Assessment of Benthic Habitats for Fisheries Management

May 2015

NCCOS points of contact: Brian Kinlan, Laura Kracker, Matt Poti

NMFS point of contact: Todd Kellison, NMFS SEFSC Beaufort Lab

URL: http://www.coastalscience.noaa.gov/projects/detail?key=226

Assessment of Benthic Habitats for Fisheries Management

INTRODUCTION

NOAA Southeast Fishery-Independent Survey (SEFIS) conducts annual surveys of hard bottom habitats of the southeast coast of the US to support reef fish stock assessments. The Beaufort Laboratory requested that NCCOS provide technical assistance to improve the understanding of managed species distribution patterns based on seafloor features. NCCOS has provided support, first, by interpreting and classifying hard bottom and other seafloor features from multibeam data acquired on the NOAA ship *Pisces*, and second, by applying predictive models to known hard bottom locations and related environmental data (Figure 1). The classification of hard bottom habitats from the ME70 multibeam data has resulted in high resolution habitat maps at 21 locations. These procedures have advanced the usefulness of the ME70 sonar system, common to all NOAA Fisheries Survey Vessels, in developing a habitat classification scheme relevant to fisheries independent monitoring efforts.

The predictive modeling portion of this effort uses statistical models to integrate information on the location of known hardbottom habitats, along with broader environmental and seafloor characteristics to predict and map the location of likely hardbottom habitats. The models are validated by comparing to information collected by video and the previously mentioned classified maps from the multibeam sonar survey data. The work completed here is improving our knowledge of South Atlantic hardbottom habitat distribution and supporting development of an expanded survey universe for fishery-independent sampling. The expansion of the fishery-independent survey universe into currently undersampled areas will lead to more accurate and representative fishery-independent information on abundance for use in stock assessments and reduce uncertainty in those assessments. The creation of a regional map of hardbottom habitat for the first time opens up the possibility of stratification of sampling by habitat, as well as improved information on EFH for use by the South Atlantic Fishery Management Council.

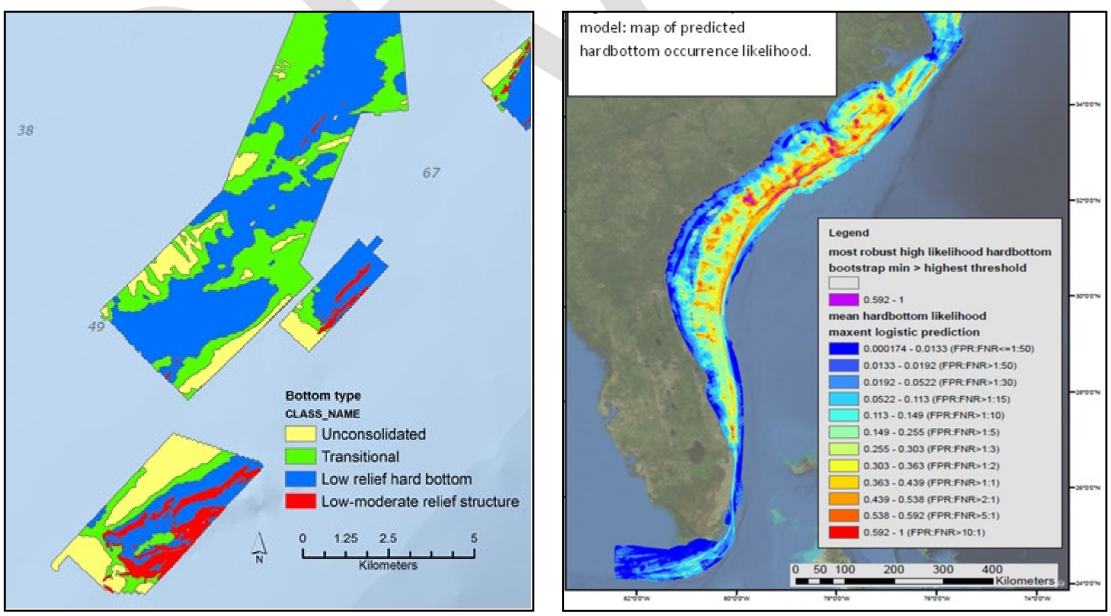


Figure 1. Fine scale benthic habitat classification map of selected marine areas derived from multibeam bathymetry and backscatter (left). Hard bottom map based on predictive model of hardbottom occurrence likelihood (right).

DRAFT

OVERVIEW

This document summarizes the methods and results of both the habitat classification mapping of targeted fish trap survey areas, as well as predictive modeling based on known hard bottom locations with validation using the classification maps produced in the first effort.

HABITAT CLASIFICATION

Habitat Classification - Methods

The data for the habitat classification component was acquired using the Simrad ME70 sonar system on the NOAA ship *Pisces.* The ME70 is a multibeam echosounder system (MBES) with 45 configurable beams and a maximum swath width of XXX. A beam configuration developed by Tom Weber (UNH CCOM) was used to collect bathymetry and backscatter data that were processed and interpreted for the habitat classification component.

Several approaches were evaluated to determine the best method for classifying ME70 backscatter and bathymetry data to identify bottom habitat types. However, since the bathymetry data collected in 2012 was not of the highest quality, this analysis relied heavily on backscatter information (Figure 2). First, a basic interpretation using a backscatter threshold (dB) to separate hard (high intensity returns) and soft bottom (low intensity returns) habitats was applied (Figure 3).

Second, an automatic feature extraction segmentation process (conducted in ENVI) was used to delineate based on backscatter characteristics. Scale and merge levels were selected to affect the size and number of features captured. These segments were then edited to better align with observable backscatter and bathymetric features by overlaying the images in ArcGIS and manually editing errant segments (Figure 4). Determining the habitat type of derived polygons depended on reviewing trap camera data (see Figure 2, upper left) and the original backscatter imagery. An attempt to apply zonal statistics to these updated polygons and use hierarchical clustering to assign polygons to classes did not provide logical results (Figure 6).

Third, an unsupervised classification (Isodata- Iterative Self-Organizing Data Analysis Technique) on backscatter data alone was produced; outputting two, three and four classes (Figure 6, 7, 8 respectively) to delineate areas. Isodata is a clustering algorithm that assigns the resultant merged pixels to the number of classes specified by the user. Trap camera data collected at SEFIS survey sites and expert opinion based on known sites within the study area were used to assist in determining which habitat type each class represents. In addition, relief areas were defined by running a 3x3 moving average window tool to calculate depth range values across the surface. The results were merged with the Isodata classification. The resulting surface had depth ranges of 0.3 to 3.0 meters which were considered low-moderate relief structure. This method of merging the unsupervised Isodata classification with depth ranges was deemed most useful because classes are suggested and the resulting class delineations were well aligned with the prior feature extraction/manual editing approach.

This approach results in a total of four classes, three of which (Unconsolidated, Transitional, Low relief hard bottom) are strongly driven by backscatter and the fourth derived from depth range (low-moderate relief structure). Two of the three backscatter classes (Unconsolidated and Low relief hard bottom) likely represent a coarse threshold of backscatter values (compare Figure 3 and Figure 6), while the third backscatter class (Transitional) appears to capture a transition zone between unconsolidated and low relief hard bottom habitats (Table 1 and Figure 7). This map was considered the best preliminary draft map and

DRAFT

was further reviewed by SEFIS scientists (Figure 9). Generating five classes (four backscatter plus depth range) revealed additional detail (Figure 8). However, whether or not this level of detail is justified and can be validated is uncertain and was rejected in favor of the four class map.

Table 1. Previous and current habitat classification legend.

Previous classification 2010 Reson 7125 Nancy Foster (Kagesten 2012)	**Current Legend PC1204**	Current classification 2012 Simrad ME70 Pisces (Kracker 2014)
Fine sediments	Yellow	Unconsolidated
Coarse sediments	Green	Transitional
Low relief hard bottom	Blue	Low relief hard bottom
Low relief structure Moderate relief structure	Red	Low-moderate relief structure
High relief structure		None

Note: To see how the current classification relates to the Coastal and Marine Ecological Classification Standard (CMECS), see Appendix 4.

DRAFT

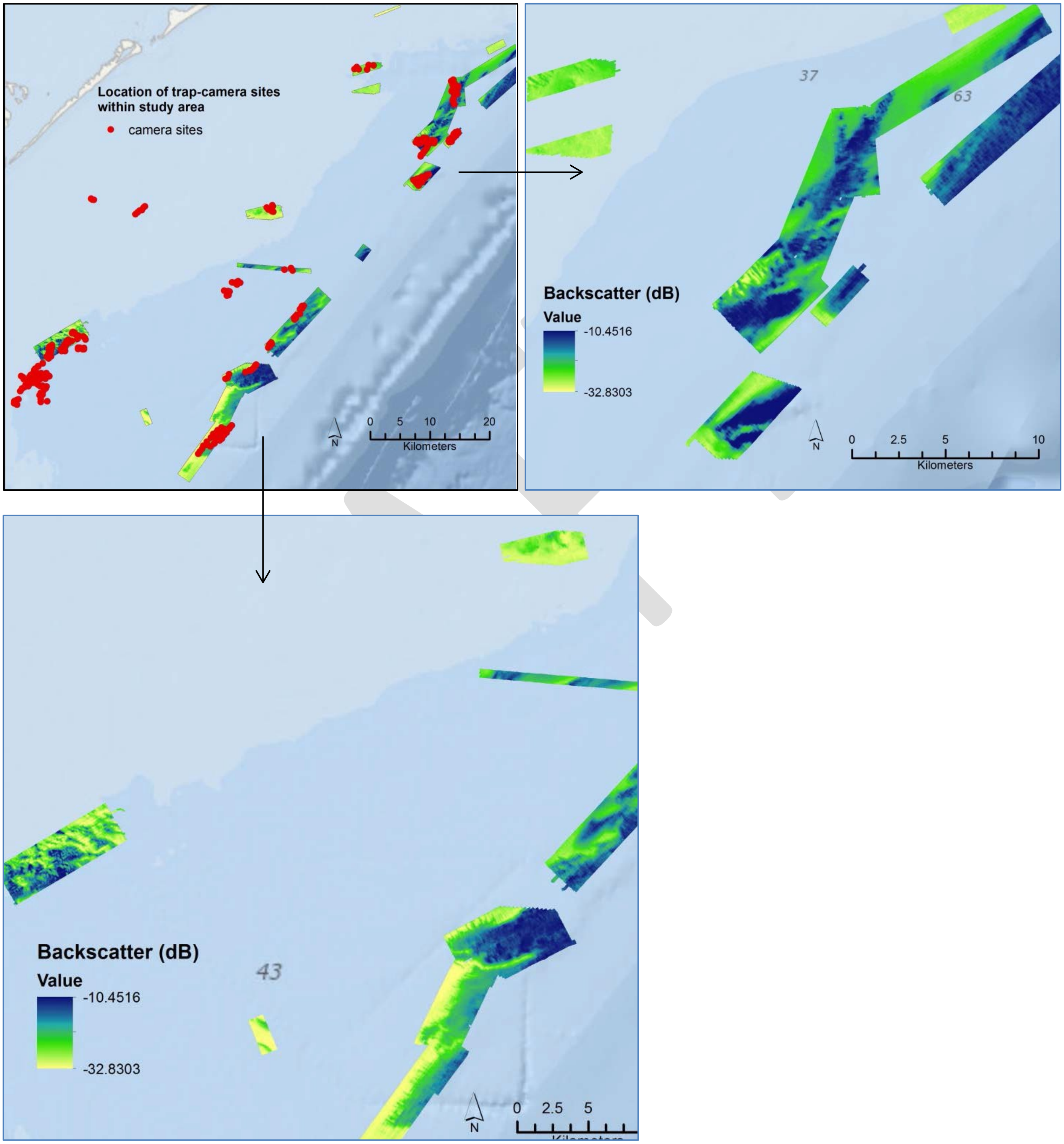


Figure 2. Raw backscatter collected with the ME70 on the NOAA ship *Pisces* in 2012.

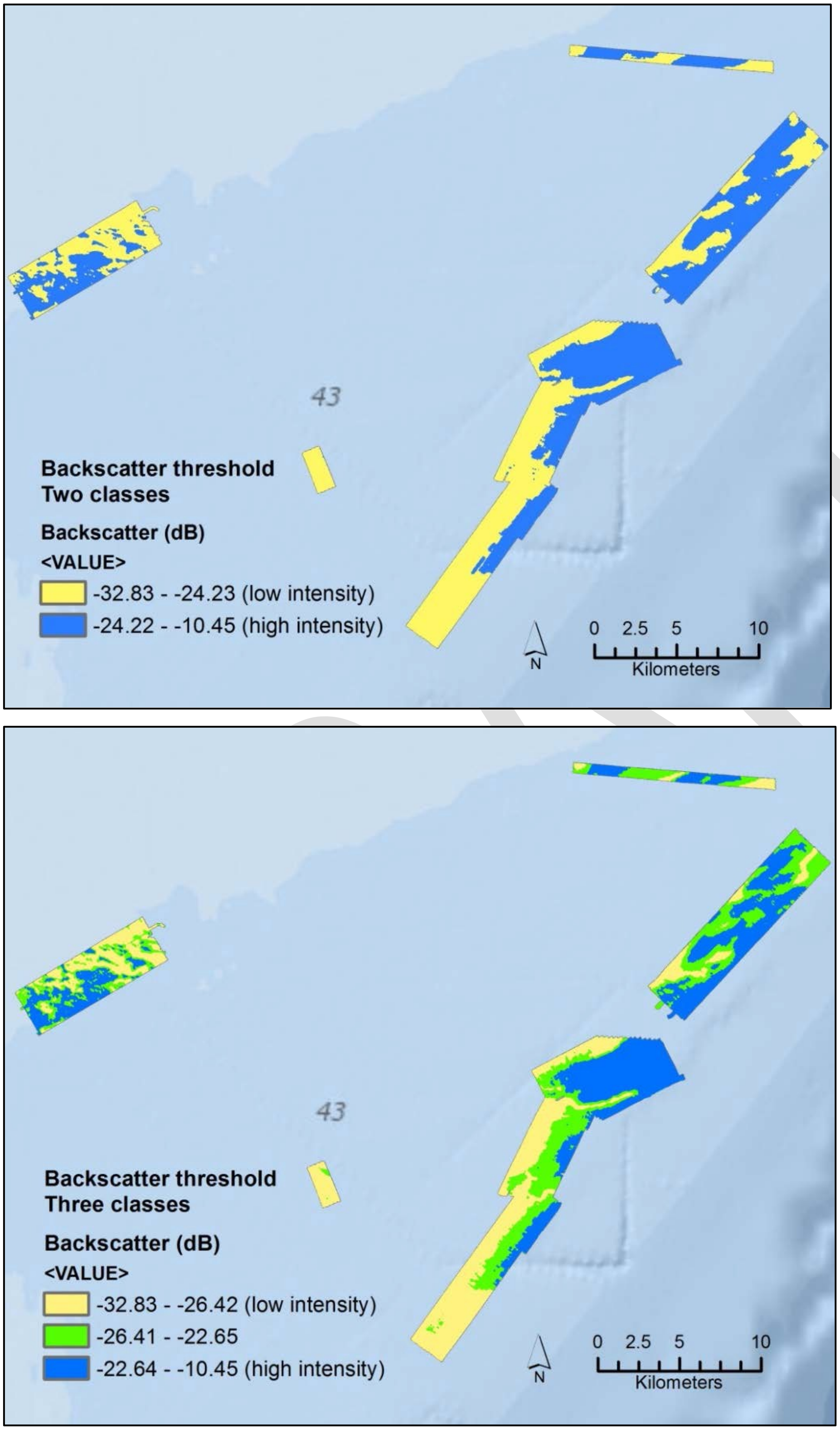


Figure 3. Backscatter with two (upper panel) and three classes (lower panel) using Jenk's natural breaks algorithm to differentiate hard and soft bottom areas.

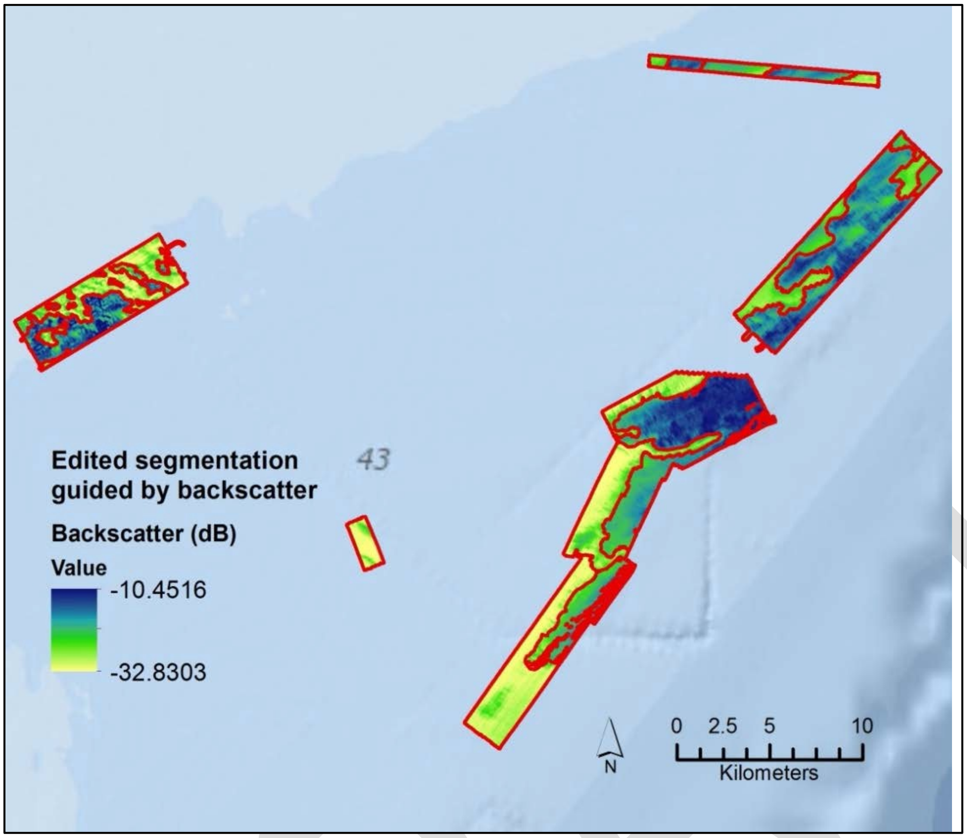


Figure 4. Edited 'Segment Only Feature Extraction' in ENVI guided by backscatter.

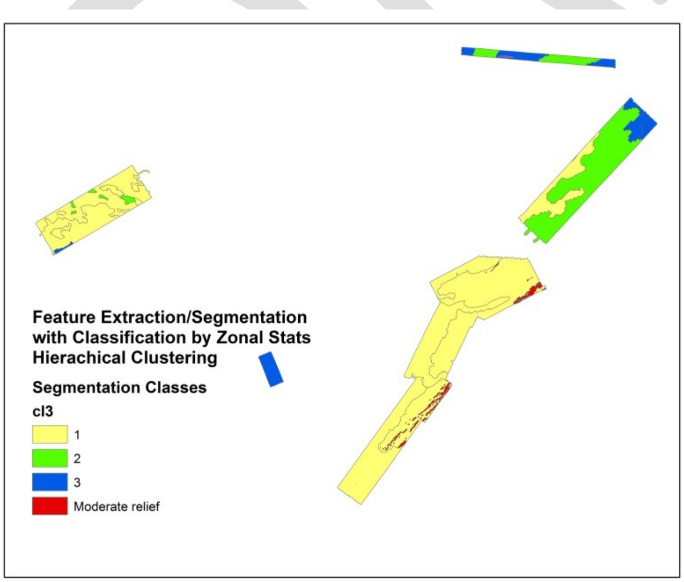


Figure 5. Poor outcome using zonal statistics of segmented polygons and Ward's hierarchical clustering to assign class membership to extracted features.

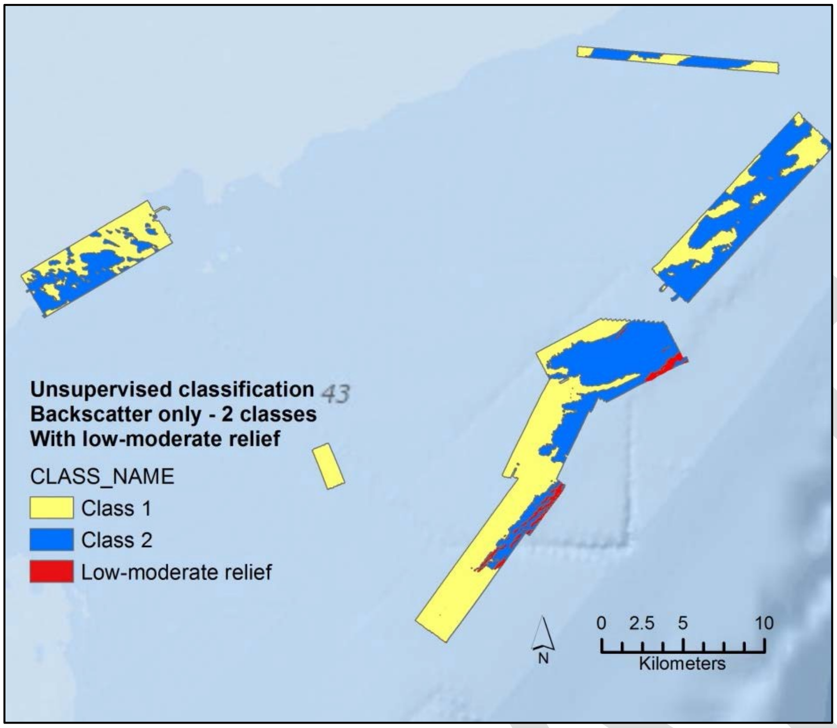


Figure 6. Unsupervised classification on backscatter only (2 classes) plus low-moderate relief

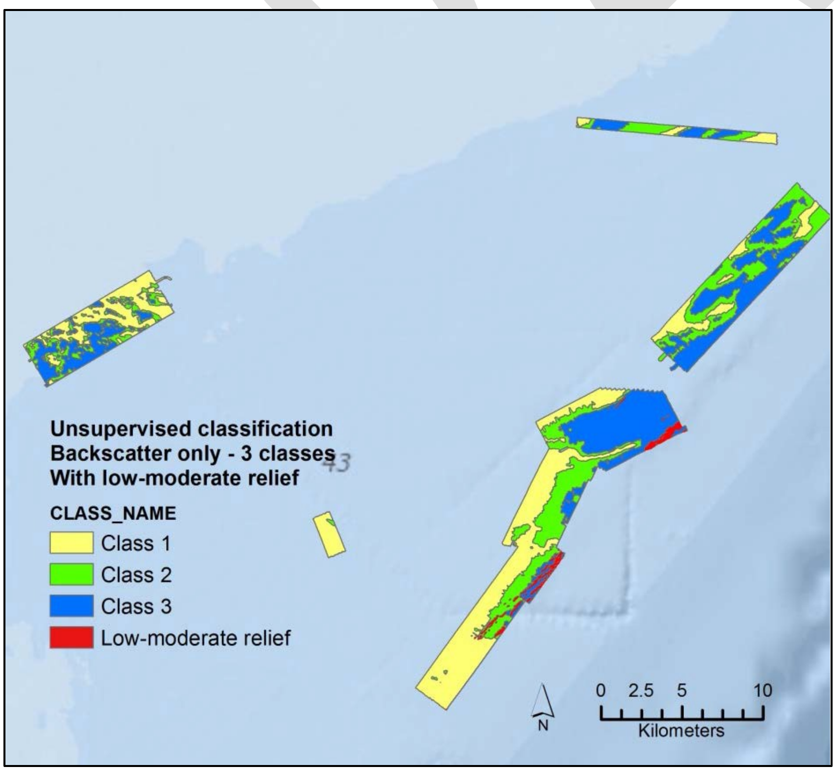


Figure 7. Unsupervised classification on backscatter only (3 classes) plus low-moderate relief.

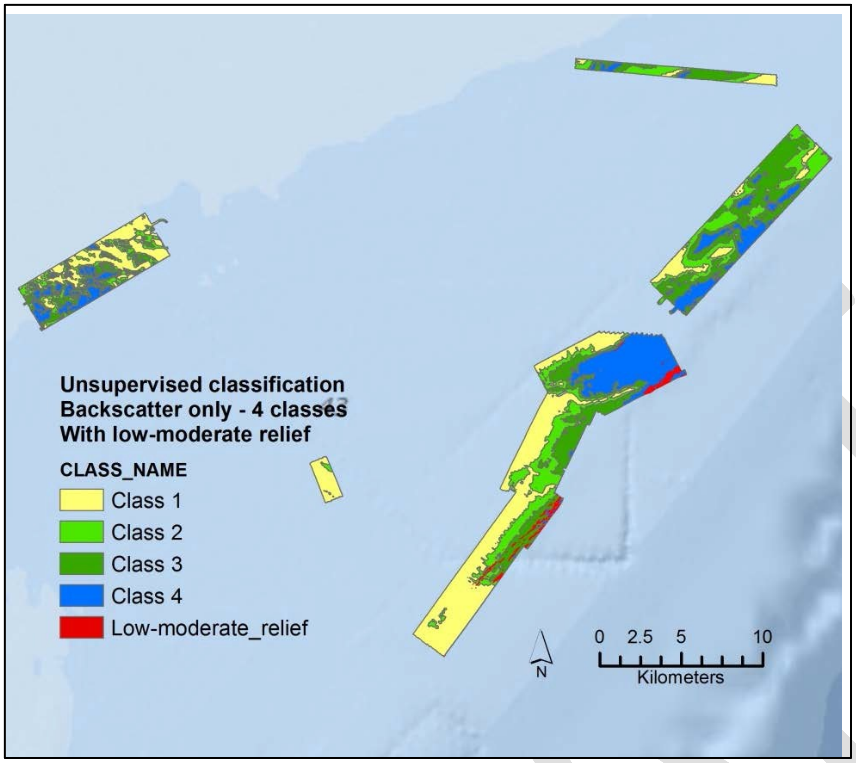


Figure 8. Unsupervised classification on backscatter only (4 classes) plus low-moderate relief.

DRAFT

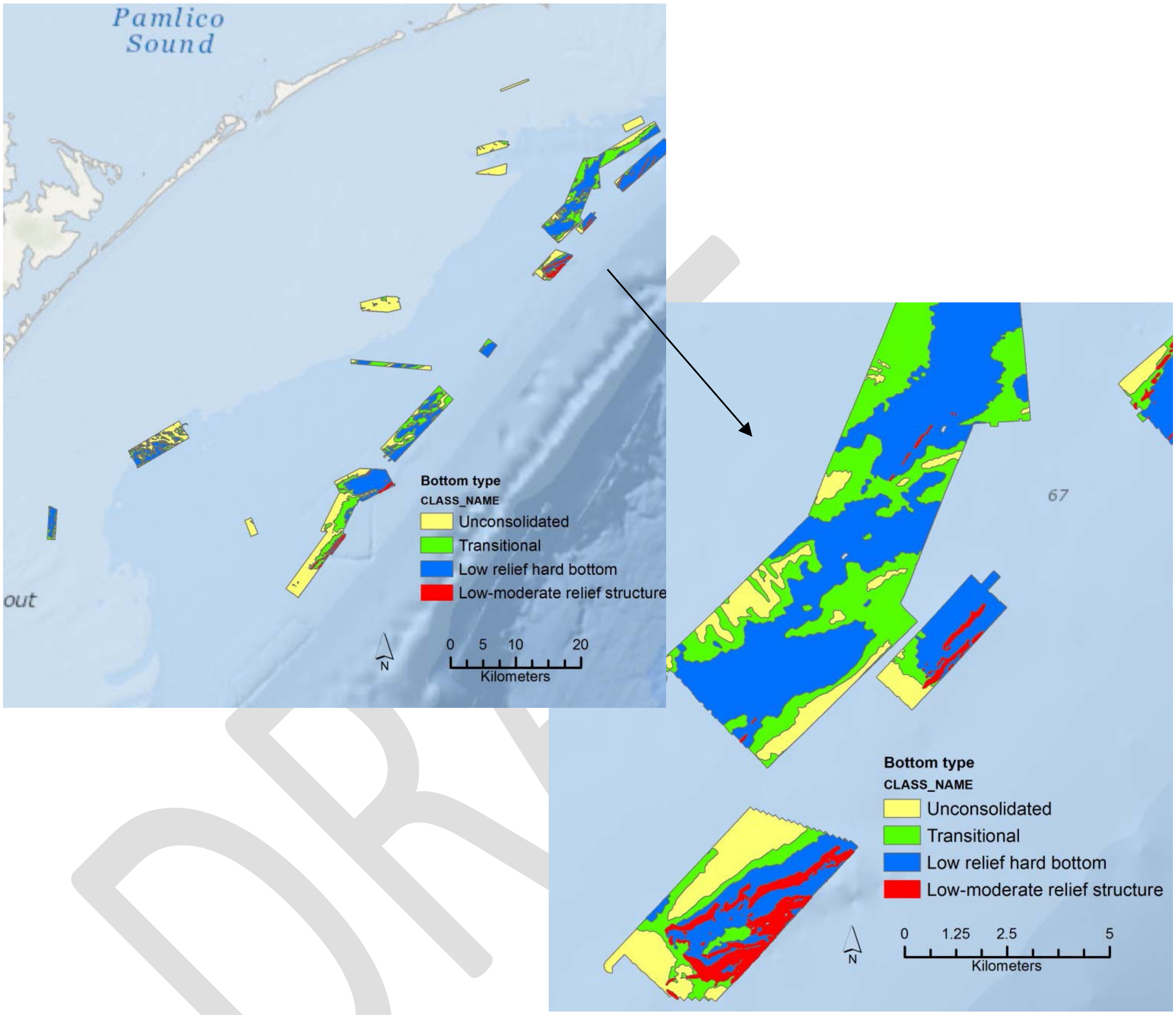


Figure 9. Unsupervised classification on backscatter into three classes plus low-moderate relief structure with final habitat classification.

DRAFT

Habitat Classification – Results

A review of the various classification options as they relate to the types and configuration of habitats in the final classification scheme was conducted in concert with SEFIS scientists to determine the most appropriate segmentation approach, the number of classes to include, and class descriptors. SEFIS scientists have conducted camera/trap surveys in the region since 2010 and are very knowledgeable about the type and variability of habitats present. That expert knowledge was critical in interpreting results of a semi-automated process such as this. The final scheme (Figure 9 and Table 1) includes a total of four classes: three classes (Unconsolidated, Transitional, and Low relief hard bottom) derived primarily from the backscatter return and one class (Low-moderate relief structure) which is based on depth range values.

Habitat Classification – Note

Approaches using derived bathymetry metrics have been useful in defining benthic habitats in other regions such as USVI (Costa et al. 2009). These same techniques were evaluated here, including derived seafloor complexity metrics from bathymetry and Principal Components Analysis as a basis for segmentation and classification. However, efforts incorporating bathymetry data were largely unsuccessful because artifacts in the data (particularly swath overlap areas) dominated the interpretation of seafloor features. While prominent ledge features (Figure 10, lower patch) are well defined, artifacts created by beam overlap dominated in low relief areas (Figure 10, upper patch).

Recent improvements in data acquisition and processing of ME70 data on the *Pisces* has improved and will greatly reduce these artifacts, making it likely that automated classification processes will be more successful in the future. It is expected that a suite of seascape metrics that are well suited to characterizing seafloor habitats of the South Atlantic Bight can be defined to comprise a standardized classification scheme for this region in the future.

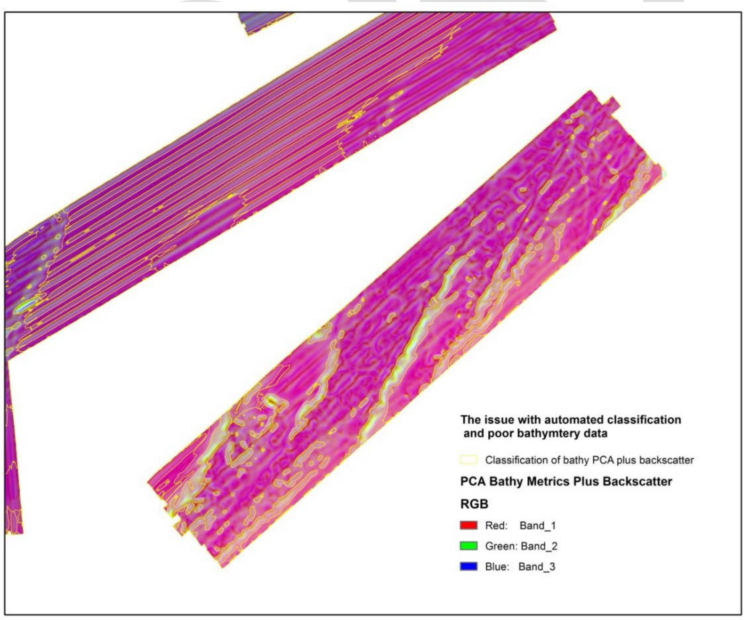


Figure 10. Bathymetric artifacts (upper patch) resulting from beam overlap issues when computing derived landscape complexity metrics.

Habitat Classification - Recommendations

The feature extraction / segmentation approach with manual classification, as well as the unsupervised classification, relies heavily on validation from trap camera data. Approximately 400 sample points for camera data were available for validating the classification of habitat type in the areas mapped here. However, these points were not evenly distributed across all habitat types (Figure 2). In addition, there is some positional uncertainty as the cameras attached to the traps are deployed from the aft deck of the ship while viewing the echosounder display. To provide a more robust set of validation points, a dedicated ground validation survey should be conducted that would allow for drop cameras deployed from a stationary platform to validate habitat type with improved positional accuracy.

While the unsupervised classification used here assigns polygons to class type, it is generated from backscatter alone. Potentially, better descriptors of habitat can be obtained with the incorporation of improved bathymetric data. Even still, the backscatter response may be an appropriate signal to focus on for this area, as there may be a limited number of habitat types within this region. Analysis of backscatter as described here may be sufficient to capture and discriminate those features (along with depth range) that are of interest to SEFIS. A study to better define classes based on backscatter thresholds would be valuable in differentiating fine sand from coarse sediments and mixed sediments and low relief with variable amounts of live bottom cover.

If a standardized classification scheme is to be implemented for the region, habitat variables recorded while reviewing camera observations could be aligned with the new scheme. In addition, an evaluation of current nomenclature used in video review and additional terms useful in defining habitat classes would lend itself to using information from the camera review as a confirmatory dataset.

As data acquisition and processing improves, it is expected that artifacts in the bathymetry data will be reduced. If so, simple landscape metrics (eg. slope of slope) applied to the bathymetry data would enhance visual interpretation of seafloor features for selecting sites for trap deployment.

REFERENCE: Costa, B.M., L.J. Bauer, T.A. Battista, P.W. Mueller and M.E. Monaco. 2009. Moderate-Depth Benthic Habitats of St. John, U.S. Virgin Islands. NOAA Technical Memorandum NOS NCCOS 105. Silver Spring, MD. 57 pp.

ACKNOWLEDGEMENT: Many thanks to Gustav Kagesten, Will Sautter, Bryan Costa, Tim Battista, Warren Mitchell, Nate Bacheler, and Todd Kellison for their assistance in completing this project.

DRAFT

PREDICTIVE MODELING OF HARDBOTTOM OCCURRENCE

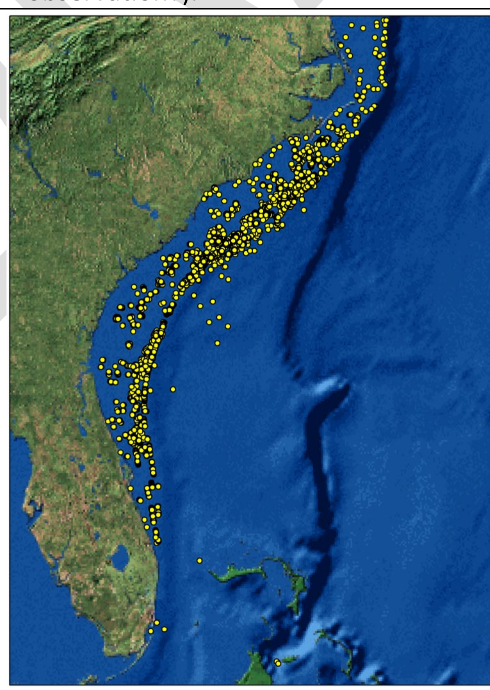
METHODS
Hardbottom Presence Data – Response Variable

Compilation of Hardbottom Presence Records

Existing data from surveys within the area of interest were compiled and merged to be used as the response variable for the predictive model (Table xx).

Table xx. Source data used to create response variable.

Dataset	Source	Description
Trap camera	SEFIS/Marmap NMFS - Beaufort	Habitat classification from underwater video
Universe Stations	SEFIS/Marmap NMFS - Beaufort	SEFIS/Marmap sampling universe deemed "good" hardbottom habitat
Observer Data	SEFIS/Marmap NMFS - Beaufort	Observer data at stations where reef indicator species were collected on commercial fishing vessels (Gulf and Atlantic Fishing Foundation). Entire fish catch and location were recorded.
Catch Data	SEFIS/Marmap NMFS - Beaufort	MARMAP catch data
TNC Hard Bottom	The Nature Conservancy	Assembled from Trawl comments, US seabed hard bottom, Valid hangs See: Greene et al 2010.
Wilmington Side scan	NOAA NCCOS – Beaufort	Wilmington East wind energy area was collected by Chris Taylor – Beaufort. Points indicating potential outcrops were identified based on (visual observation?).



DRAFT

Greene, J.K., M.G. Anderson, J. Odell, and N. Steinberg, eds. 2010. The Northwest Atlantic Marine Ecoregional Assessment: Species, Habitats and Ecosystems. Phase One. The Nature Conservancy, Eastern U.S. Division, Boston, MA. Report.

<u>De-duplication</u>

Prior to input into MaxEnt models, hardbottom presence records were de-duplicated using ENMTools (Warren et al. 2010). De-duplication reduces multiple presences with a model grid cell to a single presence observation. This is logical, as only one presence is required to confirm that a given grid cell contains suitable habitat; more occurrences do not change this conclusion. De-duplication reduces sample size, but helps to reduce the effect of sampling bias in heavily sampled areas.

Environmental Predictor Data Sources and Development

An initial set of 47 environmental data layers were produced for potential use in predictive models of hardbottom occurrence (Appendix X). These datasets included bathymetric depth and depth-derived seafloor complexity metrics, surficial sediment and substrate characteristics, and distance to shore.

<u>Bathymetric Depth</u>

The bathymetric depth dataset was derived by calculating the non-overlapping 4x4 cell block mean of a 92.6625m horizontal grid resolution bathymetry dataset created by merging bathymetry data from the National Oceanic and Atmospheric Administration (NOAA) Coastal Relief Model (CRM) (NOAA National Geophysical Data Center 2009; downloadable at http://www.ngdc.noaa.gov/mgg/coastal/crm.html) and the General Bathymetric Chart of the Oceans (GEBCO) (source; downloadable at http://www.gebco.net/data_and_products/gridded_bathymetry_data/). The block mean was calculated at 92.6625m horizontal grid resolution and bilinearly resampled to 370.65m horizontal grid resolution. The bathymetry grid resolution of 370.65m was chosen to minimize the impacts of vertical and horizontal spatial uncertainty associate with the NOAA CRM in deeper waters (Poti et al. 2012, Calder 2006), at the cost of losing resolution of some fine-scale bathymetric features.

<u>Seafloor Complexity Variables</u>

Although we were unable to use direct measurements of bottom currents as an environmental predictor, we did use a suite of seafloor complexity metrics derived at various spatial scales (370m, 1.5km, 5km, 10km, 20km) as proxies for flow and other oceanographic patterns. These scales were chosen to emphasize terrain features at a range of scales. A 92.6625m horizontal grid resolution smoothed bathymetry dataset at each scale was generated by applying a Gaussian low-pass filter to the merged 92.6625m bathymetry dataset using a custom Matlab script. Slope, slope of slope, and aspect (slope direction) were derived from the merged 92.6625m bathymetry dataset and each of the smoothed bathymetry datasets. Rugosity (ratio of surface to planar area), planiform or plan curvature, profile curvature, and total curvature were derived from the merged 92.6625m bathymetry dataset and each of the smoothed bathymetry datasets using DEM Surface Tools (Jenness 2012). Bathymetric position index (BPI) was calculated from the merged 92.6625m bathymetry dataset at all the spatial scales using an annular neighborhood with an inner radius of 1 cell and an outer radius approximately half the corresponding scales (Wright et al. 2005). The resulting seafloor complexity metrics were all bilinearly resampled and registered to the 370.65m horizontal grid resolution block mean bathymetry dataset. In addition, the non-overlapping 4x4 cell block standard deviation of the merged 92.6625m bathymetry dataset was calculated as a measure of bathymetric variation. Unless otherwise noted, all calculations of seafloor complexity were performed using ArcGIS 10.0 (ESRI 2011).

DRAFT

Substrate Variables

Seabed survey data from the usSEABED Atlantic Coast Offshore Surficial Sediment Data Release, version 1.0 parsed and extracted databases (Reid et al. 2005; downloadable at http://pubs.usgs.gov/ds/2005/118/htmldocs/usseabed.htm) were filtered to remove duplicate points and points not relating to surficial sediments and then modeled using geostatistical methods (following Poti et al. 2012) to produce gridded datasets of surficial sediment mean grain size and sediment composition (percent mud, percent sand, percent gravel). Geostatistical analyses were performed using ArcGIS 10.0 Geostatistical Analyst (source).

Modeling

MaxEnt Model Selection, Interpretation, and Accuracy Assessment

A maximum entropy (MaxEnt) model was fitted and used to predict the likelihood of hardbottom occurrence on our model grid by combining the presence-only hardbottom location data with a suite of environmental predictor datasets (Phillips et al. 2006, Phillips and Dudik 2008). The MaxEnt algorithm estimates the relative likelihood of a grid cell containing suitable conditions for hardbottom occurrence at each location in the output grid, assuming that presence locations take on the most spatially random (uniform) distribution possible under the constraint that for each environmental predictor variable the expected value from the estimated distribution matches its observed mean. We used version 3.3.3k of the MaxEnt Java software available at http://www.cs.princeton.edu/~schapire/maxent/ (Phillips et al. 2004, Phillips et al. 2006), in combination with custom Python, R (R Development Core Team) and ENMTools (Warren et al. 2010) scripts to automate an improved MaxEnt workflow that incorporates stepwise model selection, model accuracy assessment, and calibration of suitability thresholds based on cross-validation performance, as described in the following sections.

Predictor Screening

A pairwise correlation analysis was performed on the full set of potential environmental predictor datasets using the ENMTools software (Warren et al. 2010). Highly correlated pairs of environmental predictors (Spearman rank $R > 0.8$ or $R < -0.8$) were identified and the predictor that was highly correlated with the most other predictors was excluded (decisions summarized in Appendix X).

In addition, all terrain variables derived at the 92.6625m scale and some derived at the 370.65m scale were eliminated from consideration because artifacts of the bathymetry survey track lines were evident. Some terrain variables derived at the 20km scale were eliminated because edge effects were evident from the smoothing of the bathymetric data.

Categorical variables for aspect at the 370 m, 1.5 km and 10 km scales were derived by reclassifying the aspect grids at these scales into the 8 cardinal directions. Profile curvature at the 370 m and 5 km scales and plan curvature at the 370 m, 1.5 km, 5 km, 10 km, and 20 km scales were also reclassified into distinct classes and combined with reclassified slope grids at these scales to create categorical indices. The same process was used to create a categorical index for BPI and slope at the 1.5 km and 10 km scales. This was done because plan and profile curvature and BPI can all take on values of 0 in two ambiguous situations—corresponding, essentially, to flat or vertical surfaces in the measured aspect. Classifying by slope resolves these ambiguities.

As a result of the environmental data screening process, 24 environmental predictors were retained for use in the predictive models (Table X). Predictors in this set were not highly correlated (Appendix

DRAFT

X), though some correlation structure exists and should be borne in mind when attempting to interpret variable importance or selection in any given model.

Stepwise Model Selection Method

Although MaxEnt is more robust to the inclusion of large sets of environmental predictors than regression techniques (Elith et al. 2011), we sought to balance model predictive performance with model complexity since the ability of models to predict habitat suitability can be reduced when models are overly complex or overly simple (Yost et al. 2008, Warren and Seifort 2011). In addition, model interpretability is increased through parsimony (although it is still subject to the caveats of regression analyses regarding causality [Li and Wu 2004]). We, therefore, developed a simple stepwise procedure to improve parsimony of final selected MaxEnt models. While not free from the pitfalls of stepwise procedures (viz., it does not search the entire model space and does not identify an optimal model), it is a simple way to improve parsimony and interpretability of MaxEnt models.

To implement the stepwise procedure (illustrated in Figure 2), model predictive performance was measured using the mean test AUC (i.e., an out-of-bag cross-validation statistic) from repeated bootstrap iteration of a jackknife test of predictor variable importance. Model complexity was estimated using Akaike's information criterion (AICc), corrected for small sample size (Akaike 1974, Burnham and Anderson 2002), as calculated using the free, open-source ENMTools software that implements recent published advances in information theoretic model selection criteria for MaxEnt models (Warren et al. 2010).

An initial model (step 1) was built using the final set of 24 potential environmental predictors. At each iteration of the stepwise procedure, a combination bootstrap and cross-validation procedure allowed for a bootstrap assessment of model parameters and prediction error as well as a cross-validation assessment of model performance (measured by AUC and gain statistics; Phillips et al. 2006, Elith et al. 2011). Ten bootstrap models were trained using 10 randomly selected subsets of 70% of the coral presence points for each group, and then validated by testing predictions on the remaining 30% of data (70%/30% bootstrap cross-validation). The overall in-bag (training) and out-of-bag (test) model performance was estimated by combining results from these 10 bootstrap runs. Final predictions, response curves, in-bag and out-of-bag accuracy statistics and model parameters were estimated from the mean of the bootstrap distribution, with associated standard error (SE) or standard deviation (SD).

For each successive bootstrap iteration within each iteration of the stepwise procedure, jackknife and permutation procedures were also performed to assess the relative importance of environmental covariates to the fitted model in two different ways. This led to estimates (with bootstrap SE) of the mean reduction in test AUC that occurred when each environmental predictor was omitted and the model re-fitted. The environmental predictor to be dropped in the next iteration of the stepwise procedure was chosen to be the one with the smallest reduction in mean test AUC when omitted; i.e. the most redundant variable. The next step of the stepwise procedure began with running the model with a reduced set of candidate environmental predictors (i.e., omitting the most redundant variable). This was repeated until the model was run with a single remaining variable.

To balance model predictive power with model parsimony, stepwise model runs were ranked from 1 to 24 from highest bootstrap mean test AUC (best out-of-bag model predictive performance) to lowest, and 1 to 24 from lowest bootstrap mean AICc (least complex given in bag model fit) to the highest. These two ranks were then averaged, and the model run with the best averaged ranking (1 = best possible, 24 = worst possible) was selected.

Final Model Development

A final full MaxEnt model was then fitted, in which the same covariates present in the model selected by the stepwise procedure were used, but all of the coral presence records were incorporated into the training set. No bootstrapping was performed on this final model. Bootstrap parameter estimates and predictions, derived from the randomly selected subsets of 70% of the data to which the model was previously fitted, are assumed to be conservative estimates of the accuracy of all parameters, statistics, and predictions produced by the final full model, because these bootstrap estimates are derived from smaller samples.

As a final result, the full model uses a logistic transformation of the MaxEnt “raw” (i.e., exponential) output to produce an index ranging from 0 to 1. This logistic output, the default of the MaxEnt program, is related to the likelihood of habitat suitability, but is not a probability (Elith et al. 2011). It assumes that the global prevalence of the taxon being modeled is known *a priori*, and is exactly equal to 0.5. The MaxEnt logistic value can be treated as a probability and can be compared across multiple MaxEnt models for different taxa *if and only if* these extremely restrictive and, in practice, impossible assumptions hold. Therefore, it is important to treat the MaxEnt logistic value *only as a relative measure of suitability among pixels in the same model domain for the same taxon*. The mistakes of treating a MaxEnt logistic value as a probability or presenting logistic maps of different groups alongside for comparison are quite common in the peer-reviewed literature, but all such uses of the logistic value are invalid unless the global prevalence is known with certainty from independent data (Elith et al. 2011). Global prevalence can never be estimated from presence-only data alone.

Thresholded Likelihood Class Calibration

To circumvent the limitations on interpretation of the MaxEnt logistic value for presence-only data discussed above, we developed a simple but novel cross-validation calibration procedure to determine appropriate thresholds of the MaxEnt logistic value that defined hardbottom occurrence likelihood classes that had the advantage of having defined out-of-bag performance characteristics.

Using the 30% test data sets from each bootstrap/cross-validation replicate, we used R package ‘ROCR’ (Sing et al. 2005) to calculate receiver operating characteristic (ROC) curves and find the empirical optimum threshold of the MaxEnt logistic output, i.e. the threshold that minimized false positive rate (FPR) and false negative rate (FNR) for predictions on the test dataset. This implicitly assumes that false positive and false negative errors are equally costly (FPR cost:FNR cost=1:1). We also calculated the empirical optimum thresholds assuming false positives were less costly (FPR cost:FNR cost=1:50,1:30,1:20,1:15,1:10,1:8,1:5,1:3,1:2). Then, we recalculated the empirical optimum thresholds assuming false positives were increasingly more costly (FPR cost:FNR cost=2:1, 3:1,5:1,8:1, 10:1,15:1,20:1), resulting in thresholds that yield more and more conservative predictions of likely suitable habitat. This process was repeated for each bootstrap/cross-validation replicate and each FPR:FNR ratio, and the mean empirical optimum threshold for each cost ratio was calculated over the 10 bootstrap/cross-validation replicates. Thresholds were found to be consistent across replicates as indicated by low standard errors (Appendix X). Thus the mean empirical optimum thresholds for each cost ratio were used to convert MaxEnt logistic outputs into hardbottom occurrence likelihood classes.

Accuracy, Performance, and Uncertainty Assessments

In-bag model fit statistics (training AUC, training gain) and associated standard errors were calculated from the 10 randomly selected subsets of 70% of the data used for model training during the bootstrap. Out-of-bag model fit statistics (test AUC, test gain) and associated standard errors were calculated from the 10 randomly selected subsets of 30% of the data used for cross-validation during the bootstrap. The AUC statistic has been extensively discussed elsewhere (Fielding and Bell 1997). The gain is the penalized likelihood function that MaxEnt minimizes as part of its model-fitting process (Phillips et al. 2004). Exponentiating the gain gives the likelihood ratio of an average presence to an average background point (Elith et al. 2011). For example, a hypothetical test gain of 1.5 would tell us that the average presence point in the test dataset for that group was exp(1.5)=4.5 times more likely to be identified as a presence than as a background point. The higher the test gain, the better the model is, on average, at discriminating suitable habitat from the background when tested on out-of-bag data.

Spatially explicit uncertainty in model prediction was estimated by mapping the bootstrap standard error of the MaxEnt logistic value. Bootstrap standard errors were also used to represent uncertainty in estimates of all model parameters and performance statistics, and on response curves shown in Appendix X (Figures …).

Response Curves

We examined two types of response curves to evaluate the effect of a given predictor variable on likelihood of hardbottom occurrence. First, for each bootstrap replicate, a separate single variable model was fitted using each predictor variable alone, and the corresponding response curve calculated. In Appendix X, Figures X.1 to X.9 show bootstrap mean single-variable response curves on the top sub-panel, with associated bootstrap standard deviation reflecting differences in the single-variable response curve across bootstrap replicates.

Second, we examined the model response to each variable in turn, when all the other variables in the model were held constant at their average values. This leads to a marginal response curve that is influenced by any interactions between the variable of interest and all other variables in the model. In Appendix X, bootstrap mean marginal response curves (±1 SD) are shown in the middle sub-panel of each figure, for each variable included in each group model.

To gain an intuitive understanding of the response to a given variable, it is sometimes helpful to compare the relative frequency histogram of all values of an environmental variable throughout the domain (black bars in lowermost panels in Appendix X) to the relative frequency histogram of the environmental variable looking only at known hardbottom areas (i.e., presence-point locations; red bars in lowermost panels in Appendix X). Comparison of bean plots of environmental variable values at hardbottom presence points with the background distribution of the variable in the study area can also be informative (e.g., Davies and Guinotte 2011; see Appendix X). Beanplots were generated using the R package 'beanplot' (Kampstra 2008).

Variable Importance

MaxEnt provides three post-hoc assessments of the relative importance of predictor variables. First, the MaxEnt program provides a summary of how much each predictor contributes to the gain of the model, accumulated for each predictor over the course of the training algorithm. Second, the MaxEnt program randomly permutes the values of each predictor (one at a time) and determines the resulting decrease in the area under curve (AUC) statistic of the training model receiver operating characteristic (ROC) curve. This provides a measure of how strongly the model depends on each predictor. Third, the MaxEnt program estimates predictor importance using a jackknife approach, in which it re-runs the model for each predictor, first building the model with all variables except the predictor of interest, and then building the model with only the predictor of interest. If a predictor is highly correlated with other predictors, withholding it will have little impact on model performance. Therefore, an important and non-redundant predictor will have high explanatory power by itself and its omission from the model will result in a significant reduction in predictive power.

RESULTS

Hardbottom Presence Data

Candidate Set of Environmental Predictor Variables

The final candidate set of potential environmental predictors included 24 variables (Table X, Appendix X). Intercorrelation among these variables was present but generally low, with almost 90% of the pairwise Spearman rand correlations having a magnitude < 0.5 (Table X) due to the screening process (Table X). In total, 20 geomorphological (or terrain) variables, 3 substrate- (i.e., sediment) related variables, and one distance-based variable were included. Terrain variables included 5 different spatial scales of features (370m, 1.5km, 5km, 10km, 20km).

MaxEnt Model Results

Model Selection

Figure X depicts the stepwise model selection procedure using replicated (n=10) 70% training/30% test random subsample bootstrap/cross-validations. Generally, one would expect the out-of-bag predictive performance (test AUC) to increase slightly or stay the same as the first few variables are dropped, then peak and begin to decline rapidly in the last few steps when critical, non-redundant variables are dropped. In contrast, the model fit to training data relative to model complexity, as measured by AICc, would be expected to rapidly decrease in the first few steps as redundant variables are eliminated and parsimony increases. AICc then would either reach a minimum (indicating optimal model fit vs. complexity for the training dataset) or remain flat in the latter steps of the procedure, depending on the balance between numbers of parameters and fit to training data at each step. The red boxes indicate the selected model run, based on the average ranking according to these two metrics (high AUC and low AICc). This stepwise selected model included 22 of the 24 environmental predictor variables. Given the bootstrap standard errors on test AUC and training AICc, it is apparent that multiple models in the stepwise procedure would satisfy the general criterion of having high test AUC and low AICc. The average rank algorithm provides a consistent and unambiguous way to select a model, but in some cases several models have similar AUC and AICc statistics, likely due to the correlation structure among variables that leads to redundancy. In these cases it may be desirable for experts to choose one of the equivalent models based on other qualities, such as confidence and availability of the environmental predictors involved.

Performance Assessment of Selected Model

The stepwise selected model performed well in both out-of-bag (cross-validation) model performance statistics, with a test AUC of 0.845 (AUC values >0.5 indicate some predictive power and >0.7 indicate good performance; Fielding and Bell 1997) and a test gain of 0.842 (2.32x discrimination).

Final Full Model

Following model selection, the selected predictor variable subset was used to fit another, final MaxEnt model using 100% of the hardbottom presence data. This is termed the final, full model.

Based on the final full model, we mapped the MaxEnt logistic relative likelihood of hardbottom occurrence (Figure X). It is important to note that the MaxEnt logistic value is not a probability.

Threshold Calibration

The cross-validation threshold calibration procedure described above was used to derive thresholds that define hardbottom occurrence likelihood classes. Thresholds used were the averages based on 10 bootstrap/cross-validation replicates, and were found to be highly consistent across replicates (all bootstrap standard errors were <0.011422 on the MaxEnt logistic scale and 50% of the bootstrap standard errors were <0.004405 on the logistic scale). Threshold values ranged from 0.01325 (FPR cost:FNR cost=1:50) to 0.649612 (FPR cost:FNR cost=20:1).

Thresholded Final Full Model

Using the calibrated thresholds, predicted hardbottom occurrence likelihood classes were mapped (Figure X).

Variable Importance, Response Curves, and Beanplots

Table X shows three measures of variable importance: permutation tests, single-variable jackknife importance tests, and omission jackknife importance tests. Each emphasizes a different aspect of variable importance: leverage on model (permutation), amount of suitable habitat predicted solely by this variable (single-variable jackknife), and non-redundancy of the variable (omission jackknife). In Table X, the top three variables by each metric are shown in bold font, with the most important indicated by a superscript symbol (Table X).

DRAFT

APPENDICES

Appendix 1. Table 1. Surfaces used to create backscatter, bathymetry layers and habitat maps.

	Beam configuration used during data acquisition Rc= Randy Cutter Tw= Tom Weber	Bathymetric Layer Name (GRID 32 bit float)	Backscatter Layer Name (GRID 32 bit float *_Flt1) (TIFF 8 bit unsigned Int *_GS)
1	rc	box_01_3m_ban	Box_01_Bsc_Flt1
2	tw	box_06_3m_ban	Box_06_Bsc_Flt1
3	tw	box_14_3m_ban	Box_14_Bsc_Flt1
4	tw	box_15_4m_ban	Box_15_new_Bsc_Flt1
5	tw	box_32_4m_ban	Box_32_Bsc_Flt1
6	rc	box_35_3m_ban	Box_35_R_Bsc_Flt1
7	tw	box_36_3m_ban	Box_36_Bsc_Flt1 *Box_36a_GS.tif*
8	tw	nb_37N_3m_ban	Bx_37N_Bsc_Flt1
9	tw	n_37se_3m_ban	Bx_37se_Bsc_Flt1
10	rc	box_40_3m_ban	Box_40_Bsc_Flt1
11	tw	bx_40S_3m_ban	Box40S_Bsc_Flt1
12	rc	bx_42_4m_ban	Box_42_Bsc_flt1
13	tw	Pork_4m_ban	Porkins_Bsc_Flt1
14	tw	No1_4m_ban	Box_No1_Bsc_Flt1
15	rc	bx_02b_3m_ban	Box_02_Bsc_Flt1
16	rc	box_3_3m_ban	Box_03_WS_Bsc_Flt1 *(GS) box3_3m_GS*
17	tw	box_4_4m_ban	ox_04_Bsc_Flt1
18	tw	no14_4m_ban	Box_No14_Bsc_Flt1
19	tw	nNEo14_4m_ban	Box_NEo14_Bsc_Flt1
20	tw	sse17_4m_ban	Box_SEo17_Bsc_Flt1
21	rc	box_23_3m_ban	Box_23_Bsc_Flt1
22	rc	box_31_3m_ban	Box_31_Bsc_Flt1
23	rc	b_42x2_4m_ban	Box_42x2_Bsc_Flt1
24	tw	b1941_3m_ban	Box_1941_Bsc_Flt1
25	tw	b_3043_4m_ban	Box_3043_Redo_Bsc_Flt1
26	tw	SSEo0_3m_ban	Box_SSEo0_Bsc_Flt1
27	tw	Yodas_4m_ban	Box_Yodas_Bsc_Flt1
28	tw	LilCl_4m_ban	Box_LittleClam_Bsc_Flt1
29	tw	EoMon_4m_ban	EofMonitor_Bsc_Flt1
30	tw	LoRR_4m_ban	Lone_Ranger_Ridge_Bsc_Flt1

DRAFT

Appendix 2. – PROCESSING STEPS

Onboard Processing of bathymetry data in Caris and backscatter data in Fledermaus Geocoder Tool (FMGT) was completed by the 2012 SEFIS Team led by Warren Mitchell, who can provide SOPs.

1. The first step was to conduct a qualitative review of all data to look for obvious artifacts and inconsistencies.

2. Due to differences in beam configurations during data acquisition (shelf = Randy Cutter; shelf break = Tom Weber), only those collected with Tom Weber's configuration were classified here (although data from all areas were organized and reprocessed).

3. Most of the original surfaces had a large degree of overlap and 'smile' artifacts on the outer beams, as well as on nadir. Therefore, beams were trimmed during reprocessing of data (beams less than 2 degrees and greater than 60 degrees were disregarded).

4. Post-processing for the purpose of creating hard bottom maps included reprocessing backscatter data in FMGT with the parameters described here (Appendix 2. Table 1).

5. All bathymetry layers were rebuilt from ASCII text files generated from the Caris process by converting a point file to a raster in ArcGIS.

6. For both backscatter and bathymetry layers, masks were created to clip NoData from rasters (Spatial Analyst Tools – Extraction – Extract by Mask), small holes were filled with mean values at the Box level, and rasters snapped to a common layer to align surfaces. Appendix 1 Table 2 lists the names of the final clean boxes.

7. A Gaussian smooth filter was applied to both surface types after mosaicking into one surface.

8. Backscatter and depth range was further classified (backscatter threshold, automatic segmentation, and unsupervised classification) as described in this report.

DRAFT

Appendix 2.

Table 1. Post-processing of .gsf files in Fledermaus Geocoder Tool to recreate backscatter surfaces

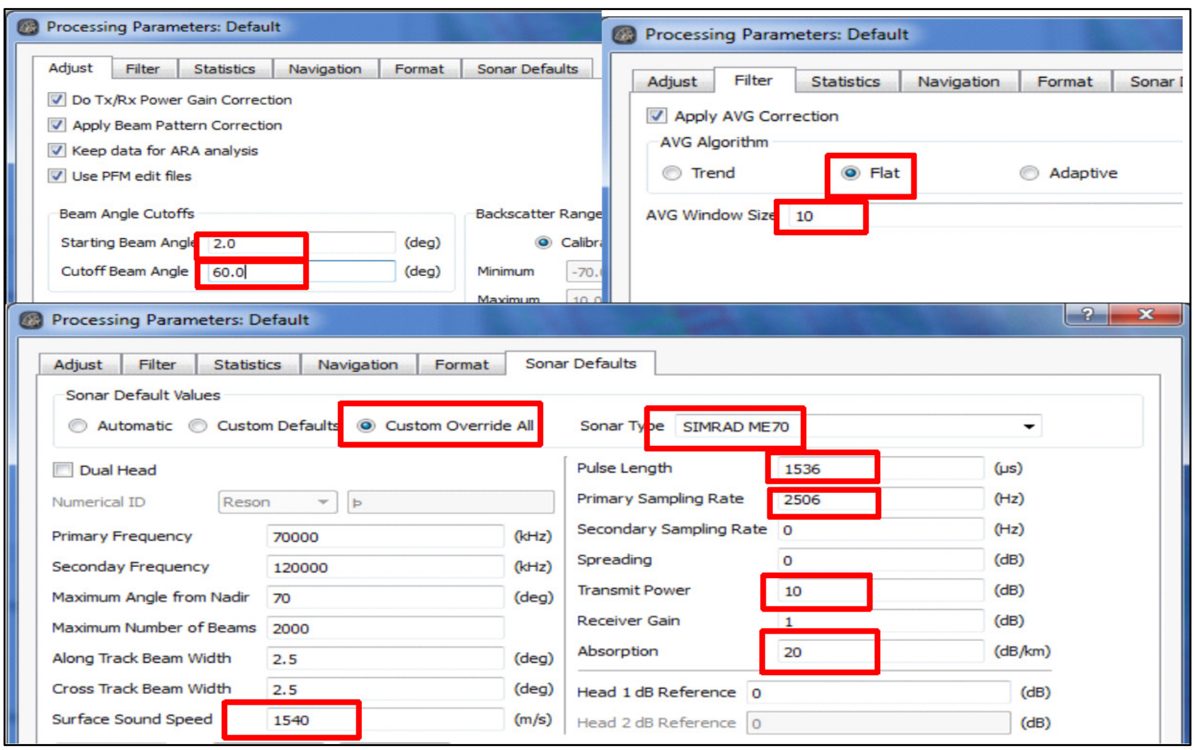
1. File –> Create Project. Leave auto compute output coordinate system
2. Adjust parameters.

    Settings –> Mosaic Parameters. Line blending 50%, No Nadir if possible, Fill gaps using adjacency

    Settings -> Processing Parameters -> Edit default.

    Change values under Adjust, Filter, and Sonar defaults Tabs as per screen captures.
3. Add Source files.

    File -> Add source files.   Find files, OK.
4. Mosaic (Manual)
5. Export as Surface (Tiff, Grid).



Appendix 3. SEFIS input into final classes

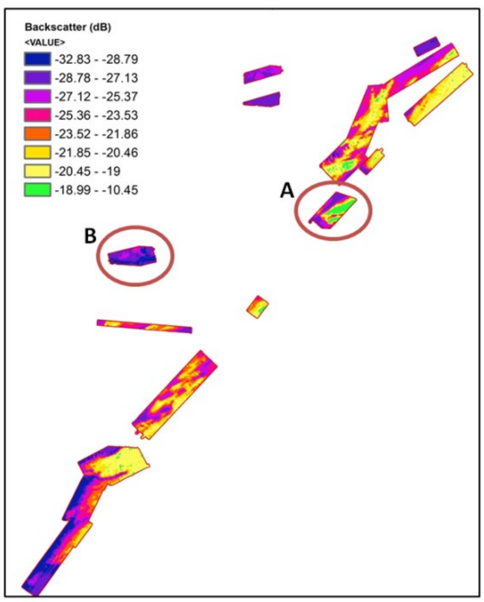
Email response regarding SEFIS PC1204 classification . **9/24/14**
From: Warren Mitchell - NOAA Affiliate warren.mitchell@noaa.gov
To: Laura Kracker - NOAA Federal laura.kracker@noaa.gov
cc: Nathan Bacheler - NOAA Federal nate.bacheler@noaa.gov

Hi Laura:

Here's the requested response from Beaufort. Thanks for sending the additional figures. Nate is Cc'd. I'll distribute a summary to Chris and Todd once I've fielded any questions generated here.

In sum: We fully support the unsupervised classification approach. Folks here prefer three classes of backscatter (e.g., Fig 8 in your draft report, slide 6 in the ppt sent this summer), and support the delineation of relief based on bathymetry (number and scale of relief categories to be determined, see *****below).

Green pixels: Green pixels in Backscatter_Scale.ppt make intuitive sense to us. We feel pretty confident the pixels depict locations of "robust" hardbottom. Looking at the screen capture below, we understand the dense green in A to be a broad area of hardbottom called, fortuitously, the "rock pile." And the relief category in slide three seems to contribute helpful, additional information. In contrast, we'd guess area B would generate neither green pixels nor significant relief. That region was very low relief, with primarily sand substrate and sparse attached biota. Region B is a location where we assume hardbottom, where present, is covered by sediment, and therefore generating weaker backscatter returns.



***Relief categories**: Nate and I are tempted to simplify relief presence to one category. Opinion is based video viewing experiences, where diverse relief is visible in a single camera view (e.g., piece of a ledge, sand patches, a large sponge, some cobbly rocks, all in the same view). Given the 4x4 m pixel size, would it make sense to define just one relief category where we'd all agree relief is moderate through high? Thinking of the video habitat classification variables, it seems signal from less rugose "low relief" habitats (e.g., Relief < 0.3 m, Biota height < 0.5 m) could get lost in noise due to grid cell summations and comparisons with variable neighboring cells. Overall, it'd seem most trustworthy to work with a more conservative, more likely predictor of relief presence. But I'm not sure what *minimum* height to put on that relief range. 0.5, or 0.75, or 1.0 m? Thoughts? We'd follow your instincts here.

Naming: We'd prefer generic titles for classifications, especially dropping references to sediment sizes. For example, use "transitional" for the mid-category and drop the "course sand, gravel" found in your draft report's Table 1. Thinking about your question, "Does the delineation identify 'real' features?", we're feeling good about the boundary delineations, but think the variability of grid cells inside classifications are tricky to characterize. Perhaps Class 1 unconsolidated, Class 2 transitional, Class 3 hard bottom, or something akin?

In general, we'd rather over-simplify in cases where data are ambiguous or conclusions are more difficult to draw. Feel free to apply that stance in future decisions needed to wrap up the reporting.

Notes from yesterday are attached. Feel free to edit, distribute, or set aside for reference.
--
Warren Mitchell
JHT Contract Fisheries Biologist
Habitat Mapping Lead, Fisheries Ecosystems Branch - SEFIS Group
NOAA Fisheries, Beaufort Laboratory
warren.mitchell@noaa.gov
252.728.8755

Appendix 4. Relationship of the current classification scheme to the Coastal and Marine Ecological Classification Standard (CMECS).

Biogeographic Setting. Realm: Temperate Northern Atlantic. Province: Warm Temperate Northwest Atlantic. Ecoregion: Carolinian

Aquatic Setting. System: Marine. Subsystem: Marine Offshore. Tidal Zone: Marine Offshore Subtidal.

Water Column Component. Water Column Layer: Marine Offshore Lower Water Column. Salinity Regime: Euhaline. Temperature Regime: Warm.

Geoform Component. Tectonic Setting: Passive Continental Margin. Physiographic Setting: Continental/Island Shelf. Geoform Origin: Geologic.

Substrate Component. Substrate Origins: Geologic Substrate, Biogenic Substrate. Substrate Classes: Rock Substrate, Unconsolidated Mineral Substrate, Coral Substrate. Substrate Subclasses: Bedrock, Coarse Unconsolidated Substrate, Fine Unconsolidated Substrate.

Biotic Component. Biotic Setting: Benthic/Attached Biota. Biotic Classes: Reef Biota, Faunal Bed

DRAFT