NEFSC_HabCam and HabCam V2

TOR 1. Review the statistical design and data collection procedures for each survey system
1.1 Imaging the seafloor
1.2 At Sea Data Acquisition
1.3 Image Annotation
1.4. Survey Design

1.1 Imaging the seafloor
The idea of taking images of the sea floor, and specifically photographically surveying scallops, dates back many decades (e.g., Posgay 1958, Hersey 1967, Wigley and Emery 1968, Caddy 1970). What is new is digital imaging and the ability to handle, process, and store huge amounts of data from towed or autonomous underwater vehicles. HabCam is a towed camera system that is towed at 5-7 knots (~3 m/s) while rapidly taking still, overlapping, digital images of the sea floor (~6/sec) and simultaneously collecting environmental data. HabCam employs 16 bit color machine vision digital cameras and high speed xenon strobes that are synchronized with the cameras to provide very short (~10 µs) exposures to eliminate motion blur even when towing at high speeds. It was specifically designed to rapidly survey large areas; a broadscale HabCam survey of the two main U.S. sea scallop regions (Mid-Atlantic and Georges Bank) takes approximately three weeks. Using an image annotation tool, we are currently able to manually identify over 450 taxonomic categories and classify substrate to 51 categories. We are currently developing automated segmentation software that will indentify substrate, scallop, finfish and other targets that can make full use of the acquired HabCam imagery.

There are currently two HabCam vehicles in operation: HabCamV2 is the original system, built through a collaboration of the Woods Hole Oceanographic Institution (WHOI), commercial fishermen and independent scientists, to address the need for non-invasive optical towed vehicle of scallop and habitat. The V2 vehicle is equipped with a single color machine vision camera and a limited environmental sensing package including a CTD. The second system HabCamV4, built by WHOI for NOAA, consists of a pair of high resolution color machine vision cameras to form a stereo imaging system coupled with a C3D side scan acoustic system to acquire synoptic information of seabed rugosity and bottom fishing activity. Its package of environmental sensors include CTD, chlorophyll, turbidity, pH, dissolved oxygen, optical plankton counter, and in-water spectrometer. Other than these sensors that are specific to HabCamV4, the use and applications of the two HabCam systems are similar. In this document we will address the general aspects of both systems and detail differences where necessary. The history of the development of these vehicles is addressed in Gallager (2013) in Appendix I of this report.

1.2 At Sea Data Acquisition

The HabCam imaging systems (Howland et al. 2006; Taylor et al. 2008; York et al. 2008) is "flown" 1.5 to 2.5 meters above the seafloor at 5 to 7 knots (~2.5 – 3 m/sec), so that a track approximately 100-120 nautical miles is imaged each 24 hours of operations. Optical imagery is collected at a width of approximately 0.75 to 1.25 meters (total ~170,000 - 260,000 square meters/day) and at a rate of 6 images per second, providing about 50% overlap to aid in mosaicing continuous strips. On the NOAA HabCamV4 system, side by side stereo pair images are fused into a single image at the time of acquisition allowing precise stereo referencing with metadata such as latitude, longitude, temperature, salinity, chlorophyll, wavelength-specific light absorption, dissolved oxygen, and other environmental data. In addition, the Teledyne Benthos interferometric C3D side scan acoustic system on the NOAA vehicle collects 3D bathymetry and backscatter to a range of ~50m on either side of the vehicle.

Images collected at sea by HabCamV2 are brought back to shore where they are color and light-field corrected. No additional image processing is conducted before human annotators begin counting and measuring scallops and identifying substrate.

The at-sea image processing workflow for HabCamV4 is considerably more extensive and consists of the following elements (refer to Fig 1.2) and described in depth in Gallager et al. 2015):

1) Image acquisition: Stereo pairs of 16 bit Bayer encoded raw images are transmitted up a fiber optic cable at 6 Hz and recorded on a 32TB RAID 5 data server as a temporally and spatially matched stereo pair.

2) Light field correction: As in all underwater imagery, HabCamV4 images have a light pattern due to the light from the four strobes not being perfectly distributed across the seafloor. To correct for this uneven illumination, as raw images are recorded, they are subsequently accessed, images are demosaiced using a Bayer decoding algorithm, and light field corrected using a procedure developed by RPI student Jason Rock and later modified by Peter Honig and Joe Futrelle at WHOI. Briefly, a light field model is built using the average of 1000 or more images over a variety of substrates and water conditions. For each new image as it arrives, it is multiplied by the light field model. This removes the low frequency variation in light unevenness very effectively. Color is then corrected for the selective attenuation of colors by sea water. The resulting light and color corrected images are then written back to the database as uncompressed png images.

3) Annotation: These png images are then set up in an assignment for annotation by trained people on the ship and back on land (see TOR 4).

4) Stereo processing (Fig. 1.3): The png images are further processed for stereo rectification using a calibration matrix of intrinsic and extrinsic parameters previously collected during calibration in a standing water tank at WHOI. The calibration matrix is applied to each image pair to provide a rectified and de-warped pair, which is subsequently recorded in the database, followed by creation of a 3D point cloud of pixels in fully three dimensional space. The 3D point cloud is then recorded to the database and used in subsequent measurements of swimming scallops, fish, and bathymetric features that cannot be measured in 2D (See TOR 2 for discussion of measurements in 2D versus 3D).

5) Seabed rugosity: In the process of calculating a 3D point cloud, the rugosity, slope, and gradient of each image pair is recorded to the database. Rugosity is defined as the ratio of the true 3D surface area calculated from the point cloud divided by the projected 2D surface area, which is the Field of View (FOV) of the camera pair (typically 0.5 to $1m^2$). Rugosity is used in further characterization of habitat (see TOR 6).

6) Sensor data: Sensors are included on HabCamV4 to measure environmental characteristics such as temperature, salinity, dissolved oxygen (SeaBird 37), high speed temperature and salinity at 10Hz (SeaBird 49), chlorophyll, turbidity, CDOM (WetLabs EcoPuck), wavelength-specific light attenuation (Hobie Labs A-Sphere), pH (Satlantic), and plankton using a high resolution, color microscope imaging system (CPICS WHOI). Data from each of these sensors is added to the database and time matched with each image as it is recorded.

7) Substrate composition: Substrate is autonomously calculated using machine vision algorithms recently published by Honig et al. (2015) and this information is recorded as the percentage cover of mud, sand, gravel, cobble, boulder, shell hash and any combination thereof (see TOR 6 for use of automated substrate characterization in habitat assessment).
Upon completion of a survey leg, the 32TB at-sea server is brought back to the lab where the processed png images are made available through a web based annotation tool developed by York (2012).

1.3 Image Annotation

Identification and measurement of scallops in HabCam V4 images from 2012-2014 have been annotated using a web-based tool developed with funding from NOAA and the Gordon and Betty Moore Funded Image Informatics Project at WHOI (Fig 1.4). The annotation tool was built by collaborative interdisciplinary teams with expertise in oceanographic imaging and informatics. The approach to development of the annotation tool was the Tetherless World Constellation's iterative development process (Fig. 1.5). The web-based annotation tool allows users to identify and measure sea scallops, other invertebrates and finfish, as well as the dominant and subdominant substrate in the image. The annotation tool can support to multiple users annotating images at the same time and at different locations. Information models and activity

diagrams were constructed based upon this use case (Fig 1.6 and 1.7). Since its initial prototype in 2012, a number of innovations have been added to the annotator tool to make annotation more efficient and effective including buttons on the GUI for commonly-seen classes or organisms and 3D on-demand, which will bring up a 3D red-cyan version of the currently viewed image on a different tab in the web browser to better provide context for identifying objects in an image.

1.4 Survey Design

Basic concepts and simulations

Because the HabCam vehicle collects a constant stream of images, data derived from the images are autocorrelated and so traditional random or stratified random survey designs are not applicable. Resource assessments from such data typically use spatial models including Generalized Linear Models (GLMs), Generalized Additive Models (GAMs) and geostatistical methods such as kriging (Rivoird et al. 2008 and see TOR-4). Literature on sampling designs for this type of survey comes primarily from acoustic surveys. With geostatistical methods, the uncertainty of the estimate at any given location in the survey domain increases with distance from the survey track. As a result, evenly spaced grids are optimal for acoustic surveys as the distance from the survey track is minimized with even spacing, assuming that the mean and variance are homogeneous throughout the modeled domain. While this assumption is traditionally applied to acoustic data, it is often incorrect for sea scallops (see TOR-4).

The HabCam sea scallop survey differs from this situation because there are persistent trends in scallop abundance due to depth, management measures and known habitat affinities. While simple geostatistical methods assume a landscape with a stationary mean, similar to Figure 1.8a, a landscape with a higher mean along the center of the landscape (Figure 1.8b) is more realistic for sea scallops, due to an underlying trend for scallop densities to be highest at intermediate depths (Figure 1.8c). In this case, it may be advantageous to increase sampling effort in the core habitats along the center of the survey area. Given a survey track of evenly-spaced transects of equal length (Figure 1.9a) and assuming an underlying variogram model, we can derive a map of kriging variances for the survey at each location in the landscape (Figure 1.9b). If the mean density is higher in the center of the landscape instead of stationary, we may assume that the standard deviation of the mean is proportional to the mean (similar to a Gamma distribution) and calculate an adjusted kriging variance for each location as:

$$AdjVar_{x,y} = KrVar_{x,y} * [e^{(CE)}]^2$$

where $AdjVar_{x,y}$ is the adjusted variance of the estimate at a given location, $KrVar_{x,y}$ is the unadjusted variance at the location and $e^{(CE)}$ is the magnitude of the center effect from Figure 1.8c.

As a proof of concept, we used geostatistical simulations to examine the effect of allowing the mean (and variance) to vary across the shelf and longitudinally along the shelf. We first simulated varying the mean across the shelf and examined how the survey variances were affected by varying (1) the proportion of the effort concentrated along the center of the survey area and (2) the length of the survey track. We modeled the cross-shelf gradient as a double-logistic with higher densities along the center of the study area and the amplitude of the center effect varying from 0 (no effect) to 1 (maximum of e^2 higher in the middle at effect = 1, Figure 1.10). To assess the effect of increasing sampling intensity along the center of the study area, we decreased the length of alternating transects (range from 0 – 100% of the total width of the study area) and increased the total number of transects to keep the total survey track length constant (i.e. Figure 1.11). We then varied the total survey track length from 1,000 to 4,000 units. For each simulation, we examined the resulting variance maps (i.e., Figure 1.12) and used the sum of the adjusted kriging variance as a relative proxy for the variance of the survey. While this is not the true variance of the survey, as the variances are correlated across the landscape, we are not aware of established methods for calculating a kriging variance for survey areas with non-stationary variances and this should be an effective relative measure for comparison purposes.

The adjusted kriging variances varied with different center effects and transect lengths (Figure 1.13). Optimal short transect lengths decreased as center effects increased and increased as total track length increased. The effects of center effect and total track length interacted on optimal short transect length. With a track length of 1,000 units, increasing the center effects from 0 and 1 decreased optimal short transect length from 67% to 30%. However, for track lengths of 4,000 pixels, varying the center effect from 0 to 1 only decreased optimal short transect length from 92% to 85%.

For a second simulation, we examined the effect of the mean and variance varying longitudinally along the survey area (i.e. zonal anisotropy, example Figure 1.14). The zonal effect was implemented by dividing the landscape into two zones (upper and lower) and adding an additional, longitudinally-oriented logistic trend to the landscape. We then varied the amplitude of the longitudinal effect (zone effect) the spacing of adjacent transects between the two zones, and total track length (Figure 1.15).

The optimal solutions for landscapes with zone effects placed more transects in the zone with higher underlying means and variances (Figure 1.16). The effect was most notable for shorter total track lengths, increasing transect density in the higher mean zone by as much as 300% over the lower mean zone.

Design Implementation

The above simulations indicate that the variance of a survey can be decreased by alternating full-length transects with shorter transects and increasing transect density in areas with known higher abundances. However, because the survey is not optimized to a single aspect of the scallop stock

(commercial biomass, recruitment, etc.), these simulation results are used informally in the design of each year's survey with the final survey design is based on researchers' knowledge of where the current stock biomass and incoming cohorts are. The two stock areas (MAB and GB) are each divided into multiple subregions, based on changes in habitat type, habitat orientation (anisotropy), and management boundaries (Figure 1.17 and 1.18). These subregions are used both for designing the survey and for abundance estimation from the resulting survey data (TOR-4). The extent of the survey area is based on an updated analysis of biomass patterns from the past surveys, Vessel Trip Reports, sea scallop observer trips, and Vessel Monitoring System data. In general, the current extent of the dredge survey was found to be very adequate for covering the scallop resource, although small areas were added to the extent of the HabCam survey to capture areas where there was evidence of significant scallop densities or commercial activity, in particular in the inshore central Mid-Atlantic region and south of Block Island, off the tip of Long Island.

The survey tracks are constructed as one long track for the MAB and three separate tracks for GB. Each track is bounded by a set of subregions. A midline, built to run along the center of biomass, is constructed to run through each set of subregions. Survey transects are centered on the midline and oriented orthogonal to the midline, approximately perpendicular to the isobaths. Survey design software prompts a user to enter the total effort (survey days) to allocate to a track, the relative lengths of the short transects on the track, and the transect density offsets for each subregion along the track. The software jitters the relative transect densities and provides a number of alternative tracks of similar lengths for the user to choose among, based on appropriate allocation of effort across the subregions, how well each track works around complex bathymetric structures, and other logistical considerations (Figure 1.19).

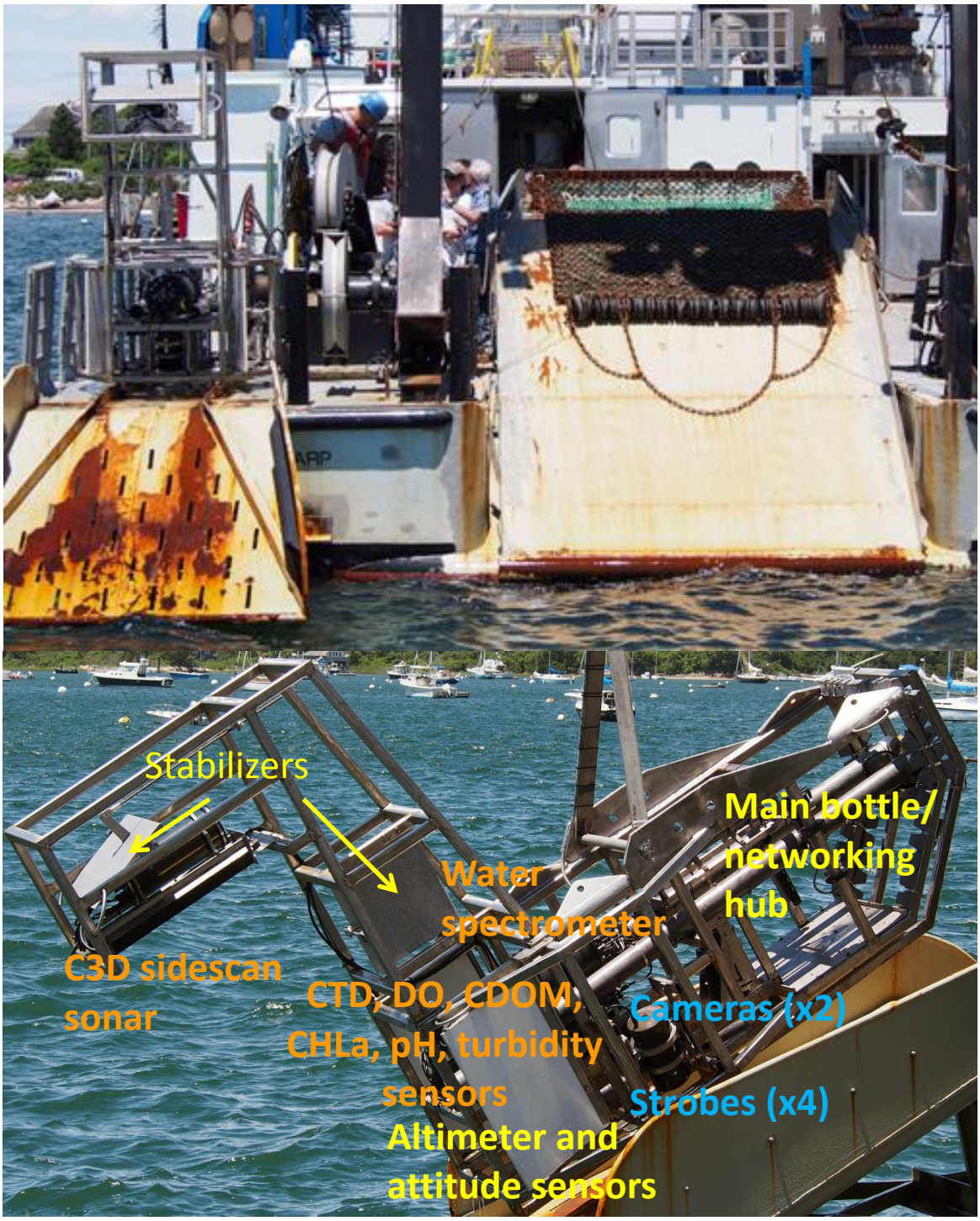


Figure 1.1. *Above*: The stern of the R/V Hugh Sharp, with HabCamV4 and the survey dredge. *Below*: The NOAA NEFSC HabCamV4 imaging system, including stereo cameras, integrated sidescan acoustics, plankton camera, light attenuation, dissolved oxygen, and pH sensors.

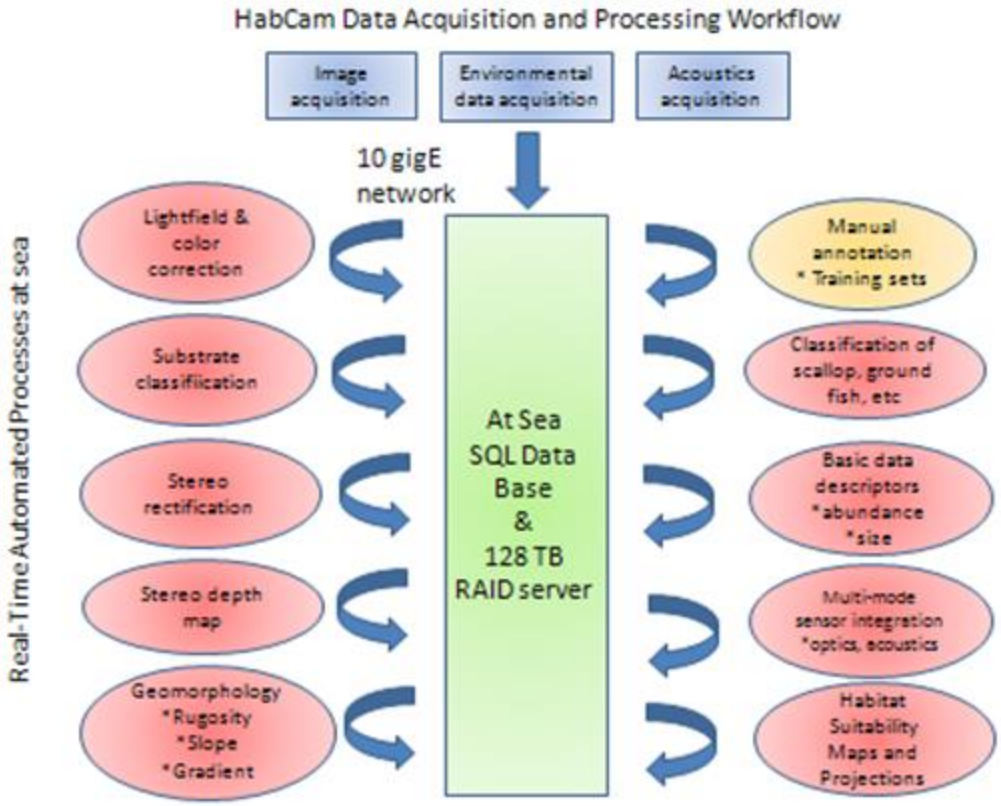


Figure 1.2. HabCamV4 data acquisition and image processing workflow.

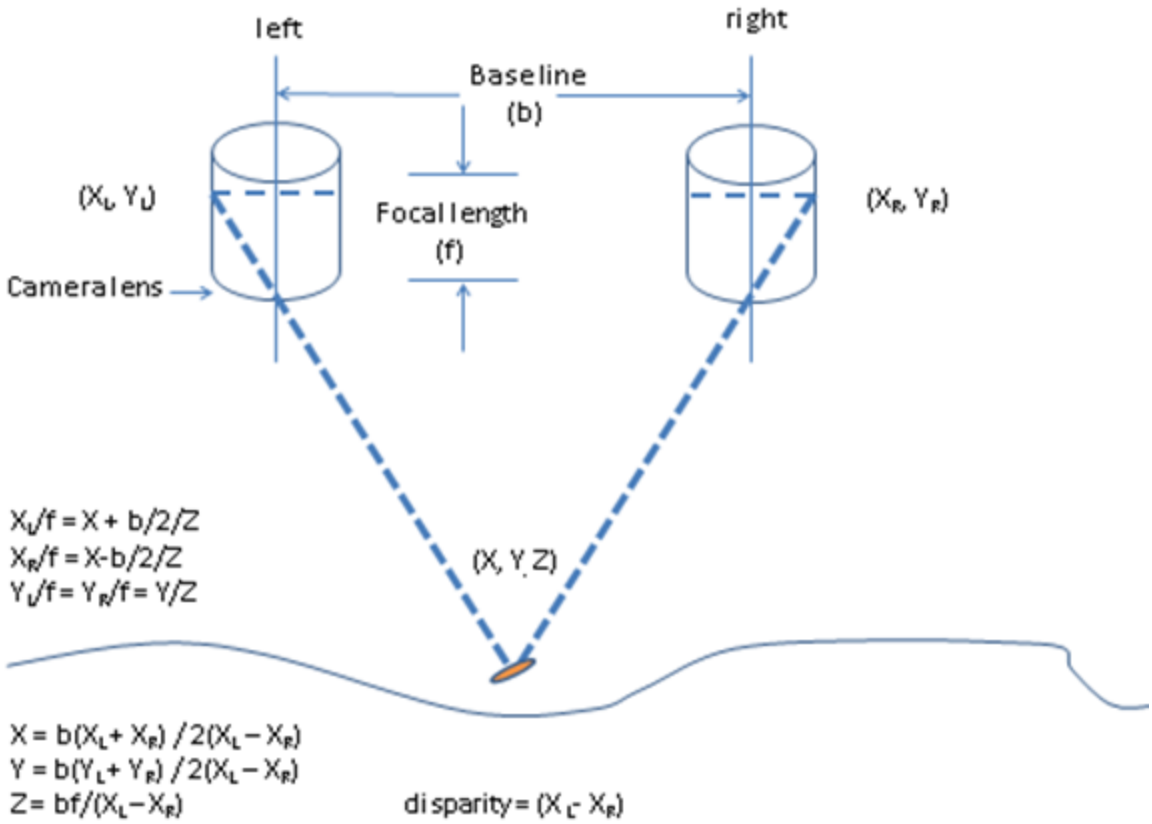


Figure 1.3. Arrangement of the stereo pair of cameras and the geometry of a 3D calculation of distance from the camera.

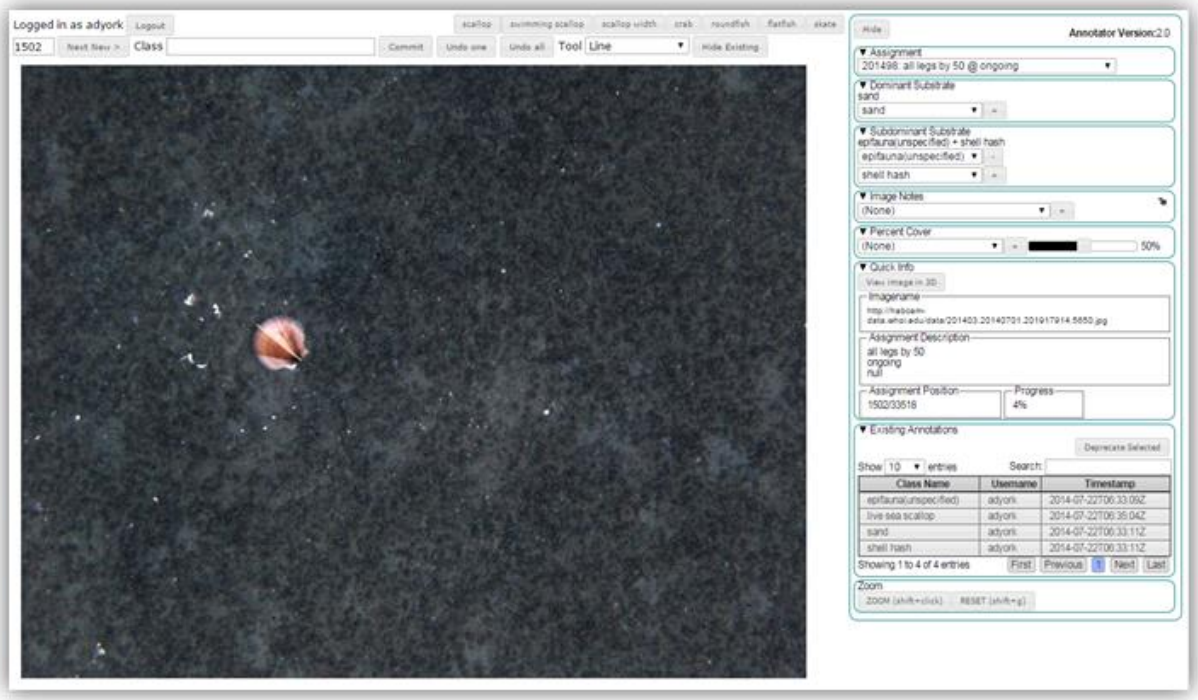


Figure 1.4. The web-based annotation tool used for NOAA HabCam V4 image annotation. Assignments of image lists are created. Scallops and groundfish are identified and measured in assigned images.

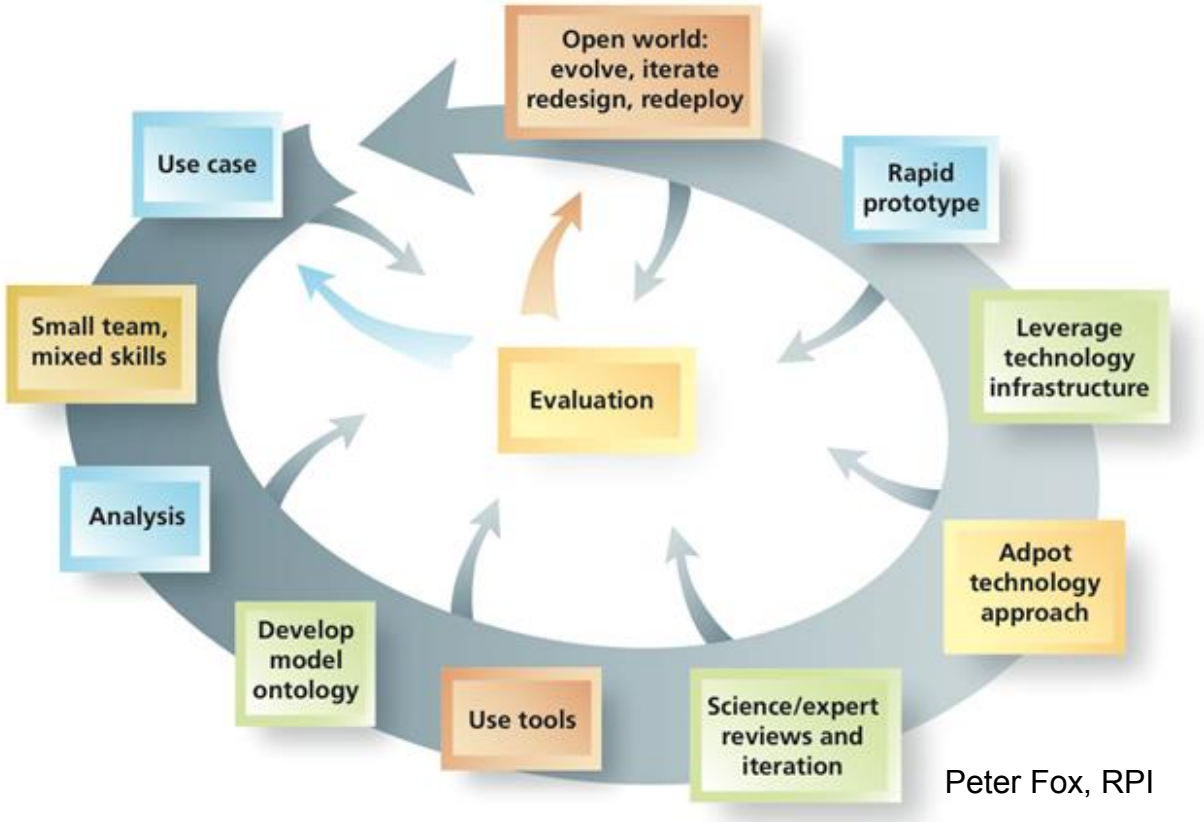


Figure 1.5. Tetherless World Iterative Model

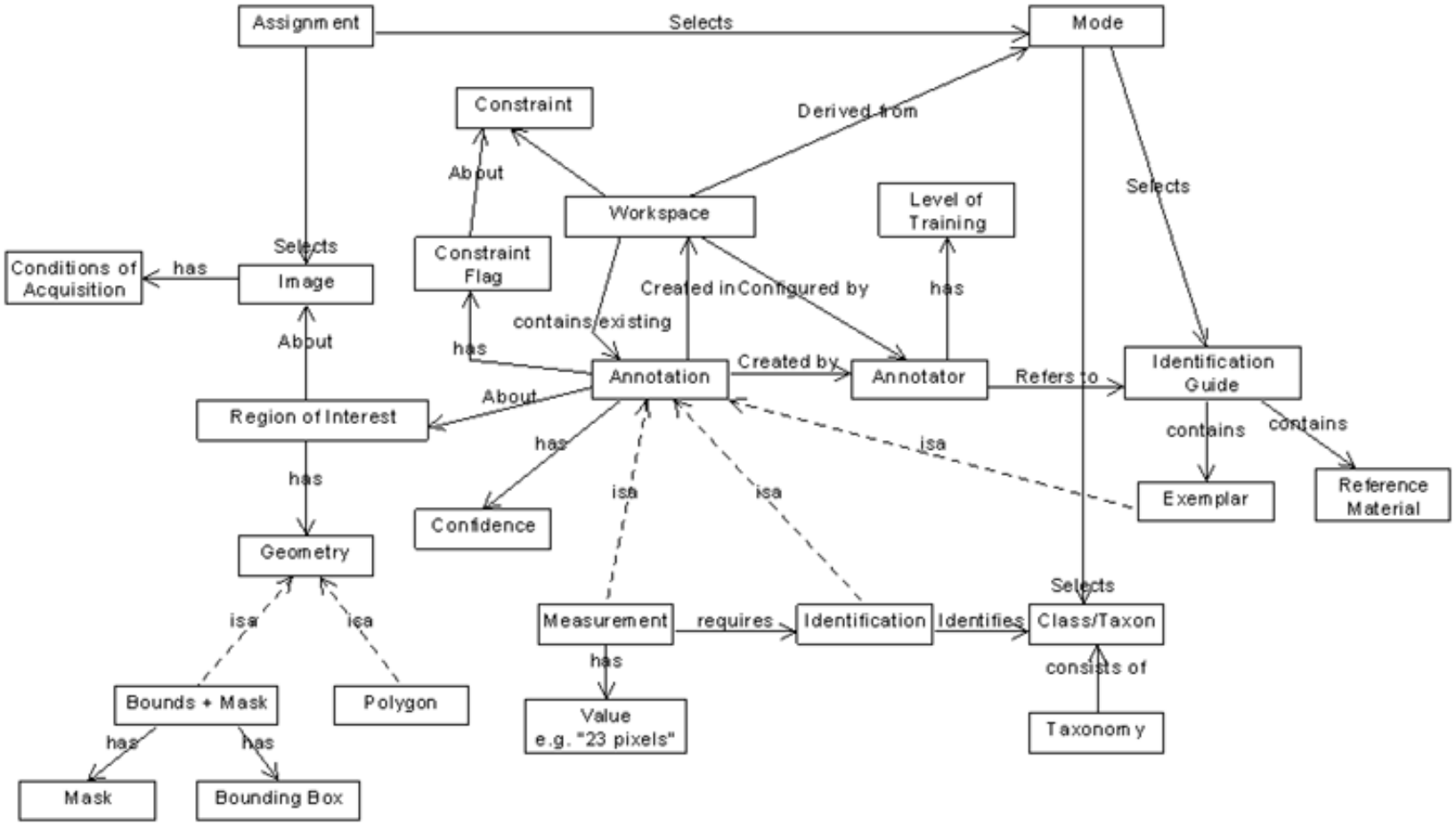


Figure 1.6. Full Information model for the Annotation Tool

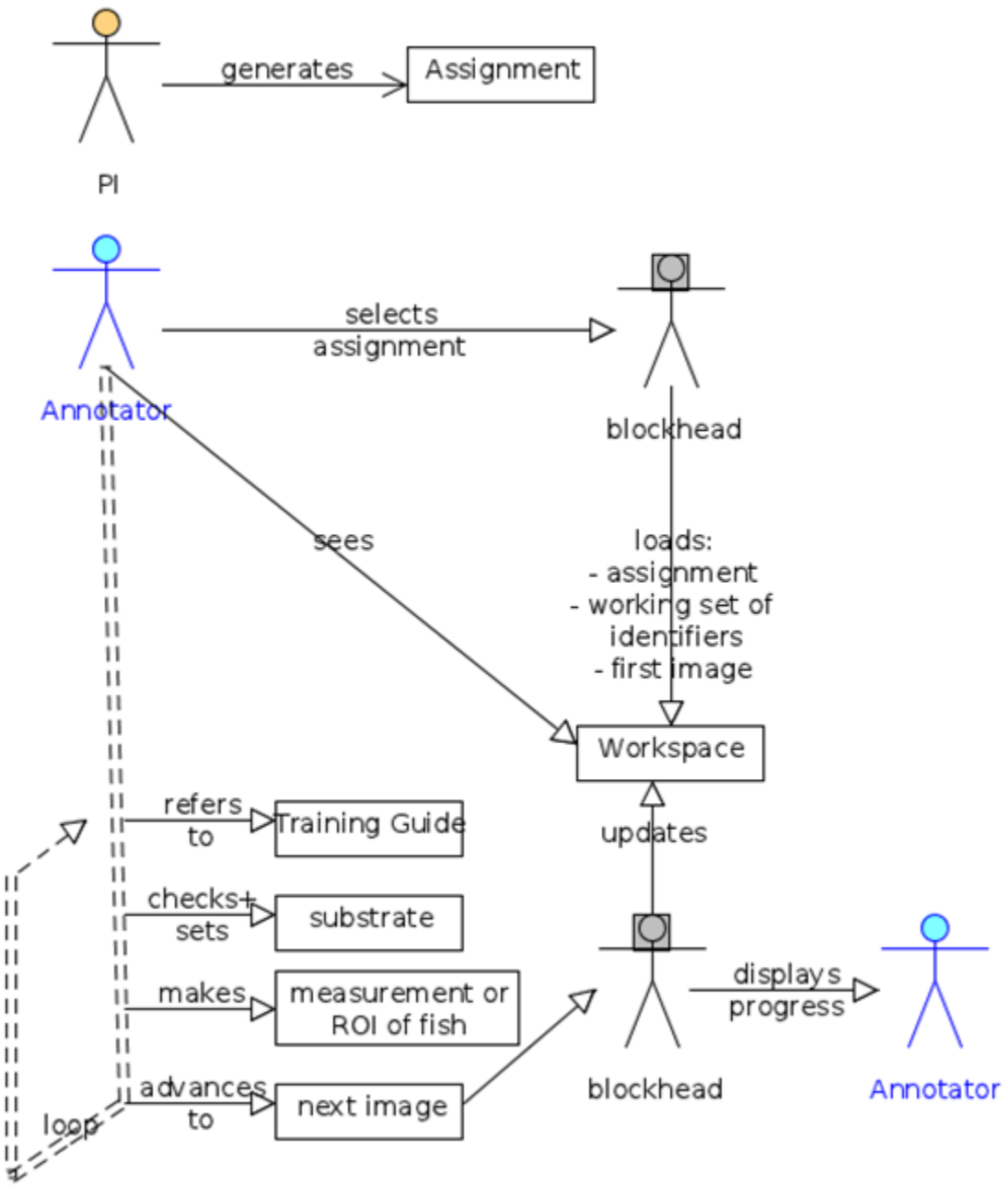


Figure 1.7. Activity diagram for the annotation tool. Shows interaction patterns during system use.

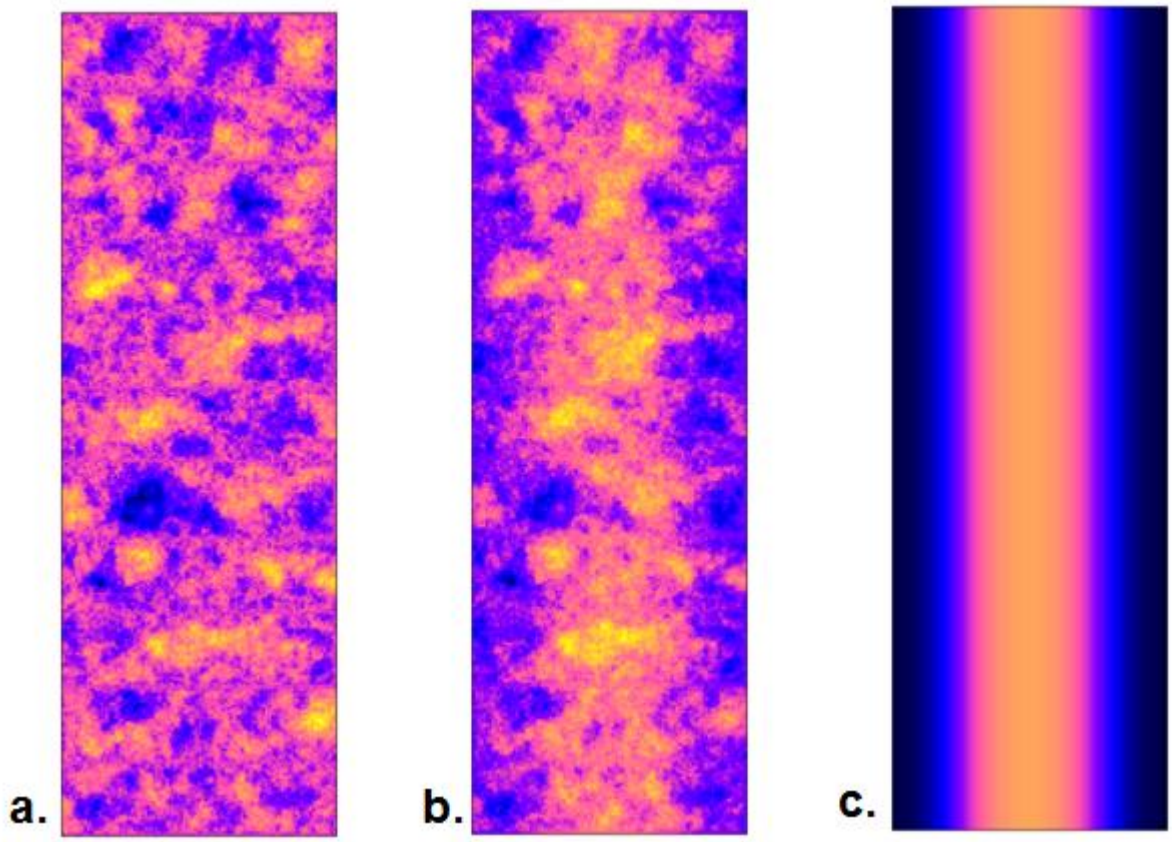


Figure 1.8. Hypothetical landscapes for (a) a landscape with a stationary mean, (b) a biased landscape with a higher mean along the center, and (c) the bias applied to landscape a to produce landscape b. In all plots, densities are higher in lighter-colored pixels.

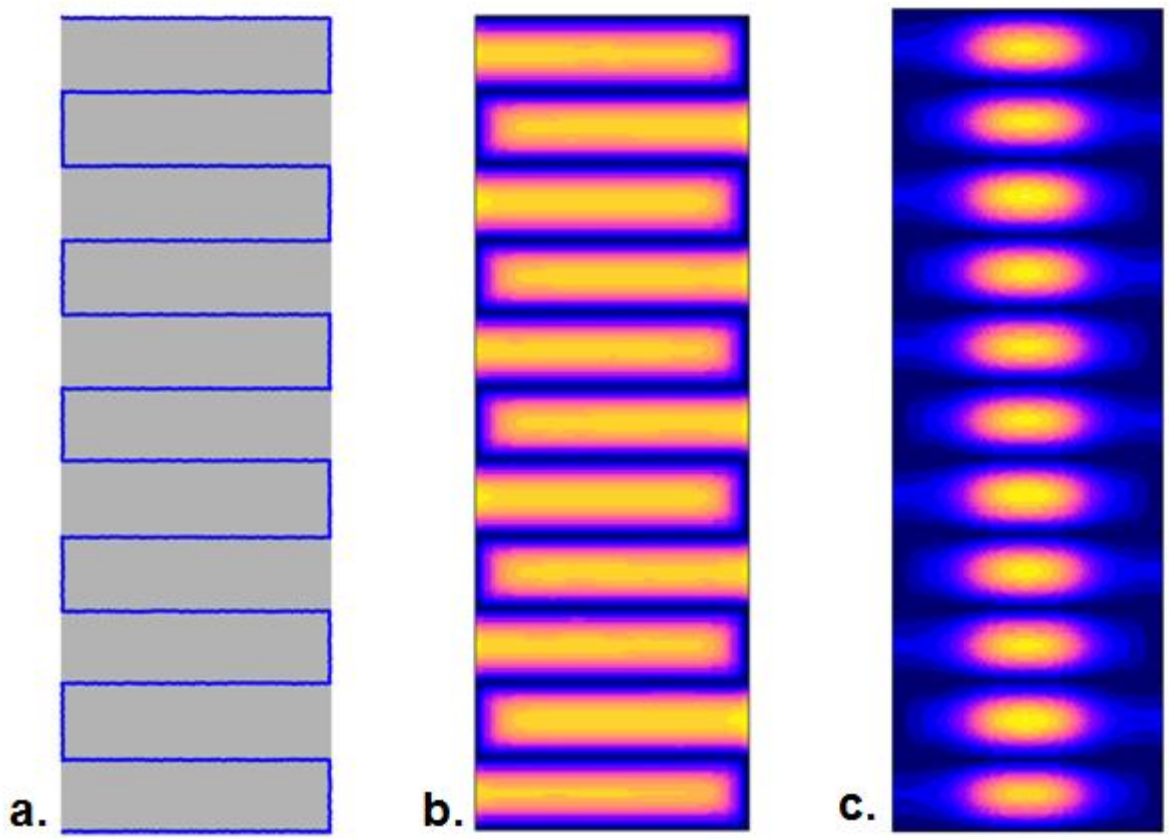


Figure 1.9 (a) A regularly spaced survey track across a rectangular survey area, (b) map of kriging variance derived from the survey track and an assumed underlying variogram model describing the data, (c) adjusted kriging variances resulting from applying the underlying trend from Figure 1c to (b).

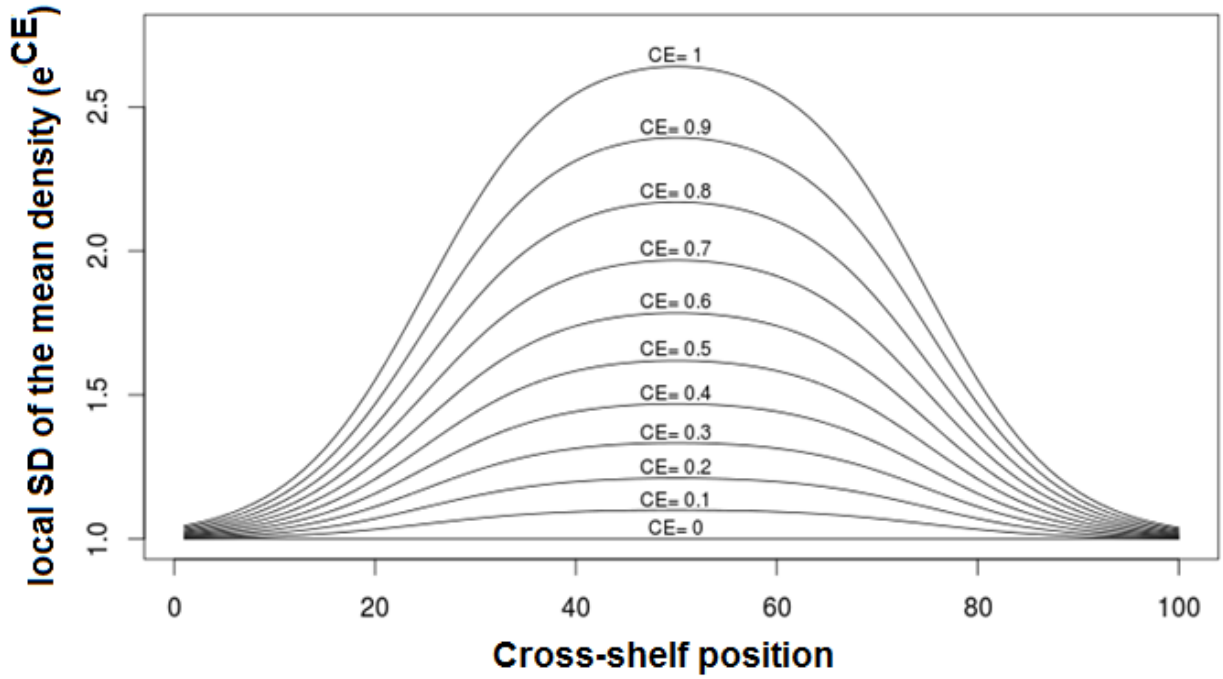


Figure 1.10. Levels of center effect used in simulations and resulting effects on the local standard deviation of the mean.

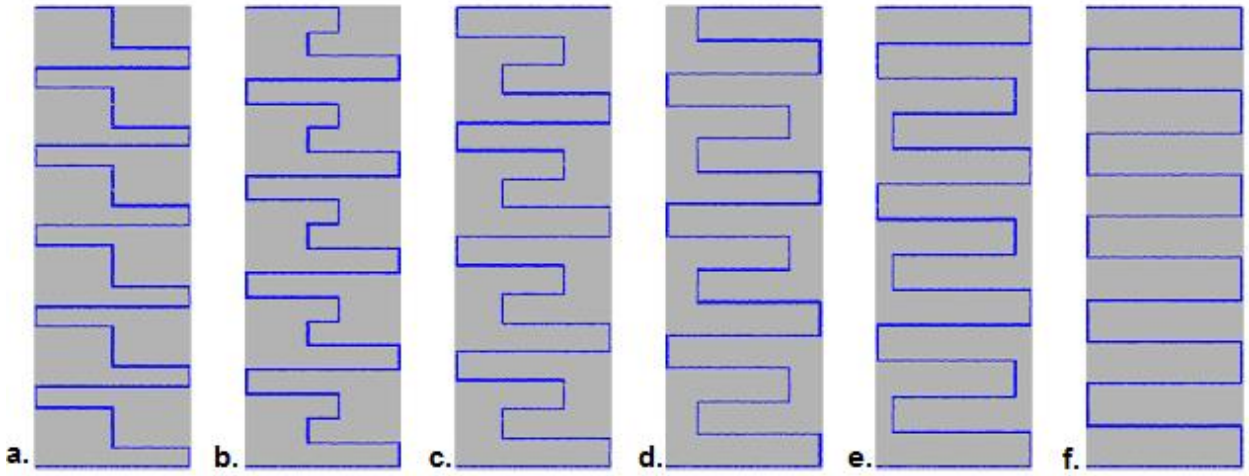


Figure 1.11 Alternative survey configurations, alternating the length of transects along the track and keeping total survey length constant. Short transect lengths are (a) 0%, (b) 20%, (c) 40%, (d) 60%, (e) 80%, and (f) 100% of the length of the long transects.

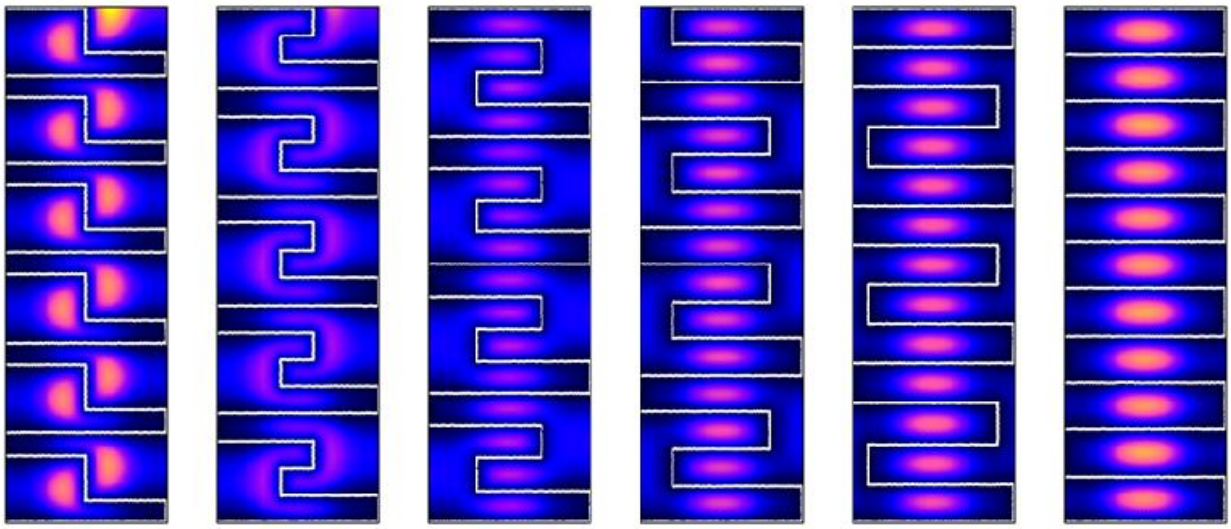


Figure 1.12. Variance maps for the survey tracks in Figure 4 with an applied center effect. Lighter colors indicate higher areas of variance.

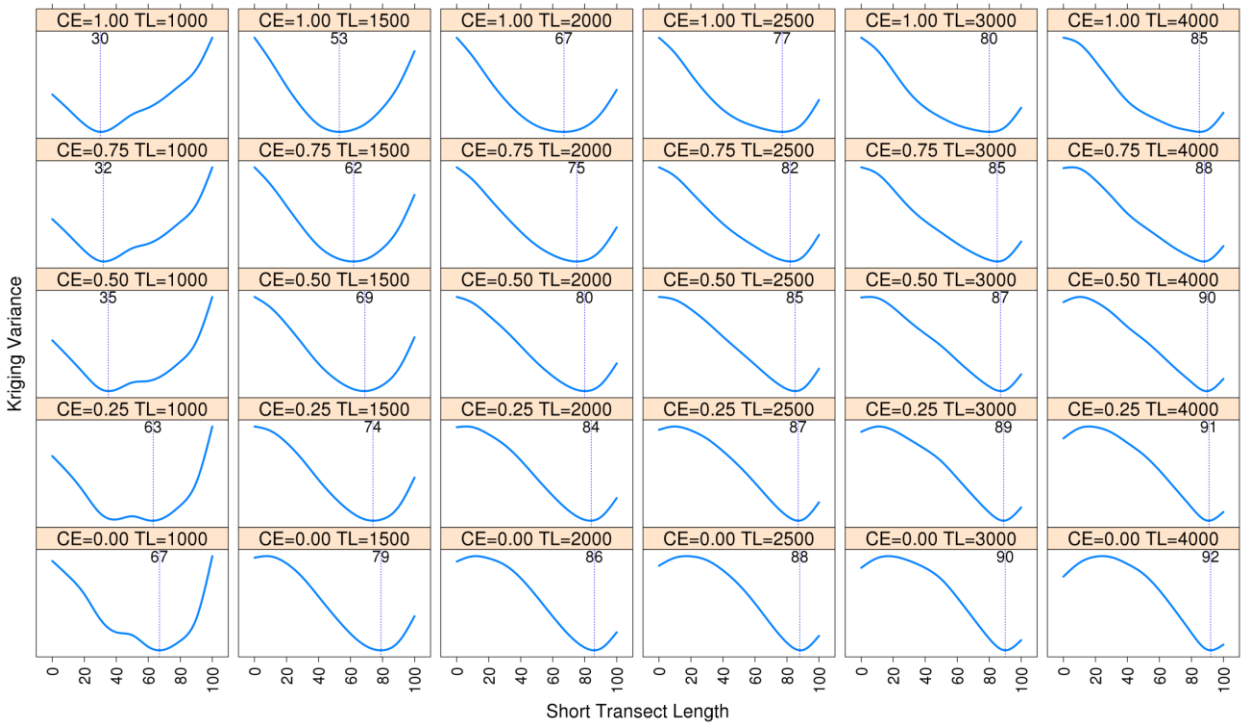


Figure 1.13. Adjusted kriging variances for different center effects (CE), short transect lengths and total survey track lengths (TL). Optimal solutions for each combination are marked with a dotted vertical line and labeled.

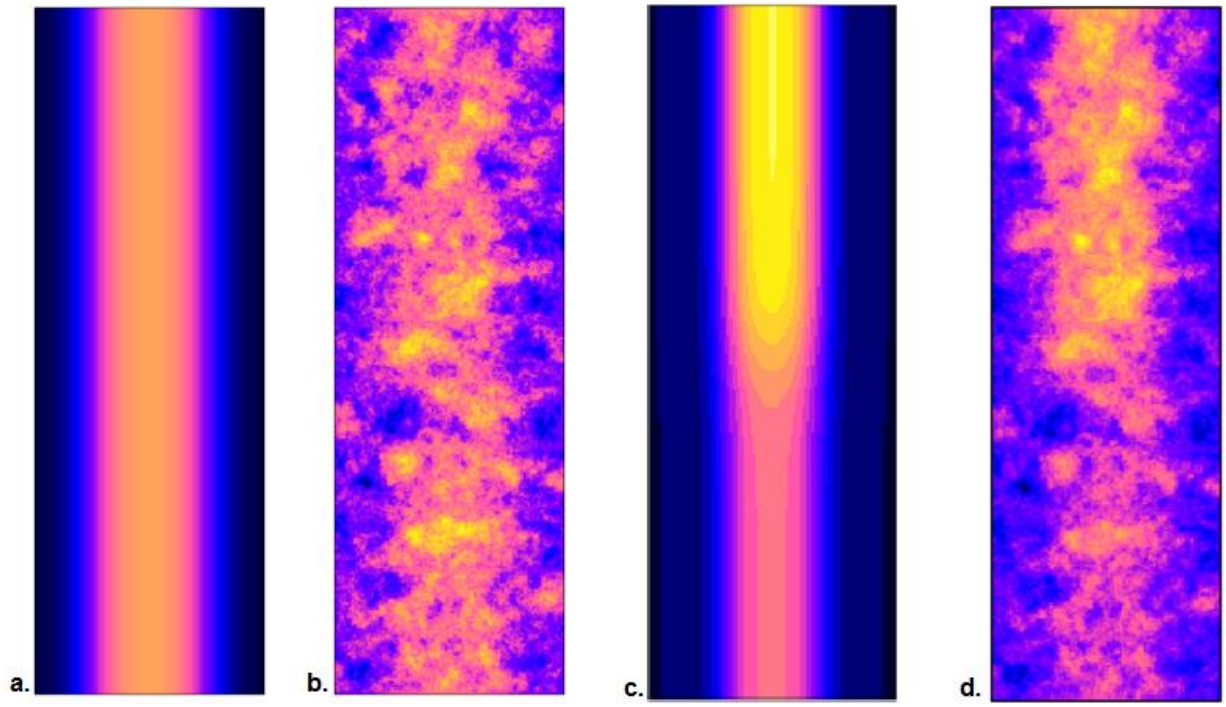


Figure 1.14. Comparison of survey landscapes without (a and b) and with (c and d) zonal anisotropy effects. Figures a and c represent the underlying trend in the mean while b and d represent the resulting simulated landscapes.

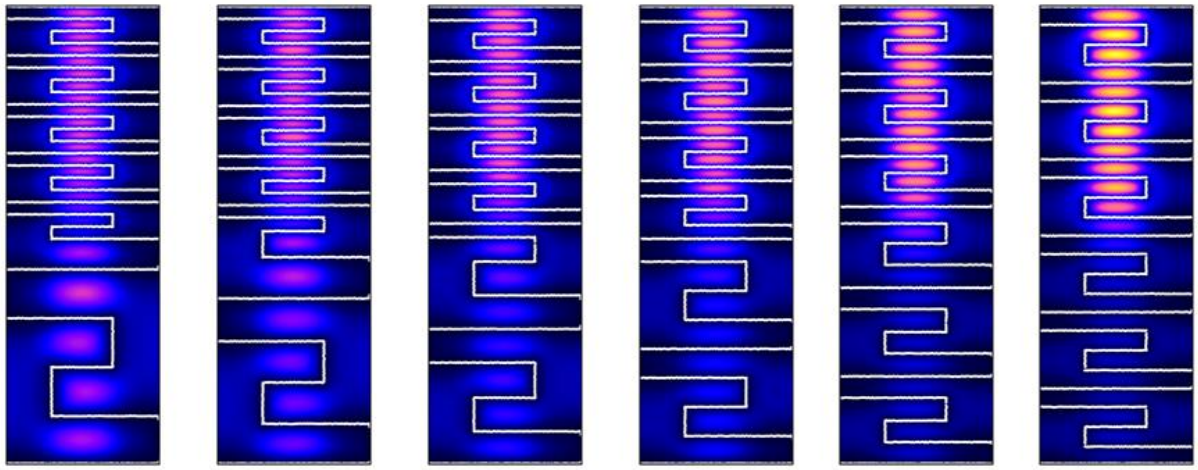


Figure 1.15. Example of varying transect density between zones and resulting variance maps for a simulated landscape with an underlying trend similar to 7c. The survey track is represented in white. Lighter colors indicate higher variances.

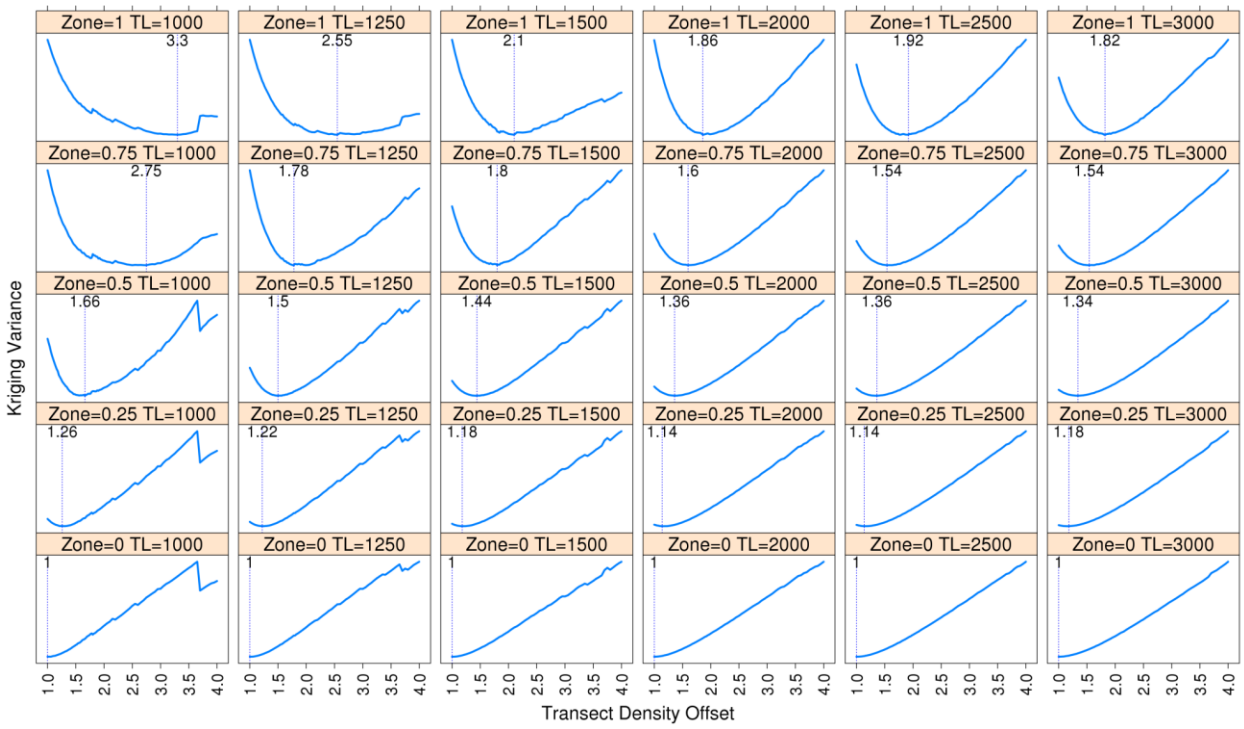


Figure 1.16. Zonal effects on transect density allocation. Higher “Transect Density Offests” represent the placement of proportionally more transects in the high density zones. Optimal solutions for each simulation set are labeled and marked with a dotted line.

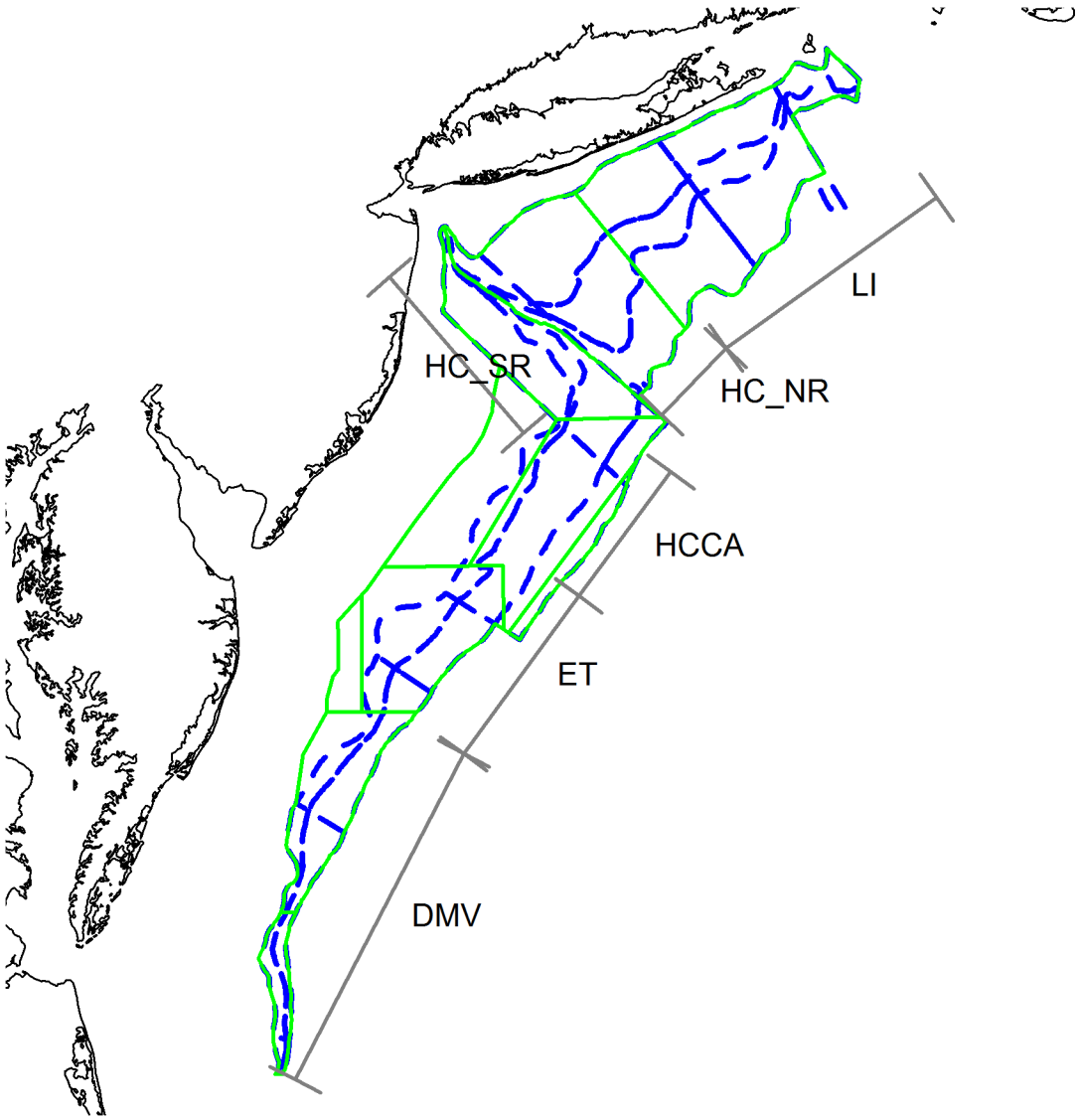


Figure 1.17. HabCam survey area (solid green line) compared to NEFSC scallop core strata (dashed blue line) in the MAB. Subregions used for allocating survey effort and abundance estimation are: LI – Long Island, HC_NR – Hudson Canyon North Rim, HC_SR Hudson Canyon South Rim, HCCA – Hudson Canyon Closed Area, ET – Elephant Trunk, DMV – DelMarVa.

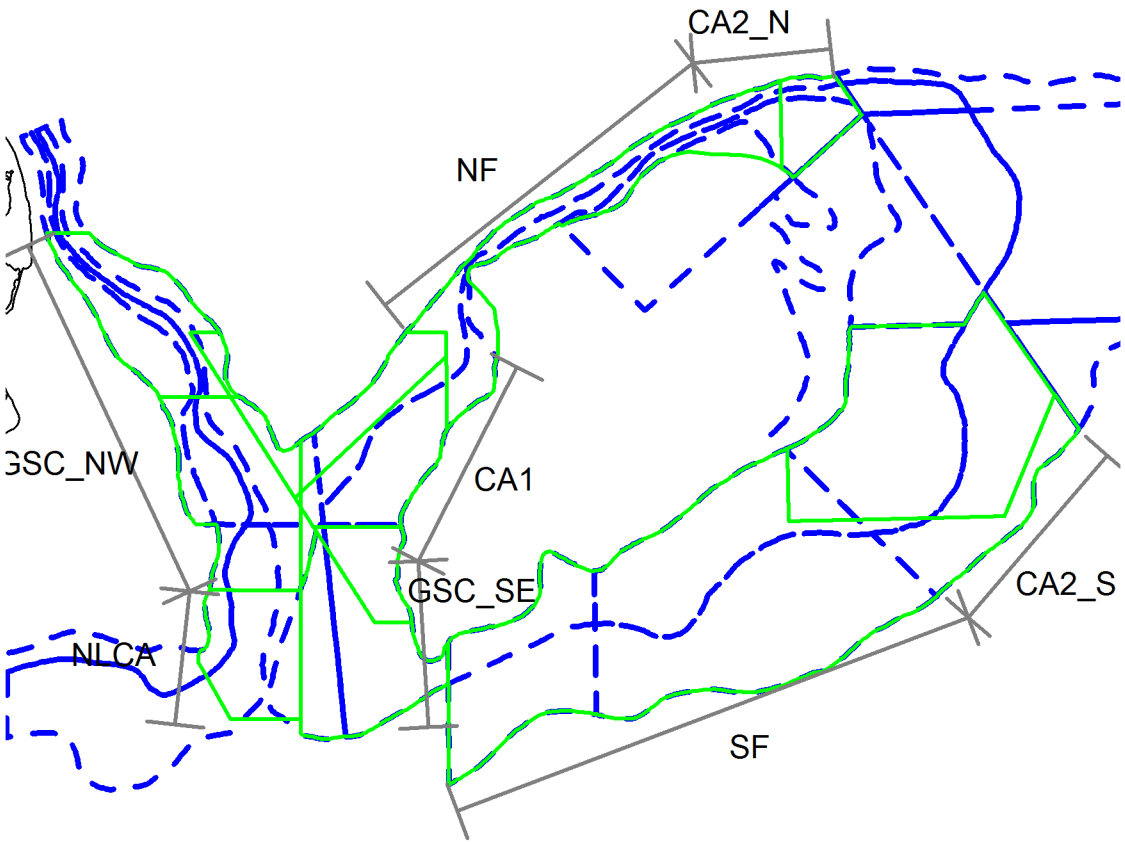


Figure 1.18. HabCam survey area (solid green line) compared to NEFSC scallop core strata (dashed blue line) for Georges Bank. Subregions used for allocating survey effort and abundance estimation are: GSC_NW – Great South Channel Northwest, NLCA – Nantucket Lightship Closed Area, GSC_SE – Great South Channel Southeast, CA1 – Closed Area 1, NF – Northern Flank, CA2_N – Closed Area 2 North, CA2_S – Closed Area 2 South, SF – Southern Flank.

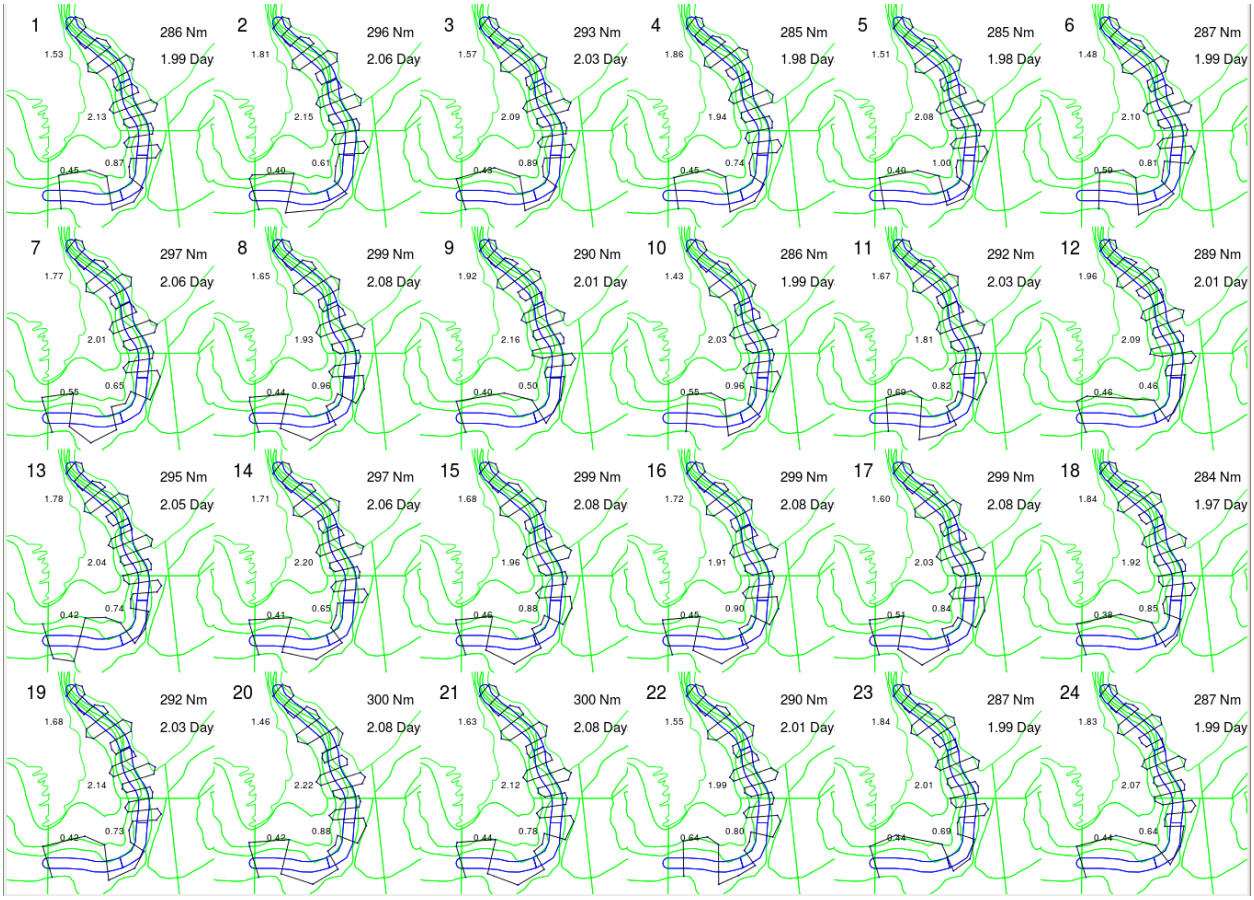


Figure 1.19. Alternative HabCam survey tracks generated by the survey software. Each track plot includes the total length (nautical miles) and projected time to completion. Small numbers are weighting factors representing the relative density of adjacent parallel transects.

TOR 2: HabCam measurement error, detection, and confounding factors

In this discussion we describe the various factors that can lead to measurement error:
2.1 Camera calibration (intrinsic factors)
2.2 Vehicle orientation and altitude (extrinsic factors)
2.3 Human annotation error
2.4 Environmental factors
2.5 2D versus 3D measurements
2.6 Conversion from shell height to width
2.7 Comparison between dredge and HabCam measured shell heights
2.8 Annotator Training

2.1 Camera calibration (intrinsic factors)
Rigorous camera calibration is primary and a fundamental procedure that must be done to minimize measurement error. HabCam cameras use a 12 mm focal length lens coupled to a ¾" (HabCamV2) or 1" Sony CCD chip (HabVamV4). The relative position between the chip sensor and the lens dictates the intrinsic calibration model for the system including estimation of in-water focal length, principle point, and pixel error, followed by image correction by employing extrinsic parameters collected for each image.

The intrinsic parameters for the HabCamV2 and V4 cameras are calculated every year before each survey and after repair or modification of the system using images of a $1m^2$ target marked off at 80mm intervals in a 4 m deep seawater tank (32psu). The HabCam vehicles are positioned above the target while the target is moved through a series of altitudes (1-3m) and orientations (roll, and pitch 0-20 degrees). At least 10 but typically 20 images representing a range of positions are used for calibration of the cameras using the Calibration Toolbox in Matlab. Each year an calibration file is prepared and stored to be used in calculating measurements during the survey. As an example of one of the intrinsic calibrations we provide the following:

Focal Length: fc = [2773.25504 2764.28859] ± [7.18117 7.13362]
Principal point: cc = [778.19667 509.00401] ± [4.13012 3.80811]
Skew: alpha_c = [0.00000] ± [0.00000] => angle of pixel axes = 90.00000 ± 0.00000 degrees
Distortion: kc = [-0.31591 0.14388 0.00070 0.00138 0.00000] ± [0.00702 0.02649 0.00038 0.00056 0.00000]
Pixel error: err = [0.53035 0.50489]

The numerical errors are approximately three times the standard deviations:
Intrinsic pixel error = +/- 1.59 pixels
Resolution range f (FOV): 0.37 – 0.89 mm/pixel
Intrinsic real-world error : 0.58 – 1.41 mm

These values provide error bounds on the resolution and accuracy of the camera system in water. Plots of the relative errors show that the camera CCD chip, lens and housing window are typically slightly out of alignment in both radial and tangential attitudes (Fig. 2.1). The pixel resolution is a function of FOV, which in turn is a function of altitude off the bottom. In calibrated screen measurement space, the overall measurement error is between 0.58 and 1.41 mm.

Distortion in each image is first corrected using the intrinsic parameters given above (Fig. 2.3).
KK = [fc(1) alpha_c*fc(1) cc(1);0 fc(2) cc(2) ; 0 0 1];
Where the KK matrix is the uncorrected image matrix.

```
r2_extreme = (nx^2/(4*fc(1)^2) + ny^2/(4*fc(2)^2));
dist_amount = 1; %(1+kc(1)*r2_extreme + kc(2)*r2_extreme^2);
fc_new = dist_amount * fc;
KK_new = [fc_new(1) alpha_c*fc_new(1) cc(1);0 fc_new(2) cc(2) ; 0 0 1];
```

KK_new s the corrected image matrix.

```
[I2] = rect(I,eye(3),fc,cc,kc,KK_new);
```

Where I is the distorted image and I2 is the undistorted image. This is accomplished for each image as it is stored in the database.

2.2 Vehicle orientation and altitude (extrinsic factors)

Extrinsic parameters relate to the combination of intrinsic parameters plus the orientation of the camera relative to the image plane. The calibration matrix is built up from the 20 or so views indicated in figure 2.4.

Calculation of extrinsic parameters

Cross over points in the calibration chart are automatically detected and their locations in pixel space extracted before calculation of extrinsic parameters (Fig. 2.4).

The extrinsic parameters are encoded in the form of a rotation matrix (**Rc_ext**) and a translation vector (**Tc_ext**). The rotation vector **omc_ext** is related to the rotation matrix (**Rc_ext**) through the Rodrigues formula: **Rc_ext = rodrigues(omc_ext)**.

Let **P** be a point space of coordinate vector **XX = [X;Y;Z]** in the grid reference frame **(O,X,Y,Z).**

Let $\mathbf{XX_c = [X_c;Y_c;Z_c]}$ be the coordinate vector of **P** in the camera reference frame $\mathbf{(O_c,X_c,Y_c,Z_c)}$.
Then **XX** and $\mathbf{XX_c}$ are related to each other through the following rigid motion equation:
$\mathbf{XX_c = Rc_ext * XX + Tc_ext}$

In addition to the rigid motion transformation parameters, the coordinates of the grid points in the grid reference frame are also stored in the matrix **X_ext**.

Each image taken by HabCam has its own unique set of extrinsic parameters.
Extrinsic parameters for an example image:

Translation vector:
Tc_ext = [-225.840216 -130.369514 608.628548]

Rotation vector:
omc_ext = [-2.148393 -2.284790 -0.123388]

Rotation matrix:

```
Rc_ext = [ -0.062925  0.996680   0.051672
           0.996448   0.059838   0.059254
           0.055966   0.055217  -0.996905 ]
```

Reprojection Pixel Error: err = [2.00116 1.26492]

Extrinsic pixel error = +/- 2 pixels
Resolution range (FOV): 0.37 – 0.89 mm/pixel

The extrinsic real-world error becomes: 1.11 – 1.78 mm (under best optical conditions)

Frequency distributions of pitch, roll, altitude, and image area (FOV) based on 129,289 images are shown in Figure 2.5. The mean pitch was -5.24 degrees indicating that, on average, the nose of the vehicle pointed down slightly. Downward pitch is part of the system design and tends to stabilize the vehicle while underway. Mean roll is -0.97 with very little variation indicating the vehicle is quite stable, laterally. Altitude measurement varied from <1 to 4.5 m off the bottom with a mean of 1.87m. Images below 1 m are out of focus and removed from the image database. Images taken higher than 3 m are typically not sufficiently clear, due to turbidity, to be useful and were also not used. Taking roll and pitch into account using the extrinsic equations present above, the FOV ranged from 0.2 to >4m^2 with a mean of 0.72 m^2 . 95% of the calculations for FOV fall between 0.4 and 1.5 m^2. Inclusion of roll and pitch in the geometric projection of the FOV has an effect of broadening and smoothing the frequency distribution of values without changing the mean.

2.3) Human annotation error

The annotator software developed specifically for annotating HabCamV4 images was described in TOR 1. Here we examine the error associated between and within individual annotators. The former will provide insight into measurement repeatability and the latter into systematic bias within an individual. To accomplish this, we assigned four identical 4.2 nm long image transects containing 4,432 images from Western Great South Channel to six annotators (raters) (Fig. 2.5). Raters measured scallops using the Annotation Tool under the same measurement rules as would be used under normal conditions (edge effects, height vs width, etc).

In most cases, mean within rater measurements were either accurate to the same number of pixels or within one pixel suggesting that within rater variability was extremely low (Table 2.1). Between rater variability was greater than within rater variability with mean values of 132, 128, 135, 127, providing a range of 135 to 127, or 8 pixels. Given the resolution range for varying FOV presented above (i.e., 0.37 – 0.89 mm/pixel), an error of 8 pixels represents a real-world error of 3.0 to 7.1 mm.

Inter and intra-Class correlations were analyzed using ICC, Intra Class Correlation analysis (McGraw and Wong, 1996). A two-way mixed effect model

$Xij = u + ri + ci = rcij + e\ ij$

where u = population mean, r = row effects, c = column effects, and e = residual effects, was used to test the hypotheses that there is no difference between scallop measurements made by the same rater four times, and that there is no difference between individual raters. While there are no statistical differences among the final results of the individual annotators, there are moderate differences (up to 7 mm) between measurements among raters (Table 2.2). However, individual rater are internally consistent, so that, for example, an annotator that tends to measure larger scallops than others does so consistently.

Total sources of error include the intrinsic camera calibration, the extrinsic calibration relative to vehicle position, and human error associated with annotation of an image (Table 2.3). By far the major source of error comes from human annotation. For example, a human error of 7.1 mm when measuring a scallop of 60 mm in height represents an overall error of 11%. We would like to improve this by better training and through automated techniques for measuring scallop shells, but 11% is where we currently lie. If human error was eliminated then the machine error of 1.78 mm represents 2.9% overall error, which is surely more acceptable.

2.4) Environmental factors

<u>Scallop Detectability</u>

A major source of shell measurement error and detection of scallops by annotators are the environmental conditions under which the images are taken. In this section we evaluate the effect of turbidity, sea state and epifauna on the detection and measurement of scallops and the ability to detect live and dead scallops.

In HabCam imagery the ability to detect scallops and identify whether they are alive or dead is critical to data accuracy and quality. Several factors influence an annotator's ability to detect a scallop including obscuration due to encrustation, turbid water, thick plankton, and high altitude (Fig 2.7.). In the presence of factors that obscure the whole image such as turbid water and plankton, the annotator is trained to throw out the entire image from the dataset by indicating the image is "unusable." If the scallop is partially obscured by encrustation, the annotator still measures the scallop, but calls the measurement "inexact" but the count is still used for scallop density calculations (Fig 2.8). If an object is completely obscured by encrustation and is the size and shape of a scallop the annotator does not call it a scallop because they cannot see any part of the shell.

Because counting of live scallops is of upmost importance, training of annotators includes rigorous practice determining whether scallops are alive or dead (See training in TOR2.8). Annotators are given guidelines for determining if a scallop is alive including visibility of tentacles, shadow around the edge of the scallop, ability to see both halves of the shell, and whether the outer edge of the scallop is clean (Fig 2.9). Not all qualities mentioned of a live scallop are visible in every image but presence of any of these qualities indicates the scallop is alive. A scallop is considered dead if the shell is flush with the seafloor, there is a hole in the shell, it is a clapper, it is the inside of one of the shells, or if the outer edge is buried by silt of debris (Fig 2.10). It should be noted that being upside down or white in color is not in itself conclusive evidence that it is dead. If an annotator is not completely confidant that a scallop is alive they can use the modifier of "probable." Figure 2.11 shows the proportion of scallops that were deemed "probable" live scallops. These probable scallops are still included in the density estimates.

<u>The effect of sea state on image annotation</u>

Sea state can have a strong impact on the ability to tow HabCam. Because, by design, the vehicle is very heavy and the cable catenary is minimal to provide low lag and high response time by the pilot controlling the HabCam winch, ship heave and pitch will be directly imposed on the cable and the vehicle. Figure 2.12 shows the relationship between variability (STD) in ships pitch and variability of vehicle pitch (STD) for a part of the 2014 survey. Although a relationship is clear, the correlation is not particularly tight because of the lag offered by the cable.

Typically if we are towing and wind speed and sea state starts to build we will continue to tow until the variation in altitude of the vehicle off the bottom becomes too severe that data can no longer be effectively collected. The result of increasing ship heave is an increase in variation in vehicle altitude and ultimately the number of images that cannot be used. Figure 2.13 shows the relationship between variation in ship heave and the percent of images out of range (i.e., vehicle altitudes < 1 and >3 m). This relationship shows that while out of range images can result even at low ship heave, once heave has increased beyond some point then in range images are a very small percentage of the total images collected.

2.5 2D versus 3D measurements

The errors associated with 2D versus 3D measurements of shell dimensions can exceed 30% due to angular displacement and up 100% due to distance from the camera for swimming scallops. Future assessments of scallops that include those that are swimming require stereo measurements to ensure accurate results. We will provide examples of the extent of the error associated with angular positioning and for swimming scallops off the bottom. Figure 2.14 shows that scallops <10mm can be swimming well off the seafloor upwards of 2 m. Older scallops >20mm are typically closer to the bottom but still can venture into the benthic boundary layer.

2.6 Conversion from shell height to width

Because scallops are not always oriented normal to the camera or may be partially obscured, scallop measurements may be either shell heights (umbo to opposite margin) or widths (lateral margins), whichever is judged to be more appropriate. Shell widths are converted to shell heights using a statistical model derived from paired measurements of scallop that were well oriented to the camera:

$$\text{Shell_height} = 3.538 + 1.034*(\text{Shell_width}) - 0.0003502*(\text{Shell_width})^2$$

Shell height is calculated in units based on the start and end coordinates of the annotated line. The size of each pixel in an image is calculated from the altitude of the associated image, based on tank calibration experiments, and this pixel size is used to convert the shell height to actual millimeters. The altitude is also used to calculate the field of view for each image for density calculations.

2.7. Comparison between dredge and HabCam measured shell heights

Comparison of region-scale normalized HabCam, and dredge shell heights are shown in Fig 2.15 (see also NEFSC dredge TOR-2). Dredge and HabCam size-frequencies were similar in 2012 in the Mid-Atlantic. The Mid-Atlantic dredge surveys were conducted about two months prior to the HabCam surveys in 2013 and 2014; this accounts for much of the disparity between the surveys. More evidence of measurement error is seen for the Georges Bank shell heights. This is plausible because the more irregular terrain on Georges Bank may induce more measurement error than the smooth sand bottoms of the Mid-Atlantic. In the 2014 sea scallop stock assessment, measurement error for HabCam and the SMAST survey was estimated by fitting the optical size-frequencies to the dredge size frequencies convolved with a normal distribution with mean zero and a fitted standard deviation *s* for all available years (through 2013) and regions. HabCam measurement error was estimated as 13 mm compared to 11 mm for the SMAST survey. Uncertainties in altitude likely account for the increased error in the HabCam survey. A new, and likely more accurate, method of estimating altitudes from stereo imagery (rather than using the altimeter) has been developed since then.

2.8 Annotator Training

Consistent and accurate quantification and measurement of scallops from imagery can be problematic due to a number of subjective decisions that individuals must make when viewing an image including:

1. The decision to accept or reject an image due to image corruption, water turbidity, unbalanced lighting or plankton in the water column obscuring the bottom.
2. The decision to accept or reject individual scallops as dead or alive based on orientation, color, and other indicators.
3. If the scallop is sufficiently inside the image to be counted.
4. The best way to measure the scallop based on orientation (i.e. shell height or shell width).

Additionally, personnel need to become familiar with recording data and navigating in the web-based annotation tool. Thus it is necessary to provide training and guidance for personnel who are new to annotating images and an annual refresher to staff who have done image annotation in previous surveys before beginning image annotation. Personnel are provided a guide to annotation with basic rules and guidance on subjective decisions (Appendix 1). The guide is updated annually to track changes in protocols and the annotation software. In 2012 and 2013, personnel completed a training set of ~300 images. Images in the training set were selected to represent the geographic coverage of our survey area, various sediment types and turbidity levels, and different scallop sizes and behaviors (swimming, partially buried, etc.). After completing the training set, each person reviewed their decisions for each image with an expert and discussed subjective decisions and techniques for dealing with difficult images.

Starting in 2014, we selected a new training set of 347 images for training and comparison. Trainee's are encouraged to consult experienced annotators and discuss difficult images while completing the training set. We developed a software routine to compare the trainee's results to an expert's results. The output from the comparison includes discrepancies in rejection of images for poor quality, misclassification of scallops and fish, and mis-measurement of scallops (Appendix 2). Trainees review the results with a more experienced annotator and review individual images to understand why their results differ from the "expert" results. Thereafter, during annotation, personnel are encouraged to seek second opinions from others as questions arise or they encounter difficult images. Images can also be flagged for review, so an experienced staff member can later return to the

image and check the work. In isolated cases where annotators were potentially doing work of questionable quality, their entire annotated image set was reviewed by NEFSC staff.

Image assignments for assessing the scallop resource typically employ a limited species list. All scallops sufficiently within the image (see Appendix 1) are identified and precisely measured. Other species of interest are only assigned general categories (i.e. flatfish, roundfish, skate, cancer crab, lobster) and marked with a bounding box so NEFSC staff can later return to them and provide a precise identification and measurement. Scallops and fish that are mostly outside the image can also be noted as “less than half”, which is useful for noting the presence of uncommon species without counting towards species density estimates.

Table 2.1. Summary statistics in pixels. N = 277 for each run. A total of 4,432 scallops were measured. KLB, ADY, PK, and DPF are initials of the four raters.

rater	**KLB**				
	run1	**run2**	**run3**	**run4**	**mean**
mean	132.41	132.21	132.58	132.28	132
STD	31.61	31.61	31.44	31.69	
SE	1.89	1.91	1.88	1.9	
rater	**ADY**				
	run1	run2	run3	run4	mean
mean	128.48	128.46	128.41	128.75	128
STD	31.42	31.55	31.44	31.78	
SE	1.88	1.89	1.88	1.9	
rater	**PK**				
	run1	run2	run3	run4	mean
mean	135.57	135.88	134.95	134.28	135
STD	31.86	31.94	31.81	31.63	
SE	1.91	1.91	1.91	1.9	
rater	**DPF**				
	run1	run2	run3	run4	mean
mean	128.36	127.39	127.48	127.56	127
STD	31.74	31.63	31.45	31.57	
SE	1.9	1.9	1.89	1.89	

Table 2.2 Results of tests for variations in annotator measurements.

ICC Type C-1: Tests the degree of consistency among measurements

r = (MSR - MSE) / (MSR + (k-1)*MSE);
F = (MSR/MSE) * (1-r0)/(1+(k-1)*r0);
df1 = n - 1;
df2 = (n-1)*(k-1);
p = 1-fcdf(F, df1, df2);

ANOVA Table

Source	SS	df	MS	F	Prob>F
Time	143.5	3	47.8	6.28	0.0003
Group	40170.4	3	13390.1	3.36	0.0182
Ineratcion	480.6	9	53.4	7.01	0
Subjects (matching)	4398738.9	1104	3984.4	522.79	0
Error	25241.8	3312	7.6		
Total	4464775.2	4431			

r = 0.9842
LB = 0.9827
UB = 0.9857
p = 0

ICC Type A-1: Test the degree of absolute agreement among measurements.

r = (MSR - MSE) / (MSR + (k-1)*MSE + k*(MSC-MSE)/n);
a = (k*r0) / (n*(1-r0));
b = 1 + (k*r0*(n-1))/(n*(1-r0));
F = MSR / (a*MSC + b*MSE);
df1 = n - 1;
df2 = (a*MSC + b*MSE)^2/((a*MSC)^2/(k-1) + (b*MSE)^2/((n-1)*(k-1)));
p = 1-fcdf(F, df1, df2);

ANOVA Table

Source	SS	df	MS	F	Prob>F
Time	143.5	3	47.8	6.28	0.0003
Group	40170.4	3	13390.1	3.36	0.0182
Ineratcion	480.6	9	53.4	7.01	0
Subjects (matching)	4398738.9	1104	3984.4	522.79	0
Error	25241.8	3312	7.6		
Total	4464775.2	4431			

r = 0.9796
LB = 0.9695
UB = 0.9856
p = 0

Table 2.3. Sources of measurement error

Source of error	**Pixel error**	**Real world unit error**
Intrinsic camera system	+/- 1.59 pixels	0.58-1.41 mm
Extrinsic camera and vehicle system	+/- 2 pixels	1.11-1.78 mm
Human annotation	+/- 4 pixels	3.0 -7.1 mm

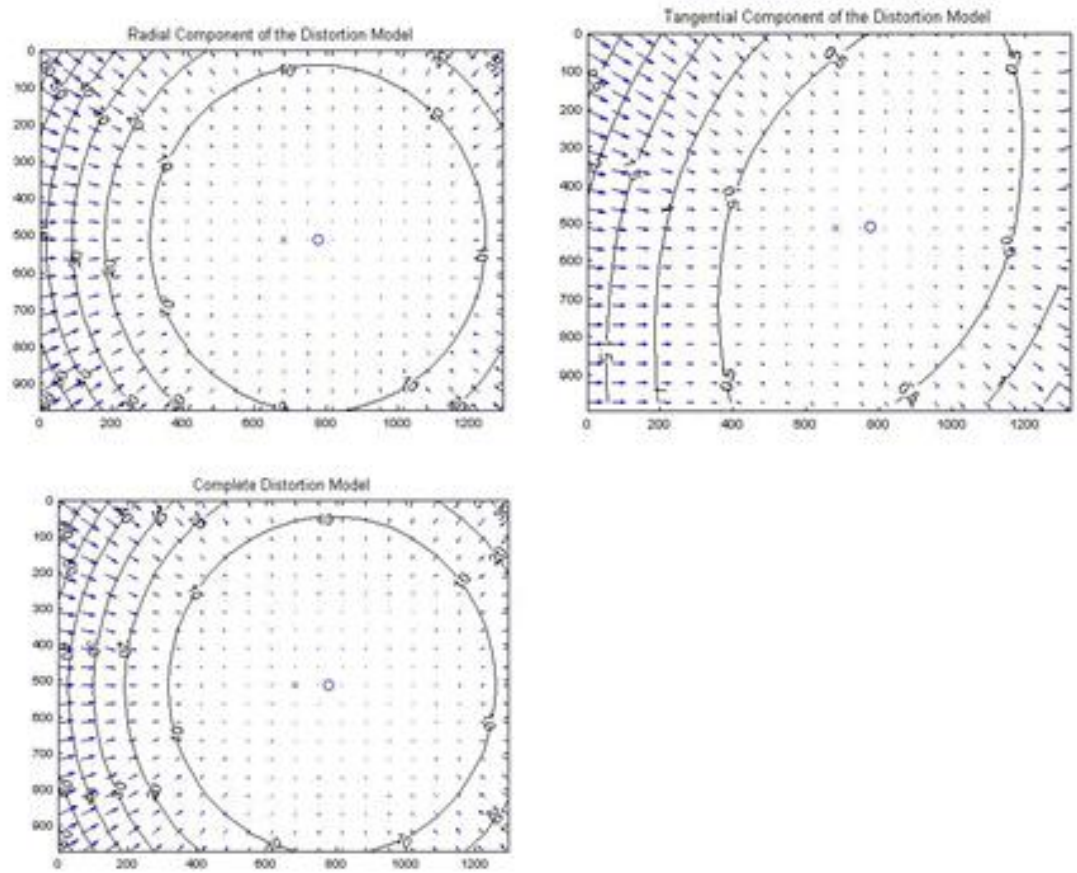


Figure 2.1. Example of radial, tangential, and complete distortion model for a HabCam camera.

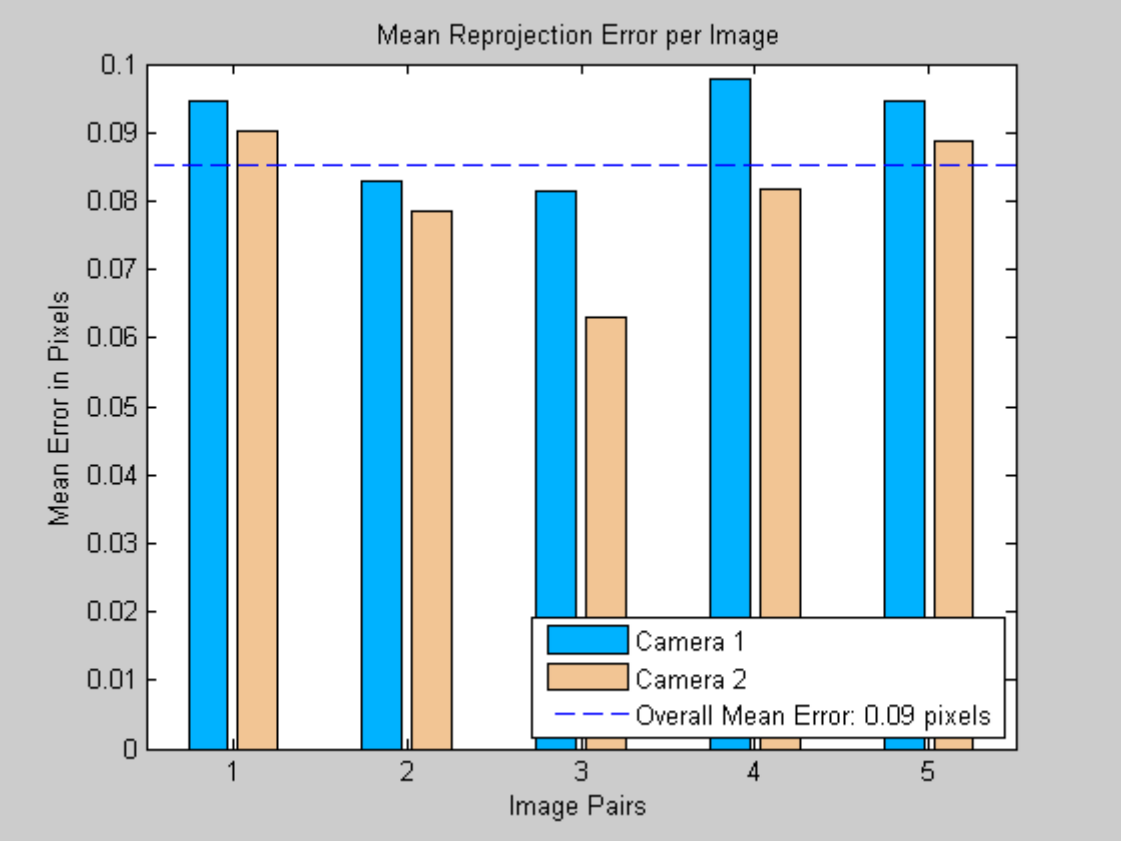


Figure 2.2. Reprojection errors associated with intrinsic calibration for stereo cameras on HabCamV4. Overall error is 0.09 pixels

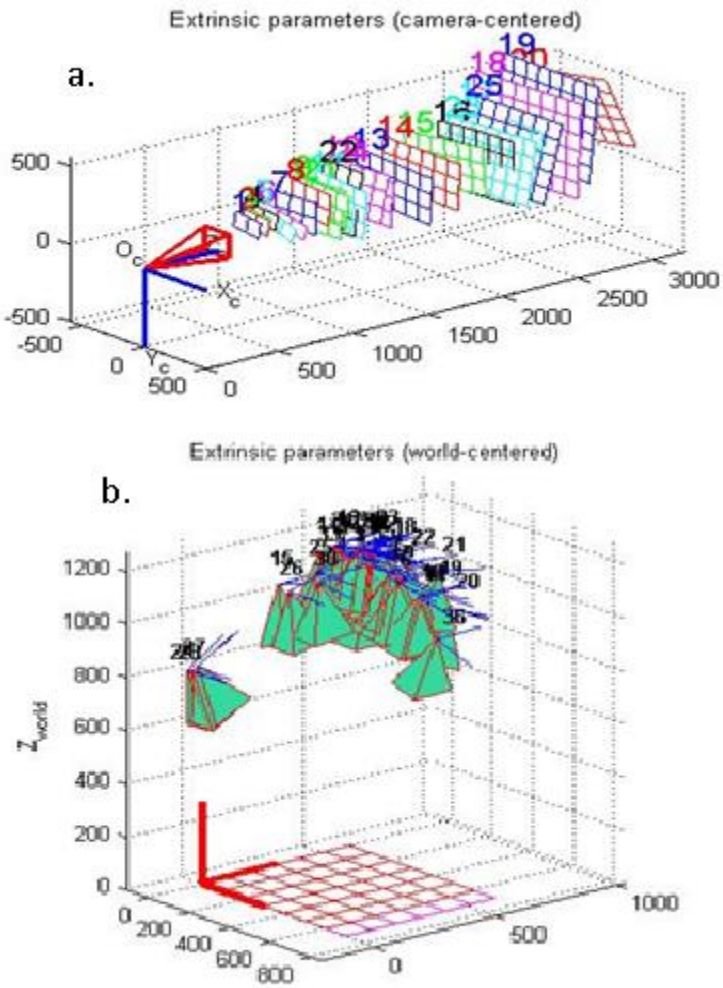


Figure 2.3. (a) Camera centric views of 28 orientations and altitudes to build extrinsic parameter list and (b) World centric views of 28 orientations and altitudes.

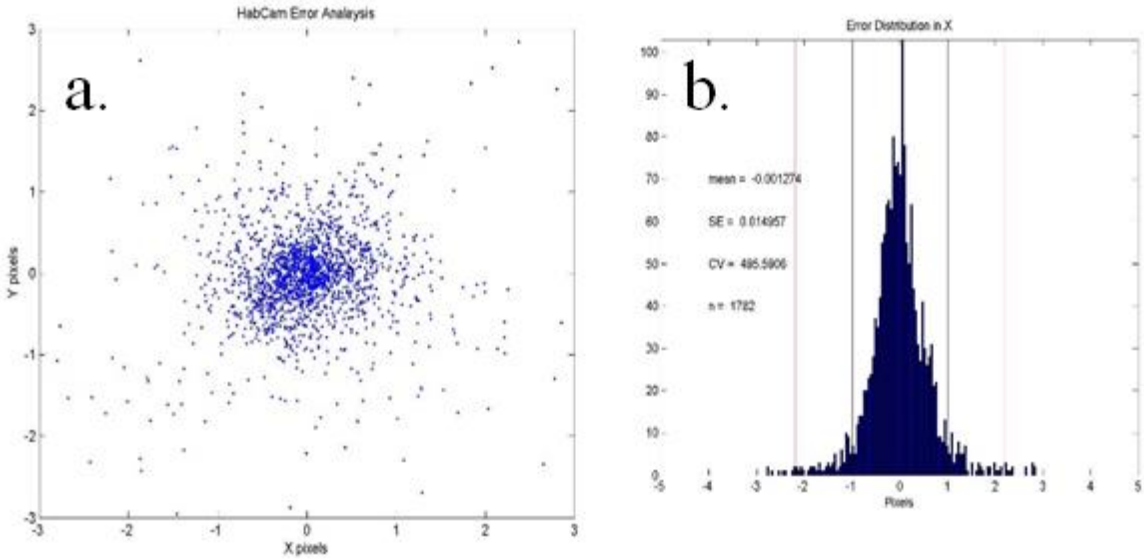


Figure 2.4. (a) Scatter plot of pixel error around the origin. (b) Frequency distribution of pixel error along x axis.

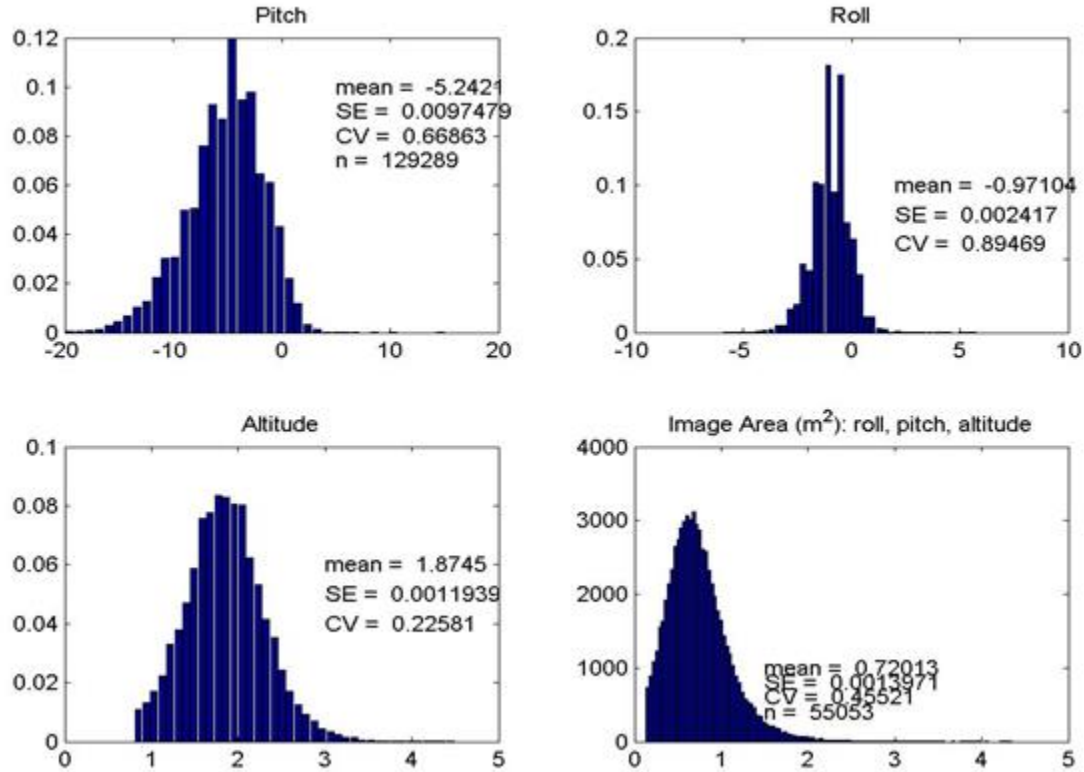


Figure 2.5. Example frequency distributions of pitch, roll, altitude, and image area (FOV)

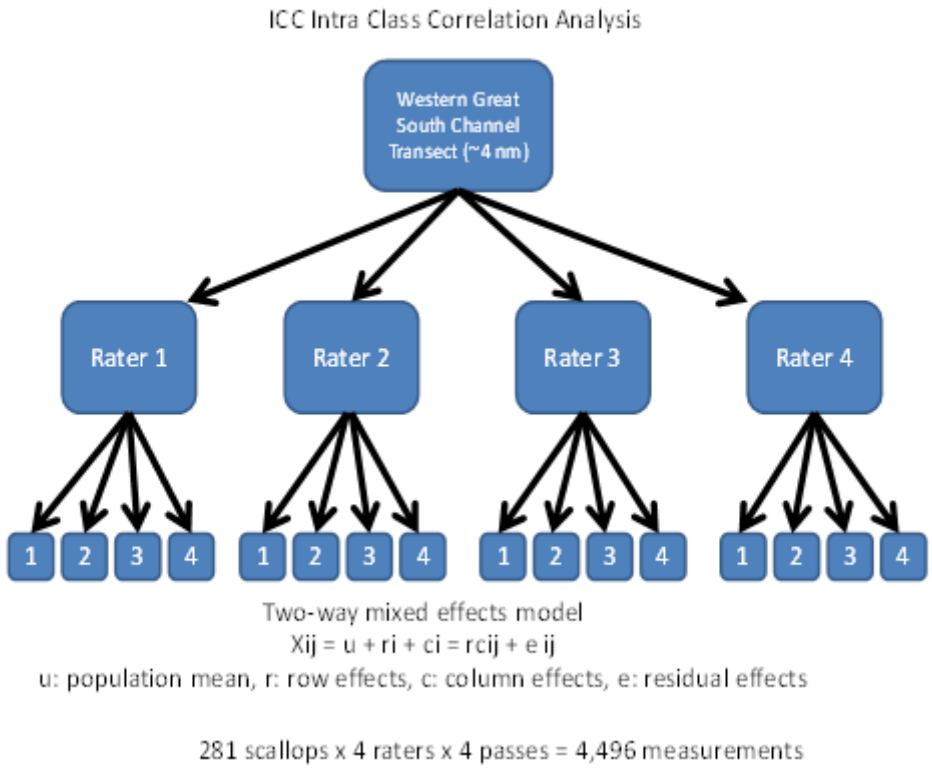


Figure 2.6. Inter Class Correlation analysis of scallop shell height measurements. Four individuals measured scallops from one transect four times.

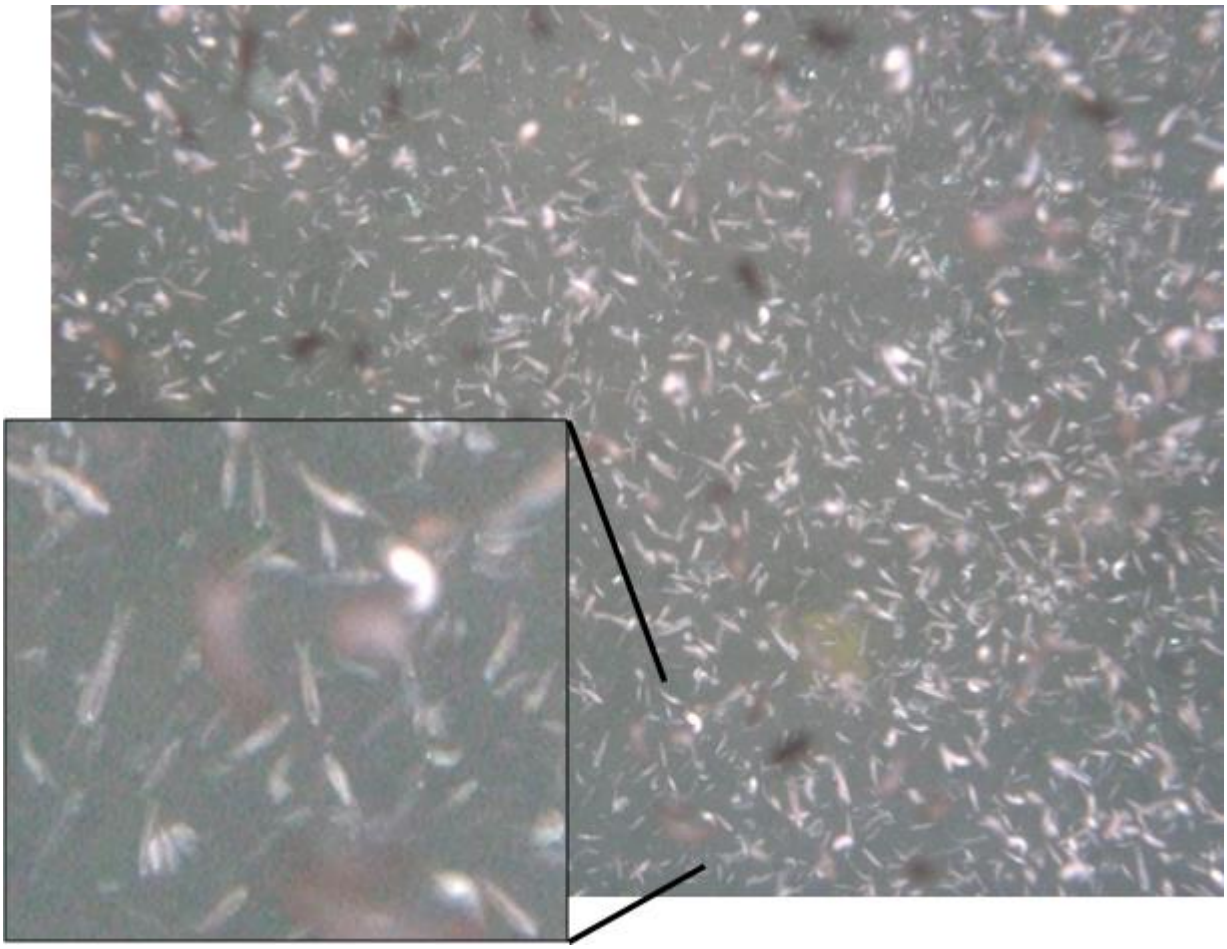
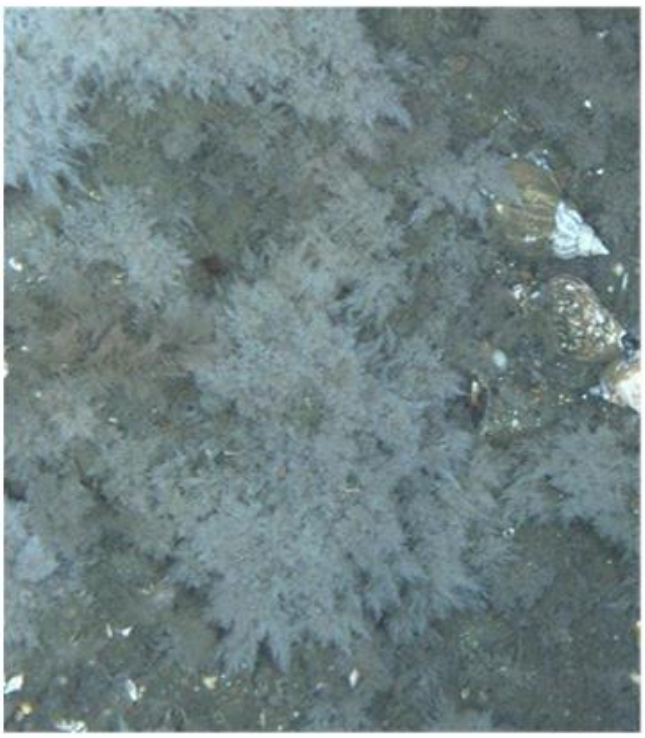
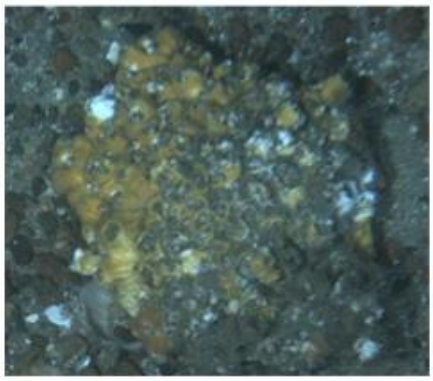


Figure 2.7. Seafloor Obscuration due to heavy concentration of plankton.

Level of encrustation that is still called a scallop



Level of epifauna encrustation that could contain a scallop.

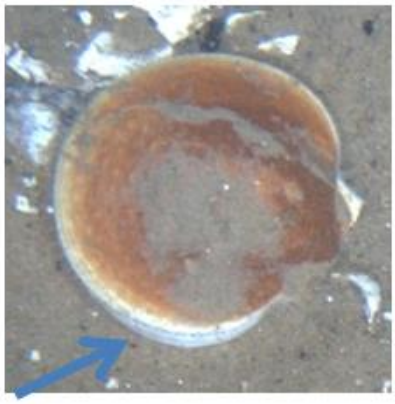


Level of encrustation that could contain a scallop.
No scallop annotated.

Figure 2.8. Scallop Encrustation; A scallop is annotated if the annotator can see part of the shell.

Shadow
Notice the shadow around scallop edge. This means the scallop has a bottom shell which raises it slightly up off the bottom.

Tentacles visible



Can see both shells
Can see both halves of shell without enough separation to be a clapper.

Clean edges
When a scallop is silted over but live, the edges are still clean. Likely due to movement and filtering.

Figure 2.9. Qualities of Live Scallops. Not all qualities are visible for every scallop but any of the qualities indicate a live scallop.

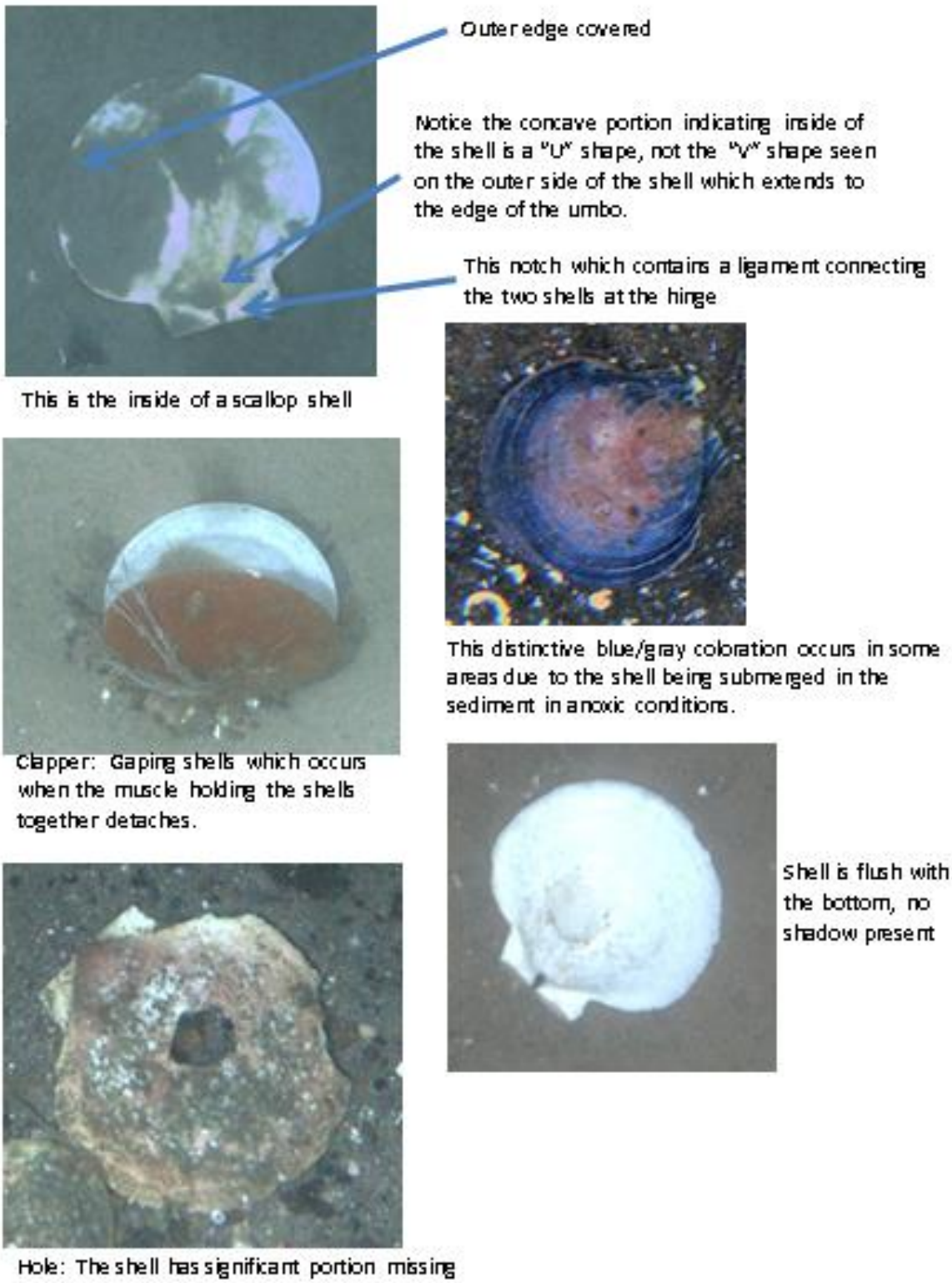


Figure 2.10. Qualities of Dead Scallops; Not all qualities are visible for every scallop but any of the qualities indicate a dead scallop.

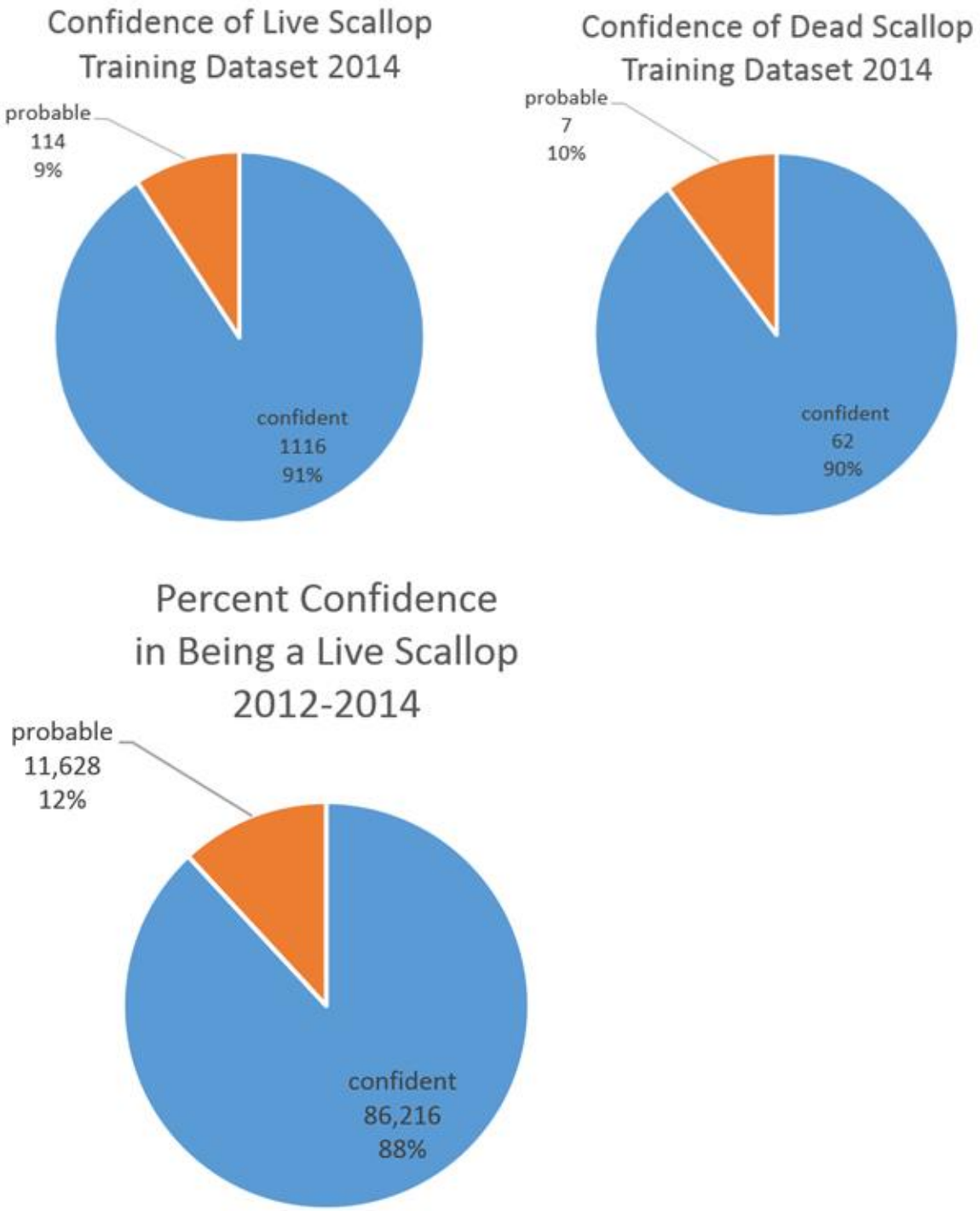


Figure 2.11. Confidence in being a live or dead scallop. Above: Results from one expert annotator upon completion of a standard image training set. Left: Results from all annotators on all images manually annotated between 2012-2014.

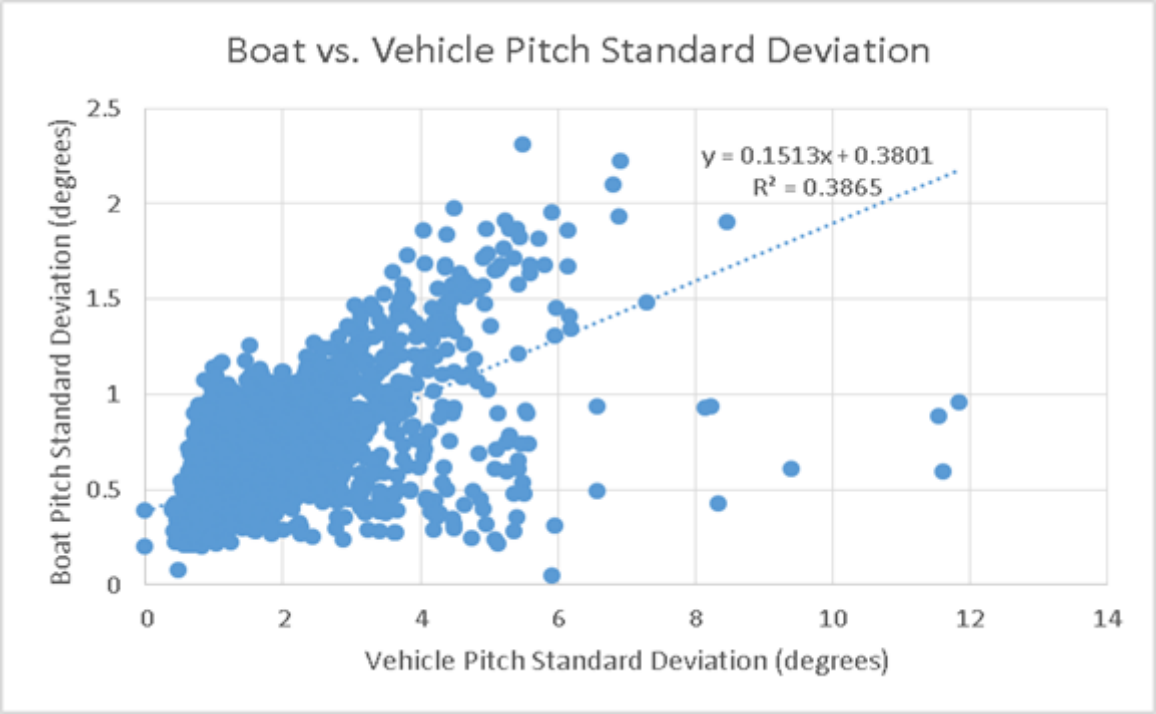


Figure 2.12. Relationship between the variation (STD) in vehicle pitch and the variation in ship heave (STD).

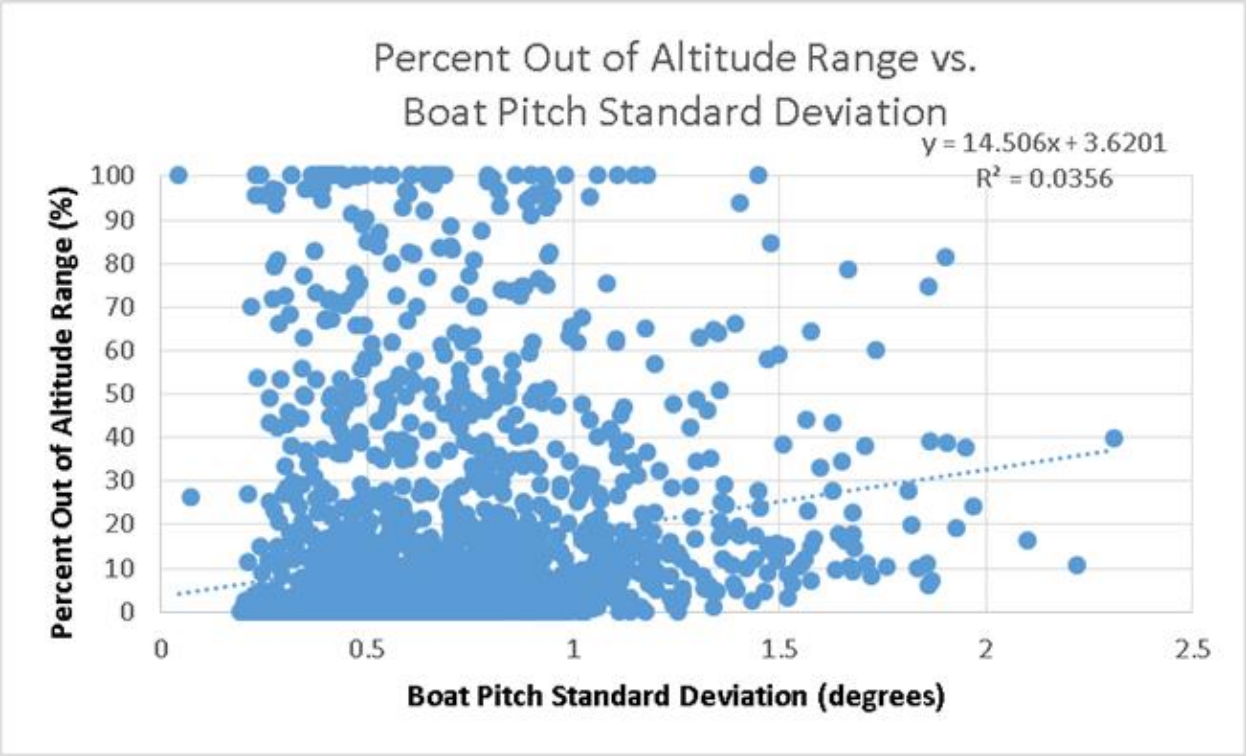


Figure 2.13. Relationship between variation in ship heave and the percentage of images that are out of good altitude range.

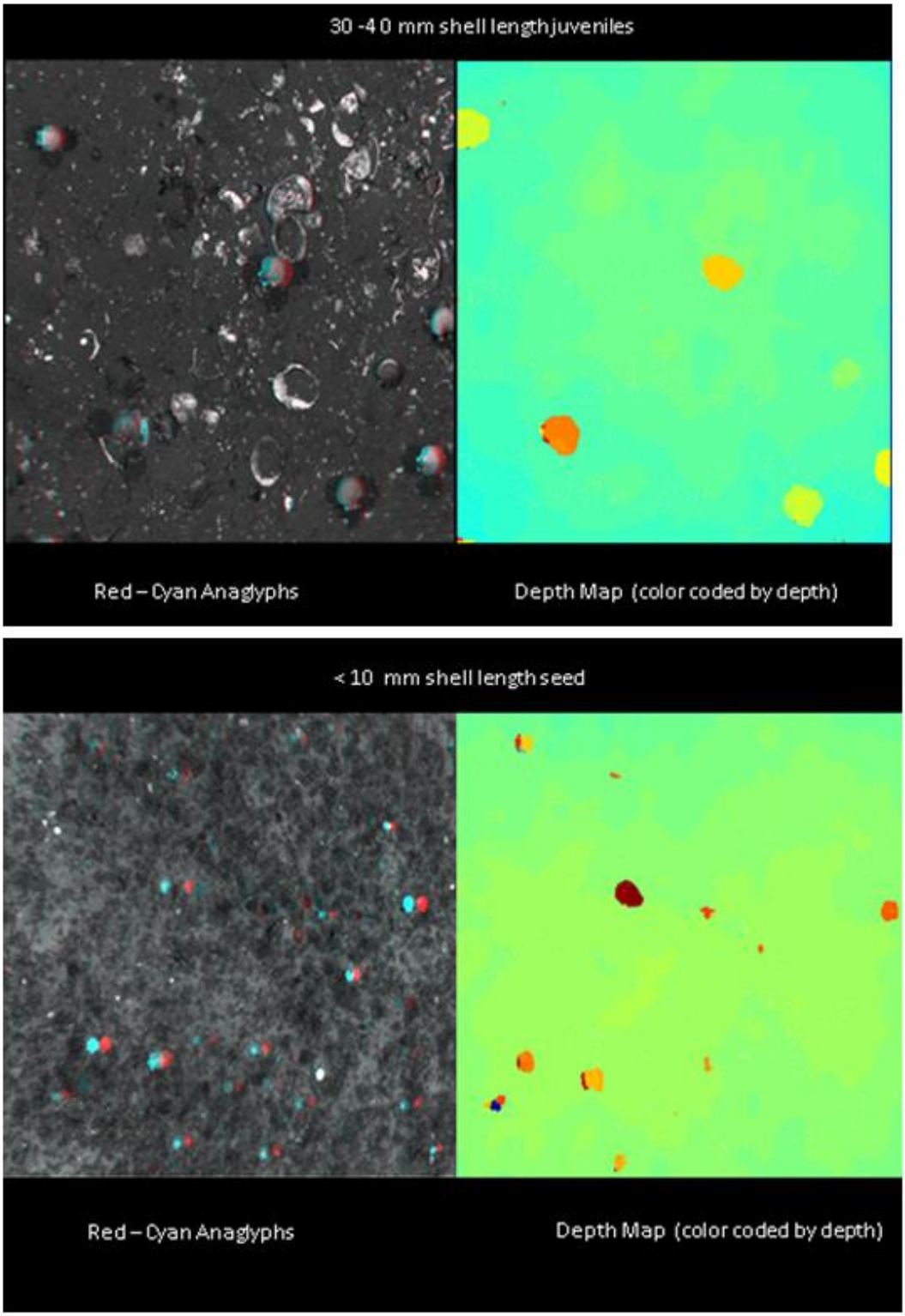


Figure 2.14. Red/Cyan Anaglyphs (left) and 3D depth map projections (right) of 2 year old scallops (top) and less than 1 year old scallops (bottom). The color of the scallop in the depth map indicates the altitude off the bottom.

Figure 2.15. Comparison between dredge and HabCam shell heights, by region and year.

Appendix 2.1. **Annotation Rules for Cruise 201403 Legs 1, 2, 3, 4**

Identifying Categories:

- sea scallops (measure with the line tool)
    - measure shell height, umbo to outer margin when possible
- "unidentified fish" (not species specific, just "unidentified fish")
    - a later assignment will speciate all the "fish" categories.
    - use box tool around the fish
- "unidentified skate"
    - use box tool around the skate
- highlights (good examples, or rare species, or unusual phenomenon)
    - highlight is an image note
    - highlight with *, ** , or *** with three stars being a more incredible image.
- "standing order species" organisms of interest that should be noted when present.
    - These are on the list right now:
        - Didemnum (measure with bounding box of some of it in the image).
        - "jonah or rock crab" measure with line tool across width of carapace.
        - "waved whelk" measure with bounding box
        - "lobster" measure with bounding box
- "dust cloud" - when there is a poof of sediment visible (with or without animal that caused it)
        - use bounding box

Subcategories (scallops):

The following can be applied to scallop class:

- "clapper": two valves together and gaping.
    - measure the distance from umbo to bottom shell in a clapper situation.
- "swimming" - when a scallop is above the seafloor, look for shadows.
- probable (explained below)
- "inexact" measurement (use this when you can't get a good length or width due to object being cut off on border or when having to estimate due to being partially obscured.
    - example: "inexact sea scallop", means that the length measurement is not exact.

Confidence:

probable (for when you are unsure of a classification but think it is what you classified it as)

- example : "sea scallop" + "probable" would be used when think it is a scallop but can't be completely sure.
- usage: will be reviewed in QAQC by a reviewer but anything left as probable after QAQC will be used in data products
- use probable + live when you think it is alive.
- use probable + dead when you think it is dead.Confidence:
probable (for when you are unsure of a classification but think it is what you classified it as)

Special Cases:

- Mark seed with a point if it is dense otherwise get a length if you can.

Substrate:

identify dominant, and any number of subdominant substrate types

example: (boulder +gravel +sand +shell hash)

Order subdominants by decreasing percent cover in the images.

For information about substrate categories EXAMPLE SUBSTRATE IMAGES and refer to the following chart:

Comparison Chart of Substrate Classifications.

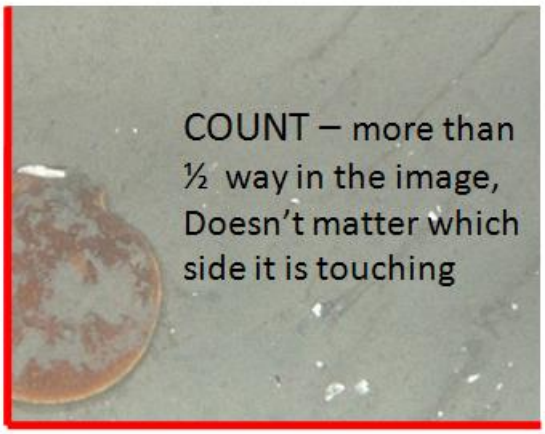
Grain size (mm)	Greene et al 1999	Valentine 2005	CMECS 2005	CMECS 2012	HabCam 2005 - 2012
< 0.004				clay	Mud
0.004- 0.0625				Silt	
< 0.06	mud	Fine grained sediment	Mud (<0.07)	Mud (<0.0625)	
0.06 – 2	Sand	Doesn't fit: ranges "Fine grained sediment "and "granules"(1-2mm)	Sand (>0.07-2)	Sand (>0.0625 – 2)	Sand
2-4	Gravel	?	Gravel	Granule	Gravel
4-24		?	Pebble (4-74)	Pebble	
24-64	Pebble	pebble			
64-256	Cobble	Cobble	Cobble (74-257)	Cobble	Cobble
256mm-3m	Boulder	Boulder	Boulder > 257	Boulder (246mm-4m)	Boulder
>3m	?				

- Note that some of the HabCam schema was because delineations in other classifications were not plausible.
  - Our "Gravel" of 2-24mm combines CMECS 2012 schema of granule and pebble because we can't tell from our imagery difference between grain sizes less than 4mm from grain sizes > 4mm.
  - We can't tell differences in fine grained sediment. Our Mud category is based on if annotators can't see the grains.
  - There is some inconsistency with annotators using mud vs. sand so in data extraction the categories of mud and sand can be combined into "fine-grained sediment."

Measuring Rules - Borders

Count all organisms that are more than half way in the image.

The border rule is only in effect if the organism is exactly half way in the image. Then only count the organisms that are half way in the top and right sides.



For Reference:

The old MIP guide that still has some relevant diagrams like border rules and examples of unreadable images, etc:

* MIP guide (PDF)

Sites:

Annotation Tool:
http://habcam-data.whoi.edu/static/annotator.html
UPDATE LOGS Has notes about annotation tool releases

Procedure

Starting from login:

1. log into site with your username and any password (later it will be password protected but not right now).
    a. You will see the “previous” and “next” buttons disappear and are left with “Next New” only.
2. Select your Assignment

For each image:

3. Measure whatever you can in the image with the proper category and hit commit.
    a. Note that the color of your annotation changes once you hit commit.
4. Check the Dominant Substrate and Subdominant substrate assignment to make sure it is correct.
5. Check if there are any notes you need to make about the image.
6. When you have marked everything in the image you want to, click the “Next New” Button. Which will:
    a. record substrate info
    b. record image notes
    c. records any annotations you haven’t already committed.
    d. takes you to the next image that hasn’t been done in your assignment.
7. Repeat steps 3-7 for every image.

* If you want to be able to use the previous and next buttons without recording data, you can open the annotation tool in a new browser window and don’t login. You will then have the next and previous buttons available to you.

Appendix 2.2. Example output from comparison of training sets between a trainee and expert. Tables and figures are intermixed for continuity.

Table A2.2.1 Tablulation of decisions to accept or reject images. The expert and trainee disagree on 8 images (shaded), 6 of which the expert rejected and trainee accepted and two that the trainee accepted and the expert rejected.

		Trainee	
		Acceptable	Rejected
Expert	Acceptable	302	2
	Rejected	6	14

Table A2.2 List of images where the trainee and expert disagreed, allowing a trainee and reviewer to easily review the discrepancies.

Imagename	ImageNum	Expert	Trainee
http://habcam-data.whoi.edu/data/201203.20120602.101501504.33105.jpg	10	Rejected	Acceptable
http://habcam-data.whoi.edu/data/201203.20120616.093605271.61466.jpg	202	Rejected	Acceptable
http://habcam-data.whoi.edu/data/201203.20120616.225416247.9956.jpg	225	Rejected	Acceptable
http://habcam-data.whoi.edu/data/201203.20120616.233328472.22962.jpg	231	Acceptable	Rejected
http://habcam-data.whoi.edu/data/201203.20120617.031825567.99648.jpg	239	Rejected	Acceptable
http://habcam-data.whoi.edu/data/201203.20120617.034809454.109425.jpg	242	Rejected	Acceptable
http://habcam-data.whoi.edu/data/201203.20120618.133045530.29539.jpg	274	Acceptable	Rejected
http://habcam-data.whoi.edu/data/201203.20120619.230742515.49051.jpg	308	Rejected	Acceptable

Table A2.2.3 Comparison of dimensions used to measure scallops: Inexact – scallop was partially obscured or off the image, Length – standard shell height umbo to opposite margin, Width – distance between lateral margins. Some disagreement is expected (shaded cells) but the trainee often uses a length measurement where the expert uses shell widths, worth reviewing. The effects of this are seen later.

		Trainee		
		Inexact	Length	Width
Expert	Inexact	8	6	5
	Length	3	185	4
	Width	6	33	42

Table A2.2.4 Comparison of scallop identification classes. Shaded cells represent critical discrepancies where the expert and trainee disagree on whether a scallop is alive or dead. All disagreements are reviewed and discussed.

		Trainee						
		dead scallop	live scallop	probable dead scallop	probable live scallop	probable swimming scallop	scallop clapper	swimming scallop
Expert	dead scallop	2	3	0	1	0	0	0
	live scallop	11	160	1	1	2	0	20
	probable dead scallop	0	1	1	0	0	0	0
	probable live scallop	1	22	1	0	0	0	0
	scallop clapper	0	11	0	0	0	4	0
	swimming scallop	0	4	0	0	0	0	12

Table A2.2.5 List of critical discrepancies in scallop class. Xpos and Ypos columns give the location of the object from the top left corner of the image.

Imagename	Image	Xpos	Ypos	Expert Class	Trainee Class
http://habcam-data.whoi.edu/data/201203.20120612.000600607.93050.jpg	175	0.30	0.68	dead scallop	live scallop
http://habcam-data.whoi.edu/data/201203.20120612.001840199.96450.jpg	176	0.94	0.90	dead scallop	live scallop
http://habcam-data.whoi.edu/data/201203.20120617.204053703.115488.jpg	261	0.56	0.06	dead scallop	probable live scallop
http://habcam-data.whoi.edu/data/201203.20120619.013242416.84757.jpg	293	0.28	0.84	dead scallop	live scallop
http://habcam-data.whoi.edu/data/201203.20120602.113527828.57060.jpg	23	0.64	0.75	live scallop	dead scallop
http://habcam-data.whoi.edu/data/201203.20120602.195811615.206775.jpg	63	0.50	0.02	live scallop	probable dead scallop
http://habcam-data.whoi.edu/data/201203.20120620.171855797.163299.jpg	330	0.86	0.36	live scallop	dead scallop
http://habcam-data.whoi.edu/data/201203.20120620.171855797.163299.jpg	330	0.92	0.04	live scallop	dead scallop
http://habcam-data.whoi.edu/data/201203.20120622.052520108.188795.jpg	335	0.53	0.12	live scallop	dead scallop
http://habcam-data.whoi.edu/data/201203.20120622.052520108.188795.jpg	335	0.83	0.10	live scallop	dead scallop
http://habcam-data.whoi.edu/data/201203.20120622.052520108.188795.jpg	335	0.63	0.34	live scallop	dead scallop
http://habcam-data.whoi.edu/data/201203.20120622.052520108.188795.jpg	335	0.32	0.73	live scallop	dead scallop
http://habcam-data.whoi.edu/data/201203.20120622.052520108.188795.jpg	335	0.26	0.45	live scallop	dead scallop
http://habcam-data.whoi.edu/data/201203.20120622.052520108.188795.jpg	335	0.30	0.26	live scallop	dead scallop
http://habcam-data.whoi.edu/data/201203.20120622.052520108.188795.jpg	335	0.06	0.40	live scallop	dead scallop
http://habcam-data.whoi.edu/data/201203.20120622.052520108.188795.jpg	335	0.10	0.54	live scallop	dead scallop
http://habcam-data.whoi.edu/data/201203.20120629.035013474.20855.jpg	342	0.33	0.86	probable dead scallop	live scallop
http://habcam-data.whoi.edu/data/201203.20120618.234404362.45812.jpg	289	0.41	0.97	probable live scallop	probable dead scallop
http://habcam-data.whoi.edu/data/201203.20120622.003556801.85055.jpg	334	0.32	0.83	probable live scallop	dead scallop
http://habcam-data.whoi.edu/data/201203.20120617.201721066.108477.jpg	260	0.30	0.35	scallop clapper	live scallop
http://habcam-data.whoi.edu/data/201203.20120629.035013474.20855.jpg	342	0.73	0.46	scallop clapper	live scallop
http://habcam-data.whoi.edu/data/201203.20120629.040531902.25891.jpg	343	0.56	0.28	scallop clapper	live scallop
http://habcam-data.whoi.edu/data/201203.20120629.040531902.25891.jpg	343	0.60	0.37	scallop clapper	live scallop
http://habcam-data.whoi.edu/data/201203.20120629.040531902.25891.jpg	343	0.39	0.48	scallop clapper	live scallop
http://habcam-data.whoi.edu/data/201203.20120629.040531902.25891.jpg	343	0.60	0.27	scallop clapper	live scallop
http://habcam-data.whoi.edu/data/201203.20120629.040531902.25891.jpg	343	0.59	0.16	scallop clapper	live scallop
http://habcam-data.whoi.edu/data/201203.20120629.040911111.27093.jpg	344	0.68	0.14	scallop clapper	live scallop
http://habcam-data.whoi.edu/data/201203.20120629.040911111.27093.jpg	344	0.55	0.62	scallop clapper	live scallop
http://habcam-data.whoi.edu/data/201203.20120629.040911111.27093.jpg	344	0.57	0.28	scallop clapper	live scallop
http://habcam-data.whoi.edu/data/201203.20120704.005608848.26991.jpg	345	0.46	0.18	scallop clapper	live scallop

Table A2.2.6 List of cases where one annotator (expert or trainee) marked a scallop that the other did not. Xpos and Ypos columns give the location of the object from the top left corner of the image.

Imagename	Image	Expert Class	Trainee Class	Xpos	Ypos
http://habcam-data.whoi.edu/data/201203.20120602.093925066.22501.jpg	9	live scallop	NA	0.72	0.37
http://habcam-data.whoi.edu/data/201203.20120602.114319884.59403.jpg	26	live scallop	NA	0.95	0.80
http://habcam-data.whoi.edu/data/201203.20120602.174712692.167768.jpg	54	live scallop width	NA	0.27	0.33
http://habcam-data.whoi.edu/data/201203.20120602.180721941.173770.jpg	57	live scallop	NA	0.12	0.03
http://habcam-data.whoi.edu/data/201203.20120602.194204339.201974.jpg	61	scallop clapper	NA	0.74	0.53
http://habcam-data.whoi.edu/data/201203.20120602.195811615.206775.jpg	63	live scallop inexact	NA	0.50	0.02
http://habcam-data.whoi.edu/data/201203.20120603.111751472.36457.jpg	72	live scallop width	NA	0.96	0.16
http://habcam-data.whoi.edu/data/201203.20120603.112707383.39216.jpg	73	live scallop	NA	0.45	0.89
http://habcam-data.whoi.edu/data/201203.20120603.112707383.39216.jpg	73	live scallop inexact	NA	0.25	0.14
http://habcam-data.whoi.edu/data/201203.20120603.112707383.39216.jpg	73	live scallop inexact	NA	0.80	0.35
http://habcam-data.whoi.edu/data/201203.20120603.200252135.192797.jpg	87	live scallop	NA	0.84	0.28
http://habcam-data.whoi.edu/data/201203.20120603.200252135.192797.jpg	87	live scallop	NA	0.48	0.44
http://habcam-data.whoi.edu/data/201203.20120603.210458269.211290.jpg	88	live scallop	NA	0.62	0.28
http://habcam-data.whoi.edu/data/201203.20120609.092603356.92919.jpg	111	live scallop	NA	0.81	0.45
http://habcam-data.whoi.edu/data/201203.20120609.193934406.80000.jpg	137	live scallop	NA	0.45	0.36
http://habcam-data.whoi.edu/data/201203.20120612.001840199.96450.jpg	176	live scallop inexact	NA	0.94	0.90
http://habcam-data.whoi.edu/data/201203.20120616.115124273.96269.jpg	209	swimming scallop	NA	0.42	0.84
http://habcam-data.whoi.edu/data/201203.20120616.131656117.117453.jpg	215	live scallop	NA	0.75	0.78
http://habcam-data.whoi.edu/data/201203.20120616.131656117.117453.jpg	215	live scallop	NA	0.23	0.38
http://habcam-data.whoi.edu/data/201203.20120602.093856658.22360.jpg	8	NA	swimming scallop inexact	0.65	0.01
http://habcam-data.whoi.edu/data/201203.20120602.111934449.52328.jpg	17	NA	live scallop	0.32	0.59
http://habcam-data.whoi.edu/data/201203.20120602.111934449.52328.jpg	17	NA	live scallop	0.60	0.84
http://habcam-data.whoi.edu/data/201203.20120602.113039521.55629.jpg	19	NA	live scallop width	0.32	0.97
http://habcam-data.whoi.edu/data/201203.20120611.080534580.55403.jpg	171	NA	live scallop	0.04	0.87
http://habcam-data.whoi.edu/data/201203.20120612.054436452.193802.jpg	187	NA	live scallop	0.74	0.42
http://habcam-data.whoi.edu/data/201203.20120612.054436452.193802.jpg	187	NA	live scallop inexact	0.02	0.81
http://habcam-data.whoi.edu/data/201203.20120616.093605271.61466.jpg	202	NA	probable live scallop	0.35	0.98
http://habcam-data.whoi.edu/data/201203.20120617.102427095.80093.jpg	254	NA	live scallop	0.12	0.28
http://habcam-data.whoi.edu/data/201203.20120618.091815036.45376.jpg	269	NA	live scallop width	0.97	0.13

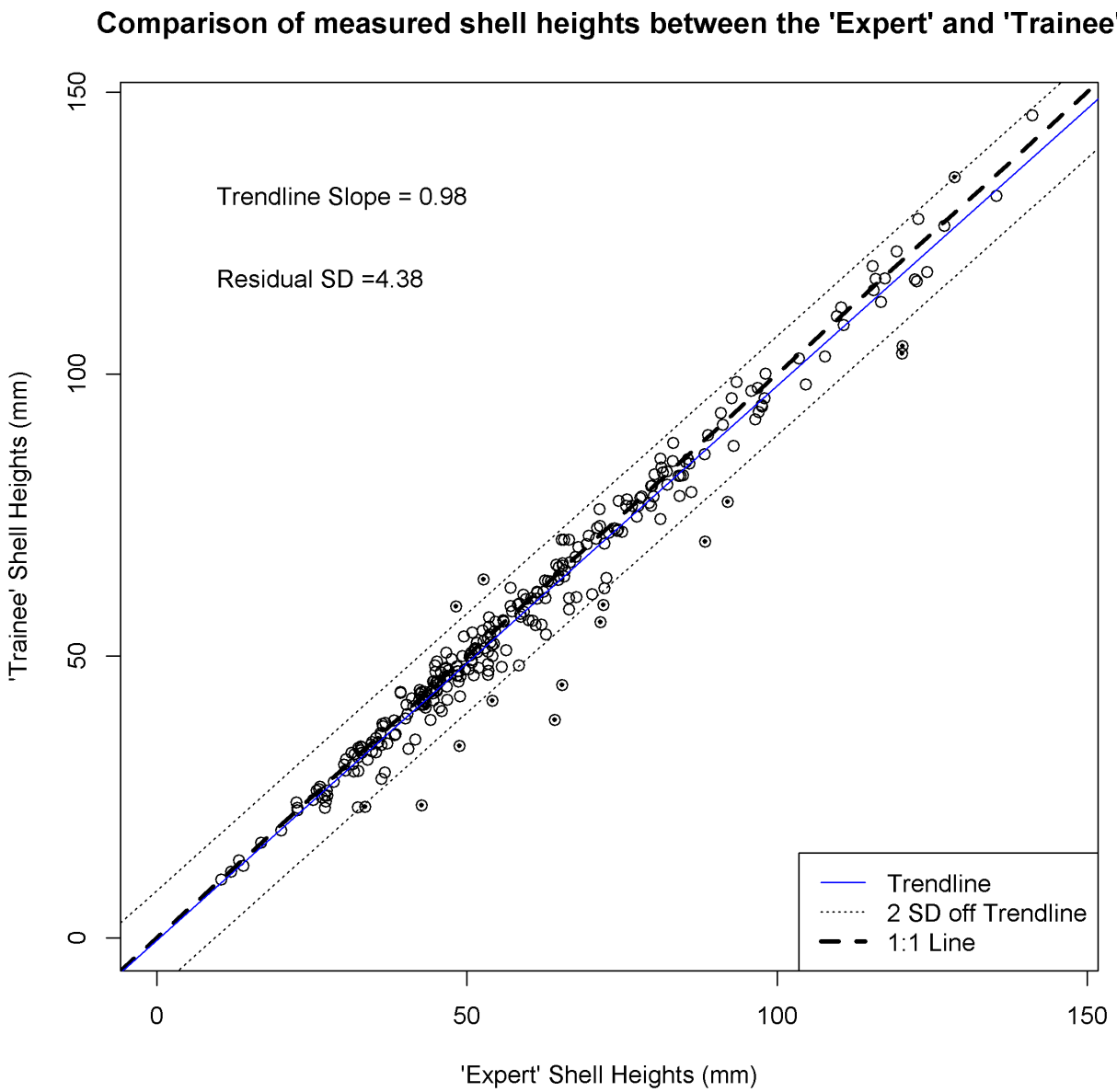


Figure A2.2.1 Shell height measurement comparison with a trendline and 1:1 line. Points outside of 2 standard deviations from the trendline are marked with a center point. The slope of less than one and pattern of negative residuals, matched with the trainee's tendency to use lengths measurement when the expert uses widths suggest that the trainee is incorrectly measuring scallops that are sitting on inclines.

Table A2.2.7 List of scallops where expert and trainee measurements by greater than standardized residuals (Figure A2.1). . Xpos and Ypos columns give the location of the object from the top left corner of the image. Ex_SH and Tr_SH give the measured Expert and Trainee shell heights. Note that most gross mis-measurements by the trainee tend to be inexact or width measures for the expert.

Imagename	Image#	Xpos	Ypos	Expert_ClassName	Trainee_ClassName	Ex_SH	Tr_SH	Resid	StdResid
http://habcam-data.whoi.edu/data/201203.20120602.114616374.60279.jpg	29	0.55	0.97	live scallop inexact	live scallop	128.64	134.96	8.93	2.04
http://habcam-data.whoi.edu/data/201203.20120602.114616374.60279.jpg	29	0.12	0.68	live scallop inexact	probable live scallop	52.64	63.60	12.26	2.80
http://habcam-data.whoi.edu/data/201203.20120602.121709949.69479.jpg	36	0.37	0.70	scallop clapper width	scallop clapper	120.13	103.66	-14.00	-3.20
http://habcam-data.whoi.edu/data/201203.20120602.124159852.76874.jpg	40	0.31	0.49	swimming scallop width	live scallop	33.53	23.34	-9.22	-2.10
http://habcam-data.whoi.edu/data/201203.20120609.033015927.7184.jpg	102	0.03	0.30	swimming scallop width	swimming scallop	42.66	23.56	-17.96	-4.10
http://habcam-data.whoi.edu/data/201203.20120612.053138676.189695.jpg	186	0.92	0.18	swimming scallop width	live scallop inexact	65.35	44.88	-18.95	-4.32
http://habcam-data.whoi.edu/data/201203.20120612.072950489.227140.jpg	189	0.40	0.43	swimming scallop width	swimming scallop	48.79	34.10	-13.45	-3.07
http://habcam-data.whoi.edu/data/201203.20120617.030050905.93865.jpg	238	0.54	0.69	swimming scallop width	live scallop	64.14	38.72	-23.91	-5.46
http://habcam-data.whoi.edu/data/201203.20120617.064647314.14971.jpg	243	0.95	0.32	live scallop width	live scallop	71.55	56.06	-13.86	-3.16
http://habcam-data.whoi.edu/data/201203.20120617.101957314.78754.jpg	253	0.83	0.11	live scallop	live scallop	120.25	104.96	-12.83	-2.93
http://habcam-data.whoi.edu/data/201203.20120619.065021380.50524.jpg	299	0.52	0.26	live scallop width	live scallop	72.00	59.12	-11.24	-2.57
http://habcam-data.whoi.edu/data/201203.20120620.045607974.16699.jpg	313	0.04	0.41	swimming scallop width	live scallop inexact	88.38	70.30	-16.17	-3.69
http://habcam-data.whoi.edu/data/201203.20120622.052811147.189817.jpg	337	0.54	0.18	live scallop width	live scallop	92.07	77.41	-12.68	-2.89
http://habcam-data.whoi.edu/data/201203.20120629.040911111.27093.jpg	344	0.80	0.12	scallop clapper width	scallop clapper	54.09	42.12	-10.64	-2.43
http://habcam-data.whoi.edu/data/201203.20120629.040911111.27093.jpg	344	0.39	0.80	live scallop	live scallop	48.20	58.82	11.85	2.70

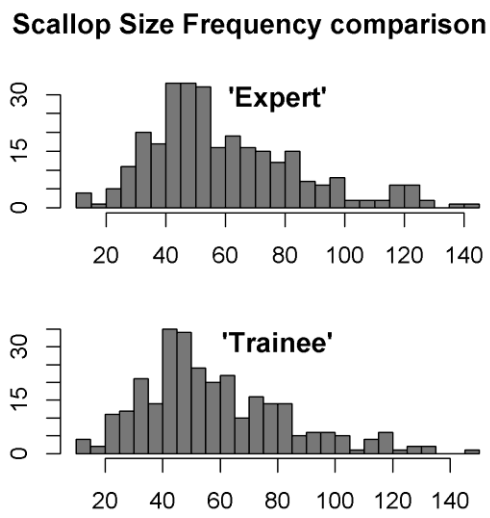


Figure A2.2.2 Histogram of shell size frequencies for expert and trainee.

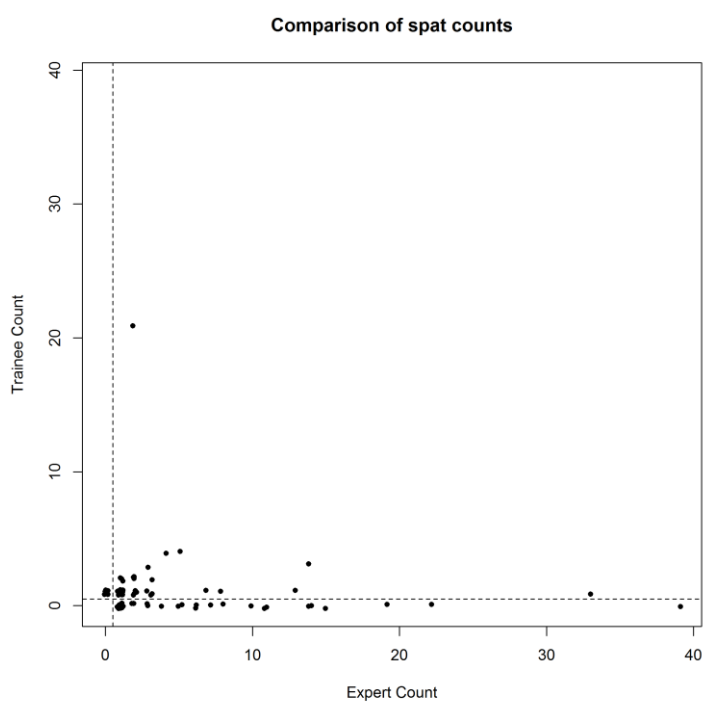


Figure A2.2.3 Numbers of spat (scallops <35mm) identified per image; expert vs trainee. We expect a lot of disagreement between annotators as spat are often difficult to positively identify in images and spat densities are not expected to be treated as robust quantitative data. However, in images with large numbers of apparent spat (>10/image) the trainee is usually missing them so further review on recognizing spat is required.

Table A2.2.8 List of images where expert and trainee spat counts differed markedly. Used for review and instruction on techniques for identifying spat in imagery.

Imagename	ImageNum	Expert count	Trainee count
http://habcam-data.whoi.edu/data/201203.20120602.111639164.51458.jpg	15	4	0
http://habcam-data.whoi.edu/data/201203.20120602.111701535.51569.jpg	16	14	3
http://habcam-data.whoi.edu/data/201203.20120602.111934449.52328.jpg	17	11	0
http://habcam-data.whoi.edu/data/201203.20120602.113039521.55629.jpg	19	3	0
http://habcam-data.whoi.edu/data/201203.20120602.113106319.55762.jpg	20	3	0
http://habcam-data.whoi.edu/data/201203.20120603.161013241.123518.jpg	81	5	0
http://habcam-data.whoi.edu/data/201203.20120603.162139509.126924.jpg	82	7	0
http://habcam-data.whoi.edu/data/201203.20120603.162323076.127438.jpg	83	5	0
http://habcam-data.whoi.edu/data/201203.20120609.095324864.101920.jpg	112	10	0
http://habcam-data.whoi.edu/data/201203.20120609.095703710.103120.jpg	113	14	0
http://habcam-data.whoi.edu/data/201203.20120609.183810507.59800.jpg	131	33	1
http://habcam-data.whoi.edu/data/201203.20120609.184338779.61600.jpg	132	39	0
http://habcam-data.whoi.edu/data/201203.20120609.185814156.66400.jpg	133	19	0
http://habcam-data.whoi.edu/data/201203.20120609.190229480.67800.jpg	135	6	0
http://habcam-data.whoi.edu/data/201203.20120616.115301322.96646.jpg	210	8	0
http://habcam-data.whoi.edu/data/201203.20120616.131546270.117070.jpg	214	22	0
http://habcam-data.whoi.edu/data/201203.20120616.131656117.117453.jpg	215	14	0
http://habcam-data.whoi.edu/data/201203.20120616.131659218.117470.jpg	216	11	0
http://habcam-data.whoi.edu/data/201203.20120616.131812349.117871.jpg	217	15	0
http://habcam-data.whoi.edu/data/201203.20120617.001551509.38157.jpg	232	15	0
http://habcam-data.whoi.edu/data/201203.20120617.005717296.53010.jpg	233	13	1
http://habcam-data.whoi.edu/data/201203.20120617.011648608.59637.jpg	234	6	0
http://habcam-data.whoi.edu/data/201203.20120629.040531902.25891.jpg	343	2	21

Table A2.2.9. Cross-tabulation of fish classifications by the expert and trainee. Shaded areas represent areas of disagreement. In this case, all classifications agree except for two roundfish that the trainee ruled as <50% on the image. Misclassification of fishes is less problematic as experts later review all fish ID's to provide precise identifications.

		Trainee				
		flatfish	roundfish	Roundfish (less than half)	skate	skate (less than half)
Expert	flatfish	11	0	0	0	0
	roundfish	0	31	2	0	0
	roundfish (less than half)	0	0	2	0	0
	skate	0	0	0	21	0
	skate (less than half)	0	0	0	0	8

Table A2.2.10. List of fish that were not noted by both the expert and trainee. Xpos and Ypos columns give the location of the fish from the top left corner of the image.

Imagename	Image#	Expert class	Trainee Class	Xpos	Ypos
http://habcam-data.whoi.edu/data/201203.20120603.175516396.154801.jpg	86	roundfish	NA	0.42	0.69
http://habcam-data.whoi.edu/data/201203.20120603.113712850.42221.jpg	75	NA	fish	0.63	0.11
http://habcam-data.whoi.edu/data/201203.20120617.034809454.109425.jpg	242	NA	skate	0.68	0.48
http://habcam-data.whoi.edu/data/201203.20120620.051439718.23342.jpg	315	NA	flatfish	0.17	0.85

Terms of Reference 3. Biological sampling aspects of the surveys.
3.1 Sub-sampling procedures
3.2 Ability to sample all size classes.
3.3 Evaluate the utility of data to detect incoming recruitment
3.4 Assess the potential ability to assess fine scale ecology (*e.g.*,Allee effect, predator-prey interactions, disturbance from fishing gear, etc.).

3.1 Subsampling procedures

Clearly, it is unnecessary to have data from every HabCam image in order to estimate abundance and describe the spatial distribution. For this reason and because there is significant autocorrelation along the survey track, we have determined the appropriate subsampling (downsampling) rate. Figure 3.1 shows the along track abundance for every other image in a 4.5km track in the northern section of the Nantucket Lightship Closed Area. The abundance ranges from 0 to >20 scallops per m^2. The patchiness of this population can be assessed using Ripley's neighbor-k statistic (Fortin and Dale 2005). Figure 3.2 shows that for this population patchiness was significant at 700-900 m and at 3.2 km scales.

To determine the appropriate subsampling rate we simulated subsampling at various levels: images from this track were downsampled by extracting abundances and the CV at intervals ranging from every 4th, 8th, 10th, 12th etc. out to every 500th image. The mean and CV remained stable up to a downsample level of every 100th image. Therefore, the required subsampling rate was determined to be every 100th image (Figure 3.3).

We also need to subsample the trackline as a function of spatial distribution and density. We select sets of images for annotation, termed "assignments", based on the spatial extent of the image set and a target image density (see TOR1 for workflow). Images are selected for assignments differently, depending on whether the annotations are being conducted at-sea during a survey cruise or on land after a survey cruise. At sea, images are selected sequentially (i.e., every 100th image) and dynamically added to the active assignment as they are collected. Annotation at-sea was first employed in the 2013 survey and at-sea annotation provided ~60% of the final annotated image set in 2014. Thus, at-sea annotation is becoming increasingly important in manually annotating images for scallop assessment.

On land, we are able to employ a more complex and iterative method that selects images spatially with a goal of having an evenly distributed effort along the survey track, including the capability of setting target spatial densities and filling gaps from previous assignments where images were rejected due to poor quality or similar issues. Based on the desired density of images to be annotated, we break the survey track into equal length segments and select one image from each segment. Individual image selection within each segment is selected in part

based on preferred vehicle altitudes (Gaussian-weighted, based on known local issues with water turbidity or other factors that affect image quality), but image selection is otherwise random. Once an assignment is completed, additional assignments for that region may be created to increase the density of annotated images or fill gaps from rejected images (Figure 3.4). In this case, all images from a buffered region around each image that has already been annotated are removed from the pool of available images before the next random subset of images is selected. The goal of this is to keep the density of annotated images consistent within subregions along the track.

3.2 Ability to sample all size classes, and detection of incoming recruitment

Recruitment events occur in specific spatio-temporal areas. For the purpose of this document, we define pre-recruits as scallops less than 35mm shell height and recruits as scallops 35-75mm in shell height, approximately representing one and two year olds, respectively.

Recruits are of sufficient size that they are readily identified in HabCam images, given their size and the resolution of the camera system (TOR1, Figure 3.5) and the size-frequency distributions for this size range are comparable to the NEFSC survey dredge with its 38mm liner (NEFSC Dredge, TOR-2). This, while a few scallop recruits may be missed in cases of turbid water or poor image quality, we expect that the efficiency of the HabCam for quantifying scallop recruits is very high.

Pre-recruits are more difficult to identify due to their size and sometimes cryptic behavior. In some cases, they may be hidden underneath pebbles or epifauna. Additionally, they are of similar size to the gravel grain size common on Georges Bank. However, image resolution is high enough that the “wings” on the hinge of the shells are often visible in images (Figure 3.6, 3.7), distinguishing them from gravel. Additionally, pre-recruits exhibit active habitat selection and thus respond differently than passive particles (gravel) to the micro-topography of their habitat and are commonly swimming in the water column, which can be confirmed by viewing the image in 3D. Thus, with some basic training and feedback, annotators can be trained to recognize pre-recruits that are visible in HabCam imagery. Typically the annotator will not measure scallops less than 30mm but will mark it as a point to identify it as a pre-recruit scallop (Figure 3.6, 3.7. The enlargement in these figures shows numerous scallops 10mm or less).

However, pre-recruits are certainly not fully detected in HabCam imagery for the reasons discussed above and resulting data on pre-recruits are only qualitative but still usually sufficient to detect large recruitment events. For example, very high numbers of pre-recruits (often > 100 m^{-2}) were observed by both HabCam vehicles on the southern flank of Georges Bank and Nantucket Shoals in 2013 (as well as dredge surveys), which corresponded to high observations of two-year recruits in this area the next year.

3.3 Assess the potential ability to assess fine scale ecology (*e.g.*,Allee effect, predator-prey interactions, disturbance from fishing gear, etc.).

As dioecious broadcast spawners, the fertilization success of sea scallops is likely positively related to density (Allee effect, Claereboudt 1999, Smith and Rago 2004). Because HabCam images overlap, one-dimensional nearest neighbor distances can be calculated, and two-dimensional nearest-neighbors can be inferred, thus allowing estimation of fertilization success based on field data.

Direct predator-prey observations can be made from HabCam images. See Figure 3.8 for crab and sea star predation on scallops.

The HabcamV4 side-scan sonar data can be combined with the imagery to infer fine-scale terrain effects. Figure 3.9 shows a section of side scan data from the C3D on HabCamV4 in the Hudson Canyon Closed Area. Sand waves at a wavelength of 0.75m are clearly delineated. Stereo images were taken down the nadir and sand dollars can be seen lining up on top of the sand waves directly in association with the acoustic image. We are working on similar analyses of fine-scale spatial distributions of scallops throughout the study region.

Figure 3.10 shows the along track rugosity measured from the side scan data along a 450m track. Small scale variation on the order of cm shows up in the power spectrum where sand waves dominate the signal. At longer wavelengths at about 85m a second mode occurs indicating that some low frequency bathymetric feature dominates. This kind of data is available for the entire survey region each year since 2012 and will continue to be collected by HabCamV4 and incorporated into habitat characterization.

Also seen in the side-scan image are dredge tracks of a commercial scalloper (Figure 3.9). We often observe dredge tracks in the side scan but it is very difficult to see any disturbance optically in the cameras. We attribute this to the compression of the substrate by the dredge such that the acoustic impedance and reflectivity is enhanced, but very little can be seen optically. A current research project involves using the side scan maps to determine which images are inside of dredge tracks and comparing these images to adjacent images outside of the dredge tracks for faunal assemblages, habitat complexity and rugosity (see NEFSC_HabCam TOR8).

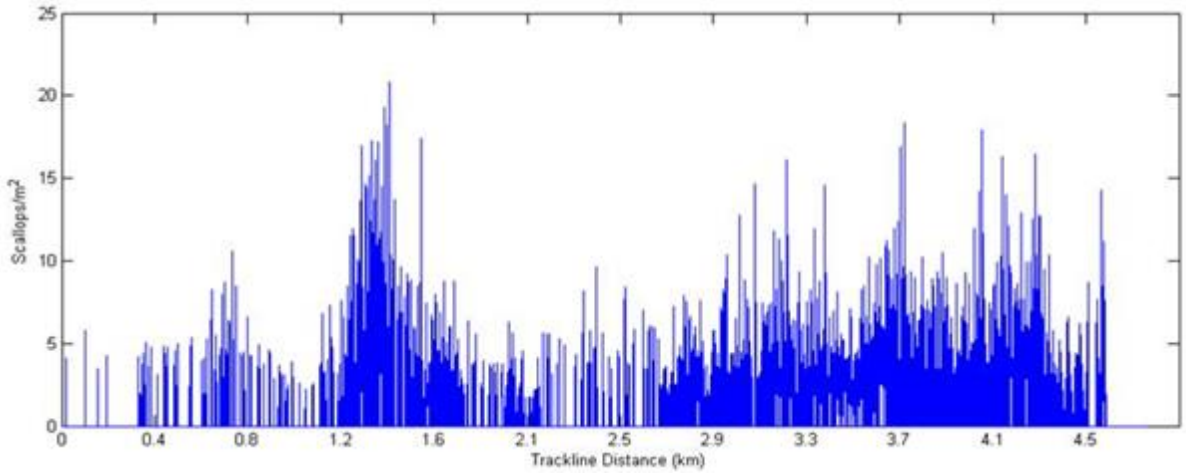


Figure 3.1. Scallop abundance in the northern section of the NLSCA survey along a 4.5 km (2 nm) track. Every other image was classified manually to establish a baseline for downsampling and patchiness. Note the very patchy distribution ranging from 0 to >20 scallops per m^2.

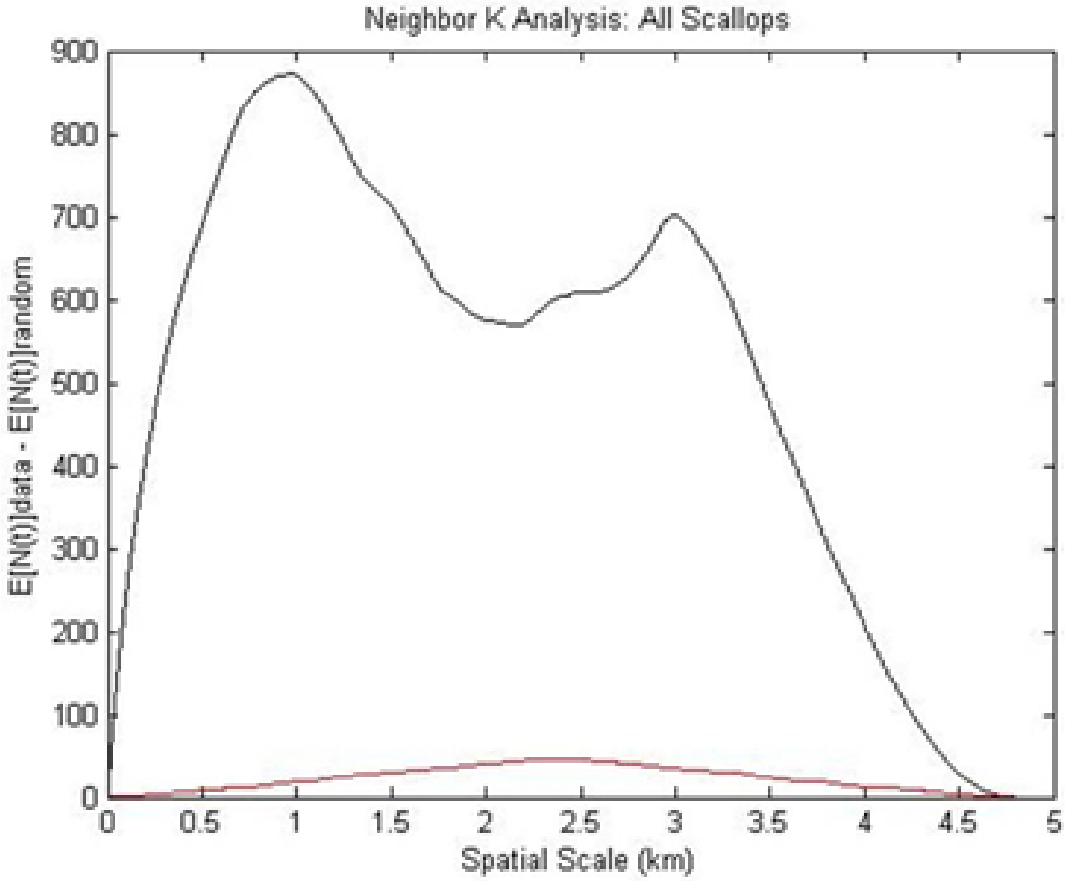


Figure 3.2. Ripley's Neighbor-k analysis of 1 dimensional patchiness of scallop distribution along the track sampled by every other image. Spatial scale is on the x axis while the residual between observed and predicted nearest neighbor distance under randomness (1000 Monte Carlo simulations) is on the y axis. The first mode indicating a characteristic patch size is located at 700-900 m, and a second at 3.2 km. Both modes are well above the red line under randomness indicating these patch dimensions to be statistically significant at the 95% confidence level.

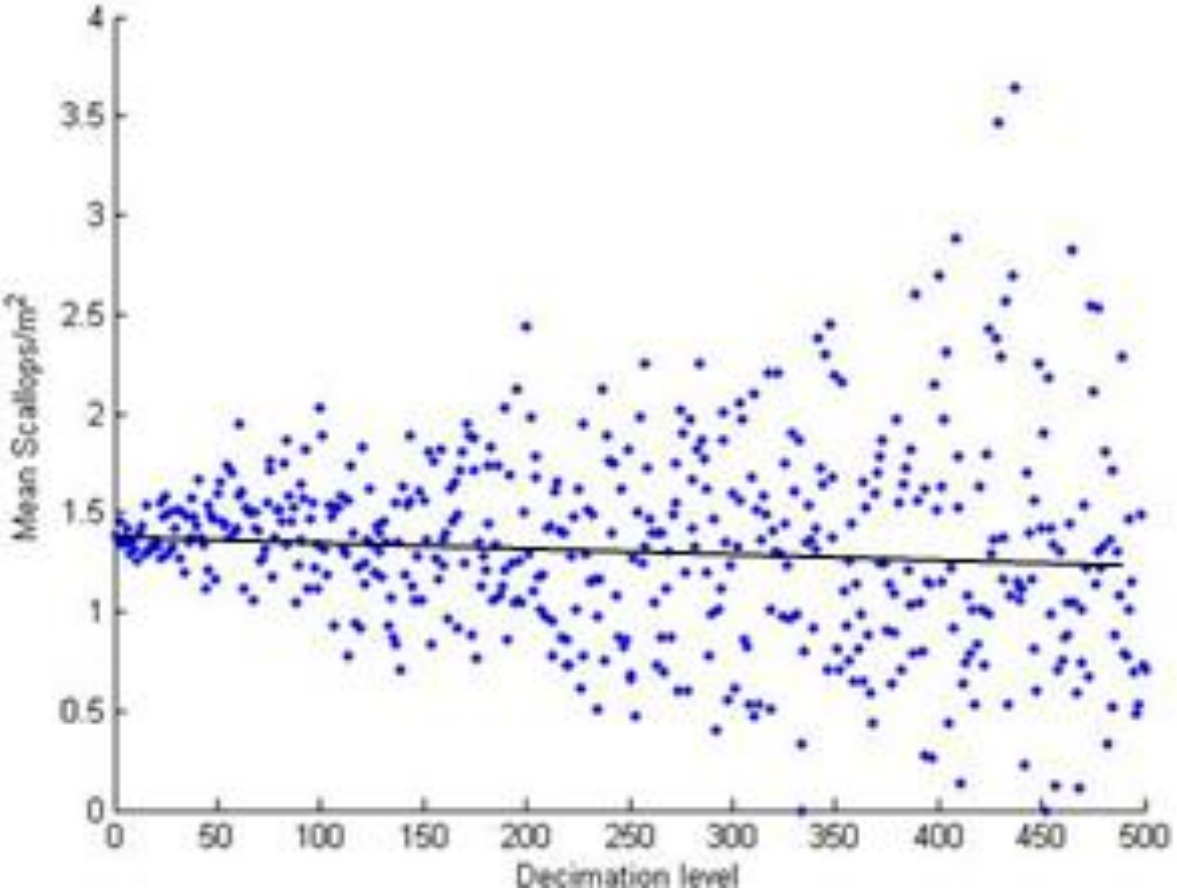


Figure 3.3. The effect of downsampling from the continuous record of scallop abundance along the trackline in Figure 3.2 Images were downsampled at every 4th, 8th, 10th, 12th… out to 500 and the scallop abundance recalculated. Cv of less than 10% is stable out to a downsample level of 100.

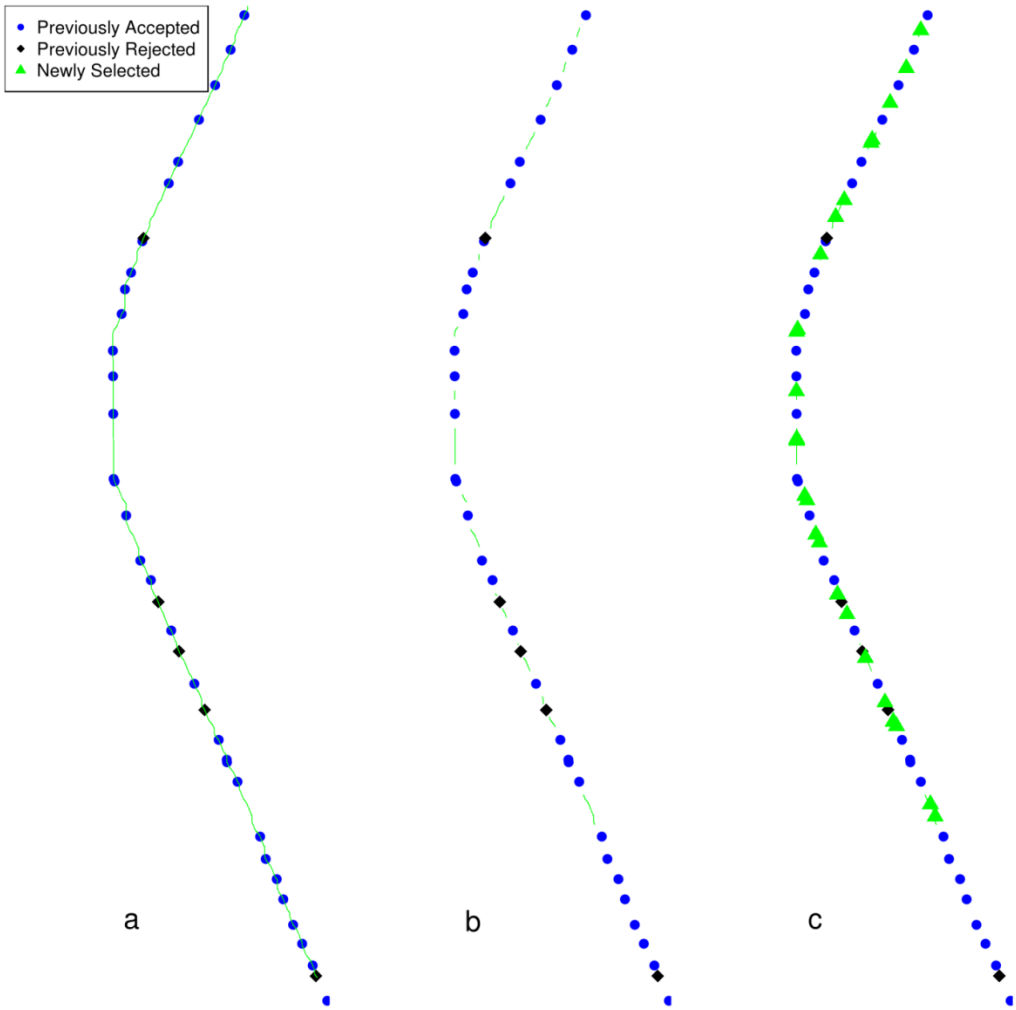
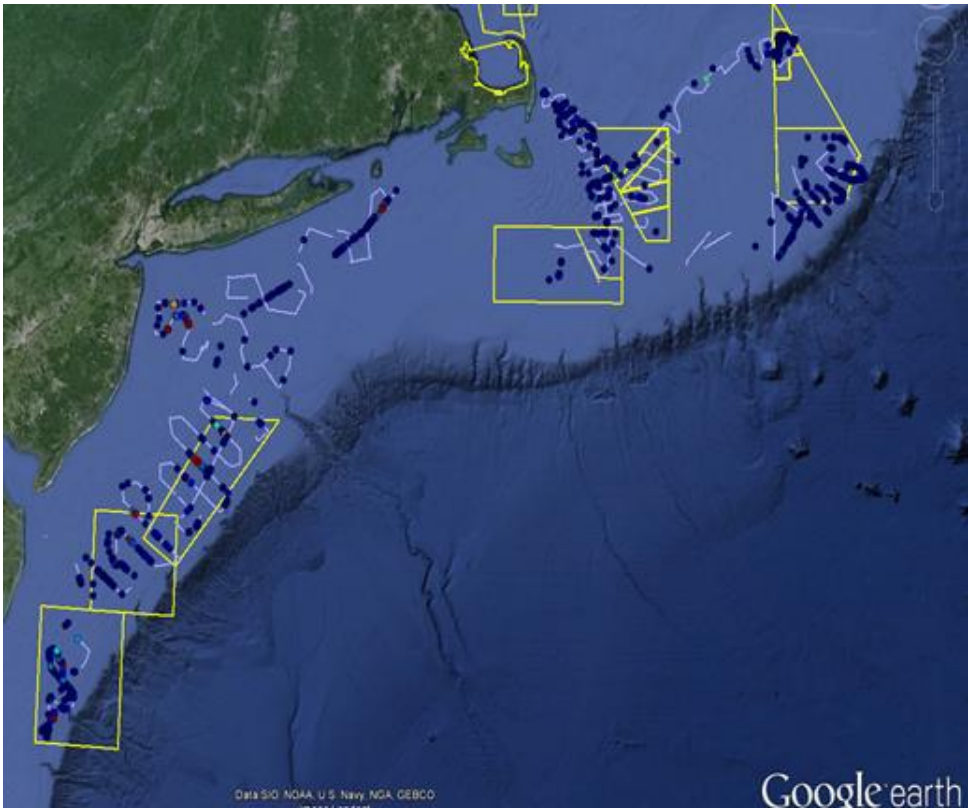


Figure 3.4. Example image selection and gap filling through iterative assignments; (a) previously annotated images (symbols) and unassigned images (green line), (b) previously annotated images with unassigned images with buffering, (c) previously and newly assigned images.

Figure 3.5. Example two-year-old recruits in HabCam imagery: *Above*: southern Georges Bank, 2014. *Below*: Hudson Canyon Rotational Area, 2012.



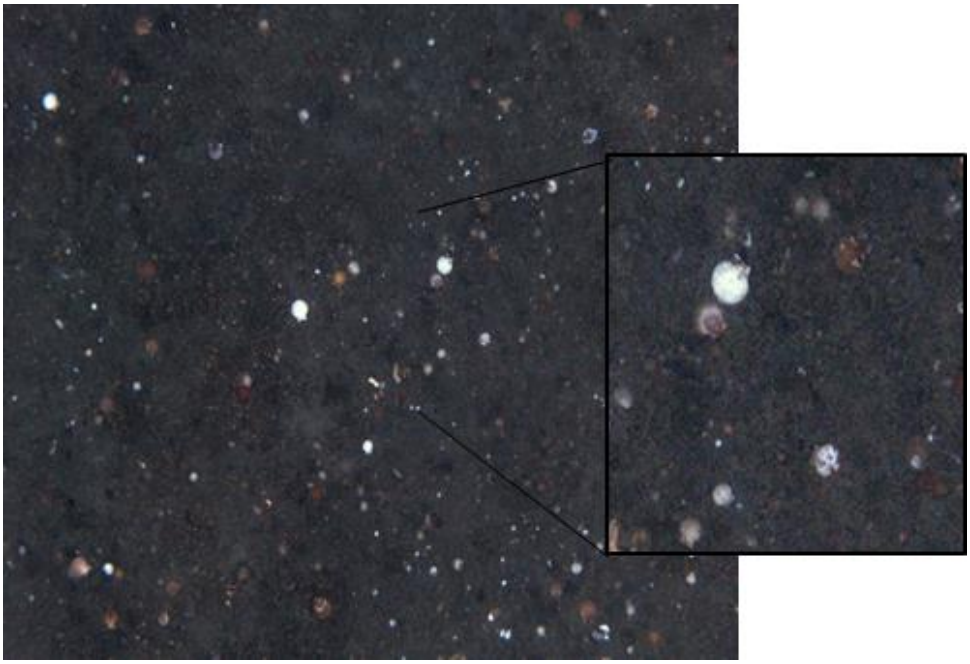


Figure 3.6. Recruitment events detected with HabCamV4 for 2012. The map (top) shows colored dots representing the number of scallops measured manually <= 30 mm per minute of acquired images. The red arrow indicates the location of the image shown (bottom). The purple line indicates the entire sampled HabCam Track V4.

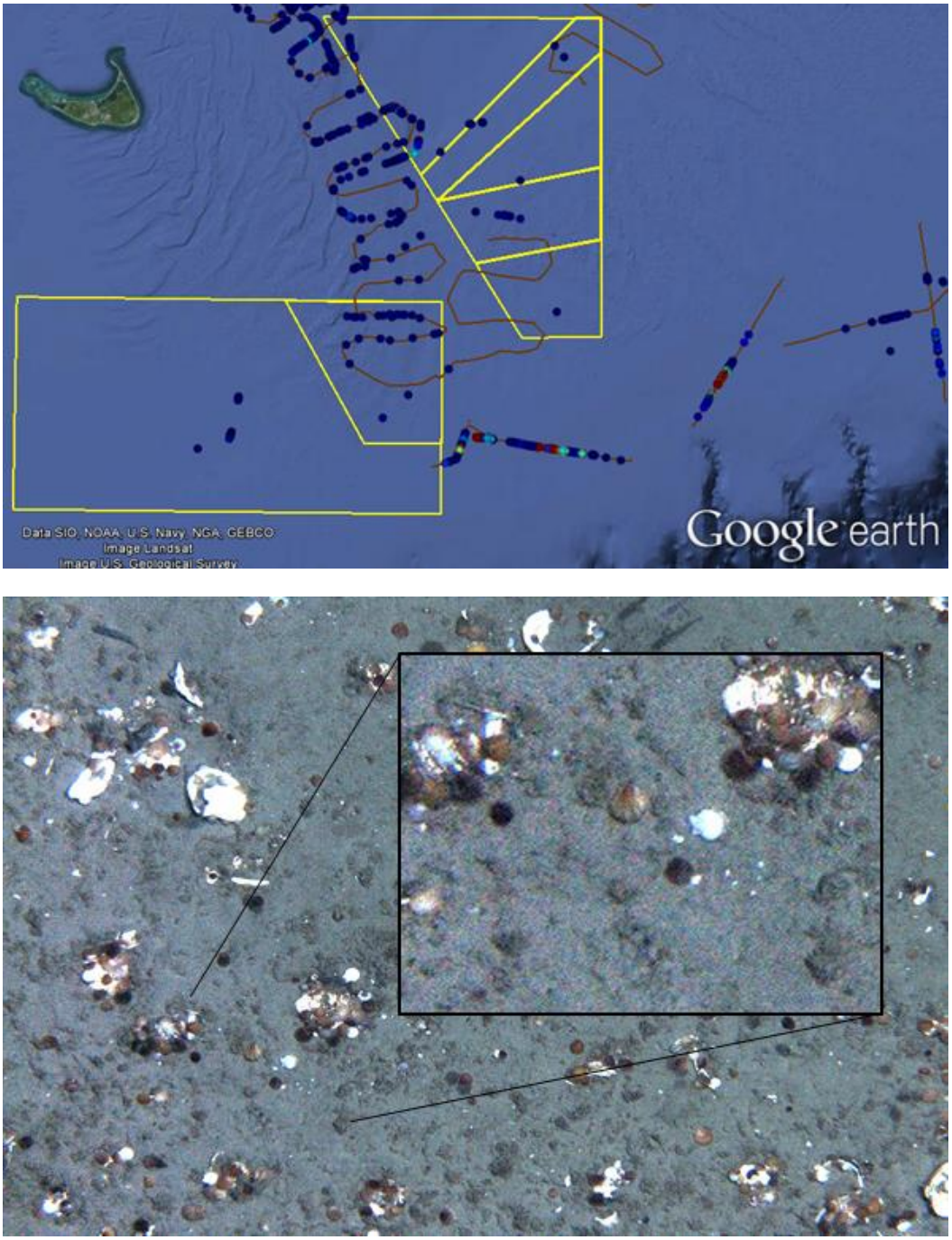


Figure 3.7. Recruitment event detected with HabCamV4 to the east of Nantucket Lightship in 2013. The map shows colored dots representing the number of scallops measured manually <= 30 mm per minute of acquired images. The red arrow indicates the location of the image shown. The orange line indicates the entire sampled HabCam Track V4. The image contains scallops <= 30 mm which were recorded by an annotator using a point count.

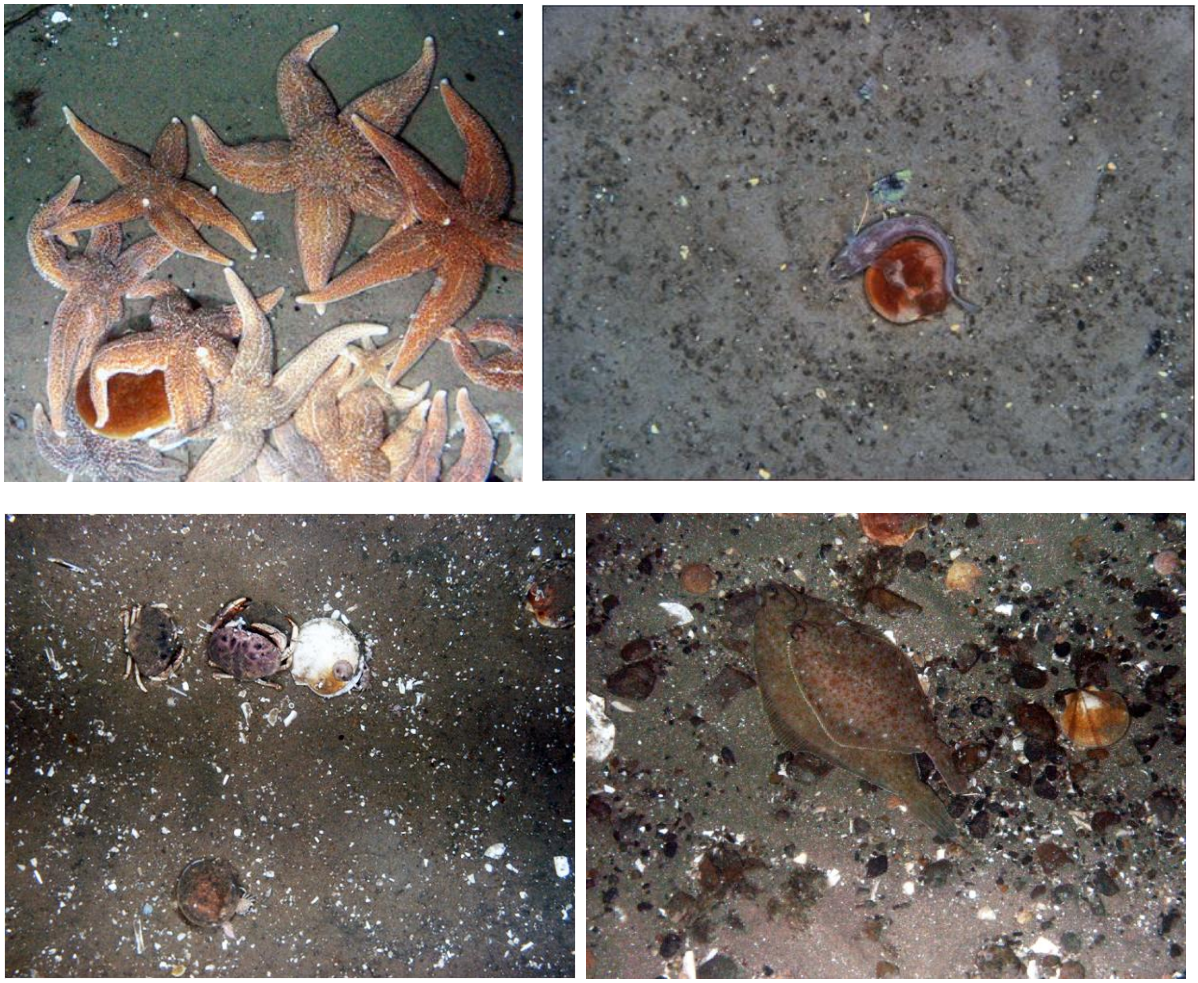


Figure 3.8. Examples of observations of fine-scale ecological interactions from HabCam images. *Above left*: Predation on a sea scallop by *Asterias vulgaris* sea stars. *Above right*: Mutualistic interaction between a sea scallop and a red hake (*Urophycis chuss*). *Below left*: Predation on a sea scallop by *Cancer* crabs. Below right: Possible yellowtail flounder (*Limanda ferruginea*) mating interaction.

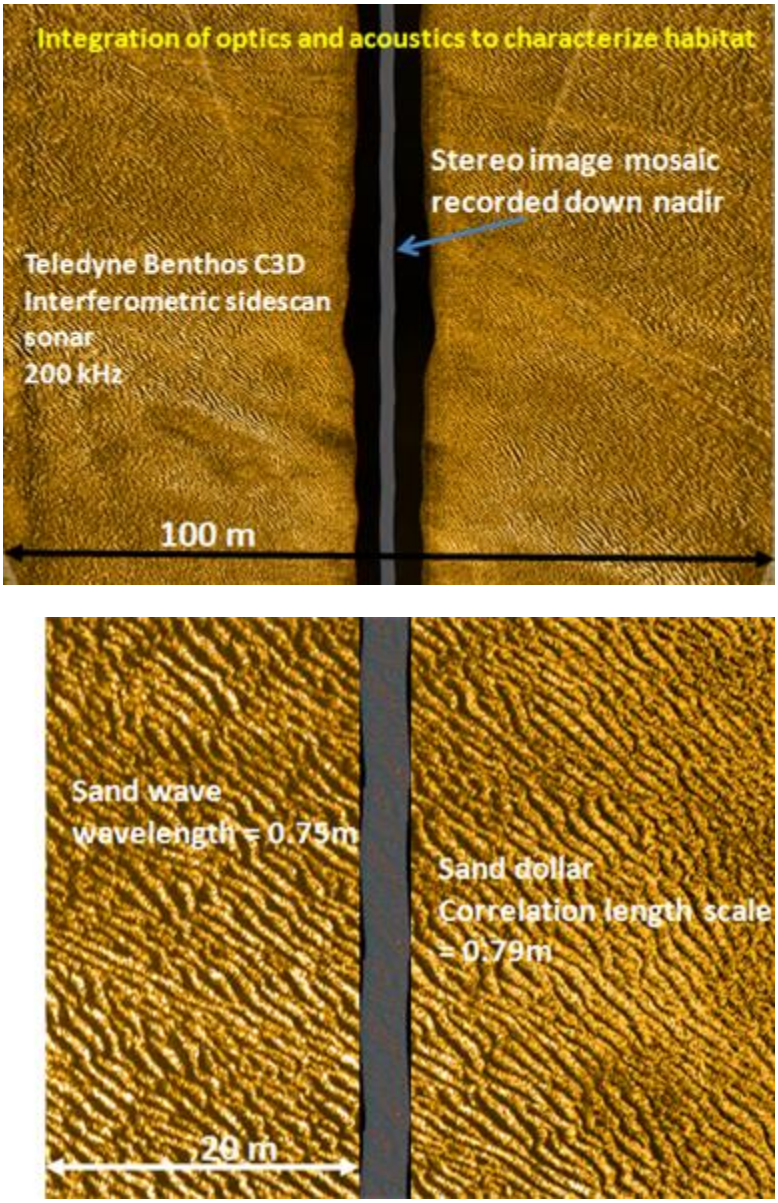


Figure 3.9. Acoustic image from the C3D on HabCamV4 full 100m swath (top) and an enlargement with 20m to the side of the vehicle (bottom).

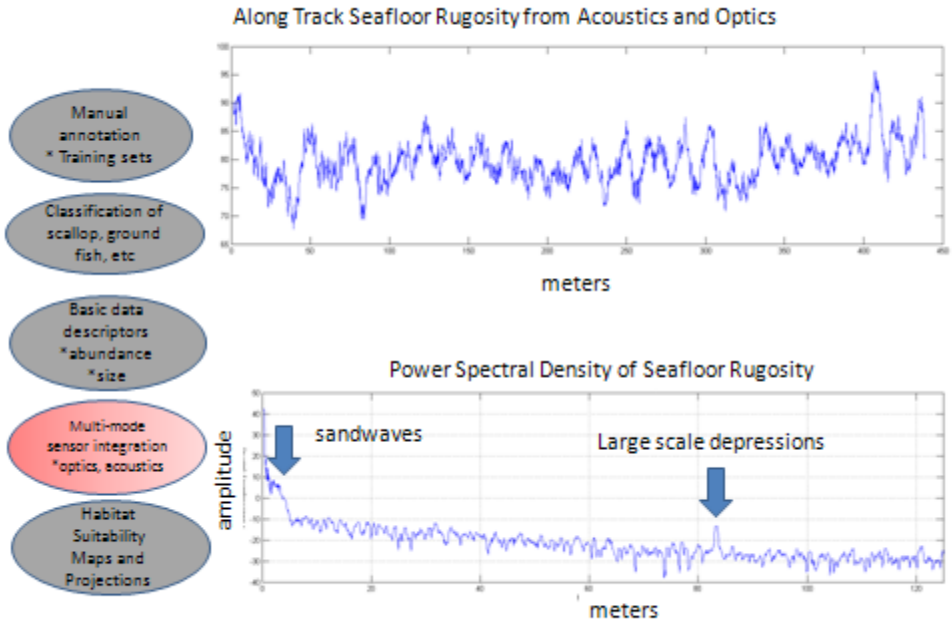


Figure 3.10. Along track rugosity collected from the stereo images (top) and the power spectral density of that signal (bottom)

TOR4 - Methods for using survey data to estimate abundance indices.

Basic modeling approaches

Population counts from HabCam data (or other continuously collected instruments such as acoustic data) are most naturally analyzed using model-based estimation, and most particularly geostastistical methods, because the data are not collected in random locations, and nearby observations tend to be highly autocorrelated. There are large-scale, often non-linear, spatial trends in sea scallop abundances. For example, scallops tend to be more abundant at intermediate depths, especially in the Mid-Atlantic. Generalized additive models (GAMs) can be used to model such trends.

Since nearby observations from HabCam are highly autocorrelated, applying a GAM model to raw HabCam counts would be inappropriate as nearby data are essentially pseudoreplicates. For that reason, neighboring data needs to either by aggregated, or fine-scale variation should be treated as random effects. HabCam data is zero inflated, even after aggregation. We accommodated this by using a "hurdle" model that model presence/absence, and biomass (or abundance) given that at least one scallop is present, separately.

While GAMs can capture large-scale trends, scallop populations are patchy at small scales, which may not be effectively modeled by the GAMs. For that reason, we combined the GAM with ordinary kriging (OK) on the GAM residuals in order to take into account small-scale autocorrelations.

Besides these model-based methods, it is useful to have a simpler population estimator. The main HabCam transects are generally parallel to the gradient of the population, which for scallops is typically parallel to the depth gradient. The edges of populations are surveyed less densely that the central areas, where the population is most concentrated (Fig 4.1). We divided the surveyed area into three strata: the central high-density area, and the marginal habitats on each edge. Data along a transect in each of these parts was averaged (for estimation of density) or summed (for estimation of biomass or abundance), and taken as a single observation, and then these aggregated data points were combined as a stratified mean. We will refer to this as the "stratified mean" method.

Simulations

As can be seen from the above discussion, there are several putative approaches to estimating abundance and biomass from HabCam data, and it is unclear which is superior. It is also unclear at what scale the data should be aggregated (for the GAM model) or be blocked as a random effect (for the mixed-effects GAM). For these reasons, simulations were conducted to evaluate the design-based method (stratified mean method, SM) and three model-based estimation methods to estimate abundance and biomass for this type of data: (1) ordinary kriging on spatially aggregated data (OK), (2) zero-inflated Generalized Additive Models on spatially aggregated data with kriged model residuals (GAM+OK), and (3) zero-inflated Generalized Additive Mixed Models where small scale variations are treated as random effects, combined with kriged model residuals (GAMM+OK). In addition to simulations, the three model-based methods were evaluated using 2013 HabCam biomass data. Co-located survey data from other gear types (dredge surveys and other image-based surveys) were used to validate model performance. The method that produce the least bias and most precise estimates was selected as the standard method to estimate abundance and biomass from 2011-2013 Habcam data (see Chang et al. for more details).

Scallop spatial distributions are non-stationary due to the influences of physical and biological environment such as substrate, depth, temperature, and predator distributions (Brand 1991; Hart 2006). The simulated scallop populations are therefore assumed to vary non-randomly according to large-scale trends (termed the “first-order effect”). For simplicity in the simulations, these trends are assumed to be possibly non-linear functions of latitude and longitude only. In real situations, depth or other factors may be important predictors of trend; longitude is treated as a surrogate for depth in the simulations. Stationary “second-order effects”, representing small-scale spatially autocorrelated variability, were added to the first order trends. Combinations of two first-order and four second-order effects were simulated, resulting in eight types of simulated population distributions (Fig 4.2). Thirty realizations were generated for each population type. Each realization was surveyed using 30 different tracks (shape and direction are the same but starting point and first turn of the track are different). Shape and direction of the simulated tracks was designed to mimic the actual HabCam survey design, with the long transects parallel to the expected scallop density gradient (Figs 4.1, 4.2). The length of these long

transects are alternated - one long transect extend to the boundary of the survey area, followed by a short transect extend to the boundary of the middle high density patches (Fig 4.1; NEFSC 2014).

The model-based and designed-based estimation methods were evaluated using percent bias and percent root mean square error (RMSE)

$$\% \text{ Bias} = \frac{\frac{\sum_{i=1}^{n}(\hat{T}_i - \mu)}{n}}{\mu} \quad (1)$$

$$\% \text{ RMSE} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(\hat{T}_i - \mu)^2}{n}}}{\mu}, \quad (2)$$

where $\hat{T}_i$ is the estimated total biomass or abundance for sample sets i, μ is the true population abundance or biomass, and n is the total number of sample sets analyzed. Percent bias of estimated CVs of the abundance or biomass estimates were also evaluated. The method that produced the least bias and most precise estimates were used to analyze the actual HabCam data.

Model-Based Estimation

OK is a standard version of the kriging models with the assumption of a stationary mean and consideration of variation and distance between sample points regarding the attributes (Webster and Oliver 2001; Hengl 2009). The total abundance or biomass (T) and its variance were estimated as

$$\hat{T} = \mathrm{A}\sum_{i=1}^{n} \hat{z}_i \ (3)$$

$$\mathrm{Var}(\hat{T}) = \mathrm{A}^2 \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(\hat{z}_i, \hat{z}_j), \ (4)$$

where $\hat{z}_i$ is the kriging estimates at location i and A is the grid size.

Regression kriging (RK) extends the OK to account for a global trend, which can be estimated using a generalized regression model (e.g., GLM or GAM) with a series of ancillary variables, and then OK is performed on the residuals of the regression model to model the second order effects (Odeh et al. 1995; Hengl 2009). Two-staged "hurdle" model were used to estimate the first-order effects, where presence/absence data are modeled separately from the biomass or abundance in non-zero segments and then combined to derive the final estimates (Barry and Welsh 2002; Smith et al. 2012, Zuur et al. 2012). The random variable of the first stage of the hurdle model is given by the distribution:

$$y = \begin{cases} 0 & \text{with probability } p \\ \varphi(\theta) & \text{with probability } 1 - p, \end{cases}$$

where p is the probability of zeros and θ is a vector of parameters to for the statistical distribution φ of positives (abundance or biomass; Smith et al. 2012). The first-order effects were estimated using a two-dimensional spline function of latitude and longitude in both the GAM and GAMM models. The spatial residuals obtained from the large-scale model were used to estimate fine-scale spatial patterns using OK. The total abundance and biomass estimated from GAM+OK and GAMM+OK model is

$$\hat{T} = \mathrm{A} \sum_{i=1}^{n} \hat{x}_i \hat{y}_i + \hat{z}_i, \qquad (5)$$

where $\hat{x}_i$ is the probability of presense estimate, $\hat{y}_i$ is the estimate of abundance or biomass given that scallops are present, and $\hat{z}_i$ is the kriged residual at location i. By assuming that $\hat{x}$ and $\hat{y}$ are independent, the variance of $\hat{T}$ can be calculated as

$$\mathrm{Var}(\hat{T}) = \mathrm{A}^2 \left(\sum_{i=1}^{n} \mathrm{E}^2(\hat{x}_i)\mathrm{Var}(\hat{y}_i) + \mathrm{E}^2(\hat{y}_i)\mathrm{Var}(\hat{x}_i) + \mathrm{Var}(\hat{x}_i)\mathrm{Var}(\hat{y}_i) + \sum_{i=1}^{n} \sum_{j=1}^{n} \mathrm{Cov}(\hat{z}_i, \hat{z}_j) \right) \qquad (6)$$

Aggregation of data for model-based estimation

The HabCam data (both simulated and real) used to build the models were grouped by segments of a given length along the tracks. The OK and hurdle GAMs were built using data averaged within segments along the tracks and the hurdle GAMMs using raw data but the fine-scale variations within track segments were treated as random effects. GAMs are based on the assumption that data points are independent; aggregating nearby data reduces the autocorrelation among the resulting aggregated data, so that they are approximately independent. This is a common technique used when analyzing fisheries acoustic data (Mello and Rose 2005). The length of the segment should be sufficient to reduce the degree of random variability and spatial autocorrelation of the data, while at the same time preserve spatial structures (Mello and Rose 2005). There is no prior knowledge on what the segment length should be used for Atlantic sea scallop and also how sensitive the segment length is for this type of analysis. Therefore, the effects of segment length to average the data or determine random effects along the tracks were evaluated. Scallop aggregations tend to occur at scales of around 1 km (NESFC 2010), so three segment lengths, 0.75, 1.5, and 2.25 km were tested. The segment lengths used in the analysis is equivalent to the grid size A, which is the grid size for interpolation.

Design-Based Estimation

For the SM method, only horizontal transects were used in the simulations. The horizontal transects were post-stratified into three strata based on high and low first-order effects (Fig 4.3). The mean and its variance of the simulated scallops (t) by segment (j) and stratum (i) were calculated as

$$\bar{t}_{i,j} = \frac{\sum_{k=1}^{n_{i,j}} t_{i,j,k}}{n_{i,j}} \quad (7)$$

$$\mathrm{Var}(\bar{t}_{i,j}) = \frac{\mathrm{Var}(t_{i,j,k})}{n_{i,j}}, \quad (8)$$

where $n_{i,j}$ is the number of images by segment and stratum. The total abundance and biomass estimates ($\hat{T}$) and its variance were estimated as (Cochran 1977)

$$\hat{T} = \mathrm{A}\sum_{i=1}^{2} \mathrm{S}_i \frac{\sum_{j=1}^{n_i} \bar{t}_{i,j}}{n_i} \quad (9)$$

$$\mathrm{Var}(\hat{T}) = \mathrm{A}^2 \sum_{i=1}^{2} \mathrm{S}_i^2 \sum_{j=1}^{n_i} \frac{\mathrm{Var}(\bar{t}_{i,j})}{n_i^2}, \quad (10)$$

where n_i is the number of segments by stratum i, and S_i is the area of stratum i.

The simulation domain could be well-stratified based on the first-order trend. However, the same precise information may not be available when dealing with the real data. The sensitivity of the SM estimates to the stratification were evaluated by enlarging (Stratified Mean Wide, SMW) and shrinking (Stratified Mean Narrow, SMN) the central high abundance or biomass stratum by 20% (Fig 4.3) and then estimating the SMs based on these less perfect stratifications.

Analysis of realized HabCam data

The GB and MAB stock region were divided into fourteen subregions based on geographic characteristics and management areas and analyzed separately because their topology, orientation, and covariance structures differ (Fig 4.4). Images taken at altitudes higher than 4 m were excluded from the analysis because of poor image quality. To match the stock assessment models, only scallops with measured shell height larger than 40 mm were used in the analysis. A summary of the HabCam data used by year from 2011 to 2013 is listed in Table 1.

Model Testing Results using Simulations

The simulation results showed no difference in preferred model types and segment length by types of simulated populations and survey tracks so the results presented here were summarized by model type and segment length (Table 4.2). Among the three model-based methods, the GAM+OK method produced the least bias and most precise abundance and biomass estimates, followed by GAMM+OK and then OK (Table 4.2). GAM+OK also produced the least biased CV estimates, followed by OK and then GAMM+OK (Table 2). The percent bias and % RMSE of the abundance and biomass estimates tend to increase with increasing segment

length (Table 4.2). The increase is most severe for the GAMs and least for the GAMMs (Table 2). These results thus indicate that GAM+OK with data averaged by 750 m (segment length) is the best method to estimate the scallop abundance and biomass among all the model-based methods. GAM+OK with 750 m aggregation length also produced the least biased CV estimates for both biomass and abundance estimates (Table 4.2).

Percent bias and % RMSE of the SM estimates were smaller than all the model-based estimates, but the CVs were substantially underestimated (Table 4.2). In addition, SM estimates were sensitive to the quality of post stratification. Improper stratification (SMW and SMN, see Fig 4.3) induced biased estimates that were worse than most of the model-based estimates (Table 4.2).

Model results on actual HabCam biomass data

Both model fittings and validations showed that no single modeling approach was consistently superior but GAM+OK performed the best in most cases, followed by GAMM+OK and then OK (Table 4.3). The models performed slightly better when segment length was smaller (Table 4.3).

Based on the simulations and the 2013 field data analysis results, the GAM+OK method, aggregated over 750 m segments, was used to estimate total abundance and biomass for each subregion in GB and MAB for 2011 to 2013. The SM estimates with careful stratifications were also provided in order to validate the model-based estimates, although the CV of SM estimates are probably understated. The average of weight or count (*t*) by image (*j*) and segment (*i*) weighed by field of view (*f*) for every 750 m along the tracks were calculated as

$$\bar{t}_i = \sum_{j=1}^{n_i} \frac{f_{i,j} t_{i,j}}{\sum_{j=1}^{n_i} f_{i,j}}, \qquad (11)$$

The $\bar{t}_i$ was weighted by both variation (s) and number of images (n) in the hurdle GAM using

$$w_i = \frac{s_i - s_{(1)}}{2(s_{(n_i)} - s_{(1)})} + \frac{n_i - n_{(1)}}{2(n_{(n_i)} - n_{(1)})} \qquad (12)$$

A hurdle GAM with a quasi-binomial distribution for the presence/absence model and quasi-Poisson distribution for the positive model were used to estimate the first-order trend with respect to latitude, longitude, and depth. Depth is correlated with latitude and/or longitude in some of the subregions. To prevent potential problems cause by collinearity, latitude and longitude were transformed into composite variables: latitude plus longitude, half of the latitude or longitude plus longitude/latitude. Twelve models with different combination of these covariates were evaluated using 10-fold cross validation and OK was performed on the residuals of the selected GAM. Depth is included in all of the candidate models because it is one of the most important variables affecting scallop distributions.

For the SM analysis, the transects were post-stratified and separated into segments based on depth and where the survey track had higher sampling. The mean count or weight and its variance by segment and stratum were calculated using equations 10 and 11. These mean and variances were weighted by total field of view (f) and length of the segment (d) to estimate the total abundance or biomass and its variance

$$\hat{T} = \mathrm{A} \sum_{i=1}^{3} \mathrm{S}_i \sum_{j=1}^{n_i} \mathrm{w}_{i,j} \bar{t}_{i,j} \quad (13)$$

$$\mathrm{Var}(\hat{T}) = \mathrm{A}^2 \sum_{i=1}^{3} \mathrm{S}_i^2 \sum_{j=1}^{n_i} \mathrm{w}_{i,j}^2 \mathrm{Var}(\bar{t}_{i,j}), \quad (14)$$

where n_i is number of segments within depth stratum i, S_i is the size of depth stratum i, and $w_{i,j}$ is the weighting factor

$$w_{i,j} = \frac{d_{i,j} - d_{i,(1)}}{2(d_{i,(n_i)}) - d_{i,(1)})} + \frac{f_{i,j} - f_{i,(1)}}{2(f_{i,(n_i)} - f_{i,(1)})} \quad (15)$$

GAM+OK and SM abundance and biomass estimates and their CVs by year are in Table 4.4, and an example of the interpolation surface for 2013 abundance of both regions is shown in Fig 4.4. GAM+OK estimates have high agreement with SM estimates. The results of simple

linear regressions of GAM+OK estimates against SM estimates for all years and subregions showed that the intercepts are not significant ($p>0.05$) and the slope is 1.00 for the abundance estimates and 0.91 for the biomass estimates. However, the CVs of the SM estimates are on average lower than the CVs of GAM+OK estimates. Mean of the CV estimates for all subregions is 0.1 for SM estimates and 0.38 for GAM+OK estimates.

References

Barry, S.C., and Welsh, A.H. 2002. Generalized additive modeling and zero inflated count data. Ecol. Model. 157(2): 179-188.

Brand, A.R. 1991. Scallop ecology: distributions and behavior. In Scallops: biology, ecology and aquaculture. Developments in aquaculture and fisheries science. Vol. 21. Edited by S.E. Shumway. Elsevier, Amsterdam. pp. 517-584.

Chang, J-H., Hart, D.R., Shank, B.V. 2015. A comparison of model-based and design-based methods to estimate Sea Scallop abundance and biomass from vessel-towed underwater camera data. Background document for this review.

Cochran, W.G. 1977. Sampling Techniques. John Wiley and Sons, New York.

Hart, D.R. 2006. Effects of sea stars and crabs on sea scallop (*Placopecten magellanicus*) recruitment in the Mid-Atlantic Bight (USA). Mar. Ecol. Prog. Ser. 306:209-221.

Hengl, T. 2009. A practical guide to geostatistical mapping. University of Amsterdam, Amsterdam, pp. 417-427.

Mello, L.G., and Rose, G.A. 2005. Using geostatistics to quantify seasonal distribution and aggregation patterns of fishes: an example of Atlantic cod (*Gadus morhua*). Can. J. Fish. Aquat.

Sci. 62(3): 659-670.

Northeast Fisheries Science Center. 2010. 50th Northeast Regional Stock Assessment Workshop: Assessment Report. Northeast Fisheries Science Center Reference Document 10-17.

Odeh, I.O.A., McBratney, A.B., and Chittleborough, D.J. 1995. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma 67(3): 215-226.

Smith, A.N., Anderson, M.J., and Millar, R.B. 2012. Incorporating the intraspecific occupancy-abundance relationship into zero-inflated models. Ecology 93(12): 2526-2532.

Webster, R., and Oliver, M.A. 2001. Geostatistics for Environmental Scientist. John Wiley and Sons, Chichester, England.

Zuur, A. F., Saveliev, A. A., and Ieno, E. N. 2012. Zero inflated models and generalized linear mixed models with R. Highland Statistics Limited, Newburgh.

Table 4.1 Summary of HabCam data for sea scallops. “Annotated Images” is the number of images analyzed. “Images w/scallops” is the number of analyzed images in which scallops were observed.

Stock	Year	Annotated Images	Images w/scallops
GB	2011	202,257	21,428
GB	2012	36,304	7,189
GB	2013	33,864	4,671
MAB	2012	20,969	2,095
MAB	2013	42,213	3,627

Table 4.2. Summary of % bias and % RMSE for mean and CV, nugget-to-sill ratio, and number of converged sample runs for biomass and abundance estimates by segment length and estimation methods.

		Biomass					Abundance				
Model Type	Segment Length	% Bias Mean	% RMSE Mean	% Bias CV	N-S Ratio	# Runs	% Bias Mean	% RMSE Mean	% Bias CV	N-S Ratio	# Runs
GAM	750	0.04	0.19	-0.12	0.28	7193	0.03	0.19	-0.11	0.22	7196
GAMM	750	0.08	0.20	0.21	0.28	5180	0.07	0.19	0.38	0.23	4209
OK	750	0.14	0.25	0.60	0.18	7195	0.14	0.24	0.58	0.12	7198
GAM	1500	0.05	0.21	0.12	0.25	7185	0.04	0.20	0.15	0.19	7184
GAMM	1500	0.08	0.20	0.95	0.27	4631	0.08	0.20	1.32	0.22	3753
OK	1500	0.15	0.26	0.49	0.12	7162	0.15	0.25	0.45	0.09	7159
GAM	2250	0.06	0.22	0.46	0.21	7182	0.06	0.21	0.88	0.17	7176
GAMM	2250	0.08	0.21	0.67	0.26	5066	0.08	0.20	1.39	0.20	4433
OK	2250	0.16	0.28	0.29	0.10	7173	0.16	0.27	0.27	0.07	7159
SM	2250	0.00	0.18	-0.61		7200	0.00	0.17	-0.66		7200
SMN	2250	0.23	0.35	-0.67		7195	0.22	0.34	-0.72		7194
SMW	2250	0.14	0.25	-0.60		7200	0.14	0.25	-0.66		7200

Table 4.3 The frequency of model types and segment lengths that could best fit the HabCam data and best validate the NEFSC dredge, SMAST, and VIMS dredge or drop camera data for the fourteen subregions in GB and MAB for 2013.

Segment Length and Model Type	HabCam	NEFSC (dredge)	SMAST (drop camera)	VIMS (dredge)	Total
500	2	1		5	8
750	3	1		3	7
1000	2	3	2	1	8
1250		5			5
1500	2	3	2		7
1750	5			1	6
GAM	5	8	2	6	21
GAMM	7	5	1	3	16
OK	2		1	1	4

Table 4.4 Summary of abundance and biomass and its CVs estimated using GAM+OK and SM methods for 2011 to 2013.

		Number (million)				Biomass (mt meats)			
Stock	Year	SM	GAM+OK	SM CV	GAM+OK CV	SM	GAM+OK	SM CV	GAM+OK CV
GB	2011	3992	3832	0.02	0.31	110204	102819	0.02	0.12
GB	2012	4003	4642	0.03	0.14	94024	94040	0.03	0.08
GB	2013	3562	4049	0.03	0.09	54683	49671	0.03	0.29
MAB	2012	4166	4902	0.03	0.13	50573	49196	0.04	0.12
MAB	2013	5064	4611	0.05	0.07	62314	61485	0.04	0.13

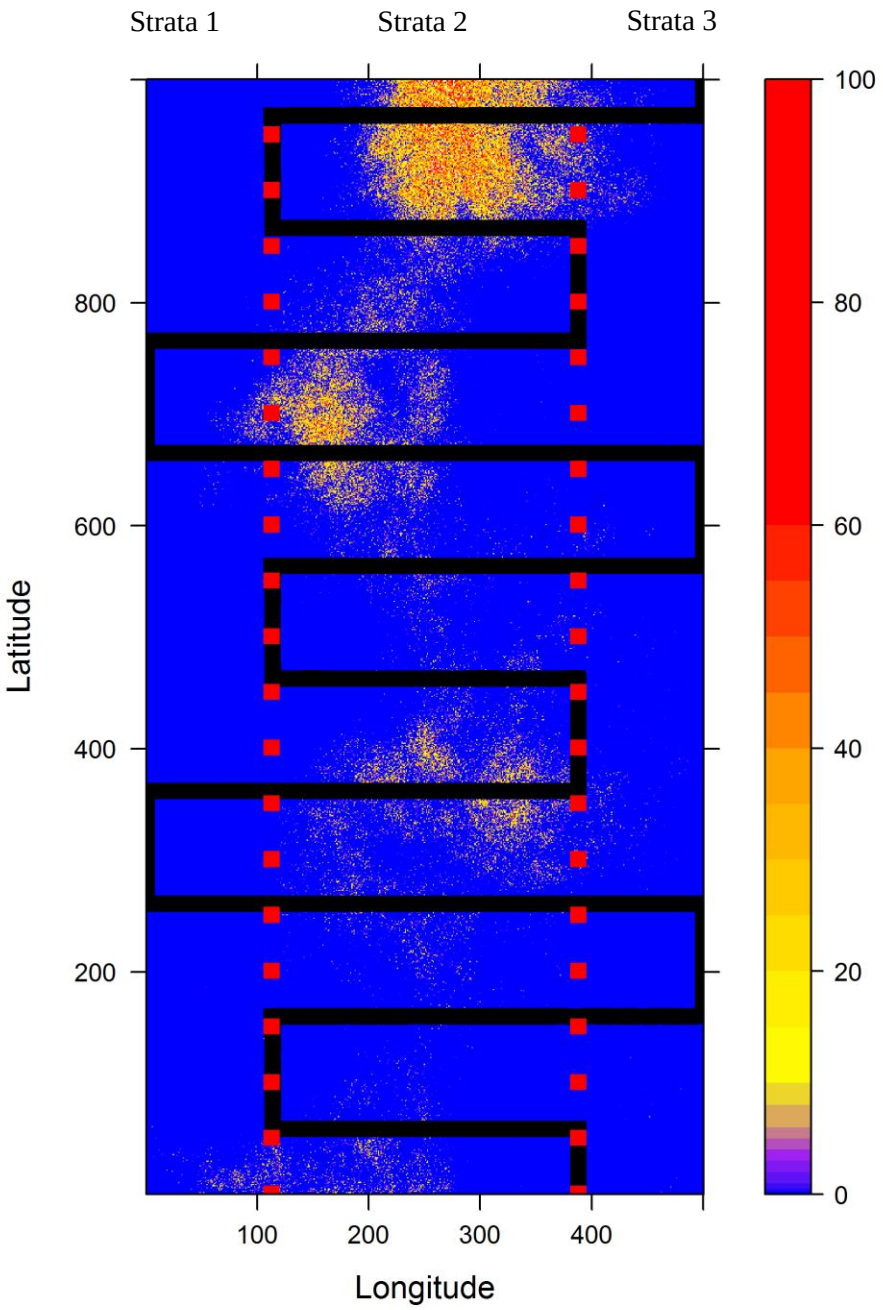


Figure 4.1. Example simulated scallop population (colored dots), HabCam track (black lines), and stratification (red dashed lines). In the stratified mean method, mean density would be estimated by first averaging within each transect and stratum, and then averaging within each stratum. In this example, the density of stratum 2 would be based on the average of 10 data points (each the average within its transect), whereas stratum 1 and 3 would be based on 4 and 5 aggregated data points, respectively.

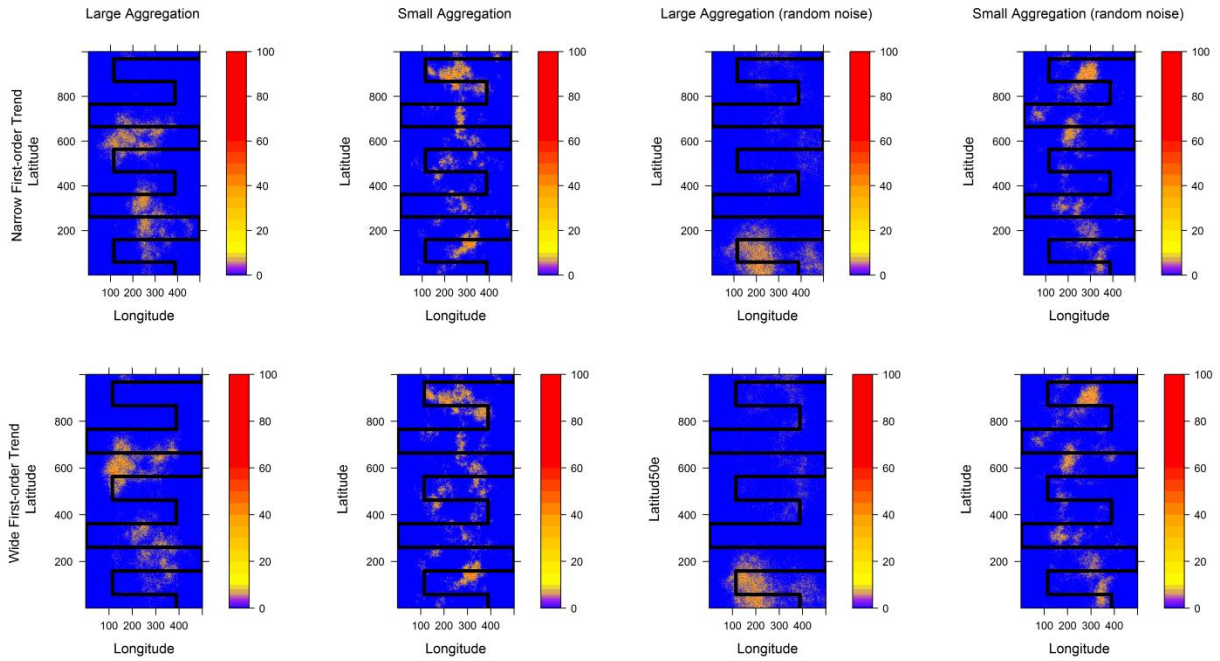


Figure 4.2: Realizations of the 8 types of simulated scallop population distribution with an overlayed sampling track.

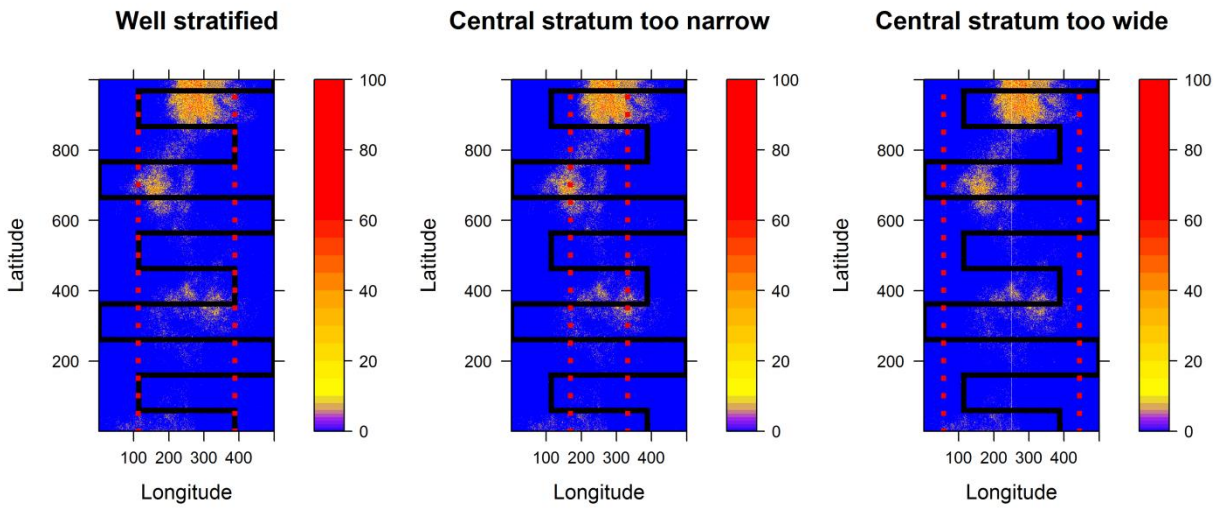


Figure 4.3: Alternative types of stratifications used for SM estimations. On the left, the stratification coincides with the area surveyed more heavily, and thus can be expected to produce unbiased estimates. If the boundaries are incorrect, however, as given in the center and right, SM estimates may be biased because sampling is uneven in at least one stratum. The stratum boundaries are given by red lines, and the colored points indicate scallop density.

(a)

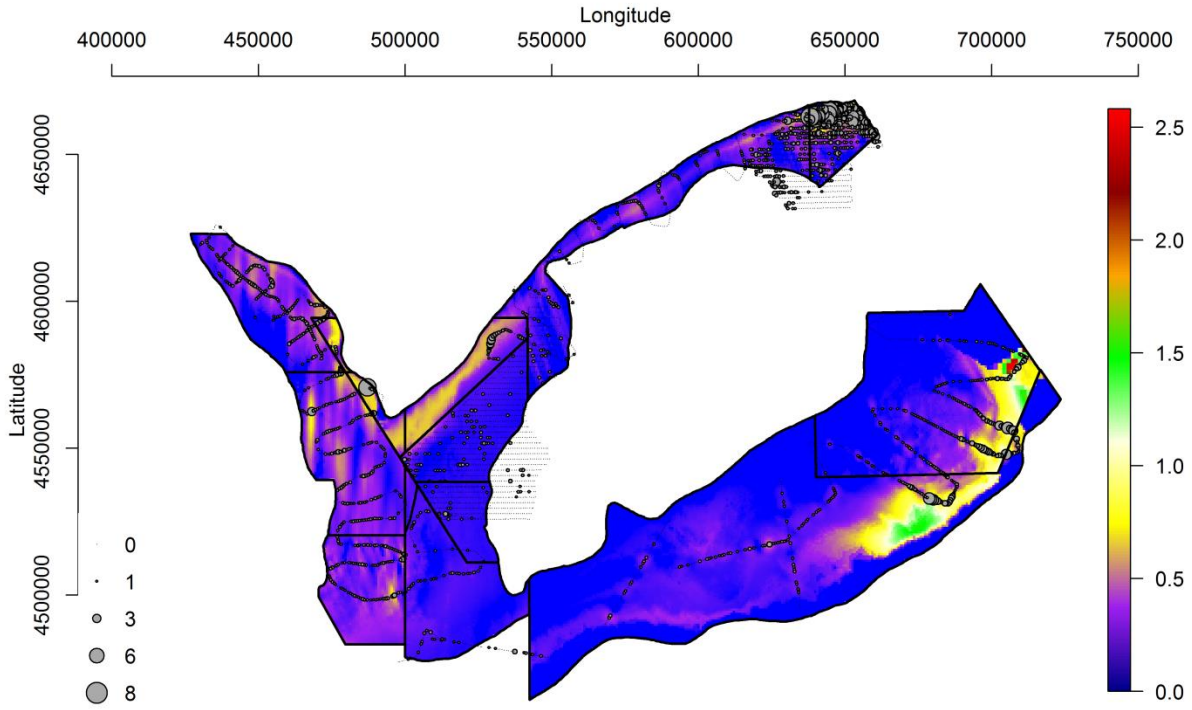


Figure 4.4 – continued next page.

(b)

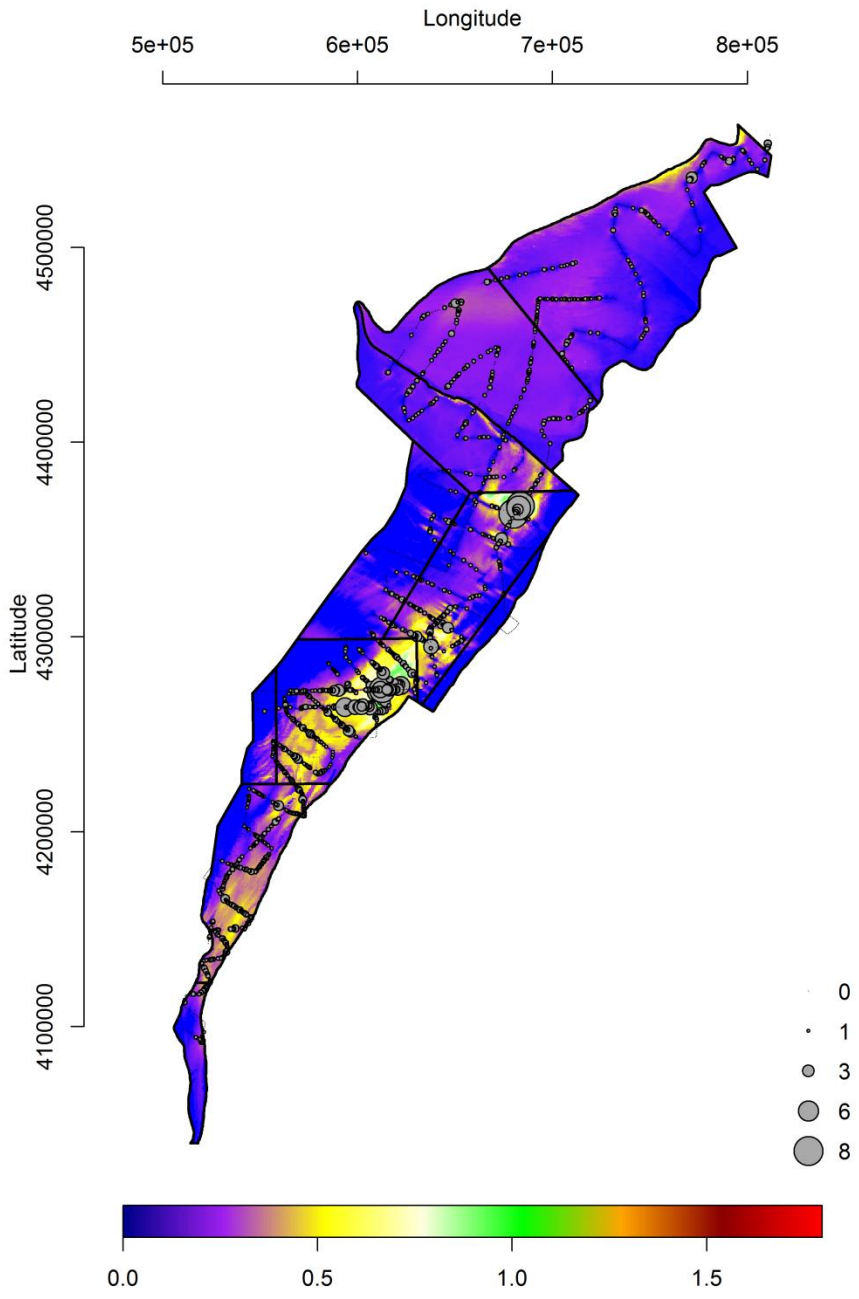


Figure 4.4. Estimated scallop densities (# m^2) on GB (a) and MAB (b) in 2013 based on HabCam data using the GAM+ OK method with data aggregated into 750 m blocks. The survey track (thin black line) and observations of scallops (circles, with size proportional to number observed) are also shown.

Term of reference 5 – CoKriging as a method for combining resource surveys, decreasing uncertainty, and mitigating bias.

Introduction

The sea scallop resource is in the uncommon situation of having multiple co-located surveys with different survey gears (Figure 1). This brings up the issue of how to better utilize and combine information from different surveys, particularly as the coverage or quality of each survey may vary regionally and inter-annually. Further, the resource is spatially managed through rotational closures and therefore, modeled at subregional scales. Thus, while region-scale surveys are often unreliable at subregional scales, there is a management need to have accurate subregional resolution when possible. Two common approaches to combining surveys are to take the arithmetic mean or an inverse-variance weighted mean (IVM) across the surveys (see NEFSC Dredge TOR5, Table 1). The arithmetic mean may not be optimal due to unequal quality (variances) of the different surveys. Inverse variance weighted means gives the minimal variance compared to other weightings, but IVM are not scale invariant (see NEFSC Dredge TOR-5). In particular, if the IVM because the variance of a survey tends to scale with the mean, inverse-variance weighting tends to down weight surveys with high means, especially at small scales. Besides these issues, there may be spatial synergies in more holistic methods of combining surveys. The spatial coverage of the combination of several surveys may be superior to any one of them.

Kriging is a geostatistical method that integrates the data values, the spatial distribution of the samples, and a variogram model to estimate the mean and uncertainty (variance) for any location in the sampling domain (Cressie 1993, Rivoirard *et al*. 2000, Bivand *et al*. 2008). The variogram is a statistical model of the spatial autocorrelation of the data based on how the variance of any two observations changes as the distance between the sampling sites increases. Block kriging, a similar method, allows for the calculation of a mean and variance of the entire sampling domain and is a common model-based methods for estimating resource abundance in fisheries. Cokriging is a similar method that estimates the mean and variance of a spatially-explicit data set, based on the data itself and its spatial cross-correlation to other data sets within the sampling domain, using cross-variograms (Figure 2). Thus, under cokriging, the estimated mean and variance of a survey at any given location is dependent on the neighboring data from the survey itself, as well as neighboring data from other surveys and how well correlated the different surveys are to each other. Therefore, this method has potential for improving one survey for a given region by borrowing information from other co-located surveys, potentially from other gear types. Here we examine the potential for cokriging as a method of combining survey estimates for Georges Bank using the 2012 data from the NEFSC and VIMS RSA dredge surveys, the SMAST drop camera survey, the NEFSC HabCam (V4) survey and the HabCam Group (V2) RSA survey.

Methods

We used the 2012 surveys on Georges Bank to demonstrate the technique. We divided Georges Bank into eight subregions based on ecological and management boundaries. For each subregion, we extracted all data for each survey within the subregion for building variogram models and all data within a 5 km buffered subregion for biomass estimation. For each variogram and cross-variogram, we log-transformed the data, fit the variogram model, and back-transformed the model parameters (Guiblin *et al*. 1995, Rivoirard *et al*. 2000). Fitting four parameters of a variogram model (nugget, sill, range and anisotropy) can be difficult for subregions and surveys where data are sparse. However, the nugget and sill parameters are characteristics of the survey gear while the range and anisotropy are characteristics of the underlying population distribution. Thus, we used the most comprehensive survey for each subregion to characterize the range and anisotropy for the subregion and fixed these parameters for the other survey variograms as well as the cross-variograms. We then used the derived variogram and cross-variogram models and the data from the buffered region to get subregional means and variances from block kriging each survey and block cokriging models across surveys to examine changes in the model means and variances.

Results

In many subregions, sampling by one or more surveys was insufficiently dense to build strong, stable models of cross-correlation. As a result, many of the cokriging models had only minor effects on survey estimates or CVs. Dredge biomass estimates changed by up to 27% (Table 2). Most dredge CVs were similar between kriging and cokriging models but three subregions decreased by >5%. Results for the Northern Flank are noteworthy with a biomass increase of 11% and a reduction in the CV of 10%. This change is largely due to a good model of cross-variance with the NEFSC HabCam survey at the eastern portion of the subregion, coupled with complimentary coverage of the central and western portion of the subregion (Figure 4). Cokriging model estimates for the SMAST survey were similar to kriging models with only two subregions changing by >5% and the CVs changing by <=4% (Table 3). The biomass estimate for Closed Area 1 decreased by 14% due to a strong cross-correlation with the dredge survey and some dredge low-biomass dredge tows in the center of the closed area (Figure 5). For the HabCam survey, biomass estimates changed more dramatically with all but one subregion changing by >=5% and two subregions changing by >30%. CVs for Cokriging models were comparatively stable with two subregions changing by >=5% (Table 4). Both of the subregions where the biomass estimates changed by >30% were areas where the subregion was not comprehensively sampled by the HabCam survey and the survey apparently sampled an area with higher than average biomasses (Figure 6).

Note that some discrepancies between the original biomass estimates (Table 1) and the kriging and cokriging model estimates (Tables 2-4) were discovered in the writing of this document. We

hope to examine the source of these changes and potentially update these numbers in time for the review meeting.

Discussion

Cokriging methods have some clear benefits for improving model-based estimates and decreasing uncertainty, particularly in situations where a survey did not effectively cover the entire resource in a subregion or sampled a biased subset of the population. However, the technique needs further development before being fully employed in resource assessment. Of particular concern is the capacity to accurately fit the models of cross-correlation between surveys as well as autocorrelation within surveys. Two of the three surveys (Dredge and SMAST) were not designed for model-based methods and often don't have sufficient sampling density, particularly at the critical short distances, to characterize fine scale autocorrelation necessary for fitting nugget parameters. Ironically, stable models of cross-correlation typically come from pairs of surveys that cover the resource at adequate sampling densities, in which case the need to borrow information on spatial distributions is not as critical.

We've considered a Generalized Additive Model (GAM) approach as a potential alternative method to cokriging that would similarly allow parallel surveys to mutually stabilize biomass estimates and decrease uncertaintly but possibly circumvent some of the problems with estimating the spatial autocorrelations and cross-correlations. Such a GAM formulation could take a form similar to:

```
Biomass ~ s(Latitude, Longitude) + s(Depth) + as.factor(SurveyType),
        Weights=varIdent(form=~1|factor(SurveyType))
```

Thus, this model would fit a single non-linear surface for the spatial distributions of scallops but allow an offset for each survey gear and different variance structures for gear types. Similar to cokriging, this method would allow for prediction of a common spatial distribution but different biomasses for each gear type, reflecting the efficiency of each survey gear. This alternative modeling framework will be an area of future research.

Table1. Estimates of scallop biomass (MT) from each survey type for Georges Bank 2012.

	Survey			
Subregion	Dredge	SMAST	HabCam	Mean
CL1-Access	4,756	6,589	3,054	4,800
CL1-No Access	1,379	6,990	4,518	4,296
CL2-Access	7,560	6,814	7,404	7,259
CL2-No Access	15,670	14,921	8,183	12,925
NLS-Access	8,274	5,201	4,434	5,970
NLS-No Access	25	2,412	No Data	1,219
South Channel	20,871	22,269	24,426	22,522
Southern Flank	1,108	2,226	7,111	3,482
Northern Flank	2,167	3,933	5,836	3,979
Total	61,810	71,355	64,966	66,044

Table2. Comparison of Kriging and Cokriging result (biomass and CVs) for dredge surveys.

	Biomass				CV		
Region	Kriging	CoKriging	□ Biomass		Kriging	CoKriging	□ CV
Closed Area 1	7.1	7.3	3%		0.18	0.17	-0.01
Closed Area 2 NoAccess	4.8	4.8	1%		0.12	0.11	-0.01
Closed Area 2 Access	4.0	5.1	27%		0.16	0.11	-0.05
Lightship	5.4	5.4	0%		0.07	0.07	0.00
South Channel_NW	20.8	21.8	5%		0.13	0.10	-0.03
South Channel_SE	4.8	4.4	-8%		0.34	0.35	0.01
Southern Flank	2.6	2.8	8%		0.34	0.27	-0.07
Northern Flank	4.8	5.3	11%		0.38	0.28	-0.10

Table3. Comparison of Kriging and Cokriging result (biomass and CVs) for the SMAST drop camera survey.

	Biomass				CV		
Region	Kriging	CoKriging	□ Biomass		Kriging	CoKriging	□ CV
Closed Area 1	10.5	9.1	-14%		0.20	0.22	0.02
Closed Area 2 NoAccess	9.6	9.9	3%		0.18	0.14	-0.04
Closed Area 2 Access	6.5	6.7	2%		0.10	0.09	0.00
Lightship	4.5	4.8	6%		0.19	0.18	-0.01
South Channel_NW	19.5	20.1	3%		0.14	0.12	-0.02
South Channel_SE	2.6	2.5	-4%		0.38	0.39	0.00
Southern Flank	5.2	5.1	-1%		0.20	0.20	0.00
Northern Flank	7.0	7.3	4%		0.27	0.25	-0.02

Table 4. Comparison of Kriging and Cokriging result (biomass and CVs) for HabCam surveys.

	Biomass				CV		
Region	Kriging	CoKriging	□ Biomass		Kriging	CoKriging	□ CV
Closed Area 1	9.0	10.1	12%		0.24	0.20	-0.03
Closed Area 2 NoAccess	6.6	7.2	9%		0.13	0.10	-0.03
Closed Area 2 Access	6.9	6.6	-5%		0.12	0.13	0.00
Lightship	4.3	4.3	0%		0.19	0.18	0.00
South Channel_NW	19.6	18.4	-6%		0.11	0.10	-0.01
South Channel_SE	3.8	2.6	-32%		0.42	0.48	0.06
Southern Flank	8.0	5.3	-34%		0.23	0.18	-0.05
Northern Flank	5.2	4.4	-15%		0.26	0.27	0.01

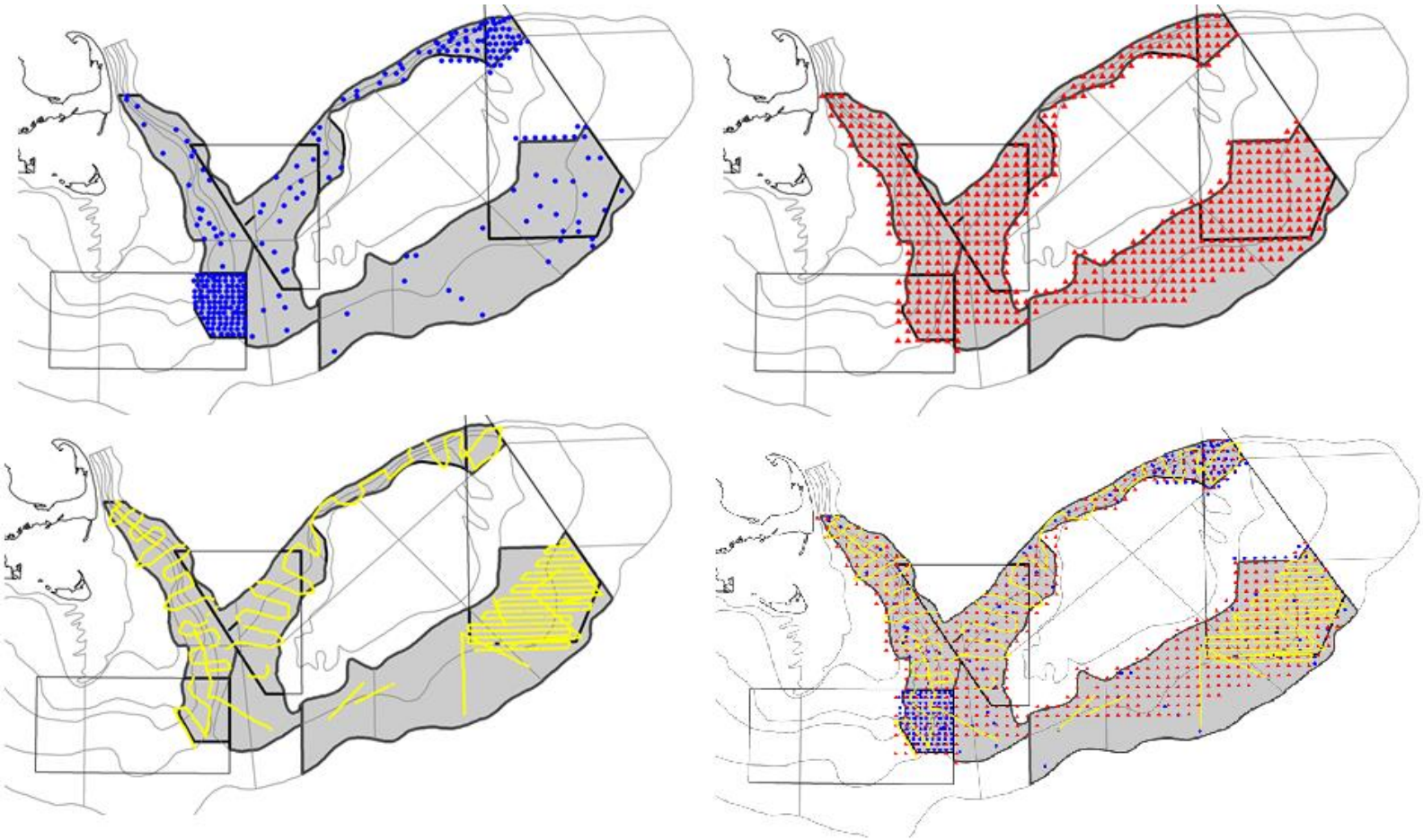


Figure 1. Sea scallop survey maps for Georges Bank in 2012 for (a) dredge surveys by the NEFSC and VIMS, (b) SMAST Drop camera, (c) NOAA HabCam (V4) and the HabCam group, and (d) all surveys combined.

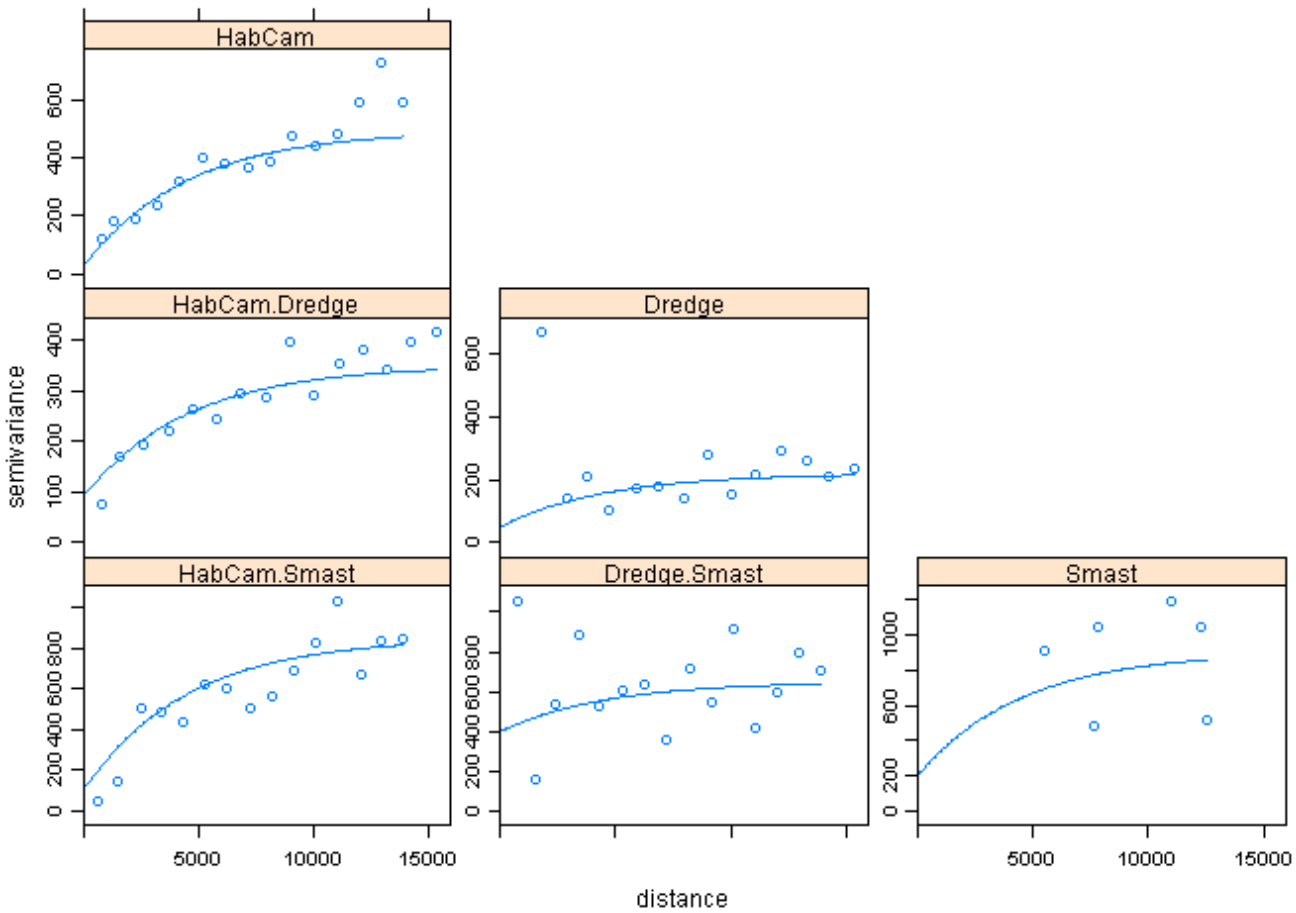


Figure 2. Example variograms and cross variograms for a subregion on Georges Bank. Open dots represent the mean variance of the data-pairs separated by a given distance and the solid lines represents the fitted variogram models. Plots on the diagonal are survey variograms while the sub-diagonals are cross-variograms for different pairs of surveys. Distance is in meters.

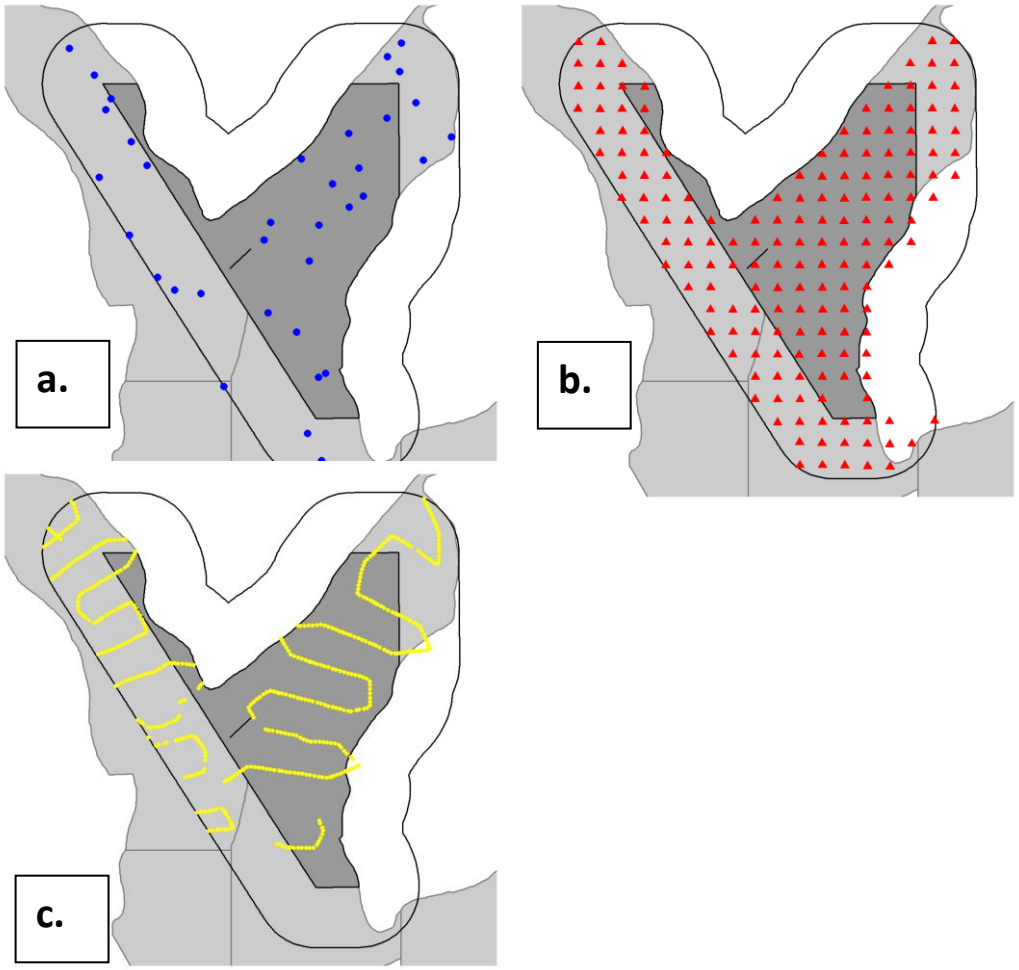


Figure 3. Example data extracted for modeling abundances in the Closed Area 2 subregion for (a) dredge surveys, (b) SMAST Drop camera and (c) HabCam surveys. The shaded area is the area to be estimated while the surrounding black line represents the neighborhood that data are drawn from for building spatial models.

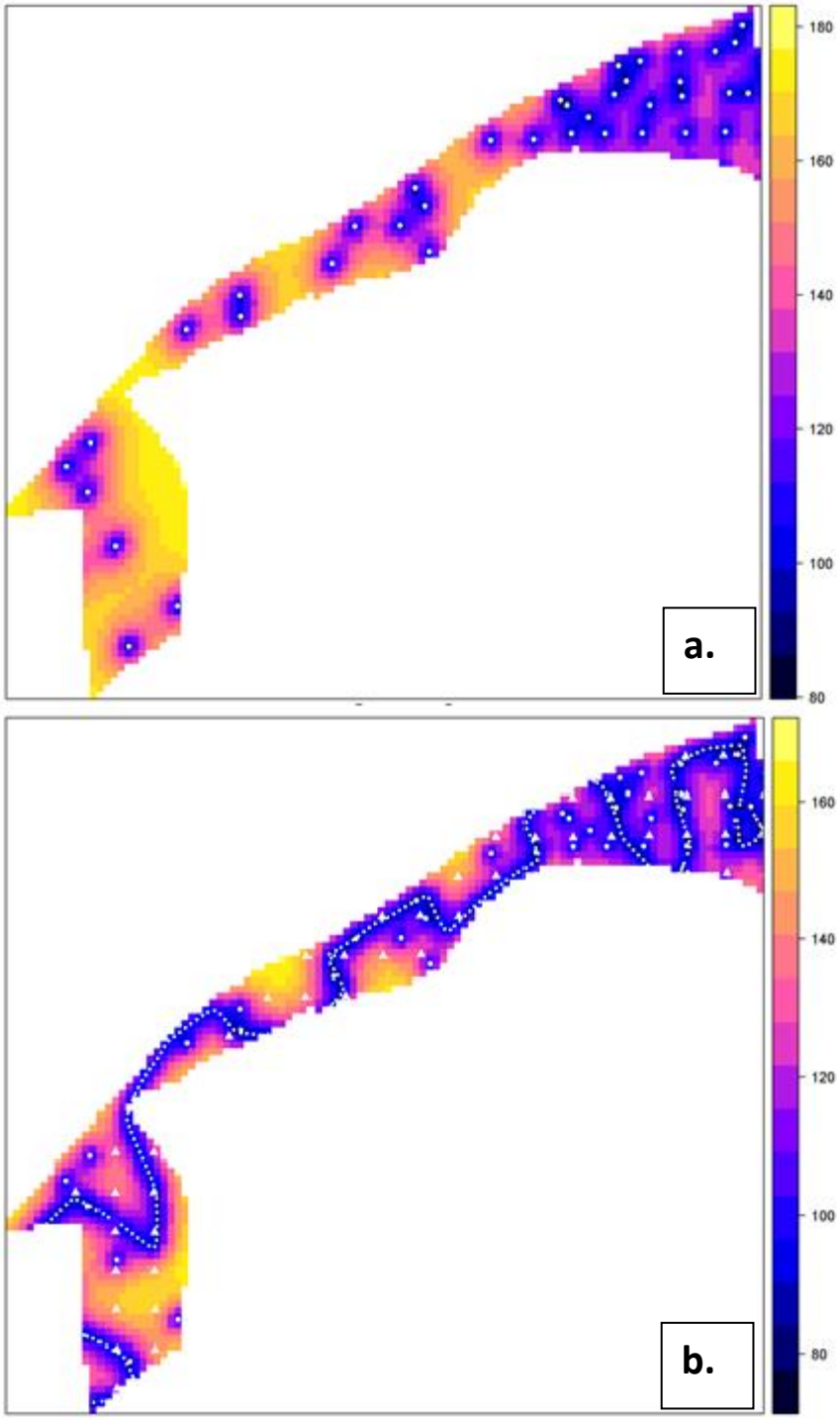


Figure 4. Kriging (a) and Cokriging (b) variance maps for the dredge survey for the northern flank. The influence of the HabCam survey is evident in the blue halos of lower variance around the HabCam track in (b).

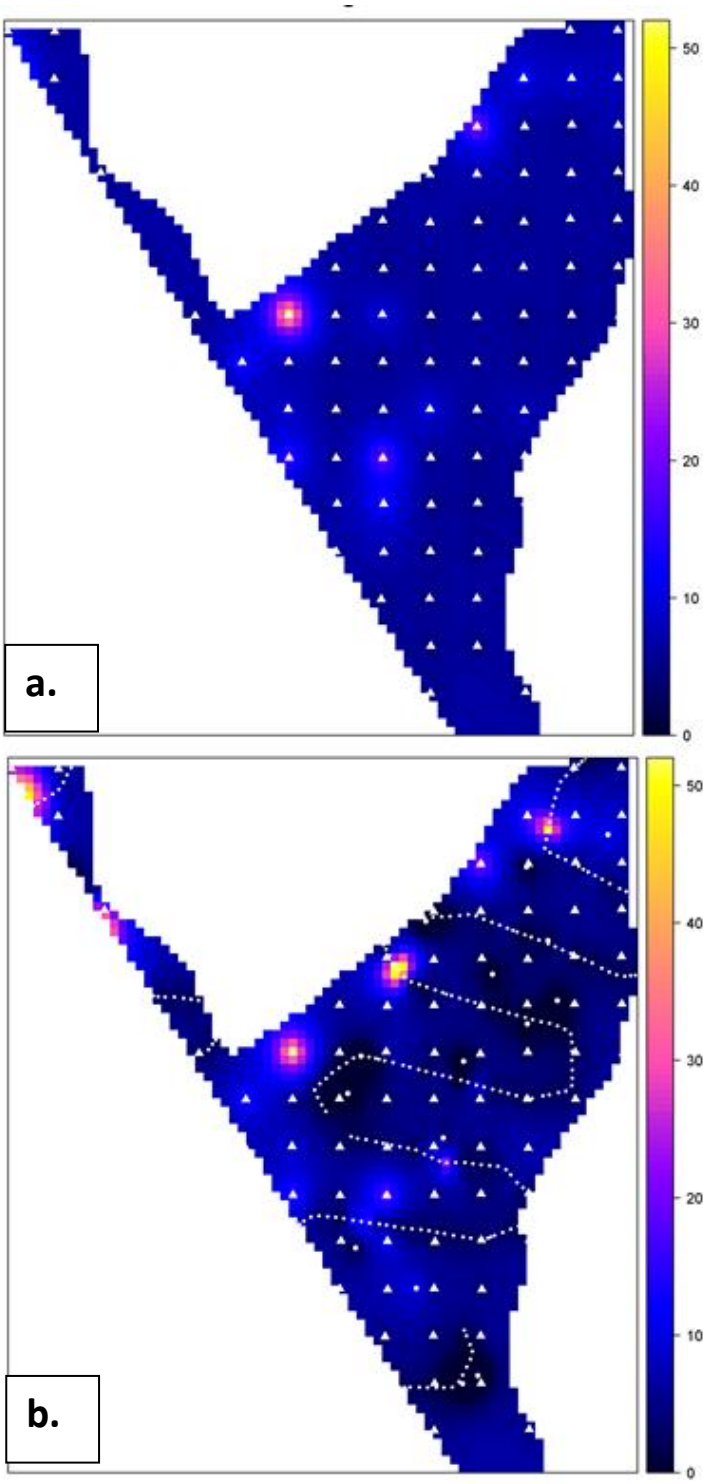


Figure 5. Biomass maps of kriging and cokriging results for the SMAST survey in Closed Area 1. The dark halos around solitary white circles are locations where the cross-correlation with the dredge survey decreased the biomass estimates.

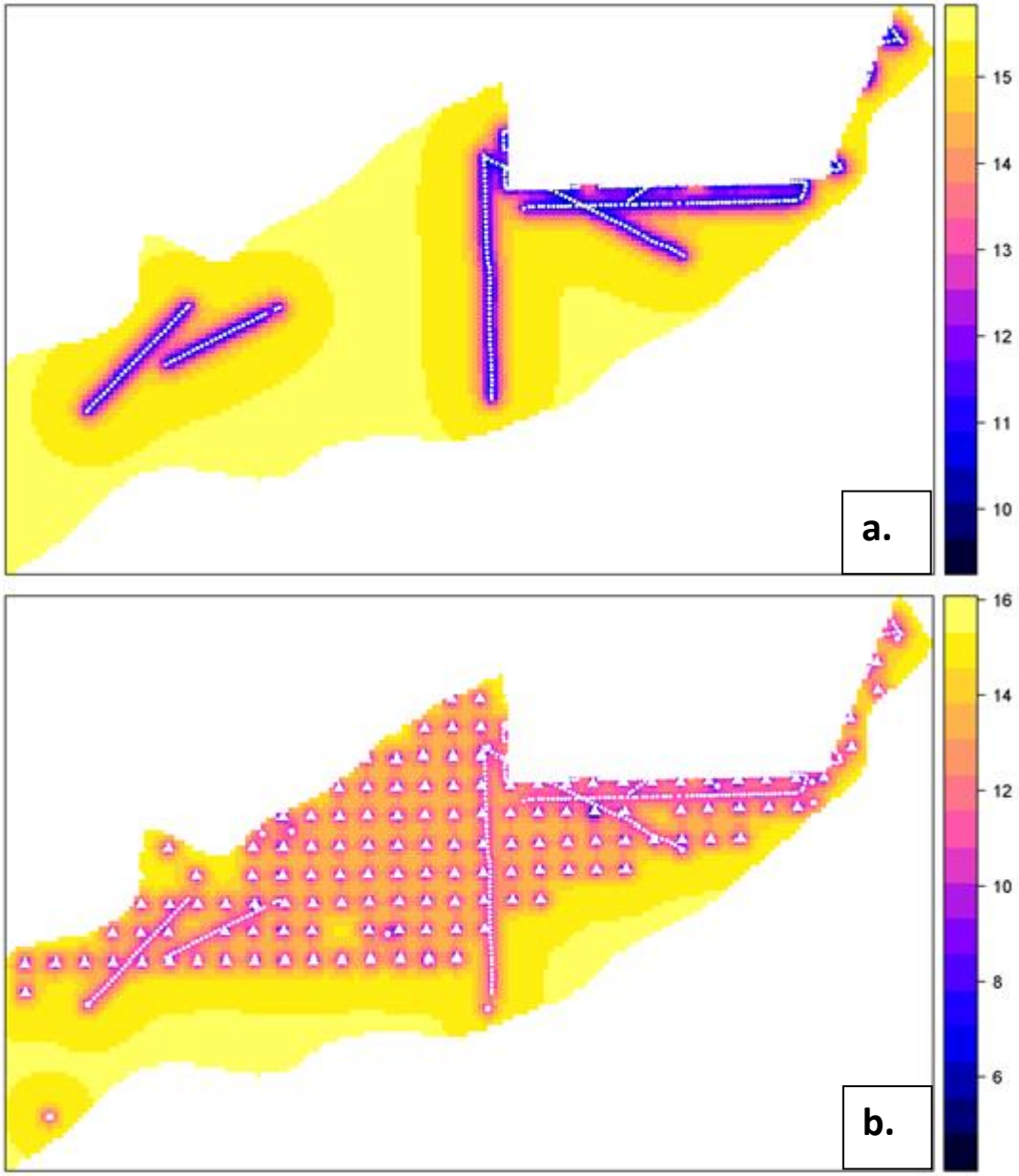


Figure 6. Kriging (a) and cokriging (b) variance maps for the southern flank for the HabCam survey. The survey is biased by over-sampling the higher density areas to the west of the subregion. The influence of the SMAST survey (white triangles) is apparent in the cokriging variance map (b) and decreased the biomass estimate by 34%.

TOR6 – Potential of HabCam surveys for non-scallop species, habitat, sea scallop ecology, ecosystem studies
6.1 Assessment of finfish
6.2 Automated finfish segmentation
6.3 Automated substrate classification
6.4 Assessment of habitat and habitat suitability modeling

6.1 Assessment of finfish and other invertebrates
Though originally designed for assessing sea scallops and benthic mapping, the HabCam vehicle has good potential for assessing some demersal finfish species that closely tend the bottom, including flatfishes, skates, and hakes, most of which are poorly sampled by traditional trawl surveys and have data-poor assessments.

Image-based surveys are commonly assumed to have an efficiency near 100% for appropriate species, providing estimates of absolute abundance rather than just an abundance index, but this assumption needs to be validated. An advantage of an imaging vehicle like HabCam is that it records the presence of fauna on the bottom but can also often capture an organism's behavior and document their reaction to the sampling gear itself, allowing for some validation or estimation of gear efficiency.

As a part of the 2014 yellowtail flounder empirical assessment, we conducted a pilot study on the ability to use HabCam imagery to estimate the efficiency of the gear and absolute abundance of yellowtail flounder in an intensively-surveyed area. From the analyzed imagery, we document diel behaviors and conditional reactions to gear and calculate preliminary estimates of gear efficiency. We also compared HabCam V2 and NEFSC scallop dredge swept-area abundances for 2010 and 2012 to estimate the efficiency of the NEFSC scallop dredge and derive an absolute abundance time series for yellowtail for the Georges Bank stock (see attached TRAC 2014 Background Documents).

A core problem with assessing finfish from HabCam surveys is that finfish are less common in images than sea scallops. Thus, the subset of survey images that are typically annotated for scallop assessment is insufficient for deriving similar estimates for most finfish species. In 2014, Dvora Hart received funding from the NOAA Office of Science and Technology to support the annotation of a much larger subset of the survey imagery to provide absolute abundance estimates for several species of flatfish, skates, and hakes. The project, initiated late in 2014, has annotated 55,000 images from the 2012 and 2013 surveys with enough remaining resources to annotate another ~450,000 images to be spread across the 2012 – 2015 surveys. We anticipate having much of the project completed and the data processed and available for upcoming benchmark assessments of hakes and skates, planned for 2016. As additional images of finfish become available from this project, NEFSC staff are performing finfish behavioral analysis, similar to the above work on yellowtail flounder to better understand gear avoidance and survey efficiency.

Also, researchers at Rutgers University are collaborating with NEFSC staff to assess the spatial distribution and fine-scale ecology of waved whelk (*Buccinum undatum*), an emerging and data-poor fishery in the region. NEFSC staff have created image assignments from archived HabCam

imagery for assessing large- and small-scale distributions of whelk in the Mid Atlantic Bight, which personnel at Rutgers are remotely analyzing using the web-based annotator.

6.2 Automated finfish assessment

The most difficult part of automated classification of fish in complex images is the segmentation of the target from the background. To that end, we have been working on developing segmentation algorithms of fish from HabCamV4 images using a variety of segmentation approaches.

One approach that we have been working with Lakshman Prasad at the Los Alamos National Laboratory (Prasad, Singh, and Gallager, submitted, Background Doc.) is to treat the images in a holistic manner and look for anomalous edges and texture. This has worked fairly well for detecting flounder on pure sand and skates on gravel background. Figure 6.1 illustrates an image with a complex background containing flounders. The associated anomaly map based on edge segments and texture shows boundaries where fish edges exist. This was produced following hierarchical segmentation followed by difference detection between segmentation layers.

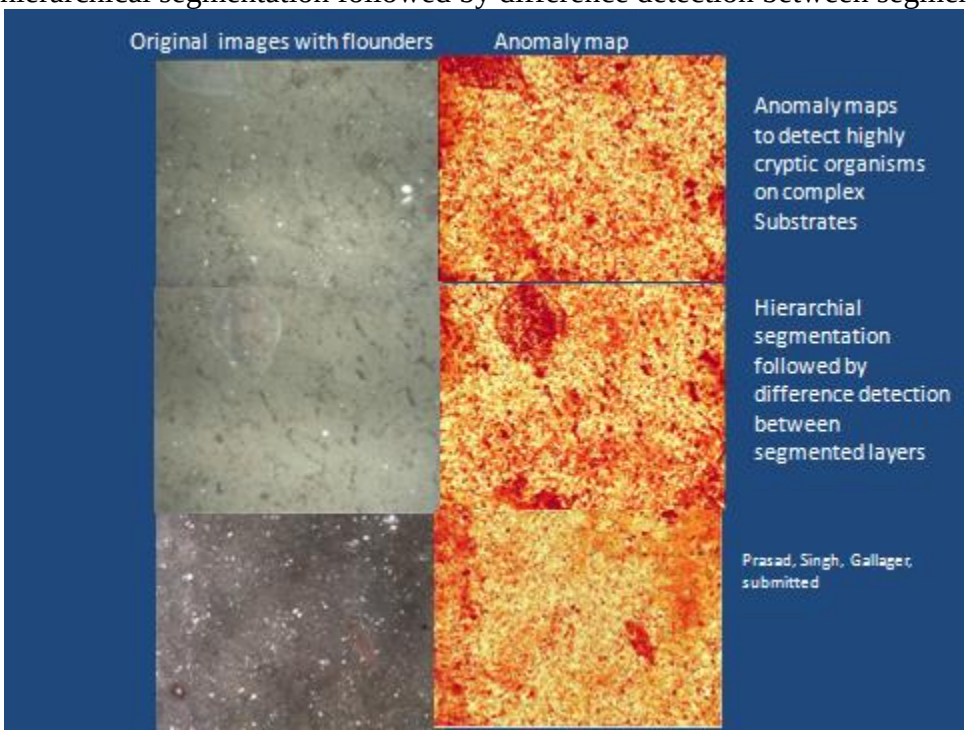


Figure 6.1 Segmentation of flat fish using hierarchical segmentation followed by differencing segmented layers from Prasad, Singh and Gallager (submitted)

Another approach we have been working on (Gallager at al, in prep) is to use the point cloud generated from rectification and disparity mapping of the stereo images and look for anomalies in the z axis. We call this Segmented Depth Maps (SDM), which clearly show distinct edges around targets that are completely buried in the sediment (Fig. 6.2 – 6.3). Figure 6.3 allows one

to see how the yellowtail are located in and near the sediment providing a discrete edge to draw contours on for classification.

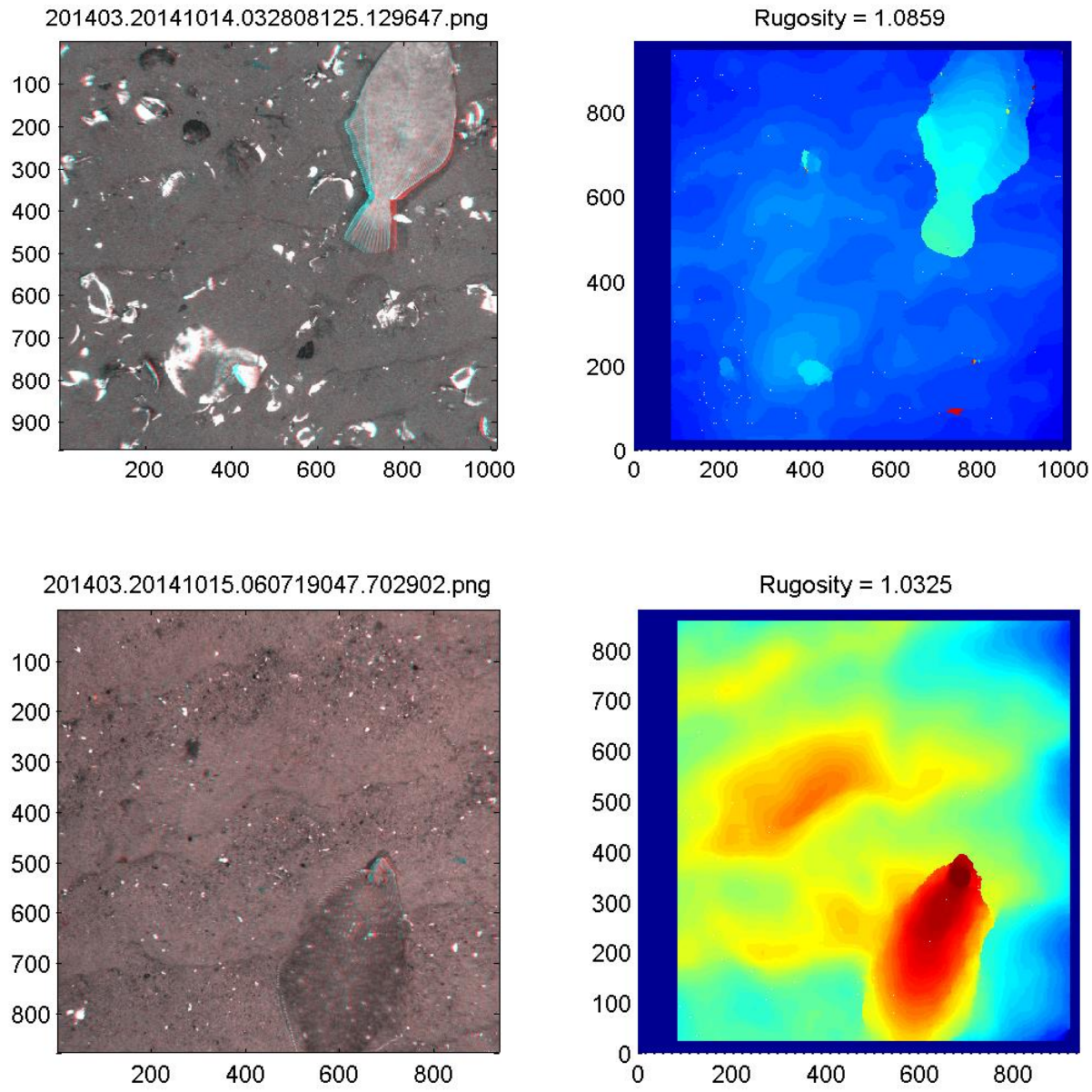


Figure 6.2 Segmentation of Yellowtail flounder from sand backgrounds using the stereo disparity and segmentation on the point cloud. Red colors are closer to the camera.

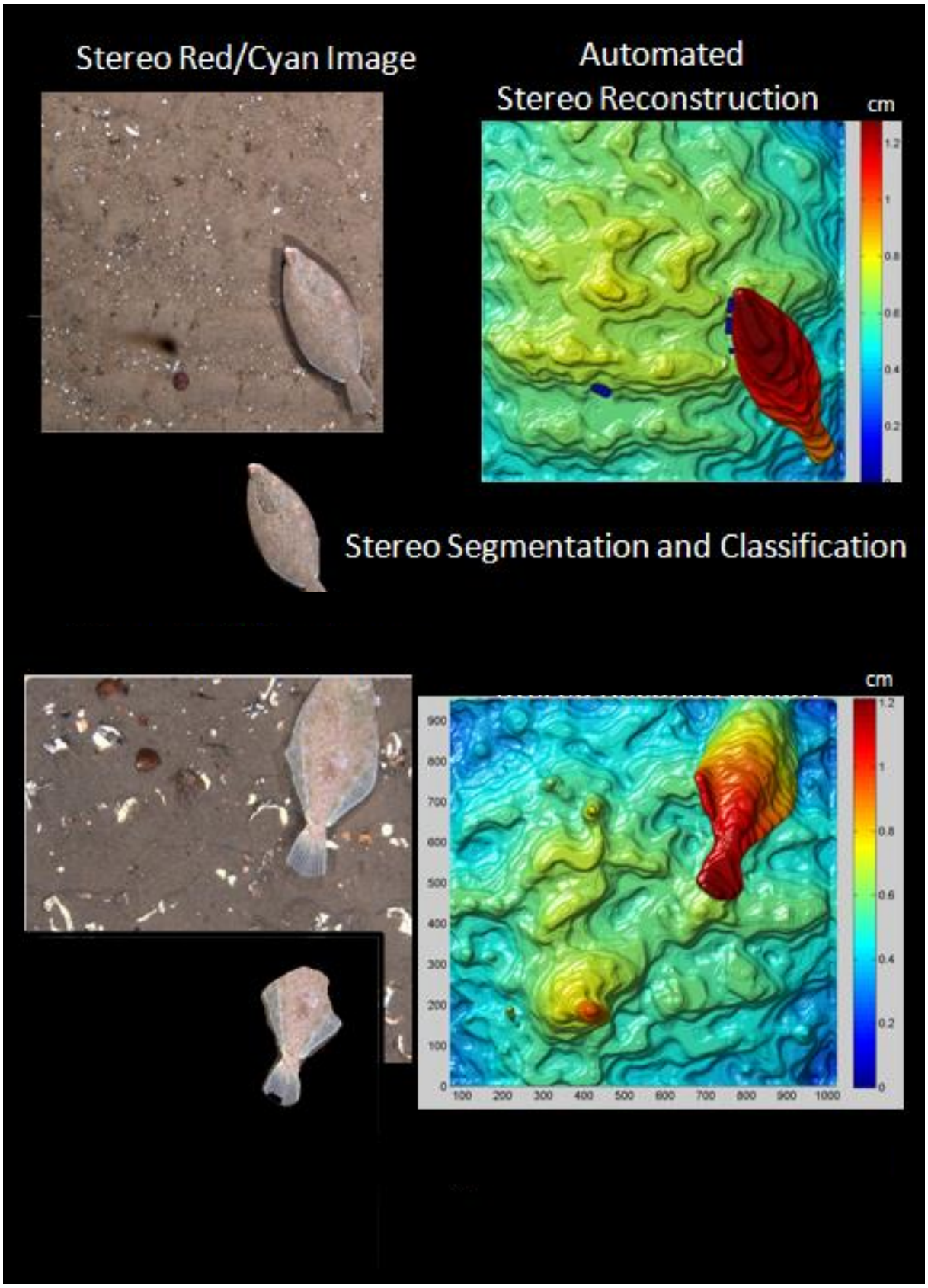


Figure 6.3 Segmentation of Yellowtail flounder from sand backgrounds using the stereo disparity and segmentation on the point cloud together with reconstruction of the 3D microstructure of bathymetry. Red colors are closer to the camera (cm). Each contour level is 1 mm in depth.

6.3 Automated substrate classification

We have been working on holistic and pixel wise classification of substrate in HabCamV4 images. A series of algorithms have been developed to break an image down into its fundamental constituents of color and texture. Each type of substrate is characterized and put into a training set. The features describing each substrate type are then used to build a classifier using random forests. The result is a mask the same size as the original image where each pixel is labeled with the substrate type the classifier chose. In most areas we are achieving 100% accuracy. The percentage composition of each substrate type is then evaluated for each image (Fig 6.4). We are currently running this classifier on the ship in real-time as well as back in the lab over all of our archived images.

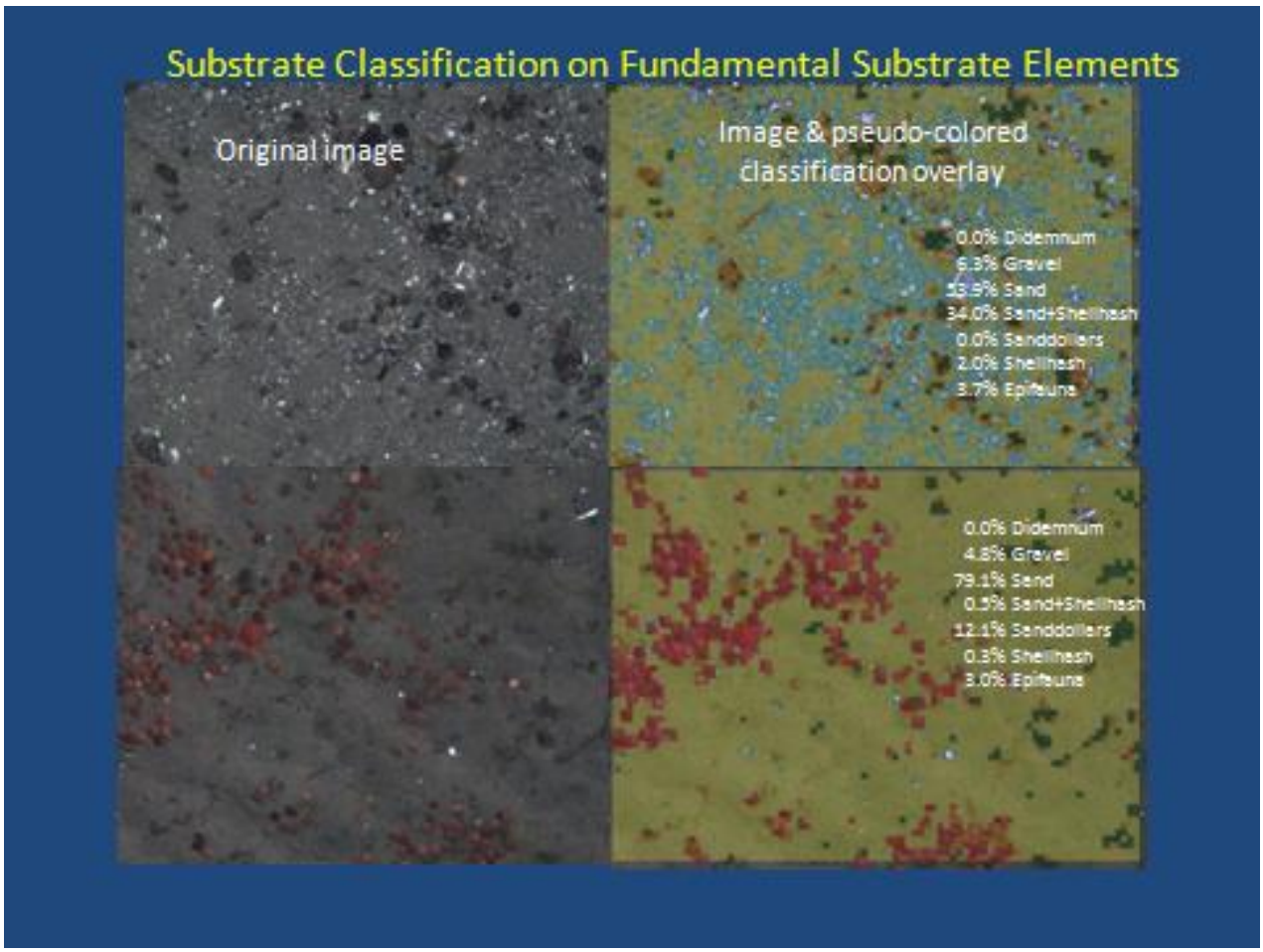


Figure 6.4 Automated substrate classification. From Honig, Stewart, York and Gallager (submitted). The submission is also available in the Appendix of this document.

6.4 Assessment of habitat and habitat suitability modeling

Data from HabCamV2 and V4 are being used to build habitat maps and to build models of habitat suitability for projection of species distributions. A habitat map is built first by assembling all of the predictive information available for a certain region and then fitting each "predictor" to a standard GIS grid. Typical predictors include: Temperature, salinity, depth, chlorophyll, greomorphology (slope, rugosity), substrate type from automated analysis, epifaunal cover such as bryozoan and tunicates, and bathymetry. Once each of these layers is interpolated to a standard grid, they may be combined with point observations of targets of interests. For example, the observations from HabCam of yellowtail flounder, cod, haddock, juvenile fish, scallop, seastar, etc. may be set on the same standard GIS grid as the predictors. The next step is to build a model of all of the predictor information where each of the targets are present. We then use both General Additive Modeling and random forests to evaluate the association between each predictor and target and develop a map of where each target should be based on the predictor information. The result are habitat suitability maps providing the probabilities of where targets like yellowtail flounder should be located on a scale of 0 to 1.

An example of habitat mapping is provided in the next few figures. Figure 6.5 shows a survey that was conducted on the northeast peak in Closed Area II HAPC in 2012. The grid spacing was 1nm to 0.5 nm at certain locations. Figure 6.6 shows the resulting predictor layers for sand, gravel, bryozoa, mussels and shell hash. The targets in this case were juvenile gadoids- cod and haddock. The distribution of juveniles were focused on the northern edge in a region where the bathymetery dropped off rapidly and in association with bryozoa and gravel beds. From these kind of habitat maps we can begin to see how population are changing distributions as function of external forcing functions such as climate change and fishing pressure.

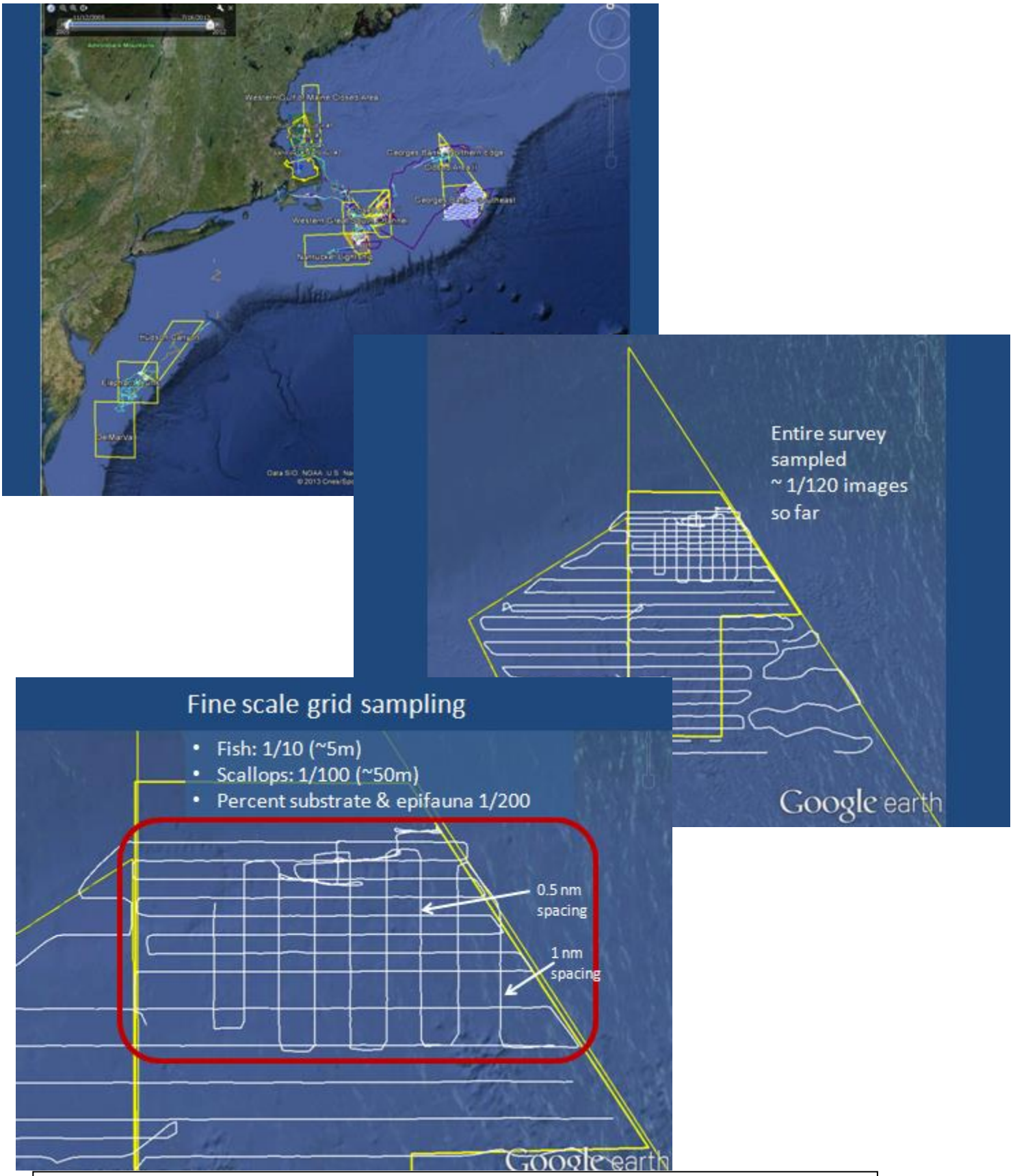


Figure 6.5 Fine scale survey grids on the northeast peak of Georges Bank in 2012 by HabCamV2. The entire Closed Area II was surveys at 1nm grid spacing the a finer grid was conducted at 0.5 nm in the northern section where it was known as cod and haddock habitat.

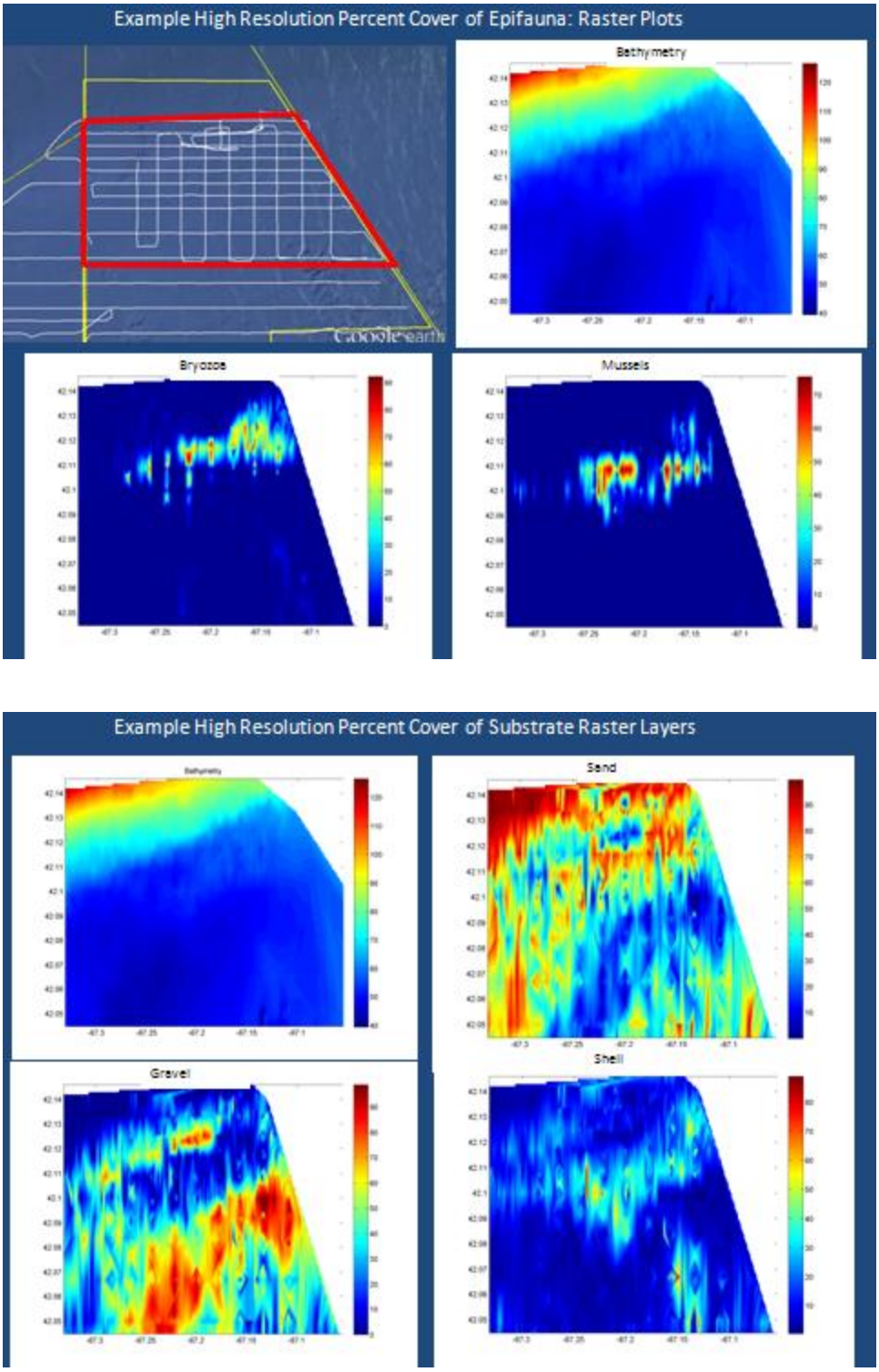


Figure 6.6 Gridded distributions of predictor layers.

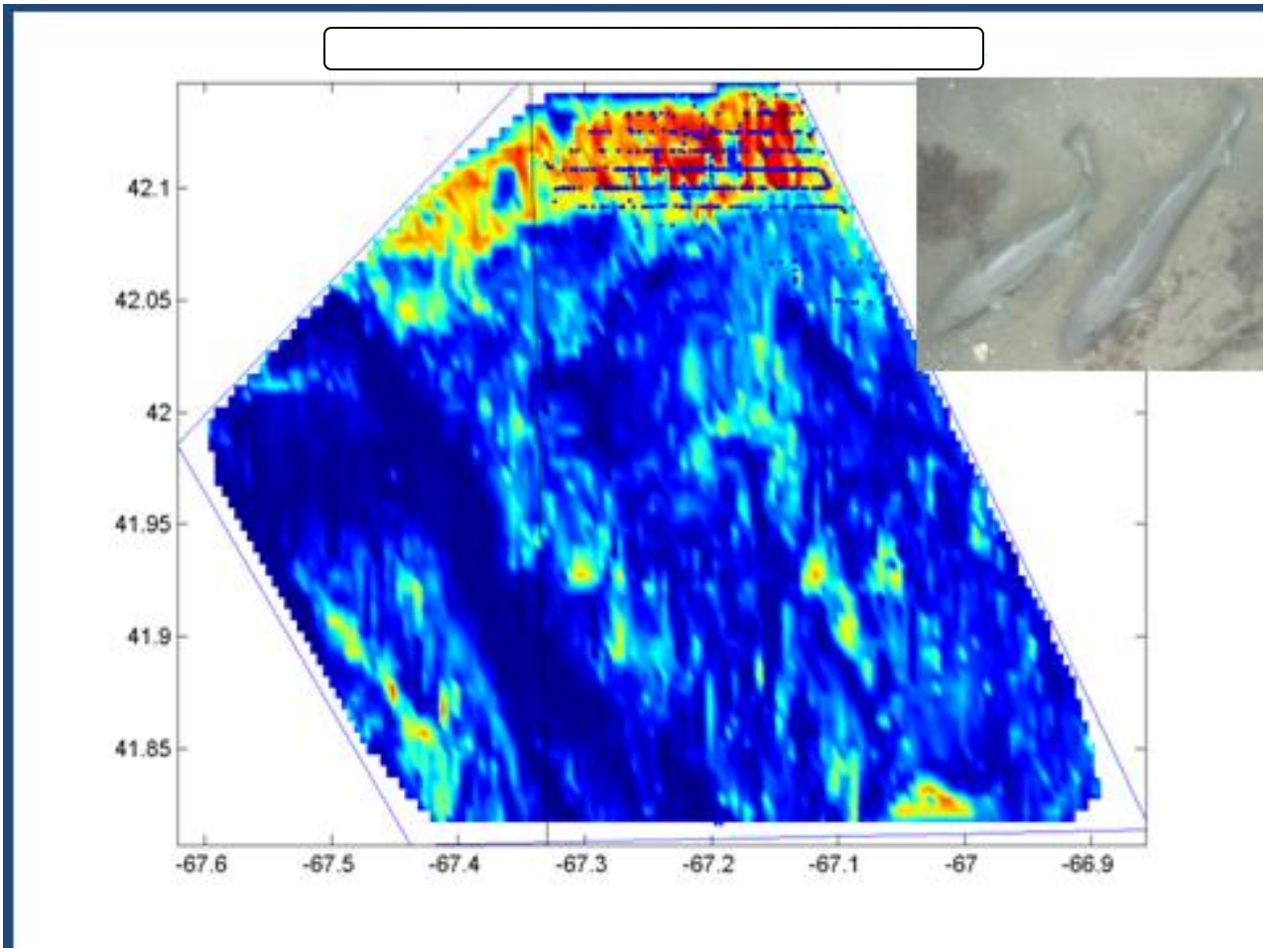


Figure 6.7 Observed (dots) and predicted colors distribution of juvelile cod and haddock in Closed Area II HAPC.

Term of Reference 7 – Frequency and combination of survey methods

Please see TOR7 in NEFSC_Dredge_corrected.pdf for discussion on the frequency and integration of NEFSC HabCam surveys with dredge surveys.

TOR8 – Future research

8.1. Methods for integration of automated identification of scallops into resource assessment.
8.2. Use of surface topography from stereo imagery for identifying scallop pre-recruits.
8.3. Finfish gear avoidance behavior and gear efficiency
8.4. Lobster habitat survey
8.5. Automated Substrate Identification
8.6. Benthic rugosity and habitat classification
8.7. Development of comprehensive maps from sediment classification and rugosity
8.8. Automated identification of scallops and finfish.
8.9. Effects of mobile fishing gear on benthic habitats
8.10. Seafloor Data Center

8.1. Methods for integration of automated identification of scallops into resource assessment.

Algorithms are rapidly being developed to identify and size scallops and fish from HabCam imagery. This is valuable as computers can potentially analyze the entire collected image set, which is far outside of what can be done manually due to the sheer volume of collected images. However, less attention has been paid to how this computer-generated data can be utilized for resource assessment. Software annotations at present have higher error rates than human annotators, and error rates vary depending on background (substrate) and image quality. Nonetheless, these annotations contain useful information. An active area of research at the NEFSC involves developing and testing various statistical methods to bias-correct automated annotations using samples of these images that have been manual annotated. An associated method would use the computer output to help optimally allocate efforts for manual annotation before beginning the manual annotation process. For example, areas that have been identified as having high densities of scallops by the automated annotators could be manually annotated at a higher rate.

8.2. Use of surface topography from stereo imagery for identifying scallop pre-recruits.

While scallop pre-recruits are small and often difficult to identify manually, they are commonly comparatively high in the water column and, thus, easily detected in digital terrain reconstructions from stereo imagery (Figure 8.1). We are currently developing methods for automatically detecting pre-recruits from the terrain reconstruction (TOR6), either using the rugosity measures or by quantifying the number of objects in an image higher than a threshold distance above the benthos. These methods can be trained by comparisons to manually annotated images where recruits were / were not available or by comparison to expected values from sediment classification or previous surveys in the area. Once developed, it would be possible to run such an algorithm on the vessel and display the output in near real-time which could be

valuable for measuring the spatial extent of recruitment events and dynamically re-designing surveys that are intended for documenting recruitment events.

8.3. Finfish gear avoidance behavior and gear efficiency.

NEFSC staff are actively working on documenting the diel patterns and avoidance behavior of finfish to the HabCam survey vehicle. The diel patterns in burying and flight behavior observed for yellowtail flounder can help explain observed diel patterns in gear efficiency for the NEFSC finfish and scallop dredge surveys. Modeling the gear avoidance behaviors provides a means to estimate the efficiency of HabCam itself, so observed densities can reliably be scaled up to absolute estimates. Once the efficiency of the imaging vehicle has been estimated, comparisons of swept-area abundances between gears, similar to the one performed for yellowtail (TOR6), can be used to estimate the efficiency of other more traditional survey gear types. Finally, information on observed diel behaviors can be helpful for people and computers alike when searching images for potentially cryptic species.

8.4. Lobster Habitat Survey

The American lobster fishery in Southern New England (SNE) consists of an inshore component, associated with complex costal habitats, and an offshore component including the outer continental shelf and submarine canyons. The long-standing paradigm for SNE lobsters is that lobsters recruit into inshore, rocky habitats and migrate offshore as adults, suggesting a dependency of the offshore fishery on inshore recruitment.

The SNE stock is in precipitous decline, partially attributed to water temperatures approaching or exceeding thermal thresholds and a recent history of recruitment failure. However, research suggests that lobsters in SNE may be moving further offshore to brood and release their eggs (B. Glenn, unpublished) and SNE fishermen are anecdotally reporting capturing very small, sublegal lobsters in offshore habitats. Thus, there is an interest in any potential for offshore habitats to provide some settlement / recruitment habitat for juvenile lobsters to help sustain the SNE fishery.

In 2014, B. Shank received funding from NOAA's Office of Science and Technology to perform a pilot survey of the offshore SNE benthos for potential lobster settlement / recruitment habitat. In conjunction with the 2015 NEFSC scallop survey, we will be using the HabCam_V4 with the C3D side scan and the vessel's multi-beam sonar to survey two regions of the outer continental shelf and upper reaches of submarine canyons. The primary goal is to find sufficiently complex, fine-scale habitat to provide refugia for lobsters through the first few years of their life. Automated sediment classification and bottom rugosity derived from stereo disparity will be used to characterize different habitats and hopefully, through multivariate cross-correlation, be scaled up to areas sampled by the side-scan and multi-beam sonar. Analysis of the side-scan and

multi-beam data, as well as large-scale habitat mapping will be accomplished through collaborating researchers at the Univeristy of New Hampshire Analysis / NOAA Center for Coastal and Ocean Mapping / Joint Hydrographic Center.

8.5. Automated Substrate Identification

We now have the ability to extract percentage cover of each substrate type from every image in real time as they are collected. This information will be used on an image by image basis to characterize habitat locally around any given target be it a fish or scallop.

Classification of seafloor substrate is fundamental to understanding and mapping habitat. While backscatter information from acoustic sidescan and multi-beam systems have used for many years as a surrogate of substrate, optical identification is necessary as not to confuse subsurface hardness with surface material available to biological activity. Given that we are now capable of producing millions of images per day from down-looking imaging systems, the need for automated substrate analysis has become fundamentally necessary for processing image information in a timely way for a variety of uses. Our submitted paper (Honig et al.) presents both scene-based bag-of-words and descriptor-based sub-image approaches to their automated identification. While the scene-based approach produces a per-image single classification of substrate (e.g. “gravel-sand with epifauna”), our sub-image approach produces a classification for each cell in a dense grid, resulting in a quantitative composite of fundamental substrate components (e.g. x% gravel, y% sand, z% epifauna, etc.). The sub-image approach also yields spatial information that is used for segmentation. Both approaches use descriptors derived from detected features as well as a scale pyramid of descriptors derived from dense grids of keypoints. For scene analysis we represent the image with a histogram of words from an exemplar descriptor vocabulary (bag-of-words), while for sub-image analysis we represent each individual grid cell with either a log-polar descriptor or a descriptor formed from a set of pixel data directly extracted from keypoints at cell locations.

For classification, we compare the use of linear support vector machines and random forests. Using these techniques, we process novel images to determine either the particular class of marine substrate (e.g. gravel-sand) in the case of scene analysis, or the quantitative composite of fundamental substrate components within the scene (e.g. x% gravel, y% sand, etc.) in the case of sub-image cell analysis. We provide a multi-scale analysis of substrate information that illustrates the need for rapid, automated processing of seafloor images.

We will now use this system to run over all of our archived images and get the substrate data for each image into our database.

8.6. Benthic rugosity and habitat classification

Benthic rugosity has been studied at scales of meters to better understand habitat and available surface area for recruitment. We now can calculate rugosity on an image by image basis with sub

image (cm) level resolution. The surface area available for recruitment and settlement of larvae of all kinds is now calculated in real time and stored along with each image in the database. This will provide an incredible library of information that may be queried from the database to extract say all the juvenile scallops that live in an area with a rugosity between 1 and 5 and a sand substrate (relatively flat, sandy environments) and compare to those in a rugosity of 5-10 (very rugose and full of depressions and hillucks) and a sand substrate. We will be able to immediately see how organisms are distributed as a function of substrate and micro-bathymetric features.

8.7. Development of comprehensive maps from sediment classification and rugosity

Shelf wide maps are now being developed of substrate and rugosity based on data from HabCamV4 and GIS mapping technologies. Soon we will be able to show the entire northeast continental shelf, the images that have been taken over time, and the associated substrate and rugosity in those images all with the click of a button. The fishing community will be able to search for specific features that they think might be important, or that may be changing over time since the maps are updated after every survey. Resource managers will be able to ask “what if” questions about how species are being redistributed as a function of climate change and where likely habitat is to support certain species.

8.8. Automated identification of scallops and finfish.

We will continue to work on developing algorithms and new approaches for segmenting and classifying fish and shellfish. As part of a collaboration with Lakshman Prasad, Los Alamos National Laboratory, we are developing a hierarchical approach to segmentation of images from a holistic perspective. In a sense you can think of this as looking at an image and deciding what is different about it compared to say 100 previous or later images. A moving average of scene descriptors provide a systematic way to detect when something is in a image that is not in the average image. This provides not only a very sensitive detection method but very rapid as well, and may be used to segment just about any target that someone might be looking for. For example, say a fisherman finds an interesting fish in his nets, he could provide a picture of that fish and we could present that to the image recognition database and it would search all images and find fish that resembled that in the test image. This would be a very powerful way to search complicated image datasets and provide answers to questions that we have not been able to address.

Another feature set that is being included into the hierarchical segmentation process is the 3D image form the stereo cameras on HabCamV4. We are now able to segment on the z-axis and detect any target that is up off the seafloor by only a few millimeters. This provides a very sensitive way to segment flounder that are actually buried but still produce a lump in the sand the same shape as their bodies.

Combining all of these techniques into one software workflow. currently we are able to detect scallops and flat fish with excellent accuracy: In a test set containing 10,319 images that contained 6,284 scallops and 63 yellowtail flounder, 97% of the scallops were found and measured and 57 yellowtails were found and measured. The idea of providing a rapid assessment of benthic conditions is rapidly approaching.

8.9. Effects of mobile fishing gear on benthic habitats

There has recently been much interest in mobile fishing gear's (trawls and dredges) impact benthic habitats, including how severe is the impact of the gear and how long does it take for the habitat to recover. but the topic remains poorly studied probably due to the difficulty in obtaining hard data on this topic. The side-scan sonar on the HabCam vehicle readily detects tracks left by scallop and clam dredges and trawl doors, making it possible to sample across these tracks and directly extract imagery of the bottom from within the tracks for paired-comparisons to adjacent habitats immediately outside the track. For the past three years, the NEFSC HabCam survey has been sampling dredge tracks created by either the NEFSC scallop dredge or VIMS RSA dredge surveys. The tracks can be reliably sampled as the locations and dates of creation of these tracks are known, the NEFSC dredge tracks are identifiable by the narrower width of the survey dredge and the VIMS tracks identifiable by paired-towing of a commercial dredge and narrower survey dredge. Thus, we are working towards better understanding both how dredging activities affects the benthos and how long such disturbance is still detectable for a variety of habitats across the continental shelf. This work can eventually be expanded to include other mobile gear types including clam dredges and otter trawls. Portions of this research are currently included in the dissertation work of a NMFS Fellow and graduate student in natural resources at Cornell University.

In a recent Research Set Aside (RSA) proposal collaborators at WHOI proposed to utilize a newly developed module for the REMUS 600 Autonomous Underwater Vehicle (AUV) to conduct a series of high resolution Before-After Control-Impact (BACI) habitat characterizations to evaluate and better understand how the Closed Area II (CLAII), Rotational Closures (RC) are working to protect habitat, scallop and groundfish populations. This will address the RSA Highest Priority of conducting an intensive industry-based survey of CLAII that may be a candidate access area in the future (i.e., open areas with high scallop recruitment or closed areas that may open to fishing in the future, such as groundfish mortality closed areas or current habitat closed areas). The project will also address the Medium Priority of Habitat characterization research by providing image transects of the bottom within scallop access areas, closed scallop areas, and in comparable fished areas that are both subject and not subject to scallop fishing utilizing the BACI design.

The northern section of CLAII, Habitat Area of Particular Concern (HAPC), has been closed to ground fishing since the early 1990's providing localized regions of communities rich in fragile epifauna such as bryozoa and lacy tube worms on a sandy gravel (Gallager et al 2012). While such areas are known to be important habitat for metamorphosing and juvenile cod and haddock as well as scallop, many scallops in the area tend to be old and large and of relatively poor commercial value. This has led the New England Fisheries Council (NEFC) to discuss the idea of providing Partial Access Areas in some of the most "over grown" regions to clean them out and provide renewed habitat for scallop and groundfish recruitment. In addition, several regions to the west in Georges Shoals Large and Georges Shoals Small, which are currently open to ground fishing, are being considered for closure to provide refugia against the constant impacts of scallop dredging.

Some of the important questions that need to be addressed before these closure policies are changed include the following:

1. What is the historical and current distribution of habitat types in the HAPC and Georges Shoals?
2. What would the impact be on the current communities (epifauna, scallops, groundfish) and habitat if regions of the HAPC were to have Partial Access?
3. How long do the mechanical impacts of scallop dredging persist as a function of substrate type (sand, sand/gravel, gravel/cobble) measured by both acoustics and optics?
4. How resilient are the different communities? (i.e., if allowed to rebound following episodes of dredging, how long would it take various communities to return to pre-impact condition)?
5. What is the distribution of the invasive tunicate species *Didemnum vexillum* in the HAPC (closed to ground fishing) and Georges Shoals (open to ground fishing) and how would providing Partial Access impact that distribution?

These are just some of the exciting questions that we can now address with HabCam data.

8.11 Seafloor Data Center

HabCam vehicle image surveys have been collecting large volumes of seafloor imagery in areas of importance to the sea scallop fishery since 2005 (Table 1). These almost 90 million seafloor images are not currently being served to the public in any easily accessible way. RAW unprocessed images are not accessible and are housed on drives in a storage cabinet. The processed images, derived from the RAW images, that are being served are not persistent and over the years have experienced drive failure bringing them down off the internet. Of the processed images that are served from habcam.whoi.edu, there is currently no way to navigate or access the stockpile of imagery. A seafloor image data center would allow all RAW and processed HabCam images to be served to the public in an easily viewable, usable, and

downloadable manner. There would also be enough storage capacity to serve other image products from other RSA funded projects including SMAST drop-camera images.

Once all images are served and a web image service has been put in place, modern color and illumination correction can be run over the entire image set. Accurate color and illumination correction, which creates an even light field, are essential for any further machine-vision processing that will be run over the images. For example, an algorithm created by Peter Honig (RPI), extracts habitat information in an automated way by calculating the percent cover of each substrate type such as sand, gravel, or shellhash. Serving the images to the public will also promote development of new algorithms for detecting sea scallops in images as well as allow for refinement of existing scallop detection algorithms.

Having a seafloor image data center will also allow for more rapid turn-around of new image data sets and data products derived from the images. Currently some processing takes place at-sea but there is still a lot of post-processing that needs to take place on land. When a cruise comes in from sea the images can be loaded into the data center and are then available for human and machine annotators to identify and measure fish and scallops in the images. A 10-gigabit network configuration will help load the images in a rapid manner.

It is imperative that image data products be dispersed to the public in a useful way. Serving the images alone is almost useless unless there is a web front end to access data. To accomplish this, small use-case meetings should be held with scallop fisherman to design a web tool that allows them to access the data they want to see in a way they request. The web front end would be finalized after input from developed use-cases but an example interfaces would be shown at these meetings to foster ideas. An example of a starting web interface is a navigational map that displays locations of seafloor images and allows users to highlight areas and be shown the resulting images. Maps of species density can be shown in a similar manner. For example, a map of yellowtail flounder abundance could be plotted with the ability to see the seafloor images along with side scan data. The maps can also be color coded to see areas of high scallop abundance with the ability to filter by time. Data download requests could be made after highlighting an area on the map and selecting a time range.

Table. 1. Summary of HabCam images collected to date and required digital storage.

Year	Number of Industry HabCam v2 Images	Number of NOAA HabCam v4* Images	Total Images	TB needed to Store RAW and Processed Images	Number of 4TB Drives Needed to Serve Images
2005-2006	3,408,578		3,408,578	18	9
2007	3,223,369		3,223,369	17	4
2008	7,121,315		7,121,315	37	9
2009	8,409,125		8,409,125	43	11
2010	5,934,323		5,934,323	31	8
2011	5,445,720		5,445,720	28	7
2012	4,422,896	12,266,918	16,689,814	86	21
2013	7,322,792	14,648,610	21,971,402	113	28
2014	3,327,738	12,448,882	15,776,620	81	20
Total	48,615,856	39,364,410	87,980,266	453 TB	118
				=906 TB of RAID 10 storage	=227 Drives
		* Produces image pairs, tallied here as two images per pair			

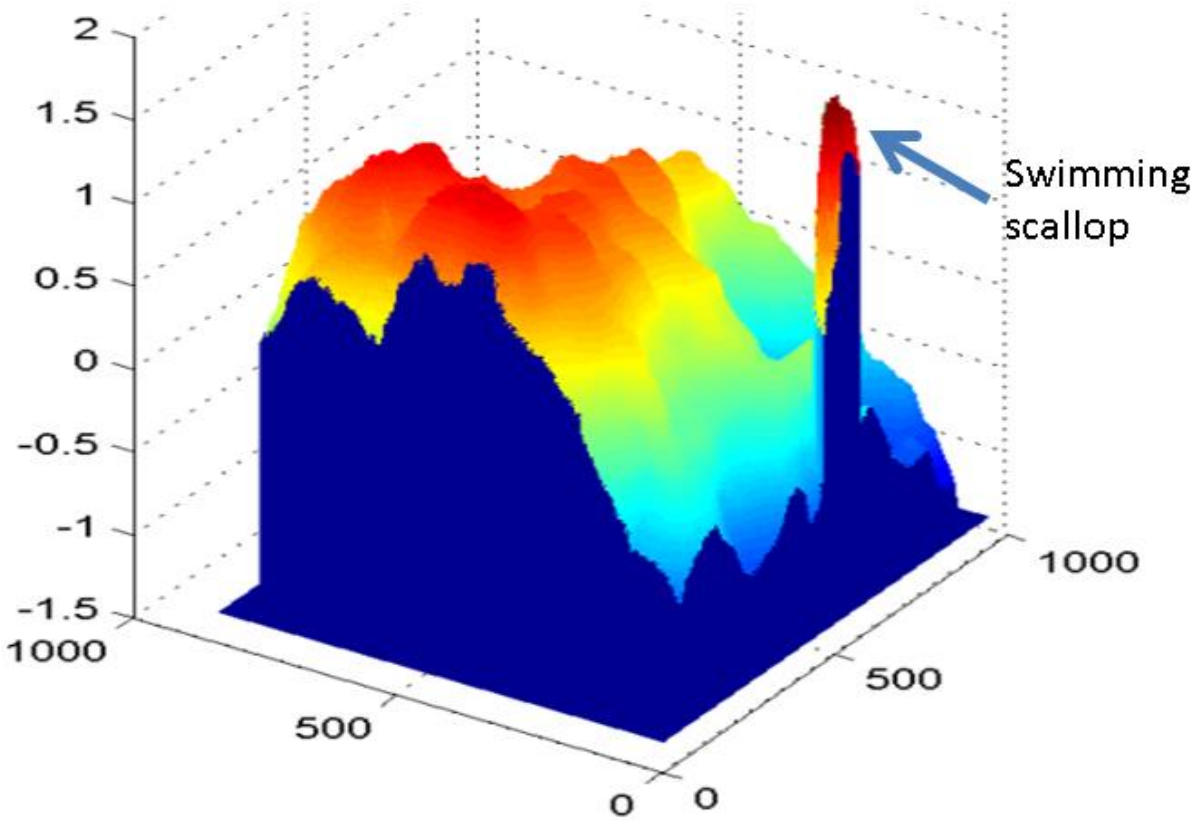


Figure 8.1. Digital terrain reconstruction from stereo imagery. Swimming scallops are evident in such reconstructions and can be automatically quantified by counting any object more than a threshold distance above the benthos.