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

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

SEDAR52-WP-01

14 September 2017

Please cite this document as:

Thompson, K.A., T.S. Switzer, M.C. Christman, S.F. Keenan, C. Gardner, and M. Campbell. 2017. Indices of abundance for Red Snapper (*Lutjanus campechanus*) using combined data from three independent video surveys. SEDAR52-WP-01. SEDAR, North Charleston, SC. 19 pp.

Indices of abundance for Red Snapper (*Lutjanus campechanus*) 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]

[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

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, 2002, and 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. However, Docs for postingcombining indices across datasets may increase predictive capabilities by allowing for the largest possible sample sizes in model fitting. Previous research has indicated that combining data across changing spatial areas and surveys and using a year only model, can yield to spurious conclusions regarding stock abundance (Campbell 2004; Ye et al. 2004). As such, we used a habitat-based approach to combining relative abundance data for generating annual trends for Red 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).

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 different potential populations of Red 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 1.

Index Construction

Habitat models

To combine the data from all three surveys into one model predicting Red 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 good, fair, or poor (G, F, 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. 2 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 Red 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 Red Snapper catches at each terminal node were then evaluated to determine the habitat characteristics defining good, fair or poor habitat. Terminal nodes with double the overall proportion of positive catches for a dataset were assigned a good habitat code. Poor sites were determined by proportion positives that were at least half of the overall proportion positive and approaching zero. The remaining sites were deemed fair, 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 only longitude showing up in all three models. The Pascagoula model showed 7 total final nodes, defined by presence/absence of seawhips, soft corals, and exposed rock, latitude, and longitude (Fig.3). The model shows one terminal node that define poor habitats for Red Snapper, and three each for fair and good habitats compared to the overall proportion positive of 0.16 (Fig. 3). The Panama City model was the simplest with only three of the continuous variables chosen, depth, latitude, and longitude (Fig. 4). Due to high proportion positive Red Snapper occurrences (0.53 overall) in this survey, habitats were primarily fair, with three terminal nodes getting that distinction and only one of each good and poor habitat nodes (Fig. 4). The FWRI CART model had depth, longitude, presence/absence of exposed rock and algae, as well as side-scanned geoform chosen as explanatory variables (Fig. 5). The final model indicated three good habitat criteria, two fair, and one poor in comparison to the 0.15 overall proportion of sites with Red Snapper observed (Fig. 5)

The site characteristics that define each node and habitat code were then used to create a habitat variable (hab: G, F, 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. 6). 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 Red 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. 7). Backwards variable selection was used and indicated that the full model preformed best, given by AIC, than models with only two or one of the potential variables.

Model diagnostics showed no discernible patterns of association between Pearson residuals and fitted values or the fitted values and the original data (Fig. 8). An examination of residuals for the model parameters (Fig. 9) 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 Red 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 Red

Snapper in the eastern GOM show initially low catches in the Pascagoula SEAMAP survey, followed by a peak in abundance in 2004 and 2005 (Fig. 10). The index included in the previous Red Snapper SEDAR indicates high catches in these years, but shows lower relative index values (Campbell et al. 2012). This disparity is due to the time period being averaged over for relativizing the index, with the previously published one including later years of higher catches beyond 2005. However, in this index the data are adjusted to the mean for three time periods given by unique survey combinations (Table 1). The higher Pascagoula catches in later years are contributing to later time periods overall predicted means, yielding a higher 2004 peak when compared to previous indices (Campbell et al. 2012). Following this peak in abundance, catches remain lower and more stable in subsequent years, given by the combination of Pascagoula and Panama City data until 2010 at which point all three surveys are contributing. Abundances show increasing trends in the last few years of the dataset 2014 on, with the highest overall number of sites with Red Snapper observed in 2016, particularly in the FWRI data (Table 4; Fig. 10).

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.

References Cited:

Campbell, M.D., K.R. Rademacher, P. Felts, B. Noble, M. Felts, and J. Salisbury. 2012. SEAMAP Reef Fish Video Survey: Relative Indices of Abundance of Red Snapper, July 2012. SEDAR31-DW08. SEDAR, North Charleston, SC. 61 pp.

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.

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 1998-2001 and 2003.

Year	FWRI	Pasc	PC	Total
1993		115		115
1994		90		90
1995		61		61
1996		133		133
1997		162		162
2002		152		152
2004		149		149
2005		274		274
2006		276	70	346
2007		319	52	371
2008		206	85	291
2009		262	99	361
2010	146	221	143	510
2011	221	337	156	714
2012	237	281	150	668
2013	184	164	94	442
2014	287	230	154	671
2015	224	152	156	532
2016	194	206	169	569
Total	1493	3790	1328	6611

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 Red Snapper presence. Note the gap in sampling for the Pascagoula lab (1998-2001, 2003).

FWRI				Pascagoula			
Year	G	P	F	Year	G	P	F
2010	0.329	0.329	0.342	1993	0.061	0.670	0.270
2011	0.213	0.588	0.199	1994	0.100	0.689	0.211
2012	0.304	0.536	0.160	1995	0.115	0.787	0.098
2013	0.332	0.380	0.288	1996	0.150	0.609	0.241
2014	0.258	0.481	0.261	1997	0.117	0.710	0.173
2015	0.375	0.362	0.263	2002	0.257	0.559	0.184
2016	0.242	0.376	0.381	2004	0.228	0.483	0.289
				2005	0.204	0.442	0.354
Panama City				2006	0.138	0.551	0.312
2006	0.371	0.414	0.214	2007	0.213	0.439	0.348
2007	0.481	0.308	0.212	2008	0.146	0.621	0.233
2008	0.424	0.424	0.153	2009	0.168	0.542	0.290
2009	0.414	0.273	0.313	2010	0.267	0.498	0.235
2010	0.448	0.140	0.413	2011	0.237	0.466	0.297
2011	0.404	0.212	0.385	2012	0.135	0.605	0.260
2012	0.387	0.147	0.467	2013	0.171	0.427	0.402
2013	0.681	0.170	0.149	2014	0.157	0.517	0.326
2014	0.597	0.162	0.240	2015	0.132	0.461	0.408
2015	0.481	0.122	0.397	2016	0.155	0.583	0.262
2016	0.497	0.124	0.379				

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	23	-9176.2			18398.47
Negative Binomial	24	-5998.3	6355.9	<2.2e-16	12044.58
Zero-inflated Poisson	46	-7112.8	2228.9	<2.2e-16	14317.5
Zero-Inflated NB	47	-5653.0	2919.5	<2.2e-16	11399.99

Table 4. Number of stations sampled (N) by year, proportion of positive sets, nominal MaxN values, standardized index with 97.5% and 2.5& CIs, and CV for the combined Gulf Red Snapper video index.

Year	Surveys	N	Proportion positive	Nominal MaxN	Std. Index	CV	LCI	UCI
1993	Pasc	115	0.0609	0.069565	0.239704	0.498121	0.058844	0.512262
1994	Pasc	90	0.0333	0.044444	0.246446	0.598245	0.023276	0.528516
1995	Pasc	61	0.0164	0.016393	0.109495	0.599260	2.9859E-09	0.342707
1996	Pasc	133	0.0752	0.090226	0.340291	0.359069	0.141126	0.612870
1997	Pasc	162	0.0309	0.104938	0.838795	0.401663	0.161499	1.469448
1998								
1999								
2000								
2001								
2002	Pasc	152	0.1579	0.355263	1.062586	0.238515	0.637627	1.625633
2003								
2004	Pasc	149	0.2013	0.852349	2.829550	0.241152	1.669985	4.188102
2005	Pasc	274	0.1971	0.700730	2.333134	0.199856	1.510920	3.343141
2006	Pasc, PC	346	0.1705	0.826590	0.782701	0.153003	0.582463	1.047074
2007	Pasc, PC	371	0.2022	1.078167	0.886568	0.132214	0.670219	1.138192
2008	Pasc, PC	291	0.2199	1.240550	1.143144	0.179072	0.762245	1.539002
2009	Pasc, PC	361	0.2881	1.678670	1.187587	0.108884	0.948297	1.426815
2010	Pasc, PC, FWRI	510	0.3157	1.352941	1.122333	0.085778	0.932927	1.316963
2011	Pasc, PC, FWRI	714	0.2843	1.138655	1.268902	0.087640	1.071675	1.502488
2012	Pasc, PC, FWRI	668	0.2021	0.544910	0.632286	0.108012	0.519038	0.778329
2013	Pasc, PC, FWRI	442	0.2670	0.873303	0.897194	0.114586	0.717073	1.128008
2014	Pasc, PC, FWRI	671	0.2385	0.678092	0.735178	0.111179	0.585609	0.892941
2015	Pasc, PC, FWRI	532	0.2594	1.001880	0.904423	0.122977	0.700209	1.121632
2016	Pasc, PC, FWRI	569	0.3480	1.448155	1.439685	0.098111	1.182992	1.740870

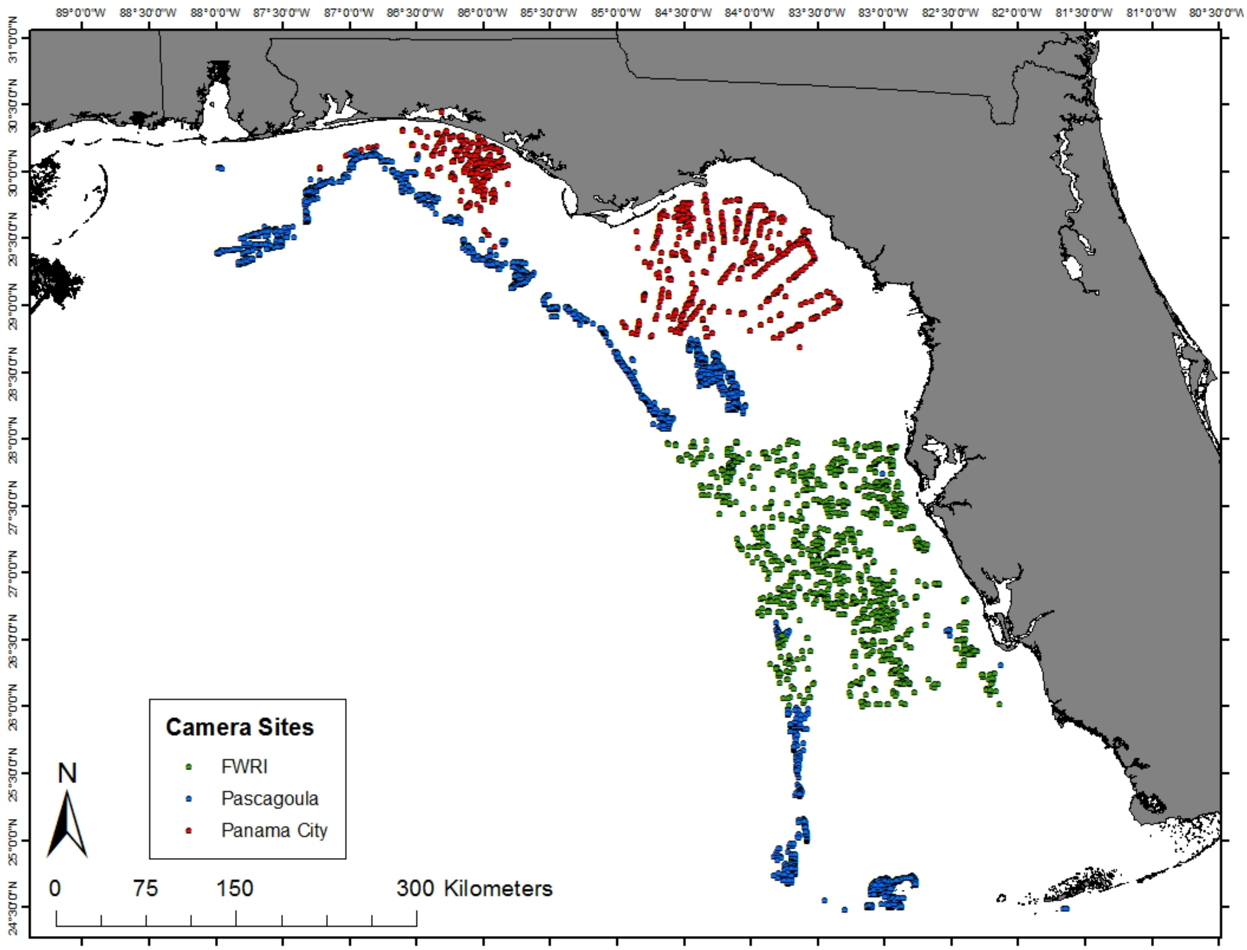

Figure 1. Map of the total video sites included in the index for each survey (by lab) across all years 1993-2016.

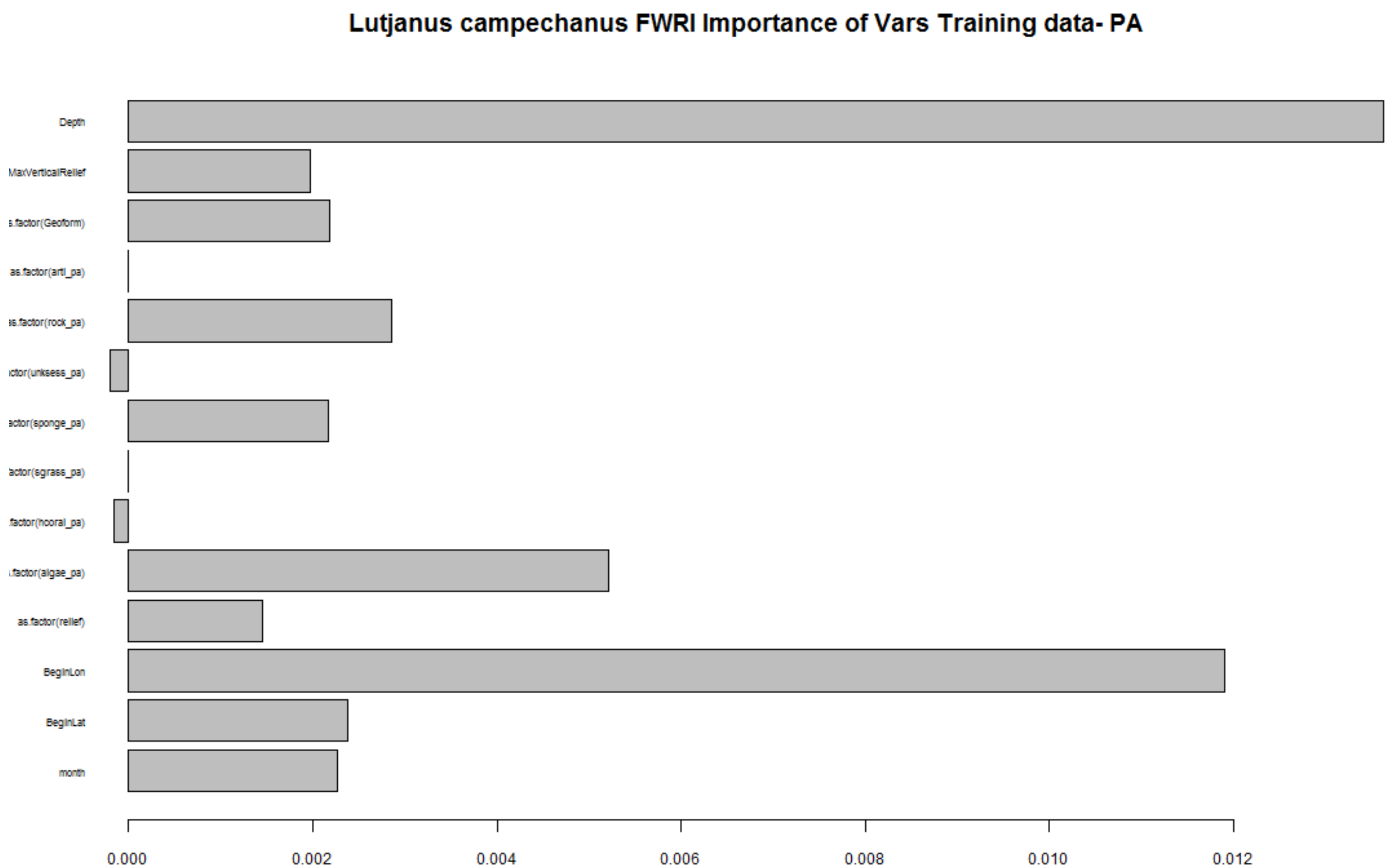

Figure 2. Random Forest generated variable importance for Red Snapper presence using FWRI survey data.

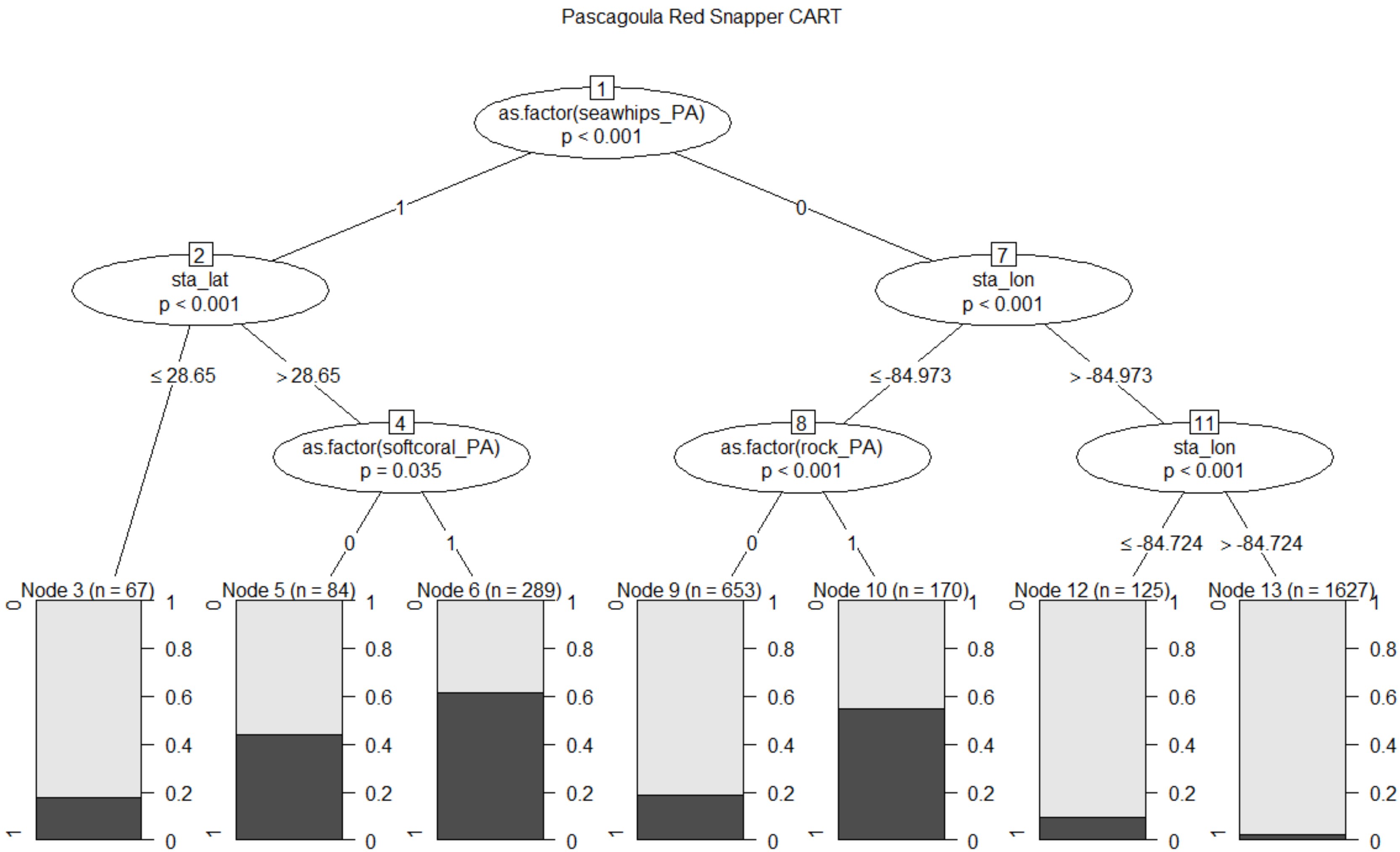

Figure 3. CART results for Red Snapper for Pascagoula's video survey. Shaded portion of the plots indicate proportion of sites given by a node where Red Snapper were observed.

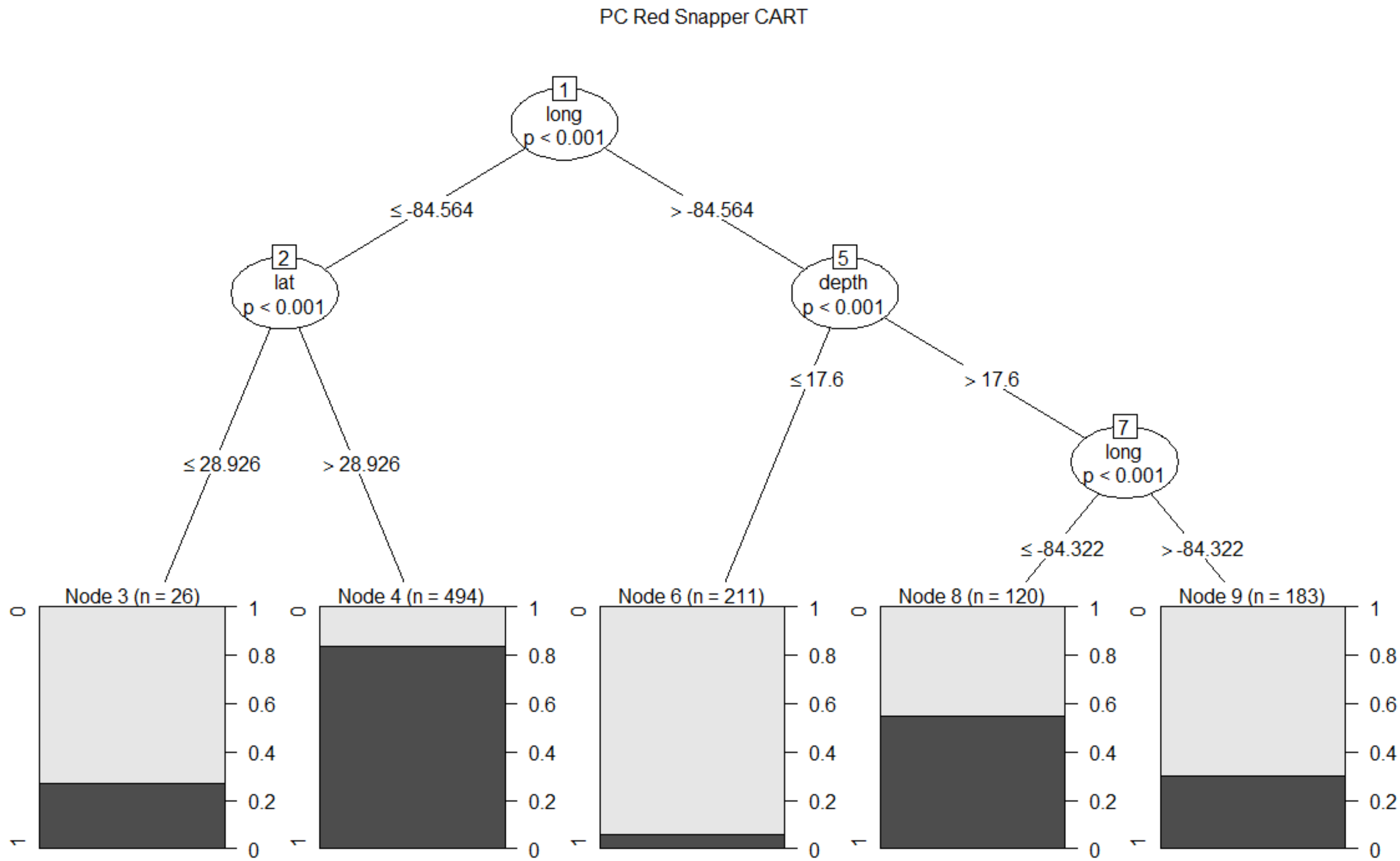

Figure 4. CART results for Red Snapper for Panama City's video survey. Shaded portion of the plots indicate proportion of sites given by a node where Red Snapper were observed.

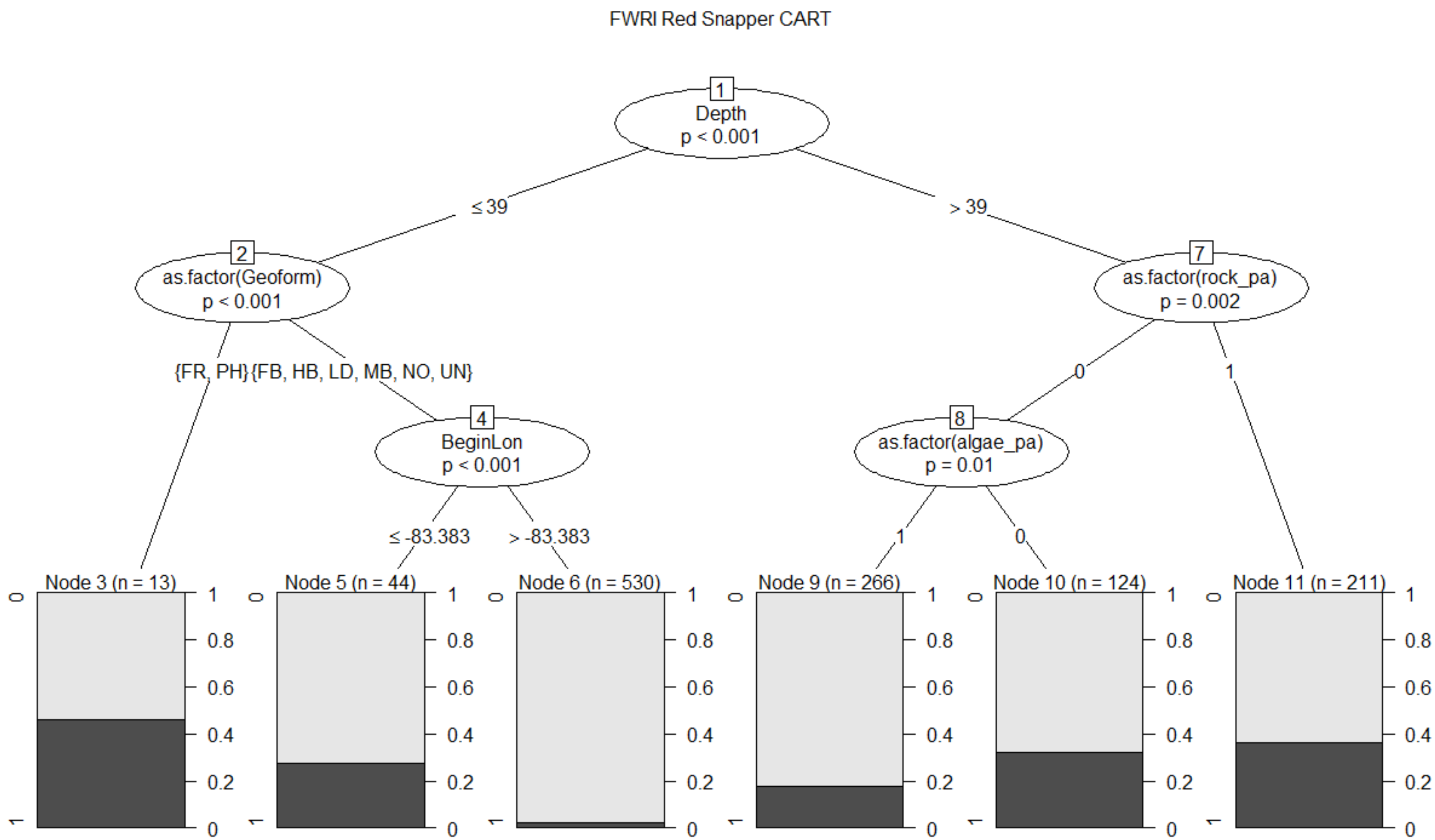

Figure 5. CART results for Red Snapper for FWRI's video survey. Shaded portion of the plots indicate proportion of sites given by a node where Red Snapper were observed.

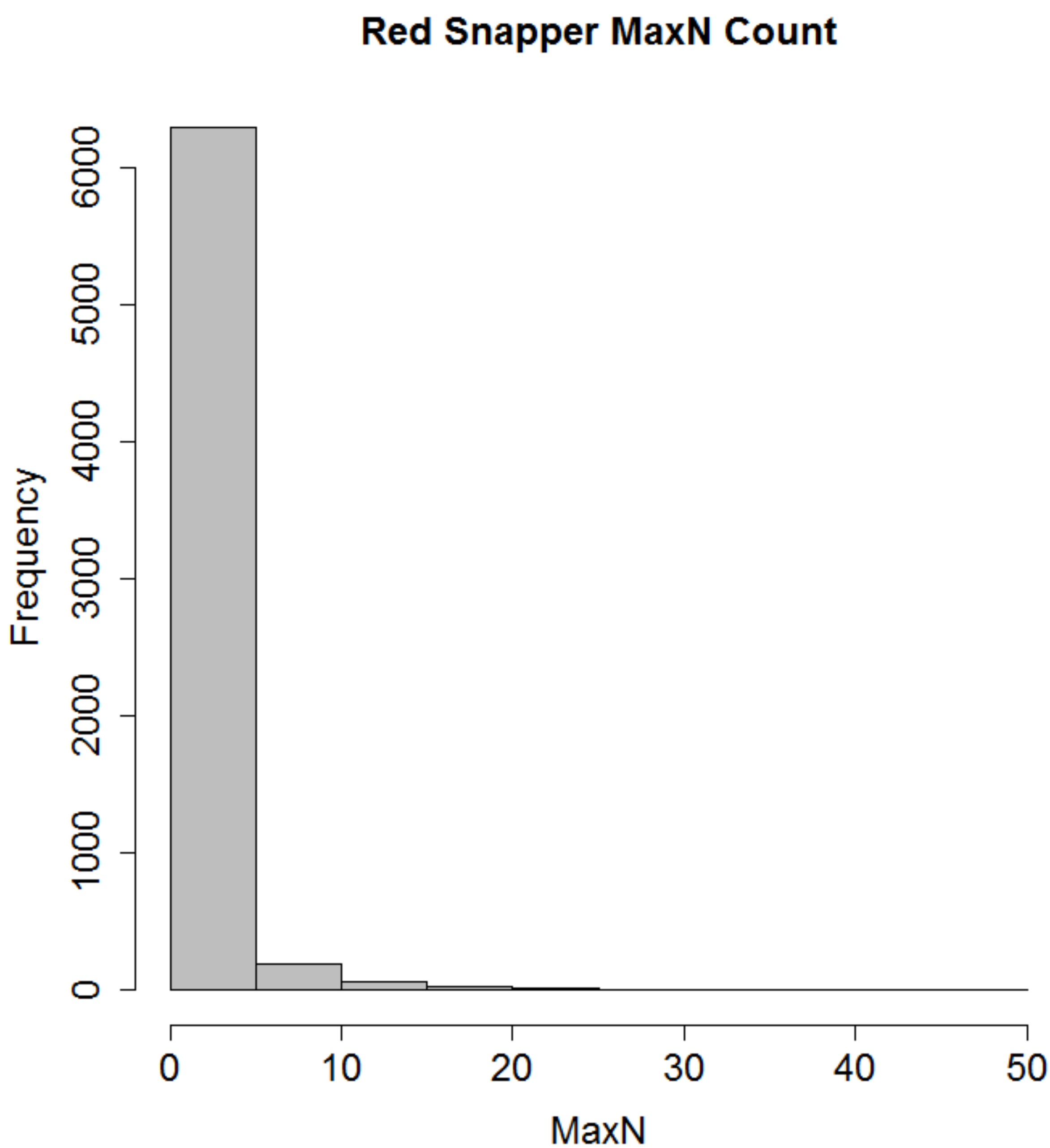

Figure 6. MaxN count distribution, showing zero inflation, for Red Snapper observed in all three video surveys used for the combined index.

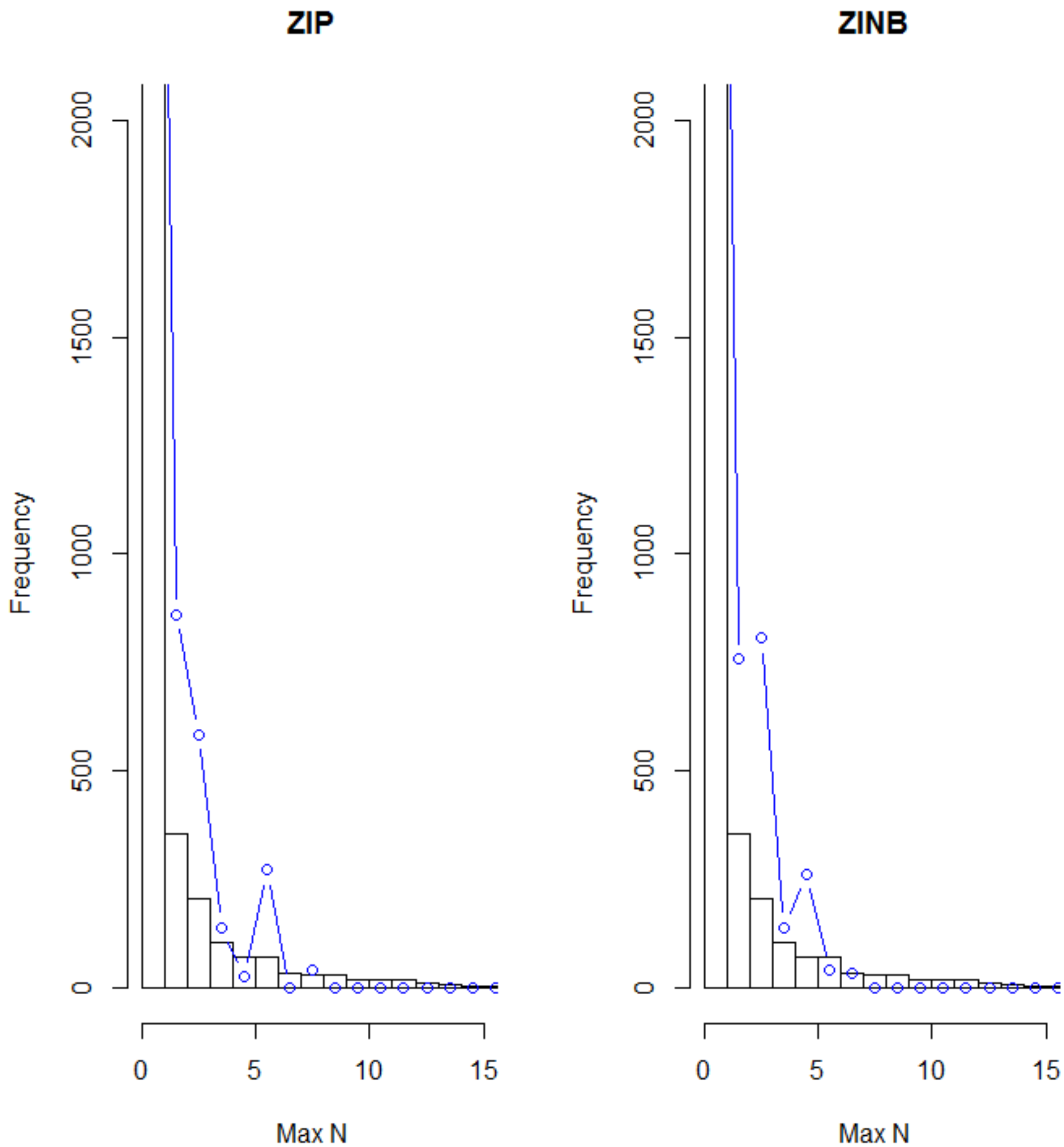

Figure 7. Combined index full model formulation comparison, with the two best models given by AIC, zero-inflated, Poisson (ZIP) and negative binomial (ZINB) fitted values plotted against the original data distribution of MaxN counts of Red Snapper.

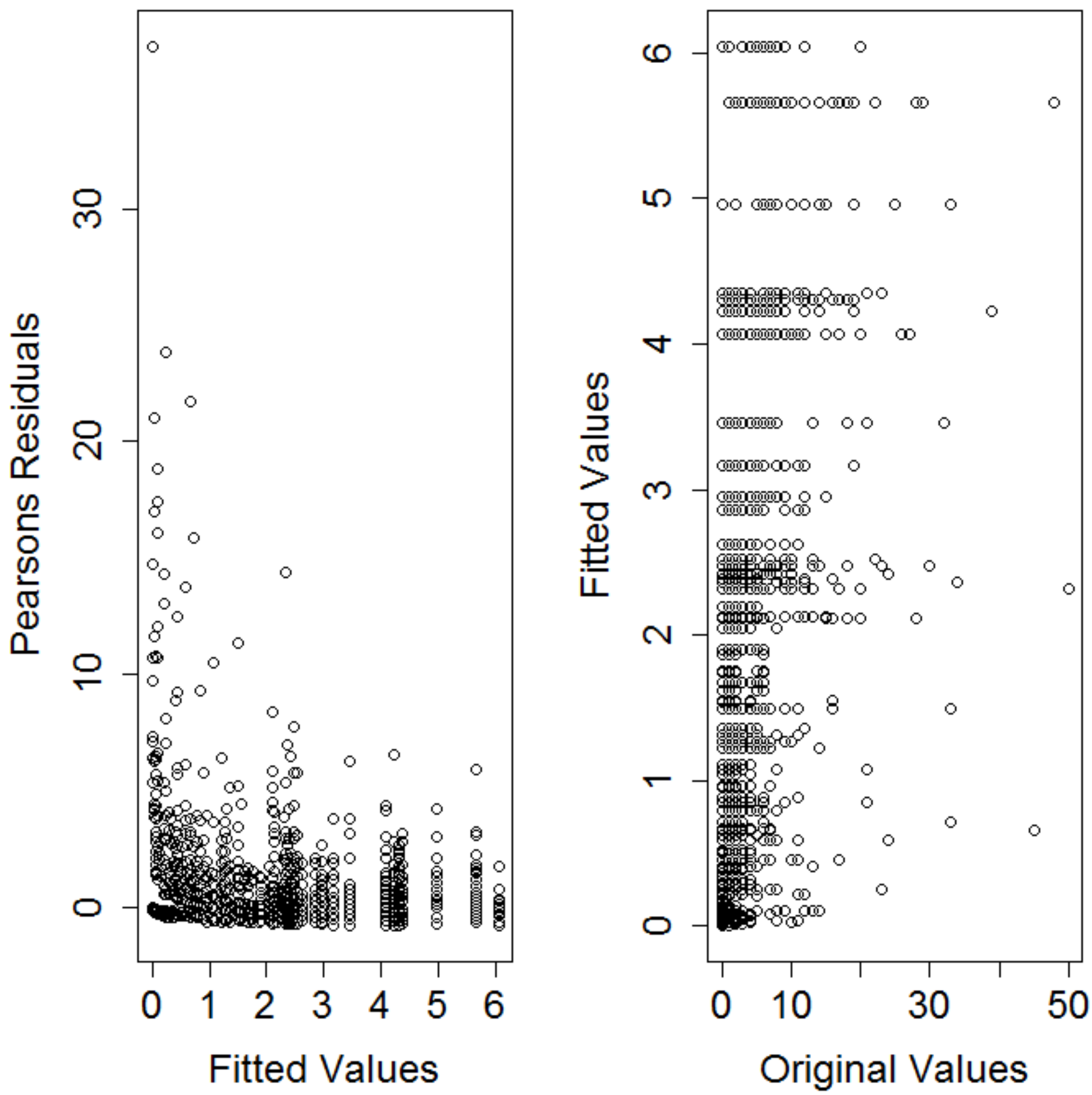

Figure 8. Model diagnostic plots showing fitted best model values against Pearson residuals (left panel) and fitted values plotted against original data values (right panel).

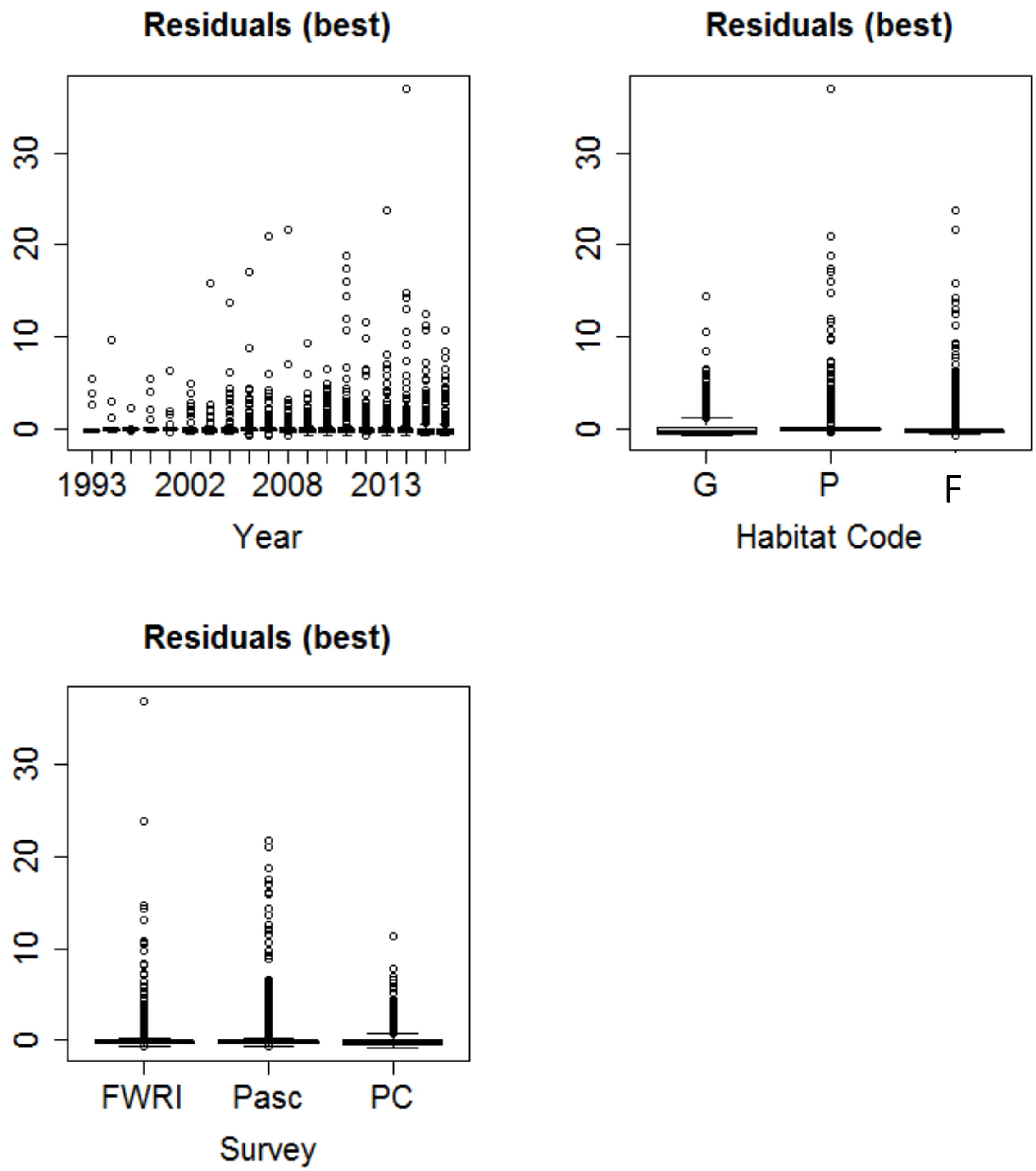

Figure 9. Model diagnostic plots showing Pearson residuals for the final (best) model plotted against spatiotemporal and environmental model parameters.

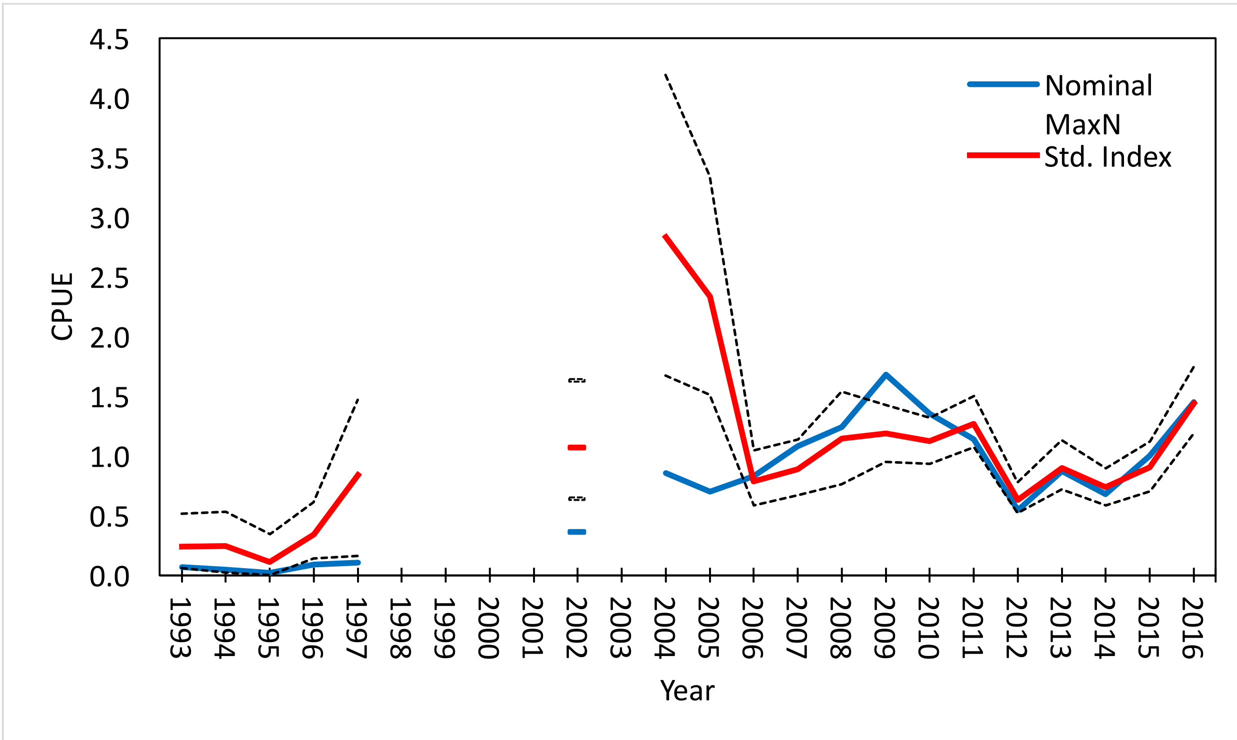

Figure 10. Relative standardized index (solid red line) with 2.5% and 97.5% confidence intervals (black dotted lines) and nominal (mean observed) MaxN values for Red Snapper CPUE using the integrated West Florida Shelf video data.