

The implementation of rare events logistic regression to predict the distribution of mesophotic hard corals across the main Hawaiian Islands (#9070)

1

First submission

Please read the **Important notes** below, and the **Review guidance** on the next page.
When ready [submit online](#). The manuscript starts on page 3.

Important notes

Editor and deadline

James Reimer / 25 Mar 2016

Files

14 Figure file(s)
8 Table file(s)
3 Raw data file(s)
1 Other file(s)

Please visit the overview page to [download and review](#) the files not included in this review pdf.

Declarations

No notable declarations are present



Please in full read before you begin

How to review

When ready [submit your review online](#). The review form is divided into 5 sections. Please consider these when composing your review:

1. BASIC REPORTING

2. EXPERIMENTAL DESIGN

3. VALIDITY OF THE FINDINGS

4. General comments

5. Confidential notes to the editor

 You can also annotate this **pdf** and upload it as part of your review

To finish, enter your editorial recommendation (accept, revise or reject) and submit.

BASIC REPORTING

-  Clear, unambiguous, professional English language used throughout.
-  Intro & background to show context. Literature well referenced & relevant.
-  Structure conforms to [PeerJ standard](#), discipline norm, or improved for clarity.
-  Figures are relevant, high quality, well labelled & described.
-  Raw data supplied (See [PeerJ policy](#)).

EXPERIMENTAL DESIGN

-  Original primary research within [Scope of the journal](#).
-  Research question well defined, relevant & meaningful. It is stated how research fills an identified knowledge gap.
-  Rigorous investigation performed to a high technical & ethical standard.
-  Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

-  Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
-  Data is robust, statistically sound, & controlled.
-  Conclusion well stated, linked to original research question & limited to supporting results.
-  Speculation is welcome, but should be identified as such.

The above is the editorial criteria summary. To view in full visit <https://peerj.com/about/editorial-criteria/>

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

Lindsay M. Veazey, Erik C. Franklin, Christopher Kelley, John Rooney, L. Neil Frazer, Robert J. Toonen

Predictive habitat suitability models are powerful tools for cost-effective, mathematically robust ecological assessment. The aim of this study was to develop a predictive habitat suitability model for two genera (*Leptoseris* and *Montipora*) of mesophotic scleractinian corals across the main Hawaiian Islands. 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, we 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. Predictions were translated to spatially independent habitat suitability maps of the main Hawaiian Islands at 25 m² resolution. Our 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.

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

Author (corresponding): Lindsay Veazey; Hawai‘i Institute of Marine Biology, 46-007 Lilipuna Rd., Kaneohe, HI, 96744. lindsayv@hawaii.edu

Co-author: Erik C. Franklin; Hawai‘i Institute of Marine Biology, School of Ocean and Earth Science and Technology, University of Hawai‘i, 46-007 Lilipuna Rd., Kaneohe, HI, 96744. erik.franklin@hawaii.edu

Co-author: Christopher Kelley; The Hawai‘i Undersea Research Laboratory, 1000 Pope Rd., Marine Science Building 303, Honolulu, HI 96822. ckelley@hawaii.edu

Co-author: John Rooney; University of Hawai‘i, Joint Institute for Marine and Atmospheric Research, 1845 Wasp Blvd., Building 176, Honolulu, HI 96818. john.rooney@noaa.gov

Co-author: L. Neil Frazer; Department of Geology and Geophysics, School of Ocean and Earth Science and Technology, University of Hawai‘i at Mānoa, 1680 East-West Road, Honolulu, HI 96822. neil@soest.hawaii.edu

Co-author: Robert J. Toonen; Hawai‘i Institute of Marine Biology, 46-007 Lilipuna Rd., Kaneohe, HI, 96744. toonen@hawaii.edu

22 Abstract

23 Predictive habitat suitability models are powerful tools for cost-effective, mathematically
 24 robust ecological assessment. The aim of this study was to develop a predictive habitat suitability
 25 model for two genera (*Leptoseris* and *Montipora*) of mesophotic scleractinian corals across the
 26 main Hawaiian Islands. The mesophotic zone (30 - 180 m) is challenging to reach, and therefore
 27 historically understudied, because it falls between the maximum limit of SCUBA divers and the
 28 minimum typical working depth of submersible vehicles. Here, we implement a logistic regression
 29 with rare events corrections to account for the scarcity of presence observations within the dataset.
 30 These corrections reduced the coefficient error and improved overall prediction success (73.6%
 31 and 74.3%) for both original regression models. Predictions were translated to spatially
 32 independent habitat suitability maps of the main Hawaiian Islands at 25 m² resolution. Our maps
 33 are the first of their kind to use extant presence and absence data to examine the habitat preferences
 34 of these two dominant mesophotic coral genera across Hawai'i.

36 Introduction

38 Consistent and pervasive deterioration of marine ecosystems worldwide highlights
 39 significant gaps in current management of ocean resources (Foley et al. 2010, Douvere 2008,
 40 Crowder and Norse 2008). One such gap is the data required for informed marine spatial
 41 planning, a management approach that synthesizes information about the location, anthropogenic
 42 use, and value of ocean resources to achieve better management practices such as defining
 43 marine protected areas and implementing harvesting restrictions (Jackson et al. 2000, Larsen et
 44 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, we employ a logistic regression with rare events corrections (King and 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 across the main Hawaiian Islands. Mesophotic coral ecosystems (MCEs), located at depths of 30 - 180 meters, are extensions of shallow reefs and are known to 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 ~~low light tolerant~~ macroalgae, sponges, and hard corals (Lesser et al. 2009).

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 our 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, we model the habitat associations of mesophotic scleractinian corals because of both their intrinsic biological value as well as their potential to recolonize ~~such~~ globally threatened shallow reef areas and ~~offer~~ refuge to mobile reef organisms (Bongaerts et al. 2010, Kahng et al. 2014).

Materials and methods

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 our analyses. This imagery came from 19 dives conducted using submersibles including remotely operated vehicles (ROV), autonomous underwater vehicles (AUV), and tethered optical assessment devices (TOAD) in the ‘Au‘au Channel, Maui (13 dives) and two other geographically distinct regions: south O‘ahu and southeast Kaua‘i (6 dives). We analyzed dive video using the Coral Point Count with Excel extensions (CPCe) tool (Kohler and 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 our use of it ensures database consistency with regions processed prior to this study. In an effort to evaluate the accuracy of RFPOM, we counted all corals in 200 randomly selected screengrabs and found that this method underestimates  by 2.4% in these images.

We 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 ‘Au‘au Channel, 3517 new snapshots were collected from the additional dives across south O‘ahu and Kaua‘i (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, we analyzed 2757 unprocessed images using the RFPOM (Table 1).

We selected our environmental covariates 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) (Table 2). We defined summer and winter seasons as May – September and October – April, respectively (Kay 1994, Rooney et al. 2010). We delineated significant wave height estimates and mean current velocities by season and direction. We extracted and averaged significant wave height data from 144 days per season of twenty-four hour PacIOOS Regional Ocean Modeling System values for 2011 - 2015 (see website: <http://oos.soest.hawaii.edu/las/>). This model has a 4 km horizontal resolution with 30  vertical levels across seafloor terrain. 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. We 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:

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

where *x* represents the measured Chlorophyll *a* concentrations (mg/m³) 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 m x 25 m consistent with a conservatively estimated ± 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 ± 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 ≤ 5 m. Addition of a conservative 5 m observation area buffer to the location error area produces an observational data position of ± 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 ≥ 1 presence observation as "present" cells and all cells with only absence observations as "absent" cells. 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:

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

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

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

is the linear sum of predictor variables, $x_1, x_2, \dots, x_n$, with intercept β_0 and regression coefficients $\beta_1, \beta_2, \dots, \beta_n$. 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 ($y = 0 \gg y = 1$) (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:

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

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:

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

where the correction factor C_i is given by:

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

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 mesophotic *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 $\text{expit}(\theta) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n$ 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., $\Delta\text{AICc} \geq 2$), indicating strong preference for the best model (e.g., Hayward et al. 2007).

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:

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

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):

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

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.

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 $\Delta AICc$). We want to maximize true positives (TP) and true negatives (TN) while minimizing false positives (FP) and false negatives (FN). The

245 sensitivity-specificity sum maximization approach (Cantor et al. 1999) therefore maximizes

$$246 \quad SS_{\max} = \frac{TP}{TP + FN} + \frac{TN}{TN + FP}, \quad (9)$$

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

252 ROC curves plot the true positive test rate against the false positive test rate across
 253 different theta cutoff points (Hadley and McNeil 1982). We calculated values for sensitivity and
 254 specificity for threshold increments of 0.005 ± 1 standard deviation of the rounded mean for each
 255 model. Because each theta threshold value varied based on the genus and model, the threshold-
 256 independent area under the curve (AUC) test statistics best reflect the predictive accuracy of the
 257 model.

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

$$260 \quad OPS = \frac{TP + TN}{TP + TN + FP + FN}. \quad (10)$$

261 Overall prediction success is a measure of total correct classification of both present and absent
 262 observations. While this is a good final assessment of model classification error, consideration of
 263 the prediction success alone is not a viable evaluation method when binary data is highly
 264 imbalanced, as a value given by this method may primarily represent model success in
 265 identifying the most common observation type (Fielding and Bell 1997). We plotted our
 266 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 MHI and ran a hotspot analysis using the ArcGIS Getis-Ord G_i^* Hotspot Analysis tool. We constructed a polygon fishnet composed of 1 km² cells which encompassed all islands. We summed each 25 m² 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:

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

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 ; $\bar{X} = \frac{1}{n} \sum_{j=1}^n x_j$; and $S = \sqrt{\frac{1}{n} \sum_{j=1}^n x_j^2 - (\bar{X})^2}$.

Results

For our *Leptoseris* dataset, Geary's $C = 0.990$; for our *Montipora* dataset, Geary's $C = 0.996$. For our *Leptoseris* dataset, Moran's $I = 0.006$; for our *Montipora* dataset, Moran's $I = 0.003$. These values do not indicate any local clustering or global spatial autocorrelation within either dataset. We observed negligible levels of autocorrelation up to ~100 m for *Montipora* (Fig.

3). The absence of spatial autocorrelation within the data allows us to utilize a much less computationally intensive model than would be required if autocorrelation were present. The mixed models required to accommodate spatial autocorrelation can take hours or days to run (e.g., Breslow et al. 1993).

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. Our 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 (Supplementary material, Tables A1 – A4).

Predicted presence probability values (θ) averaged 0.051 for *Leptoseris* and 0.040 *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, we 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). We selected an appropriate threshold for each model (Table 4) in order to maximize $SS_{\max}$ (Liu et al. 2005).

Our final hotspot maps show the results of our analysis for *Leptoseris*, *Montipora*, and both genera combined across all islands (Figs. 7 - 9). We show hotspots of habitat suitability for both coral genera in red for areas of highest suitability and blue for areas of lowest suitability. We 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 ≥ 1.96 or G_i^* statistic ≤ -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.

Discussion

In this study, we 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 mesophotic *Montipora* 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.

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 separate mesophotic reefs into three distinct depth sections: upper (30 - 50 m), branching/plate dominated (50 - 80 m), and *Leptoseris* dominated (≥ 80 m). The depth at which suitability peaks for *Leptoseris* occurs at a range where steep ridges and drop-offs are plentiful in our 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 our 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 and Laboute 1986, Bridge and Guinotte 2013). The deeper distribution of *Leptoseris* may protect it from damage related to wave intensity, allowing it to colonize slopes. 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



356 also makes the coral vulnerable to smothering by sediment accumulation. The success of
 357 *Leptoseris* corals in areas of moderate current flow may be related to the improbability of
 358 sediment settlement and accumulation. While the model did not capture the same effect of
 359 current flow on *Montipora* distribution, we recognize that the morphology of some *Montipora*
 360 species is extremely similar to that of *Leptoseris*. We do not expect either genus to readily
 361 colonize highly turbid, stagnant regions, especially given that certain species of heterotrophic
 362 *Montipora* are known to exploit strong currents to meet their energy requirements (Rooney et al.
 363 2010).



364 Substrate hardness, a variable known to influence coral colonization, was notably absent
 365 from each model. Substrate hardness values were derived from acoustic backscatter imagery
 366 readings. The base resolution of these readings (50 m x 50 m) was not sufficiently detailed for
 367 purposes of this analysis. We noted plentiful coral colonization along larger surfaces like lava
 368 fingers, the hardness of which would be detectable by backscatter surveys, as well as across
 369 small rock fragments strewn across a sand flat, which would be obscured by the softness of the
 370 surrounding benthos. We can conclude that measurements of benthic hardness are not detailed
 371 enough for predictive modeling purposes at a 25 m² resolution.



372 We emphasize that the purpose of this study was to build apan-Hawai'i predictive habitat
 373 map for two dominant mesophotic coral genera. Because the scope of this study included all
 374 main Hawaiian Islands, we were constrained by the coarseness of available full-coverage
 375 environmental data. As we build on this analysis, we plan to use our maps to identify areas of
 376 interest for further study at higher resolution and to include additional variables currently only
 377 available in certain regions, such as light intensity and temperature at depth. For example, our
 378 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 m²) exist for this region, and incorporation of these data into new analyses of subsets of our study area may refine our conclusions.

Error sources and model reliability

We examined two types of error (false negatives and false positives) and analyzed our 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 and 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. We assumed coral detectability was constant across the study region and that we 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).

We 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). Our careful exclusion of overlapping observation points within each 25 m² rectified this sampling design flaw as much as possible and eliminated pseudoreplication.

Distinctions between coral genera

Our *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 Ni‘ihau, where the average winter significant wave height equaled 1.78 meters, almost double the mean significant wave height of our 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 and 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 slope of available substrate and exposure to wave

425 energy.

426 We found no records of *Montipora* presence when processing our O‘ahu dataset, which
 427 probably contributed to the very low predicted mean probability of *Montipora* occurrence there
 428 (0.1%). We believe this is due in part to the sampling pattern across south O‘ahu; we recorded
 429 62.3% of all observations processed for this region at a depth of 75 m or greater. *Montipora*
 430 prevalence is greater in the upper-to-middle mesophotic zone, and the relative deepness of the
 431 O‘ahu dives likely influenced their nonappearance in this portion of the dataset. The mean
 432 significant wave height across the mesophotic zone was lower across the southern and western
 433 coasts (1.50 m) than that observed across the northern and eastern coasts (2.37 m) of the island.
 434 As at Ni‘ihau, we assume that this high northern and eastern average height, coupled with the
 435 absence of *Montipora* presences in O‘ahu in the training dataset, greatly impacted our model's
 436 ability to detect areas of suitable habitat around the island. The results of our Getis-Ord G_i^*
 437 Hotspot Analysis corroborate the findings of Costa et al. (2015), who used Maximum Entropy
 438 software to predict the highest occurrence probability of *Leptoseris* and *Montipora* in the middle
 439 and mid-coastal ‘Au‘au Channel, respectively (Costa et al. 2015).

440 The factors influencing the distribution of coral species in shallow and mesophotic
 441 habitats differ. One of the fundamental drivers of the occurrence and abundance of coral species
 442 on shallow reefs in Hawaiian waters is wave stress (Dollar 1982, Grigg 1983, Franklin et al.
 443 2013). Given the depth range of MCEs, wave stress is unlikely to serve as a direct influence on
 444 coral occurrence but may provide secondary effects as wave events lead to debris reaching
 445 MCEs (Kahng 2014). Furthermore, the decoupled effects of environmental drivers ~~between~~ on
 446 shallow and mesophotic zones extend between the islands. In shallow reef communities
 447 *Montipora* species become ~~more relatively~~ dominant from Hawaii Island to Ni‘ihau (Franklin et

al. 2013), but appear to peak in occurrence in the mesophotic zone of Maui Nui. While the distribution of shallow corals are influenced by strong environmental drivers, 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.

Conclusions

We implemented a rare events corrected logistic regression to determine the most influential environmental predictors of mesophotic *Montipora* and *Leptoseris* colonization. Habitat preference differences between dominant mesophotic genera appear distinct and multifaceted. *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.

Acknowledgments

Authors thank the many staff members of HURL, PIFSC, and SOEST who helped acquire and process our data. L.V. thanks Dr. David Hondula (Researcher at Arizona State University) for his

dedicated GIS instruction. We dedicate this manuscript to the memory of our coauthor, Dr. John Rooney. *A hui hou*, John.

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.

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.

495 Bridge, T. and Guinotte, J. 2013. Mesophotic coral reef ecosystems in the Great Barrier Reef
496 world heritage area: their potential distribution and possible role as refugia from disturbance.
497 Research Publication no. 109. Great Barrier Reef Marine Park Authority. Townsville, QLD. 57
498 pp.
499

500 Calcagno, V. and Mazancourt, Claire D. 2010. glmulti: An R Package for Easy Automated Model
501 Selection with (Generalized) Linear Models. *The Journal of Statistical Software* 34: 1 - 29.
502

503 Cantor, S. B., Sun, C. C., Torturer-Luna, G., Richards-Koru, R., and Follen, M. 1999. A
504 comparison of C/B ratios from studies using receiver operating characteristic curve analysis.
505 *Journal of Clinical Epidemiology* 52: 885 - 892.
506

507 Choirat, C., Honaker, J., Imai, K., King, G. and Lau, O. 2015. Zelig: Everyone's Statistical
508 Software, Version 5.0-3, URL: <http://ZeligProject.org>.
509

510 Costa, B., Kendall, M., Rooney, J., Chow, M., Lecky, J., Parrish, F.A., Montgomery, A. and
511 Spalding, H. 2012. Prediction of Mesophotic Coral Distributions in the Au'au Channel, Hawai'i.
512 NOAA Technical Memorandum NOS NCCOS 149. Prepared by the NCCOS Center for Coastal
513 Monitoring and Assessment Biogeography Branch. Silver Spring, MD. 44 pp.
514

515 Costa, B., Kendall, M.S., Parrish, F.A., Rooney, J., Boland, R.C., Chow, M., Lecky, J.,
516 Montgomery, A. and Spalding, H., 2015. Identifying Suitable Locations for Mesophotic Hard
517 Corals Offshore of Maui, Hawai'i. *PloS one*, 10.

518

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

521

522 Cramer, W., Bondeau, A., Woodward, F. I., Prentice, I. C., Betts, R. A., Brovkin, V., Cox, P. M.,
523 Fisher, V., Foley, J. A., Friend, A. D., Kucharik, C., Lomas, M. R., Ramankutty, N., Sitch, S.,
524 Smith, B., White, A. and Young-Molling, C. 2001. Global response of terrestrial ecosystem
525 structure and function to CO₂ and climate change: results from six dynamic global vegetation
526 models. *Global Change Biology* 7: 357 - 373. DOI:10.1046/j.1365-2486.2001.00383.x.

527

528 Cramer, J. S. 2003. Logit models: from economics and other fields. Cambridge Univ. Press, pp.
529 66 - 67.

530

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

534

535 Dancey, C. and Reidy, J. 2004. Statistics Without Maths for Psychology. England: Pearson
536 Education Limited.

537

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

541

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

543

544 Douvère, F. 2008. The importance of marine spatial planning in advancing ecosystem-based sea
545 use management. *Marine Policy* 32: 762 - 771. DOI:10.1016/j.marpol.2008.03.021.

546

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

549

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

555

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

558

559 Geary, R. 1954. The contiguity ratio and statistical mapping. *The Incorporated Statistician* 5: 115
560 - 146.

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

563

564 Grigg, R. W. 1983. Community structure, succession, and development of coral reefs in Hawaii.

565 *Marine Ecology Progress Series* 11: 1 - 14.

566

567 Grigg, R. W. 1998. Holocene coral reef accretion in Hawai'i: a function of wave exposure and

568 sea level history. *Coral Reefs* 17: 263 - 272.

569

570 Grigg, R. W. 2006. Depth limit for reef building corals in the Au'au Channel, S.E. Hawaii. *Coral*

571 *Reefs* 25: 77 - 84.

572

573 Hadley, J. and McNeil, B. 1982. The meaning and use of the area under a receiver operating

574 characteristic (ROC) curve. *Radiology* 143: 29 - 36.

575

576 Harmelin-Vivien M. L. and Laboute, P. 1986. Catastrophic impact of hurricanes on atoll outer

577 reef slopes in the Tuamotu (French Polynesia). *Coral Reefs* 5: 55 - 62.

578

579 Hayward, M. W., Tores, P. J. D., Dillon, M. J. and Banks, P. B. 2007. Predicting the occurrence

580 of the quokka, *Setonix brachyurus* (Macropodidae: Marsupialia), in Western Australia's northern

581 jarrah forest. *Wildlife Research* 34: 194 - 199.

582

583 Hosmer, D., Hosmer, T., Cessie, S. and Lemeshow, S. 1997. A comparison of goodness-of-fit

584 tests for the logistic regression model. *Statistics in Medicine* 16: 965 - 980.

585

586 Hosmer, D. and Lemeshow, S. 2004. Applied Logistic Regression. Hoboken, New Jersey: John
587 Wiley and Sons, Inc.
588

589 Hurley, K.K.C, Timmers, M. A., Godwin, L.S., Copus, J.M., Skillings, D.J. and Toonen, R.J.
590 2016. An assessment of shallow and mesophotic reef brachyuran crab assemblages on the
591 south shore of O‘ahu, Hawai‘i. *Coral Reefs*, online early. DOI:10.1007/s00338-015-1382-z.
592

593 Hurvich, C. M. and Tsai, C. L. 1989. Regression and time series model selection in small
594 samples. *Biometrika* 76: 297 - 307.
595

596 Imai, K., King, G. and Lau, O. 2008. Toward A Common Framework for Statistical Analysis and
597 Development. *Journal of Computational and Graphical Statistics* 17: 892 - 913.
598

599 Jackson, L., Trebitz, A. and Cottingham, K. 2000. An Introduction to the Practice of Ecological
600 Modeling. *BioScience* 8: 694 - 706.
601

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

604 Kahng, S.E., Garcia-Sais, J.R., Spalding, H.L., Brokovich, E., Wagner, D., Weil, E., Hinderstein,
605 L. and Toonen, R.J. 2010. Community ecology of mesophotic coral reef ecosystems. *Coral Reefs*
606 29: 255 - 275.

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

609 Kay, E. Alison. 1994. A Natural History of the Hawaiian Islands: Selected Readings II.
610 Honolulu, Hawai'i: University of Hawai'i Press.

611

612 Kane, C., Kosaki, R. and Wagner, D. 2014. High levels of mesophotic reef fish endemism in the
613 Northwestern Hawaiian Islands. *Bull. Mar. Sci.* 90: 693 - 703. DOI:10.5343/bms.2013.1053.

614

615 Kendall, M. S., Jensen, O. P., Alexander, C., Field, D., McFall, G., Bohne, R. and Monaco, M. E.
616 2005. Benthic mapping using sonar, video transects, and an innovative approach to accuracy
617 assessment: a characterization of bottom features in the Georgia Bight. *Journal of Coastal*
618 *Research* 21: 1154 - 1165.

619

620 King, G. and Zeng, L. 2001. Logistic regression in rare events data. *Political Analysis* 9: 137 -
621 163.

622

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

626

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

630

631 Larsen, P., Barker, S., Wright, J. and Erickson, C. 2004. Use of Cost Effective Remote Sensing to
632 Map and Measure Marine Intertidal Habitats in Support of Ecosystem Modeling Efforts:
633 Cobscook Bay, Maine. *Northeastern Naturalist* 11: 225 - 242.

634

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

638

639 Lesser, M., Slaterry, M. and Leichter, J. 2009. Ecology of mesophotic coral reefs. *Journal of*
640 *Experimental Marine Biology and Ecology* 375: 1 - 8.

641

642 Lesser, M., Slaterry, M., Stat, M., Ojimi, M., Gates, R. and Grottoli, A. 2010.
643 Photoacclimatization by the coral *Montastraea cavernosa* in the mesophotic zone: light, food, and
644 genetics. *Ecology* 91:990 - 1003.

645

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

648

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

652

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).

Menard, S. 2000. Coefficients for Determination in Multiple Logistic Regression Analysis. *The American Statistician* 54: 17 - 24.

Peng, C., Lee, K. and Ingersoll, G. 2002. An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research* 96: 3 - 14.

Pacific Islands Benthic Habitat Monitoring Center (PIBHMC) 2015.
<ftp://ftp.soest.hawaii.edu/pibhmc/website/webdocs/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.

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.

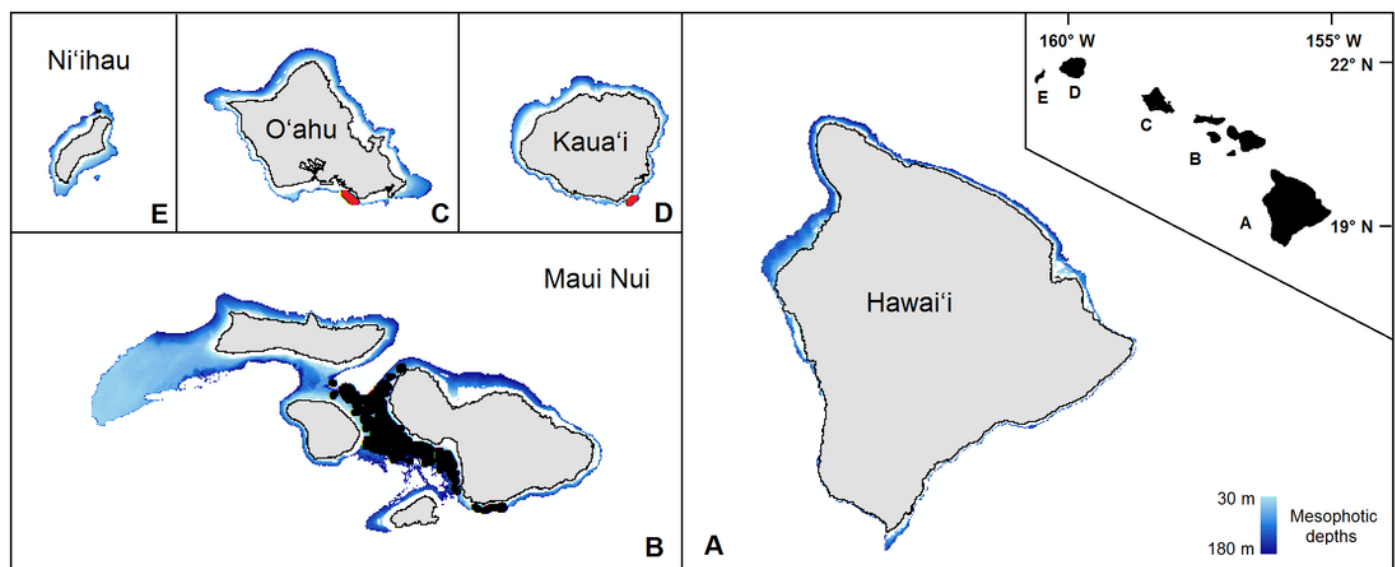
- Rooney, J., Donham, E., Montgomery, A., Spalding, H., Parrish, F., Boland, R., Fenner, D.,
Gove, J. and Vetter, O. 2010. Mesophotic coral ecosystems in the Hawaiian Archipelago. *Coral
Reefs* 29: 361 - 367.
- Ruiz, G., Fofonoff, P. and Hines, A. 1999. Non-indigenous species as stressors in estuarine 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., Vanwalleghe, 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.

699 Webster, J.M., Clague, D.A., Riker-Coleman, K., Gallup, C., Braga, J.C., Potts, D., Moore, J.G.,
 700 Winterer, E.L. and Paull, C.K. 2004. Drowning of the 150 m reef off Hawaii: A casualty of
 701 global meltwater pulse 1A? *Geology* 32: 249 - 252.
 702
 703 Wright, D.J., Pendleton, M., Boulware, J., Walbridge, S., Gerlt, B., Eslinger, D., Sampson, D. and
 704 Huntley, E. 2012. ArcGIS Benthic Terrain Modeler (BTM), Environmental Systems Research
 705 Institute, NOAA Coastal Services Center, Massachusetts Office of Coastal Zone Management.
 706
 707 Yokoyama, Y., Purcell, A., Lambeck, K., and Johnston, P. 2001. Shore-line reconstruction
 708 around Australia during the last glacial maximum and late glacial stage. *Quaternary*
 709 *International* 83: 9 - 18.

1

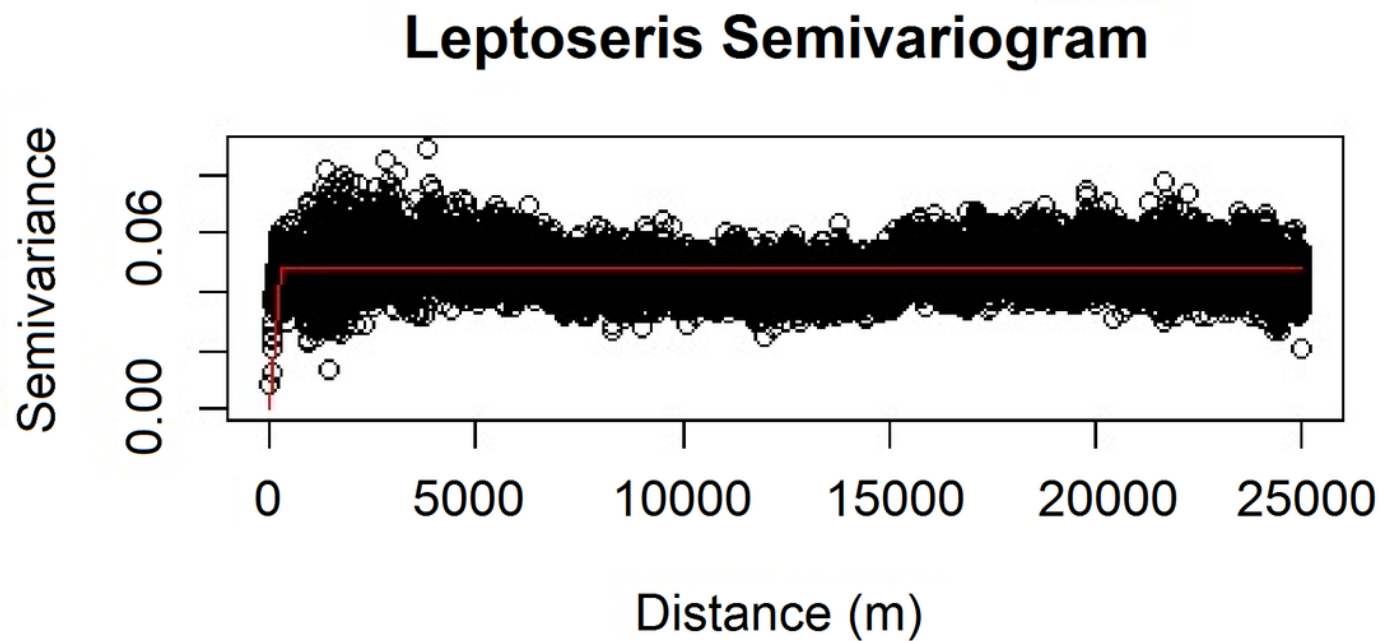
The mesophotic zone of the main Hawaiian Islands

The study domain, demarcated in blue, encompasses the mesophotic zone (30 - 180 m in depth) of the main Hawaiian Islands. Black circles are the observations from the pre-existing Maui Nui dataset. Red circles are the previously unprocessed observations in south O'ahu and southeast Kaua'i.



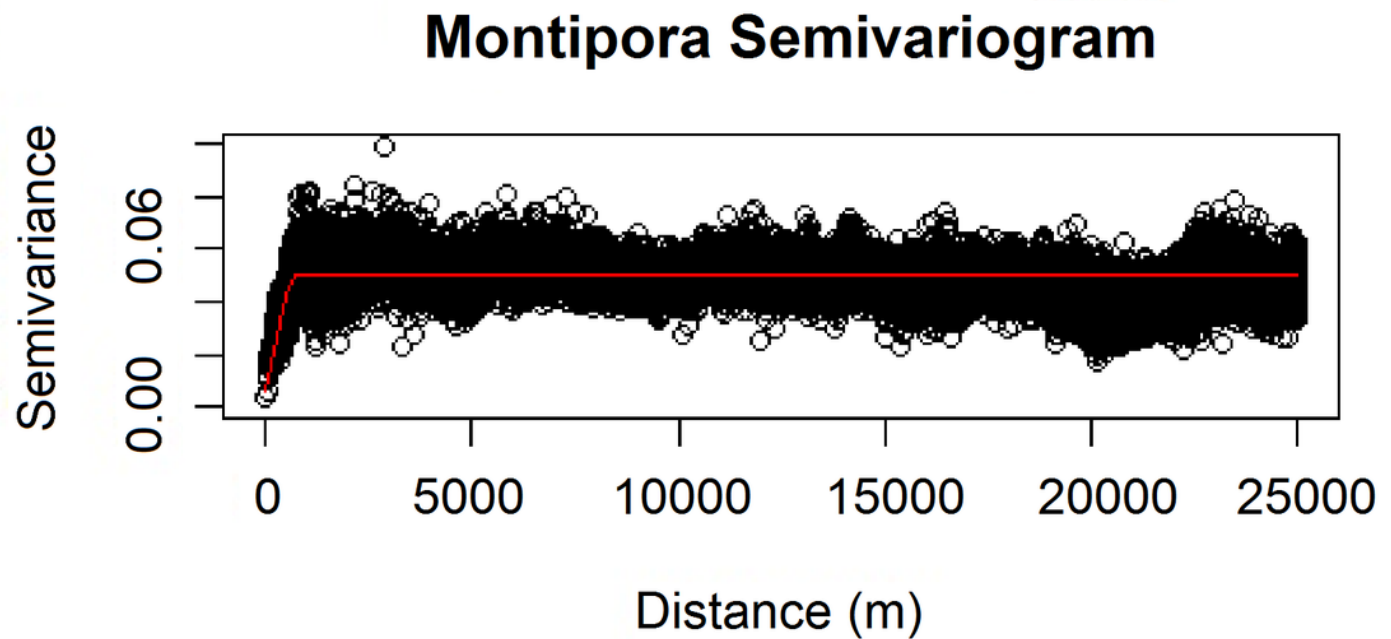
2

Modeled spherical semivariogram for *Leptoseris*.



3

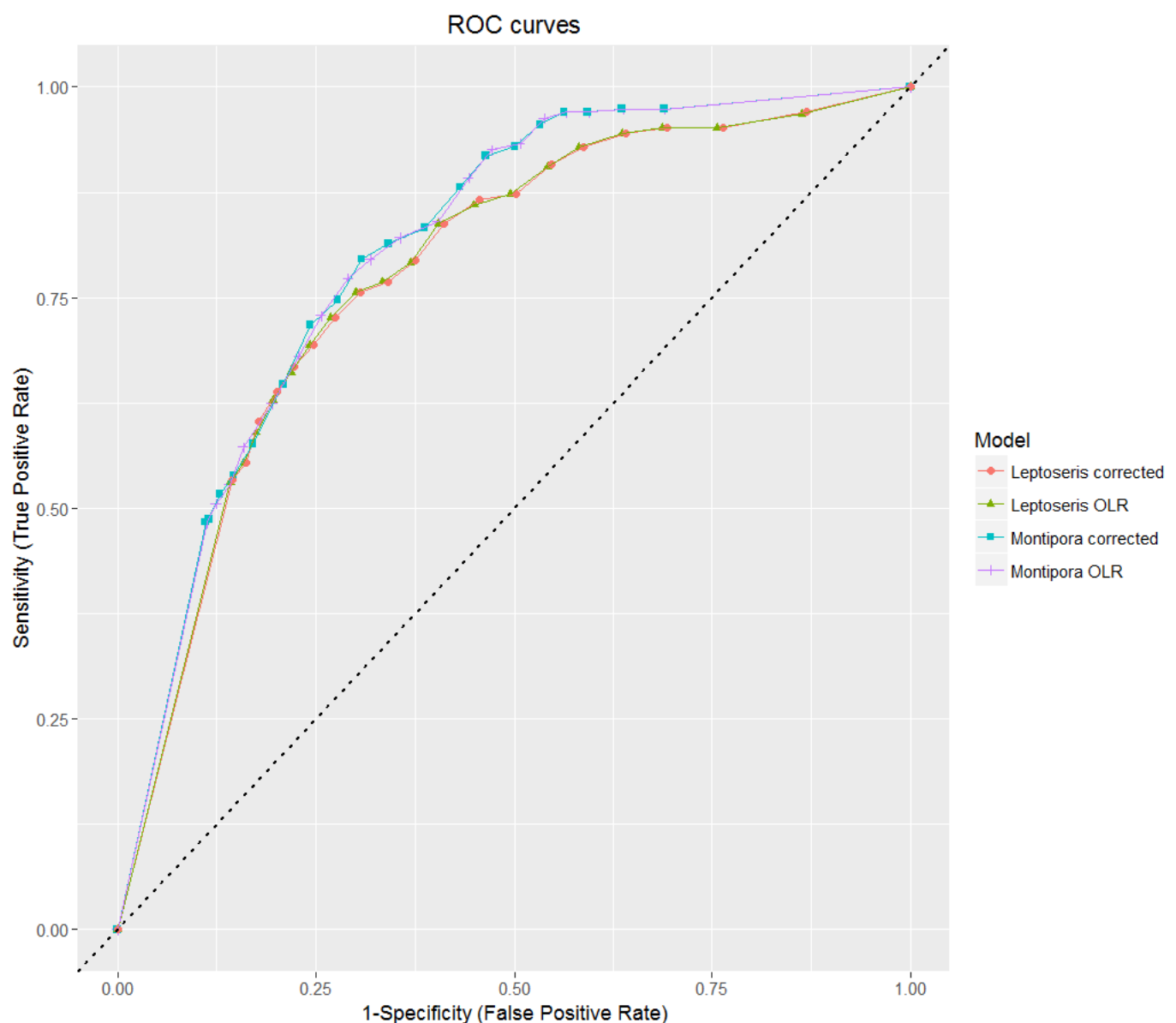
Modeled spherical semivariogram for *Montipora*.



4

ROC curves for all models.

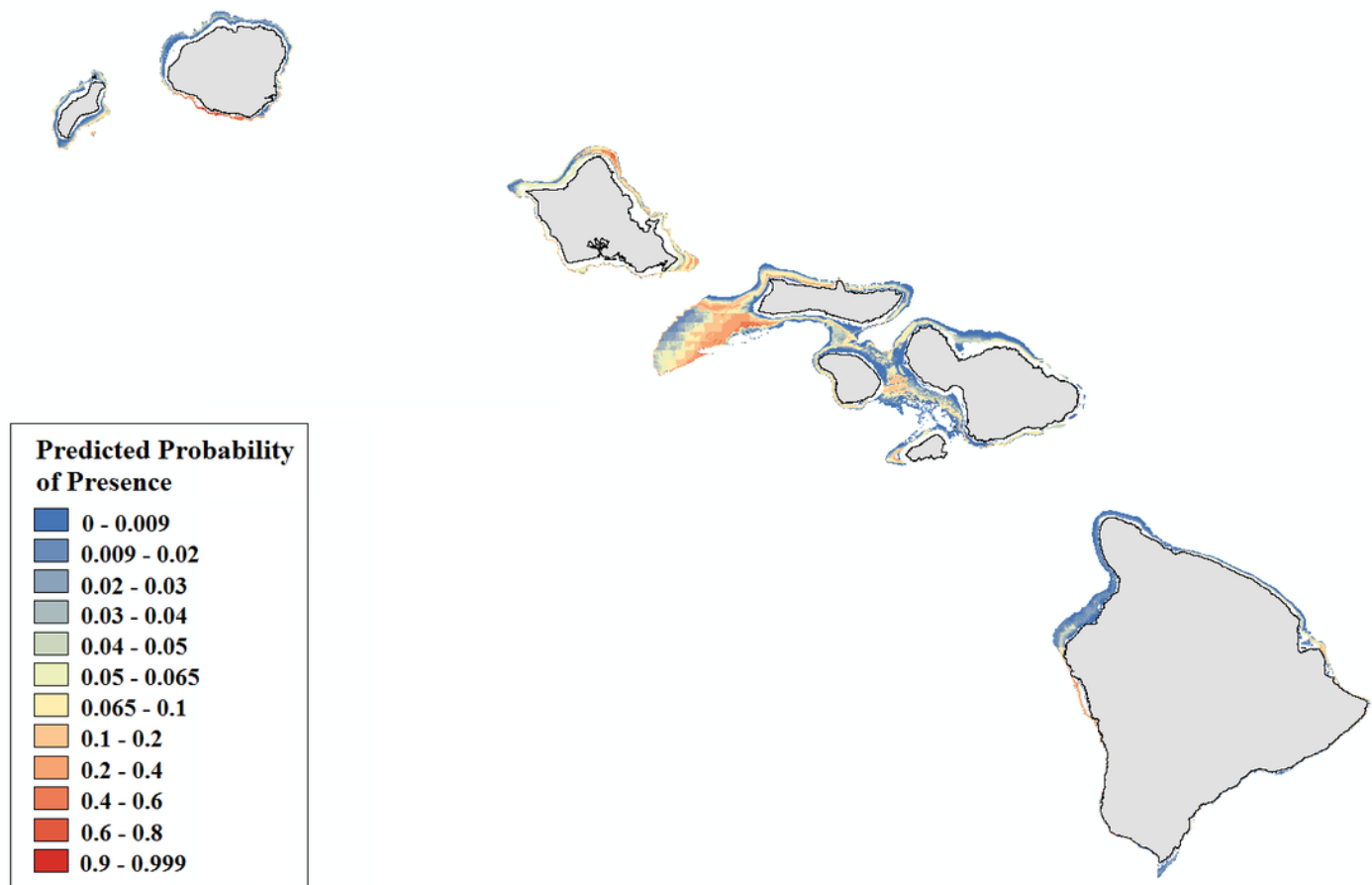
AUC values for all models fall in between 0.7 and 0.9, which indicates predictive reliability. The dashed line from (0, 0) to (1, 1) indicates the null threshold at which model performance is considered unacceptable (< 0.5).



5

Modeled area of suitable habitat for *Leptoseris*.

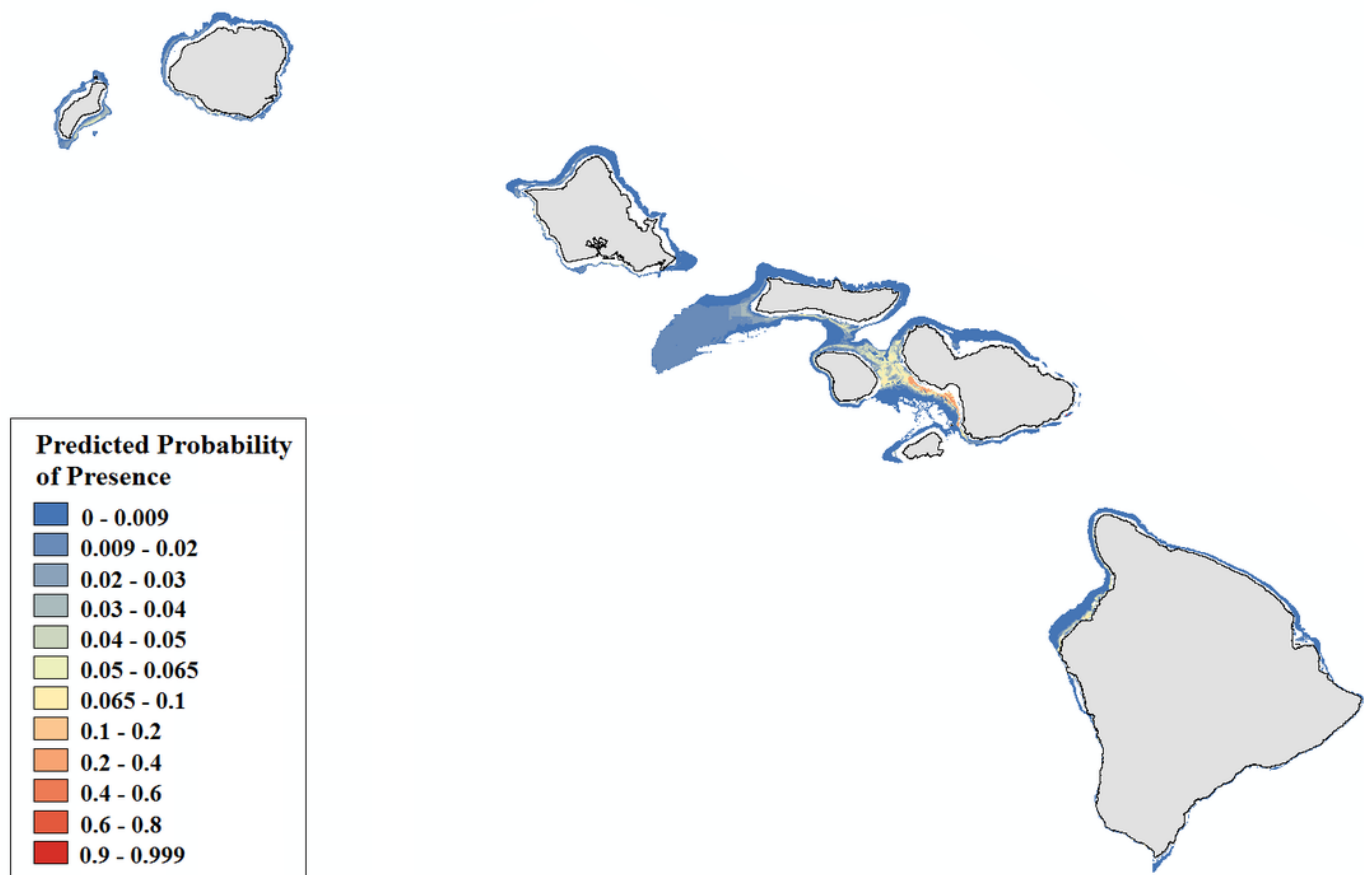
Probability of presence is depicted along a color gradient ranging from red (1; most suitable) to blue (0; least suitable).



6

Modeled area of suitable habitat for *Montipora*.

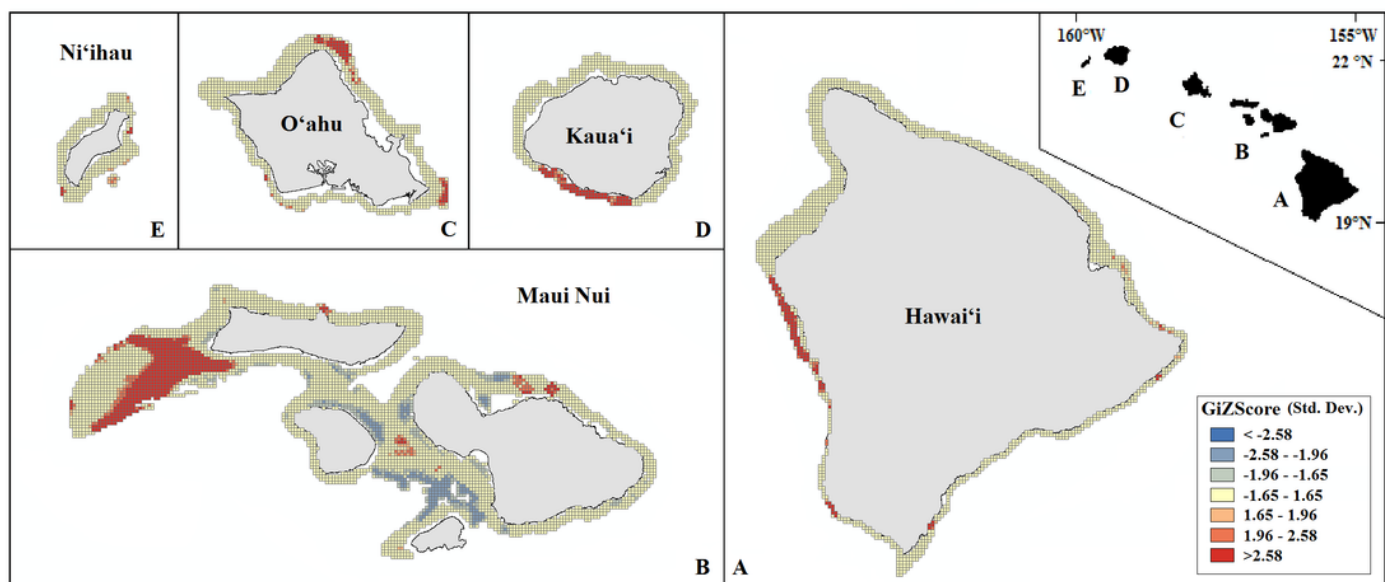
Probability of presence is depicted along a color gradient ranging from red (1; most suitable) to blue (0; least suitable).



7

Mapped result of our Getis-Ord G_i^* hotspot analysis performed for probability estimates of *Leptoseris* occurrence.

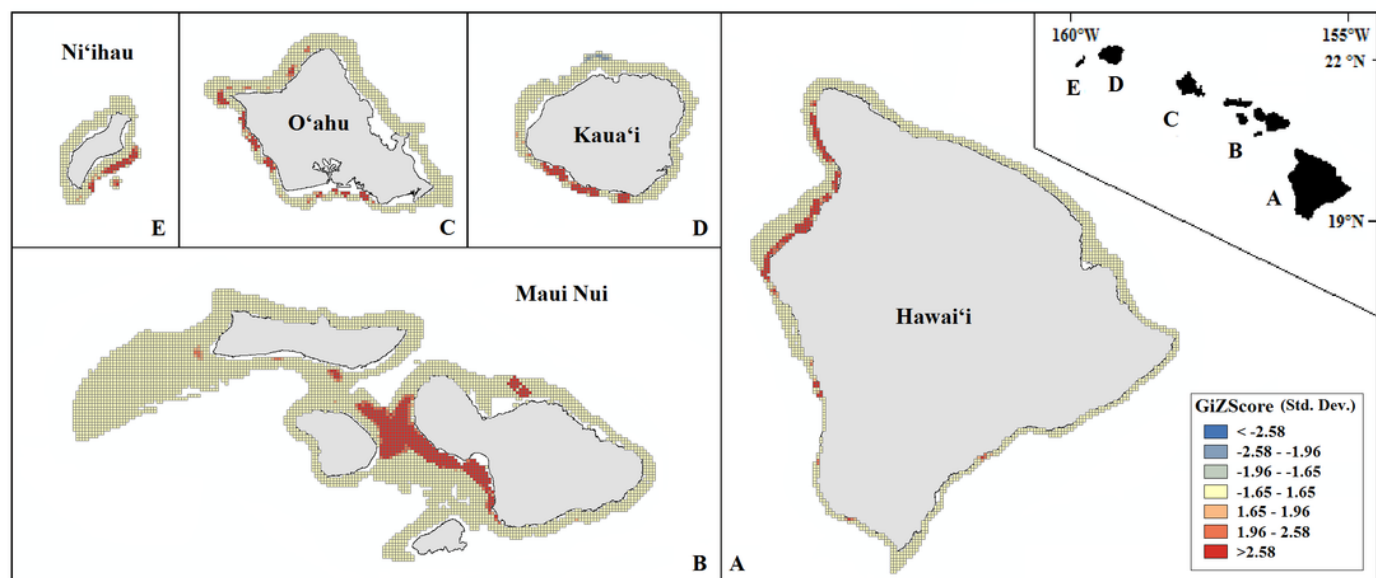
A significant hotspot is < -1.96 or > 1.96 ; here, all hotspots are shown in red (> 1.96) or blue (< -1.96).



8

Mapped result of our Getis-Ord G_i^* 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).



9

Mapped result of our Getis-Ord G_i^* 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).

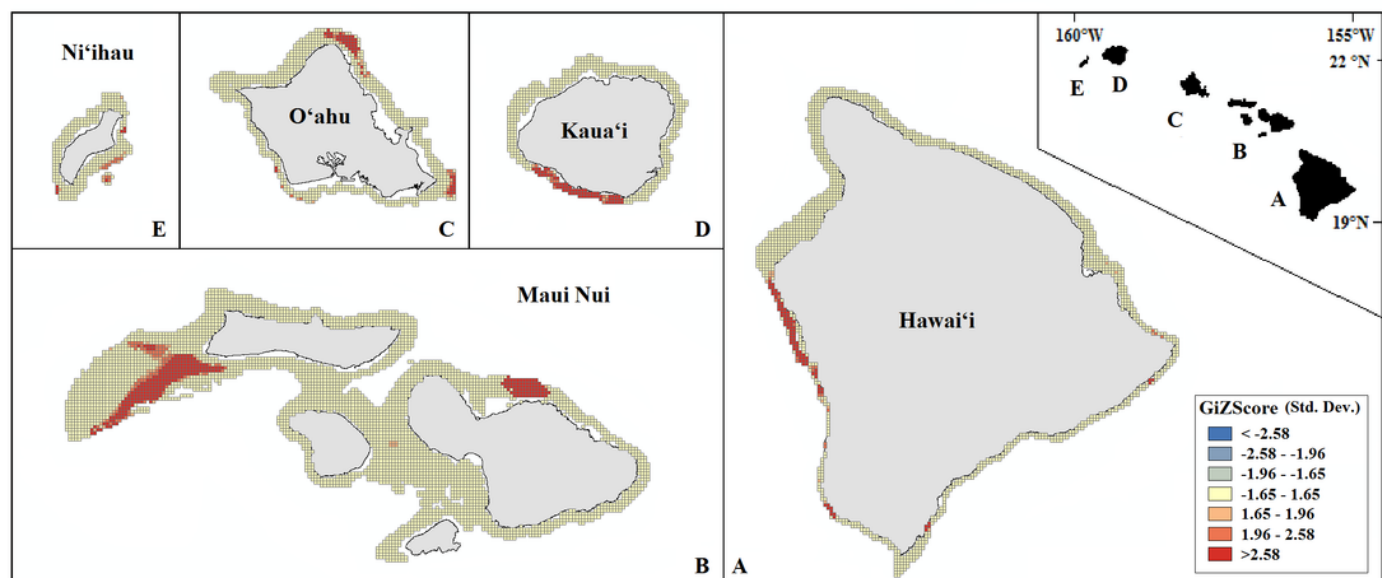


Table 1 (on next page)

Number of field observations for each coral genus.

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

Table 2(on next page)

List of all variables considered for inclusion in our analyses.

1
2
3
4

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	<i>Leptoseris</i>
					<i>Montipora</i>
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 km 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 m x 50 m resampled to 25 m 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 Hawaii Regional Ocean Model	4 km 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 Hawaii SWAN Wave Model	1 km x 1 km	Sig. wave height (summer) (m)
					Sig. wave height (winter) (m)

Table 3(on next page)

Summary statistics for theoretical semivariogram models.

Genus	Sum of squares	Input σ^2	Input ψ	Actual σ^2	Actual ψ	Actual τ^2
<i>Leptoseris</i>	2940.671	0.055	218	0.051	206.909	0
<i>Montipora</i>	14013.610	0.020	390	0.032	390.000	0.003

1

Table 4(on next page)

Predictive models 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.

1

Genus	θ threshold	TP	TN	FP	FN	TP _{cv}	TN _{cv}	FP _{cv}	FN _{cv}	Sensitivity	Specificity	SS _{max}	OPS	OPS _{cv}	AUC
<i>Leptoseris</i>	$\theta_{OLR} = 0.065$	223	4133	1525	84	219	3757	1894	94	0.7264	0.7305	1.4569	73.0%	66.7%	0.782
	$\theta_{corr} = 0.067$	220	4168	1490	87	182	4000	1651	131	0.7166	0.7367	1.4533	73.6%	70.1%	0.780
<i>Montipora</i>	$\theta_{OLR} = 0.064$	200	4299	1536	69	159	4453	1406	86	0.7435	0.7368	1.4803	73.7%	75.6%	0.808
	$\theta_{corr} = 0.0625$	198	4336	1499	71	165	4406	1453	80	0.7361	0.7465	1.4792	74.3%	74.9%	0.809