Indices of abundance for Scamp (*Mycteroperca phenax*) using combined data from three independent video surveys

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

SEDAR68OA-WP-02

28 March 2022

Please cite this document as:

Thompson, Kevin, A., Theodore S. Switzer, Mary C. Christman, Sean F. Keenan, Christopher Gardner, Katherine E. Overly, Matt Campbell. 2022. Indices of abundance for Scamp (*Mycteroperca phenax*) using combined data from three independent video surveys. SEDAR68OA-WP-02. SEDAR, North Charleston, SC. 22 pp.

Indices of abundance for Scamp (*Mycteroperca phenax*) 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], Katherine E. Overly[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

Historically, three different stationary video surveys were conducted for reef fish in the northern Gulf of Mexico (GOM). The NMFS SEAMAP reef fish video survey, carried out by NMFS Mississippi Laboratory (Pascagoula), has the longest running time series (1993-1997, 2002, and 2004+), followed by the NMFS Panama City lab survey (PC; 2005+), with the most recent survey being the Florida Fish and Wildlife Research Institute video survey (FWRI, starting year 2010; Table 1). While the surveys use standardized deployment, camera field of view, and fish abundance methods 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. Traditionally the surveys have submitted independent indices for each survey, however, combining indices across datasets likely increases predictive capabilities by allowing for the largest possible sample sizes in model fitting and encompassing a greater proportion of the distribution of the stock. Previous research has indicated that combining data across changing spatial areas and surveys and using a year only model, can yield spurious conclusions regarding stock abundance (Campbell 2004; Ye et al. 2004). As such, we used a habitat-based approach to combine relative abundance data for generating annual trends for Scamp (*Mycteroperca phenax*) throughout the GOM. The methods presented here and throughout are the same as used in the research track assessment for Scamp in SEDAR 68 (Thompson et al. 2020).

Survey Comparisons

Survey design

The Pascagoula survey primarily targets high-relief topographic features along the continental shelf from south Texas to south Florida (Fig. 1). 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 selected at random from known reef areas identified through habitat mapping (multi-beam and side-scan 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). Because of differences in spatial extent, habitat types and availability, and potential variation habitat association across regions, the east and west regions of this survey were treated as two surveys. This was done to yield more appropriate habitat models as well as appropriate weighting values in the final index values.

The Panama City video survey targets the inner shelf of the northeast GOM (5-60 m depth) ranging from NMFS, SEFSC statistical zone 6 through 10 (Fig. 1). 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. Two known reef sites, a minimum of 250 m apart within each selected block are randomly selected. 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 also include NMFS, SEFSC 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 due to the short time series available (Fig. 1). Sites are initially randomly selected and mapped using side scan sonar over a 2.1 km^2 area (Switzer et al. 2020). Video deployment sites are then randomly assigned proportionally across region and depth zones (Thompson et al. 2017). Relative contribution of each survey by area and habitat observed is given in Table 2.

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 (i.e. MaxN, MinCount). 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).

Fish length measurement

Fish length measurements have varied through time for the surveys, starting with the Pascagoula survey in 1995 fish lengths were measured from video using lasers attached on the camera system with known geometry (Campbell et al. 2017). Panama City survey also used this laser-based approach from 2007 to 2009. However, the frequency of hitting targets with the laser is low and to increase sample size any measurable fish during the video read was measured (i.e. not just at the mincount), and fish could have potentially been measured twice. Subsequent years from (2008 in Pascagoula and 2010 in Panama City) used a stereo-video approach, which is the only method used in the entirety of the FWRI dataset. Vision Measurement System (VMS, Geometrics Inc.) was used to estimate size of fish up to 2014 for all three surveys and all switched to SeaGIS software (SeaGIS Pty. Ltd.) and have used them for the remainder of the timeseries. Length composition data was compared across the surveys to check that similar sized fish were targeted by each survey and to determine the best methods for combining them for the index (Fig. 2).

Data reduction

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

For this assessment, initial scoping calls indicated that data should be combined for Scamp and the congener Yellowmouth Grouper (*Mycteroperca interstitialis*) due to difficulties distinguishing them apart across all gears and surveys. As such, the MaxN values used were the sum of Scamp and Yellowmouth Grouper. However, counts of Yellowmouth Grouper were rare in the PC and FWRI survey (less than 3 observations total) so they were excluded for those datasets. Pascagoula's survey had more occurrences of Yellowmouth Grouper or fish deemed to be either Yellowmouth or a Scamp (45 occurrences for a total of 52 fish) and therefore these observations were included in analyses.

Index Construction

Habitat models

To develop a single index of abundance for Scamp the data from all three surveys was, a habitat variable was created that included each of the separate survey individual variables that could be applied to all the data. This was done so final index models can account for changing sampling effort and habitat allocation through time rather than limiting the model to be predicted only by year and survey. 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 this method accounts for correlations among variables and allows both continuous and categorical data to be included. 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). CART models were not refit for this assessment and the data for habitat classes was applied to the new terminal year of 2019 using the same habitat models used in SEDAR 68, also presented here.

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. Predictor variables included the habitat metrics coded on the video reads (reduced to presence/absence), the latitude and longitude of each site and depth for all three survey sets. For FWRI and Panama City's data, side-scan geoform was also included as a landscape-level habitat variable, with values derived using a modified version of the Coastal and Marine Ecological Classification Standard (CMECS) classification approach. Geoform was not included as a predictor variable for the analysis of Pascagoula survey data because the habitat mapping for that survey has primarily been conducted utilizing multibeam sonar, and at present, comparable habitat classification is not possible using the PASCAGOULA survey multibeam data. 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 analysis, each survey was modeled separately for the entirety of that dataset. The random forest analysis fitted 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 survey dataset.

From the random forest analysis, approximately 50% of the potential variables were retained for each survey given by the importance values for a final CART model. The final model was created by fitting the presence of Scamp 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 Scamp catches at each terminal node was then evaluated to determine the habitat characteristics defining good, fair or poor habitat. Terminal nodes with double (2X) the overall proportion of positive catches for a dataset were assigned a good habitat code. Poor sites were identified as those determined by proportion positives that were at least half (50%) of the overall proportion positive and were generally 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. Scamp habitat models indicated an association with factors commonly attributed to reef or rugose habitats, including rock, relief, and soft coral as well as Geoform for FWRI and PC (Figs. 4-7). Scamp were found to be in a relatively low proportion of sites for FWRI survey (13.4%) with higher occurrence rates for PC (26.1%) and Pascagoula's areas (24.5% in the east and 26.5% for the west).

The site characteristics that define each node and habitat code were then used to create a habitat variable (i.e., 'hab' and coded as: G or F or P) that was then back applied to each site for each of the three survey datasets. The datasets were then combined for the index model. The final proportion of sites in the three habitat categories for each individual survey set and year are shown in Table 3.

Index model fitting and diagnostics

The final model used to index abundance was fit using a negative binomial distribution (given the high occurrence of zero's in the data; Figure 8) with the formula:

*MaxN = Y*Hab *Survey*

Where Hab is the CART derived habitat code and survey represents the survey that collected the data for each site. Backwards variable selection was used and indicated that the full model performed best, given by AIC, compared to models with only one or two of the potential variables.

Model diagnostics indicated no discernible patterns of association between Pearson residuals and fitted values or the fitted values and the original data (Fig. 9), indicating correspondence to underlying model assumptions (Zuur et al. 2009).

The index was fit in SAS using the Proc GLIMMX procedure. To account for the variation in survey area, differences in area mapped with known habitat, and the distribution of Fair, Good, and Poor habitats by survey by year, the estimated MaxN means provided by the GLM were adjusted. The known potential survey universe for each of the three was first multiplied by the proportion of habitat mapping grids that had reef habitat to provide an area weight. This was then multiplied by each year x Survey X hab combination (up to 12 for the final years with three surveys and three habitat levels), providing a weighting factor for each of the mean estimates. Area weighting factors are provided in Table 2. Weighted index values were then standardized to the grand mean.

Length Compositions

Previous SEDARs in which these data have been combined and used for, the panel has suggested length compositions should account for survey differences when possible. Scamp has enough of a sample size to use a model-based, multinomial approach to generate annual length comps. The methods are originally outlined in Walter et al. 2017. For this approach three models were fit and compared using AIC values using the length data from video sets included in the index. Models fit were:

1) LenBin~ year + survey + hab, 2) LenBin~ year + survey , 3) LenBin~ year.

AIC values indicated that the second model that included a year and survey effect only was the most parsimonious. Final model-based, annual length comps for the combined survey were then calculated following that model form.

Results and Discussion:

Annual standardized index values for Scamp in the GOM, including coefficients of variation, are presented in Table 4. The model CV's indicate a good model fit, with highest values in earlier years ~20-35%, but steadily decreasing CV's as additional surveys are added and continue with CV's in the range of ~15% in the final years. CVs and confidence limits were found to be high in the years of 1995. Biomass trends for Scamp in the GOM show low and variable numbers early in the time series, followed by a peak in abundance in 2004 and subsequent decline to low but stable numbers in since 2012 (Table 4; Fig. 10). Multinomial length compositions are shown in Fig. 11.

References Cited:

Campbell, R.A. 2004. CPUE standardization and the construction of indices of stock abundance in a spatially varying fishery using general linear models. Fisheries Research 70: 209-227.

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.

Carruthers, T.R., J.F. Walter, M., M.K. McAllister, and M.D. Bryan. 2015. Modelling age-dependent movement: an application to red and gag groupers in the Gulf of Mexico. Canadian Journal of Fisheries and Aquatic Sciences 7: 1159-1176.

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/.

Switzer, T.S, A.J. Tyler-Jedlund, S.F. Keenan, and E.J. Weather. 2020. Benthic habitats, as derived from classification of side-scan sonar mapping data, are important determinants of reef-fish assemblage structure in the Eastern Gulf of Mexico. Marine and Coastal Fisheries. 12:21-32.

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.

Thompson, Kevin, A., Theodore S. Switzer, Mary C. Christman, Sean F. Keenan, Christopher Gardner, Matt Campbell. 2018. Indices of abundance for Red Grouper (Epinephelus morio) using combined data from three independent video surveys. SEDAR61-WP-03. SEDAR, North Charleston, SC. 18 pp.

Thompson, Kevin A. Theodore S. Switzer, Mary C. Christman, Sean F. Keenan, Christopher Gardner, Katherine E. Overly, Matt Campbell. 2019. Indices of abundance for Vermilion Snapper (Rhomboplites aurorubens) using combined data from three independent video surveys. SEDAR67-WP-03. SEDAR, North Charleston, SC. 17 pp.

Walter, J., Kevin Thompson and Ted Switzer. 2020. Model-based size composition of vermilion snapper obtained from three visual surveys. SEDAR67-WP-16. SEDAR, North Charleston, SC. 15 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, East and West regions, and NMFS Panama City. No data were available or used from any survey from 1998-2001; 2003.

Year	FWRI	Pascagoula East	Pascagoula West	PC	Total
1993		123	57		180
1994		99	61		160
1995		69	56		125
1996		140	172		312
1997		162	134		296
2002		152	108		260
2004		149	51		200
2005		274	140		414
2006		288	162	95	545
2007		330	192	63	585
2008		208	131	90	429
2009		265	183	107	555
2010	158	223	114	145	640
2011	222	349	105	158	834
2012	237	283	202	150	872
2013	185	167	145	97	594
2014	287	235	113	164	799
2015	224	152	59	168	603
2016	195	206	178	171	750
2017	154	223	211	150	738
2018	127	213	201	101	642
2019	183	310	279	108	880
Total	1972	4620	3054	1767	11413

Table 2. Proportion of sites for each habitat level (**F**air**, G**ood**, P**oor) as determined by individual survery categorical regression trees (CARTs) for Greater Amberjack presence. Note the gap in sampling for the Pascagoula lab (1998-2002 and 2003).

Pascagoula East				Pascagoula West			
Year	F	G	P	Year	F	G	P
1993	0.74	0.08	0.18	1993	0.40	0.40	0.19
1994	0.77	0.04	0.19	1994	0.28	0.41	0.31
1995	0.54	0.12	0.35	1995	0.29	0.36	0.36
1996	0.64	0.15	0.21	1996	0.55	0.27	0.18
1997	0.62	0.13	0.25	1997	0.32	0.50	0.18
2002	0.40	0.30	0.30	2002	0.27	0.54	0.19
2004	0.53	0.25	0.22	2004	0.29	0.31	0.39
2005	0.54	0.18	0.28	2005	0.35	0.35	0.30
2006	0.53	0.14	0.33	2006	0.31	0.20	0.49
2007	0.49	0.17	0.34	2007	0.26	0.30	0.44
2008	0.46	0.11	0.43	2008	0.33	0.16	0.51
2009	0.48	0.18	0.34	2009	0.22	0.26	0.52
2010	0.50	0.15	0.35	2010	0.29	0.20	0.51
2011	0.48	0.22	0.30	2011	0.24	0.38	0.38
2012	0.41	0.15	0.44	2012	0.38	0.24	0.38
2013	0.59	0.20	0.21	2013	0.34	0.24	0.42
2014	0.46	0.18	0.36	2014	0.31	0.27	0.42
2015	0.59	0.14	0.27	2015	0.32	0.27	0.41
2016	0.40	0.16	0.45	2016	0.30	0.39	0.31
2017	0.44	0.09	0.47	2017	0.26	0.25	0.49
2018	0.50	0.11	0.39	2018	0.29	0.18	0.53
2019	0.63	0.07	0.29	2019	0.27	0.24	0.49

FWRI				Panama City			
Year	F	G	P	Year	F	G	P
2010	0.57	0.10	0.33	2006	0.16	0.24	0.60
2011	0.43	0.08	0.49	2007	0.40	0.13	0.48
2012	0.57	0.07	0.36	2008	0.30	0.18	0.52
2013	0.60	0.11	0.29	2009	0.48	0.16	0.36
2014	0.56	0.10	0.34	2010	0.57	0.14	0.28
2015	0.57	0.07	0.36	2011	0.46	0.11	0.43
2016	0.54	0.05	0.41	2012	0.55	0.07	0.39
2017	0.48	0.05	0.47	2013	0.80	0.00	0.20
2018	0.50	0.06	0.44	2014	0.57	0.13	0.30
2019	0.46	0.19	0.34	2015	0.57	0.04	0.39
				2016	0.62	0.04	0.34
				2017	0.40	0.17	0.43

2018	0.22	0.09	0.69
2019	0.49	0.12	0.39

Table 3. The habitat weighting used with the annual distribution of Fair, Good, Poor habitats to adjust estimated model means to account for sampling variation across surveys.

Survey	Total Universe Area (km2)	Proportion of grids with habitat	Total Universe Area X Prop transects	Area Weighting values (1993-2005)	Area Weighting values (2006-2009)	Area Weighting values (2010-2019)
Pascagoula E	34490	0.81	27936.9	0.707	0.514	0.429
Pascagoula W	31258	0.37	11565.46	0.293	0.213	0.177
PC	22104	0.67	14860.9	0.000	0.273	0.228
FWRI	37290	0.29	10814.09	0.000	0.000	0.166

Table 4. Number of stations sampled (N) by survey and year, proportion of positive sets, standardized index, and CV for the annual FWRI Scamp video index of the Gulf of Mexico.

Year	N	Prop present	Std. Index	Std. Nominal	CV	Lower 95% CI	Upper 95% CI
1993	180	0.217	0.888	0.981	0.174	0.666	1.111
1994	160	0.150	0.508	0.543	0.236	0.335	0.680
1995	125	0.224	0.577	0.646	0.257	0.364	0.790
1996	312	0.218	0.794	0.776	0.176	0.593	0.995
1997	296	0.236	0.659	0.911	0.135	0.531	0.787
2002	260	0.419	1.795	1.898	0.143	1.426	2.163
2004	200	0.305	2.031	2.315	0.177	1.515	2.547
2005	414	0.290	1.530	1.677	0.135	1.233	1.827
2006	545	0.169	0.961	0.905	0.170	0.726	1.196
2007	585	0.287	1.563	1.479	0.122	1.289	1.837
2008	429	0.273	1.155	1.186	0.149	0.908	1.401
2009	555	0.265	1.254	1.229	0.129	1.022	1.485
2010	640	0.239	1.094	1.095	0.125	0.897	1.291
2011	834	0.254	1.206	1.280	0.098	1.036	1.377
2012	872	0.182	0.687	0.749	0.121	0.568	0.806
2013	594	0.215	0.744	0.785	0.119	0.617	0.872
2014	799	0.208	0.894	0.856	0.119	0.742	1.046
2015	603	0.226	0.958	0.866	0.133	0.775	1.140
2016	750	0.245	0.806	0.895	0.103	0.687	0.925
2017	738	0.225	0.766	0.825	0.117	0.638	0.895
2018	642	0.181	0.566	0.662	0.109	0.478	0.655
2019	880	0.166	0.564	0.633	0.102	0.482	0.647

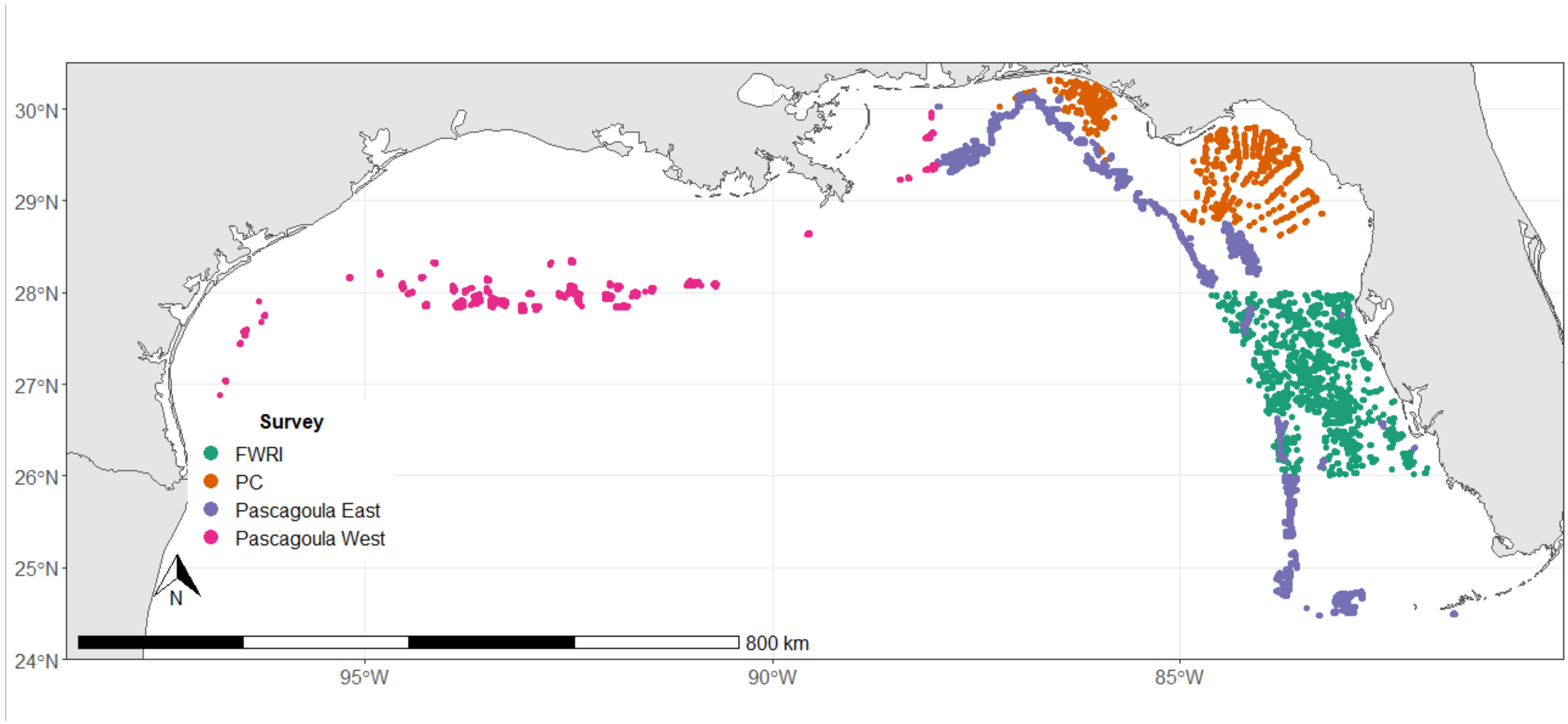

Figure 1. Map of all video sites included in the index for each survey across all years 1993-2018.

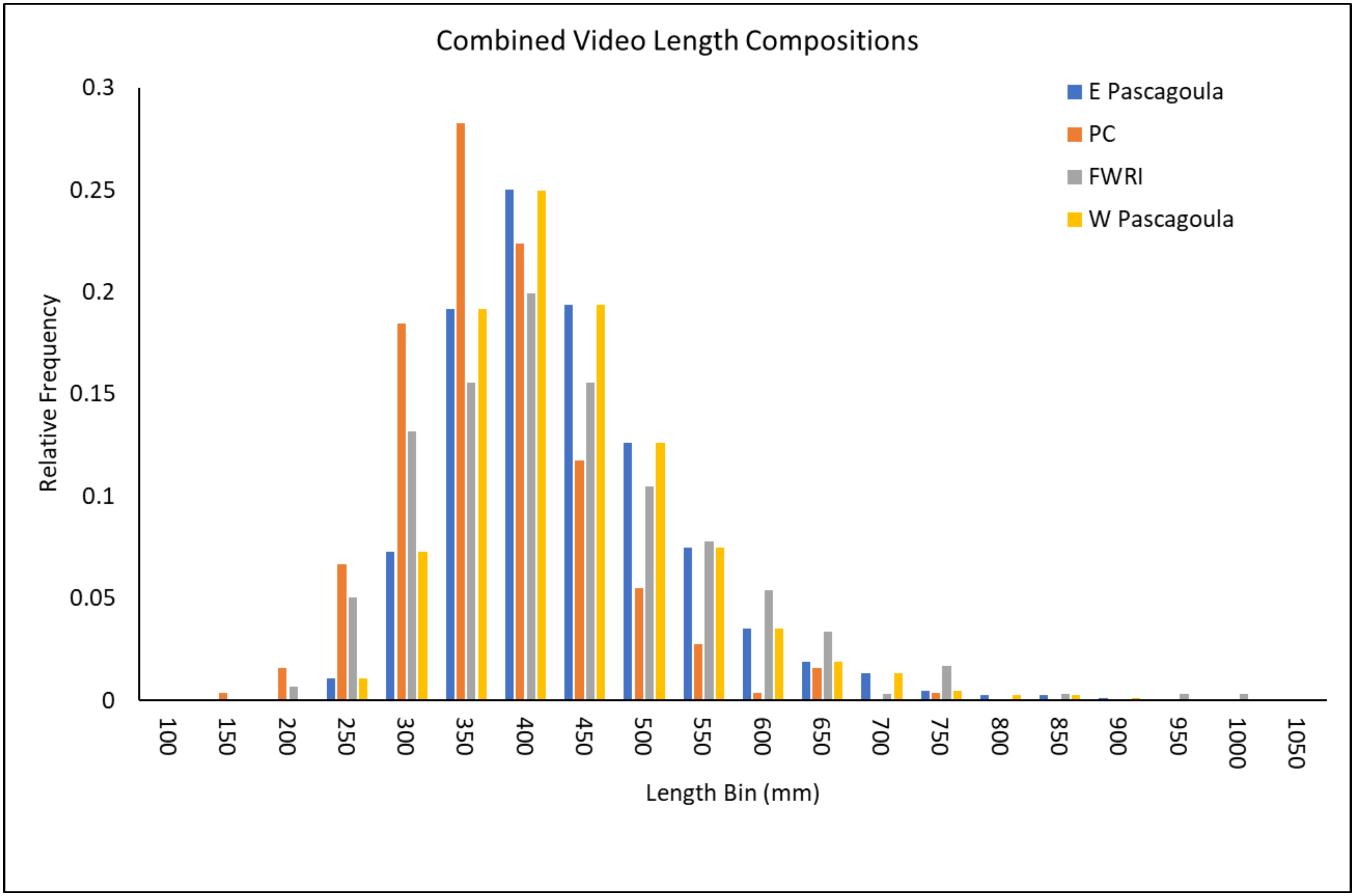

Figure 2. Nominal length compositions of the three surveys used in the combined index for Scamp.

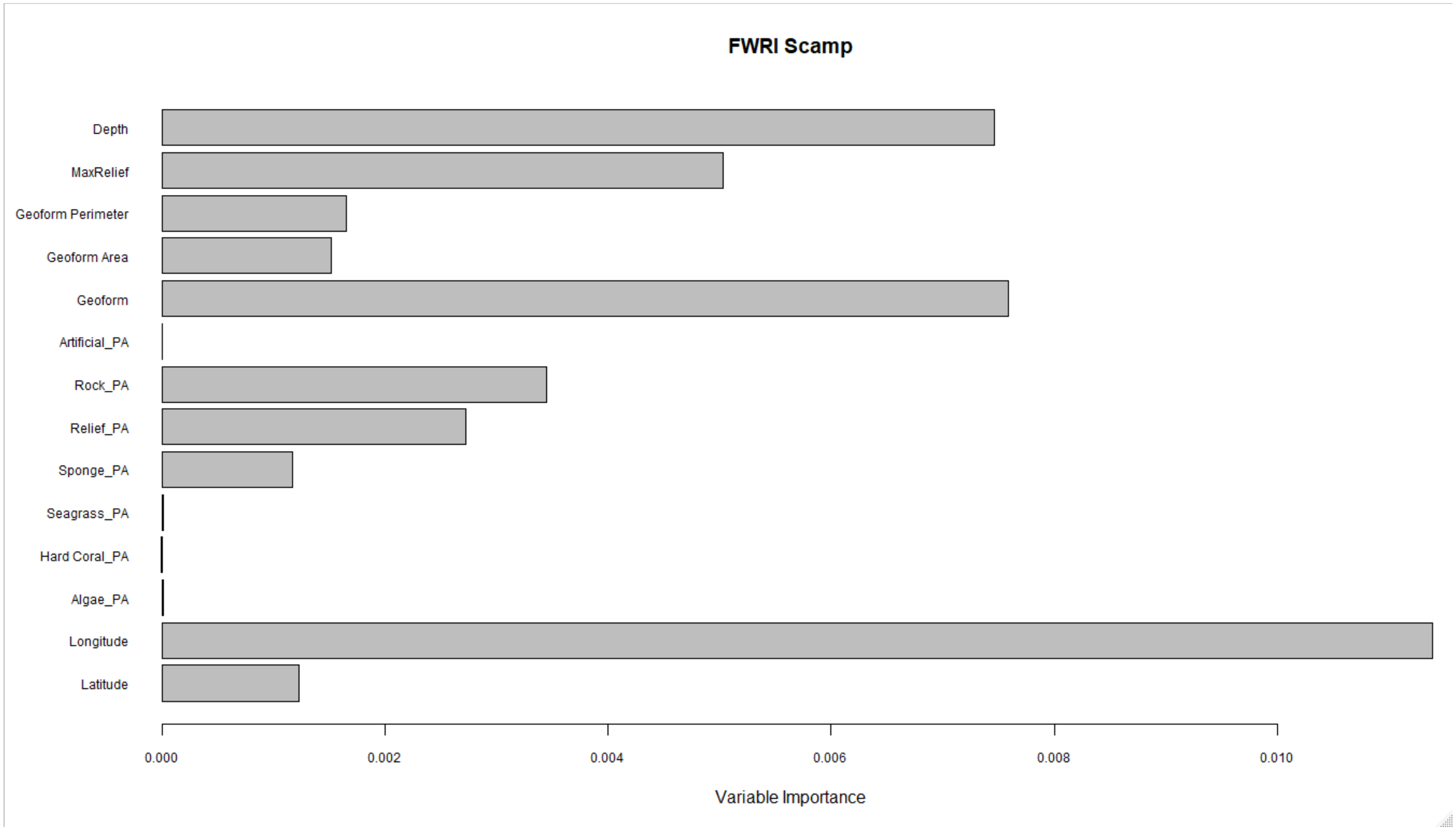

Figure 3. Random Forest generated variable importance for Scamp presence using FWRI survey data.

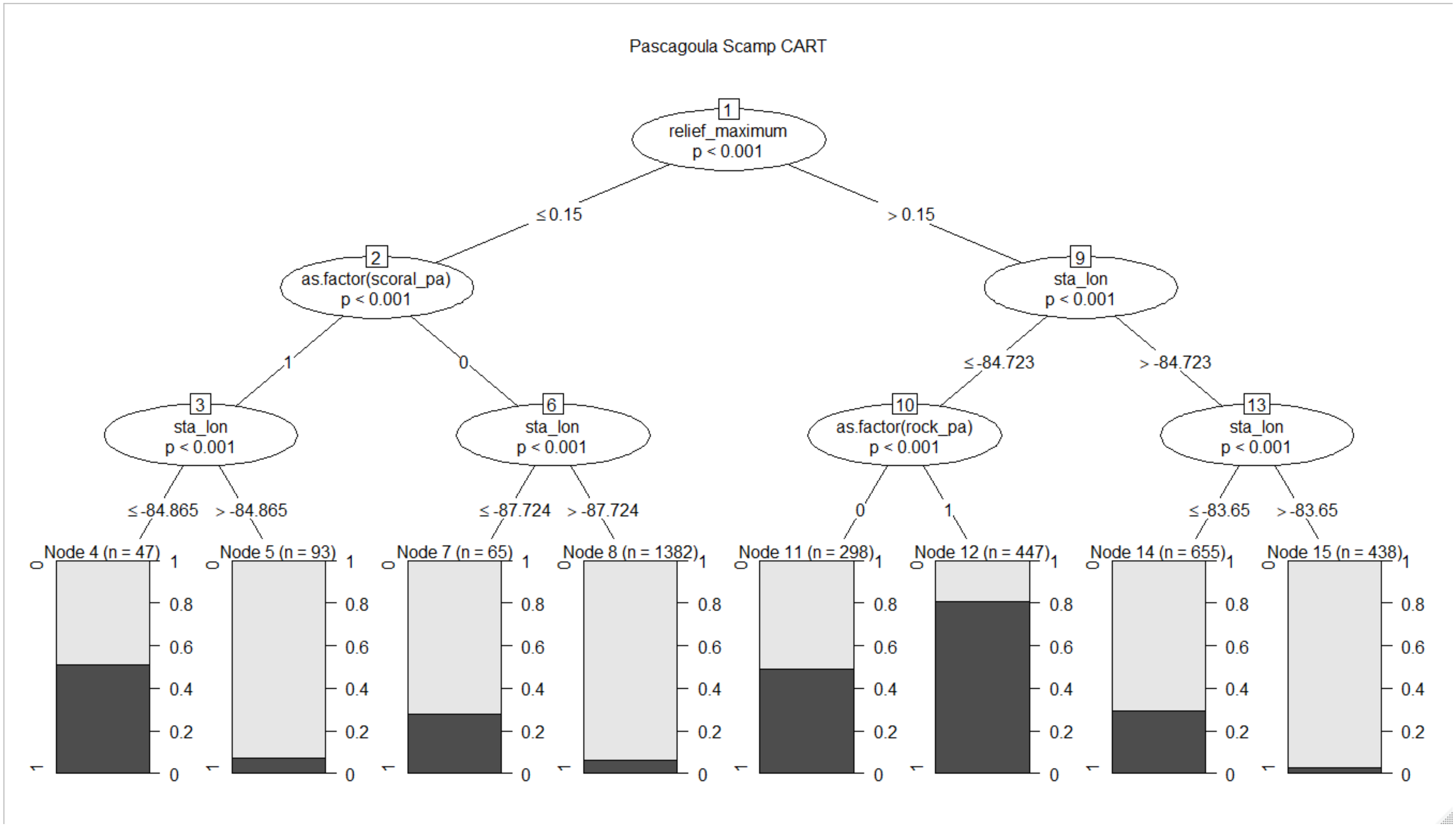

Figure 4. CART results for Scamp for Pascagoula video survey for the eastern Gulf region. Shaded portion of the plots indicate proportion of sites given by a node where Scamp were observed (24.5% of sites had Scamp present overall).

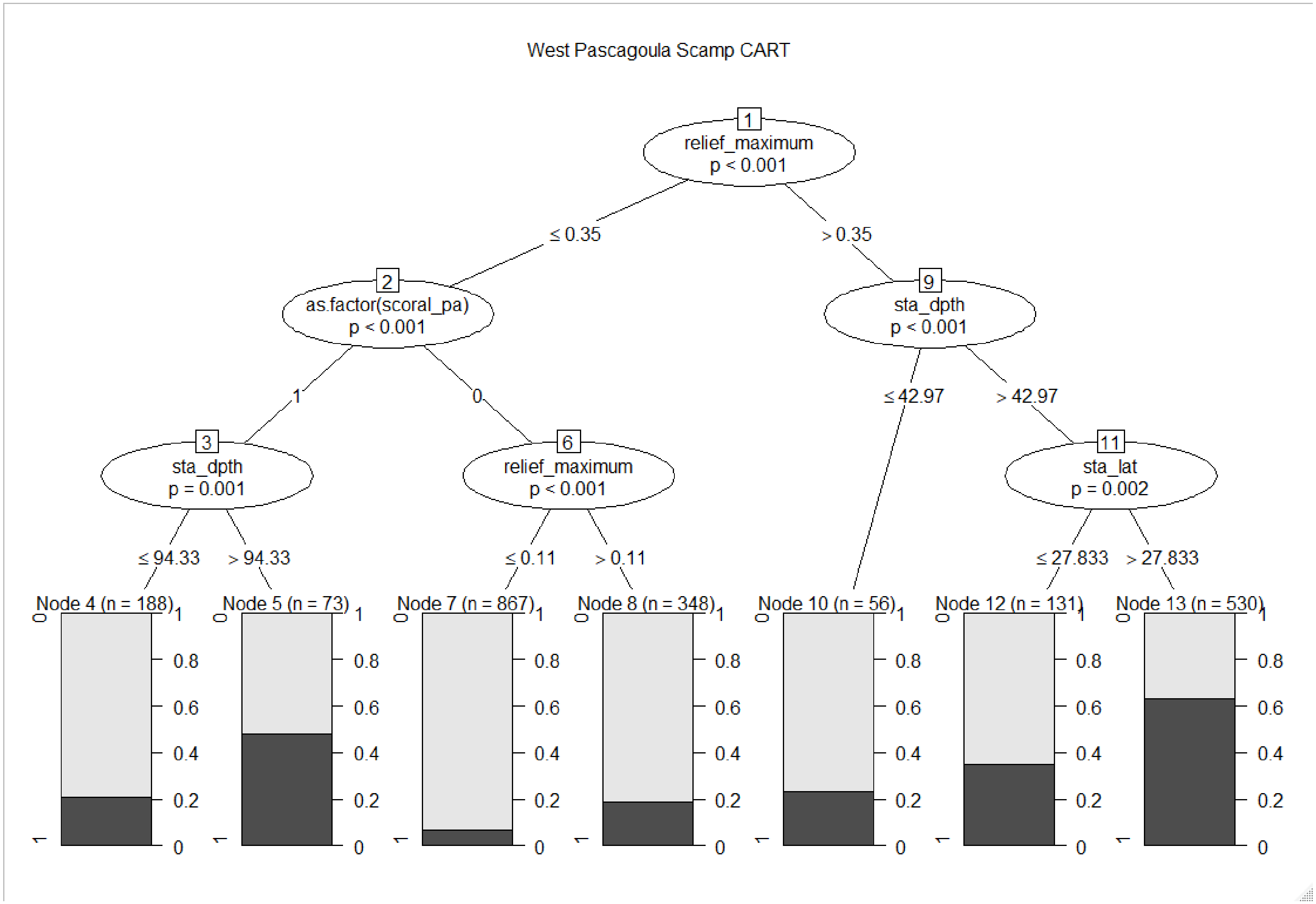

Figure 5. CART results for Scamp for Pascagoula video survey for the western Gulf region. Shaded portion of the plots indicate proportion of sites given by a node where Scamp were observed (26.5% of sites had Scamp present overall).

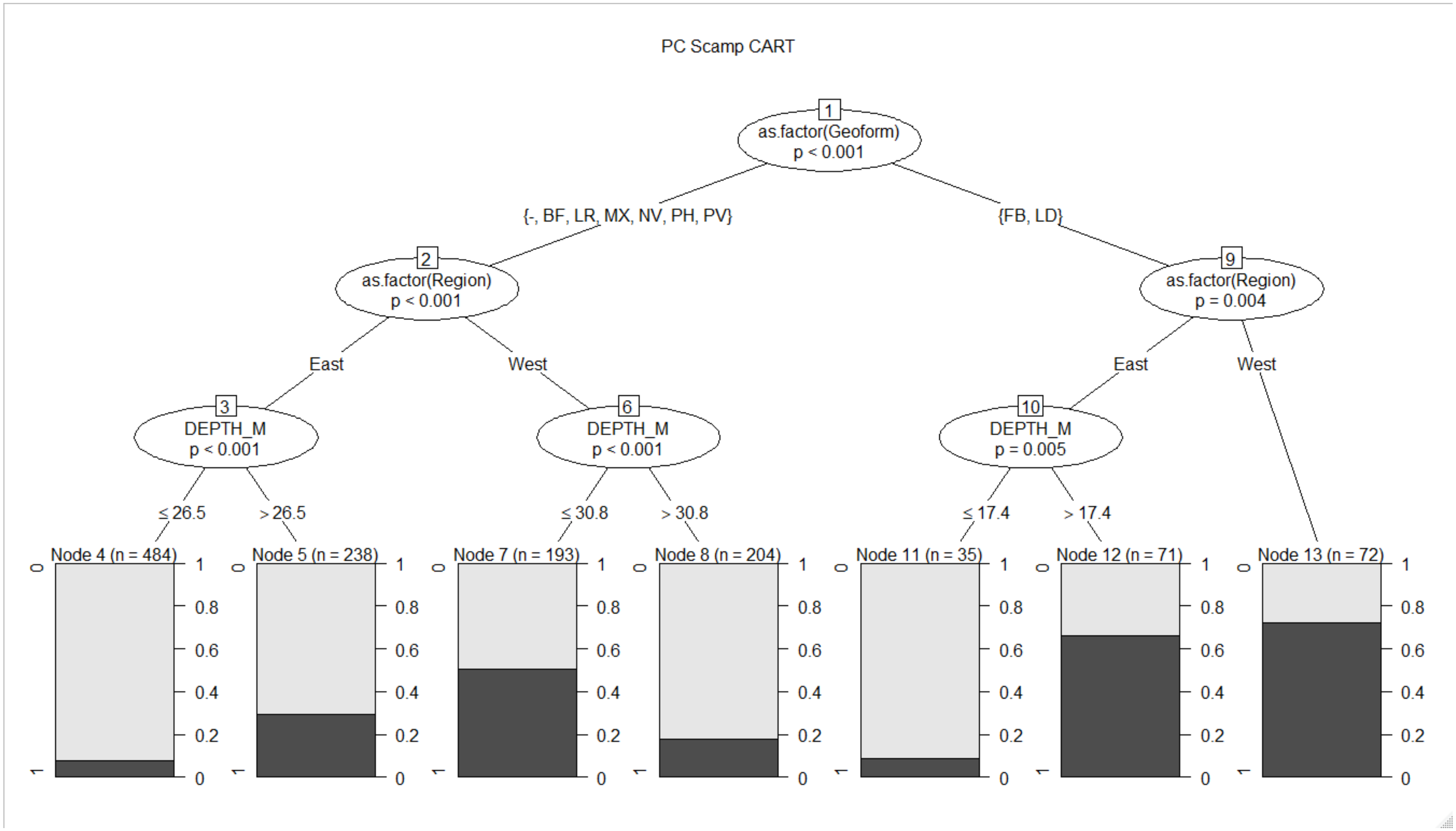

Figure 6. CART results for Scamp for Panama City's video survey. Shaded portion of the plots indicate proportion of sites given by a node where Scamp were observed (26.1% of sites had Scamp present overall)

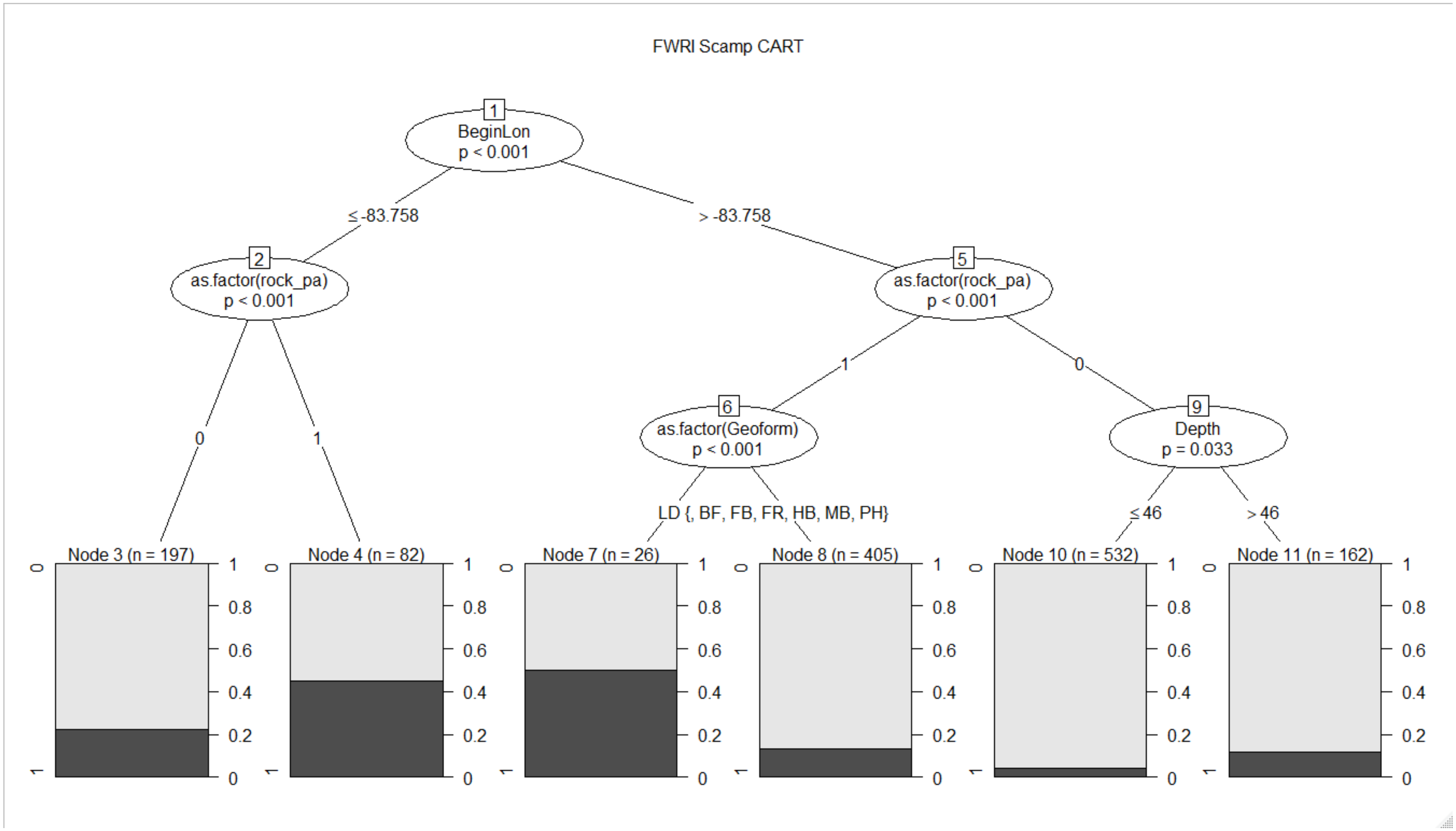

Figure 7. CART results for Scamp for FWRI's video survey. Shaded portion of the plots indicate proportion of sites given by a node where Scamp were observed (13.4% of sites had Scamp present overall).

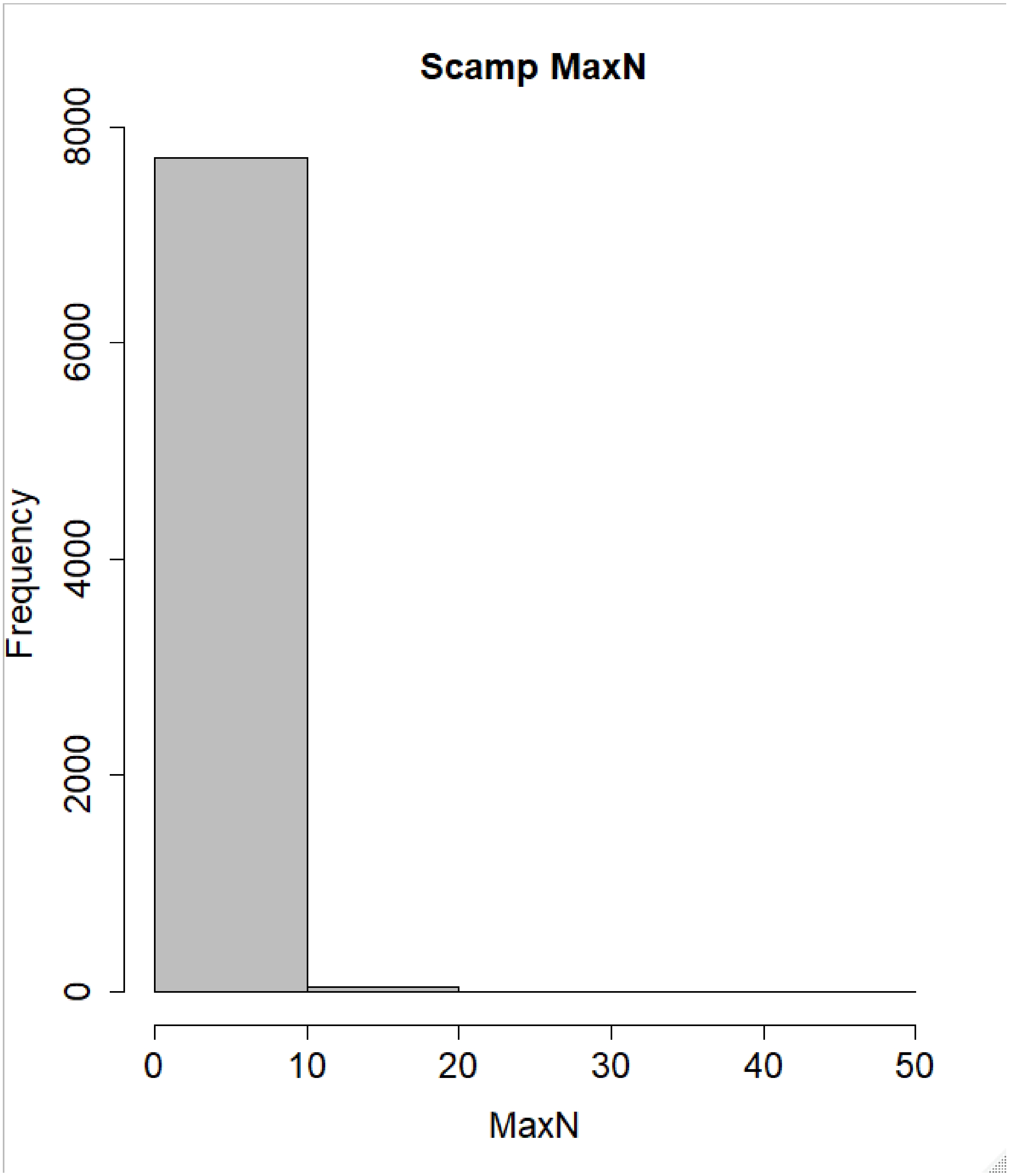

Figure 8. MaxN count distribution for Scamp observed in all four surveys used for the combined GOM index.

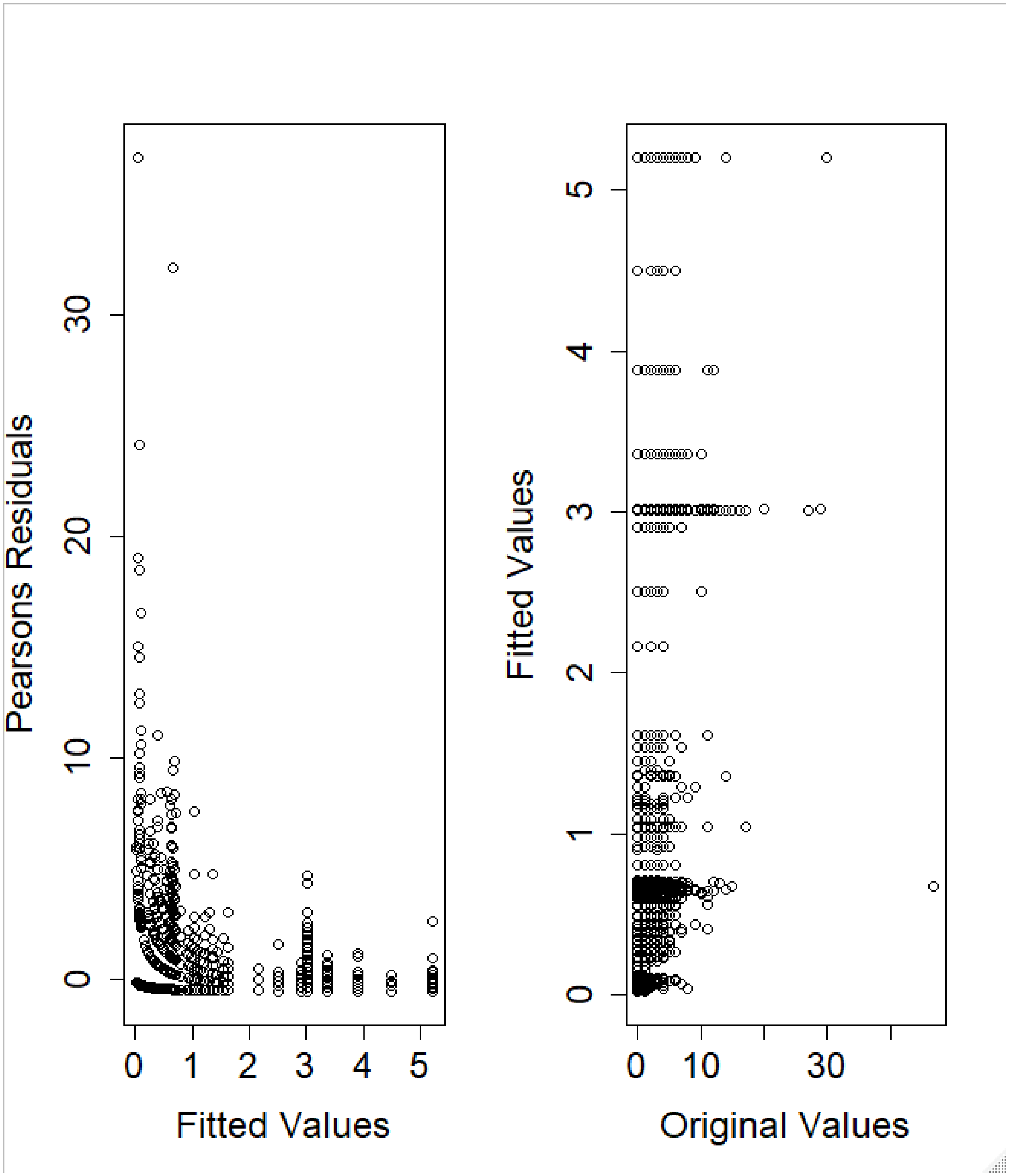

Figure 9. Model diagnostic plots showing fitted best model values against Pearson residuals (left panel) and fitted values plotted against original data values (right panel).

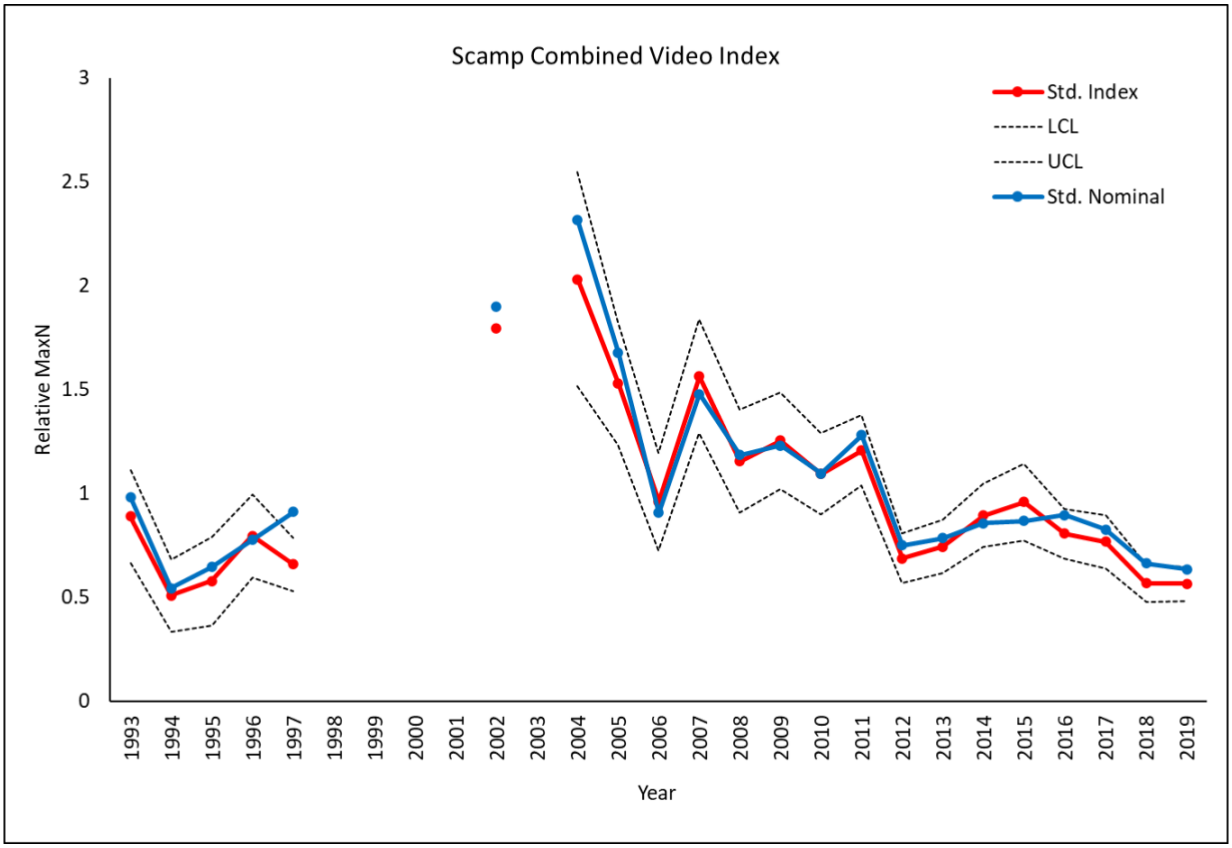

Figure 10. Standardized index (solid red line) with 2.5% and 97.5% confidence intervals (black dotted lines) and nominal index (solid blue line) for Scamp CPUE (MaxN) using the integrated West Florida Shelf video data.

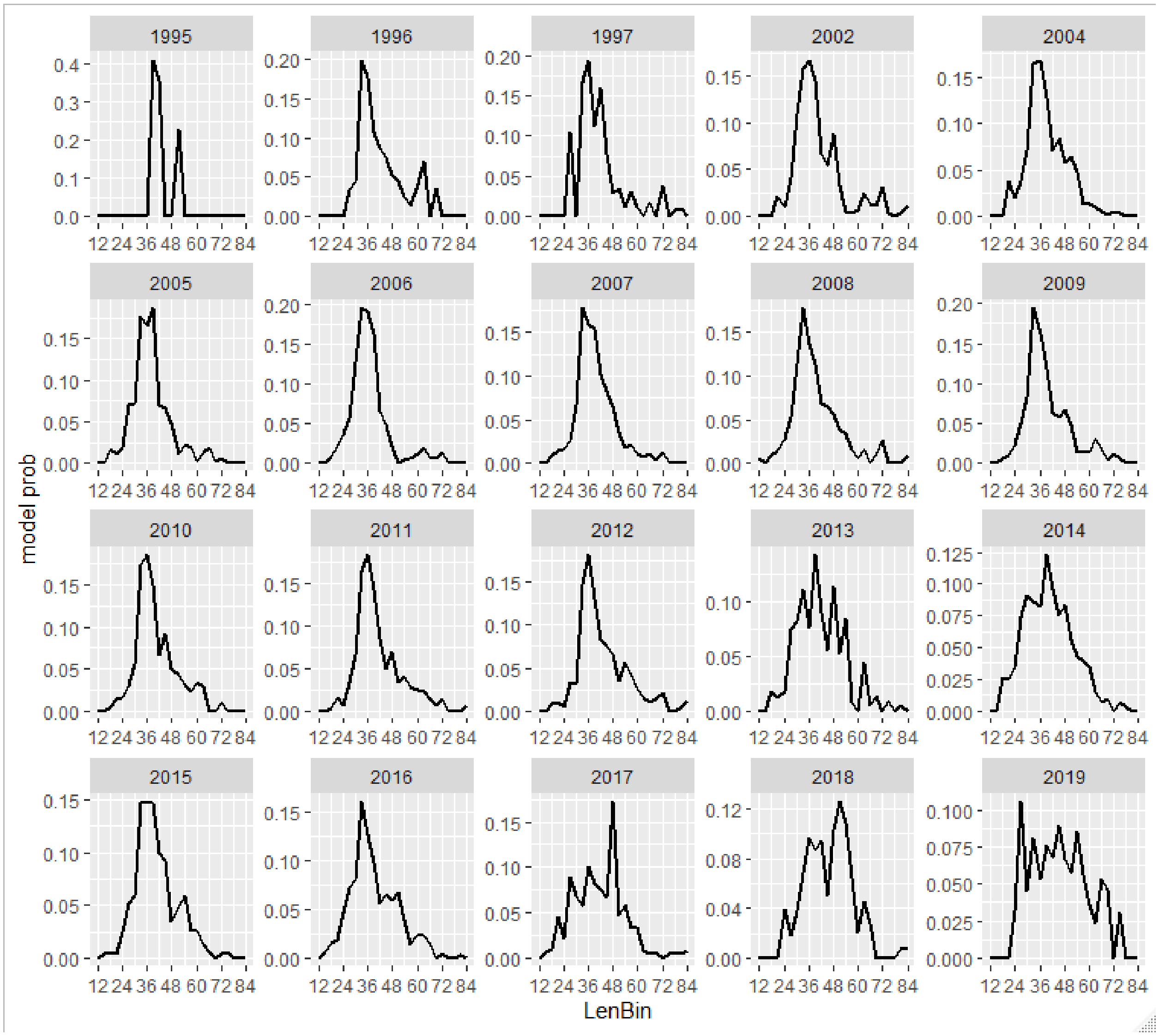

Figure 11. Final multinomial model fitted length composition data for the combined video index.