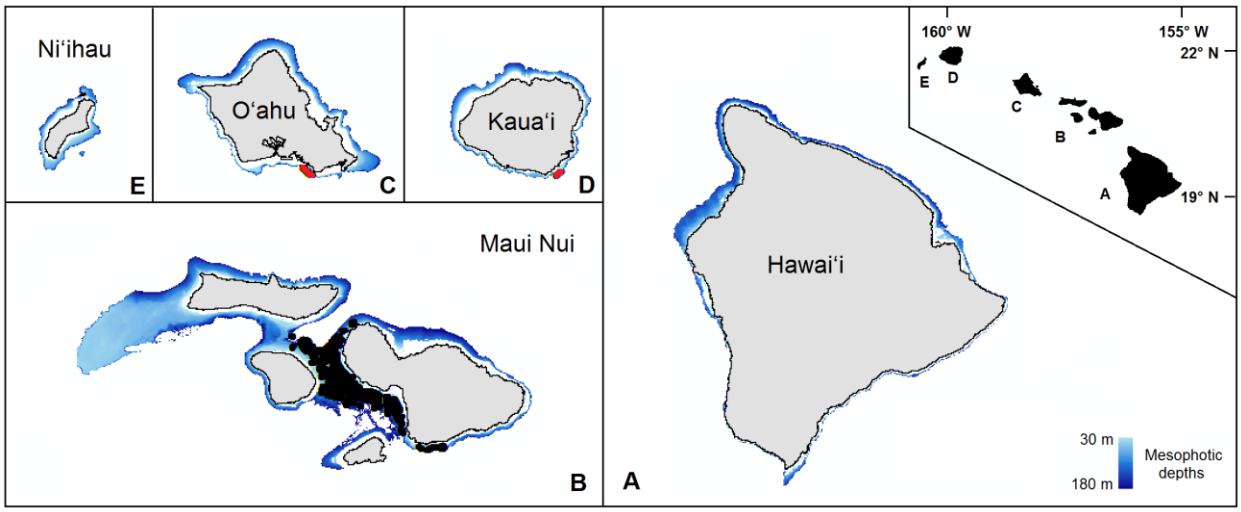

Figure 1. The study domain, demarcated in blue, encompasses the mesophotic zone (30 – 180 m in depth) of the lower main Hawaiian Islands. Black circles are the observations from the pre-existing Maui Nui dataset. Red circles are the previously unprocessed observations in south Oʻahu and southeast Kauaʻi.

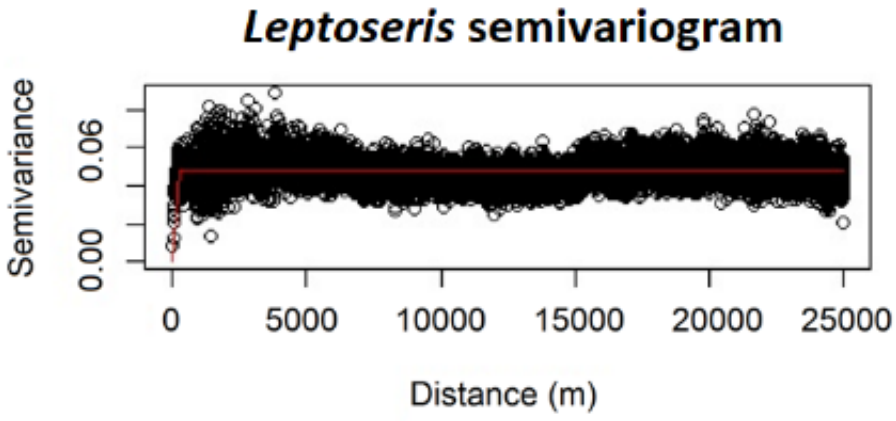

Figure 2. Modeled spherical semivariogram for *Leptoseris*.

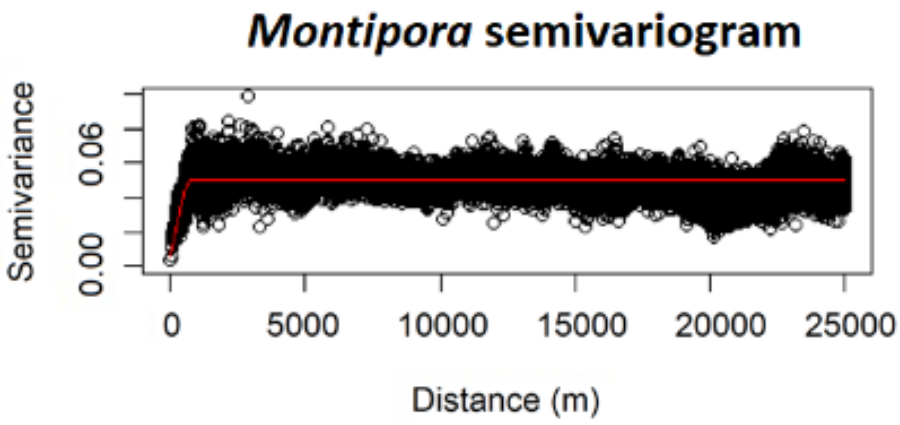

Figure 3. Modeled spherical semivariogram for *Montipora*.

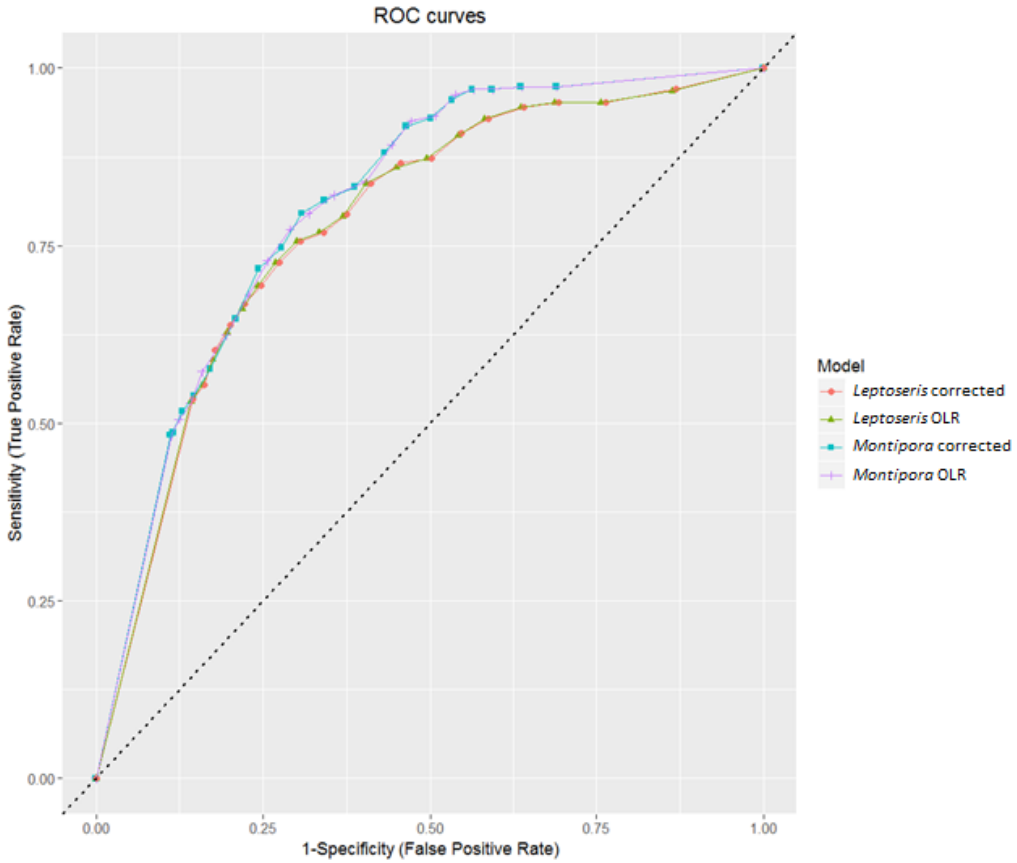

Figure 4. ROC curves for all models. AUC values for all models fall in between 0.7 and 0.9, which indicates predictive reliability. The dashed line from (0, 0) to (1, 1) indicates the null threshold at which model performance is considered unacceptable (< 0.5).

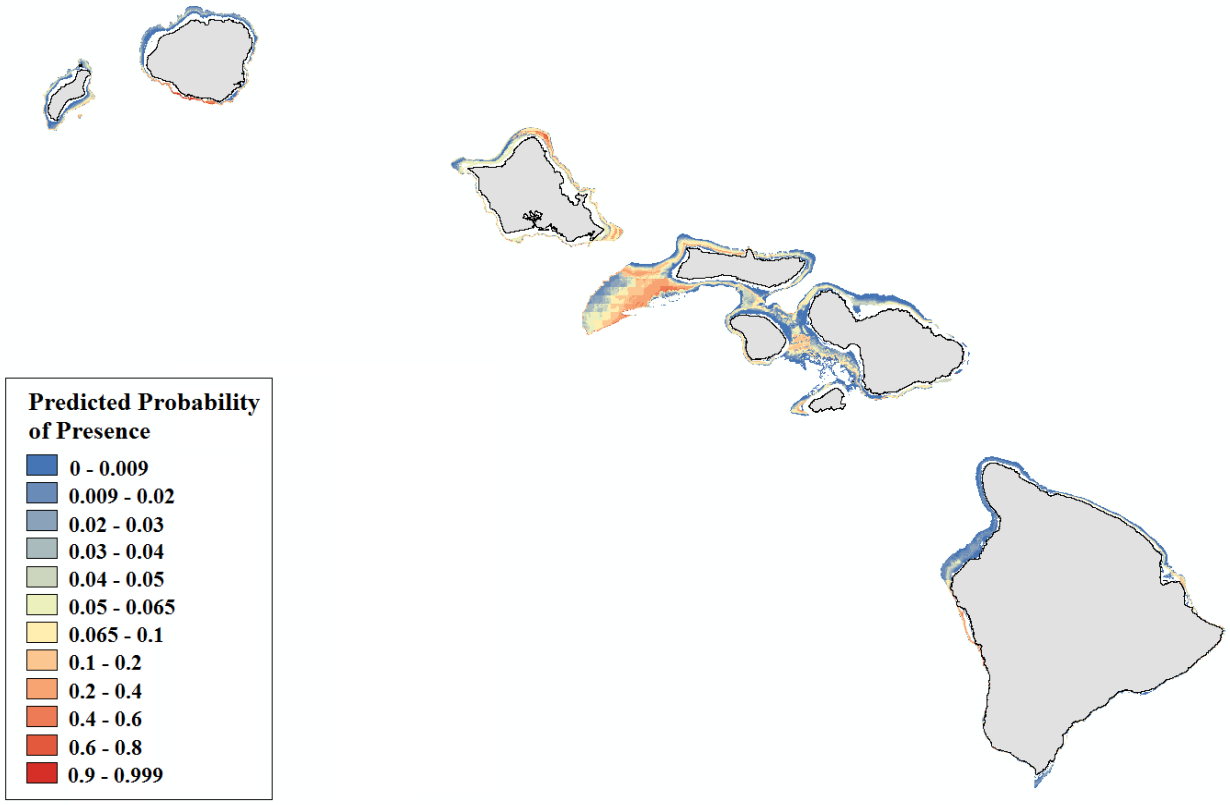

Figure 5. Modeled area of suitable habitat for *Leptoseris*. Probability of presence is depicted along a color gradient ranging from red (1; most suitable) to blue (0; least suitable).

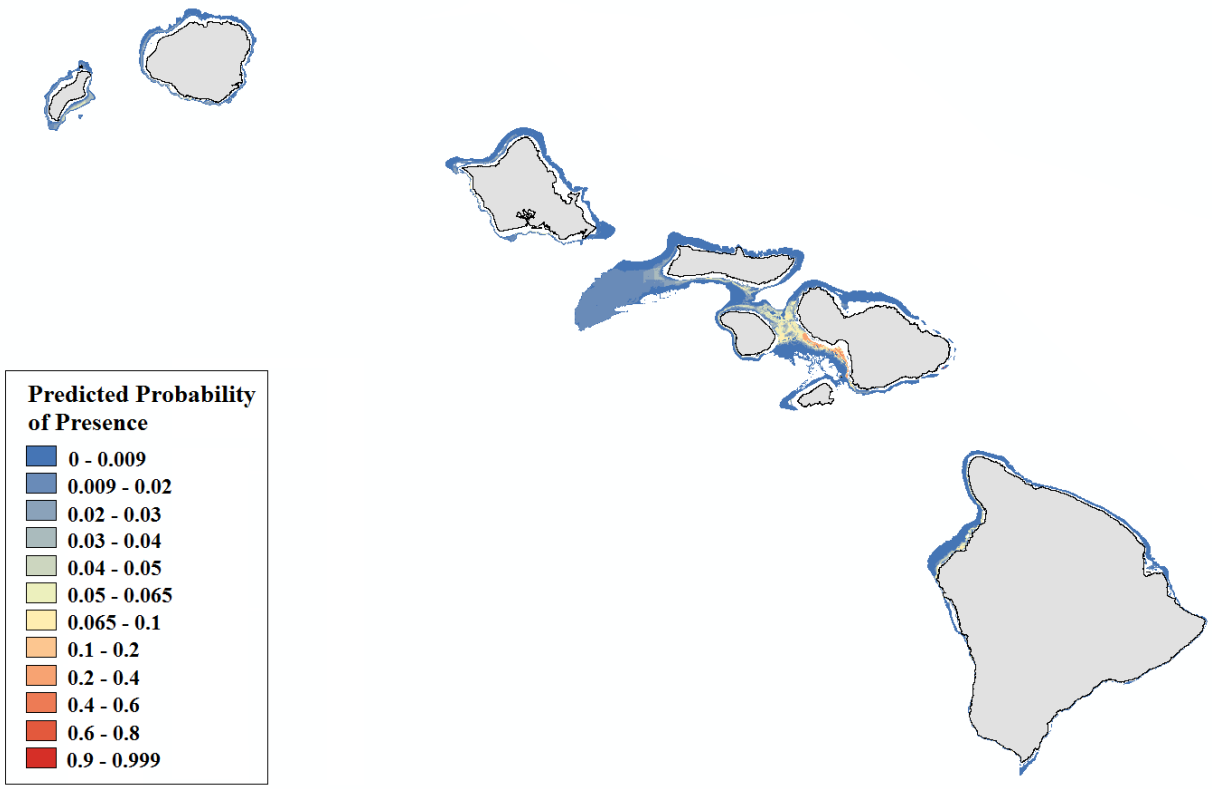

Figure 6. Modeled area of suitable habitat for *Montipora*. Probability of presence is depicted along a color gradient ranging from red (1; most suitable) to blue (0; least suitable).

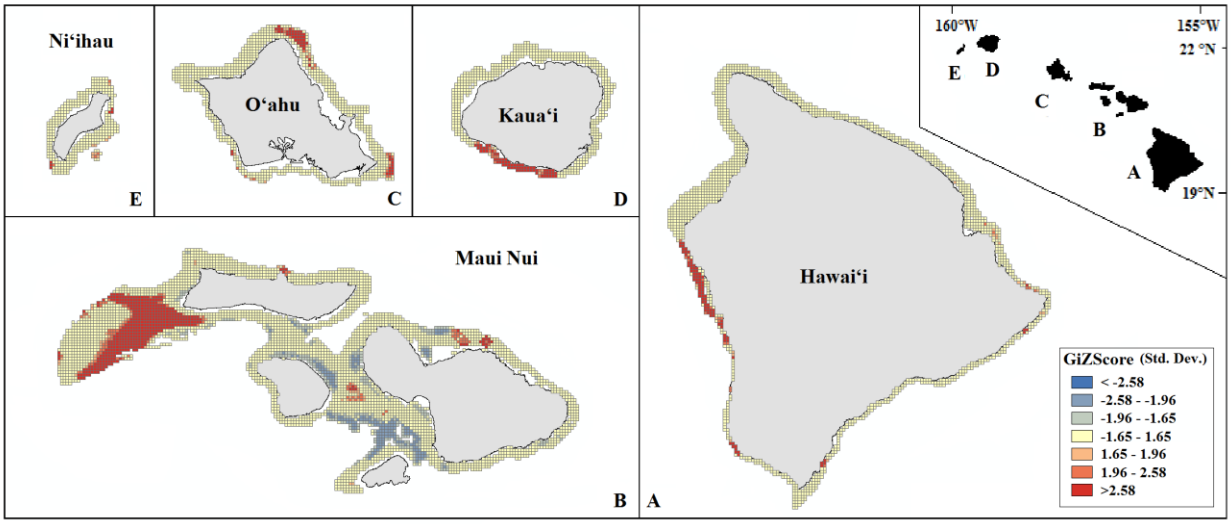

Figure 7. Mapped result of my Getis-Ord Gi* hotspot analysis performed for probability

estimates of *Leptoseris* occurrence. A significant hotspot is < -1.96 or > 1.96; here, all hotspots are shown in red (> 1.96) or blue (< -1.96).

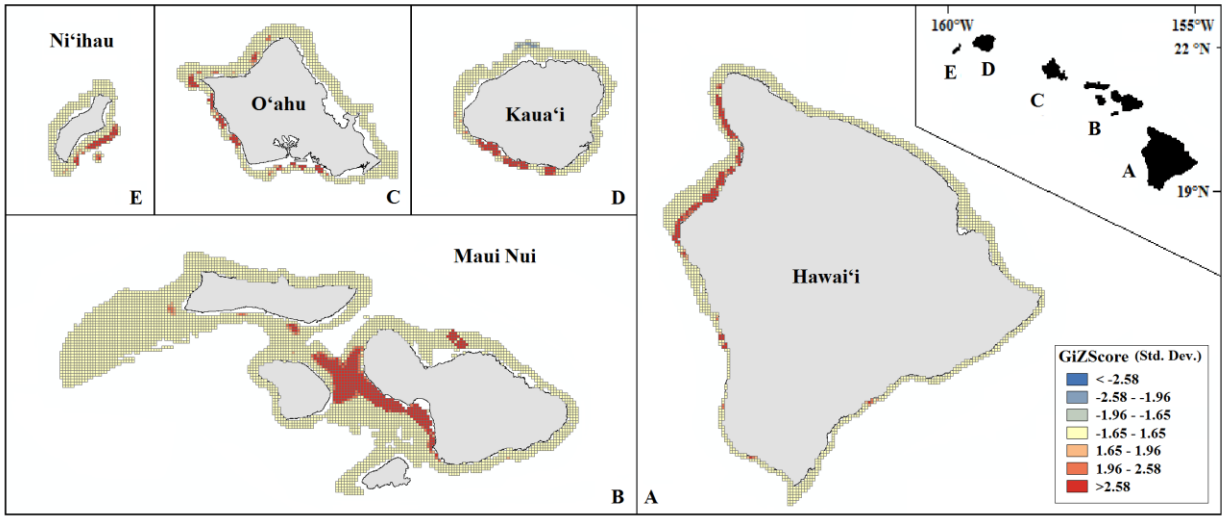

Figure 8. Mapped result of my Getis-Ord Gi* hotspot analysis performed for probability estimates of *Montipora* occurrence. A significant hotspot is ≤ -1.96 or ≥ 1.96; here, all hotspots are shown in red (≥ 1.96) or blue (≤ -1.96).

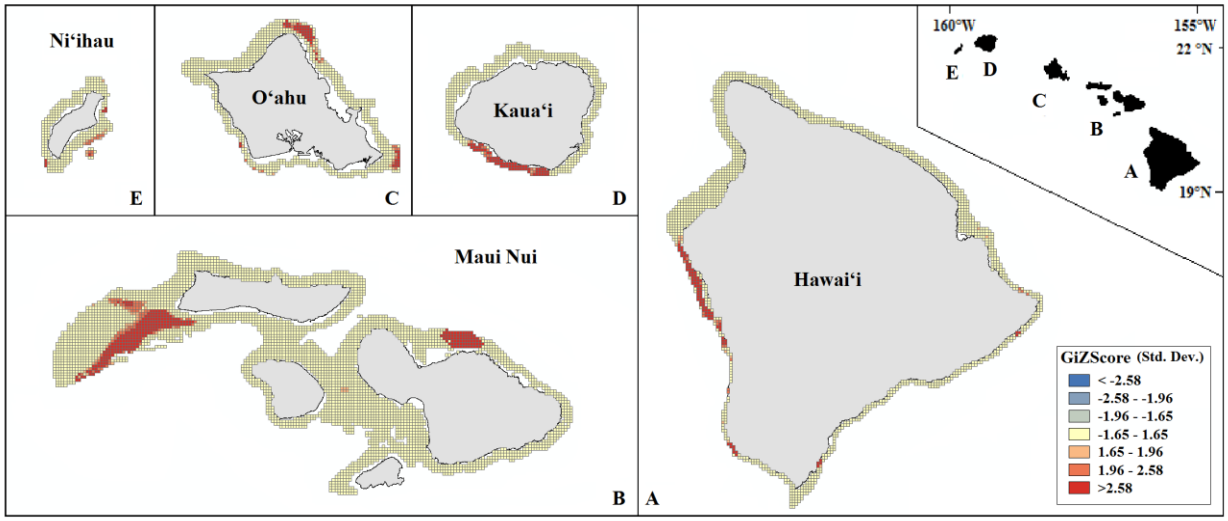

Figure 9. Mapped result of my Getis-Ord Gi* hotspot analysis performed for summed probability estimates of *Leptoseris* and *Montipora* occurrence. A significant hotspot is ≤ -1.96 or ≥ 1.96; here, all hotspots are shown in red (≥1.96).

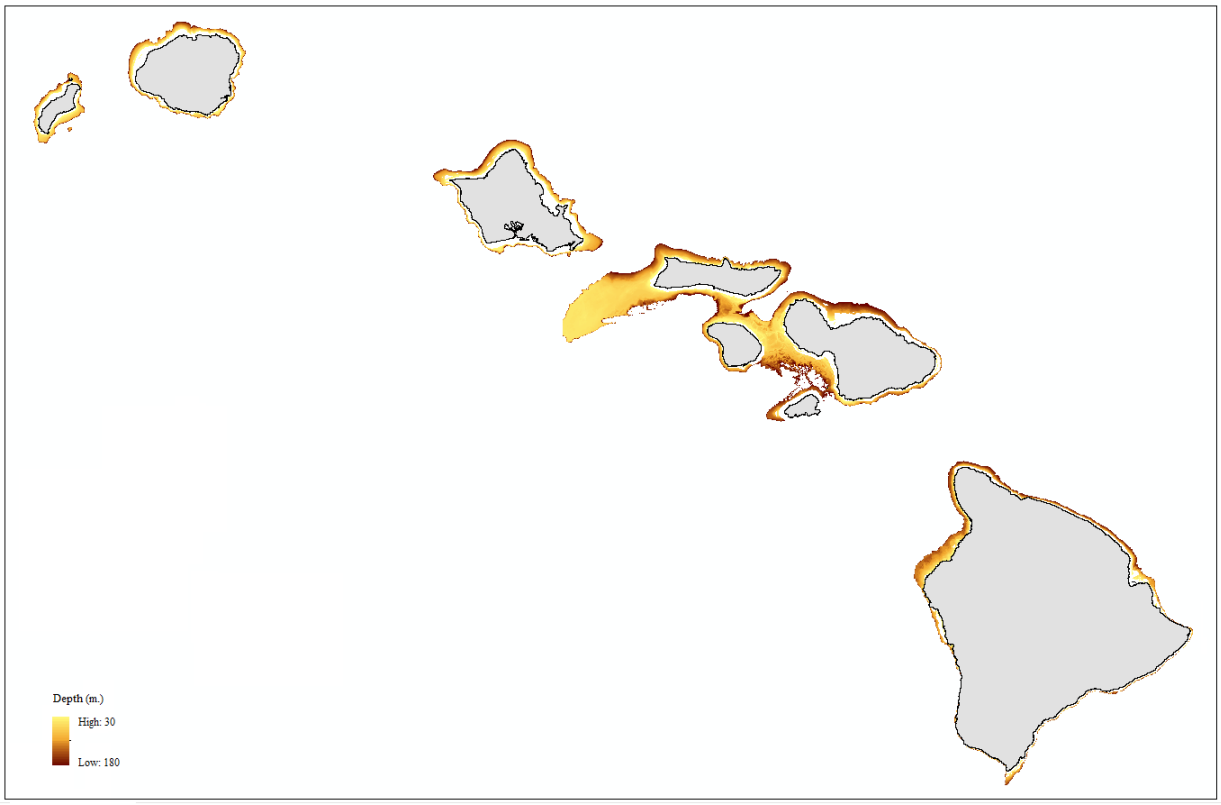

Figure A1. Map of 25 m resolution bathymetry values.

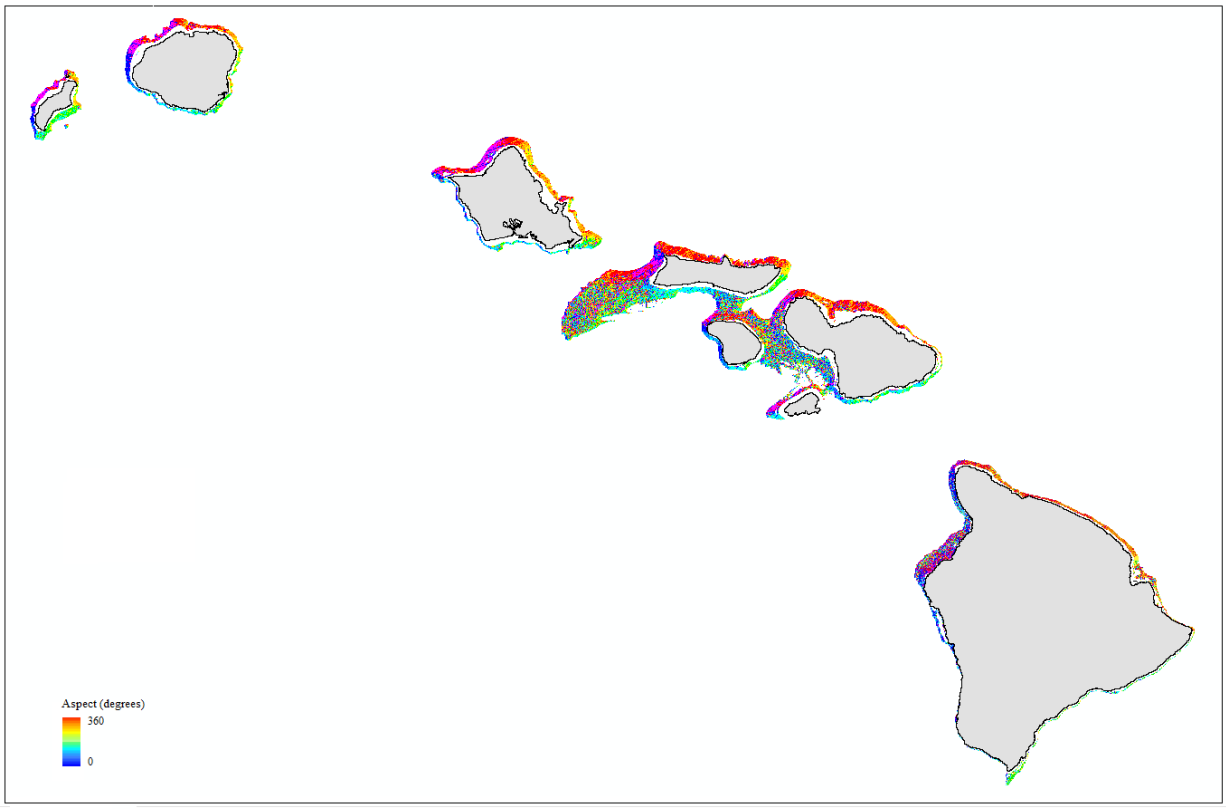

Figure A2. Map of 25 m resolution aspect values.

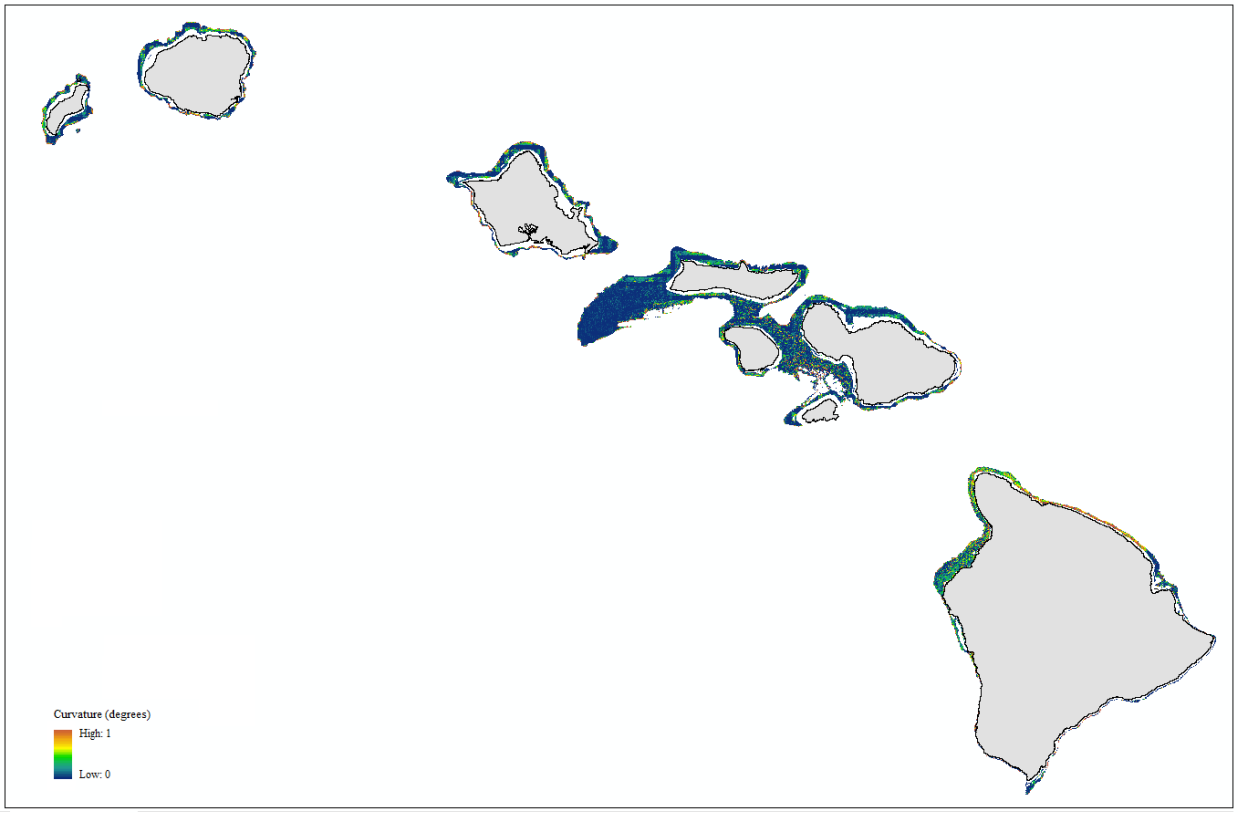

Figure A3. Map of 25 m resolution curvature values.

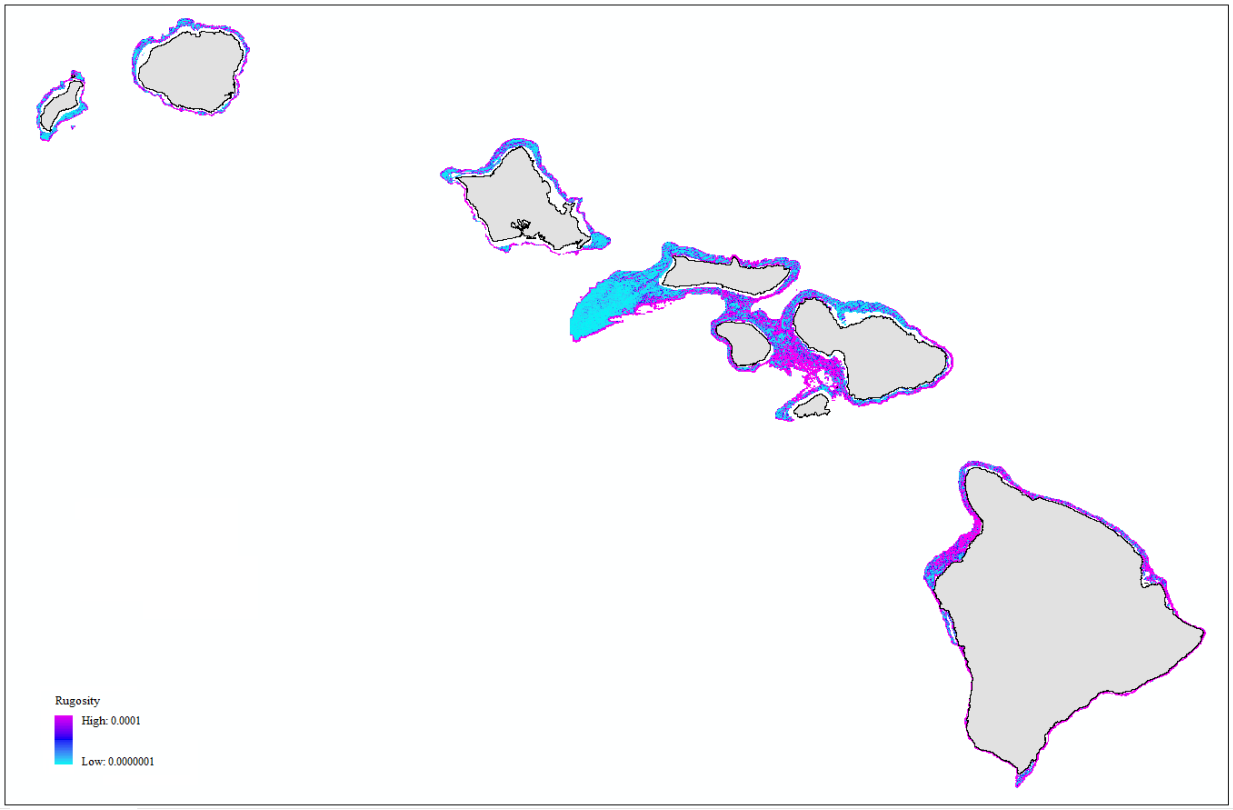

Figure A4. Map of 25 m resolution rugosity values.

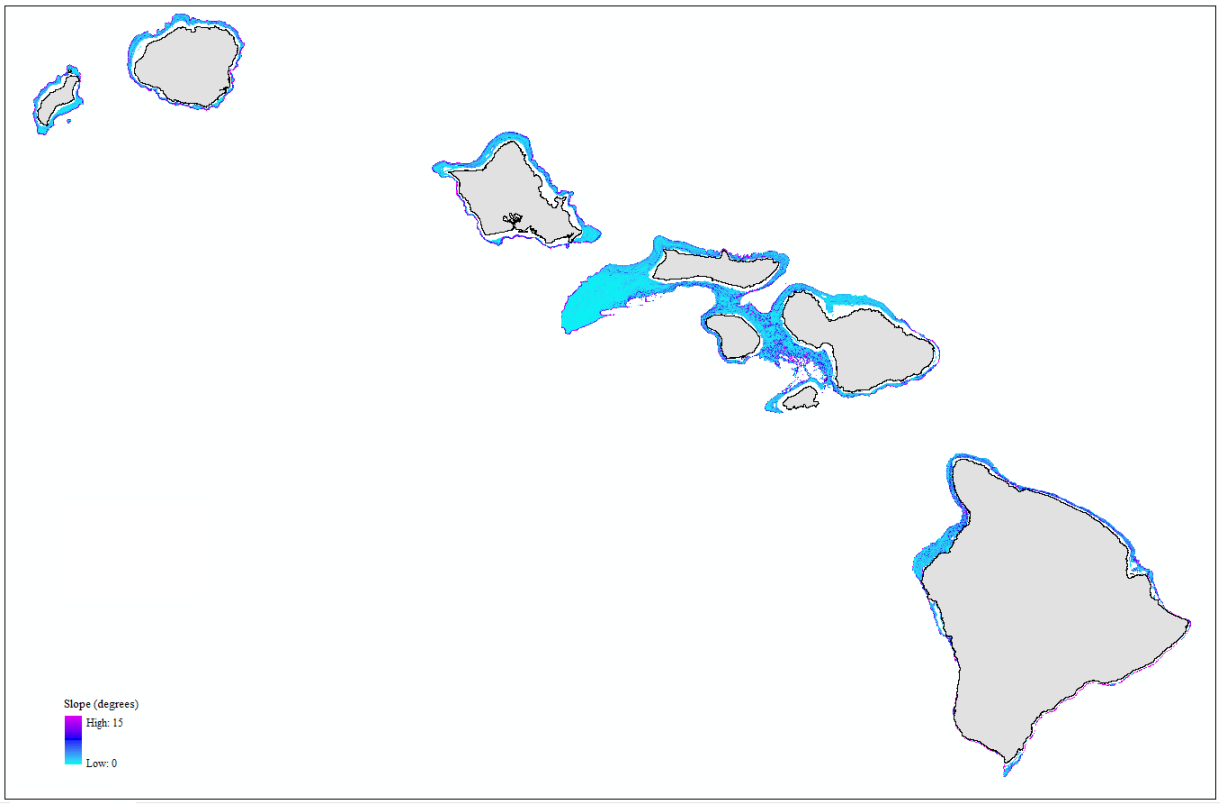

Figure A5. Map of 25 m resolution slope values.

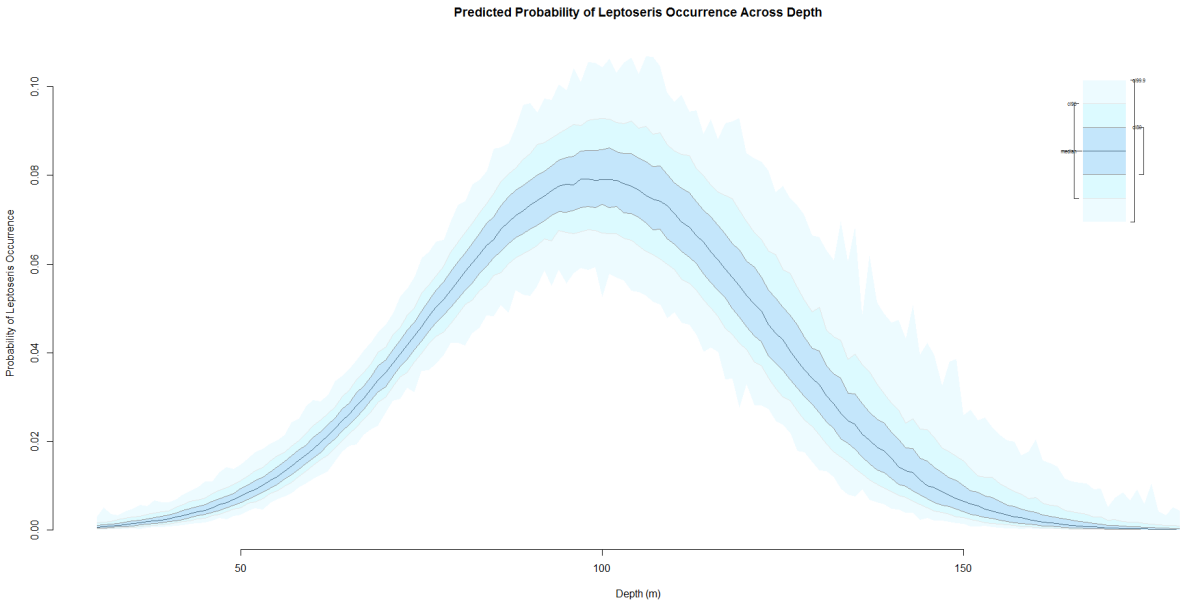

Figure A6. Effect of depth on *Leptoseris* probability values.

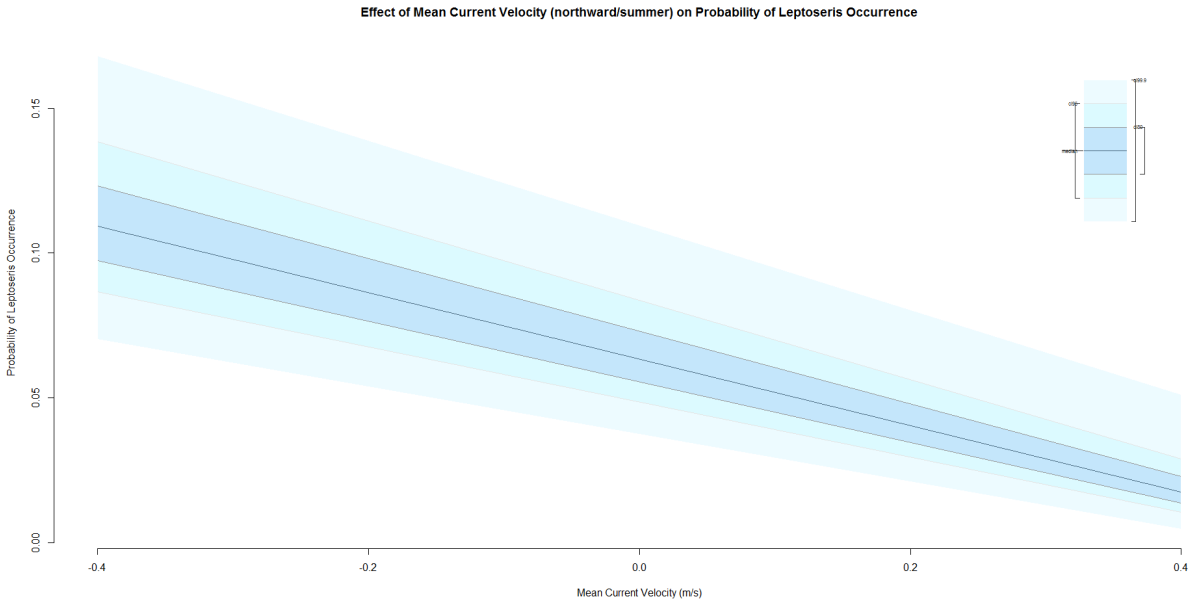

Figure A7. Effect of current velocity on *Leptoseris* probability values.

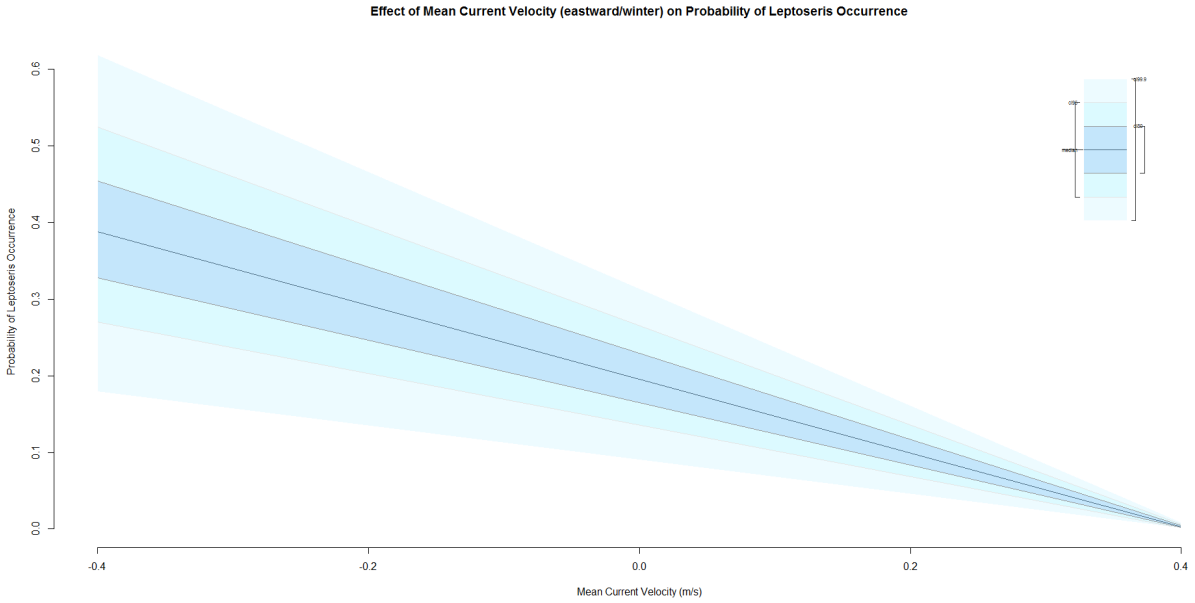

Figure A8. Effect of current velocity on *Leptoseris* probability values.

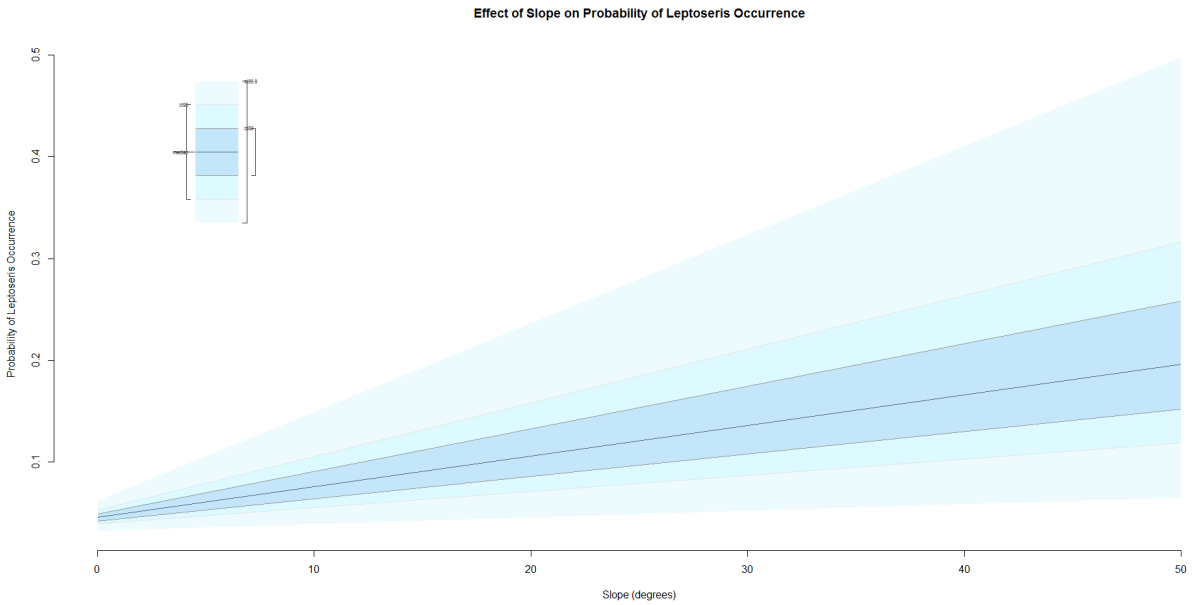

Figure A9. Effect of slope on *Leptoseris* probability values.

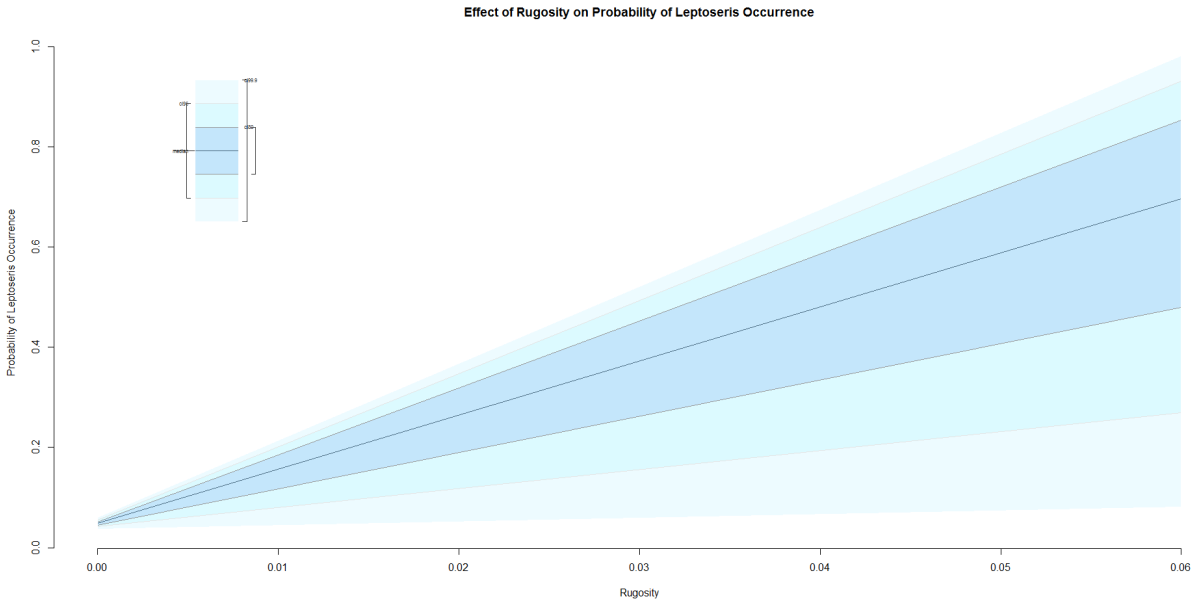

Figure A10. Effect of rugosity on *Leptoseris* probability values.

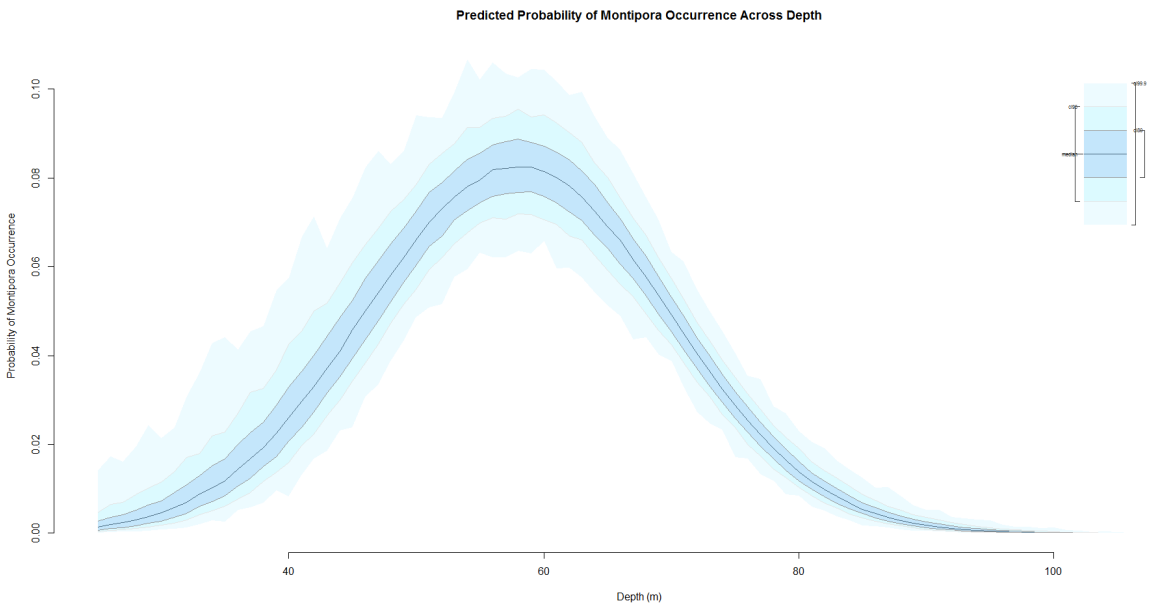

Figure A11. Effect of depth on *Montipora* probability values.

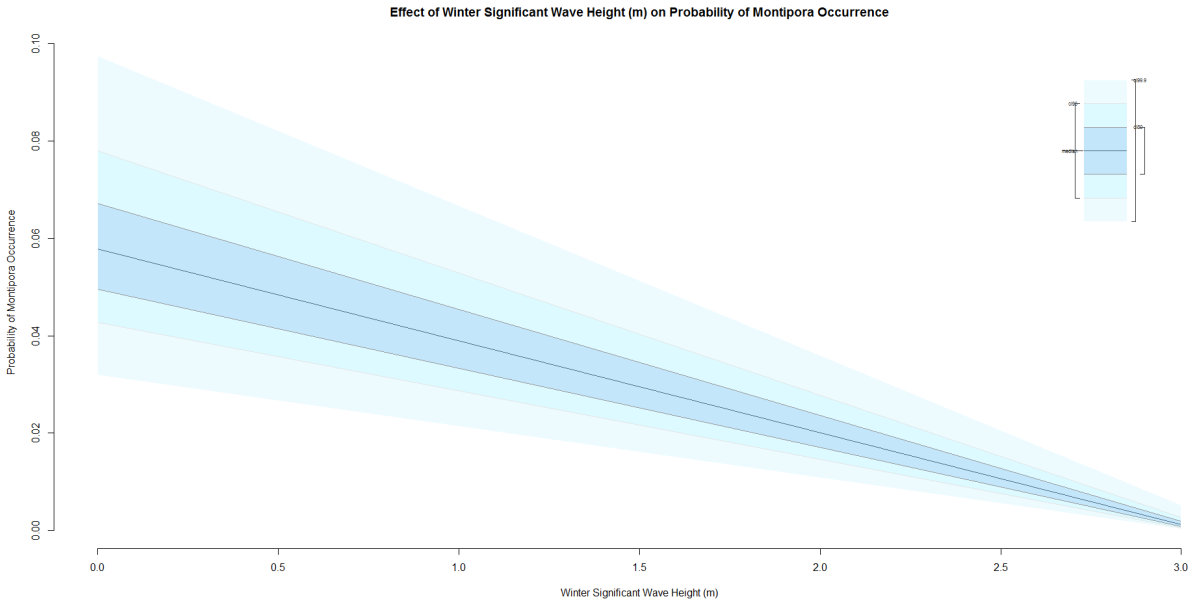

Figure A12. Effect of winter significant wave height on *Montipora* probability values.

2.9 Tables

Table 1. Number of field observations for each coral genus.

Source	No. observations	*Leptoseris*	*Montipora*
O‘ahu	2645	192	0
Kaua‘i	112	38	3
Maui	19957	708	791
Total	22714	938	794

Table 2. List of all variables considered for inclusion in my analyses.

Variable	Category	Variable description	Source	Resolution	Variable
Biological (response)	Hard coral	Presence/ absence between 30 - 180 m in depth	PIFSC, HURL optical validation data	NA	*Leptoseris*
					Montipora
Environmental (predictor)	Light availability	The depth of the euphotic zone (PAR 1%) determined using the Morel method (2007)	NOAA Oceanwatch Live Access Server; NASA, 2014	4 x 4 km	Mean euphotic depth (m)
	Topography	Seafloor complexity calculated with the ArcGIS BTM Terrain Ruggedness tool	USGS, 1998; University of Hawaii SOEST, 2014	50 x 50 m resampled to 25 x 25 m	Rugosity (unitless)
		Depth of seafloor			Depth (m)
		Rate of change calculated with the ArcGIS BTM Slope tool			Slope (degrees)

		Curvature of the seafloor calculated using the ArcGIS Curvature tool			Curvature (degrees of degrees)
		Hardness of seafloor detected by acoustic backscatter			Substrate hardness (unitless)
		Distance of observation point to nearest coastline			Distance to coastline (m)
		Compass direction of maximum slope calculated using the ArcGIS Aspect tool			Aspect (degrees)
	Waves/currents	Mean current velocity data obtained per season (winter/summer) for depths: 200, 150, 125, 100, 75, 50, 30 m	PacIOOS Hawai‘i Regional Ocean Model	4 x 4 km	Mean cur. vel. (northward/summer) (m s-1)
					Mean cur. vel. (northward/winter) (ms-1)
					Mean cur. vel. (eastward/summer) (m s-1)
					Mean cur. vel. (eastward/winter)(m s-1)
		Sea surface mean significant wave height	PacIOOS Hawai‘i SWAN Wave Model	1 x 1 km	Sig. wave height (summer) (m)
					Sig. wave height (winter) (m)

Table 3. Summary statistics for theoretical semivariogram models.

Genus	Sum of squares	Input	Input	Actual	Actual	Actual
Leptoseris	2940.671	0.055	218	0.051	206.909	0
Montipora	14013.610	0.020	390	0.032	390.000	0.003

Table 4. Predictive model output. Results by genus: theta threshold subscripts indicate model type and training and validation (c-v) outputs. Sensitivity and specificity totals apply to training data only.

	FP	FN	$TP_{c\text{-}v}$	$TN_{c\text{-}v}$	$FP_{c\text{-}v}$	$FN_{c\text{-}v}$	Sensitivity	Specificity	SS_{max}	OPS	$OPS_{c\text{-}v}$	AUC
	1525	84	219	3757	1894	94	0.7264	0.7305	1.4569	73.0%	66.7%	0.782
	1490	87	182	4000	1651	131	0.7166	0.7367	1.4533	73.6%	70.1%	0.780
	1536	69	159	4453	1406	86	0.7435	0.7368	1.4803	73.7%	75.6%	0.808
	1499	71	165	4406	1453	80	0.7361	0.7465	1.4792	74.3%	74.9%	0.809

Genus	θ threshold	TP	TN
Leptoseris	θ_{OLR} = 0.065	223	4133
	θ_{corr} = 0.067	220	4168
Montipora	θ_{OLR} = 0.064	200	4299
	θ_{corr} = 0.0625	198	4336

Table A1. Summary output for *Leptoseris* OLR model.

Covariate	Coefficient estimate	Std. error
Intercept	-12.920	1.143
Depth	0.201	0.02575
Depth*Depth	-0.001	0.0001413
Mean current velocity: northward, summer	-2.392	0.5673
Mean current velocity: eastward, winter	-6.625	0.8725
Slope	0.033	0.008194
Rugosity	64.910	18.250

Table A2. Summary output for *Montipora* OLR model.

Covariate	Coefficient estimate	Std. error
Intercept	-11.890	2.160
Depth	0.3617	0.06851
Depth*Depth	-0.004	0.0005437
Significant wave height: winter	-1.303	0.1926

Table A3. Summary output for *Leptoseris* rare events corrected model.

Covariate	Coefficient estimate	Std. error
Intercept	-12.740	1.142
Depth	0.197	0.02572

Depth*Depth	-0.001	0.0001
Mean current velocity: northward, summer	-2.396	0.5666
Mean current velocity: eastward, winter	-6.591	0.8715
Slope	0.033	0.008185
Rugosity	64.560	18.230

Table A4. Summary output for *Montipora* rare events corrected model.

Covariate	Coefficient estimate	Std. error
Intercept	-13.990	2.158
Depth	0.4407	0.06847
Depth*Depth	-0.004	0.0005433
Significant wave height: winter	-1.300	0.1925

2.10 References

Adams, S. M. (2005). Assessing cause and effect of multiple stressors on marine systems. *Marine Pollution Bulletin* 51: 649 – 657.

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics* 2: 370 – 374.

Allison, P.D. (2001). Logistic regression using the SAS System: Theory and application. *Wiley Interscience*, New York. 288 pp.

Bak, R.P.M., Nieuwland, G., and Meesters, E.H. (2005). Coral reef crisis in deep and shallow reefs: 30 years of constancy and change in reefs of Curacao and Bonaire. *Coral Reefs* 24: 475 – 479.

Bongaerts, P., Ridgway, T., Sampayo, E. M., and Hoegh–Guldberg, O. (2010). Assessing the 'deep reef refugia' hypothesis: focus on Caribbean reefs. *Coral Reefs* 29: 309 – 327.

Bongaerts, P., Frade, P., Hay, K., Englebert, N., Latijnhouwers, K., Bak, R., Vermeij, M., and Hoegh–Guldberg, O. (2015). Deep down on a Caribbean reef: lower mesophotic depths harbor a specialized coral–endosymbiont community. *Scientific Reports* 5: 7652.

Breslow, N. E., and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 88: 9 – 25.

Bridge, T., and Guinotte, J. (2013). Mesophotic coral reef ecosystems in the Great Barrier Reef world heritage area: their potential distribution and possible role as refugia from disturbance. Research Publication no. 109. Great Barrier Reef Marine Park Authority. Townsville, QLD. 57 pp.

Calcagno, V., and Mazancourt, C. D. (2010). glmulti: An R package for easy automated model selection with (Generalized) Linear Models. *The Journal of Statistical Software* 34: 1 – 29.

Cantor, S. B., Sun, C. C., Torturer–Luna, G., Richards–Koru, R., and Follen, M. (1999). A comparison of C/B ratios from studies using receiver operating characteristic curve analysis. *Journal of Clinical Epidemiology* 52: 885 – 892.

Choirat, C., Honaker, J., Imai, K., King, G., and Lau, O. (2015). *Zelig: Everyone's statistical software, Version 5.0–3,* URL: http://ZeligProject.org.

Costa, B., Kendall, M., Rooney, J., Chow, M., Lecky, J., Parrish, F.A., Montgomery, A. and Spalding, H. (2012). Prediction of mesophotic coral distributions in the ʻAuʻau Channel, Hawaiʻi. NOAA Technical Memorandum NOS NCCOS 149.Prepared by the NCCOS Center for Coastal Monitoring and Assessment Biogeography Branch. Silver Spring, MD. 44 pp.

Costa, B., Kendall, M.S., Parrish, F.A., Rooney, J., Boland, R.C., Chow, M., Lecky, J., Montgomery, A., and Spalding, H. (2015). Identifying suitable locations for mesophotic hard corals offshore of Maui, Hawaiʻi. *PLoS One* 10:e0130285.

Crain, C., Kroeker, K., and Halpern, B. (2008). Interactive and cumulative effects of multiple human stressors in marine systems. *Ecology Letters* 11: 1304 – 1315.

Cramer, W., Bondeau, A., Woodward, F. I., Prentice, I. C., Betts, R. A., Brovkin, V., Cox, P. M., Fisher, V., Foley, J. A., Friend, A. D., Kucharik, C., Lomas, M. R., Ramankutty, N., Sitch, S., Smith, B., White, A., and Young–Molling, C. (2001). Global response of terrestrial ecosystem structure and function to CO_2 and climate change: results from six dynamic global vegetation models. *Global Change Biology* 7: 357 – 373. DOI:10.1046/j.1365–2486.2001.00383.x.

Cramer, J. S. (2003). Logit models: from economics and other fields. Cambridge: Cambridge Univ. Press, pp. 66 – 67.

Crowder, L., and Norse, E. (2008). Essential ecological insights for marine ecosystem–based management and marine spatial planning. *Marine Policy* 32: 772 – 778. DOI:10.1016/j.marpol.2008.03.012.

Dancey, C., and Reidy, J. (2004). Statistics without maths for psychology. England: Pearson Education Limited.

Dolan, M., Grehan, A., Gunhan, J., and Brown, C. (2008). Modelling the local distribution of cold–water corals in relation to bathymetric variables: Adding spatial context to deep–sea video data. Deep Sea Research Part I: *Oceanographic Research Papers* 56: 1564 – 1579.

Dollar, S. J. (1982). Wave stress and coral community structure in Hawaii. *Coral Reefs* 1: 71 – 81.

Douvere, F. (2008). The importance of marine spatial planning in advancing ecosystem–based sea use management. *Marine Policy* 32: 762 – 771. DOI:10.1016/j.marpol.2008.03.021.

Fielding, A. H., and Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/ absence models. *Environmental Conservation* 24: 38 – 49.

Foley, M., Halpern, B., Fiorenza, M., Armsby, M., Caldwell, M., Crain, C., Prahler, E., Rohr, N., Sivas, D., Beck, M., Carr, M., Crowder, L., Duffy, J., Hacker, S., McLeod, K., Palumbi, S., Peterson, C., Regan, H., Ruckelshaus, M., Sandifer, P., and Steneck, R. (2010). Guiding ecological principles for marine spatial planning. *Marine Policy* 34: 955 – 966. DOI:10.1016/j.marpol.2010.02.001.

Forsman, Z. H., Concepcion, G. T., Haverkort, R. D., Shaw, R. W., Maragos, J. E., and Toonen, R. J. (2010). Ecomorph or endangered coral? DNA and microstructure reveal Hawaiian species complexes: *Montipora dilatata/flabellata/turgescens* & *M. patula/verrilli. PLoS One*; 5(12):e15021.

Franklin, E. C., Jokiel, P. L., and Donahue, M. J. (2013). Predictive modeling of coral distribution and abundance in the Hawaiian Islands. *Marine Ecology Progress Series* 481: 121 – 132.

Fricke, H.W., Vareschi, E., and Schlichter, D. (1987). Photoecology of the coral *Leptoseris fragilis* in the Red Sea twilight zone (an experimental study by submersible). *Oecologia* 73:371 – 381.

Geary, R. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician* 5: 115 – 146.

Gobin, A., Campling, P., and Feyen, J. (2001). Logistic modelling to identify and monitor local land management systems. *Agricultural Systems* 67: 1 – 20.

Goreau, T., and Goreau, N. (1973). The ecology of Jamaican coral reefs II. Geomorphology, zonation, and sedimentary phases. *Marine Science Bulletin* 23: 399 – 464.

Grigg, R. W. (1981). Coral reef development at high latitudes in Hawai‘i. In: *Proceedings of the 4th International Coral Reef Symposium* 1:687 – 693.

Grigg, R. W. (1983). Community structure, succession, and development of coral reefs in Hawaii. *Marine Ecology Progress Series* 11: 1 – 14.

Grigg, R. W. (1998). Holocene coral reef accretion in Hawai‘i: a function of wave exposure and sea level history. *Coral Reefs* 17: 263 – 272.

Grigg, R. W. (2006). Depth limit for reef building corals in the ‘Au‘au Channel, S.E. Hawaii. *Coral Reefs* 25: 77 – 84.

Grottoli, A. G., Rodrigues, L. J., and Palardy, J. E. (2006). Heterotrophic plasticity and resilience in bleached corals. *Nature* 440: 1186 – 1189.

Hadley, J., and McNeil, B. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29 – 36.

Harmelin–Vivien M. L., and Laboute, P. (1986). Catastrophic impact of hurricanes on atoll outer reef slopes in the Tuamotu (French Polynesia). *Coral Reefs* 5: 55 – 62.

Hayward, M. W., Tores, P. J. D., Dillon, M. J., and Banks, P. B. (2007). Predicting the occurrence of the quokka, *Setonixbrachyurus* (Macropodidae: Marsupialia), in Western Australia's northern jarrah forest. *Wildlife Research* 34: 194 – 199.

Hosmer, D., Hosmer, T., Cessie, S., and Lemeshow, S. (1997). A comparison of goodness–of–fit tests for the logistic regression model. *Statistics in Medicine* 16: 965 – 980.

Hosmer, D., and Lemeshow, S. (2004). Applied logistic regression. Hoboken, New Jersey: John Wiley and Sons, Inc.

Hurley, K. K. C, Timmers, M. A., Godwin, L. S., Copus, J. M., Skillings, D. J., and Toonen, R. J. (2016). An assessment of shallow and mesophotic reef brachyuran crab assemblages on the

south shore of O‘ahu, Hawai‘i. *Coral Reefs* 35: 103 – 112. DOI:10.1007/s00338–015–1382–z.

Hurvich, C. M., and Tsai, C. L. (1989). Regression and time series model selection in small samples. *Biometrika* 76: 297 – 307.

Imai, K., King, G., and Lau, O. (2008). Toward a common framework for statistical analysis and development. *Journal of Computational and Graphical Statistics* 17: 892 – 913.

Jackson, L., Trebitz, A., and Cottingham, K. (2000). An introduction to the practice of ecological modeling. *BioScience* 8: 694 – 706.

Jian, X., Olea, R. A., and Yu, Y. S. (1996). Semivariogram modeling by weighted least squares. *Computers and Geosciences* 22: 387 – 397.

Jokiel, P. L., and Coles, S. L. (1977). Effects of temperature on the mortality and growth of Hawaiian reef corals. *Marine Biology* 43: 201 – 208.

Jokiel, P. L., Brown, E. K., Friedlander, A., Rodgers, S. K., and Smith, W. R. (2004). Hawai‘i coral reef assessment and monitoring program: spatial patterns and temporal dynamics in reef coral communities. *Pacific Science* 58: 159 – 174.

Kahng, S. E., and Kelley C. D. (2007). Vertical zonation of megabenthic taxa on a deep photosynthetic reef (50 – 140 m) in the ‘Au‘au Channel, Hawai‘i. *Coral Reefs* 26: 679 – 687.

Kahng, S. E., Garcia–Sais, J. R., Spalding, H. L., Brokovich, E., Wagner, D., Weil, E., Hinderstein, L., and Toonen, R. J. (2010). Community ecology of mesophotic coral reef ecosystems. *Coral Reefs* 29: 255 – 275.

Kahng, S.E., Hochberg, E. J., Apprill, A., Wagner, D., Luck, D. G., Perez, D., and Bidigare, R. R. (2012). Efficient light harvesting in deep–water zooxanthellate corals. *Marine Ecology Progress Series* 455: 65 – 77.

Kahng, S. E., Copus, J. M., and Wagner, D. (2014). Recent advances in the ecology of mesophotic coral ecosystems (MCEs). *Current Opinion in Environmental Sustainability* 7: 72 – 81.

Kay, E. A. (1994). A Natural History of the Hawaiian Islands: Selected Readings II. Honolulu, Hawai‘i: University of Hawai‘i Press.

Kane, C., Kosaki, R., and Wagner, D. (2014). High levels of mesophotic reef fish endemism in the Northwestern Hawaiian Islands. *Bulletin of Marine Science* 90: 693 – 703. DOI:10.5343/bms.2013.1053.

Kendall, M. S., Jensen, O. P., Alexander, C., Field, D., McFall, G., Bohne, R., and Monaco, M. E. (2005). Benthic mapping using sonar, video transects, and an innovative approach to accuracy assessment: a characterization of bottom features in the Georgia Bight. *Journal of Coastal Research* 21: 1154 – 1165.

King, G., and Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis* 9: 137 – 163.

Kohler, K. E., and Gill, S. M. (2006). Coral Point Count with Excel extensions (CPCe): A Visual Basic program for the determination of coral and substrate coverage using random point count methodology. *Computers and Geosciences* 32: 1259 – 1269. DOI:10.1016/j.cageo.2005.11.009.

Lambert, P., Sutton, A., Burton, P., Abrams, K., and Jones, D. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine* 24: 2401 – 2428.

Larsen, P., Barker, S., Wright, J., and Erickson, C. (2004). Use of cost effective remote sensing to map and measure marine intertidal habitats in support of ecosystem modeling efforts: Cobscook Bay, Maine. *Northeastern Naturalist* 11: 225 – 242.

Lenihan, H., Peterson, C., Kim, S., Conlan, K., Fairey, R., McDonald, C., Grabowski, J., and Oliver, J. (2003). Variation in marine benthic community composition allows discrimination of multiple stressors. *Marine Ecology Progress Series* 261: 63 – 73.

Lesser, M., Slattery, M., and Leichter, J. (2009). Ecology of mesophotic coral reefs. *Journal of Experimental Marine Biology and Ecology* 375: 1 – 8.

Lesser, M., Slattery, M., Stat, M., Ojimi, M., Gates, R., and Grottoli, A. (2010). Photoacclimatization by the coral *Montastraea cavernosa* in the mesophotic zone: light, food, and genetics. *Ecology* 91: 990 – 1003.

Liu, C., Berry, P., Dawson, T., and Pearson, R. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28: 385 – 393.

Luck, D. G., Forsman, Z. H., Toonen, R. J., Leicht, S. J., and Kahng, S. E. (2013). Polyphyly and hidden species among Hawai'i's dominant mesophotic coral genera, *Leptoseris* and *Pavona* (Scleractinia: Agariciidae). *PeerJ* 1:e132.

Marcelino, L. A., Istneat, M. W., Stoyneva, V., Henss, J., Rogers, J. D., Radosevich, A., Turzhitsky, V., Siple, M., Fang, A., and Swain, T. D. (2013). Modulation of light–enhancement to symbiotic algae by light scattering in corals and evolutionary trends in bleaching. *PLoS One* 8:e61492.

McLeod, K., Lubchenco, J., Palumbi, S., and Rosenberg, A. (2005). Scientific consensus statement on marine ecosystem–based management. Communication Partnership for Science and the Sea (COMPASS). http://compassonline.org/marinescience/solutions_ecosystem.asp

Nyström, M., Folke, C., and Moberg, F. (2000). Coral reef disturbance and resilience in a human–dominated environment. *Trends in Ecology & Evolution* 15: 413 – 417.

Pacific Islands Benthic Habitat Monitoring Center (PIBHMC) 2015. ftp://ftp.soest.hawaii.edu/pibhmc/Website/Ibdocs/documentation/Optical–Proc–Overview.pdf.

Puglise, K., Hinderstein, L., Marr, J., Dowgiallo, M., and Martinez, F. (2009). Mesophotic coral ecosystems research strategy: International workshop to prioritize research and management needs for Mesophotic Coral Ecosystems, Jupiter, Florida, 12–15 July 2008. Silver Spring, MD: NOAA National Centers for Coastal Ocean Science, Center for Sponsored Coastal Ocean Research, and Office of Ocean Exploration and Research, NOAA Undersea Research Program. NOAA Technical Memorandum NOS NCCOS 98 and OAR OER 2.24 pp.

Rogers, C. S. (1990). Responses of coral reefs and reef organisms to sedimentation. *Marine Ecology Progress Series* 62: 185 – 202.

Rooney, J., Fletcher, C., Engels, M., Grossman, E., and Field, M. (2004). El Niño control of Holocene reef accretion in Hawai‘i. *Pacific Science* 58: 305 – 324.

Rooney, J., Donham, E., Montgomery, A., Spalding, H., Parrish, F., Boland, R., Fenner, D., Gove, J., and Vetter, O. (2010). Mesophotic coral ecosystems in the Hawaiian Archipelago. *Coral Reefs* 29: 361 – 367.

Ruiz, G., Fofonoff, P., and Hines, A. (1999). Non–indigenous species as stressors in estuarine and marine communities: Assessing invasion impacts and interactions. *Limnology and Oceanography* 44: 950 – 972.

Sokal, R.R., and Oden, N.L. (1978). Spatial autocorrelation in biology. *Biological Journal of the Linnean Society* 10: 199 – 228.

Tomz, M., King, G., and Zeng, L. (2003). ReLogit: Rare events logistic regression. *Journal of Statistical Software* 8: i02.

Tabachnick, B., and Fidell, L. (1994). Using multivariate statistics. England: Pearson.

Tyedmers, P., Watson, R., and Pauly, D. (2005). Fueling global fishing fleets. *AMBIO* 34: 635 – 638.

Van Den Eeckhaut, M., Vanwalleghem, T., Poesen, J., Govers, G., Verstraeten, G., and Vandekerckhove, L. (2006). Prediction of landslide susceptibility using rare events logistic regression: A case–study in the Flemish Ardennes (Belgium). *Geomorphology* 76: 392 – 410.

Webster, J. M., Clague, D. A., Riker–Coleman, K., Gallup, C., Braga, J. C., Potts, D., Moore, J. G., Winterer, E. L. and Paull, C. K. (2004). Drowning of the 150 m reef off Hawaii: A casualty of global meltwater pulse 1A? *Geology* 32: 249 – 252.

White, K. N., Ohara, T., Fujii, T., Kawamura, I., Mizuyama, M., Montenegro, J., Shikiba, H., Naruse, T., McClelland, T. Y., Denis, V., and Reimer, J. D. (2013). Typhoon damage on a shallow mesophotic reef in Okinawa, Japan. *PeerJ* 1:e151.

Wright, D. J., Pendleton, M., Boulware, J., Walbridge, S., Gerlt, B., Eslinger, D., Sampson, D., and Huntley, E. (2012). ArcGIS Benthic Terrain Modeler (BTM), Environmental Systems Research Institute, NOAA Coastal Services Center, Massachusetts Office of Coastal Zone Management.

Yokoyama, Y., Purcell, A., Lambeck, K., and Johnston, P. (2001). Shore–line reconstruction around Australia during the last glacial maximum and late glacial stage. *Quaternary International* 83: 9 – 18.

Chapter 3: Distribution of Hawaiian mesophotic coral *Leptoseris* spp. and calcified green alga *Halimeda kanaloana* influenced by seafloor slope, curvature, and aspect

3.1 Abstract

Mesophotic coral ecosystems (MCEs), named for their light-limited nature, extend from 30 m to 180 m below the sea surface and are inhabited by specialized coral and algal species tolerant of low light conditions (Kahng et al. 2010). The relative inaccessibility of MCEs

contributes to the dearth of survey data and incomplete knowledge about their ecological community structure. To better understand the abiotic factors that influence the distribution of mesophotic coral and algal communities, I constructed multiple geographically-explicit species distribution models for two primary ecosystem architects, the corals of genus *Leptoseris* and the endemic calcified alga *Halimeda kanaloana*. Models are based on observations from submersible dives and environmental data collected across the mesophotic zone of eastern Penguin Bank, Molokaʻi, Hawaiian Islands during 2006 to 2010. Models performed consistently better (predictive performance) and consistently indicated the influence of certain covariates at high and moderate spatial scales, though the area under the curve (AUC) values of all models were comparable. Across all spatial scales, corals were positively influenced by seafloor slope, aspect, and curvature. *H. kanaloana* meadows were predicted across regions with minimal slope and aspect and perennial moderate current flux. This study furthers scientific understanding of the drivers of mesophotic community structure by examining the influence of different environmental covariates over spatial scales. Additionally, these are the first distribution models of an endemic mesophotic alga in Hawaiʻi.

3.2 Introduction

Mesophotic coral ecosystems (MCEs) are challenging to access due to their depth range (30 – 180 m), which extends beyond the recreational SCUBA limit but is too shallow to survey using costly submersibles. Scientific interest in MCEs has steadily increased in the past few decades largely due to the hypothesis that MCEs may act as refugia for organisms driven out of degraded shallow reefs (Glynn 1996). Sponges, algae, and zooxanthellate corals that have adapted to very low light levels dominate MCEs (Lesser et al. 2009). In the Northwestern Hawaiian Islands, some MCEs are known to harbor extremely high levels of endemism, up to 100% across certain reef patches (Kane et al. 2014; Kosaki et al. 2017). Clear water and availability of rugose or sloped hard substrate are thought to influence coral colonization in Hawaiian MCEs (Costa et al. 2015; Veazey et al. 2016; Pyle et al. 2016). Water clarity and substrate type are thought to influence mesophotic algal communities (Spalding 2012).

Ecological research on mesophotic organisms is challenging to conduct based on the costly, difficult logistics required to survey populations within this depth range. Prior efforts to characterize mesophotic coral communities in Hawaiʻi and around the world have required the use of species distribution models to generate maps of probable coral occurrence (e.g., Costa et

al. 2015; Veazey et al. 2016) or have made broad inferences from the aggregation of multiple, fragmentary surveys (e.g., Lesser et al. 2009; Wagner et al. 2014; Pyle et al. 2016). Minimal sampling and taxonomic uncertainty has slowed the creation of any maps to date pertaining to the distribution patterns of mesophotic algae. No prior work exists regarding the predicted distribution of mesophotic algal communities in Hawai‘i or around the world (Kahng et al. 2010).

Isolated corals of the genus *Leptoseris* are scattered across Penguin Bank, a submerged volcanic shield southwest of the island of Moloka‘i (Kahng & Maragos 2006). Previous work has pointed to the region as a possible hotspot for *Leptoseris* colonization (Veazey et al. 2016), possibly due to the occurrence of strong, organic matter-carrying currents. The plate-like morphology, specialized gastrovascular cavity, and conical, light-refracting microskeletal structure of certain *Leptoseris* corals makes them uniquely suited to the light-starved mesophotic zone (Fricke et al. 1987; Schlichter & Fricke 1991; Schlichter et al. 1992). Because these corals are heterotrophic (i.e., they feed from the water column and photosynthesize), I predict that they may colonize high-relief regions exposed to moderate current flow to maximize feeding and photosynthesizing and minimize smothering via sedimentation. Other deep water heterotrophic corals (i.e., *Lophelia pertusa, Paragorgia arborea, Primnoaresedae formis*) colonize on steep slopes or in the vicinity of strong currents (Mortensen et al. 2001; Leverette & Metaxas 2005; Locker et al. 2010).

The episodic influxes of nitrogen and phosphorus associated with strong currents may promote the growth of macroalgae (Leichter et al. 2003; Pérez-Mayorga et al. 2011). Smith et al. (2004) suggest that occurrences of dense *Halimeda tuna* beds across Conch Reef, Florida, are associated with periodic, localized increases in nutrient levels and decreases in temperature that may be indicative of internal tidal action. Similarly, exposure to current flux influences the growth and distribution of *Halimeda* meadows in the northern section of the Great Barrier Reef (Drew & Abel 1988; Wolanski et al. 1988). Spalding (2012) noted the high abundance of a psammophytic (i.e., thrives only in sand) species, *Halimeda kanaloana,* across the Maui Nui complex in Hawai‘i. The endemic *H. kanaloana* is a calcareous green alga which can grow in thick, patchy meadows of 100% cover in certain areas to ~ 90 m depths (Spalding 2012). Relatively little information exists about the ecology or distribution of *H. kanaloana* across Penguin Bank.

To better understand the abiotic drivers that influence the distributions of mesophotic benthic organisms, I explored relationships between coral and algae observations and abiotic covariates to create a suite of ecological niche models. These models were then used to produce predictive geographic maps of the probable occurrence of beds of the scleractinian corals in the genus *Leptoseris* and meadows of the algae *H. kanaloana* across eastern Penguin Bank, Molokaʻi. To better understand the role that spatial scale played in the niche models, I generated my analysis for corals and algae at multiple spatial resolutions (5 m, 25 m, and 100 m).

3.3 Methods

3.3.1 Study site

Penguin Bank is located off the southwestern coast of the Hawaiian island of Molokaʻi (Fig. 1) is approximately 50 km long and 19 km wide with an average depth of 70 m across the bank. Along the perimeter, the depth rapidly drops to over 500 m (Gregory & Kroenke 1982). The bank is known for such features as endemic algae and corals, strong currents, and large, carbonate sandbanks (Norris et al. 1995; Kahng & Maragos 2006; Kelman et al. 2012; Sabine & Mackenzie 1995). It an important humpback whale calving area and is an important habitat for the commercial bottomfish fishery (Baker & Herman 1981; Haight et al. 1993; Sackett et al. 2014).

3.3.2 Biological and environmental data

The Hawaiʻi Undersea Research Laboratory (HURL) and the Pacific Islands Fisheries Science Center (PIFSC) provided video from 39 submersible dives conducted from 2006 – 2010 across Penguin Bank. I collected 10,782 frame grab images by pausing each video track every 30 seconds that the sub was in motion. A sampling interval of 30 seconds has precedent in the literature (e.g., Costa et al. 2015; Veazey et al. 2016) and is the standard protocol for efficient optical validation data collection by PIBHMC. A subset of these images (4,304) were not used as they failed to fall inside my study range, lacked image clarity, or geographically overlapped with another observation. If more than one observation occurred in the same pixel, I assigned the pixel the maximum value of all observations (i.e., all zeroes designated the cell as an absence; any positive identification, denoted by a value of one, designated the cell as a presence). A further 2,211 observations were excluded from my model-building dataset due to missing covariate data, leaving a final dataset of 4,267 observations for inclusion in my 5 m analyses (Figs. 2, 3). I resampled these data for my 25 m and 100 m analyses and recorded 2,571

observations and 726 observations, respectively (Table 1). Seventy percent of my data was randomly selected for model building and 30% was set aside for cross-validation (Araujo et al. 2005).

I considered oceanographic and benthic geomorphological covariates that may influence the distribution of mesophotic *Leptoseris* corals and green alga *Halimeda kanaloana* in Hawai‘i (Table 2). Substrate hardness values, based on the strength of the returned sonar signal, were categorized as rocky (≥ 187; Kahng & Kelley 2007), moderately rocky, mostly sandy with scattered rocks or gravel, and mostly soft sand. My bin ranges were determined by subtracting one standard deviation from the next-highest classification. I also aimed to construct novel predictive species distribution models for *H. kanaloana* across eastern Penguin Bank. *H. kanaloana* colonizes deep, soft sediment, which differs from the hard substrata typically inhabited by hermatypic corals (Spalding 2012). In addition to substrate hardness, I selected variables that characterize the seafloor, including rugosity, slope, curvature, and compass direction.

To characterize the water column, I included variables that would indicate water clarity and current flux. I used seasonal (summer = May – September and winter = October – April; Kay 1994) mean, max, mean maximum, and maximum variance values of the diffuse attenuation coefficient at 490 nm (the wavelength for blue light) to calculate proportional downwelled irradiance at depth, such that

$$\frac{E_d(z)}{E_d(0^-)} = \exp(-K_d z).$$

Here, $E_d(0\text{-})$ represents downwelled irradiance just below the sea surface (Franklin et al. 2013). I also included estimates of the mean euphotic depth, or the depth at which light levels are sufficient for the rate photosynthesis to match the rate of respiration. I sourced my K_d and chlorophyll *a* values from the NOAA Coral Reef Watch Satellite Oceanography and Climatology Division.

Water movements were modeled with estimates of the mean and variance of directional, seasonal current velocities. I sourced these data from the PacIOOS Hawai‘i Regional Ocean Data Assimilating Model (2007 – 2012).

3.3.3 Data cleaning and model specification

I followed a straightforward data exploration and cleaning protocol recommended by Zuur et al. (2010); this procedure included checking for outliers and collinearity among the covariates and independence of my response variable. A preliminary exploration of my covariates revealed strong correlation between several variables (≥ 0.6) or complete separation of the variable values (i.e., most values at either end of a given variable's range occurred when observational data values were all equivalent to 0 or 1, but not both) (Figs. 4 – 9). Subsequently, I refined my lists of considered covariates (Tables 3, S1).

I chose best-fit generalized linear models (GLMs) for my coral and algal observations and environmental covariates at each spatial scale based on Δ AICc values of ≥ 2. This value meets the threshold (value = 2) of "strong support" as identified by Burnham and Anderson (2002), and the significant change in AICc values between the best model and the "next best" set of predictors indicates that the selected set of predictors significantly impact the model's predictive capacity.

The generalized linear model, also known as a GLM, generalizes the permissible distributions for the regression residuals beyond the normal through the use of a link function. A GLM with a logistic regression function that uses a logit link is widely used in studies about habitat suitability, epidemiology, environmental science, etc. (e.g., Kennedy et al. 2002; Peduzzi et al. 1996; Ohlmacher & Davis 2003). The equation for this GLM is given as

$$\text{logit}(\theta) = \mu = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n + \varepsilon_i .$$

Here, theta represents the probability that an organism is present at a given pixel; betas are the intercept and regression coefficients, respectively; and the *x* variables are the linear sum of covariates. One assumption inherent in this equation is the independence of residuals, which are the differences in the observed value and the model-predicted value at a given pixel *i*.

Spatial autocorrelation can result in spurious ecological inferences (Legendre 1993), so I evaluated both global-scale (i.e., Moran's I; expectation in the absence of autocorrelation = 0; Moran 1950) and local-scale (i.e., Geary's; expectation = 1; Geary 1954) spatial autocorrelation for models at all spatial scales. The values of global and local autocorrelation indicated spatial autocorrelation at different scales was present in the data (Moran 1950; Geary 1954). The presence of both localized and large-scale geographic clustering in the data necessitated my specification of spatially explicit models, as this clustering violates the assumption of independence of residuals.

To account for spatial autocorrelation in the data, I ran generalized linear mixed models (GLMMs) at all resolutions (McCulloch & Neuhaus 2001). GLMMs account for spatial autocorrelation in the data through inclusion of a random effect, so they can be considered an extension of the GLM, which includes only fixed effects (hence, GLMMs are "mixed"). A GLMM is expressed as

$$\mu = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n + \tau_n z_n + \varepsilon_i .$$

Here, tau represents the random complement to the fixed beta coefficients and *z* represents the random complement to the fixed *x* predictor variables. The non-normality of my residuals (i.e., a sample of the predicted values of my dependent variable subtracted from the observed values) as indicated by an assortment of quantile plots guided my selection of spatial models. I adapted the methodology outlined by Dormann et al. (2007) to specify my GLMMs. I fit my *Leptoseris* GLMMs by specifying all variables as fixed effects and included a dummy variable as my random effect, which incorporated my coordinate data, and thus took into account any spatial autocorrelation, in the estimation of the coefficient values. In addition, I fit *H. kanaloana* GLMMs at all resolutions to create spatial predictive layers of probable meadows across eastern Penguin Bank. The coefficients of my GLMMs are provided in Tables S2 – S13.

The probability of the occurrence of an organism in an environment is most often not equivalent to 50%, which has long been the default probability, or theta, threshold value for interpreting model predictions during cross-validation with binary response variable data (Liu et al. 2005). Theta threshold selection is a technique that requires the calculation of correctly predicted presences (sensitivity) and absences (specificity) during cross-validation given different thresholds (i.e., above a certain threshold, a probability value equals 1; below the threshold, it is taken as 0) (e.g., Liu et al 2005; Veazey et al. 2016). I used theta threshold selection to pick the most appropriate occupancy probability threshold for each model at each resolution. The best threshold value was identified as the value occurring at the point on the receiver operating characteristic (ROC) curve furthest from the line of chance (i.e., at which the model is predicting correctly as often as it is predicting incorrectly). I calculated the area under the curve (AUC), which is the area between the ROC curve and the line of chance; larger AUC values indicate better-performing, more trustworthy models.

All data manipulation was performed using R statistical software (R Core Team, 2017) and ESRI ArcGIS v.10.1 (ESRI 2012).

3.4 Results

The aspect, curvature, and slope of the seafloor consistently positively influenced the distribution of *Leptoseris* corals (Table 3). Substrate hardness also correlates positively with coral distribution, as noted previously by Taylor & Wilson (2003). All models consistently indicated some influence of mean maximum proportional downwelled irradiance during summer on coral distribution, though the nature of this influence is uncertain when examining results across spatial scales. My 5 m *Leptoseris* model performed well in the identification of coral presences and absences (all AUC values > 0.75; all cross-validated sensitivity and specificity values > 0.7). The AUC (0.77) was high enough to be classified as "very trustworthy" in its predictive capabilities (Hosmer & Lemeshow 2004). Despite slightly better AUC values, my moderate (25 m) and coarse (500 m) resolution *Leptoseris* models did not perform as well as my 5 m models during cross-validation. Cross-validation of the 25 m GLMM revealed low specificity (0.47). Surprisingly, my 100 m model fared better than the 25 m model (GLMM sensitivity = 0.73; specificity = 0.88). Theta threshold values for my *Leptoseris* models were universally 0.05.

The aspect and slope of the seafloor consistently negatively influenced *H. kanaloana* distribution at all resolutions. Mapped predictions for highly suitable *H. kanaloana* habitat did not exclusively overlay sites with low backscatter values, which represent soft sediment readings, a known habitat requirement for colonization by this alga (Spalding 2012). My *H. kanaloana* 5 m GLMM correctly predicted 88% of presence observations and 91% of absence observations in the testing dataset, and the model was trustworthy in its predictive capacity (AUC = 0.95). The model performed well following cross-validation (sensitivity = 0.96; specificity = 0.99). The 25 m GLMM performed relatively poorly, especially during cross-validation (AUC = 0.81; cross-validation sensitivity = 1; cross-validation specificity = 0.662). Similar to the performance of the *Leptoseris* models, I found that my 100 m alga model performed better than the 25 m model, but worse than the 5 m model (AUC = 0.89; sensitivity = 0.96; specificity = 0.76). Tables 4 – 6 outline the performance of all models.

The 5 m *Leptoseris* GLMM map predicted suitable habitat across the southern slopes of east Penguin Bank (Fig. 10). The 25 m *Leptoseris* model agreed with this general trend and predicted the highest suitability across the south and along slopes to the north (Fig. 11). The 100 m *Leptoseris* model showed less of a distinct region of suitability, but the output demarcating the

southern flank of the bank as most suitable agrees with the output of the other models; in addition, the 100 m model specified more suitable habitat closer to shore in the north (Fig. 12). All *Leptoseris* maps identify the southern portion of east Penguin Bank as most suitable for coral colonization, while the middle and far northern regions of the bank were not identified as suitable.

The 5 m *H. kanaloana* model predicted high habitat suitability across the middle to southeastern portion of Penguin Bank (Fig. 13). The 25 m *H. kanaloana* also predicts highest habitat suitability across the southeastern flank and nearshore, with moderate habitat across the middle of the bank (Fig. 14). The 100 m alga map identified the southeastern quadrant of the bank as most suitable for *H. kanaloana* (Fig. 15). Striping is present in these maps and is an artifact of the raw covariate data layers.

I implemented a *G*-test for goodness of fit for all models based on their identified best threshold value (Fig. 16). Almost all models displayed a very slight tendency to overpredict suitability, with the coarser resolution *H. kanaloana* model faring worse than the finer resolution models.

3.5 Discussion

I implemented spatially explicit models to predict the distribution of *Leptoseris* spp. corals and *H. kanaloana* meadows across Penguin Bank, Molokaʻi at three resolutions: 5 m, 25 m, and 100 m. The predictive maps created at 5 m resolution are the most finely resolved Hawaiʻi-specific mesophotic coral and algal habitat suitability maps to date.

3.5.1 Consistently influential environmental covariates at varying resolutions

Leptoseris habitat suitability values were highest (third quantile; probabilities $\geq$ 0.04) in my 5 m model at a median depth of 128 m. The clarity of oligotrophic Hawaiian waters extends the potential suitable habitat for these corals far offshore and at depths as great as 153 m (Kahng & Maragos 2006; Kahng 2013; Dinesen 1980; Costa et al. 2015).

The steepness (slope) of a pixel only slightly impacted the coral habitat suitability values generated by my 5 m model, with the third quantile of suitability values assigned to pixels with slope values between 0 and 90 degrees, though the mean slope value within this higher suitability subset only equaled 4.85 degrees. Across all of the relatively flat expanse of eastern Penguin Bank, the mean slope value was approximately 2.28 degrees. Similarly, the curvature of the seafloor only slightly influenced the spread of the habitat suitability values, with the highest

probabilities assigned to pixels with somewhat laterally convex curvature. My prior analyses (Veazey et al. 2016) indicated that *Leptoseris* corals may colonize areas of high relief in order to heighten their exposure to suspended organic matter, but across the plane of Penguin Bank, my models are unable to expand on this hypothesis.

Aspect values encompassed flat terrain (-1) and all values from 0 to 360 (due north) within pixels exhibiting habitat suitability values higher than the threshold value, though the mean aspect value within the high suitability subsets for all models fell between 146 – 189 degrees (south-southeast). Ralston et al. (1986) and Haight et al. (1993) conducted bottomfish surveys in Johnston Atoll and Penguin Bank, respectively, and each study concluded that the deflection of stronger deep currents along the perimeter of the banks may influence the increased abundance of fish species around the bank edges due to the boost in suspended planktonic biomass. My models predict some increase in *Leptoseris* habitat suitability across the southern portion of the east bank, which borders the Kalohi Channel to the southwest. Given the presence of mean winter northerly or easterly current flux values in all coral models, I speculate that the topographical feature of aspect is present in my models as a potential proxy for the heightened habitat suitability in areas exposed to deflected, upwelled, nutrient-rich currents around the perimeter of the bank. Patzert et al. (1970) recorded a consistent northerly current flowing across the bank, which is consistent with the high habitat suitability predicted across southern-facing surfaces.

In addition, year-round sunlight exposure is highest across southern aspects, with the midday winter solstice solar azimuth value for 2017 calculated as 170°. Given that *Leptoseris* corals are zooxanthellate (i.e., their symbionts photosynthesize), I expect that both *H. kanaloana* and *Leptoseris* may colonize southern-facing surfaces to maximize productivity.

I recorded *Halimeda kanaloana* present from approximately 66 to 74 m depth and from 4.8 to 6.5 km distance from the Molokaʻi coastline. My 5 m model suggests that the most suitable habitat for this macroalgae exists across the middle and southeastern portion of Penguin Bank. Areas predicted by my 5 m model as highly suitable (i.e., third quantile of suitability values; probability $\geq$ 0.09) included depths between 30 m and 75 m with a mean seafloor aspect of 171 degrees (south-southeast). The alga model and the *Leptoseris* models indicated that prime colonizable habitat exists across south-tilting substrate. Despite this commonality, swaths of

habitat predicted as suitable for *H. kanaloana* were predominantly separate and southeastward of the highest *Leptoseris* probabilities.

The characteristics of all environmental predictor variables at all resolutions are summarized in Tables S14 – S16.

3.5.2 Comparison of coral models

All 5 m models correctly identified at least 70% of true presence observations and at least 75% of true absences in the training datasets. Both the 5 m and 25 models identified the southern slopes of eastern Penguin Bank as most suitable for *Leptoseris* colonization. The 100 m model did not offer markedly different spatial predictions in comparison to the highly- and moderately-resolved models, though the nearshore portion of the bank was identified as suitable for coral habitation. All three models demarcated the middle-northern side of the bank as most unsuitable for *Leptoseris* colonization.

3.5.3 Comparison of alga models

The *H. kanaloana* models identified the southeastern flank of Penguin Bank as likely locations for extensive meadows. The eastern (i.e., coastal) and middle portion of Penguin Bank was identified as moderately suitable for colonization. Substrate hardness was featured in my 5 m models and was slightly, but highly significantly, negatively correlated with alga occurrence (i.e., meadows are more likely to be found in soft, sandy substrate). Aspect, slope, and seafloor curvature were featured in all models; however, the influence of these variables was slight. Moderate current flux minimally, but significantly, influenced alga distribution in my finely resolved model (Tables S2 – S3, S6 – S7, S10 – S11). From this, I conclude that *H. kanaloana* meadows are found in locations with flat, sandy substrate that are exposed to moderate year-round current flux.

The 5 m alga model performed better than the 100 m model, which performed better than the 25 m model, based on AUC values. While AUC is a valuable performance metric, the differences observed in this metric for these models cannot stand alone as a measure of their overall trustworthiness. Numerous artifacts are present in the mapped output of certain models, and these should be considered in a balanced evaluation of overall model predictive performance.

These maps, specifically at the 5 m and 100 m resolutions, contain extensive artifactual striping. I strongly suspect that these artifacts are a result of the addition of the coarse resolution

(4 km) variable “winter euphotic mean depth”, which featured in both models, but did not feature in the artifact-free 25 m model. As the environmental covariates were regressed at each resolution, I observed variations in the magnitude of their interactions with each other and their effects on the dependent variable. I wasn’t surprised by the appearance of these interactive oscillations, but I did not expect the emergence of such major striping artifacts as a result of the inclusion of coarse resolution covariates in certain models. As I build on this work, I will further investigate these interactions by rerunning the 5 m and 100 m models without the coarse covariates. I am also interested in developing a coarser model (i.e., $\geq$ 1 km resolution) in order to consider all potential covariates and reduce the possibility of striping within the mapped output.

3.5.4 Limitations and continuing research

Notably, seafloor hardness was a consistently influential variable for determining coral presence or absence at 5 m and 25 m only after I binned values based on standard deviation. Based on the appearance of this variable as a predictor for benthic organismal occurrence after categorization, I caution modelers against prematurely restricting their map output area as delimited exclusively by raw backscatter values; instead, I suggest that researchers carefully consider their overall research goals, and apply backscatter maps to fit those goals. For example, the spatial extent of distribution maps of a marine species that has some known relationship with substrate type (e.g., scleractinian corals or psammophytic algae) should not be restricted based solely on backscatter data. However, submersible or technical dive operations may be able to more efficiently isolate regions of interest, like hard, coral-rich bioherms, by identifying particularly hard substrate. Because backscatter is density dependent, we can assume that corals in dense aggregations will reflect a higher regional backscatter value.

The amount of data associated with my high-resolution (5 m) layers pushed computational limitations and forced us to subdivide my data multiple times to complete the necessary analyses, which introduced the possibility of human error and further slowed my work. While the findings of this work offer novel insight on the effect of different habitat covariates on the distribution of certain prominent mesophotic organisms, I urge spatial analysts to carefully consider their pixel size based on the massive processing power required for fine-scale analyses.

3.6 Acknowledgments

Funding for this research was provided by the University of Hawai‘i Sea Grant College Program grant no. NA14OAR4170071 and NOAA Fisheries no. NA10NMF4520163. I thank J. Potemra and E. Geiger for their assistance in providing environmental covariate data.

3.7 Figures

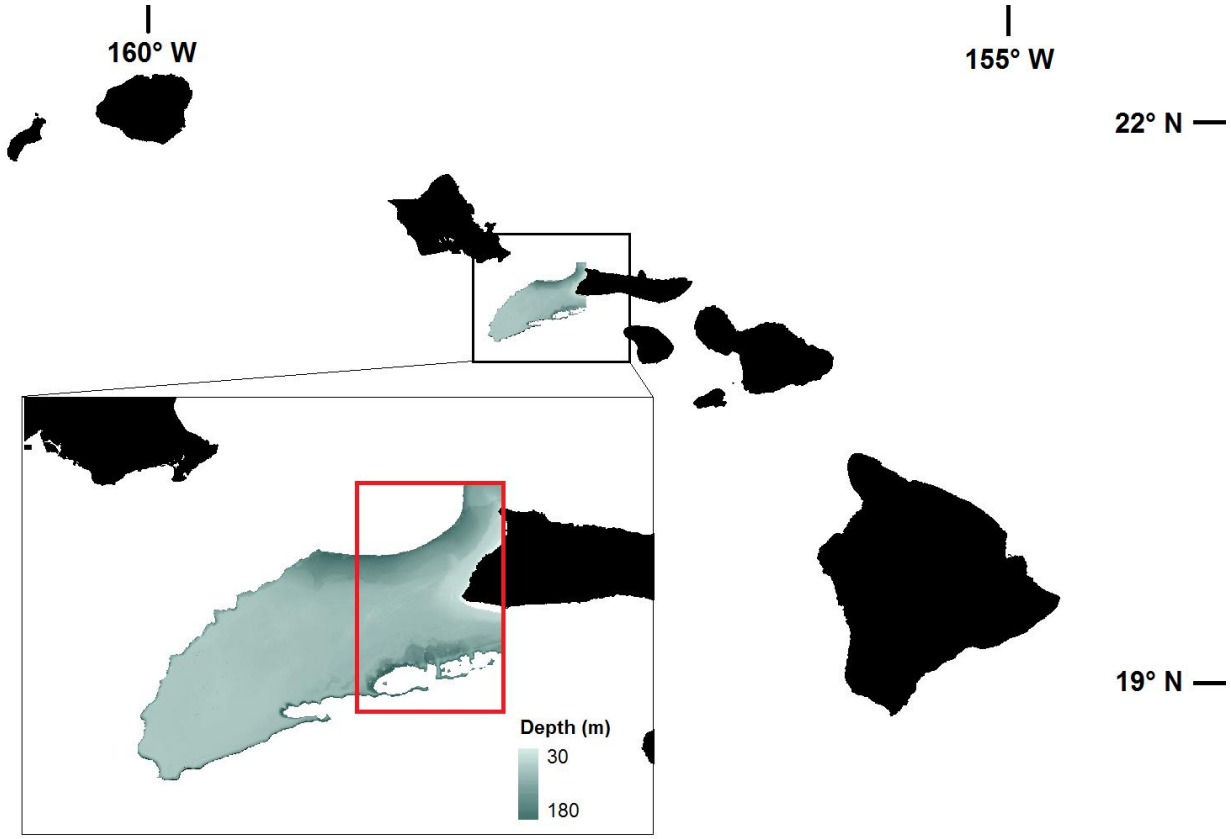

Fig. 1. The geographical extent of the main Hawaiian Islands. The study area of eastern Penguin Bank, Moloka‘i, is visible in the red box in the inset.

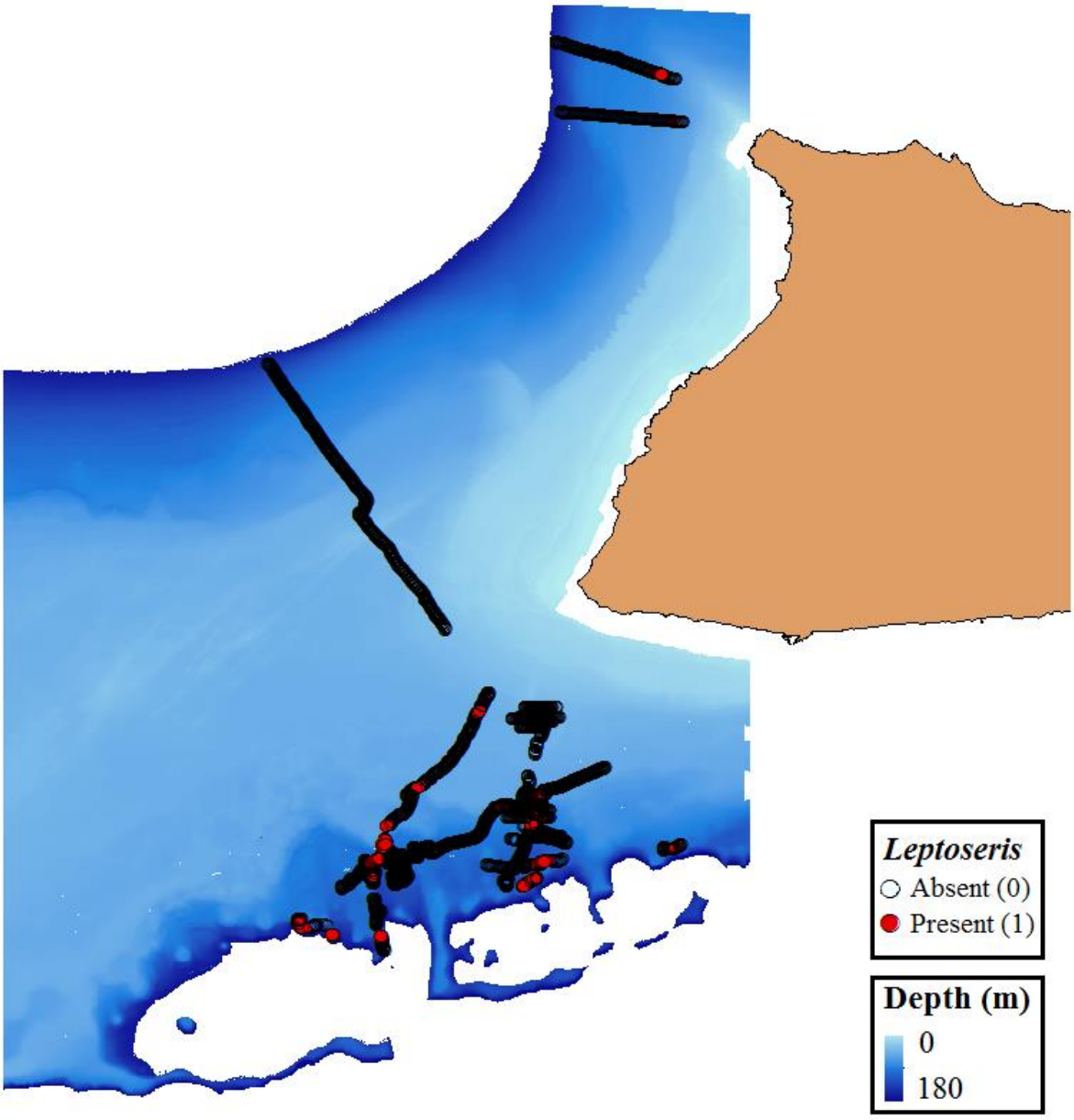

Fig. 2. HURL submersible and ROV survey tracks overlaying the study area. *Leptoseris* sp. presences (red) and absences (open black circle) denoted.

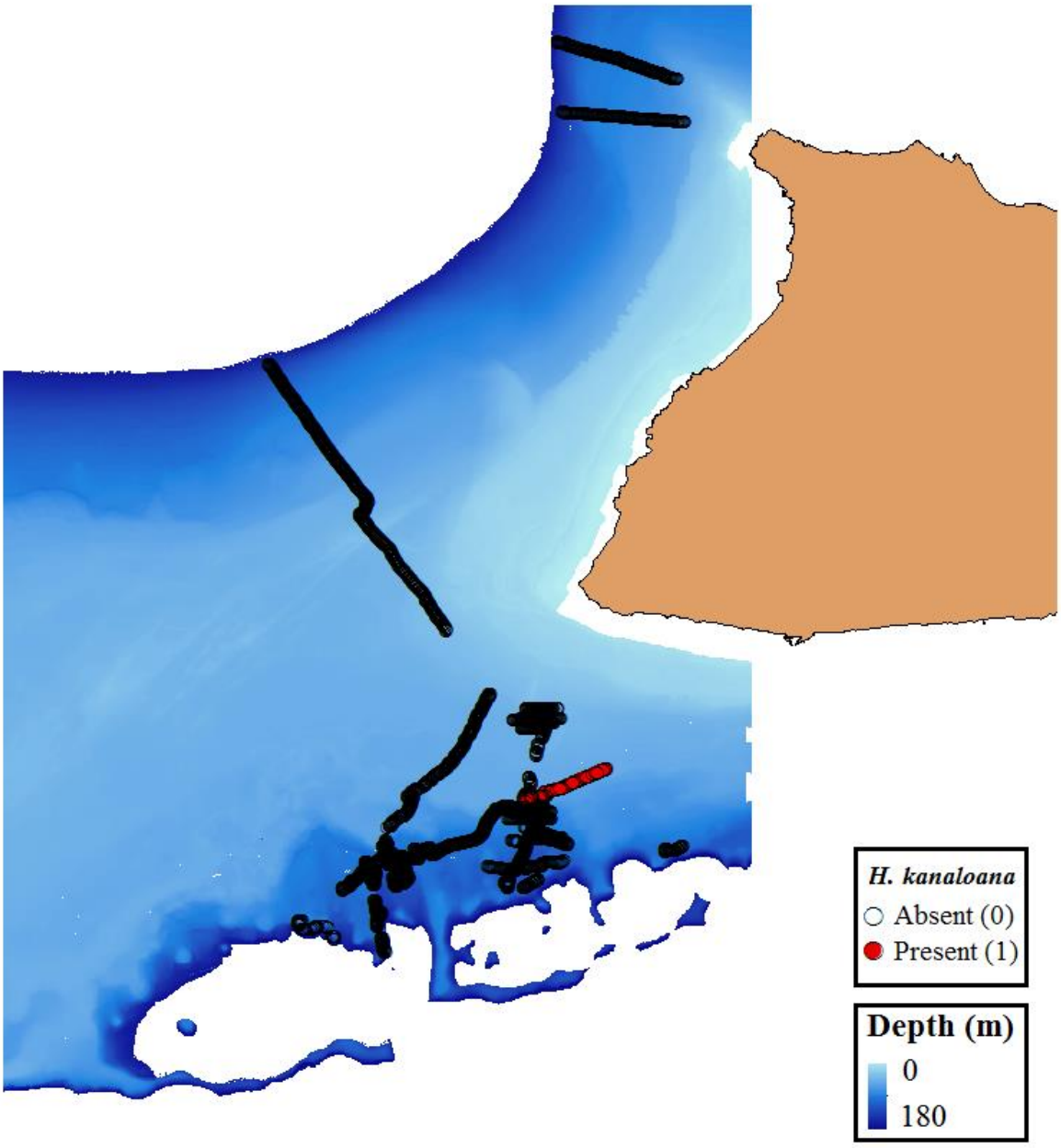

Fig. 3. HURL submersible and ROV survey tracks overlaying the study area. *Halimeda kanaloana* presences (red) and absences (open black circle) denoted.

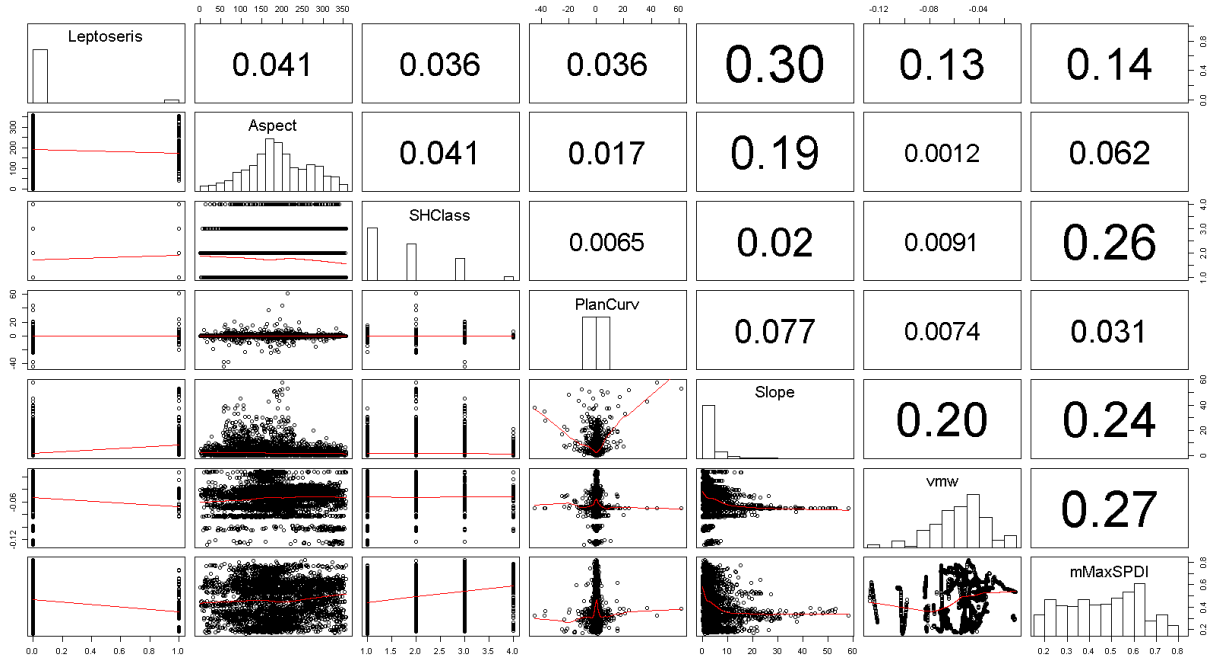

Fig. 4. Correlation scatterplot showing the relationship between the response variable (*Leptoseris*) and covariates included in my final models at 5 m resolution. On the left of the diagonal, locally weighted scatterplot smoothing lines display the relationships between the values of the corresponding variables. On the right of the diagonal, a numerical value expresses the degree of correlation between the corresponding variables.

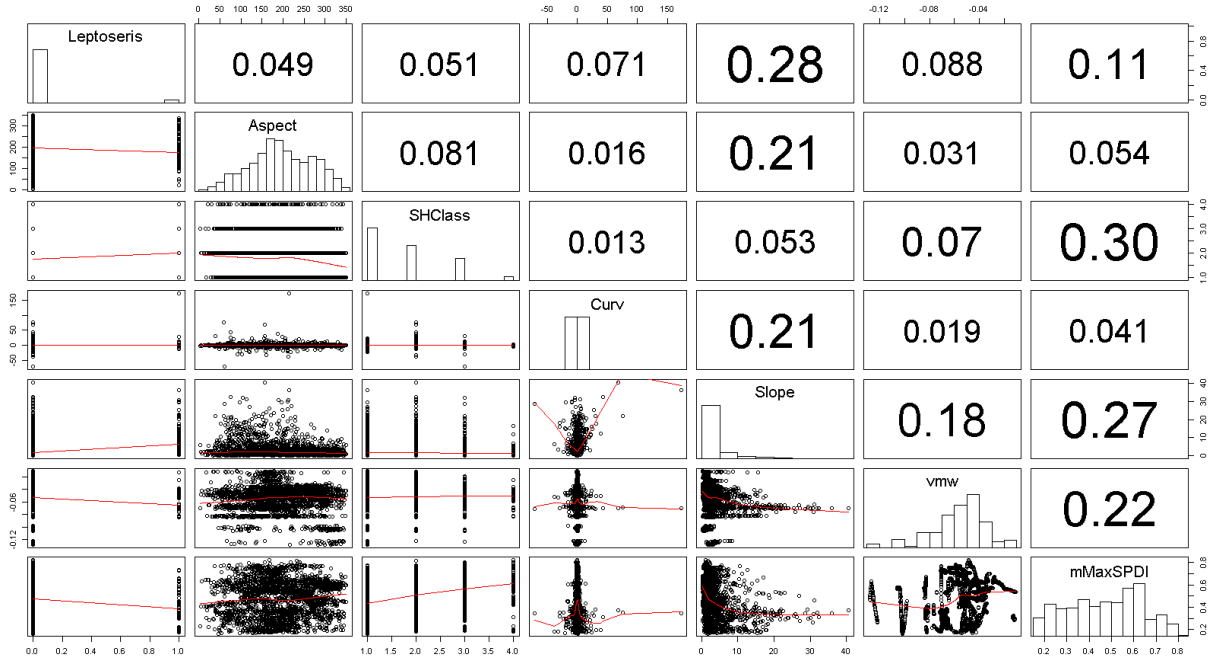

Fig. 5. Correlation scatterplot showing the relationship between the response variable (*Leptoseris*) and covariates included in my final models at 25 m resolution.

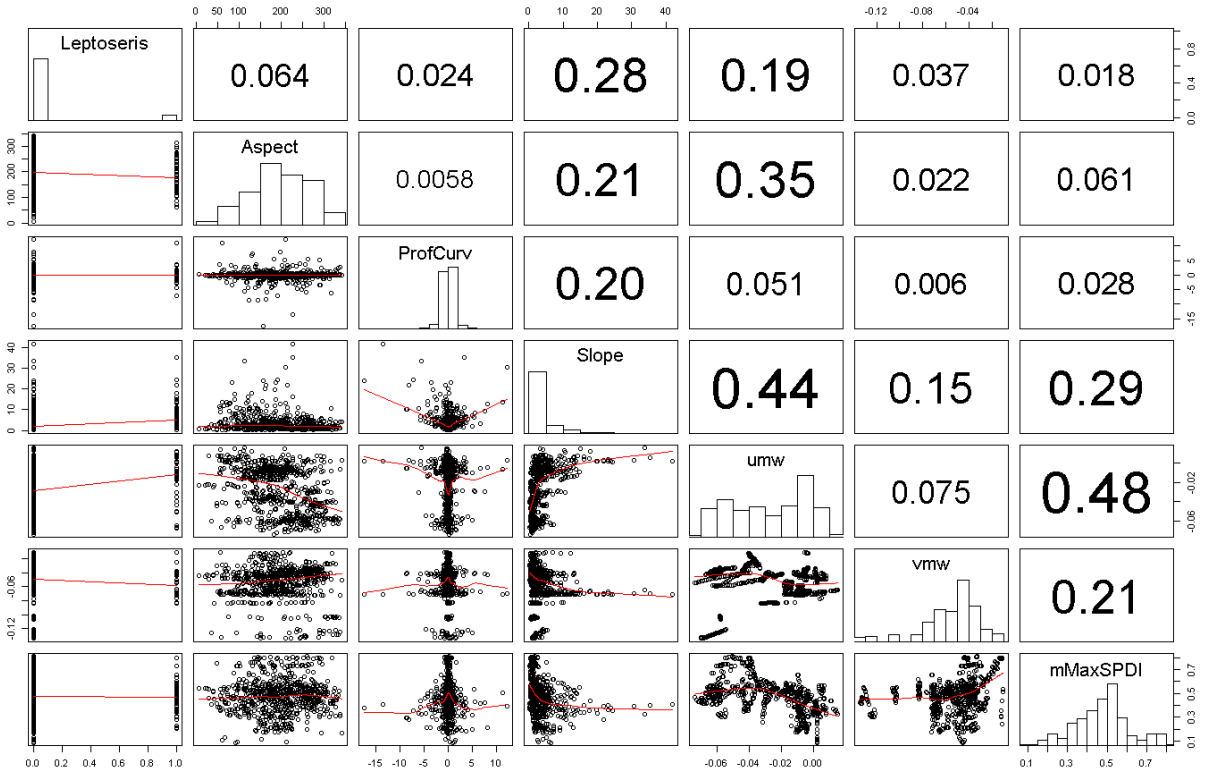

Fig. 6. Correlation scatterplot showing the relationship between the response variable (*Leptoseris*) and covariates included in my final models at 100 m resolution.

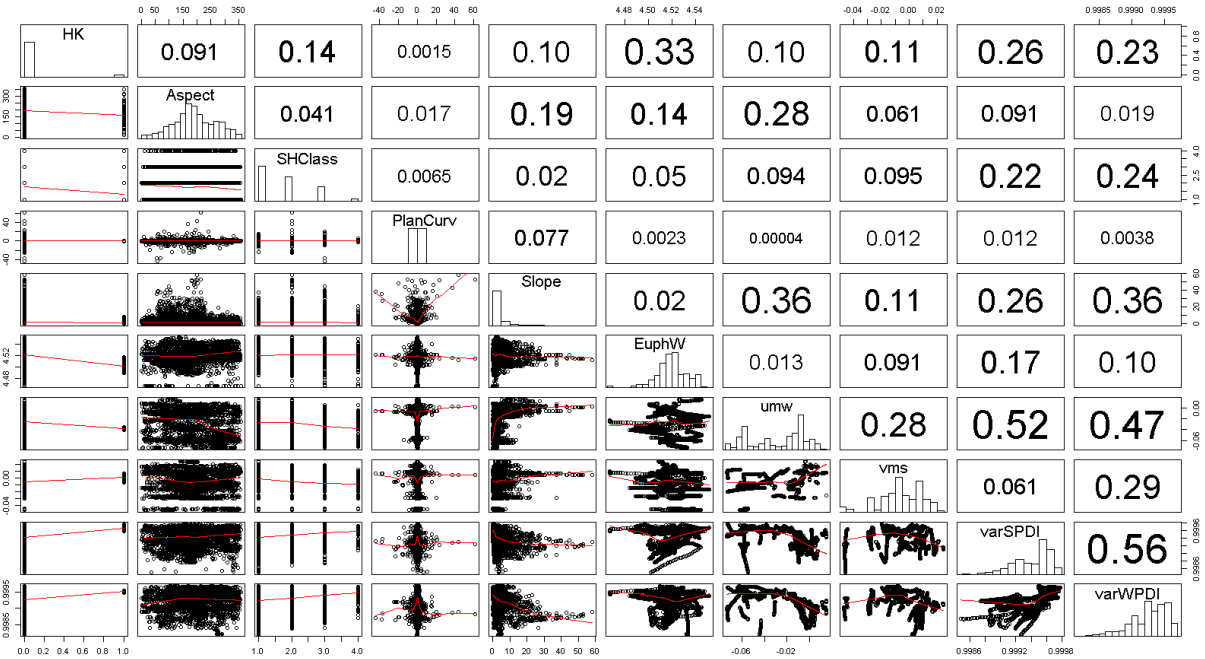

Fig. 7. Correlation scatterplot showing the relationship between the response variable (*H. kanaloana*) and covariates included in my final models at 5 m resolution.

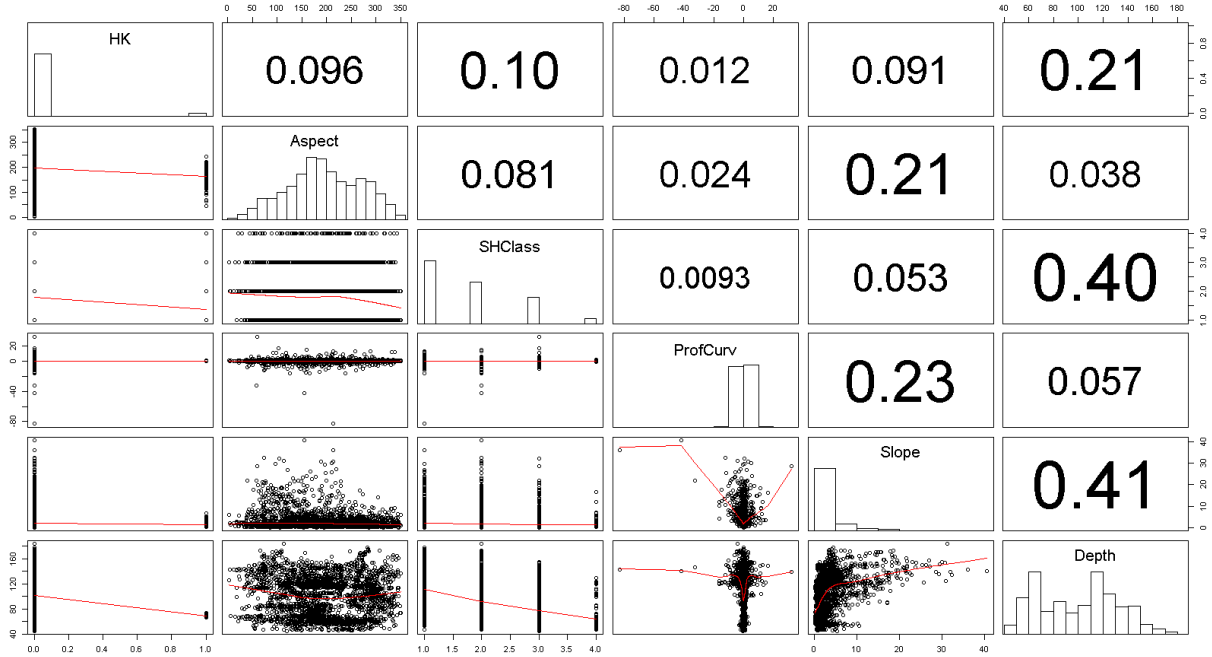

Fig. 8. Correlation scatterplot showing the relationship between the response variable (*H. kanaloana*) and covariates included in my final models at 25 m resolution.

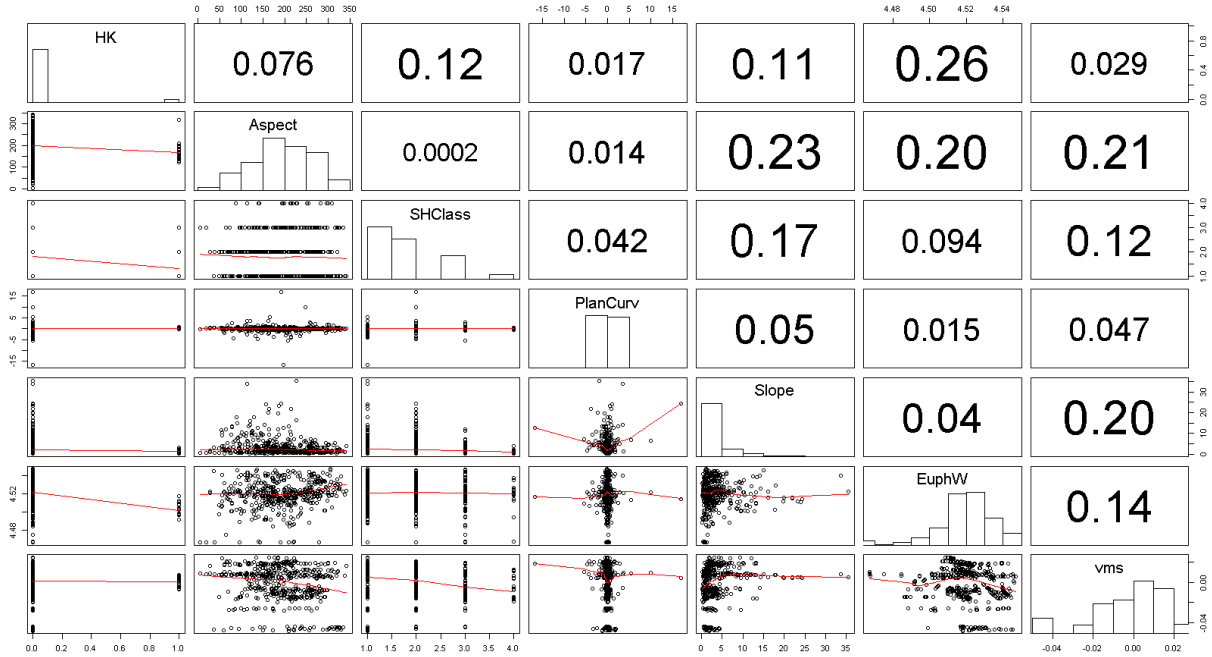

Fig. 9. Correlation scatterplot showing the relationship between the response variable (*H. kanaloana*) and covariates included in my final models at 100 m resolution.

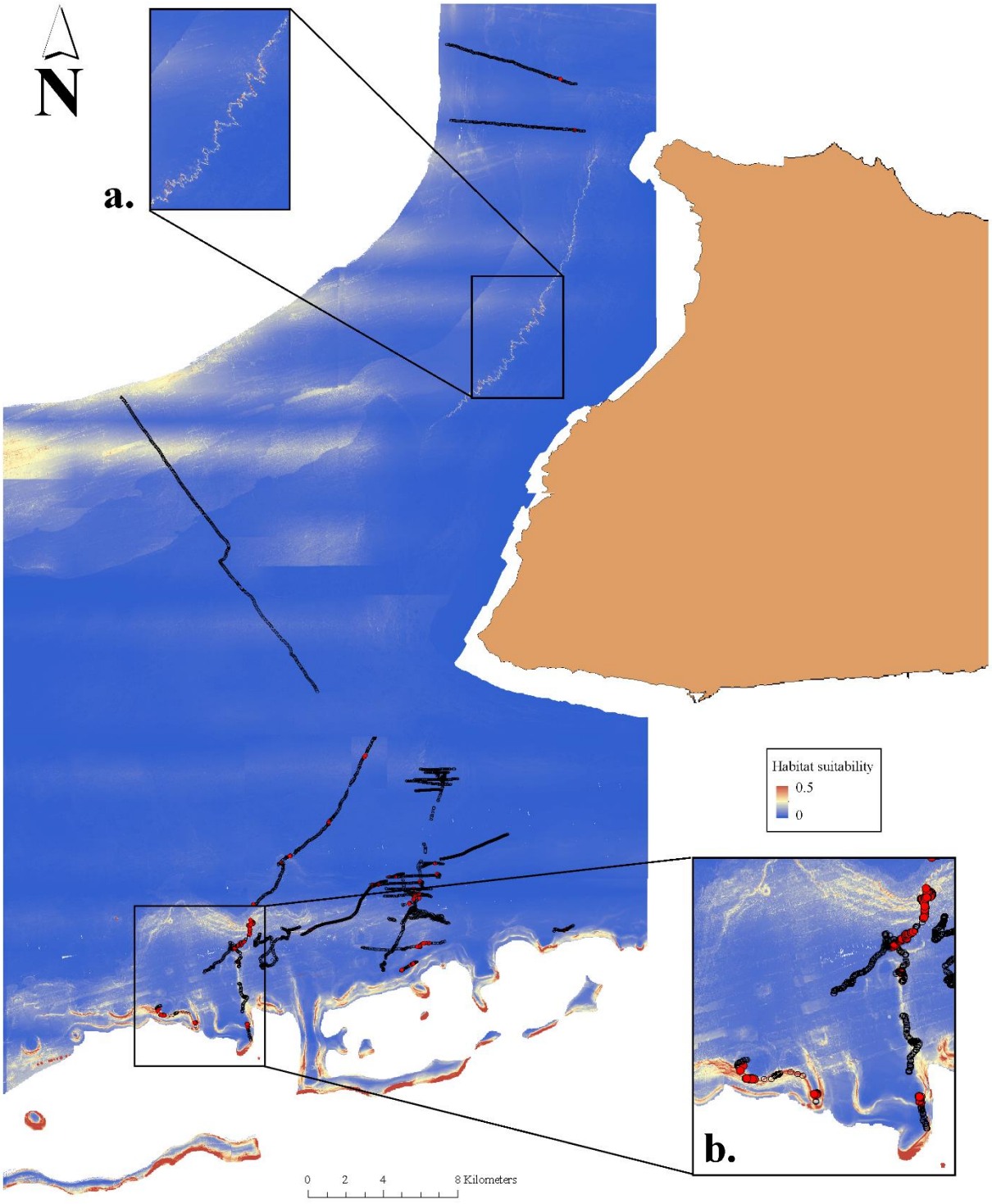

Fig. 10. Predicted habitat suitability for *Leptoseris* sp. across eastern Penguin Bank at 5 m resolution. **a)** Moderate and high suitability was reserved mostly for the southern slopes of the bank, but regions across the bank with high relief and a southern aspect were also predicted to have high suitability for *Leptoseris* colonization. **b)** The southern slopes of the bank are exposed to high current flow from the Kalohi Channel; my model predicted swaths of moderate to high suitability in this region.

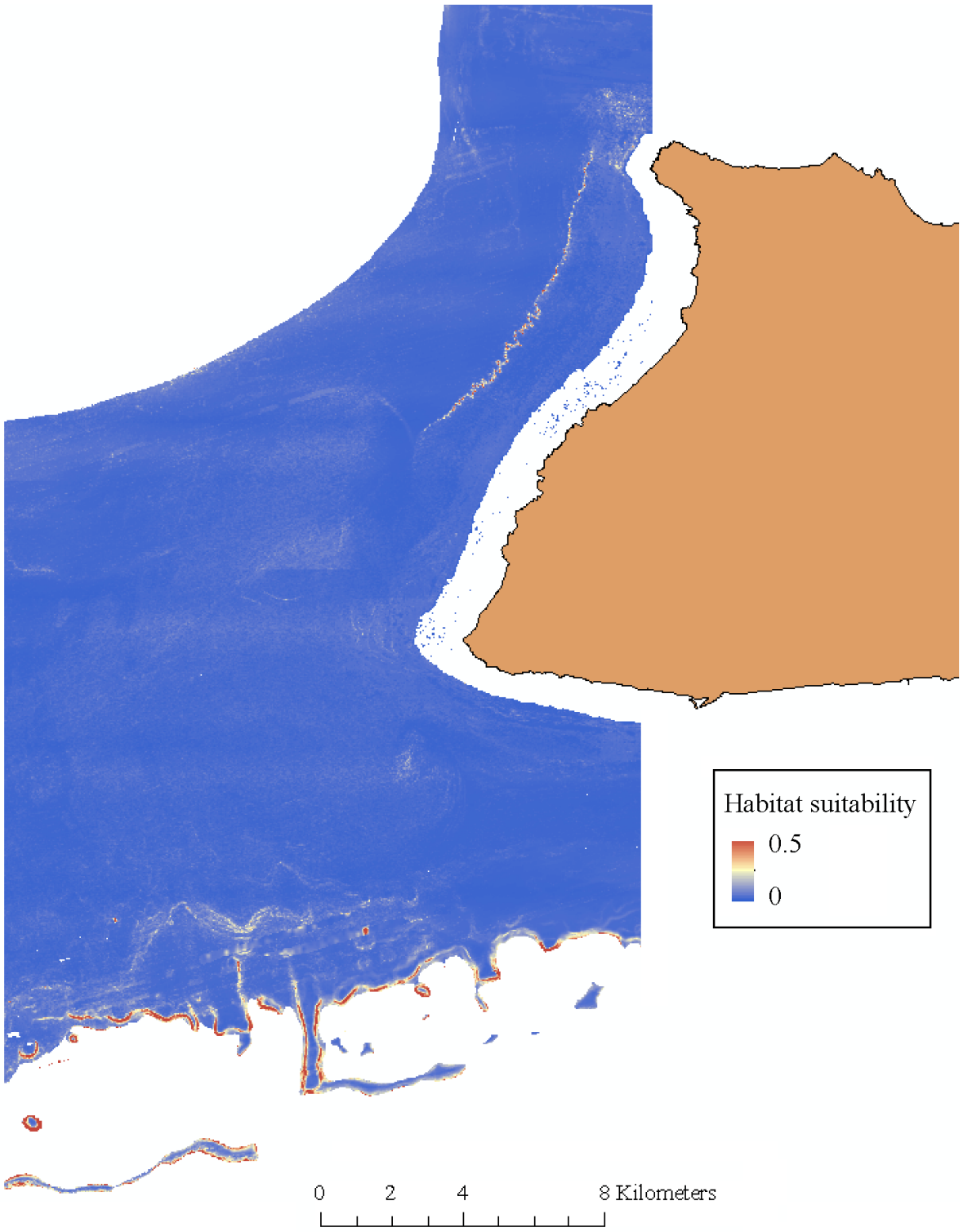

Fig. 11. Predicted habitat suitability for *Leptoseris* sp. across eastern Penguin Bank at 25 m resolution. The model predicted highest suitability across the sloping southern portion of the bank.

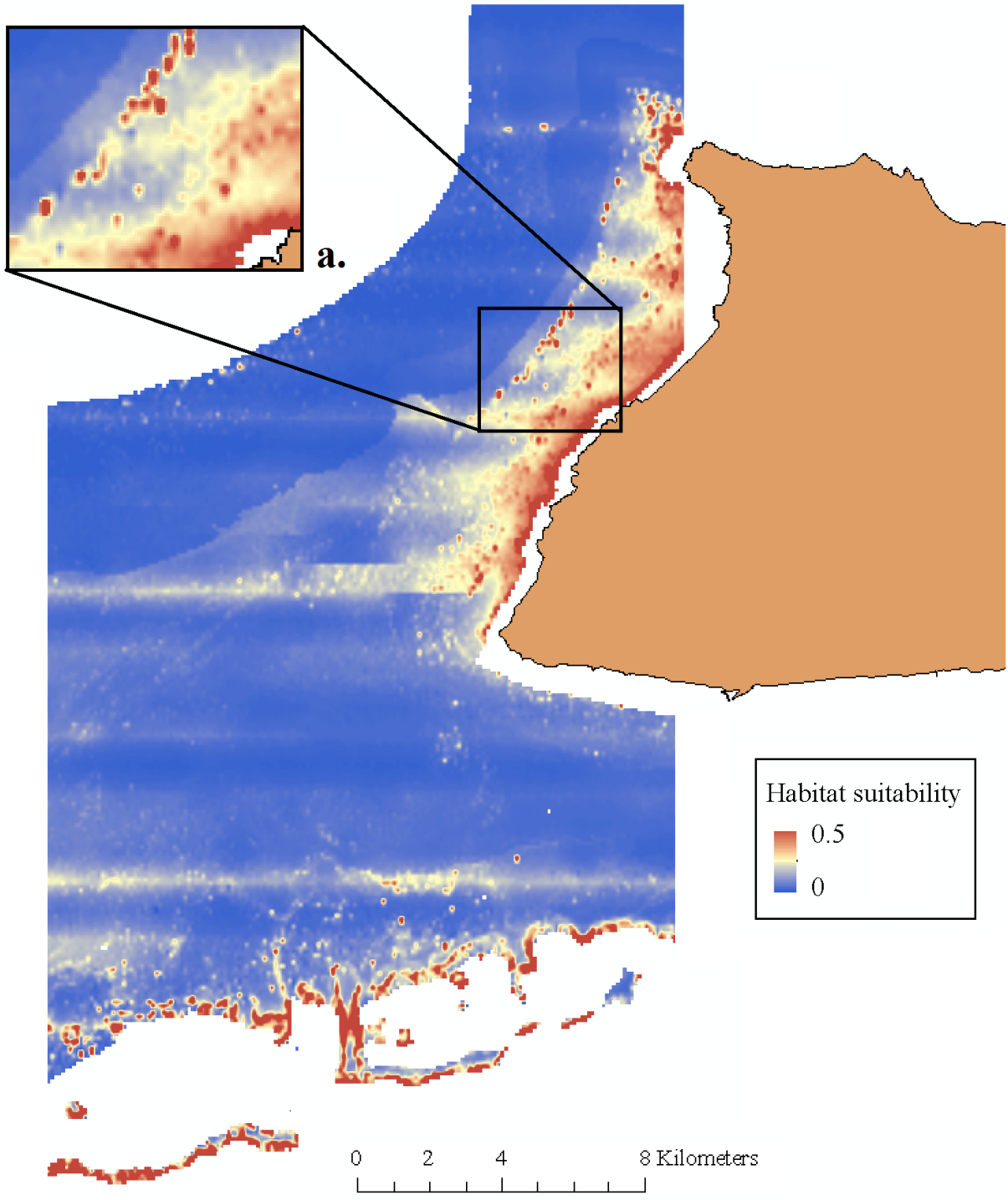

Fig. 12. Predicted habitat suitability for *Leptoseris* sp. across eastern Penguin Bank at 100 m resolution. The highest habitat suitability is predicted along the southern portion of the bank. **a)** The inset shows the higher suitability predicted by this model around nearshore Penguin Bank.

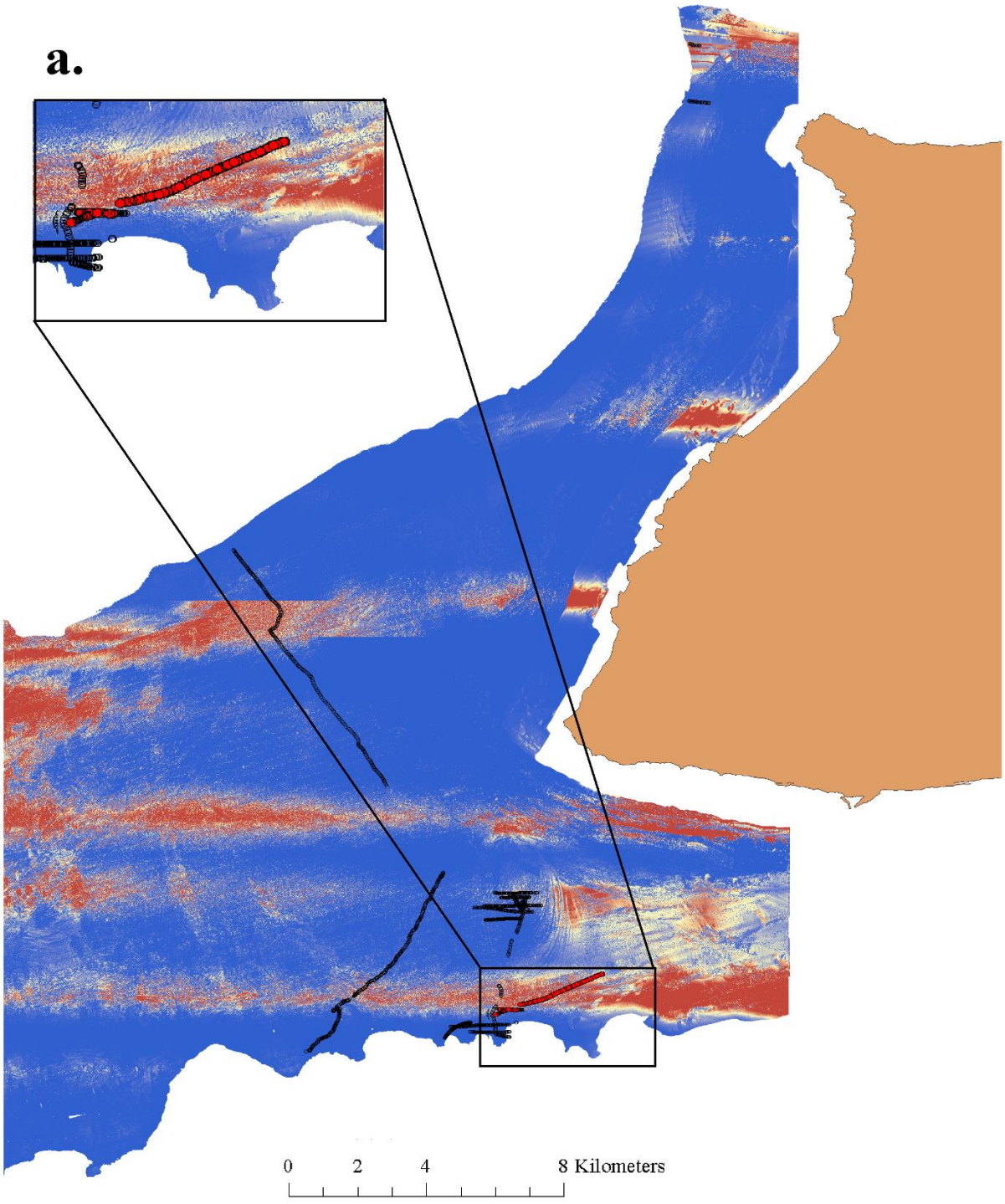

Fig. 13. Predicted habitat suitability for *H. kanaloana* across eastern Penguin Bank at 5 m resolution. **a)** The model predicted high suitability across the flat, sandy portions of the bank.

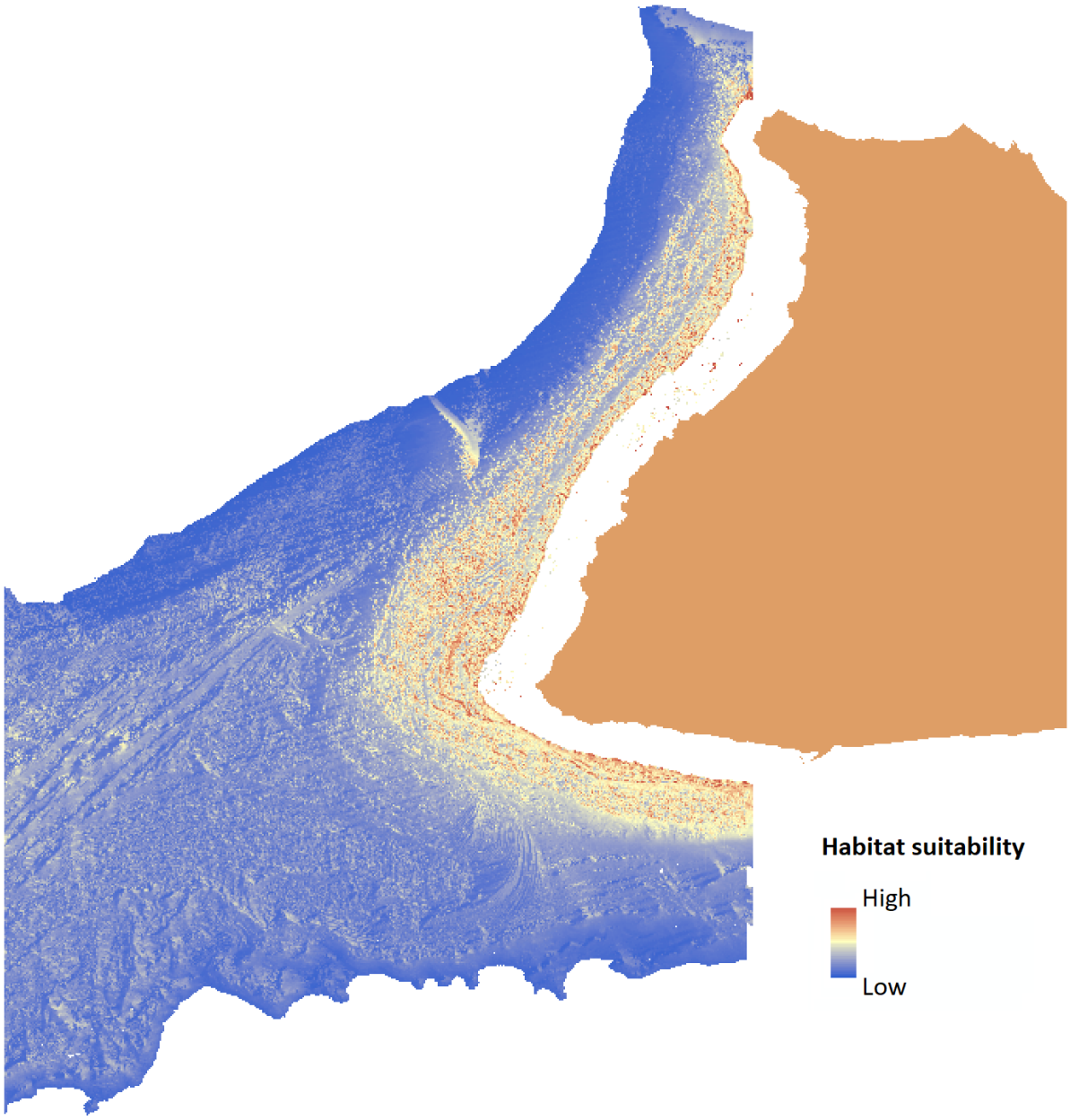

Fig. 14. Predicted habitat suitability for *H. kanaloana* across eastern Penguin Bank at 25 m resolution. Highest suitability was predicted nearshore and in the middle portions of the bank.

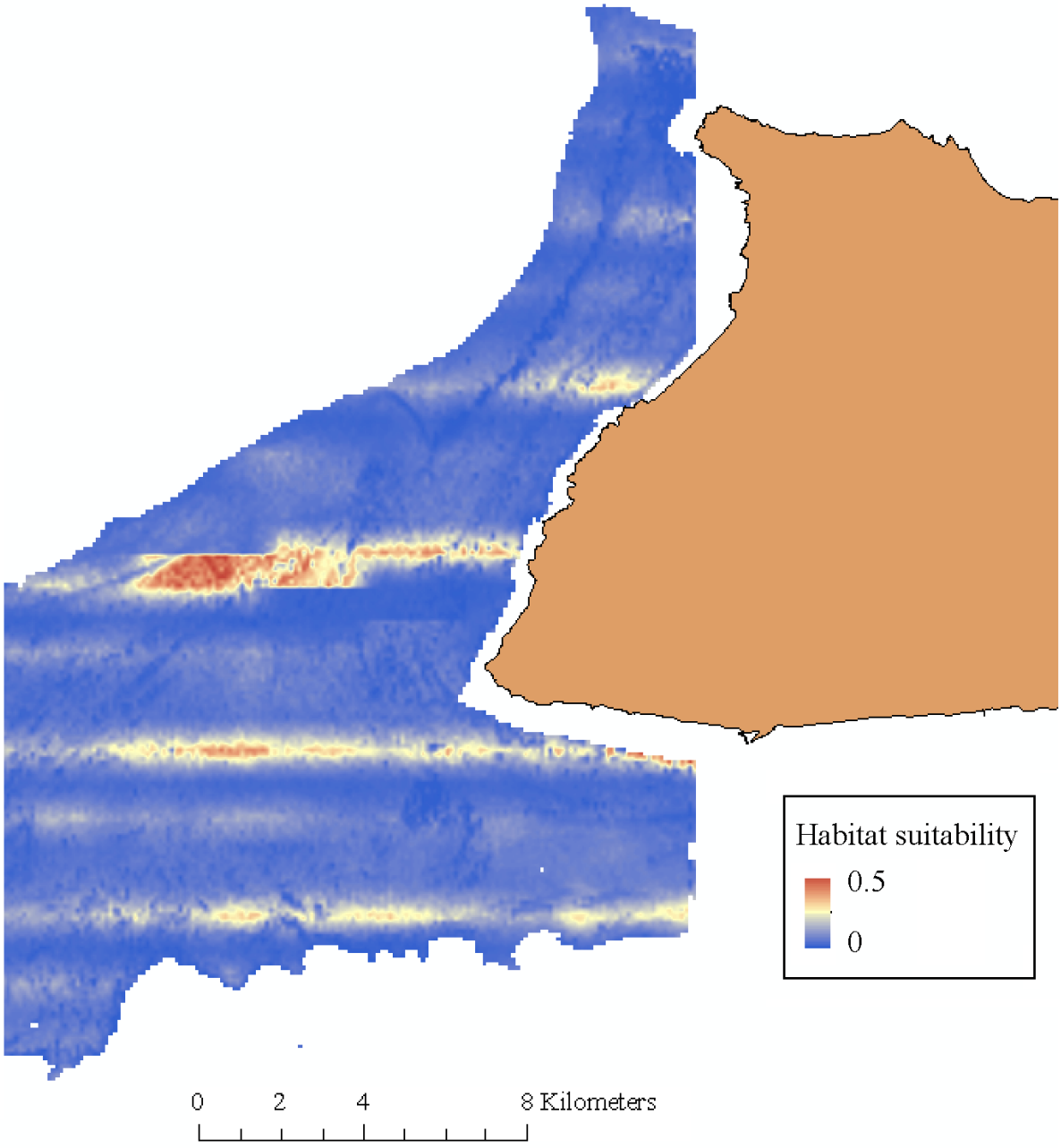

Fig. 15. Predicted habitat suitability for *H. kanaloana* across eastern Penguin Bank at 100 m resolution. Patches of high suitability are predicted across the middle of the bank.

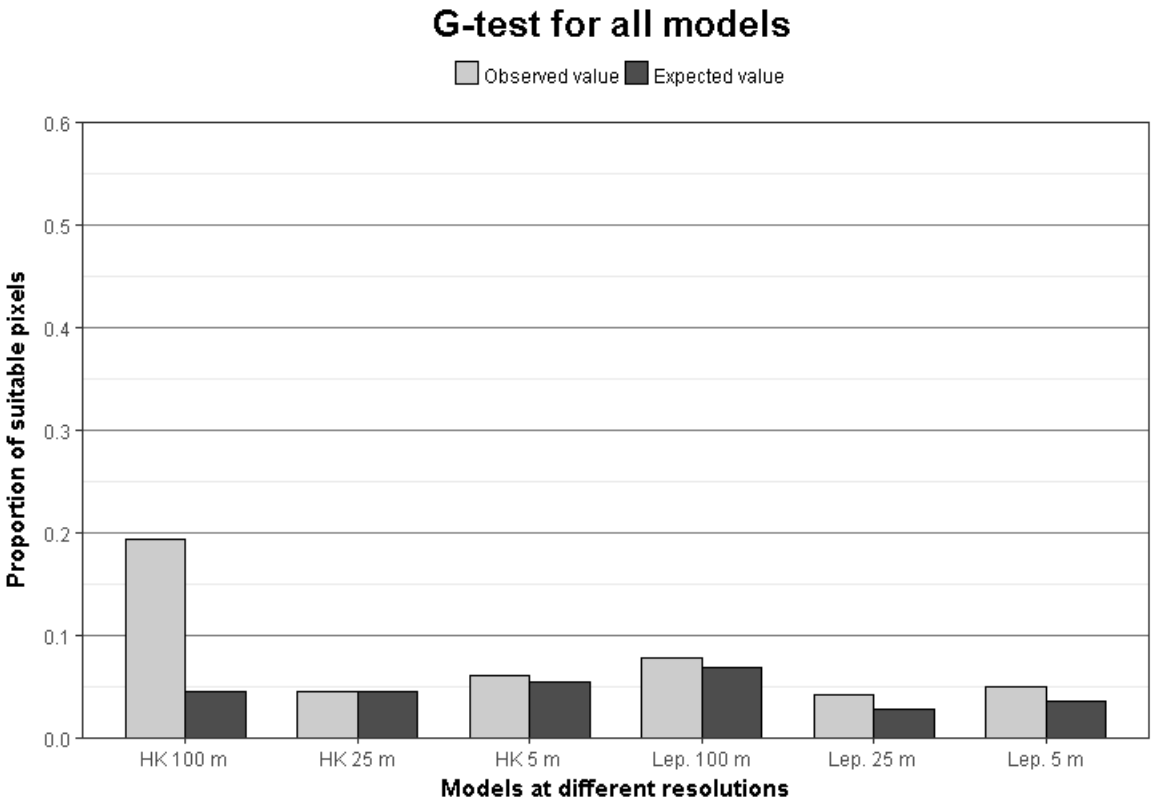

Fig. 16. Results of a *G*-test for goodness of fit for all models based on their identified best threshold value.

3.8 Tables

Table 1. Observational presence/absence data information.

Response variable	Resolution (m)	Number of presence points	Number of absence points	Total
Leptoseris sp.	5	217	4050	4267
	25	120	2451	2571
	100	57	669	726
Halimeda kanaloana	5	242	4025	4267
	25	112	2459	2571
	100	33	693	726

Table 2. List of variables considered for my analyses.

Variable type	Category	Variable description	Source	Resolution	Variable

Biological (response)	Hard coral	Presence/ absence between 10 – 180 m in depth	PIFSC, HURL optical validation data	NA	*Leptoseris* spp.
	Calcareous algae	Presence/ absence between 30 – ~90 m in depth			*Halimeda kanaloana*
Environmental (predictor)	Topographic	Seafloor complexity calculated with the ArcGIS BTM Terrain Ruggedness tool	University of Hawai‘i SOEST, 2016	5 m	Rugosity (unitless)
		Depth of seafloor			Depth (m)
		Rate of change calculated with the ArcGIS BTM Slope tool			Slope (degrees)
		Lateral convexity of seafloor calculated using the ArcGIS Curvature tool			Planform curvature (degrees of degrees)
		Concavity of seafloor calculated using the ArcGIS Curvature tool			Profile curvature (degrees of degrees)
		Curvature of the seafloor calculated using the ArcGIS Curvature tool			Curvature (degrees of degrees)
		Hardness of seafloor detected by acoustic backscatter			Substrate hardness (unitless)
		Seafloor hardness categorized based on subtraction of one standard deviation per bin			Categorized substrate hardness classified from 1 (soft) to 4 (hard) (unitless)
		Compass direction of maximum slope calculated using the ArcGIS Aspect tool			Aspect (degrees)
	Oceanographic	The depth of the euphotic zone (PAR 1%) per season (summer/winter) determined using the Morel method (2007)	NOAA Coral Reef Watch: Satellite Oceanography and Climatology Division (2012 – 2017)	750 m	Mean euphotic depth (m)
		Mean and variance of water column turbidity per season (winter/summer); i.e., the rate at which light at 490 nm attenuates with depth			Diffuse attenuation coefficient (K_d) at 490 nm (m^{-1})
		Mean, variance, mean max, and max variance of proportional downwelled irradiance at depth (z)			
		Mean and variance of directional current velocities per season (winter/summer) for depths: 150, 100, 50 m	PacIOOS Hawai‘i Regional Ocean Model: Data Assimilation (2007 – 2017)	4 km	Mean and variance of current vel. (northward/summer) (m s^{-1})
					Mean and variance of current vel. (northward/winter) (m s^{-1})
					Mean and variance of current vel. (eastward/summer) (m s^{-1})
					Mean and variance of current vel. (eastward/winter) (m s^{-1})

Table 3. Variables included in my analyses following refinement of the covariate list based on high correlation with other variables or evidence of complete separation. Covariates that were most strongly correlated with the dependent variable were reserved and are denoted with a checkmark; covariates that correlated strongly ($\geq$ 0.6) with these variables were removed.

Covariates that appeared highly significant (significance level < 0.05) in my GLMM output per analysis are denoted with an asterisk.

Variable	Inclusion in analyses					
	Lep. (5 m)	*Lep.* (25 m)	*Lep.* (100 m)	*H. k.* (5 m)	*H. k.* (25 m)	*H. k.* (100 m)
Aspect	✓	✓	✓*	✓*	✓*	✓
Curvature		✓				
Current velocity (winter eastward mean)			✓*	✓*		
Current velocity (winter northward mean)	✓*	✓	✓			
Current velocity (summer northward mean)				✓*		
Depth					✓*	
Proportional downwelled irradiance (summer mean max)	✓*	✓*	✓*			
Proportional downwelled irradiance (summer variance)				✓		
Proportional downwelled irradiance (winter variance)				✓		
Euphotic mean depth (winter)				✓*		✓*
Planform curvature	✓		✓*	✓		✓
Profile curvature			✓*		✓	
Slope	✓*	✓*	✓*	✓*	✓	✓*
Substrate hardness (categorized)	✓*	✓*		✓*	✓	

Table 4: Performance statistics of 5 m spatial models.

Model (5 m resolution)	Threshold	Specificity	Sensitivity	AUC
***Lep.* GLMM**	5.00E-02	7.50E-01	7.29E-01	7.69E-01
***Lep.* GLMM (c.v.)**	5.00E-02	7.76E-01	7.37E-01	-
***H. kanaloana* GLMM**	1.00E-01	9.08E-01	8.82E-01	9.45E-01
***H. kanaloana* GLMM (c.v.)**	1.00E-01	9.93E-01	9.59E-01	-

Table 5: Performance statistics of 25 m spatial models.

Model (25 m resolution)	Threshold	Specificity	Sensitivity	AUC
***Lep.* GLMM**	5.00E-02	8.68E-01	6.81E-01	8.06E-01
***Lep.* GLMM (c.v.)**	5.00E-02	8.59E-01	4.69E-01	-
***H. kanaloana* GLMM**	5.00E-02	6.80E-01	1.00E+00	8.08E-01
***H. kanaloana* GLMM (c.v.)**	5.00E-02	6.62E-01	1.00E+00	-

Table 6: Performance statistics of 100 m spatial models.

Model (100 m resolution)	Threshold	Specificity	Sensitivity	AUC
***Lep.* GLMM**	5.00E-02	8.84E-01	5.77E-01	7.82E-01
***Lep.* GLMM (c.v.)**	5.00E-02	8.78E-01	7.27E-01	-
***H. kanaloana* GLMM**	5.00E-02	7.61E-01	9.56E-01	8.90E-01
***H. kanaloana* GLMM (c.v.)**	5.00E-02	7.68E-01	1.00E+00	-

Table S1. Covariates considered for, but not included in, my final models.

Variable
Current velocity (summer eastward mean)
Current velocity (winter eastward variance)
Current velocity (summer eastward variance)
Current velocity (winter northward variance)
Current velocity (summer northward variance)
Kd490 (winter mean)
Kd490 (summer mean)
Kd490 (winter variance)
Kd490 (summer variance)
Proportional downwelled irradiance (summer mean)
Proportional downwelled irradiance (winter mean)
Proportional downwelled irradiance (winter mean max)
Proportional downwelled irradiance (summer variance max)

Proportional downwelled irradiance (winter variance max)
Euphotic mean depth (summer)
Rugosity
Substrate hardness (raw)

Table S2: *H. kanaloana* 5 m GLM descriptive statistics.

Variable	Estimate	Adjusted Std. Error	Pr(>|z|)
Intercept	2.50E+01	1.39E+00	< 2E-16
Aspect	3.70E-04	5.78E-05	< 2E-16
SHClass	-4.49E-02	4.88E-03	< 2E-16
Slope	-3.90E-03	8.19E-04	1.90E-06
EuphW	-5.49E+00	3.07E-01	< 2E-16
umw	-1.43E+00	1.95E-01	< 2E-16
vms	2.11E+00	2.97E-01	< 2E-16
PlanCurv	4.72E-04	1.71E-03	7.82E-01

Table S3: *H. kanaloana* 5 m GLMM descriptive statistics.

Variable	Estimate	Adjusted Std. Error	Pr(>|z|)
Intercept	5.49E+02	3.75E+01	< 2E-16
Aspect	-7.50E-03	1.34E-03	< 2E-16
SHClass	-1.97E+00	1.89E-02	< 2E-16
Slope	-2.68E-01	6.68E-02	1.00E-04
EuphW	-1.22E+02	8.30E+00	< 2E-16
umw	-4.17E+01	5.50E+00	< 2E-16
vms	7.17E+01	1.02E+01	< 2E-16
PlanCurv	-8.97E-02	1.55E-01	5.62E-01

Table S4: *Leptoseris* 5 m GLM descriptive statistics.

Variable	Estimate	Adjusted Std. Error	Pr(>|z|)
Intercept	-3.08E+00	5.00E-01	< 2E-16
SHClass	3.20E-01	1.11E-01	4.00E-03
Slope	8.24E-02	1.00E-02	< 2E-16
vmw	-1.48E+01	4.98E+00	3.00E-03
mMaxSPDI	-4.50E+00	7.12E-01	< 2E-16
Aspect	6.06E-04	1.28E-03	6.37E-01
PlanCurv	8.18E-03	1.95E-02	6.75E-01

Table S5: *Leptoseris* 5 m GLMM descriptive statistics.

Variable	Estimate	Adjusted Std. Error	Pr(>|z|)
Intercept	-3.10E+00	5.22E-01	< 2E-16
SHClass	3.21E-01	1.14E-01	5.05E-03
Slope	8.25E-02	1.10E-02	< 2E-16
vmw	-1.47E+01	5.11E+00	3.98E-03
mMaxSPDI	-4.51E+00	7.32E-01	< 2E-16
Aspect	5.98E-04	1.32E-03	3.21E-01
PlanCurv	8.06E-03	2.00E-02	6.88E-01

Table S6: *H. kanaloana* 25 m GLM descriptive statistics.

Variable	Estimate	Adjusted Std. Error	Pr(>|z|)
Intercept	1.89E+00	5.58E-01	7.00E-04
Aspect	-7.20E-03	2.00E-03	4.00E-05
Depth	-4.30E-02	7.00E-03	< 2E-16
Slope	-1.04E-01	8.00E-02	1.93E-01

ProfCurv	3.48E-02	1.36E-01	7.98E-01

Table S7: *H. kanaloana* 25 m GLMM descriptive statistics.

Variable	Estimate	Adjusted Std. Error	Pr(>|z|)
Intercept	1.89E+00	4.12E-01	< 2E-16
Aspect	-7.20E-03	1.30E-03	< 2E-16
Depth	-4.30E-02	4.60E-03	< 2E-16
Slope	-1.04E-01	5.90E-02	7.86E-02
ProfCurv	3.40E-01	9.83E-02	7.11E-01

Table S8: *Leptoseris* 25 m GLM descriptive statistics.

Variable	Estimate	Adjusted Std. Error	Pr(>|z|)
Intercept	-3.79E+00	5.98E-01	< 2E-16
SHClass	4.17E-01	1.48E-01	4.90E-03
Slope	1.37E-01	1.89E-02	< 2E-16
mMaxSPDI	-2.11E+00	8.73E-01	1.59E-02
vmw	-7.13E+00	6.74E+00	2.90E-01
Aspect	8.57E-04	1.78E-03	6.31E-01
Curv	4.41E-03	9.60E-03	6.46E-01

Table S9: *Leptoseris* 25 m GLMM descriptive statistics.

Variable	Estimate	Adjusted Std. Error	Pr(>|z|)
Intercept	-3.84E+00	4.75E-01	< 2E-16
SHClass	3.75E-01	1.35E-01	5.50E-03
Slope	1.32E-01	1.71E-02	< 2E-16

mMaxSPDI	-1.87E+00	8.11E-01	2.20E-02
vmw	1.00E-01	6.55E+00	9.88E-01
Aspect	7.34E-04	1.69E-03	6.64E-01
Curv	-4.00E-03	9.85E-03	6.85E-01

Table S10: *H. kanaloana* 100 m GLM descriptive statistics.

Variable	**Estimate**	**Adjusted Std. Error**	**Pr(>|z|)**
Intercept	2.16E+02	5.83E+01	4.21E-02
Slope	-4.14E-01	2.02E-01	4.07E-02
EuphW	-4.85E+01	1.17E+01	3.90E-05
Aspect	3.61E-03	3.37E-02	2.86E-01
PlanCurv	2.22E-01	3.20E-01	4.88E-01

Table S11: *H. kanaloana* 100 m GLMM descriptive statistics.

Variable	**Estimate**	**Adjusted Std. Error**	**Pr(>|z|)**
Intercept	2.16E+02	3.80E+01	< 2E-16
Slope	-4.10E-01	1.76E-01	8.00E-03
EuphW	-4.84E+01	8.51E+00	< 2E-16
Aspect	3.61E-03	2.46E-03	1.44E-01
PlanCurv	2.22E-01	2.34E-01	3.44E-01

Table S12: *Leptoseris* 100 m GLM descriptive statistics.

Variable	**Estimate**	**Adjusted Std. Error**	**Pr(>|z|)**
Intercept	-9.30E-02	5.33E-02	8.18E-02
ProfCurv	1.74E-02	7.12E-03	1.46E-02

Slope	2.08-02	3.29E-03	< 2E-16
umw	1.34E+00	6.11E-01	2.91E-02
mMaxSPDI	2.78E-01	9.46E-02	3.42E-03
vmw	-4.341E-01	5.02E-01	3.88E-01
Aspect	6.49E0-05	1.79E-04	7.16E-01

Table S13: *Leptoseris* 100 m GLMM descriptive statistics.

Variable	**Estimate**	**Adjusted Std. Error**	**Pr(>|z|)**
Intercept	-4.95E+00	1.01E+00	< 2E-16
ProfCurv	1.94E-01	9.21E-02	3.67E-02
Slope	1.53-01	4.06E-02	2.00E-04
umw	3.39E+01	1.27E+01	7.76E-03
mMaxSPDI	4.80E+00	9.46E-02	3.42E-03
vmw	-4.341E-01	5.02E-01	3.88E-01
Aspect	6.49E0-05	1.75E+00	6.50E-03

Table S14: Summary statistics for environmental predictor variables for each training dataset at 5 m resolution.

Variable (5 m resolution)	**Min**	**Mean**	**Max**
Longitude	666712	673283	679172
Latitude	2323341	2330626	2350626
Leptoseris spp.	0	0.051	1
Halimeda kanaloana	0	0.057	1
Aspect	0.704	190.793	357.257
Curvature	-72.908	0.075	120.15
Current velocity (winter eastward mean)	-0.074	-0.028	0.015
Current velocity (summer eastward mean)	-0.088	-0.034	0.042
Current velocity (winter northward mean)	-0.128	-0.055	-0.0111
Current velocity (summer northward mean)	-0.047	-0.005	0.025
Current velocity (winter eastward variance)	0.012	0.023	0.038
Current velocity (summer eastward variance)	0.011	0.024	0.039
Current velocity (winter northward variance)	0.005	0.034	0.073
Current velocity (summer northward variance)	0.004	0.031	0.07
Depth	44.11	100.57	177.43

Diffuse attenuation coefficient at 490 nm (Kd490) (winter mean)	0.008	0.013	0.023
Diffuse attenuation coefficient at 490 nm (Kd490) (summer mean)	0.002	0.006	0.012
Diffuse attenuation coefficient at 490 nm (Kd490) (winter variance)	5.27E-06	7.34E-06	1.21E-05
Diffuse attenuation coefficient at 490 nm (Kd490) (summer variance)	2.28E-06	5.83E-06	1.06E-05
Proportional downwelled irradiance (summer mean)	0.264	0.58	0.895
Proportional downwelled irradiance (winter mean)	0.053	0.287	0.623
Proportional downwelled irradiance (summer mean max)	0.164	0.461	0.819
Proportional downwelled irradiance (winter mean max)	0.001	0.082	0.329
Proportional downwelled irradiance (summer variance)	0.999	0.999	0.999
Proportional downwelled irradiance (winter variance)	0.998	0.999	1
Proportional downwelled irradiance (summer variance max)	0.997	0.999	1
Proportional downwelled irradiance (winter variance max)	0.995	0.999	1
Euphotic mean depth (winter)	4.467	4.52	4.552
Euphotic mean depth (summer)	4.534	4.563	4.582
Planform curvature	-44.431	-0.018	61.428
Profile curvature	-58.722	-0.094	30.339
Rugosity	0	0.001	0.075
Slope	0.017	3.583	58.125
Substrate hardness (raw)	6.799	113.494	252.449
Substrate hardness (categorized)	1	1.806	4

Table S15: Summary statistics for environmental predictor variables for each training dataset at 25 m resolution.

Variable (25 m resolution)	**Min**	**Mean**	**Max**
Longitude	666712	673173	679162
Latitude	2323341	2331408	2350616
Leptoseris spp.	0	0.047	1
Halimeda kanaloana	0	0.044	1
Aspect	3.298	195.525	351.005
Curvature	-70.312	0.198	172.26
Current velocity (winter eastward mean)	-0.074	-0.029	0.015
Current velocity (summer eastward mean)	-0.088	-0.033	0.042
Current velocity (winter northward mean)	-0.128	-0.056	-0.011
Current velocity (summer northward mean)	-0.047	-0.006	0.025
Current velocity (winter eastward variance)	0.012	0.024	0.038
Current velocity (summer eastward variance)	0.011	0.023	0.039
Current velocity (winter northward variance)	0.005	0.035	0.073
Current velocity (summer northward variance)	0.004	0.032	0.07
Depth	44.85	99.92	177.8
Diffuse attenuation coefficient at 490 nm (Kd490) (winter mean)	0.008	0.013	0.023
Diffuse attenuation coefficient at 490 nm (Kd490) (summer mean)	0.002	0.006	0.012
Diffuse attenuation coefficient at 490 nm (Kd490) (winter variance)	5.29E-06	7.40E-06	1.20E-05
Diffuse attenuation coefficient at 490 nm (Kd490) (summer variance)	2.29E-06	5.77E-06	1.06E-05
Proportional downwelled irradiance (summer mean)	0.264	0.588	0.896

Proportional downwelled irradiance (winter mean)	0.053	0.293	0.624
Proportional downwelled irradiance (summer mean max)	0.165	0.472	0.82
Proportional downwelled irradiance (winter mean max)	0.001	0.083	0.332
Proportional downwelled irradiance (summer variance)	0.999	0.999	1
Proportional downwelled irradiance (winter variance)	0.998	0.999	1
Proportional downwelled irradiance (summer variance max)	0.997	0.999	1
Proportional downwelled irradiance (winter variance max)	0.995	0.999	1
Euphotic mean depth (winter)	4.467	4.52	4.551
Euphotic mean depth (summer)	4.535	4.564	4.582
Planform curvature	-38.41	0.057	89.366
Profile curvature	-82.894	-0.139	31.902
Rugosity	0	0.001	0.119
Slope	0.131	3.234	40.527
Substrate hardness (raw)	21.98	115.82	239.26
Substrate hardness (categorized)	1	1.809	4

Table S16: Summary statistics for environmental predictor variables for each training dataset at 100 m resolution.

Variable (500 m resolution)	**Min**	**Mean**	**Max**
Longitude	666712	672784	679212
Latitude	2323341	2332553	2350841
Leptoseris spp.	0	0.079	1
Halimeda kanaloana	0	0.047	1
Aspect	5.457	191.823	341.654
Curvature	-24.911	-0.070	31.273
Current velocity (winter eastward mean)	-0.072	-0.028	0.015
Current velocity (summer eastward mean)	-0.087	-0.033	0.042
Current velocity (winter northward mean)	-0.134	-0.054	-0.011
Current velocity (summer northward mean)	-0.048	-0.002	0.024
Current velocity (winter eastward variance)	0.012	0.020	0.038
Current velocity (summer eastward variance)	0.011	0.021	0.039
Current velocity (winter northward variance)	0.005	0.035	0.079
Current velocity (summer northward variance)	0.004	0.031	0.076
Depth	46.36	98.60	178.01
Diffuse attenuation coefficient at 490 nm (Kd490) (winter mean)	0.008	0.014	0.023
Diffuse attenuation coefficient at 490 nm (Kd490) (summer mean)	0.002	0.005	0.012
Diffuse attenuation coefficient at 490 nm (Kd490) (winter variance)	5.29E-06	7.48E-06	1.21E-05
Diffuse attenuation coefficient at 490 nm (Kd490) (summer variance)	2.30E-06	5.76E-06	1.03E-06
Proportional downwelled irradiance (summer mean)	0.265	0.594	0.889
Proportional downwelled irradiance (winter mean)	0.044	0.298	0.616
Proportional downwelled irradiance (summer mean max)	0.090	0.481	0.808
Proportional downwelled irradiance (winter mean max)	2.00E-05	0.091	0.321
Proportional downwelled irradiance (summer variance)	0.999	0.999	1
Proportional downwelled irradiance (winter variance)	0.998	0.999	1

Proportional downwelled irradiance (summer variance max)	0.997	0.999	1
Proportional downwelled irradiance (winter variance max)	0.996	0.999	1
Euphotic mean depth (winter)	4.467	4.520	4.550
Euphotic mean depth (summer)	4.535	4.564	4.581
Planform curvature	-16.79	-0.106	17.77
Profile curvature	-17.48	-0.036	8.121
Rugosity	4.40E-07	6.57E-04	4.77E-03
Slope	0.196	3.254	41.897
Substrate hardness (raw)	20.99	116.91	216.31
Substrate hardness (categorized)	1	1.834	4

3.9 References

Araujo, M. B., Pearson, R. G., Thuiller, W., and Erhard, M. (2005). Validation of species–climate impact models under climate change. *Global Change Biology* 11(9): 1504 – 1513.

Baker, C. S., and Herman, L. M., (1981). Migration and local movement of humpback whales (*Megaptera novaeangliae*) through Hawaiian waters. *Canadian Journal of Zoology* 59(3): 460 – 469.

Burnham, K. P., and Anderson, D. R. (2002). Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, Second Edition. New York: Springer-Verlag.

Costa, B., Kendall, M. S., Parrish, F. A., Rooney, J., Boland, R. C., Chow, M., Lecky, J., Montgomery, A., and Spalding, H. (2015). Identifying suitable locations for mesophotic hard corals offshore of Maui, Hawai‘i. *PLoS ONE* 10: e0130285.

Dinesen, Z. D. (1980). A revision of the coral genus *Leptoseris* (Scleractinia: Fungiina: Agariciidae). *Memoirs of the Queensland Museum* 20(1): 181 – 235.

Dormann, F. C., McPherson, J. M., Araújo, M. B., Bivand, R., Bolliger, J., Carl, G., Davies, R. G., Hirzel, A., Jetz, W., Kissling, W. D., Kühn, I., Ohlemüller, R., Peres-Neto, P. R., Reineking, B., Schröder, B., Schurr, F. M., and Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30: 609 – 628.

Drew, E. A., and Abel, K. M. (1988). Studies on *Halimeda*. The distribution and species composition of *Halimeda* meadows throughout the Great Barrier Reef Province. *Coral Reefs* 6: 195 – 205.

ESRI (2012). ArcGIS Desktop: Release 10.1. Redlands, CA: Environmental Systems Research Institute.

Franklin, E. C., Jokiel, P. L., and Donahue M. J. (2013). Predictive modeling of coral distribution and abundance in the Hawaiian Islands. *Marine Ecology Progress Series* 481: 121 – 132.

Fricke, H.W., Vareschi, E., and Schlichter, D. (1987). Photoecology of the coral *Leptoseris fragilis* in the Red Sea twilight zone (an experimental study by submersible). *Oecologia* 73(3): 371 – 381.

Geary, R. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician* 5: 115 – 146.

Glynn, P. W. (1996). Coral reef bleaching: facts, hypotheses and implications. *Global Change Biology* 2: 495 – 509.

Gregory III, A. E., and Kroenke, L. W. (1982). Reef development on a mid-oceanic island: reflection profiling studies of the 500-meter shelf south of Oahu. *AAPG Bulletin* 66(7): 843 – 859.

Haight, W. R., Parrish, J. D., and Hayes, T. A. (1993). Feeding ecology of deepwater lutjanid snappers at Penguin Bank, Hawaii. *Transactions of the American Fisheries Society* 122.3: 328 – 347.

Hosmer, D., and Lemeshow, S. (2004). Applied logistic regression. Hoboken, New Jersey: John Wiley and Sons, Inc.

Kahng, S.E. (2013). Growth rate for a zooxanthellate coral (*Leptoseris hawaiiensis*) at 90m. *Galaxea, Journal of Coral Reef Studies* 15(2): 39 – 40.

Kahng, S. E., Garcia-Sais, J. R., Spalding, H. L., Brokovich, E., Wagner, D., Weil, E., Hinderstein, L., and Toonen, R. J. (2010). Community ecology of mesophotic coral reef ecosystems. *Coral Reefs* 29(2): 255 – 275.

Kahng, S. E., and Kelley, C. D. (2007). Vertical zonation of megabenthic taxa on a deep photosynthetic reef (50 – 140 m) in the 'Au'au Channel, Hawaii. *Coral Reefs* 26(3): 679 – 687.

Kahng, S. E., and Maragos, J. E. (2006). The deepest zooxanthellate, scleractinian corals in the world? *Coral Reefs* 25: 25.

Kane, C., Kosaki, R., and Wagner, D. (2014). High levels of mesophotic reef fish endemism in the Northwestern Hawaiian Islands. *Bull. Mar. Sci*. 90: 693 – 703.

Kelman, D., Posner, E. K., McDermid, K. J., Tabandera, N. K., Wright, P. R., and Wright, A. D. (2012). Antioxidant activity of Hawaiian marine algae. *Marine Drugs* 10(2): 403 – 416.

Kennedy, T. A., Naeem, S., Howe, K. M., Knops, J. M., Tilman, D., and Reich, P. (2002). Biodiversity as a barrier to ecological invasion. *Nature* 417(6889): 636 – 638.

Legendre, P. (1993). Spatial autocorrelation: trouble or new paradigm? *Ecology* 74(6): 1659 – 1673.

Leichter, J. J., Stewart, H. L., and Miller, S. L. (2003). Episodic nutrient transport to Florida coral reefs. *Limnology and Oceanography* 48(4): 1394 – 1407.

Lesser, M. P., Slattery, M., and Leichter, J. J. (2009). Ecology of mesophotic coral reefs. *Journal of Experimental Marine Biology and Ecology* 375(1): 1 – 8.

Leverette, T. L., and Metaxas, A. (2005). Predicting habitat for two species of deep-water coral on the Canadian Atlantic continental shelf and slope. In Freiwald, A., and Roberts, J.M. (eds) *Cold-Water Corals and Ecosystems* 467 – 479. Berlin: Springer-Verlag.

Liu, C., Berry, P., Dawson, T., and Pearson, R. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28: 385 – 393.

Locker, S. D., Armstrong, R. A., Battista, T. A., Rooney, J. J., Sherman, C., and Zawada, D. G. (2010). Geomorphology of mesophotic coral ecosystems: current perspectives on morphology, distribution, and mapping strategies. *Coral Reefs* 29(2): 329 – 345.

McCulloch, C. E., and Neuhaus, J. M. (2001). Generalized linear mixed models. New York: John Wiley and Sons, Ltd.

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika* 37: 17 – 23.

Mortensen, P. B., Hovland, T., Fosså, J. H., and Furevik, D. M. (2001). Distribution, abundance and size of *Lophelia pertusa* coral reefs in mid-Norway in relation to seabed characteristics. *Journal of the Marine Biological Association of the UK* 81(04): 581 – 597.

Norris, J. N., Abbott, I. A., and Agegian, C. R. (1995). *Callidictyon abyssorum*, gen. et sp. nov. (Rhodophyta), a new deep-water net-forming alga from Hawai‘i. *Pacific Science* 49: 192 – 201.

Ohlmacher, G. C., and Davis, J. C. (2003). Using multiple logistic regression and GIS technology to predict landslide hazard in northeast Kansas, USA. *Engineering Geology* 69(3): 331 – 343.

Patzert, W. C., Wyrtki, K., and Santamore, H. J. (1970). Current measurements in the central North Pacific Ocean. Hawaii Institute of Geophysics HIG-70-31:1 - 26.

Peduzzi, P., Concato, J., Kemper, E., Holford, T.R., and Feinstein, A.R., 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology* 49(12): 1373 – 1379.

Pérez-Mayorga, D. M., Ladah, L. B., Zertuche-González, J. A., Leichter, J. J., Filonov, A. E., and Lavín, M. F. (2011). Nitrogen uptake and growth by the opportunistic macroalga *Ulva lactuca* (Linnaeus) during the internal tide. *Journal of Experimental Marine Biology and Ecology* 406(1): 108 – 115.

Pyle, R. L., Boland, R., Bolick, H., Bowen, B. W., Bradley, C. J., Kane, C., Kosaki, R. K., Langston, R., Longenecker, K., Montgomery, A., Parrish, F. A., Popp, B. N., Rooney, J., Smith, C. M., Wagner, D., and Spalding, H. L. (2016). A comprehensive investigation of mesophotic coral ecosystems in the Hawaiian Archipelago. *PeerJ* 4:e2475.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ralston, S. T., Gooding, R. M., and Ludwig, G. M. (1986). An ecological survey and comparison of bottom fish resource assessments (submersible versus handline fishing) at Johnston Atoll. *Fishery Bulletin* 84(1): 141 – 156.

Sabine, C. L., and Mackenzie, F. T. (1995). Bank-derived carbonate sediment transport and dissolution in the Hawaiian Archipelago. *Aquatic Geochemistry* 1(2): 189 – 230.

Sackett, D. K., Drazen, J. C., Moriwake, V. N., Kelley, C. D., Schumacher, B. D., and Misa, W. F. (2014). Marine protected areas for deepwater fish populations: an evaluation of their effects in Hawai'i. *Marine Biology* 161(2): 411 – 425.

Schlichter, D., and Fricke, H. W. (1991). Mechanisms of amplification of photosynthetically active radiation in the symbiotic deep-water coral *Leptoseris fragilis*. *Coelenterate Biology: Recent Research on Cnidaria and Ctenophora* 389 – 394. Springer Netherlands.

Schlichter, D. (1992). A perforated gastrovascular cavity in the symbiotic deep-water coral *Leptoseris fragilis*: A new strategy to optimize heterotrophic nutrition. Helgolander Meeresuntersuchungen. Hamburg, 45(4): 423 – 443.

Smith, J. E., Smith, C. M., Vroom, P. S., Beach, K. L., and Miller, S. (2004). Nutrient and growth dynamics of *Halimeda tuna* on Conch Reef, Florida Keys: Possible influence of internal tides on nutrient status and physiology. *Limnology and Oceanography* 49(6): 1923 – 1936.

Spalding, H. L. (2012). *Ecology of mesophotic macroalgae and* Halimeda kanaloana *meadows in the main Hawaiian Islands*. Doctoral dissertation. University of Hawai‘i at Mānoa.

Taylor, P. D., and Wilson, M. A. (2003). Paleoecology and evolution of marine hard substrate communities. *Earth-Science Reviews* 62(1): 1 – 103.

Veazey, L. M., Franklin, E. C., Kelley, C., Rooney, J., Frazer, L. N., and Toonen, R. J. (2016). The implementation of rare events logistic regression to predict the distribution of mesophotic hard corals across the main Hawaiian Islands. *PeerJ* 4:e2189.

Wagner, D., Kosaki, R. K., Spalding, H. L., Whitton, R. K., Pyle, R. L., Sherwood, A. L., Tsuda, R. T., and Calcinai, B. (2014). Mesophotic surveys of the flora and fauna at Johnston Atoll, Central Pacific Ocean. *Marine Biodiversity Records* 7:1 – 10.

Wolanski, E., Drew, E., Abel, K., and O’Brien, J. (1988). Tidal jets, nutrient upwelling and their influence on the productivity of the alga *Halimeda* in the Ribbon Reefs, Great Barrier Reef. *Estuarine, Coastal and Shelf Science* 25: 169 – 201.

Zuur, A. F., Ieno, E. N., and Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution* 1(1): 3 – 14.

Chapter 4: Present-day distribution and potential spread of the invasive green alga *Avrainvillea amadelpha* around the main Hawaiian Islands

4.1 Abstract

Algal assemblages are critical components of marine ecosystems from the intertidal to mesophotic depths; these plants are primary producers, nutrient cyclers, and substrate providers. Coastal reef ecosystems can be disrupted by stressors such as storm events, effluent inundation, sudden temperature shifts, and non-native invaders. *Avrainvillea amadelpha*, also known as leather mudweed, is an invasive green alga first recorded in the main Hawaiian Islands on the island of Kauaʻi, prompting concern due to extreme competitiveness with native algae and seagrasses. This alga has spread rapidly around parts of Oʻahu, decreasing the biodiversity of the benthos from shorelines to ~90 m depth. Here I employ a boosted regression tree (BRT) modeling framework using past and current distribution data to create maps identifying highly vulnerable regions prone to *Avrainvillea* invasion. My models indicate that regions exposed to minimal bottom currents and at least an annual maximum of eight degree heating weeks (DHW) are particularly susceptible to *Avrainvillea* colonization, especially in areas adjacent to some coastal development, fishing pressure, and shipping traffic. I identified Mokuleʻia, eastern Oʻahu (southward from Kahuku through Waimānalo), and much of the Honolulu metro area (eastward from ʻEwa through Maunalua Bay) as leather mudweed "hotspots". Additionally, I extrapolated the model to the main Hawaiian Islands and forecasted how a 25% increase in statewide annual maximum DHW may change habitat suitability for *Avrainvillea.* I identified hotspots along eastern and western Hawaiʻi, all coasts of Maui, south Molokaʻi and Lānaʻi, northern and eastern Kauaʻi, and south Niʻihau as susceptible regions of particular note for resource managers and conservationists.

4.2 Introduction

Disturbances can dramatically alter the community composition of coral reef ecosystems. The replacement of coral cover by macroalgae is known as a "phase shift" (Littler & Littler 1984; MacManus & Polsenberg 2004), which can occur in response to pressures such as storm events (Rogers & Miller 2006), coral disease outbreaks (Aronson & Precht 2001; Porter et al. 2001), nutrient input (McCook 1999), or removal of herbivores (Hughes et al. 2007). Often, these stressors occur simultaneously or in sequence as one stress event lowers the threshold of

coral resilience, thereby magnifying total coral mortality (Brandt et al. 2013; Redding et al. 2013; Vega-Thurber et al. 2014; Casey et al. 2014).

In Hawai‘i, shallow coral reef ecosystems endure common reef stresses, including overuse and physical degradation (i.e., trampling), overextraction (i.e., overfishing of important food fish, such as parrotfish (*uhu*) and jacks (*ʻōmilu*), chemical degradation from freshwater input, effluent, or sunscreen, and invasion by harmful bacteria, viruses, or competitors (Rodgers & Cox 2003; Stamoulis et al. 2017; Takesue & Storlazzi 2017; Friedlander et al. 2018;). Invasive algae can become opportunistic when some environmental change facilitates their spread or allows them to outcompete native species. They are a strong ecological and economic concern in Hawai‘i (Schaffelke et al. 2006), where eutrophication due to runoff likely influenced the spread of invasive *Hypnea musciformis* on Maui (Smith et al. 2002) and *Dictyosphaeria cavernosa* in O‘ahu (Stimson et al. 1996).

One particularly troublesome competitor that has garnered statewide attention in the past several decades is the invasive green macroalga *Avrainvillea amadelpha* sensu Brostoff (*Avrainvillea* herein). This macroalga, also known as “leather mudweed” due to its thick, paddle-like blades (Fig. 1), was first discovered in Hawai‘i in 1981 (Brostoff 1989) and has been observed from the intertidal to the mesophotic zone (30 – 90 m) (Spalding 2012; Pyle et al. 2016; Cox et al. 2017). Its origin and phylogeny remain uncertain, but it is not present in historical records of shallow marine habitats around the main Hawaiian Islands (MHI) (Brostoff 1989). To date, this highly successful macroalga has spread across O‘ahu (Fig. 2) and has been detected in several locations around Kaua‘i (Smith et al. 2002; Wade & Sherwood, submitted). In the mesophotic zone, video records show *Avrainvillea* beds competing with meadows of a native *Udotea* species. (Spalding 2012) (Fig. 3).

Avrainvillea is of particular concern in Hawai‘i because of the weedy characteristics it exhibits, such as benthos alteration (Martinez et al. 2009; Sansone et al. 2017), tolerance to environmental extremes (e.g., light availability, exposure, etc.), and unpalatability to native herbivores (Van Heukelem 2016; S. Chulakote, pers. comm.). The alga is also presumed to have successful vegetative propagation via fragmentation, and holdfast siphon viability like many other siphonous green algae (Walters & Smith 1994). This alga may also change the community diversity and structure across the reefs it invades. Langston & Spalding (2017) recorded a higher abundance of fishes above open sand versus within *Avrainvillea* canopy; Longenecker et al.

(2011) determined that removal of *Avrainvillea* returned invertebrate communities to compositions observed in unaffected regions. All of these traits combine to make *Avrainvillea* a formidable competitive presence across native coral reefs.

The state has spent over $2 million to remove 1.32 million kilograms of *Avrainvillea* across a 90,000 m^2 swath of Maunalua Bay, Oʻahu (Kittinger et al. 2013). Much effort has been invested in observing and recording *Avrainvillea* invasions, as well as stymying the progress of ongoing invasions (Hawaiʻi Department of Aquatic Resources, unpublished data). However, to date, no projections exist regarding the potential spread of this invasive macroalga across Hawaiʻi's reef systems. When used as a management tool, habitat suitability models may provide critical information about the current or probable future distribution of opportunistic competitors. Habitat suitability models can also improve understanding of dispersal methods and provide early warning about regions that are particularly vulnerable, but not yet invaded. This work will shed light on areas of concern around Oʻahu and across the MHI, and provide guidelines for resource managers monitoring the spread of this botanical invader.

4.3 Methods

4.3.1 Observational data

I sourced my observational data from video obtained during submersible and ROV dives and shore-based or snorkeling surveys (Fig. 4; Table 1). The Hawaiʻi Undersea Research Laboratory (HURL) provided records of Pisces submersible and RCV-150 ROV dives conducted during November 2006. To obtain mesophotic records of *Avrainvillea* occurrence, I processed 34 hours of footage from 7 dives across all coastlines around Oʻahu (Table 2). I collected 3,796 new snapshots of the benthos by pausing each video track in 30 second intervals, but excluded 2,441 photo stills due to blurriness, stationarity of the vessel, or positioning outside the depth maximum (> 90 m).

Shallow (~5 – 10 m) subtidal observations of *Avrainvillea* were collected during multiple snorkeling and free-diving trips conducted from July 2015 to November 2017. I combined these data with two existing subtidal survey datasets, which included data from 2015 – 2017 provided by the Hawaiʻi Department of Aquatic Resources. The Our Project in Hawaiʻi's Intertidal (OPIHI) school-based monitoring program acquired intertidal shore-based survey data through transect sampling at low tides (0 – -0.15 m) from March – June 2017.

I removed 8,040 observations from the raw dataset of 33,286 due to overlap within pixels or failure to fall within my depth range. All shoreline survey point data were inspected for geographical precision and manually moved to the closest classified marine pixel if needed. I combined all observational data into one dataset spanning depths from 0 – 90 m and covering all coastlines across O‘ahu. I resampled all data to a grid size of 250 x 250 m, resulting in a final dataset of 276 observations of *Avrainvillea* presence (207) or absence (69).

4.3.2 Environmental data

Fifteen environmental predictor variables were included in my model (Table 3). I represented all covariates as 250 m raster grids across the study domain. These covariates may have a direct influence on the metabolic constraints and, therefore, the distribution of *Avrainvillea.* I broadly categorized predictors into three classifications: 1) topographic, which included seafloor slope, aspect (i.e., compass direction of seafloor), rugosity (i.e., roughness), and bathymetric position index; 2) oceanographic, which included values for seafloor current velocity (annual mean and standard deviation), surface chlorophyll a concentration (annual mean), and surface turbidity (annual mean); and 3) anthropogenic, which included indices describing fishing pressure, nearshore coastal development, anomalous sea surface warming, shipping traffic, nutrient and sediment inputs, and proximity to human population. Table 3 includes further details about considered environmental covariates.

I first examined the spatial coverage of all covariate layers. Due to insufficient coverage, I excluded the layers displaying proximity to human population, annual mean turbidity, and sediment influx. To account for similar lack of spatial coverage of my nutrient flux and coastal development layers, I extrapolated the closest offshore values and applied these values to offshore observational data points outside the layer coverage.

Collinearity of predictor variables may preclude suitable model fitting (i.e., preventing the model from converging or contributing to high standard errors in coefficient estimates). I created a correlation scatterplot to examine the relationships between all covariates and the response variable (i.e., *Avrainvillea* occurrence) (Fig. 5). Following Dancey and Reidy (2004) and Tabachnick and Fidell (1996), I removed any covariates that exceeded a correlation cutoff threshold of 0.7 or did not show a correlation with the response variable. We included seven covariates in my initial model (seafloor slope, annual mean seafloor current velocity, maximum

degree heating weeks (DHW), nutrient flux, shipping traffic, shoreline development percentage, and scaled annual fishing catch).

4.3.3 Model development

Boosted regression tree (BRT) models fit an ensemble of statistical models by combining the power of boosting (a technique that blends multiple models to improve predictions) and regression trees (models that identify links between response variables and covariates via recursive binary splits; Elith et al. 2008). Per Elith et al. (2008), I built one BRT model and applied it to my entire dataset. To start, I retained stochasticity in my model (bag fraction = 0.5) and specified a slow learning rate (lr = 0.0025), which 1) minimized the contributions of each tree to the final model, 2) minimized the number of very small, random subsets of data selected for fitting each tree, and 3) decreased the chance of fitting abnormal trees.

Binomial deviance was used for my loss function to enhance the robustness of the predictions (Hastie et al. 2001; Elith et al. 2008). I selected my final parameters by systematically varying each parameter and examining the change in residual deviance, or goodness of fit, of the model (Elith & Leathwick 2017). The regression trees partition the study space into sets of rectangles in order to detect important effects, thereby incorporating geographical effects within the model.

The relative importance of each predictor variable was established using formulae within the gbm library (Friedman 2001; Friedman & Meulman 2003). These values are determined by the number of times each variable is selected in a split, and weighted by the subsequent model improvement based on that split (weights are squared to give higher value to predictors that improve the model more considerably). I scaled predictor contributions such that the sum equated to 100, with higher values indicating greater influence on the response variable, i.e., occurrence of *Avrainvillea*.

4.3.4 Evaluation of model performance

I randomly divided my dataset 70/30 to develop my model (i.e., my "training" and "testing" datasets). Per Elith et al. (2008), I initially set nt_0 = 50 and ran 10 unique BRTs. I subsequently increased the value of nt by 10 (i.e., nt_1 = 60, nt_2 = 70, and so on) and ran new models until I found a combination that produced higher average mean predictive performance and lower average standard errors than the previous set of models. I kept my lr very low (lr = 0.0025) to ensure precise estimation of the optimal nt = 80 (cross-validated optimal nt = 80).

I used R statistical software (R Core Team 2017) and ESRI ArcGIS v.10.4 (ESRI 2017) to perform all data analyses. I specified my BRTs using the dismo package v. 1.1 – 4 (Hijmans et al. 2017).

4.4 Results

4.4.1 Performance of models

My initial model (193 observations, 7 predictors) fit 1620 trees. I then reran my initial model using *k*-fold cross-validation ($k = 10$) to simulate a fixed number of five variable drops. The ideal number of two variable drops resulted in the removal of mean nutrient flux and seafloor slope from my simplified model (nt = 1760; 193 observations, 5 predictors) (Fig. 6). Though model simplification resulted in no significant change in total deviance or total residual deviance, and cross-validated calibration predictive accuracy increased slightly (+0.03 change to 0.955) (Table 4).

I cross-validated the model using withheld validation data (83 observations). The cross-validated AUC (0.961) indicated high predictive accuracy and model trustworthiness, with an average of 89% of observations identified correctly (Hosmer & Lemeshow 2004).

4.4.2 Influential environmental covariates

My model indicated that the five variables exhibiting the most influence on the probability Avrainvillea sp. occurrence across O‘ahu coastlines include percentage of shoreline development (32%), scaled maximum degree heating weeks (27%), scaled fishing catch (17%), mean current velocity at the seafloor (13%), and shipping traffic (11%) (Table 5). We produced partial dependence plots (Fig. 7) to visually interpret the relative importance of each predictor variable in determining the distribution of *Avrainvillea*.

4.4.3 Interactions between environmental variables

I identified interactions between 1) seafloor current and shipping traffic, and 2) seafloor current and DHW. These interactions indicated that relatively calm regions exposed to shipping traffic above the 50th percentile (~500 transits annually) and $\geq$ 8 DHW may be more susceptible to *Avrainvillea* invasion.

4.4.4 Areas highlighted as hotspots

Areas around O‘ahu identified by the model as most susceptible to *Avrainvillea* invasion include Mokule‘ia in north O‘ahu, eastern O‘ahu (southward from Kahuku through Waimānalo), and much of the Honolulu metro area (eastward from ‘Ewa through Maunalua Bay) (Fig. 8),

most of which is already inundated with Avrainvillea (Fig. 2). I extended the model to predict areas of high Avrainvillea suitability around the surrounding MHI. The results indicate that managers should concentrate surveys for Avrainvillea along the southern and western coasts of Maui, western and eastern Hawai‘i, south Moloka‘i and Lāna‘i, northern and eastern Kaua‘i, and south Ni‘ihau (Fig 9). Model predictions extend approximately 1.5 – 2 km offshore and encompass the shoreline to the mesophotic zone (to 90 m depth). The model indicated that approximately one quarter (423 km2) of the modeled area (1691 km2) is highly suitable Avrainvillea habitat ($\theta > 0.91$, or third quantile cutoff value).

4.5 Discussion

This is the first study to predictively model and map the present-day and potential distribution of invasive leather mudweed across O‘ahu and extrapolate results to surrounding islands in the event of potential future climate conditions. I highlight this approach as one possible management tool for prioritizing management and conservation efforts and in the face of changing anthropogenic and environmental drivers and future climate conditions.

4.5.1 Traits of imperiled reefs

Marine regions identified as most susceptible to *Avrainvillea* colonization experience minimal annual mean current flux (< 0.005 m/s). These predictions agree with the observational data used to train and test my final model. Littler et al. (2004) hypothesized that *Avrainvillea* mounds in Belize may strategically colonize protected, calm, and shallow embayments in order to maximize productivity. I caution that my current flux values were generally very low (range = 0 – 0.02 m/s) and the mean current flux value for pixels containing *Avrainvillea* did not differ significantly from the mean value for pixels without *Avrainvillea* (mean difference < 0.01 m/s). Additionally, little is known of the success of tidal current dispersing asexual fragments or any details for sexual reproduction among O‘ahu populations.

My model also indicated some influence of degree heating weeks (DHW) on the susceptibility of habitat to *Avrainvillea* colonization. DHW is a metric that couples the duration and intensity of anomalously warm regional sea surface temperatures, and Eakin et al. (2010) note that thresholds of 8 and 12 DHW indicate high and very high risks of bleaching-related coral mortality, respectively. My model indicated that regions experiencing a maximum of at least 8 DHW during 2013 – 2016 were more suitable for *Avrainvillea* invasion. Cox et al. (2017) speculate that recent warming events in 2015 caused a die-back in abundant native intertidal

macroalgae, facilitating the rapid spread of leather mudweed in this intertidal environment. They observed an opportunistic increase in abundance of *Avrainvillea* following the reduction in brown algae (*Padina sanctae-crucis* and *Dictyota* spp.), which dominate multiple intertidal sites around O‘ahu (Cox et al. 2017). This invasion notably followed a period of unseasonably warm, calm seas and a series of severe bleaching events from 2014 – 2017. The impact of warmer ocean temperatures and the resulting die-back of native macroalgae may also increase the success of more thermally resistant invasive species. I hypothesize that reefs destabilized from repeated heat stress are more susceptible to incursion of invasive species like *Avrainvillea*. Peyton (2009) noted that *Avrainvillea* in Maunalua Bay hosts epiphytes in its canopy that may act as “sunscreen” during periods of high light exposure, which suggests some level of shade tolerance in this alga.

Finally, my model suggested that anthropogenic stresses on land and in coastal regions may increase the possibility of *Avrainvillea* invasion. Regions adjacent to coastal development or higher-than-average shipping traffic were identified as more susceptible to colonization by the invasive alga. Additionally, the model suggests that regions exposed to moderate to heavy (> 50th quantile; scaled values, > 0.2) fishing pressure are more prone to *Avrainvillea* invasion. Prior survey data (K. Peyton, Hawai‘i Conservation Conference; unpublished data) shows some correlation with the removal of herbivorous fish and urchins and the subsequent appearance of *Avrainvillea*. Though phase shifts may occur across overfished reefs (e.g., Hughes 2007), researchers are uncertain about the role of grazing pressure in controlling the spread of an unpalatable invader like *Avrainvillea* (Van Heukelem 2016). However, scaled fishing catch was one of the six factors that best explains the presence of *Avrainvillea*, which argues that the role of reduced herbivory must also be considered. Although herbivorous coral reef fish populations around O‘ahu are generally low (Williams et al. 2016), and the influence of broadly reduced herbivory (as many species have been reduced or lost) on the increase of *Avrainvillea* is unknown, the correlation between fishing pressure and spread of this alien invasive merits further study.

4.5.2 Looking to the future

My model pinpointed much of O‘ahu’s east coast, the Honolulu metro area, and Mokule‘ia (on the north shore) as regions particularly susceptible to *Avrainvillea* invasion (Fig. 8). Binary representation of the absence (0) or presence (1) of coastal modification (i.e., change

of geomorphology due to human activity; Wedding et al. 2018) highlights the correlation of model-identified highly vulnerable regions and developed regions (Fig. 10). Given this visible association between vulnerability and coastal development, I urge managers to consider concentrating resources across reef systems adjacent to expanding urban centers.

In addition to using the model to highlight important stressors and the most vulnerable coastal regions island-wide based on present-day conditions, I used my framework to forecast how reef vulnerability might change following a 25% increase in statewide maximum DHW. With all other predictors held constant (i.e., held at present-day values), my model predicted changes in habitat suitability per 250 m pixel ranging from decreases of up to 31% to increases as high as 58%. The Waiʻanae (west) coast of Oʻahu and much of the southern coast experienced little to no increases, or even slight decreases, in overall vulnerability to *Avrainvillea* invasion. Portions of the eastern coast (from Lāʻie wrapping around south through Hanauma), patches of the Honolulu metro area, and Kaʻena Point through Mokuleʻia (north shore) displayed the most dramatic increases in vulnerability (Fig. 11, inset C). Across the other MHI, I observed increased vulnerability around the southern and western shores of Kauaʻi and Maui Nui, and around northwestern and eastern Hawaiʻi, including Hilo Bay (Fig. 11).

4.6 Conclusions

We constructed a series of boosted regression trees to predictively map marine habitat vulnerability to invasion and dominance by the alien invasive mudweed *Avrainvillea* as a method to highlight possible management actions in the face of a changing climate. Our findings indicate that conservationists and state agencies concerned about the spread of this alga may consider focusing monitoring or removal resources across shallow reef systems adjacent to coastal development activities and subject to high fishing pressure or shipping traffic, particularly where *Avrainvillea* has not yet been or has only sporadically been observed. Conversely, areas less susceptible to invasion by *Avrainvillea* may be considered as a higher priority for conservation measures. The long-term success of *Avrainvillea* removal efforts once this alga invades a system remains uncertain.

4.7 Acknowledgments

Funding for this research was provided by the University of Hawaiʻi Sea Grant College Program grant no. NA14OAR4170071. I thank M. Winston, B. Neilson, K. Tejchma, and C. Kelley for their assistance in acquiring critical observational data. Submersible surveys were funded in part

by grants from the NOAA Undersea Research Program's Hawai'i Undersea Research Laboratory, NOAA Coral Reef Conservation Program, and NOAA Office of Ocean Exploration program. OPIHI surveys were funded by Sea Grant.

4.8 Figures

Figure 1. Video still of dense *Avrainvillea amadelpha* meadow at 42 m depth off the coast of southwest O'ahu observed during submersible dive RCV-369 in November 2006.

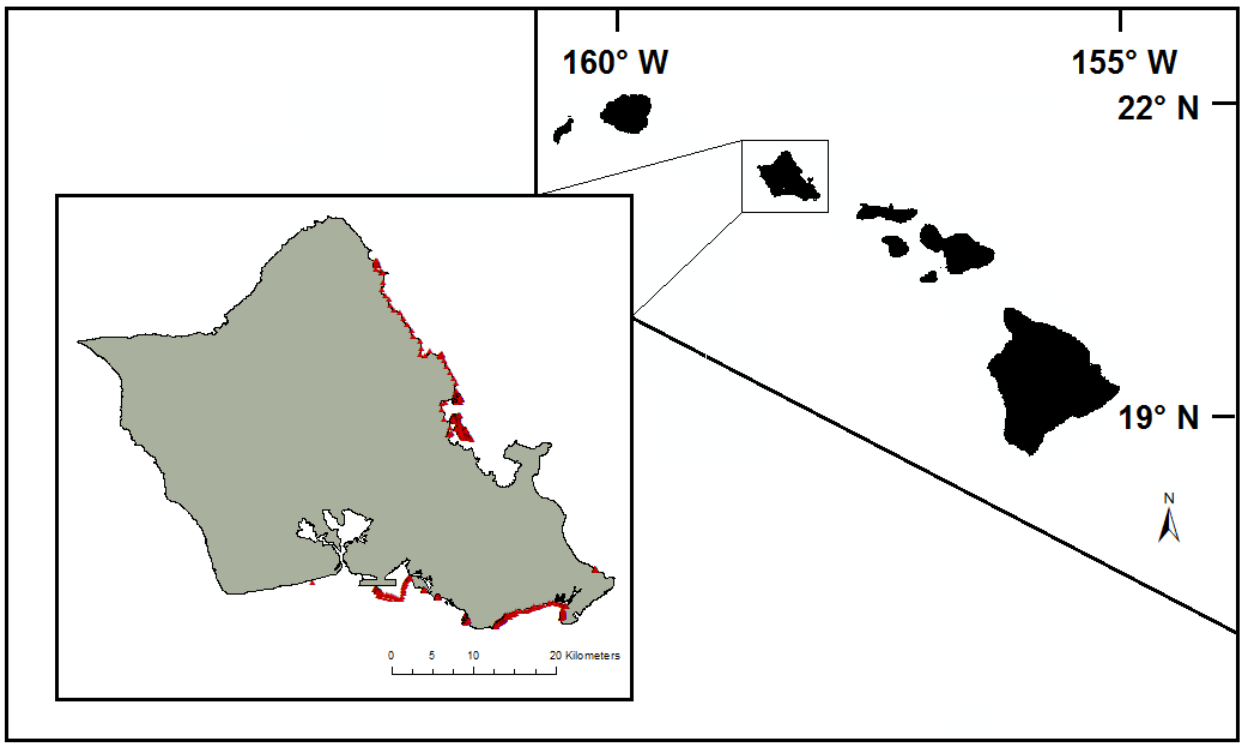

Figure 2. Geolocated survey data denoting 23,421 observations of *Avrainvillea amadelpha* occurrence around O‘ahu, Hawai‘i, from 2015 – 2017. Observations demarcated with red triangles.

Figure 3. Video stills of *Avrainvillea amadelpha* overtaking the lime green blades of native *Udotea* alga. Dives RCV-369 (top) and P4-188 (bottom) were conducted off the southwestern and southern coasts, respectively, of O‘ahu in November 2006. Stills were taken at depths of 40 – 50 m in the upper mesophotic zone.

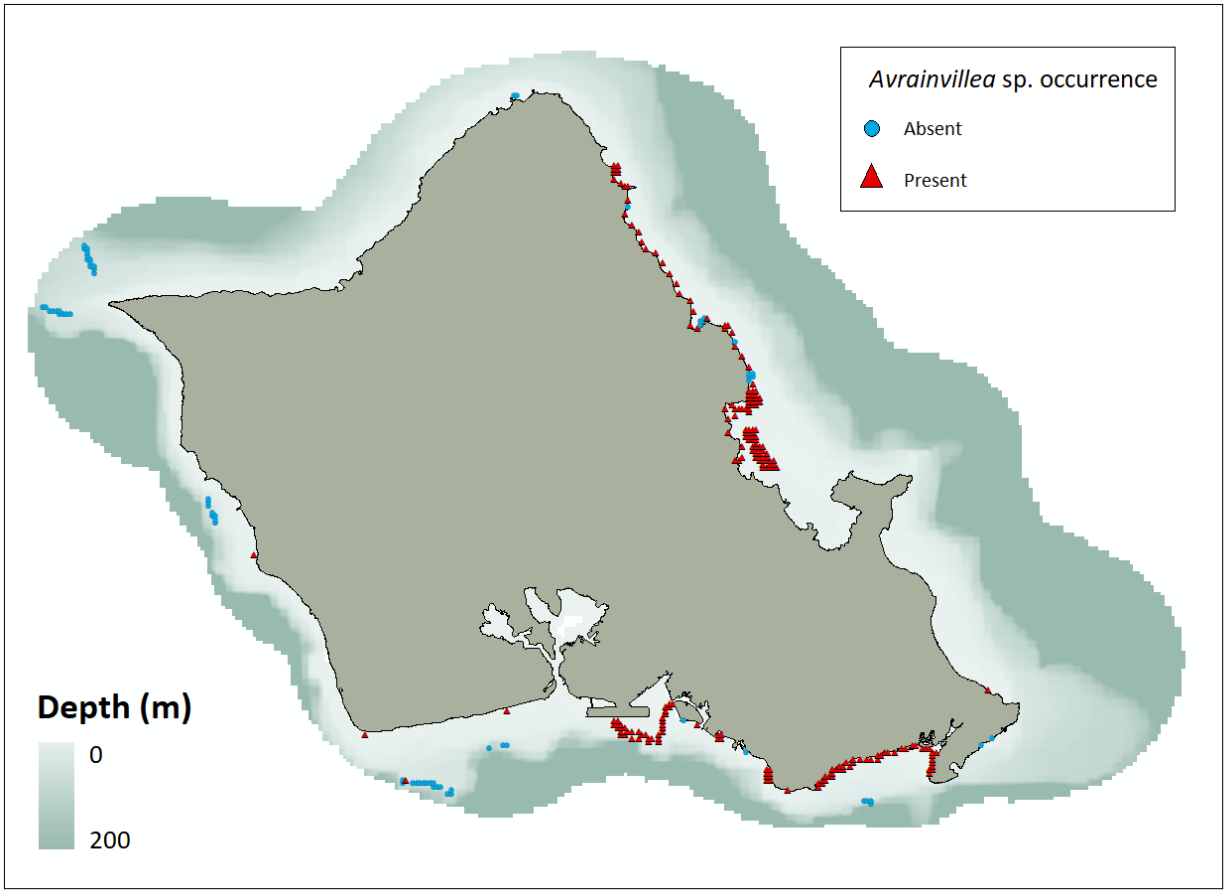

Figure 4. Map of *Avrainvillea amadelpha* presence (red triangles) or absence (blue circles) across O‘ahu from depths 0 – 200 m. Observations edited for overlap at 250 m raster resolution scale and constrained to depths ≤ 90 m.

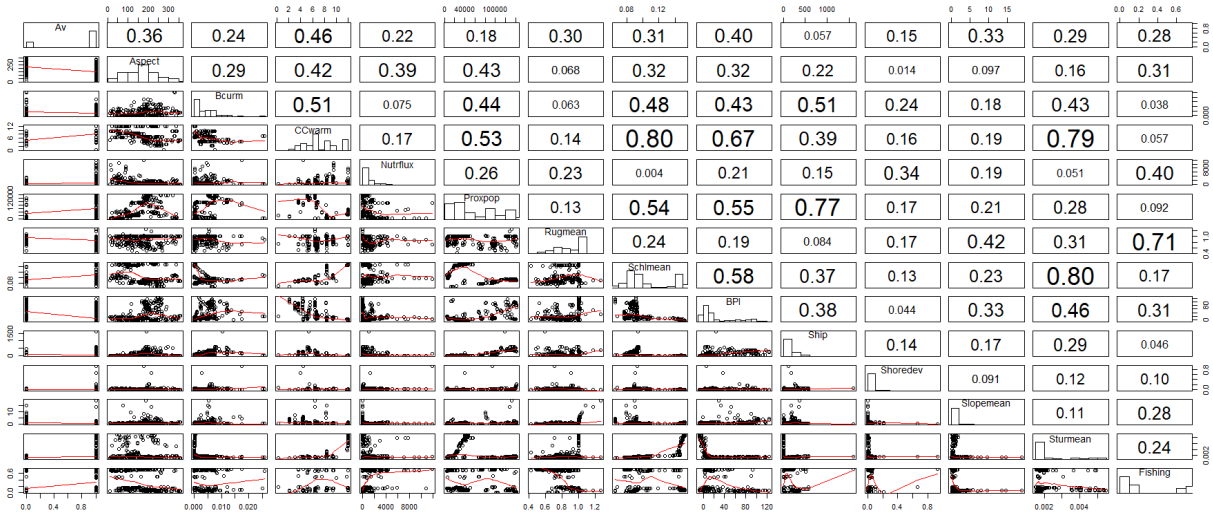

Figure 5. Correlation scatterplot displaying relationships between covariates and response variable.

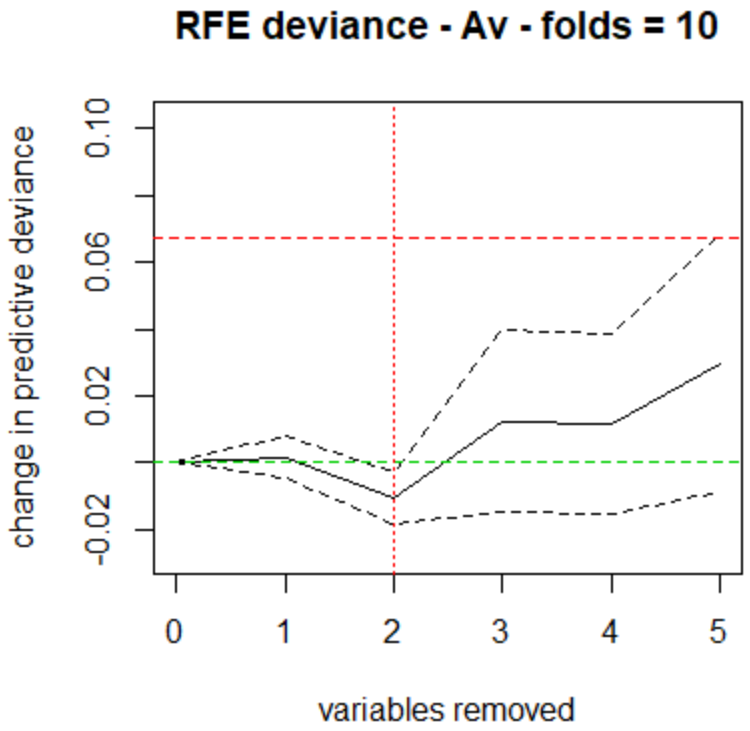

Figure 6. Plotted changes in predictive deviance of model following *k*-fold cross-validation and model simplification.

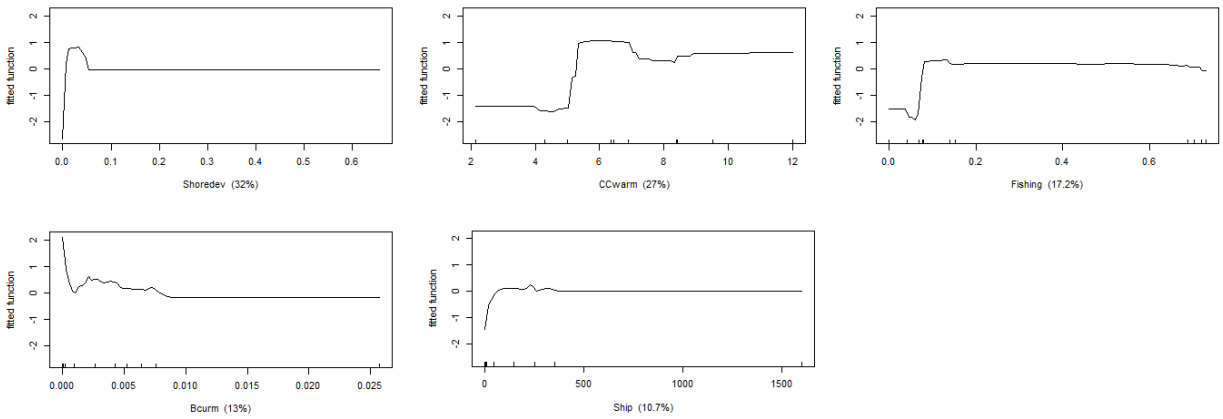

Figure 7. Partial dependence plots indicating the influence of each predictor in the final simplified model.

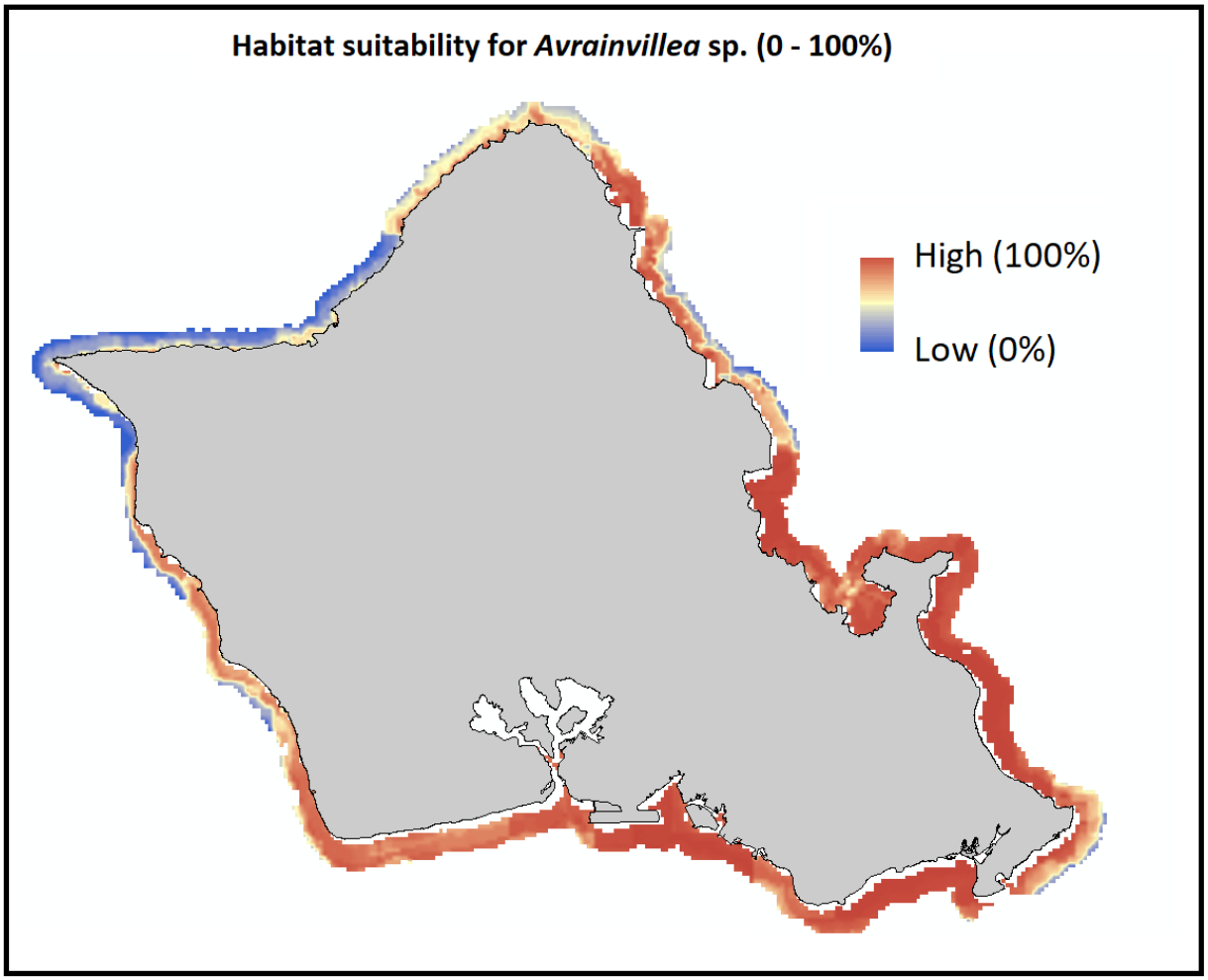

Figure 8. Mapped model predictions around nearshore O‘ahu represented at 250 m raster resolution. Blue pixels indicate habitat less suitable for *Avrainvillea* invasion; red pixels indicate regions identified as highly suitable for *Avrainvillea* invasion.

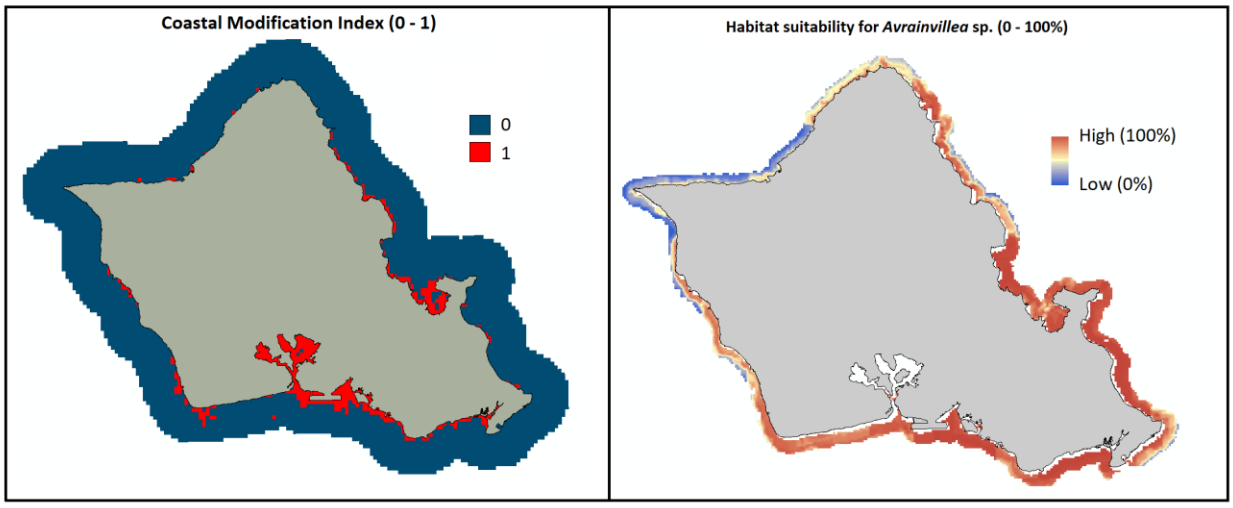

Figure 9. Side-by-side comparison of coastal modification index values across nearshore O‘ahu (left) and model predictions of *Avrainvillea* habitat suitability. Modification was classified on a binary scale, such that areas without modification were marked “0”. Modified areas (“1”) may contain structures, altered coastline, or dredged habitat. See Lecky (2016) for further information.

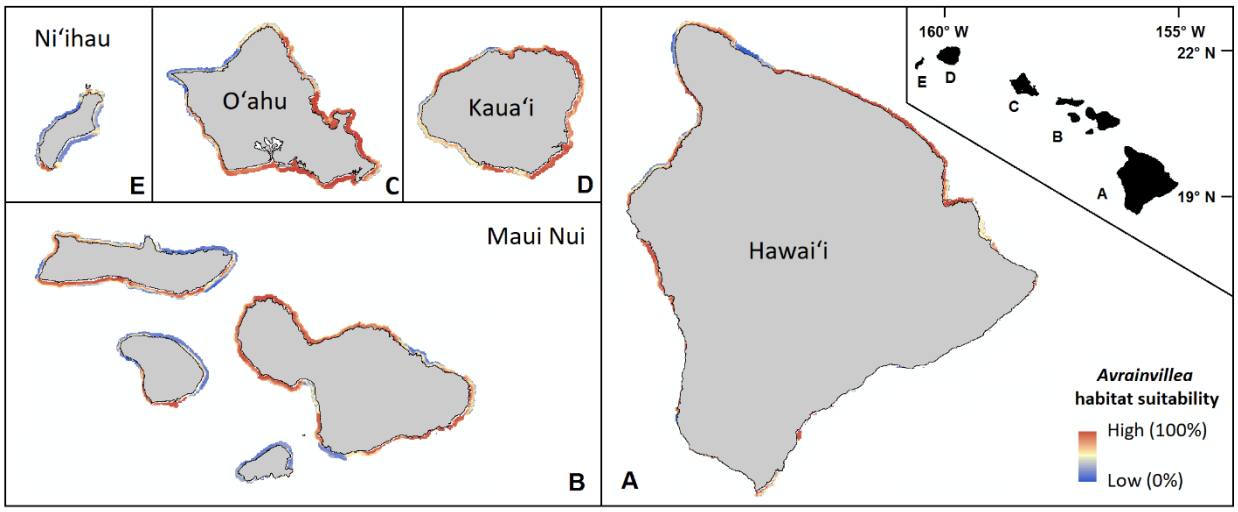

Figure 10. Mapped model predictions around the main Hawaiian Islands represented at 250 m raster resolution. Cool colors indicate low habitat suitability for *Avrainvillea*; warm colors indicate high suitability.

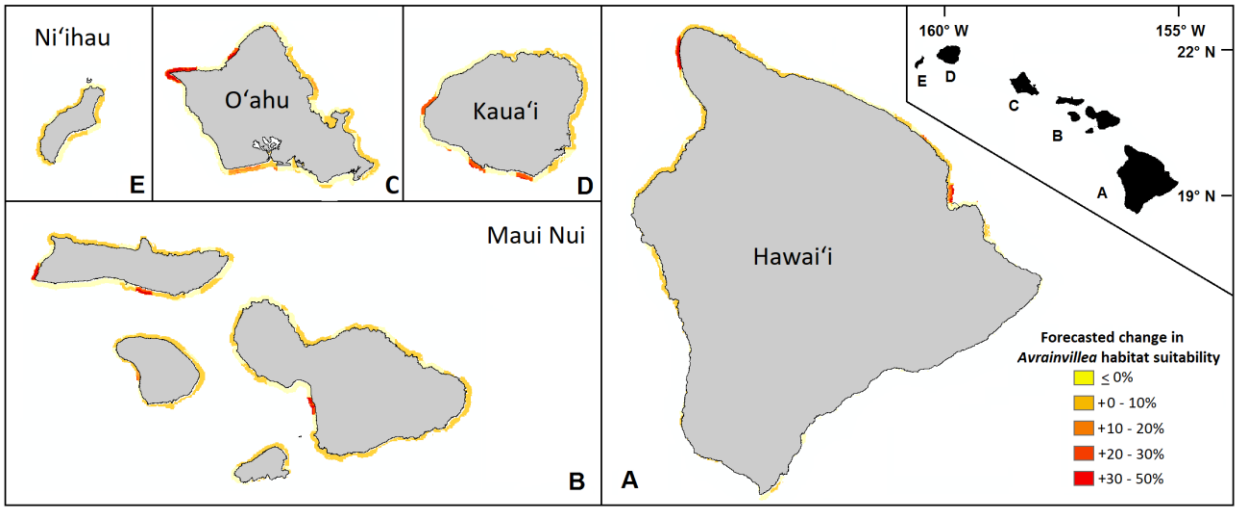

Figure 11. Forecasted changes in habitat vulnerability subsequent to a theoretical 25% increase in maximum DHW across all coasts of the main Hawaiian Islands. Yellow tones indicate slight decreases or no changes in susceptibility to *Avrainvillea* invasion, while orange and red colors indicate increases in vulnerability compared to present day (Fig. 10) predictions.

4.9 Tables

Table 1. Raw observational data survey information.

Survey type	Survey dates	Number of presence points	Number of absence points
Mesophotic	11/2006	124	1874
Snorkel	6/2015 - 11/2017	31,223	56
Shore	3/2017 - 6/2017	4	5

Table 2. Detailed mesophotic survey information.

Latitude (start)	Longitude (start)	Vessel	Dive	Date	Max Depth (m)	Location
21.5739	-158.3373	RCV	371	6-Nov-06	67	Oahu N
21.3208	-158.1458	RCV	369	15-Nov-06	66	Oahu SW
21.4333	-158.2067	RCV	370	15-Nov-06	120	Oahu W
21.6214	-158.2998	P4	183	16-Nov-06	179	Oahu N
21.2662	-158.1003	P4	188	27-Nov-06	117	Oahu S
21.2833	-158.0167	RCV	376	27-Nov-06	82	Oahu W
21.2483	-157.7588	RCV	377	27-Nov-06	112	Oahu S

Table 3. List of variables considered for our analyses.

Variable type	Category	Variable description	Source	Resolution	Variable
Biological (response)	Macroalgae	Presence/ absence between 0 – 90 m depth	Hawai‘i Undersea Research Lab optical validation data; snorkel and Our Project in Hawai‘i’s Intertidal (OPIHI) shore-based surveys	NA	*“Avrainvillea amadelpha”*
Environmental (predictor)	Topographic	Seafloor complexity	NOAA National Centers for Coastal Ocean Science Marine Biogeographic Assessment of the Main Hawaiian Islands (2016)	360 m	Rugosity (unitless)
		Mean rate of change			Slope (degrees)
		Compass direction of maximum slope			Aspect (degrees)
		Focal mean analysis of bathymetry and slope; 5 km neighborhood			Bathymetric position index (m)
	Oceanographic	Seafloor current velocity (annual mean and standard deviation)			Bottom current velocity (annual mean, std. dev.) (m/s)

		derived from Hybrid Coordinate Ocean Model (HYCOM) output			
		Annual mean surface chlorophyll *a* concentration as proxy for phytoplankton biomass; data from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite			Sea surface concentration of chl. *a* (annual mean) (mg/m^3)
		Annual mean turbidity derived from quantification the light reflected by organic and inorganic matter in the 547 nm wavelength			Sea surface turbidity (annual mean) (sr^{-1})
	Anthropogenic	Scaled (0 – 1) average annual fisheries haul from non-commercial shore-based fishing, and non-commercial boat-based fishing using line, net, and spear fishing gear	Wedding et al. (2018); Lecky (2016)	100 m	Scaled (0 – 1) average annual reef fish catch of fin fish spp. (excluding *akule* and '*opelu*)
		Development of land to hard, man-made surface from 2005 – 2010; decay function implemented to represent sediment dispersal from coastal construction; values scaled from 0 – 1 to represent relative magnitude of development			Nearshore new development impact, 2005 – 2011
		Maximum observed degree heating weeks from 2013 – 2016; data from NOAA Coral Reef Watch 5 km products; values scaled 0 – 12			Maximum degree heating weeks (unitless)
		Addition of sediment due to land use changes (i.e., coastal construction, agricultural activities); calculated by Falinski (2016)			Sediment input (t/yr/ha)
		Addition of nitrogen, phosphorus, and other nutrients from sewage released from submarine disposal systems (i.e., cesspools and septic tanks)			On site waste disposal system nutrient input ($g/day/km^2$)
		Ship location data from August 2011 – August 2012 was sourced from the Pacific Islands Ocean Observing System. The ArcGIS line density tool was used to calculate ship track density within pixels.			Ship-based shipping (estimated annual number of ship crossings)
		Sum of human population in 15 km neighborhood			Human population in 15 km radius (number of people)

Table 4. Model performance details.

Model	Number of predictors	Total Deviance (mean)	Residual deviance (mean)	AUC (training)	AUC (testing)	Standard error (testing AUC)
Initial model (av.mod)	7	1.122	0.21	0.998	0.925	0.022
Simplified model (av.simp)	5	1.122	0.205	0.997	0.955	0.014

Table 5. Ranked influence of predictors included in simplified model.

Variable	Scaled importance (0-100)
Shoreline development index (0-100%)	32.03097
Degree Heating Weeks (DHW)	27.01103
Scaled annual mean fishing catch	17.20106
Mean seafloor current (m/s)	13.03637
Shipping traffic	10.72057

4.10 References

Aronson, R. B., and Precht, W. F. (2001). White-band disease and the changing face of Caribbean coral reefs. In Porter, J.W. (ed.) *The Ecology and Etiology of Newly Emerging Marine Diseases*. *Hydrobiologia* 460: 25 – 38.

Brandt, M. E., Smith, T. B., Correa, A. M., and Vega-Thurber, R. (2013). Disturbance driven colony fragmentation as a driver of a coral disease outbreak. *PLoS One* 8(2): e57164.

Brostoff, W. N. (1989). *Avrainvillea amadelpha* (Codiales, Chlorophyta) from Oahu, Hawaii. *Pacific Science* 43: 166 – 169.

Bruno, J. F., Slatman, H., Precht, W. F., Selig, E. R., and Schutte, V. G. (2009). Assessing evidence of phase shifts from coral to macroalgal dominance on coral reefs. *Ecology* 90(6): 1478 – 1484.

Casey, J. M., Ainsworth, T. D., Choat, J. H., and Connolly, S. R. (2014). Farming behavior of reef fishes increases the prevalence of coral disease associated microbes and black band disease. In *Proceedings of the Royal Society* 281(1788): 20141032. The Royal Society.

Costa, B., and Kendall, M. (2016). Marine Biogeographic Assessment of the Main Hawaiian Islands: Synthesized physical and biological data offshore of the Main Hawaiian Islands from 1891-01-01 to 2015-03-01 (NCEI Accession 0155189). Version 1.1. NOAA National Centers for Environmental Information. Dataset. DOI:10.7289/V56H4FG9.

Cox, T. E., Spalding, H. L., and Foster, M. S. (2017). Spatial and temporal variation of diverse inter-tidal algal assemblages in southwest O‘ahu. *Marine Ecology* 38(3).

Dancey, C. and Reidy, J. (2004). Statistics without maths for psychology. England: Pearson Education Limited.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics* 7:1 – 26.

Elith, J., Leathwick, J. R., and Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology* 77(4): 802 – 813.

Elith, J., and Leathwick, J. (2017). Boosted regression trees for Ecological Modeling. http://www.lcis.com.tw/paper_store/paper_store/brt(5)-2015115131033846.pdf

ESRI (2017). ArcGIS Desktop: Release 10.4. Redlands, CA: Environmental Systems Research Institute.

Falinski, K. (2016). *Predicting sediment export into tropical coastal ecosystems to support ridge to reef management*. Doctoral dissertation. University of Hawai‘i at Mānoa.

Friedlander, A. M., Donovan, M. K., Stamoulis, K. A., Williams, I. D., Brown, E. K., Conklin, E. J., DeMartini, E. E., Rodgers, K. S., Sparks, R. T. and Walsh, W. J. (2018). Human-induced gradients of reef fish declines in the Hawaiian Archipelago viewed through the lens of traditional management boundaries. *Aquatic Conservation: Marine and Freshwater Ecosystems* 28:146 - 157.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics* 1189 – 1232.

Friedman, J. H., and Meulman, J. J. (2003). Multiple additive regression trees with application in epidemiology. *Statistics in Medicine* 22(9): 1365 – 1381.

Guisan, A., Thuiller, W., and Zimmermann, N. E. (2017). *Habitat suitability and distribution models: with applications in R*. Cambridge: Cambridge University Press.

Hastie, T., Tibshirani, R., and Friedman, J.H. (2001). *The elements of statistical learning: data mining, inference, and prediction.* New York: Springer-Verlag.

Hillis, D. M., and Bull, J. J. (1993). An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* 42(2): 182 – 192.

Hosmer, D. and Lemeshow, S. 2004. Applied Logistic Regression. Hoboken, New Jersey: John Wiley and Sons, Inc.

Hughes, T. P., Rodrigues, M. J., Bellwood, D. R., Ceccarelli, D., Hoegh-Guldberg, O., McCook, L., Moltschaniwskyj, N., Pratchett, M. S., Steneck, R. S., and Willis, B. (2007). Phase shifts, herbivory, and the resilience of coral reefs to climate change. *Current Biology* 17(4): 360 – 365.

Kittinger, J.N., Bambico, T.M., Minton, D., Miller, A., Mejia, M., Kalei, N., Wong, B., and Glazier, E.W. (2016). Restoring ecosystems, restoring community: socioeconomic and cultural dimensions of a community-based coral reef restoration project. *Regional Environmental Change* 16(2): 301 – 313.

Langston, R. C., and Spalding, H. L. (2017). A survey of fishes associated with Hawaiian deep-water *Halimeda kanaloana* (Bryopsidales: Halimedaceae) and *Avrainvillea* sp. (Bryopsidales: Udoteaceae) meadows. *PeerJ* 5:e3307.

Lecky, J. H. (2016). *Ecosystem vulnerability and mapping cumulative impacts on Hawaiian reefs.* Master's thesis. University of Hawai'i at Mānoa. http://hdl.handle.net/10125/51453.

Littler, M. M., and Littler, D. S. (1984). Models of tropical reef biogenesis: The contribution of algae. *Progress in Phycological Research* 3: 323 – 363.

Longenecker, K. R., Bolick, H., and Kawamoto, R. (2011). *Macrofaunal invertebrate communities on Hawaii's shallow coral-reef flats: Changes associated with the removal of an invasive alien alga.* Honolulu: Bishop Museum Press.

McCook, L. J. (1999). Macroalgae, nutrients and phase shifts on coral reefs: scientific issues and management consequences for the Great Barrier Reef. *Coral Reefs* 18(4): 357 – 367.

MacManus, J. W., and Polsenberg, J. F. (2004). Coral–algal phase shifts on coral reefs: ecological and environmental aspects. *Progress in Oceanography* 60: 263 – 279.

Martinez, J. A., Smith, C. M., & Richmond, R. H. (2012). Invasive algal mats degrade coral reef physical habitat quality. *Estuarine, Coastal and Shelf Science* 99: 42 – 49.

Peyton, K. A. (2009). *Aquatic invasive species impacts in Hawaiian soft sediment habitats*. Doctoral dissertation. University of Hawai‘i at Mānoa.

Peyton, K. (2011). What Are the Factors Contributing to the Expansion of the Invasive Macroalga *Avrainvillea amadelpha* Around Kalaeloa: Pollution, Grazing or Both? Presentation, Hawai‘i Conservation Conference. Retrieved at https://vimeo.com/44835154.

Porter, J. W., Dustan, P., Jaap, W. C., Patterson, K. L., Kosmynin, V., Meier, O. W., Patterson, M. E., and Parsons, M. (2001). Patterns of spread of coral disease in the Florida Keys. In Porter, J.W. (ed.) *The Ecology and Etiology of Newly Emerging Marine Diseases*. *Hydrobiologia* 460: 1 – 24.

Pyle, R. L., Boland, R., Bolick, H., Bowen, B. W., Bradley C. J., Kane, C., Kosaki, R. K., Langston, R., Longenecker, K., Montgomery, A., Parrish, F. A., Popp, B. N., Rooney, J., Smith, C. M., Wagner, D., and Spalding, H. L. (2016) A comprehensive investigation of mesophotic

coral ecosystems in the Hawaiian Archipelago. *PeerJ* 4:e2475 DOI:10.7717/peerj.2475.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Redding, J. E., Myers-Miller, R. L., Baker, D. M., Fogel, M., Raymundo, L. J., and Kim, K. (2013). Link between sewage-derived nitrogen pollution and coral disease severity in Guam. *Marine Pollution Bulletin* 73(1): 57 – 63.

Rodgers, K. U., Cox, E., and Newtson, C. (2003). Effects of mechanical fracturing and experimental trampling on Hawaiian corals. *Environmental Management* 31(3): 0377 – 0384.

Rogers, C. S., and Miller, J. (2006). Permanent 'phase shifts' or reversible declines in coral cover? Lack of recovery of two coral reefs in St. John, US Virgin Islands. *Marine Ecology Progress Series* 306: 103 – 114.

Sansone, F. J., Spalding, H. L., and Smith, C. M. (2017). Sediment biogeochemistry of mesophotic meadows of calcifying macroalgae. *Aquatic Geochemistry* 23(3): 141 – 164.

Schaffelke, B., Smith, J. E., and Hewitt, C. L. (2006). Introduced macroalgae–a growing concern. *Journal of Applied Phycology* 18(3-5): 529 – 541.

Smith, J. E., Hunter, C. L., and Smith, C. M. (2002). Distribution and reproductive characteristics of nonindigenous and invasive marine algae in the Hawaiian Islands. *Pacific Science* 56(3): 299 – 315.

Spalding, H. L. (2012). *Ecology of mesophotic macroalgae and* Halimeda kanaloana *meadows in the main Hawaiian Islands*. Doctoral dissertation. University of Hawai'i at Mānoa.

Stamoulis, K. A., Friedlander, A. M., Meyer, C. G., Fernandez-Silva, I., and Toonen, R. J. (2017). Coral reef grazer-benthos dynamics complicated by invasive algae in a small marine reserve. *Scientific Reports* 7: 43819.

Stimson, J., Larned, S., and McDermid, K. (1996). Seasonal growth of the coral reef macroalga *Dictyosphaeria cavernosa* (Forskål) Børgesen and the effects of nutrient availability, temperature and herbivory on growth rate. *Journal of Experimental Marine Biology and Ecology* 196(1-2): 53 – 77.

Tabachnick, B., and Fidell, L. (1994). Using Multivariate Statistics. England: Pearson.

Takesue, R. K., and Storlazzi, C. D. (2017). Sources and dispersal of land-based runoff from small Hawaiian drainages to a coral reef: Insights from geochemical signatures. *Estuarine, Coastal and Shelf Science* 188: 69 – 80.

Van Heukelem, L. (2016). Does the initial diet of hatchery-reared *Tripneustes gratilla* Linnaeus) impact their effectiveness as a biocontrol for invasive seaweeds. Master's thesis, University of Hawai'i at Mānoa. http://hdl.handle.net/10125/51357.

Vega-Thurber, R. L., Burkepile, D. E., Fuchs, C., Shantz, A. A., McMinds, R., and Zaneveld, J. R. (2014). Chronic nutrient enrichment increases prevalence and severity of coral disease and bleaching. *Global Change Biology* 70(2): 544 – 554.

Wade, R. M., and Sherwood, A. R. Updating *Plakobranchus* cf. *ianthobapsus* (Gastropoda, Sacoglossa) host use: diverse algal-animal interactions revealed by NGS with implications for invasive species management. *Molecular Phylogenetics & Evolution* (submitted).

Walters, L. J., and Smith, C. M. (1994). Rapid rhizoid production in *Halimeda discoidea* fragments – a mechanism for survival after separation from adult thalli. *Journal of Experimental Marine Biology and Ecology* 175: 105 – 120.

Wedding, L. M., Lecky, J. H., Gove, J., Walecka, H. R., Donovan, M. K., Williams, G. J., Jouffray, J., Crowder, L. B., Erickson, A., Falinski, K., Friedlander, A., Kappel, C., Kittinger, J., McCoy, K., Norström, A., Nyström, M., Oleson, K. L., Stamoulis, K. A., White, C., Selkoe, K. A. (2018) Advancing the integration of spatial data to map human and natural drivers on coral reefs.. *PLoS ONE* 13(3): e0189792.

Williams, I. D., White, D. J., Sparks, R. T., Lino, K. C., Zamzow, J. P., Kelly, E. L., and Ramey, H. L. (2016). Responses of herbivorous fishes and benthos to 6 years of protection at the Kahekili Herbivore Fisheries Management Area, Maui. *PLoS One* 11(7): e0159100.

Chapter 5: Synthesis and conclusions

5.1 Overview of dissertation objectives

As I wrap up this dissertation, I will address how I met each of my initially stated objectives; when I began this work, I wanted to:

1. *Generate the first maps of hard coral distribution across the mesophotic zone (30 – 180 m) of the entire span of the MHI*: In Chapter 2, I outline the novel methodology applied to a newly processed dataset of occurrence data for scleractinians of genera *Montipora* and *Leptoseris* to create the first pan-Hawai‘i predictive distribution maps across the mesophotic zone.
2. *Explore the influence of different environmental predictors at varying spatial scales on the distribution of mesophotic corals and algae*: In Chapter 3, I examine how the influence of different environmental drivers on the distribution of *Leptoseris* sp. corals and calcified alga *Halimeda kanaloana* changes when modeled at different spatial scales (fine, moderate, and coarse). I chose Penguin Bank, Moloka‘i, as my study site for this chapter based on the results of my models in the second chapter, which seemed to pinpoint this area as a potential hotspot for *Leptoseris* sp. colonization.
3. *Generate the first map of predicted native invasive* Avrainvillea amadelpha *distribution and identify “hotspot” regions of concern across nearshore to mesophotic (0 – 90 m) O‘ahu and surrounding MHI*: Knowing that the predictive map of *Halimeda kanaloana* distribution was the first continuous layer of habitat suitability for any alga in Hawai‘i, I wanted to create a product that would have management applications statewide. In Chapter 4, I outline the methodology applied to a dataset of *Avrainvillea* occurrence across the intertidal to mesophotic ($\leq$ 90 m) zones

around O‘ahu to create the first pan-Hawai‘i map of predicted habitat suitability for this invasive macroalgae. Other than a couple of lone sightings in Kaua‘i, researchers have not spotted *Avrainvillea* adjacent to any other MHI. I produced maps of habitat suitability for all MHI for present-day and “warm seas” (+1°C annual mean seafloor temperature) conditions in order to better inform managers about the spatial allocation of their biosecurity resources.

I successfully achieved all of my objectives listed in the introduction of this dissertation.

5.2 Next steps for MCE research, management, and conservation in Hawai‘i

In the approximately five years since I conceptualized this dissertation project, I have *not* seen any meaningful change in the management practices of the state in regards to MCEs. I suspect that MCEs will, for the time being, fall victim to the “out of sight, out of mind” adage, as do all but nearshore marine habitats, at least in the public eye. I am not discouraged by this lack of attention for my chosen study zone. I’ve successfully defended the scientific, economic, ecological, and intrinsic importance of the mesophotic biota highlighted in this dissertation to grant agencies and scientific and public audiences. At this point, it has become second nature to draw parallels between shallow and mesophotic reefs as a way to hook the attention of a given audience, 99% of whom have never seen and will never see the mesophotic zone. One of my deep dives a few years ago took me beyond the 30 m upper limit of the mesophotic zone; the descent into the dim quiet was peaceful and decidedly anticlimactic- a nice reminder that this delineation between marine zones is arbitrary. All natural spaces are important and intrinsically valuable, even if humans do not directly and regularly use them. As I conducted my dissertation research, this worldview came to mind quite often.

However, the cold truth is that intrinsic value doesn’t speak- money does, and so my work has consistently provided strong and direct management applications with the economic benefits of such applications highlighted in each study. My hope is that my rigorous application of the scientific method to the various ecological datasets that were painstakingly collected and processed by me and my fellow scientists has produced useful, informative, and genuinely interesting results in the form of habitat suitability maps and related, novel inferences.

I would like to see this work inform resource managers and also act as a springboard for new projects, particularly related to habitat modeling in Hawai‘i. Models are only as good as the data used to build them, so I am confident that the researchers who come after me will be able to improve upon and expand this work. My wishlist for that improvement includes more

interdisciplinary collaboration between ecologists and mathematicians and more funding (of course) for biological sampling and model validation.

I feel privileged to have devoted my graduate career to researching the deep(er) ocean, one of the last, relatively minimally human-impacted regions on our planet, and I look forward to what's next.