FINAL REPORT: Fisheries-independent pilot survey for Golden (*Lopholatilus chamaelonticeps*) & Blueline (*Caulolatilus microps*) Tilefish throughout the range from Georges Bank to Cape Hatteras

AUTHORS: Michael G. Frisk, Jill A. Olin, Robert M. Cerrato, Paul Nitschke and Laurie Nolan

Key findings:

*Abundance and distribution**:***

- Golden Tilefish showed a core area of abundance approximately from south of the Hudson Canyon near Toms Complex to southern Georges Bank near Veatch Canyon.
- Catches were patchy throughout the range.
- Depth strata 3 dominated catches, none were captured in depth strata 1 (41–44.9/75–82.1 fathoms/meters).
- Catches of Blueline Tilefish were low and patchy.
- Larger hooks failed to capture a greater number of large Tilefish of both species; however, small hooks captured a greater number of small Tilefish.

Environmental preferences:

- Golden Tilefish occupied a very narrow temperature range and relatively narrow depth, oxygen and salinity range.
    - Possible limitation for range expansion.
    - Sensitive to environment change.
- Blueline Tilefish environmental analysis results were not significant; however, the species also displayed a limited temperature and depth range.

Survey design analysis:

- Proportional and optimum allocation of samples increased survey precision compared to simple random sampling.
- For Golden Tilefish and the overall survey, it seems possible to obtain a cv of 10% or better by shifting sampling effort to strata with larger mean abundance, variance and area.
- Revenue generated by selling fish can reduce the survey cost by 2-10%.

Survey Design Recommendations:

- Considering statistical and biological concerns we recommend that future surveys continue with proportional sampling (i.e., survey (stratified random pilot with min 3 hauls per stratum) or proportional (stratified random) allocations designs) of the ‘expanded’ range at a similar effort level and regional coverage sampled in the pilot survey. See section “Survey Design Recommendations” for full list of recommendations.

Objective 1: *Establish a comprehensive fishery-independent bottom long-line survey for the Golden and Blueline Tilefish along the Atlantic coast.*

Pilot Survey Design:

A stratified random sampling design was used in the pilot survey with a target range of about 200 stations. The survey was initially proposed to consist of sampling stations representing the “core” fishing areas of Tilefish based on commercial catch and a shallower and deeper “expanded” region to evaluate areas outside of the traditional fishery and better define the species range and abundance. The study area was divided into 9 north-south regions (NS codes 1-9) using the NEFSC bottom trawl survey latitudinal strata boundaries and 4 depth ranges (depth codes 1-4), developed at a meeting with the fishing industry, that considered both Golden and Blueline Tilefish depth distributions. Stratification was based on the following depth ranges (in fathoms/meters): 1 = 41-44.9/75-82.1, 2 = 45-53.9/82.3-98.6, 3 = 54-137.9/98.8-252.2 and 4 = 138-166/252.4-303.6. The N-S areas are labeled 01 to 09 and the depths 1-4, so the label for an individual strata was coded as for example 05-1 (Figure 1).

Stations were initially allocated to strata approximately in proportion to area, with few modifications (Table 1). Except for stratum 02-3, no stations were proposed for N-S strata 01 and 02, since Tilefish were not expected to be caught in this northern region. The two outer depth ranges (01 & 04) were allocated three samples per strata to allow the calculation of a standard deviation. The total number of proposed stations in this “expanded” region was 42. Depth range 02 was originally allocated 35 samples. After assigning samples based on area, three additional samples were added so that there were at least three in each strata for a total of 38 stations. The eight strata in depth range 03 were originally allocated 123 samples. After assigning samples based on area, three additional stations were created so that there were at least five in each strata for a total of 126 stations. Overall, the target survey had a total of 206 stations. This number was reduced slightly during the survey due to logistical considerations.

We conducted two cruises to complete the 2017 fisheries-independent pilot survey for Golden and Blueline Tilefish. Cruise 1 was conducted July 19th-July 28th in the southern portion of the project range and Cruise 2 was conducted August 5th-August 17th in the northern portion of the project range. F/V Sea Capture personal included Captain John Nolan and crew members Brent Davis, Al Ellis, Stephen Doyle and Aaron Smith. Scientific crew included Paul Nitschke on cruise 1 (NEFSC-NOAA) and Jill Olin (SoMAS) on cruises 1 and 2.

Table 1. Distribution and allocation of stations by latitude-depth strata in the survey.

Strata	Area (km^2)	% Total Area	# Stations	Strata	Area (km^2)	% Total Area	# Stations
01--1	433.3	1.2	0	05--3	2720.4	7.7	22
01--2	589.4	1.7	0	05--4	208.6	0.6	3
01--3	817.3	2.3	0	06--1	734.9	2.1	3
01--4	91.1	0.3	0	06--2	630.7	1.8	3
02--1	1168.3	3.3	0	06--3	727.7	2.1	6
02--2	2653.5	7.5	0	06--4	57.3	0.2	3
02--3	3684.9	10.4	30	07--1	314.6	0.9	3

02--4	237.3	0.7	0	07--2	374.1	1.1	3
03--1	1519.1	4.3	3	07--3	1551.0	4.4	12
03--2	2320.7	6.5	10	07--4	98.0	0.3	3
03--3	3184.3	9.0	26	08--1	182.9	0.5	3
03--4	177.3	0.5	3	08--2	708.0	2.0	3
04--1	1592.4	4.5	3	08--3	550.2	1.6	5
04--2	2167.4	6.1	10	08--4	62.2	0.2	3
04--3	2538.4	7.2	20	09--1	191.4	0.5	3
04--4	240.7	0.7	3	09--2	331.9	0.9	3
05--1	977.5	2.8	3	09--3	336.1	0.9	5
05--2	1236.1	3.5	6	09--4	48.1	0.1	3

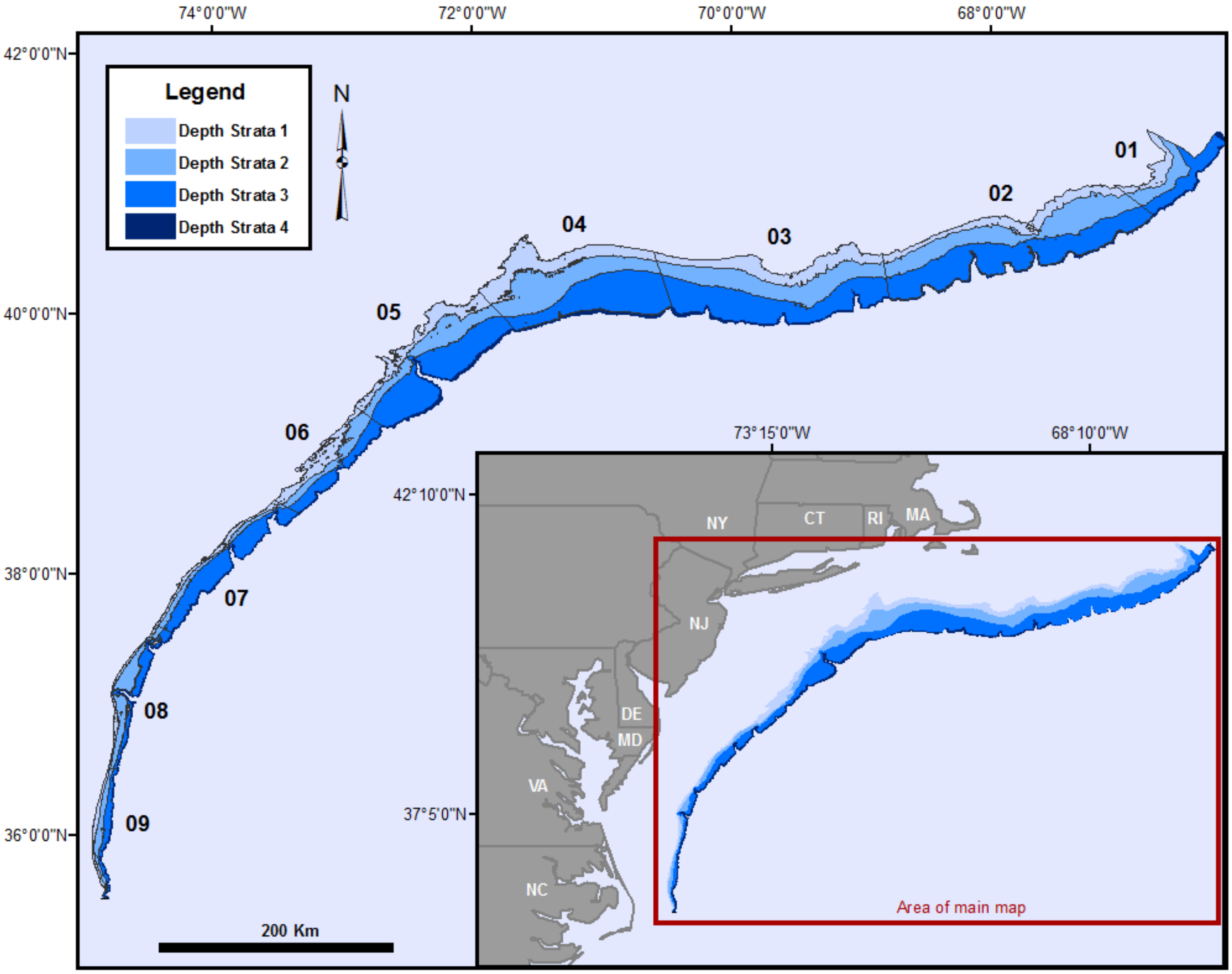


Figure 1. Stratified random sampling design with strata identified as 9 north-south regions (NS codes 01-09) and 4 depth ranges (depth codes 1-4).

Gear and deployment:

We used bottom long-lines that consisted of a one-nautical mile (1,852 m) mainline equipped with 150 evenly spaced gangions. Our original survey design proposed to use 300 evenly spaced gangions over one nautical mile for each station. However, after conducting the first stations using 300 hooks it became apparent that a reduction in the number of hooks was required due to time constraints needed for deployment, soak, retrieval and sample processing. There is a logistic tradeoff between the number of stations that can be conducted per day and the number hooks per station. There was no significant relationship between the number of hooks per set and the total catch per hook set ($F_{1,192} = 0.34$, $P = 0.556$; Supplemental Figure 1), maximum catch at any set was 41 individual fishes. Given this, hook saturation did not appear to be an issue with the use of 150 hooks per set. We chose to conduct more random stations per day to meet our target number of stations for the survey instead of achieving fewer stations using 300 hooks.

We deployed three different offset circle hook sizes (small = 8/0, regular = 12/0, large = 14/0), distributed at a ratio of 20-60-20. Bait presence and catch were recorded by hook number and hook size for each set. A standard bait size was used for all hooks to provide a consistent attraction potential for all hook sizes being compared. The original project design included deploying hook timers on 10% of the regular hooks for each set (30 per set). This protocol was implemented on the 1st day of the cruise where we conducted three stations. However, activation of the hook timer failed, likely because Tilefish captured did not provide enough force for timer activation. None of the hook timers were activated despite capturing 50 fish. Additionally, the hook timers slowed the deployment and haul speed and were cumbersome for the crew. As such, the scientific crew reduced the number of hook timers deployed per line to ~5-10. Following survey completion, a total of three hook timers were activated by Tilefish. The hook timers did indicate a duration of 22-30 minutes of fishing before catching.

Current meters were attached at each end of the long-line and data are currently being processed by Vitalii Sheremet, University of Rhode Island/Woods Hole Institute. The CTD was cast for a total of 188 stations (see summary by strata in Table 2); missing CTD station casts resulted from poor weather conditions.

Table 2. Summary (mean ± SD) of surface and bottom water temperature (°C), salinity (psu) and dissolved oxygen (mgL^{-1}) for the four depth strata (fathoms–fa; meters–m) in the Tilefish survey.

		Depth		Surface			Bottom		
Strata	*n*	fa	m	Temp	Sal	DO	Temp	Sal	DO
1	22	41.0–44.9	75–82.1	24.3 ± 2.6	32.0 ± 2.6	7.1 ± 0.6	10.0 ± 1.9	34.2 ± 0.8	6.9 ± 0.7
2	33	45.0–53.9	82.3–98.6	23.4 ± 2.4	32.6 ± 1.2	6.9 ± 0.6	11.3 ± 1.3	34.7 ± 0.6	6.6 ± 0.6
3	118	54.0–137.9	98.8–252.2	23.3 ± 2.0	33.0 ± 3.0	6.5 ± 0.8	12.4 ± 0.9	35.3 ± 0.4	5.9 ± 0.6
4	21	138.0–166.0	252.4–303.6	24.4 ± 2.0	32.6 ± 1.0	6.7 ± 0.9	10.5 ± 0.9	35.3 ± 0.2	4.7 ± 0.5

All attempts to maintain a consistent soak duration were made. However, to accommodate the number of stations in the survey and the steam time between locations, soak time ranged from 30 minutes to 4 hours with the average being 40 minutes. This range was necessary as multiple lines are deployed in different locations to maximize the number of stations completed per day. The effect of soak time on catch rates was not significant ($F_{1,187} = 0.005$, $P = 0.944$; Supplemental Figure 2). All fishing occurred in daylight hours, with the first line set no earlier than sunrise and the last no later than 30 minutes before sunset.

Objective 2: *Quantify relative abundance, biomass and size-structure of the two species.*

Abundance and distribution:

Catch was recorded from all strata sampled during the survey (Table 3). A total of 1,392 individuals were collected during the survey and included 21 species (Supplemental Table 1). Of the catch, 75 individuals were Blueline and 619 individuals were Golden Tilefish (Table 3) and their depth (Figure 2) and spatial distribution across the survey differed (Figure 3). Golden Tilefish showed a broader distribution, but were in highest abundance in depth strata 03 (98.8-252.2 m; Figure 2) in the northern portion of the range (Figure 3), whereas Blueline Tilefish showed a more restricted distribution, being caught in highest abundance in depth strata 02 (82.3-98.6 m; Figure 2) and generally in the southern portion of the range (Figure 3).

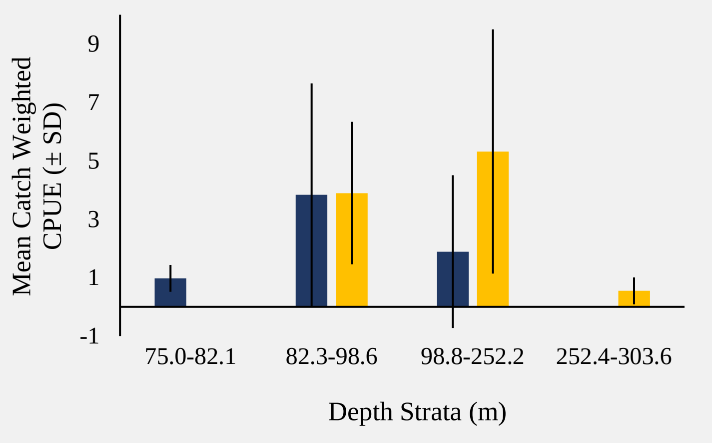


Figure 2. Blueline (dark blue bar) and Golden (yellow bar) Tilefish CPUE by depth. Data are mean (± SD).

Table 3. Total catch of Golden and Blueline Tilefish by latitude-depth strata.

Strata	Total Catch	Total Blueline	Total Golden	Strata	Total Catch	Total Blueline	Total Golden
02--3	143	0	0	06--3	23	0	0
03--1	25	0	0	06--4	4	0	1
03--2	52	0	29	07--1	1	0	0
03--3	271	0	141	07--2	9	0	0
03--4	14	0	1	07--3	63	22	2
04--1	13	0	0	07--4	21	0	5
04--2	52	0	11	08--1	26	1	0
04--3	210	1	174	08--2	14	11	0
04--4	12	0	7	08--3	14	0	9
05--1	9	0	0	08--4	7	0	0
05--2	23	0	0	09--1	4	3	0
05--3	311	2	235	09--2	51	35	1
05--4	9	0	3	09--3	2	0	0
06--1	1	0	0	09--4	3	0	0
06--2	8	0	0				

The core region of abundance for Golden Tilefish ranged from the southern edge of the Hudson Canyon to Veatch Canyon on Georges Bank (Figure 3). The distribution was patchy in the core area and the majority of captures were in depth strata 03 (98.8-252.2 m) and 04 (252.4-303.6). Stations placed in shallow regions did not produce large abundances of Golden Tilefish; however, it's possible the use of a similar bait size on all hooks may have limited capture of small Tilefish that were hypothesized, by participants in the fishery, to occur in shallow habitat. Further sampling may identify additional areas of abundance for Golden Tilefish that were not detected. Blueline Tilefish were primarily distributed south of the Hudson Canyon and catches were low and patchy and showed a similar distribution to the observer data (Figure 3). Additional sampling is needed to improve the delineation of Blueline Tilefish distribution in the survey area. It appears both species have a patchy distribution and occur in a relatively narrow depth range.

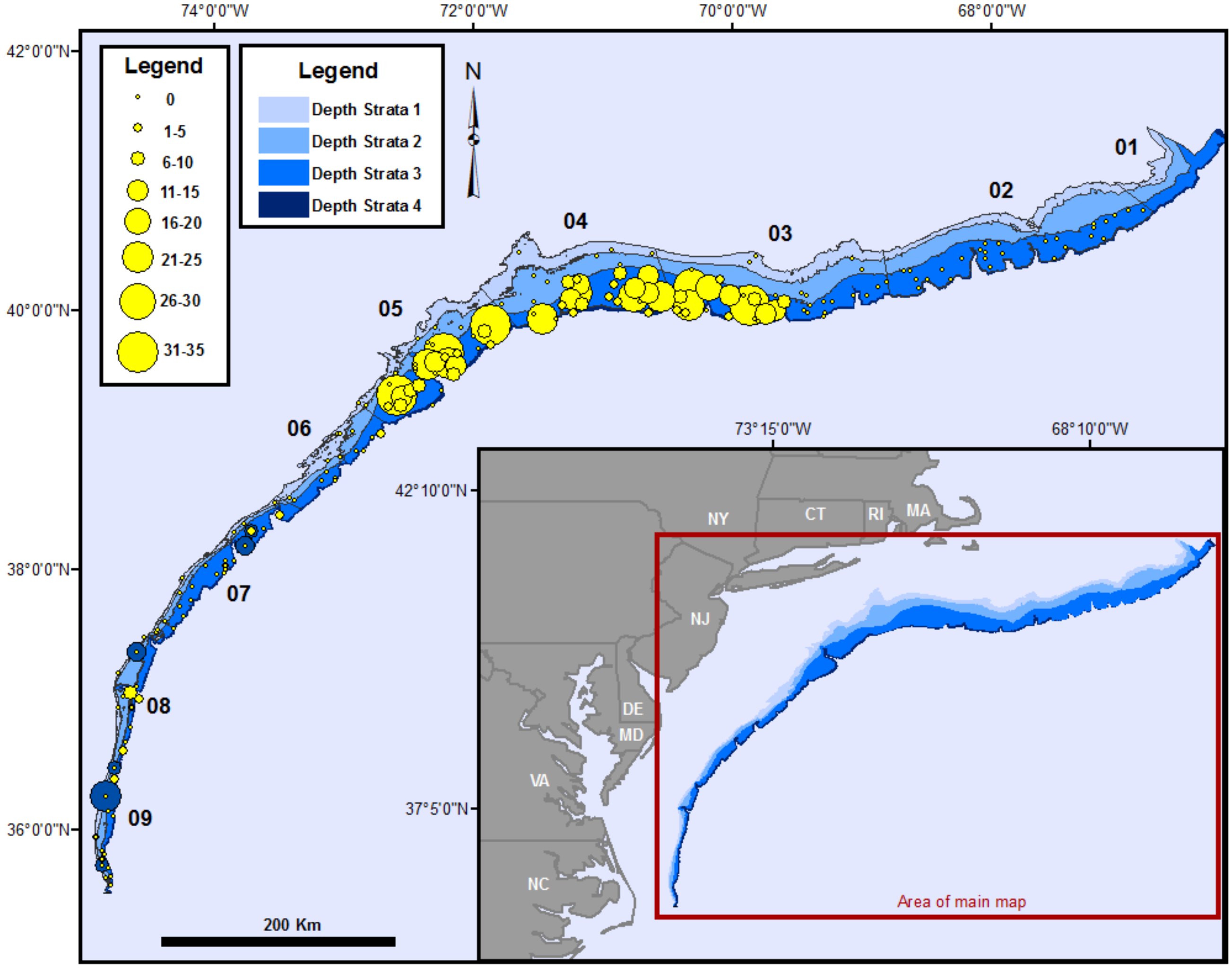


Figure 3. Station locations and distribution of Golden (yellow) and Blueline (dark blue) Tilefish caught (number of individuals) during the survey.

Size-structure and maturity:

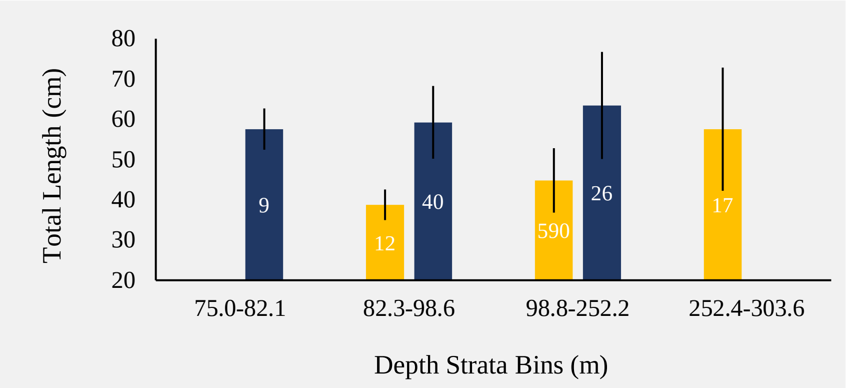


Figure 4. Catch of Blueline (dark blue bar) and Golden (yellow bar) Tilefish by total length. Data are mean (± SD).

Tilefish ranged in size from 25 to 110 cm and weighed 0.5 to 22.1 kg. The survey was dominated by catches of Golden Tilefish that averaged 45 cm in length and Blueline Tilefish that averaged 60 cm in length (Figure 4). Smaller individuals of both species were generally caught in shallower depth strata (Figure 4). There appears to be a trend of increasing Golden Tilefish length with depth; however, confidence intervals were overlapping and no further tests were conducted. Blueline Tilefish did show a similar trend with depth and size, although the trend is less apparent compared to the Golden Tilefish distribution (Figure 4) in the survey.

Gonads were classified as immature and mature for all Tilefish individuals caught in the survey. Immature classes including developing gonads. Mature classes included ripe and resting gonads. These classifications followed the criteria outlined in Idelberger (1985). The proportion of immature and mature Golden Tilefish was very similar across all depth strata (Figure 5, left panel). The overall catch of Golden Tilefish was dominated by immature individuals (Figure 5). In contrast, immature Blueline Tilefish were only captured in one depth strata (98.8-252.2 m; Figure 5, right panel) and contributed only a small proportion to total catch. Sample size was much lower for Blueline Tilefish and additional sampling is needed to determine whether the result is a sampling artifact.

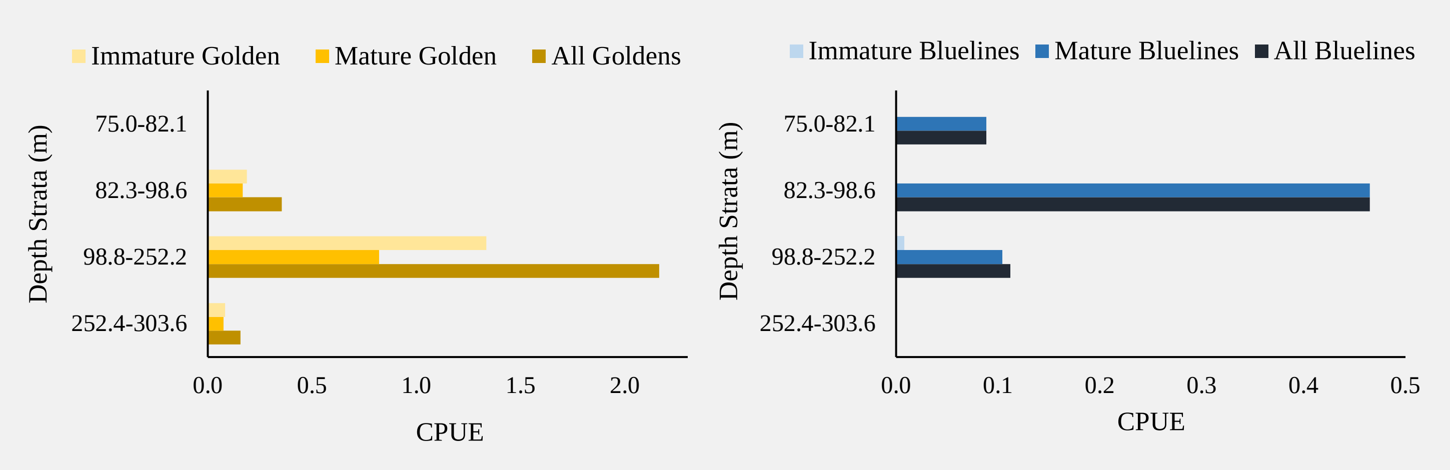


Figure 5. CPUE of maturity classes of Golden (left panel) and Blueline (right panel) Tilefish by depth strata.

***Biological sampling*:**

The survey provided an opportunity to collect samples for studies the research team is planning to develop and include analysis of stable isotopes, maturation, genetics and ageing. A total of 554 Tilefish, including both species were sub-sampled for a range of tissues. These included fin, reproductive, muscle, liver, stomach and otoliths. All tissues are currently being stored by M. Frisk for future analyses.

Gear Selectivity:

The distribution of catch across hook sizes were similar among Tilefish species with differences in the CPUE of small individuals between hook sizes (Figures 6). The large hook (14/0) does not appear to select for larger Blueline or Golden Tilefish compared to the regular and small hook sizes (Figure 6). However, the large hook did not catch the smallest individuals for either species. In contrast, the small hook (8/0) caught the most Tilefish in the overall survey. Specifically, the small hook captured a greater number of small Tilefish, chiefly Golden Tilefish.

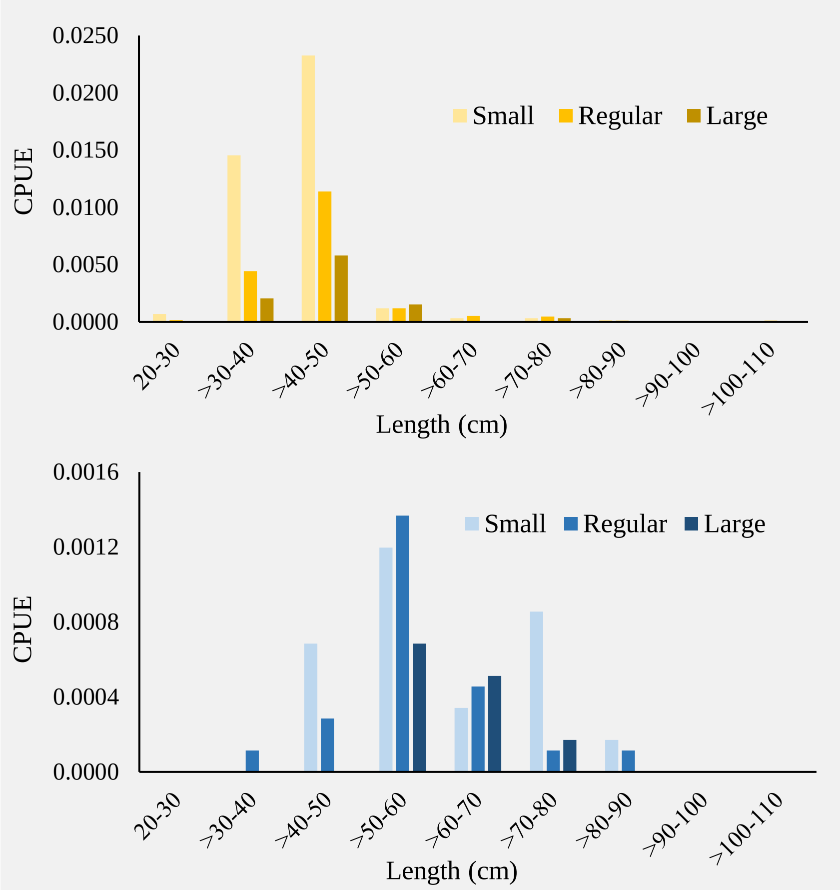


Figure 6. Length distributed CPUE (catch adjusted for proportion of hook size deployed) of Golden (upper panel) and Blueline (lower panel) Tilefish by hook size.

The observed length data for Golden Tilefish did not follow a common statistical distribution and appears in two modes between 36–50 cm total length (Figure 7). The two modes were most apparent in the small and regular hook sizes and may have originated from cohorts in the population. Because the data did not follow a common distribution, and appeared bi-model, typical analyses comparing means were not utilized. Instead, the observed length distributions were analyzed to determine if they originated from the same population using the non-parametric Kruskal-Wallis test. Significant differences in the distributions were estimated by hook size ($X^2 = 14.343$, df = 2, $P = 0.001$). The post-hoc Dunn's test indicated that the large and regular hook sizes originated from significantly different distributions ($Z = 2.74$, $P = 0.009$), as was the regular vs. small ($Z = 2.74$, $P = 0.009$) and large vs. regular was non-significant ($Z = 0.799$, $P = 0.428$). However, a review of the cumulative distribution functions of catch at length for small, regular and large hooks did not indicate large differences in cumulative catch by length class (Supplemental Figure 3). In this case, large differences between the distributions were not noted (Supplemental Figure 3).

Catch-Per-Unit-Effort data was not normal and contained a large proportion of zeros. Catch-Per-Unit-Effort of immature and mature Golden Tilefish were compared by hook size with the estimation of bootstrapped distributions of the means. The procedure utilized simple bootstrapping of the mean CPUE for each hook size and the 2.5th and 97.5th percentiles were reported (Figure 8). A total length of 40 cm was used to approximate the size of maturation (based on visual determination of individuals in the survey). The small hook size had the highest CPUE values for both immature and mature Golden Tilefish (Figure 8).

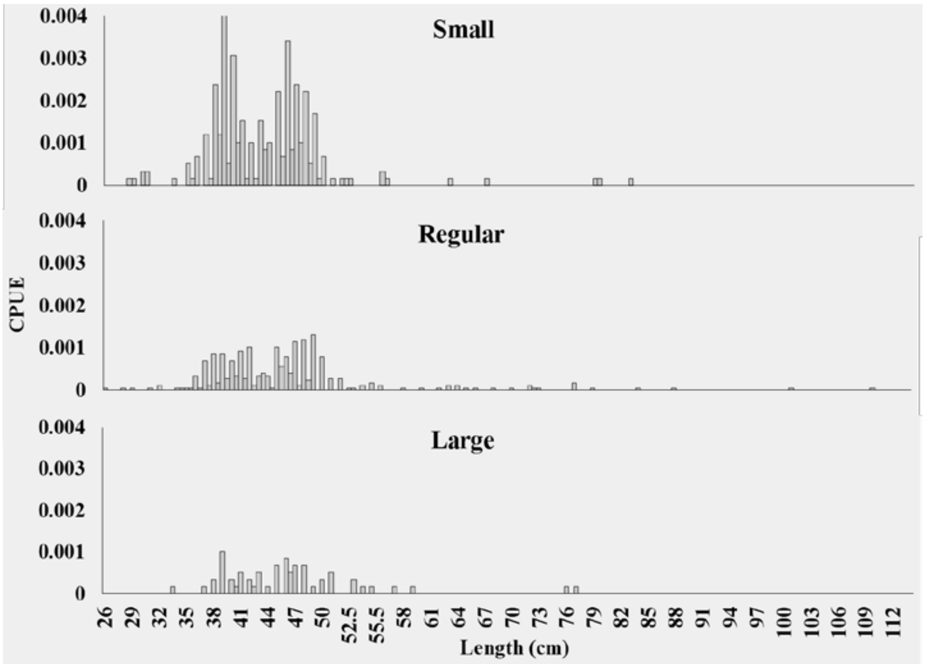


Figure 7. CPUE-length distribution of Golden Tilefish by hook size.

The data collected on the selectivity of each hook size has potential implications for the use of a domed shaped selectivity function in the stock assessment. Here, more individuals were captured by the small hooks and fewer individuals were captured by the large hooks and is not consistent with a domed shape curve. However, it should be noted that domed shaped selectivity could be determined by factors other than hook size and additional research is warranted.

One potential confounding issue for comparison of selectivity for the small hook, was the use of the same bait size for all three hook sizes in the survey. Use of consistent bait size across hooks was an attempt to standardize attraction potential across hook sizes to reduce potential bias. It is possible that Tilefish < 30 cm had difficulty taking the bait or were able to consume bait without biting the hook; thus, the potential exists that our values for the number of small Tilefish captured on small hooks are biased. Future surveys could experiment with bait size to test for a potential bias, as well as determine if Tilefish < 30 cm are within the survey area. Finally, we did not see an increase in large Tilefish captured by large hooks; potentially providing preliminary information that a doomed shaped selectivity curve based on gear hook size selectivity is not supported.

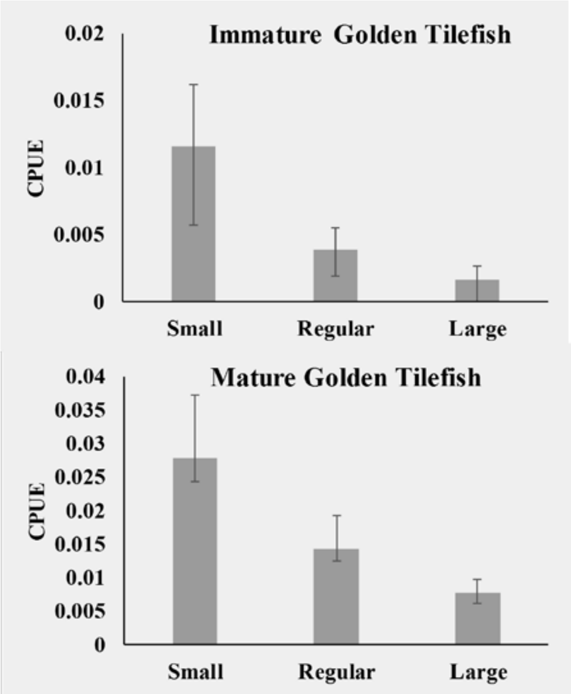


Figure 8. Golden Tilefish CPUE by hook size. Error bars indicate the 2.5th and 97.5th percentiles of the bootstrapped distribution of the mean.

Objective 3 and 4: *Determine the spatial distribution of both species and identify preferred depth strata across size range; and evaluate the role of environmental variables in driving the observed spatial distribution patterns.*

Environmental preferences of both Tilefish species were estimated using two approaches. First, environmental preferences were estimated following the non-parametric method developed by Perry & Smith (1994) and second using generalized additive models (GAMs) to predict species abundance and presence. The method developed by Perry & Smith (1994) provides a descriptive method of defining a species habitat preference by estimating the differences between available and occupied habitat through comparison of the cumulative distributions (Dunton et al. 2010; Sagarese et al. 2014). Habitat variables used in the analysis include temperature, salinity, dissolved oxygen and depth. First, the cumulative distribution function (CDF) of the available habitat $f(t)$ adjusted for unequal sampling effort within strata (W_h/n_h) was estimated with the following function:

(1) $$f(t) = \sum_h \sum_i \frac{W_h}{n_h} I(x_{hi})$$

where

$$I(x_{hi}) \begin{cases} 1, & \text{if } x_{hi} \leq t \\ 0, & \text{otherwise} \end{cases}$$

and where W_h is the proportion of the survey in stratum h ($h = 1, \ldots, L$), n_h is the number of stations in stratum h, x_{hi} is the measurement for a habitat variable (e.g., temperature) in station i of stratum h ($i = 1, \ldots, n_h$), and I is the indicator function where t represents an index ranging from the lowest to the highest value of the habitat variable. Equation 1 was calculated over all values of t for each habitat measurement (x_{hi}) available. Second, the CDF of occupied habitat $g(t)$ was estimated with the following function:

(2) $$g(t) = \sum_h \sum_i \frac{W_h}{n_h} \frac{y_{hi}}{\bar{y}_{st}} I(x_{hi})$$

where y_{hi} is the number of individuals caught in station i and stratum h, and $\bar{y}_{st}$ is the stratified mean catch (Perry & Smith 1994). Note that Equation 2 specifies the catch-weighted distribution of the habitat variable. For each habitat variable, the 5th, 50th (median) and 95th percentiles were determined. If species are randomly distributed with respect to the habitat covariate (x_{hi}), $f(t)$ between catch and habitat could be determined as the degree of difference between occupied ($g(t)$) and available ($f(t)$) habitat, with a Kolmogorov–Smirnov type test statistic (TS) for the absolute maximum vertical difference between the two CDFs:

(3) $$\max|g(t) - f(t)| = \max\left| \sum_h \sum_i \frac{W_h}{n_h} \left(\frac{y_{hi} - \bar{y}_{st}}{\bar{y}_{st}}\right) I(x_{hi})\right|$$

The estimated TS from Equation 3 was then compared with a pseudo-population of 10,000 randomized test statistics (PPTS) obtained by randomizing pairings of the following,

(4) $$\frac{W_h}{n_h} \left(\frac{y_{hi} - \bar{y}_{st}}{\bar{y}_{st}}\right)$$

and x_{hi} for all h and i across the entire survey (Perry & Smith 1994). Significance was estimated as,

$$(5) \qquad p = \left(\frac{\#\text{PPTS} > \text{TS}}{\text{Total PPTS}}\right).$$

The distributions of Golden and Blueline Tilefish were also modeled separately using generalized additive models (GAMs; Hastie & Tibshirani 1990; Wood 2006), a semiparametric extension of the generalized linear model (GLM). GAMs utilize a smoothing function (Wintle et al. 2005) that can easily handle nonlinear relationships and uncover hidden structure between variables missed by traditional linear methods (Hastie & Tibshirani 1990; Guisan et al. 2002). GAM analyses are often data-driven and are predictive in nature (Yee & Mitchell 1991; Fewster et al. 2000; Guisan et al. 2002). Two models were constructed for each species. The first predicted the probability of occurrence (PA) using a logit link function and a binomial error distribution. The second predicted the abundance (ABUN based on CPUE data) using a log link function and a lognormal error distribution (ABUN). All GAMs were built in R (R Core Development Team 2017) with the package “mgcv” (Wood 2011) using cubic regression splines and 5 knots ($k = 5$).

Model selection was based on Akaike’s information criterion (AIC; Akaike 1973) with small-sample bias adjustment (AIC_c; Hurvich & Tsai 1989). In determining model AIC_c values, all variables were counted as unique parameters and the number of observations used to compute the log-likelihood were used in calculating AIC_c. Models were ranked and compared using AIC_c weights and ΔAIC_c where AIC_c weights measure the weight in support of the model given the data and ΔAIC_c is the relative difference between the top-ranked model and each alternative model. In most cases, the model with the lowest AIC_c value was considered the best-supported model. However, when the AIC_c of several models differed by ≤ 2, we considered these models to be equally possible. Additionally, if the number of parameters (df) in comparative models differed by 1, then model selection was based on the log-likelihood, with the best-supported model having the lower log-likelihood (Burnham & Anderson 2002). Akaike weights (w_i) were calculated to interpret the weight of evidence for the best-fitting model with evidence ratios used to compare among models (Johnson & Omland 2004).

Unbiased estimates of each optimal model’s predictive performance were obtained through a validation evaluation. PA models were tested for discrimination and accuracy in using the packages “pROC” (Robin et al. 2011) and “Presence-Absence” (Freeman 2008), respectively. The ability of the model to discriminate between presence and absence sites was described using AUC (Brotons et al. 2004; Leathwick et al. 2006), with values between 0.7 and 0.9 considered reasonable and values >0.9 good, as the true positive rate was high relative to the false positive rate (Swets 1988; Pearce & Ferrier 2000). The ability to correctly predict the proportion of stations with a species given an occupied environmental profile was determined by calibration plots, with perfect calibration indicated by a line with a slope = 1 and an intercept = 0 (Wintle et al. 2005; Heinanen et al. 2008). Validation of ABUN models was assessed using model performance estimators, including calibration, correlations and mean error (Potts & Elith 2006; Heinanen et al. 2008). Calibration was measured with a simple linear regression between observed and predicted values, with the intercept term indicative of bias and the slope reflective of the consistency in the predictions (Potts & Elith 2006). The strength of the relationship between observed and predicted values was assessed using Pearson's correlation coefficient (*r*), although a perfect correlation ($r = 1.0$) may still display bias in a consistent direction (Potts & Elith 2006; Heinanen et al. 2008). The similarity between ranks of observed and predicted values was assessed using Spearman's rank correlation (r_{sp}), with a high value indicating a correct order of predictions (Potts & Elith 2006). Lastly, both root mean square error of prediction (RMSE) and average error (AVE) were calculated.

Cumulative Distribution Function Analysis:

Survey-specific cumulative distribution functions (CDFs) for available and occupied habitat with salinity, temperature, dissolved oxygen and depth profiles are shown in Figure 9 and median (50th), 2.5th and 95th percentiles are provided in Table 4. All CDFs were significant for Golden Tilefish while none were significant for Blueline Tilefish (Table 4). Salinities of occupied areas were similar to available habitat in all surveys for both species (Table 4, Figure 9A). Golden Tilefish associated with a relatively narrow temperature range with the majority of individuals (80%) captured between 9.8 and 12.4°C (Figure 9B upper). Blueline Tilefish similarly showed association with water temperatures that ranged between 9.9 and 12.8°C, although this association was not significant (Figure 9B lower). Golden Tilefish showed significant associations with depth, occurring in deeper water than available habitat (occupied: 113-165 m compared to available: 74-165 m) and a relatively narrow range. The CDFs for Blueline Tilefish showed steps and jumps in the distributions, indicating that a few stations with large catches impacted the shape of the function. This occurs with limited sample size and unusually large catches and indicates that more data is needed to resolve the associations between Blueline Tilefish and environmental variables.

The CDF analysis provided a powerful univariate approach for delineating habitat associations and produced significant results for Golden Tilefish. However, data was not sufficient to model the habitat associations of Blueline Tilefish and additional data is needed to better delineate habitat associations in the species.

Table 4. Habitat associations of Golden and Blueline Tilefish in the mid-Atlantic.

	Survey			Golden						Blueline					
Variable	5th	50th	95th	5th	50th	95th	DIF	TS	*P*	5th	50th	95th	DIF	TS	*P*
Salinity (psu)	33.4	35.2	35.6	33.6	34.7	35.5	0.05-0.41	0.33	**0.003**	33.6	34.3	35.5	0.11-0.78	0.18	0.978
Temperature (°C)	8.2	12.2	14.4	9.8	12.1	12.4	0.05-0.41	0.31	**0.007**	9.9	11.0	12.8	0.12-0.71	0.23	0.873
Dissolved Oxygen (mgL^{-1})	4.8	6.1	7.7	5.9	6.1	6.6	0.05-0.42	0.30	**0.011**	6.1	6.8	7.4	0.13-0.76	0.32	0.532
Depth (m)	74.2	84.2	165	113	128	165	0.07-0.69	0.83	**0.000**	90	100	120	0.16-0.99	0.46	0.885

Note: Data shown include habitat percentiles (5th, 50th, 95th), DIF (range of absolute vertical distance between distributions), TS (test statistic) and *P* value (probability). Significance (bolded) is based on an a priori α = 0.05.

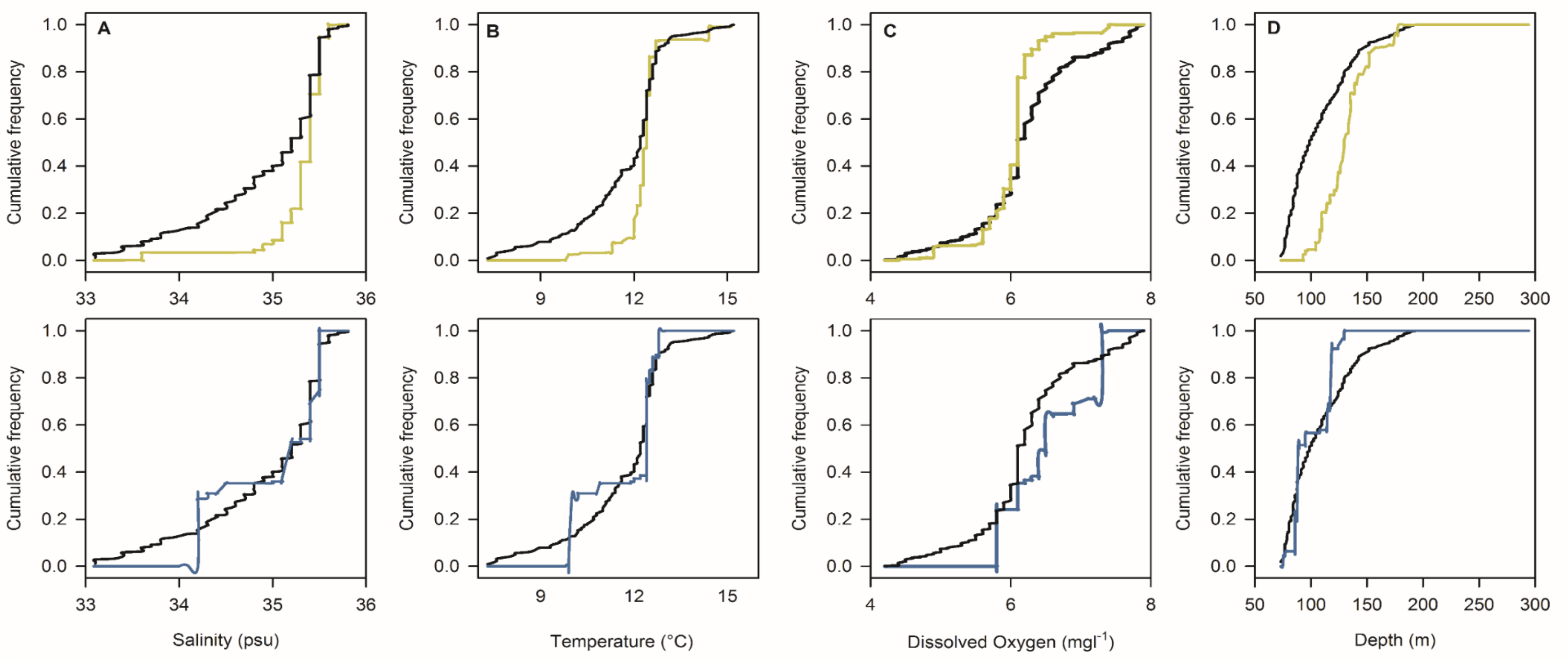


Figure 9. Cumulative distributions of available (black line) and occupied habitat for Golden (yellow line; upper panel) and Blueline (blue line; lower panel) Tilefish from the mid-Atlantic. **A** salinity, **B** temperature, **C** dissolved oxygen and **D** depth.

Generalized Additive Model Analysis:
The distribution of Golden and Blueline Tilefish throughout the mid-Atlantic differed. For Golden Tilefish, both probability of occurrence and abundance GAMs showed affinity for temperature, salinity, depth and dissolved oxygen (Table 5, Figure 10, 11) and models accounted for 34-45.4% of the deviance. Figures 10-13 can be interpreted as the effect of a single variable on the response variable (occurrence or abundance) if all other predictive variables are held constant. Thus, positive values over a range of the x-axis indicates a positive effect on the response variable. The probability of occurrence and abundance increased with temperatures between 10-12°C and at depths > 120 m (Figure 10, 11). For Blueline Tilefish, the best model supporting the probability of occurrence included salinity, depth and dissolved oxygen, whereas the best model supporting abundance included all four environmental variables. Both models indicated higher probability presence and abundance of Blueline Tilefish in shallow depths (Figure 12, 13). Similarly, catch of Blueline was more likely in temperatures between 10 and 14°C (Figure 13).

The ability of GAMs for Blueline and Golden Tilefish models to predict presence and absence for each station scored “good” according to AUC values (Table 6; Brotons et al. 2004; Leathwick et al. 2006). This implies the true positive rate was high compared to false positives (Swets 1988; Pearce & Ferrier 2000). Calibration plots were used to access the ability of the models to accurately predict sites with positive catches (Wintle et al. 2005; Heinänen et al. 2008). The best case scenario is a linear regression with a slope of 1.0 and intercept of 0. All GAM models showed poor performance with slopes ranging from 0.38 to 0.51 with low correlation coefficients with the exception of $r_{sp} = 0.45$ for Golden Tilefish.

GAMs are a powerful tool that can identify drivers of presence and abundance; but, are data hungry and require a balance between complexity and parsimony. Here we used 5 knots and a range of variables to identify important drivers of the species habitat preferences. The models were able to identify important variables with reasonable fits; however, more data is needed to improve model calibration and predictive performance to better define habitat preferences. The research team intends to collate data from the fishery observer program and NOAA annual surveys to developed additional models to delineate habitat preferences of both species.

Table 5. Summary of the optimal model selected using Akaike's Information Criterion (AIC_c), weight (w_i) is ratio of ΔAIC_c values for each model relative to the whole set of candidate models and the deviance [Dev (%)] explained for the occurrence (PA; using binomial distribution) and catch (ABUN; using negative binomial) models for each Tilefish species.

		Occurrence GAM				
Species	Common Name	Model	*n*	AICc	w_i	Dev (%)
Lopholatilus chamaelonticeps	Golden	PA ~ *s*(Temperature) + *s*(Salinity) + *s*(Depth) + *s*(DO)	188	175.91	0.49	34.0
Caulolatilus microps	Blueline	PA ~ *s*(Salinity) + *s*(Depth) + *s*(DO)	188	83.44	0.59	22.7
		Abundance GAM				
Species	Common Name	Model				
Lopholatilus chamaelonticeps	Golden	ABUN ~ *s*(Temperature) + *s*(Salinity) + *s*(Depth) + *s*(DO)	188	477.90	0.99	45.4
Caulolatilus microps	Blueline	ABUN ~ *s*(Temperature) + *s*(Salinity) + *s*(Depth) + *s*(DO)	188	112.63	0.59	66.7

Table 6. Validation measures for the optimal occurrence (PA) and abundance (ABUN) models for Tilefish species based on independent test datasets.

		Occurrence GAM					Abundance GAM						
Species	Common Name	AUC (%)	AUC CI (%)	*m*	*b*	P	*r*	r_{sp}	*m*	*b*	RMSE	AVE	P
Lopholatilus chamaelonticeps	Golden	94.1	89.2-99.0	0.55	0.12	0.000	0.10	0.48	0.25	2.74	7.2	1.6	0.36
Caulolatilus microps	Blueline	95.5	90.4-100	0.04	0.31	0.000	-0.02	0.10	-1.69	10.29	60.59	9.49	0.84

Note: AUC = area under the receiver operating characteristic curve; AUC CI = 95% confidence intervals around AUC; *m* = slope and *b* = y intercept of the fitted calibration line: observed = *m*(predicted) + *b*; r = Pearson's correlation coefficient; rsp = Spearman's rank correlation coefficient; RMSE = root mean square error of prediction; and AVE = average error.

Figure 10. GAM plots identifying the effects of the variables from the optimal models on the probabilities of occurrence (PA) for Golden Tilefish from the Northwest Atlantic in 2017. Hatched lines represent 95% confidence intervals.

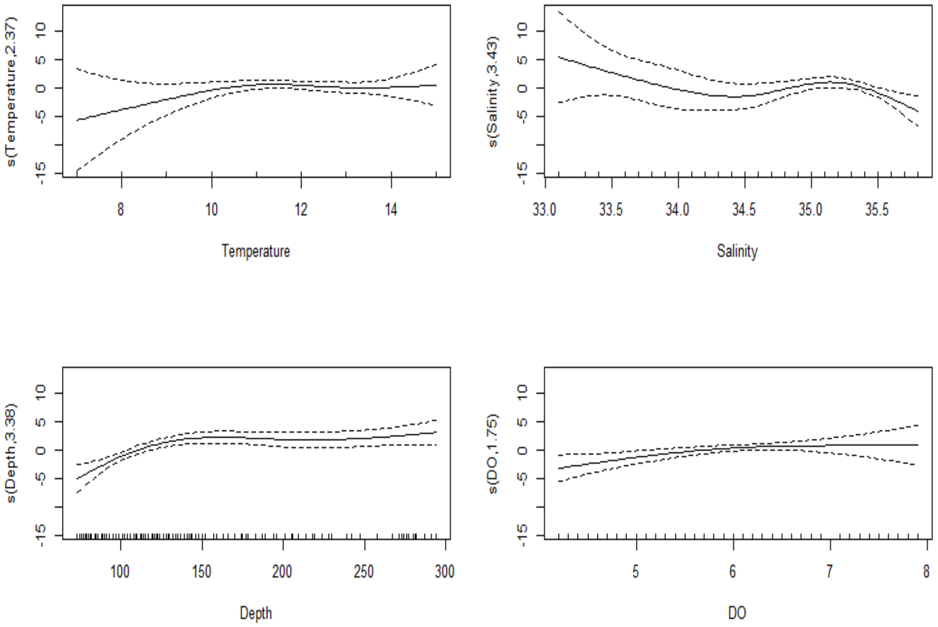


Figure 11. GAM plots identifying the effects of the variables from the optimal models on the probabilities of catch (ABUN) for Golden Tilefish from the Northwest Atlantic in 2017. Hatched lines represent 95% confidence intervals.

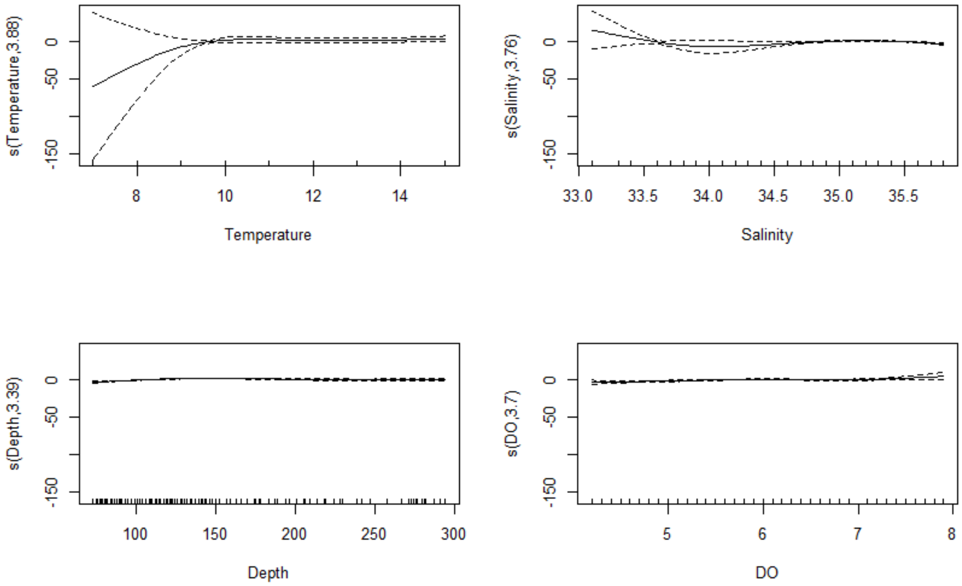


Figure 12. GAM plots identifying the effects of the variables from the optimal models on the probabilities of occurrence (PA) for Blueline Tilefish from the Northwest Atlantic in 2017. Hatched lines represent 95% confidence intervals.

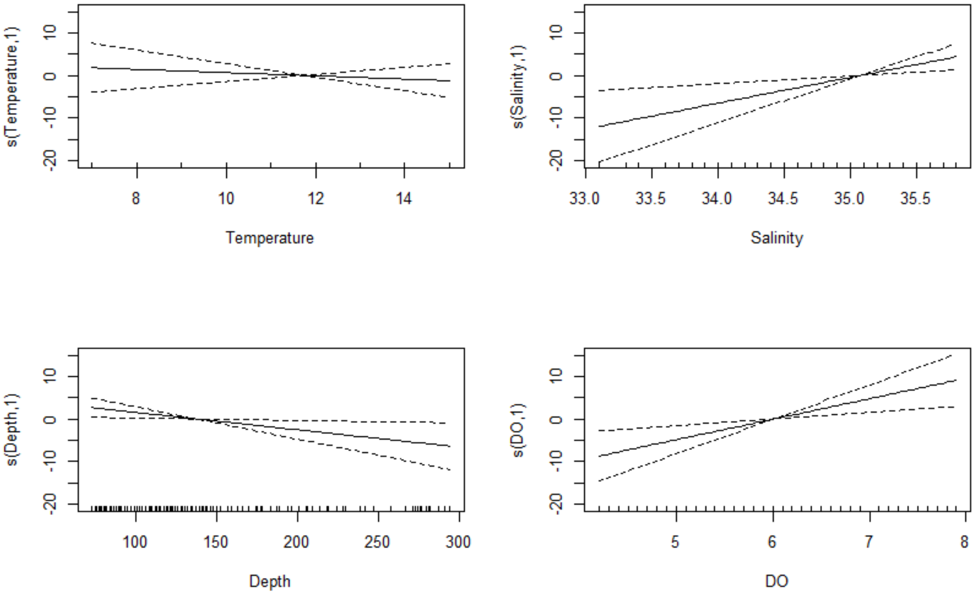
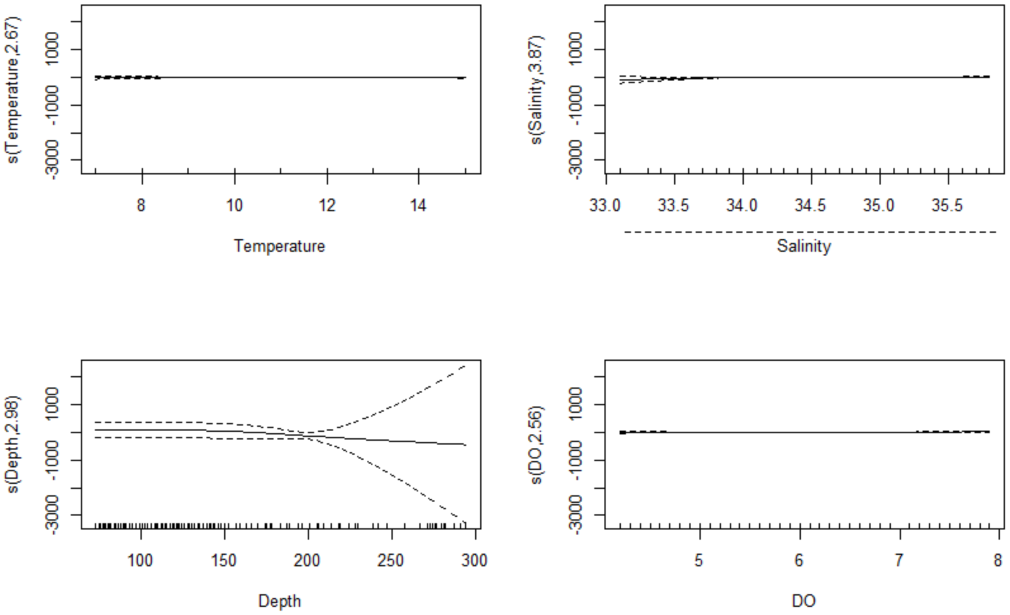


Figure 13. GAM plots identifying the effects of the variables from the optimal models on the probabilities of catch (ABUN) for Blueline Tilefish from the Northwest Atlantic in 2017. Hatched lines represent 95% confidence intervals.



Objective 5: *Evaluate proposed survey design (cost, proposed sampling intensity and statistical power).*

Sampling estimates

Sampling estimates were based on Cochran (1977). The mean catch per station for stratified random sampling ($\bar{y}_{st}$) was estimated as

$$\bar{y}_{st} = \sum_{h=1}^{L} W_h \bar{y}_h \tag{1}$$

where W_h is the fraction of the study area in stratum h ($h = 1,2,\ldots,L$) and

$$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h} \tag{2}$$

is the sample mean catch per station at stratum h. In the above equation, y_{hi} is the number of Tilefish caught in stratum h from sample i , with $i = 1,2,\ldots,n_h$. The number of samples collected in stratum h is n_h and the total number of samples in the survey is n, where

$$n = \sum_{h=1}^{L} n_h \ . \tag{3}$$

Assuming that finite population corrections can be ignored (i.e., the number of possible samples in each strata is large compared to the n_h, an unbiased estimate of the variance of $\bar{y}_{st}$ is

$$s^2(\bar{y}_{st}) = \sum_{h=1}^{L} \frac{W_h^2 s_h^2}{n_h} \tag{4}$$

where s_h^2 is the sample variance for stratum h. The quantity $s(\bar{y}_{st})$ is the standard error of the mean catch per station for stratified random sampling ($\bar{y}_{st}$). Assuming that $\bar{y}_{st}$ is approximately normally distributed, a confidence interval for this estimate may be computed as

$$CI = \bar{y}_{st} \pm t s(\bar{y}_{st}) \tag{5}$$

where t is the $\left(1 - \frac{\alpha}{2}\right)$ quantile of the t-distribution. Since no homogeneity of variance assumptions can be made for the s_h, the distribution of $s(\bar{y}_{st})$ is complex; however, an approximate solution for the confidence interval can be obtained by estimating an adjusted or “effective” degrees of freedom for t as (Cochran 1977; Satterthwaite 1946)

$$n_e = \frac{\left(\sum_{h=1}^{L} g'_h s_h^2\right)^2}{\sum_{h=1}^{L} \frac{\left(g'_h s_h^2\right)^2}{(n_h - 1)}} \tag{6}$$

where

$$g'_h = \frac{W_h^2}{n_h} \ . \tag{7}$$

The effective degrees of freedom n_e lies between the smallest $(n_h - 1)$ and the total degrees of freedom across all strata. The normality assumption for $\bar{y}_{st}$ made by Cochran (1977) is based on invoking the Central Limit Theorem for large sample sizes in each stratum. To check the adequacy of this approximation, the confidence interval was also estimated by bootstrapping the survey results (Hastie et al. 2001). Bootstrapped confidence intervals were generated using the "boot" package in R (Canty & Ripley 2017) and 5,000 replicates.

Optimum Allocation:

The efficacy of the survey was evaluated by comparing the uncertainty of the estimated $\bar{y}_{st}$ to alternative survey designs. One such alternative was optimal allocation where the financial cost of sampling is incorporated into the selection of the number of samples in each strata. The intent of this survey scheme is to balance statistical power, catch levels and financial cost. The optimum allocation approach was modified from Cochran (1977) by including a term reducing the cost of the survey by an amount equal to the value of the Tilefish sold to the market.

Cost was estimated as fixed costs plus sampling costs minus the wholesale value of the fish:

$$c = c_0 + \sum_{h=1}^{L} c_h n_h - \sum_{h=1}^{L} p_h \bar{y}_h n_h \tag{8}$$

where c_0 is fixed costs for supplies and travel between Montauk and the sampling area. This cost did not include the science party, their travel expenses, their supplies, overhead, etc., although these could be added as fixed costs if necessary. Fixed costs were taken to be \$5,000 for the vessel supplies required by the fishing effort plus two travel days (\$12,000).

Sampling cost was calculated as the sum over all strata of the cost per sample in stratum h, c_h, times n_h the number of samples in the stratum. The cost per sample was kept simple in the current analysis by allowing it to be constant across all stratum, and by taking it to be the day rate charter (\$6,000) divided by the average number of samples per day (8.8 per day from the survey after excluding the partial first and last sampling days). Therefore, c_h=\$682. This approach assumed that the sampling cost was dominated by the cost of sample collection and not the travel cost between sampling stations. Since the average distance between stations was only about 4.2 nmi, this simple cost estimate was thought to be adequate for the survey carried out.

The last term in equation (8) is the revenue generated by selling the Tilefish, where p_h is the wholesale price of a fish caught in stratum h. Although this value can vary by size and species, and therefore among strata, a constant value of \$10 per fish was assumed to be adequate for the analysis. The cost per fish is multiplied by the number of fish caught within a stratum ($\bar{y}_h n_h$) and then summed across all stratum.

The optimal proportion of samples in each strata was estimated by minimizing the product of equations (4) and (8) with respect to the n_h (Cochran 1977). This was done by differentiation with respect to n_h and setting the results to zero to yield:

$$\left[\frac{n_h}{n}\right]_{opt} = \frac{\frac{W_h s_h}{\sqrt{c_h - p_h \bar{y}_h}}}{\sum_{h=1}^{L} \frac{W_h s_h}{\sqrt{c_h - p_h \bar{y}_h}}} \tag{9}$$

The final step in optimum allocation was to determine n by either 1) choosing a fixed variance and estimating a minimum cost, or 2) choosing a fixed cost and estimating a minimum variance (Cochran 1977). The former was selected; equation (9) was substituted into equation (4) and solved for n, giving

$$n = \frac{1}{V}\left(\sum_{h=1}^{L} W_h s_h \sqrt{c_h - p_h \bar{y}_h}\right)\left(\sum_{h=1}^{L} \frac{W_h s_h}{\sqrt{c_h - p_h \bar{y}_h}}\right) \tag{10}$$

As a practical matter, V was determined by choosing a coefficient of variation ($cv = standard\ error/\bar{y}_{st}$) as a fraction of the mean (e.g., 0.05, 0.1, 0.2) and then calculating $V = (cv \cdot \bar{y}_{st})^2$.

Estimated Relative Precision of Survey Designs:

The fishing survey actually carried out was assessed by comparing the estimated precision of the survey relative to other potential sampling designs, each using the same sample size n. To make the comparisons easily interpretable, the estimated precision of the survey was expressed in the form of a coefficient of variation, using the statistics from equations (2) and (4):

$$cv_{survey} = \frac{s(\bar{y}_{st})}{\bar{y}_{st}}\,. \tag{11}$$

This estimated precision was compared to cv's from three different sampling designs:

$$cv_{ran} = \frac{se_{ran}}{\bar{y}_{st}} = \frac{\frac{s}{\sqrt{n}}}{\bar{y}_{st}} \tag{12}$$

where s is the sample standard deviation ignoring any stratification,

$$cv_{prop} = \frac{se_{prop}}{\bar{y}_{st}} = \frac{\sqrt{\frac{\sum_{h=1}^{L} W_h s_h^2}{n}}}{\bar{y}_{st}} \tag{13}$$

$$cv_{opt} = \frac{se_{opt}}{\bar{y}_{st}} = \frac{\sqrt{\frac{\left(\sum_{h=1}^{L} W_h s_h\right)^2}{n}}}{\bar{y}_{st}} \tag{14}$$

The three estimated cv's in equations (12), (13), and (14) are for simple random sampling, proportional allocation, and optimal allocation, respectively. It should be noted that the fishing survey carried out was not strictly proportional allocation since samples were added to strata to keep sample sizes above a minimum level, and the survey collected fewer than the planned number of samples. Thus, the comparison between the actual survey and proportional allocation is expected to be close but not identical. All estimates ignore the finite population correction. Presentation of the results will focus on comparing the three allocation designs being considered for future surveys: survey (pilot), proportional and optimal.

Evaluation of survey design:

Golden Tilefish Core Area

In the core area for Golden Tilefish (03-2, 03-3, 03-4, 04-2, 04-3, 04-4, 05-2, 05-3, 05-4), the estimated mean catch from stratified sampling ($\bar{y}_{st}$) was 5.34 individuals per line with a standard error ($s(\bar{y}_{st})$) of 0.72 individuals per line. A total of 94 stations were sampled within the core area. The coefficient of variation ($s(\bar{y}_{st})$/ $\bar{y}_{st}$) was 0.14. An approximate 95% confidence interval for the mean was [3.90, 6.78] individuals per line based on the t-distribution with 65 effective degrees of freedom. The field survey cost (equation 8) for this area was about $75K, with about $81K in sampling costs and $6K in revenue.

The lower and upper bounds of the 95% bootstrapped confidence intervals differed by 1.3% or less from the t-distribution approximation, and the bootstrap analysis generally supported the use of this approximation. Estimated mean catch (5.34 individuals per line) and standard error (0.70) from the bootstrap analysis were almost identical to the survey estimates. The distribution of the bootstrap estimates was reasonably symmetric (Figure 14), although its shape was somewhat leptokurtotic (Figure 15), and hence it differed from normal (Shapiro-Wilk normality test, W = 0.995, *P* = 0.0015).

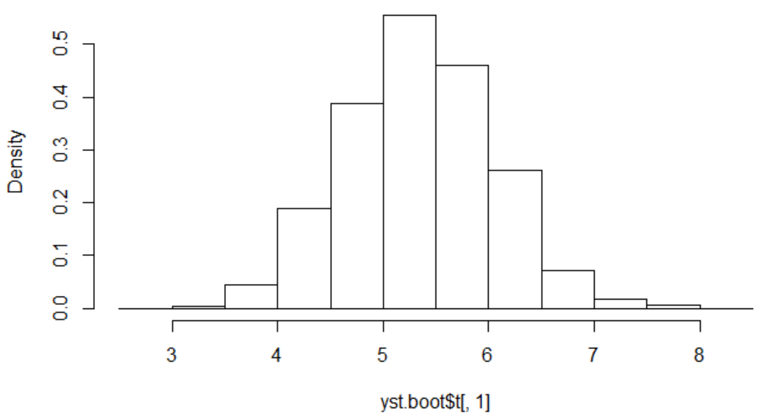


Figure 14. The distribution of bootstrap estimates of mean catch

An optimum allocation strategy suggested increasing the fraction of samples in strata 03-3, 04-3, and 05-3 relative to the survey and proportional allocation (Figure 16). These were strata with the largest mean catches per line, standard deviations, and potential revenue. Under an optimum allocation scheme, the coefficient of variation could be reduced from 0.14 to 0.12 for the current survey of 94 samples (Table 7). Revenue from the sale of Golden Tilefish would increase from $6K to almost $8K, and this revenue would cover 9.6% of the survey cost. To lower the coefficient of variation to 0.10 (i.e., equivalent to a 95% confidence interval that was about ±20% of the mean), would require increasing sampling effort by 46% (Figure 17) and the sampling cost by about 32% (Figure 18).

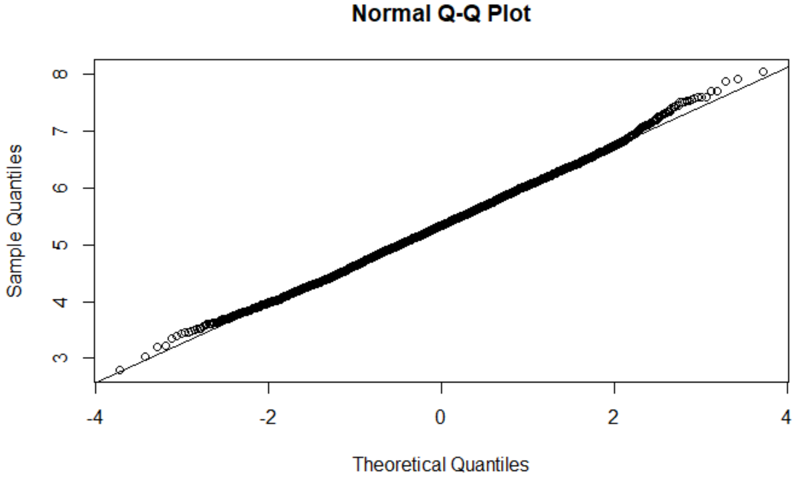


Figure 15. Quantile-Quantile plot comparing the bootstrapped estimates of mean catch from stratified sampling to the normal distribution in the Golden Tilefish core area.

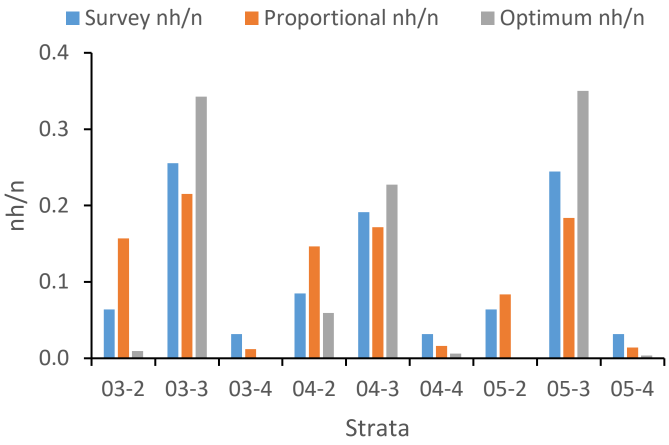


Figure 16. Survey, proportional, and optimum sampling fractions (nh/n) for the Golden Tilefish core area.

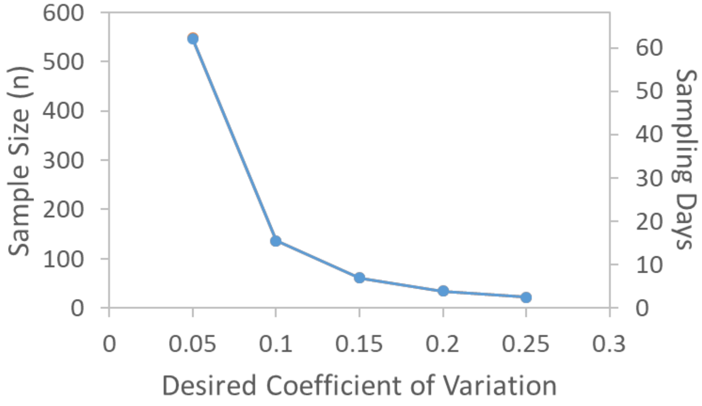


Figure 17. Estimated sample size required to obtain a desired coefficient of variation with optimum allocation for the Golden Tilefish core area.

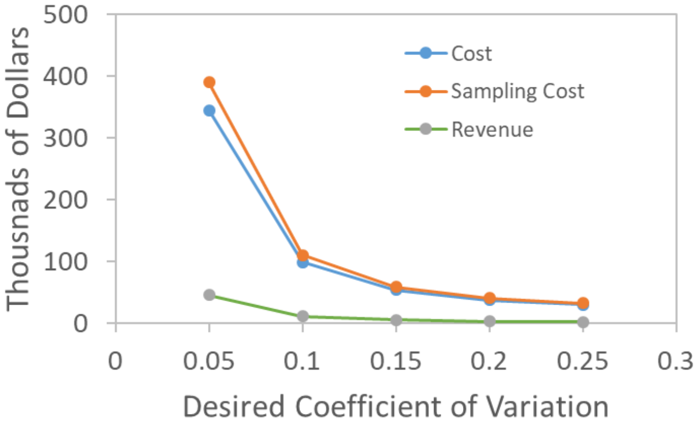


Figure 18. Estimated cost to obtain a desired coefficient of variation with optimum allocation for the Golden Tilefish core area. Cost is estimated as fixed costs plus sampling costs minus the wholesale value of the fish.

Table 7. Golden Tilefish core area survey results, sampling fractions (n_h/n) and estimated coefficients of variation for the survey, proportional allocation, and optimum allocation. Estimates assume a sample size n=94 and mean $\bar{y}_{st}$=5.34.

Strata	W_h (relative area)	Survey $\bar{y}_h$	Survey s_h	Survey n_h/n	Proportional Allocation n_h/n	Optimum Allocation n_h/n
03-2	0.157	0.17	0.41	0.053	0.157	0.005
03-3	0.215	7.04	10.04	0.266	0.215	0.259
03-4	0.012	0.33	0.58	0.032	0.012	0.001
04-2	0.147	1.38	2.67	0.085	0.147	0.034
04-3	0.172	9.67	8.17	0.191	0.172	0.217
04-4	0.016	2.33	2.52	0.032	0.016	0.004
05-2	0.084	0.00	0.00	0.064	0.084	0.000
05-3	0.184	11.70	11.70	0.245	0.184	0.477
05-4	0.014	1.73	1.73	0.032	0.014	0.002
$\bar{y}_{st}$				5.34	5.34	5.34
n				94	94	94
standard error (se)				0.72	0.80	0.64
coefficient of variation (cv)				0.14	0.15	0.12

Blueline Tilefish Core Area

In the core area for Blueline Tilefish (07-2, 07-3, 08-2, 08-3, 09-2, 09-3), the estimated mean catch from stratified sampling ($\bar{y}_{st}$) was 2.34 individuals per line with a standard error ($s(\bar{y}_{st})$) of 1.00 individual per line. A total of 32 stations were sampled within this core area. The coefficient of variation ($s(\bar{y}_{st})/\ \bar{y}_{st}$) was 0.43. An approximate 95% confidence interval for the mean was [-0.11, 4.79] individuals per line based on the t-distribution with 7 effective degrees of freedom. The field survey cost (equation 8) for this area was about $38K, with about $39K in sampling costs and less than $1K in revenue.

While the estimated mean catch from the bootstrap analysis (2.34 individuals per line) agreed with the survey, suggesting little bias in the bootstrap mean, the bootstrap estimated standard error (0.86) was substantially lower than the survey estimate (1.00). The distribution of the bootstrap estimates is slightly skewed and clearly leptokurtotic (Figures 19, 20), and it differed from normal (Shapiro-Wilk normality test, W = 0.996, P < 0.0001). These outcomes are the result of a

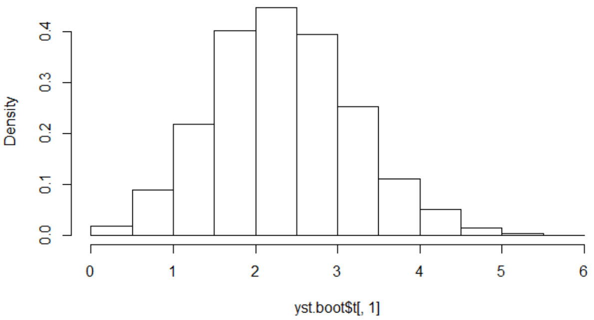


Figure 19. The distribution of bootstrap estimates of mean catch from stratified sampling ($\bar{y}_{st}$) in the Blueline Tilefish core area.

small number of samples (32) spread out over 6 strata and with nonzero catch in only 8 of the 32 samples. Overall, the assumption that $\bar{y}_{st}$ is approximately normally distributed is suspect, and any confidence interval generated from the Blueline catch data is unreliable.

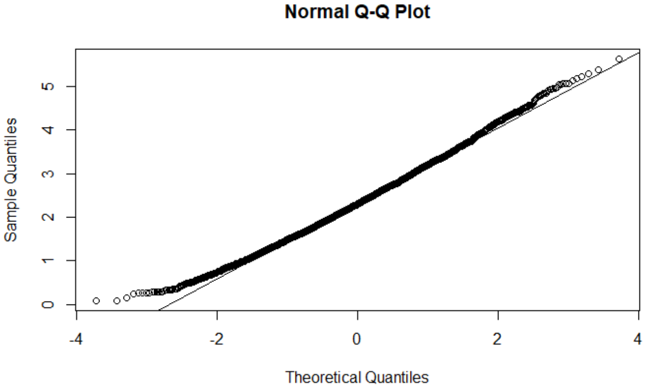


Figure 20. Quantile-Quantile plot comparing the bootstrapped estimates of mean catch from stratified sampling to the normal distribution in the Blueline Tilefish core area.

An optimum allocation strategy suggested increasing the fraction of samples in strata 07-3, 08-2, and 09-3 relative to the survey and proportional allocation (Figure 21). With the exception of one individual fish, these were the only strata in this core area (07-2, 07-3, 08-2, 08-3, 09-2, 09-3) where Blueline were caught. Under an optimum allocation scheme, the coefficient of variation could be reduced from 0.43 to 0.29 for the current survey of 32 samples (Table 8). Revenue from the sale of Blueline Tilefish would increase slightly from $630 to $1370, and this revenue would only cover 3.5% of the sampling cost. To lower the coefficient of variation to 0.10 (i.e., equivalent to a 95% confidence interval that was about ±20% of the mean), would require increasing sampling effort by 860% (Figure 22) and the sampling cost by about 500% (Figure 23). Presumably, a greatly increased sampling effort would also produce an approximately normally distributed $\bar{y}_{st}$.

Table 8. Blueline Tilefish core area survey results, sampling fractions (n_h/n) and estimated coefficients of variation for the survey, proportional allocation, and optimum allocation. Estimates assume a sample size n=32 and mean $\bar{y}_{st}$=2.34.

Strata	W_h (relative area)	Survey $\bar{y}_h$	Survey s_h	Survey n_h/n	Proportional Allocation n_h/n	Optimum Allocation n_h/n
07-2	0.097	0.00	0.00	0.094	0.097	0.000
07-3	0.403	2.00	4.38	0.344	0.403	0.447
08-2	0.184	3.67	6.35	0.094	0.184	0.300
08-3	0.143	0.17	0.41	0.188	0.143	0.015
09-2	0.086	9.67	10.26	0.094	0.086	0.238
09-3	0.087	0.00	0.00	0.188	0.087	0.000
$\bar{y}_{st}$				2.34	2.34	2.34
n				32	32	32
standard error (se)				1.00	0.87	0.68
coefficient of variation (cv)				0.43	0.37	0.29

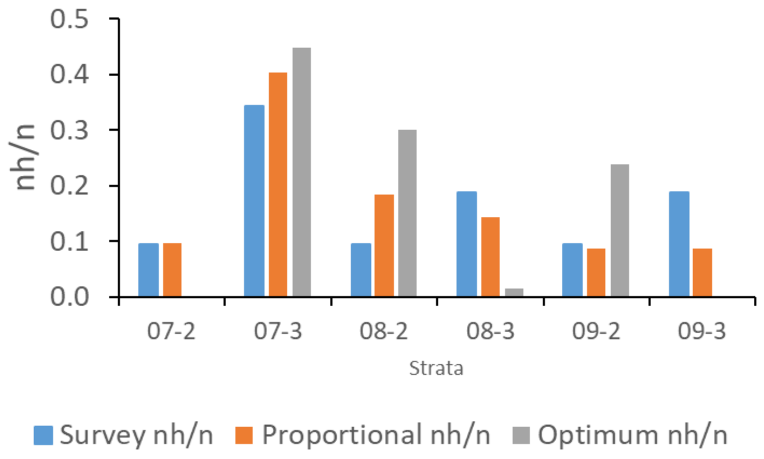


Figure 21. Survey, proportional, and optimum sampling fractions (nh/n) for the Blueline Tilefish core area.

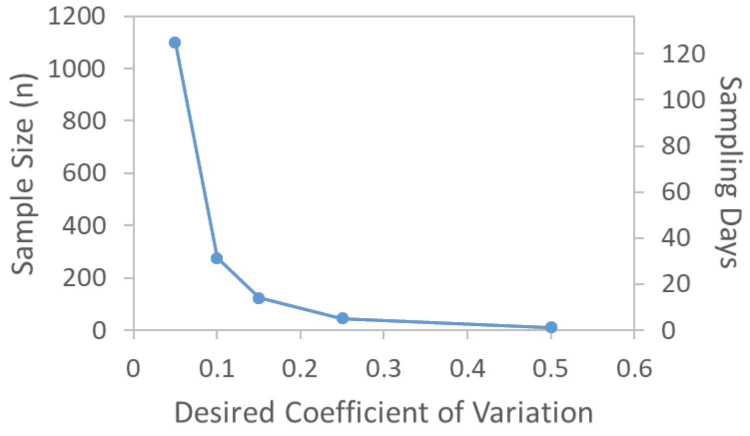


Figure 22. Estimated sample size required to obtain a desired coefficient of variation with optimum allocation for the Blueline Tilefish core area.

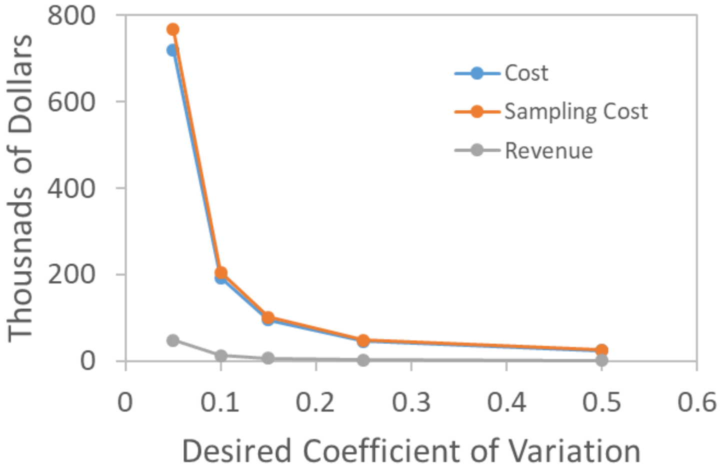


Figure 23. Estimated cost to obtain a desired coefficient of variation with optimum allocation for the Golden Tilefish core area. Cost is estimated as fixed costs plus sampling costs minus the wholesale value of the fish.

Golden + Blueline Combined for All Strata

With the catch of Golden and Blueline combined and with all strata, the estimated mean catch from stratified sampling ($\bar{y}_{st}$) was 3.06 individuals per line with a standard error ($s(\bar{y}_{st})$) of 0.39 individuals per line. A total of 194 stations were sampled. The coefficient of variation ($s(\bar{y}_{st})/\,\bar{y}_{st}$) was 0.13. An approximate 95% confidence interval for the mean was [2.29, 3.83] individuals per line based on the t-distribution with 72 effective degrees of freedom. The field survey cost (equation 8) for this area was about $142K, with about $132K in sampling costs and $7K in revenue.

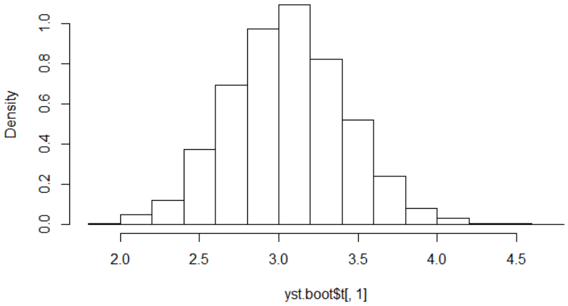


Figure 24. The distribution of bootstrap estimates of mean catch for Golden + Blueline Tilefish from stratified sampling ($\bar{y}_{st}$) in the study area.

The lower and upper bounds of the 95% bootstrapped confidence intervals differed by 2.7% or less from the t-distribution approximation, and the bootstrap analysis generally supported the use of this approximation. Estimated mean catch (3.06 individuals per line) and standard error (0.37) from the bootstrap analysis were almost identical to the survey estimates. The distribution of the bootstrap estimates was reasonably symmetric (Figure 24), although its shape was somewhat leptokurtotic (Figure 25), and hence it differed from normal (Shapiro-Wilk normality test, W = 0.999, *P* = 0.0015).

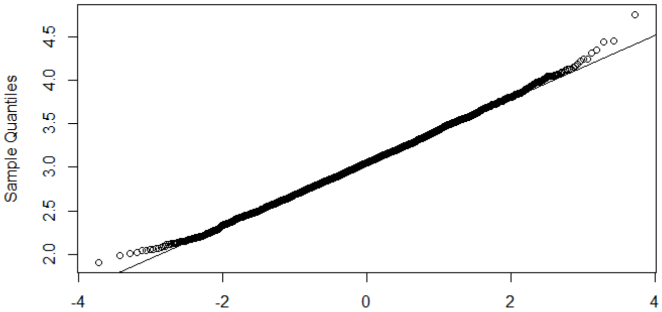


Figure 25. Quantile-Quantile plot comparing the bootstrapped estimates of mean catch for Golden + Blueline Tilefish from stratified sampling to the normal distribution in the study area.

An optimum allocation strategy suggested increasing the fraction of samples in strata 03-3, 04-3, 05-3, 07-3, 08-2, and 09-2 relative to the survey and proportional allocation (Figure 26). This outcome essentially combines the two individual Tilefish core area sampling strategies. It would essentially eliminate sampling in 13 of the 29 strata sampled (02-3, 03-1, 04-1, 05-1, 05-2, 06-1, 06-2, 06-3, 07-1, 07-2, 08-1, 08-4, 09-4) since no Tilefish were caught in these strata. Under this optimum allocation scheme, the coefficient of variation could be reduced from 0.13 to 0.09 for the current survey of 194 samples (Table 9). Revenue from the sale of Golden Tilefish would increase from $7K to almost $15K, and this revenue would cover 10.0% of the survey cost. A coefficient of variation of 0.10 (i.e., equivalent to a 95% confidence interval that was about ± 20% of the mean) could be obtained by reducing sampling effort by 24% (Figure 27) and the sampling cost by about 12% (Figure 28).

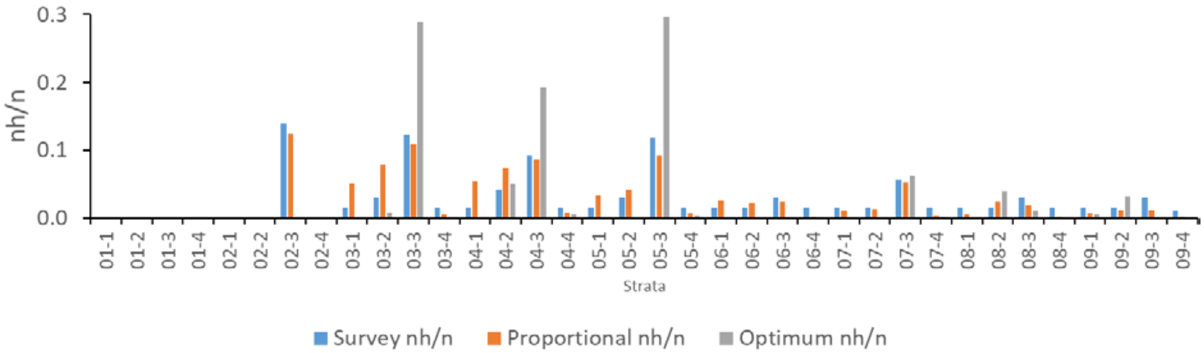


Figure 26. Survey, proportional, and optimum sampling fractions (nh/n) for the combined Golden + Blueline Tilefish survey area.

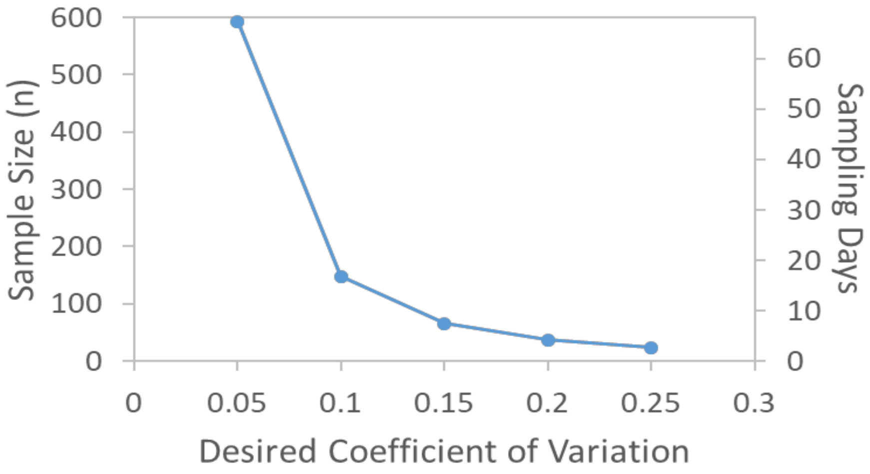


Figure 27. Estimated sample size required to obtain a desired coefficient of variation with optimum allocation for the combined Golden + Blueline Tilefish survey area.

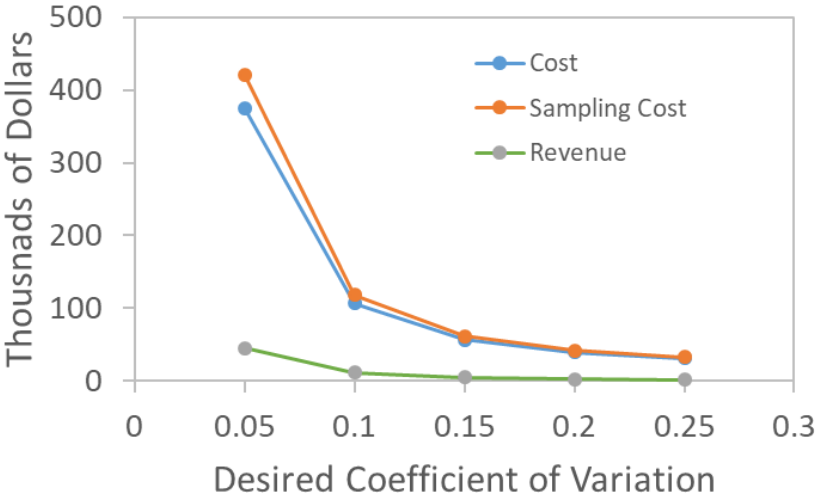


Figure 28. Estimated cost to obtain a desired coefficient of variation with optimum allocation for the combined Golden + Blueline Tilefish survey area. Cost is estimated as fixed costs plus sampling costs minus the wholesale value of the fish.

Table 9. Combined Golden + Blueline Tilefish survey results, sampling fractions (n_h/n) and estimated coefficients of variation for the survey, proportional, and optimum allocation. Estimates assume a sample size n = 194 and mean $\bar{y}_{st}$=3.06. Note that no samples were collected in strata 01-1, 01-2, 01-3, 01-4, 02-1, 02-2, and 02-4, so their $\mathrm{W_h}$ values were set to zero.

Strata	$\mathrm{W_h}$ (relative area)	Survey $\bar{\mathrm{y}}_\mathrm{h}$	Survey $\mathrm{s_h}$	Survey $\mathrm{n_h/n}$	Proportional Allocation $\mathrm{n_h/n}$	Optimum Allocation $\mathrm{n_h/n}$
01-1	0.000	0.00	0.00	0.000	0.000	0.000
01-2	0.000	0.00	0.00	0.000	0.000	0.000
01-3	0.000	0.00	0.00	0.000	0.000	0.000
01-4	0.000	0.00	0.00	0.000	0.000	0.000
02-1	0.000	0.00	0.00	0.000	0.000	0.000
02-2	0.000	0.00	0.00	0.000	0.000	0.000
02-3	0.125	0.00	0.00	0.139	0.125	0.000
02-4	0.000	0.00	0.00	0.000	0.000	0.000
03-1	0.052	0.00	0.00	0.015	0.052	0.000
03-2	0.079	0.17	0.41	0.031	0.079	0.008
03-3	0.108	7.04	10.04	0.124	0.108	0.290
03-4	0.006	0.33	0.58	0.015	0.006	0.001
04-1	0.054	0.00	0.00	0.015	0.054	0.000
04-2	0.074	1.38	2.67	0.041	0.074	0.050
04-3	0.086	9.72	8.20	0.093	0.086	0.193
04-4	0.008	2.33	2.52	0.015	0.008	0.005
05-1	0.033	0.00	0.00	0.015	0.033	0.000
05-2	0.042	0.00	0.00	0.031	0.042	0.000
05-3	0.092	10.30	11.73	0.119	0.092	0.297
05-4	0.007	1.00	1.73	0.015	0.007	0.003
06-1	0.025	0.00	0.00	0.015	0.025	0.000
06-2	0.021	0.00	0.00	0.015	0.021	0.000
06-3	0.025	0.00	0.00	0.031	0.025	0.000
06-4	0.002	0.33	0.58	0.015	0.002	0.000
07-1	0.011	0.00	0.00	0.015	0.011	0.000
07-2	0.013	0.00	0.00	0.015	0.013	0.000
07-3	0.053	2.18	4.60	0.057	0.053	0.062
07-4	0.003	1.67	2.89	0.015	0.003	0.002
08-1	0.006	0.00	0.00	0.015	0.006	0.000
08-2	0.024	3.67	6.35	0.015	0.024	0.040
08-3	0.019	1.67	2.25	0.031	0.019	0.011
08-4	0.002	0.00	0.00	0.015	0.002	0.000
09-1	0.006	3.00	3.00	0.015	0.006	0.005
09-2	0.011	9.67	10.26	0.015	0.011	0.032
09-3	0.011	0.17	0.41	0.031	0.011	0.001
09-4	0.002	0.00	0.00	0.010	0.002	0.000

$\bar{y}_{\mathrm{st}}$				3.06	3.06	3.06
n				194	194	194
standard error (se)				0.39	0.42	0.27
coefficient of variation (cv)				0.13	0.14	0.09

The survey as implemented was successful from a statistical standpoint for the Golden core area and overall survey, in that it generated an estimated coefficient of variation that was between that expected for proportional and optimum allocation; however, the coefficient of variation for the Blueline tilefish core area was somewhat worse than random sampling. This result for the Blueline area was due to both low sample size (n = 32) in the core area, and the fact that Blueline tilefish were caught at only 10 of those 32 sampling stations. Estimated precision of the stratified mean was always better with proportional and optimum allocation over random sampling, suggesting that future surveys at least maintain depth stratified proportional allocation.

A coefficient of variation of 10% (i.e., 95% CI of ± 20%) is an achievable goal for survey precision in the Golden core area and the overall survey, but not for a survey in the Blueline core area without substantially increasing sampling effort. Improving precision would require shifting sampling effort from strata where few or no tilefish where caught to strata with the greatest catch (03-3, 04-3, 05-3, 07-3, 08-2, 09-2). For the Golden core area, both optimum allocation and a modest increase of sampling effort by 46% would be required to obtain a coefficient of variation of 10%. For the overall survey, optimum allocation alone with no increased sampling effort would be sufficient to obtain a coefficient of variation of 10%. For the Blueline tilefish core area, optimum allocation along with an 860% increase of sampling effort would be required. This would require increasing sampling effort in this core area alone from n = 32 to n = 275, and this is probably an unrealistic goal. It is important to note that these estimates are for static tilefish populations and will be sensitive to any changes in their geographic distribution or abundance. In addition, decreasing sampling in strata with no catch during the current survey could bias geographic range estimated during future surveys.

Revenue generated by the sale of tilefish caught during the survey can offset the sampling cost by 2-7% depending on the area surveyed. The highest revenue generated was in the Golden core area ($6K), and the lowest non-zero revenue was in the Blueline tilefish core area ($630). One curious feature of the optimum allocation analysis is that average wholesale value of the fish caught per sample in a strata ($p_h\bar{y}_h$) cannot equal or exceed the cost per sample in the strata (c_h). If this occurs for any strata, the terms in the numerator and denominator of equation (9) for n_h/n blow up or become imaginary. If this were to occur at one strata, the remaining terms in the denominator become negligible and $n_h/n \rightarrow 1$ for that strata and zero elsewhere. Therefore, to break even or make a profit from sampling, the optimum allocation solution is to employ the fishers' strategy, i.e., mainly ignore all strata except for the one that breaks even or is profitable.

It should be noted that sampling cost in equation (8) was assumed to be dominated by the cost of sample collection and not the travel cost between sampling stations. Since the average distance between stations was only about 4.2 nmi, this simple cost estimate was thought to be adequate for the survey carried out. The alternative would have been to subdivide sampling costs into separate sample collection and travel costs (sampling costs = sample collection costs + between sample travel costs). Travel would then have to be estimated for different sample sizes by linear programming methods that solve the "traveling salesman problem" (i.e., the shortest path through many points). There is a considerable literature on various algorithms to solve this problem numerically (e.g., Beardwood et al. 1959; Lawler et al. 1985; Gutin & Punnen 2002), and there is at least one R library (Hahsler & Hornik 2012) to do these calculations. Had the average distances between stations been 10-20 nmi, then travel time would have been important to consider separately.

***Survey Design Recommendations*:**

1.0 Considering statistical and biological concerns we recommend that future surveys continue with proportional sampling (i.e., survey (pilot) or proportional allocation designs) of the ‘expanded’ range at a similar effort level and regional coverage sampled in the pilot survey. The design resulted in reasonable CV’s and uncertainty ranges for abundance estimates. Cost savings resulting from the optimum sampling do not out-weight the benefits of sampling a geographic range that extends to depths outside of each species core range. If future surveys employ the optimal allocation strategy it lowers the ability to detect range expansions or contractions. This is important for species that are distributed in an extremely patchy manner. Continual evaluation of survey design could be used to reduce the geographic range sampled as additional data is obtained; however, given analysis of the pilot survey we feel a proportional design similar to the pilot survey is recommended. This is also supported by observations of the fishing community that have noted small Tilefish in the shallow depths that the optimal strategy removes.

2.0 A smaller scaled survey targeting Golden Tilefish could also be successfully employed at a much lower cost and produce reasonable CVs by utilizing any of the three evaluated designs (Survey (pilot), proportional or optimal). Here again, we recommend continuing the ‘expanded’ range to detect potential distributional shifts. The pilot survey results for Blueline Tilefish did not provide adequate data to evaluate the best survey design.

3.0 The pilot survey did not produce a large amount of revenue from sold fish. However, we recommend that future surveys continue to sell Tilefish to offset survey costs for two primary reasons: first, some years may produce large revenues and second discarded fish have very low survival and would be wasted.

4.0 We recommend the continuation of using the three hook sizes in order to track cohorts and inform assessment models (i.e., domed shaped catchability).

5.0 The current project benefited from one unpaid participant and future surveys will require one additional person to assist is cruise and data analyses. Considering the current implementation of the survey this could be achieved by a graduate student and a modest increase in PI effort.

6.0 The spatially comprehensive data collected from this pilot survey is valuable for the design of a future potential long-term industry based survey under the desired goals for indexing either both tilefish species or an individual tilefish stock with the known funding constraints.

References:

Akaike H. 1973. Information theory as an extension of the maximum likelihood principle. In: Petrov BN, Csaki F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, Hungary, 267–281 pp.

Beardwood J, Halton JH, Hammersley JM. 1959. The shortest path through many points. Proceeding of the Cambridge Philosophical Society 55: 299-327.

Bland JM, Altman DG. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 327(8476):307–310.

Brotons L, Thuiller W, Araujo MB, Hirzel AH. 2004. Presence–absence versus presence-only modelling methods for predicting bird habitat suitability. Ecography, 27(4):437–448.

Burnham KP, Anderson DR. 2002. Model selection and multimodel inference: A practical information—theoretic approach. Springer-Verlag, New York, NY, 488 pp.

Canty A, Ripley B. 2017. boot: Bootstrap R (S-Plus) Functions. R package version 1.3-20

Cochran WG. 1977. Sampling Techniques, 3rd edition. Wiley and Sons, New York, NY, 428 pp.

Dunton KJ, Jordaan A, McKowen KA, Conover DO, Frisk MG. 2010. Abundance and distribution of Atlantic sturgeon (*Acipenser oxyrinchus*) within the Northwest Atlantic Ocean, determined from five fishery-independent surveys. Fisheries Bulletin 108:450–465.

Fewster RM, Buckland ST, Siriwardena GM, Baillie SR, Wilson JD. 2000. Analysis of population trends for farmland birds using generalized additive models. Ecology 81(7):1970–1984.

Freeman A, Moisen G. 2008. Presence-Absence: An R Package for Presence Absence Analysis. Journal of Statistical Software 23:11.

Guisan A, Edwards TC, Hastie T. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157:89–100.

Gutin G, Punnen AP (eds.). 2002. The Traveling Salesman Problem and Its Variations, Volume 12 of Combinatorial Optimization. Kluwer, Dordrecht.

Hahsler M, Hornik K. 2012. TSP - Infrastructure for the Traveling Salesperson Problem. Journal of Statistical Software 23: 1-21.

Hastie TJ, Tibshirani RJ. 1990. Generalized Additive Models. Chapman & Hall, New York, NY, 352 pp.

Heinanen S, Ronka M, Von Numers M. 2008. Modelling the occurrence and abundance of a colonial species, the arctic tern *Sterna paradisaea* in the archipelago of SW Finland. Ecography 31(5):601–611.

Hurvich CM, Tsai C. 1989. Regression and time series model selection in small samples. Biometrika 76:297–307.

Idelberger CF. 1985. Reproductive biology of the Tilefish, *Lopholatilus chamaeleonticeps,* in the Middle Atlantic Bight. M.S. thesis, Rutgers University, New Brunswick, New Jersey, 51 pp.

Johnson JB, Omland KS. 2004. Model selection in ecology and evolution. Trends in Ecology and Evolution 19:101–108.

Lawler EL, Lenstra JK, Rinnooy Kan AHG, Shmoys DB (eds.). 1985. The Traveling Salesman Problem. Wiley, New York.

Leathwick JR, Elith J, Hastie T. 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modelling of species distributions. Ecological Modelling 199(2):188–196.

Pearce J, Ferrier S. 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling 133(3):225–245.

Perry RI, Smith SJ. 1994. Identifying habitat associations of marine fishes using survey data: an application to the Northwest Atlantic. Canadian Journal of Fisheries and Aquatic Sciences 51(3): 589–602.

Potts JM, Elith J. 2006. Comparing species abundance models. Ecological Modelling 199(2):153–163.

R Core Development Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J, Muller M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12:77. r-project.org/web/packages/mgcv/index.html.

Sagarese SR, Frisk MG, Miller TJ, Sosebee KA, Musick JA, Rago PJ. 2014. Influence of environmental, spatial, and ontogenetic variables on habitat selection and management of spiny dogfish in the Northeast (US) shelf large marine ecosystem. Canadian Journal of Fisheries and Aquatic Sciences 71:567–580.

Satterthwaite FE. 1946. An approximate distribution of estimates of variance components. Biometrics 2:110–114.

Swets JA. 1988. Measuring the accuracy of diagnostic systems. Science 240(4857):1285–1293.

Wintle BA, Elith J, Potts JM. 2005. Fauna habitat modelling and mapping: A review and case study in the Lower Hunter Central Coast region of NSW. Austral Ecology 30(7):719–738.

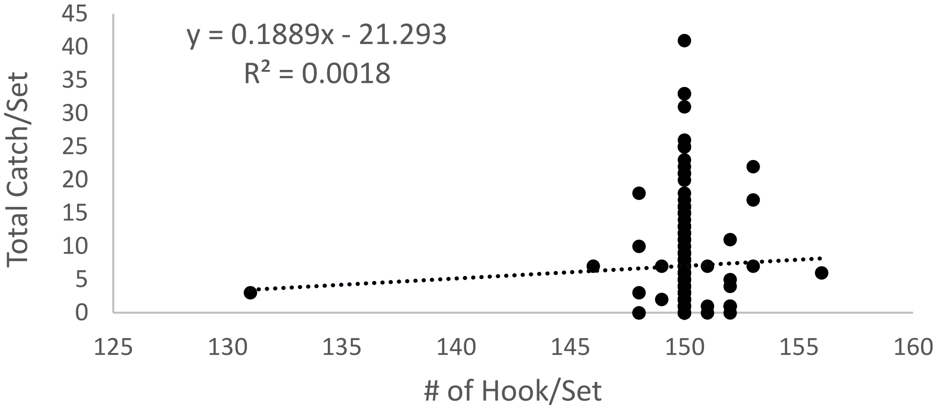
Wood SN. 2006. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC, Boca Raton, Florida.

Wood SN. 2011. Package ‘mgcv’. [online]. Available from http://cran.

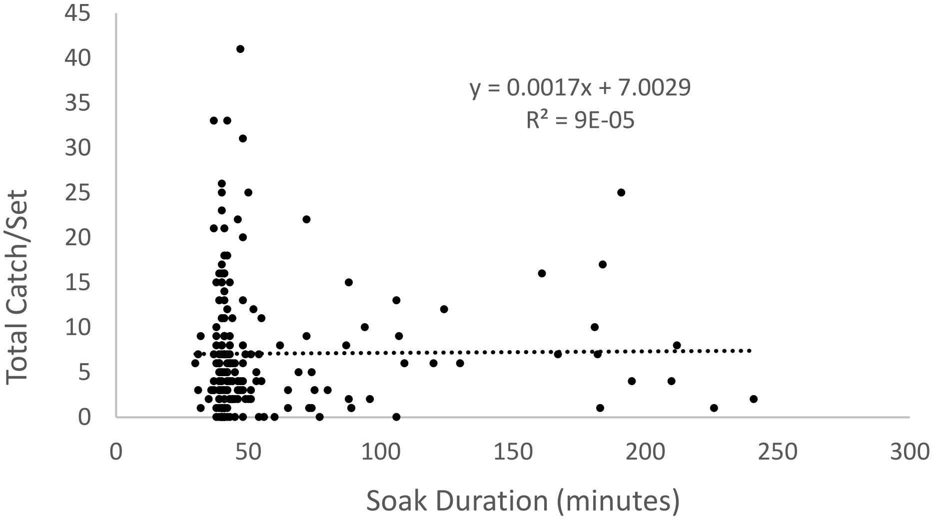
Yee TW, Mitchell ND. 1991. Generalized additive models in plant ecology. Journal of Vegetation Science 2:587–602.

Supplemental Section:

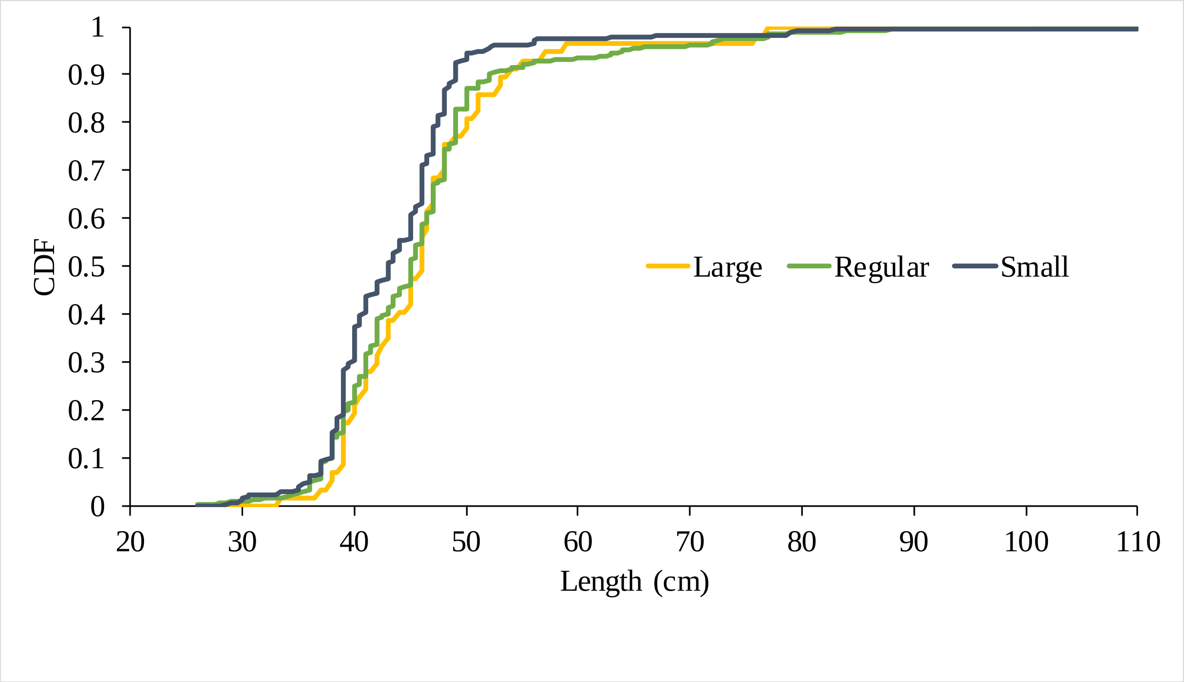
Supplemental Figure 1: Relationship between number of hooks per set and total number of fish caught per set.



Supplemental Figure 2. Relationship between the soak duration and the total number of fish caught per set.



Supplemental Figure 3. Cumulative distributions of small, regular and large hook sizes.



Supplemental Table 1. Taxa and number of individuals of each taxon collected by hook size in the survey.

		Hook Size			
Species	**Common Name**	**Small**	**Regular**	**Large**	**Total**
Centropristis striata	Black Sea Bass	3	5	1	9
Helicolenus dactylopterus	Black Bellied Rose		1	1	2
Caulolatilus microps	Blueline Tilefish	20	40	8	75
Scyliorhinus retifer	Chain Dogfish	13	37	2	52
Congridae	Conger Eel	2	5		7
Carcharhinus obscurus	Dusky Shark		1	1	2
Lopholatilus chamaelonticeps	Golden Tilefish	238	323	58	619
Myxine glutinosa	Hagfish	2			2
Cancer borealis	Jonah Crab	1	7	3	11
Leucoraja erinacea	Little Skate	4	16	4	24
Scomber scombrus	Mackerel	1			1
Coryphaena hippurus	Mahi Mahi		1		1
Carcharhinus signatus	Night Shark			1	1
Merluccius albidus	Offshore Hake	12	7	2	21
Echeneidae	Remora		1		1
Sphyrna lewini	Scalloped Hammerhead			1	1
Mustelus canis	Smooth Dogfish	1	3	1	5
Carcharhinus brevipinna	Spinner Shark		1		1
Squalus acanthias	Spiny Dogfish	1	3		4
Urophycis regia	Spotted Hake	167	344	41	552
Paralichthys dentatus	Summer Flounder	1			1
	Total	**466**	**802**	**124**	**1392**

Supplemental Table 2. Catch-weighted estimates for Tilefish.

Strata	Area (km^2)	% Survey area (Wh)	Proposed	Actual (nh)	Wh/nh	Blueline Catch	Blueline Weighted	Golden Catch	Golden Weighted
01--1	433.3	1.2	0	0					
01--2	589.4	1.7	0	0					
01--3	817.3	2.3	0	0					
01--4	91.1	0.3	0	0					
02--1	1168.3	3.3	0	0					
02--2	2653.5	7.5	0	0					
02--3	3684.9	10.4	30	27	0.38				
02--4	237.3	0.7	0	0					
03--1	1519.1	4.3	3	3	1.43				
03--2	2320.7	6.5	10	5	1.31			1	1
03--3	3184.3	9.0	26	25	0.36			169	470
03--4	177.3	0.5	3	3	0.17			1	6
04--1	1592.4	4.5	3	3	1.50				
04--2	2167.4	6.1	10	8	0.76			11	14
04--3	2538.4	7.2	20	18	0.40	1	3	174	437
04--4	240.7	0.7	3	3	0.23			7	31
05--1	977.5	2.8	3	3	0.92				
05--2	1236.1	3.5	6	6	0.58				
05--3	2720.4	7.7	22	23	0.33	2	6	235	704
05--4	208.6	0.6	3	3	0.20			3	15
06--1	734.9	2.1	3	3	0.69				
06--2	630.7	1.8	3	3	0.59				
06--3	727.7	2.1	6	6	0.34				
06--4	57.3	0.2	3	3	0.05			1	19
07--1	314.6	0.9	3	3	0.30				
07--2	374.1	1.1	3	3	0.35				
07--3	1551.0	4.4	12	10	0.44	22	50	2	5
07--4	98.0	0.3	3	4	0.07			5	72
08--1	182.9	0.5	3	4	0.13				
08--2	708.0	2.0	3	3	0.67	11	17		
08--3	550.2	1.6	5	5	0.31	1	3	9	29
08--4	62.2	0.2	3	3	0.06				
09--1	191.4	0.5	3	3	0.18	9	50		
09--2	331.9	0.9	3	5	0.19	29	155		
09--3	336.1	0.9	5	4	0.24			1	4
09--4	48.1	0.1	3	2	0.07				

Photos:

Photo 1: Crew bringing in first haul (J. Nolan; B. Davis; A. Ellis; S. Doyle; P. Nitschke).

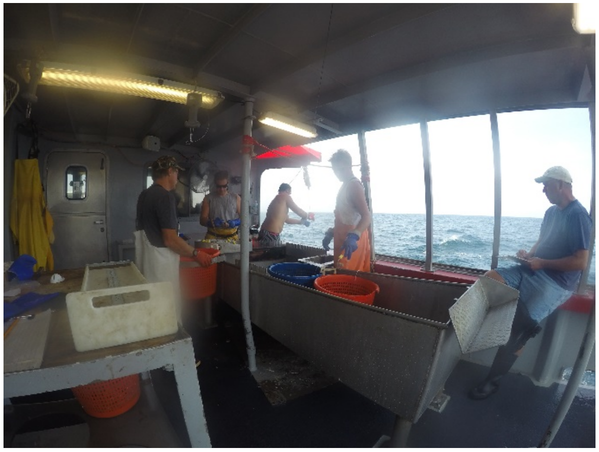


Photo 2: Catch ready for data collection.

Photo 3: Golden Tilefish caught in survey (A. Smith; J. Olin).