Video index of abundance for red snapper along the United States Southeast Atlantic coast,

2010 – 2012: am evaluation using multiple modeling approaches

Nathan M. Bacheler and Lewis Coggins

Summary Delta GAM

* 2010-2012 SEFIS and MARMAP video data
* Reconnaissance stations included if they were included in sampling universe the following year (MARMAP “recon” stations excluded because I couldn’t determine which recon stations were included, and the fact that the number of recon stations was low)
* All latitudes and soak times
* Depth < 100 m
* Removed missing depth, sampling durations, and bottom temps
* Binomial model for presence-absence
* 4th root Gaussian model for positive MeanCount data (fit better than lognormal, but still not very good)
* Final binomial and Gaussian models included variables if they were included in either model based on AIC and GCV
* Will eventually include calibration information to adjust MeanCount in 2010, since a different camera was used in that year
* 2013 will also eventually be added when it becomes available

# *Full data subsetting:*

rs <- rs[rs$Station\_Type!="Recon",] # remove recon stations

rs <- rs[rs$A.Video.Readable == "Yes",] # remove invalid videos

rs <- subset(rs, rs$Start\_Depth > 0) # remove NA in depth

rs <- subset(rs, rs$Start\_Depth < 100) # remove < 100 m deep

rs <- subset(rs, rs$LastOfTemp > 0) # remove blank water temps

rs <- subset(rs, rs$Turbidity != "Unknown") # remove unknown turbidity values

*# Variable names:*

Year = Year of the sample

Turbidity = Water turbidity

Current\_Direction: Direction of current relative to camera

Current\_Magnitude: Magnitude of the water current

Substrate\_Cat: Categorical variable for percent of bottom that is hardbottom

Relief: Maximum relief of substrate

Size: Mean size of hardbottom

Biotic\_Density\_Cat: Percent of substrate covered in attached biota

Biotic\_Type: Predominant type of attached biota

Biotic\_Height: Maximum height of attached biota

Start\_Depth: Bottom depth

Julian: Day of the year of the sample

Start\_Latitude: Latitude of the sample

LastOfTemp: Bottom temperature of sample

TOD: Time of day of the sample

Table 1. Annual SouthEast Reef Fish Survey sampling information for videos included in the analyses between North Carolina and Florida, 2011 – 2012.

|  |  |  |  |
| --- | --- | --- | --- |
|  | 2010 | 2011 | 2012 |
| Number of videos included | 225 | 611 | 1061 |
| Date range sampled | 28 Jul – 27 Oct | 19 May – 25 Oct | 24 Apr – 10 Oct |
| Latitude range sampled (°N) | 28.71 – 31.74 | 27.23 – 34.32 | 27.23 – 35.01 |
| Depth range sampled (meters) | 16 – 64 | 15 – 93 | 15 – 18 |
| Bottom water temperature range (°C) | 12.40 – 29.14 | 14.80 – 28.81 | 12.92 – 27.82 |

Table 2. Categorical variables considered for inclusion in the red snapper abundance index models, along with the proportion of sampling that occurred within each level for each year. [NMB note: we might want to do something with variables that have some levels that were rarely encountered, e.g., Substrate relief = Unknown] . Depth, Latitude, Julian Day, Time of Day, and Temperature were treated as continuous variables in the Delta Generalized Additive Model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **2010** | **2011** | **2012** | **Overall** |
| **Depth** |  |  |  |  |
| *d1* (15-28m) | 0.19 | 0.26 | 0.29 | 0.27 |
| *d2* (28-38m) | 0.31 | 0.22 | 0.24 | 0.24 |
| *d3* (38-52m) | 0.33 | 0.25 | 0.24 | 0.25 |
| *d4* (52-98m) | 0.17 | 0.27 | 0.23 | 0.24 |
| **Latitude** |  |  |  |  |
| *l1* (27.2-29.7ºN) | 0.28 | 0.32 | 0.20 | 0.25 |
| *l2* (29.7-31.3ºN) | 0.50 | 0.25 | 0.19 | 0.25 |
| *l3* (31.3-32.8ºN) | 0.22 | 0.31 | 0.22 | 0.25 |
| *l4* (32.8-35.0ºN) | 0.00 | 0.11 | 0.38 | 0.25 |
| **Water Temperature** |  |  |  |  |
| *t1* (12.4-19.6 ºC) | 0.19 | 0.32 | 0.22 | 0.25 |
| *t2* (19.6-22.2 ºC) | 0.16 | 0.29 | 0.25 | 0.25 |
| *t3* (22.2-24.8 ºC) | 0.09 | 0.19 | 0.31 | 0.25 |
| *t4* (24.8-29.1 ºC) | 0.56 | 0.20 | 0.22 | 0.25 |
| **Julian Day** |  |  |  |  |
| *t1* (115-171) | 0.00 | 0.24 | 0.31 | 0.25 |
| *t2* (171-209) | 0.01 | 0.35 | 0.25 | 0.25 |
| *t3* (209-235) | 0.24 | 0.18 | 0.28 | 0.25 |
| *t4* (235-300) | 0.74 | 0.23 | 0.16 | 0.25 |
| **Time of Day** |  |  |  |  |
| *tod1* (0.44-0.52) | 0.13 | 0.28 | 0.26 | 0.25 |
| *tod2* (0.52-0.65) | 0.27 | 0.25 | 0.25 | 0.25 |
| *tod3* (0.65-0.77) | 0.30 | 0.22 | 0.25 | 0.25 |
| *tod4* (0.77-0.94) | 0.30 | 0.25 | 0.24 | 0.25 |
| **Water clarity** |  |  |  |  |
| *Clear* (horizon visible) | 0.55 | 0.46 | 0.52 | 0.50 |
| *Moderate* (bottom visible) | 0.32 | 0.42 | 0.41 | 0.40 |
| *Turbid* (bottom not visible) | 0.13 | 0.12 | 0.08 | 0.10 |
| **Current direction** |  |  |  |  |
| *Away* (current heading away from camera) | 0.16 | 0.24 | 0.28 | 0.26 |
| *Sideways* (current heading sideways from camera) | 0.40 | 0.48 | 0.49 | 0.47 |
| *Towards* (current heading into camera) | 0.18 | 0.19 | 0.23 | 0.21 |
| *Unknown* (current direction cannot be estimated) | 0.26 | 0.09 | 0.00 |  |
| **Current magnitude** |  |  |  |  |
| *Weak* (fish swimming not affected) | 0.92 | 0.95 | 0.86 | 0.90 |
| *Strong* (fish swimming affected) | 0.04 | 0.03 | 0.14 | 0.09 |
| *Unknown* (current magnitude cannot be estimated) | 0.04 | 0.01 | 0.00 | 0.01 |
| **Attached biota density** |  |  |  |  |
| *None* (no attached biota present) | 0.21 | 0.10 | 0.09 | 0.10 |
| *Low* (0.5 – 9.9% of substrate covered) | 0.35 | 0.24 | 0.30 | 0.29 |
| *Moderate* (10.0 – 39.9% of substrate covered) | 0.14 | 0.16 | 0.26 | 0.21 |
| *High* (≥ 40% of substrate covered) | 0.26 | 0.44 | 0.34 | 0.36 |
| *Unknown* (density of attached biota cannot be estimated) | 0.04 | 0.07 | 0.02 | 0.04 |
| **Predominant attached biota type** |  |  |  |  |
| *NA* (no attached biota present) | 0.22 | 0.10 | 0.12 | 0.13 |
| *Coral* and sponge (most attached biota is coral and sponges | 0.48 | 0.52 | 0.50 | 0.50 |
| *Algae* (most attached biota is algae) | 0.14 | 0.31 | 0.22 | 0.24 |
| *Unknown* (cannot identify attached biota) | 0.16 | 0.08 | 0.16 | 0.13 |
| **Attached biota height** |  |  |  |  |
| *NA* (attached biota not present) | 0.20 | 0.10 | 0.09 | 0.10 |
| *Low* (maximum height is ≤ 0.5 m) | 0.59 | 0.55 | 0.39 | 0.46 |
| *High* (maximum height is > 0.5 m) | 0.21 | 0.29 | 0.51 | 0.40 |
| *Unknown* (cannot estimate height of attached biota) | 0.00 | 0.07 | 0.02 | 0.03 |
| **Amount of hardbottom substrate** |  |  |  |  |
| *None* (no hardbottom present) | 0.34 | 0.14 | 0.29 | 0.25 |
| *Low* (0 – 9.9% substrate is hardbottom) | 0.19 | 0.36 | 0.36 | 0.34 |
| *Moderate* (10.0 – 39.9% substrate is hardbottom) | 0.10 | 0.12 | 0.14 | 0.13 |
| *High* (≥ 40% substrate is hardbottom) | 0.38 | 0.32 | 0.20 | 0.26 |
| *Unknown* (cannot see bottom) | 0.00 | 0.05 | 0.02 | 0.03 |
| **Substrate relief** |  |  |  |  |
| *Low* (maximum relief of substrate is < 0.3 m) | 0.73 | 0.63 | 0.68 | 0.67 |
| *Moderate* (maximum relief of substrate is 0.3 – 1.0 m) | 0.16 | 0.24 | 0.22 | 0.22 |
| *High* (maximum relief of substrate is ≥ 1.0 m) | 0.10 | 0.08 | 0.08 | 0.08 |
| *Unknown* (cannot see bottom) | 0.00 | 0.05 | 0.02 | 0.03 |
| **Hardbottom substrate size** |  |  |  |  |
| *NA* (no hardbottom present) | 0.34 | 0.14 | 0.29 | 0.25 |
| *Coarse* (most hardbottom is ≤ 1.0 m in diameter) | 0.34 | 0.37 | 0.30 | 0.33 |
| *Continuous* (most hardbottom is > 1.0 m in diameter) | 0.32 | 0.43 | 0.40 | 0.40 |
| *Unknown* (cannot see bottom) | 0.00 | 0.05 | 0.02 | 0.03 |

Table 3. Model selection for red snapper binomial (presence-absence) and positive MeanCount submodels for video data collected by the SouthEast Reef Fish Survey, 2010 – 2012. Degrees of freedom are shown for categorical terms, and estimated degrees of freedom are shown for smoothed terms. Asterisks denote significance at the following alpha levels: \* = 0.10, \*\* = 0.05, \*\*\* 0.01; na: covariate was excluded from that particular model; GCV: generalized cross validation score; AIC: Akaike information criterion; Dev: percent deviance explained by the model. Base is: where *x* is red snapper MeanCount, *f*s are factors, *g*s are smoothed variables, *y* is the year, *wc* is water clarity, *cd* is current direction, *cm* is current magnitude, *sc* is percent of substrate that is hard bottom, *sr* is relief of the substrate, *ss* is size of hard bottom substrate, *bd* is attached biota density, *bt* is the predominant attached biota, *bh* is the maximum height of the attached biota, *d* is bottom depth, *t* is day of the year, *lat* is latitude, *temp* is water bottom temperature, *tod* is time of the day, and is the random error. Only the four best models are shown for each submodel.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | GCV | AIC | Dev | *y* | *wc* | *cd* | *cm* | *sc* | *sr* | *ss* | *bd* | *bt* | *bh* | *d* | *t* | *lat* | *temp* | *tod* |
| **Binomial submodel** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| base – *ss – sr – bh – wc – bt – tod* | -0.16019 | 1593.1 | 28.3 | 2 | na | 3\*\*\* | 2 | 4\*\*\* | na | na | 4 | na | na | 4.3\*\*\* | 6.4\* | 8.3\*\*\* | 2.2\*\*\* | na |
| base – *ss – sr – bh – wc – bt* | -0.15979 | 1593.9 | 28.3 | 2 | na | 3\*\*\* | 2 | 4\*\*\* | na | na | 4 | na | na | 4.3\*\*\* | 6.4\* | 8.3\*\*\* | 2.2\*\*\* | 1.0 |
| base – *ss – sr – bh – wc – tod* | -0.15921 | 1595.0 | 28.5 | 2 | na | 3\*\*\* | 2 | 4\*\* | na | na | 4 | 3 | na | 4.4\*\*\* | 6.5\* | 8.3\*\*\* | 2.3\*\*\* | na |
| base – *ss – sr – bh – bt – tod* | -0.15908 | 1595.2 | 28.4 | 2 | 2 | 3\*\*\* | 2\* | 4\*\*\* | na | na | 4 | na | na | 4.2\*\*\* | 6.4\*\* | 8.3\*\*\* | 2.2\*\*\* | na |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Positive submodel** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| base – *cm – tod – sr – temp – ss – sc – wc* | 0.03991 | -179.2 | 27.5 | 2\*\*\* | na | 3\*\*\* | na | na | na | na | 4\*\*\* | 3\*\*\* | 3\*\* | 3.6\*\* | 5.3 | 6.5\*\*\* | na | na |
| base – *cm – tod – sr – temp – ss – sc* | 0.03999 | -178.5 | 27.9 | 2\*\*\* | 2 | 3\*\*\* | na | na | na | na | 4\*\*\* | 3\*\*\* | 3\*\* | 3.6\*\* | 5.0 | 6.2\*\*\* | na | na |
| base – *cm – tod – sr – ss – sc – wc* | 0.04001 | -177.4 | 27.5 | 2\*\*\* | na | 3\*\*\* | na | na | na | na | 4\*\*\* | 3\*\*\* | 3\*\* | 3.6\*\* | 5.1 | 6.4\*\*\* | 1.0 | na |
| base – *cm – tod – temp – ss – sc – wc* | 0.04014 | -176.8 | 27.8 | 2\*\*\* | na | 3\*\*\* | na | na | 3 | na | 4\*\*\* | 3\*\*\* | 3\*\* | 3.7\*\* | 5.4 | 6.4\*\*\* | na | na |

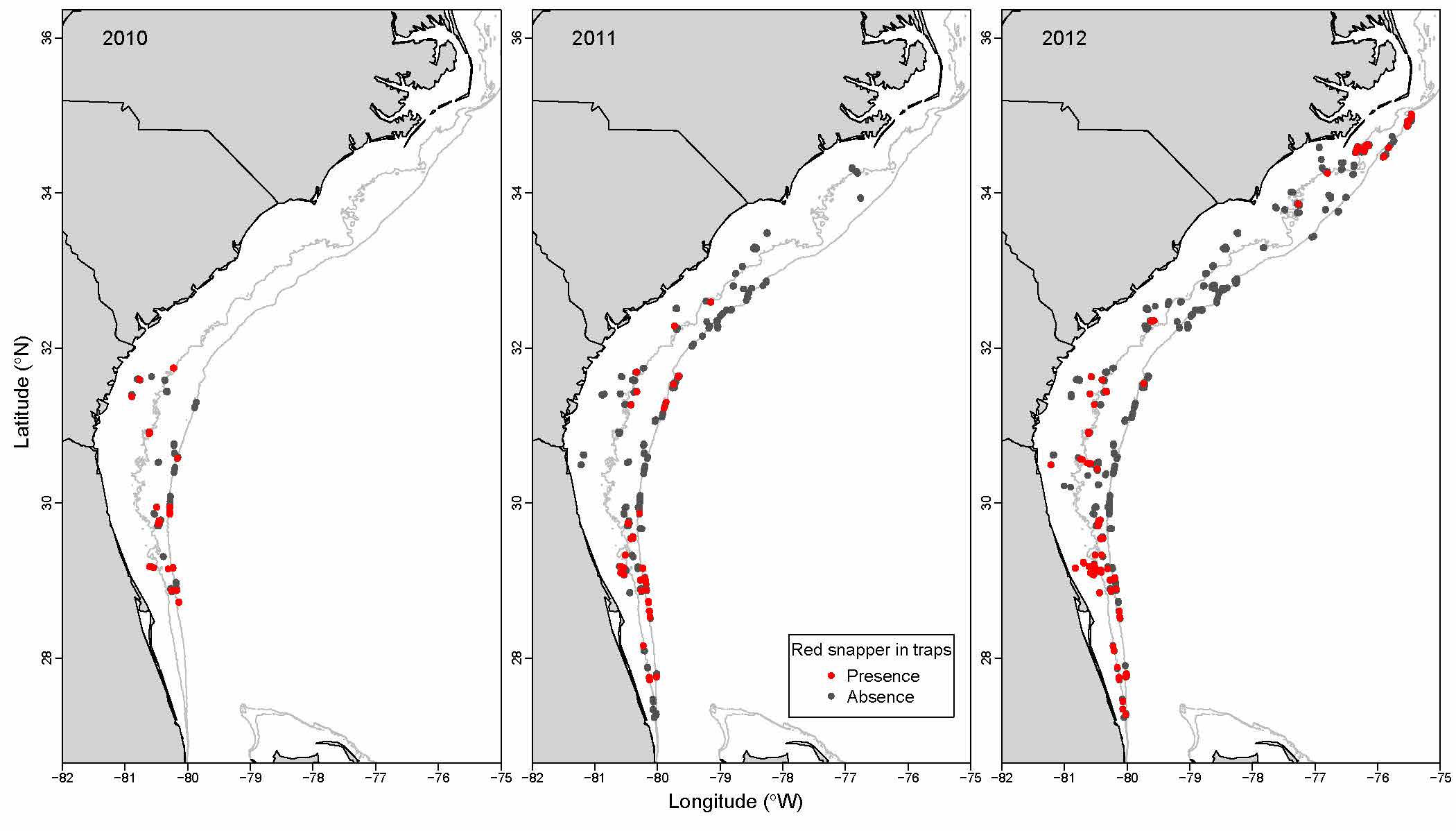


Figure 1. Map of red snapper presence-absence on videos collected by the SouthEast Reef Fish Survey, 2010 – 2012.

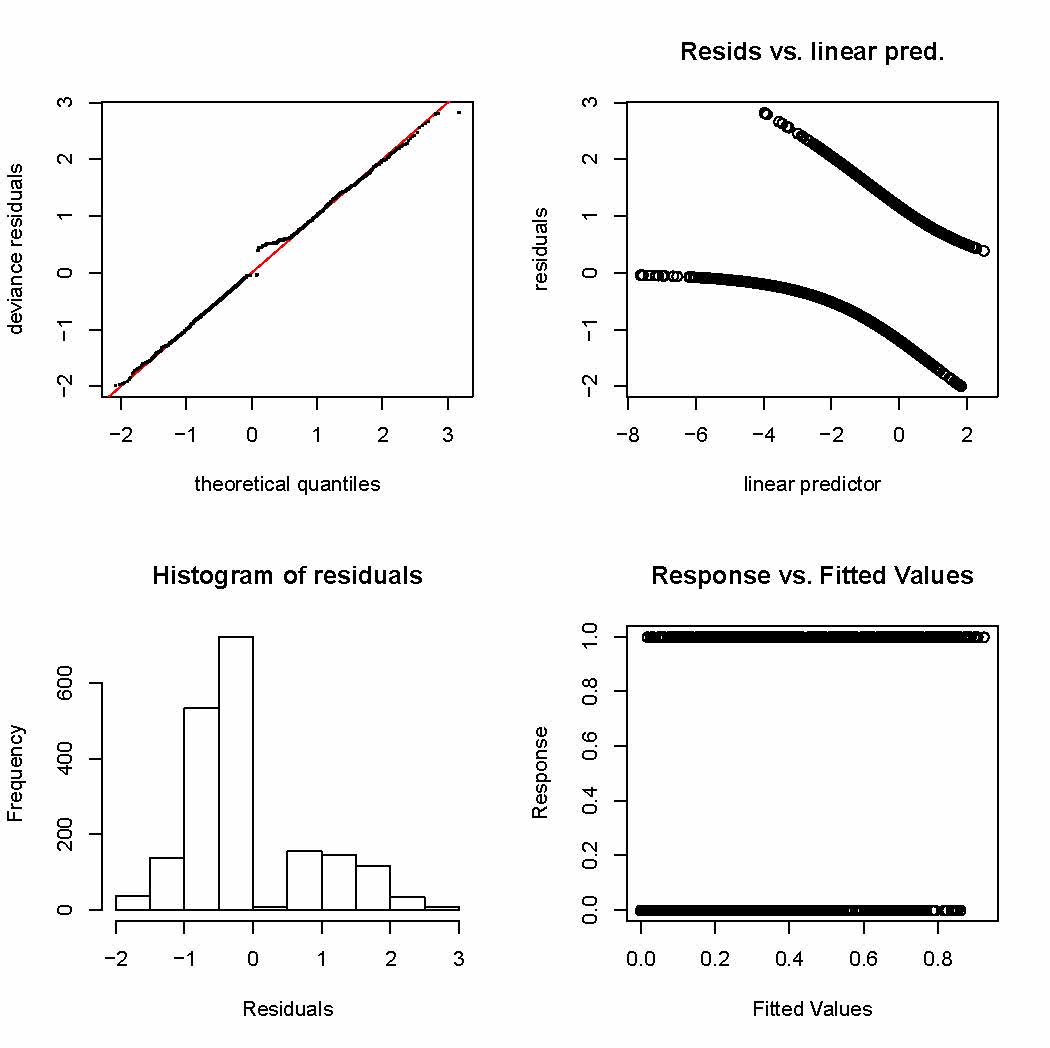


Figure 2. Model diagnostics for the binomial (presence-absence) generalized additive model for red snapper based on videos collected by the SouthEast Reef Fish Survey, 2010 – 2012. The plots are (top left) standard quantile-quantile probability plot, (top right) residuals versus linear predictor plot, (bottom left) histogram of residuals, and (bottom right) the response versus fitted values.

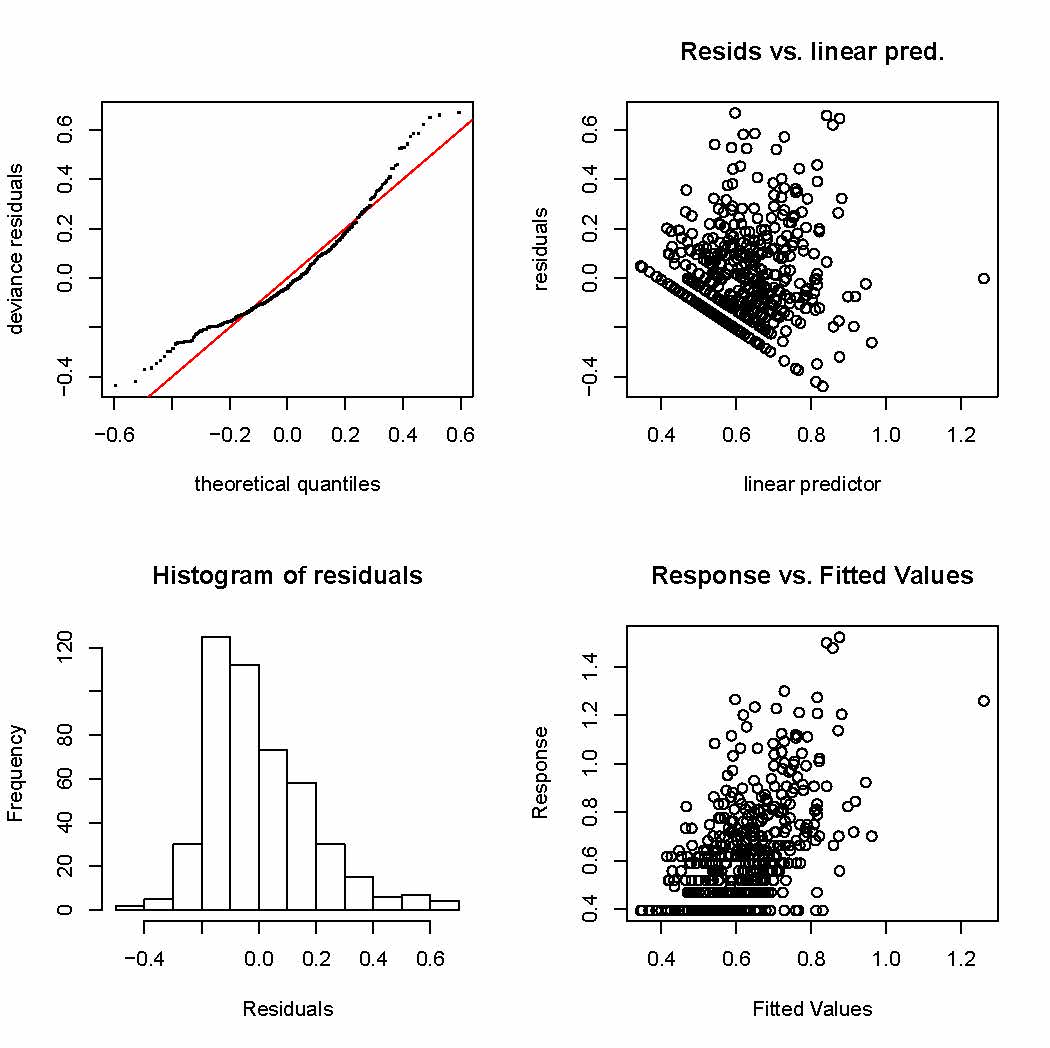


Figure 3. Model diagnostics for the positive MeanCount generalized additive model for red snapper based on videos collected by the SouthEast Reef Fish Survey, 2010 – 2012. The plots are (top left) standard quantile-quantile probability plot, (top right) residuals versus linear predictor plot, (bottom left) histogram of residuals, and (bottom right) the response versus fitted values. [NMB note: not a very good fit, but better than lognormal fit.]

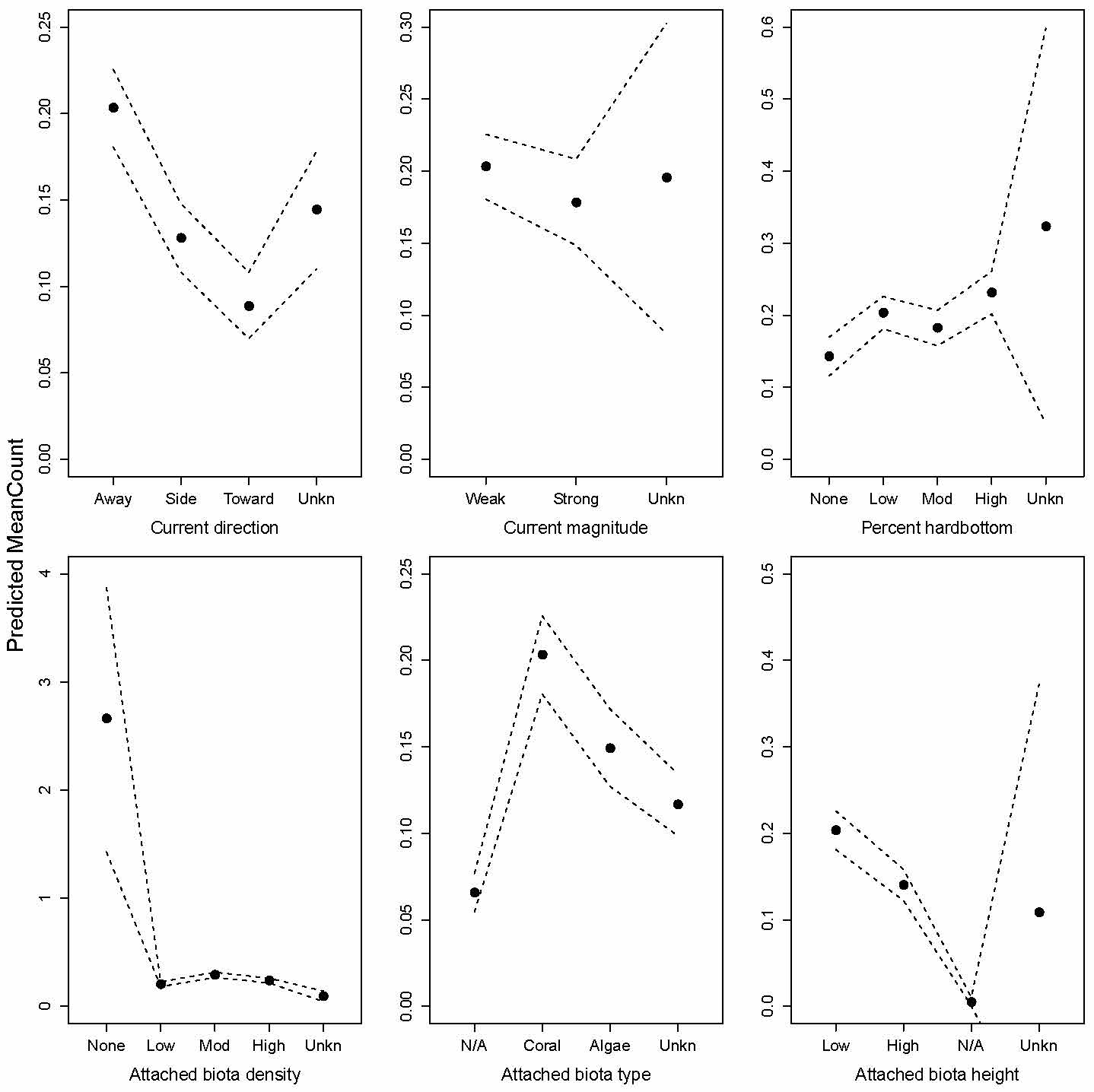


Figure 4. Predicted red snapper video MeanCount for the six categorical variables included in the delta-generalized additive model based on data from the SouthEast Reef Fish Survey, 2010 – 2012. Filled circles are mean predictions and dashed lines are 95% confidence intervals.

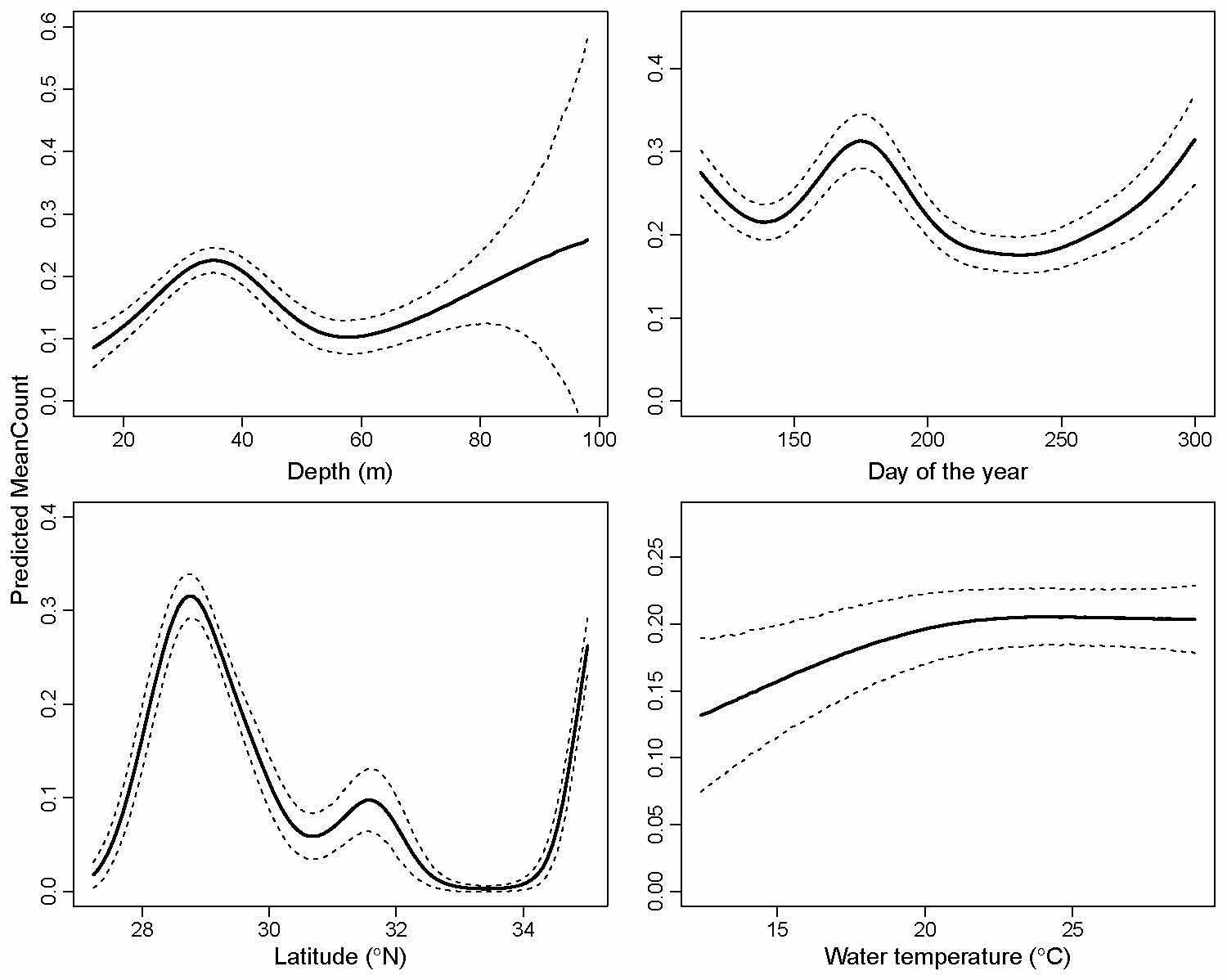


Figure 5. Predicted red snapper video MeanCount for the four continuous variables included in the delta-generalized additive model based on data from the SouthEast Reef Fish Survey, 2010 – 2012. Solid black lines are mean predictions and dashed lines are 95% confidence intervals.

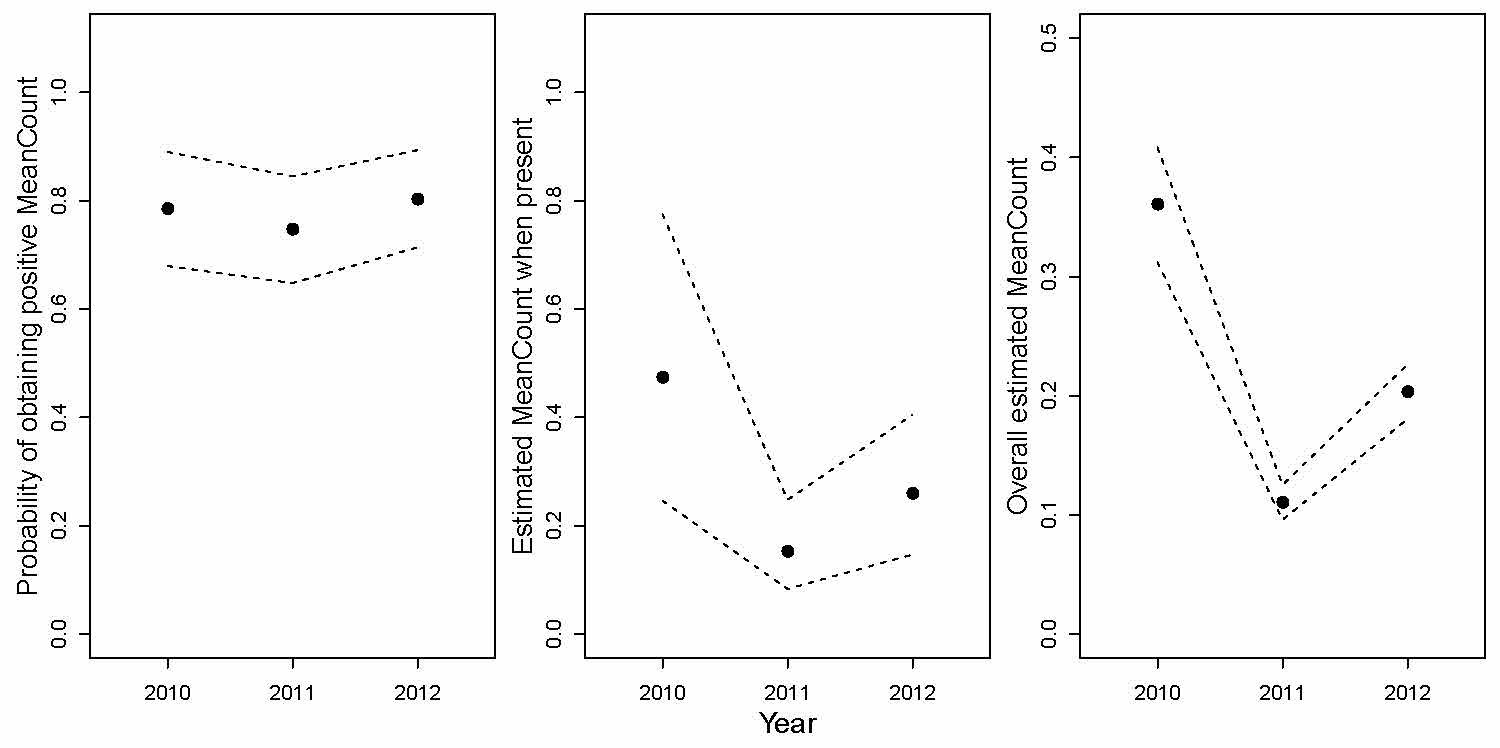


Figure 6. Predicted red snapper video MeanCount for the three years (2010 – 2012) included in the delta-generalized additive model analysis constructed from data from the SouthEast Reef Fish Survey. Filled circles are mean predictions and dashed lines are 95% confidence intervals. A binomial model was used to estimate the probability of obtaining a positive MeanCount (presence-absence) and a Gaussian model was used to estimate catch when present. Overall estimated MeanCount (i.e., red snapper index of abundance) was determined through the combination of binomial and Gaussian model outputs.

Summary of Delta GLM

* 2010-2012 SEFIS and MARMAP video data
* Reconnaissance stations included if they were included in sampling universe the following year (MARMAP “recon” stations excluded because I couldn’t determine which recon stations were included, and the fact that the number of recon stations was low)
* All latitudes and soak times
* Depth < 100 m
* Removed missing depth, sampling durations, and bottom temps
* Binomial model for presence-absence
* Both lognormal and gamma error distributions were considered to model the MeanCount Data. The lognormal distribution fit the data best based on AIC.
* Lognormal model for positive MeanCount data
* Model selection was accomplished considering AIC in a backward stepwise procedure.

# *Full data subsetting:*

rs <- rs[rs$Station\_Type!="Recon",] # remove recon stations

rs <- rs[rs$A.Video.Readable == "Yes",] # remove invalid videos

rs <- subset(rs, rs$Start\_Depth > 0) # remove NA in depth

rs <- subset(rs, rs$Start\_Depth < 100) # remove < 100 m deep

rs <- subset(rs, rs$LastOfTemp > 0) # remove blank water temps

rs <- subset(rs, rs$Turbidity != "Unknown") # remove unknown turbidity values

rs <- subset(rs, rs$No.Readable.Frames >19) # remove all samples with no readable frames

# *Categorization of continuous variables:*

The continuous variables (Start\_Depth, Start\_Latitude, LastOfTemp, and Julian) were converted to categorical variables and classified based on the quartiles of the data distributions. The data ranges of each of these variables are listed in Table 1 and the categorical definitions and proportions of observations within each category are reported in Table 2.

*# Variable names:*

Year = Year of the sample

Turbidity = Water turbidity

Current\_Direction: Direction of current relative to camera

Current\_Magnitude: Magnitude of the water current

Substrate\_Cat: Categorical variable for percent of bottom that is hardbottom

Relief: Maximum relief of substrate

Size: Mean size of hardbottom

Biotic\_Density\_Cat: Percent of substrate covered in attached biota

Biotic\_Type: Predominant type of attached biota

Biotic\_Height: Maximum height of attached biota

Start\_Depth: Bottom depth

Julian: Day of the year of the sample

Start\_Latitude: Latitude of the sample

LastOfTemp: Bottom temperature of sample

TOD: Time of day of the sample

Table 4. Delta GLM model selection for red snapper binomial (presence-absence) and positive MeanCount submodels for video data collected by the SouthEast Reef Fish Survey, 2010 – 2012. Degrees of freedom are shown each of the factors. Asterisks denote significance at the following alpha levels: \* = 0.10, \*\* = 0.05, \*\*\* 0.01; na: covariate was excluded from that particular model; AIC: Akaike information criterion; Dev: deviance explained by the model. Base is: where *x* is red snapper MeanCount, *f*s are factors, *y* is the year, *wc* is water clarity, *cd* is current direction, *cm* is current magnitude, *sc* is percent of substrate that is hard bottom, *sr* is relief of the substrate, *ss* is size of hard bottom substrate, *bd* is attached biota density, *bt* is the predominant attached biota, *bh* is the maximum height of the attached biota, *d* is bottom depth, *t* is Julian day of the year, *lat* is latitude, *temp* is water bottom temperature, *tod* is time of the day, and is the random error. The positive submodel uses the logarithm of the MeanCount as the dependent variable ( log(). Only the four best models are shown for each submodel.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | AIC | *y* | *wc* | *cd* | *cm* | *sc* | *sr* | *ss* | *bd* | *bt* | *bh* | *d* | *t* | *lat* | *temp* | *tod* |
| **Binomial submodel** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| base –*wc* –*cm* –*sc* –*sr* –*temp* –*tod* | 1804.4 | 2\* | na | 3\*\*\* | na | na | na | 3\*\*\* | 4\* | 3\*\*\* | 3\*\* | 3\*\*\* | 3\* | 3\*\*\* | na | na |
| base –*wc* –*cm* –*sc* –*sr* –*tod* | 1805.3 | 2\* | na | 3\*\*\* | na | na | na | 3\*\*\* | 4\* | 3\*\*\* | 3\*\* | 3\*\*\* | 3\* | 3\*\*\* | 3\* | na |
| base –*wc* –*cm* –*sc* –*sr* | 1807.5 | 2\* | na | 3\*\*\* | na | na | na | 3\*\*\* | 4\* | 3\*\*\* | 3\*\* | 3\*\*\* | 3\* | 3\*\*\* | 3\* | 3 |
| base –*cm* –*sc* –*sr* | 1810.2 | 2\* | 2 | 3\*\*\* | na | na | na | 3\*\*\* | 4\* | 3\*\*\* | 3\*\* | 3\*\*\* | 3\* | 3\*\*\* | 3\* | 3 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Positive submodel** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| base – *cm – sc – sr– ss –d –temp* –*tod* | 1497.0 | 2\*\*\* | 2 | 3\*\*\* | na | na | na | na | 4\* | 3\* | 3\*\* | na | 3 | 3\* | na | na |
| base – *cm – sr – ss –d –temp* –*tod* | 1498.9 | 2\*\*\* | 2 | 3\*\*\* | na | 4 | na | na | 4\* | 3\* | 3\*\* | na | 3 | 3\* | na | na |
| base – *cm – sr –d –temp* –*tod* | 1500.8 | 2\*\*\* | 2 | 3\*\*\* | na | 4 | na | 3 | 4\* | 3\* | 3\*\* | na | 3 | 3\* | na | na |
| base – *cm – sr –d –temp* | 1504.2 | 2\*\*\* | 2 | 3\*\*\* | na | 4 | na | 3 | 4\* | 3\* | 3\*\* | na | 3 | 3\* | na | 3 |

Figure XX. Diagnostics of the delta GLM positive submodel showing model Top panel shows the histogram of empirical log CPUE, with the normal distribution (empirical mean and variance) overlaid. Bottom panel shows the quantile-quantile plot of residuals from the fitted model.

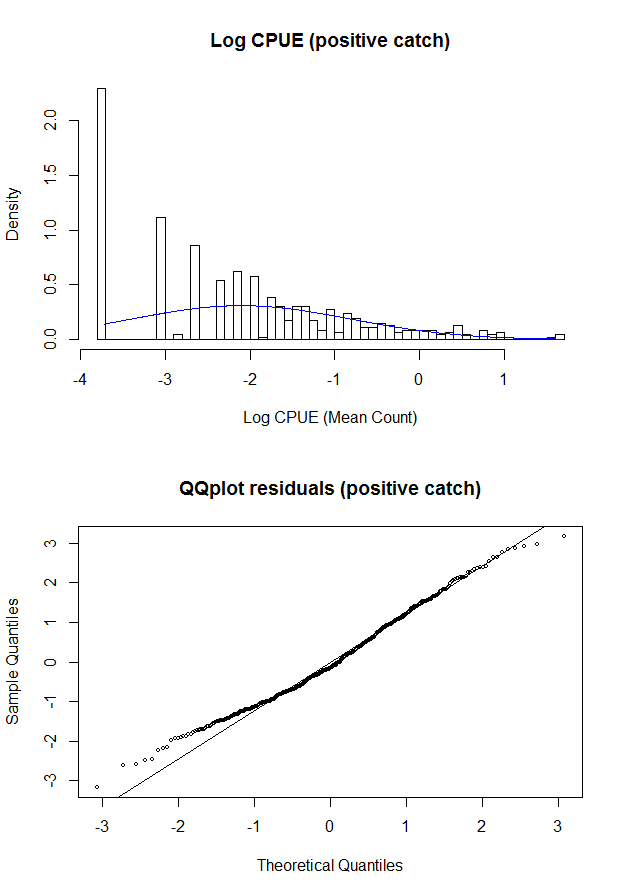


Figure XX. Residual plots for the delta GLM positive submodel. Box-and-whisker plots give first, second (median), and third quartile, as well as limbs that extend approximately one interquartile range beyond the nearest quartile, and outliers (circles beyond the limbs).

Figure XX. Residual plots for the delta GLM positive submodel. Box-and-whisker plots give first, second (median), and third quartile, as well as limbs that extend approximately one interquartile range beyond the nearest quartile, and outliers (circles beyond the limbs).

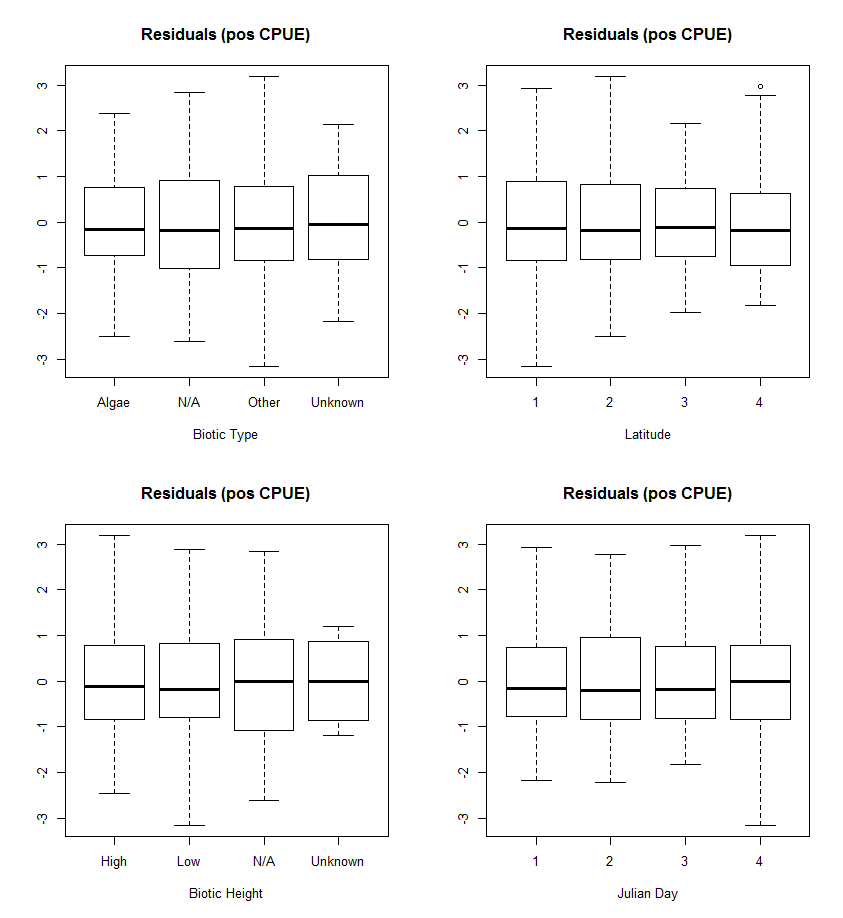


Figure XX. Residual plots for the delta GLM binary submodel. Box-and-whisker plots give first, second (median), and third quartile, as well as limbs that extend approximately one interquartile range beyond the nearest quartile, and outliers (circles beyond the limbs).

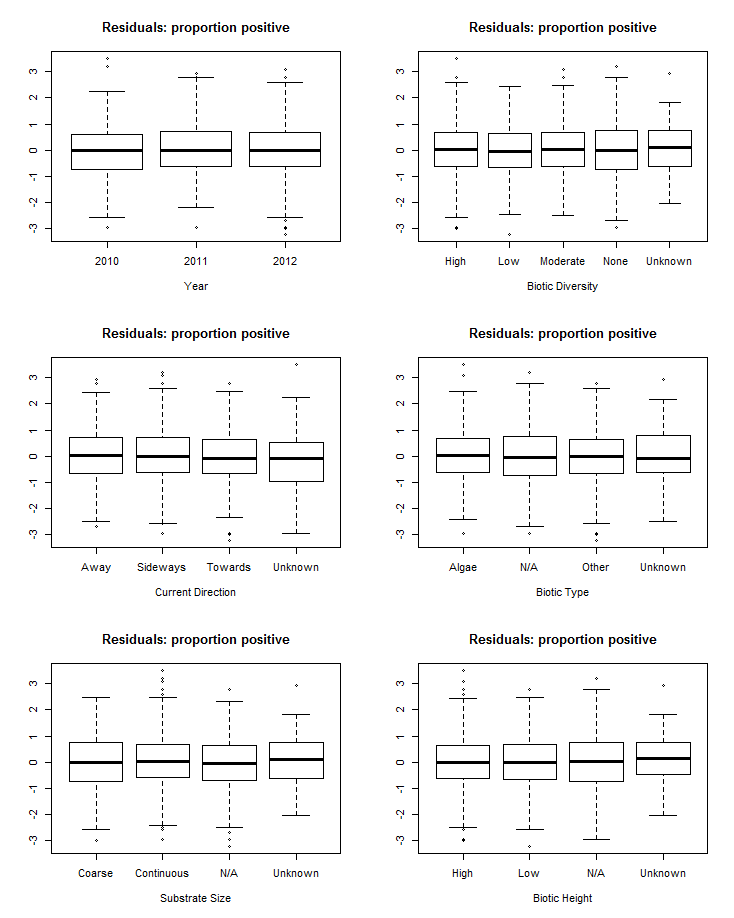


Figure XX. Residual plots for the delta GLM binary submodel. Box-and-whisker plots give first, second (median), and third quartile, as well as limbs that extend approximately one interquartile range beyond the nearest quartile, and outliers (circles beyond the limbs).

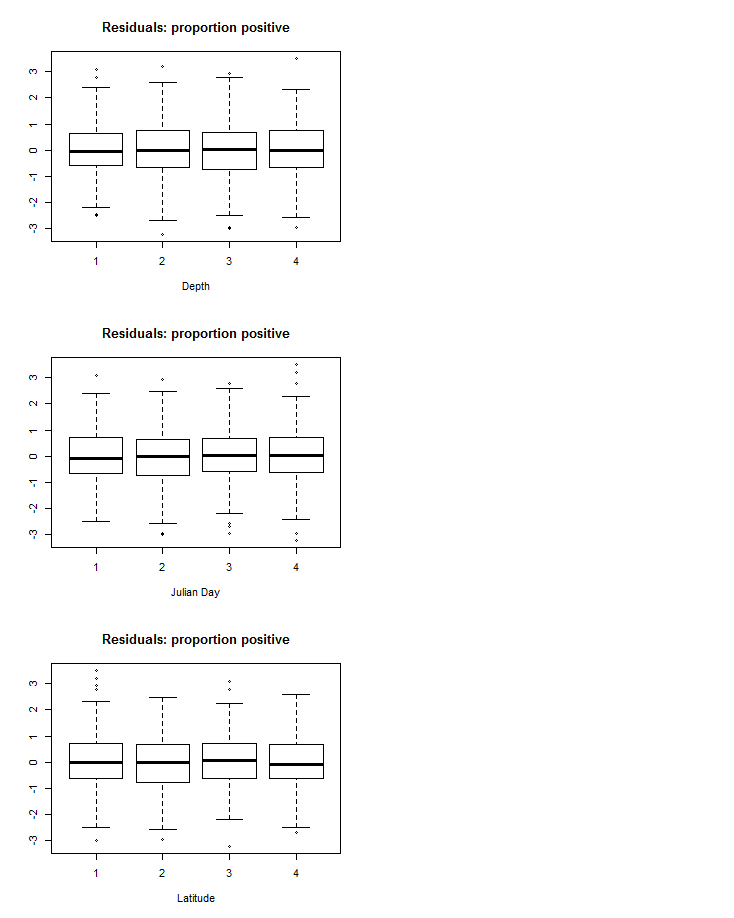
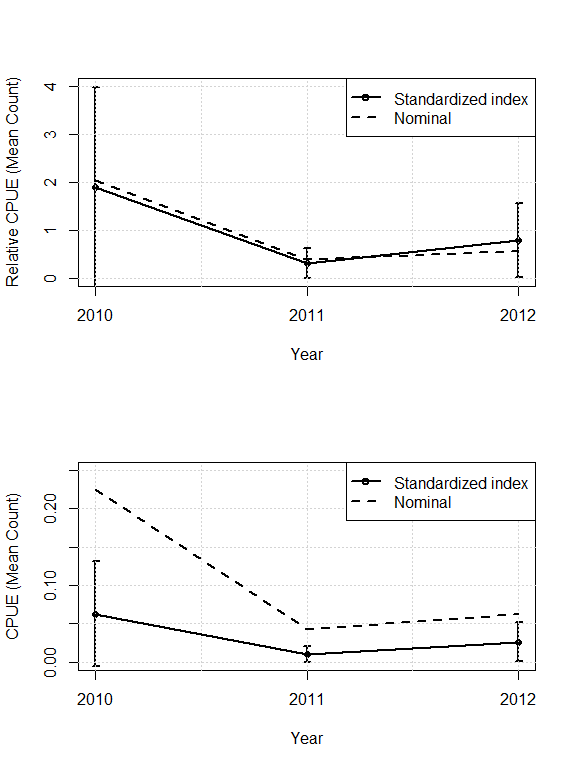


Figure XX. Red snapper standardized and nominal CPUE from the video data on either relative (top panel) or absolute (bottom panel) scale. 

Summary of Zero Inflated Models

* 2010-2012 SEFIS and MARMAP video data
* Reconnaissance stations included if they were included in sampling universe the following year (MARMAP “recon” stations excluded because I couldn’t determine which recon stations were included, and the fact that the number of recon stations was low)
* All latitudes and soak times
* Depth < 100 m
* Removed missing depth, sampling durations, and bottom temps
* Response variable is the total number of Red Snapper observed on all video frames examined (TotalCount).
* Removed all samples that did not have 41 readable video frames.
* Both the Zero Inflated Poisson (ZIP) and the Zero Inflated Negative Binomial (ZINB) were considered. The ZINB model fit the data best based on AIC.
* Model selection was accomplished considering AIC in a backward stepwise procedure following Zuur et al. 2007 Analyzing Ecological Data).

# *Full data subsetting:*

rs <- rs[rs$Station\_Type!="Recon",] # remove recon stations

rs <- rs[rs$A.Video.Readable == "Yes",] # remove invalid videos

rs <- subset(rs, rs$Start\_Depth > 0) # remove NA in depth

rs <- subset(rs, rs$Start\_Depth < 100) # remove < 100 m deep

rs <- subset(rs, rs$LastOfTemp > 0) # remove blank water temps

rs <- subset(rs, rs$Turbidity != "Unknown") # remove unknown turbidity values

rs <- subset(rs, rs$No.Readable.Frames ==41) # retain samples with 41 readable video frames

# *Categorization of continuous variables:*

The continuous variables (Start\_Depth, Start\_Latitude, LastOfTemp, and Julian) were converted to categorical variables and classified based on the quartiles of the data distributions. The data ranges of each of these variables are listed in Table 1. and the categorical definitions and proportions of observations within each category are reported in Table 2.

*# Variable names:*

Year = Year of the sample

Turbidity = Water turbidity

Current\_Direction: Direction of current relative to camera

Substrate\_Cat: Categorical variable for percent of bottom that is hardbottom

Biotic\_Density\_Cat: Percent of substrate covered in attached biota

Start\_Depth: Bottom depth

Julian: Day of the year of the sample

Start\_Latitude: Latitude of the sample

LastOfTemp: Bottom temperature of sample