Indices of abundance for Gray Snapper (*Lutjanus griseus*) using combined data from three independent video surveys

Kevin A. Thompson, Theodore S. Switzer, Mary C. Christman, Sean F. Keenan, Christopher Gardner, Matt Campbell, Adam Pollack

SEDAR51-DW-15

15 June 2017

Please cite this document as:

Thompson, Kevin A., Theodore S. Switzer, Mary C. Christman, Sean F. Keenan, Christopher Gardner, Matt Campbell, Adam Pollack. 2017. Indices of abundance for Gray Snapper (*Lutjanus griseus*) using combined data from three independent video surveys. SEDAR51-DW-15. SEDAR, North Charleston, SC. 20 pp.

Indices of abundance for Gray Snapper (*Lutjanus griseus*) using combined data from three independent video surveys

Kevin A. Thompson[1], Theodore S. Switzer[1], Mary C. Christman[2], Sean F. Keenan[1], Christopher Gardner[3], Matt Campbell[4], Adam Pollack[5]

[1]Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute, St. Petersburg, FL

[2] MCC Statistical Consulting LLC, Gainesville, FL

[3] Southeast Fisheries Science Center, Panama City Laboratory, Panama City, FL

[4]Southeast Fisheries Science Center, Mississippi Laboratories, Pascagoula, MS

[5]Riverside Technology, Inc., Southeast Fisheries Science Center, Mississippi Laboratories, Pascagoula, MS

Introduction

Currently there are three different stationary video surveys for reef fish in the Gulf of Mexico (GOM). The NMFS SEAMAP video survey, carried out by NMFS Pascagoula lab, has the longest running time series (1992-1997, 2004+), followed by the NMFS Panama City lab survey (2005+), with the most recent survey being the Florida Fish and Wildlife Research Institute SEAMAP survey (FWRI, starting year 2008). While the surveys share many commonalities regarding the use of stationary cameras to assess fish abundancies on reef or structured habitat, there are variations in survey design and habitat characteristics collected in addition to the time period and area sampled. In previous SEDAR data workshops and meetings, it was recommended that a combined index be developed for the entire GOM for assessment purposes. Combining indices across datasets may increase predictive capabilities by allowing for the largest possible sample sizes in model fitting. However, as previous research has indicated, combining data across changing spatial areas and surveys, combined with the use of a year only model, can yield to spurious conclusions regarding stock abundance (Campbell 2004; Ye et al. 2004). As such, we developed a novel, habitat-based approach to combining relative abundance data for generating annual trends for Gray Snapper throughout the eastern GOM.

Survey Comparisons

Survey design

The Pascagoula lab survey primarily targets high-relief topographic features along the continental shelf from south Texas to south Florida. Sites are selected using a stratified, random design with strata determined by region and total proportion of reef area in a sampling block (10 minute latitude X 10 minute longitude blocks). Sites are described by multi-beam sonar. This survey uses the Mississippi river delta as a geographic feature separating the west and east regions of the GOM (Campbell et al. 2017).

The Panama City video survey targets the inner shelf of the northeast GOM. Survey design has changed through time, but since 2010 a two-stage unequal probability design has been used. Blocks are

5 minutes x 5 minutes in size with sites randomly, proportionally allocated by region, sub-region and depth. This survey is broken up into eastern and western regions by Cape San Blas in the Florida Panhandle. Sites are described using side-scanning before video deployment (Gardner et al. 2017).

The FWRI survey initially focused on the regions offshore of Tampa Bay and Charlotte Harbor, FL (NMFS statistical zones 4 and 5) with habitats either inshore (10-36 m depth) or offshore (37-110 m depth). The survey has since expanded to include statistical zones 9 and 10 off the Florida Panhandle in 2014 with additional sites added in 2016 to cover the entirety of the West Florida Shelf from statistical zones 2-10, although only data from statistical zones 4 and 5 are included in these analyses. Sites are initially mapped using side scan sonar over a 0.1 nm x 0.3 nm area. Video deployment sites are then randomly assigned proportionally across region and depth zones (Thompson et al. 2017).

Video reads

All three surveys use paired stereo-imaging cameras at each site. All videos are read to identify the maximum number of individuals of each species viewed in a single frame within a 20 minute time frame (MaxN). Habitat characteristics on video are also noted with the percentage or presence/absence of abiotic and biotic habitat types that may contribute to fish biomass (e.g. sponge, algae, and corals), although some categories are not shared among all labs (Campbell et al. 2017; Gardner et al. 2017; Thompson et al. 2017). All three labs also have length measurements for a subset of the reef fish of interest using either VMS or SeaGIS software depending on year of the survey. The length distributions for each lab were compared, and for Gray Snapper, were generally similar, with minor variation in median by lab, indicating that combining data for an index was appropriate (Fig. 1)

Data reduction

For all surveys, video reads were excluded if they were unreadable due to turbidity or deployment errors. For the Pascagoula Lab, data included in this index are from 1993 and on, due to different counting methods in 1992. Furthermore, Pascagoula data was only included from the region east of the Mississippi delta due to very low catches of Gray Snapper in the western GOM. The entire spatial extent of the Panama City data was used from 2006 on with 2005 excluded because of an incomplete survey. The FWRI data was limited to 2010 and on due to the previous year's not including side-scan geoform as a variable which was determined to be potentially important. FWRI data were spatially limited to zones 4 and 5 due to the other areas of the WFS not having sufficient years of sampling. Final sample sizes by lab and year can be found in Table 1 and spatial coverage is shown in Figure 2.

Index Construction

Habitat models

To combine the data from all three surveys into one model predicting Gray Snapper relative CPUE throughout the time series, we created a habitat variable that included each lab's individual variables that could be applied to all the data. This was done so final index models can account for changing effort and habitat allocation through time rather than limiting the model to be predicted only by year and lab. We first determined the percentage of sites that occurred on optimal, suboptimal, or poor (O, S, P) habitats for each survey independently. For this we used a categorical regression tree approach (CART) because it can account for correlations among variables and can include both continuous and categorical data. It has been previously demonstrated to be a useful tool in fisheries ecology and specifically in describing fish-habitat associations (De'Ath and Fabricus 2000; Yates et al. 2016).

For these initial analyses, MaxN for each site was reduced to a presence and absence variable and was used as the response variable for habitat designations. We first used a random forest approach to reduce the number of potential variables to be selected from in the final model for each lab's dataset to reduce redundant or correlated variables used in the final indexing model. For the random forest, each lab was modeled separately with the entirety of that lab's dataset. The random forest runs fit 2000 CARTS to the data and then determined each variables importance, a scale less number used to indicate the number of final models each variable occurred in and its significance therein. An example of output is given in Fig. 3 for the FWRI dataset.

We retained approximately 50% of the potential variables for each lab given by the random forest importance values for a final CART model. The final model was created by fitting the presence of Gray Snapper at site to the independent variables for a training dataset of 80% of the data. The remaining 20% of the data were retained in a test dataset to determine misclassification rates for each of the three models. The proportion of sites with positive Gray Snapper catches at each terminal node were then evaluated to determine the habitat characteristics defining optimal, suboptimal or poor habitat. Terminal nodes with double the overall proportion of positive catches for a dataset were assigned an optimal habitat code. Poor sites were determined by proportion positives that were approaching zero. The remaining sites were deemed suboptimal, and included the range of the overall proportion positive. All analyses were carried out using R version 3.0.2 (R Core Team 2014) and the Party package for CART (Hothorn et al. 2006).

CART results varied by lab with respect to the final variables chosen, with no variables showing up in all three models. The Pascagoula model showed 8 total final nodes, defined by presence/absence of sponge and algae, depth, latitude, relief, and longitude (Fig.4). The model shows three terminal nodes that define poor habitats for Gray Snapper, and two each for suboptimal and optimal habitats compared to the overall proportion positive of 0.12 (Fig. 4). The Panama City model was the simplest with only two variables chosen, depth and region (east and west of Cape San Blas; Fig. 5). Habitats were only found to be either suboptimal (4 terminal nodes) or optimal (one terminal node) with no sites found to have significantly lower proportions positive when compared to the overall dataset (Fig. 5). The FWRI CART model had presence/absence of sponge, rock, and algae as well as side-scanned geoform and latitude chosen as explanatory variables (Fig. 6). Most sites were found to be in suboptimal habitats with 3 nodes with the reaming sites either in poor habitats (2 nodes) or optimal (1 node, Fig. 6).

The site characteristics that define each node and habitat code were then used to create a habitat variable (hab: O, S, P) that was then back-applied to each site for each lab's dataset. The datasets were then combined for the index model. The final proportion of sites in the three hab categories for each lab and year are shown in Table 2.

Index model fitting and diagnostics

Like the individual survey indices, the combined dataset remained zero-inflated and therefore didn't conform to assumptions of normality (Fig. 7). Due to the count nature of the data, and the possibility of inflation of the zero counts we used four different error distribution models to construct preliminary evaluation models (i.e., Poisson, Negative Binomial, Zero-inflated Poisson, and Zero-inflated Negative Binomial). The zero inflated approaches model the zero counts using two different processes, a binomial and a count process (Zuur et al. 2009).

Initially, four full (all potential variables) general linear models (GLM) were considered utilizing both a Poisson (P) and Negative binomial (NB) error distribution and both Zero-inflated Poisson (ZIP) and zero inflated Negative Binomial (ZINB) formulations.

(1) *MaxN = Y+ Hab + Lab*

Where Hab is the CART derived habitat code and Lab represents the survey that collected the data for each site.

We compared the variance structure of each model formulation using likelihood ratio tests (Zuur et al. 2009) and Aikaike's information theoretic criterion (AIC; Zuur et al. 2009) to determine the most appropriate model formulation for the development of a video index for Gray Snapper in the Eastern GOM. Results of the likelihood ratio test indicate that the negative binomial models fit the data better than the Poisson, with the zero-inflated NB model the most appropriate (Table 3). The fitted values of the full zero-inflated negative binomial model (ZINB) also matched the MaxN values more closely than the non-inflated model, which performed poorly when predicting higher values of MaxN (Fig. 8).

Model diagnostics showed no discernible patterns of association between Pearson residuals and fitted values or the fitted values and the original data (Fig. 9). An examination of residuals for the model parameters (Fig. 10) showed no clear patterns of association, indicating correspondence to underlying model assumptions (Zuur et al. 2009). Confidence intervals were then determined by bootstrapping the model fitting over 5000 iterations. CPUE trends were adjusted to be relativized to 1 for each of the three time periods defined by differing number of surveys included (i.e. Pasc only, Pasc + PC, and Pasc + PC + FWRI, Table 1). Relativizing CPUE is standard practice in indices for SEDAR and prevents the addition of a dataset in the time series from artificially increasing estimated biomass. Modeling was conducted using the zeroinfl function of the pscl package in R (Jackman 2008).

Results and Discussion:

Annual standardized index values for Gray Snapper in the Eastern Gulf of Mexico, including coefficients of variation, are presented in Table 4. The model CV's indicate a good fit, with values higher in earlier years and steadily decreasing CV's as the surveys are added and continue. Biomass trends for Gray Snapper in the eastern GOM are relatively stable through the years (Fig. 11). Peaks in abundance occurred in 2008-201 and later in 2014. Interestingly, these peaks occur a year or two after peaks in nearshore, age-1 abundance, indicating these independent datasets in the GOM for this species are tracking the same population and trends (Flaherty-Walia et al. 2017).

We feel the method presented here with habitat models done for each survey and then combining the data for a unified index provides a reliable way to reduce the numbers of indices being submitted for assessment, while including each survey's individual MaxN observations and habitat data. Model trends were similar to the individual indices presented as part of this SEDAR, with CVs of similar or lower value. We recommend continuing research into these methods by applying it to future assessments. We also plan to explore methods to weight MaxN values by habitat code for each survey in the modelling process, rather than just use the habitat codes as an explanatory variable.

References Cited:

Campbell, M.D., Kevin R. Rademacher, Michael Hendon, Paul Felts, Brandi Noble, Ryan Caillouet, Joseph Salisbury, and John Moser. 2017. SEAMAP Reef Fish Video Survey: Relative Indices of Abundance of Grey Snapper. SEDAR51-DW-07. SEDAR, North Charleston, SC. 31 pp.

Flaherty-Walia, K.E., Theodore S. Switzer, and Amanda J. Tyler-Jedlund. 2017. Gray Snapper Abundance Indices from Inshore Surveys of Northeastern Gulf of Mexico estuaries. SEDAR51- DW-11. SEDAR, North Charleston, SC. 61 pp.

Gardner, C.L., D.A. DeVries, K.E. Overly, and A.G. Pollack. 2017. Gray snapper Lutjanus griseus Findings from the NMFS Panama City Laboratory Camera Fishery-Independent Survey 2005- 2015. SEDAR51-DW-05. SEDAR, North Charleston, SC. 25 pp.

Hothorn, T, K. Hornik, and A. Zeileis. 2006. Unbiased Recursive Partitioning: A Conditional Inference Framework. Journal of Computational and Graphical Statistics 15: 651-674.

Jackman, S. 2008. Pack: Classes and methods for R developed in the political science computational laboratory, Stanford University. Department of Political Science, Stanford University, Stanford, CA.

R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Thompson, K.A., Theodore S. Switzer, and Sean F. Keenan. 2017. Indices of abundance for Gray Snapper (Lutjanus griseus) from the Florida Fish and Wildlife Research Institute (FWRI) video survey on the West Florida Shelf. SEDAR51-DW-10. SEDAR, North Charleston, SC. 22 pp.

Yates KL, Mellin C, Caley MJ, Radford BT, Meeuwig JJ (2016) Models of Marine Fish Biodiversity: Assessing Predictors from Three Habitat Classification Schemes. PLoS ONE 11(6): e0155634. https://doi.org/10.1371/journal.pone.0155634

Zuur, A.F., E.N. Ieno, N.J. Walkder, A.A. Saveliev, and G.M. Smith. 2009. Mixed effects models and extensions in ecology with R. Spring Science and Business Media, LLC, New York, NY.

Table 1. Summary of sample sizes by year for each of the three included video surveys, Florida Fish and Wildlife Research Institute (FWRI), NMFS Pascagoula (PASC), and NMFS Panama City (PC). No data were available or used from any survey from 1998-2003.

	Survey			
Year	FWRI	PASC	PC	Total
1993		123		123
1994		98		98
1995		69		69
1996		140		140
1997		161		161
2004		149		149
2005		274		274
2006		288	71	359
2007		330	51	381
2008		208	85	293
2009		265	99	364
2010	146	223	143	512
2011	222	348	156	726
2012	237	283	150	670
2013	183	167	94	444
2014	286	235	164	685
2015	224	152	157	533
Total	1298	3513	1170	5981

Table 2. Proportion of sites for each habitat level (**O**ptimal, **P**oor, **S**uboptimal) as determined by individual lab categorical regression trees (CARTs) for Gray Snapper presence. Note the gap in sampling for the Pascagoula lab (1998-2003).

FWRI				Pascagoula			
Year	O	P	S	Year	O	P	S
2010	0.062	0.767	0.171	1993	0.455	0.244	0.301
2011	0.140	0.491	0.369	1994	0.235	0.541	0.224
2012	0.173	0.451	0.376	1995	0.464	0.130	0.406
2013	0.175	0.224	0.601	1996	0.236	0.321	0.443
2014	0.255	0.161	0.584	1997	0.335	0.174	0.491
2015	0.250	0.192	0.558	2004	0.295	0.436	0.268
				2005	0.120	0.562	0.318
Panama City				2006	0.194	0.583	0.222
2006	0.127	0.000	0.873	2007	0.106	0.630	0.264
2007	0.176	0.000	0.824	2008	0.202	0.519	0.279
2008	0.094	0.000	0.906	2009	0.223	0.555	0.223
2009	0.071	0.000	0.929	2010	0.148	0.587	0.265
2010	0.063	0.000	0.937	2011	0.141	0.580	0.279
2011	0.058	0.000	0.942	2012	0.191	0.477	0.332
2012	0.033	0.000	0.967	2013	0.198	0.497	0.305
2013	0.043	0.000	0.957	2014	0.170	0.579	0.251
2014	0.030	0.000	0.970	2015	0.237	0.487	0.276
2015	0.025	0.000	0.975				

Table 3. Likelihood ratio comparisons and AIC values for the combined video survey index for the four potential distributions initially explored.

	Df	Likelihood	χ^2	*p*-value	AIC
Poisson	21	-10107.1			20256.16
Negative Binomial	22	-5156.4	9901.5	<2.2e-16	10356.71
Zero-inflated Poisson	42	-6561.9	2811.1	<2.2e-16	13207.86
Zero-Inflated NB	43	-4900.4	3323.0	<2.2e-16	9886.9

Table 4. Number of stations sampled (N) by survey and year, proportion of positive sets, standardized index, and CV for the annual FWRI Gray Snapper video index of the West Florida Shelf.

Year	Surveys	N	Prop. positive	Std. Index	CV
1993	Pasc	123	0.1707	0.9960	0.27004
1994	Pasc	98	0.1735	1.2119	0.31257
1995	Pasc	69	0.2754	0.7804	0.26544
1996	Pasc	140	0.2000	1.3367	0.26765
1997	Pasc	161	0.2236	1.0878	0.21132
1998					
1999					
2000					
2001					
2002					
2003					
2004	Pasc	149	0.1409	0.6125	0.30286
2005	Pasc	274	0.0693	0.9747	0.27983
2006	Pasc, PC	359	0.1643	0.8913	0.17034
2007	Pasc, PC	381	0.1129	0.4397	0.19663
2008	Pasc, PC	293	0.2048	1.3786	0.16796
2009	Pasc, PC	364	0.2610	1.2904	0.14294
2010	Pasc, PC, FWRI	512	0.1895	1.2539	0.12659
2011	Pasc, PC, FWRI	726	0.1653	0.7204	0.11940
2012	Pasc, PC, FWRI	670	0.1836	0.9801	0.13120
2013	Pasc, PC, FWRI	444	0.2117	0.8685	0.13260
2014	Pasc, PC, FWRI	685	0.2686	1.1143	0.10295
2015	Pasc, PC, FWRI	533	0.2964	1.0628	0.10721

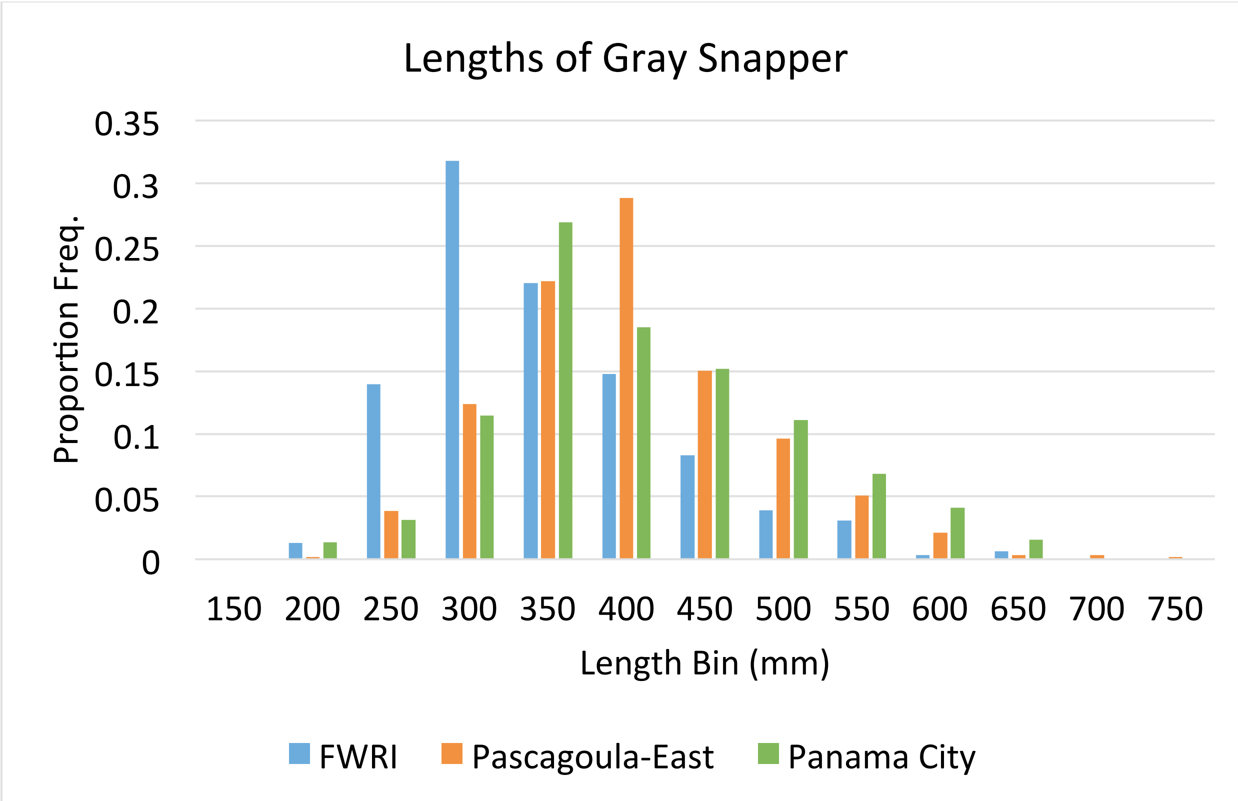

Figure 1. Length frequencies of Gray Snapper observed on video from the three surveys using VMS and SeaGIS.

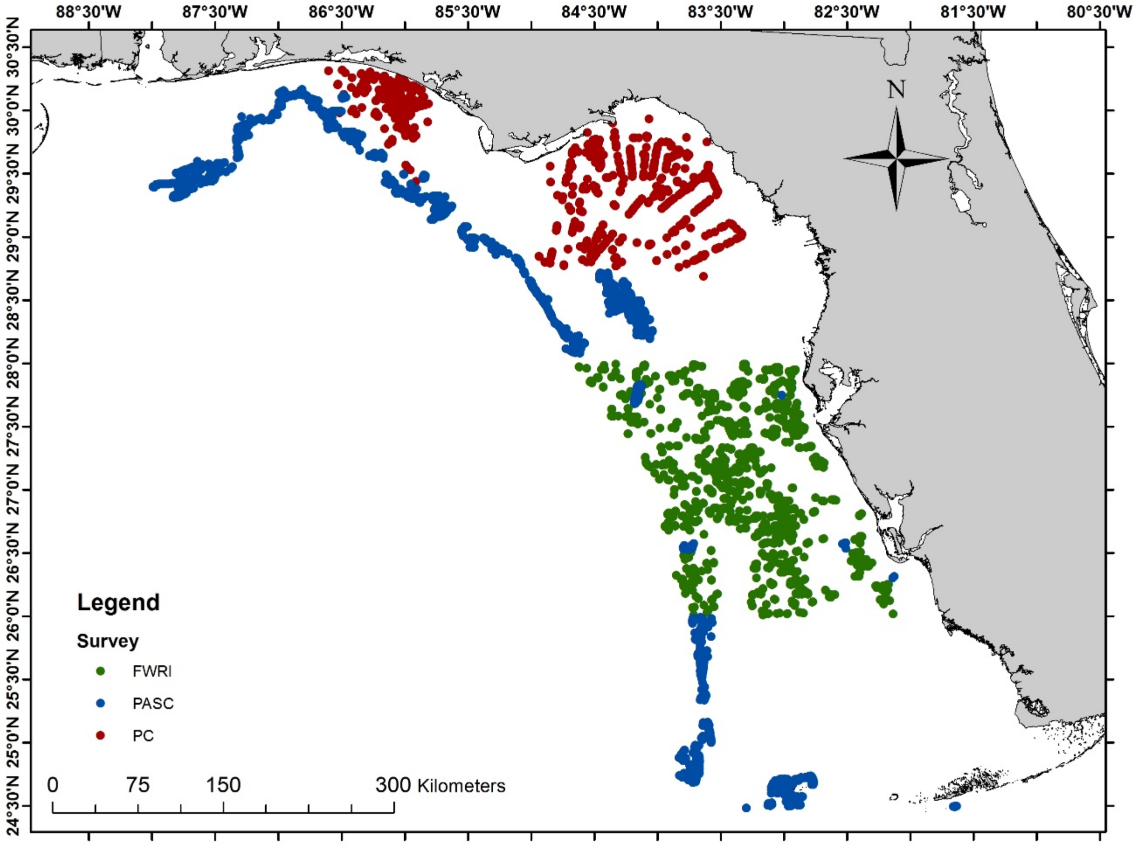

Figure 2. Map of the total video sites included in the index for each survey (by lab) across all years 1993-2015.

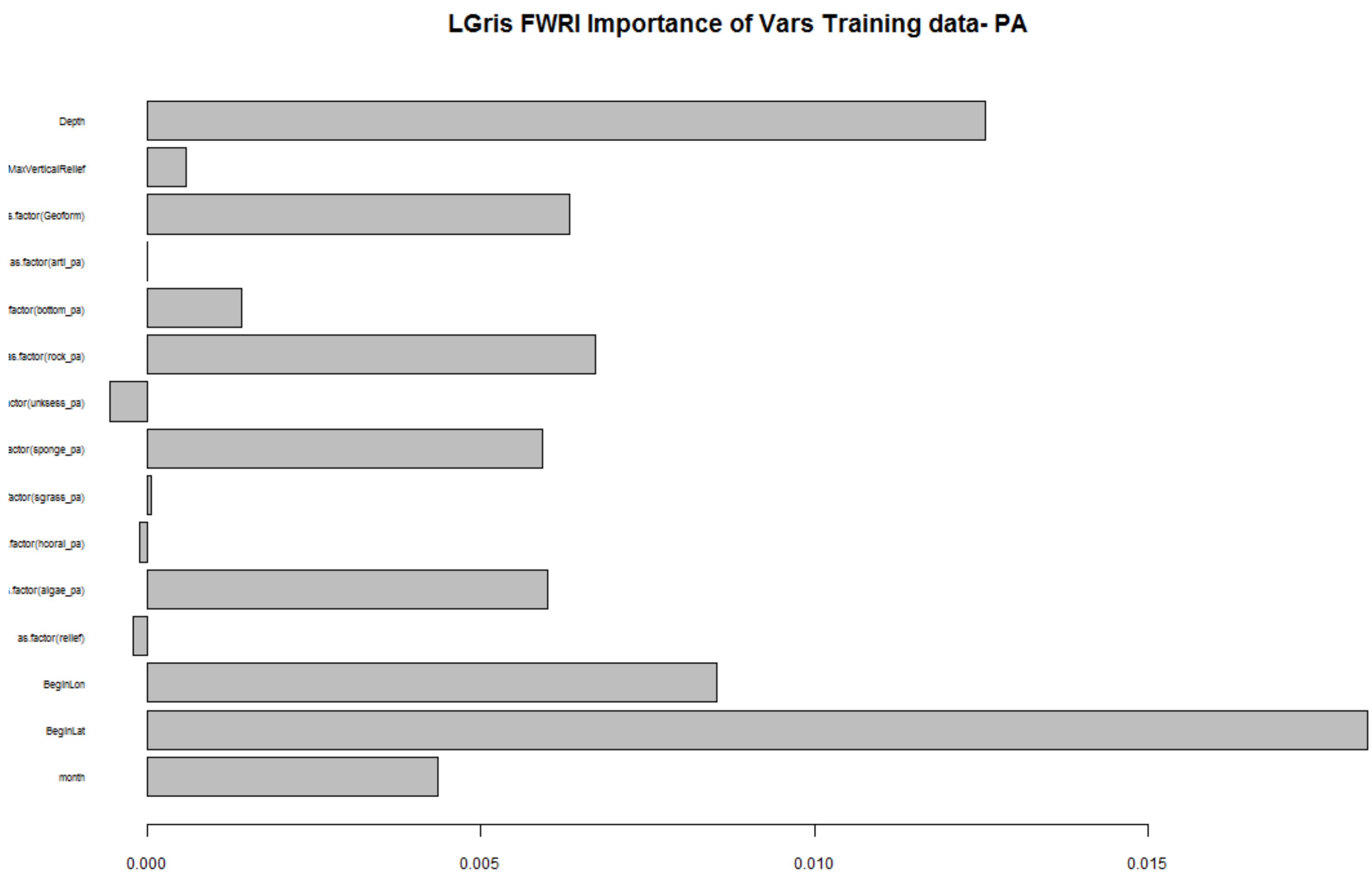

Figure 3. Random Forest generated variable importance for Gray Snapper presence using FWRI survey data.

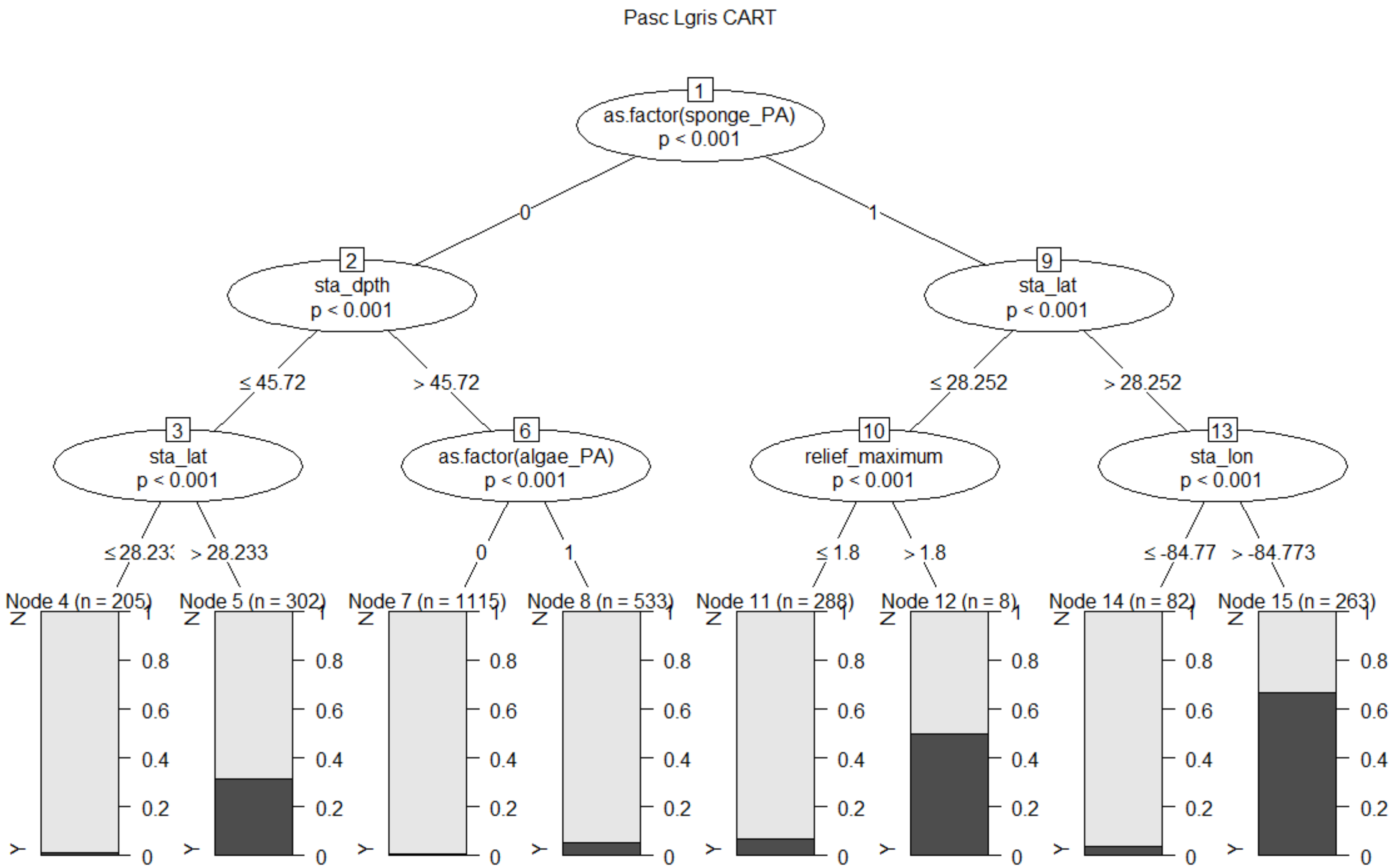

Figure 4. CART results for Gray Snapper for Pascagoula's video survey. Shaded portion of the plots indicate proportion of sites given by a node where Gray Snapper were observed.

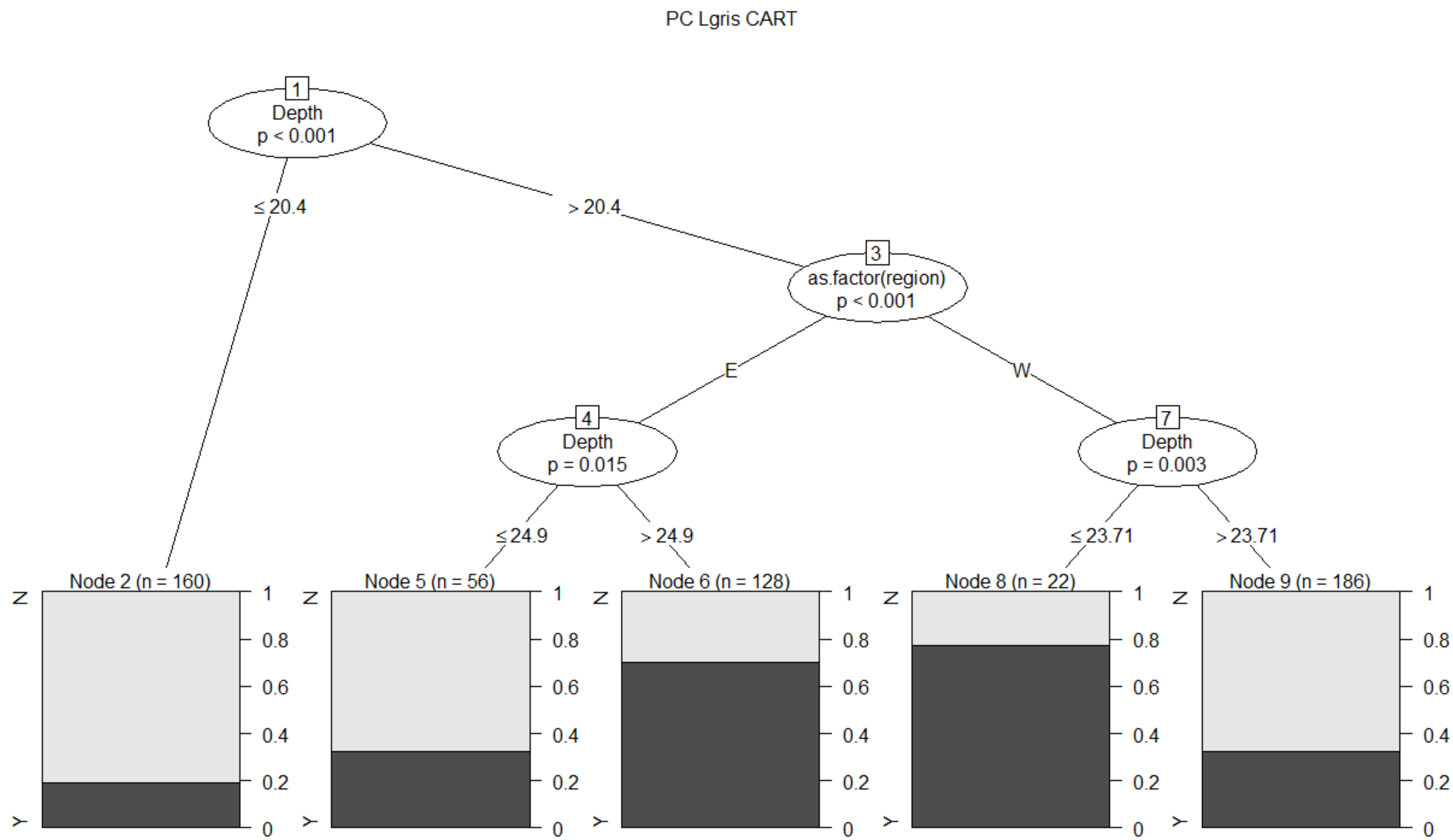

Figure 5. CART results for Gray Snapper for Panama City's video survey. Shaded portion of the plots indicate proportion of sites given by a node where Gray Snapper were observed.

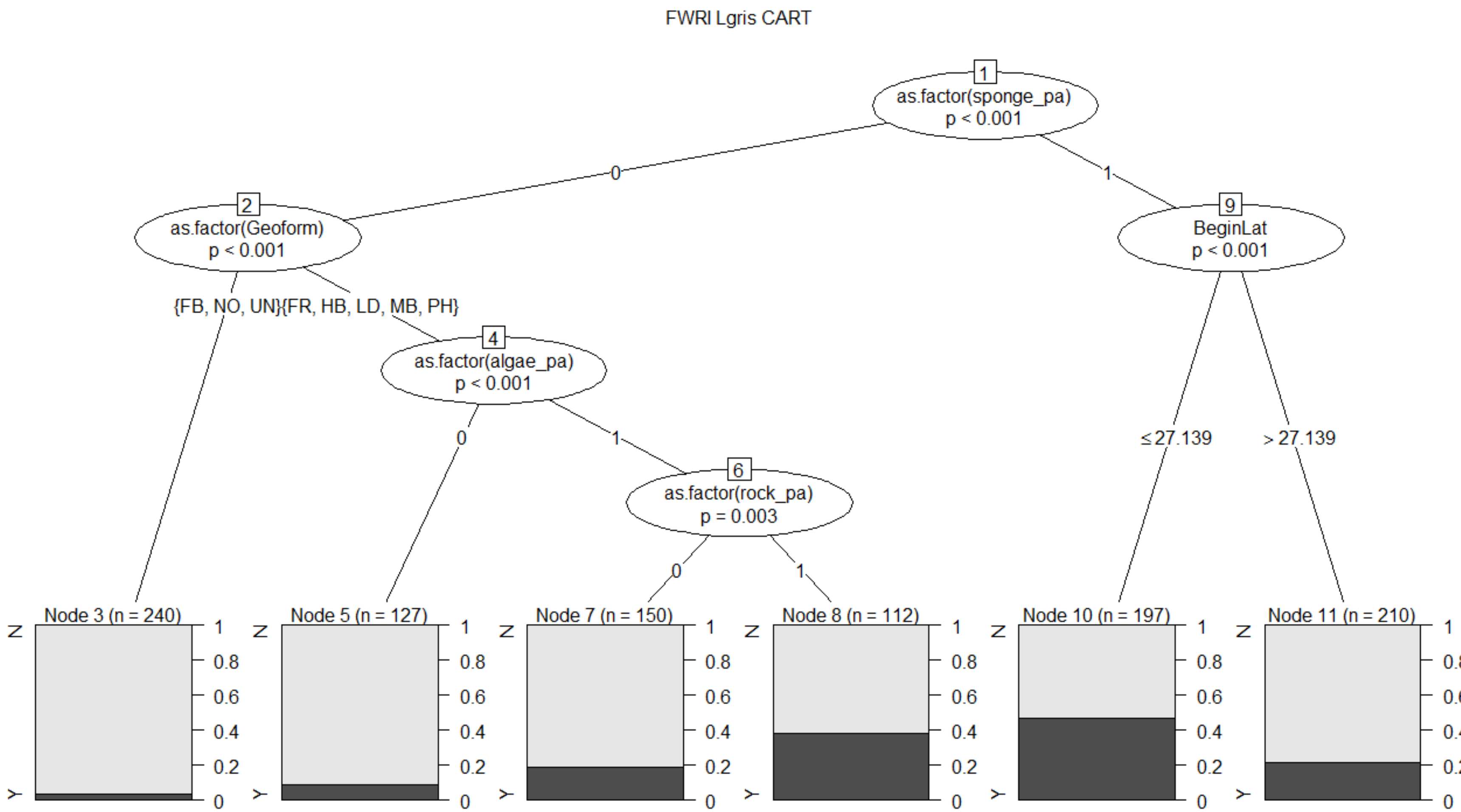

Figure 6. CART results for Gray Snapper for FWRI's video survey. Shaded portion of the plots indicate proportion of sites given by a node where Gray Snapper were observed.

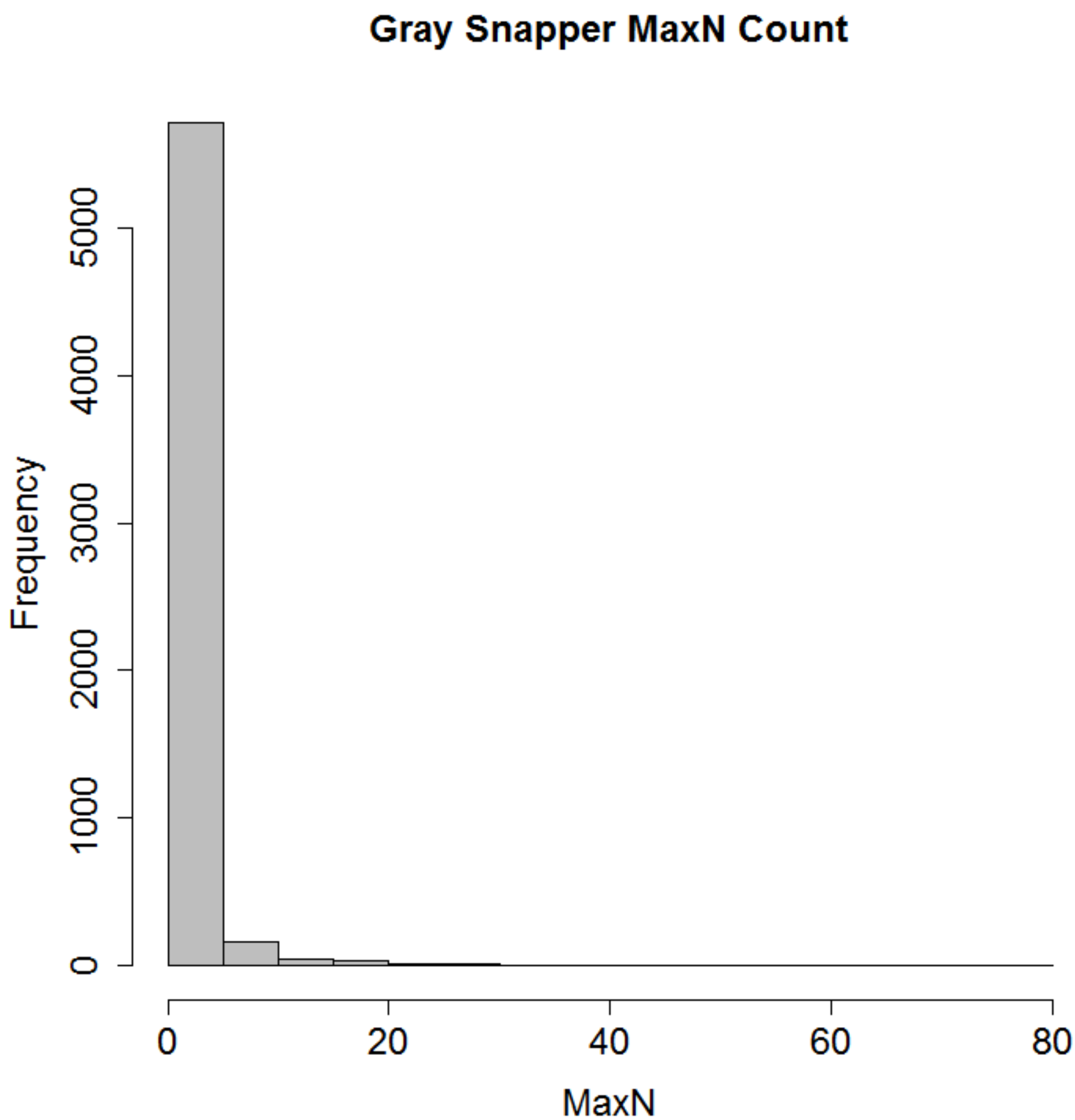

Figure 7. MaxN count distribution for Gray Snapper observed in all three video surveys on the West Florida Shelf used for the combined index.

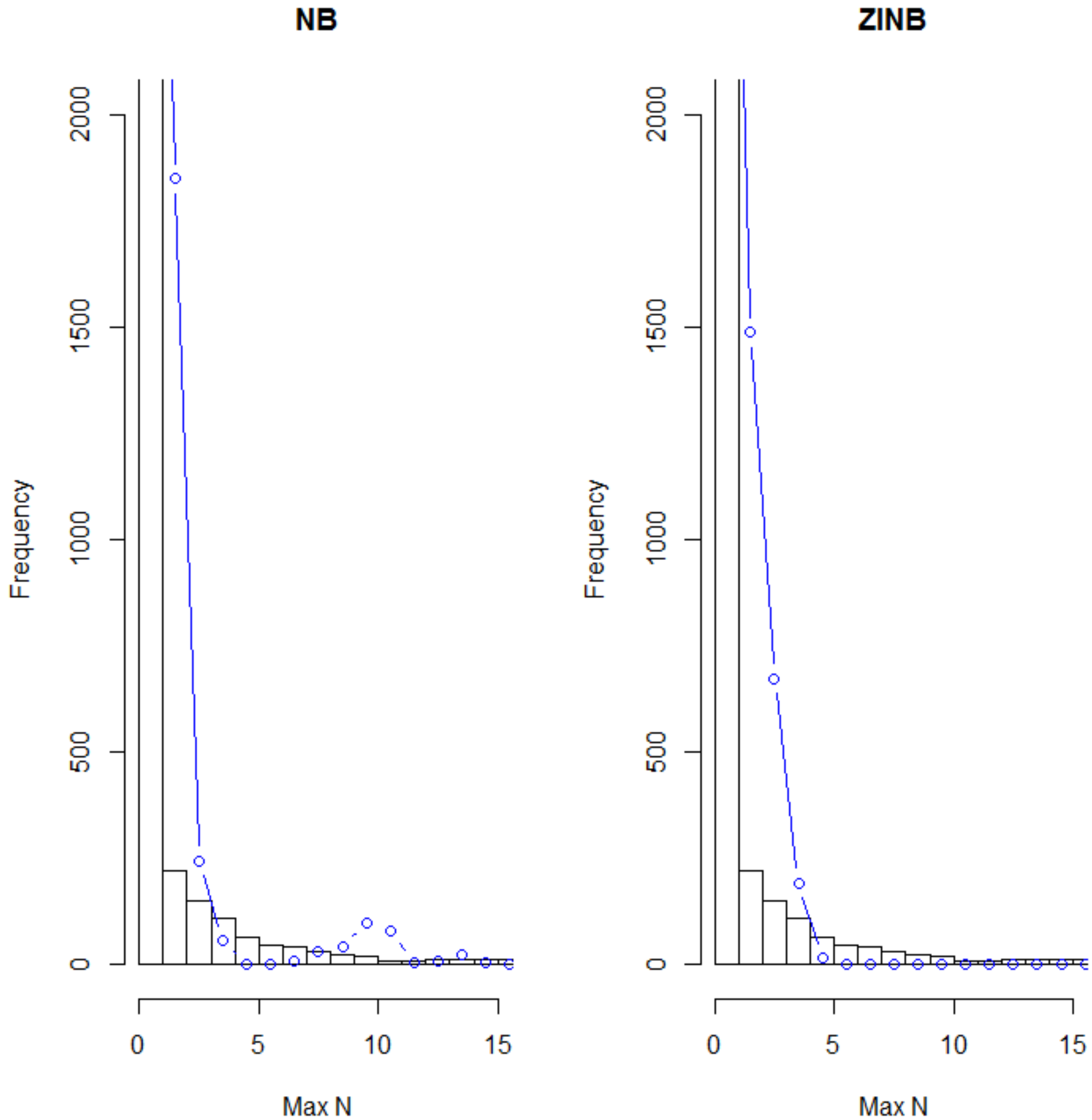

Figure 8. Combined index full model formulation comparison, with the two best models given by AIC, negative binomial (NB) and zero-inflated negative binomial (ZINB) fitted values plotted against the original data distribution.

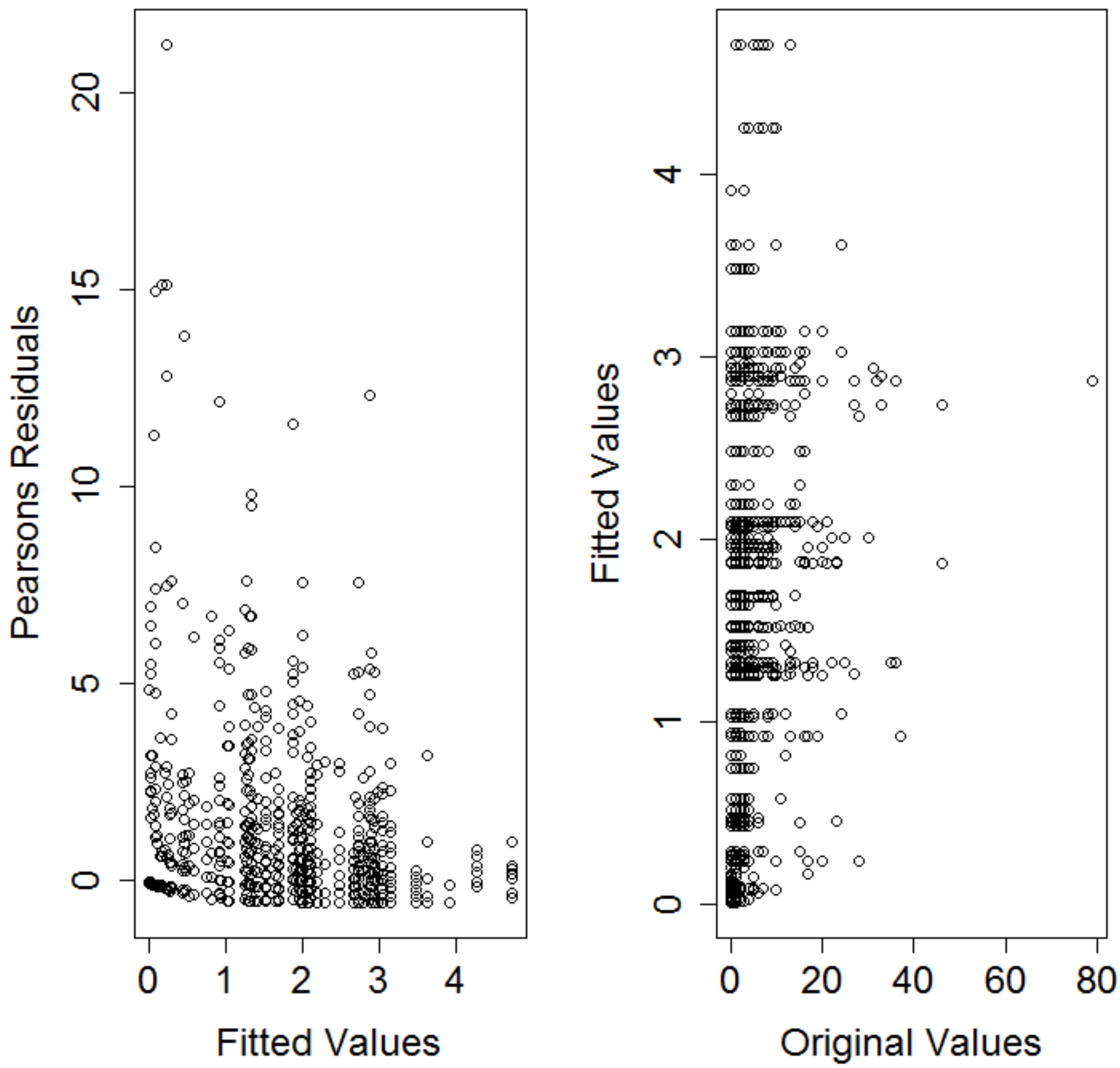

Figure 9. Model diagnostic plots showing fitted best model values against Pearson residuals (left panel) and fitted values plotted against original data values (right panel).

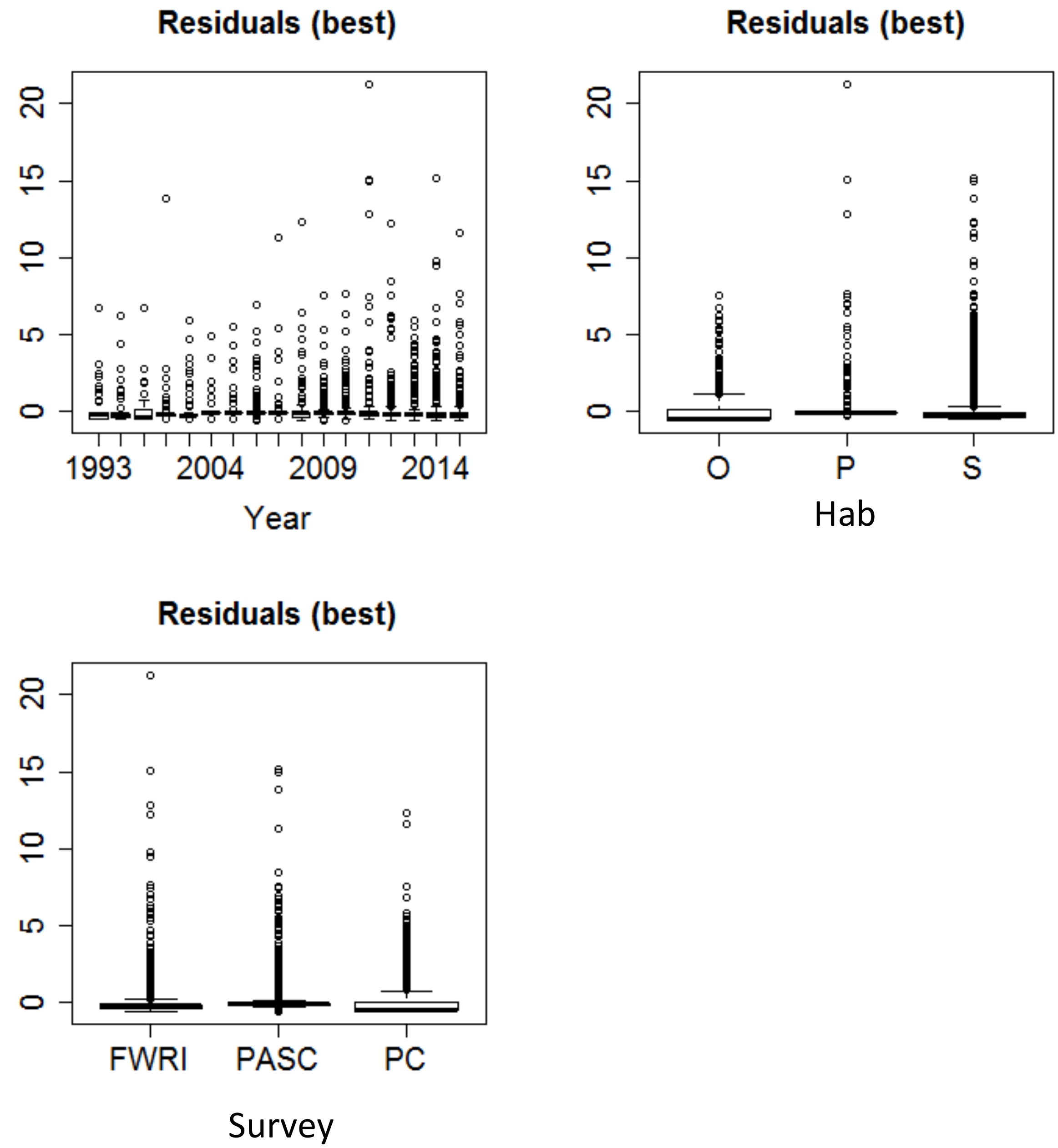

Figure 10. Model diagnostic plots showing Pearson residuals for the final (best) model plotted against spatiotemporal and environmental model parameters.

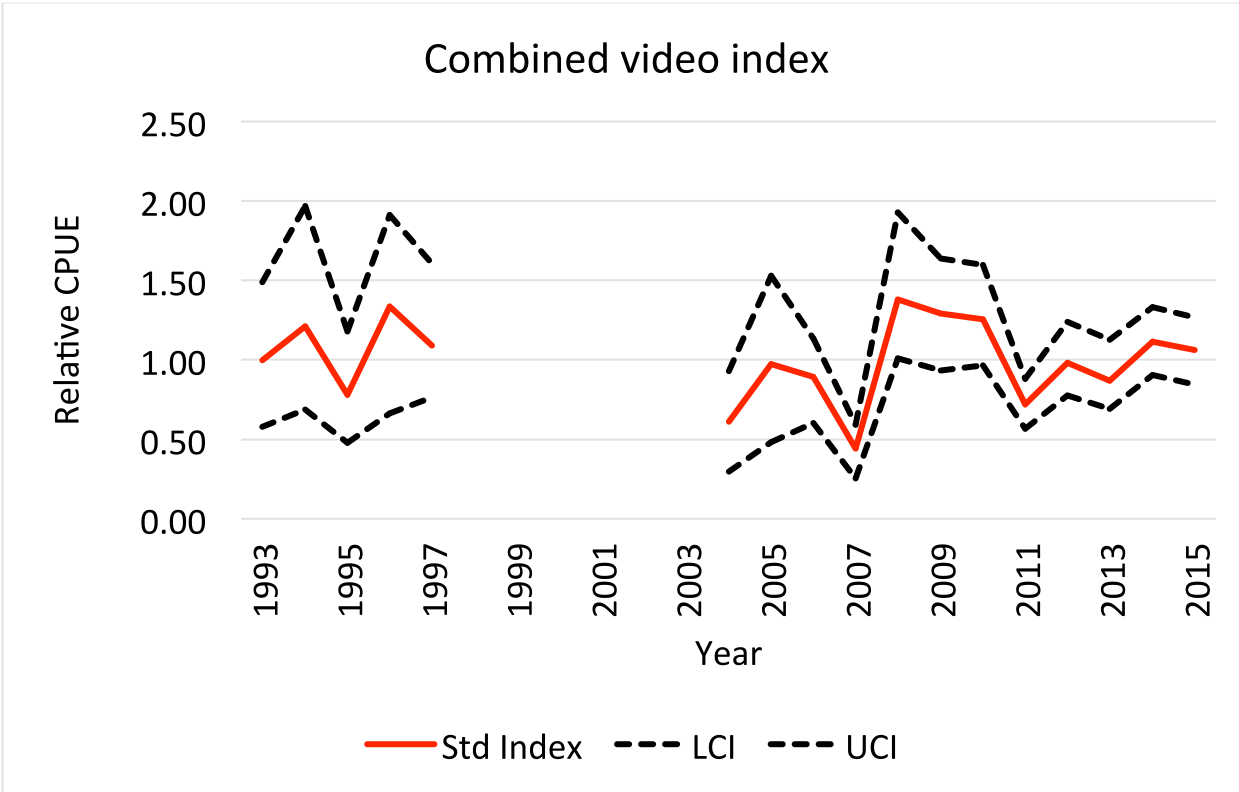

Figure 11. Relative standardized index (solid red line) with 2.5% and 97.5% confidence intervals (black dotted lines) for Gray Snapper CPUE (MaxN) using the integrated West Florida Shelf video data.