Modeling Juvenile Atlantic cod and yellowtail flounder abundance on Georges Bank and in the Gulf of Maine using 2-stage generalized additive models

A final report to the Closed Area Technical Team of the New England Fisheries Management Council
April 12, 2013

Prepared by
Samuel Truesdell
University of Maine, Orono ME, 04469
207-581-4405
samuel.truesdell@maine.edu

Sections

1. EXECUTIVE SUMMARY

1.1 Value of the models to management

Management strategies, especially for species or communities in changing ecosystems, should be grounded in ecology; in other words regulations should be designed considering ecological explanations. Statistical models that estimate the combined effects of such explanations (i.e. that use ecological variables) are thus a natural fit to serve as a foundation for management. However, because of the complexity and nonlinear nature of natural environments, especially flexible models are often necessary to explain the relationships observed within these systems. Generalized additive models are in many cases well suited for use in these situations because they are highly adaptable and unbounded by the linear assumptions of traditional statistical models, and we use this class of model here to explain relationships between juvenile groundfish and their habitat. The outputs from the additive models include the linear or nonlinear relationships between each of the explanatory variables and the model response, the residuals for sampled locations, and the predicted values at those locations.

The generalized additive models are able to identify important habitat characteristics that can be used by managers, but they are constrained to the available variables and the statistical assumptions of the models. These models together with empirical methods like the spatial cluster analyses that were conducted separately by members of the New England Fisheries Management Council provide a useful parallel examination of juvenile groundfish habitat; the value of this parallel process lies in that the approaches are different. The cluster analyses are completely observational and thus represent a thoroughly empirical technique for identifying critical habitat, and although they cannot explain ecological associations or processes (useful in the backing of management decisions) like the generalized additive models they provide an excellent check on the soundness of the additive models. The analysis of groundfish critical habitat benefits greatly from the combination of these two approaches.

1.2 Short summary of findings

The final generalized additive models were decided upon using a backwards selection algorithm (section 3) beginning with a full model including physical and environmental variables such as depth, bottom characteristics, temperature, and zenith angle. Once a final model was developed it was evaluated using model diagnostics, the critical habitat variables were identified, and predictions were produced.

The habitat variables that (qualitatively) proved most important in determining the distribution of the juvenile groundfish stocks we examined were depth and bottom temperature and both had generally negative effects on abundance (i.e. expected abundance decreased with increasing depth or temperature). Season, sediment, and the shape of the seabed were also important, but the particular effects were not as consistent across the stocks (and in the case of sediment could not be compared across all three). Zenith angle was also an important variable for standardizing catch in some cases; it can remove variation in fish catchability that is related to circadian rhythms.

Juvenile cod on Georges Bank were predicted to occur mostly off Cape Cod, in the Great South Channel, and along the northern edge of Georges. In the Gulf of Maine the region of highest expected juvenile cod catch was in Massachusetts Bay, and elsewhere the model predicted the highest abundances along the Maine coast. High predictions for Georges Bank yellowtail were scattered, though they were more common on the southeast part of Georges and in the Nantucket Lightship area.

2. MODELING RATIONALE

Two-stage generalized additive models were used to describe the relationship between the explanatory variables and the counts of juvenile groundfish.

2.1 Generalized additive models

We used generalized additive models because of their flexibility which is often a critical attribute when describing ecological phenomena. This class of model is an extension of generalized linear models in that they can accept the various error distributions from within the exponential family and the explanatory variables are related to the predicted value through a "link function." The difference is that the additive models are capable of including nonlinear effects, so no assumption of linearity is required when relating the model terms to the response. Within the modeling process the relationships between the continuous variables and the response are described by nonlinear smooth functions, so each of these relationships can change across values of the continuous independent variables.

2.2 Two-stage models

An oft-encountered difficulty in modeling fisheries data is the presence of an excessive number of zeros. If the ratio of zeros to non-zeros is too large then the response cannot be modeled effectively using a common error distribution. Various strategies exist for dealing with this problem but the one we used was a two-stage model. Two models were developed: one estimating the simple presence or absence of a species and another modeling the data conditional on presence. Predictions can be made by multiplying the expected values of the two models together.

For the presence-absence model we used a binomial error distribution and for the conditional presence model we logged the response and used a Gaussian error distribution with an identity link function, meaning that we assumed the residuals to be distributed normally and used no transformation between the scale of the model fitting and the scale of the response.

3. MODEL SELECTION ALGORITHM

Final candidate models were found using a backwards-selecting algorithm that employs a combination of likelihood ratio tests and model significance p-values to choose reasonable models.

3.1 Details of the model selection algorithm

Each iteration of the model selection algorithm has four steps. They are:

(1) Begin with a full model with *n* terms.

(2) Remove each model term one-at-a-time, creating *n* new models with *n*-1 terms each.

(3) Use a likelihood ratio test to determine which of the sub-models provides the least new information (i.e. which likelihood ratio test of sub-model against the full model is the least significant; this identifies which term adds the least to the model's explanatory power).

(4) Remove that term and use the rest as an updated "full" model.

This algorithm is repeated until two conditions are met:

(1) All model terms are significant based on the specified p-value significance threshold for significant model terms; and

(2) Removing any of the remaining terms produces a significant model difference based on the specified p-value significance threshold for the likelihood ratio tests.

3.2 Rationale for p-value thresholds

P-value significance thresholds for both the model term significance and the likelihood ratio tests were set at p=0.25. With respect to the model term significance, this generous threshold ensures that even marginally significant variables are retained in the final model. Should any of these variables be considered unimportant or unusable for management they are easily discarded and the model can be updated. Similarly, the relatively high threshold p-value for the likelihood ratio tests encourages the algorithm to stop when only marginally significant differences are found because it is easier for two models to be significantly different when the p-value is set relatively high.

We selected "generous" p-value thresholds because we did not want the selection algorithm to remove variables that were important even in a very small way; this selection is better left as a qualitative analysis by experts in juvenile groundfish ecology.

3.3 Interaction terms

Interaction terms were not included in the saturated model that fed into the backwards selection algorithm. Already there were many single terms in the model relative to the amount of data, especially for the presence models on Georges Bank (only 176 data points). Since each categorical variable removes at least two degrees of freedom and each continuous variable in these models typically used between 1 and 7 degrees of freedom, including interaction terms at the start often led to candidate models that were not possible to run.

We did, however, manually include interaction terms after the algorithm was complete. We chose each set of significant terms in the final model and added them to the saturated model singly and evaluated their significance. We used a less generous significance threshold of 0.05 for interaction terms because they are more difficult to explain and thus to justify for inclusion in management measures. None of these terms had p-values lower than 0.05 and so none were included in the final models. We did not use likelihood ratio tests for interaction term models.

3.4 Likelihood ratio as opposed to AIC

The algorithm used likelihood ratio tests as opposed to AIC (Akaike Information Criterion). The difference is that AIC includes a penalty for the number of parameters estimated in the model. In this case we were not particularly interested in the most parsimonious model, which is why we set our model term significance and likelihood ratio test p-value thresholds high at 0.25. Since these models will be used or adapted by managers who have an expert understanding of the biology of the species we felt the best approach was to err on the side of a more inclusive model that could be reduced further if need be. AIC encourages parsimony and so would risk removing important terms.

4. VARIABLES

The response variables for the binomial additive models were the presence/absence of juvenile cod or yellowtail flounder and for the count models the response was the logged tow abundance. Juvenile cod were defined as those less than or equal to 35cm in fall and 25cm in spring, while juvenile yellowtail were defined as less than or equal to 15cm year-round.

The candidate variables to explain variability in the catch of juvenile cod and yellowtail were:

(1) Bottom temperature: collected from survey tows;

(2) Average tow depth: collected from survey tows;

(3) Seabed Form: A combination of slope and "Land Position Index" from TNC that indicates the type of bottom e.g. "depression" or "high slope;"

(4) Dominant sediment type: from Harris and Stokesbury (2010) with categories such as mud and sand [available on Georges Bank only];

(5) Sediment coarseness: indicates the grain size of the sediment (Harris and Stokesbury 2010) [available on Georges Bank only];

(6) Shear stress: benthic boundary layer shear stress from Harris et al. 2012 [available on Georges Bank only];

(7) Substrate: categorical variable indicating substrate type from TNC

(8) Season: spring or fall;

(9) Purpose code: indicates what survey the data come from (spatial and seasonal survey coverage may be found in appendix 2); and

(10) Zenith angle: can help account for diel behavioral changes in catchability (courtesy L. Jacobson and J. Tang; http://nefsc.noaa.gov/publications/crd/crd1114/index.html).

The substrate variable (7) overlaps with substrate oriented variables on Georges Bank from Harris and Stokesbury (2010; 4-5) and so was not used for the Georges Bank data since the resolution was coarser. However, this finer scale sediment data along with shear stress (6) were not available outside Georges Bank, so the coarse sediment data were used to model Gulf of Maine cod. Additional information on the variables can be found in tables 1-4 of appendix 1.

5. MODELING RESULTS AND INTERPRETATION

The data, models, predictions and diagnostics for all three stocks are summarized below.

5.1 Georges Bank cod

5.1.1 Data

The general saturated model for Georges Bank cod was:

$$\hat{J} = SEA + PC + SBF + SD + s(SC) + s(STR) + s(T) + s(Z) + s(D)$$

Where SEA is season, PC is purpose code (survey type), SBF is seabed form, SD is dominant sediment type, SC is sediment coarseness, STR is shear stress, T is temperature, Z is zenith angle at tow-time, and D is depth. $\hat{J}$, the expected value of the response, was zero or one for the presence-absence model and the logged measured juvenile abundance for the conditional presence model.

Before the modeling stage began, all these data were investigated to examine their relationship with juvenile abundance and check for outliers. Figures including histograms for the variables and plots of each against total juvenile abundance and abundance conditioned on presence may be found in appendix 3. The available data, including the proportion of positive tows are in Fig. 1. The resolution of the grid in Fig. 1, as in all the similar figures including residual plots is 0.09 x 0.09 min., or approximately 10 km^2 (referenced in the north-south direction).

Cooperative research surveys for goosefish and cod (purpose codes 4 and 5) were excluded for this analysis because these surveys had little overlap with the regions of interest on Georges Bank; the goosefish survey was excluded because there was only one positive tow in the overlapping area, and the cod survey excluded because there were only 3 tows overall in the region (Table 1).

5.1.2 Correlations among continuous variables

No variables were removed from the cod data set based on their correlation. The one potential candidate was to remove either sediment coarseness or shear stress. While the relationship was clear and positive there was still considerable variability within the overall correlation (Fig. 2). Both terms were left in the model. Both shear stress and coarseness remained in the final model and since coarseness was only marginally significant it may be reasonable to remove this term from the final model.

5.1.3 Model results

5.1.3.1 Presence-absence model

Following model selection, the significant terms for the presence-absence model were purpose code, season, sediment coarseness, shear stress, zenith, temperature and average depth. Shear stress and zenith angle were marginally significant, but the rest had p-values less than 0.01 (Table 3). There were 901 data points used and the model explained 31.8% of the deviance.

Spring had a negative effect on the probability of presence and the Massachusetts Department of Marine Fisheries survey (purpose code 11) had a positive effect relative to the NFMS bottom trawl survey (purpose code 10). The model output smooth plots for the continuous variables are given in figure 13. They show sediment coarseness to have a positive linear effect; shear stress to have a negative effect between values of 1 and 3; bottom temperature to have a highly negative almost linear effect; zenith angle to have a slightly positive linear effect; and depth to have a positive effect between approximately 5 to 35 meters and then a strong negative effect between depths of about 35 to 80 meters. A general summary of the effects are given in tables 2 and 3 and the smooth plots for continuous variables are given in Fig. 3.

Model diagnostics (Fig. 4) showed the presence-absence model to be somewhat reasonable (for an ecological data set). The residuals and quantiles showed a slightly skewed distribution that lacks small positive values and has too many small negative values. The high number of small negatives probably comes from observed values of zero and very small predictions. While the observed data are actual discrete counts, since the model expected values are not they are unlikely to predict a response of exactly zero. But since they predict close to zero, when the residuals are calculated (observed minus predicted) the result is an overrepresentation of residuals that are negative but close to zero.

5.1.3.2 Conditional presence model

The conditional presence model proved to explain much less variance at only 6.11%. The only significant effect in the model was shear stress and it was marginal at p = 0.03 (Table 3). The effect was negative and linear, so expected abundance decreased with increasing shear stress, but the residuals show much scatter around the trend line (Fig. 5). Season and purpose code were forced into the model as standardizing variables though neither were statistically significant. Spring had a negative effect relative to fall and the Massachusetts Department of Marine Fisheries survey (purpose code 11) had a positive effect relative to the NFMS bottom trawl survey (purpose code 10). There were many fewer observations available for the conditional model, with only 176 locations. A summary of the effects is given in tables 2 and 3.

The conditional presence model had mixed diagnostics (Fig. 6). There was some skew in the residuals and some increasing variance in the residuals versus linear predictors but these patterns were not overly concerning. On the other hand the plot of the response versus fits (each observation plotted against its fitted value) indicates that the model does not fit particularly well.

5.1.3.3 Residuals

Spatial plots of residuals and standardized residuals (residual divided by the mean) are provided for the final output, i.e. the product of the presence-absence and conditional presence models, for each scenario. These types of residual plots are an important diagnostic for ecological data sets with a spatial component. They show the range of the departure from the expected values; but, more importantly, they indicate whether there are spatial patterns in the residuals. Spatial patterns in the residuals indicate that there are likely to be other important variables that are not defined in the model.

The Georges Bank cod residuals are generally positive on the western part, especially around Cape Cod, and negative across the rest of Georges Bank (Figs. 7 and 8). This indicates that there are other sources of variability within the models that are not taken into account and that cause this spatial pattern in the residuals.

5.1.4 Predictions

The overall predictions (Fig. 9) for Georges Bank cod show the highest expected abundance off Cape Cod and east of Nantucket throughout the Great South Channel. There are also higher predicted values along the northern edge of Georges Bank. Throughout the rest of the area the predictions are mostly mixed, but typically predict an expected survey catch of less than one fish per tow.

The spring and fall predictions (Figs. 10 and 11) also show concentrations around Cape Cod and in the Great South Channel. They differ, however, in that on Georges Bank itself in the spring the model predicts relatively more cod in the center of the bank area while in the fall they are confined to the outskirts.

5.2 Gulf of Maine cod

5.2.1 Data

The general saturated model for Gulf of Maine cod was:

$$\hat{J} = SEA + PC + SBF + SED + s(T) + s(Z) + s(D)$$

Where SEA is season, PC is purpose code (survey type), SBF is seabed form, SED is sediment type, T is temperature, Z is zenith angle at tow-time, and D is depth. $\hat{J}$, the expected value of the response, was zero or one for the presence-absence model and the logged measured juvenile abundance for the count model.

Before the modeling stage began, all these data were investigated to examine their relationship with juvenile abundance and check for outliers. Figures including histograms for the variables and plots of each against total juvenile abundance and abundance conditioned on presence may be found in appendix 4.

Only the cooperative research goosefish survey (purpose code 4) was excluded for this analysis; it was eliminated because there were zero positive tows, again due to lack of overlap with the region of interest. The other data sets had reasonable numbers of positive records (Table 4). The spatial distribution of the data we used, including where juvenile cod were actually caught, is given in Fig. 12.

5.2.2 Correlations among continuous variables

While some trends are evident in the relationships among continuous variables for the Gulf of Maine cod data, there is too much variability to warrant any exclusion among the one relationship that is approximately linear on average, zenith angle and depth (Fig. 13). All continuous variables were retained for the saturated model.

5.2.3 Model results

5.2.3.1 Presence absence model

The variables that best explain the presence of juvenile cod were sediment type, seabed form, temperature and depth; all these p-values were less than 0.01 (Table 3). The model explained 20.7% of the deviance and was based on 4030 data points. Out of the sediment types, mud had a very negative effect and the smallest sand category as well as the largest sand category also had negative effects though they was weaker. The “high flat” seabed form category had a strong positive effect, as did the high slope. Relative to the Maine-New Hampshire inshore trawl survey (purpose code 1), the industry-based cod cooperative survey (purpose code 5) had a positive effect, the NMFS bottom trawl survey (purpose code 10) had a negative effect, and the Massachusetts Department of Marine Fisheries survey (purpose code 11) had a positive effect. Only this final survey was statistically different from the Maine-New Hampshire survey. Season insignificant, but spring had a negative effect relative to fall. Temperature and depth both had highly significant, negative effects on abundance (Table 3; Fig. 14). The temperature effect shows a sharp decline at values less than about five, followed by a more gradual decline between 5 and 11 degrees, then a steeper decline again at temperatures higher than 11 (though there is relatively less data at these higher temperatures). On average, abundance is highest at depths between approximately 0 and 80 meters, then declines rapidly after that. The partial residuals (the residuals with respect to a single term after the intercept and the effects of the other model terms have been removed; Wood 2006), however, show two modes: one being this decline and another (much smaller) an increase in abundance with depth (Fig. 14). These residuals were mapped but there was no obvious spatial pattern that would explain the second mode.

Similarly to the Georges Bank cod residuals, the Gulf of Maine presence-absence residuals show a break in the distribution at small positive values (Fig. 15). Otherwise the residuals are fairly normal. The response against the fits show more misclassifications than with the Georges Bank cod model; especially there were more fitted values close to 1 (expected presence) where in fact juveniles were absent in the observed data set.

5.2.3.2 Conditional presence model

The conditional presence model explained only 11.3% of the deviance, and was based on 1277 data points. Most important to describing the abundance of cod in this model were sediment type, temperature, depth and season. Mud had a negative effect on measured juvenile abundance, while large and medium sand sizes had a positive, marginally significant effect (Tables 2 and 3). Spring had a highly significant, positive effect and the effect of large-sized sand was also positive. Relative to the Maine-New Hampshire inshore trawl survey (purpose code 1), the industry-based cod cooperative survey (purpose code 5), the NMFS bottom trawl survey (purpose code 10), and the Massachusetts Department of Marine Fisheries survey (purpose code 11) each had negative effects. Temperature and depth again both had significant effects (Table 3). Abundance increased slightly with temperature from 0 to 10 degrees, then showed a marked decline, though there were only very few data points above 10 degrees. The depth effect was slightly negative and linear, and zenith remained in the model but the effect direction was not clear (Fig. 16).

Residuals for the conditional presence model are not entirely symmetrical about zero but do not indicate a concerning departure from normality (Fig. 17). The residuals against the linear predictor do not show terribly increasing variance, but again the response versus fitted values leaves much to be desired as the trend is barely discernible.

5.2.3.3 Residuals

The residuals and standardized residuals show underpredictions in Massachusetts Bay and in eastern Maine and generally slight overpredictions across the rest of the sample area (Figs. 18 and 19).

5.2.4 Predictions

The 2-stage model predicts most juvenile cod in the Gulf of Maine to be found close to the coast and on Stellwagen Bank (Fig. 20). There is also a cluster of positive predictions in the eastern Gulf of Maine at the edge of the sampling area. Unlike for the Georges Bank juvenile cod, the spring and fall predictions in the Gulf of Maine do not appear to differ measurably (Figs. 21 and 22).

5.3 Georges Bank Yellowtail Flounder

5.3.1 Data

The general saturated model for Georges Bank yellowtail was:

$$\hat{J} = SEA + PC + SBF + SD + s(SC) + s(STR) + s(T) + s(Z) + s(D)$$

Where SEA is season, PC is purpose code (survey type), SBF is seabed form, SD is dominant sediment type, SC is sediment coarseness, STR is shear stress, T is temperature, Z is zenith angle at tow-time, and D is depth. $\hat{J}$, the expected value of the response, was zero or one for the presence-absence model and the logged measured juvenile abundance for the conditional presence model.

Before the modeling stage began, all these data were investigated to examine their relationship with juvenile abundance and check for outliers. Figures including histograms for the variables and plots of each against total juvenile abundance and abundance conditioned on presence may be found in appendix 5.

All surveys except the NMFS bottom trawl and Massachusetts Marine Fisheries trawl (purpose codes 10 and 11) were excluded for this analysis. The most positive records (77) came from the NMFS survey, so despite the low ratio of tows in which yellowtail were actually caught it was included (Table 5). The Massachusetts Marine fisheries survey had a small sample size at 75, but 20% of those tows caught juvenile yellowtail. The spatial distribution of the data we used, including where juvenile yellowtail flounder were actually caught, is given in Fig. 23.

5.3.2 Correlations among continuous variables

These data were almost identical to those used in the Georges Bank cod analysis, and so the same description follows as found in section 5.1.2. No variables were removed from the cod data set based on their correlation. The one potential candidate was to remove either sediment coarseness or shear stress. While the relationship was clear and positive there was still considerable variability within the overall correlation (Fig. 24). Both terms were left in the model.

5.3.3 Model results

5.3.3.1 Presence-absence model

The presence-absence model explained 23.3% of the variance and was based on 915 sample locations. Spring had a positive and significant effect as did zenith angle (Tables 2 and 3; Fig. 25). The Massachusetts Department of Marine Fisheries survey (purpose code 11) had a positive effect relative to the NMFS bottom trawl survey (purpose code 10). Seabed form, sediment coarseness and depth all remained in the model although their significance was only marginal, though a small positive effect was noted for “high flat” areas relative to depressions. Sediment coarseness increased slightly across values less than about 2.2 and decreased slightly at values larger than about 2.5 but these effects were small. Estimated abundance increased slightly with depth until about 85 meters, after which it declined. Zenith angle had a highly significant, positive, almost linear effect indicating that more yellowtail are caught at night. Season also had a highly significant, posiotive effect.

The model produced close to no residuals between zero and one using these data, indicating that it is not doing a sufficient job capturing the variability in the response. Large observations are underpredicted leading to the cluster of positive residuals greater than one. Many zero catches were slightly overpredicted which results in the skewed count between zero and negative one (Fig. 26). Extreme outliers are evident in the plot of residuals against the linear predictor and there are almost no locations that predict presence at a probability greater than 0.5. The poor model diagnostics question both the model predictions and the effects of the significant variables.

5.3.3.2 Conditional presence model

The conditional presence model explained 52.9% of the variance using 90 tow locations where juveniles were caught. The unfixed terms remaining in the model were sediment coarseness, temperature, and depth (Table 3). The standardizing variable season had a negative though non-significant effect for spring relative to fall, and the Massachusetts Division of Marine Fisheries survey (purpose code 11) had a negative effect relative to the NFMS bottom trawl survey (purpose code 10). The temperature effect was marginally significant and positive between 4 and 7 degrees where most of the data lay, and then declined at higher values. The depth effect was significant (Table 3) and negative linear and sediment coarseness was also significant but inconclusive in direction (Fig. 27).

The diagnostics for this model were much better (Fig. 28). The residuals appear normally distributed and no patterns are evident in the plot of residuals against the linear predictor. The fitted values look to be highly correlated with the response. However, due to the small number of data points it is possible (and perhaps likely) that this model is overspecified and the diagnostics are misleading. Care should be taken that the overall predictions are closely examined to be sure they are realistic.

5.3.3.3 Residuals

No spatial patterns are particularly evident in the residuals for yellowtail on Georges Bank (Figs. 29 and 30). There seems to be some underprediction just off the northern tip of Cape Cod (more evident in the standardized residuals; Fig. 30), but other than that no clustering is evident.

5.3.4 Predictions

The overall model predictions for Georges Bank yellowtail are somewhat scattered at this scale of spatial grouping (Fig. 31). The clusters, though they are not very tight, look to be in the Nantucket Lightship area and on the eastern part of Georges Bank. There are scattered high predictions in the Great South Channel and elsewhere on Georges Bank. Some clusters of positive tows on eastern Georges Bank and in the Nantucket Lightship area are visible in spring (Fig. 32), but the patterns look somewhat more random in fall (Fig. 33).

6. EXPLANATION OF APPENDICES

Appendix 1 is an extension of section 4 and contains additional information about the candidate variables and their sources. The tables were prepared by M. Bachman.

Appendix 2 shows the spatial and seasonal distribution of the fisheries surveys that were used in the modeling. These figures were prepared by M. Bachman.

Appendices 3-5 contain preliminary analyses for each of the stocks. Included are (1) Histograms for those candidate variables that are continuous; (2) barplots for those that are discrete; (3) scatterplots with loess smooths for each continuous variable against the logged juvenile counts for all tows and also for only the tows in which juveniles of the species were present; and (4) boxplots of logged juvenile counts conditioned on each category of the discrete variables also for both all tows and only the tows where juveniles of the species were present.

Appendix 6 contains the generalized additive model output from R (package mgcv).

7. ACKNOWLEDGEMENTS

The author would like to thank Larry Jacobson and Jiashen Tang for providing the code we used to calculate zenith angle, Jui-Han Chang for playing the primary role in developing the model selection algorithm and both Jui-Han and Stefan Zhand for their comments on the analyses. Members of the Closed Area Technical Team's comments on 3/7/2013 were extremely helpful, and comments on drafts of this report by Andy Applegate and Michelle Bachman of the New England Fisheries Management Council were much appreciated. Michelle Bachman prepared the tables in appendix 1 and the figures in appendix 2.

SOURCES CITED

Harris, BP and KDE Stokesbury. 2010. The spatial structure of local surficial sediment characteristics on Georges Bank, USA. Continental Shelf Research 30: 1840-1853.

Harris, BP, GW Cowles and KDE Stokesbury. 2012. Surficial sediment stability on Georges Bank, in the Great South Channel and on eastern Nantucket Shoals. Continental Shelf Research 49: 65-72.

Wood, SN. 2006. Generalized additive models an introduction with R. Taylor & Francis : New York, 391 pp.

Table 1: Tow counts for all survey types in the Georges Bank cod data set.

Data Type	Purpose Code			
	4	5	10	11
Conditional Presence	1	2	144	48
All data	56	3	983	72
Ratio	0.018	0.67	0.15	0.67

Table 2: Summary of parameter effects for all models. +/++ = positive/very positive effect; -/-- = negative/very negative; ~ = complicated spline relationship; 0 = significant term but spline relationship questionable. Purpose code is not included because it is too inconsistent across the various data sets; since different data sets were used for each analysis the effects are not meaningful as a comparison.

Variable	(Relative to)	GB Cod		GOM Cod		GB Yellowtail	
		P/A	P	P/A	P	P/A	P
DEPTH		— —		— —	—	0	— —
TEMPERATURE		— —		— —	~		~
ZENITH		+			0	+ +	
Sed Coarseness		+ +		NA		0	~
Shear Stress		—	—	NA			
Season – Spring	Fall	— —			+ +	+ +	
SB Form – High Flat	Depression			+ +		+	
SB Form – High Slope	Depression			+ +			
SB Form – Low Slope	Depression						
SB Form – Mid Flat	Depression						
SB Form – Side Slope	Depression						
Dominant Sed – Sand	Silt/Mud			NA			
Dominant Sed – Pebble	Silt/Mud			NA			
Dominant Sed – Cobble	Silt/Mud			NA			
Dominant Sed – Boulder	Silt/Mud			NA			
Sediment – SandXL	Gravel	NA		—			NA
Sediment – SandLarge	Gravel	NA			+		NA
Sediment – SandMed	Gravel	NA					NA
Sediment – SandSmall	Gravel	NA		—			NA
Sediment – Silt/Mud	Gravel	NA		— —	—		NA

Table 3: P-values for the effects included in the final models. Purpose code is not divided into separate categories because the categories vary by data set, so the minimum p value relative to the reference level is reported.

Variable	(Relative to)	GB Cod		GOM Cod		GB Yellowtail	
		P/A	P	P/A	P	P/A	P
DEPTH		<0.001		<0.001	<0.001	0.043	0.006
TEMPERATURE		<0.001		<0.001	<0.001		0.032
ZENITH		0.0034			0.098	<0.001	
Sed Coarseness		<0.001		NA	NA	0.063	0.001
Shear Stress		0.098	0.027	NA	NA		
Season – Spring	Fall	<0.001	0.380	0.113	<0.001	<0.001	0.242
SB Form – High Flat	Depression			<0.001		0.018	
SB Form – High Slope	Depression			0.022		0.919	
SB Form – Low Slope	Depression			0.764		1	
SB Form – Mid Flat	Depression			0.132		0.109	
SB Form – Side Slope	Depression			0.870		1	
Dominant Sed – Sand	Silt/Mud			NA	NA		
Dominant Sed – Pebble	Silt/Mud			NA	NA		
Dominant Sed – Cobble	Silt/Mud			NA	NA		
Dominant Sed – Boulder	Silt/Mud			NA	NA		
Sediment – SandXL	Gravel	NA	NA	0.090	0.392	NA	NA
Sediment – SandLarge	Gravel	NA	NA	0.143	0.023	NA	NA
Sediment – SandMed	Gravel	NA	NA	0.955	0.061	NA	NA
Sediment – SandSmall	Gravel	NA	NA	0.010	0.469	NA	NA
Sediment – Silt/Mud	Gravel	NA	NA	<0.001	0.009	NA	NA
Purpose Code	NA	<0.001	0.304	<0.001	<0.001	<0.001	<0.001

Table 4: Tow counts for all survey types in the Gulf of Maine cod data set.

Data Type	Purpose Code				
	1	4	5	10	11
Conditional Presence	616	0	39	219	462
All data	2005	117	115	1461	763
Ratio	0.31	0	0.34	0.15	0.61

Table 5: Tow counts for all survey types in the Georges Bank yellowtail flounder data set.

Data Type	Purpose Code						
	4	5	6	10	11	40	60
Conditional Presence	0	0	0	77	15	0	7
All data	58	15	149	997	75	7	2018
Ratio	0	0	0	0.08	0.20	0	0.003

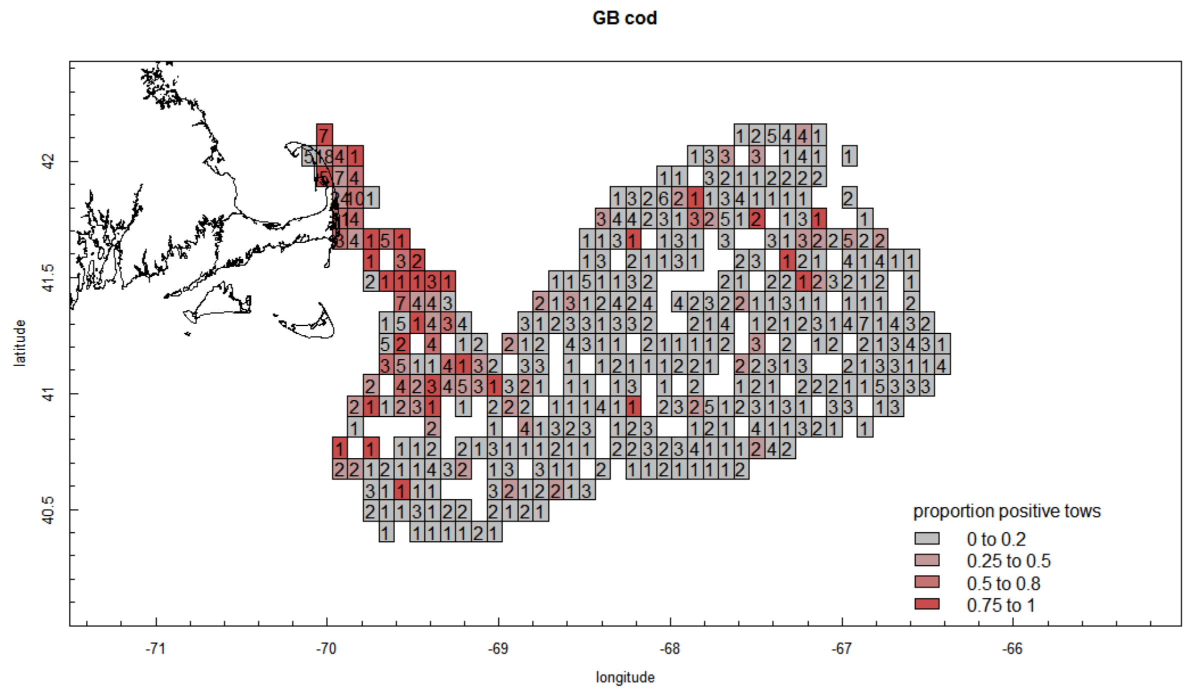


Figure 1: Number of tows per grid square and the proportion of those tows where juvenile cod were caught. The resolution of the grid, as in all the similar figures including the residual plots is 0.09 x 0.09 min., or approximately 10 km^2 (referenced in the north-south direction).

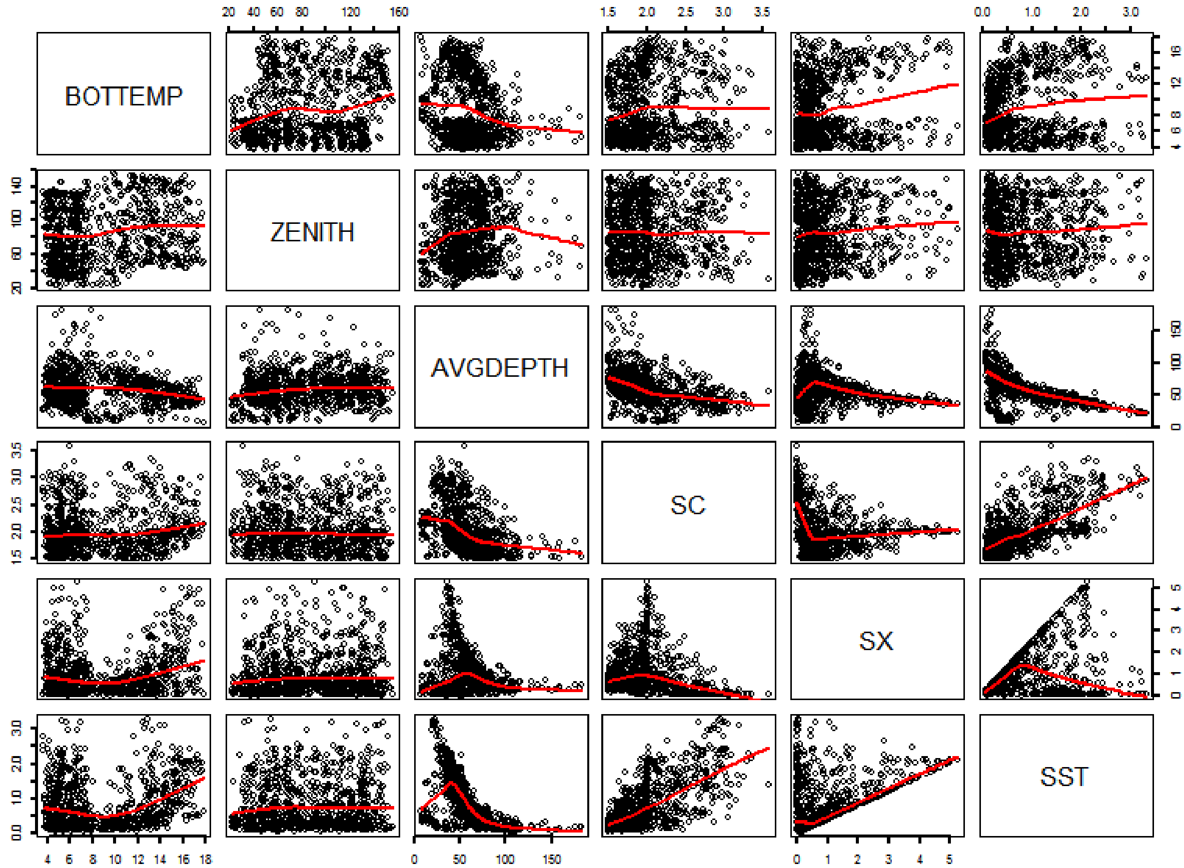


Figure 2: Correlations among continuous variables for the Georges Bank cod dataset.

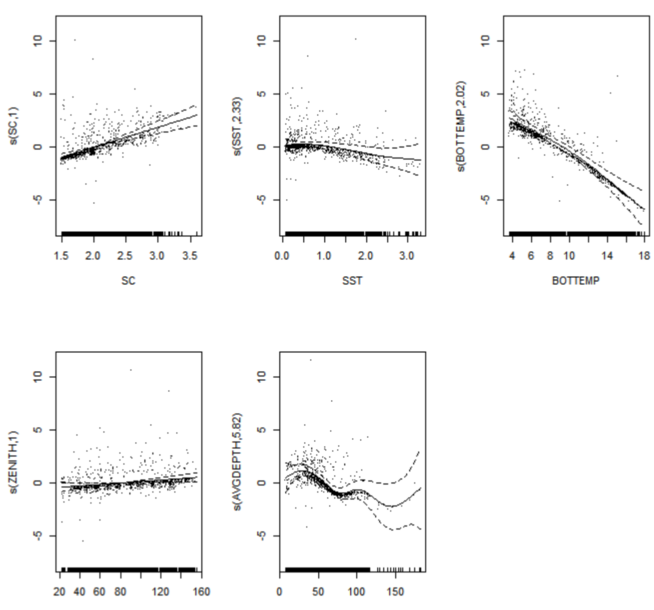


Figure 3: GAM smooth plots for the Georges Bank cod presence-absence model

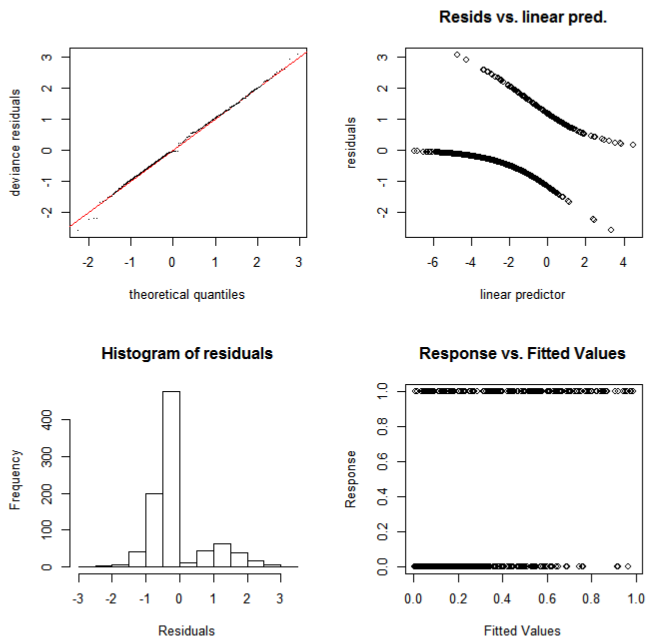


Figure 4: Diagnostic plots of presence absence model for Georges Bank cod

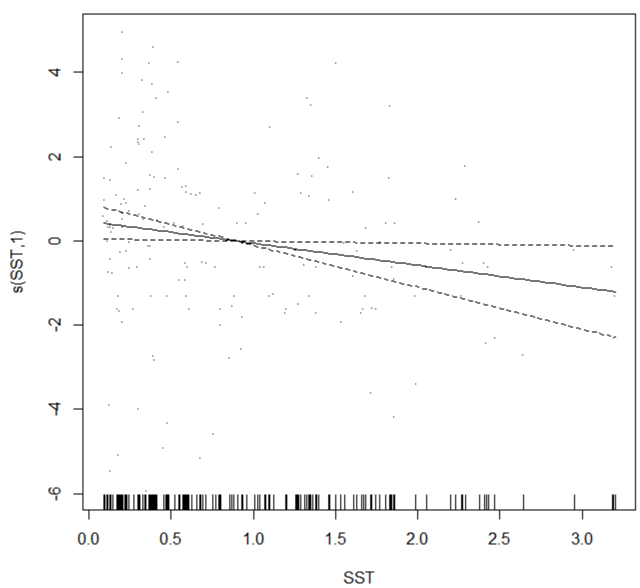


FIGURE 5: GAM smooth plot for the Georges Bank cod conditional presence model

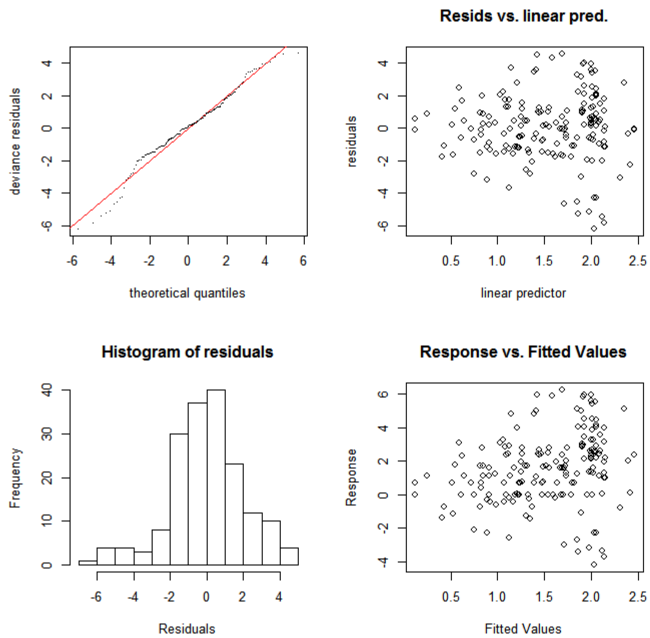


Figure 6: Diagnostic plots of conditional presence model for Georges Bank cod

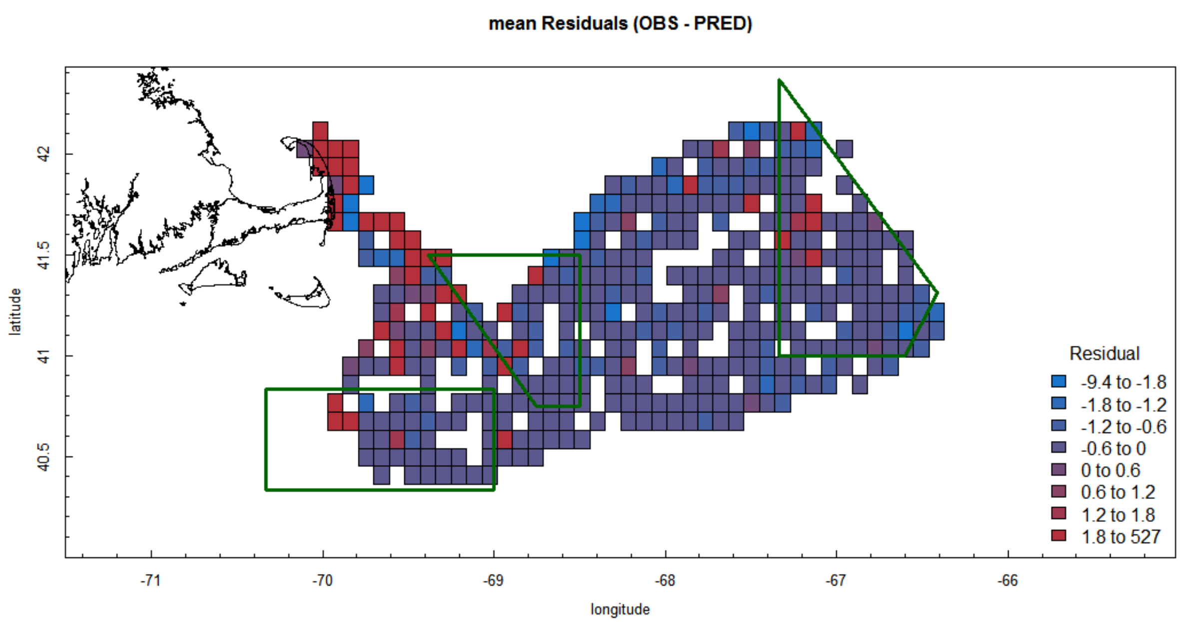


Figure 7: Mean residuals per square bin for Georges Bank cod

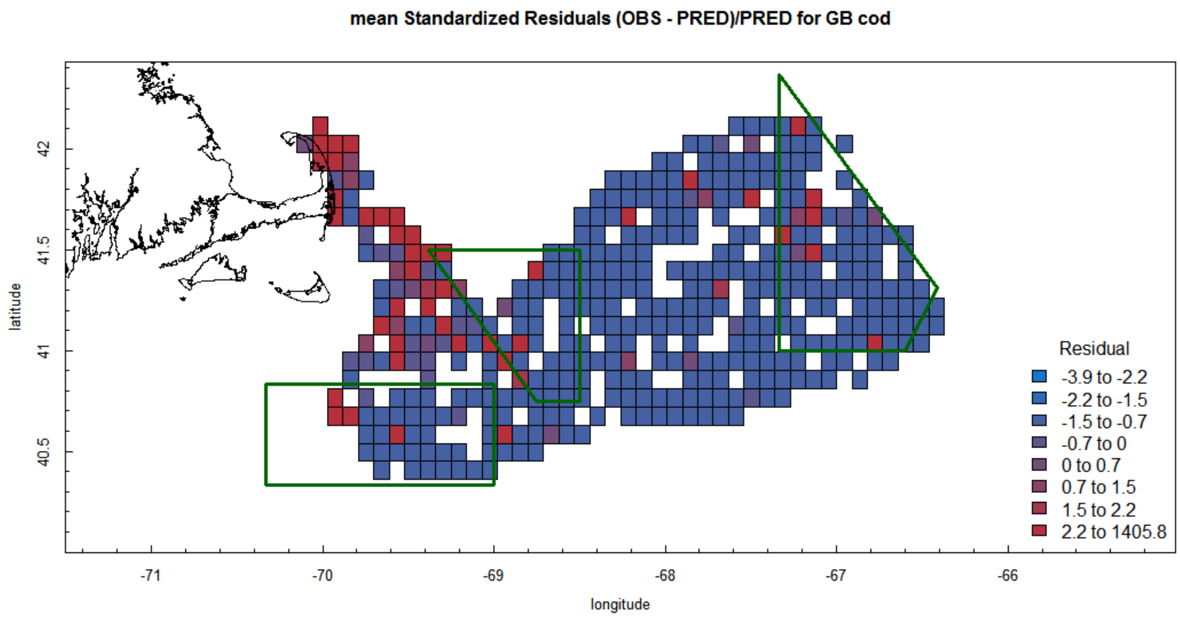


Figure 8: Mean residuals standardized by predictions per square bin for Georges Bank cod.

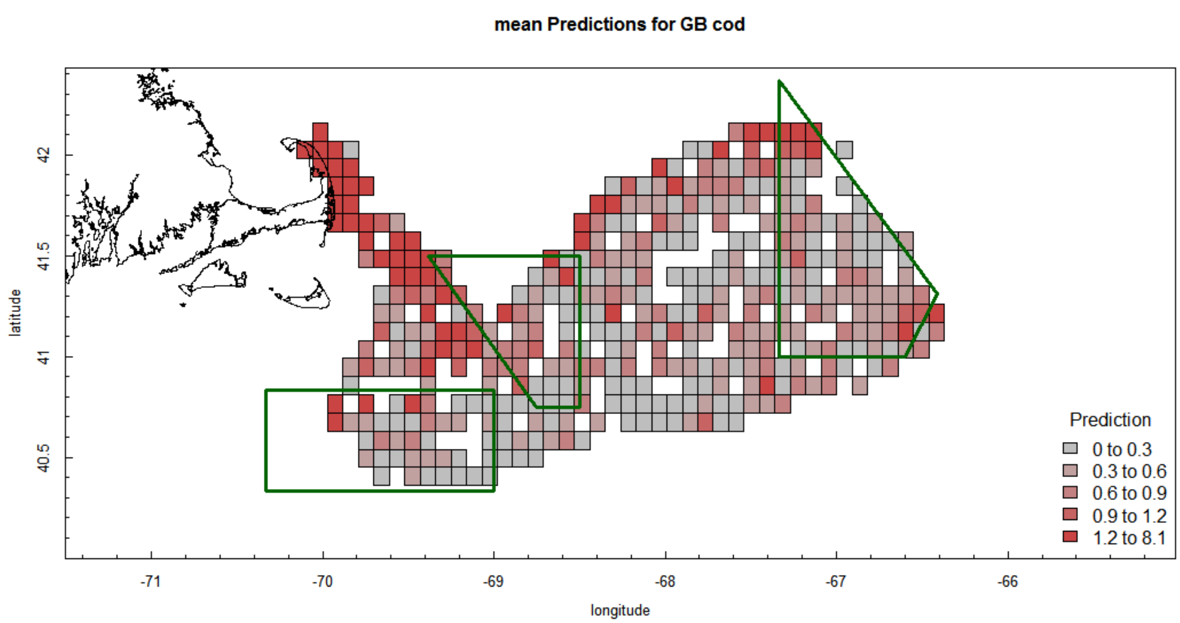


Figure 9: Mean overall predictions for Georges Bank cod.

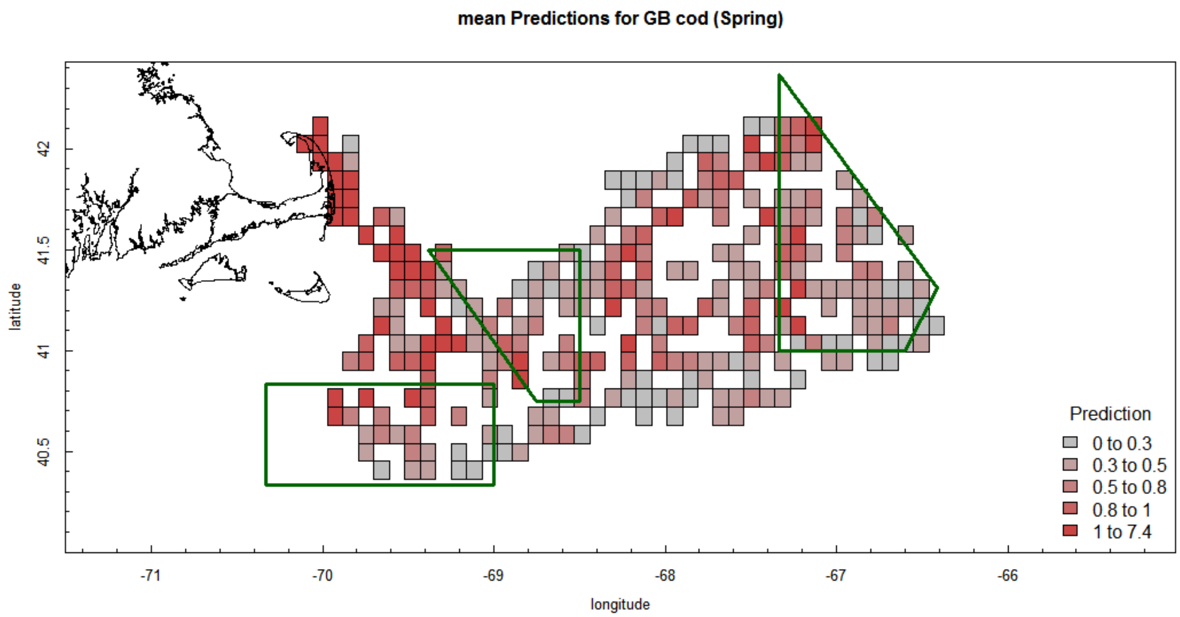


Figure 10: Mean predictions for Georges Bank cod in spring.

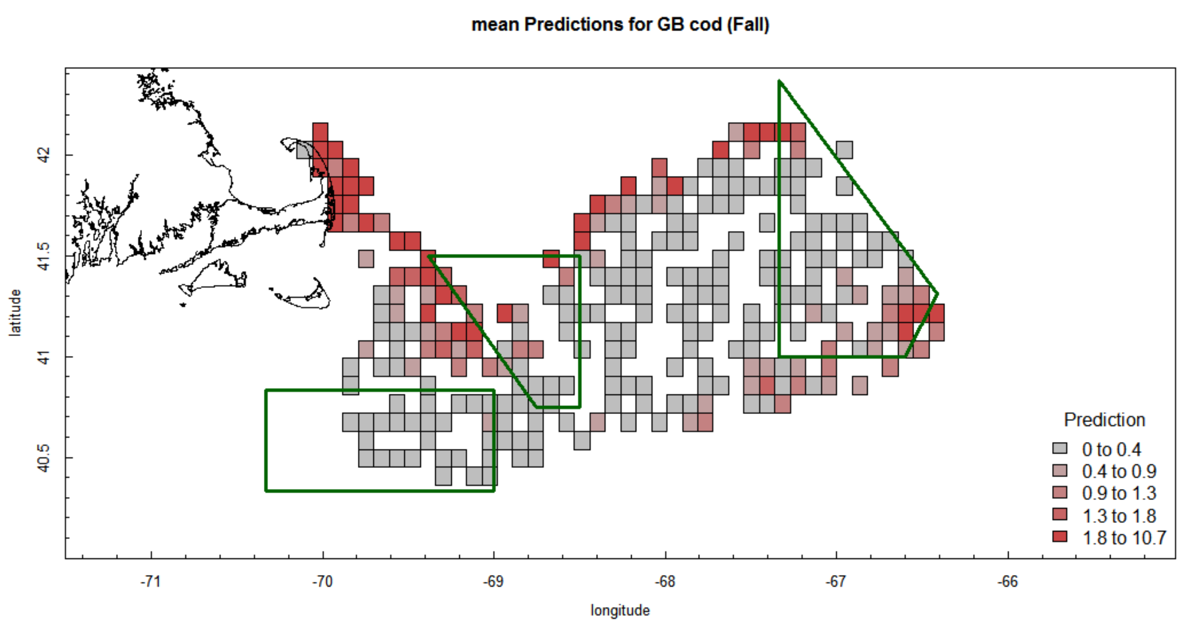


Figure 11: Mean predictions for Georges Bank cod in fall.

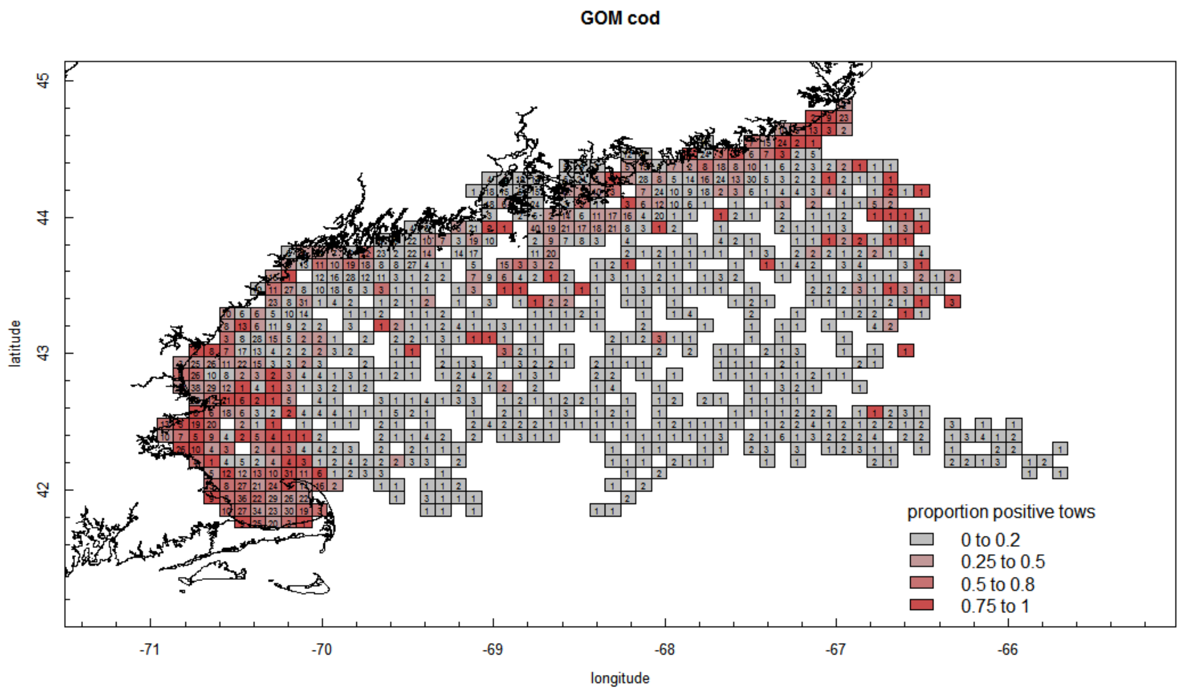


Figure 12: Number of tows per grid square and the proportion of those tows where juvenile cod were caught.

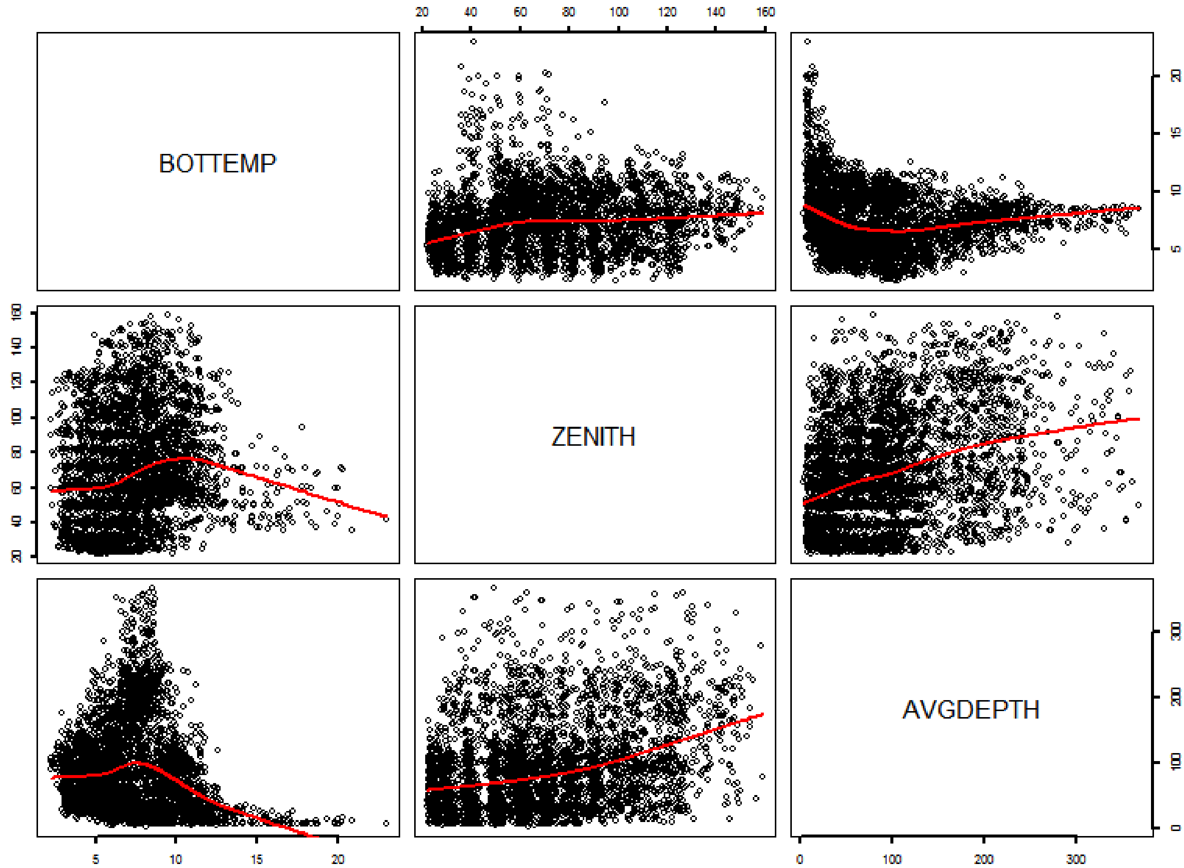


Figure 13: Correlations among continuous variables for the Gulf of Maine cod dataset.

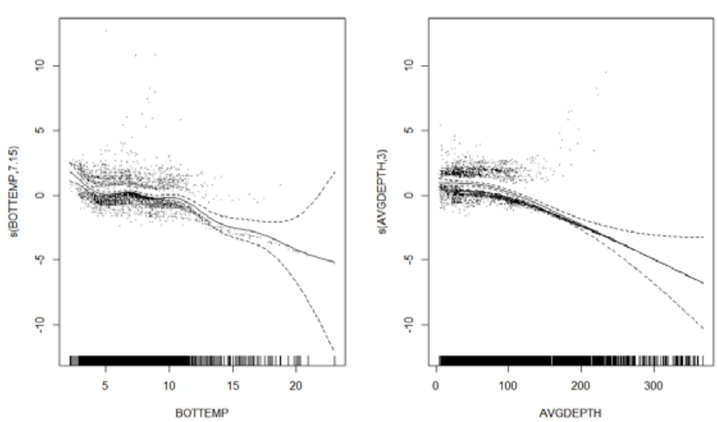


Figure 14: GAM smooth plots for the Gulf of Maine cod presence-absence model

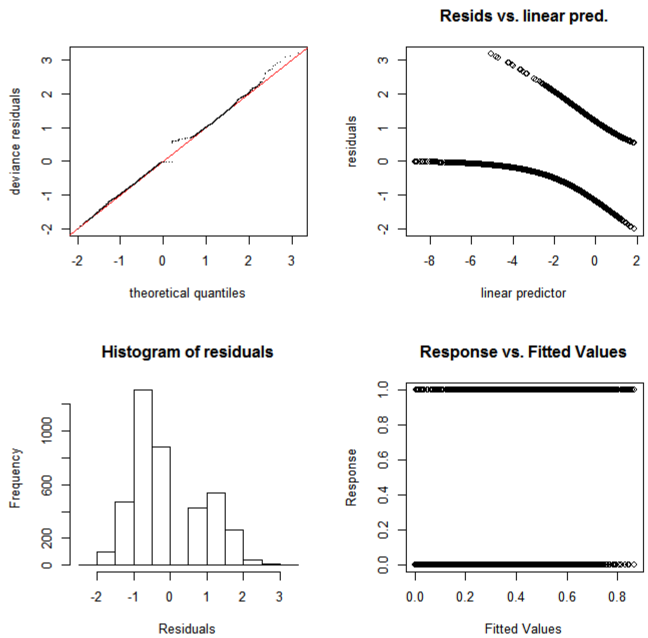


Figure 15: Diagnostic plots of presence absence model for Gulf of Maine cod.

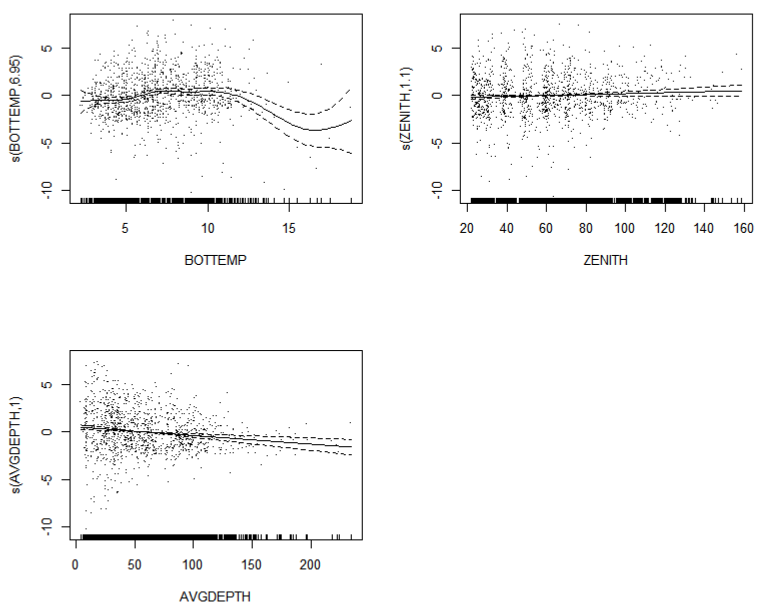


Figure 16: GAM smooth plots for the Gulf of Maine cod conditional presence model.

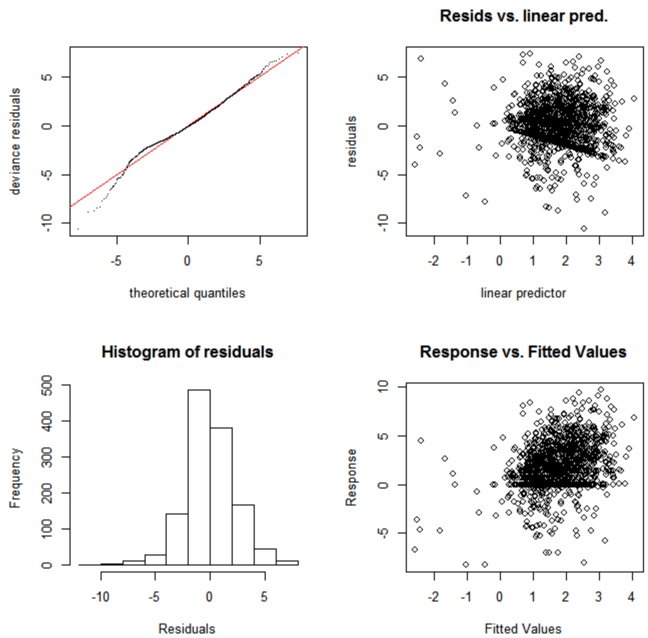


Figure 17: Diagnostic plots of conditional presence model for Gulf of Maine cod.

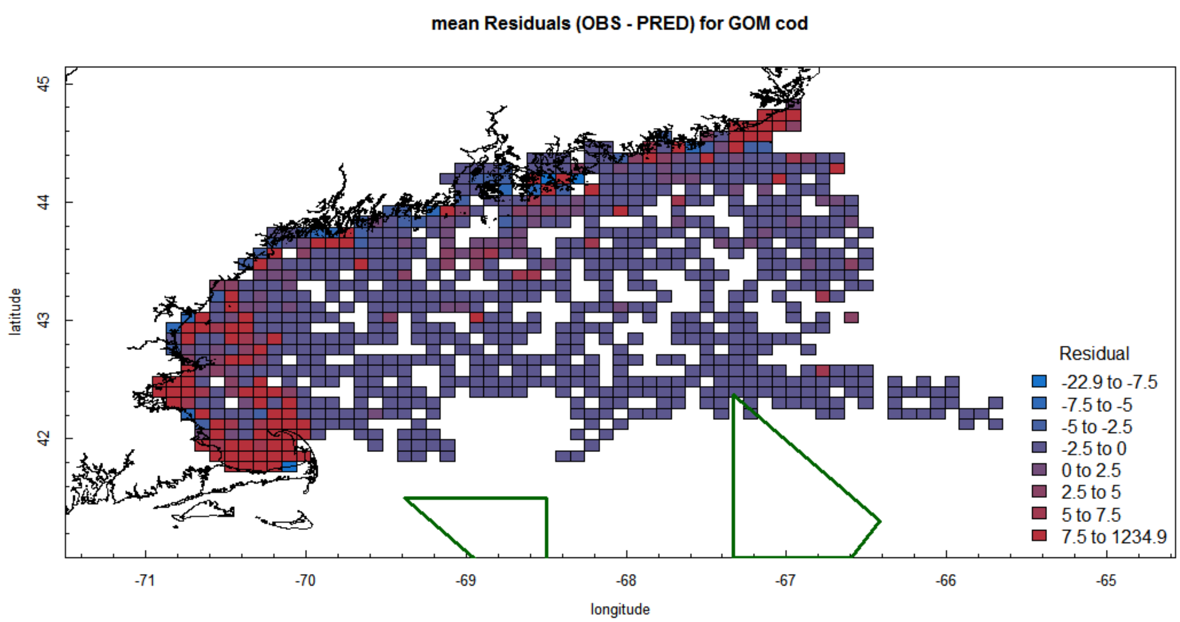


Figure 18: Mean residuals per square bin for Gulf of Maine cod.

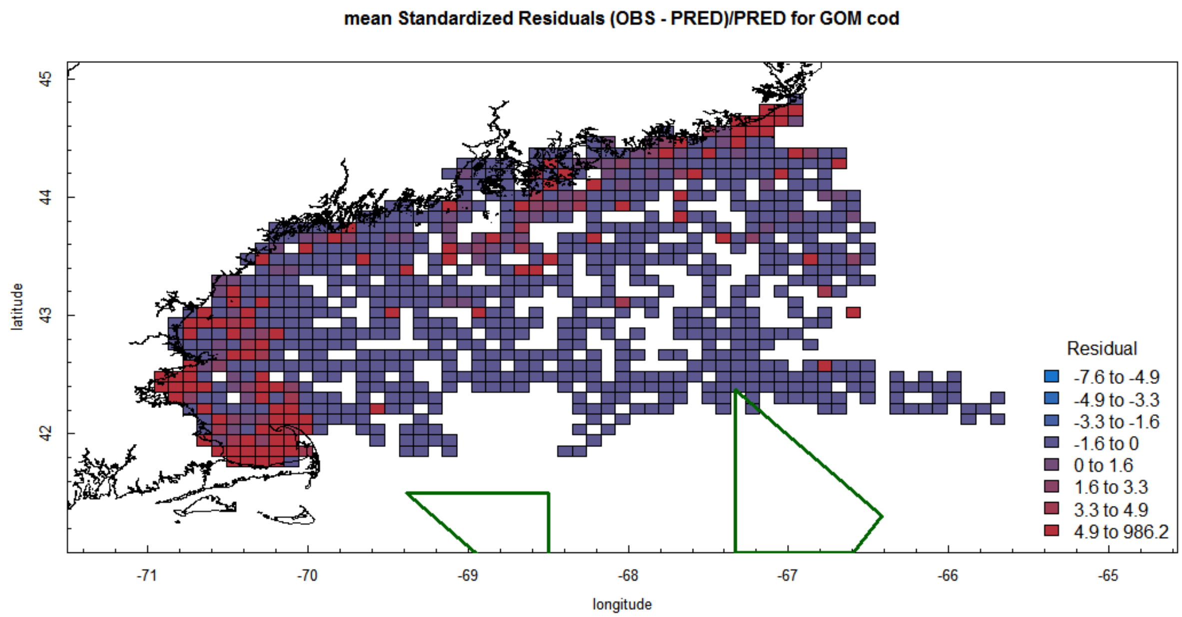


Figure 19: Mean residuals standardized by predictions per square bin for Gulf of Maine cod

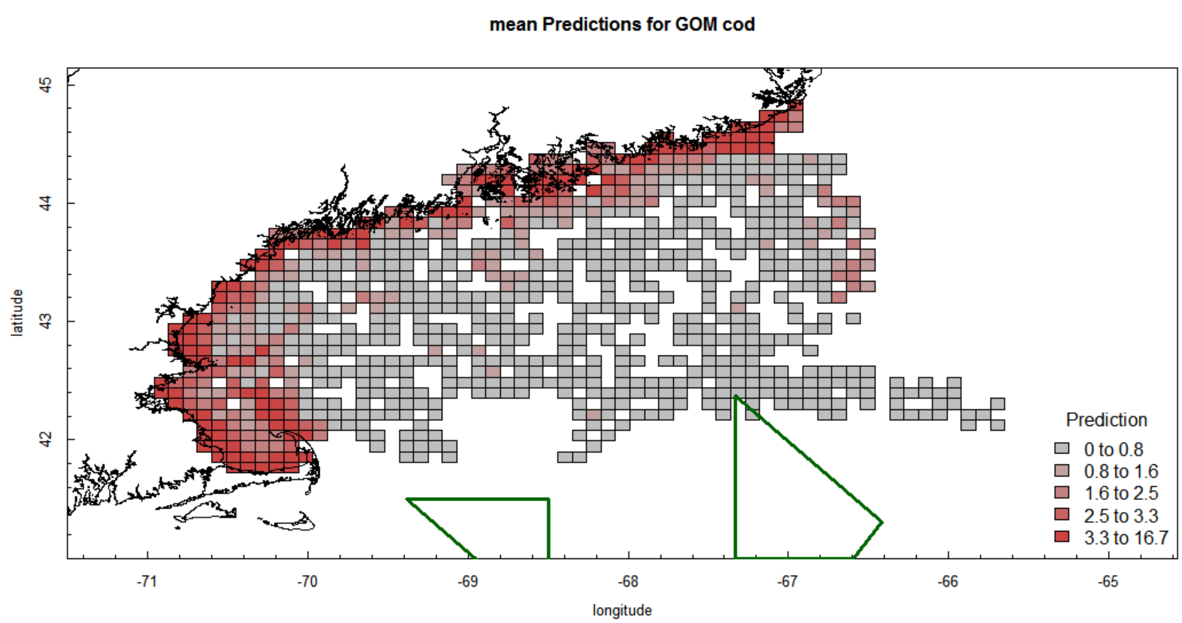


Figure 20: Mean overall predictions for Gulf of Maine cod

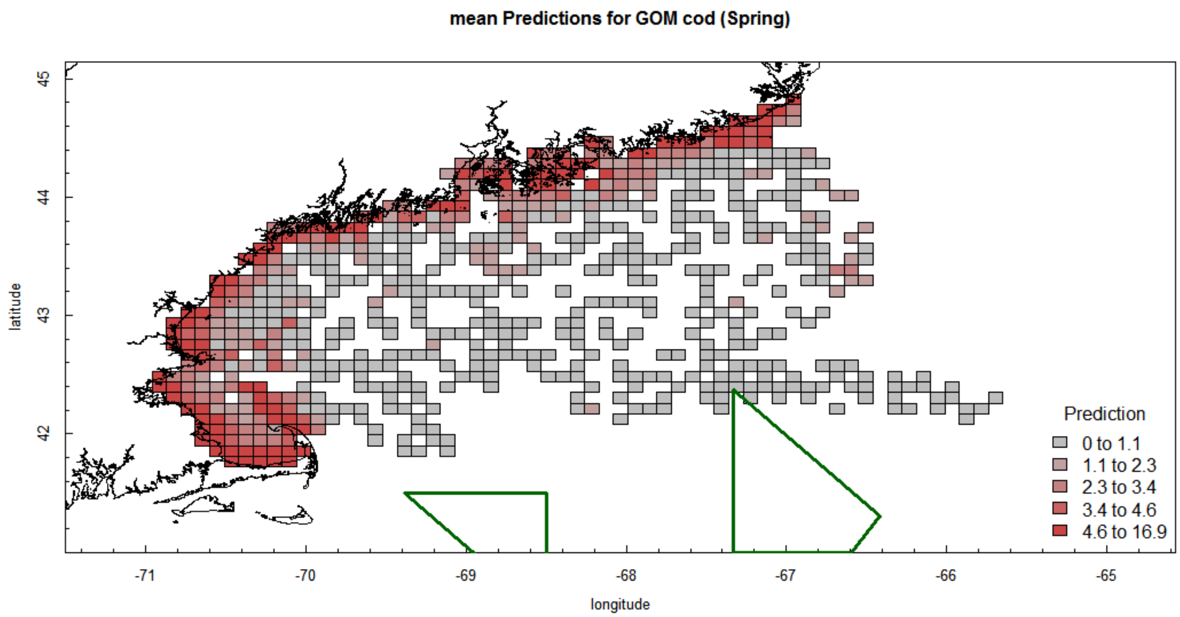


Figure 21: Mean predictions for Gulf of Maine cod in spring

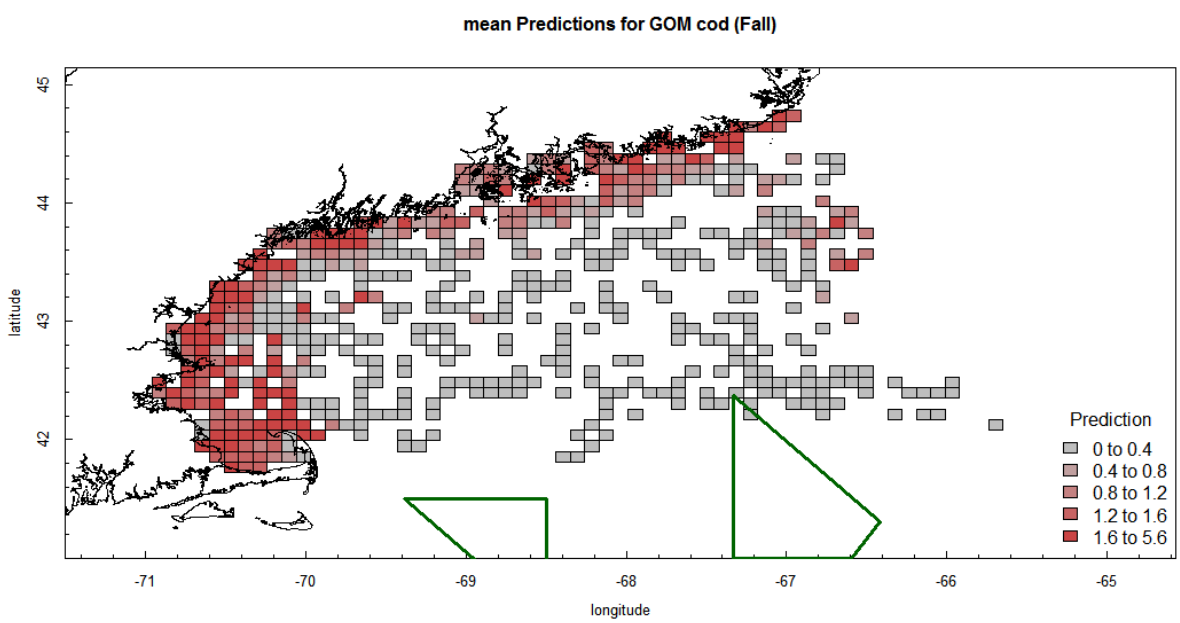


Figure 22: Mean predictions for Gulf of Maine cod in fall.

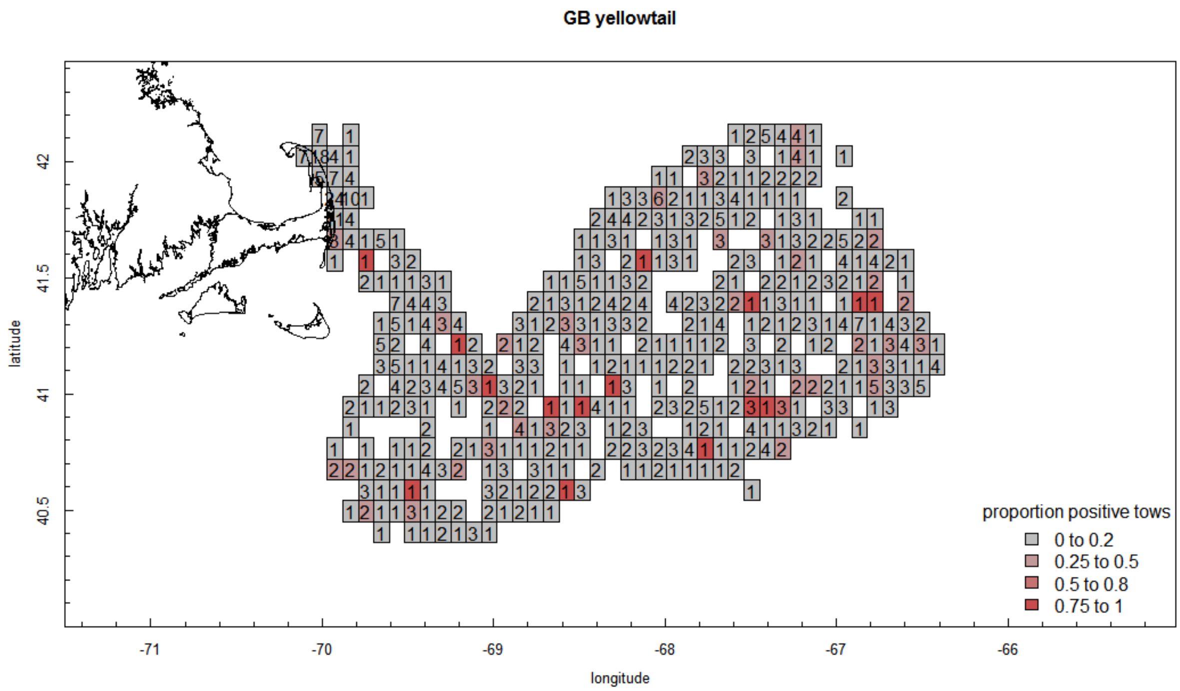


Figure 23: Number of tows per grid square and the proportion of those tows where juvenile yellowtail were caught.

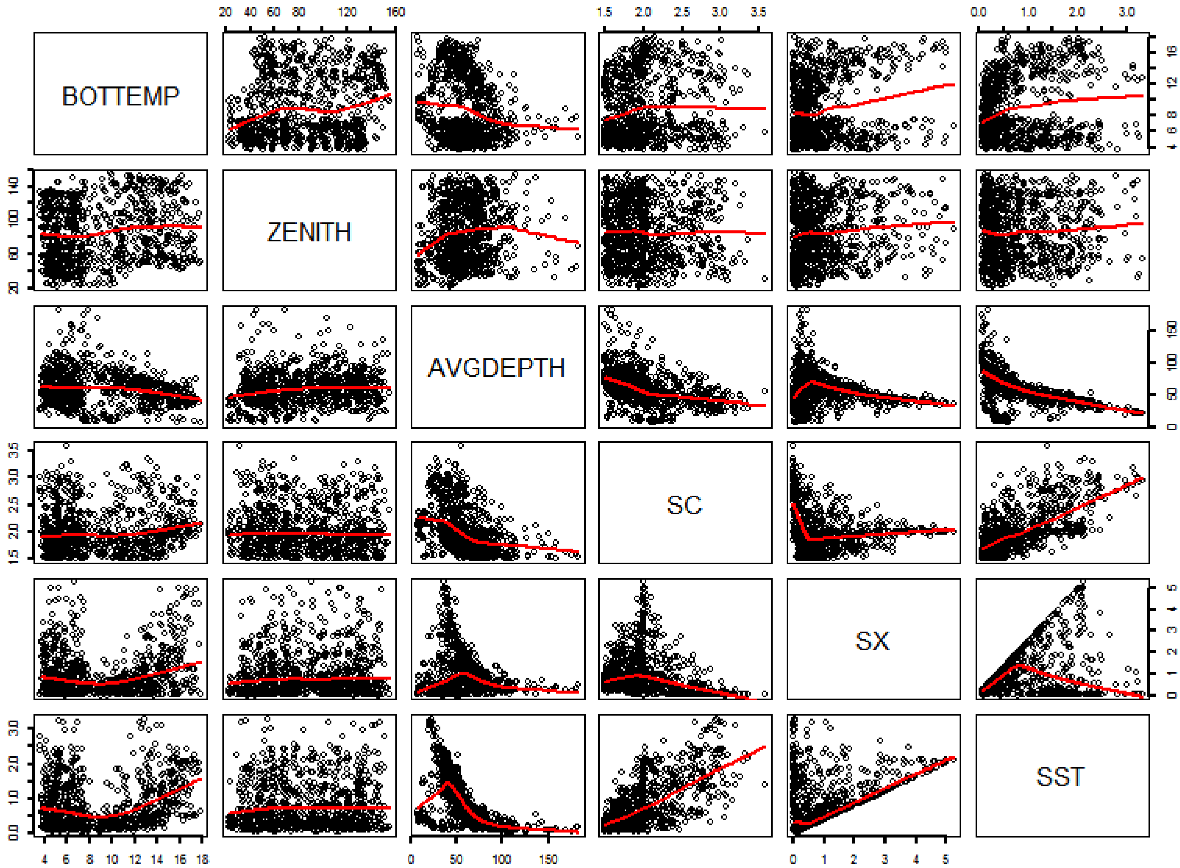


Figure 24: Correlations among continuous variables for the Georges Bank yellowtail flounder dataset.

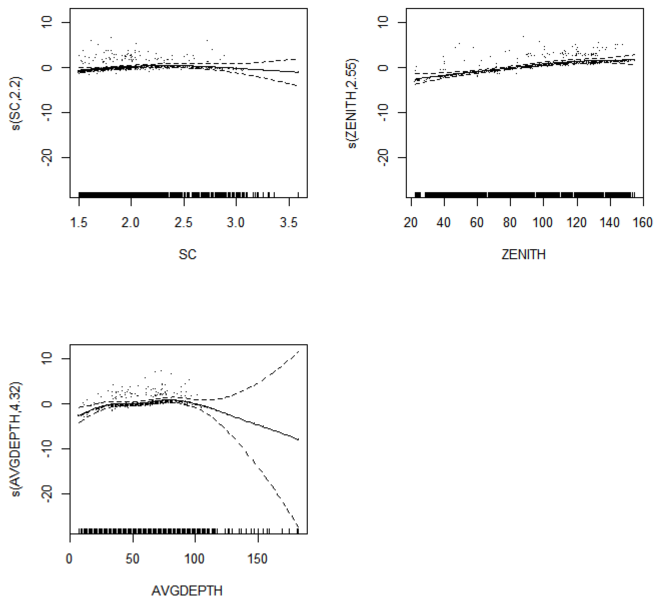


Figure 25: GAM smooth plots for the Georges Bank yellowtail flounder presence-absence model.

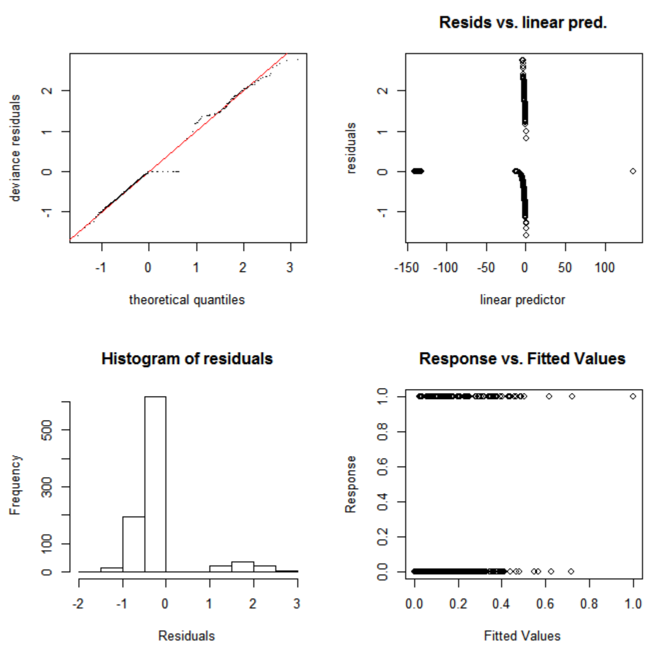


Figure 26: Diagnostic plots of presence absence model for Georges Bank yellowtail flounder.

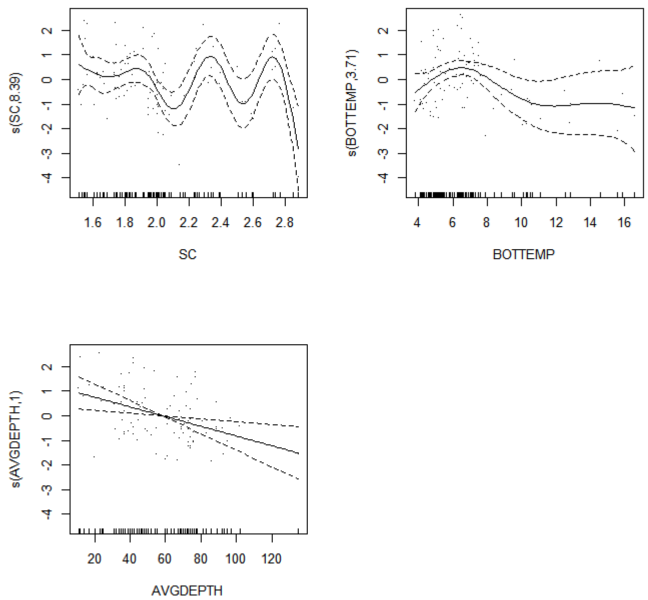


Figure 27: GAM smooth plots for the Georges Bank yellowtail flounder conditional presence model

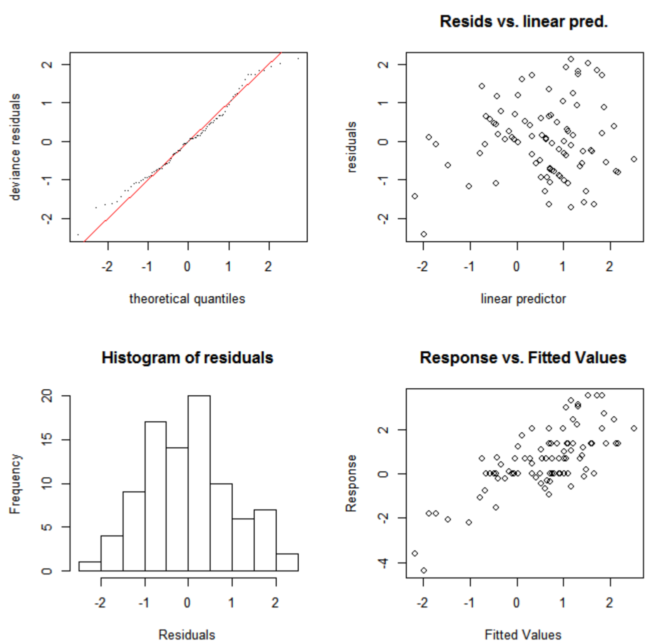


Figure 28: Diagnostic plots of conditional presence model for Georges Bank yellowtail flounder

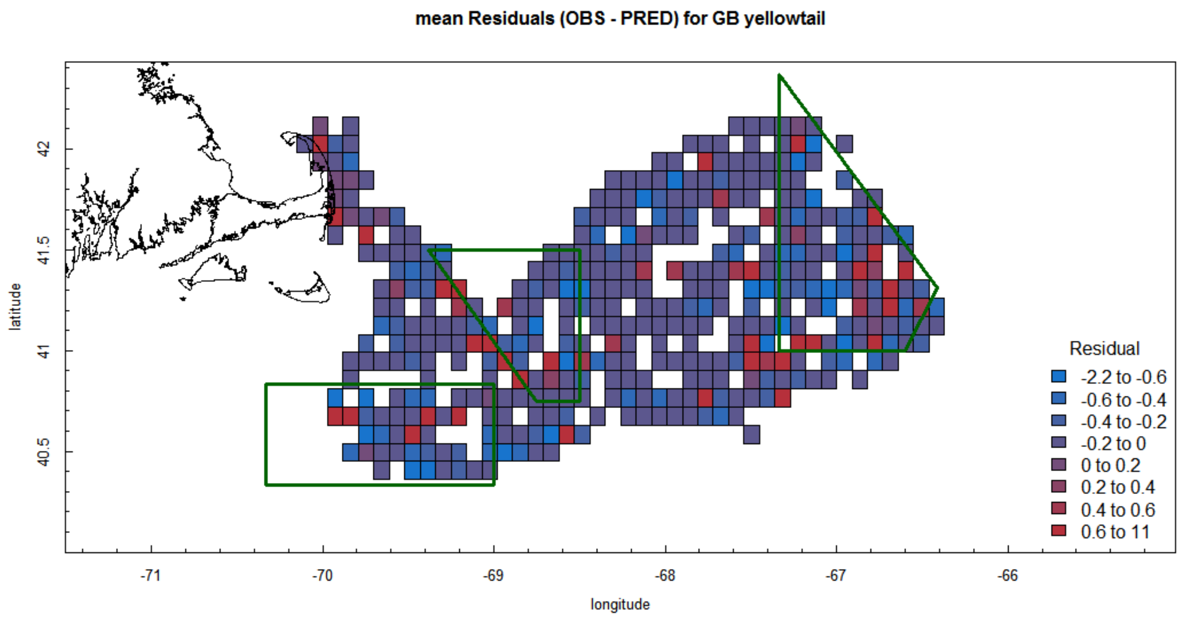


Figure 29: Mean residuals per square bin for Georges Bank yellowtail flounder.

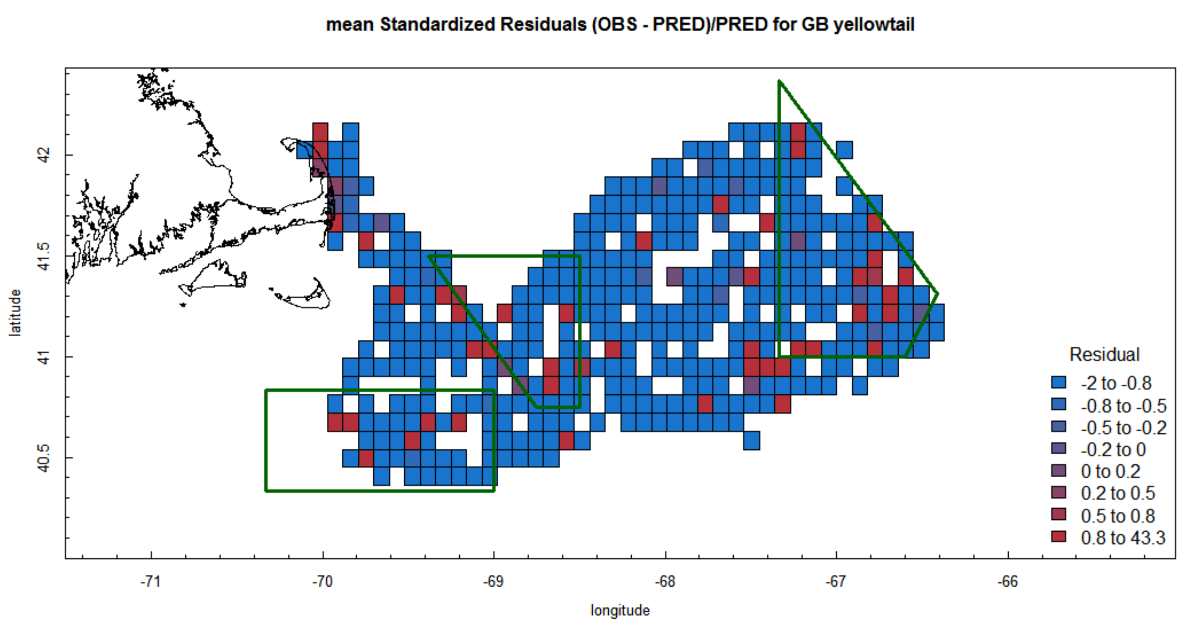


Figure 30: Mean residuals standardized by predictions per square bin for Georges Bank yellowtail flounder.

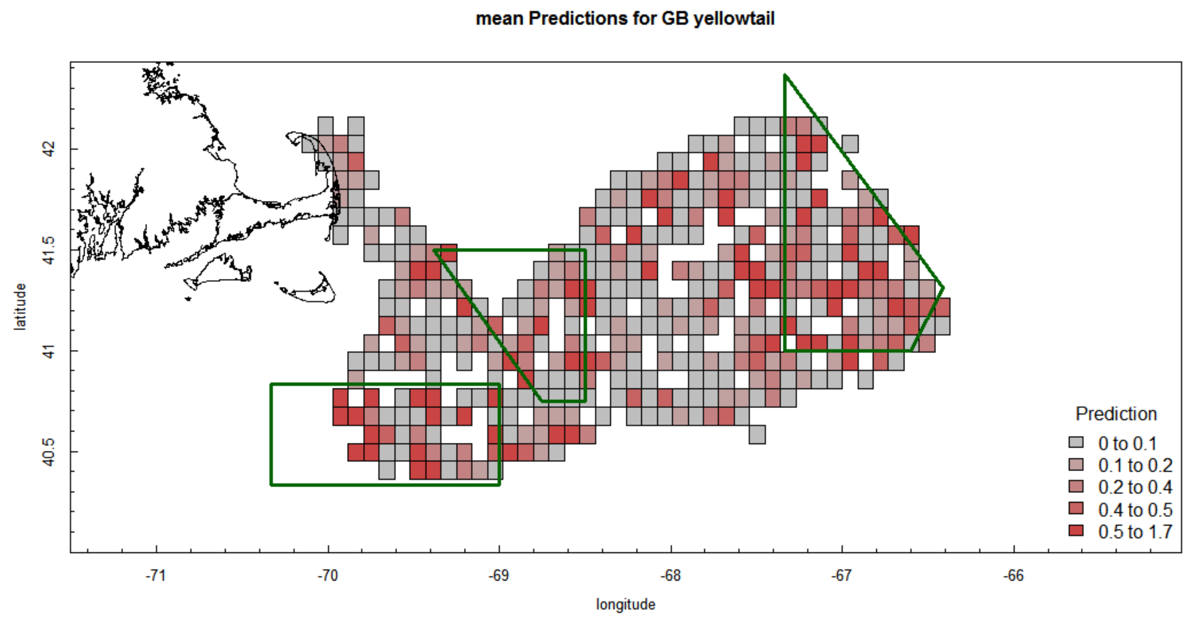


Figure 31: Mean overall predictions for Georges Bank yellowtail flounder.

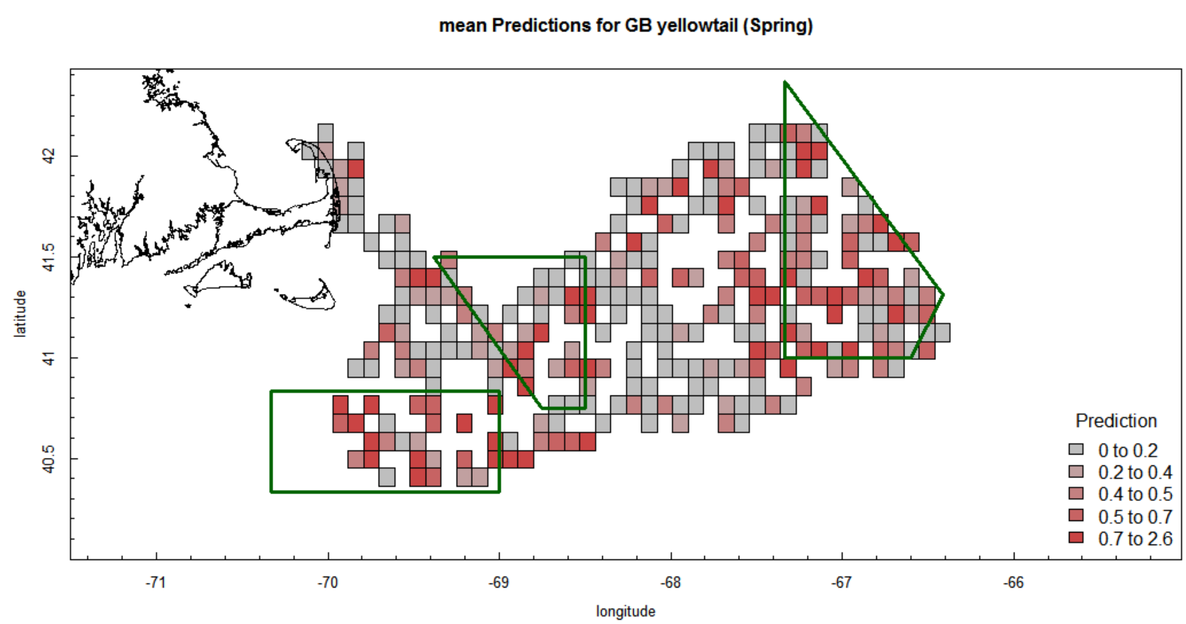


Figure 32: Mean predictions for Georges Bank yellowtail flounder in spring.

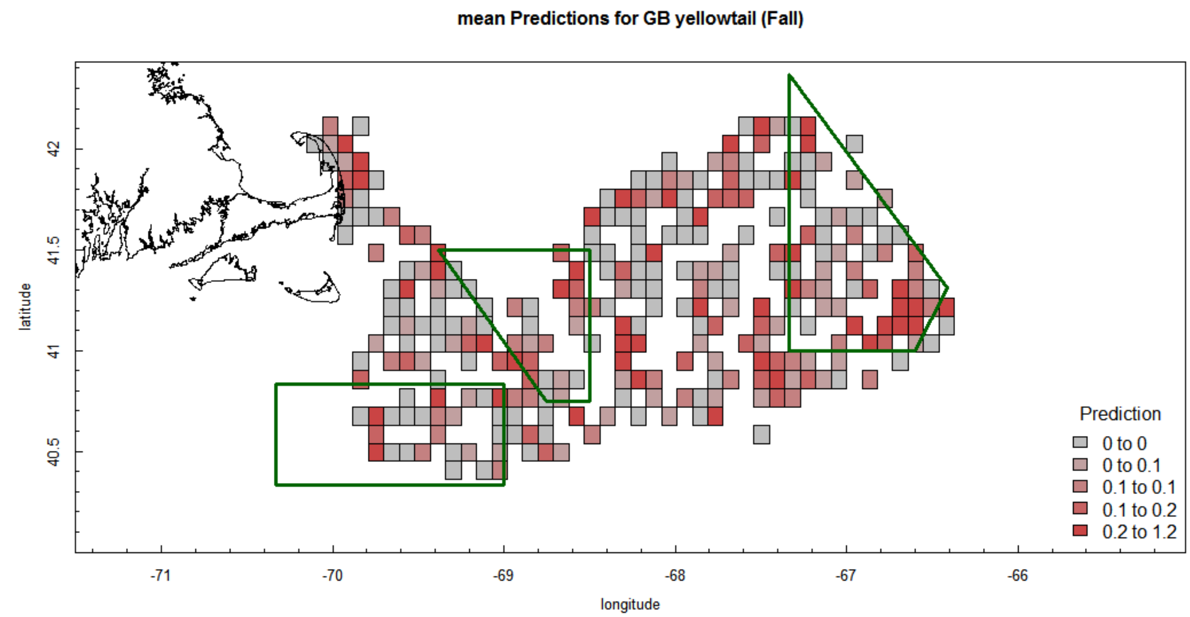


Figure 33: Mean predictions for Georges Bank yellowtail flounder in fall.

APPENDIX 1: Additional information on the candidate variables

Table 1 – Length thresholds analyzed for small fish. The thresholds were selected using age/length keys based on fall and spring NMFS trawl survey data to capture most of the age 0 and 1 juveniles. All lengths were rounded to the nearest 5 cm.

Species	Survey season	Juvenile max length
Atlantic cod	Spring	25
	Fall	35
Yellowtail flounder	Spring	15
	Fall	15

Table 2 - Survey purpose codes

Purpose code	Description	Notes
1	Maine New Hampshire trawl survey	Separate data file
4	Cooperative research survey – goosefish	Data from 2004 and 2009
5	Cooperative research survey – IBS cod	Data from 2003-2007
6	Cooperative research survey – IBS yellowtail	Data from 2003-2005, SNE-MAB
9	Cooperative research survey – paired trawl	
10	NMFS NEFSC bottom trawl survey	Spring, summer, fall, winter (winter through 2009, all other years 2002-2012)
11	MA DMF bottom trawl survey	Fall and spring, off MA coast
40	NMFS NEFSC shrimp survey	GOM, summer survey
60	NMFS NEFSC sea scallop survey	GB and MAB, summer survey

Table 3 – Habitat data in first data sets distributed

Data type	Data source	Coverage	Variable type	Notes
Depth	Fish survey data 2002-2012.	Same as catch data - each station has a depth	Continuous integer	Should probably use coastal relief model depth if we need a surface to predict to – working on joining this data set. Because depth is not expected to vary between years, CRM or survey depth should be fairly consistent.
Bottom temperature	Fish survey data 2002-2012.	Same as catch data - each station has a bottom temp	Continuous integer	Hard to come up with a single average bottom temperature layer by season – varies by year. Best info will be the temperature at the time of the tow.
Substrate	usSEABED, as processed forTNC ecoregional assessment	Entire coast to about 2500 m	Categorical- interpolated polygons of average grain size. 5 bins – 1 mud, 3 subdivisions of sand, 1 gravel. Polygons spatially joined to midpoint of tows.	Have other data sources for substrate as well but this one is the easiest to work with/most spatially comprehensive. Will provide additional data for yellowtail and cod for GB only.

Data type	Data source	Coverage	Variable type	Notes
Substrate	State of Maine	Inshore Maine coast – just beyond 3 nm boundary.	Categorical - interpolated polygons based on multibeam backscatter – sand, rock, gravel, mud. Polygons spatially joined to midpoint of tows.	Can be used as an alternative for MENH catch data. Does not cover entire footprint of MENH survey so there will be some tows without a substrate attribute if using these data
Seabed form	Derived from TNC depth and position index	Entire coast to about 2500 m	Publically available as a raster, 83 m resolution. Categorical variable – 9 combos of low/mid/high position combined with flat/moderate/steep slope.	Would need to join spatially to survey data set – having issues extracting raster to points. Trying to include these data and will send an updated data set.

Table 4 - Sediment and sediment stability data from Harris and Stokesbury 2010 and Harris et al 2012

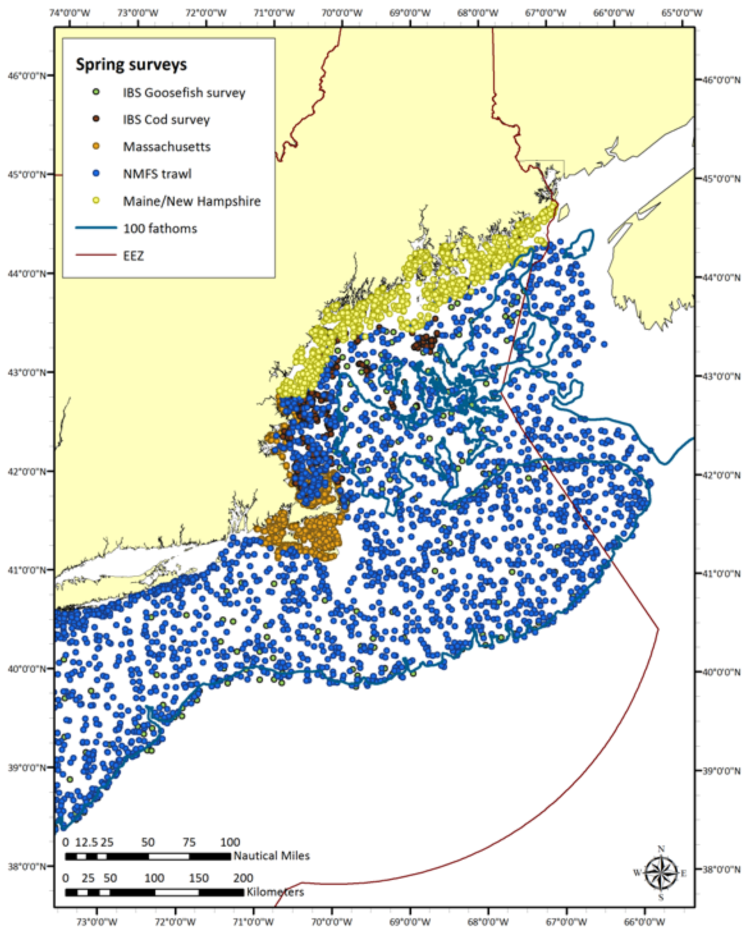
Field	Description
Long	Sediment Map Grid Longitude
Lat	Sediment Map Grid Latitude
Sm	Maximum Size Sediment Type Values: 1 = Silt/Mud, 2 = Sand, 3 = Granule/Pebble, 4 = Cobble, 5= Boulder Details on page 1842 - 1843 of Harris and Stokesbury 2010
Sd	Dominant Sediment Type (Most commonly occurring type in four replicate samples per station). Values: 1 = Silt/Mud, 2 = Sand, 3 = Granule/Pebble, 4 = Cobble, 5= Boulder Details on page 1842 - 1843 of Harris and Stokesbury 2010
Sc	Sediment Coarseness Values ≤ 2 = Smooth, >2 but <4 = Intermediate, ≥ 4 = Coarse Details on page 1842 - 1843 of Harris and Stokesbury 2010
Sx	Sediment Stability Index Values ≥ 1 = unstable. Values < 1 = Stable Details in section 2.3 of Harris et al 2012
Sst	Benthic boundary shear stress (N m^{-2}, annual mean max M_2+S_2 tidal = bi-weekly) Details in section 2.1 of Harris et al 2012

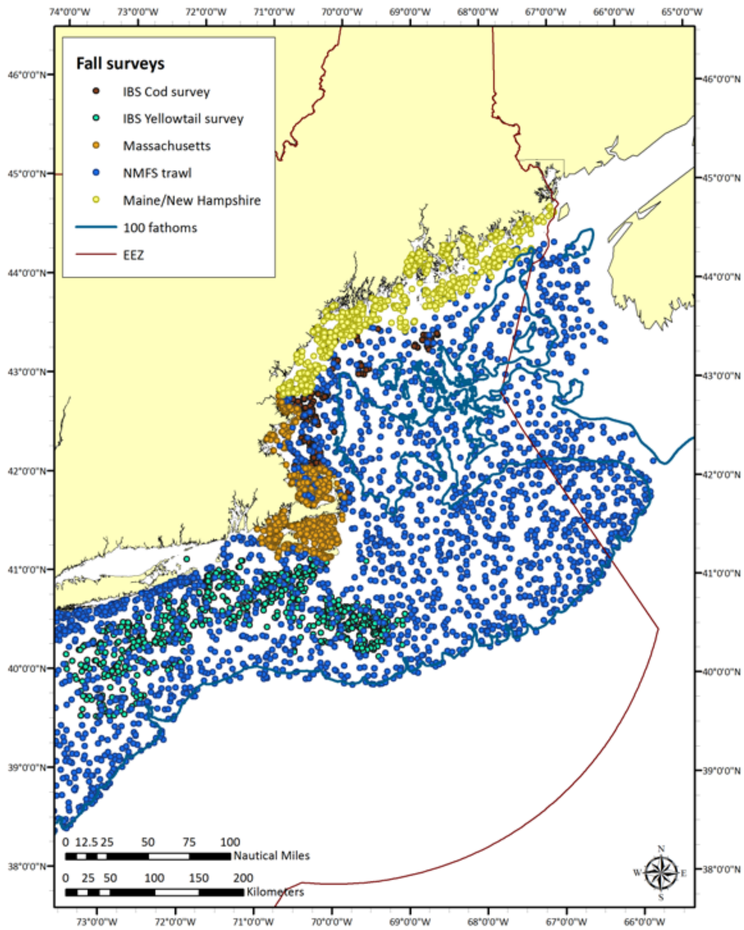
Table 5 - Seabed forms data

SLOPE	C_SLOPE	LPI	C_LPI	SEABEDFORM	SB_form
0 - 0.015%	1	Low Land Position	1	Depression	1
0 - 0.015%	1	Low Land Position	2	Depression	1
0 - 0.015%	1	Mid Land Position	3	Mid Flat	2
0 - 0.015%	1	Mid Land Position	4	Mid Flat	2
0 - 0.015%	1	High Land Position	5	High Flat	3
0 - 0.015%	1	High Land Position	6	High Flat	3
0.015 - 0.05%	2	Low Land Position	1	Depression	1
0.015 - 0.05%	2	Low Land Position	2	Depression	1
0.015 - 0.05%	2	Mid Land Position	3	Mid Flat	2

SLOPE	C_SLOPE	LPI	C_LPI	SEABEDFORM	SB_form
0.015 - 0.05%	2	Mid Land Position	4	Mid Flat	2
0.015 - 0.05%	2	High Land Position	5	High Flat	3
0.015 - 0.05%	2	High Land Position	6	High Flat	3
0.05 - 0.8	3	Low Land Position	1	Low Slope	4
0.05 - 0.8	3	Low Land Position	2	Low Slope	4
0.05 - 0.8	3	Mid Land Position	3	Side Slope	6
0.05 - 0.8	3	Mid Land Position	4	Side Slope	6
0.05 - 0.8	3	High Land Position	5	High Slope	5
0.05 - 0.8	3	High Land Position	6	High Slope	5
0.8 -8%	4	Low Land Position	1	Low Slope	4
0.8 -8%	4	Low Land Position	2	Low Slope	4
0.8 -8%	4	Mid Land Position	3	Side Slope	6
0.8 -8%	4	Mid Land Position	4	Side Slope	6
0.8 -8%	4	High Land Position	5	High Slope	5
0.8 -8%	4	High Land Position	6	High Slope	5
>8%	5	Low Land Position	1	Steep	7
>8%	5	Low Land Position	2	Steep	7
>8%	5	Mid Land Position	3	Steep	7
>8%	5	Mid Land Position	4	Steep	7
>8%	5	High Land Position	5	Steep	7
>8%	5	High Land Position	6	Steep	7

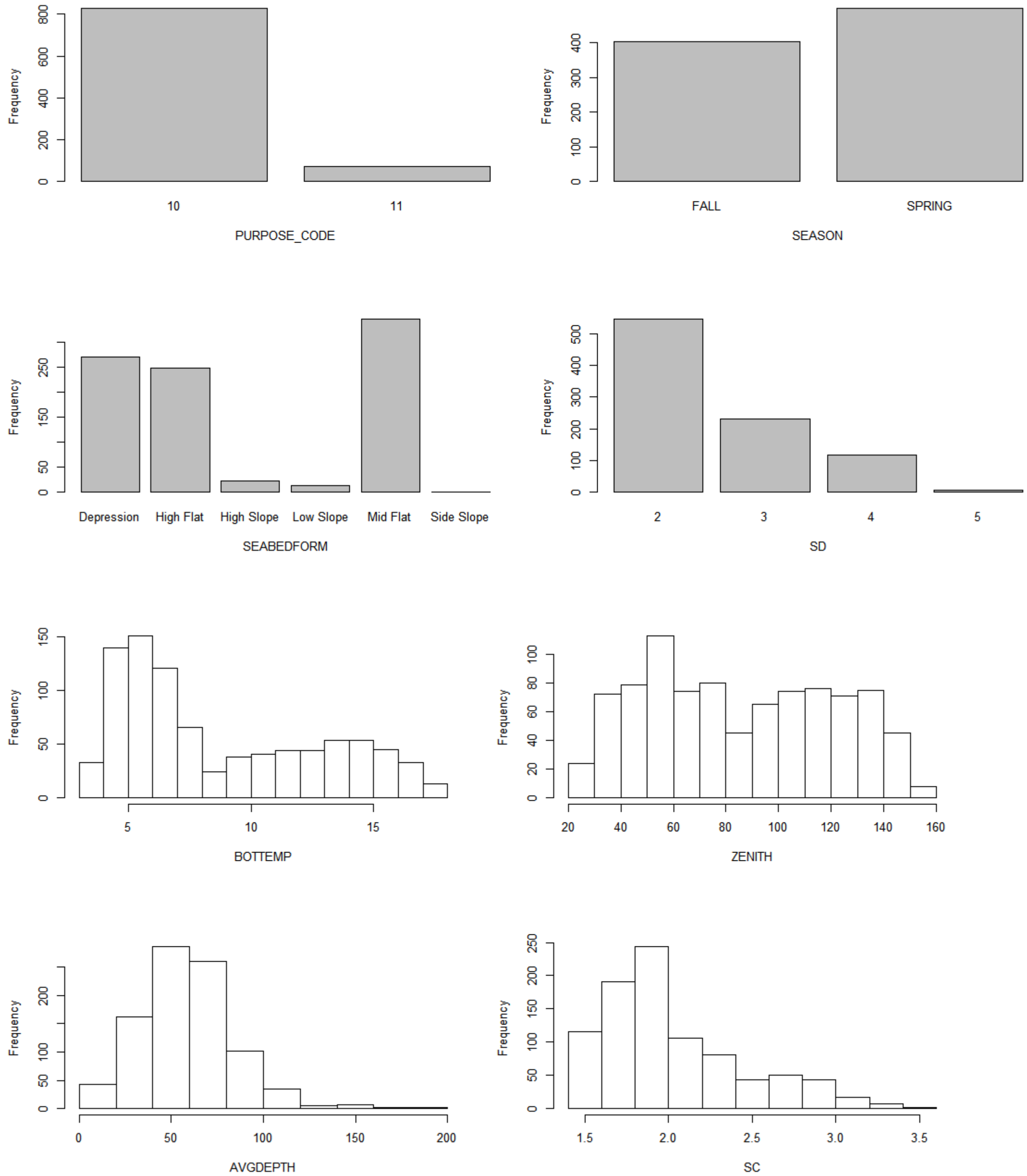
APPENDIX 2: Spatial and seasonal distribution of the fisheries surveys that were used in the modeling

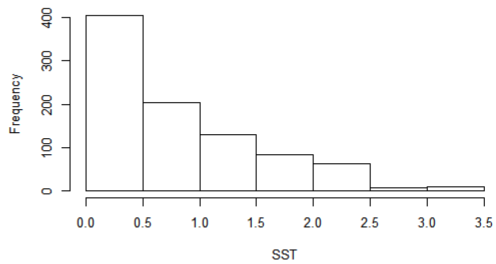


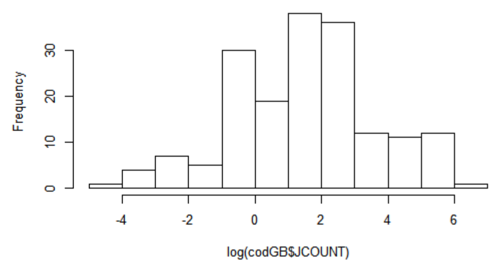


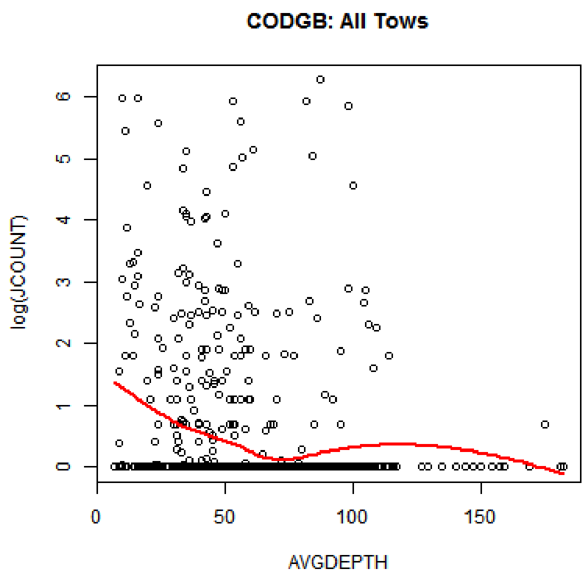
Fall surveys
IBS Cod survey
IBS Yellowtail survey
Massachusetts
NMFS trawl
Maine/New Hampshire
100 fathoms
EEZ
0 12.5 25 50 75 100
Nautical Miles
0 25 50 100 150 200
Kilometers

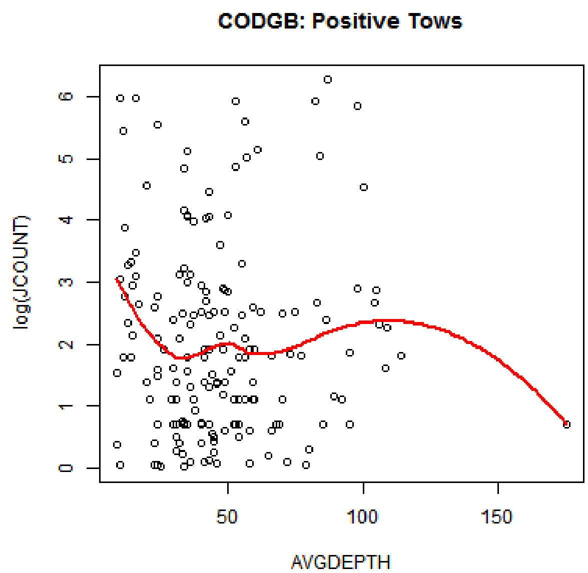
APPENDIX 3: Premodeling Georges Bank cod analysis

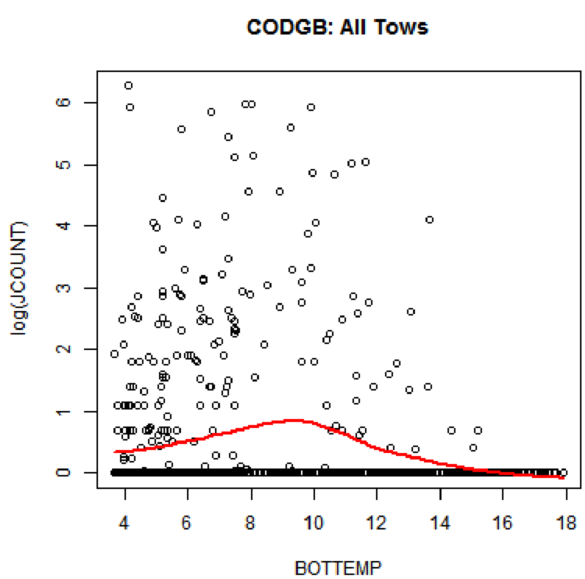


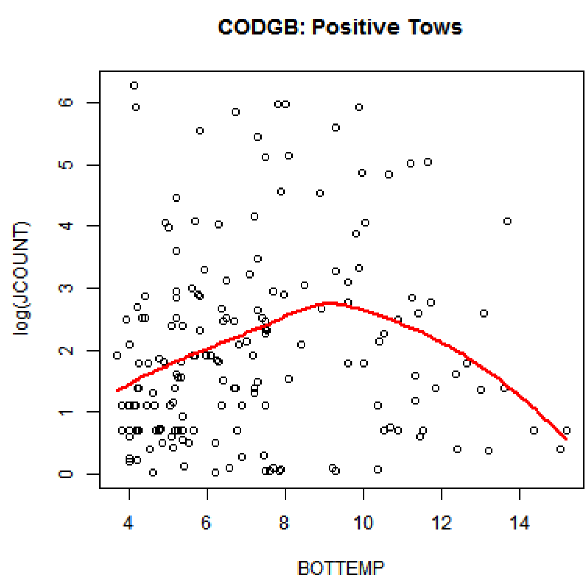

Frequency
SST

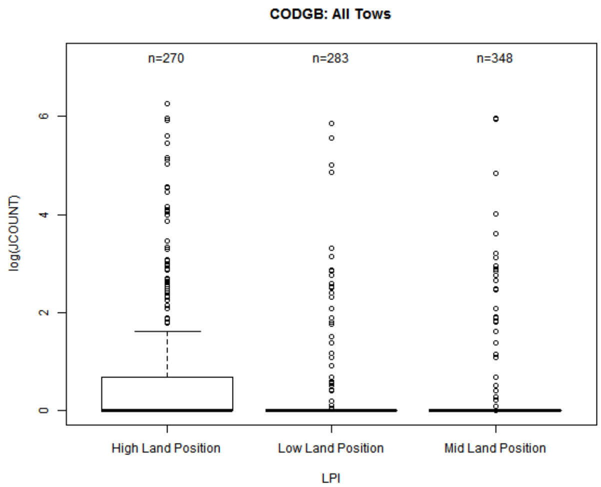

Frequency
log(codGB$JCOUNT)

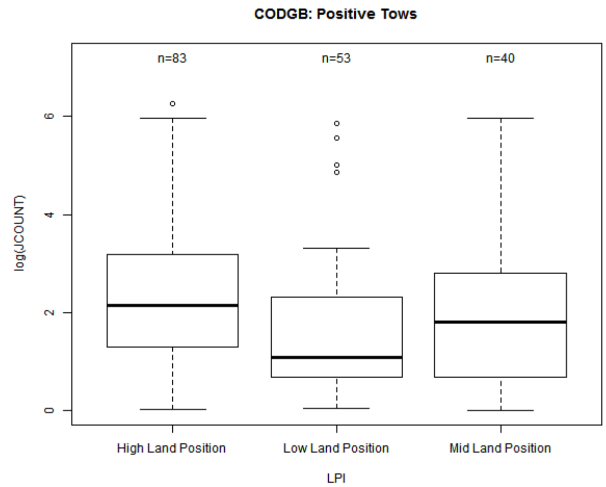

CODGB: All Tows
log(JCOUNT)
AVGDEPTH

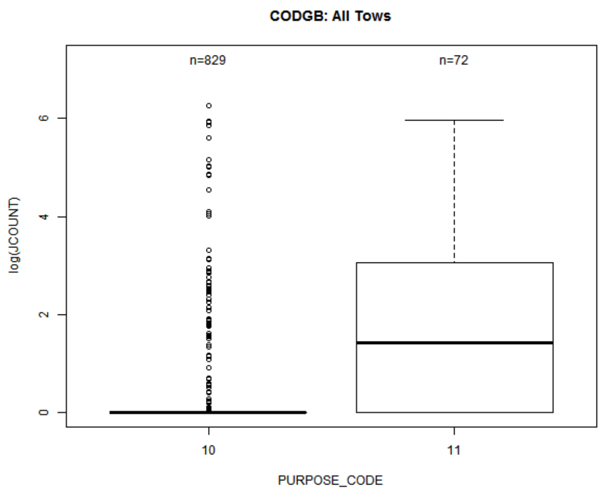

CODGB: Positive Tows
log(JCOUNT)
AVGDEPTH

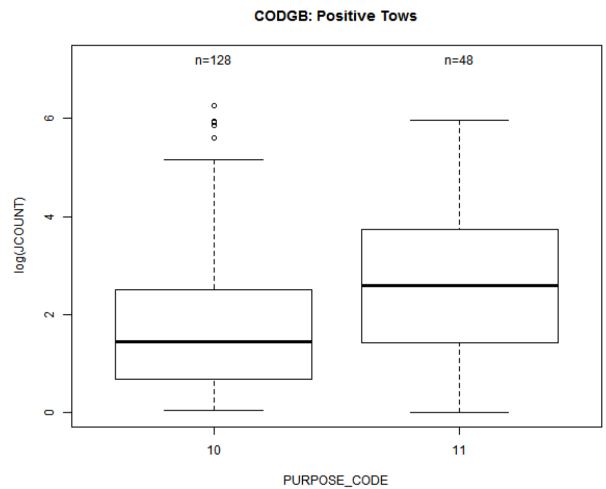

CODGB: All Tows
log(JCOUNT)
BOTTEMP

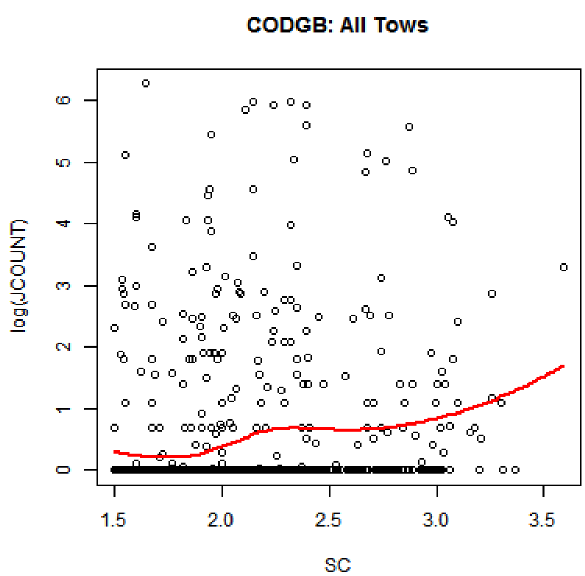

CODGB: Positive Tows
log(JCOUNT)
BOTTEMP

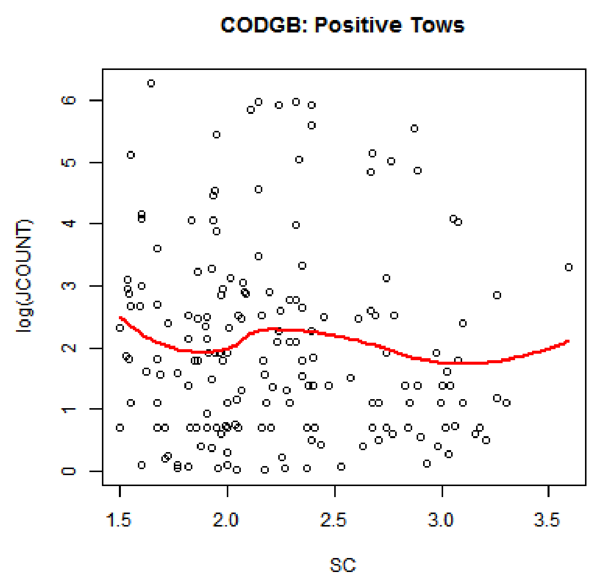

CODGB: All Tows
n=270
n=283
n=348
log(JCOUNT)
0
2
4
6
High Land Position
Low Land Position
Mid Land Position
LPI

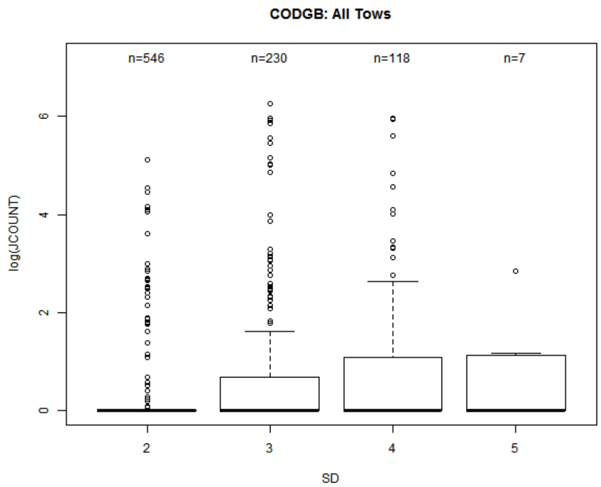

CODGB: Positive Tows
n=83
n=53
n=40
log(JCOUNT)
0
2
4
6
High Land Position
Low Land Position
Mid Land Position
LPI

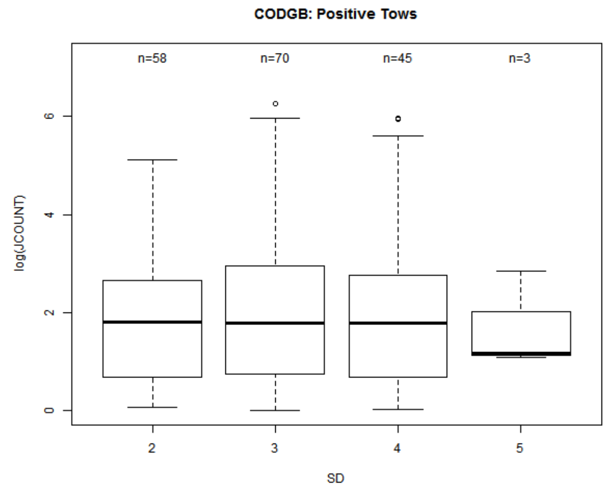

CODGB: All Tows
n=829
n=72
log(JCOUNT)
0
2
4
6
10
11
PURPOSE_CODE

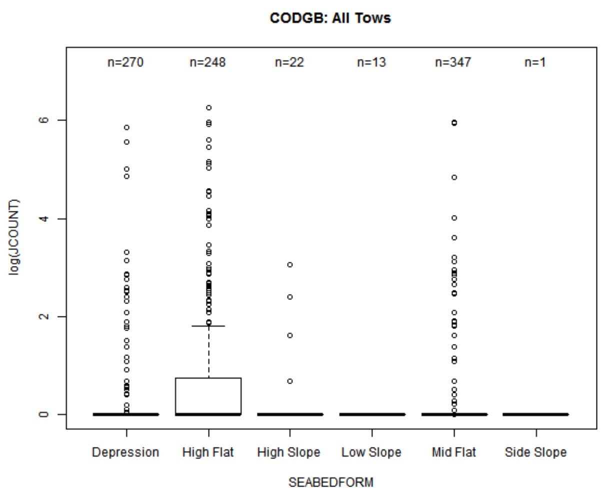

CODGB: Positive Tows
n=128
n=48
log(JCOUNT)
0
2
4
6
10
11
PURPOSE_CODE

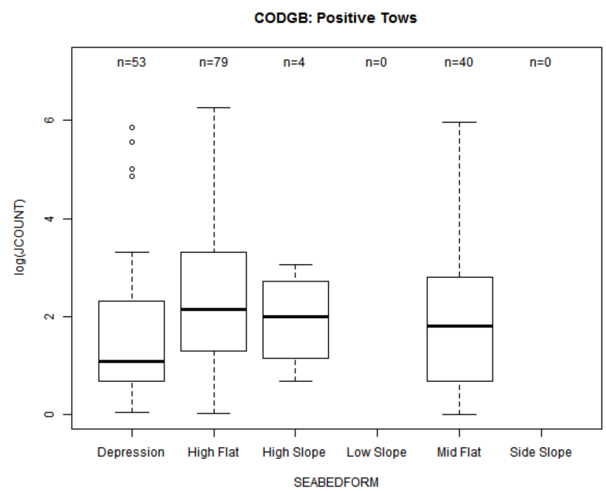

CODGB: All Tows
log(JCOUNT)
0
1
2
3
4
5
6
1.5
2.0
2.5
3.0
3.5
SC

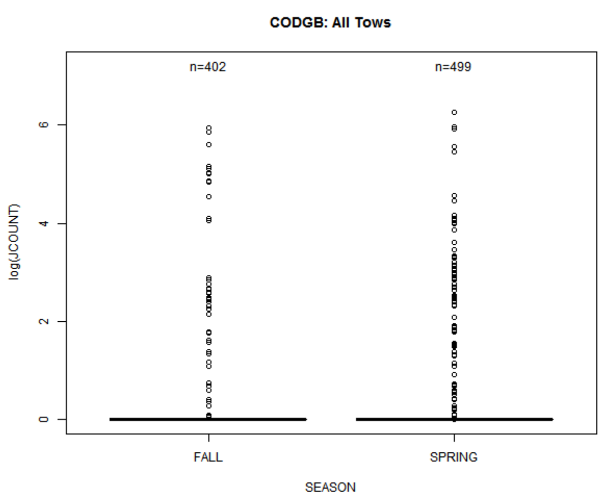

CODGB: Positive Tows
log(JCOUNT)
0
1
2
3
4
5
6
1.5
2.0
2.5
3.0
3.5
SC

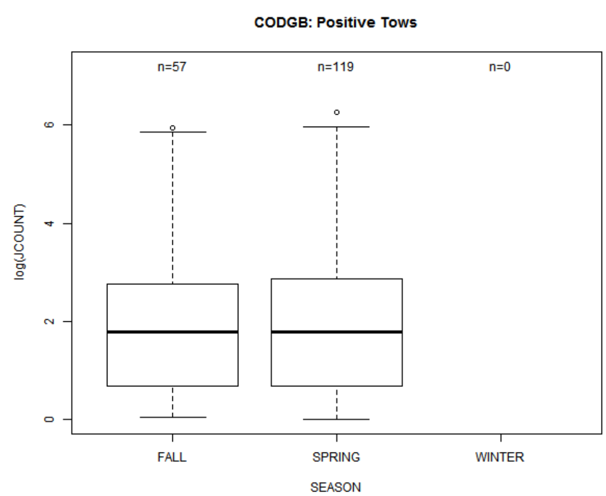

CODGB: All Tows
n=546
n=230
n=118
n=7
log(JCOUNT)
0
2
4
6
2
3
4
5
SD

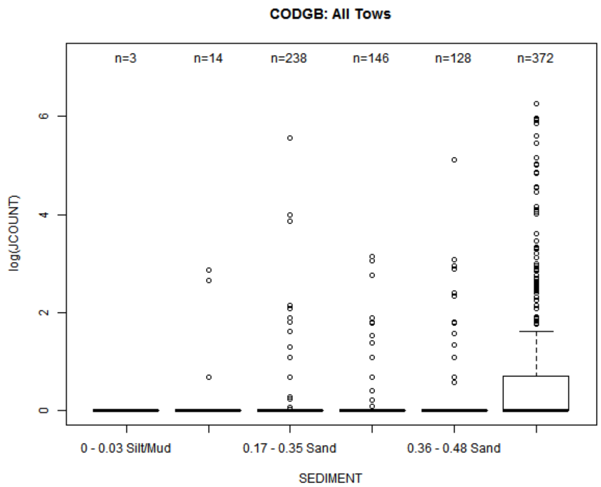

CODGB: Positive Tows
n=58
n=70
n=45
n=3
log(JCOUNT)
0
2
4
6
2
3
4
5
SD

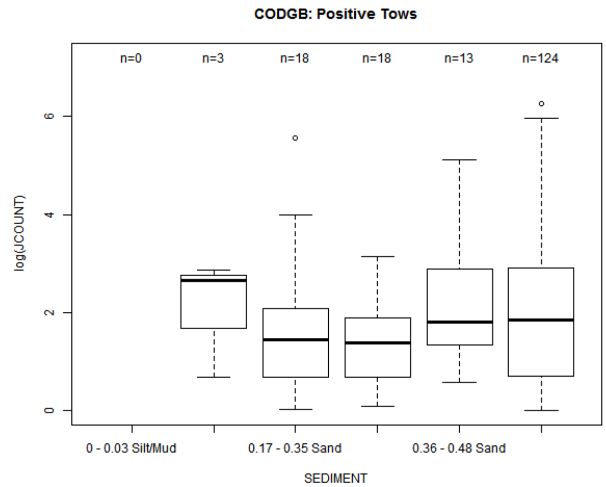

CODGB: All Tows
n=270
n=248
n=22
n=13
n=347
n=1
log(JCOUNT)
0
2
4
6
Depression
High Flat
High Slope
Low Slope
Mid Flat
Side Slope
SEABEDFORM

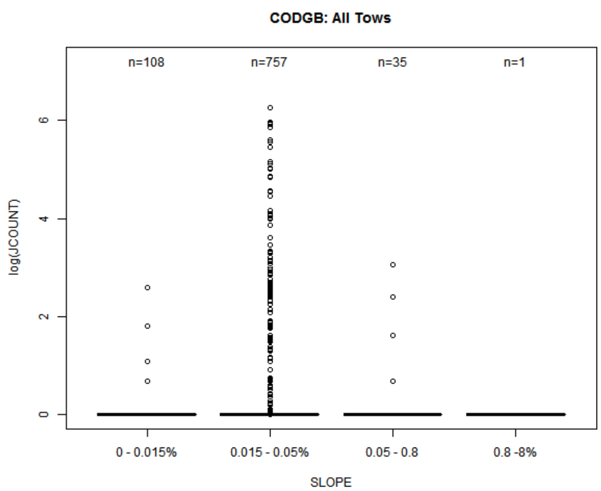

CODGB: Positive Tows
n=53
n=79
n=4
n=0
n=40
n=0
log(JCOUNT)
0
2
4
6
Depression
High Flat
High Slope
Low Slope
Mid Flat
Side Slope
SEABEDFORM

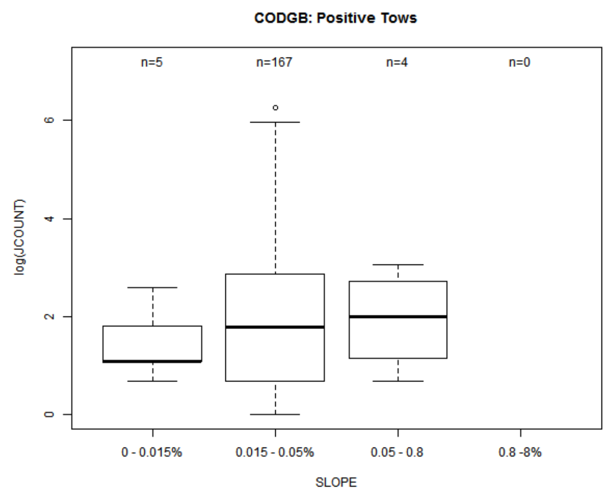

CODGB: All Tows
n=402
n=499
log(JCOUNT)
0
2
4
6
FALL
SPRING
SEASON

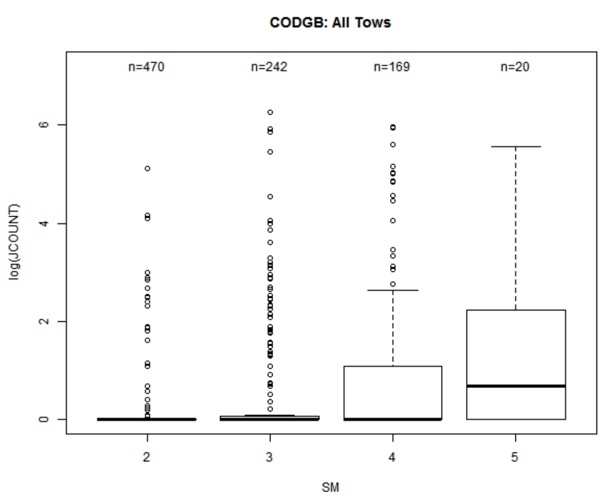

CODGB: Positive Tows
n=57
n=119
n=0
log(JCOUNT)
0
2
4
6
FALL
SPRING
WINTER
SEASON

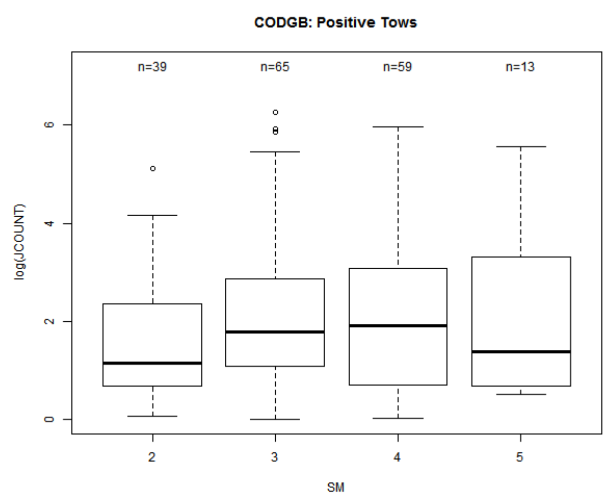

CODGB: All Tows
n=3
n=14
n=238
n=146
n=128
n=372
log(JCOUNT)
0
2
4
6
0 - 0.03 Silt/Mud
0.17 - 0.35 Sand
0.36 - 0.48 Sand
SEDIMENT

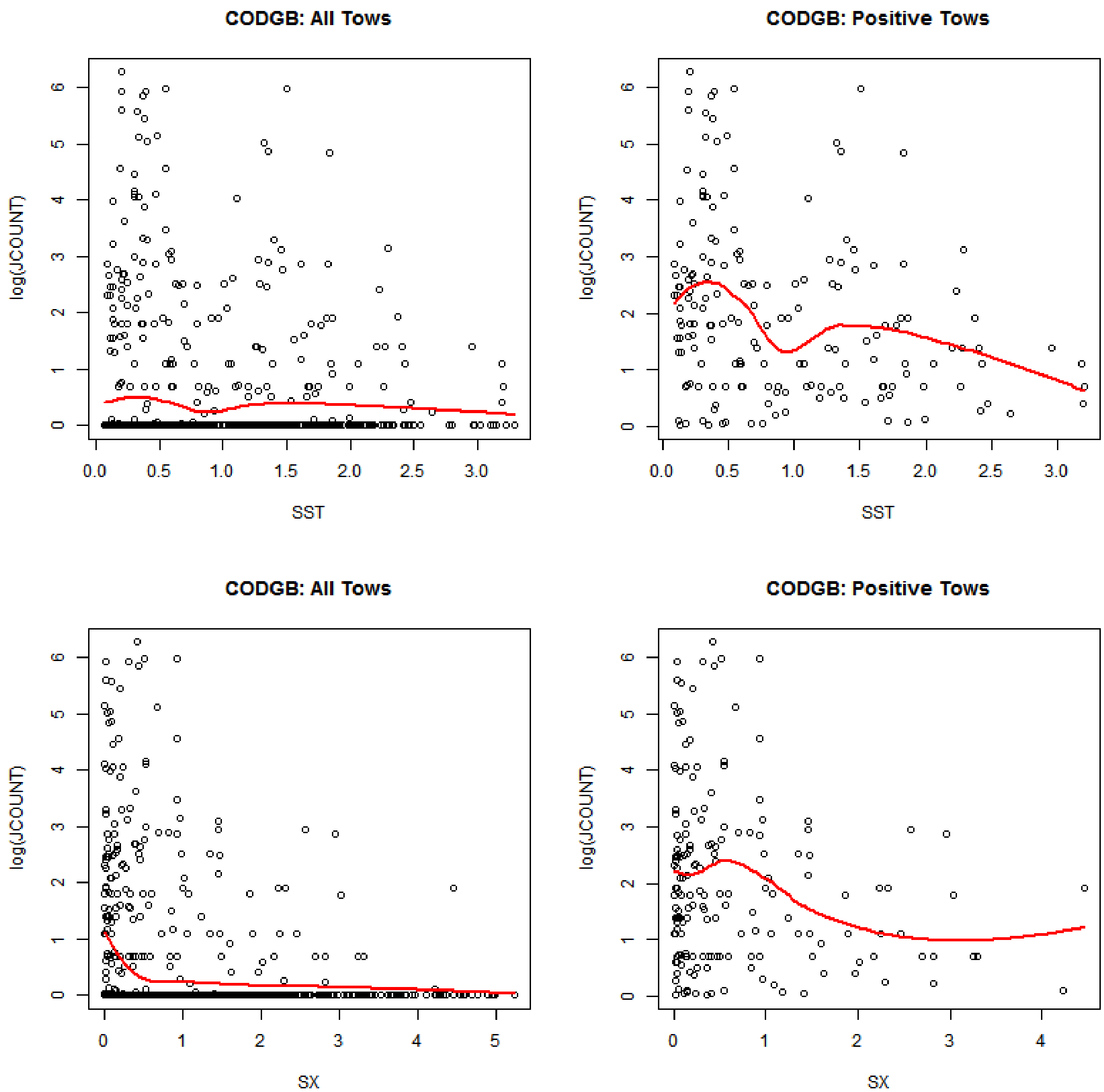

CODGB: Positive Tows
n=0
n=3
n=18
n=18
n=13
n=124
log(JCOUNT)
0
2
4
6
0 - 0.03 Silt/Mud
0.17 - 0.35 Sand
0.36 - 0.48 Sand
SEDIMENT

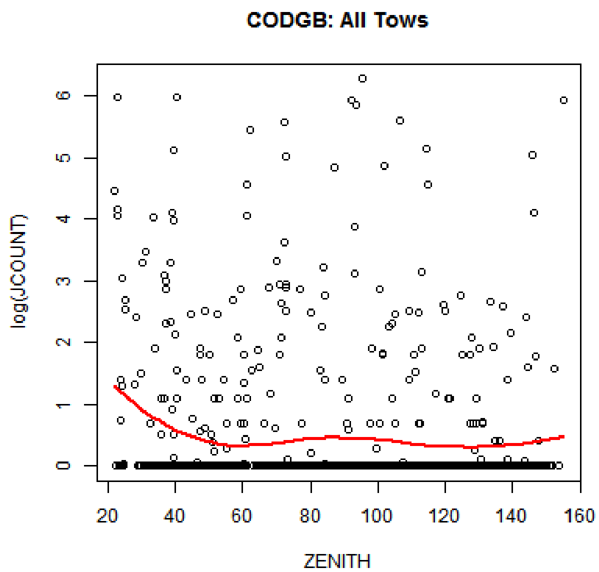

CODGB: All Tows
n=108
n=757
n=35
n=1
log(JCOUNT)
0
2
4
6
0 - 0.015%
0.015 - 0.05%
0.05 - 0.8
0.8 -8%
SLOPE

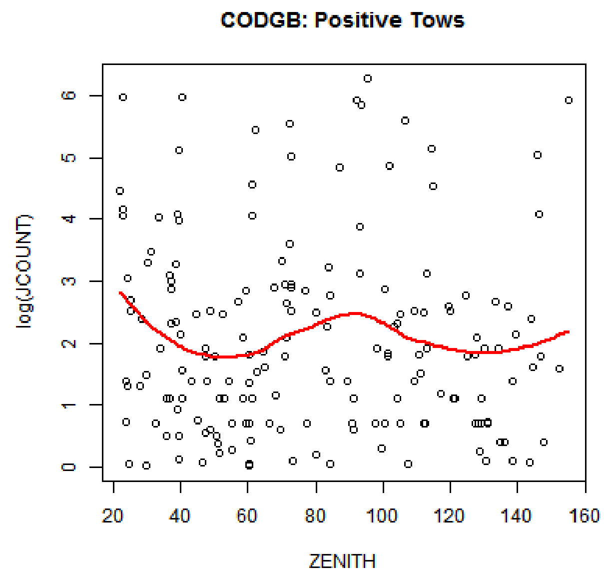

CODGB: Positive Tows
n=5
n=167
n=4
n=0
log(JCOUNT)
0
2
4
6
0 - 0.015%
0.015 - 0.05%
0.05 - 0.8
0.8 -8%
SLOPE


CODGB: All Tows
n=470
n=242
n=169
n=20
log(JCOUNT)
0
2
4
6
2
3
4
5
SM

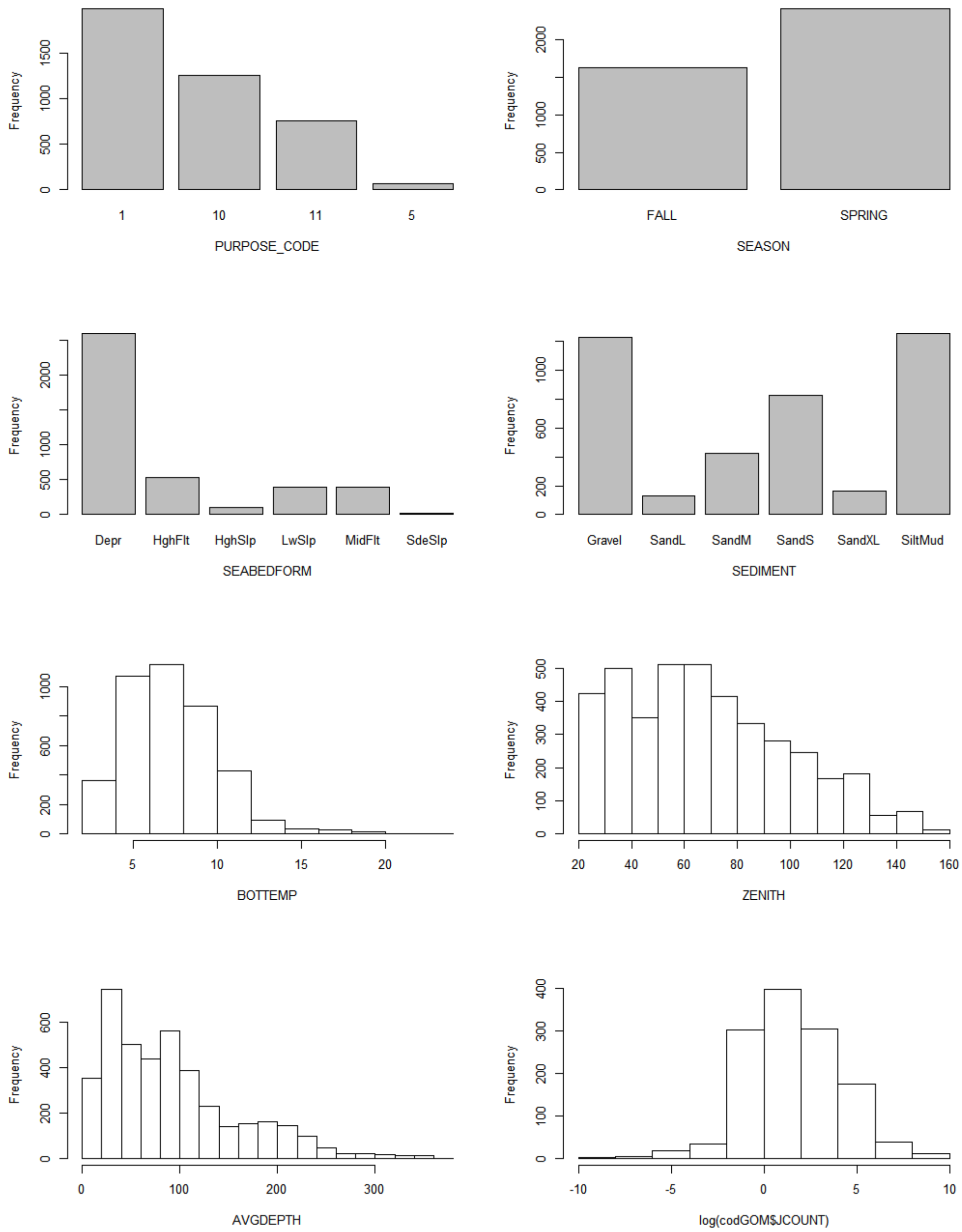

CODGB: Positive Tows
n=39
n=65
n=59
n=13
log(JCOUNT)
0
2
4
6
2
3
4
5
SM

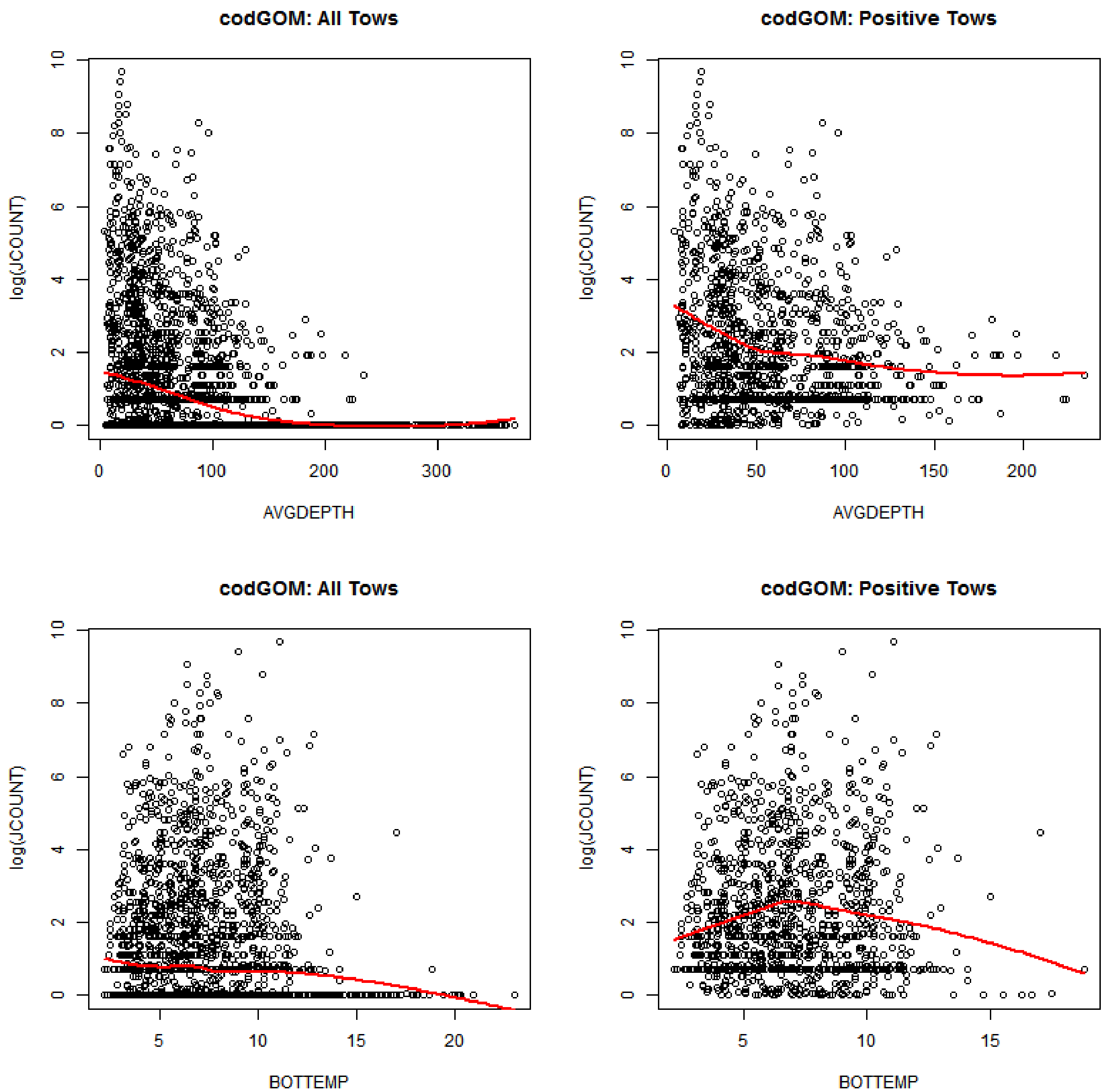

CODGB: All Tows
log(JCOUNT)
SST
CODGB: Positive Tows
log(JCOUNT)
SST
CODGB: All Tows
log(JCOUNT)
SX
CODGB: Positive Tows
log(JCOUNT)
SX


CODGB: All Tows
log(JCOUNT)
ZENITH
0
1
2
3
4
5
6
20
40
60
80
100
120
140
160

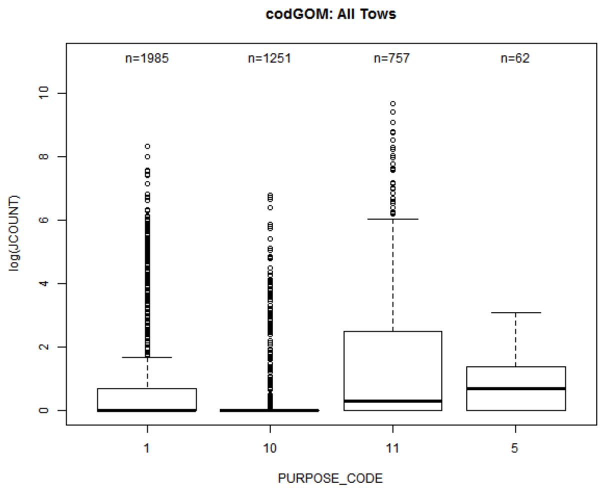

CODGB: Positive Tows
log(JCOUNT)
ZENITH
0
1
2
3
4
5
6
20
40
60
80
100
120
140
160

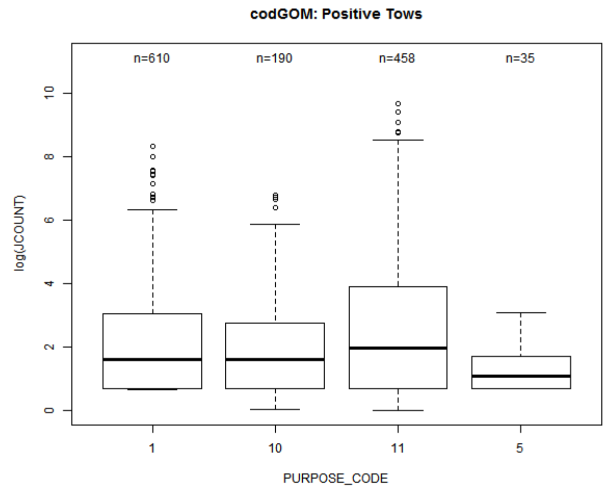
APPENDIX 4: Premodeling Gulf of Maine cod analysis

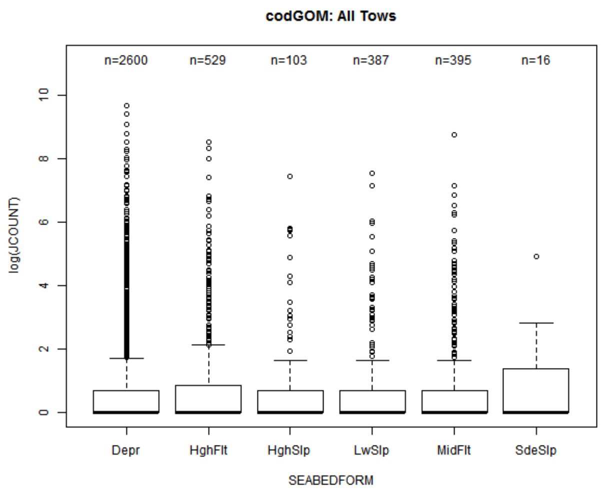


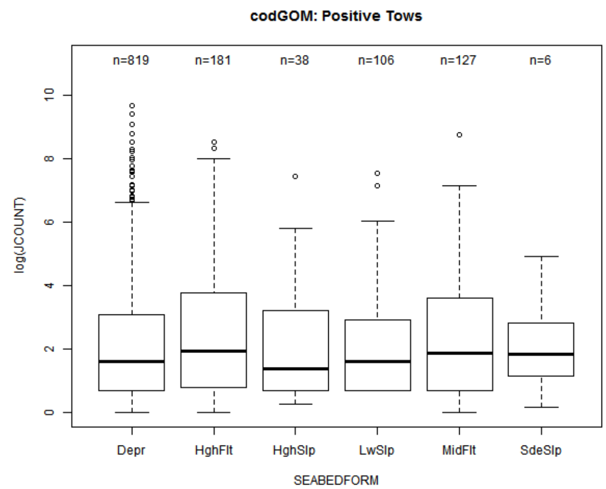


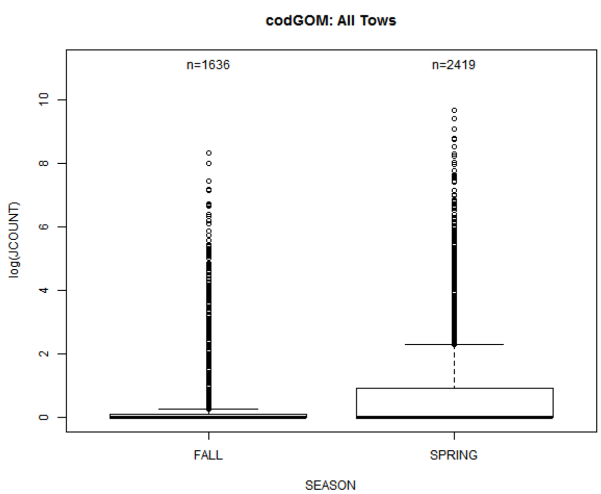
codGOM: All Tows
codGOM: Positive Tows
log(JCOUNT)
AVGDEPTH
codGOM: All Tows
codGOM: Positive Tows
log(JCOUNT)
BOTTEMP

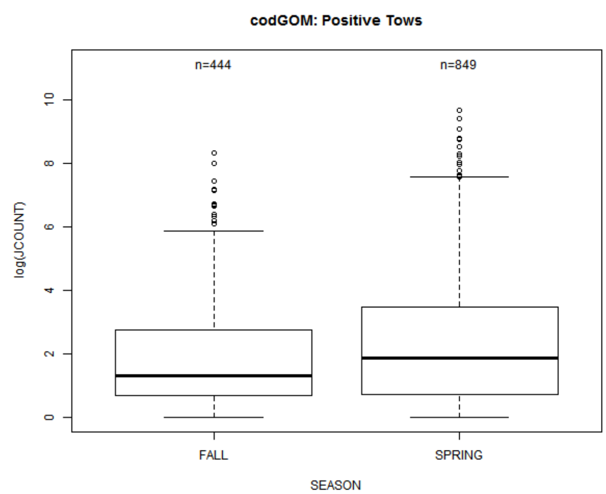

codGOM: All Tows
n=1985
n=1251
n=757
n=62
log(JCOUNT)
1
10
11
5
PURPOSE_CODE

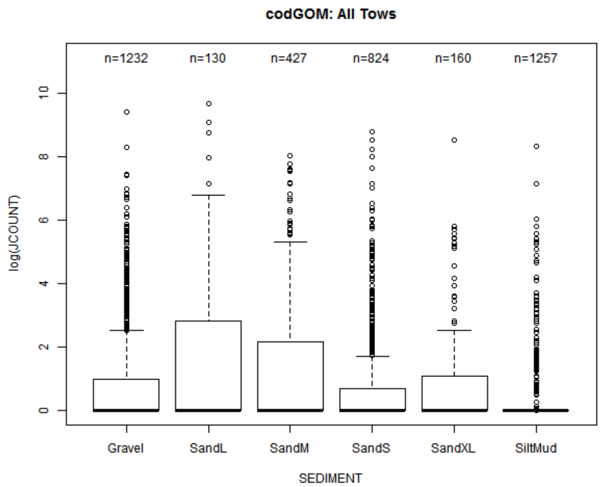

codGOM: Positive Tows
n=610
n=190
n=458
n=35
log(JCOUNT)
1
10
11
5
PURPOSE_CODE

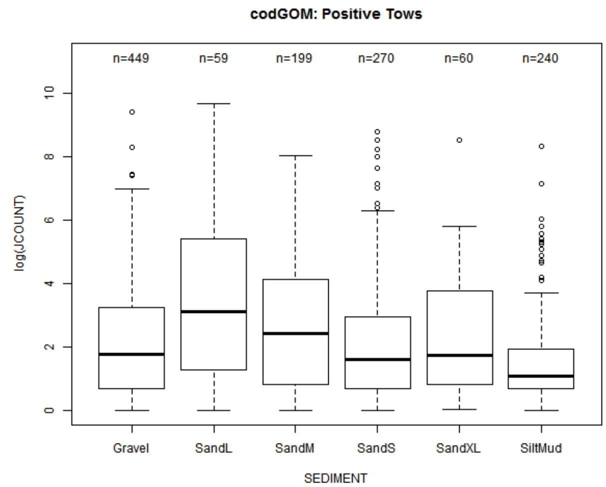

codGOM: All Tows
n=2600
n=529
n=103
n=387
n=395
n=16
log(JCOUNT)
Depr
HghFlt
HghSlp
LwSlp
MidFlt
SdeSlp
SEABEDFORM

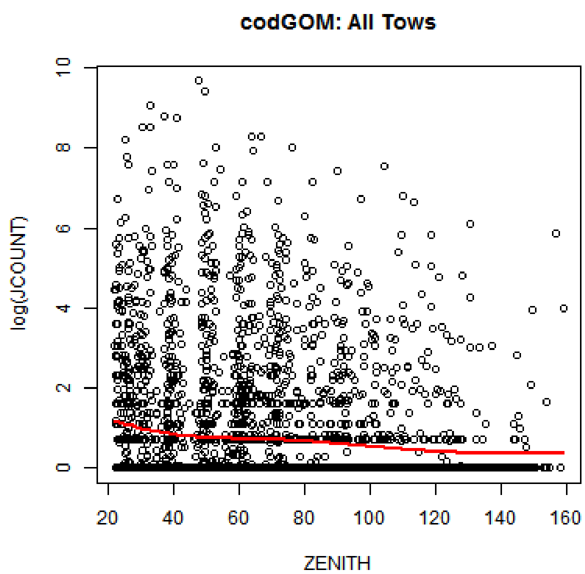

codGOM: Positive Tows
n=819
n=181
n=38
n=106
n=127
n=6
log(JCOUNT)
Depr
HghFlt
HghSlp
LwSlp
MidFlt
SdeSlp
SEABEDFORM

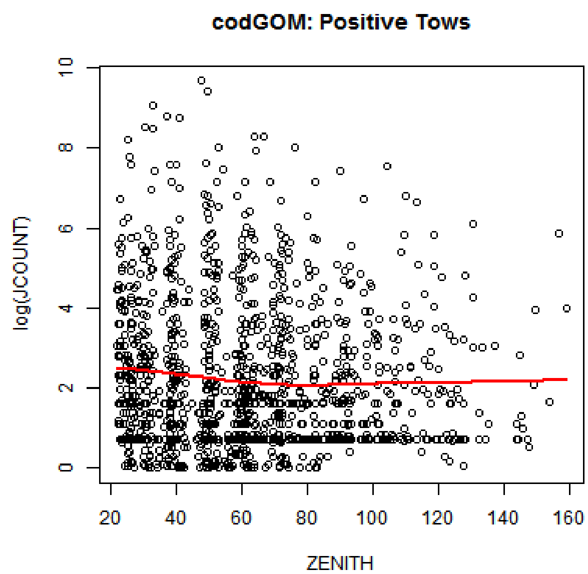

codGOM: All Tows
n=1636
n=2419
log(JCOUNT)
FALL
SPRING
SEASON


codGOM: Positive Tows
n=444
n=849
log(JCOUNT)
FALL
SPRING
SEASON

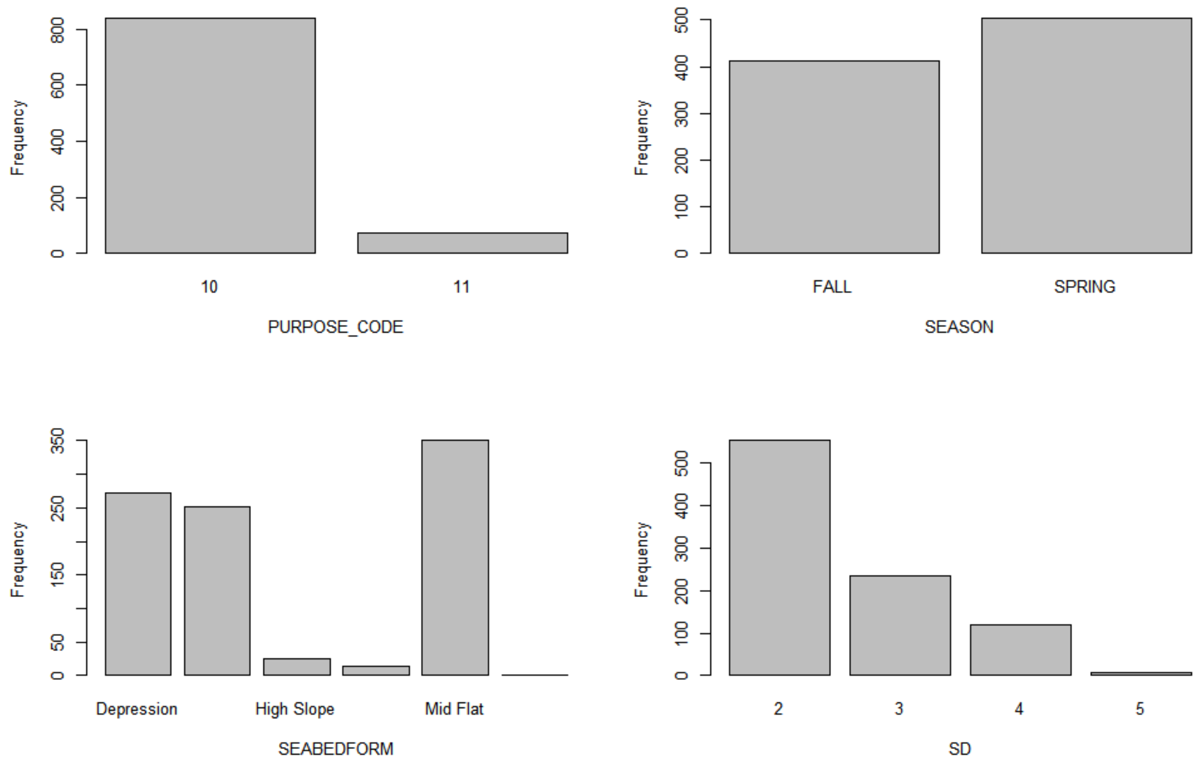

codGOM: All Tows
n=1232
n=130
n=427
n=824
n=160
n=1257
log(JCOUNT)
Gravel
SandL
SandM
SandS
SandXL
SiltMud
SEDIMENT

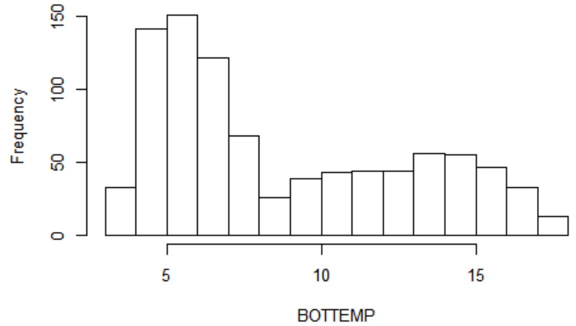

codGOM: Positive Tows
n=449
n=59
n=199
n=270
n=60
n=240
log(JCOUNT)
Gravel
SandL
SandM
SandS
SandXL
SiltMud
SEDIMENT

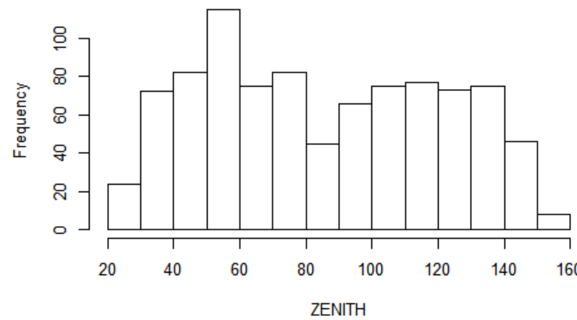

codGOM: All Tows
log(JCOUNT)
ZENITH

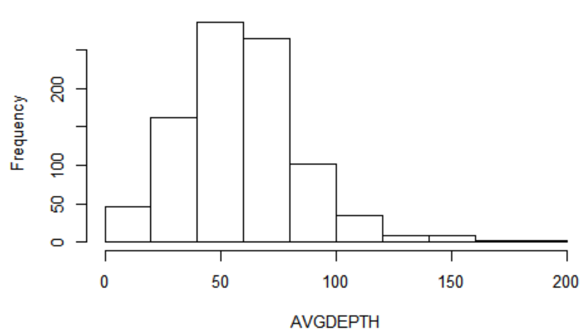

codGOM: Positive Tows
log(JCOUNT)
ZENITH

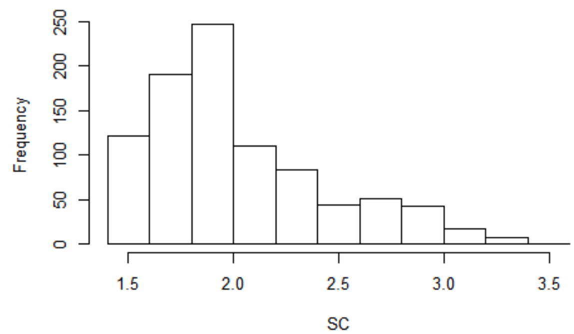
APPENDIX 5: Premodeling Georges Bank yellowtail flounder analysis

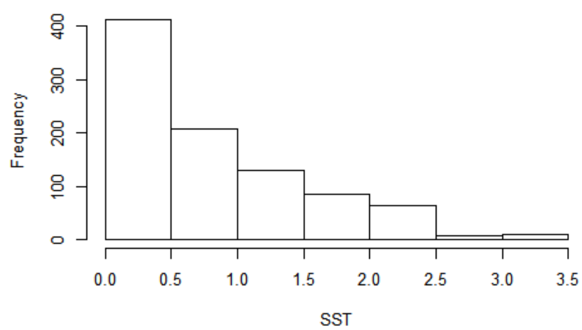


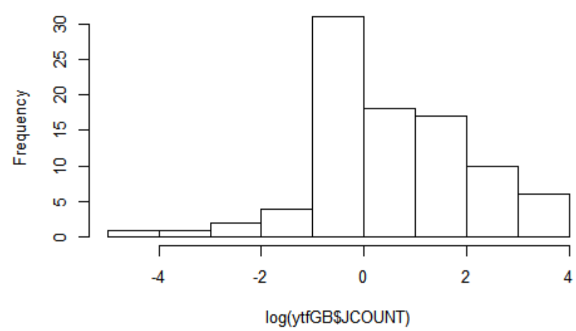

Frequency
0
50
100
150
5
10
15
BOTTEMP

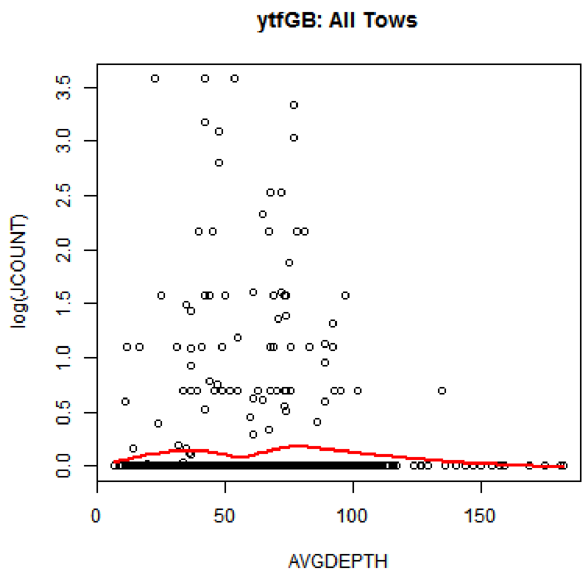

Frequency
0
20
40
60
80
100
20
40
60
80
100
120
140
160
ZENITH

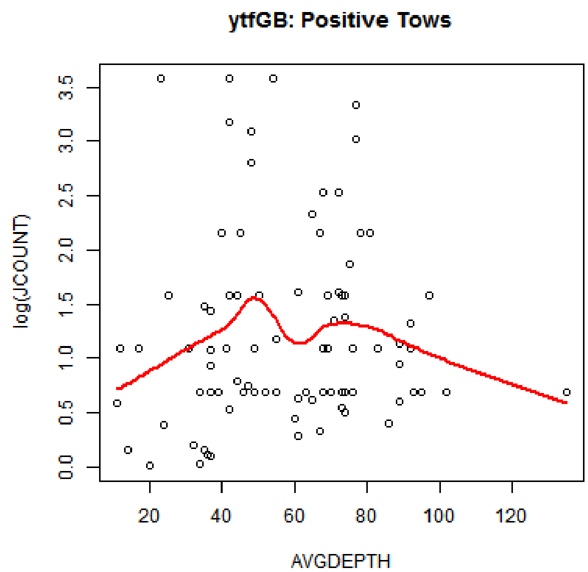

Frequency
0
50
100
200
0
50
100
150
200
AVGDEPTH

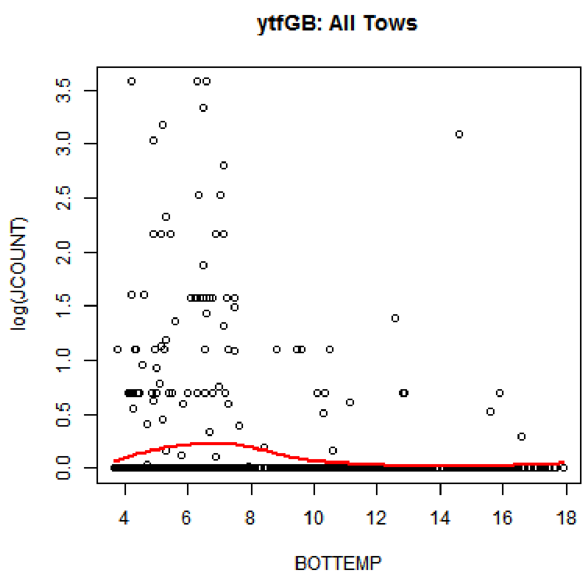

Frequency
0
50
100
150
200
250
1.5
2.0
2.5
3.0
3.5
SC

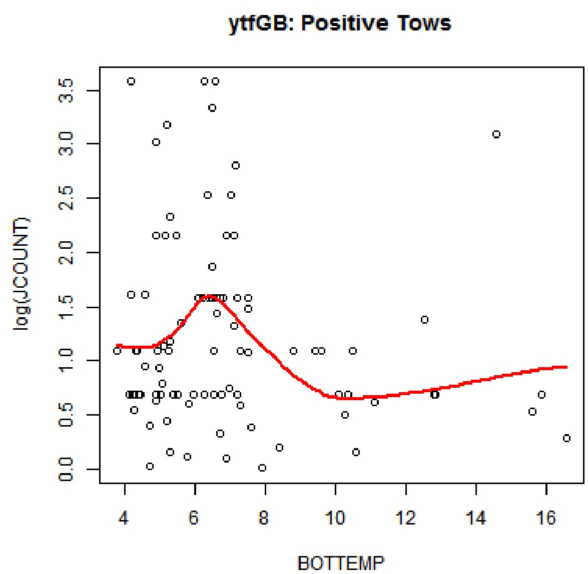

Frequency
0
100
200
300
400
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
SST

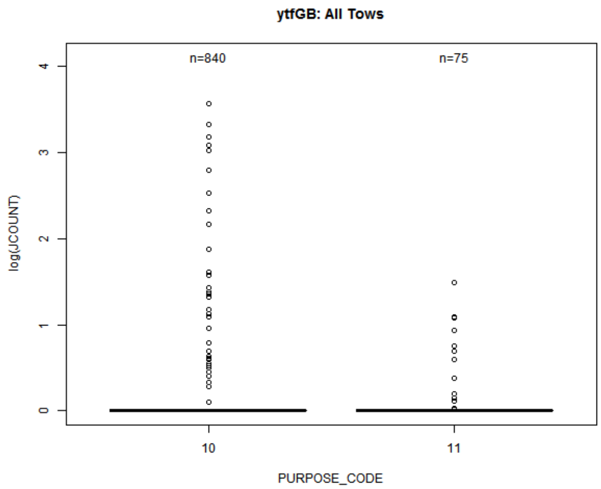

Frequency
0
5
10
15
20
25
30
-4
-2
0
2
4
log(ytfGB$JCOUNT)

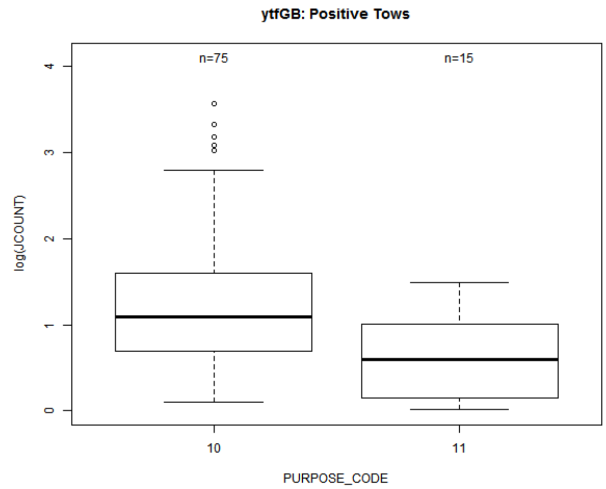

ytfGB: All Tows
log(JCOUNT)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
50
100
150
AVGDEPTH

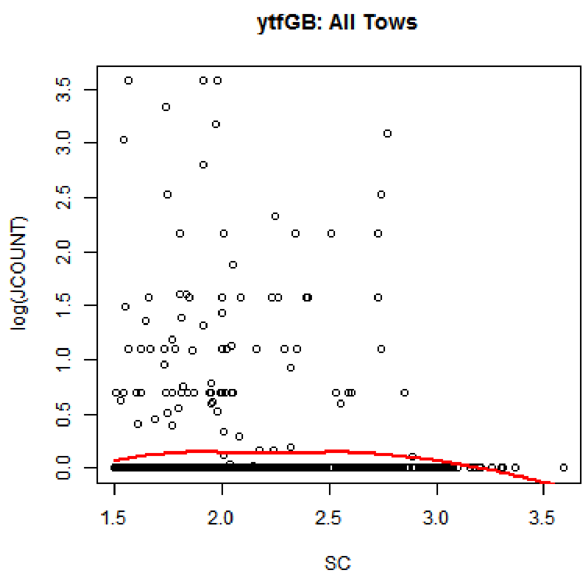

ytfGB: Positive Tows
log(JCOUNT)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
20
40
60
80
100
120
AVGDEPTH

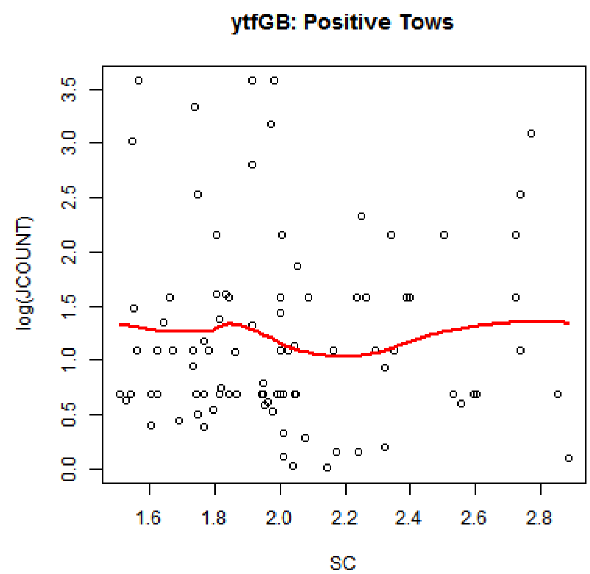

ytfGB: All Tows
log(JCOUNT)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4
6
8
10
12
14
16
18
BOTTEMP

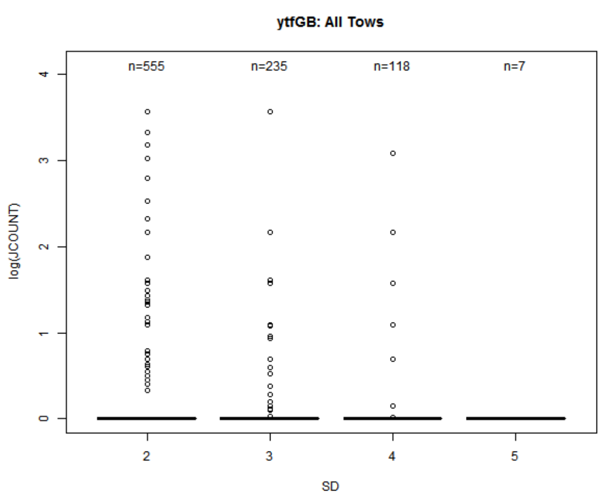

ytfGB: Positive Tows
log(JCOUNT)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4
6
8
10
12
14
16
BOTTEMP

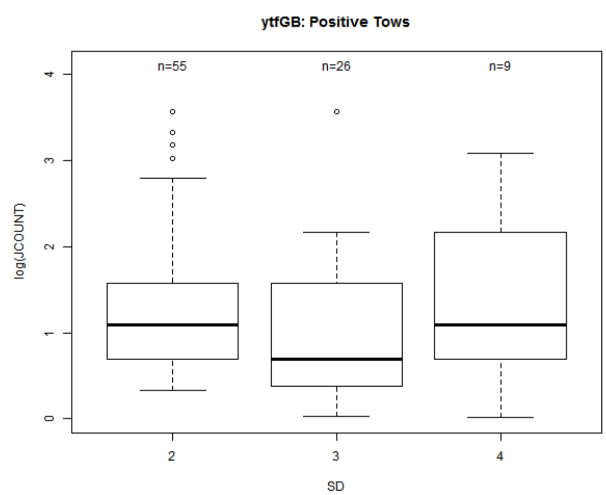

ytfGB: All Tows
n=840
n=75
log(JCOUNT)
PURPOSE_CODE

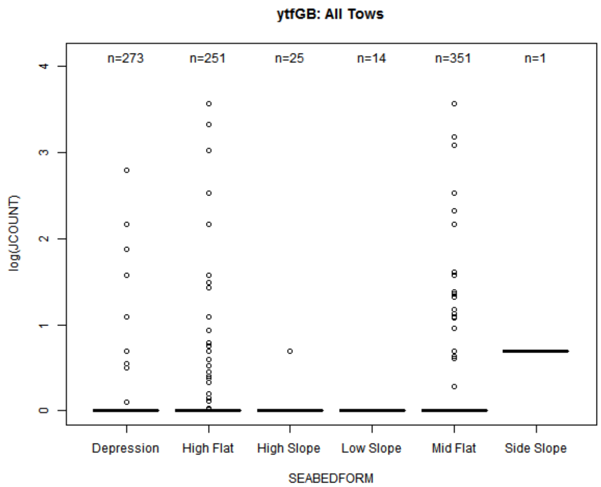

ytfGB: Positive Tows
n=75
n=15
log(JCOUNT)
PURPOSE_CODE

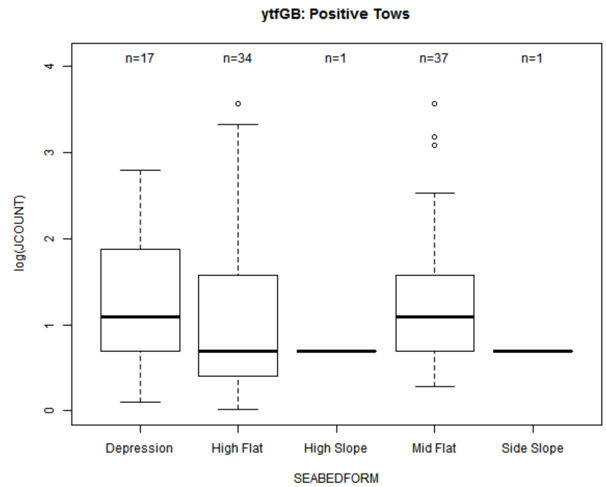

ytfGB: All Tows
log(JCOUNT)
SC

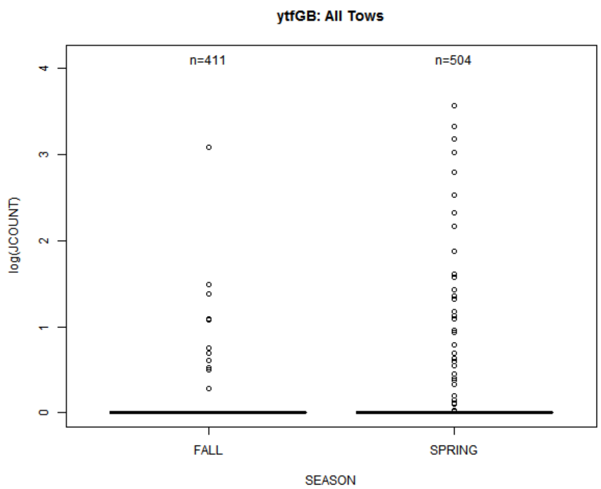

ytfGB: Positive Tows
log(JCOUNT)
SC

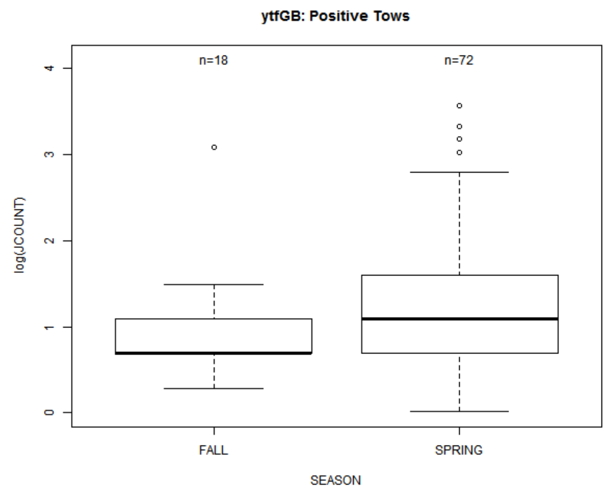

ytfGB: All Tows
n=555
n=235
n=118
n=7
log(JCOUNT)
SD

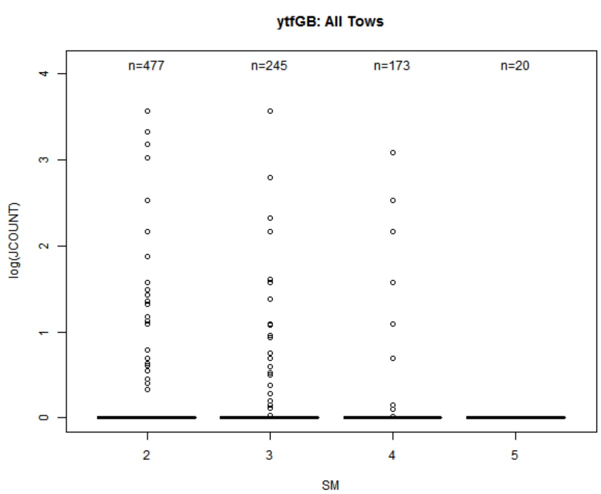

ytfGB: Positive Tows
n=55
n=26
n=9
log(JCOUNT)
SD

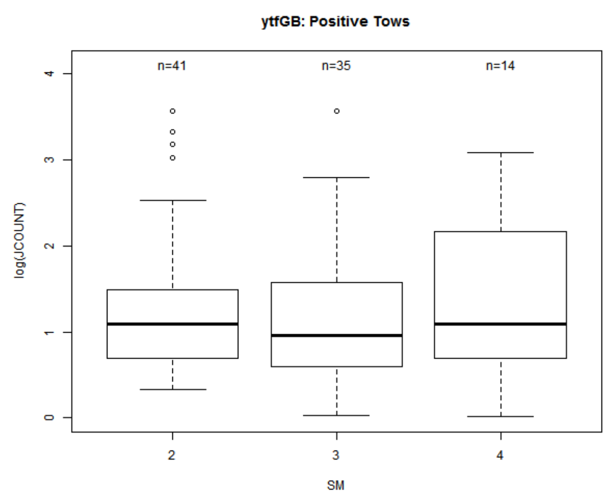

ytfGB: All Tows
n=273
n=251
n=25
n=14
n=351
n=1
log(JCOUNT)
0
1
2
3
4
Depression
High Flat
High Slope
Low Slope
Mid Flat
Side Slope
SEABEDFORM

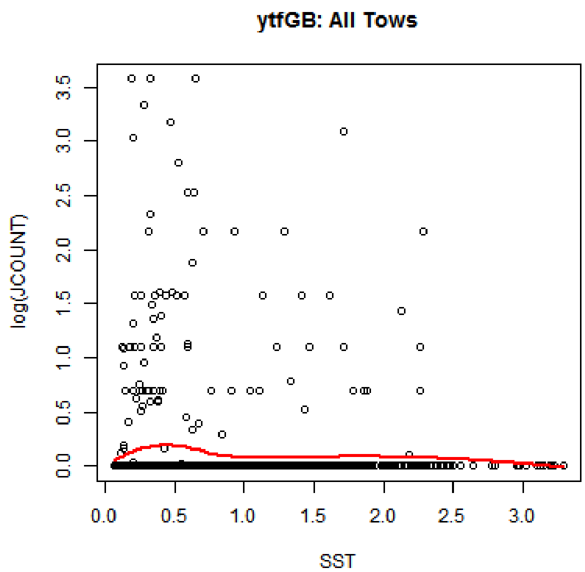

ytfGB: Positive Tows
n=17
n=34
n=1
n=37
n=1
log(JCOUNT)
0
1
2
3
4
Depression
High Flat
High Slope
Mid Flat
Side Slope
SEABEDFORM

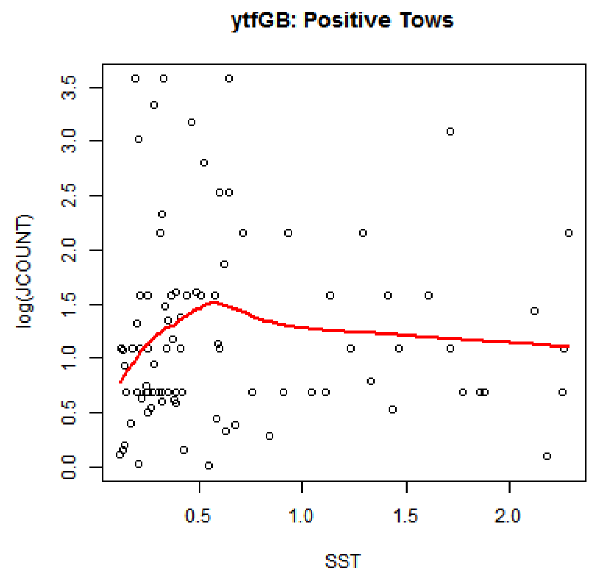

ytfGB: All Tows
n=411
n=504
log(JCOUNT)
0
1
2
3
4
FALL
SPRING
SEASON

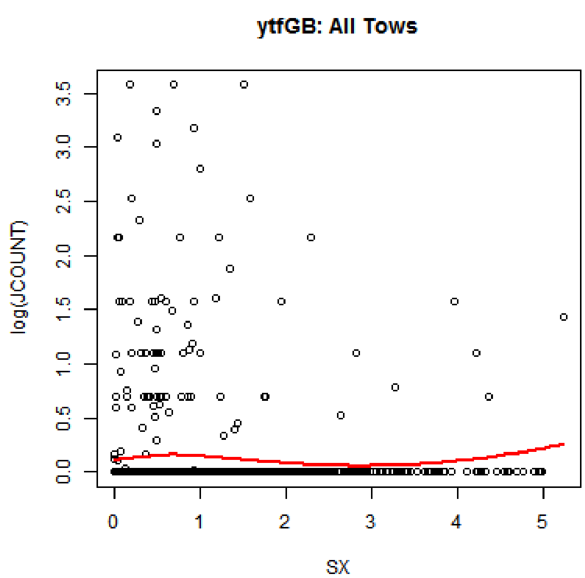

ytfGB: Positive Tows
n=18
n=72
log(JCOUNT)
0
1
2
3
4
FALL
SPRING
SEASON

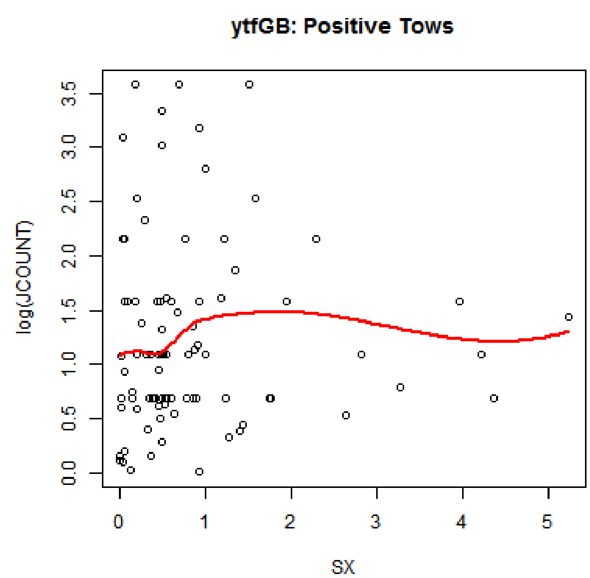

ytfGB: All Tows
n=477
n=245
n=173
n=20
log(JCOUNT)
0
1
2
3
4
2
3
4
5
SM

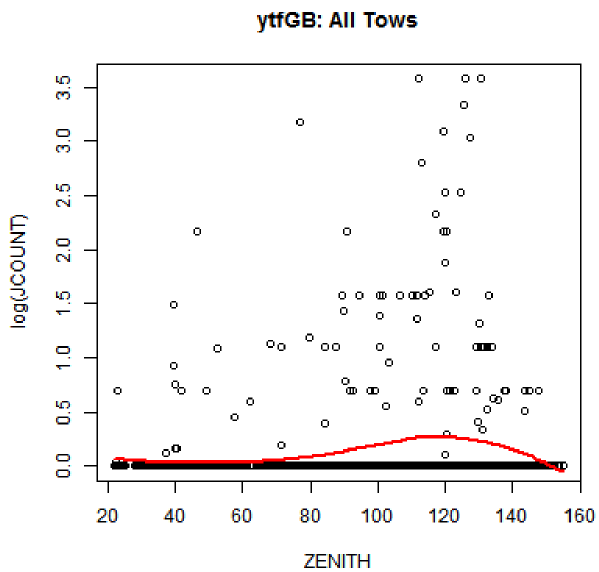

ytfGB: Positive Tows
n=41
n=35
n=14
log(JCOUNT)
0
1
2
3
4
2
3
4
SM


ytfGB: All Tows
log(JCOUNT)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
SST


ytfGB: Positive Tows
log(JCOUNT)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0.5
1.0
1.5
2.0
SST


ytfGB: All Tows
log(JCOUNT)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
1
2
3
4
5
SX


ytfGB: Positive Tows
log(JCOUNT)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
0
1
2
3
4
5
SX



ytfGB: All Tows
log(JCOUNT)
ZENITH
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
20
40
60
80
100
120
140
160
ytfGB: Positive Tows
log(JCOUNT)
ZENITH
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
20
40
60
80
100
120
140

APPENDIX 5: R output for Generalized Additive Models

GB COD

Presence-absence:

```
Family: binomial
Link function: logit

Formula:
JPA ~ s(SC) + s(SST) + s(BOTTEMP) + s(ZENITH) + s(AVGDEPTH) +
    PURPOSE_CODE + SEASON
<environment: 0x000000000cfa3a40>

Parametric coefficients:
                Estimate Std. Error  z value   Pr(>|z|)
(Intercept)    -0.737896   0.252449 -2.92295  0.0034673 **
PURPOSE_CODE11  2.999776   0.555045  5.40457 6.4966e-08 ***
SEASONSPRING   -2.938084   0.479922 -6.12200 9.2409e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                edf  Ref.df   Chi.sq    p-value
s(SC)       1.00003 1.00006 36.94210 1.2176e-09 ***
s(SST)      2.33007 2.94928  6.21400 0.09800112 .
s(BOTTEMP)  2.02435 2.55942 66.22891 6.9660e-14 ***
s(ZENITH)   1.00011 1.00021  4.48607 0.03418393 *
s(AVGDEPTH) 5.82263 6.90603 24.75804 0.00080294 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =   0.33   Deviance explained = 31.8%
UBRE score = -0.29305  Scale est. = 1         n = 901
```

Conditional presence:

```
Family: gaussian
Link function: identity

Formula:
LJCOUNT ~ s(SST) + PURPOSE_CODE + SEASON
<environment: 0x000000000d3f7218>

Parametric coefficients:
                Estimate Std. Error  t value   Pr(>|t|)
(Intercept)     1.634190   0.274850  5.94576 1.4956e-08 ***
PURPOSE_CODE11  0.420564   0.407628  1.03173    0.30365
SEASONSPRING   -0.310542   0.352346 -0.88135    0.37936
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
       edf Ref.df       F  p-value
s(SST)   1      1 4.96494 0.027137 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.0448   Deviance explained = 6.11%
GCV score = 4.3217  Scale est. = 4.2235    n = 176
```

GOM COD

Presence-absence:

```
Family: binomial
Link function: logit

Formula:
JPA ~ SEDIMENT + SEABEDFORM + s(BOTTEMP) + s(AVGDEPTH) + PURPOSE_CODE +
    SEASON
<environment: 0x000000000d3a6100>

Parametric coefficients:
                   Estimate  Std. Error  z value   Pr(>|z|)
(Intercept)       -1.02655076  0.13251625 -7.74660 9.4384e-15 ***
SEDIMENTSandL     -0.33211052  0.22657209 -1.46581  0.1427014
SEDIMENTSandM      0.00755436  0.13386025  0.05643  0.9549955
SEDIMENTSandS     -0.28345951  0.10980305 -2.58153  0.0098364 **
SEDIMENTSandXL    -0.33175598  0.19585904 -1.69385  0.0902936 .
SEDIMENTSiltMud   -0.78397190  0.10351551 -7.57347 3.6338e-14 ***
SEABEDFORMHghFlt   0.70570896  0.12627017  5.58888 2.2854e-08 ***
SEABEDFORMHghSlp   0.54682707  0.23807505  2.29687  0.0216263 *
SEABEDFORMLwSlp    0.04026009  0.13453752  0.29925  0.7647508
SEABEDFORMMidFlt   0.20738550  0.13766122  1.50649  0.1319410
SEABEDFORMSdeSlp   0.10661293  0.64902397  0.16427  0.8695213
PURPOSE_CODE5      0.05166306  0.32139607  0.16075  0.8722936
PURPOSE_CODE10    -0.18276372  0.12173152 -1.50137  0.1332606
PURPOSE_CODE11     1.06381109  0.11463685  9.27984 < 2.22e-16 ***
SEASONSPRING      -0.22420977  0.14147846 -1.58476  0.1130203
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                edf  Ref.df   Chi.sq    p-value
s(BOTTEMP)  7.15089 8.02022  99.5606 < 2.22e-16 ***
s(AVGDEPTH) 2.99783 3.79881 154.6007 < 2.22e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.233   Deviance explained = 20.7%
UBRE score = 0.0031764  Scale est. = 1         n = 4030
```

Conditional presence:

```
Family: gaussian
Link function: identity

Formula:
LJCOUNT ~ SEDIMENT + s(BOTTEMP) + s(ZENITH) + s(AVGDEPTH) + PURPOSE_CODE +
    SEASON
<environment: 0x000000000c858308>

Parametric coefficients:
                 Estimate Std. Error  t value   Pr(>|t|)
(Intercept)      1.248910   0.191071  6.53637 9.1365e-11 ***
SEDIMENTSandL    0.784456   0.343634  2.28283  0.0226067 *
SEDIMENTSandM    0.381939   0.203525  1.87662  0.0608012 .
SEDIMENTSandS   -0.132415   0.182710 -0.72473  0.4687536
SEDIMENTSandXL   0.274475   0.320227  0.85713  0.3915384
SEDIMENTSiltMud -0.503095   0.192048 -2.61963  0.0089084 **
PURPOSE_CODE5   -0.566216   0.482631 -1.17319  0.2409434
PURPOSE_CODE10  -0.169725   0.214314 -0.79195  0.4285412
PURPOSE_CODE11  -0.710416   0.171520 -4.14188 3.6745e-05 ***
SEASONSPRING     1.201488   0.214426  5.60328 2.5816e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                edf  Ref.df        F    p-value
s(BOTTEMP)  6.94784 8.01288  6.29133 5.0284e-08 ***
s(ZENITH)   1.10171 1.19648  2.60141   0.097887 .
s(AVGDEPTH) 1.00000 1.00000 15.67671 7.9248e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
R-sq.(adj) =     0.1   Deviance explained = 11.3%
GCV score = 5.3659  Scale est. = 5.2859    n = 1277
```

GB YELLOWTAIL FLOUNDER

Presence-absence:

```
Family: binomial
Link function: logit

Formula:
JPA ~ SEABEDFORM + s(SC) + s(ZENITH) + s(AVGDEPTH) + PURPOSE_CODE +
    SEASON
<environment: 0x000000000c45d8e0>

Parametric coefficients:
                        Estimate  Std. Error  z value   Pr(>|z|)
(Intercept)          -4.69930e+00 4.32743e-01 -10.85934 < 2.22e-16 ***
SEABEDFORMHigh Flat   8.47769e-01 3.58846e-01   2.36249  0.018153 *
SEABEDFORMHigh Slope  1.13286e-01 1.11495e+00   0.10161  0.919070
SEABEDFORMLow Slope  -1.29852e+02 1.79356e+07  -0.00001  0.999994
SEABEDFORMMid Flat    5.44139e-01 3.39981e-01   1.60050  0.109488
SEABEDFORMSide Slope  1.45514e+02 6.71089e+07   0.00000  0.999998
PURPOSE_CODE11        3.44180e+00 6.54014e-01   5.26258 1.4205e-07 ***
SEASONSPRING          1.57767e+00 3.07405e-01   5.13223 2.8633e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                edf  Ref.df   Chi.sq    p-value
s(SC)       2.20239 2.76797  6.89979   0.063309 .
s(ZENITH)   2.54518 3.21132 44.74094 2.2056e-09 ***
s(AVGDEPTH) 4.32232 5.18676 11.78614   0.042646 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.151   Deviance explained = 23.3%
UBRE score = -0.46944  Scale est. = 1         n = 915
```

Conditional presence:

```
Family: gaussian
Link function: identity

Formula:
LJCOUNT ~ s(SC) + s(BOTTEMP) + s(AVGDEPTH) + PURPOSE_CODE + SEASON
<environment: 0x000000000bd95f58>

Parametric coefficients:
                Estimate Std. Error  t value   Pr(>|t|)
(Intercept)     1.541077   0.455266  3.38500  0.0011424 **
PURPOSE_CODE11 -2.700304   0.430554 -6.27169 2.1756e-08 ***
SEASONSPRING   -0.625619   0.530054 -1.18029  0.2416687
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                edf  Ref.df       F   p-value
s(SC)       8.39251 8.89552 3.44734 0.0013499 **
s(BOTTEMP)  3.71048 4.56325 2.68667 0.0317813 *
s(AVGDEPTH) 1.00000 1.00000 7.98539 0.0060276 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.432   Deviance explained = 52.9%
GCV score = 1.4111  Scale est. = 1.1586    n = 90
```