Vol. 377: 1–11, 2009
doi: 10.3354/meps07839

MARINE ECOLOGY PROGRESS SERIES
Mar Ecol Prog Ser

Published February 26

OPEN ACCESS

FEATURE ARTICLE

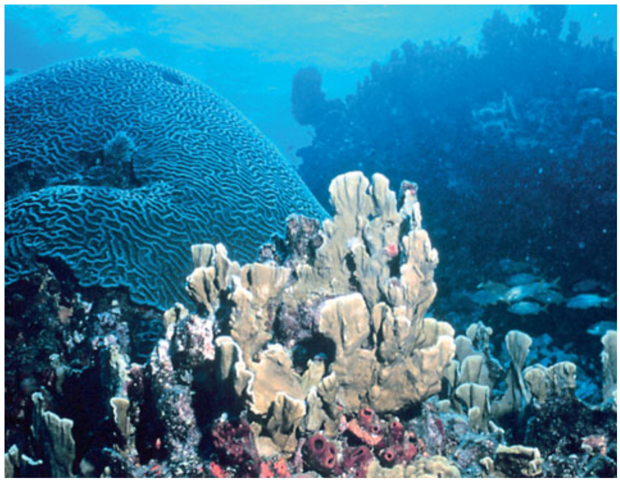
Rugosity-based regional modeling of hard-bottom habitat

Daniel C. Dunn*, Patrick N. Halpin

Marine Geospatial Ecology Lab, Nicholas School of the Environment, Box 90328, Duke University, Levine Science Research Center, Durham, North Carolina 27708, USA

ABSTRACT: Systematic conservation planning is most often directed at the representation and protection of marine biodiversity. However, direct observation and sampling of marine biodiversity is extremely time consuming and expensive. Due to these constraints, marine conservation planners have sought proxies for marine biodiversity to use in their models. Hard-bottom habitats support high levels of biodiversity and are frequently used as a surrogate for it in marine spatial planning. Rugosity (i.e. the roughness of the seafloor) is an indicator of hard-bottom habitat. In the present study, we expand on previous analyses of the relationship between rugosity and hard-bottom and create the first data-driven regional rugosity model to predict hard-bottom habitat. We used logistic regression to create an empirical model and compare it to other pre-formulated definitions of rugosity with receiver operator characteristic curves. Our model performed better than all other models and was able to correctly predict the presence or absence of hard-bottom habitat with ~70% accuracy. This model offers a fast and inexpensive alternative to more traditional survey methods, and should be of value to regional conservation planners and fisheries managers as an initial predictor of hard-bottom habitat. By testing this model with low-resolution (90 m) bathymetry data, we demonstrate that this type of information may be used in marine conservation plans in regions such as developing countries, where high-resolution data is not currently available. Further, our model offers a proxy for marine habitat diversity in non-coastal areas, an underrepresented sector in marine conservation planning.

KEY WORDS: Hard-bottom habitat · Rugosity · Mesoscale modeling · Remote sensing · Marine spatial planning · Biodiversity · Proxy



Hard-bottom habitat harboring high levels of biodiversity can be predicted at regional scales by modeling of remotely sensed bathymetric data.

Photo: Florida Keys National Marine Sanctuary

INTRODUCTION

Significant loss of marine biodiversity due to a variety of threats (pollution, resource extraction, climate change, etc.) has led to a call for better conservation and management planning in the marine environment (Hughes 1994, Jackson et al. 2001, POC 2003, USCOP 2004). This has been reflected in a steady increase in the number of spatially explicit marine planning efforts undertaken over the last 2 decades (Leslie 2005). Although conservation planning must be done at a variety of scales (Lourie & Vincent 2004), mesoscale planning efforts have garnered more attention recently, as regional managers attempt to prioritize their conser-

*Email: daniel.dunn@duke.edu

vation efforts and dollars. Such systematic conservation planning is based on explicit goals, and is most commonly directed at the representation and protection of marine biodiversity (Margules & Pressey 2000, Beck 2003, Leslie 2005). However, direct observation and sampling of marine biodiversity is extremely time consuming and expensive. Due to these constraints, marine conservation planners have sought out proxies for marine biodiversity to use in their models (Ward et al. 1999, Margules & Pressey 2000).

Hard-bottom habitats (i.e. live coral, rock/coral rubble, exposed low-profile carbonate and phosphorite substrates, thinly covered hard substrate with emergent growth, or artificial structures; SEAMAP-SA 2001) support some of the highest levels of biodiversity in the world (Connell 1978, Knowlton 2001). Hard-bottom habitats are also vital to the life history of many endangered species, including the napoleon wrasse *Cheilinus undulatus*, the banggai cardinalfish *Pterapogon kauderni*, the hawksbill sea turtle *Eretmochelys imbricata* (IUCN 2004), and important commercial and recreational fisheries (e.g. snapper/grouper and rockfish complexes; see www.safmc.net). Due to the high correlation between hard-bottom habitat and high biodiversity, maps of hard-bottom are often used as proxies for marine biodiversity in regional conservation planning and in siting marine reserves (DeBlieu et al. 2005, Ferdaña et al. 2006). Thus, the identification of hard-bottom habitat is a priority for marine conservation planners.

The most common methods employed to map hard-bottom habitats are the use of fixed towed video cameras, single or multibeam acoustic sonar, aerial photography, and satellite imagery (e.g. Landsat, SPOT, IKONOS, Quickbird). Although video cameras and sonar are quite accurate, they are labor intensive and have very small swaths or sampling extents. These methodologies can also be expensive due to technology costs for sonar and fuel and labor costs for both sonar and video. Towed cameras are further limited in their ability to collect data by the turbidity of the water. These types of high-resolution data are also typically not available for the vast majority of marine environments in the developing world and much of the developed world. Aerial photography and satellite images offer a much greater extent, but are also limited by water depth (≤25 m) and turbidity, as well as cloud cover, surface reflectance, and water turbulence. Due to these problems, few mesoscale datasets have been produced of hard-bottom habitat to assist regional conservation planning. One alternative gaining attention is the integration of these techniques into regional datasets (Todd & Greene 2007). Another option may be the use of regional bathymetry datasets to create a rugosity model to predict hard-bottom habitat.

Rugosity may be defined as the roughness of the physical structure of the seafloor, and is a key element and indicator of benthic habitat complexity (S. K. Wilson et al. 2007)[1]. Numerous studies have shown a strong correlation between either rugosity or benthic complexity and reef fish assemblages (Luckhurst & Luckhurst 1978, Roberts & Ormond 1987, McCormick 1994, Friedlander & Parrish 1998, Gratwicke & Speight 2005, S. K. Wilson et al. 2007), reviewed by Knudby et al. (2007), rockfish assemblages (Yoklavich et al. 2000, Williams & Ralston 2002, (NCCOS 2003), and gastropod abundance and diversity (Beck 2000). Rugosity has also been used to identify and classify benthic habitats using light detection and ranging, LIDAR, also known as airborne laser swath mapping (ALSM) (Brock et al. 2004, 2006, Kuffner et al. 2007), single or multibeam sonar (Lundblad et al. 2006, M. F. J. Wilson et al. 2007), and remote sensing based on satellite imagery (Pittman et al. 2007, Purkis et al. 2008). Recently, studies have combined the use of remote sensing and *in situ* observations to create multi-scale predictive models of reef fish diversity (Pittman et al. 2007, Purkis et al. 2008), or multi-scale benthic habitat models (M. F. J. Wilson et al. 2007).

Although explanatory models based on the relationship between rugosity and fish abundance and richness have been constructed previously, most of these models were created at fine resolutions of 1 to 10 m over small areas (Luckhurst & Luckhurst 1978, Friedlander & Parrish 1998, Gratwicke & Speight 2005) and may not be applicable at other scales. Pittman et al. (2007) identified this issue and used a multi-scale approach to determine that remotely sensed measures of rugosity alone may suffice to accurately predict fish species richness. They specifically identified fairly coarse grain rugosity (7225 m^2) as the primary explanatory variable in the model. Purkis et al. (2008) found similar, though weaker association, beginning at ≤5030 m^2. In the present study, we built on these findings and provided a coarse-grained, remotely sensed, regional predictor of hard-bottom habitat for the South Atlantic Bight, to aid efforts at regional marine conservation planning. This is the first regional hard-bottom dataset produced from bathymetry that we are aware of. Although Pittman et al. (2007), Purkis et al. (2008), and the other studies mentioned previously attempted to define the relationship between fish abundance or

[1]For the sake of clarity, we use only the term 'rugosity' when referring to the roughness of the sea floor. 'Topographic complexity' and 'benthic complexity' are often used synonymously with 'rugosity'; however, there are shades to their exact definitions (see Ferdaña et al. 2006). Our use of 'rugosity' in the present paper refers to changes in the degree and direction of relief and does not fully encompass other measures of topographic or benthic complexity.

diversity and rugosity, a regional model cannot assume to do this due to lack of data. Instead, with our model, we attempted to predict hard-bottom habitat as a proxy for fish abundance and diversity.

While Pittman et al. (2007) made major strides in attempting to bridge the gap between previous high-resolution analyses and the needs of marine conservation planners (i.e. data for coarse-grain, large-extent studies), the analysis was limited by the definition of rugosity as the standard deviation of depth within a specified extent. This is one of many algorithms for creating a rugosity index within a geographic information system (GIS). Other analyses (Brock et al. 2004, Jenness 2004, Gratwicke & Speight 2005, Kuffner et al. 2007, Purkis et al. 2008) have used a technique for defining rugosity grounded in Dahl's surface index (SI), which characterizes rugosity as the ratio of the surface area to the planimetric area (Dahl 1973). Rugosity indices have been formulated in still other analyses using the line density of the acceleration of the slope (i.e. the slope of the slope) (Ardron & Sointula 2002). The authors have also previously analyzed the use of bathymetric curvature, and simple combinations of aspect variety and depth difference in preliminary rugosity models. All of these methods are valid and useful. However, many have only been applied and tested on very high resolution (≤1 m^2) local datasets with limited extents and all are based on subjective, preconceived notions of how to describe rugosity. Ideally, since marine spatial planning needs to be performed at a variety of scales, both local and larger regional models would be incorporated into the planning process.

The lack of an empirical, mesoscale model to predict hard-bottom habitat remains a gap between researchers and managers, especially in developing countries and other data-poor regions. For such regions, we present here a method that can provide an initial predictor of hard-bottom habitat and fill an important data gap. We have created a data-driven model to predict areas of hard-bottom habitat by including several possible components of rugosity as predictor variables within a logistic regression framework. The model is based on a publicly available, coarse grain (~8100 m^2) bathymetric dataset, the National Geophysical Data Center Coastal Relief Model (Divins & Metzger 2003), so as to meet the needs of regional planners (e.g. Fishery Management Councils, conservation NGOs, and state environmental protection agencies).

MATERIALS AND METHODS

Study area. For the present study we used the Southeast Area Monitoring and Assessment Program (SEAMAP) database (SEAMAP-SA 2001), which describes the location and extent of hard-bottom reef habitats throughout the South Atlantic Bight. We selected 5 study sites between Jacksonville and Palm Beach on the Atlantic coast of Florida in the United States based on the availability of high-quality data from the SEAMAP hard-bottom dataset. We chose these locations (N = 8091) for the consistency of the sampling regimes spatially and by the gear type employed (both within each location and amongst the locations) to serve as the 'observed dataset' in our model. Sample sizes varied between sites (range = 1037 to 2349 observations), as did the percentage of hard-bottom observations (range = 8.2 to 16.3%). Locations that had >1 sample were removed, as were all 'potential hard-bottom', 'artificial hard-bottom' and 'artificial reef' points, limiting the total dataset used in the analysis to 7264 observations.

The sample sites ranged in size from 202 to 969 km^2 and were located in relatively close proximity to the shore (mean ± SD distance of sampled locations = 11.1 ± 6.4 km) on the upper continental shelf. The average depth across all sites was 15.66 ± 5.11 m.

Predictor variables. We used variations on (e.g. using different neighborhood sizes), and interactions between, 12 continuous predictor variables. The variables included were: slope, SD of the depth, flow accumulation (see http://support.esri.com/ for a detailed description of the flow accumulation function), aspect variety, slope difference, SD of the slope, distance to high slope, curvature, fine-, and broad-scale bathymetric position index (Wright et al. 2005), and the ratio of the surface area to the planimetric area. The variables flow accumulation and distance to high slope resulted in values several orders of magnitude larger than those of the other variables. These 2 variables were log transformed to help standardize errors and estimates. We derived all predictor variables used in the logistic regression from a bathymetric dataset: the National Geophysical Data Center (NGDC) Coastal Relief Model (Divins & Metzger 2003), which provides depth data for the coast of the United States at a resolution of 3 arc-seconds or ~90 m (~8100 m^2). We downloaded a custom grid of the bathymetric data encompassing the study sites (from 31 to 27° N, from the coast to 79.75° W) from the NGDC website, and manipulated it in ArcMAP 9.1 (ESRI 2005). Although there was significant and strong correlation between some of the predictor variables, we included all variables in the initial stepwise model as we did not plan on analyzing coefficient estimates and errors, but simply wanted the model with the best predictive ability.

We created all variables in ArcMap (ESRI 2005) as raster grids. Slope, aspect, curvature, and flow accumulation rasters were created from the bathymetry grid using functions within the Spatial Analyst Extension

(ESRI 2005). Rasters of SD of depth and slope, as well as the variety of aspects, were generated by passing a moving 3 × 3 cell window over slope, depth, or aspect grids and calculating a summary statistic for the neighborhood (i.e. the variety or SD). We produced the slope difference raster by subtracting the mean of a 9-cell moving window from the value of the center cell. The distance to high slope areas was calculated by reclassifying the slope grid into 2 classes using Jenks Natural Breaks, and then taking the Euclidean distance to the nearest high slope area for each grid cell. The ratio of surface area to planimetric area was calculated using the ArcView 3.x extension, Surface Areas and Ratios from Elevation Grid v. 1.2 (Jenness 2004). Both fine-scale and broad-scale bathymetric position index (BPI; see Wright et al. 2005, Lundblad et al. 2006) grids were constructed using the Benthic Terrain Modeler (BTM) extension for ArcMap 9.x (Wright et al. 2005).

Analytical framework. We modeled the presence or absence of hard-bottom habitat as a function of possible components of rugosity using a multiple logistic regression (Hosmer & Lemeshow 1989, Nelder & McCullagh 1989). Generalized linear models (GLMs) have frequently been used in presence/absence models and are easily interpretable (Guisan & Zimmerman 2000). We tested variables derived from a coarse (~90 m) bathymetric dataset to determine if they could be used to reasonably predict the presence or absence of hard-bottom habitat. We performed all statistical analyses using the R open-source software package (R-Development-Core-Team 2004), and used the 'glm' function from the base 'stats' package in R (Venables & Ripley 1999, R-Development-Core-Team 2004) and the 'stepAIC' function from the MASS package (Venables & Ripley 1999) to implement the logistic regression. As our response variable was binary (presence or absence of hard-bottom habitat), we selected the binomial family and a logistic link function for the GLM. We performed model selection by using Akaike's information criterion (AIC; Akaike 1973, 1974), as well as through the comparison of the area under the curve (AUC) of the receiver operator characteristic (ROC) curves (Green & Swets 1974, Hanley 1982, Lusted 1984). All stepwise regression model selection was based on which model had the lowest AIC score.

We completed the analysis in 2 stages: (1) creation and selection of a mesoscale rugosity model; (2) comparison of the empirical model to models based on existing formulations of rugosity.

(1) We created and compared 3 models to predict hard-bottom habitat. First, we ran a 'maximal model' that included all variables and second-order interaction terms. We assumed this model would overfit the data, but that it would represent the highest level of prediction achievable using these variables. Next, we included all variables from the maximal model in a backward and forward stepwise regression (referred to as the 'stepwise model'). The stepwise regression model was then stripped of all variables that were not significant ($p > 0.05$). After reiterating the stepwise regression model without these terms, several other terms failed to exhibit significant results ($p > 0.05$) and were consequently removed. We included the remaining variables in the final 'stripped model'.

(2) We compared the predictive ability of the best model from the first stage of the analysis with 3 commonly used models of rugosity: Dahl (1973)/Jenness (2004), Wright et al. (2005), and Pittman et al. (2007). We also included one further model using SD of depth and the variety of aspects, as we had used this proxy for rugosity previously. The Dahl/Jenness model included only the surface area ratio, and the Pittman model included only SD of depth. The Wright model (using the BTM) included slope, surface area ratio, and BPI at both fine and broad scales. We reproduced each of the other 4 models as a regression model using the same presence/absence data for hard-bottom habitat used to create our empirical models.

We further evaluated the 'best' model by analyzing ROC curves using the ROCR package (Sing et al. 2005). We constructed ROC curves for the best model and analyzed them with respect to the model's sensitivity or the true positive rate (number of true positive predictions/number of positive samples), and specificity or the true negative rate (number of true negative predictions/number of negative samples). The intersection of these 2 graphs maximizes the percent of both presence and absence (i.e. hard-bottom habitat and non-hard-bottom habitat) that may be correctly predicted at the same time. This intersection also corresponds to a cutoff point of the response variable (y) in the logistic regression that may be used to achieve the stated level of simultaneous prediction.

Our interpretation of AUC follows Hosmer & Lemeshow (1989) in considering an AUC of 0.5 as 'no discrimination', 0.7 to 0.8 as 'acceptable discrimination', 0.8 to 0.9 as 'excellent discrimination', and >0.9 as 'outstanding discrimination'.

Finally, we created predictive maps of the likelihood of hard-bottom habitat across the South Atlantic Bight by reproducing the equation of the best regression model in the Map Algebra tool within ArcMap.

RESULTS

We compared AIC and AUC values for the maximal, stepwise and stripped models. Models with low AIC values and high AUC values were considered to be better than those with higher AIC values and lower AUC

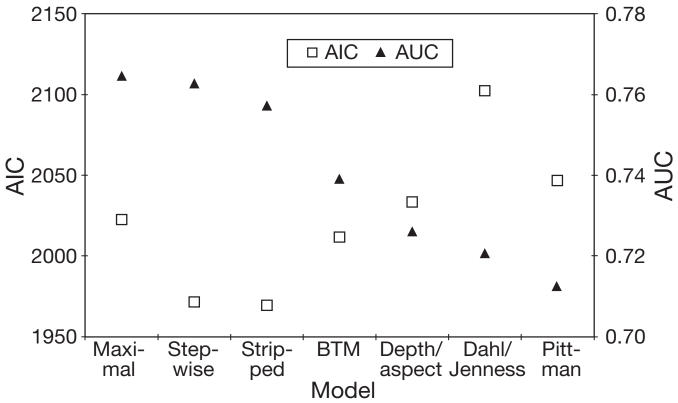


Fig. 1. Regression model results including area under the curve (AUC) and Akaike's information criterion (AIC). BTM: Benthic Terrain Modeler

values. All 3 models exhibited AUC values on the high end of the 'acceptable' range (Fig. 1). Although the maximal model had the highest AUC (AUC = 0.765), the AIC value for the maximal model was higher than both the stepwise regression model and the stripped stepwise regression model (AIC = 2022.5, 1971.4, and 1969.3, respectively). AUC values and AIC values for the stepwise and stripped models were very similar (0.763 and 0.757, respectively). In an effort to create a simpler, more easily interpretable and reproducible model for managers to use to predict hard-bottom habitat, we used the stripped model in the next stage of the analysis.

Stripped model explanatory variables

The stripped model used a large number of explanatory variables (Table 1). We attempted to substitute for, or eliminate, a number of the second-order interaction terms, but found significant increases in AIC and decreases in AUC each time. Six first-order terms were included in the final model: (1) variety of aspects, (2) slope difference from a 3 × 3 neighborhood, (3) SD of the slope, (4) log of distance to high slope areas, (5) surface area ratio, and (4) fine-scale BPI. Two of these, SD of the slope and neighborhood slope difference were highly correlated, and their coefficient estimates were not interpreted. The variety of aspects and the surface area ratio both reflect measures of the degree of rugosity in an area. Increases in these values led to a higher likelihood of hard-bottom habitat. The model indicated an inverse relationship between distance from high slope areas and likelihood of hard-bottom habitat, indicating a decrease in the probability of hard-bottom habitat further away from high slope areas. There was also an inverse relationship between fine-scale BPI and probability of hard-bottom habitat, though no immediate explanation for this relationship was found.

The second-order terms included in the best model were not easily explained. It is clear that SD of slope, variety of aspects, surface area ratio, and fine-scale BPI all played significant roles as interaction terms. We conjecture that areas of high slope deviation were

Table 1. Predictor variables and coefficients retained in the logistic regression model that 'best' predicted hard-bottom habitat (the Stripped Stepwise Regression Model). Variables were selected through backward and forward stepwise regression from an initial maximal model containing all first- and second-order terms. All non-significant values were stripped from the final model. Pr(>|z|) is the two-tailed p-value corresponding to the *z* ratio (*z* score) based on a normal distribution expressed in units of its distribution's standard deviation. ***p = 0.001; **p = 0.01; *p = 0.05; (–) p = 0.1. SD: standard deviation; BPI: bathymetric position index

Predictor variable	Estimate ± SE	*z*	Pr(>|*z*|)	p
(Intercept)	$(-5.121 \pm 1.160) \times 10^{4}$	–4.413	1.02×10^{-5}	***
Relative aspect variety	$(4.279 \pm 1.400) \times 10^{4}$	3.057	2.234×10^{-3}	**
Slope difference	$(-5.084 \pm 2.124) \times 10^{4}$	–2.393	1.6698×10^{-2}	*
SD of slope	$(6.468 \pm 1.621) \times 10^{4}$	3.989	6.62×10^{-5}	***
Log of distance to high slope	(-1.509 ± 0.317)	–4.761	1.92×10^{-6}	***
Surface area	$(5.121 \pm 1.160) \times 10^{4}$	4.413	1.02×10^{-5}	***
Fine-scale BPI	(-3.893 ± 1.145)	–3.401	6.70×10^{-4}	***
Relative aspect variety:slope	$(2.001 \pm 0.603) \times 10^{1}$	3.321	8.98×10^{-4}	***
Relative aspect variety:surface area	$(-4.279 \pm 1.400) \times 10^{4}$	–3.057	2.234×10^{-3}	**
Relative aspect variety:log of flow accumulation	$(-6.958 \pm 3.480) \times 10^{-1}$	–1.999	4.5564×10^{-2}	*
SD of slope:slope	$(3.251 \pm 1.550) \times 10^{1}$	2.097	3.5994×10^{-2}	*
Slope:SD of depth	$(-3.254 \pm 1.245) \times 10^{1}$	–2.613	8.987×10^{-3}	**
Difference from mean slope:SD of slope	(-9.474 ± 5.139)	–1.843	6.5260×10^{-2}	(–)
Difference from mean slope:surface area	$(5.084 \pm 2.124) \times 10^{4}$	2.393	1.6700×10^{-2}	*
SD of slope:log_slope_dist	(3.572 ± 0.857)	4.169	3.06×10^{-5}	***
SD of slope:SD of depth	$(1.373 \pm 0.388) \times 10^{1}$	3.539	4.02×10^{-4}	***
SD of slope:surface area	$(-6.469 \pm 1.621) \times 10^{4}$	–3.990	6.62×10^{-5}	***
Log of distance to high slope:fine-scale BPI	$(8.770 \pm 3.388) \times 10^{-1}$	2.589	9.631×10^{-3}	**
Fine-scale BPI:SD of depth	(3.206 ± 1.117)	2.872	4.083×10^{-3}	**
Fine-scale BPI:broad-scale BPI	$(7.293 \pm 2.799) \times 10^{-1}$	2.606	9.173×10^{-3}	**

even more important when they coincided with large differences in depth. This combination relates to the degree of relief and complexity in the benthic environment. The correlations between SD of depth, SD of slope, slope, and surface area make further speculation about the role of most of the second-order variables in the model inappropriate.

Empirical versus pre-formulated model comparisons

The stripped regression model was better at predicting hard-bottom habitat than the 4 commonly used rugosity models (Fig. 1). The other 4 models produced 'acceptable' discrimination levels (AUC > 0.7), but were not as good as the stripped stepwise regression model (AUC = 0.757). AIC values for these models (AIC > 2011) were also substantially worse than those for the stripped model (AIC = 1969.3). The stripped stepwise regression model was thus the best model for predicting hard-bottom habitat, and its performance was more closely evaluated to better understand its true predictive ability.

'Best' model evaluation

The AUC for the stripped model (Fig. 2a) is considerably better than random (AUC = 0.5), and corresponds to the middle portion of the 'acceptable discrimination' range. When both the true positive rate and the true negative rate are simultaneously maximized, the model correctly predicts areas of both hard-bottom and non-hard-bottom habitat with 70% accuracy (Fig. 2b). The cutoff (i.e. the y-value from the regression equation) associated with this level of prediction is 0.035. Although the model was created using only 'hard-bottom' data points from the SEAMAP dataset, the model appears to also predict 'partial hard-bottom' areas well (Fig. 3). A map of the probability of existence of hard-bottom areas in the South Atlantic Bight was created using this model (Fig. 4).

DISCUSSION

Filling data gaps in marine conservation planning

Although there has been an increase in the number of regional marine planning efforts over the last 2 decades, models to support the site prioritization tools that drive these plans have lagged behind. Limited data availability in developing countries has exacerbated this issue. Systematic conservation planning to represent and protect marine biodiversity requires spatially consistent, large-area datasets to allow comparisons to be drawn between sub-regions. However, most marine research is done over small areas and at high resolution, in an effort to control for unknowns and to better differentiate between the effects of variables. The needs of marine managers and researchers have often been at odds, and the gap between the products of one and the needs of the other has steadily widened (but see Valentine et al. 2005, Pitcher 2007). In the present study, we attempted to fill this gap by using the knowledge generated in small-scale studies

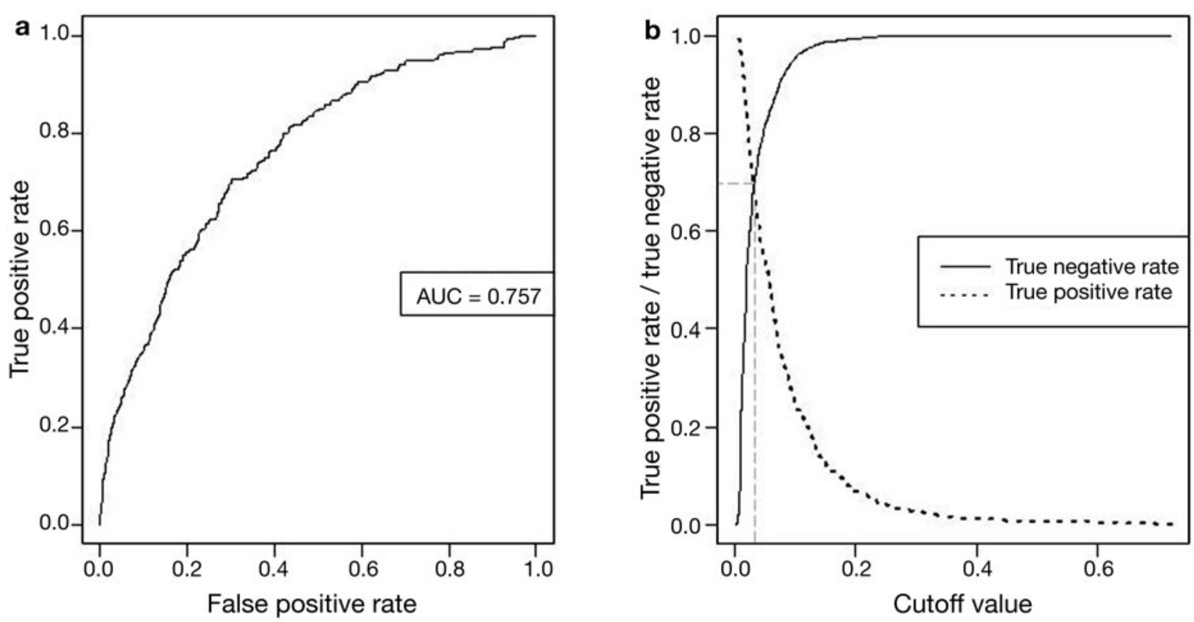


Fig. 2. (a) Receiver operator characteristic curve and (b) sensitivity/specificity curves for the 'stripped' regression model

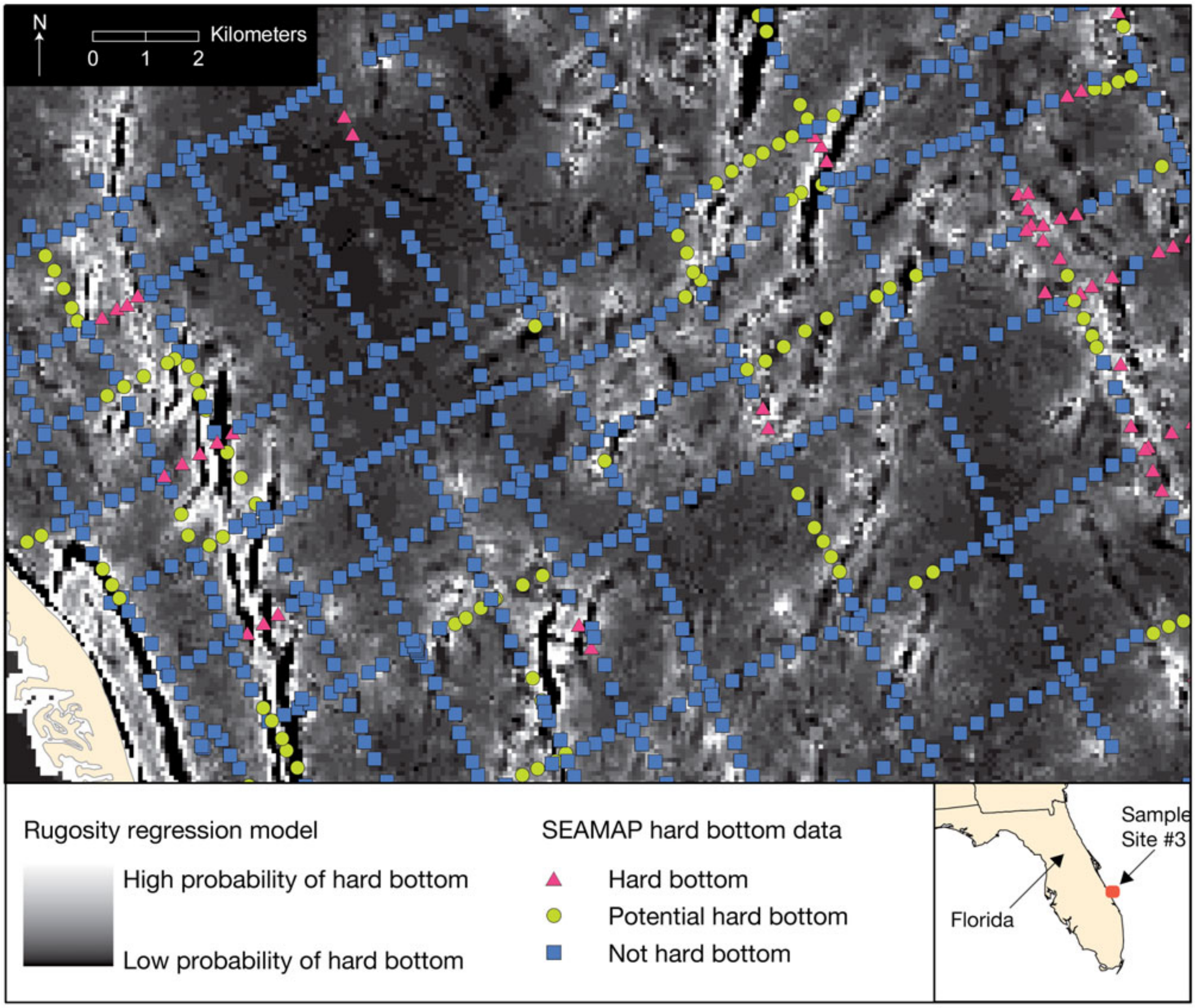


Fig. 3. Detail of predicted likelihood of hard-bottom habitat and observed samples from the SEAMAP hard-bottom dataset (including 'potential hard-bottom' sites) based on the stripped regression model

to create an empirically predicted dataset at a resolution and scale useable by marine resource managers. The regional hard-bottom dataset created in the present study is unique, and the methodology used to create it can be applied to any other marine environments and regions. Further, by employing a fairly low-resolution (90 m) bathymetry dataset to predict hard-bottom habitat, we have opened up to developing countries that may have no access to high-resolution datasets the possibility of incorporating this type of information into planning exercises.

The present study focussed on one more data gap in regional marine resource management: data availability in non-coastal areas. Most previous regional marine planning efforts have been highly focused on the land–sea interface because, at least in part, it contains the most easily accessible data. Recently, several national and international programs have begun to concentrate on non-coastal mapping (e.g. the European Union's Mapping European Seabed Habitat program and Australia's National Marine Bioregionalisation program). However, most models continue to focus on readily available population density and road density data, land-use classifications and estuarine habitat maps, and other datasets related to land-based anthropogenic effects on the marine environment. Few marine species habitat maps have been available on a regional scale, and thus incorporating them into such systematic planning schemes is difficult. Although we limited our model to depths of <60 m, this encompasses most of the continental shelf in the southeastern United States and thus dramatically enlarges the area over which managers can reasonably extend their planning efforts. In addition, although the applicability

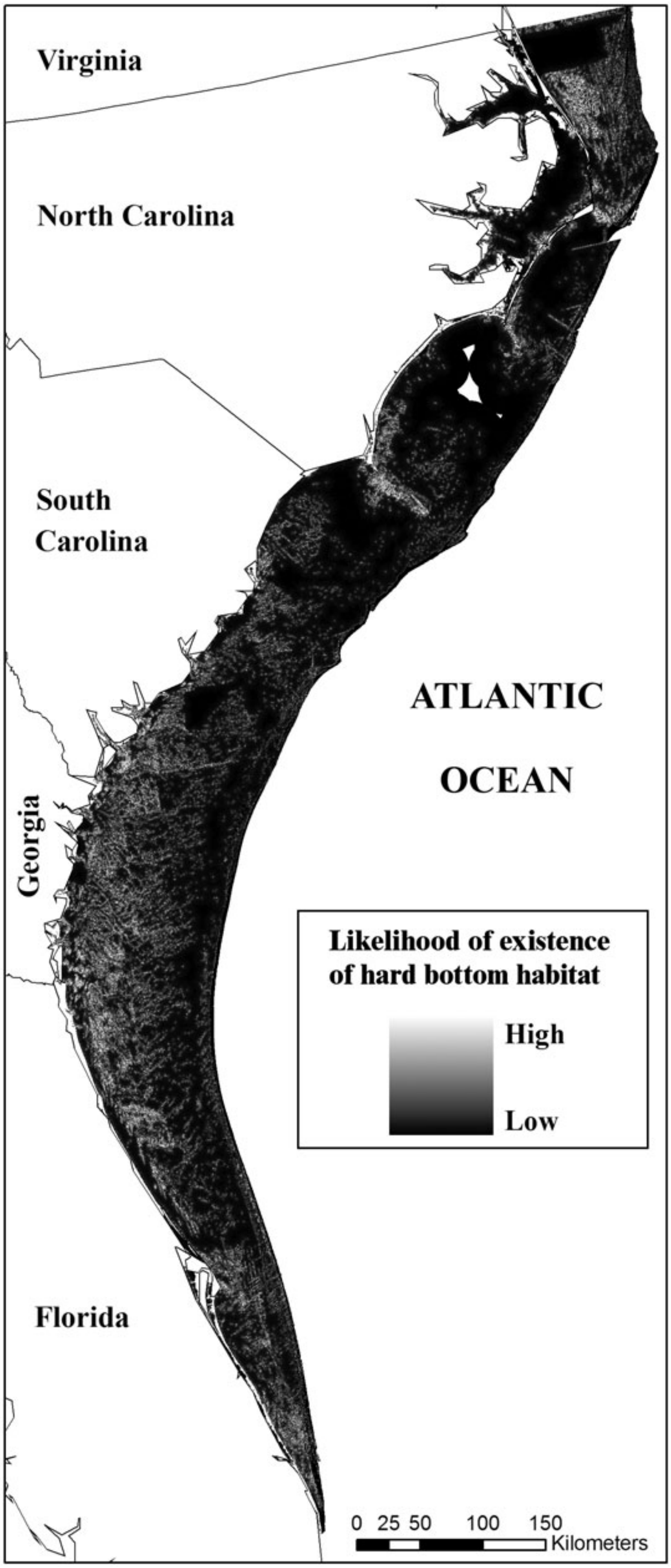


Fig. 4. Final predicted map of likelihood of hard-bottom habitat along the South Atlantic Bight based on the 'stripped' regression mode

of this model to other bathymetric datasets should be tested, the present study shows that the methodology is sound and should be transferable to predicting hard-bottom habitat on the continental slope (see M. F. J. Wilson et al. 2007) and the abyssal plain.

Model applicability

Fish species richness, and marine biodiversity more generally, have repeatedly been shown to be related to both hard-bottom habitat and rugosity. Pittman et al. (2007) directly suggested that, in the absence of data on habitat types, rugosity could be a useful surrogate for reef fish diversity. Our map of predicted hard-bottom habitat (Fig. 4) is the logical conclusion of our attempt to apply this knowledge on a regional scale. By presenting this map and the model it is based on, we hope to offer regional marine resource planners in both developing and developed countries a regional proxy for hard-bottom habitat and an initial indicator of marine biodiversity. This dataset could be used in site prioritization algorithms such as Marxan (Possingham et al. 2000), or to quantify and compare the likely amount of hard-bottom habitat being protected in existing marine reserves. It might also prove useful as an input in analyses of coral patch connectivity for areas where reef locations have not been identified. In this scenario, areas predicted to have hard-bottom habitat would be used as spawning locations in larval dispersal and network connectivity models (see Treml et al. 2008).

This method will also be useful to USA state fisheries agencies and regional fisheries councils as they attempt to comply with the essential fish habitat (EFH) standards in the re-authorized Magnuson-Stevens Fishery Conservation and Management Act and move towards an ecosystem approach to fisheries (EAF). As traditional fisheries management continues to fail to prevent overfishing in numerous fisheries (POC 2003, USCOP 2004, NOAA 2007), governmental bodies are also turning more and more to marine protected areas (MPAs) as a possible means of regulating fisheries and increasing fishery resilience (Sumaila et al. 2000, Halpern 2003). All of these new standards and management measures require a regional dataset of benthic habitat types. The current study presents an efficient method for these administrative bodies to take a first step towards incorporating habitat type and marine biodiversity into their planning, using free and readily available datasets.

This type of model may also prove useful as a comparison to low data-density areas in mosaic datasets created from a variety of data sources (e.g. the dataset used in this analysis: SEAMAP-SA 2001). These datasets may have numerous problems that raise uncertainty in their results (e.g. irregular spatial sampling, the variety of different methods/gears employed, the assumption of habitat presence based on capture of obligate species, and inconsistencies between sampling results at the same location). Evaluation of our model results in relation to these data-deficient areas will help increase confidence in the results and use of such mosaic datasets.

The authors are implementing a workflow that links all of the software tools needed to produce a predicted hard-bottom map for regions where this model is not directly applicable. Such a model would require similar bathymetric and presence/absence hard-bottom datasets for that region. Tools already exist to create several of the explanatory variable layers from the bathymetry data, including surface area and fine- and broad-scale BPI (Wright et al. 2005), and work is underway to create tools for the remaining variables. Once complete, these tools will be linked to spatially explicit statistical modeling software such as Marine Geospatial Ecology Tools (Roberts et al. 2008), which can fit regression models, create predicted presence/absence maps, and evaluate model performance using ROC curves and other metrics.

Model performance

All the empirical models we tested performed better than the pre-formulated algorithms. AIC values for all pre-formulated models except the BTM model were considerably higher than our models. AUC values showed less difference, but, again, our models performed better in this respect. The pre-formulated algorithms of rugosity have never been tested in a regression model. It is likely that the coefficients and equation formulated by the regression model significantly improved the results of these models for the purpose of predicting hard-bottom habitat. For this reason we recommend that regression analysis (e.g. GLMs, generalized additive models, classification, regression trees) be used regardless of which and how variables are chosen to be included in future attempts to build rugosity models on any scale.

Although, in our study area, the other models tested in this analysis did not discriminate between hard-bottom and non-hard-bottom areas as effectively as the empirical regression models, this in no way invalidates their prior use in other analyses. The stripped stepwise regression model was the 'best' model to predict this particular regional dataset, with its low relief, coarse resolution, and large extent. All of the other models were used to analyze different datasets, generally at much higher resolutions and over much smaller areas. However, from the perspective of predictive ability as interpreted by AUC, the pre-formulated models in the present study still provided 'acceptable' results. The simplicity of these models and the low data-processing requirements associated with them make them an interesting alternative or an even faster 'first-look' at where hard-bottom habitat is likely to exist.

Model limitations and future directions

Application of this methodology should work for a wide variety of benthic terrain types. However, this model itself is only applicable to low relief, nearshore, large-area studies. Use of this model without adjustment to predict hard-bottom habitats outside of the continental shelf of the southeastern USA is recommended only if strong similarities between the benthic environment in the 2 areas can be proven (i.e. habitat types, amount of relief, etc.) and the resolution of the input bathymetric dataset is similar. In all other circumstances, we recommend using the methodology set out in this paper to formulate a new predictive model.

Questions still remain regarding the ability of mesoscale models to directly predict species diversity, richness, and abundance. While our model does not directly predict these indices, it is meant to predict an indicator of them (i.e. hard-bottom habitat). Due to this, further groundtruthing of this model must be done at individual sites to determine if it not only predicts hard-bottom habitat, but is also a direct indicator of species diversity, richness, or abundance. The cost and time associated with such a project, however, precluded its inclusion in the present study. Macroscale studies have often found depth and latitude to be important factors in predicting diversity, richness, and abundance, as well as fish assemblages (Yoklavich et al. 2000, Williams & Ralston 2002, NCCOS 2003). These factors should also be included in marine spatial planning to help ensure representation of different species assemblages and to refine or weight biodiversity proxies.

The present study was limited by the coarse resolution of the bathymetric dataset and the errors inherent in the response variable dataset (i.e. the SEAMAP hard-bottom dataset). It is very likely that the model could be significantly improved if either of these 2 issues were resolved. Users should also be aware of data gaps in the coastal relief model bathymetry dataset. There are clearly anomalous areas and data-deficient areas within this dataset, and the predictive ability of our model is inhibited by this.

Numerous potential components of rugosity were included in this analysis as explanatory variables.

However, we did not attempt to find threshold values for the variables or test the use of the continuous variables against binary versions. As we saw in the final model from the inclusion of the variable representing the distance from the high slope, but not the variable for the slope itself, variables using thresholds may be more relevant than the straight variable in predicting hard-bottom habitat. We also did not look at variables at multiple scales. In particular, BPI could have been examined over different neighborhoods (as in M. F. J. Wilson et al. 2007). These factors, and other methods for transforming the predictor variables, should be taken into account in future rugosity models.

Acknowledgements. We thank the Southeast Area Monitoring and Assessment Program (SEAMAP) for supplying the SEAMAP hard-bottom dataset. We thank J. Moore for his contributions and advice regarding the analytical approach. We also appreciate the time, effort, and feedback given by J. Roberts, Z. Ferdaña, and C. Kot to review this manuscript. This work stems from previous research and we acknowledge the assistance of B. Best, C. Mankoff, and K. Eschelbach with the initial study. As always, we are indebted to our family and friends for their support of our endeavors.

LITERATURE CITED

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F (eds) Proc 2nd Int Symp Information Theory, Tsahkadsor, Armenia. Akademiai Kiado, Budapest, p 267–281

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Contr 19:716–723

Ardron J, Sointula BC (2002) A GIS recipe for determining benthic complexity: an indicator of species richness. In: Breman J (ed) Marine geography: GIS for the oceans and seas. Environmental Systems Research Institute, Redlands, CA, p 169–175

Beck MW (2000) Separating the elements of habitat structure: independent effects of habitat complexity and structural components on rocky intertidal gastropods. J Exp Mar Biol Ecol 249:29–49

Beck MW (2003) The sea around us: conservation planning in marine regions. In: Groves CR (ed) Drafting a conservation blueprint. Island Press, Washington, DC, p 319–344

Brock JC, Wright CW, Clayton T, Nayegandhi A (2004) LIDAR optical rugosity of coral reefs in Biscayne National Park, Florida. Coral Reefs 23:48–59

Brock JC, Wright CW, Kuffner IB, Hernandez R, Thompson P (2006) Airborne LIDAR sensing of massive stony coral colonies on patch reefs in the northern Florida reef tract. Remote Sens Environ 104:31–42

Connell JH (1978) Diversity in tropical rain forests and coral reefs. Science 199:1302–1310

Dahl AL (1973) Surface area in ecological analysis: quantification of benthic coral-reef algae. Mar Biol 23:239–249

DeBlieu J, Beck MW, Dorfman D, Ertel P (2005) Conservation in the Carolinian ecoregion: an ecoregional assessment. The Nature Conservancy, Arlington, VA

Divins DL, Metzger D (2003) NGDC coastal relief model. US Department of Commerce, National Oceanic and Atmospheric Administration, Washington, DC

ESRI (Environmental Systems Research Institute) (2005) ArcGIS. ESRI, Redlands, CA

Ferdaña Z, Beck MW, Dorfman D (2006) Improving methods for marine regional assessments: examples from the Pacific Northwest. The Nature Conservancy, Arlington, VA

Friedlander AM, Parrish JD (1998) Habitat characteristics affecting fish assemblages on a Hawaiian coral reef. J Exp Mar Biol Ecol 224:1–30

Gratwicke B, Speight MR (2005) The relationship between fish species richness, abundance and habitat complexity in a range of shallow tropical marine habitats. J Fish Biol 66:650–667

Green DM, Swets JA (1974) Signal detection theory and psychophysics. Krieger, Huntington, NY

Guisan A, Zimmerman NE (2000) Predictive habitat distribution models in ecology. Ecol Model 135:147–186

Halpern BS (2003) The impact of marine reserves: Do reserves work and does reserve size matter? Ecol Appl 13:117–137

Hanley JA (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29–36

Hosmer DW, Lemeshow S (1989) Applied logistic regression. John Wiley & Sons, New York

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

IUCN (International Union for Conservation of Nature) (2004) 2004 IUCN red list of threatened species. Available at: www.iucnredlist.org (accessed on 23 September 2008)

Jackson JBC, Kirby MX, Berger WH, Bjorndal KA and others (2001) Historical overfishing and the recent collapse of coastal ecosystems. Science 293:629–637

Jenness JS (2004) Calculating landscape surface area from digital elevation models. Wildl Soc Bull 32:829–839

Knowlton N (2001) The future of coral reefs. Proc Natl Acad Sci USA 98:5419–5425

Knudby A, LeDrew E, Newman C (2007) Progress in the use of remote sensing for coral reef biodiversity studies. Prog Phys Geogr 31:421–434

Kuffner IB, Brock JC, Grober-Dunsmore R, Bonito VE, Hickey TD, Wright CW (2007) Relationships between reef fish communities and remotely sensed rugosity measurements in Biscayne National Park, Florida, USA. Environ Biol Fishes 78:71–82

Leslie HM (2005) A synthesis of marine conservation planning approaches. Conserv Biol 19:1701–1713

Lourie SA, Vincent ACJ (2004) Using biogeography to help set priorities in marine conservation. Conserv Biol 18: 1004–1020

Luckhurst BE, Luckhurst K (1978) Analysis of the influence of substrate variables on coral reef fish communities. Mar Biol 49:317–323

Lundblad ER, Wright DJ, Miller J, Larkin EM and others (2006) A benthic terrain classification scheme for American Samoa. Mar Geod 29:89–111

Lusted LB (1984) ROC recollected. Med Decis Making 4: 131–135

Margules CR, Pressey RL (2000) Systematic conservation planning. Nature 405:243–253

McCormick MI (1994) Comparison of field methods for measuring surface topography and their associations with a tropical reef fish assemblage. Mar Ecol Prog Ser 112:87–96

NOAA (National Oceanic and Atmospheric Administration) (2007) 2006 Annual report to Congress on the status of US fisheries. US Department of Commerce, Silver Spring, MD

NCCOS (National Centers for Coastal Ocean Science) (2003) A biogeographic assessment off North/Central California:

to support the joint management plan review for Cordell Bank, Gulf of the Farallones, and Monterey Bay National Marine Sanctuaries: Phase I — marine fishes, birds and mammals. NOAA NCCOS biogeography team in cooperation with the National Marine Sanctuary Program. US Dept of Commerce, Silver Spring, MD

Nelder JA, McCullagh PP (1989) Generalized linear models, 2nd edn. Chapman & Hall, London

Pitcher CR (2007) Seabed biodiversity on the continental shelf of the Great Barrier Reef World Heritage Area. CSIRO Marine and Atmospheric Research, Cleveland, Qld

Pittman SJ, Christensen JD, Caldow C, Menza C, Monaco ME (2007) Predictive mapping of fish species richness across shallow-water seascapes in the Caribbean. Ecol Model 204:9–21

POC (2003) America's living oceans: charting a course for sea change; summary report: recommendations for a new ocean policy. Pew Oceans Commission, Arlington, VA

Possingham HP, Ball IR, Andelman S (2000) Mathematical methods for identifying representative reserve networks. In: Ferson S, Burgman M (eds) Quantitative methods for conservation biology. Springer-Verlag, New York, p 291–305

Purkis S, Graham N, Riegl B (2008) Predictability of reef fish diversity and abundance using remote sensing data in Diego Garcia (Chagos Archipelago). Coral Reefs 27: 167–178

R-Development-Core-Team (2004) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Roberts CM, Ormond RFG (1987) Habitat complexity and coral reef fish diversity and abundance on Red Sea fringing reefs. Mar Ecol Prog Ser 41:1–8

Roberts JJ, Best BD, Dunn DC, Treml EA, Halpin PN (2008) Marine geospatial ecology tools: an integrated framework for ecological geoprocessing with ArcGIS, Python, R, MATLAB, and C++. Available at: http://mgel.env.duke.edu/tools

SEAMAP-SA (Southeast Area Monitoring and Assessment Program–South Atlantic) (2001) Distribution of bottom habitats on the continental shelf from North Carolina through the Florida Keys. SEAMAP-SA Bottom Mapping Workgroup, Atlantic States Marine Fisheries Commission, Washington, DC

Sing T, Sander O, Beerenwinkel N, Lengauer T (2005) ROCR: visualizing classifier performance in R. Bioinformatics 21: 3940–3941

Sumaila UR, Gue'nette S, Alder J, Chuenpagdee R (2000) Addressing ecosystem effects of fishing using marine protected areas. ICES J Mar Sci 57:752–760

Todd BJ, Greene HG (eds) (2007) Mapping the seafloor for habitat characterization. Spec Pap 47, Geological Association of Canada, St. John's, NF

Treml E, Halpin P, Urban D, Pratson L (2008) Modeling population connectivity by ocean currents, a graph-theoretic approach for marine conservation. Landsc Ecol 23:19–36

USCOP (2004) An ocean blueprint for the 21st century. Final report, US Commission on Ocean Policy, Washington, DC

Valentine PC, Todd BJ, Kostylev VE (2005) Regional habitat classification as applied to marine sublittoral environments of northeastern North America. In: Barnes PW, Thomas JP (eds) Benthic habitats and the effects of fishing. American Fisheries Society, Bethesda, MD

Venables WN, Ripley BD (1999) Modern applied statistics with S, 4th edn. Springer, New York

Ward TJ, Vanderklift MA, Nicholls AO, Kenchington RA (1999) Selecting marine reserves using habitats and species assemblages as surrogates for biological diversity. Ecol Appl 9:691–698

Williams EH, Ralston S (2002) Distribution and co-occurrence of rockfishes (family: Sebastidae) over trawlable shelf and slope habitats of California and southern Oregon. Fish Bull 100:836–855

Wilson MFJ, O'Connell B, Brown C, Guinan JC, Grehan AJ (2007) Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope. Mar Geod 30:3–35

Wilson SK, Graham NAJ, Polunin NVC (2007) Appraisal of visual assessments of habitat complexity and benthic composition on coral reefs. Mar Biol 151:1069–1076

Wright DJ, Lundblad ER, Larkin EM, Rinehart RW, Murphy J, Cary-Kothera L, Draganov K (2005) ArcGIS benthic terrain modeler. Oregon State University, Corvallis,OR

Yoklavich MM, Greene HG, Sullivan DE, Lea RN, Love MS (2000) Habitat associations of deep-water rockfishes in a submarine canyon: an example of a natural refuge. Fish Bull 98:625–641

Editorial responsibility: Romuald Lipcius, Gloucester Point, Virginia, USA

Submitted: June 23, 2008; Accepted: November 19, 2008
Proofs received from author(s): January 29, 2009