MISS LINN  SEKUND (Orcid ID : 0000-0002-3218-2158)

Received Date : 24-Jun-2016

Revised Date   : 08-Feb-2017

Accepted Date : 12-Feb-2017

Article type      : Regular Paper

Corresponding Author email:  linn.sekund@gmail.com

# Explaining island-wide geographical patterns of Caribbean fish diversity: A multi-scale seascape ecology approach

Linn Sekund[1], Simon Pittman[2, 3]

[1]Erasmus Mundus Marine Biodiversity and Conservation (EMBC), Ghent University, Belgium
[2]U.S. National Oceanic & Atmospheric Administration Biogeography Branch, 1305 East-West Highway, Silver Spring, Maryland 20910, United States of America
[3]Centre for Marine and Coastal Policy Research, Marine Institute, Plymouth University, Plymouth, PL4 8AA, United Kingdom

**Key words**: multi-scale, fish species richness, regression trees, environmental predictors, seascape ecology, landscape ecology

## Abstract

Geographical patterning of fish diversity across coral reef seascapes is driven by many interacting environmental variables operating at multiple spatial scales. Identifying suites of variables that explain spatial patterns of fish diversity is central to ecology and informs prioritization in marine conservation, particularly where protection of the highest biodiversity coral reefs is a primary goal. However, the relative importance of conventional within-patch variables versus the spatial patterning of the surrounding seascape is still unclear in the ecology of fishes on coral reefs. A multi-scale seascape approach derived from landscape ecology was applied to quantify and examine the explanatory roles of a wide range of variables at different spatial scales including: 1.) within-patch structural attributes from field data (5 x 1 $m^2$ sample unit area); 2.) geometry of the seascape from seafloor maps (10 m – 50 m radius seascape units); and wave exposure from a hydrodynamic model (240 m resolution) for 251 coral reef survey sites in the U.S. Virgin Islands. Nonparametric statistical learning techniques using single classification and regression trees (CART) and ensembles of boosted regression trees (TreeNet) were used to: 1.) model interactions; and 2.) identify the most influential environmental predictors from multiple data types (diver surveys, terrain models, habitat maps) across multiple spatial scales (1 $m^2$ to 196,350 $m^2$). Classifying the continuous response variables into a binary category and instead predicting the presence and absence of fish species richness hotspots (top 10% richness) increased the predictive performance of the models. The best CART model predicted fish richness hotspots with 80% accuracy. The statistical interaction between abundance of living scleractinian corals measured by SCUBA divers within 1 $m^2$ quadrats and the topographic complexity of the surrounding seafloor terrain (150 m radius seascape unit) measured from a high resolution terrain model best explained geographical patterns in fish richness hotspots. The comparatively poor performance of models predicting continuous variability in fish diversity across the seascape could be a result of a decoupling of the diversity-environment relationship due to structural degradation leading to a widespread homogenisation of coral reef structure.

## Introduction

The conservation importance of species richness hotspots goes beyond the focus on the number of species, because diverse coral reefs are positively correlated with fish biomass (Duffy *et al*. 2015), and confer greater resilience to disturbance and disease than low diversity reefs (Raymundo et al. 2009, Rogers 2012). Prioritizing protection of the highest diversity coral reefs is often a key objective of coral reef management strategies, such as the establishment of marine protected areas, because diverse coral reefs perform important functional roles in the maintenance of ecosystem health and the provisioning of valuable ecosystem services (Holmlund & Hammer 1999; Harborne *et al.* 2006).

Identifying the key environmental drivers of the spatial patterning of diversity across coral reefs is important in identifying patches of resilience to environmental change and in prioritizing coral reefs for conservation actions (Rogers 2013). Coral reefs, however, exhibit complex topographic and compositional structure at a range of spatial scales presenting a major challenge to explaining patterns in biodiversity. For instance, the finest spatial scale(s) conventionally measured in fish ecology are the within-patch measurements of leaf length, height or coral cover typically measured in centimetres or percentage cover within a handheld quadrat or with point counts along a video transect. At broader spatial scales, measurements can include the surrounding seascape composition (the abundance and variety of patch types), seascape configuration (spatial arrangement of patches), or terrain morphology (e.g., topographic complexity of a 3D surface) measured from digital models. In turn, these structural patterns are influenced by even broader scale patterns and processes such as the hydrodynamic regime, the geomorphology of the coast and other bio-physical and chemical patterns and processes such as freshwater inflow and salinity. The relative importance of within-patch structural heterogeneity, patch-mosaic and terrain morphology and broader patterning are rarely known because studies tend to focus at a single scale measuring variables that are finer than the spatial scales routinely traversed by many fishes. Even where multi-scale studies are conducted, scale selection is often too narrow, with few studies using biological reasons for the scale of measurements (Bird Jackson & Fahrig 2014). In addition, although cross-scale environmental interactions are known to be important in ecology (Holling 1994), statistical interactions between environmental variables are also rarely examined across scales (Cushman & McGarigal 2004).

What ecologists choose to measure when quantifying environmental variability and the spatial scales of those measurements have important consequences for the way that variable importance is interpreted in ecology (Wiens 1989; Meetenmeyer 1989). This has implications for our understanding of the drivers of biodiversity, forming a major knowledge gap in marine ecology and in conservation science (Wolman 2006). In the absence of information on a single scale with which to measure the environment, we propose that it is ecologically meaningful to apply an exploratory multi-scale approach (Wiens 1989; Kotliar & Wiens 1990; Schneider 2001; Pittman & McAlpine 2003). A multi-scale analysis offers a way to identify an optimal spatial scale(s) where seascape structure most strongly correlates with a specific response variable. This is particularly relevant when determining the drivers of assemblage diversity where species perceive and respond differently and at different spatial scales to patterning in the surrounding seascape (Pittman *et al*. 2007). In landscape ecology, this is referred to as an organism-centered approach (Pearson *et al*. 1996; Betts *et al*. 2014) where environmental heterogeneity is perceived as a nested spatial hierarchy of structures (Kotliar & Wiens 1990).

In addition to the problem of selecting a single ecologically ambiguous scale for measurement, many studies also select, often through convention or convenience, only a single data type for analyses. For example, both response and predictor variables can usually be measured as either a continuous metric or a categorical metric or index. Similarly, should we measure structural attributes of the three dimensional terrain or the two-dimensional patch mosaic (McGarigal *et al*. 2009; Pittman & Olds 2015) or adopt a pluralistic approach? (Price *et al*. 2009). It is rarely known *a priori* which is likely to be most ecologically meaningful and the preferred data type is rarely explicitly stated in the research questions. As such, an exploratory multi-scale and multi-model approach may better capture the complex assemblage-seascape associations, especially with diverse assemblages with diverse life history characteristics and mobility (Pittman & Knudby 2015).

The general approach developed here integrates conventional field measurements with novel landscape ecology techniques to quantify vertical (3D) and horizontal (2D) benthic structure at multiple spatial scales using a combination of continuous and categorical environmental variables. Scale selection was guided by a review of home range sizes for a range of common Caribbean reef-associated fishes. Both single tree models and ensembles of regression trees were used to determine the contribution that each predictor type makes in explaining spatial patterns of fish diversity across the coral reefs of St. John, U.S. Virgin Islands in the Eastern Caribbean. In addition to explaining number of fish species (referred to as fish species richness) we also quantify taxonomic diversity.

Three primary research questions were addressed to examine the contribution of environmental predictors in models of fish species diversity patterns across coral reefs of St. John in the U.S. Virgin Islands:

1. Does within-patch structure explain more of the variability in fish species richness and taxonomic diversity than the surrounding seascape structure?
2. Which spatial scale(s) of seascape patterning best explain fish species richness and taxonomic diversity?
3. Which interacting environmental variables best characterize fish richness hotspots?

## Material and methods

### Study area

St. John is one of three main islands of the United States Virgin Islands (USVI) located on the Puerto Rican shelf in the Eastern Caribbean (N 18° 20' 12.521'' W 64° 43' 41.1427'') (Figure 1). St. John's topography consists of steep slopes, exposed cliffs and dense vegetation. The near-shore seascape of

St. John supports a complex mosaic of habitat types including seagrass, mangrove and coral reef (including colonized pavement, linear and patch reefs). Coral reefs and associated habitats of the Virgin Islands provide important economic, cultural, social, and environmental values and benefits to people. The economic value of coral reefs in the USVI has been estimated at USD $187 million annually (van Beukering 2011). In the past five decades, regardless of protected area designation, coral reef communities in the U.S. Virgin Islands have declined in structural complexity and ecological integrity due to a variety of environmental stressors including climate change, disease, coastal development and fishing (Rogers & Beets 2001). Around St. John, live hard coral cover ranges from 0 to 86 % (mean 10.41 SE+/- 0.7). Recent comparative analyses on MPA performance revealed no significant difference in fish species richness or reef condition inside versus outside of St. John MPAs (Pittman *et al.* 2014).

*Data collection*

Fish surveys

Survey missions were conducted annually in July from 2002 to 2011 around the island of St. John as part of NOAA's Coral Reef Ecosystem & Assessment Monitoring project (NCREM) operated by NOAA National Centres for Coastal Science' (NCCOS) Biogeography Branch in collaboration with the U.S. National Park Service. The number of fish species were quantified from underwater visual surveys of fish at spatially random locations over hardbottom habitat classes as represented in the NOAA benthic habitat map (Kendall 2001). The depth range of samples was 11 ±0.4 m with a minimum of 0.18 m and maximum of 27.6 m.

At each sample location a trained observer on SCUBA swam along a 25 m long by 4 m wide belt transect (100 $m^2$ sample unit area) for fifteen minutes maintaining a constant speed while identifying and counting the abundance of all fish observed, including in the water column. The species abundance was recorded in 5 cm size class increments using the visual estimation of fork length (Friedlander *et al*. 2013). A total of 251 samples from hardbottom habitat within 4 km of the coastline were used in the analysis (Figure 2). Fish data are available online at http://www8.nos.noaa.gov/bpdmWeb/queryMain.aspx.

Taxonomic diversity is a proxy for functional diversity where a greater taxonomic diversity usually represents a greater variety of fundamentally different life histories and functional groups (Warwick & Clarke 1995). Taxonomic diversity was calculated in the computer software package PRIMER version 6 (Plymouth Routines In Multivariate Ecological Research) from an untransformed species-abundance matrix (Clarke & Warwick 2001). Taxonomic diversity is defined as the average weighted path length between every pair of individuals (Equation 1). Weightings were put on shorter branch length in order to have more weight on classes and order than species and genus.

$$\Delta = \frac{\Sigma\Sigma_{i<j} W_{ij} X_i X_j + \Sigma_i 0.\, x_i (x_i - 1) / 2}{\Sigma\Sigma_{i<j} x_i x_j + \Sigma_i x_i (x_i - 1) / 2}$$

*Equation 1*

*Where $X_iX_j$ is the abundance of i and j species and $W_{ij}$ is the distinctness weight given to the path length linking i and j in the hierarchical classification (*Clarke & Warwick 2001).

### *Within-patch structure*

Along each fish transect, fine scale benthic habitat composition was quantified within 1 m² quadrats at five locations to subsample the 100 $m^2$ fish sample unit area. Estimates of the percentage cover of benthic species and types of biogenic structure such as algae, gorgonians, live coral and sponges were measured in the field by trained scientific divers. From these data additional metrics were calculated to include the coral-macroalgal ratio, maximum hard coral cover, maximum crustose cover and species richness of scleractinian corals. The quadrat locations were pre-selected to ensure a sample was collected at least once within each 5 m length of the transect. The quadrat was divided into 100 (10 x 10 cm) smaller squares to help estimate cover. The percentage of cover was estimated within the quadrat in a two-dimensional plane perpendicular to the observer's line of vision (Friedlander *et al.* 2013). The topographic complexity (rugosity) was measured using the chain-tape method (McCormick 1994). Two six-meter long chains were draped over the substratum along the transect and the distance of the chain along the horizontal tape was recorded by the fish observer. Mean and maximum rugosity were calculated for each survey location.

### *Seascape terrain complexity*

Water depth and the topographic complexity of the seafloor were quantified at multiple spatial scales from a high resolution (3 x 3 m) bathymetric terrain model derived from airborne hydrographic LIDAR (Light Detection and Ranging). The LiDAR sensor measures the difference in the time of reflectance for pulses of high energy laser to return to the aircraft from the seafloor and the water surface to estimate water depth (Pittman *et al.* 2013). Topographic complexity was measured by applying a slope-of-the-slope morphometric to the digital terrain using a Geographical Information System (GIS). Slope-of-the-slope, a measure of terrain roughness (Figure 3), was calculated by creating an initial slope surface from the bathymetry and then calculating slope of the initial slope surface to create a second derivative surface (Pittman *et al.* 2009). Slope-of-the-slope has been described as the maximum rate of slope change between neighbouring cells (Pittman *et al.* 2009).

Mean water depth and the maximum slope-of-the-slope were then quantified at multiple spatial scales (seascape sample unit areas of: 10 m radius = 314 $m^2$; 25 m radius = 1,964 $m^2$; 50 m radius = 7854 $m^2$, 100 m radius = 31,416 $m^2$; 150 m radius = 70,686 $m^2$; 200 m radius = 125,664 $m^2$ and 250 m radius = 196,350 $m^2$) surrounding each sample point using a moving window analysis in ArcGIS (ArcGIS Spatial Analyst Neighbourhood Tool) (ESRI Inc., http://www.esri.com/).

*Ecological scale selection*

The spatial scale range selected (seascape sample unit area) for quantification of mapped seascape structure was guided by existing data on the known size of fish home ranges for 11 common Caribbean reef associated fishes (mean body length =14.2 ±SE11.7). A literature review revealed that home ranges varied widely in size (mean= 752 $m^2$ ± 1098), from 0.5 $m^2$ for redlip blenny to 2,874 $m^2$ for red hind (Table 1). Our analyses was designed to ensure that the scale range selected for the analytical window size overlapped the scale domains for the selected fish home range sizes and extended to encompass the broader surrounding seascape (Figure 4).

| **Species name** | **Home range (m²)** | **Fork length (cm)** | **Technique used** | **Study duration** | **Location** | **References** |
|---|---|---|---|---|---|---|
| Redlip blenny (*Ophioblennius atlanticus*) | 0.5 | 5.6 | Visual census and mark-recapture | 1 month | Barbados and Curaçao | Nursall 1977 |
| Juvenile blue parrotfish (*Scarus coeruleus*) | 35.8 | 4-9.5 | Focal observation | 9 month | Florida Keys | Overholtzer & Motta 1999 |
| Sharpnose puffer (*Canthigaster rostrata*) | 71 | 5 | Mark and recapture | 1 month | Panama | Sikkel 1990 |
| Stripped parrotfish (*Scarus iserti*) | 80 | NA | Territories marked at regular intervals | 2 month | Belize | Mumby &Wabnitz 2002 |
| Red band parrotfish (*Sparisoma aurofrenatum*) | 112 |  | Focal observation | 1 month | Florida Keys | Catano et al. 2015 |
| Caribbean wrasses (*H.bivittatus, H.garnoti, H.maculipinna, H.poeyi, T.bifasciatum*) (Labridae) | 132 ± 21.2 | 2-4 | Visual census mark and recapture | 6 month | St Croix, USVI | Jones et al. 2005 |
| Spanish hogfish (*Bodianus rufus*) | 148 | 22-40 | Visual census | 2 month | Panama | Hoffman 1983 |
| Redfin parrotfish (*Sparisoma rubripinne*) | 784 | NA | Territories marked at regular intervals | 2 month | Belize | Mumby & Wabnitz 2002 |
| Schoolmaster snapper (*Lutjanus apodus*) | 1290 | >24 | Acoustic telemetry | 24 hours | St John, USVI | Hitt et al. 2011 |
| Bluestripped grunt (*Haemulon scirius*) | 2778 | >24 | Acoustic telemetry | 24 hours | St John, USVI | Hitt et al. 2011 |
| Red hind grouper (*Epinephelus guttatus*) | 2874 | NA | Acoustic telemetry | 152 days | Puerto Rico | Shapiro et al. 1994 |

Table 1: Home range area for 15 species of Caribbean reef fish.

### *Seascape patch-mosaic composition & patch proximity*

Using techniques from landscape ecology, the seascape mosaic composition (proportion of major patch types) was quantified from a benthic habitat map at multiple spatial extents surrounding each fish sample point. The seascape sample unit was defined using circular buffers (same radii as for terrain metrics) within a GIS for each survey site using the Diversity Calculator tool (ESRI script Diversity Calculator) developed by NOAA Biogeography Branch (http://arcscripts.esri.com/details.asp?dbid=15258).

Patch-mosaic composition was quantified from the NOAA Biogeography Branch Shallow-Water Benthic Habitat map (Zitello *et al.* 2009) which was mapped with a 100 m resolution minimum mapping unit. The major cover types of interest were seagrass percentage cover with a map user accuracy of 91.5% (n=71) and percentage of high complexity hardbottom habitat (a combination of aggregate reef, aggregate patch reef, individual reef, pavement, pavement with sand channels and spur and groove) with a map accuracy of 86.1%. Overall map accuracy for all hard bottom areas was 96 % (user's accuracy for hard bottom areas=97.3%, n=299) (Zitello *et al.* 2009). In addition to quantifying the area of major habitat types (patch types) we also created a categorical variable of habitat type whereby each individual class of habitat type was assigned a unique value.

To consider the potential effects on fish diversity of connectivity between coral reefs and complementary non-reef habitat types (Pittman *et al*. 2007b; Olds *et al*. 2012), we quantified proximity between fish survey sites on coral reefs and the nearest seagrass and mangrove patches, a raster surface of patch proximity was created using Euclidean distance across the insular shelf.

### *Wave exposure*

Wave exposure can have a considerable influence on the structuring of coral reef fish assemblages (Fulton & Bellwood 2005; Brown *et al*. 2016). Here we quantified the spatial pattern of wave exposure from points predicted across a 240 m resolution grid using the CariCOOS (Caribbean Coastal Ocean Observing System) Nearshore Wave Model (CNWM) (Canals-Silander in press) based on the Simulating WAve Nearshore SWAN model (Booij *et al.* 1999). The model was designed to be applied in shallow waters with ambient currents, bays, estuaries and channels and provides an annual average of wave power. SWAN is an adaptation of the WAM wave model for deep and intermediate water by incorporation of depth-induced wave breaking and triad wave-wave interactions (Salmon & Holthuijsen 2015).

*Data analysis*

The data for three response variables and 36 explanatory variables were checked for outliers, homogeneity and normality of variance (Table 2). Collinearity among predictors and spatial autocorrelation were examined according to the method suggested by Zuur *et al.* (2010).

| | Variables | Data source | How it was calculated | Resolution | Mean ± SE (Min-max values) |
|---|---|---|---|---|---|
| **Unique ID** | Survey Index | Transect data | 251 hardbottom surveys sites, with a unique ID and geographic coordinates | 100 m² | |
| **Response variables** | Species richness | | Fish visual census, at each sample location a trained observer on SCUBA swam along a 25 m long by 4 m wide belt transect for fifteen minutes maintaining a constant speed while identifying and counting the abundance of all fish observed. The species abundance was recorded in 5 cm size class increments using the visual estimation of fork length | | 23.6±0.4<br>(8-44) |
| | Taxonomic diversity | | Calculated in the software PRIMER version 6. Weights were put on shorter branch length in order to have more weight on classes and order than species and genus $\Delta = \frac{\Sigma\Sigma_{i<j} w_{ij} x_i x_j + \Sigma_i 0.\ x_i (x_i-1)/2}{\Sigma\Sigma_{i<j} x_i x_j + \Sigma_i x_i (x_i-1)/2}$ | | 58±0.9<br>(11.6-75.5) |
| | Richness hotspots | | The top 10% of species richness was categorized as 1 and the remaining 90% as 0 | | |
| **Within-patch structures** | Live coral + Gorgonians | Quadrat data | Sum of the maximum values of live coral and gorgonian cover from the 5 quadrats on the transect for each location | 100 m² | 10.29±0.5<br>(0.7-44) |
| | Biogenic structure | | Sum of sponges, coral and gorgonian cover. The sum of sponges correspond to the sum of maximum percentage cover of encrusting and upright sponges from the 5 quadrats | | 13.45±0.6<br>(0.1-48) |
| | Algae | | Sum of maximum percentage cover of turf algae, microalgae, crustose algae and cyanobacteria from the 5 quadrats | | 54.26±1.7<br>0-99 |
| | Live hard coral | | The maximum values of hard coral and crustose cover, hard coral species | | 10.41±0.7 (0-86) |

|  |  |  |  |  |  |
|---|---|---|---|---|---|
|  | cover |  | richness and coral:macroalgal ratio from the 5 quadrats on the transect |  |  |
|  | Crustose cover |  |  |  | 6.18±0.6 (0-70) |
|  | Hard coral species richness |  |  |  | 7±0.2 (0-18) |
|  | Coral : Macroalgal ratio |  |  |  | 1.36±0.5 (-1-111) |
|  | Total maximum holes |  | Sum of the highest number of small and big holes from the 5 quadrats |  | 13±0.8 (0-70) |
|  | Maximum rugosity |  | Rugosity index measured with the chain tape method, by placing a 6-m chain at two randomly selected start positions ensuring no overlap along 25-m belt transect. The chain was placed such that it follows the relief along centerline of the belt transect. Two divers measured the straight-line horizontal distance covered by the chain |  | 0.28±0.008 (0.025-0.71) |
| **Seascape composition** | High hard complexity | 2009 NOAA Biogeography Branch Shallow Water Benthic Habitat map | Quantified at a range of spatial scales (10m, 50m, 100m, 150m, 200m, 250m radii) using the Diversity Calculator tool in ArcGIS. Detailed geomorphological structures such as aggregate reef, aggregated patch reef, individual patch reef, Pavement, Pavement with sand channels, Spur and Groove were considered as having a high hard structural complexity | 100 m | 73±1.5 (0-100) |
|  | Percentage seagrass cover |  | The major cover of interest was quantified at a range of spatial scales (10m, 50m, 100m, 150m, 200m, 250m radii) using the Diversity Calculator tool in ArcGIS |  | 2.8±0.5 (0-61) |

| | | | | | |
|---|---|---|---|---|---|
| | Habitat type | 2001 NOAA Biogeography Branch Shallow Water Benthic Habitat map | Nine hard-bottom habitats (aggregated reef, aggregate patch reef, individual patch reef, pavement, pavement with sand channels, reef rubble, rock outcrop, sand, sand and scattered coral and rock, spur and groove) | 300 m | |
| **Seascape complexity** | Bathymetry mean (10m, 50m, 100m, 150m, 200m, 250m radii) | Airborne hydrographic LIDAR (Light Detection and Ranging) | Quantified at a range of spatial scales (10m, 50m, 100m, 150m, 200m, 250m radii) using Focal statistics within the Spatial analyst tool | 3 m | 11±0.4 (0.18-27.6) |
| | Maximum Slope-of-the-slope (10m, 50m, 100m, 150m, 200m, 250m radii) | | Slope-of-the-slope was calculated by creating a slope surface from the bathymetry and then calculating slope of the initial slope surface. Quantified at a range of spatial scales (10m, 50m, 100m, 150m, 200m, 250m radii) using focal statistics within the Spatial analyst tool | | 41.75±1 (6.12-82.2) |
| | Bathymetry classes | | The continuous variables were extracted to points using ArcGIS Spatial Analyst tool and then classified into 5 and 7 categories using ArcGIS Reclass tool | | |
| | Slope-of-the-slope classes | | | | |
| **Wave** | Wave exposure | Simulating WAve | Quantified from points predicted using the CariCOOS (Caribbean Coastal Ocean Observing System) | 240 m | 1.12±0.04 (0.1-3.2) |

| | | | | | |
|---|---|---|---|---|---|
| | Wave exposure classes | Nearshore SWAN model | The continuous variable was extracted to points using ArcGIS Spatial Analyst tool and then classified into 7 categories using ArcGIS Reclass tool | | |
| **Patch proximity** | Distance to mangrove and seagrass | Airborne hydrographic LIDAR (Light Detection and Ranging) | To quantify patch proximity between fish survey sites on coral reefs and the nearest seagrass and mangrove patches a raster of patch proximity was created using Euclidean distance across the insular shelf. | 3 m | Mangrove: 1399±69 (63-3826)<br>Seagrass: 258±16 (0-1230) |
| | Distance to mangrove and seagrass classes | | The continuous variable was extracted to points using ArcGIS Spatial Analyst tool and then classified into 4 and 5 categories using ArcGIS Reclass tool | | |

Table 2 Descriptions of explanatory variables.

*Modeling algorithms*. Machine learning algorithms using single classification and regression trees (CART$^{TM}$ Breiman *et al.* 1984) and stochastic gradient boosted classification and regression trees (TreeNet$^{TM}$) (Friedman 2003) were applied to determine variable interactions and measure variable importance in models of both fish species richness and taxonomic diversity. These nonparametric techniques are more suitable than conventional linear models for exploring complex data that may have multiple structures rather than a single dominant structure (Elith *et al*. 2006, Hastie *et al.* 2009). Using the same explanatory variable in different parts of the tree, these machine learning algorithms deal effectively with nonlinear relationships and higher order interactions that are expected in large complex and multi-scale ecological datasets. Combining CART and TreeNet in exploratory analyses provides high interpretability through the simplicity of CART together with the ability to model higher order interactions using the full suite of variables with TreeNet. In addition, both continuous and categorical data can be incorporated in these models. All models were fitted in Salford Predictive Modeler® software suite (Salford Systems http://www.salford-systems.com/).

*CART*

Fish species richness and taxonomic diversity were analyzed as both a continuous and a categorical variable. The response was transformed into a binary categorical variable with even split (n=251) (high species richness> 24 and low species richness <24) and into fish species richness hotspots defined as the top 10 percent of fish species richness (Ceballos & Ehrlich 2006). CART models were implemented using a standard error rule and a minimum cost tree. The splitting criterion used the Gini index for classification and least squares for regression trees. Trees were generated using a 10-fold cross-validation for testing. The minimum node sample size was set at 3. The optimal tree was the smallest tree with the lowest error. The model starts with the largest tree fitting all the data and then prunes it until it reaches a balance between the smallest numbers of nodes and the smallest error.

*TreeNet$^{TM}$ Stochastic Gradient Boosting*

TreeNet (Salford Systems Inc.) is a machine learning algorithm using ensembles of many simple small least squares regression trees or classification trees that are combined through averaging to give improved estimation accuracy (Friedman 2003; Elith *et al.* 2008). Boosted regression trees have been demonstrated to outperform many commonly used algorithms (i.e. Generalised Linear and Additive Models), for predictive modeling (Elith *et al.* 2006; Friedman 2003; Pearson 2015). Boosted trees can model multiple interactions between predictors and are robust to irrelevant predictors and overfitting.

Like CART models, trees were generated using 10-fold cross-validation. A very slow learning rate (lr) was set (0.001) and the tree complexity (tc) was 6 nodes and a maximum of 10,000 trees allowed.

### *Variable importance*

The relative importance to the optimal model of each variable was estimated from its ranked contribution and the loss of performance when removed from the model. For CART models only, primary splitters were reported here. Surrogates were not taken into account because our data did not have a pattern of missing data. The percentage contribution of each predictor variable was based on the number of times a variable was selected for splitting, weighted by the squared improvement to the model as a result of each split and averaged over all trees (Freidman & Meulman 2003). The relative contribution of each environmental variable on the response was measured so that the sum adds to 100, with a higher number indicating stronger influence on the response (Elith *et al.* 2008). Using partial-dependence plots, the effect of a variable on the response after accounting for the average effects of all other variables in the model is represented (Elith *et al.* 2008). The interactions between predictors were also evaluated using a function testing each possible pair of predictors.

### *Model performance*

The predictive performance of the final models was evaluated using the coefficient of determination ($R^2$) for regression analysis and the Area-Under-the-Curve (AUC) of the Receiver Operating Characteristic curve (ROC) for classification trees (Muñoz & Felicísimo 2004). AUC ranges between 0 and 1 with higher values indicating a better performance. An AUC value of 0.7-0.8 is considered an acceptable prediction; 0.8-0.9 is excellent and >0.9 is outstanding. A value of 0.5 is defined as the predictive ability that could be obtained by chance (Hosmer *et al.* 2000).

### *Statistical analysis*

Data did not meet the assumptions for parametric ANOVA, therefore we applied a non-parametric Kruskal-Wallis test to examine differences between categories of variables and spatial scales in their ability to explain the spatial patterns in fish species richness and taxonomic diversity. A multiple pairwise comparison was computed using Dunn's test (Dunn 1964). Spearman correlations were used to examine the relationship between fish species richness and different categories of variables using the percentage variable importance scores from the predictive models. The mean species richness per mapped habitat type was also investigated to determine differences that may be attributed to the type of habitat rather than its actual physical characteristics. The tests of statistical difference were computed using GraphPAD Prism (Version 6.05).

## Results

### Spatial autocorrelation

Fish species richness data exhibited spatial autocorrelation (Moran's I index: 0.005, p=0.0023) meaning that fish richness were spatially clustered. However, analysis of spatial autocorrelation in the model residuals from CART and TreeNet models were not significant (Moran's I index CART: 0.01, p>0.1, Moran's I index TreeNet: -0.008, p>0.1) indicating that the model errors were not significantly affected by the spatial autocorrelation in the response data and will not bias coefficient estimates (Kühn & Dormann 2012) (Figure 5).

### Model performance

Both CART and TreeNet models showed higher performance when the response variable of fish species richness was classified into richness hotspots (top 10 percent) (AUCCART=0.8, AUCTreeNet=0.77). Predictive performance of richness hotspots was higher for the CART models than for TreeNet models (Overall correct CART= 80.2%; overall correct TreeNet= 71.7%), with low misclassification rates (missclassCART=6%, missclassTreeNet=9% respectively). The primary splitters selected by the CART model were live hard coral cover from quadrats, modeled wave exposure and topographic complexity measured by slope-of-the-slope within the 25 m radius sample unit (1,963.5$m^2$), 50 m (7854 $m^2$) and 150 m (70,686 $m^2$) radius (Figure 6). Models for fish species richness performed better than models for taxonomic diversity (Table 3) and therefore we have focused our results and discussion on the best performing models for fish species richness.

| Responses | Species richness | Taxonomic | Richness hotspots |
|---|---|---|---|

| | | | **diversity** | **(Top 10%)** |
|---|---|---|---|---|
| **Model Performance** | | R² | | AUC |
| **TreeNet** | | 0.24 | 0.15 | 0.77 |
| **CART** | | 0.22 | 0.12 | 0.8 |
| | | Percentage contribution | | |
| **CART** | Live hard coral cover | 100 | 100 | 100 |
| **Primary splitters** (Figure 5) | Slope-of-the-slope (25m,50m,150m radius) | 50 | 41 | 50 |

*Table 3: TreeNet and CART model performance (R² and AUC) and primary splitters contribution to the CART model. Species richness and taxonomic diversity are described as continuous data while richness hotspots (Top 10%) is classified as categorical data.*

*Hypothesis testing*

1. Does within-patch structure explain more of the variability in fish species richness and taxonomic diversity than the surrounding seascape structure?

When comparing the strength of correlations for groups of environmental variables, fish species richness displayed a stronger positive correlation with within-patch variables (Rho= 0.25, p=0.0005) than with patch proximity (Rho= -0.02, p=0.0139) and seascape composition (Rho= -0.03, p=0.0002) (Figure 7a), but showed a similar strength of association as estimated for seascape terrain complexity Overall, seascape terrain complexity variables yielded a significantly (p=0.0006) stronger association with fish species richness than did seascape composition variables.

In addition to a simple correlation coefficient, the mean percentage contribution of groups of environmental variables as calculated by the TreeNet and CART models, was tested to determine if any variable group contributed more than any other (Figure 7b). The between-group contributions to the TreeNet models were significantly (p=0.0057) different. Following similar relative patterns as the correlations, pairwise tests revealed that within-patch variables contributed significantly more (p=0.0241) to the model than seascape composition variables and wave exposure, therefore highlighting the importance of within-patch variables for predicting fish species richness.

At the level of individual variables, the most important within-patch environmental predictor was live hard coral cover, which displayed a significant (Rho= 0.45, p=8.604E$^{-13}$) positive correlation with fish species richness. In contrast, a weak negative correlation existed between live hard coral cover and taxonomic diversity (Rho=-0.26, p=2.065E$^{-5}$). In fact, taxonomic diversity was more weakly correlated with all environmental variables than species richness. Comparing ranked variable importance scores from both CART and TreeNet models revealed that out of the 41 predictors the most influential single variable for richness hotspots was the amount of live hard coral cover (AUC CART= 0.8, AUC TreeNet= 0.77) followed by slope-of-the-slope at 25 m, 50 m and 150 m radii. In the best CART model, highest fish richness was predicted for coral reefs with coral cover greater than 8.13 % and a topographic complexity (slope-of-the-slope) lower than 53.3 degrees. Examination of partial-dependence plots (Figure 8) show a sharp increase in partial dependence between fish species richness and the amount of live hard coral cover above approximately 8 % tapering off at approximately 25 % cover. For slope-of-the-slope, partial dependence with fish species richness increases sharply from 56 degrees to approximately 60 degrees.

To examine the utility of benthic habitat type as a spatial proxy for fish species richness we grouped fish samples by mapped benthic habitat types and tested for differences which revealed that mean fish species richness was significantly (p=0.0049) different between habitat types. More specifically, mean fish richness was significantly (p=0.0034) higher over aggregated patch reef than pavement, colonized pavement and linear reef (Figure 9).

2. Which spatial scale(s) of seascape patterning best explained fish species richness and taxonomic diversity?

The most influential single spatial scale for seascape patterning was the 25 m radius seascape equivalent to a sample unit area of 1,963.5 m² (Figure 10). The mean contribution to the two models was significantly (p<0.05) different between scales. Topographic complexity quantified using slope-of-the-slope within the 25 m radius seascape sample unit was the most important predictor for fish richness in the TreeNet model (Figure 8) and was a primary splitter in the CART model (Figure 6). The three highest mean percentage contributions were identified with the finer scale seascape sample units of 10 to 50 m radii.

Although fish species richness was more strongly correlated (Rho= 0.37, p=6.223E$^{-10}$) with slope-of-the-slope at the 25 m radius scale, this scale of correlation was only statistically higher than the broadest slope-of-the-slope at 250 m radius (Rho= 0.10, p= 0.1) (Figure 11).

3. Which interacting environmental variables best characterize fish richness hotspots?

The optimal CART model predicted highest fish richness for sites characterized by live hard coral cover greater than 8.13 %. Where live hard coral cover was less than 8.13 %, the highest fish richness was found on hard bottom areas where slope-of-the-slope was greater than 53.3 degrees. Lowest fish richness was found on the least topographically complex sites, with lowest live hard coral cover.

The TreeNet model for fish richness hotspots revealed that interactions between several variables were important in the models (Figure 12 & 13). When examining only the mapping variables, the interaction between slope-of-the-slope at the spatial scale of 25 m and habitat type play an important role in explaining the richness hotspots (Figure 12). The optimal TreeNet model, which allowed multiple interactions between variables, predicted that highest fish richness was found on aggregated patch reefs with highest slope of the slope values (Figure 14).

## Discussion

Identifying the key environmental variables that explain spatial patterns of species richness is an ongoing challenge in ecology and can provide useful information to guide spatial planning and prioritize management actions. Rarely, however, do studies of fish assemblages consider the structure of the seascape surrounding the sampled reef. Our multi-scale seascape ecology approach examined the relative influence of within-patch structural attributes, as well as the surrounding seascape geomorphology, habitat type and area, proximity of complementary habitat types and a key hydrodynamic variable (wave energy) known to influence fish ecology. We found that some of the finest scale measurements (within-patch attributes) explained more of the spatial variation in fish species richness than the surrounding seascape composition (area of habitat types), wave action, and distance to seagrasses and mangroves. However, the best models incorporated interactions between both within-patch and seascape variables, represented by the amount of live scleractinian coral cover measured by SCUBA divers, and the topographic complexity of the seafloor terrain measured with an airborne laser (LiDAR). The results add additional evidence that the most reliable biophysical characteristics for explaining patterns of fish species richness across coral reefs are the amount of live coral and reef structural complexity (Bell & Galzin 1984, Gratwicke & Speight 2005, Pittman *et al*. 2009, Graham & Nash 2013, Coker *et al*. 2014).

For coral reef conservation, these results highlight the importance of prioritizing actions to focus on enhancing and protecting the amount of live coral and the topographic complexity of reefs, in order to conserve functionally diverse coral reef communities. For future attempts at predictive modeling, our results indicate that higher model performance is likely to be attained by integration of maps of live hard coral cover together with high-resolution bathymetry of the seafloor.

*Model performance*

CART modeling provided a clear model in a single small tree, while TreeNet models allowed interactions among multiple variables in a complex solution combining many hundreds of small trees. Therefore, the simultaneous use of these two algorithms provides the advantage of maintaining the simplicity of interpretation with a CART model, while benefiting from the flexibility provided by ensembles of trees with TreeNet (Fahrnkopf 2015). The CART model is typically used as an exploratory precursor to the more powerful boosted regression trees, but in this study, it was shown that the predictive performance of the CART model was greater when applied to address the question of predicting richness hotspots (AUCCART= 0.8) than with the continuous variable of species richness ($R^2$CART= 0.22). With regard to model performance, TreeNet provided no advantage over CART for modeling the binary classification of presence and absence of a richness hotspot (AUCCART= 0.8, AUCTreeNet= 0.77). Not only did the CART model provide good performance in predicting sites with the highest fish richness, it also provided a numerical description of the primary environmental variables that make a coral reef suitable habitat for maintaining high biodiversity. In addition, the breakpoints on the environmental variables that determined the groupings of fish survey sites have potential to help identify thresholds, or ecological tipping points, beyond which species richness abruptly declines or increases (Lintz *et al*. 2011).

*Which habitat variables and scales most influence fish diversity?*

Our multi-scale analyses demonstrated that within-patch variables (live hard coral cover) and remotely sensed topographic complexity (slope-of-the-slope) contributed to the best models of fish species richness. A high number of fish species (mean species richness 26.8 per 100 $m^2$) was predicted for reefs with live hard coral cover greater than 8.13 %. In our study region, approximately 42 % of reef sites had at least one quadrat (1 $m^2$) with live hard coral cover greater than 8.13 %. This is relatively low when compared with other regions of the Caribbean and reflects a substantial decline since the 1980s (Gardner *et al*. 2003). It has been widely recognized that live hard coral cover has a strong positive correlation with fish species richness (Carpentier 1981, Bell & Galzin 1984) because of coral's provisioning of food, settlement substratum and shelter for a wide range of fishes (Wilson *et al*. 2006, Coker *et al*. 2014).. In fact, studies on coral reefs have shown that even a very small increase in live hard coral cover (< 2 %) can result in significant increases in the total number of fish species (Bell & Galzin 1984). Conversely, however, declines in live hard coral result in declines in fish abundance and species richness (Wilson *et al*. 2006). Identifying and modeling the link between the amount of live hard coral and the biodiversity of coral reefs is particularly significant because of the documented recent declines in live coral cover due to a wide range of stressors (Alvarez-Filip *et al*. 2011). This presents a technical challenge in applied predictive modelling, whereby the difficulty in deriving reliable maps of live coral cover currently hinders efforts to develop spatially continuous predictors for mapping coral-associated biological distributions. Further experimental investigations using remote sensing data are urgently required to identify high spatial resolution data suitable for mapping either directly, or through proxies, fine-scale environmental variables, such as live coral cover and topographic complexity (Leiper *et al*. 2014, Hedley *et al*. 2016).

At broader spatial scales (25 m radius) surrounding fish survey sites, topographic complexity measured using slope-of-the-slope (2nd derivative of bathymetry), served as a good predictor of fish species richness. The utility of spatial metrics of terrain topographic complexity are increasingly being demonstrated through spatial modeling studies for predicting distributions in a wide range of reef organisms and communities from both tropical and temperate environments (Pittman *et al.* 2009, Pittman & Brown 2011, Cameron *et al*. 2014, Young & Carr 2015). Slope and curvature of the seafloor terrain functions as a proxy for biogenic structural complexity and also influences current flow (Mohn & Beckmann 2002), which potentially increase food supply for benthic species (Wilson *et al.* 2007). Around the island of St. John, USVI, reef edges are associated with high topographic complexity (Figure 14), a biogeomorphological pattern that is also associated with high coral cover and fish species richness. The ecological significance of topographic complexity as a key contributor to the geographical distribution of fish richness highlights serious ecological consequences for the long-term capacity of reefs to support high diversity given the recent widespread declines in the topographic complexity of Caribbean reefs (Alvarez-Filip *et al.* 2009; Pittman *et al.* 2010; Rogers *et al*. 2014).

In general, seascape predictors played an important contribution only at relatively fine spatial scales across the scale range. For instance, fish species richness was more strongly correlated with slope-of-the-slope at the 25 m radius scale than slope-of-the-slope at 250 m radius. Our review of fish home range areas confirms that the 25 m radius seascape sample unit encompassed many of the home range areas reported for common Caribbean fish species present in the study area. It is plausible that many of the daily ecological processes for these common species occurred within this space (i.e., the ecological neighborhood [sensu Addicott *et al*. 1989]).

However, coral reefs are highly heterogeneous in time and space and not all coral reefs sites surveyed hosted the same amount of fish species even if they did exhibit similar amounts of live hard coral cover and topographic complexity. Fish assemblages comprise very complex interactions between physical, chemical and biological processes (Longmore 2014) that challenge modeling efforts despite including multiple types of environmental variables at multiple spatial scales. Several of our environmental variables made only insignificant contributions to the best models. For instance, patch proximity metrics representing distance from reef to nearest seagrasses and mangroves made a relatively minor contribution to models of fish richness and diversity. This is despite the fact that many species on coral reefs utilize a mosaic of patch types through the life history. Connectivity between coral reefs, seagrasses and mangroves has been shown to influence the structure and function of coral reef fish assemblages where the spatial arrangement of patches and particularly the proximity of patch types influence the strength of interactions (Mumby *et al*. 2004, Dorenbosch *et al*. 2004, Pittman *et al*. 2007, Olds *et al*. 2012, Nagelkerken *et al.* 2015). The inclusion of a wide range of reef sites in this study together with assemblages including many species not closely associated with seagrasses and mangroves likely explains the weak influence of proximity. Alternatively, inclusion of patch size attributes together with patch proximity could be examined i.e., close proximity to large area of seagrass may influence species richness on coral reefs more than equal proximity to only very small patches. Indices that integrate seascape metrics in a functionally meaningful way should be examined in future studies.

Furthermore, classifying highly heterogeneous continuous environmental variables (depth, slope-of-the-slope, wave exposure, distance to patch) into several discrete classes of values to simplify variability did not enhance their contribution to model performance compared with using the original continuous variables. However, we did find that the benthic habitat map classes performed well when allowed to interact with topographic complexity quantified by slope-of-the-slope. While cost effective and widely available for relatively broad geographical areas, remotely sensed data may not capture sufficient ecological variability to explain the complex patterns of biological distributions for coral reef fishes. Additional types of environmental variables could also be explored such as diver defined habitat classes, higher resolution terrain models and outputs from connectivity models (Yates *et al*. 2016).

### *Limitations of the study and future research*

Although live hard coral cover exhibited the highest influence on fish species richness and taxonomic diversity, the strength of the bivariate correlation between the amount of live coral and fish response was moderate (Rho Fish richness= 0.45, p=8.604E$^{-13}$) and negatively weak (Rho Taxonomic diversity= -0.26, p=2.065E$^{-5}$). In addition to the inevitability of missing variables, several other factors related to the fish data collection could result in a weaker than expected association with structural variables, including temporal variability in the fish richness data which was collected over a period of 10 years. All fish visual censuses took place during daylight hours and therefore we were unable to account for diel cycles and surveys were biased to non-cryptic species. Little is known about the effectiveness of daytime visual surveys as surrogates for multi-phyla diversity patterns, but where destructive sampling has attempted to census the complete fish assemblage on Caribbean coral reefs it is clear that daytime visual census is only reporting on a moderate and visually distinctive proportion of the fishes present (Smith-Vaniz *et al*. 2006; Harborne *et al.* 2012). For example, at Buck Island Reef National Monument in St. Croix, U.S. Virgin Islands, when visual census data and rotenone samples of fishes were compared only 36 % of the 228 species sampled with rotenone were detected through visual census (Smith-Vaniz *et al*. 2006).

One key limitation influencing the application of our models for mapping fish species richness across the study area is that the detailed within-patch variables measured i.e. live coral cover are not available as spatially continuous data. Rarely are benthic maps produced which accurately represent the distribution and amount of live coral, although attempts have been made and methods are continually evolving to tackle the challenge (Mumby *et al.* 2004; Joyce *et al.* 2013). However, the results from the study show that live hard coral cover and slope-of-the-slope across spatial scales are positively correlated (Rho=0.5, p=1.87E$^{-11}$) suggesting that topographic complexity alone may provide a useful spatial proxy map with which to predict fish species richness as has been demonstrated in southwestern Puerto Rico (Pittman & Brown 2011).

At broader spatial extents, questions arise regarding the possible homogenization of the seascape around St. John, U.S. Virgin Islands and the wider Caribbean. Phase shifts to algal dominated reefs, declines in structural complexity *i.e*. reef flattening, and loss of faunal diversity result in declining structural and functional heterogeneity. Sites will become less different in community composition and diversity, metrics less variable and correlations decoupled. As reefs degrade, it is expected that fish communities will have fewer specialist species and a greater proportion of generalist species (Alvarez-Filip *et al.* 2015). In the last ten years, fish abundance has declined through the entire Caribbean region (Paddack *et al.* 2009), particularly large bodied fishes (Stallings *et al.* 2009). Hurricanes, ocean acidification, declining water quality and physical damage from fishing gears are some of the main causes of seascape homogenization causing reef flattening and a general decline in architectural complexity (Alvarez-Filip 2009; Alvarez-Filip *et al.* 2011). This phenomenon has a negative impact on animal diversity (Smokorowski & Pratt 2007; Seiferling *et al.* 2014). Massicotte *et al.* (2015) found a strong positive relationship between environmental heterogeneity and fish species richness. Further studies should address the ecological consequences of seascape homogenization with a particular focus on the possible alteration of habitat function and the implications for understanding seascape biodiversity patterns.

## Acknowledgements

We offer our gratitude to the many NOAA data providers including scientific divers and remote sensing analysts at the National Centers for Coastal and Ocean Science' Biogeography Branch. We are also grateful to the Marine Institute of Plymouth University for providing research facilities to L. Sekund during a visiting fellowship as part of the International Master of Science in Marine Biodiversity and Conservation. S. Pittman was funded by NOAA Coral Reef Conservation Program.

## References

Alvarez-Filip L., Dulvy N.K., Gill J.A., Côté I.M., Watkinson A.R. (2009) Flattening of Caribbean coral reefs: regionwide declines in architectural complexity. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 3019-3025.

Alvarez-Filip L., Gill J.A., Dulvy N.K., Perry A.L., Watkinson A.R., Côté I.M. (2011) Drivers of region-wide declines in architectural complexity on Caribbean reefs. *Coral Reefs*, **30**, 1051-1060.

Alvarez-Filip L., Paddack M.J., Collen B., Robertson D.R., Côté I.M. (2015) Simplification of Caribbean reef-fish assemblages over decades of coral reef degradation. *PloS One*, **10**, e0126004.

Bell J., Galzin R. (1984) Influence of live coral cover on coral-reef fish communities. *Marine Ecology Progress Series*, **15**, 265-274.

Betts M.G., Fahrig L., Hadley A.S., Halstead K.E., Bowman J., Robinson W.D., Lindenmayer D.B. (2014) A species centered approach for uncovering generalities in organism responses to habitat loss and fragmentation. *Ecography*, **37**, 517-527.

Booij N., Ris R., Holthuijsen L.H. (1999) A third-generation wave model for coastal regions: 1. Model description and validation. *Journal of Geophysical Research: Oceans (1978–2012)*, **104**, 7649-7666.

Breiman L., Friedman J., Olshen R., Stone C. (1984) *Classification and Regression Trees.* Pacific Grove: Wadsworth, 1984.

Brown C.J., Harborne A.R., Paris C.B., Mumby P.J. (2016) Uniting paradigms of connectivity in marine ecology. *Ecology*, **97**(9), 2447-2457.

Carpenter K., Miclat R., Albadalejo V., Corpuz V. (1981) The influence of substrate structure on the local abundance and diversity of Philippine Reef Fishes. *Proceedings of the Fourth International Coral Reef Symposium, Manila, Vol.2.*

Catano L.B., Gunn B.K., Kelley M.C., Burkepile D.E. (2015) Predation risk, resource quality, and reef structural complexity shape territoriality in a coral reef herbivore. *PloS One*, **10**, e0118764.

Ceballos G., Ehrlich P.R. (2006) Global mammal distributions, biodiversity hotspots, and conservation. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 19374-19379.

Clark K.R., Warwick R.M. (2001) A further biodiversity index applicable to species lists: variation in taxonomic distinctness. *Marine Ecology Progress Series,* **216**, 265-278 [373].

Coker D.J., Wilson S.K., Pratchett M.S. (2014) Importance of live coral habitat for reef fishes. *Reviews in Fish Biology and Fisheries*, **24**(1), 89-126.

Cushman S., McGarigal K. (2004) Patterns in the species–environment relationship depend on both scale and choice of response variables. *Oikos*, **105**, 117-124.

Dorenbosch M., Van Riel M., Nagelkerken I., Van der Velde G. (2004) The relationship of reef fish densities to the proximity of mangrove and seagrass nurseries. *Estuarine, Coastal and Shelf Science*, **60**, 37-48.

Duffy J.E., Lefcheck J.S., Stuart-Smith R.D., Navarrete S.A., Edgar G.J. (2016) Biodiversity enhances reef fish biomass and resistance to climate change. *Proceedings of the National Academy of Sciences*, **113**(22), 6230-6235.

Dunn O.J. (1964) Multiple comparisons using rank sums. *Technometrics*, **6**, 241–252.

Elith J., Graham C.H., Anderson R.P., Dudík M., Ferrier S., Guisan A., Hijmans R.J., Huettmann F., Leathwick J.R., Lehmann A., Li J., Lohmann L., Loiselle B.A (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129-151.

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

Fahrnkopf K. (2015) TreeNet Gradient Boosting and CART Decision Trees: A Winning Combination. [online]. Available from http://info.salford-systems.com/cart-and-treenet-a-winning-combination.

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

Friedlander A.M., Jeffrey C., Hile S., Pittman S.J., Monaco M., Caldow C. (2013) Coral reef ecosystems of St. John, US Virgin Islands: Spatial and temporal patterns in fish and benthic communities (2001-2009). NOAA Technical Memorandum 152. Silver Spring, MD. 150 pp

Fulton C.J., Bellwood D.R. (2005) Wave-induced water motion and the functional implications for coral reef fish assemblages. *Limnology and Oceanography*, **50**(1), 255-264.

Gardner T.A., Côté I.M., Gill J.A., Grant A., Watkinson A.R. (2003) Long-term region-wide declines in Caribbean corals. *Science*, **301**(5635), 958-960.

Graham N.A.J., Nash K.L. (2013) The importance of structural complexity in coral reef ecosystems. *Coral Reefs*, **32**, 315-326.

Gratwicke B., Speight M.R. (2005) The relationship between fish species richness, abundance and habitat complexity in a range of shallow tropical marine habitats. *Journal of Fish Biology*, **66**, 650-667.

Harborne A.R., Mumby P.J., Micheli F., Perry C.T., Dahlgren C.P., Holmes K.E., Brumbaugh D.R. (2006) The functional value of Caribbean coral reef, seagrass and mangrove habitats to ecosystem processes. *Advances in Marine Biology*, **50**, 57-189.

Harborne A.R., Jelks H.L., Smith-Vaniz W.F., Rocha L.A. (2012) Abiotic and biotic controls of cryptobenthic fish assemblages across a Caribbean seascape. *Coral Reefs*, **31**, 977-990.

Hastie T., Tibshirani R., Friedman J. (2009) Unsupervised learning (pp. 485-585). Springer New York

Hedley J.D., Roelfsema C.M., Chollett I., Harborne A.R., Heron S.F., Weeks S., Skirving W.J., Strong A.E., Eakin C.M., Christensen T.R., Ticzon V. (2016) Remote sensing of coral reefs for monitoring and management: a review. *Remote Sensing*, **8**(2), 118.

Hitt S., Pittman S.J., Nemeth R.S (2011) Diel movements of fishes linked to benthic seascape structure in a Caribbean coral reef ecosystem. *Marine Ecology Progress Series*, **427**, 275-291.

Hoffman S.G. (1983) Sex-related foraging behaviour in sequentially hermaphroditic hagfishes (*Bodianus* spp.). *Ecology*, **64**(4), 798-808.

Holling C.S. (1994) Cross-scale morphology, geometry, and dynamics of ecosystems. *In Ecosystem Management*, 351-423. Springer New York.

Holmlund C.M., Hammer M. (1999) Ecosystem services generated by fish populations. *Ecological Economics*, **29**, 253-268.

Hosmer Jr D.W., Lemeshow S., Sturdivant R.X. (2013) *Applied Logistic Regression*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Jackson H. B., Fahrig L. (2015) Are ecologists conducting research at the optimal scale?. *Global Ecology and Biogeography*, *24*(1), 52-63.

Jones K.M.M. (2005) Home range areas and activity centres in six species of Caribbean wrasses (Labridae). *Journal of Fish Biology*, **66**, 150-166.

Joyce K.E., Phinn SR., Roelfsema C.M. (2013) Live coral cover index testing and application with hyperspectral airborne image data. *Remote Sensing,* **5**, 6116-6137.

Kendall M., Monaco M., Buja K., Christensen J., KRUER C., Finkbeiner M., Warner R. (2001) Methods Used to Map the Benthic Habitats of Puerto Rico and the US Virgin Islands. U.S. National Oceanic and Atmospheric Administration, National Ocean Service. Center for Coastal Ocean Science Biogeography Program.

Kendall M.S., Miller T.J., Pittman S.J. (2011) Patterns of scale-dependency and the influence of map resolution on the seascape ecology of reef fish. *Marine Ecology Progress Series*, **427**, 259-274.

Kotliar N.B., Wiens J.A. (1990) Multiple scales of patchiness and patch structure: a hierarchical framework for the study of heterogeneity. *Oikos*, **59**, 253-260.Kühn I., Dormann C.F. (2012) Less than eight (and a half) misconceptions of spatial analysis. *Journal of Biogeography*, **39**(5), 995-998.

Leiper I.A., Phinn S.R., Roelfsema C.M., Joyce K.E., Dekker A.G. (2014) Mapping coral reef benthos, substrates, and bathymetry, using compact airborne spectrographic imager (CASI) data. *Remote Sensing*, **6**(7), 6423-6445.

Longmore A. (2014) Spatial and temporal scales of key ecological processes in marine protected areas. *Fisheries Victoria Science Report Series No. 4* Published by the Victorian Government, Department of Environment and Primary Industries, May 2014.

Massicotte P., Proulx R., Cabana G., Rodríguez M.A. (2015) Testing the influence of environmental heterogeneity on fish species richness in two biogeographic provinces. *PeerJ*, **3**, e760

McCormick M.I. (1994) Comparison of field methods for measuring surface topography and their associations with a tropical reef fish assemblage. *Marine Ecology Progress Series***, 112**, 87-96.

McGarigal K., Tagil S., Cushman S.A. (2009) Surface metrics: an alternative to patch metrics for the quantification of landscape structure. *Landscape Ecology*, **24**, 433-450.

Meentemeyer V. (1989) Geographical perspectives of space, time, and scale. *Landscape Ecology*, **3**, 163-173.

Mohn C., Beckmann A. (2002) Numerical studies on flow amplification at an isolated shelfbreak bank, with application to Porcupine Bank. *Continental Shelf Research*, **22**, 1325-1338.

Muñoz J., Felicísimo Á.M. (2004) Comparison of statistical methods commonly used in predictive modelling. *Journal of Vegetation Science*, **15**, 285-292.

Mumby, P. J., Hedley, J. D., Chisholm, J. R. M., Clark, C. D., Ripley, H., & Jaubert, J. (2004). The cover of living and dead corals from airborne remote sensing. *Coral Reefs*, **23**, 171-183.

Mumby P.J., Edwards A.J., Arias-González J.E., Lindeman K.C., Blackwell P.G., Gall A., Renken H. (2004) Mangroves enhance the biomass of coral reef fish communities in the Caribbean. *Nature*, **427**, 533-536.

Mumby P.J. (2006) Connectivity of reef fish between mangroves and coral reefs: algorithms for the design of marine reserves at seascape scales. *Biological Conservation*, **128**, 215-222.

Mumby P.J., Wabnitz C.C. (2002) Spatial patterns of aggression, territory size, and harem size in five sympatric Caribbean parrotfish species. *Environmental Biology of Fishes*, **63**(3), 265–279.

Nagelkerken, I., Sheaves, M., Baker, R., & Connolly, R. M. (2015) The seascape nursery: a novel spatial approach to identify and manage nurseries for coastal marine fauna. *Fish and Fisheries*, **16**, 362-371.

Nursall J.R. (2009) Territoriality in redlip blennies (Ophioblennius atlanticus-Pisces: Blenniidae). *Journal of Zoology*, **182**, 205-223.

Olds A.D, Connolly R.M., Pitt K.A., Maxwell P.S. (2012) Primacy of seascape connectivity effects in structuring coral reef fish assemblages. *Marine Ecology Progress Series*, **462**, 191-203.

Overholtzer K., Motta P. (1999) Comparative resource use by juvenile parrotfishes in the Florida Keys. *Marine Ecology Progress Series*, **177**, 177-187.

Paddack M.J., Reynolds J.D., Aguilar C., Appeldoorn R.S., Beets J., Burkett E.W., Fonseca A.C. (2009) Recent region-wide declines in Caribbean reef fish abundance. *Current Biology*, **19**, 590-595.

Pearson S.M., Turner M.G., Gardner R.H., O'Neill R.V. (1996) An organism-based perspective of habitat fragmentation. Biodiversity in managed landscapes: theory and practice. Oxford University Press, New York, 77-95.

Pearson R. (2015) Training course on species distribution modelling taught at University College London. [online]. Available from https://www.youtube.com/watch?v=jeOi9r2evIc.

Pittman S., McAlpine C. (2003) Movements of marine fish and decapod crustaceans: process, theory and application. *Advances in Marine Biology*, **44**, 205-294.

Pittman S.J., Christensen J.D., Caldow C., Menza C., Monaco M.E. (2007a) Predictive mapping of fish species richness across shallow-water seascapes in the Caribbean. *Ecological Modelling*, **204**, 9-21.

Pittman S.J., Caldow C., Hile S.D., Monaco M.E. (2007b) Using seascape types to explain the spatial patterns of fish in the mangroves of SW Puerto Rico. *Marine Ecology Progress Series*, **348**, 273-284.

Pittman S.J., Costa B.M., Battista T.A. (2009) Using Lidar bathymetry and boosted regression trees to predict the diversity and abundance of fish and corals. *Journal of Coastal Research*, **53**, 27 - 38.

Pittman S.J., Costa B., Jeffrey C. F. Caldow C. (2010, November) Importance of seascape complexity for resilient fish habitat and sustainable fisheries. In Proceedings of the Gulf and Caribbean Fisheries Institute (Vol. 63, pp. 420-426). Gulf and Caribbean Fisheries Institute, c/o Harbor Branch Oceanographic Institution, Inc. Fort Pierce FL 34946 United States.

Pittman S.J., Brown K.A. (2011) Multi-scale approach for predicting fish species distributions across coral reef seascapes. *PLoS One*, **6**, e20583.

Pittman S.J., Costa B., Wedding L.M. (2013) LiDAR applications. *Coral Reef Remote Sensing*. Springer 145-174.

Pittman S.J., Monaco M.E., Friedlander A.M., Legare B., Nemeth R.S., Kendall M.S., Caldow C. (2014) Fish with chips: tracking reef fish movements to evaluate size and connectivity of Caribbean marine protected areas. *PLoS One*, **9**, e96028.

Pittman S.J., Bauer L., Hile S.D., Jeffrey C.F.G., Davenport E., Caldow C. (2014) Marine protected Areas of the U.S. Virgin Islands: Ecological Performance Report. NOAA Technical Memorandum NOS NCCOS 187. Silver Spring, MD. 89 pp.

Pittman S.J., Knudby A. (2014) Spatial predictive modeling of coral reef fish species and assemblages. In Bortone SA, Mueller M: Interrelationships between coral reefs and fisheries. CRC Press, 31 pages.

Pittman S.J., Olds A. (2015) Ecology of fishes on coral reefs: Seascape ecology of fishes on coral reefs. In C Mora (ed.) Ecology of Fishes on Coral Reefs. Cambridge University Press, pp.274-282.

Price B., Kutt A.S., McAlpine C.A., Phinn S.R., Pullar D.V. et al. (2009) Continuum or discrete patch landscape models for savanna birds? Towards a pluralistic approach. *Ecography*, **32**, 745–756.

Raymundo L.J., Halford A.R., Maypa A.P., Kerr A.M. (2009) Functionally diverse reef-fish communities ameliorate coral disease. *Proceedings of the National Academy of Sciences* **106**:17067-17070.

Roberts C.M., McClean C.J., Veron J.E., Hawkins J.P., Allen G.R., McAllister D.E., Mittermeier C.G., Schueler F.W., Spalding M., Wells F., Vynne C. (2002) Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science*, **295**(5558) 1280-1284.

Rogers A., Blanchard J.L., Mumby P.J. (2014) Vulnerability of coral reef fisheries to a loss of structural complexity. *Current Biology*, **24**, 1000-1005.

Rogers C. S., Beets J. (2001) Degradation of marine ecosystems and decline of fishery resources in marine protected areas in the US Virgin Islands. *Environmental Conservation*, **28**(04), 312-322.

Rogers C.S. (2013) Coral reef resilience through biodiversity. *ISRN Oceanography*, **2013.**

Salmon J.E., Holthuijsen L.H. (2015) Modeling depth-induced wave breaking over complex coastal bathymetries. *Coastal Engineering*, **105**, 21-35.

Schneider D.C. (2001) The Rise of the Concept of Scale in Ecology: The concept of scale is evolving from verbal expression to quantitative expression. *BioScience*, **51**, 545-553.

Seiferling I., Proulx R.L., Wirth C. (2014) Disentangling the environmental-heterogeneity-species diversity relationship along a gradient of human footprint. *Ecology*, **95**, 2084-2095.

Shapiro D.Y., Garcia-Moliner G., Sadovy Y. (1994) Social system of an inshore stock of the red hind grouper, *Epinephelus guttatus* (Pisces: Serranidae). *Environmental Biology of Fishes*, **41**, 415-422.

Sikkel P.C. (1990) Social organization and spawning in the Atlantic sharpnose puffer, *Canthigaster rostrata* (Tetraodontidae). *Environmental Biology of Fishes*, **27**, 243-254.

Smith-Vaniz W.F., Jelks H.L., Rocha L.A. (2006) Relevance of cryptic fishes in biodiversity assessments: a case study at Buck Island Reef National Monument, St. Croix. *Bulletin of Marine Science*, **79**(1), 17-48.

Smokorowski K., Pratt T. (2007) Effect of a change in physical structure and cover on fish and fish habitat in freshwater ecosystems-a review and meta-analysis. *Environmental Reviews*, **15**, 15-41.

Stallings C.D. (2009) Fishery-independent data reveal negative effect of human population density on Caribbean predatory fish communities. *PLoS One*, **4**, e5333.

Van Beukering, P. Brander, L., Van Zanten, B., Verbrugge, E., Lems, K. (2011) The Economic Value of the Coral Reef Ecosystems of the United States Virgin Islands. IVM Report number: R-11/06.

Warwick R., Clarke K. (1995) New 'biodiversity' measures reveal a decrease in taxonomic distinctness with increasing stress. *Marine Ecology Progress Series*, **129**, 301-305.

Wiens J.A. (1989) Spatial scaling in ecology. *Functional Ecology*, **3**, 385-397.

Wiens J.A., Milne B.T. (1989) Scaling of ’landscapes’ in landscape ecology, or, landscape ecology from a beetle’s perspective. *Landscape Ecology*, **3,** 87-96.

Wilson S.K., Graham N.A.J., Pratchett M.S., Jones G.P., Polunin N.V.C. (2006) Multiple disturbances and the global degradation of coral reefs: are reef fishes at risk or resilient? Global Change Biology **12**(11):2220-2234.

Wilson M.F.J., O'Connell B., Brown C., Guinan J.C., Grehan A.J. (2007) Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. *Marine Geodesy*, **30**(1-2), 3-35.

Wolman A.G. (2006) Measurement and meaningfulness in conservation science. *Conservation Biology*, **20**, 1626-1634.

Yates K.L., Mellin C., Caley M.J., Radford B.T., Meeuwig J.J. (2016) Models of marine fish biodiversity: Assessing predictors from three habitat classification schemes. *PLoS One*,**11**(6), e0155634.

Young M., Carr M.H. (2015) Application of species distribution models to explain and predict the distribution, abundance and assemblage structure of nearshore temperate reef fishes. *Diversity and Distributions*, **21**(12) 1428-1440.

Zitello A.G., Bauer L.J., Battista T.A., Mueller P.W., Kendall M.S., Monaco M.E. (2009) Shallow-Water Benthic Habitats of St. John, U.S. Virgin Islands. NOAA Technical Memorandum NOS NCCOS 96. Silver Spring, MD. 53 pp.

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

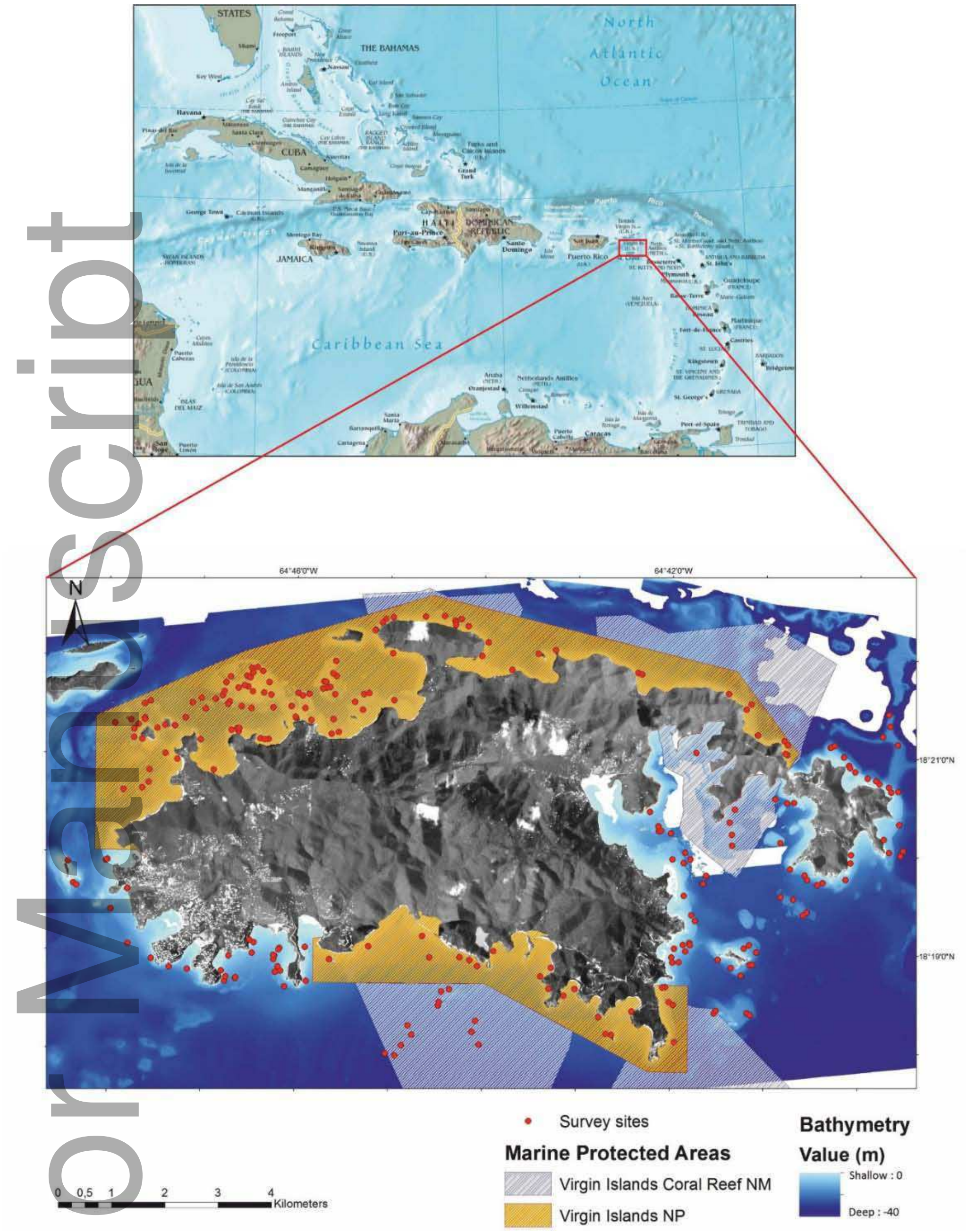


*Figure 1: Locations of survey points and marine protected areas around St. John, USVI. Source: NOAA Biogeography Branch.*

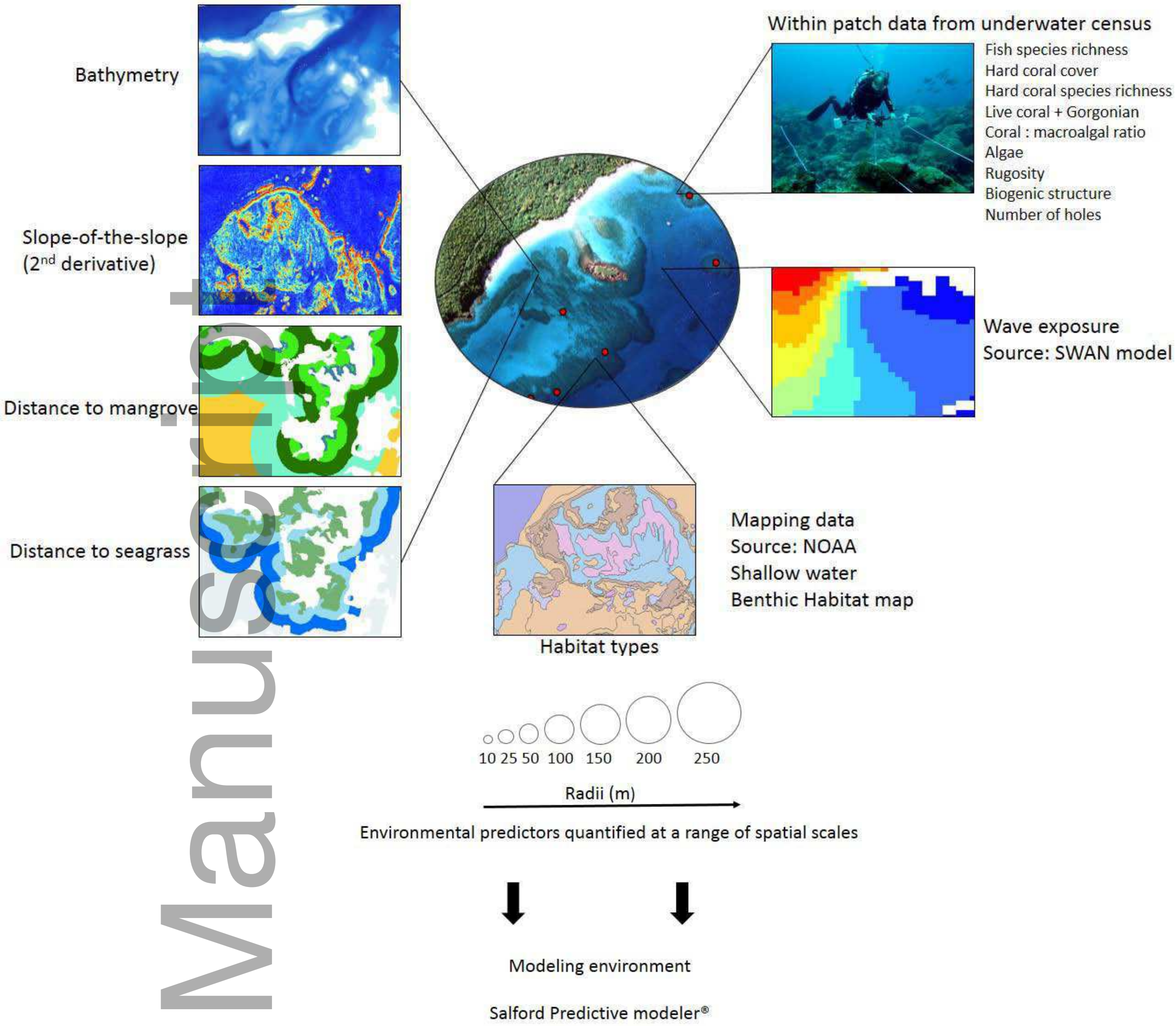
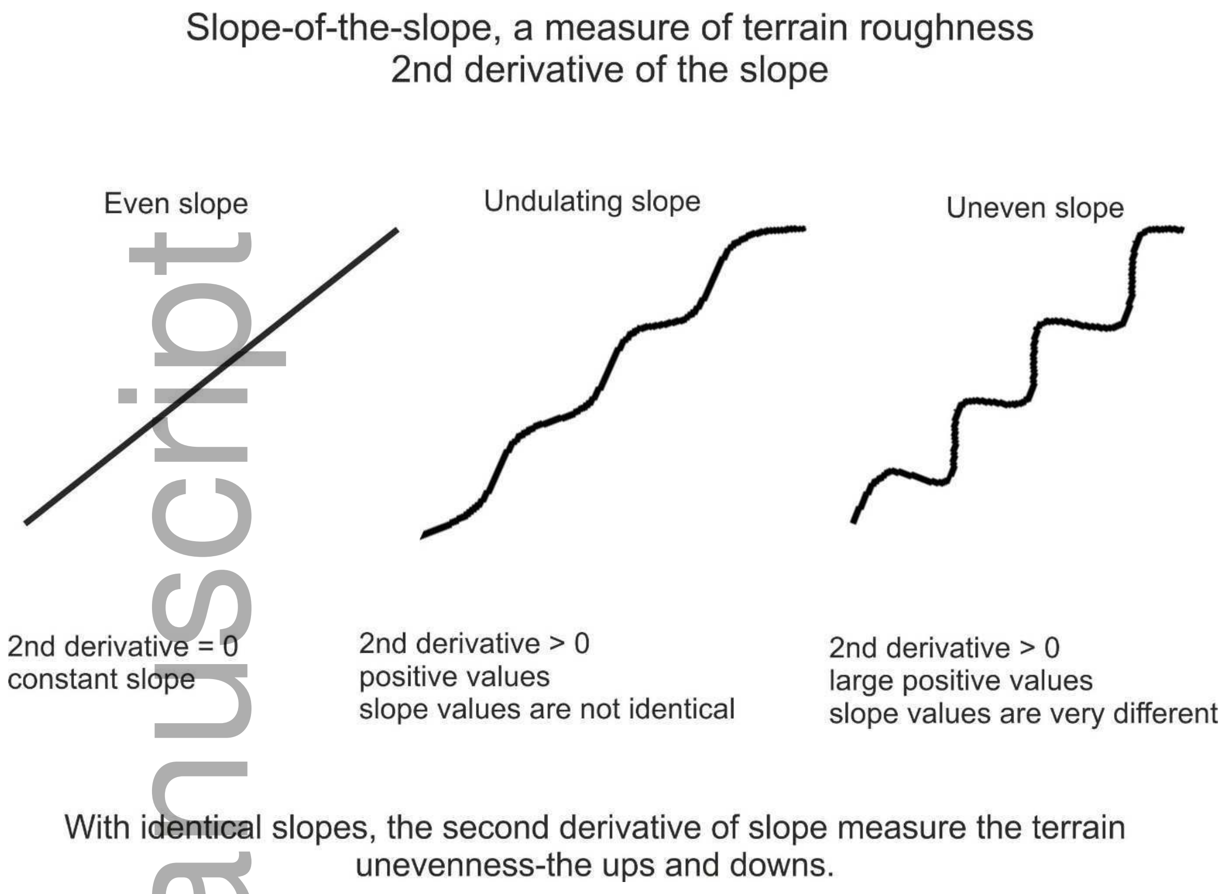


*Figure 2: Data collection and analytical process prior to implementation in the model. Bathymetry, slope-of-the-slope, habitat types* were quantified at multiple spatial scales 10m, 25m, 50m, 100m, 150m, 200m, 250m radius surrounding each fish sampling point. All environmental predictors were incorporated in models using Salford Predictive modeler.



With identical slopes, the second derivative of slope measure the terrain unevenness-the ups and downs.

Figure 3: Slope-of-the-slope, a measure of terrain roughness

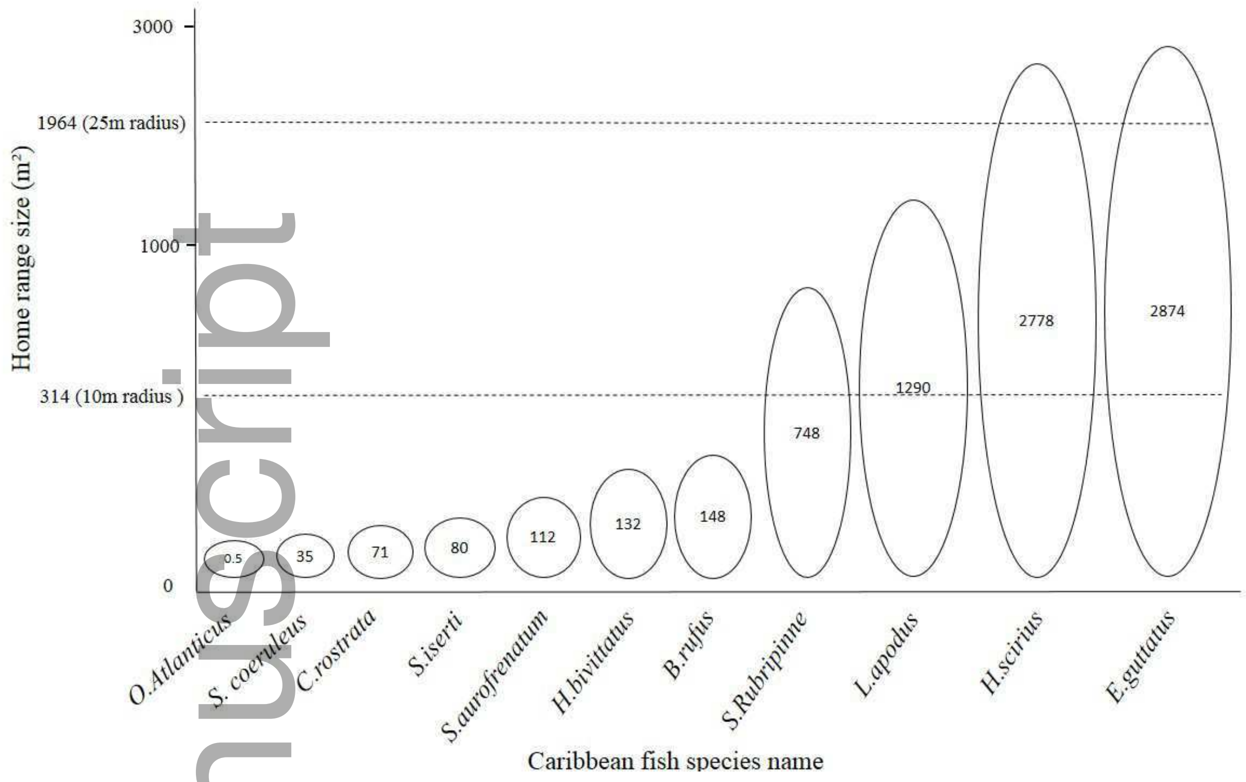


*Figure 4: Home range sizes (m²) of Caribbean reef associated fish species based on existing data. The ecological scale selection (10m, 25m, 50m, 100m, 150m, 200m, 250m radius) adopted in this study was based upon these home range sizes and extended to include larger areas.*

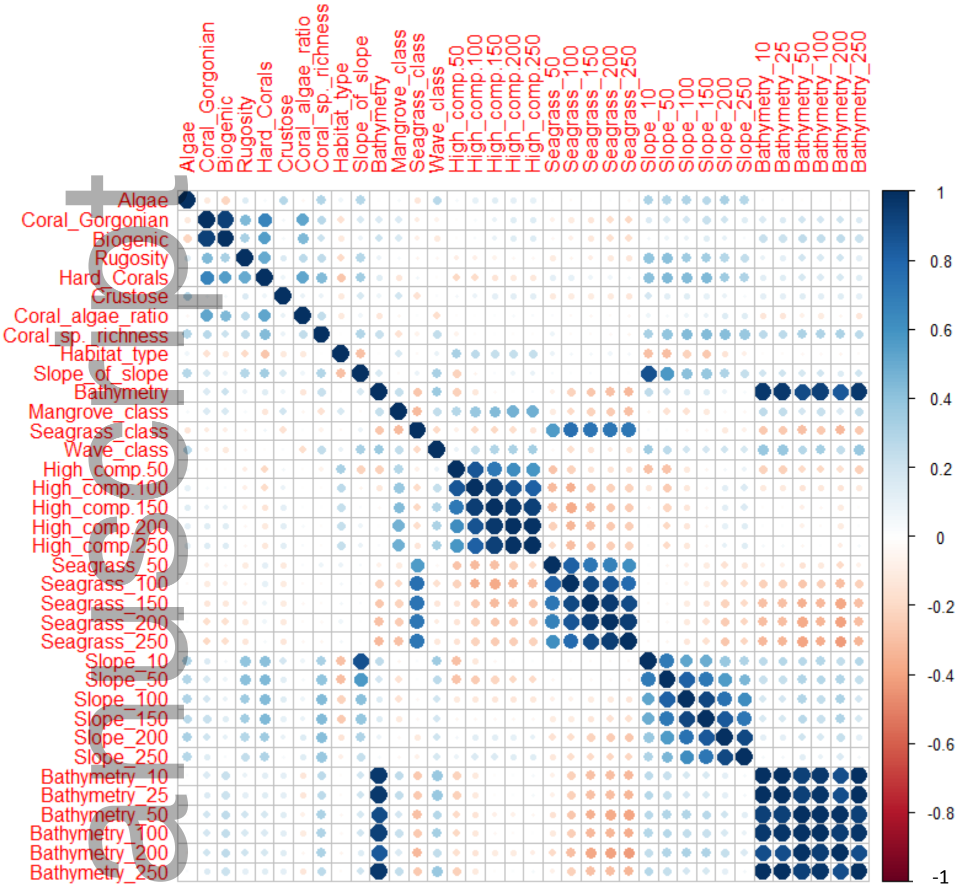


*Figure 5: Multicollinearity matrix between all predictor variables (refer to appendix table).*

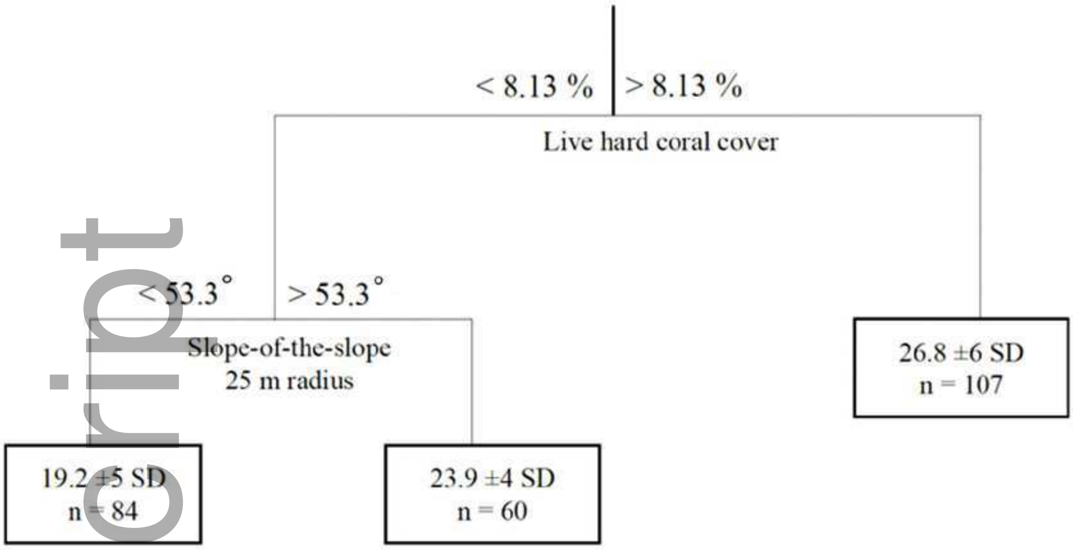


*Figure 6: Tree diagram of the CART model revealed that the primary splitters are live hard coral cover and slope-of-the-slope 25m radius. The CART diagram gives a qualitative indication on the most important environmental variables for fish species richness out of the 36 explanatory variables studied. With live hard coral cover greater than 8.13%, the mean species richness is 26.8 and with a slope-of-the-slope within 25m radius greater than 53.3°, the mean species richness is 23.9.*

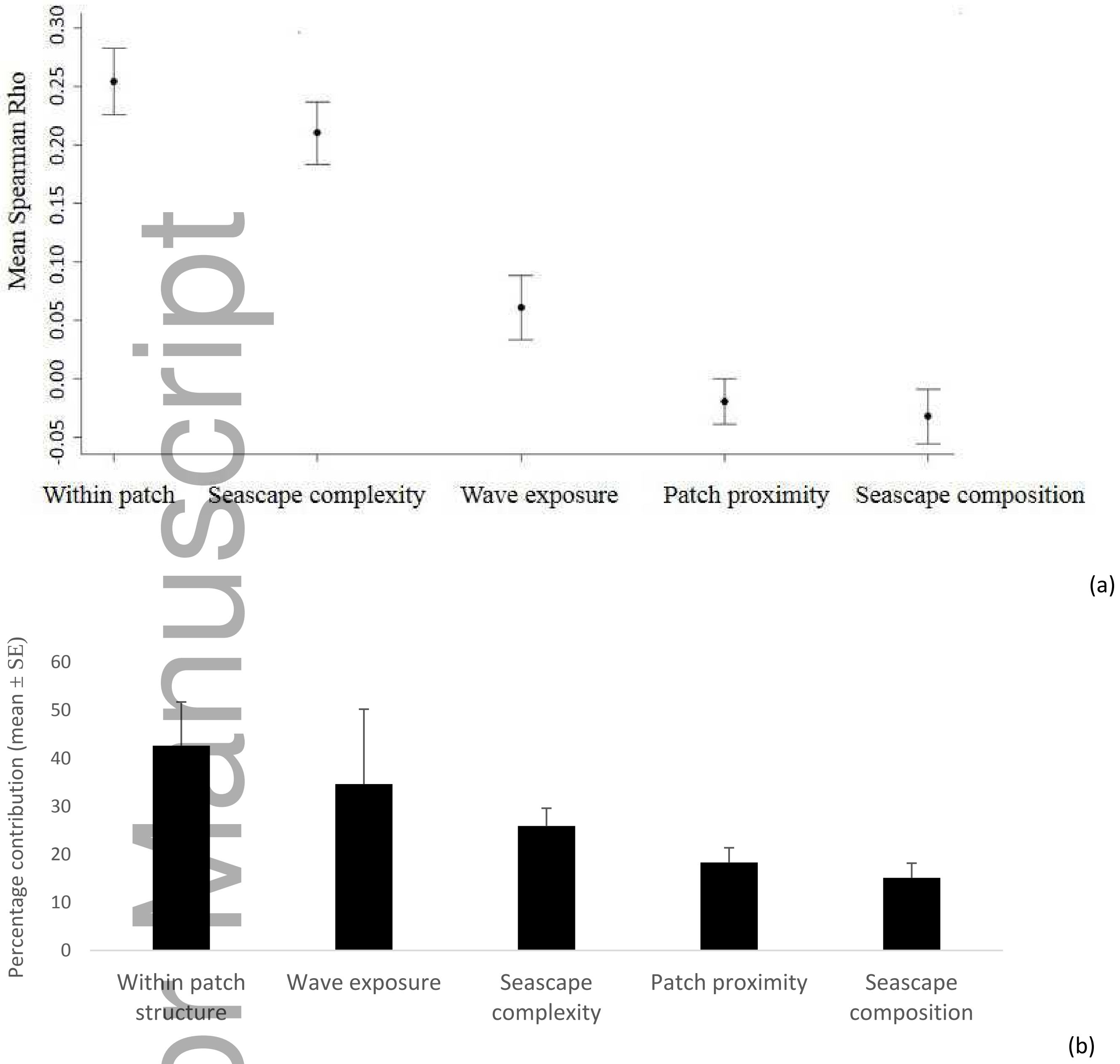


*Figure 7: (a) Spearman correlation coefficient (mean Rho±SE) between fish species richness and the five variable groups; within patch, seascape complexity, wave exposure, patch proximity and seascape composition (b) Percentage (mean±SE) contribution of variables by group to the TreeNet models.*

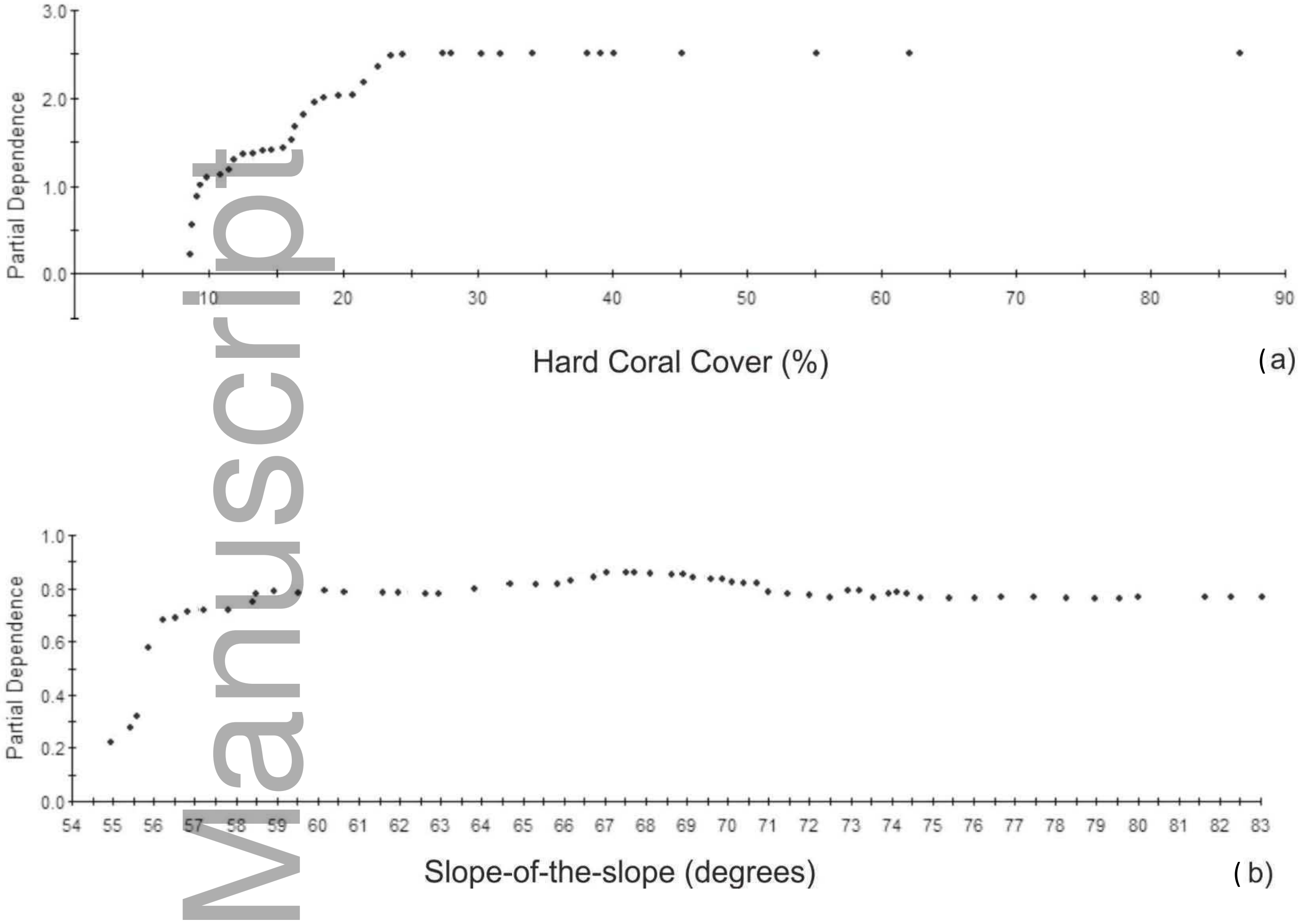


*Figure 8: (a) Fish species richness increase until reaching a stabilizing threshold at approximately 25% of maximum hard coral cover (b) Similar observation for slope-of-the-slope at 25m radius. The response variable increases before stabilizing at 60 degrees of slope-of-the-slope. Source: TreeNet model*

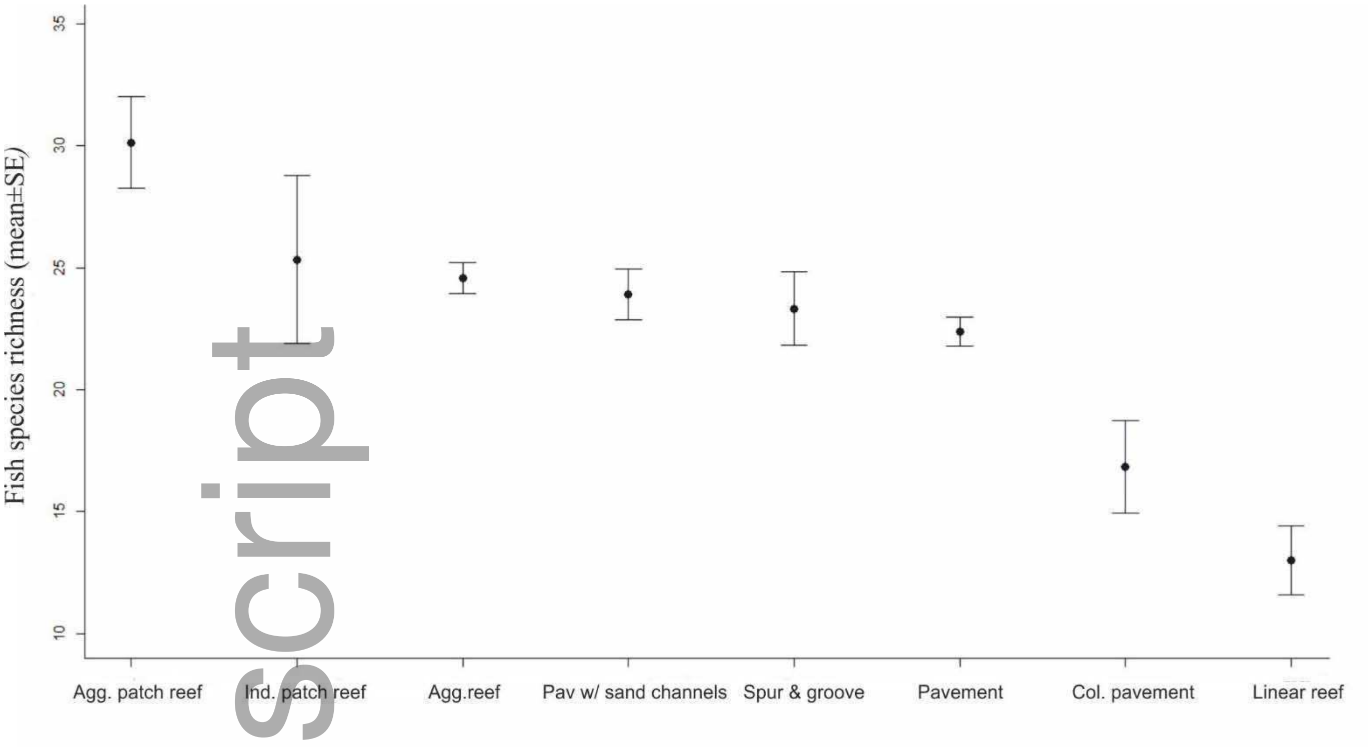


*Figure 9: Fish species richness (mean±SE) within each habitat type, eg: Aggregated patch reef, Individual patch reef, Aggregated reef, Pavement with sand channels, Spur and groove, Pavement, Colonized pavement and Linear reef.*

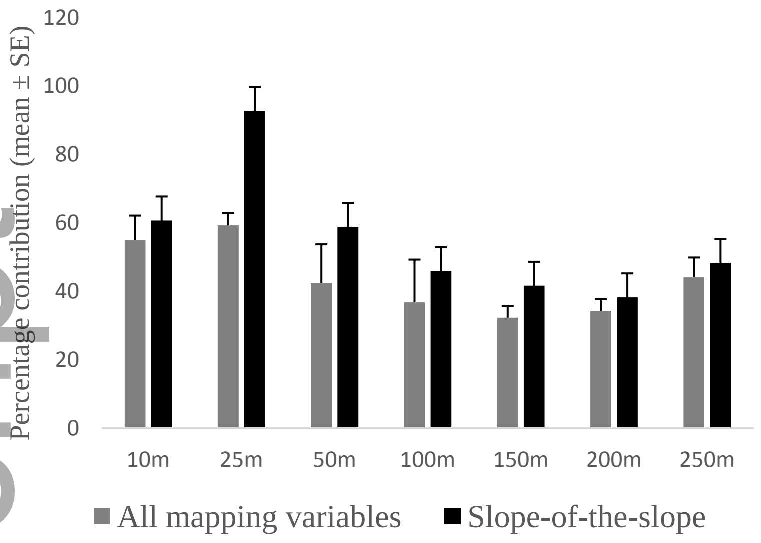


*Figure 10: Spatial scales percentage contribution to the TreeNet model (mean±SE) for all mapping variables and slope-of-the-slope. The seascape complexity variable slope-of-the-slope within 25m radius shows the highest percentage contribution to the TreeNet model.*

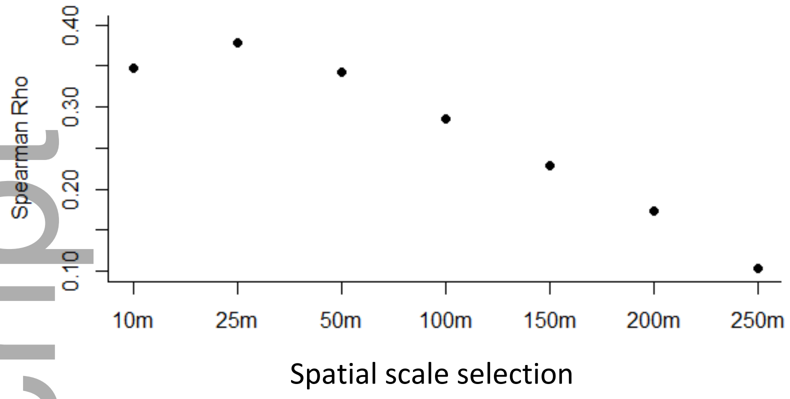


*Figure 11: Spearman correlation coefficients (Rho) for the seven spatial scales of slope-of-the- slope. Fish species richness revealed a stronger correlation with the topographic complexity variable slope-of-the-slope within 25m radius than within 250m radius.*

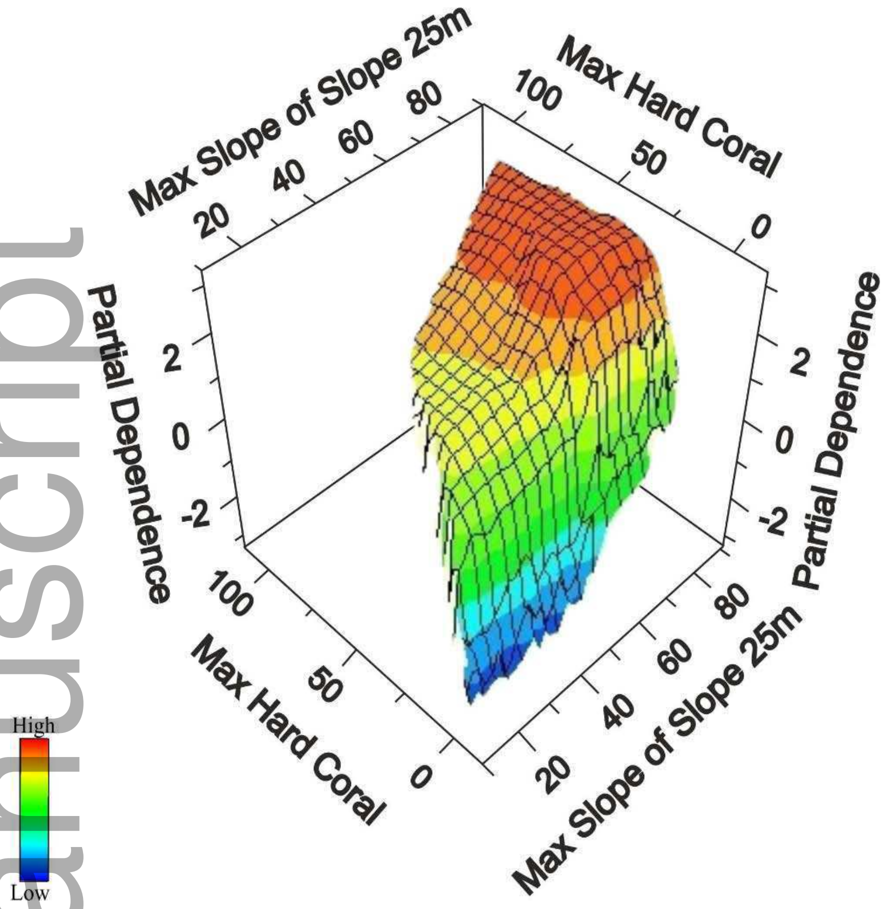


*Figure 12: Visualization of the interaction between hard coral cover and slope-of-the-slope at 25m radius. High species richness will be found in areas with hard coral cover between 25 and 86% with a slope-of-the-slope between 60 and 80 degrees within 25m radius. Source: TreeNet model*

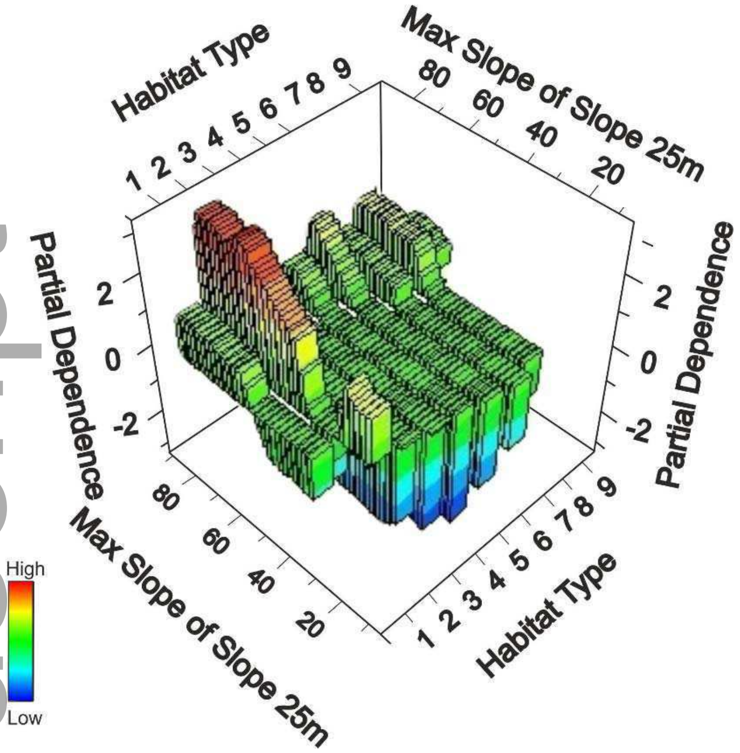


*Figure 13: Visualization of the interaction between the two most influential mapping variables for richness hotspots. Higher fish species richness was found in aggregated patch reefs (2) with a slope-of-the-slope between 60 and 80 degrees at a spatial scale of 25m. Source: TreeNet model.*

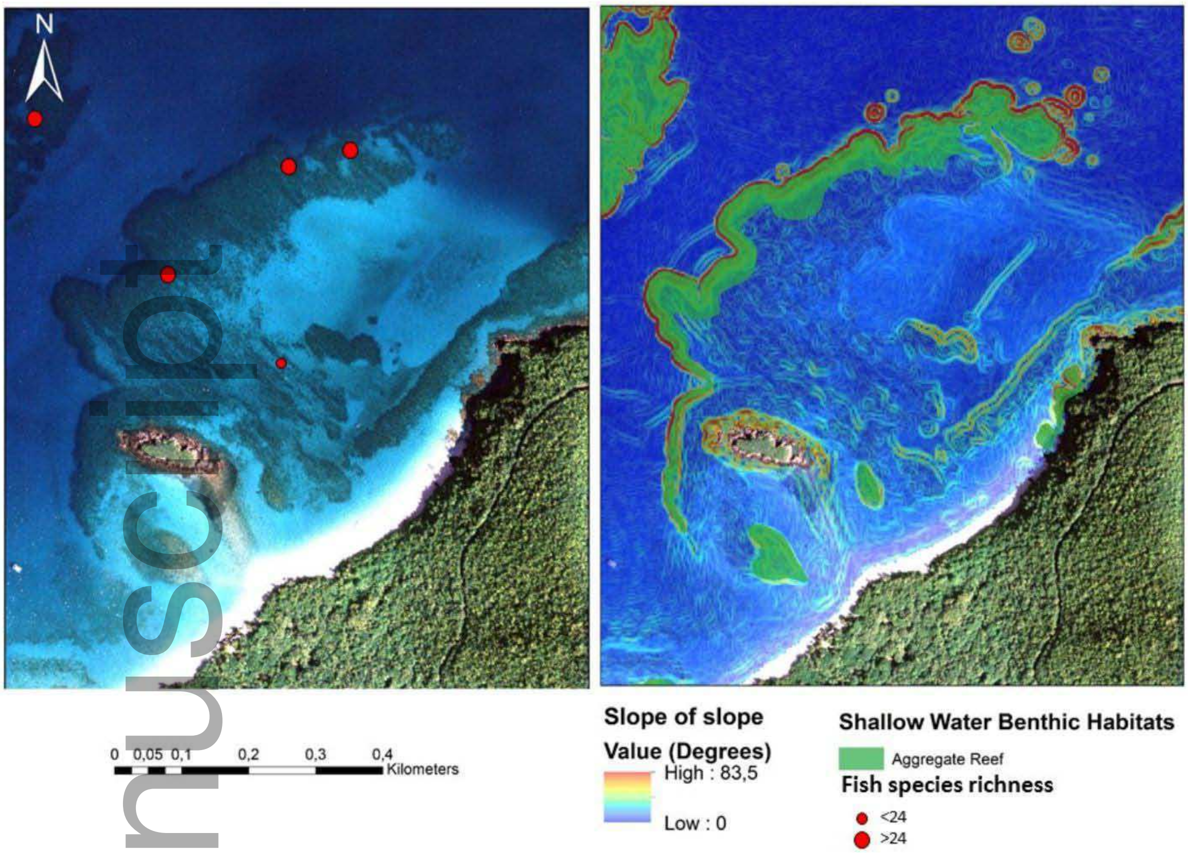


*Figure 14: On the left, an aerial image of St John, USVI, where reefs are clearly visible. The red dots correspond to data collection sites. On the right, the same aerial image with the Shallow Water Benthic Habitat map on top. Aggregated reefs are strongly associated with high values of slope-of-the-slope (Northeast region of St John, USVI). Source: Aerial imagery by NOAA Biogeography Branch (30cm resolution).*