Integrating Fish Trawl, Water Quality, and Benthic Mapping Data in the Peconic Estuary

Final Report to
New York State Department of Environmental Conservation

by
Robert Cerrato, Roger Flood, Michael Frisk, Adrian Jordaan, Tyler Abruzzo, Alison Flanagan, and Michael White

School of Marine and Atmospheric Sciences
Stony Brook University
Stony Brook, NY 11794-5000

Introduction

A number of extensive environmental and faunal data sets are available for the Peconic Estuary system. These include high resolution multibeam and side-scan sonar (e.g., Flood 2004), grain-size and benthic fauna data (e.g., Cerrato and Maher 2007, Cerrato et al 2009, 2010), water quality data (e.g., Nuzzi and Waters 1998), and finfish and mobile invertebrates (e.g., Weber et al. 1998). The availability of these data sets presents a unique opportunity to examine and better understand habitat-animal relationships at the ecosystem level in the Peconics.

The overall goal of this project was to carry out an integrated analysis of existing finfish, benthic fauna, water quality, grain-size, and sonar data to better understand habitat-animal relationships in the Peconic Estuary system. Specific objectives included: (a) assemble and georeference existing data, (b) develop and compare analytical techniques for classifying bottom type based sonar data, (c) conduct a spatial/temporal analysis of existing trawl survey data, and (d) link existing faunal and environmental data in an integrated analysis. Much of this work was carried out as part of student thesis research and draws much of its content from Abruzzo (2015), Flanagan (2016), and White (2015). This report is structured around the original project objectives, but additional information and detail about a topic may be found in the student theses.

Objective 1: Assemble and georeference existing data for the Peconics

Existing trawl data, benthic faunal data, multibeam and side scan data, grain-size data, and water quality data were assembled into spreadsheets using a common set of identifiers. Data were then georeferenced and loaded into a GIS. This objective was straightforward but very labor intensive. Spreadsheets and geodatabases formed the fundamental data source for subsequent analyses.

Methods:

Trawl data was obtained from the New York State Department of Environmental Conservation (NYSDEC). NYSDEC has conducted a fishery-independent trawl survey since 1987 in accordance with the Atlantic States Marine Fisheries Commission's (ASMFC) Fishery Management Plan for weakfish. The survey included not only weakfish but all finfish and mobile invertebrates captured and has been used to evaluate the abundance of commercially important species (Weber et al. 1998). The survey area consisted of Flanders Bay, Great Peconic Bay, Little Peconic Bay, and several smaller bays (Southold Bay, Noyack Bay, and Shelter Island Sound) bordering Shelter Island not including the North and South channels around Shelter Island. Allocation of stations is based on 77 1-minute latitude and 1-minute longitude sampling blocks (Figure 1-1). Each week from May through October, the survey samples 16 randomly chosen stations, with some annual start dates beginning at the end of April and some end dates spilling over into November. Exceptions to this design plan included no trawls in 2005, no trawls until mid-July in 2006, the beginning of August in 2008, and the beginning of June in 2010. From 1987-2012, a 10.7 meter workboat, the David H. Wallace, collected samples using a 4.9 meter semi-balloon otter trawl with a 3.2cm mesh codend and a small mesh liner (1.3 cm). Tows were set for 10 minutes at an approximate speed of 2.5 knots. For more information on gear and survey design see Weber et al. (1998).

The general start locations of each trawl usually occurred at the center of each trawl sampling block. From 1987 to 1991, starting and ending locations were recorded using Loran C navigation. In 1992, SatNav which also recorded Loran coordinates was added and was used as the primary navigation system. Therefore, Loran coordinates were recorded for each trawl from 1987-2000. Loran coordinates from 1987-2000 were converted to latitude and longitude by using Andren SeaMarks 8.2 (Andren Software, Indialantic FL). After 2000, a GPS navigation system was used which recorded locations of each trawl in latitude and longitude. At the start of each tow, surface and bottom temperature, depth, salinity, dissolved oxygen and secchi disc depth were recorded. Depth was also recorded at the end of a tow.

Chlorophyll measurements from 1987-2012 were obtained from the Suffolk County Department of Health Services (SCDHS 2011). Stations 60113, 60114, 60130 and 60170 were sampled continuously from 1987-2012; therefore only chlorophyll measurements from these stations were used (Figure 1-2). Typically, the SCDHS measured fluorescence using a YSI 6600 Probe 6025. For more information, chlorophyll calibration, and QA/QC, see SCDHS (2010).

Two climatic indices were also used in the analysis: the Atlantic Multi-Decadal Oscillation Index (AMO) (Figure 1-3) and the North Atlantic Oscillation Index (NAO) (Figure 1-4). These indices were obtained from the National Oceanic and Atmospheric Administration at the following addresses:

ftp://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/nao_index.tim

http://www.esrl.noaa.gov/psd/data/correlation/amon.us.data

For more information on these indices see www.esrl.noaa.gov and www.cpc.noaa.gov.

Sonar and grain-size data were obtained from a series of projects whose goal was to map the habitats of the Peconic Estuary (Flood 2004, Cerrato and Maher 2007, Cerrato et al 2009, 2010). Sonar survey data were visually segmented into acoustic provinces using backscatter as a proxy for bottom type (Figure 1-5a, b). Sampling stations were randomly positioned within each acoustic province (Figure 1-5c). Bottom sampling was conducted aboard the R/V Pritchard operated by Stony Brook University using a modified van Veen grab (0.04 m^2). Sediment grain-size analyses measured percent composition by weight of major size-fractions (gravel, sand, silt, clay). A total of 380 sediment samples from 315 stations were analyzed for grain-size.

Results:

Trawl start and end locations were entered into a spreadsheet for 9,511 trawl samples collected from 1987-2012. Approximately two-thirds of these locations were listed on hardcopy data sheets as Loran coordinates. These were converted using Andren SeaMarks 8.2. Location data were georeferenced and locations verified by comparing results to the trawl sampling block. Average grain size characteristics attributed to each acoustic province were assigned to each trawl whose midpoint occurred within the bottom type (6,496 trawl tows).

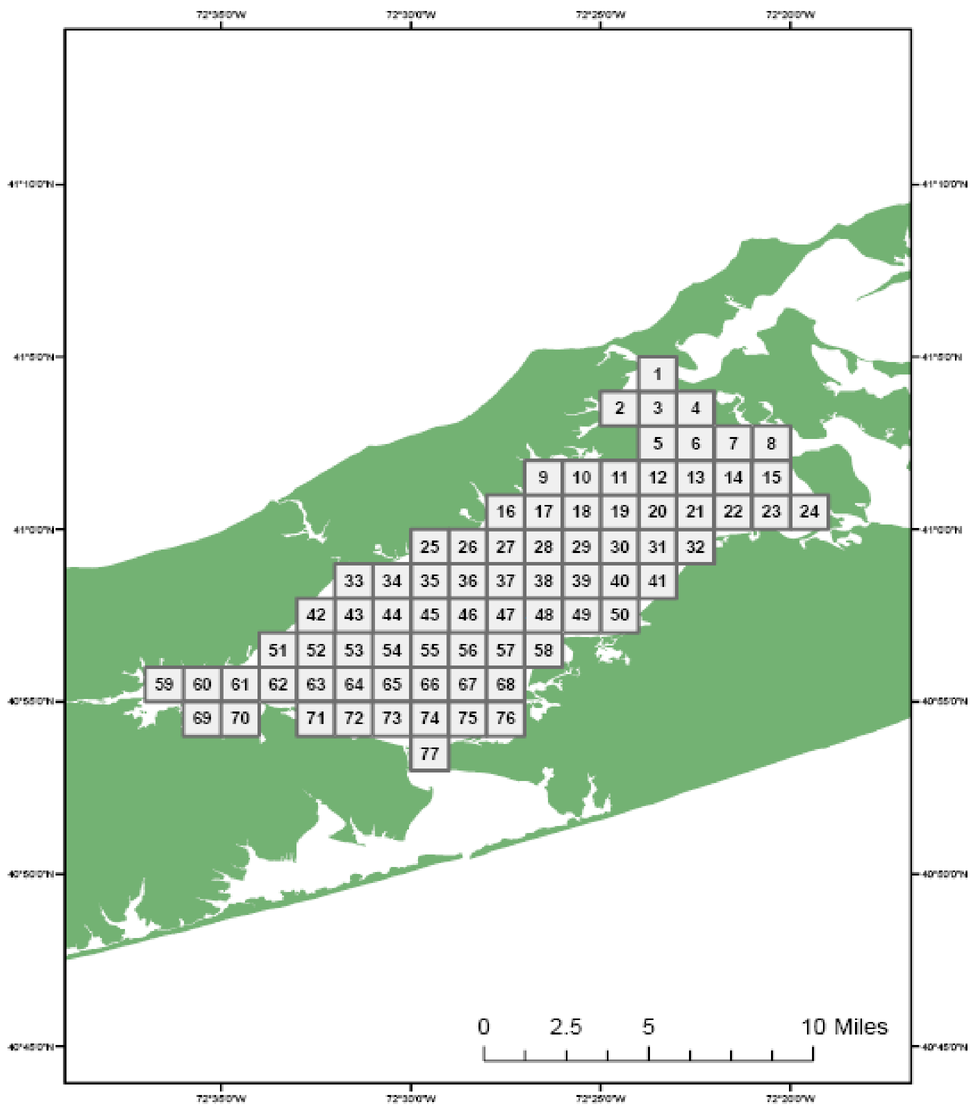

Figure 1-1. Survey stations used in the Peconic Trawl Survey. Stations were based on a 1' latitude and 1' longitude grid. Recreated from Weber et al. (1998).

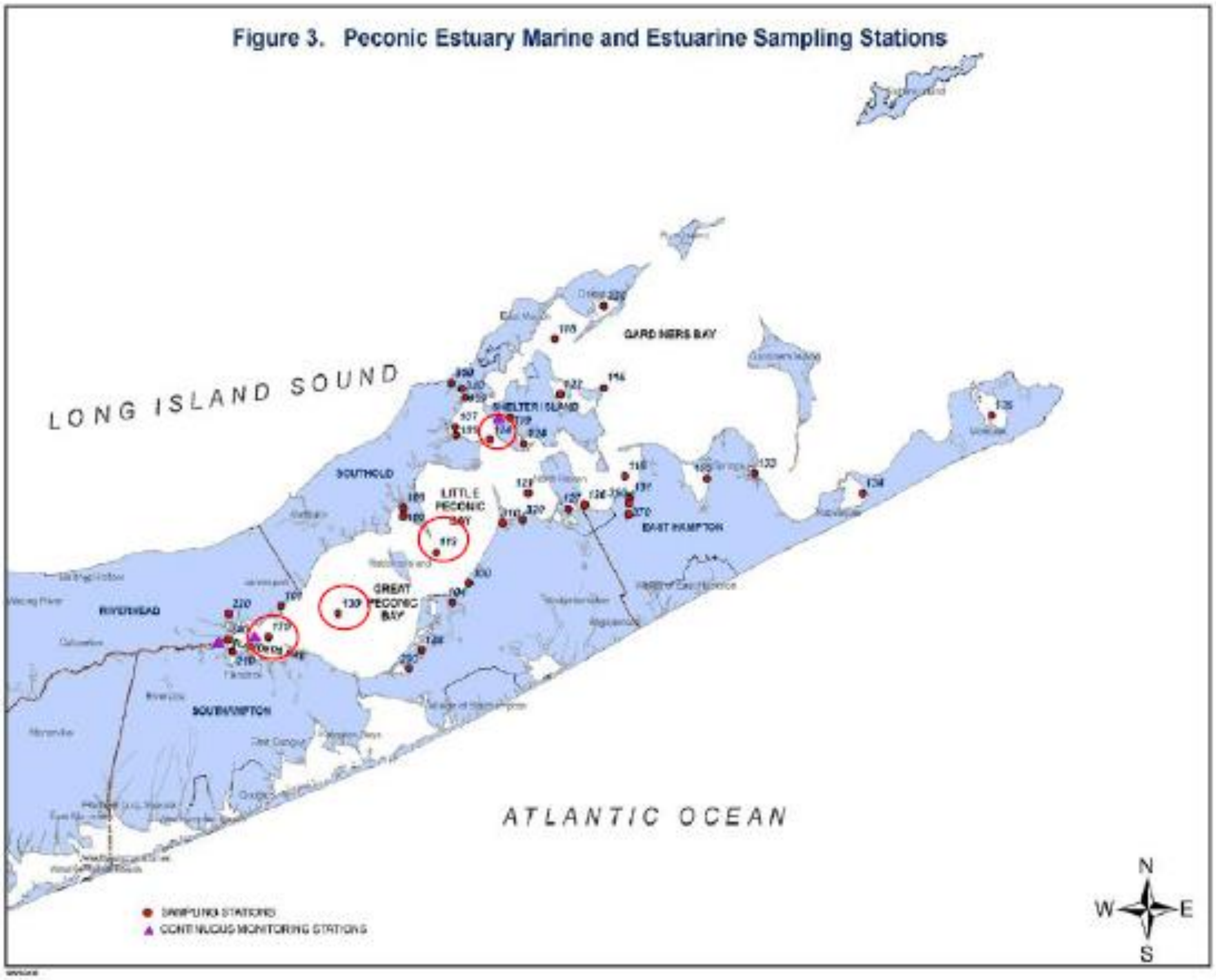

Figure 1-2. Map of sites where water quality sampling is performed by Suffolk County Department of Health Services. Chlorophyll values from stations circled in red were used in the current study (St. 133, 114, 130 and 170). Map taken from SCDHS (2011).

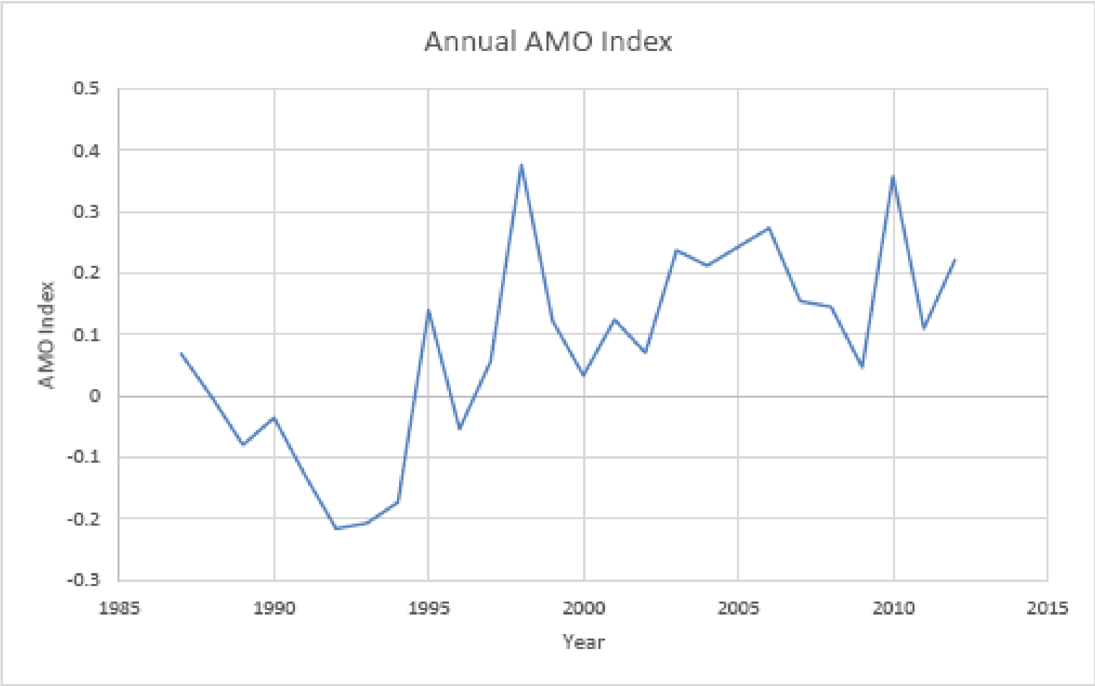

Figure 1-3. Annual AMO Index. Averaged from monthly values taken from the National Oceanic & Atmospheric Association (NOAA).

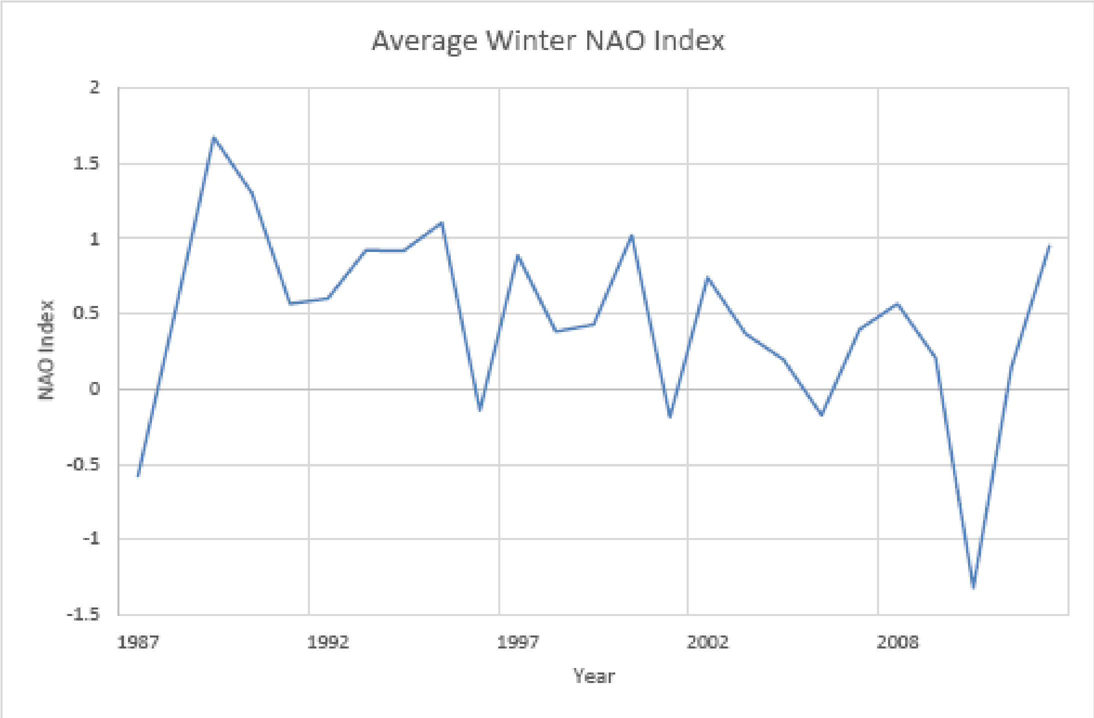

Figure 1-4. Winter (Jan-Mar) NAO Index. Averaged from monthly values taken from the National Oceanic & Atmospheric Association (NOAA).

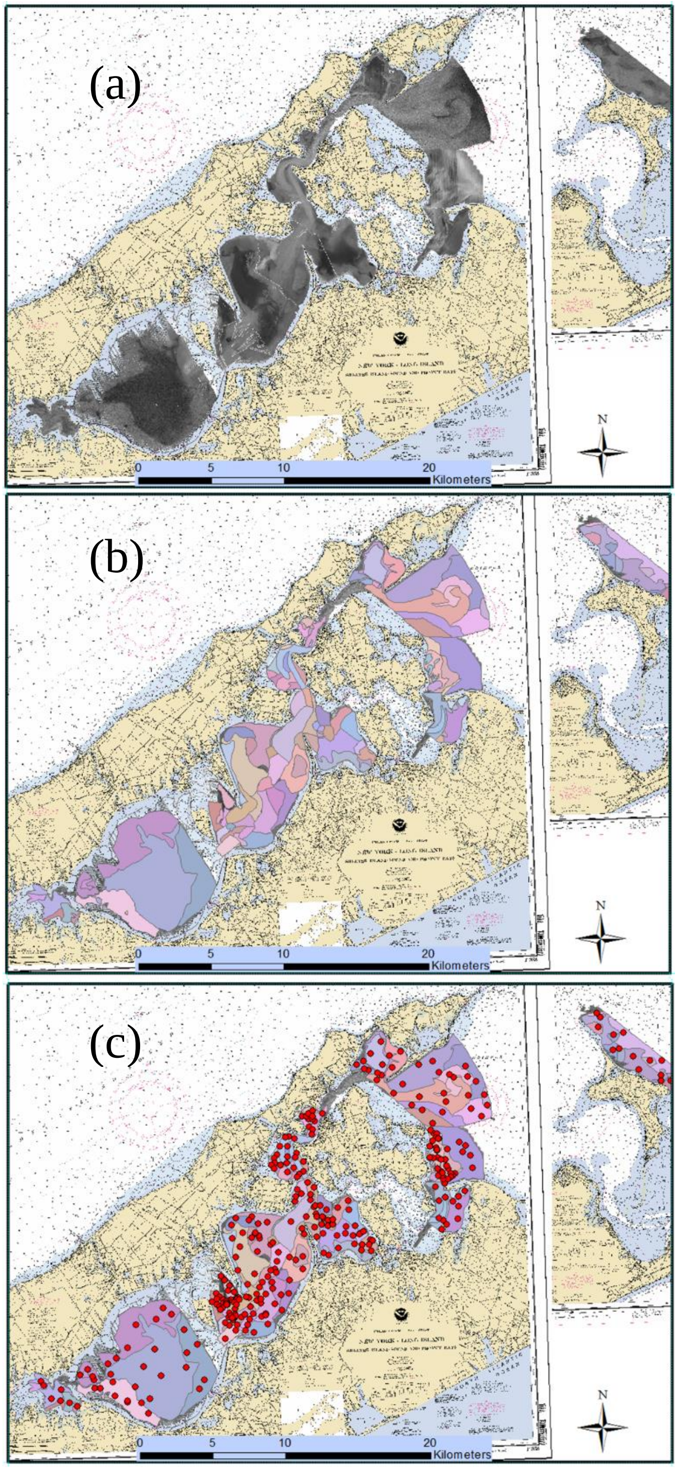

Figure 1-5. Sonar, acoustic provinces, and sampling locations from the benthic habitat mapping project (Flood 2004, Cerrato and Maher 2007, Cerrato et al 2009, 2010). (a) Sonar backscatter. (b) Acoustic provinces. (c) Sampling locations.

Objective 2: Develop and compare analytical techniques for classifying bottom type based sonar data

A variety of techniques were explored to classify bottom type based on multibeam and side scan sonar data. These included attempts to improve the quality of sonar mosaics, examining new techniques to segment the seafloor into homogeneous areas, deriving additional parameters from backscatter data, and developing and utilizing a statistical approach that explicitly involved spatial and temporal variation.

Task 1: Improve the quality of sonar mosaics

The initial objective of this task was to reprocess old sonar data to obtain higher resolution, higher quality bottom maps. New developments in the processing and analysis of backscatter had occurred since the original acquisition of the sonar data. In particular, the Fledermaus software suite from Quality Positioning Services (QPS) was thought to have the potential to significantly improve the quality of the backscatter mosaics created and allow additional sonar parameters to be derived from the backscatter data. This analysis focused on multibeam data, and used data sets from Great Peconic Bay and Little Peconic Bay as examples. It tested the idea that successful reprocessing of older surveys using new techniques was possible, and if validated, a substantial amount of archived acoustic data could provide valuable new information to researchers. Full details of this process can be found in White (2015) (Appendix A).

Methods:

The Fledermaus software suite from QPS (104 Congress St., Portsmouth, NH) contains several modules for interactive geospatial processing, visualization, and analysis of sonar data. This software suite has several advantages for the purpose of acoustic analysis, including flexible but advanced backscatter mosaic creation and wizard-based importing and exporting of data and products.

While initial reprocessing using this software seemed promising, certain obstacles began to present themselves when applied to the multibeam data for Great Peconic Bay and Little Peconic Bay. Certain parts of the data in the layers created were being left out indiscriminately because the computers using the new software simply did not have enough processing memory. Transferring the data to a new workstation helped resolve this problem. After improving the hardware associated with reprocessing, there continued to be issues loading raw data into the Fledermaus modules. The multibeam data were collected with a Kongsberg dual head Simrad 3000 multibeam that writes raw.all files during data acquisition. The dual head raw.all files, however, were not compatible with any Fledermaus modules. As a result, half of the survey area was deleted during loading, erroneous backscatter values and incorrect depth calculations occurred (e.g., >6000m in the Peconic Estuary), and the software effectively ceased all reprocessing of archived multibeam data.

These problems were attributed eventually to data reading errors in the software. Communication with QPS and troubleshooting determined that the dual head files were being read incorrectly. Ultimately, the software did not count heads but individual beams. The Feldermaus software had not been used to process the specific set-up and datagram structure of the Kongsberg instrument used in the Peconics to acquire the sonar data. QPS developers fixed the bug and issued a new software version that was released to all users. Once this version was installed, dual head Simrad 3000 data were successfully reprocessed.

Each acoustic data set went through a rigorous manual editing process. Track lines were manually turned on and off in the software and evaluated to determine if they contained quality backscatter data. During acoustic surveys, erroneous lines can occur as a result of many variables, such as gas bubbles under the transducer, transit lines from port to survey areas where the ship was traveling too fast to acquire good quality data, and obstacles like crab traps in the track line that obstruct the survey. These can produce artifacts. By comparing individual track lines to their neighbors and examining the raw image using the Beam Pattern Viewer in the software, erroneous lines were identified and removed. Lines were blended at 50% based on dB mean and the option of "no nadir is possible" was selected for the smoothing of near nadir values. Several mosaics were created for both Great Peconic Bay and the Little Peconic Bay. The final mosaics were chosen based on visual comparison and exported as SD files. These are Fledermaus specific viewing files which contain the raster data. In the Fledermaus viewer module, backscatter values for each sample site were assigned for both the individual pixel backscatter and the mean pixel backscatter. Finally, the mosaics were exported as ArcView Grid Files and loaded as raster features in ArcCatalog 10.1 (ESRI, Redlands, CA).

Results:

The original and reprocessed backscatter for the Great Peconic Bay is shown in Figure 2-1. Each pixel on the reprocessed mosaic represents 1 m^2 of seafloor. For the Little Peconic Bay, the reprocessed backscatter is displayed in Figure 2-2. Each pixel represents 2 m^2 of seafloor. The granule texture present in the original mosaic (noise) was removed and boundaries between dissimilar acoustic areas became clearer. The reduction in noise was a result of applied radiometric and geometric corrections. Radiometric corrections included the removal of variable acquisition gains, power levels, pulse widths and incidence angles. Corrections for slant-range distortion and transducer altitude were also applied.

<u>Task 2: Compare image segmentation/classification methods to identify bottom provinces from sonar data</u>

The goal of this task was to explore representative unsupervised and supervised image analysis techniques to determine if image segmentation algorithms can replace the subjective, visual approach used to classify areas into homogeneous bottom types (Cerrato and Maher 2007, Cerrato et al 2009, 2010). Three software approaches were considered: classic image analysis tools that carry out image processing such as filtering and edge detection, multiresolution segmentation, and a supervised maximum likelihood classification method.

Methods:

Image processing was carried out on the sonar original sonar mosaics produced by Flood (2004) since the reprocessed backscatter data were not available at the time this task was undertaken. The first method applied was through a software collection of computer imaging tools called CVIPtools developed at Southern Illinois University at Edwardsville. This software carries out a number of image enhancement techniques (e.g., filters, transformations, sharpening, smoothing) along with image processing analyses such as edge detection, pattern classification, and segmentation.

The second method examined was multiresolution or hierarchal segmentation. This method produces a set of image segmentations at two spatial scales by first partitioning the image into small areas, called image objects, based on a procedure that maximizes homogeneity within image objects. It then merges groups of image objects together to form a final set of spatially larger areas. This analysis was carried out using a trial version of eCognition Developer (Trimble Geospatial, Westminster, CO). A full version of the software, costing over $5,000 was not feasible.

The third method, supervised maximum likelihood classification, was carried out in ArcGIS 10.1 (ESRI, Redlands, CA). In supervised classification the image processing software is guided by the user who specifies the number of distinct types of image classes present and provides small representative examples of each. The software then applies this “training set” or “signature set” to the entire image matching unclassified areas to the training set properties. This supervision by the user allows potentially more complex structures to be defined than with unsupervised computer analysis, while also being more objective than a user visually segmenting an image.

Results:

A variety of image analysis algorithms for edge detection, segmentation, and filtering using CVIPtools were applied to backscatter data. Some of the segmentation algorithms were potentially useful, but the end result was not substantially better than the original backscatter image (Figure 2-3).

Multiresolution segmentation using eCognition Developer was successful in partitioning sonar backscatter images into small image object areas (Figure 2-4). Unfortunately, further application of the software did not lead to an acceptable final set of larger areas. This may have been because the trial software had limited options available. Overall, this method appeared promising but would require a significant investment in the software and extensive operator training to be effective.

The maximum likelihood classification algorithm was reasonably successful in differentiating broad areas of the seafloor apparent in the backscatter data (Figure 2-5). It did not always result in sharp boundaries between bottom types, possibly because the spectral properties at transitions was gradational making it difficult for the classification algorithm to distinguish between areas.

Task 3: Derive new variables from sonar data

Biologically relevant variables had been obtained previously from a detailed analysis of multibeam backscatter and bathymetry data (Maher and Cerrato 2004, Maher 2006, Cerrato and Maher 2007). Examples include mean, standard deviation, Pace features, and grey-level concurrence matrices derived from the backscatter and rugosity derived from the bathymetry in the neighborhood of sampling stations. The goal of this task was to determine if new environmental variables are available that may be useful in the context of habitat classification.

Methods:

Recently, angular range analysis (ARA) had been proposed (Fonseca and Mayer 2007) as a method to estimate surficial seafloor properties; therefore this task focused on testing this set of measures. ARA derives measures from the relationship between backscatter strength and grazing angle. In particular, ARA analysis estimates the strength of backscatter at different grazing angles as well as the slope of the relationship at those grazing angles. This multivariate information is then compared to empirically derived measurements acoustic impedance of seafloor sediments and ARA patches are assigned indices based on grain size. The Fledermaus software was used to carry out Fonesca and Mayer's (2007) approach, and ARA mosaics were created for both Great Peconic Bay and the Little Peconic Bay.

Results:

ARA results were interesting but did not appear to provide information that was substantially different than the original backscatter data (Figure 2-6). The advantage of this approach appeared to be mainly in remotely predicting grain-size but not necessarily valuable in assisting in image segmentation. Full details of this process can be found in White (2015) (Appendix A).

Task 4: Couple ordination and spatial statistics to evaluate performance of environmental variables

Field-based sampling methods for examining spatial structure in benthic communities have traditionally involved *in situ* sampling. The distance interval between these samples defines the smallest scale over which the data set can inform species-environmental relationships. Below that distance, small-scale variations are present (Legendre and Legendre 2012). These are presumably dominated by biotic interactions (Legendre 1993), especially at scales approaching the organisms themselves (e.g., Shumchenia and King 2010), along with small-scale variability in measured environmental variables, other unmeasured environmental factors that become important at small scales, and measurement error.

With the proliferation of techniques for the analysis of sonar data (e.g., acoustic segmentation via visual, supervised, and/or unsupervised classification), acoustically-defined habitats (referred to as "provinces" here) or areas of the seafloor consisting of apparently homogeneous geophysical conditions can be readily derived and used in community analysis. Beyond that obvious "seascape" scale, other scales at which sonar data should be analyzed to provide useful information on community structure is unclear. Scale selection for the purpose of analysis can be as critical as observational scale is for identifying relationships (Dungan et al. 2002).

Multiscale ordination (MSO) couples spatial statistics to examining the spatial structure of biotic-environmental relationships by inserting multivariate regression results from a direct gradient method such as redundancy analysis (RDA) into an empirical variogram (Wagner 2003, 2004, Wagner and Fortin 2005). MSO is distance-based, where the spatial structure in a data set is scrutinized using the empirical variogram, which is a plot of ecological "distance" against geographic for pairs of stations (Legendre 1993; Wagner 2003). This analysis uses the output of a nonspatial regression analysis to partition the community variation that is explained by environmental factors (explanatory variables) from residual (unexplained) variability at different spatial scales or distances in geographic space. The overall analysis also explicitly tests for the presence of spatial structure in residuals due to small-scale autocorrelation or missing environmental factors.

The goal of this task was to use multiscale ordination to examine the spatial structure and biotic-environmental relationships of infaunal benthic communities. Analysis of the data focused on quantifying the fraction of community variation that is below the observational scale of the data sets, estimating the amount of community variation explained by commonly used *in situ* and seascape-scale explanatory variables, determining whether explanatory variables identified in a model selection process remove spatial structure in residuals or whether unresolved structure remained, and providing guidance for designing future studies. The analysis below was part of a broader study by Flanagan (2016). Full details of the broader study can be found in Appendix B.

Methods:

Within the Peconics, benthic data analysis for this task focused on areas near Robins Island and Shelter Island collected as part of the benthic mapping project in Cerrato and Maher (2007). Seascape or habitat-scale environmental data were generated by segmenting each site into regions of the seafloor with apparently homogeneous geophysical conditions. This was based primarily on visual interpretation of backscatter data from multibeam and sidescan sonar surveys (Figure 2-7 and 2-8). These homogeneous regions will be referred to as acoustic provinces. Faunal and sediment samples were collected *in situ* using a modified Van Veen grab (20×20cm). Sampling locations were random, but stratified by province, to insure full coverage of the range of habitats present. Surficial percent cover data were obtained from analysis of images extracted from seabed surface videos using a 2 megapixel Seatrex HD underwater camera mounted on an aluminum tripod. For further methods, see Cerrato and Maher (2007) and Flanagan (2016).

a) Empirical variograms

The spatial structure of the species assemblage was examined by constructing an empirical variogram of the multivariate faunal data for each site (Wagner 2003):

$$\gamma(h) = \sum_{i=1}^{S} \frac{1}{2n_h} \sum_{a,b|h_{ab}\approx h} (x_{ia} - x_{ib})^2 \qquad (1)$$

where $\gamma(h)$ is the empirical variance of the faunal data at distance class h, x_{ia} and x_{ib} are measured values for species i ($i = 1\ to\ s$) in samples a and b, respectively, and the inner summation is over all pairs of samples separated by a geographic distance of approximately h. The measured values used for each species (x_{ia} and x_{ib}) were Hellinger transformed abundances. This transformation takes the square root of the relative abundance of each species in a sample (Legendre and Gallagher 2001). As such, it focuses the analysis on compositional differences and downplays the influence of highly abundant species to prevent them from dominating the analysis. In addition, when used in conjunction with Euclidian distance, the ecological distance measure utilized here and in the multivariate regression analyses presented below, it produces good representations of ecological dissimilarity (Legendre and Gallagher 2001). Summing equation (1) for all pairs of samples, instead of a distance class subset, yields s^2 the total sample variance (Bachmaier and Backes 2008).

A plot of $\gamma(h)$ vs. distance classes h is called a variogram plot. Distance intervals were selected for each data set to have plots with about 10 $\gamma(h)$ values. This allowed the distance classes to usually contain >30 sample pairs, as suggested by Journal and Huijbregts (1978). Important features of the multivariate variogram include the *sill*, which is the asymptotic value in the variogram plot (if it exists) and the *nugget* or *nugget effect*, the value of $\gamma(h)$ at $h = 0$ (Wagner 2003; Legendre and Legendre 2012). The nugget effect is of particular interest in this study because it represents the community variation occurring at scales below that of the sampling interval (Legendre and Legendre 2012). Variograms were created using the rda(), mso(), and msoplot() functions in the vegan package of R (R Foundation for Statistical Computing, Vienna, Australia). The code for mso() and msoplot() was created and first published by Wagner (2004).

Empirical variograms created for each site were fit to a number of common models (e.g., exponential, spherical, gaussian, piece-wise linear) utilized in spatial statistics in order to obtain estimates of the nugget effect (i.e., the intercept). Models were fit by weighted least squares, as suggested by Cressie (1993), with weighting factors defined by the number of sample pairs in each distance class. Weighted least squares was carried out using function nls() in the stats package of R (R Foundation for Statistical Computing, Vienna, Australia). Model selection was achieved using Akaike's Information Criterion (AIC) (Akaike 1973).

b) Multiscale ordination

Multiscale ordination (MSO) extends spatial statistics to examining the spatial structure of biotic-environmental relationships by inserting regression results into the variogram (Wagner 2003, 2004, Wagner and Fortin 2005). It does this by partitioning the x_{ia} and x_{ib} pairs in equation (1) into fitted and residual parts ($\hat{x}_{iafit} + \hat{x}_{iares}$) and ($\hat{x}_{ibfit} + \hat{x}_{ibres}$), respectively. Substituting these into equation (1) leads to:

$$\gamma(h) = \sum_{i=1}^{s} \frac{1}{2n_h} \sum_{a,b|h_{ab}\approx h} \Big[\big(\hat{x}_{iafit} - \hat{x}_{ibfit}\big)^2 + (\hat{x}_{iares} - \hat{x}_{ibres})^2 + 2\big(\hat{x}_{iafit} - \hat{x}_{ibfit}\big)(\hat{x}_{iares} - \hat{x}_{ibres})\Big]$$

$$= \gamma_{fit}(h) + \gamma_{res}(h) + \gamma_{cross}(h) \quad (2)$$

The first two terms on the right hand side are variograms of the fitted and residual values. The third term is twice the covariance between the fitted and residual differences for distance class h (Wagner 2003). Multivariate regression estimates of species-environmental relationships can be obtained by redundancy analysis (RDA) for small to moderate environmental gradients or canonical correspondence analysis (CCA) for large gradients (Wagner 2004). In assessing whether the RDA can account for spatial structure in the data, $\gamma_{res}(h)$ should remain constant over most of the geographic range and $\gamma_{cross}(h)$ should not be significantly different from 0 at any distance class.

Estimates of predicted and residual Hellinger transformed abundance values were obtained by RDA. RDA is a multivariate method that combines multiple linear regression with ordination. A parsimonious set of explanatory environmental variables was identified by sequentially adding variables in a forward selection process (Jongman et al. 1995). Candidate variables for each site included water depth, grab penetration depth, apparent RPD depth, grain size (% gravel, sand, and mud), surficial percent cover (shell, seaweed, and other materials observed in bottom images), and categorical variables as binary 1/0 values representing acoustic provinces. Prior to analysis, results from multivariate regression tree (MRT) analysis (De'ath 2002), explained in detail in Objective 3, was used to guide the formation of larger groups of contiguous and/or noncontiguous acoustic provinces with similar faunal assemblages to avoid evaluating all possible 2^n - 1 unique combinations. At each step in the forward selection RDA process, the environmental variable explaining the largest amount of faunal variability was selected, and its effect removed before the next best fitting variable was considered. Variables identified by forward selection were trimmed back to a smaller set by the AICc stopping criterion (Burnham and Anderson 2002). Forward selection in RDA was carried out in Canoco 4.5 (Microcomputer Power, Ithaca, NY, USA).

Variogram measures were recomputed to quantify the extent the RDA models explained local vs. habitat-level community variation. This analysis also assessed the ability of acoustic provinces to serve as proxies for habitats within the study areas. The variogram function mso() was modified slightly to replace the geographic distance matrix between pairs of samples with a matrix of two "distance" classes: within (1) and between (2) acoustic provinces. The output provided the variogram values for $\gamma(within/between)$, $\gamma_{fit}(within/between)$, and $\gamma_{res}(within/between)$. Assuming that small-scale variation in the data cannot inform species-environmental relationships, estimates of the nugget effect were subtracted from residuals to assess model performance.

Results:

a) Empirical variograms

The empirical variance $\gamma(h)$ increased with geographic distance class across both study locations, indicating that spatial structure was present in the species assemblages (Figure 2-9). At small distance classes, $\gamma(h)$ was dominated by pairs of samples that were within the same acoustic province, and $\gamma(h)$ reached a sill once it became dominated by between-province sample pairs at large distance classes. Based on estimates of the nugget effect in the variogram models, small-scale variability represented 36% of the overall community variance at Robins Island and 37% at Shelter Island. A piecewise linear model was selected for Robins Island and a spherical model for Shelter Island.

b) Multiscale ordination

The forward selection process in RDA resulted in the selection of 6-8 explanatory variables and r^2 values representing 45-53% of the total variance (Figure 2-10). Acoustic provinces dominated the analysis at both sites, but other environmental variables identified in forward selection and subsequently retained by the AICc model selection criterion were water depth (Robins Island and Shelter Island), percent sand (Shelter Island), and the percent cover classes from image analysis including % mud cover, % shell fragment cover, and % *M. porifera* cover at Robins Island.

MSO results indicated that the spatial dependence in the faunal data was effectively captured by the explanatory variables selected in the RDA, and that the residuals contained no detectible spatial structure (Figure 2-11). Variograms formed from RDA predictions, $\gamma_{fit}(h)$ had evident spatial structure that paralleled the empirical variogram $\gamma(h)$. In contrast, $\gamma_{res}(h)$ was fairly constant across all distance classes, indicating that the residual relationships were stationary across the distance classes and no unknown environmental factor(s) were present influencing the spatial structure of the residuals.

Subtracting the nugget estimates of small-scale variability from RDA regression residuals altered the perception of RDA model performance both overall and at local vs. habitat levels (Figure 2-12). With small-scale variability removed from the residuals, the nonspatial RDA explained >71% of the remaining variance in community structure. When comparisons were restricted to sample pairs within the same acoustic province, the RDA model explained 42% for Robins Island and 23% for Shelter Island. When comparisons were restricted to sample pairs in different acoustic provinces, the RDA model explained 83% of the remaining variance in community structure for Robins Island and 73% for Shelter Island. Thus, within-province explained variance was notably weak while between-province explained variance was notably strong.

Discussion:

The analysis in Task 4 suggested that visual segmentation did well in defining bottom types with distinct benthic assemblages. While image segmentation software may improve boundary locations, the large fraction of explained, between-province variance (> 73%) indicated that habitat differences in community structure were effectively explained by acoustic province, depth, grain-size, and percent cover variables.

In contrast, the weak link in modeling biotic-environmental relationships seemed to be in explaining within-province variation. Here, RDA explained only 23-42% of the nugget-corrected variance. Improvements in image backscatter resolution (Task 1) along with using new image analysis software (Tasks 2-3) could potentially produce substantial improvements. In particular, the image objects resulting from the first stage of the segmentation analysis in eCognition Developer seem promising as candidates for further analysis. These within-province patches may be characterizing smaller-scale seafloor variability, and measures derived from them, such as backscatter mean, standard deviation, Pace features, and grey-level concurrence and bathymetric rugosity, should be considered as explanatory variables in future analyses.

Nugget estimates of small-scale variability represented 36-37% of the total community variation at the Robins and Shelter Island sites. In Flanagan (2016), four additional sites in the brackish and freshwater portions of the Hudson River were also analyzed, and the range of nugget estimates expanded to 36-59%. Thus, a large fraction of the community variation was below the resolution of the surveys. Potential causal factors for this small-scale variation include biotic interactions (Legendre 1993) and randomness in settlement, especially at a scale approaching the organisms themselves (e.g., Shumchenia and King 2010), small-scale variability in measured environmental variables, other unmeasured environmental factors that become important at small scales, and measurement error. This resolution problem combined with the weak within-province results highlight further challenges to understanding heterogeneity in benthic communities. On the other hand, spatial structure in the infaunal community at the habitat scale was effectively explained by commonly collected environmental variables, and therefore, the surveys were well suited to detect community structure differences driven by large-scale, habitat level environmental changes. This is a particularly important result because it provides a clear path for identifying and managing benthic habitats via sonar and traditional benthic fauna sampling.

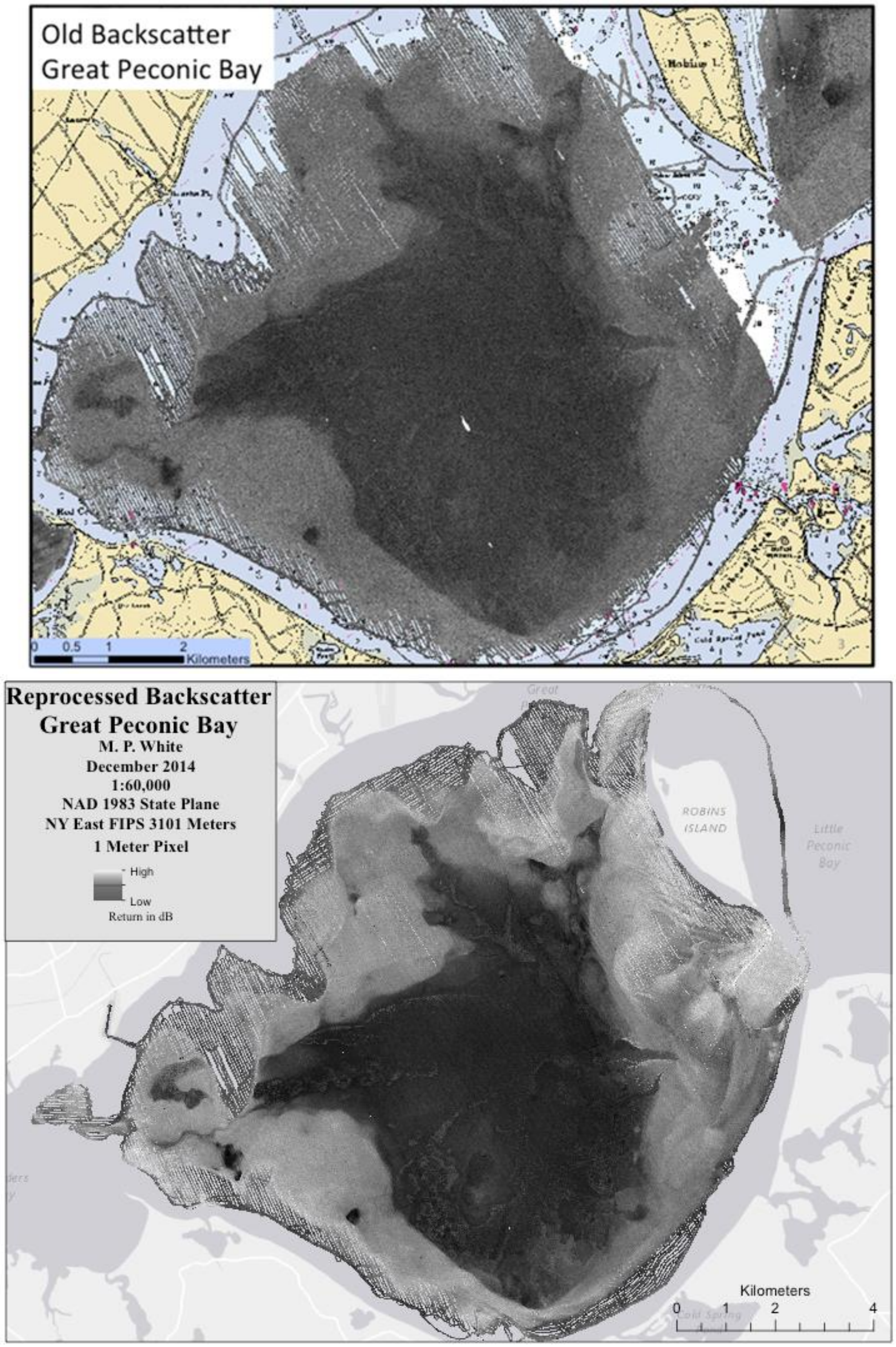

Figure 2-1: Original and reprocessed backscatter mosaic for the Great Peconic Bay. Each pixel for the reprocessed mosaic represents $1m^2$. Brighter areas represent more intense backscatter; darker areas represent less intense backscatter. Mosaic was created using all of the data available for the Great Peconic Bay; therefore, values were normalized over the entire area.

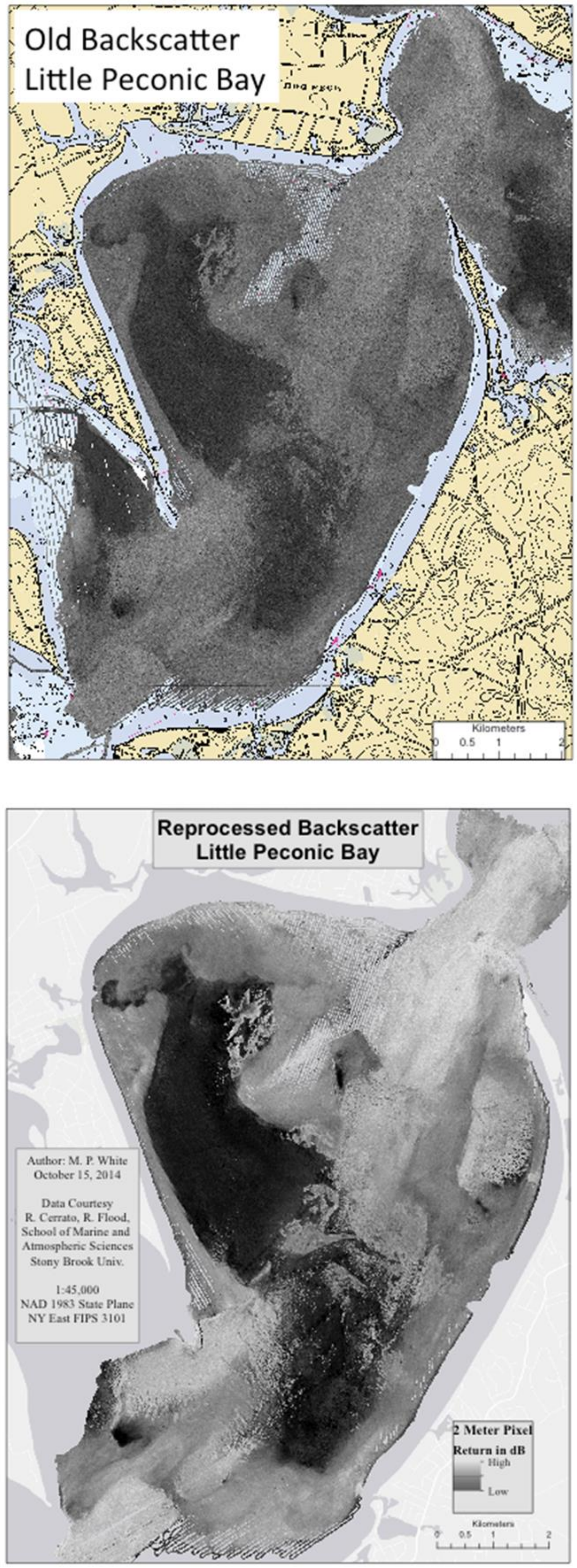

Figure 2-2: Original and reprocessed backscatter for the Little Peconic Bay. Each pixel for the mosaic represents 2 m^2. Brighter areas represent more intense backscatter; darker areas represent less intense backscatter.

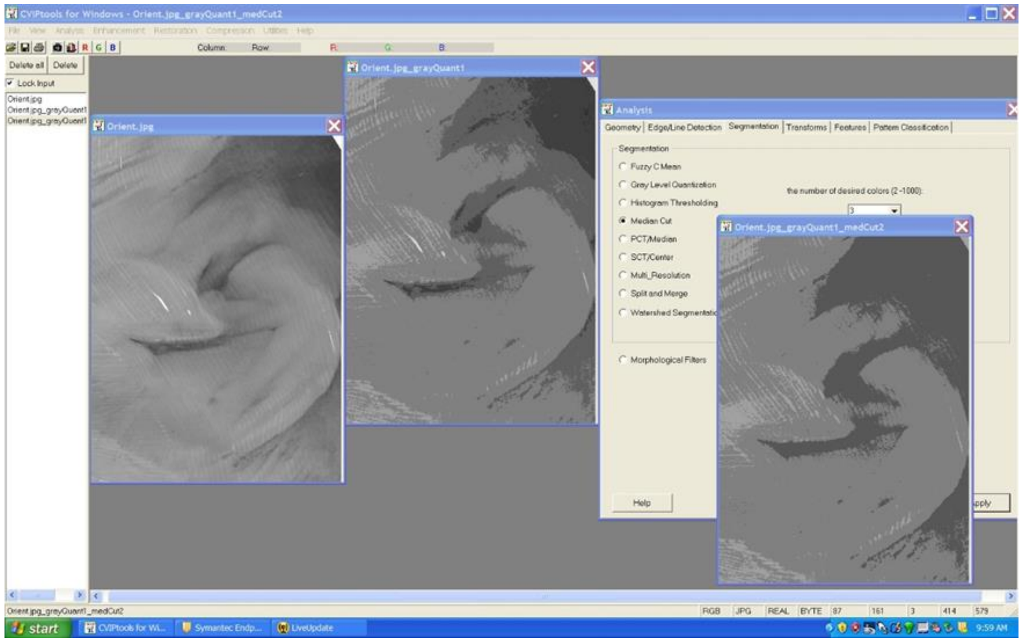

Figure 2-3: An example of image analysis using CVIPtools. The original backscatter mosaic is displayed in the left panel and is a section of seafloor outside Orient Harbor. Brighter areas represent more intense backscatter; darker areas represent less intense backscatter. The other two images are the result of applying a grey-scale filter (middle panel) followed by a median cut image segmentation algorithm (right panel).

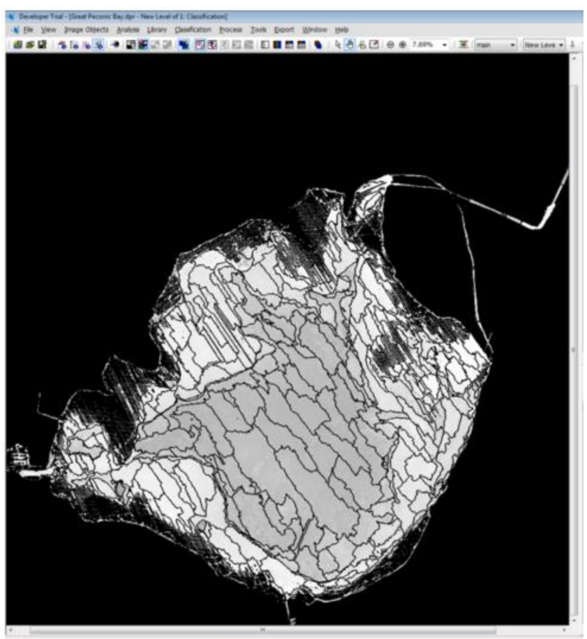

Figure 2-4: An example of image analysis of Great Peconic Bay backscatter data using eCognition Developer. Brighter areas represent more intense backscatter; darker areas represent less intense backscatter. Image objects resulting from the first stage of the segmentation analysis are shown.

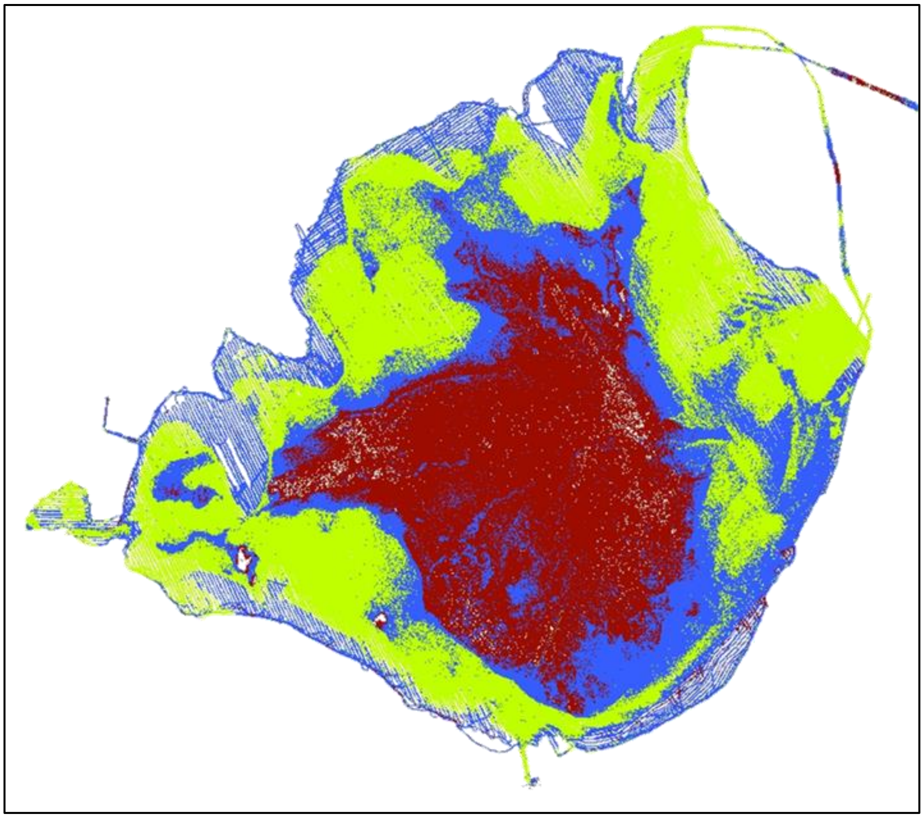

Figure 2-5: An example of image analysis using maximum likelihood classification. The area segmented is Great Peconic Bay.

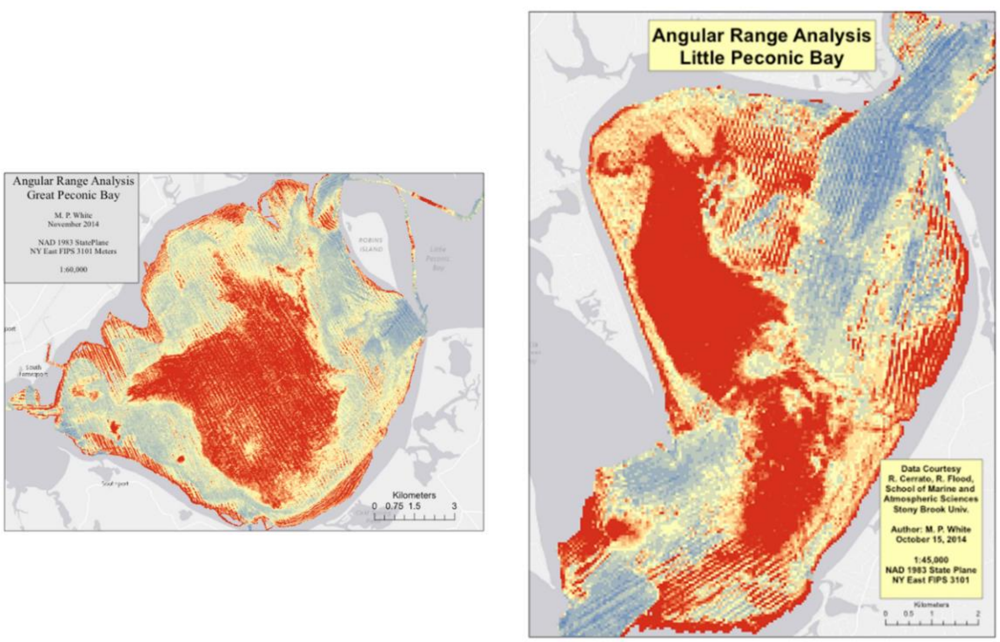

Figure 2-6. ARA layers for Great Peconic Bay and Little Peconic Bay. Segmented areas with the same ARA indices are assumed to have similar sediment grain sizes. Indices are represented by color, with reds being finer grained material and blues coarser grained material.

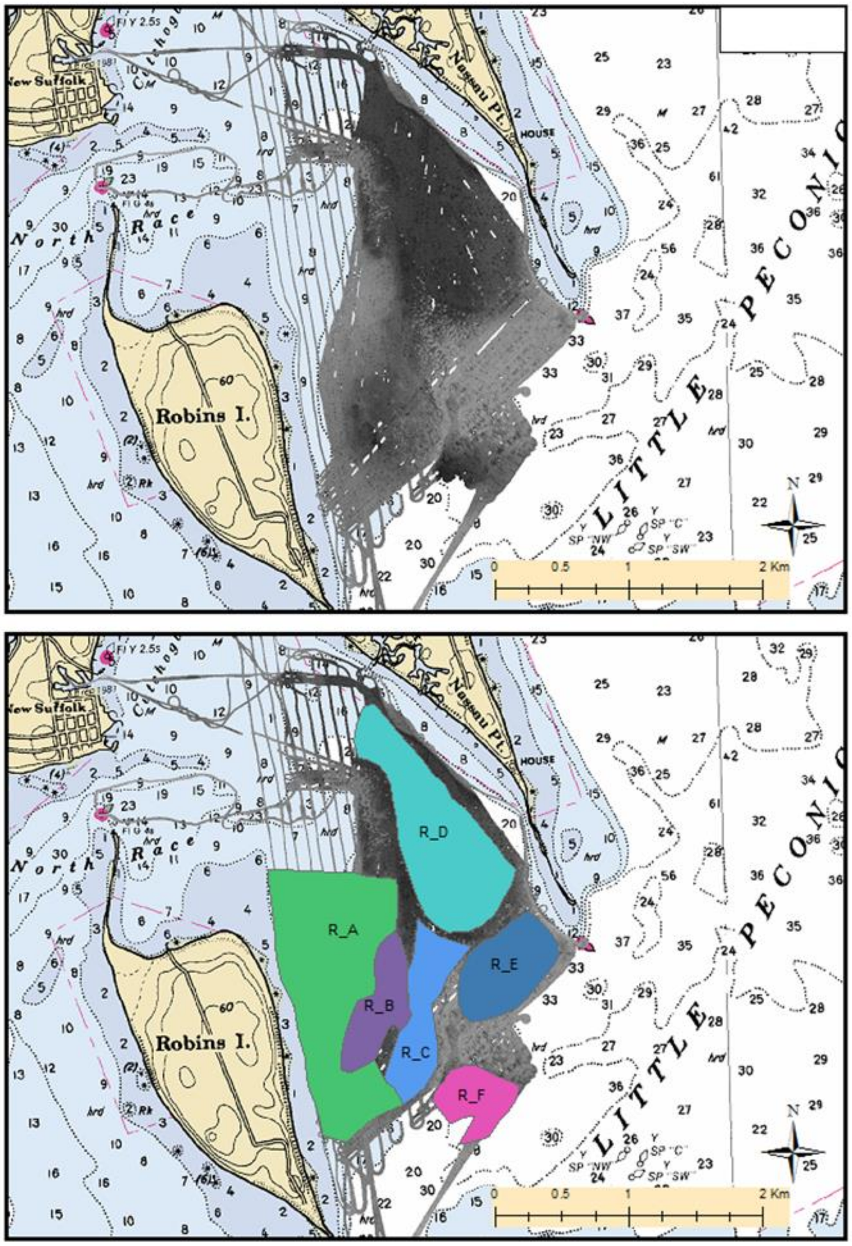

Figure 2.7. Multibeam sonar data (top) and visual interpretation of acoustic provinces (bottom) at Robins Island.

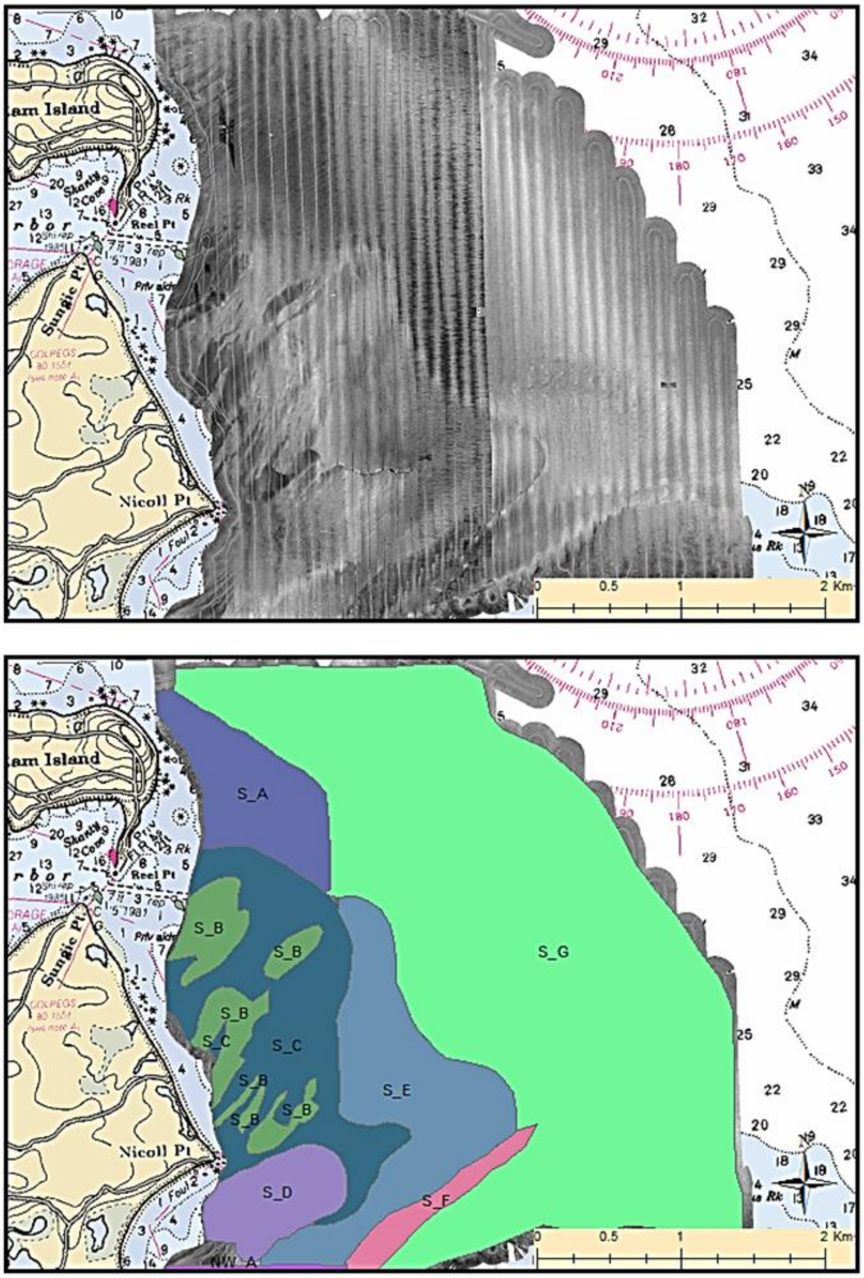

Figure 2-8. Side scan sonar data (top) and visual interpretation of acoustic provinces (bottom) at Shelter Island.

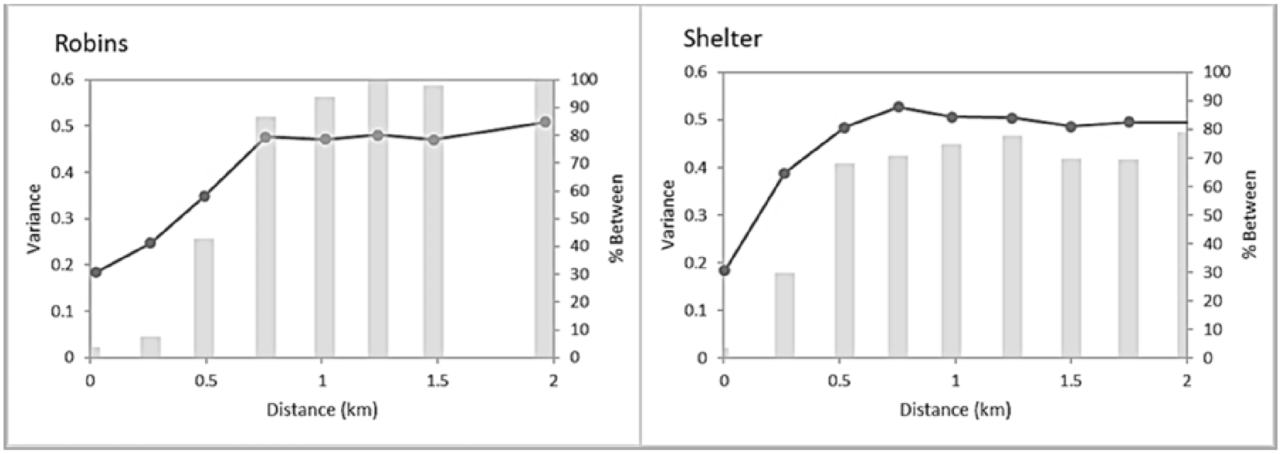

Figure 2-9. Multivariate variogram (points) of community data and the percentage of stations (bars) within and between acoustic provinces against increasing distance in geographic space. Distance intervals for grouping station pairs were 0.25 km increments.

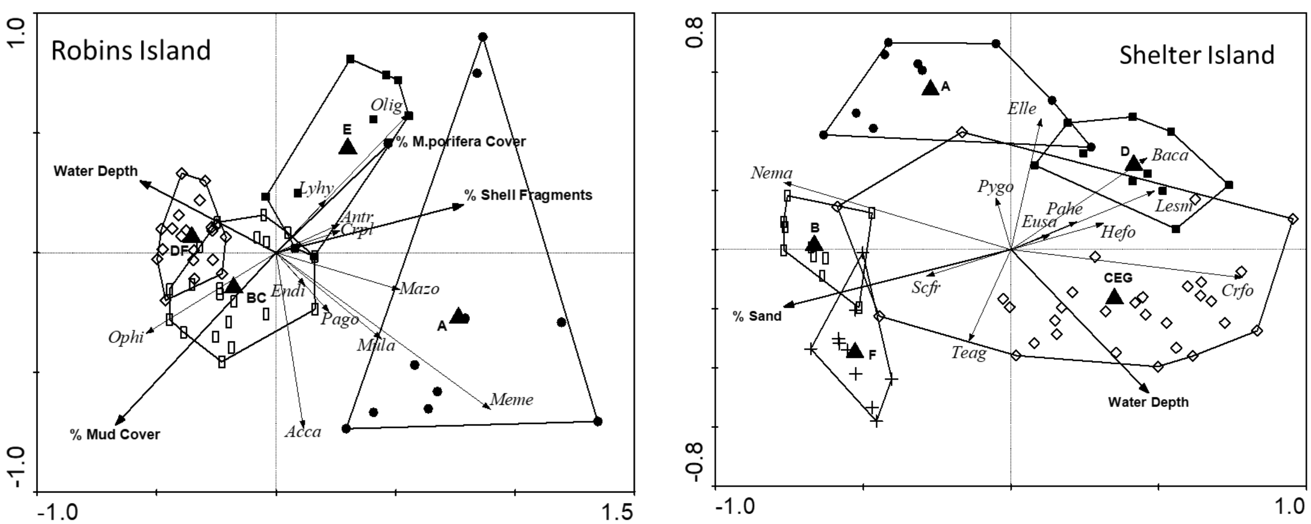

Figure 2-10. Ordination triplots of nonspatial RDA results. Scores for individual samples plotted as points. Sample points close to one another tend to have similar faunal structure than those further apart. Different point shapes represent samples collected from different acoustic provinces. Continuous explanatory variables and individual taxa are plotted as vectors. The vector arrowheads represent high, the origin average, and the tail (when extended through the origin) low values of the selected environmental variables or taxa. Projections of sample points onto an individual taxa vector approximate the Hellinger transformed abundances for that taxon. For clarity, only species with the highest amounts of explained variance (typically those with > 10%) are plotted. Categorical acoustic province variables are designated as letters and are plotted as triangular points. Provinces were arbitrarily labeled as A, B, C, etc.

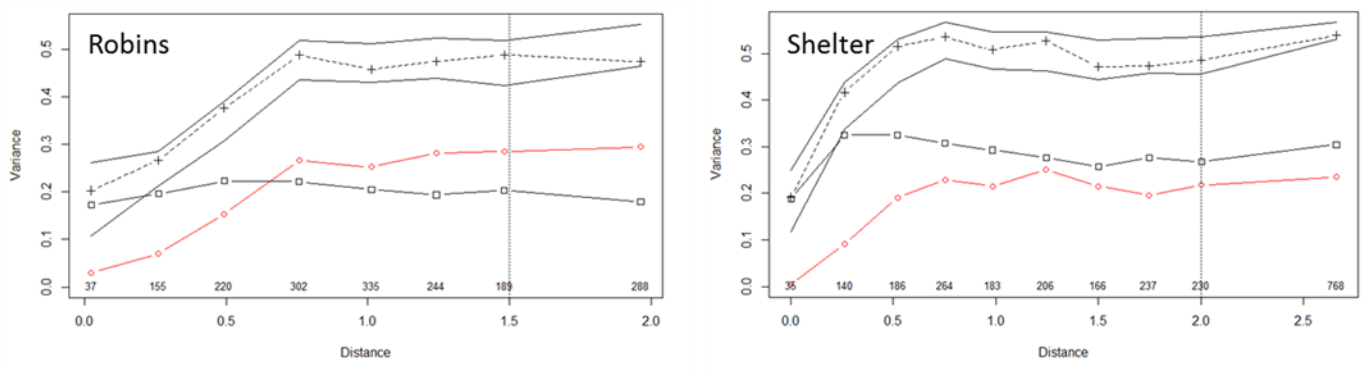

Figure 2-11. MSO plots for each study area. Crosshairs are the sum of fitted and residual variograms, solid lines are Bonferroni-corrected point confidence intervals around the empirical variogram, open diamonds represent the fitted variogram, and open squares are the residual variogram. See the explanation of equation 2 for definitions of these quantities. Distance intervals are in kilometers.

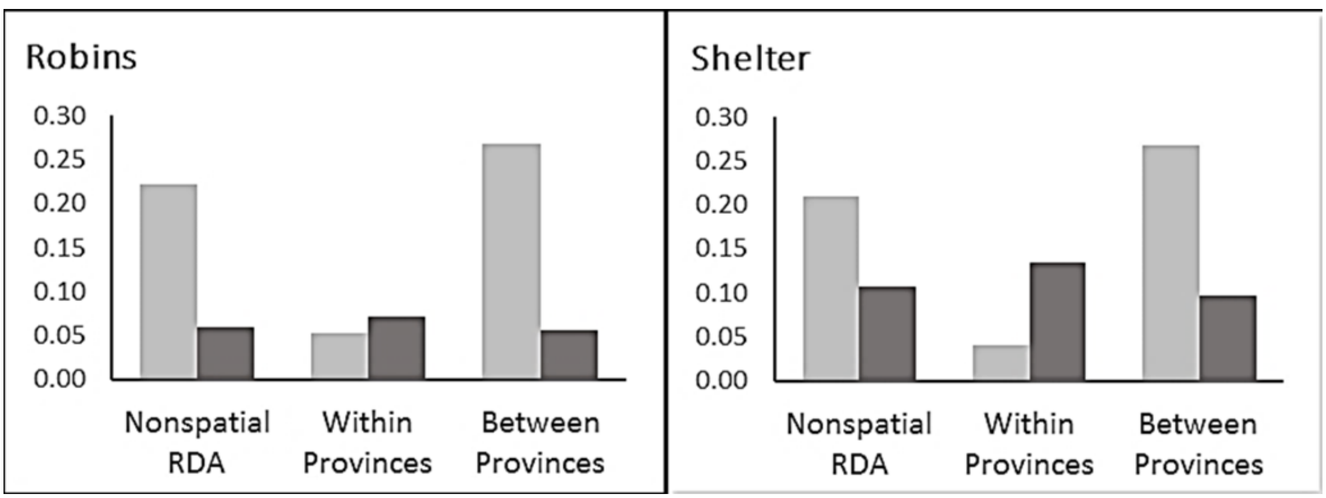

Figure 2-12. Community variance explained by RDA models (light grey) compared to nugget corrected residual variance (dark grey). For each data set, comparisons were made for the entire nonspatial RDA and for results broken down into components representing variation within and between acoustic provinces.

Objective 3: Conduct a spatial/temporal analysis of existing trawl survey data

The goal of this objective was to analyze changes in the Peconic Estuary's community of finfish and macro-invertebrates using a suite of multivariate statistical techniques. This study primarily used data generated by the New York State Department of Environmental Conservation's (NYSDEC) annual juvenile trawl survey. Initial attempts to identify community patterns on a per trawl basis proved unproductive because of the large variation in the data. Structure was clearly present, including very strong seasonal patterns and spatial relationships, but the patterns were especially difficult to relate to environmental drivers. Based on these preliminary analyses, this objective was focused, therefore, on annual variation in order to understand broad-scale temporal patterns as a way to meaningfully approach this complex set of data. Full details of this work can be found in Abruzzo (2015) (Appendix C).

Methods:

Characteristics of trawl data, benthic data, multibeam and side scan data, grain-size data, and water quality data were described under Objective 1 and will not be repeated here. In the trawl data, rare species were deleted to restrict the study to common species. A species was considered rare if it occurred in less than 10% of all tows collected from 1987-2012. After rare species were removed, the relative abundance of each species was calculated for each tow and Hellinger transformed (Legendre & Gallaher 2001). Annual transformed relative abundance was then calculated as the average of tows for that year. The Hellinger transformation down weights the importance of highly abundant species, and combined with Euclidean distance, produces a good multivariate representation of ecological data (Legendre & Gallaher 2001). Several taxa were combined into larger functional groups because of their similar ecological roles: (1) anchovy (bay anchovy *Anchoa mitchilli* and striped anchovy *Anchoa hepsetus*), (2) spider crab (six spine spider crab *Libinia dubia* and nine spine spider crab *Libinia emarginata*) and (3) searobin (northern searobin *Prionotus carolinus* and striped searobin *Prionotus evolans*). Twenty-two taxa were included in the analysis (Table 1).

To capture the structure of temporal patterns in the finfish and shellfish community, we utilized a variety of statistical techniques including empirical variograms, multivariate regression trees (MRT), and redundancy analysis (RDA). A description of empirical variograms and RDA is contained in Task 3 of Objective 2 and will not be repeated here. Multivariate regression trees (MRT) classifies samples into groups by repeatedly splitting the data based on criteria obtained from explanatory variables (De'ath, 2002). The end result of the MRT is a tree with a top node representing all data and subsequent branches representing subgroups of the data chosen to minimize the within group and maximize the between group variation (De'ath 2002). Trees are overgrown initially, but they are then "pruned" back based on a cross-validation process which calculates the predicted mean square error for each number of splits. The simplest tree whose predicted mean square error is within one standard error of the minimum error is selected as the final pruned result (De'ath 2002). To investigate whether large-scale, temporal community shifts occurred, MRT was utilized to analyze temporal variation in community structure using "Year" as an explanatory variable. The analysis was carried out with the 'mvpart' package in R.

Results:

The five most common taxa, each occurring in >50% of the tows, were lady (or calico) crabs *(Ovalipes ocellatus*), winter flounder (*Pseudopleuronectes americanus*), scup (*Stenotomus chrysops*), spider crabs (*Libinia spp.*), and bay anchovy (*Anchoa mitchilli*). Spearman rank correlation with year revealed 8 taxa that had significantly increased or decreased ($p<0.001$) during the 26 year time series (Figure 3-1). Summer flounder, scup, and smallmouth flounder significantly increased during the time period, while windowpane flounder, horseshoe crab, winter flounder, oyster toadfish, and lady crab all significantly decreased.

The fish and mobile invertebrate community structure in the Peconic Estuary is nonstationary with assemblages becoming increasingly different in time (Figure 3-2). Community structure is positively correlated over time intervals from 1 to about 7 years, with the strength of the correlation progressively decreasing. Assemblages separated by about 8-10 years are uncorrelated from one another. They then become negatively autocorrelated for time intervals of 10 years or more, suggesting that regime shifts, i.e., major changes on community structure, occur on that temporal scale.

Forward selection RDA produced a model with four explanatory variables: the annual AMO lagged by 2 years, the average winter NAO lagged by 1 year, the annual AMO, and average annual chlorophyll (Figures 3-3). These four variables explained 53.1% of the Hellinger transformed community variance, and the first two axes of the redundancy analysis displayed 52.0% of the total community variance and 98.0% of the species-environmental relationship. Notable environmental variables used in the analysis but not selected included annual average values of temperature, seasonal temperature, salinity, dissolved oxygen, secchi depth, and the remaining AMO and NAO variables. Plots of observed and predicted Hellinger transformed abundances for each of the 22 taxa are given in Figure 3-4.

Three major transitions have occurred dividing the Peconic Estuary trawl survey time series into four distinct periods: 1987-1989, 1990-1999, 2000-2009, and 2010-2012 (Figure 3-5). The most distinctive of these transitions was 1999-2000. Secondary transitions occurred at 1989-1990 and 2008-2009. During the 1999-2000 transition, for example, scup, summer flounder, and smallmouth flounder increased, while winter flounder, windowpane flounder, horseshoe crab, oyster toadfish, and lady crab decreased in abundance (Figure 3-6). This change shifted community composition from a flatfish/benthic invertebrate community to one more dominated by demersal fish. Additionally, the secondary periods 1987-1989 and 2010-2012 were characterized by high anchovy, squid, and smooth dogfish populations. When relating these assemblage changes to environmental variables, regional indicators associated with the winter North Atlantic Oscillation and the Atlantic Multidecadal Oscillation explained 48% of the variance in community composition, while of the local variables examined (surface and bottom temperature, salinity, dissolved oxygen, secchi depth, and chlorophyll) only chlorophyll contributed, increasing explained variance by an additional 4%. Strong shifts in community structure similar to those in the Peconics Estuary have also been found in Long Island Sound (Howell and Auster 2012) and Narragansett Bay (Collie et al 2008).

Discussion:

Trends in community composition indicated the dominate fish and mobile invertebrate taxa abruptly changed at three time points: 1989-1990, 1999-2000, and 2008-2009. Whether these changes reflect an altered community dynamics under alternate stable states is debatable, but the changes within the context of the dataset qualifies as a regime shift, i.e., a sudden, non-linear changes of abundance at multiple trophic levels. The exact mechanisms as to why these changes occurred are not known, but the results suggest linkage to regional climatic indices rather than local environmental conditions. Changes in species abundances tend to be due to biotic processes such as prey availability, survival (both natural and fishery), growth, and/or recruitment success. Climatic shifts that involve temperature or wind patterns may induce physical changes within the ocean or coastal areas that in turn affect productivity, migratory movements, trophic interactions, and population dynamics (Hunt Jr. et al. 2002, Möllmann et al. 2008). AMO and NAO should, therefore, be considered as proxies for other processes regulating the finfish and macroinvertebrate community.

Scientific Name	Common Name	Percent Occurrence (%)
Ovalipes ocellatus	Calico Crab	84.3
Pseudopleuronectes americanus	Winter Flounder	67.0
Stenotomus chrysops	Scup	57.2
Libinia dubia + Libinia emarginata	Spider Crab spp.	55.2
Anchoa mitchilli + Anchoa hepsetus	Anchovy spp.	53.9
Limulus polyphemus	Horseshoe Crab	44.6
Scophthalmus aquosus	Windowpane Flounder	41.7
Prionotus evolans + Prionotus carolinus	Searobin spp.	40.9
Loligo pealei	Long-Finned Squid	40.4
Squilla empusa	Mantis Shrimp	40.2
Paralichthys dentatus	Summer Flounder	35.1
Cynoscion regalis	Weakfish	34.7
Syngnathus fuscus	Pipefish	32.5
Sphoeroides maculatus	Puffer	30.9
Tautoga onitis	Tautog	24.7
Opsanus tau	Oyster Toadfish	20.0
Peprilus triacanthus	Butterfish	19.0
Menidia spp.	Silverside spp.	18.2
Callinectes sapidus	Blue Crab	16.3
Etropus microstomus	Smallmouth Flounder	14.9
Mustelus canis	Smooth Dogfish	14.6
Gobiosoma ginsburgi	Seaboard Goby	13.6

Table 1. List of taxa that occurred in more than 10% of all tows collected from 1987-2012.

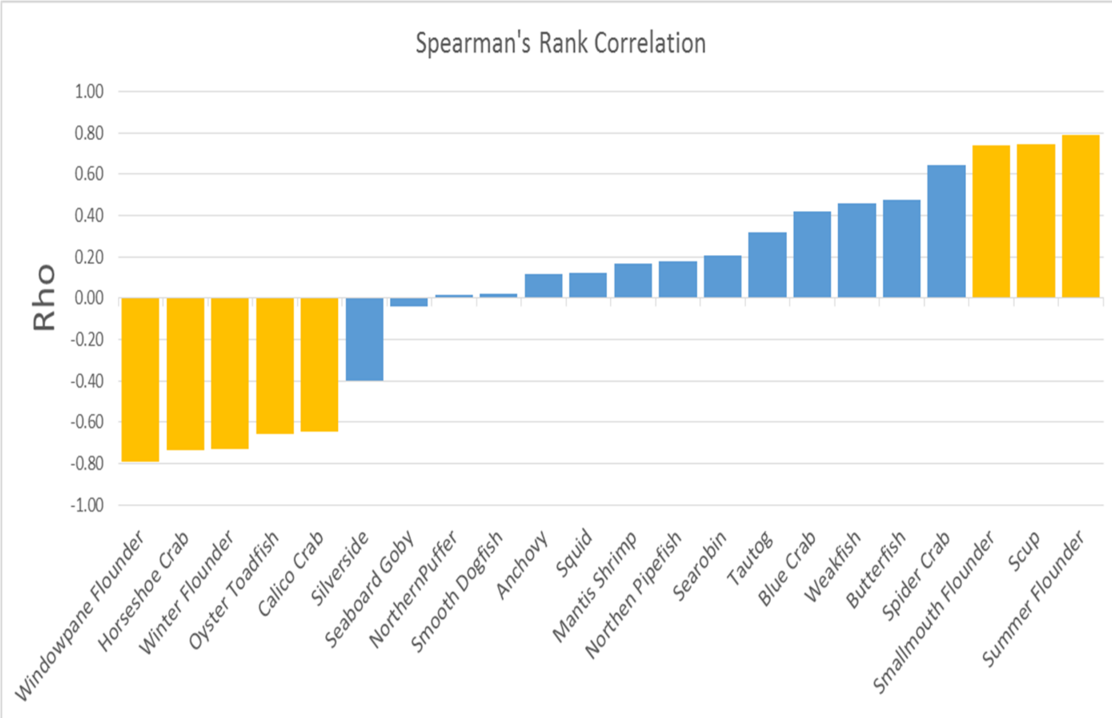

Figure 3-1. Spearman's Rank Correlation between years and Hellinger transformed abundance. Rho is a measure of the correlation between the year and the species relative abundance based on ranking. Bars that appear in yellow are significant at the p<0.001 level.

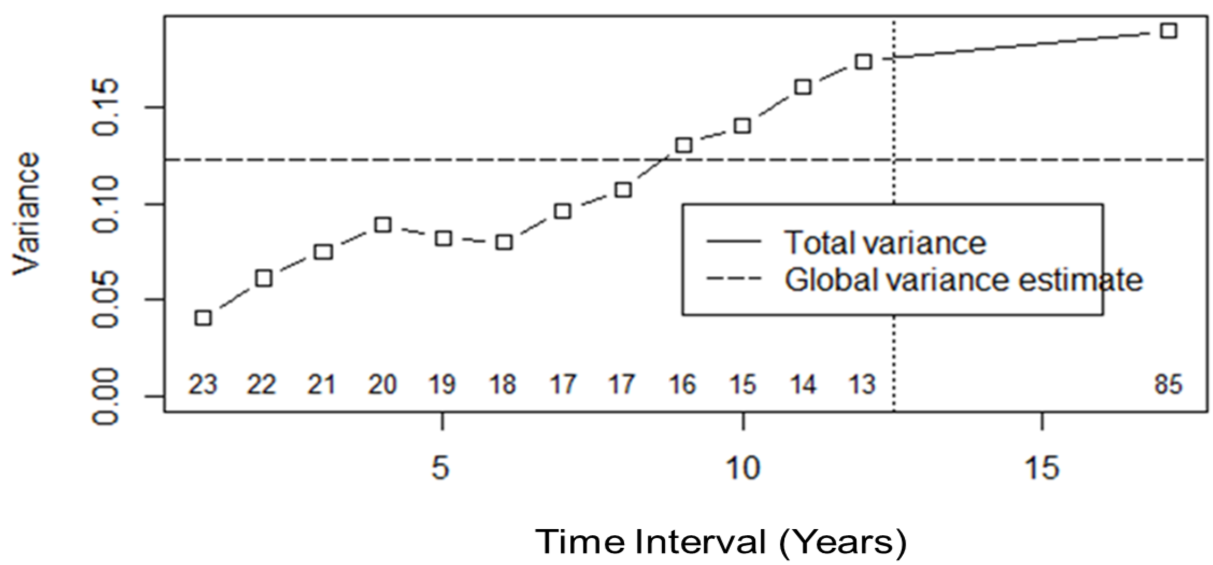

Figure 3-2. Multivariate variogram (points) of community versus increasing time interval between sampling. Numbers at the bottom are the number of pairs of comparisons averaged with a time interval. The dashed line is the oveall sample variance.

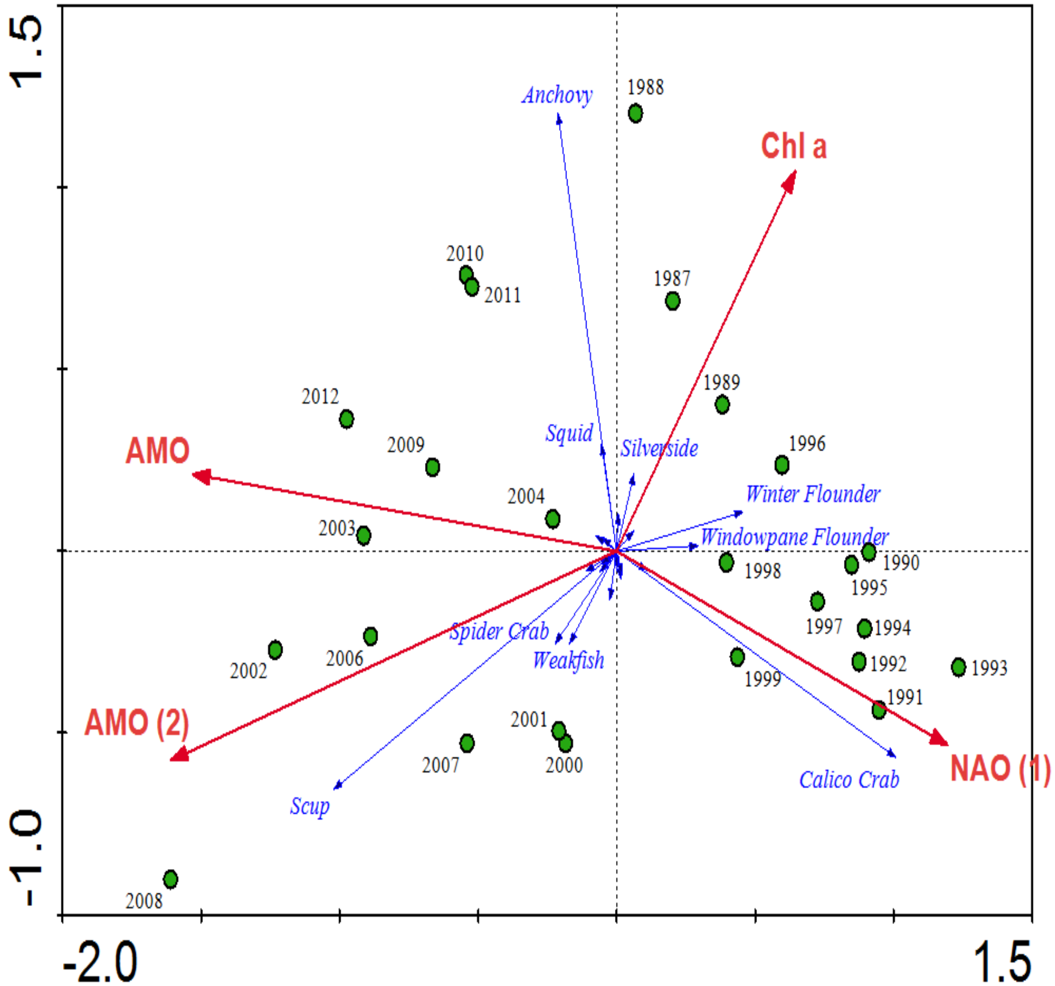

Figure 3-3. Redundancy analysis (RDA) results. The first two RDA axes are plotted. The green circles represent annual transformed species abundances. The blue lines represent species and the red lines represent environmental variables. Only the environmental variables that were included in the minimum AICc model are plotted. Only a subset of taxa are displayed for clarity.

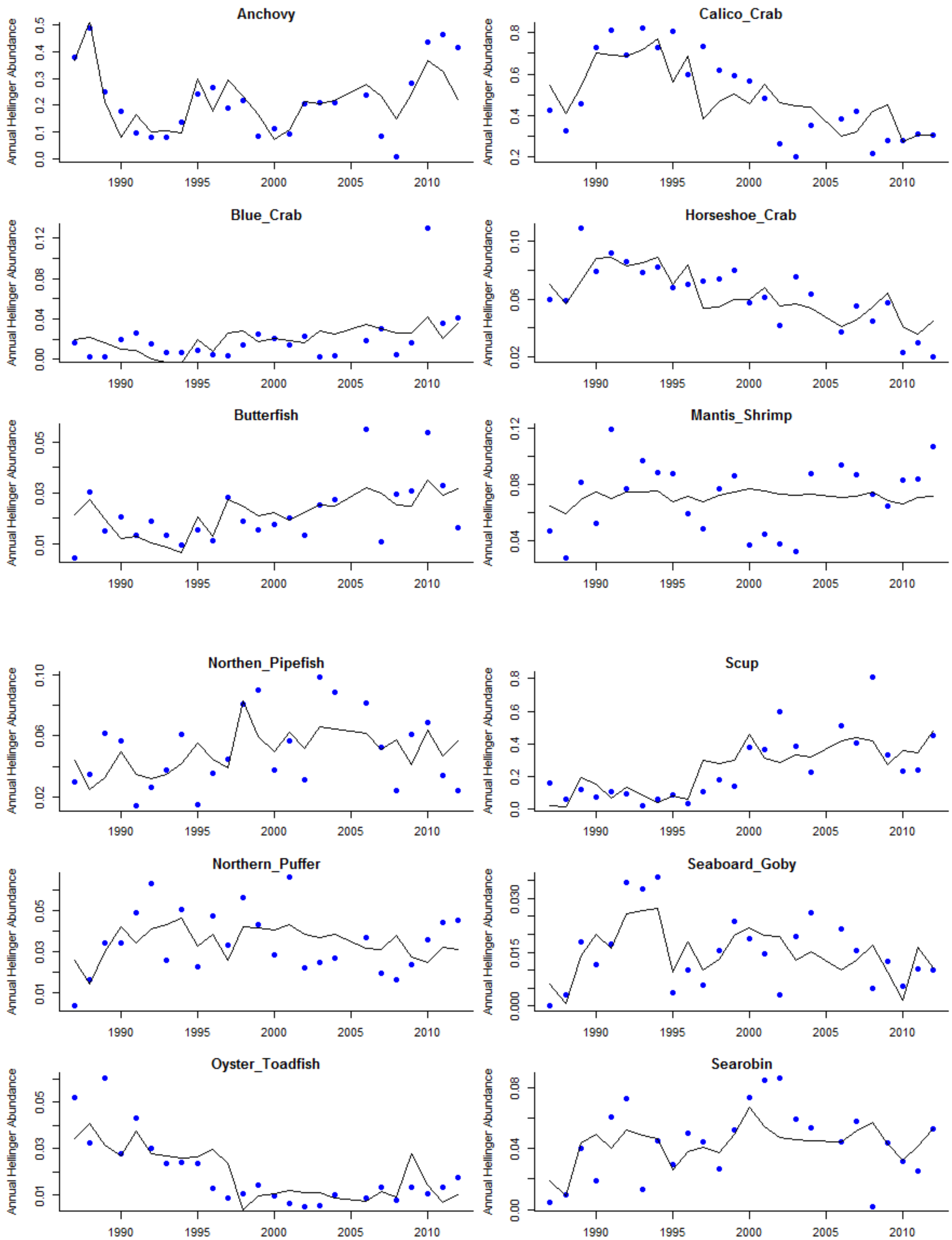

Figure 3-4. Observed (points) and predicted (lines) annual Hellinger transformed abundance based on the results of the RDA model.

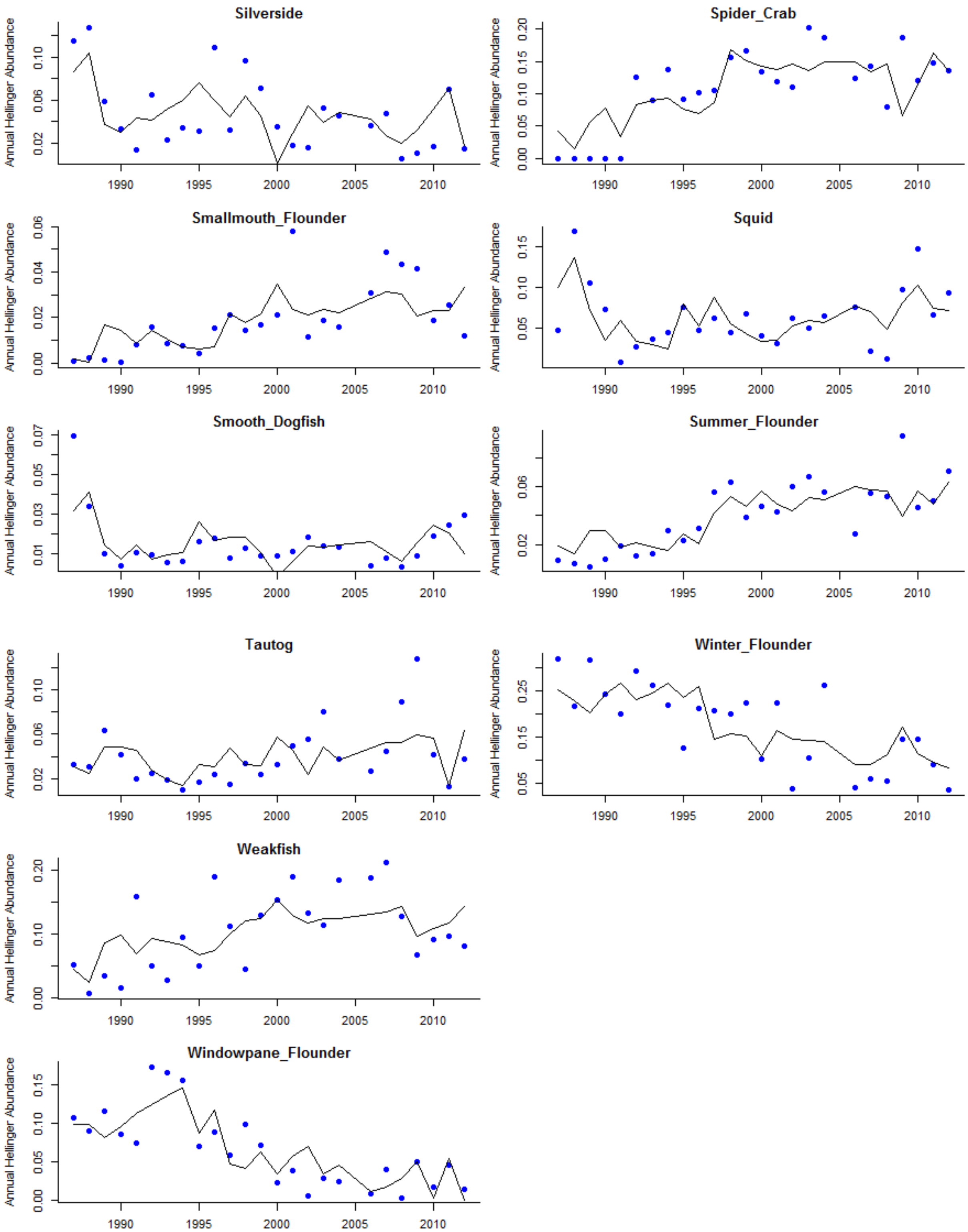

Figure 3-4 (continued). Observed (points) and predicted (lines) annual Hellinger transformed abundance based on the results of the RDA model.

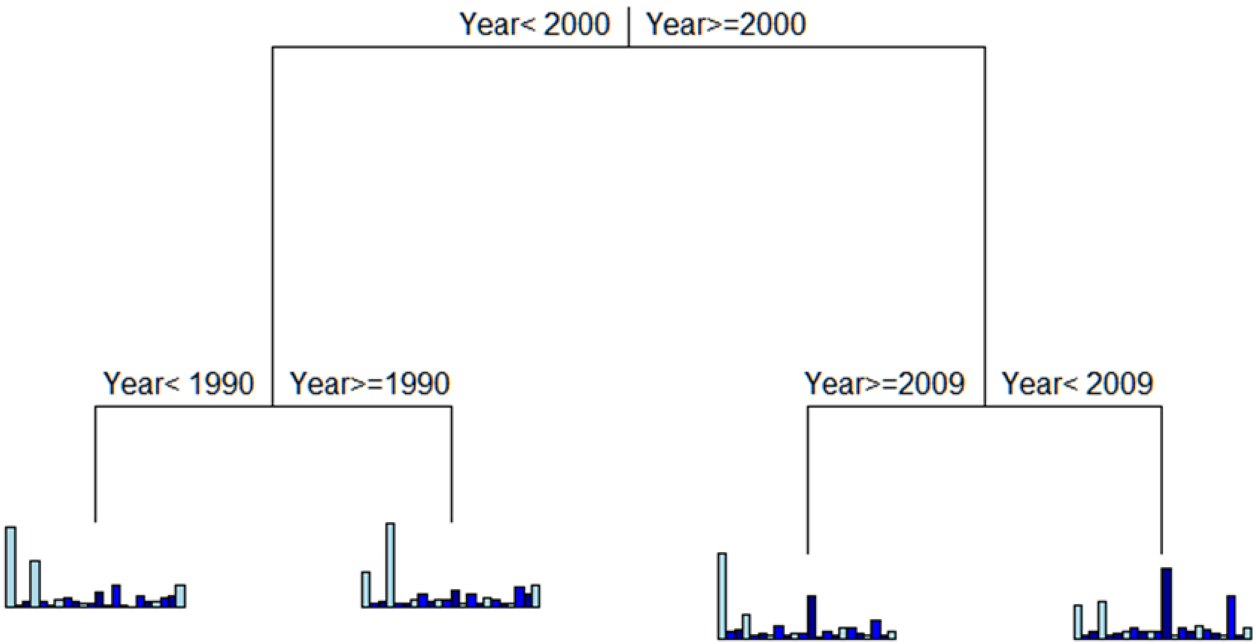

Figure 3-5. MRT analysis of the Peconic trawl data with year as an explanatory variable.

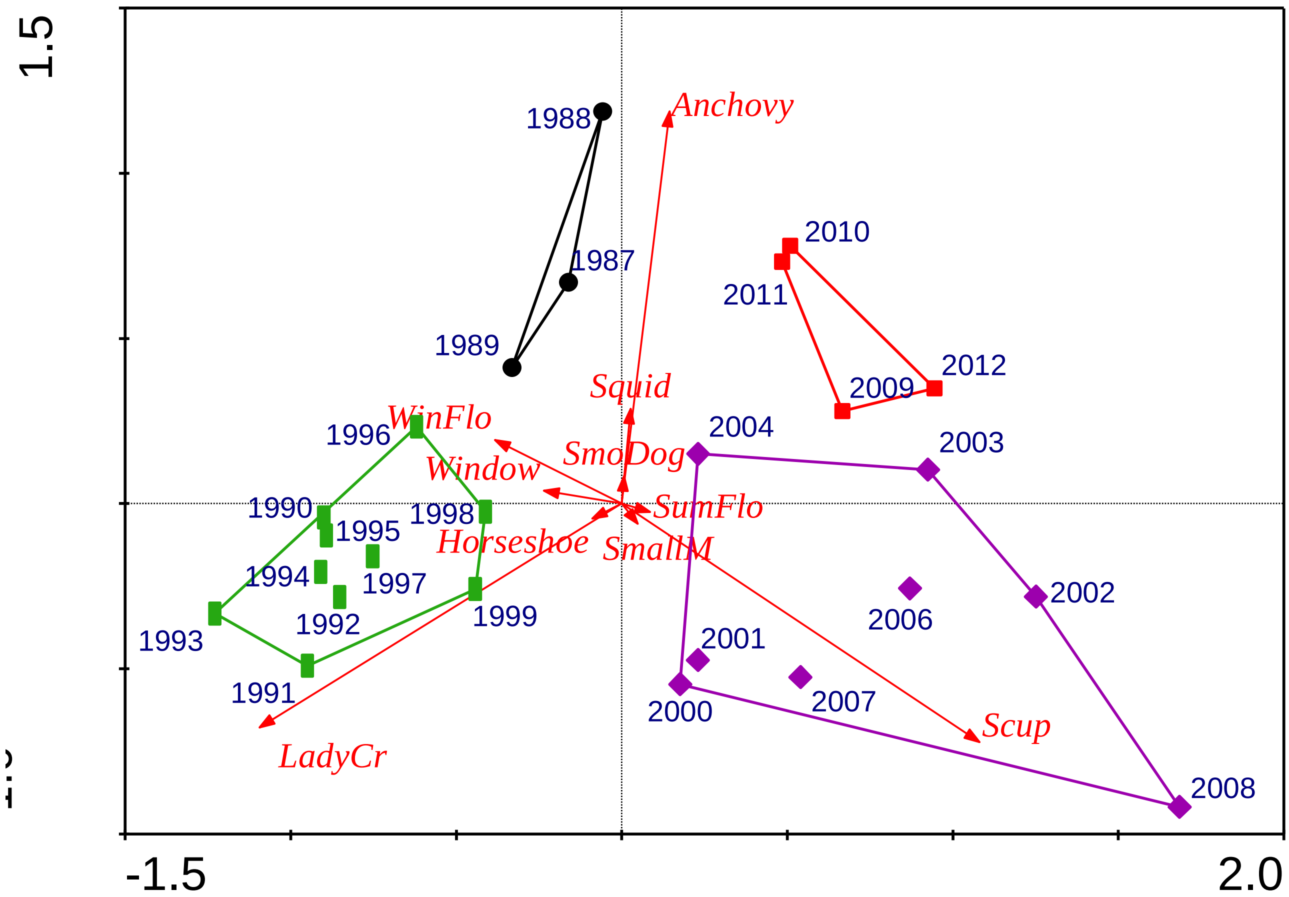

Figure 3-6. RDA and MRT results combined. Envelopes contain groups of years identified by MRT.

Objective 4: Link existing faunal and environmental data in an integrated analysis

The focus of this objective was to use existing faunal and environmental data to identify areas of specific use by finfish and mobile invertebrates in the Peconics Estuary and to create spatial habitat models for four species of YOY fish that are important to the Peconic Bay community (northern puffer, scup, weakfish, and winter flounder). To identify use areas, the New York State Department of Environmental Conservation's (NYSDEC) annual juvenile trawl survey data were used to map areas of high species richness and diversity. Further, GIS maps of young of the year (YOY) abundance were created to identify and visualize essential areas that might disproportionally contribute to the success of YOY. Trawl data, benthic faunal data, multibeam and side scan data, grain-size data, and water quality data were then utilized to construct generalized additive models to describe abiotic habitat preferences of the four YOY species. Full details of this process can be found in Abruzzo (2015) (Appendix C).

Methods:

a) Mapping

Shannon diversity and species richness were calculated for each tow to distinguish areas of the Bay that were high diversity areas for YOY fish. Diversity and species richness calculations included YOY northern puffer (Sphoeroides maculatus; <101mm), scup (Stenotomus chrysops; <84mm), weakfish (Cynoscion regalis; <153mm), winter flounder (Pseudopleuronectes americanus; <91mm), silversides (Menidia menidia; <61mm), bay anchovy (Anchoa mitchilli; <43mm), smooth dogfish (Mustelus canis; <391mm), tautog (Tautoga onitis; <51mm), butterfish (Peprilus triacanthus; <91mm), northern searobin (Prionotus carolinus; <51mm), striped searobin (Prionotus evolans; <71mm), smallmouth flounder (Etropus microstomus; <51mm), summer flounder (Paralichthys dentatus; <327mm), and windowpane flounder (Scophthalmus aquosus; <201mm). Trawl survey data from 2006-2012 were utilized since length data prior to 2006 were not available in database form to distinguish YOY from older fish.

Contour maps of diversity, species richness, and seasonal species abundance of four selected species (northern puffer, scup, weakfish, and winter flounder) were created in ArcGIS 10.1. Because the Peconic Estuary is broken up by spits and islands, a continuous interpolation method was deemed inaccurate; therefore, the kernel smoothing method was selected to carry out interpolation with barriers. An outline of the study area was used as the "barrier" in the contouring process. This outline also bounded the edge of the contours as ~305m (1,000ft) from land since the trawl survey was not designed to represent inshore fish populations.

b) Generalized additive models (GAM)

A two-stage GAM was created for each of the four YOY species (northern puffer, scup, weakfish, and winter flounder) and for each season. This two-stage process was necessary due to the large number of zero counts in the data (Welsh et al. 1996, Jensen et al. 2005, Sagarese et al. 2014). The first stage model (PA) was based on occurrence (presence-absence) and utilized a logit link function and binomial errors. The second stage model (ABUN) included only those tows that contained $\geq$ 1 individual of that particular species and was based on abundance expressed as the number individuals caught per tow (CPUE). This model utilized a log link function and negative binomial errors. Winter flounder YOY were rarely caught in the fall months, and therefore a GAM could not be created for this season due to the lack of data. Northern puffer, scup, and weakfish did not enter the Peconic Estuary system until late June, so spring season GAMs were not created for these species. In total, each species had PA and ABUND GAMs created for two seasons.

Explanatory variables taken from the trawl survey data included bottom temperature, bottom salinity, bottom dissolved oxygen, depth and secchi depth. Additionally, grain size data estimated during acoustic benthic mapping research was included in the analyses (Flood 2004; Cerrato & Maher 2007; and Cerrato et al. 2009, 2010). The sonar mapping study used side scan and multibeam sonar data to delineate bottom types. Each bottom type was subsequently sampled to estimate percent gravel, percent sand, percent mud, and sediment organic content (SOM). Average sediment characteristics of each bottom type region were estimated utilizing a spatial join in ArcMap of the bottom type polygons with the grain size point feature data. A second spatial join was then applied to assign the average grain size characteristics attributed to each bottom type to each trawl whose midpoint occurred within the bottom type polygon. Tows whose midpoint did not fall into bottom type areas were not used in the creation of the GAMs. Additionally, tows were only used in the GAM if they contained information for each explanatory variable (temperature, salinity, dissolved oxygen, depth and secchi depth, and grain size).

GAMs were created using the rationale in Sagarese et al. (2014) and the packages “mgcv” (Wood 2011) and “MuMIn” (Bartoń 2014) in R (R Core Development Team 2010). Cubic regression splines were used in each model, along with 5 pre-specified knots or 5 degrees of freedom (k=5). To validate how well the GAM was able to predict the occurrence or abundance of a particular species, tows were randomized and divided into a training set that contained 70% of the data to create the models. The other 30% was used later as “test” data to estimate how well the model predicted a species’ occurrence or abundance (Fielding & Bell 1997, Brotons et al. 2004, Sagarese et al. 2014). All GAMs were first built using all of the variables. PA and CPUE models took the form of:

$$p = s(Temp) + s(Sal) + s(Depth) + s(DO) + s(Sec) + s(Sand)+s(Mud)+s(SOM)$$

where p is the estimated probability of a species occurrence or abundance, s is the cubic regression spline, Temp is the bottom temperature (°C), Sal is the bottom salinity (ppt), Depth is the depth (meters), DO is the bottom dissolved oxygen (mg/L), Sec is the secchi depth (meters), Sand is the percentage of sand that constitutes the sediment, Mud is the percent mud found in the sediment, and SOM is the organic matter contained in the sediment (percent loss on ignition). Percent gravel was not used in the GAMs to prevent co-linearity problems. The “dredge” function in the R package “MuMIn” was then applied to model every potential combination of explanatory variables. The model with the lowest Akaike Information Criterion (AIC) was selected as the best model.

Results:

a) Mapping

There were 1,888 tows (443 spring tows, 741 summer tows, and 704 fall tows) used in the creation of the species abundance and diversity maps. Across all seasons, the spring months were the least diverse in terms of species richness and the Shannon diversity index (Figures 4-1, 4-2). Areas that contained the highest species richness (1-2 species) in spring included the middle and western parts of the Great Peconic Bay, as well as the western portion of the Little Peconic Bay. The summer months contained the highest species richness and diversity values relative to other seasons. Areas with the highest species richness (4-5) occurred around Nassau Point. The same can be said for diversity, along with other peaks in Little Peconic Bay, northern portion of the Southold Bay, and southern parts of Noyack Bay.

Species richness and diversity in the fall were intermediate between spring and summer values. The highest species richness values (3-4) were observed in the middle portion of the Great Peconic Bay, in the center portion of the Little Peconic Bay, and areas in Southold and Noyack Bay. Diversity contours revealed the same general pattern found in the species richness results.

Northern puffer YOY were collected throughout the Peconics Estuary (Figure 4-3). In the summer, they were caught in higher abundances in Cutchogue Harbor, Nassau Point, Hog Neck Bay, and the southern regions of the Great and Little Peconic Bay. They were also found in Flanders Bay, Noyack Bay, and Southold Bay, but in relatively less abundance compared to other areas. In the fall, YOY northern puffer were less abundant than what was observed in the summer. They were found in higher abundance in Noyack Bay and just west of Jessup Neck, as well as in southeastern Great Peconic Bay.

YOY scup were rarely found before the beginning of July. In the summer, juveniles were found throughout the Peconics, from Shelter Island to Flanders Bay (Figure 4-4). However, greater abundances (~15-20) of juveniles were found east of Robins Island. In Great Peconic Bay, more juveniles were caught in the southern portion and around the edges, than in the northern and central areas. In the fall, juveniles were still in high abundance throughout the bay. A greater number (~10-15) were found in the eastern portion of the system compared to Great Peconic Bay and Flanders Bay (~6-8).

Weakfish distribution showed distinct offshore areas of occurrence (Figure 4-5). In the summer and the fall weakfish YOY were caught in almost every tow from the middle of the Great Peconic Bay and Little Peconic Bay. They were also caught around in Noyack Bay and Southold Harbor. Weakfish were found in higher abundances in the summer compared to the fall.

Winter flounder were more abundant in the spring months and inhabited the western portions of the Peconics (Figures 4-6). In the spring, winter flounder occurred in the central portion of the Great Peconic Bay, Cutchogue Harbor, and along the southern portions of Little Peconic Bay and Noyack Bay. In the summer, fewer winter flounder were found, and they were primarily caught in the northwest region of Great Peconic Bay. A few individuals were also found in locations where they were abundant in the spring.

b) Generalized additive models (GAMs)

Presence-absence and abundance models for the four taxa yielded distinctly different results. The AIC selected presence-absence models had from 3-6 environmental variables with an average r^2 value of 0.24 $\pm$ 0.07 (1 sd). Temperature (6 of 8 models), salinity (6/8), depth (5/8), and % Mud (5/8) were the most frequently selected variables. DO (2/8) and Secchi depth (2/8) were the least selected. The AIC selected abundance models had consistently lower r^2 values (0.14 $\pm$ 0.08) and from 1-5 environmental variables. Salinity (5/8) and depth (4/8) were the environmental variables selected most frequently. DO (0/8) and Secchi depth (2/8) were the least frequently chosen. Unlike presence-absence, there was a seasonal component to the variables chosen for abundance. The models for summer abundance were dominated by sediment variables (% sand, % mud, and SOM), while the models for fall were dominated by temperature, salinity, and depth.

Discussion:

Mapping results indicated very distinct spatial patterns in species richness, diversity, and the distribution of puffer, scup, and in particular weakfish and winter flounder. Temperature, salinity, depth, and grain-size appear to drive local spatial distributions in the bay based on the GAM results.

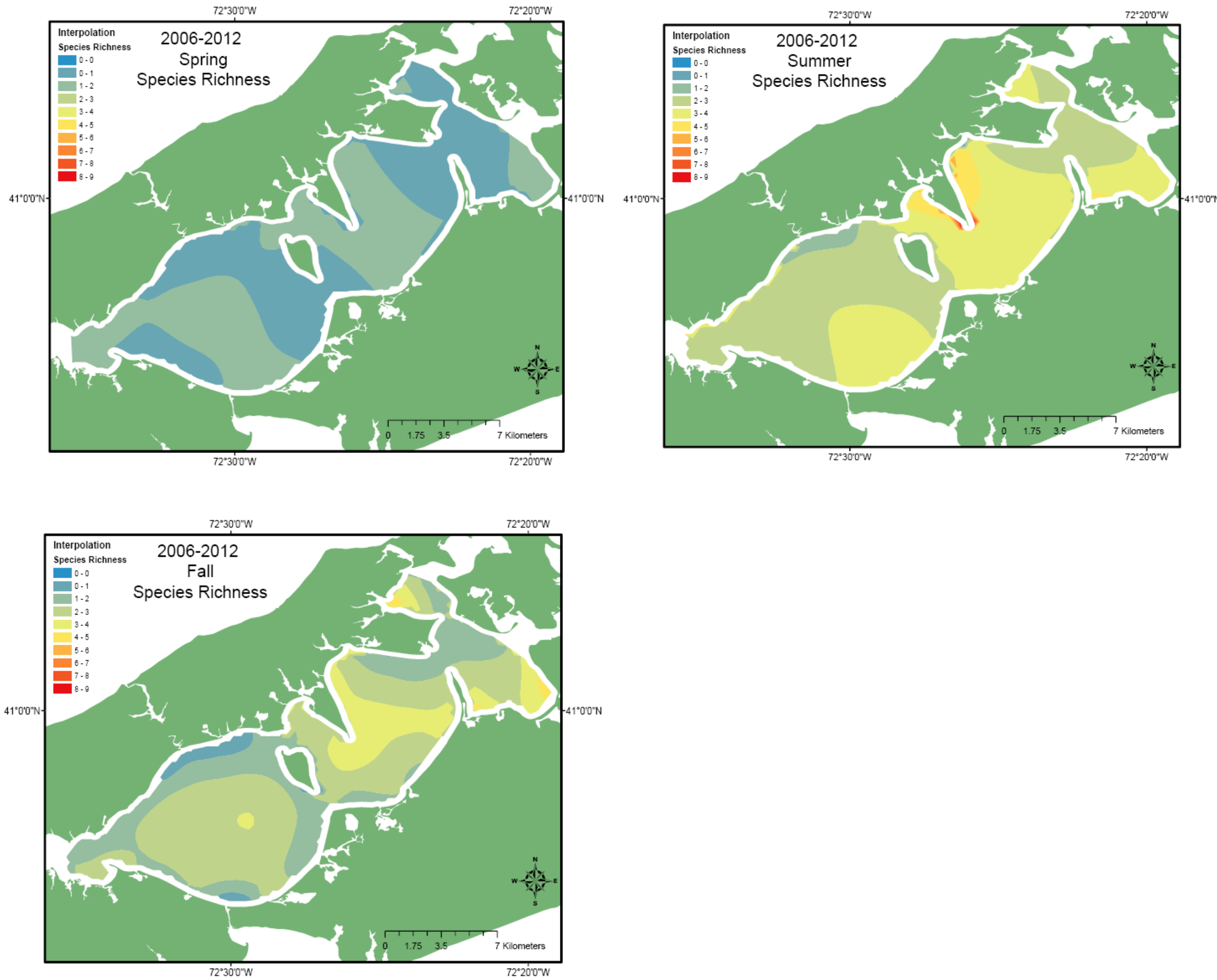

Figure 4-1. Seasonal species richness from 2006-2011.

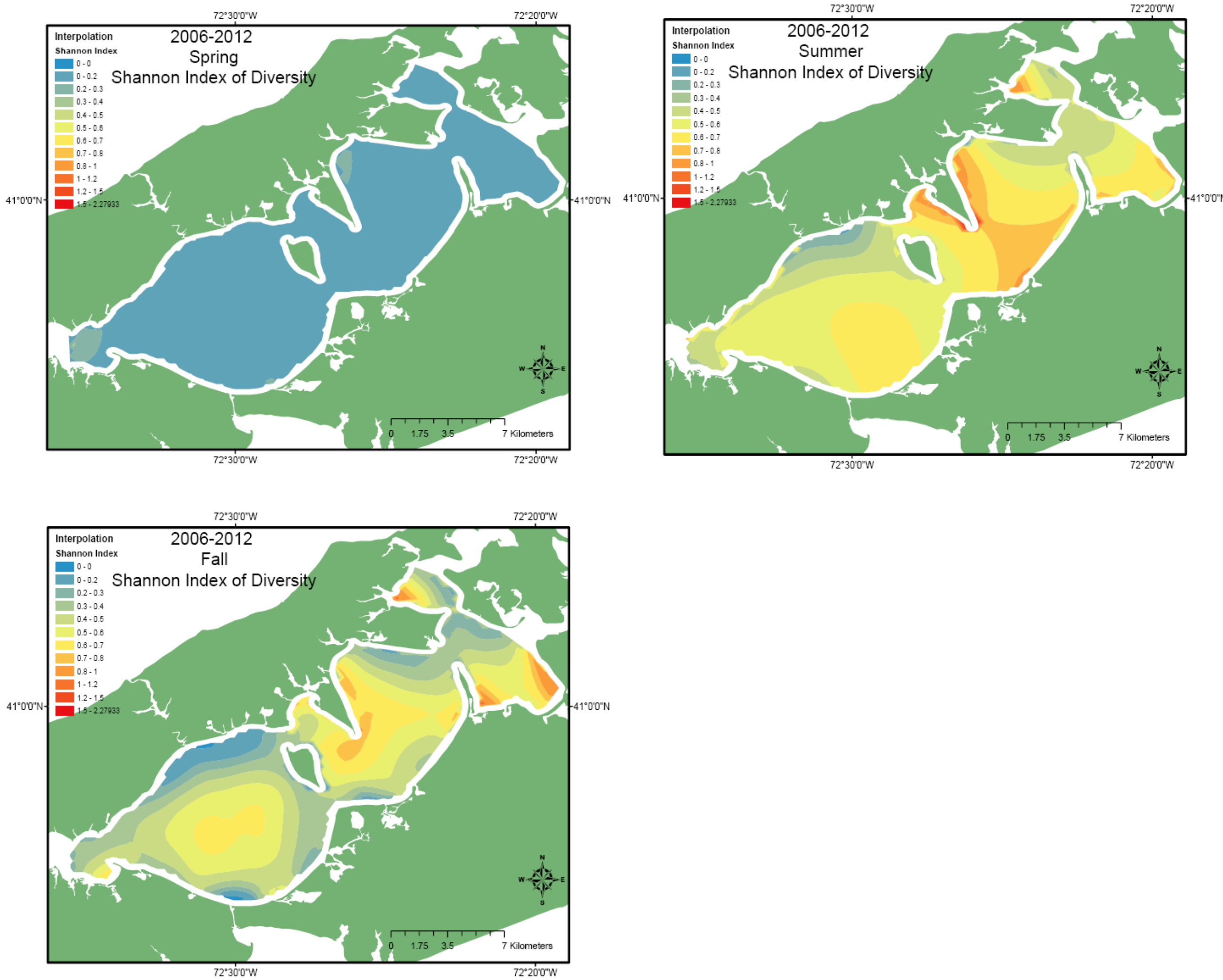

Figure 4-2. Seasonal Shannon diversity from 2006-2011.

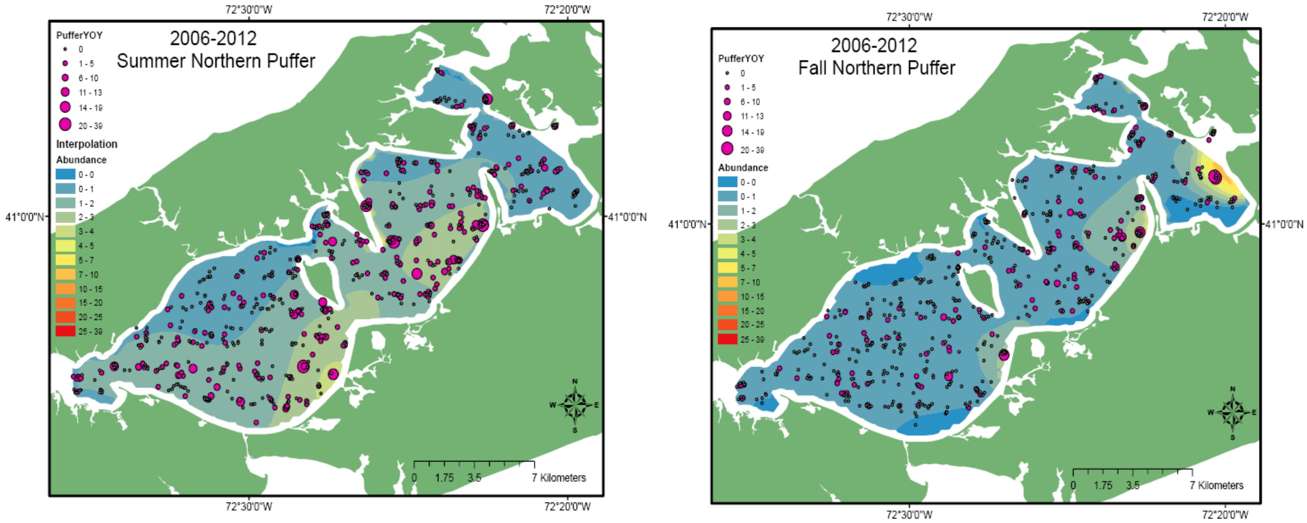

Figure 4-3. Seasonal northern puffer YOY abundance from 2006-2011. Trawls represented as small grey dots contained no YOY. Purple circles specify the abundance of YOY that were caught in that tow. Interpolation results are also mapped.

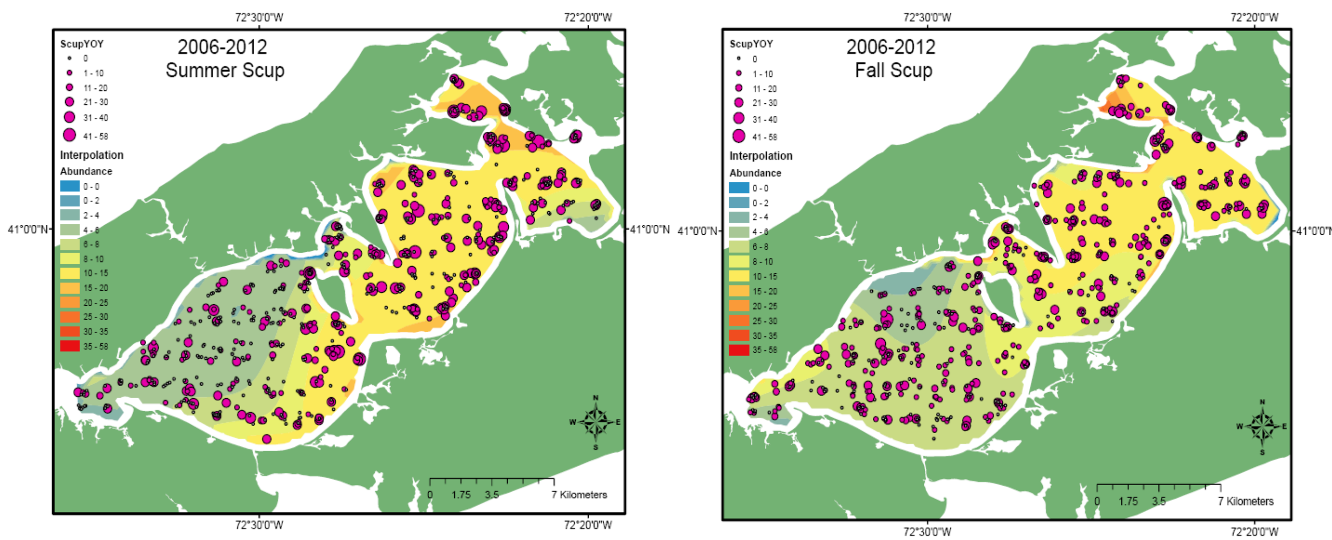

Figure 4-4. Seasonal scup YOY abundance from 2006-2011. Trawls represented as small grey dots contained no YOY. Purple circles specify the abundance of YOY that were caught in that tow. Interpolation results are also mapped.

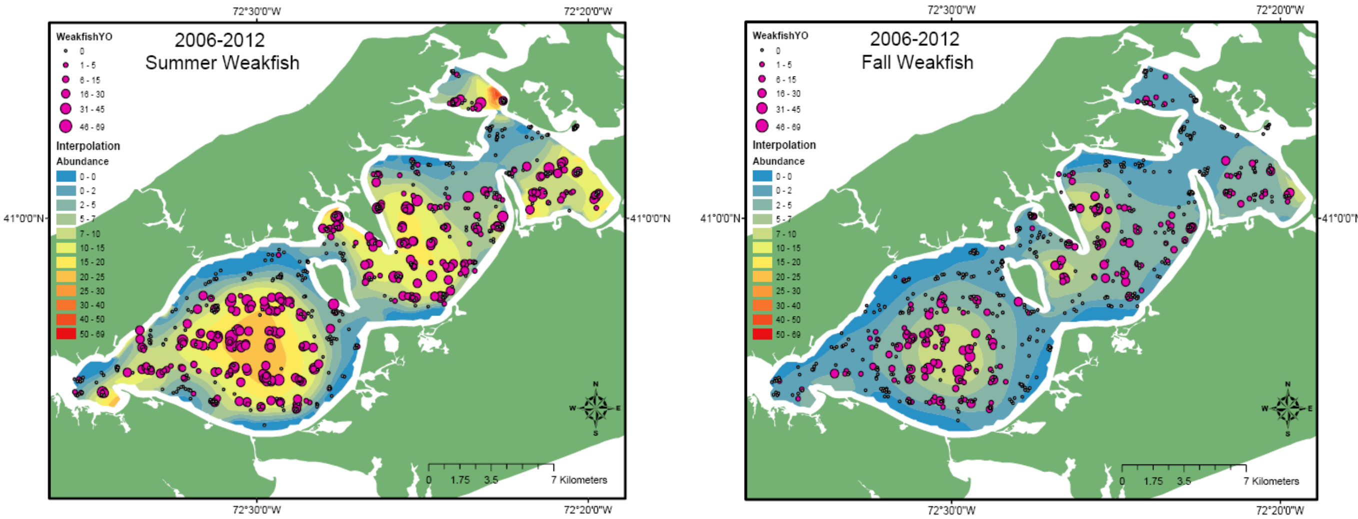

Figure 4-5. Seasonal weakfish YOY abundance from 2006-2011. Trawls represented as small grey dots contained no YOY. Purple circles specify the abundance of YOY that were caught in that tow. Interpolation results are also mapped.

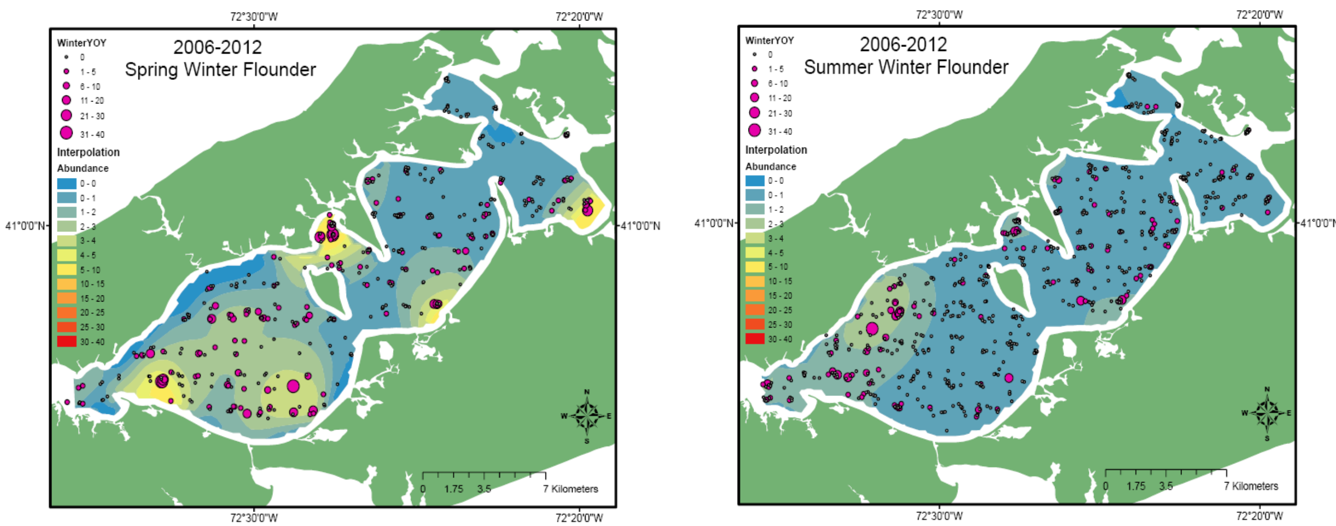

Figure 4-6. Seasonal winter flounder YOY abundance from 2006-2011. Trawls represented as small grey dots contained no YOY. Purple circles specify the abundance of YOY that were caught in that tow. Interpolation results are also mapped.

Summary

The results of this study have a number of clear management implications. Analyses of benthic macrofaunal communities demonstrated that sonar mosaics and benthic faunal sampling was very effective in identifying benthic habitats. A simple set of depth, grain-size, cover, and acoustic province variables explained 83% of the between-province variance in community structure for Robins Island and 73% for Shelter Island after adjusting for the small-scale variability that was below the resolution of the surveys. The process does, therefore, differentiate benthic habitats with high accuracy; however, it also requires ground truth data to be effective. The model selection process combined some acoustic provinces into larger groups. For example, contiguous acoustic provinces B and C were combined into one faunal assemblage and so were the noncontiguous acoustic provinces D and F (Figures 2-7, 2-10). Visual segmentation of the acoustic data was conservative, therefore, and resulted in more provinces than were justified by the community structure results. In addition, a large fraction of community variation (36-37%) was below the resolution of the surveys and within-province variation was only moderately explained (42% for Robins Island and 23% for Shelter Island). This variability would not impact initial management decisions regarding habitat identification, but it does highlight the challenges that exist to sensing change in monitoring data and in understanding the heterogeneity present in benthic communities. Progress in improving the quality of sonar mosaics, refining image segmentation/classification methods, and targeting sampling to better characterize small-scale and within-province variation should help.

The fish and mobile invertebrate community sampled in the juvenile trawl survey is nonstationary and seems to change on a decadal timescale (Figures 3-2, 3-6). This pattern of change is only evident from analysis of time series data emphasizing the value of long-term monitoring to understand the structure of the ecosystem. Change was primarily driven by regional factors that also influenced Long Island Sound and Narragansett Bay. Although NAO and AMO indices were important in explaining shifts in community structure, they should be regarded as proxies for other processes regulating the finfish and macroinvertebrate community. The regulatory drivers could have both bottom-up and top-down origins. It is just as likely, for example, that species are responding to shifts in temperature as it is that they are responding to changes in predators and competitors. Management decisions within the Peconic Estuary and even at the bay/harbor level should, therefore, incorporate some element of the regional context of the community structure and recognition of the high degree of species associations present.

There were spatial hotspots or concentrations of species richness, diversity, and abundances of individual species present in the 2006-2012 YOY data. For example, high Shannon diversity was found near Nassau Pont and in several locations in Southold and Noyack Bays. Winter flounder YOY had distinct spatial distributions suggesting localized recruitment areas, and weakfish were distributed in the middle of the Great Peconic Bay, Little Peconic Bay, Noyack Bay, and Southold Harbor. Whether these patterns persist prior to or beyond the 2006-2012 period is not known, but they may be important in identifying areas of significance.

A variety of follow-up studies can be suggested utilizing the existing data. At the population level, inputting length data for 1987-2006 would greatly expand YOY analyses and make it possible to better understand the temporal and spatial patterns present in the data and their relationship to environmental factors. Analysis of spatial and temporal patterns of abundant invertebrates such as lady crabs, spider crabs, and horseshoe crabs would also prove interesting. It would also be useful to extend data analysis to finer temporal scales. For example, species specific timing of movement patterns into and out of the Peconics with respect to temperature changes would offer insights for migratory species. The rich spatial-temporal structure in the juvenile trawl survey data would also allow detailed analysis of abundance-occupancy (AO) relationships in individual species. These AO relationships coupled with life history characteristics such as longevity, growth, and age at maturity, fishery status, level of exploitation, migratory behavior, temperature range, feeding guild, etc. may allow identification of species most vulnerable to further environmental change. Although not using existing data, analysis of genetic variations in YOY winter flounder at abundance hotspots may suggest whether patterns of site fidelity exist. If found, the existence of generic structure would impact local management of this species.

At the community level, examination of OA changes in species assemblages could provide insights into the stability of different assemblages in relation to environmental variation. This analysis could be carried out for each time period identified in the juvenile trawl data (i.e., 1987-1989, 1990-1999, 2000-2009, and 2010-2012) to potentially identify the mechanisms creating the change. It could also be developed at small spatial-temporal scales through second stage community analysis (Clark et al 2006) which would provide an approach to derive measures of variation for each of the 77 grid units in the trawl survey. The trawl survey data set would also be amenable to dynamic spatial-temporal trophic modeling through Ecopath with Ecosim.

At the habitat level, results of the spatial analysis of benthic macrofauna indicated that progress is needed in improving the quality of sonar mosaics, refining image segmentation/classification methods, and targeting sampling to better characterize small-scale and within-province variation. Exploration of the backscatter and bathymetry characteristics of image objects identified by the eCognition analysis would be a good starting point. Analysis of the benthic macrofauna also suggested that the empirical variogram for each site (Figures 2-9 to 2-12) should be regarded as a complex composite formed from multiple within and between habitat spatial relationships. It is possible that community assemblages within each province have unique and perhaps quite different spatial structures leading to differences in variogram characteristics. Decomposing a large-scale variogram containing multiple habitats into its component parts could examine whether this occurs. The degree to which spatial relationships are unique to or common among specific habitats would provide insights into fundamental ecological relationships.

References:

Abruzzo TR. 2015. Analyzing Spatial and Temporal Trends in the Community Structure of the Peconic Bay Estuary. M.S. thesis, School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY.

Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. 2nd International Symposium on Information Theory (B. N. Petrov and F. Cztiki, eds), pp. 267-281. Budapest: Akademiai Kiado.

Bachmaier M and Backes M. 2008. Variogram or semivariogram? Understanding the variances in a variogram. Precision Agriculture 9: 173-175.

Burnham KP and Anderson DR. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd Edition. Springer, New York.

Cerrato RM and Maher NP. 2007. Benthic Mapping for Habitat Classification in the Peconic Estuary: Phase I Ground Truth Studies. MSRC Special Report Number 134. Marine Sciences Research Center, Stony Brook University, Stony Brook, NY.

Cerrato RM, Flood RD, and Holt LC. 2009. Benthic Mapping for Habitat Classification in the Peconic Estuary: Phase II Ground Truth Studies. Final Report to Suffolk County Office of Ecology. Marine Sciences Research Center Special Report No. 136, Stony Brook University, Stony Brook, NY, 93 pp.

Cerrato RM, Flood RD, and Holt LC. 2010. Benthic Mapping for Habitat Classification in the Peconic Estuary: Phase III Ground Truth Studies. Final Report to Suffolk County Office of Ecology. Marine Sciences Research Center Special Report No. 137, Stony Brook University, Stony Brook, NY, 96 pp.

Cressie NAC. 1993. Statistics for Spatial Data. John Wiley and Sons, New York. 900 pp.

De'Ath G. 2002. Multivariate regression trees: a new technique for modeling species-environment relationships. Ecology 83:1105-1117.

Dungan JL, Perry JN, Dale MRT, Legendre P, Citron-Pousty S, Fortin M-J, Jakomulska A, Miriti M, Rosenberg MS. 2002. A balanced view of scale in spatial statistical analysis. Ecography 25: 626–640.

Flanagan AM. 2016. Quantitative Benthic Community Models: The Relationship Between Explained Variance and Scale. Ph.D. thesis, School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY.

Flood RD. 2004. Benthic Mapping for Habitat Classification in the Peconic Estuary: Side Scan Sonar and Multibeam Mapping. Final Report to The Nature Conservancy. Marine Sciences Research Center, Stony Brook University, Stony Brook, NY.

Fonseca L and Mayer L. 2007. Remote estimation of surficial seafloor properties through the application Angular Range Analysis to multibeam sonar data. Marine Geophysical Research 28: 119-26.

Hunt Jr, GL, Stabeno P, Walters G, Sinclair E, Brodeur RD, Napp J M, & Bond NA. 2002. Climate change and control of the southeastern Bering Sea pelagic ecosystem. Deep Sea Research II 49:5821-5853.

Jensen OP, Seppelt R, Miller TJ, & Bauer LJ. 2005. Winter distribution of blue crab Callinectes sapidus in Chesapeake Bay: application and cross-validation of a two-stage generalized additive model. Marine Ecology Progress Series 299:239-255.

Jongman RHG, ter Braak CJF, and Van Tongeren OFR. 1995. Data analysis in community and landscape ecology. Cambridge University Press, New York.

Jordaan A, Frisk MG, Wolff NH, Incze LS, Hamlin L, and Chen Y. 2007. Structure of fish assemblages along the Northeastern United States based on trawl survey data: indicators of biodiversity and a basis for ecosystem and area-based management. ICES CM 2007/A:05

Journal AG, Huijbregts CJ. 1978. Mining Geostatistics. Academic Press, New York, 600 pp.

Legendre P. 1993. Spatial autocorrelation: Trouble or new paradigm? Ecology 74(6): 1659-1673.

Legendre P and Gallagher ED. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia 129: 271-280.

Legendre P and Legendre L. 2012. Numerical Ecology, 3rd English Edition. Elsevier, Amsterdam. 990 pp.

Maher NP. 2006. A New Approach for Benthic Biotope Identification and Mapping. Ph.D. thesis, School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY.

Möllmann C, Müller-Karulis B, Kornilovs G, & St John MA. 2008. Effects of climate and overfishing on zooplankton dynamics and ecosystem structure: regime shifts, trophic cascade, and feedback loops in a simple ecosystem. ICES Journal of Marine Science 65:302-310.

Nuzzi R, Waters RM. 1998. Surface Water Quality Monitoring Report 1976-1996. Volume I, Narrative; Volume II, Data. Suffolk County Department of Health Services, Office of Ecology, Riverhead NY.

Sagarese SR, Frisk MG, Cerrato RM, Sosebee KA, Musick JA, & Rago PJ. 2014. Application of generalized additive models to examine ontogenetic and seasonal distributions of spiny dogfish (Squalus acanthias) in the Northeast (US) shelf large marine ecosystem. Canadian Journal of Fisheries and Aquatic Sciences 71:1-31.

Shumchenia and King J. 2010. Comparison of methods for integrating biological and physical data for marine habitat classification. Continental Shelf Research 30:1717-1729.

Suffolk County Department of Health Services (SCDHS). 2011. Surface water quality monitoring data, SCDHS Office of Ecology, Yaphank, N.Y.

Wagner HH. 2003. Spatial covariance in plant communities: Integrating ordination, geostatistics, and variance testing. Ecology 84(4): 1045-1057.

Wagner HH. 2004. Direct multi-scale ordination with canonical correspondence analysis. Ecology 85(2): 342-351.

Wagner HH, Fortin MJ. 2005. Spatial analysis of landscapes: Concepts and statistics. Ecology 86: 1975-1987.

Weber A, Grahn C, & Havens B. 1998. Species Composition, Seasonal Occurrence and Relative Abundance of Finfish and Macroinvertebrates Taken by Small-Mesh Otter Trawl in Peconic Bay, New York. New York State Department of Environmental Conservation, Division of Fish, Wildlife and Marine Resources, Marine Finfish Unit, East Setauket, New York, 127 pp

Welsh AH, Cunningham RB, Donnelly CF, & Lindenmayer D B. 1996. Modelling the abundance of rare species: statistical models for counts with extra zeros. Ecological Modelling 88:297-308.

White MP. 2015. Marine Acoustics and Habitat Mapping in the Peconic Estuary, NY. M.S. thesis, School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY.

Appendices

Appendix A:

White MP. 2015. Marine Acoustics and Habitat Mapping in the Peconic Estuary, NY. M.S. thesis, School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY.

Appendix B:

Flanagan AM. 2016. Quantitative Benthic Community Models: The Relationship Between Explained Variance and Scale. Ph.D. thesis, School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY.

Appendix C:

Abruzzo TR. 2015. Analyzing Spatial and Temporal Trends in the Community Structure of the Peconic Bay Estuary. M.S. thesis, School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY.