Combined indices of abundance for Red Grouper (*Epinephelus morio*) in the eastern Gulf of Mexico using data from three historic video surveys and unified G-FISHER program

Justin P. Lewis, Heather M. Christiansen, Theodore S. Switzer, Sean F. Keenan, Kate E. Overly, Matthew D. Campbell

SEDAR88-WP-16

21 March 2024

Please cite this document as:

Lewis, Justin P., Heather M. Christiansen, Theodore S. Switzer, Sean F. Keenan, Kate E. Overly, Matthew D. Campbell. 2024. Combined indices of abundance for Red Grouper (*Epinephelus morio*) in the eastern Gulf of Mexico using data from three historic video surveys and unified G-FISHER program. SEDAR88-WP-16. SEDAR, North Charleston, SC. 21 pp.

Combined indices of abundance for Red Grouper (*Epinephelus morio*) in the eastern Gulf of Mexico using data from three historic video surveys and unified G-FISHER program

Justin P. Lewis[1], Heather M. Christiansen[1], Theodore S. Switzer[1], Sean F. Keenan[1], Kate E. Overly[2], Matthew D. Campbell[3]

[1]Florida Fish and Wildlife Conservation Commission, Fish and Wildlife Research Institute, St. Petersburg, FL
[2]Southeast Fisheries Science Center, Panama City Laboratory, Panama City, FL
[3]Southeast Fisheries Science Center, Mississippi Laboratories, Pascagoula, MS

Introduction

Historically, three independent stationary video surveys were conducted in the northern Gulf of Mexico (GOM) to derive fishery-independent abundance estimates of important reef fish stocks. The longest running survey was the SEAMAP reef fish video (SRFV) survey initiated by the NMFS Mississippi Laboratory in 1992, followed in 2005 by the NMFS Panama City laboratory (PC) survey, and finally the Florida Fish and Wildlife Research Institute (FWRI) video survey, which started in 2010. Each survey used standardized sampling and data processing procedures. However, there remained subtle variations in video annotation protocols as well as survey design and spatial coverage (Figure 1) that presented obvious challenges from an assessment perspective. As such, a new survey initiative was undertaken, the Gulf Fishery Independent Survey of Habitat and Ecosystem Resources (G-FISHER) program, using funds provided by the NOAA RESTORE science program to integrate the three historic surveys under a single, unified design from 2020 onward (Figure 2).

The initial approach to integrate data from these independently conducted surveys into assessment models was to calculate individual indices for each survey. Alternatively, by combining indices across datasets, one would likely improve predictive performance by allowing for the largest possible sample sizes in model fitting. However, 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). We therefore used a habitat-based approach to combine relative abundance data for generating annual trends for Red Grouper (*Epinephelus morio*) throughout the eastern GOM (Thompson et al. 2022).

Methods

Historic survey designs:

The SRFV survey primarily targeted high-relief topographic features along the continental shelf from south Texas to south Florida. Site selection followed a stratified random design with strata determined by region and total proportion of reef area in a sampling block of 10’ latitude X 10’ longitude in size. Sites were selected at random from known reef areas identified through habitat mapping (multi-beam and side-scan sonar). Historic indices developed from the survey designate the Mississippi river delta as a geographic feature separating the west and east regions of the GOM (Campbell et al. 2017), with data from the western GOM being excluded from index calculations for Red Grouper.

The PC survey targeted 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. The survey area was divided into eastern and western sub-regions by Cape San Blas in the Florida Panhandle and further gridded into sampling blocks, which were 5' latitude X 5' longitude in size. Sites were randomly selected and proportionally allocated by region, sub-region, and depth (Gardner et al. 2017).

The FWRI survey initially focused on the regions offshore of Tampa Bay and Charlotte Harbor, FL (i.e., NMFS statistical zones 4 and 5) that were partitioned into inshore (10-36 m) and offshore (37-110 m) depth strata. The initial survey domain was later expanded in 2014 to include NMFS statistical zones 9 and 10 off the Florida Panhandle, and the offshore depth strata extended out to a depth of 180 m. This was followed by another spatial expansion in 2016 that included additional sites to cover the entirety of the West Florida Shelf from NMFS statistical zones 2-10 from depths of 10 to 180 m (Figure 1). Sites were first randomly selected and mapped using side scan sonar over a 2.1 km^2 area (Switzer et al 2020; Keenan et al 2022) and then video deployment sites were then randomly assigned proportionally across region and depth zones (Thompson et al. 2017).

G-FISHER survey design:

From 2020 onward, all video surveys were conducted under the G-FISHER program. A single set of sites were selected annually with sampling effort conducted using standardized G-FISHER gear and protocols. While the G-FISHER data set is generated by all three labs, it was treated as an extension of the FWRI survey for two reasons. First, spatial coverage in the eastern GOM is nearly identical to the 2016 expansion of FWRI's video survey. Second, the survey design and standardized protocols adopted by G-FISHER were largely modeled after the approaches of FWRI's survey; the most notable differences being the breadth of habitat video annotations, which are now more comprehensive and standardized among all three lab partners. Relative contribution of each survey by area and habitat observed is given in Table 1.

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, often referred to as MaxN or 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. rock, sponge, algae, and corals). While some categories were not historically recorded by all three labs (Campbell et al. 2017; Gardner et al. 2017; Thompson et al. 2017), the habitat annotation procedures adopted by G-FISHER are more comprehensive and include those habitat variables recorded during any of the three historic surveys.

Fish length measurements:

The methods used to obtain fish length information from video records have also evolved over time. Length measurements from the SRFV and PC surveys were initially estimated using parallel lasers attached to the camera system (Campbell et al. 2017; Gardner et al. 2017). However, these fixed mounted lasers resulted in very few usable laser contacts needed to obtain individual length measurement and higher probability of repeated measurements of the same individual. Therefore, both surveys adopted stereo-video methods (2008 and 2010 for SRFV and PC surveys respectively). From the onset, the FWRI survey used stereo-video methods to obtain length measurements. Length estimates from all three surveys were obtained from Vision Measurement System (VMS, Geometrics Inc.) through 2014. From 2015 to 2022, all length measurements were obtained from the SeaGIS software (SeaGIS Pty. Ltd.).

Data reduction:

For all surveys, video reads were excluded if they were unreadable due to turbidity or deployment errors. Data from the SRFV survey collected in 1992 were excluded from index calculations because of differences in counting methods in this first year, and no survey data are available for years 1998-2001 and 2003 (Table 2). Data from the remaining years from 1993 to 2019 were further restricted to the region east of the Mississippi delta because of potential demographic differences between Red Grouper in the western GOM. The entire spatial extent of the Panama City data was used from 2006 to 2019; data from 2005 was excluded because of an incomplete survey. Data from FWRI included nine years of data collection prior to G-FISHER (2010-2019) and the three years of data collection under the G-FISHER program (2020-2022). Data from all nine statistical zones were deemed sufficient for subsequent analyses. Final sample sizes by lab and year can be found in Table 2 and spatial coverage is shown in Figure 1 and Figure 2. Length measurements observed using stereo cameras were also compared to confirm that the three surveys have been sampling the same size and age fish (Figure 3), indicating that combining length data from these surveys is appropriate.

Index Construction

Habitat classification:

To produce a single index of abundance for Red Grouper using data from all three surveys, a common categorial habitat variable was generated for all three surveys that was based on the available habitat information from each survey. This was done so the final index model 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 habitats using a categorical regression tree (CART) approach independently for each survey. We selected this approach because it accounts for correlations among variables, can accommodate both continuous and categorical, and has clear utility to describe fish-habitat associations (De'Ath and Fabricus 2000; Yates et al. 2016).

For these initial analyses, Red Grouper MaxN at each site was converted to presence/absence and used as the response variable for defining each habitat class. Predictor variables included the habitat characteristics derived from video reads, which were reduced to presence/absence values, and the latitude, longitude, and depth of each site for all three survey datasets. Models using survey data from FWRI also included side-scan geoform 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 SRFV or PC survey data because side-scan sonar data was not available. However, a general habitat category variable derived from video reads was included in models using PC survey data.

We first used a random forest approach to reduce the number of potential variables to be selected in the final model for each lab's dataset by eliminating redundant or correlated variables from the initial suite of variable used to inform our habitat classification criteria. Using the full dataset from each survey, a random forest analysis fitted 2000 CART models to the data and then determined each variable's 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 Figure 4 for the FWRI survey dataset.

From each lab-specific random forest output, approximately 50% of the variables were selected based on their ranked importance for inclusion in the lab-specific CART models. The final model was created by fitting the presence of Red Grouper at a 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 lab-specific CART. Each terminal node was then assigned one of three habitat classifications (i.e., good, fair, or poor) based on the proportion of sites with positive Red Grouper catches. Terminal nodes criteria were used to define good habitat if the proportion positive was more than twice the overall proportion positive, poor habitat if the proportion positive was less than or equal to the overall average, and the criteria for any remaining terminal nodes used to define fair habitat. The lab-specific criteria associated with each habitat class was used to define a new variable ("Hab") for each site in their respective MaxN data sets for Red Grouper. All analyses were carried out using the partykit package (Hothorn and Zeileis 2015) in R version 4.3.1 (R Core Team 2023).

Index model fitting and diagnostics

The model used to standardize CPUE and provide an index of abundance was fit using a negative binomial with the formula:

$$\mathrm{MaxN} \sim \mathrm{Year} + \mathrm{Hab} + \mathrm{Lab}$$

where Hab is the CART derived habitat code and Lab 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 showed no discernible patterns of association between Pearson residuals and fitted values or the fitted values and the original data (Figure 8 and 9). An examination of residuals for the model parameters (Figure 9) showed no clear patterns of association, 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 Hab classes 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 Lab x Hab combination (e.g., up to 7 for years 2010-2019), providing a weighting factor for each of the mean estimates. Area weighting factors for each time period are provided in Table 1. Weighted index values were then standardized to the grand mean.

Results and Discussion:

Red Grouper were found at a relatively high proportion of sites in all three surveys and ranged from 0.28 for the SRFV survey to 0.35 for the PC survey. The variables used to assign each habitat class based on CART model outputs varied somewhat among surveys (Figure 5-7). Habitat classes for the SRFV survey were defined by the presence of soft coral, substrate max relief, depth, and geographic location (Figure 5). This was also the only CART model that included terminal nodes that met our proportion positive threshold to define good habitat. For this data set, the habitat of a sampling site was designated as good if it met the criteria of node 8 (i.e., no soft coral observed, max substrate relief > 0.33 m, and at longitudes > -84.747°) or node 12 (i.e., soft coral was observed, depth ≥ 50.7 m, and at latitudes > 24.693°). Poor habitat for the SRFV survey was assigned to sites that met the criteria defining node 4 (i.e., no soft coral observed and max substrate relief less than 0.05 m) and the remaining node criteria were used to define the sampled habitat as fair. The PC survey habitat classification was based on general habitat category (obtained from video reads), sampling month, site depth, rock presence/absence, and relief presences/absence (Figure 6). The habitat of a given sampling site was considered fair if it met the criteria defined by node 3 (general habitat category flat bottom, ledge, mixed, or potholes and sampled in May or June), 6 (general habitat category flat bottom, ledge, mixed, or potholes, sampled in July-November, and rock was observed), or 11 (general habitat category low relief or none visible, depth > 16.5 m, and relief present). The remaining node criteria were used to define fair habitat. In the case of the FWRI data set, the habitat at a given site was classified as fair if no sponge was observed, algae was observed, and site longitude was > -85.775 (node 8) or if sponge was observed and site longitude > -85.303 (nodes 12 and 13; Figure 7). The remaining criteria of the remaining terminal nodes were used to define poor habitat. The overall proportion of sites in each habitat category for each survey are shown in Table 3.

Annual standardized index values for Red Grouper in the Eastern Gulf of Mexico, including coefficients of variation, are presented in Table 4. The model CVs indicate a good fit to very good fit, with the highest CV values of ~20%-25% in the early years of the time series steadily declining as surveys are added to the lowest CV value of 6% in years 2019-2022. Abundance estimates of Red Grouper in the eastern GOM were relatively low at the beginning of the time series but showed a steady increase through 2005 (Figure 10). This was followed by a decline in abundance over the next two years, an increasing trend from 2007 to 2009, and a downward trend from 2009 to 2015. From 2015 onward, the abundance estimates have continued to trend upward though still below the peak estimated abundance observed in 2009 (Table 4; Figure 10).

References

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(2-3), pp.209-227.

Campbell, M.D., K. R. Rademacher, M. Hendon, P. Felts, B. Noble, R. Caillouet, J. Salisbury, and J. Moser. 2017. SEAMAP Reef Fish Video Survey: Relative Indices of Abundance of Grey Snapper. SEDAR51-DW-07. SEDAR, North Charleston, SC. 31 pp.

De'ath, G. and Fabricius, K.E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology, 81(11), pp.3178-3192.

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. and Zeileis, A., 2015. partykit: A modular toolkit for recursive partytioning in R. The Journal of Machine Learning Research, 16(1), pp.3905-3909.

Keenan, S.F., Switzer, T.S., Knapp, A., Weather, E.J. and Davis, J., 2022. Spatial dynamics of the quantity and diversity of natural and artificial hard bottom habitats in the eastern Gulf of Mexico. Continental Shelf Research, 233, p.104633.

R Core Team. 2023. 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., T.S. Switzer, and S.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, K.A., T.S. Switzer, M.C. Christman, S.F. Keenan, C.L. Gardner, K.E. Overly, and M.D. Campbell. 2022. A novel habitat-based approach for combining indices of abundance from multiple fishery-independent video surveys. Fisheries Research 247: 106178.

Yates, K.L., C. Mellin, M.J. Caley, B.T. Radford, and J.J. Meeuwig. 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. Walker, 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. Estimated amount of total reef habitat within each survey domain, and resultant survey-specific habitat weighting factors. These weights were multiplied by the total percentage of habitat class, within each survey (see Table 3) to define final habitat weights.

	Survey					
	SRFV (1993-2019)	PC (2006-2019)	FWRI (2010-2013)	FWRI (2014-2015)	FWRI (2016-2019)	GFISHER (2020-2022)
Total universe area (km^2)	31247	22105	46286	58970	144403	144403
Area x proportion of mapped with reef	23977	14861	10161	11463	27939	27939
Time-specific habitat weights						
1993-2005	1					
2006-2009	0.62	0.38				
2010-2013	0.51	0.32	0.17			
2014-2015	0.48	0.30		0.23		
2016-2019	0.36	0.22			0.42	
2020-2022						1

Table 2. Annual sample sizes for each survey data sets used to estimate Red Grouper abundance. No survey data are available for 1998-2001 and 2003. Data collected from 2020 to 2022 under the G-FISHER program were treated as an extension of the FWRI time series and labeled as such in this table.

Year	SRFV	PC	FWRI	Total
1993	115			115
1994	90			90
1995	60			60
1996	133			133
1997	162			162
1998				
1999				
2000				
2001				
2002	152			152
2003				
2004	148			148
2005	274			274
2006	277	89		366
2007	319	53		372
2008	206	83		289
2009	262	105		367
2010	221	134	155	510
2011	337	157	222	716
2012	280	144	236	660
2013	163	84	185	432
2014	230	158	348	736
2015	152	151	388	691
2016	205	142	729	1076
2017	221	145	628	994
2018	212	83	696	991
2019	277	84	907	1268
2020			760	760
2021			991	991
2022			936	936
Total	4496	1612	7181	13289

Table 3. Proportion of sites assigned to each habitat category (Good, Fair, or Poor) as determined by lab-specific categorical regression trees (CARTs) for Red Grouper presence/absence.

Lab/Survey	Good (G)	Fair (F)	Poor (P)
SRFV	0.16	0.57	0.27
PC		0.84	0.16
FWRI		0.64	0.36

Table 4. Total number of sites sampled (N), proportion of sites with Red Grouper present, nominal CPUE, standardized index of abundance, and index CVs for each year of the combined survey data set. Both nominal CPUE and index are standardized to their respective overall mean.

Year	N	Prop Positive	Std. Nominal	Std. Index	CV
1993	115	0.25	0.80	0.72	0.20
1994	90	0.30	0.78	0.64	0.20
1995	60	0.32	0.87	0.54	0.25
1996	133	0.29	0.84	0.75	0.16
1997	162	0.38	1.14	0.97	0.12
1998					
1999					
2000					
2001					
2002	152	0.36	1.15	0.95	0.13
2003					
2004	148	0.41	1.42	1.24	0.12
2005	274	0.35	1.22	1.32	0.09
2006	366	0.31	0.93	1.13	0.10
2007	372	0.20	0.66	0.76	0.12
2008	289	0.30	0.92	1.13	0.10
2009	367	0.40	1.32	1.59	0.08
2010	510	0.33	1.05	1.15	0.08
2011	716	0.41	1.36	1.39	0.06
2012	660	0.34	1.15	1.17	0.07
2013	432	0.34	1.18	1.06	0.09
2014	736	0.27	0.89	0.79	0.08
2015	691	0.22	0.64	0.59	0.10
2016	1076	0.26	0.91	0.82	0.06
2017	994	0.29	0.94	0.92	0.06
2018	991	0.22	0.75	0.97	0.08
2019	1268	0.24	0.81	1.09	0.06
2020	760	0.34	1.12	1.05	0.06
2021	991	0.33	1.11	1.14	0.06
2022	936	0.31	1.04	1.12	0.06

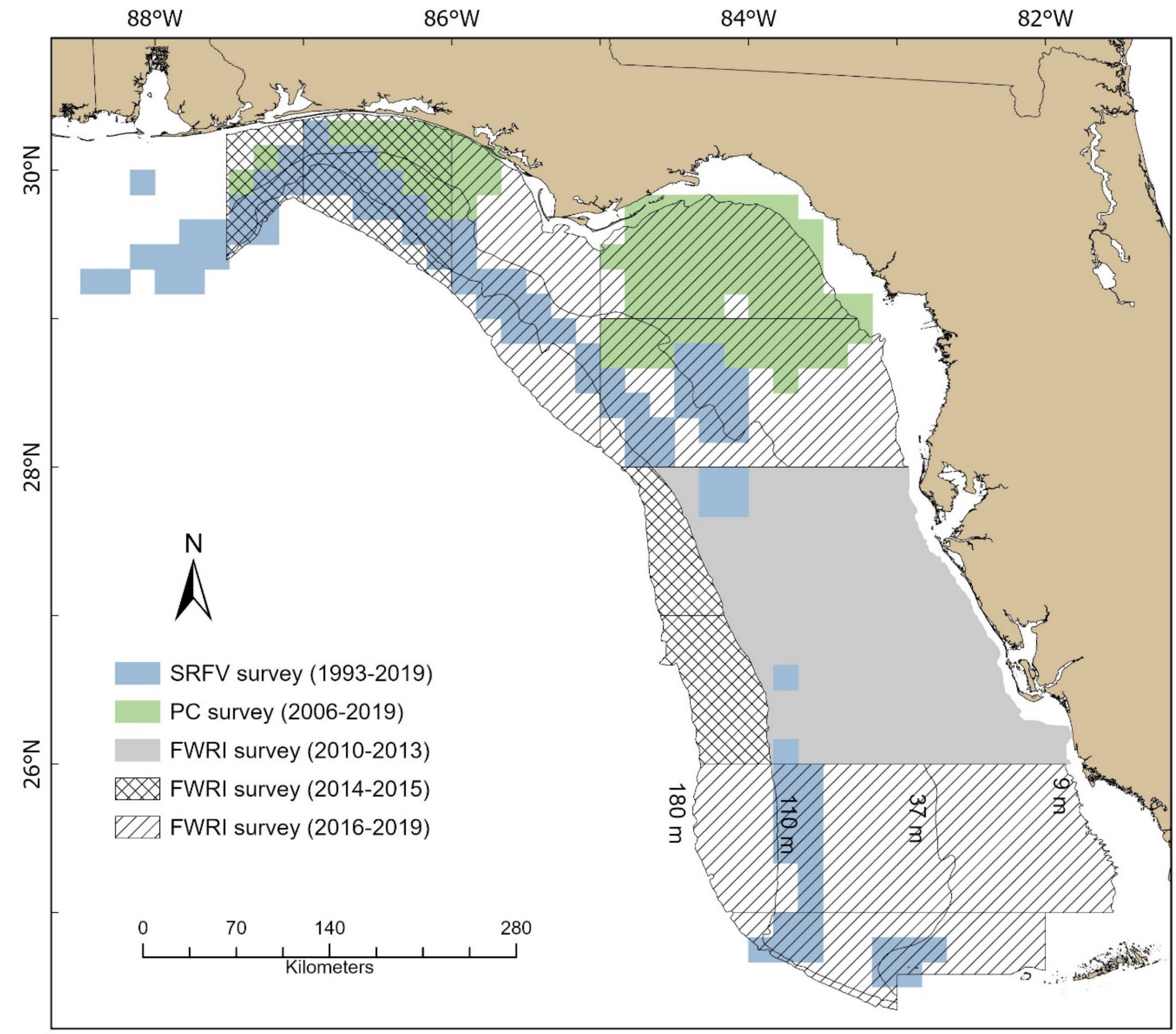

Figure 1. Spatial coverage of three historic reef fish video surveys in the eastern Gulf of Mexico from 1992 to 2019. All changes in spatial coverage of the FWRI survey represent spatial expansions that include all areas previously sampled.

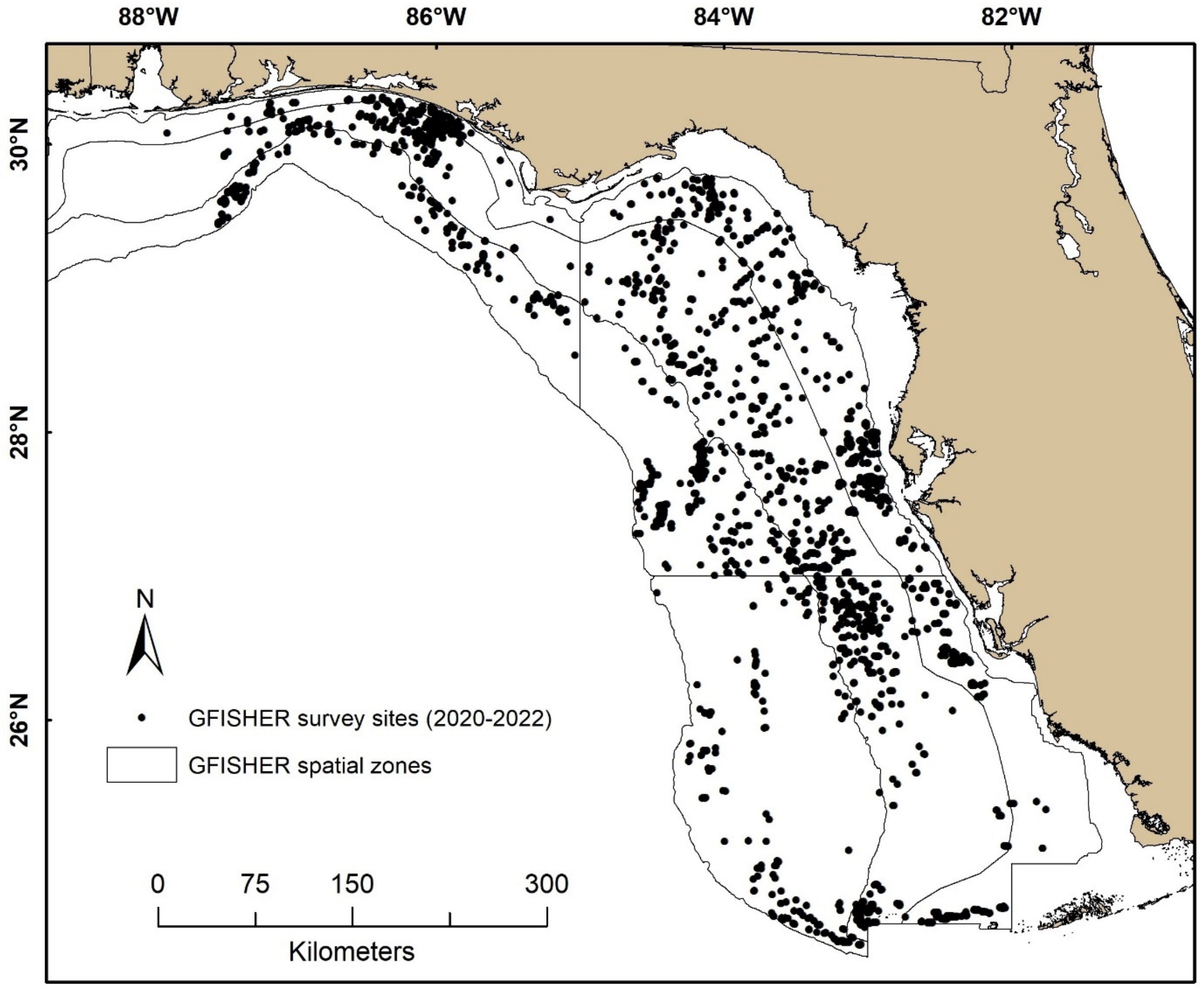

Figure 2. Distribution of sampling effort in the eastern Gulf of Mexico under the G-FISHER program from 2020 to 2022.

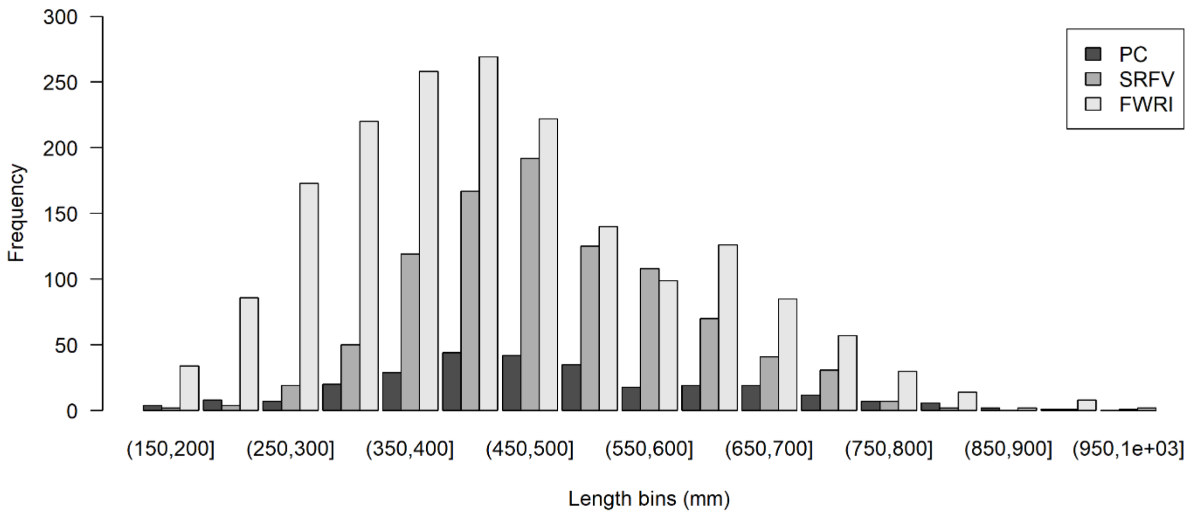

Figure 3. Length frequencies of Red Grouper observed on video from the three surveys using VMS and SeaGIS.

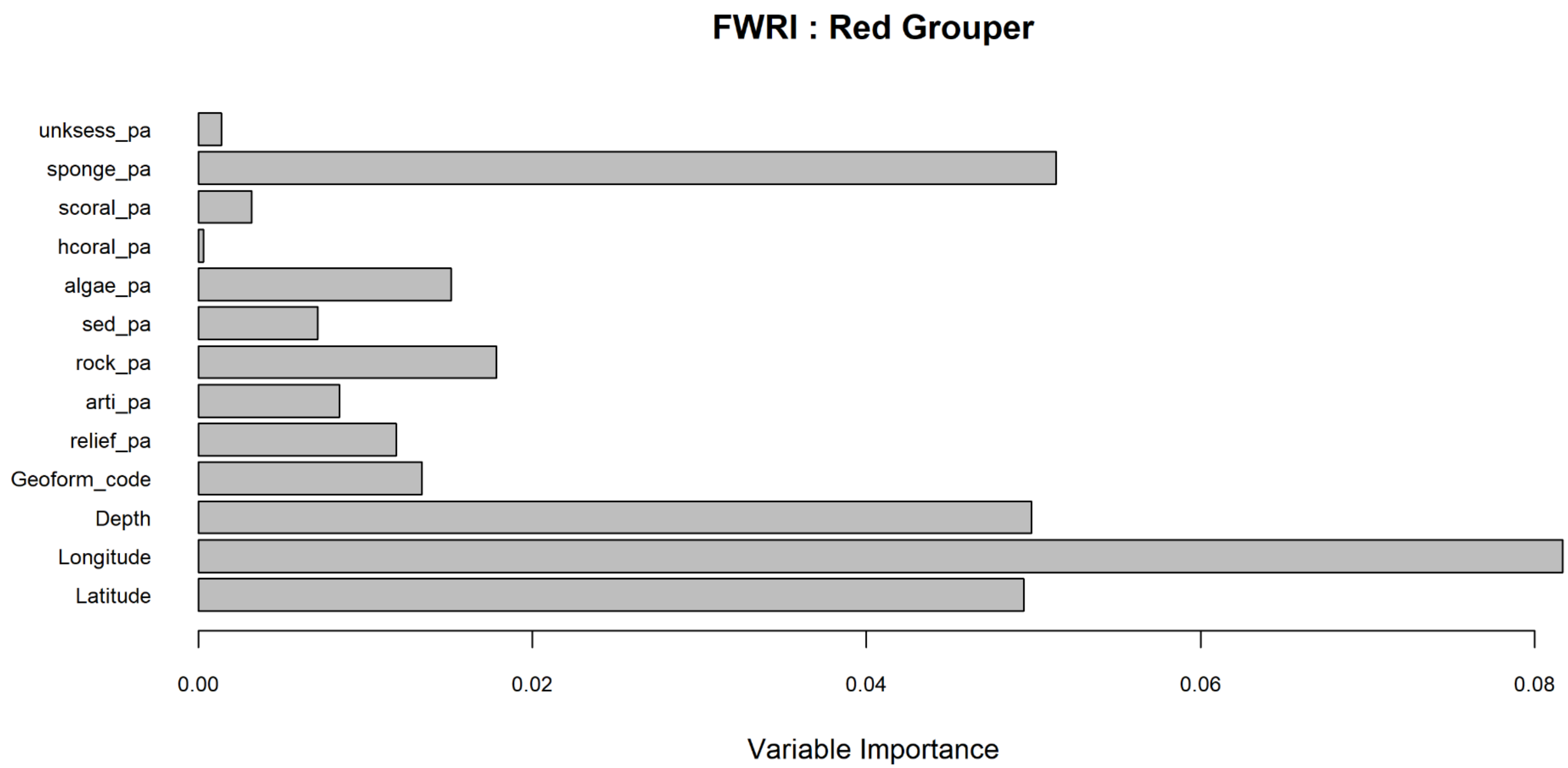

Figure 4. Random Forest generated variable importance for Red Grouper presence using FWRI survey data (2010-2019) and G-FISHER data (2020-2022).

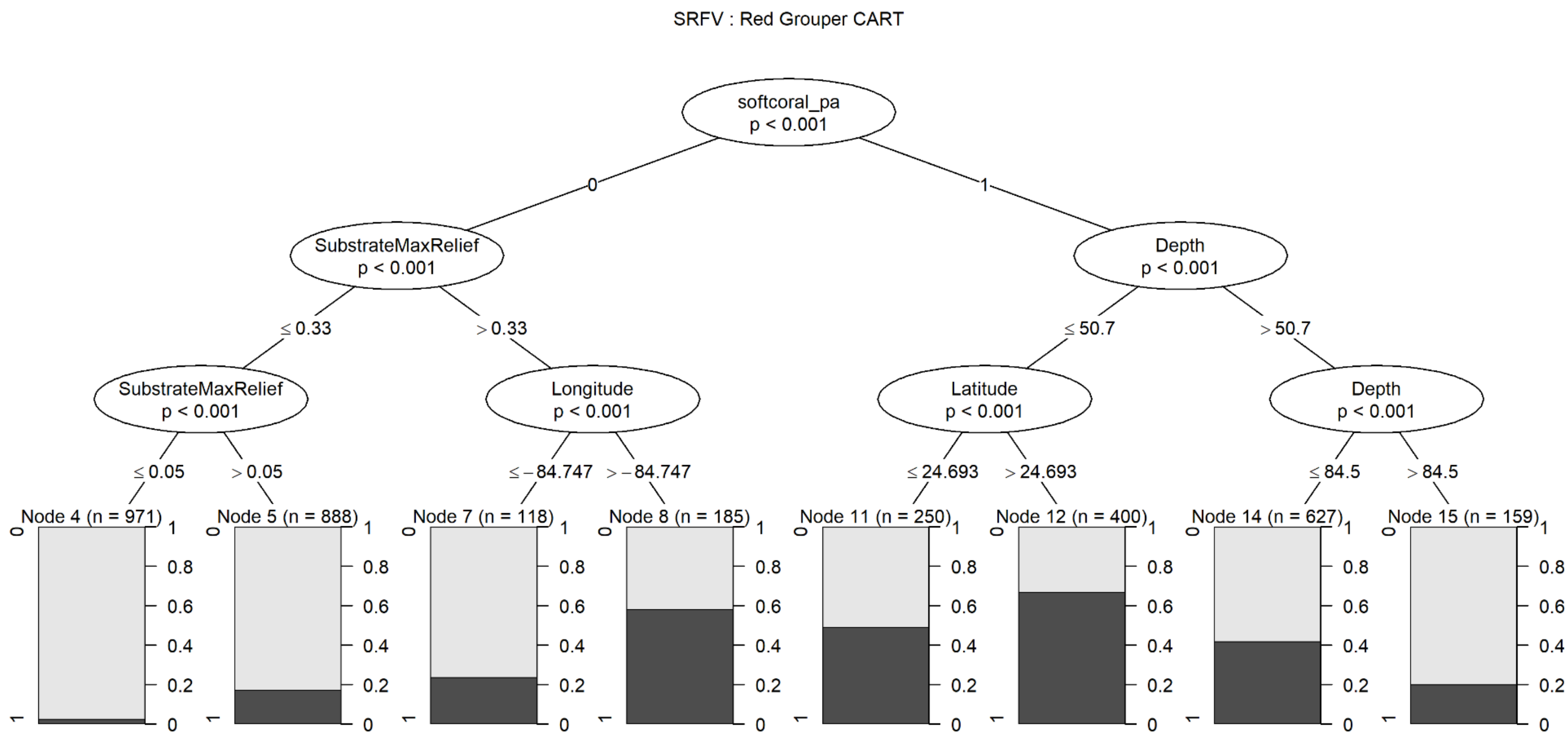

Figure 5. CART results for Red Grouper for the SEAMAP reef fish video (SRFV) survey conducted by the NMFS lab in Pascagoula, MS. Shaded portion of the plots indicate proportion of sites given by a node where Red Grouper were observed. Overall proportion positive = 0.28, misclassification rate = 0.24.

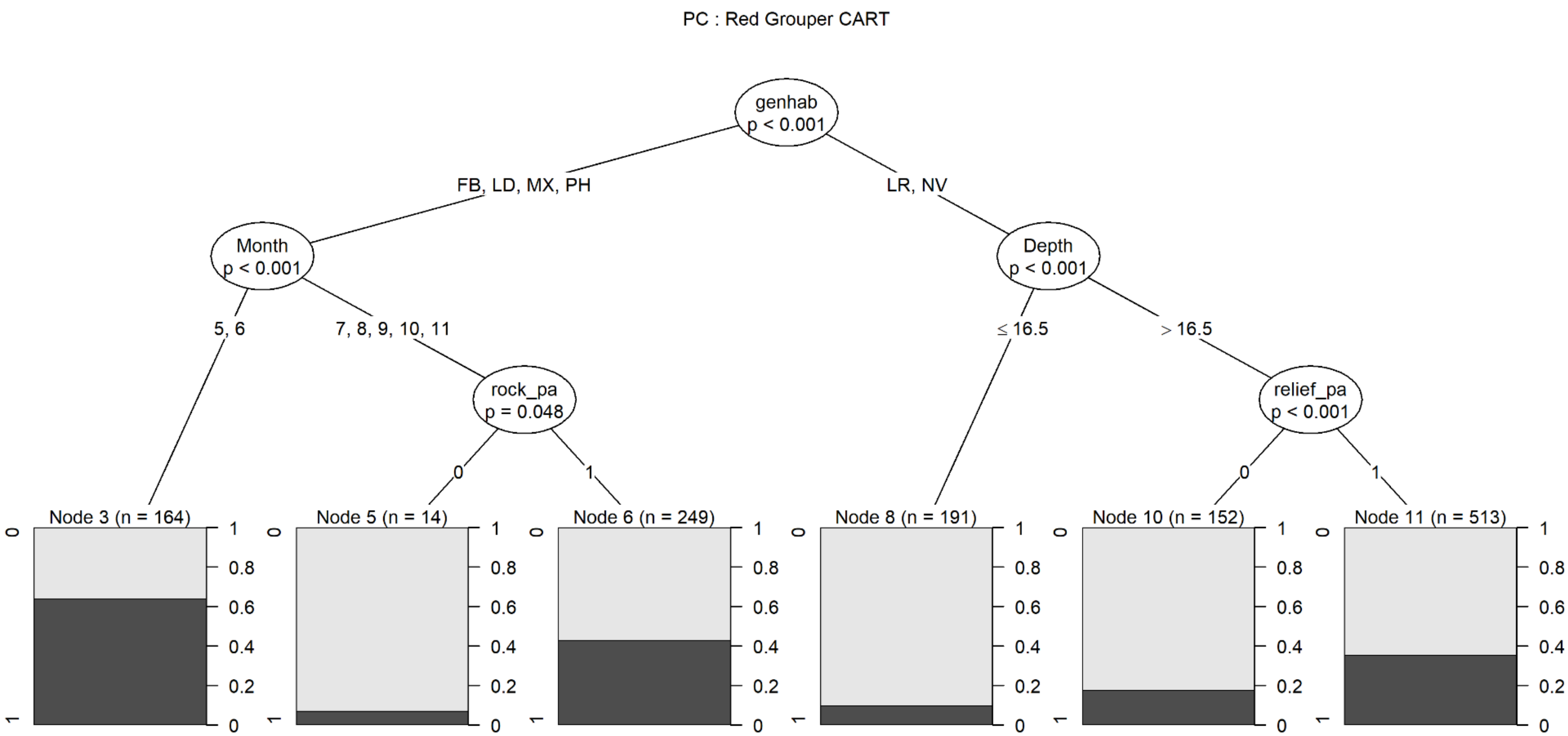

Figure 6. CART results for Red Grouper for Panama City's video survey. Shaded portion of the plots indicate proportion of sites given by a node where Red Grouper were observed. Overall proportion positive = 0.35, misclassification rate = 0.36.

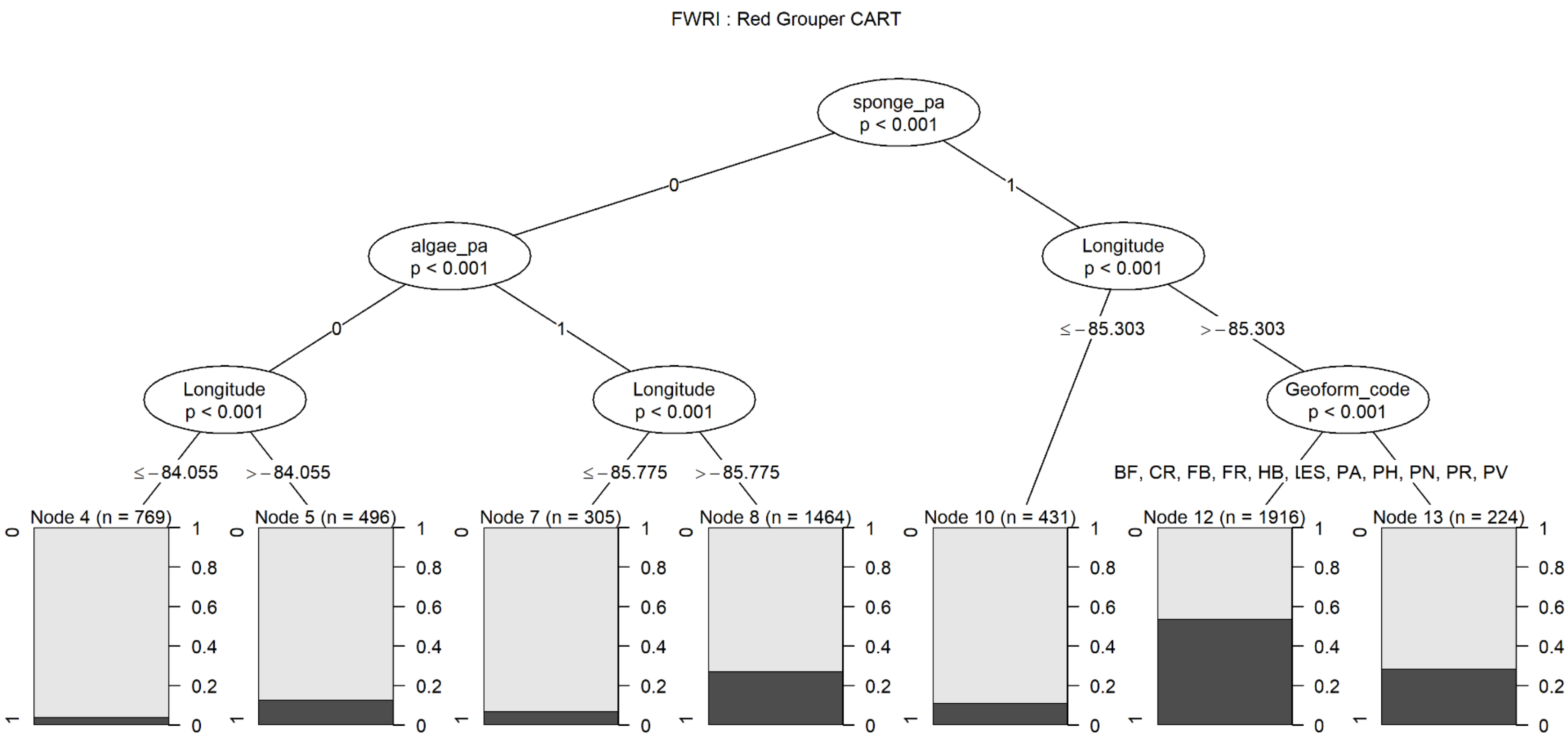

Figure 7. CART results for Red Grouper for FWRI's video survey. Shaded portion of the plots indicate proportion of sites given by a node where Red Grouper were observed. Overall proportion positive = 0.30, misclassification rate = 0.30.

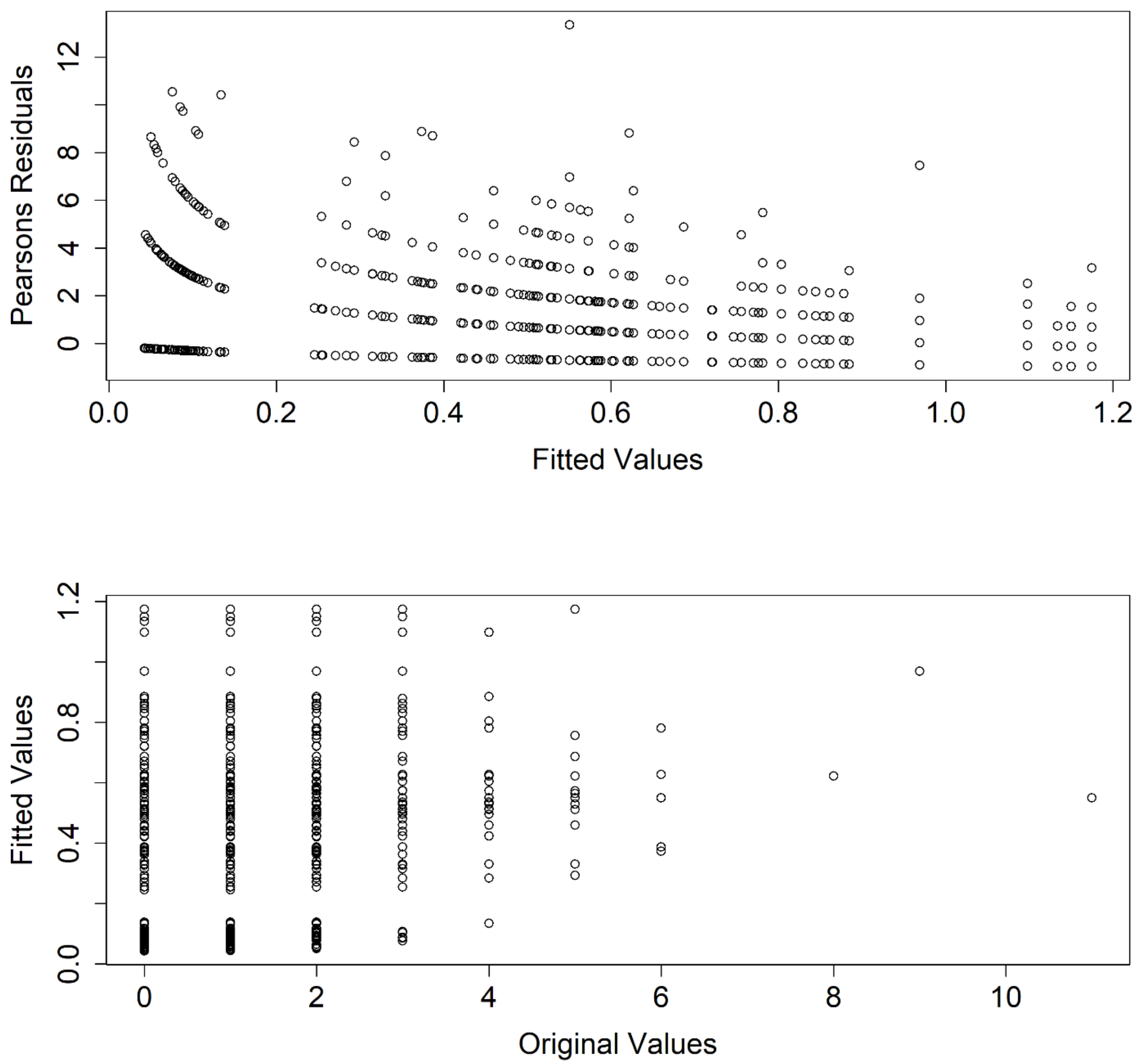

Figure 8. Model diagnostic plots showing fitted best model values against Pearson residuals (top panel) and fitted values plotted against original data values (bottom panel).

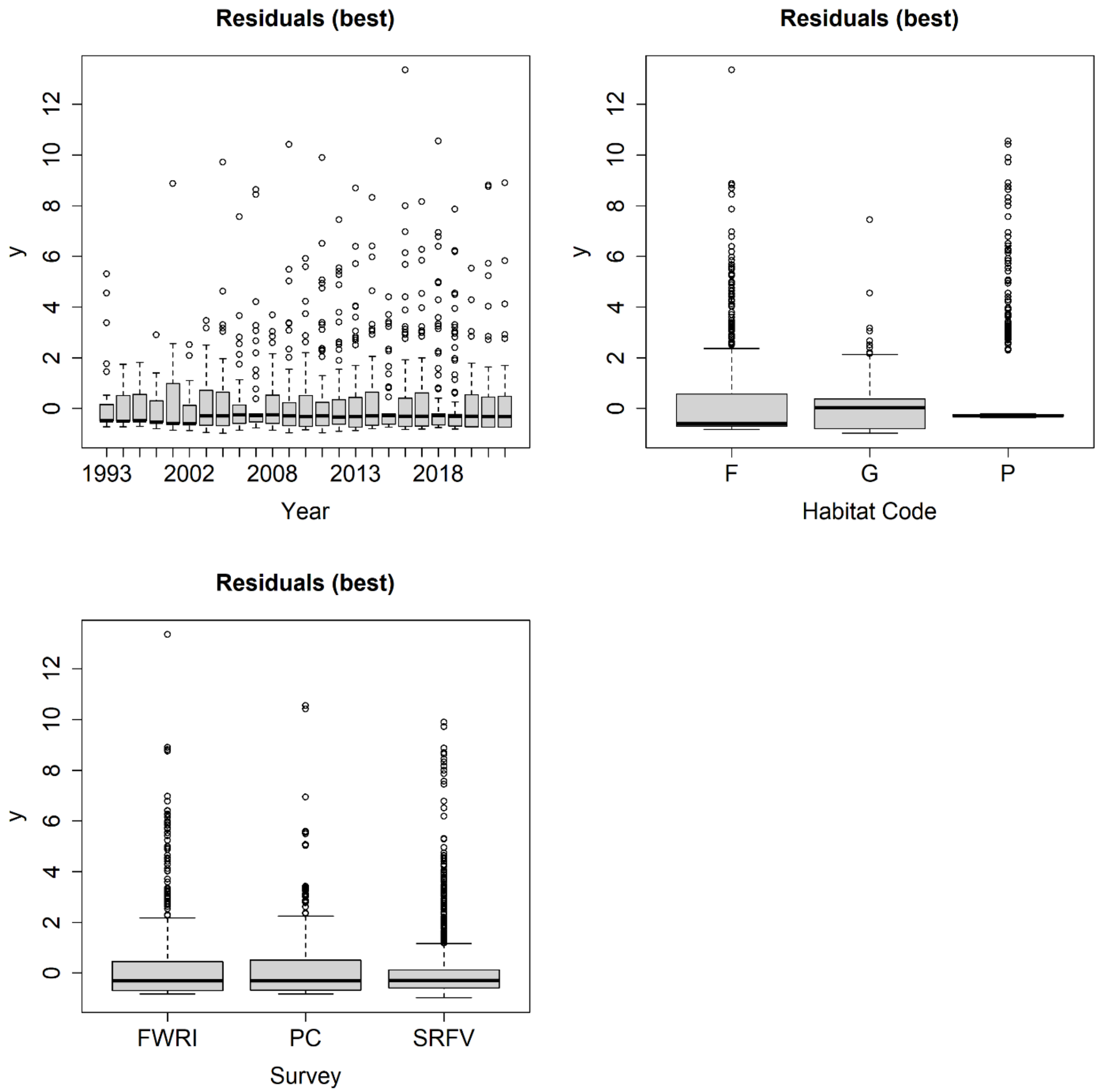

Figure 9. Model diagnostic plots showing Pearson residuals for the final (best) model plotted against model parameters.

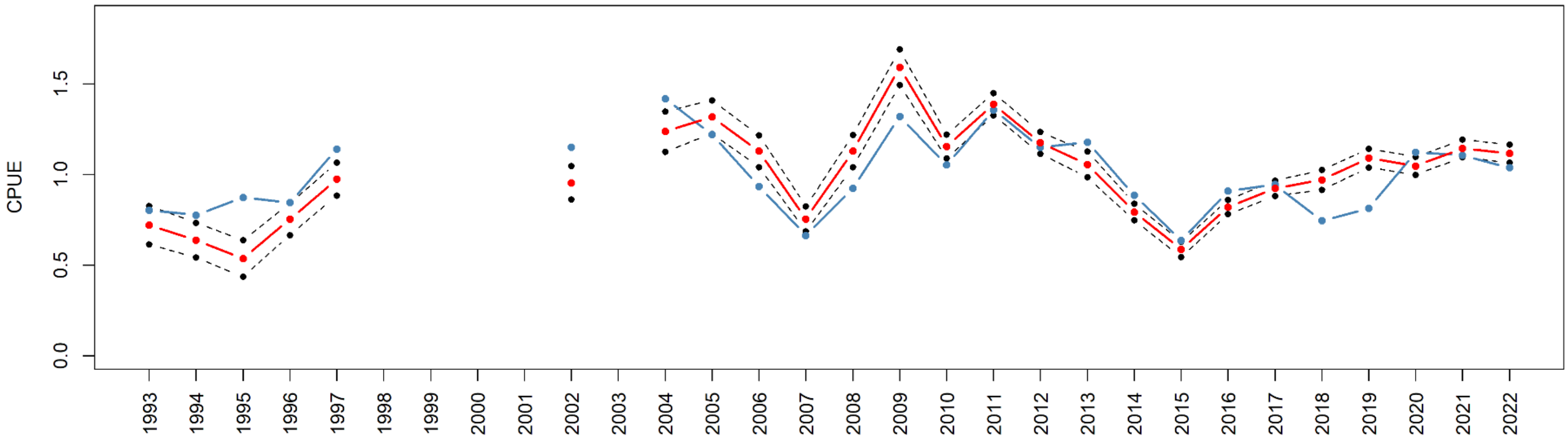

Figure 10. Standardized index of abundance (solid red line) with 2.5% and 97.5% confidence intervals (black dotted lines) and nominal CPUE (solid blue line) for Red Grouper in the eastern Gulf of Mexico.