Standardized index of Gulf of Mexico cobia (*Rachycentron canadum*) from headboat data

Eric Fitzpatrick

2019 SEDAR28-WP-05

1 June 2020

Please cite this document as:

Fitzpatrick, E. 2020. Standardized index of Gulf of Mexico cobia (*Rachycentron canadum*) from headboat data. 2019-S28Update-WP-05. SEDAR, North Charleston, SC. 27 pp.

Standardized index of Gulf of Mexico cobia (*Rachycentron canadum*) from headboat data.

Sustainable Fisheries Branch, National Marine Fisheries Service,
Southeast Fisheries Science Center,
101 Pivers Island Rd, Beaufort, NC 28516

Eric Fitzpatrick

February 2020

Abstract

A standardized index of abundance was developed using cobia count data from the Southeast headboat survey trip records (logbooks) for 1986-2018. The analysis included areas from north Florida through Texas (GoM). The index is meant to describe population trends of fish in the size/age range of fish landed by headboat vessels. Data filtering and subsetting steps were applied to the data to model trips that were likely to have directed cobia effort.

Background

The headboat fishery in the Gulf of Mexico includes for-hire vessels. The fishery uses hook and line gear, generally targets hard bottom reefs as the fishing grounds, and generally targets multiple species in the snapper-grouper complex. One of the key characteristics defining a headboat from other recreational fishing such as charter boats is the number of anglers.

Headboats in the South Atlantic and Gulf of Mexico are sampled from North Carolina to Texas. Data have been collected since 1972, but logbook reporting did not start until 1973. In addition, only North Carolina and South Carolina were included in the earlier years of the data set. In 1976, data were collected from North Carolina, South Carolina and northern Florida. Full coverage began in 1978 with data collected from southern Florida while Georgia didn't have active vessels in the fishery until 1993. Sampling in the Gulf of Mexico began in 1986 (Figure 1). Variables reported in the data set include year, month, day, area, location, trip type, number of anglers, species, catch, and vessel id. Biological data and discard data were recorded for some trips in some years.

A bag limit of 2 cobia/person/day has been in place since August 20 1990 in federal waters and most recently (February 2018) a bag limit of 1 cobia/person/day has been in place in Florida waters.

Issues

Cobia are a semi-large coastal migratory pelagic species and thus not ideally targeted by snapper-grouper bottom fishing gear.

However, not all anglers on a headboat use typical bottom fishing gear (e.g. the surface line fishers off the stern). So, one unknown is what proportion of cobia caught on headboats are caught on bottom rigs versus surface line rigs? If the predominant cobia interaction is on surface line rigs, then how often are surface lines deployed on headboat trips? We have no data to examine this gear issue.

A necessary assumption is that the probability of encountering a cobia on a headboat is fairly constant with respect to surface line use. Let's ignore that issue and assume that the proportion of surface lines deployed is constant on all headboat trips. We know that cobia are a seasonal migrant in the South Atlantic (potentially less of an issue in the Gulf of Mexico) and thus we expect the encounter rate with headboats to change with seasons. We also know that cobia migrate spatially, so that the area is also a contributing factor to determining the probability of catching a cobia on a headboat. Ideally, we can use the history of cobia catches on all headboats to determine the time and space strata where cobia have a reasonable chance of being caught. However, there is an underlying assumption that you have a period of time in the data where cobia were abundant enough to fully fill out the time and space strata (i.e. identify the time/space strata). With low catch rates of cobia and the history of fishing for the species, this assumption may be violated.

On any given headboat trip the captain makes a choice once they leave the dock as to where the boat is going to fish. Catching cobia is unlikely to factor into that decision, rather the choice is going to be more likely based on maximizing the catch of bottom fishes. So, another assumption is that within the space/time strata where cobia are caught the captain's location selection has the same probability of capturing cobia. This may depend on the site variability within a captains fishing area.

Another issue with cobia is they are a thigmotaxic species that are drawn to structure. To what degree is the catch rate of cobia on a headboat affected by this behavior?

Identifying a period of time (seasonally and temporally) where these issues are relatively stable would be necessary in the development of an index of abundance for cobia.

Exploratory Data Analysis

Headboat records were examined to determine if sufficient data exists to develop a standardized index of abundance for south Atlantic cobia.

Positive cobia trips represent a small fraction of the overall composition in the south Atlantic headboat fishery (~1-7% annually) (Table 1). Since 1986, an average 1,986 cobia were captured per year in the Gulf of Mexico headboat fishery (FL-TX). Data filtering steps were applied to identify trips that likely had directed cobia effort.

<u>*Data Exclusions*</u>

1. Outlier removal

Extreme values occur more frequently in self-reported data because there are limited methods for validating data. Recent SEDAR stock assessments have removed values at the extreme upper tail of distribution for cpue and associated fields for self-reported fishery-dependent data. For snapper grouper, typically the largest 0.5% are excluded but due cobia being a rare event species that might be recorded more accurately the largest 0.1% values were excluded. The number of anglers on a trip can also influence cpue when calculated as fish/angler-hour. Trips with the largest 0.5% values for reported anglers were removed. Figure 2 shows the excluded trips based on outlier definitions by region. Removing a small percentage of the trips with extreme values is an unbiased method to correct for potential errors in self-reported data.

2. Cutoff for number of trips per vessel and number of anglers

Logbooks submitted by vessels that participated infrequently in the fishery are likely to be less accurate and may add noise to the data. Even if a vessel fished infrequently for one year, the number of trips should be greater than 30. We removed vessels that had fewer than 30 trips in the logbook database. It is rare for a headboat to fish with few anglers. There is anecdotal information that headboats would sometimes fish with just the crew and that logbooks for these trips were submitted. Experienced crew are likely to be more efficient at catching fish than paying customers. Captains may also limit distance to reduce fuel costs for trips with few paying customers. Trips with 6 or fewer anglers were excluded.

3. Core Vessels

To identify headboat trips that best characterize the cobia fishery, vessels that consistently caught cobia were selected (82 headboats representing 90% (prior to any filtering) of cobia effort and landings). This analysis was split up into two regions (Gulf of Mexico to the Florida Keys and the east coast of Florida) due to such a large area being considered in the index to make sure regions were not excluded from the final analysis. Cobia trips from these ‘core’ vessels increased from 7% (all data) to 17% (full and half day trips, model input) (Table 1). Prior to the analysis, proportion positive averaged near 7% from 1986-2018. Additional subsetting methods were explored (identify vessels using a proportion positive cutoff (Figure 3)) but led to a reduction in positive trips and convergence issues since this approach identified vessels that had fewer overall trips but more cobia interactions. Selecting data using a core group of vessels while removing vessels that inconsistently or never reported cobia more appropriately reflects directed cobia effort in the headboat fishery.

4. Starting year

Figure 4 illustrates when the ‘core’ headboats entered the fleet from the late 1980s to 2018. The top portion of the figure includes the vessels from the Gulf of Mexico while the lower portion of the figure includes the vessels from the east coast of Florida. The percentage of ‘core’ vessels reporting cobia in the fishery ranges between 40% and 60% (Figure 5). Figure 6 shows consistency in cobia cpue over time from the 1980s to the more recent years. Unlike SEDAR 58 in the South Atlantic where desirability may have shifted in the late 1980s, the relative consistency in cpue through time suggests desirability was stable but may need to be explored further during the next research track assessment for this species.

1986 was chosen as the start year for the Gulf of Mexico cobia headboat index.

5. Terminal year - spawning closure exclusion

2018 was chosen as the terminal year.

6. Trip types

The relatively few multi-day trips and the 3/4-day trips were filtered. Figure 7 shows the variability associated with these trip types by region. Trips by region for full and half day trips are presented in Figure 8.

Evaluation of explanatory variables

YEAR - Year was necessarily included, as standardized catch rates by year are the desired outcome. Years modeled were 1986-2018.

TRIP TYPE (t) – Full and half day trips were included in the standardization.

SEASON (s)– All months were included due to consistent cobia trips throughout the year. Four levels for season are in the model. The seasonal pattern in cpue across months seems consistent across regions (Figure 9).

REGION (r)- Three regions modeled include (east coast FL (eFL), west coast FL (including Alabama and Mississippi (very few positive trips)) (wFL), Texas to Louisiana (TXLA) (Figure 1). After filtering, a reduction in positive trips from the wFL region was noted but included in the final analysis in order to model the total headboat effort throughout the range. This may need further exploration during the next research track assessment for cobia but is unlikely to make a difference in the final index due to the low cobia catches.

VESSEL SIZE (v) - A factor was created for the vessel size using the quartiles of the maximum number of anglers across all trips as breaks for the factors. The proxy for vessel size is the maximum anglers reported over all trips for a vessel (Figure 10). Due to limited data and convergence issues, vessel size was modified to two levels: ‘small’ or ‘large’.

PERCENT FULL (pf)
The number of anglers reported for a trip was divided by the maximum number of anglers for a vessel to obtain an estimate of crowding and subsequently a proxy for a trips ability to either ‘target’ cobia that are visible or land cobia successfully depending on the boats capacity. This was then divided into 4 equally spaced factors but subsequently led to convergence issues due to low sample sizes and therefore was modified to two levels: ‘partial’ or ‘full’. The density of percent full by region and the density of cpue associated with each factor are shown in figure 11.

PARTY SIZE (ps)
The number of anglers reported for a trip was divided into 4 equally spaced factors but led to convergence issues due to low sample sizes and therefore was modified to two levels: ‘small’ or ‘large’.

Following filtering and subsetting, trips retained for model input are presented in Figures 12 and 13.

Standardization

Zero-inflated models are valuable tools for modeling distributions that do not fit standard error distributions due to excessive number of zeroes. These data distributions are often referred to as "zero-inflated" and are a common condition of count based ecological data. Zero inflation is considered a special case of over-dispersion that is not readily addressed using traditional transformation procedures (Hall 2000). Due to the high proportion of zero counts found in our data set (Figure 14), we used a zero-inflated mixed model approach that accounts for the high occurrence of zero values, as well as the positive counts. The model does so by combining binomial and count processes (Zuur et al. 2009).

The modeling approached used here was similar to that used in SEDAR58 for cobia headboat index. We initially considered a full null model (1) using both a zero-inflated Poisson (ZIP) and a zero-inflated negative binomial (ZINB) formulation,

$$Count = y + r + t + s + ps + v + pf \mid y + r + t + s + ps + v + pf \quad (1)$$

In this formulation, variables to the left of the "|" apply to the count sub-model, and variables to the right apply to the binomial sub-model. In this analysis, we favored a simpler null model because of the relatively small proportion of positive counts for cobia,

$$Count = y \mid y \quad (2)$$

which allowed us to add covariates using a step-wise forward selection process (rather than the backward selection). However, prior to adding covariates we compared ZIP and ZINB formulations. We compared the variance structure of each model formulation using AIC and likelihood ratio tests (Zuur et al 2009) to determine the most appropriate model error structure for the development of a cobia headboat index. The results of these tests (Table 2) support the ZINB formulation (similar results were obtained when using the full null model). These results concur with our expectations based on the over dispersion within the headboat data. A comparison between the fitted and original data for the ZIP and ZINB model formulations is shown in Figure 15. The rootogram (Kleiber and Zeileis 2017) in the lower panels of Figure 16 extends the Tukey (1977) rootogram to regression models. These plots are useful as diagnostics specific to overdispersion and/or excess zeros in count data models.

We used a step-wise forward model selection procedure to systematically include important covariates in our model formulation. In this procedure, we added each explanatory variable one at a time, alternating between the count (negative binomial) and binomial components. The variable with the largest ΔAIC was added, and the process repeated until no variable resulted in ΔAIC>2. The final cobia ZINB model formulation included year, vessel size, trip type, season, region and party size in the negative binomial component, and year, region, season, trip type and vessel size in the binomial component,

$$Count = y + v + t + s + r + ps|y + r + s + t + v \quad (3)$$

Diagnostics of the final model showed no clear patterns of association between Pearson's residuals and fitted values, or between the fitted values and original data (Figure 17) indicating acceptable model choice (Zuur et al 2009).

All data manipulation and analysis was conducted using R version 3.5.0 (R Core Team 2018). Modeling was executed using the **zeroinfl** function in the **pscl** package (Kleiber and Zeileis 2017), available from the Comprehensive R Archive Network (CRAN).

Uncertainty

Uncertainty in the index was computed using a bootstrap procedure with *n*=1000 replicates. In each replicate, a data set of the original size was created by drawing observations (rows) at random with replacement. This was done by year, to maintain the same annual sample size as in the original data. The model (Equation 3) was fitted to each data set, and uncertainty (CVs) was computed from those fits that converged.

Results and discussion

Annual standardized index values for cobia including CVs are presented in Table 3. The relative nominal index fell within the 2.5% and 97.5% confidence intervals of the standardized index and tracked closely with the standardized index (Figure 18). The ZINB standardized index is plotted with the SEDAR 28 delta-GLM index look similar except for a few years in the late 1990s and early 2000s along. In the most recent years the ZINB and the delta-GLM indices have a similar downward trend but the ZINB is consistently higher.

Literature cited

Hall, D. B. 2000. Zero-Inflated Poisson binomial regression with random effects: a case study. Biometrics, 56: 1030-1039.

Stephens, A., and A. MacCall. 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fish. Res. 70:299-310.

Tukey J.W. 1977. Exploratory Data Analysis. Addison-Wesley Publishing Company, Phillipines.

Zeileis A. and C. Kleiber. 2017. countreg: Count Data Regression. R package version 0.2-0/r34, URL http://R-Forge.R-project.org/projects/countreg/.

Zeileis A., C. Kleiber, and S. Jackman. 2008. "Regression Models for Count Data in R." Journal of Statistical Software, 27(8), 1–25. doi: 10.18637/jss.v027.i08.

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. Total number of headboat trips and positive cobia trips in the Gulf of Mexico (and east coast of FL) by year for the raw data compared to the model input.

	Raw Data				Core Vessel			
Year	cobia trips	total trips	N.fish	%pos	cobia trips	total trips	N.fish	%pos
1986	903	13,712	1,366	7%	690	7,876	1,007	9%
1987	958	14,125	1,457	7%	651	7,116	926	9%
1988	905	14,413	1,501	6%	648	6,335	1,020	10%
1989	773	14,444	1,270	5%	531	5,966	841	9%
1990	904	18,064	1,547	5%	645	6,556	1,099	10%
1991	1,004	16,179	1,911	6%	787	5,929	1,552	13%
1992	1,575	20,466	3,025	8%	1,276	8,049	2,491	16%
1993	1,739	20,102	4,003	9%	1,287	7,245	2,912	18%
1994	1,612	18,715	3,611	9%	1,234	6,859	2,569	18%
1995	1,450	16,822	3,598	9%	1,105	6,164	2,358	18%
1996	1,150	13,886	2,826	8%	841	4,147	1,928	20%
1997	1,256	14,537	2,966	9%	965	4,624	2,318	21%
1998	1,133	12,760	2,242	9%	772	3,750	1,541	21%
1999	760	9,411	1,500	8%	612	2,760	1,270	22%
2000	821	10,317	1,388	8%	630	3,174	1,120	20%
2001	924	9,963	1,751	9%	746	3,238	1,415	23%
2002	879	9,213	1,367	10%	675	2,947	1,077	23%
2003	719	9,183	1,181	8%	536	3,009	932	18%
2004	810	9,929	1,350	8%	586	3,209	988	18%
2005	1,015	9,666	1,548	11%	652	2,750	1,058	24%
2006	938	8,882	1,636	11%	777	3,282	1,433	24%
2007	1,010	9,640	1,732	10%	794	3,287	1,402	24%
2008	892	12,518	1,601	7%	638	3,510	1,168	18%
2009	1,305	15,087	2,162	9%	944	5,255	1,613	18%
2010	1,227	13,845	2,356	9%	964	5,067	1,843	19%
2011	1,160	15,183	2,032	8%	823	4,964	1,502	17%
2012	1,315	16,184	2,483	8%	975	5,596	1,800	17%
2013	1,255	17,188	2,009	7%	995	5,884	1,646	17%
2014	1,298	20,933	2,174	6%	1,040	7,244	1,743	14%
2015	1,095	21,945	1,739	5%	823	7,271	1,312	11%
2016	1,086	20,744	1,804	5%	837	6,002	1,321	14%
2017	853	16,185	1,335	5%	578	3,824	919	15%
2018	716	14,786	1,073	5%	497	3,560	766	14%

Table 2: Preliminary model error structure comparison

	df	**Likelihood**	**AIC**	χ^2	**df**	***p*-value**
ZIP	66	-110579	221290			
ZINB	67	-104214	208562	12730	1	<0.001

Table 3: The relative nominal *Count*, proportion positive, standardized index, and CV for the SEDAR 28 Update (2020) cobia index.

Year	Nominal	Proportion Positive	Standardized Index	CV
1986	0.409	9%	0.487	0.033
1987	0.416	9%	0.466	0.050
1988	0.515	10%	0.610	0.046
1989	0.451	9%	0.527	0.041
1990	0.536	10%	0.679	0.059
1991	0.837	13%	0.922	0.037
1992	0.990	16%	1.022	0.019
1993	1.285	18%	1.241	0.046
1994	1.198	18%	1.087	0.034
1995	1.223	18%	1.055	0.041
1996	1.487	20%	1.194	0.046
1997	1.603	21%	1.325	0.030
1998	1.314	21%	1.050	0.049
1999	1.471	22%	1.095	0.012
2000	1.128	20%	0.837	0.039
2001	1.397	23%	1.082	0.025
2002	1.169	23%	0.962	0.024
2003	0.990	18%	0.763	0.031
2004	0.985	18%	0.818	0.071
2005	1.230	24%	1.044	0.057
2006	1.396	24%	1.132	0.064
2007	1.364	24%	1.177	0.034
2008	1.064	18%	1.261	0.030
2009	0.982	18%	1.123	0.033
2010	1.163	19%	1.487	0.055
2011	0.968	17%	1.229	0.045
2012	1.029	17%	1.502	0.028
2013	0.895	17%	1.203	0.038
2014	0.769	14%	1.200	0.028
2015	0.577	11%	0.818	0.035
2016	0.704	14%	0.962	0.025
2017	0.768	15%	0.877	0.055
2018	0.688	14%	0.761	0.054

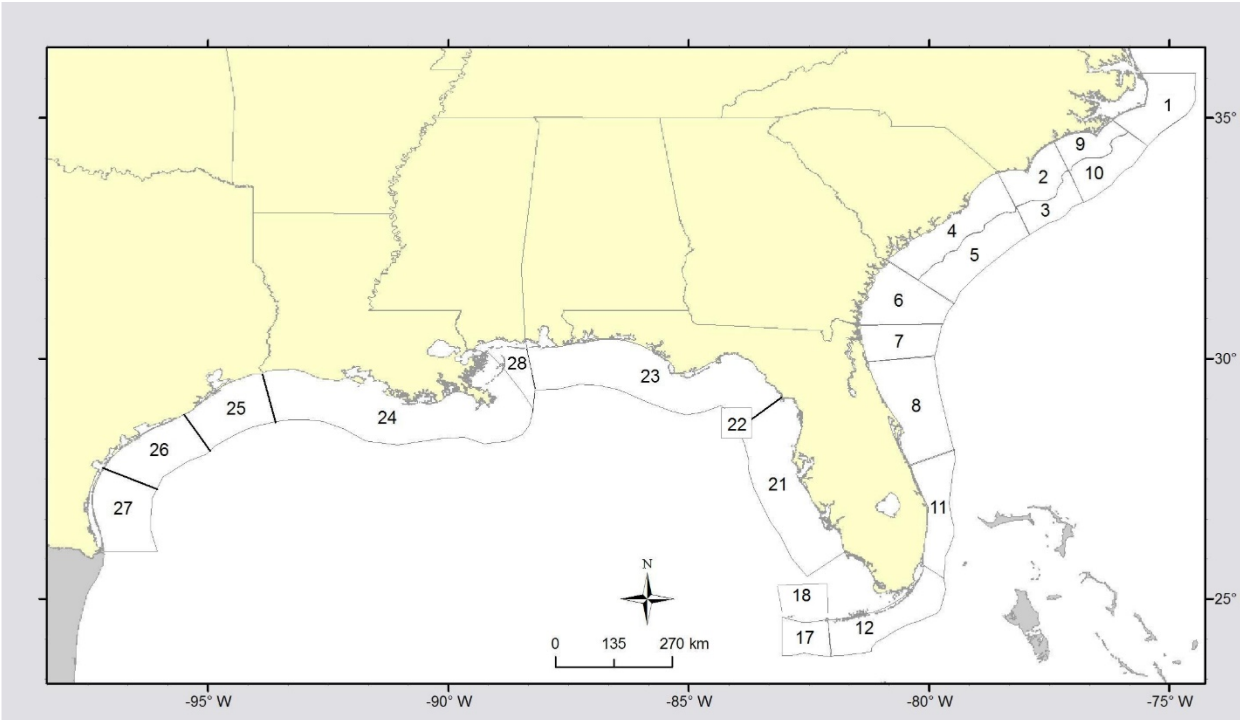


Figure 1. Map of headboat sampling area definition.

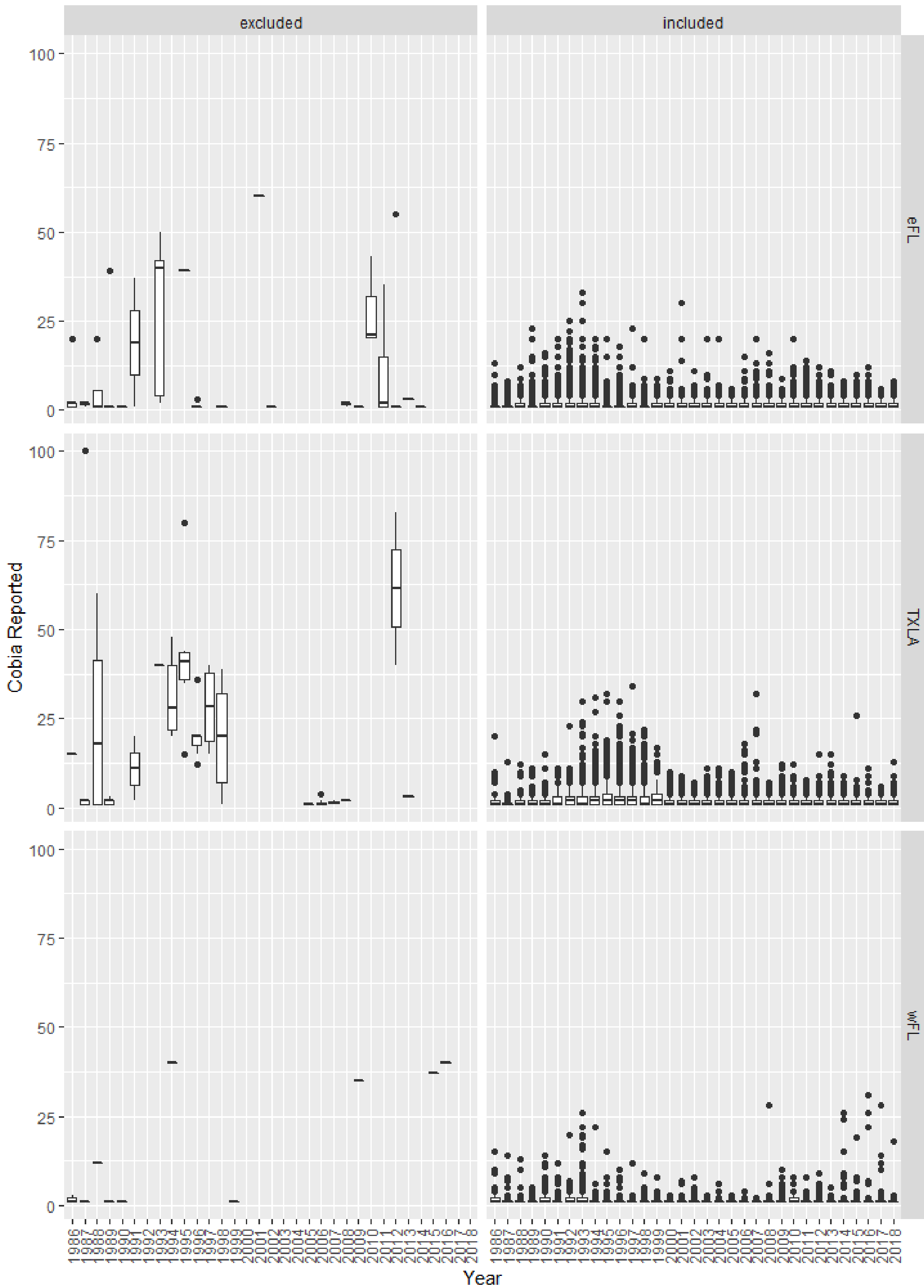


Figure 2. Records determined as outliers (excluded) based on removal of values above the 99.5th percentile for anglers and 99.9th percentile for number of fish caught.

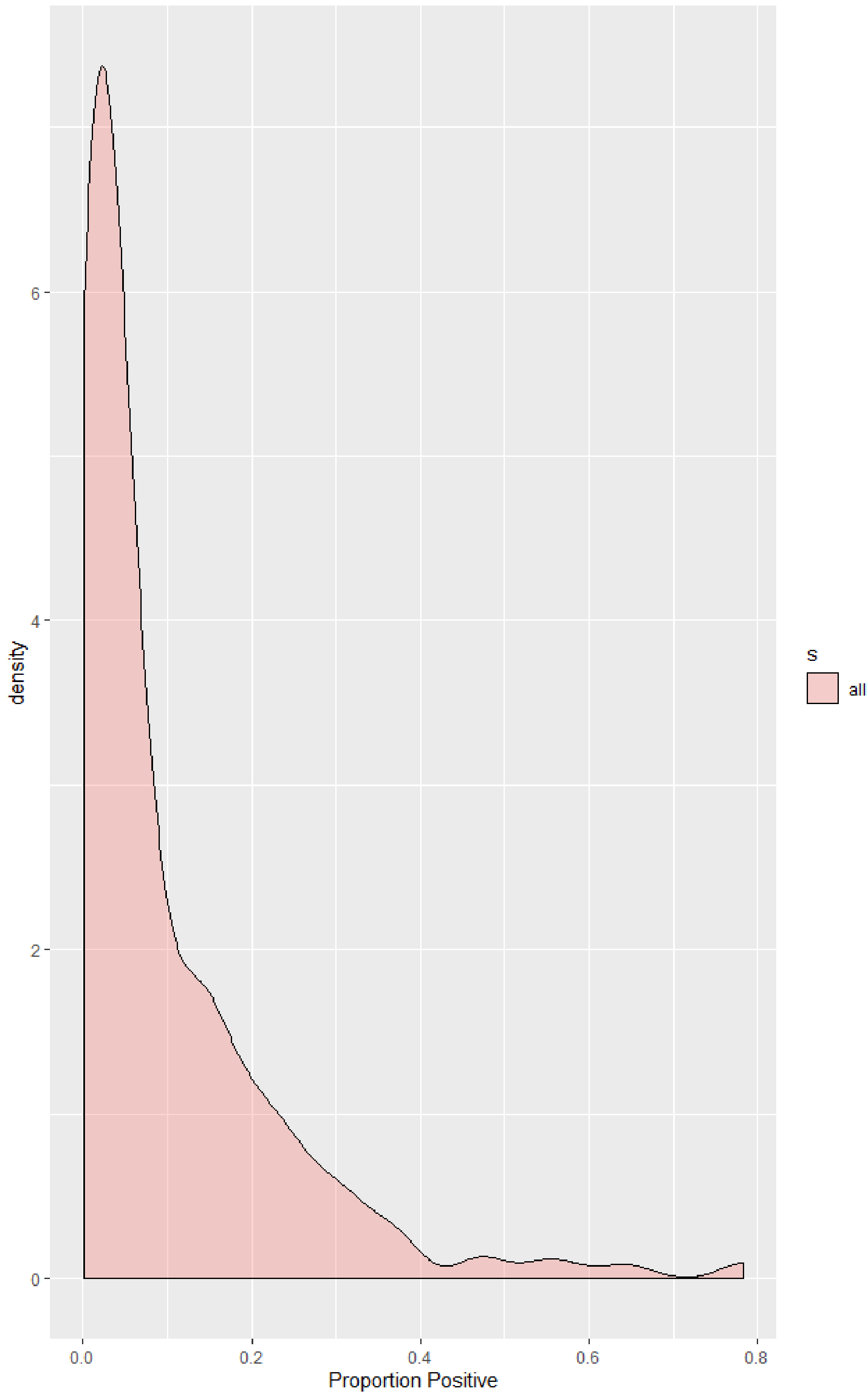


Figure 3. Density of proportion positive of positive cobia trips among vessels.

Figure 4. Each series of horizontal dots represents a ‘core’ vessel participation in the headboat fishery (blank=not active, red=in fishery-at least one positive cobia trips and blue=in fishery-no cobia.

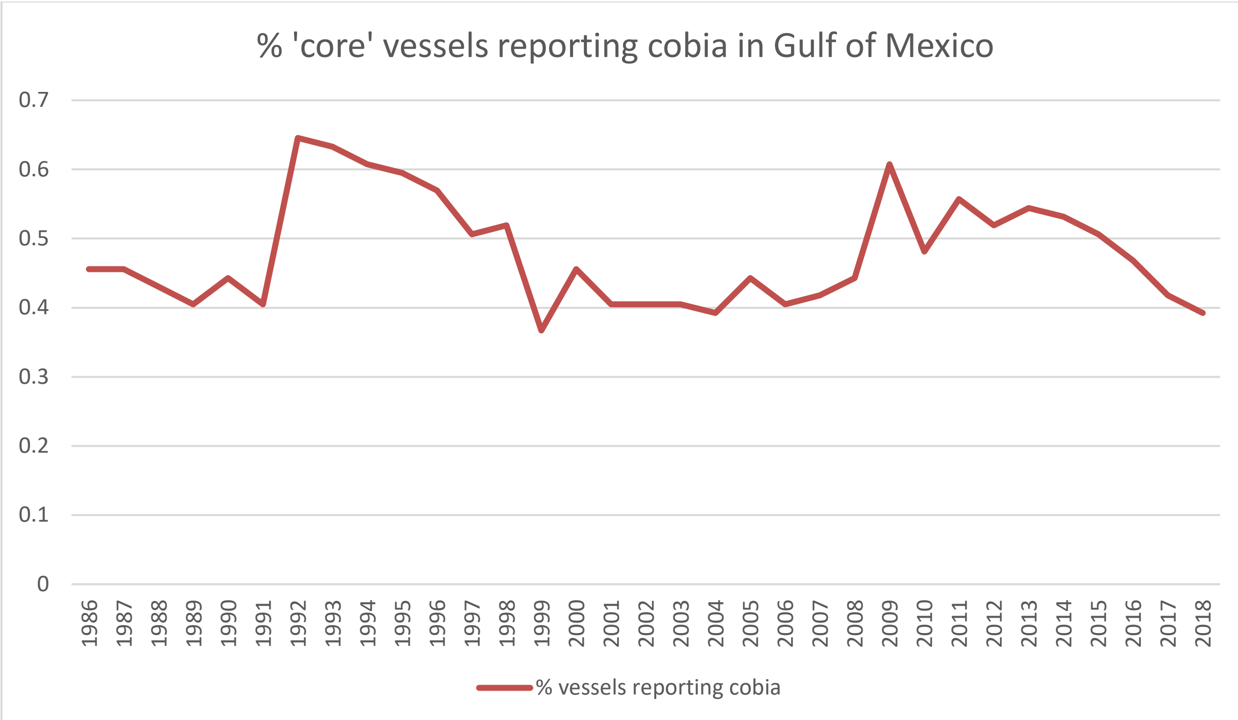


Figure 5. Proportion of active ‘core’ vessels participating in the headboat fishery by year.

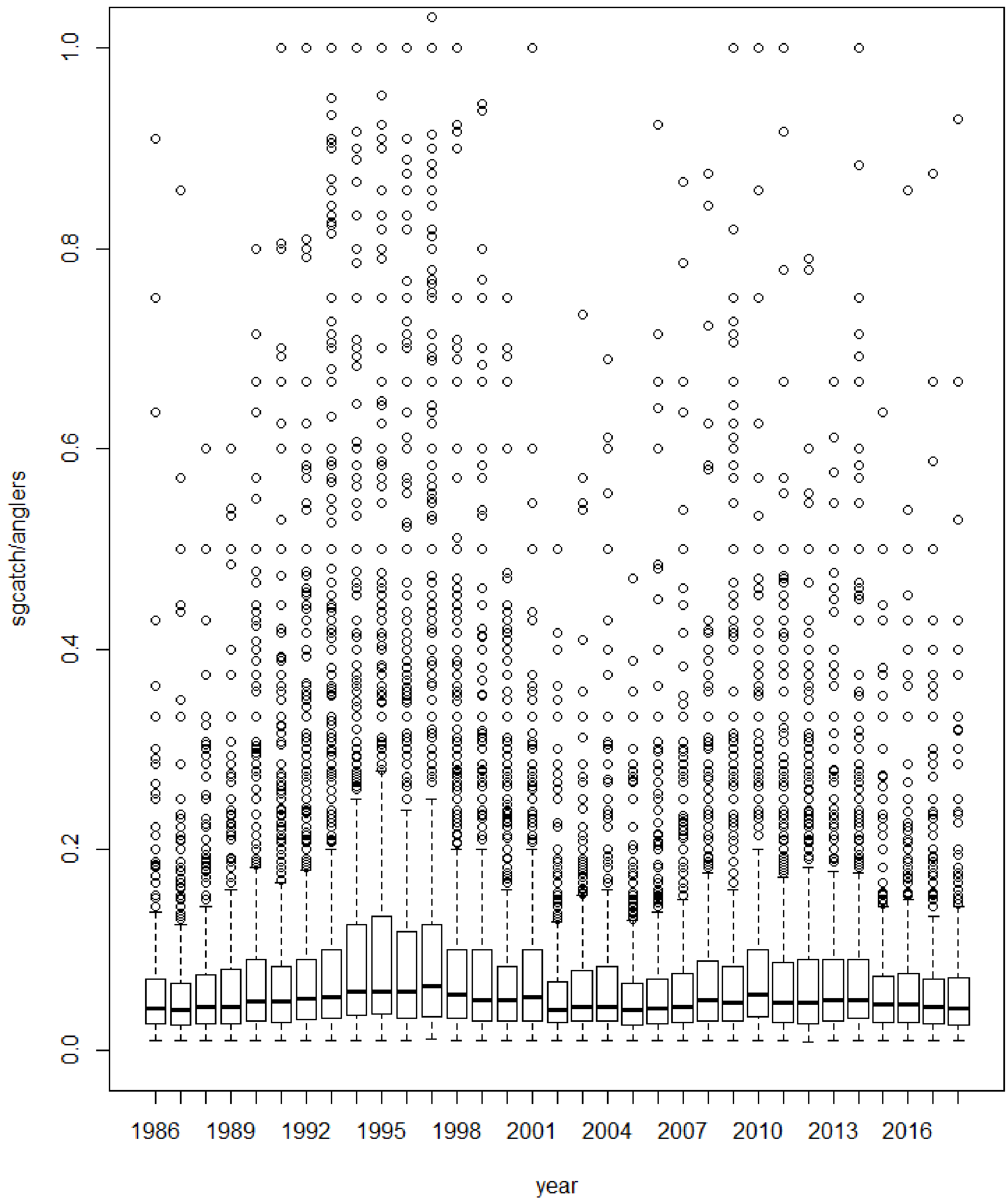


Figure 6. Box plot of cpue by year from headboat logbook records from 1986-2018 from Florida to Texas.

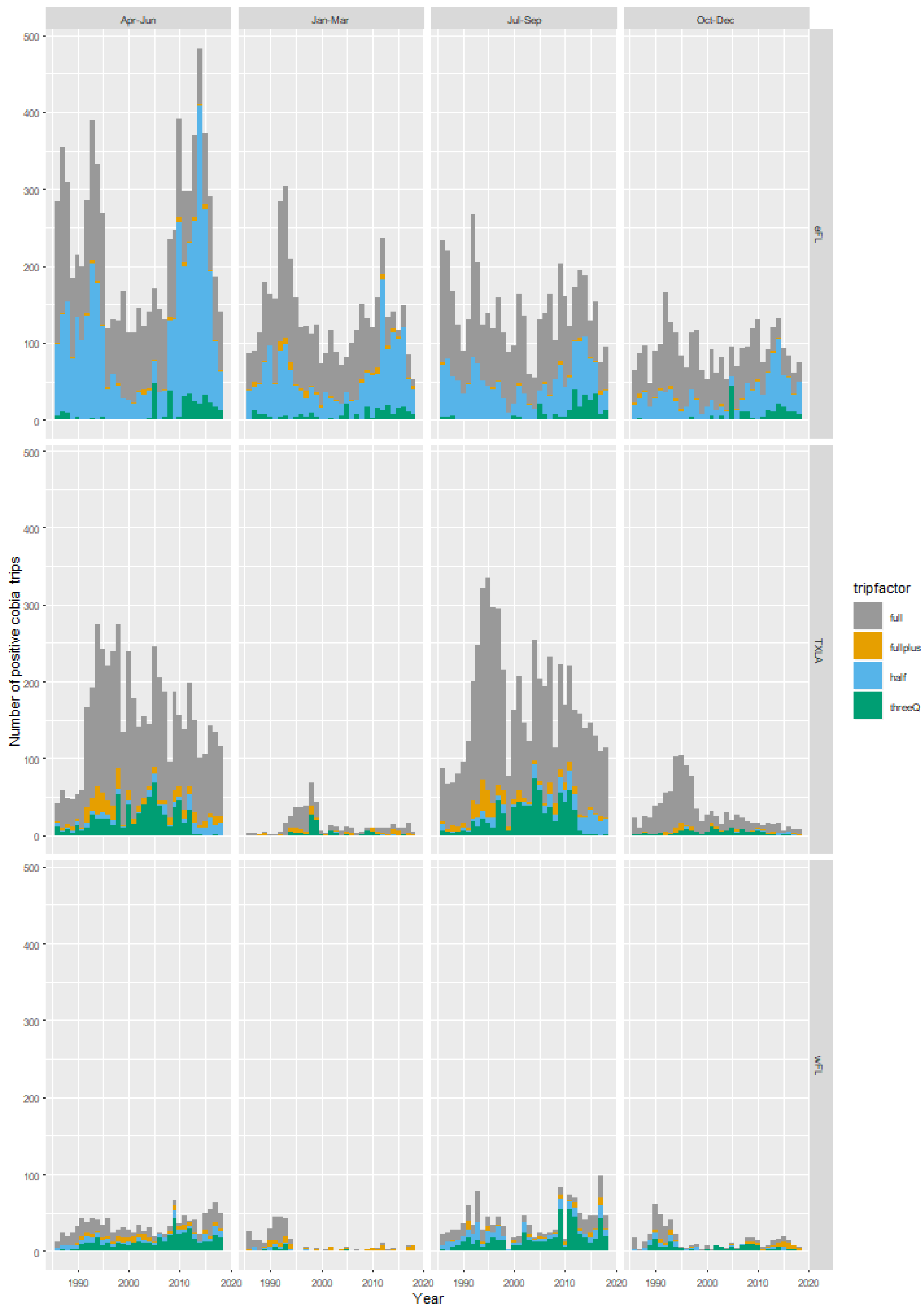


Figure 7. Unfiltered positive cobia trips by region and season prior to model input filtering.

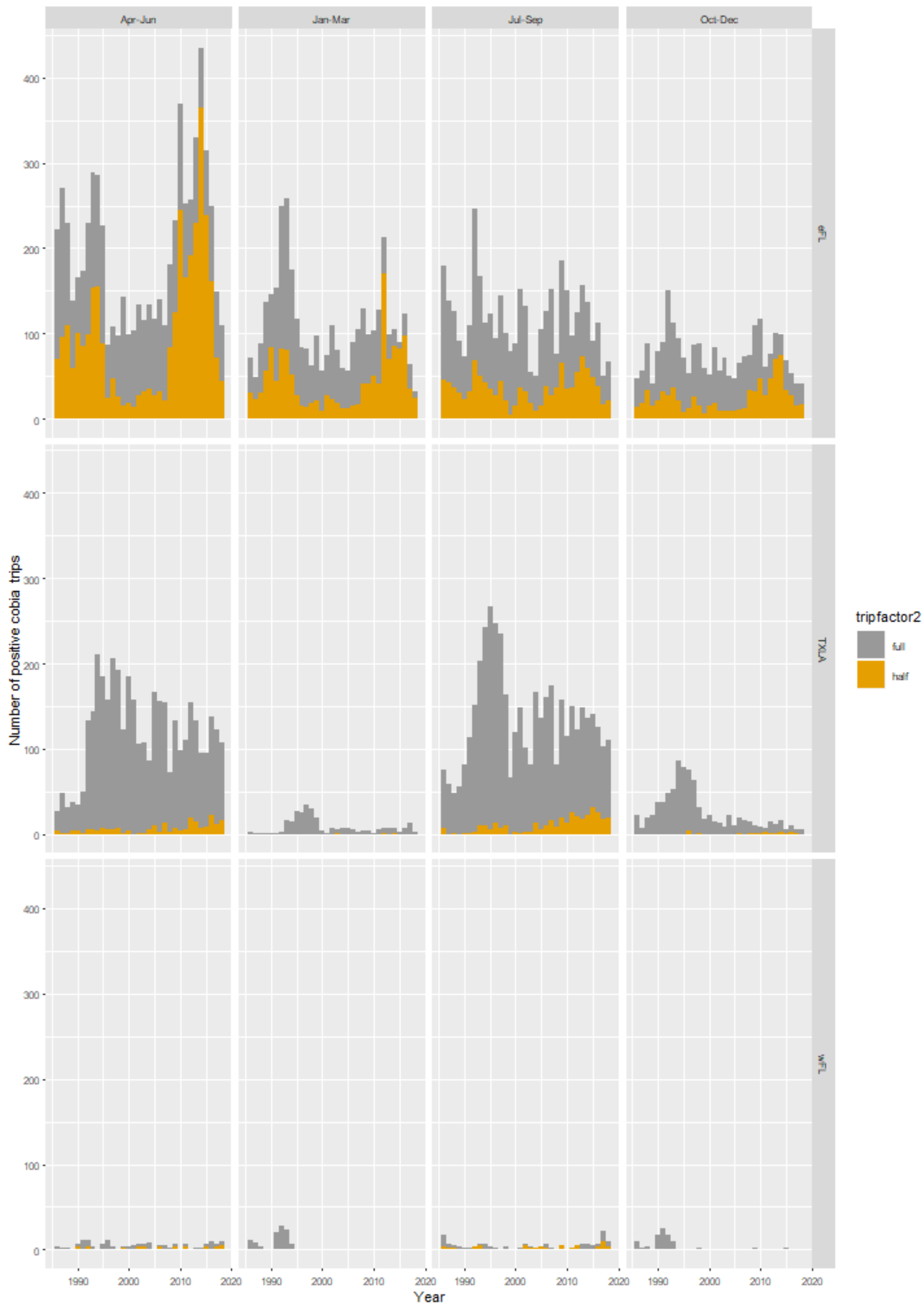


Figure 8. Model input of positive cobia trips by region and season (full and full plus trips only) (Regions: east coast FL, west coast FL, Alabama-Texas).

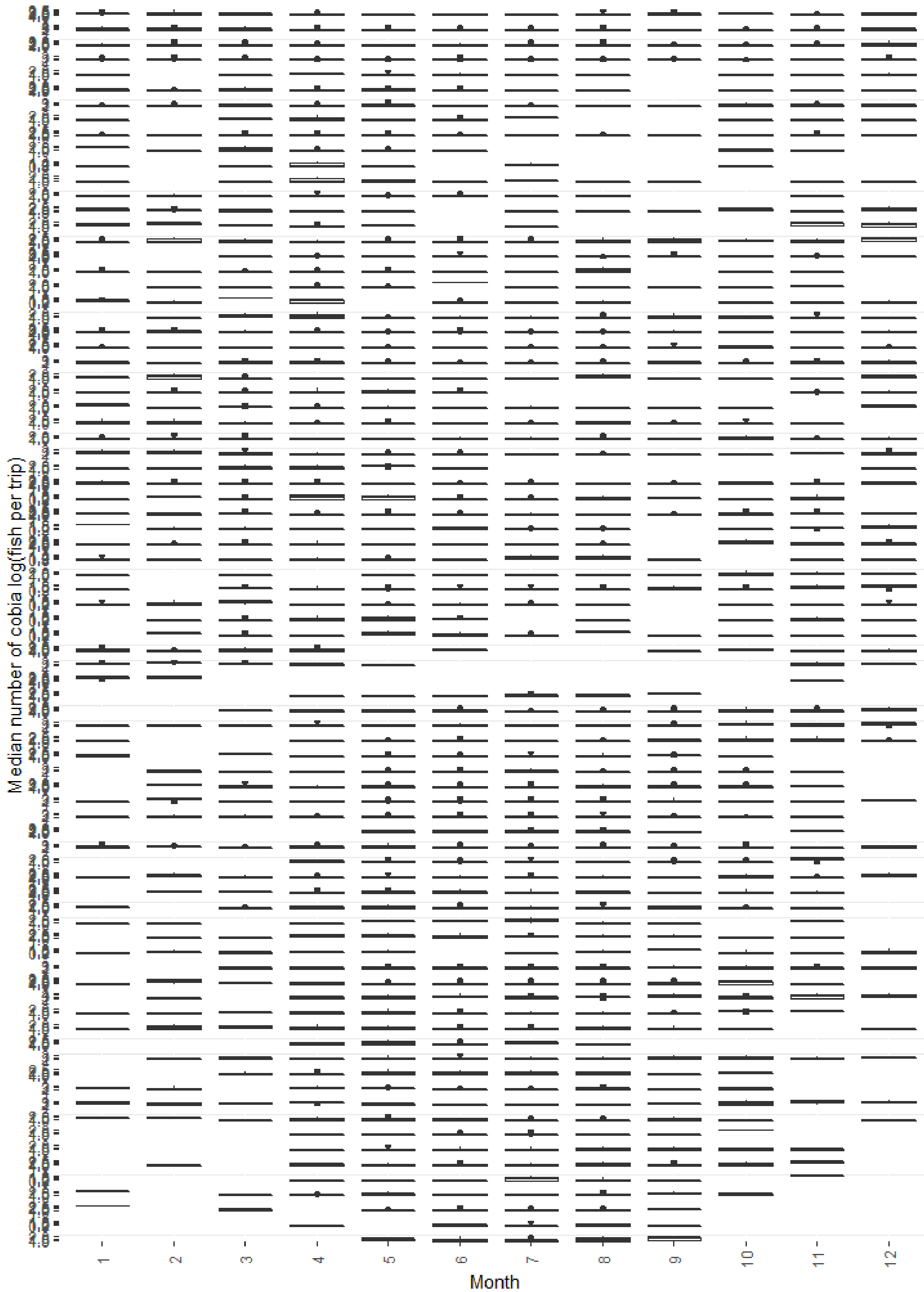


Figure 9. Boxplot of monthly mean cobia caught by 'core' headboats (prior to filtering months for model input). The intent of this boxplot is to show that all months should be retained.

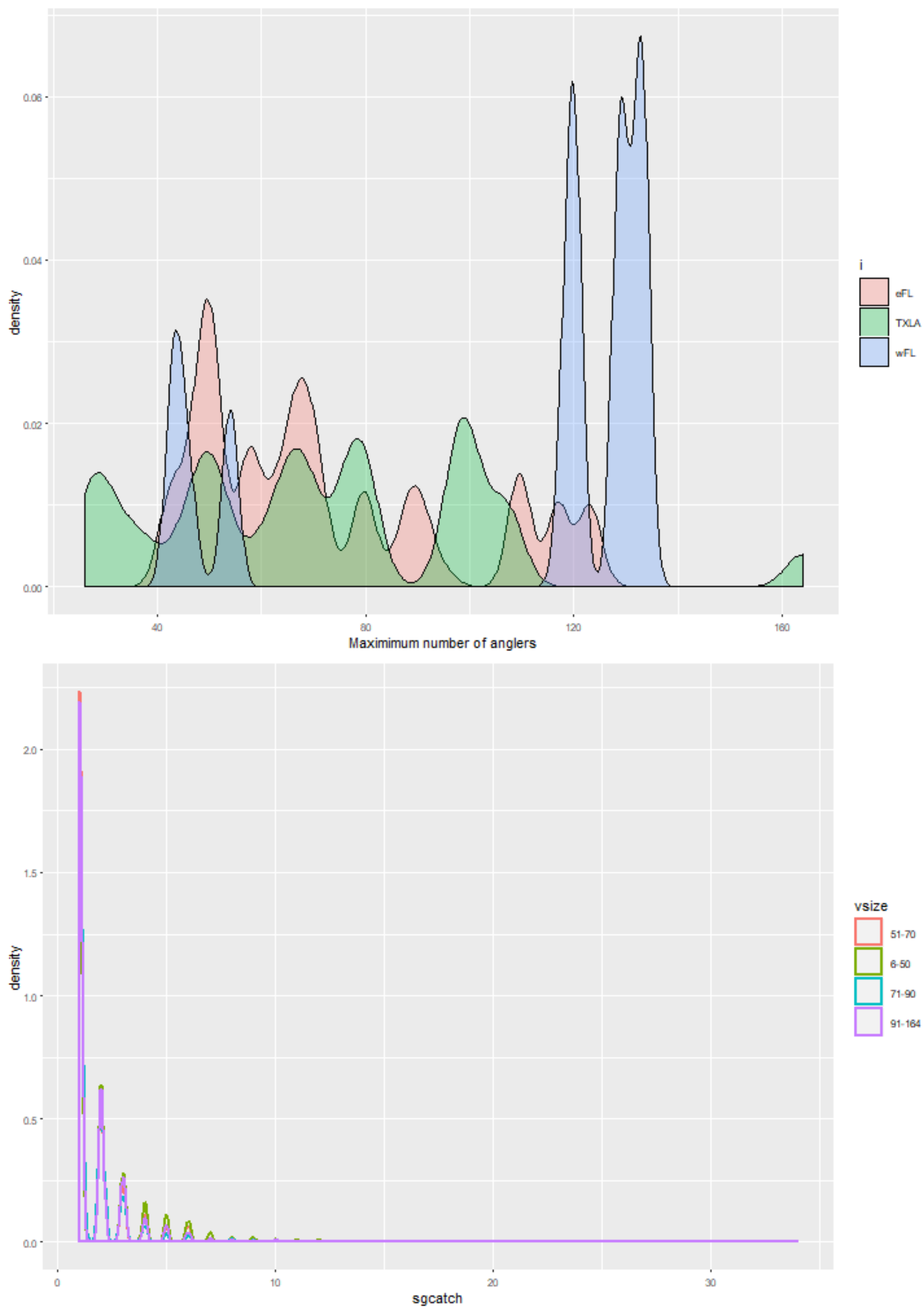
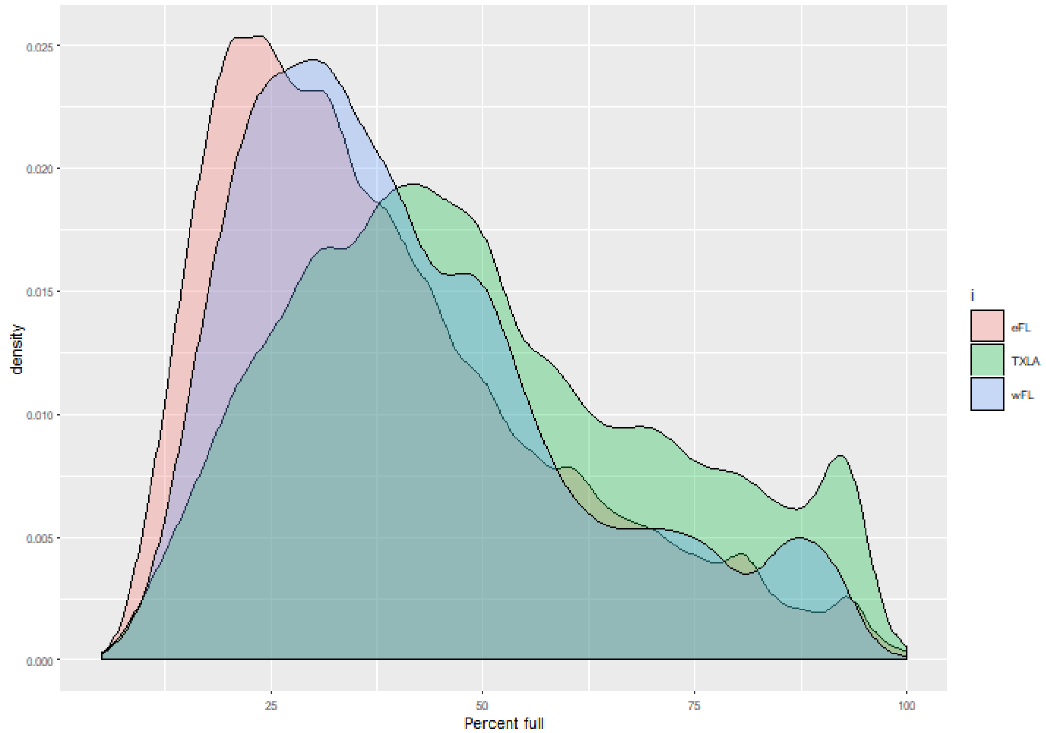


Figure 10. Density of maximum number of anglers across regions and cpue associated with the factors for maximum anglers as a proxy for vessel size. Due to convergence issues the factor vessel size was simplified to 'small' and large'.



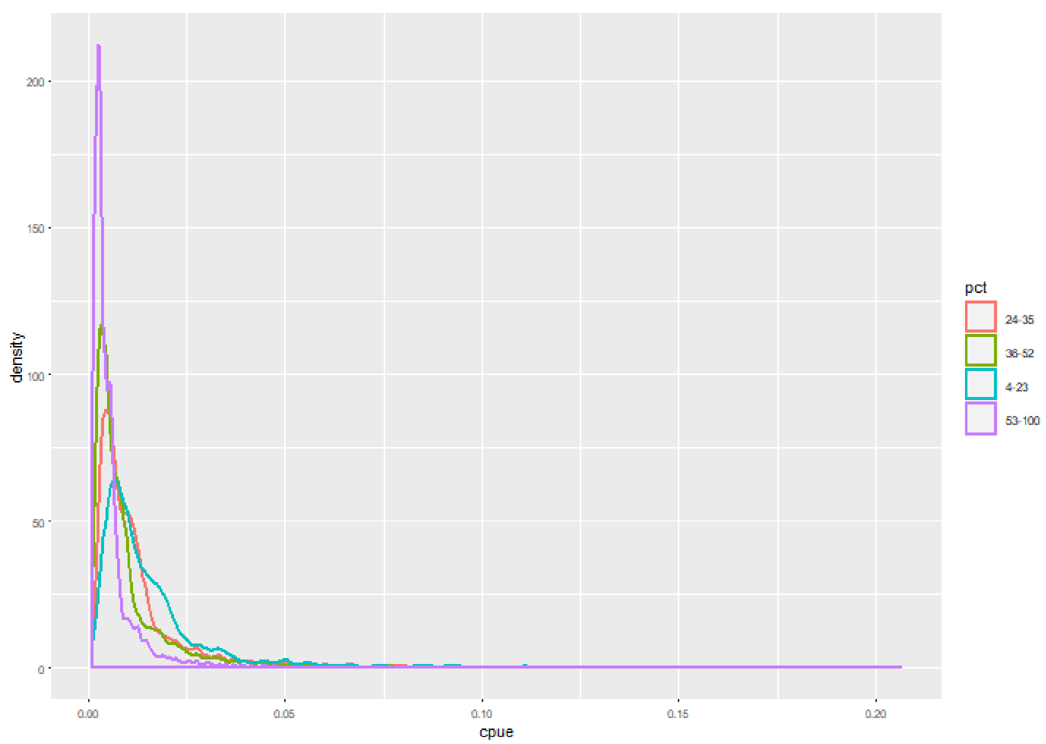


Figure 11. Density of percent full across regions and cpue associated with the factors for percent full. Due to convergence issues the factor percent full was simplified to ‘half and full.

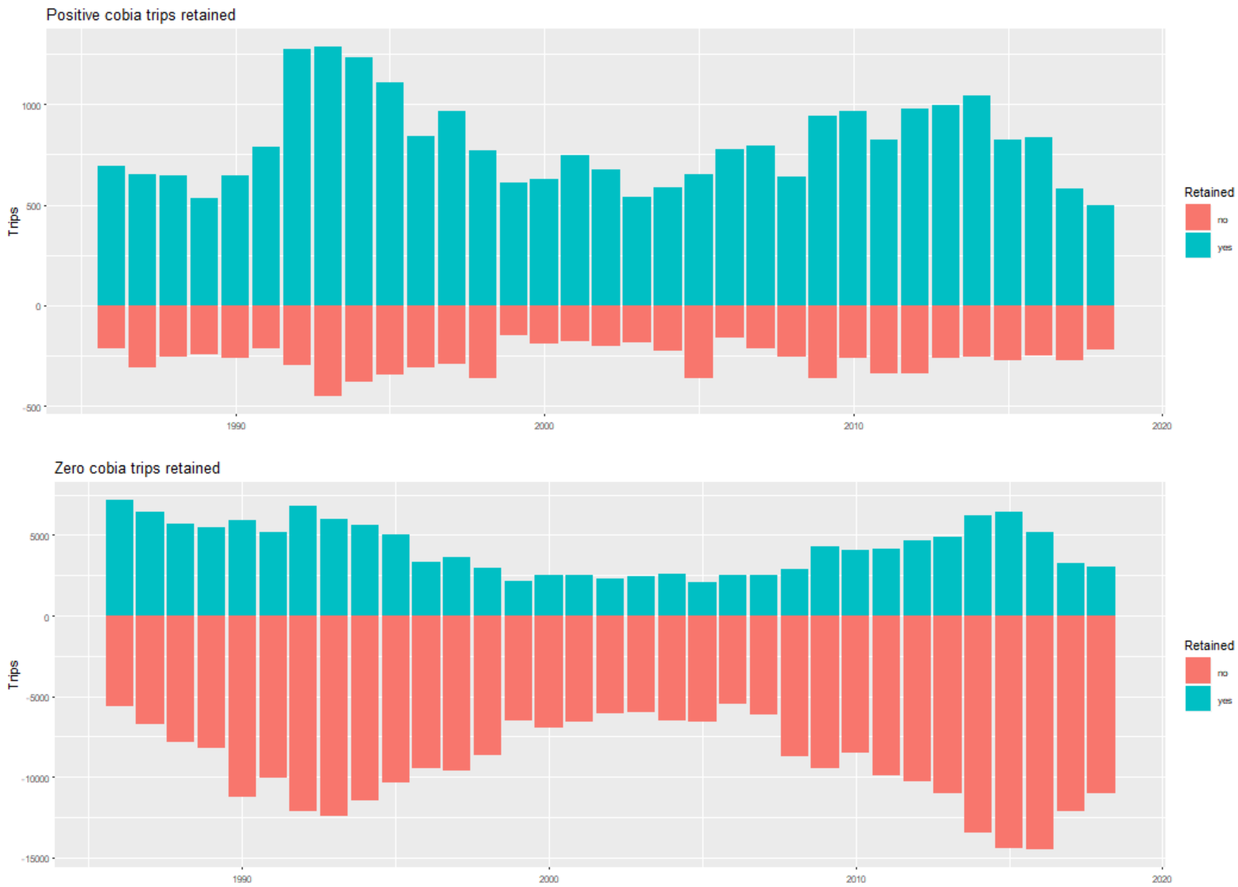


Figure 12. Positive and zero cobia trips retained after subsetting using 'core' vessels by year.

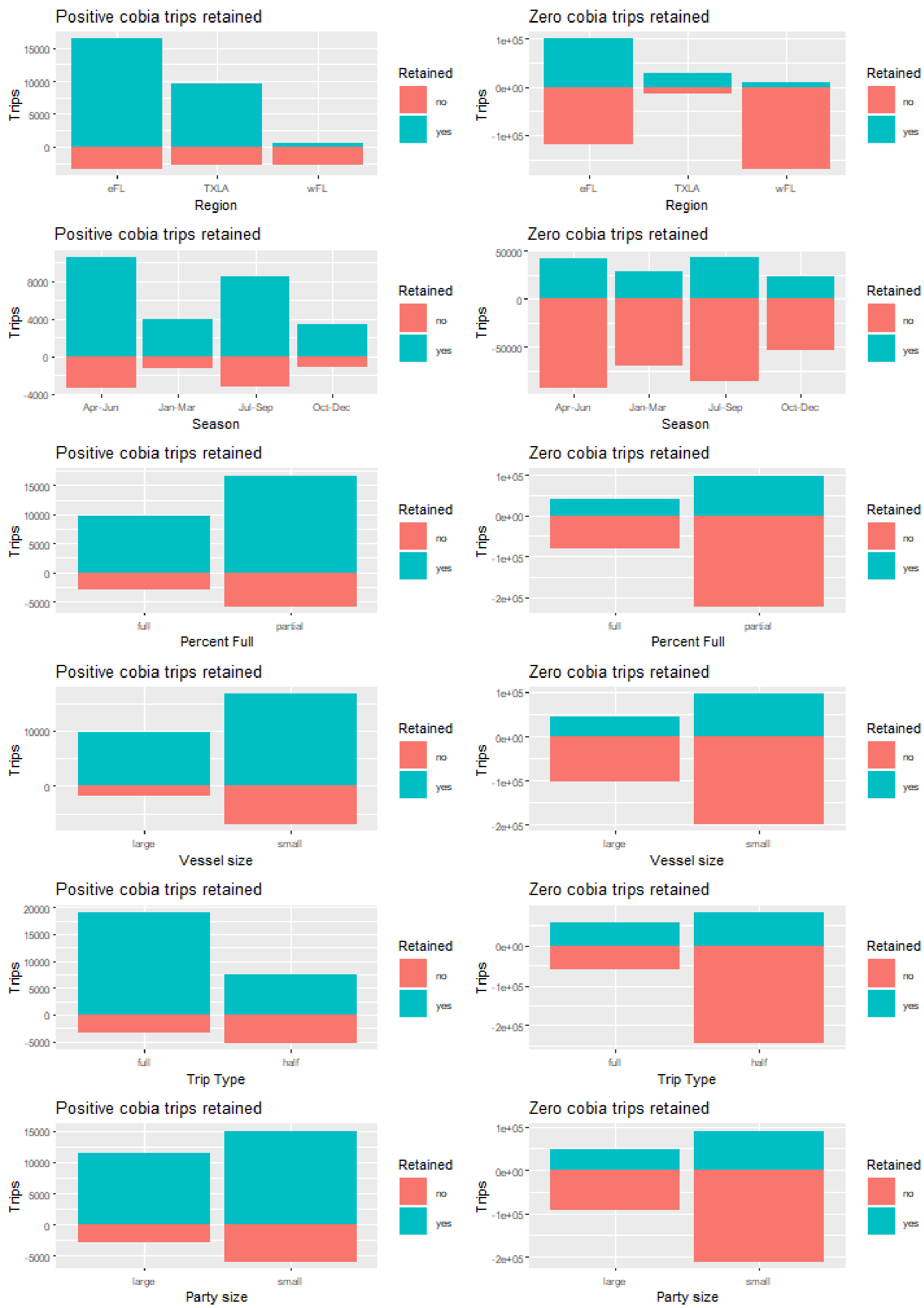


Figure 13. Positive and zero cobia trips retained after subsetting using 'core' vessels by factor. Note:

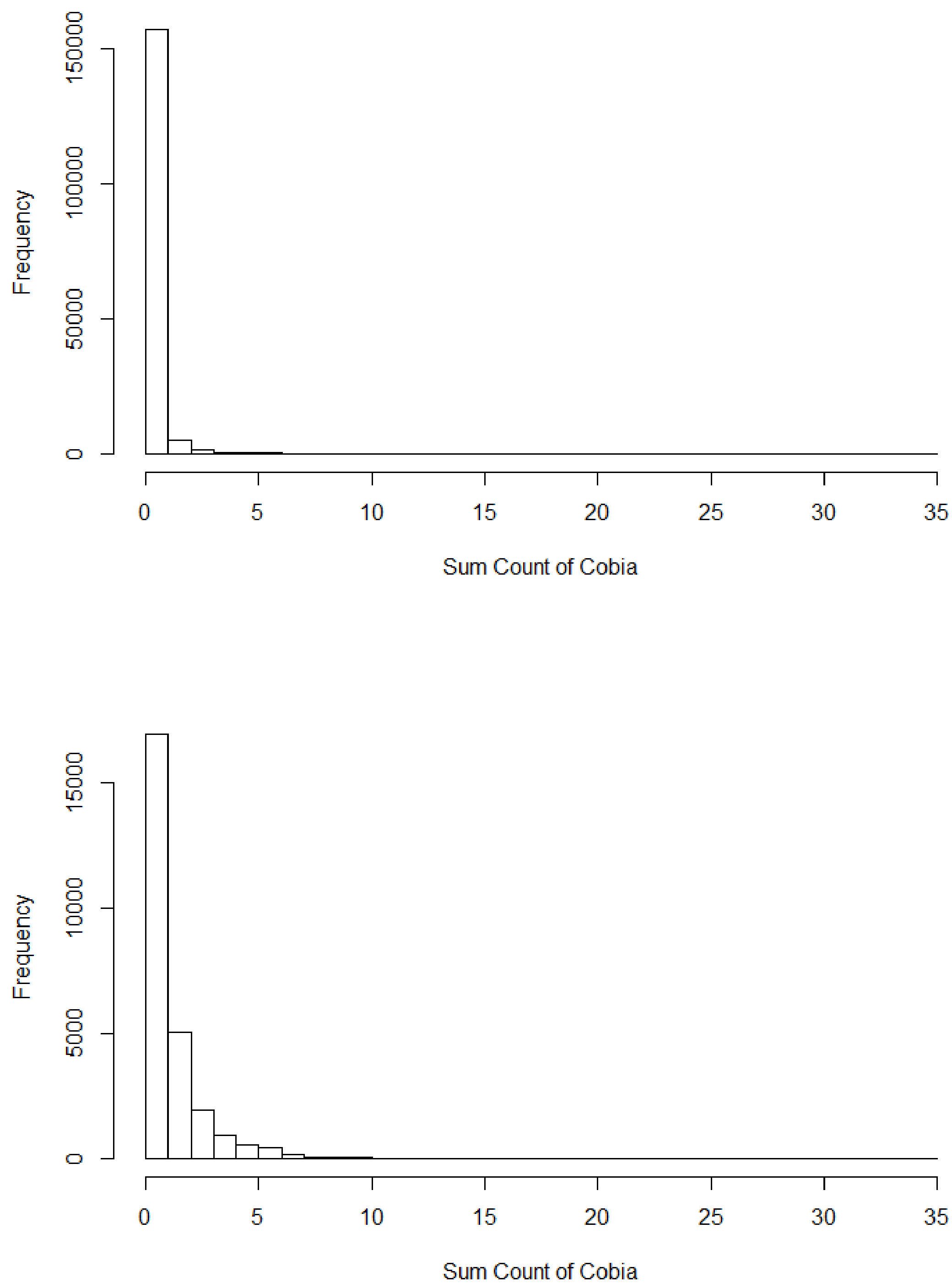


Figure 14. Count distribution of all 'core' cobia trips (top) and positive 'core' cobia trips (bottom).

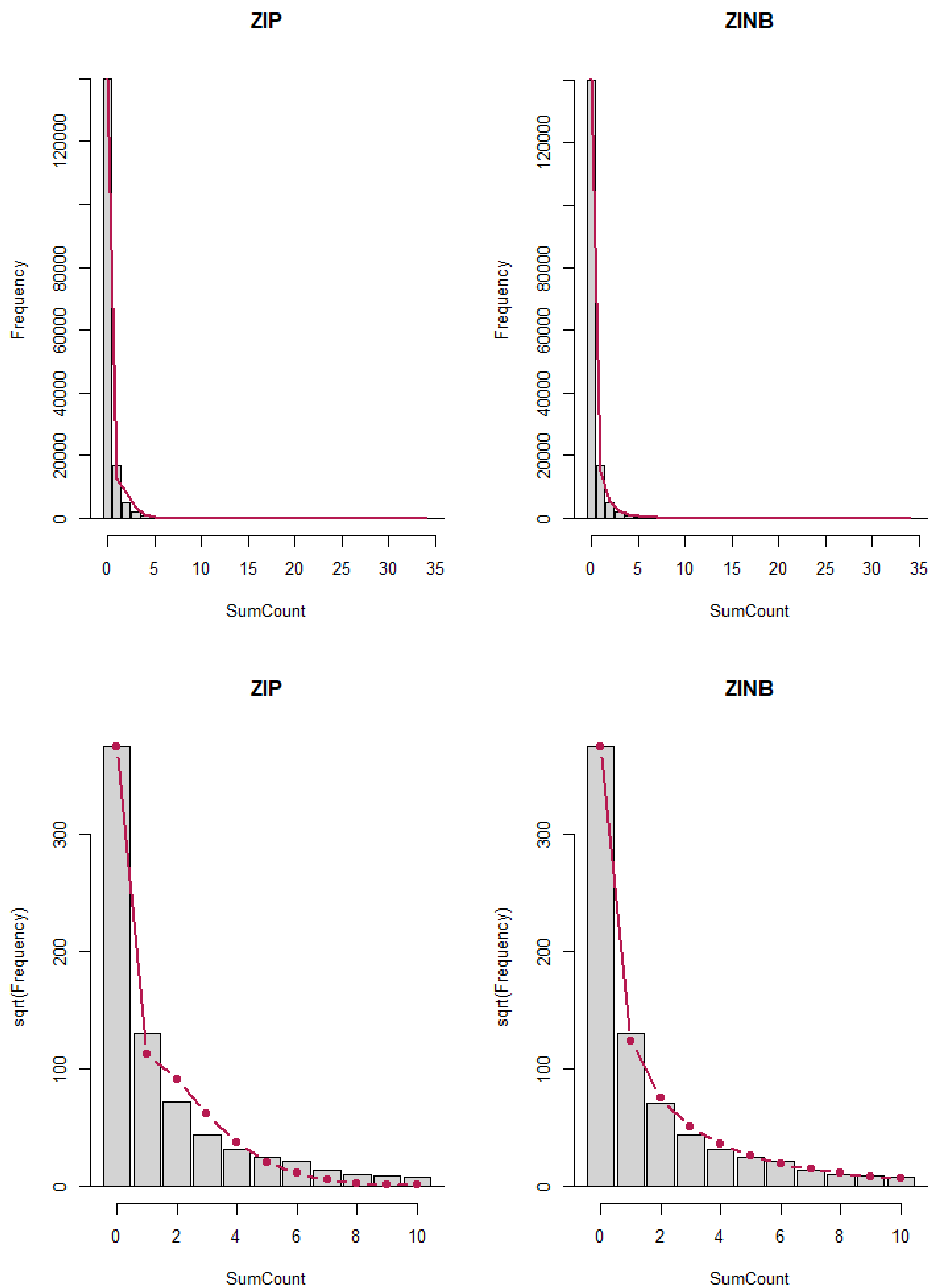
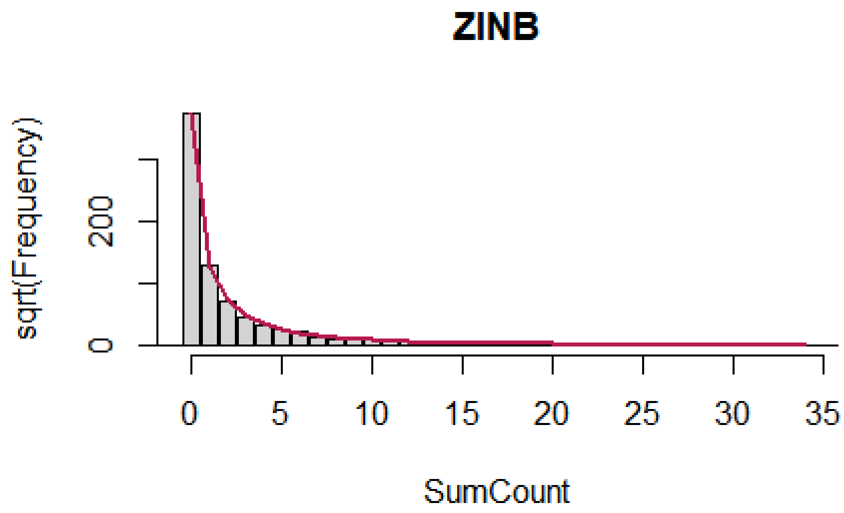


Figure 15. Model formulation comparison, with ZIP (left) and ZINB (right) fitted values plotted against the original data distribution with all covariates included. The lower panels are square root transformed and truncated at 10 fish for inspection of goodness of fit over the range of values for the bulk of the data.



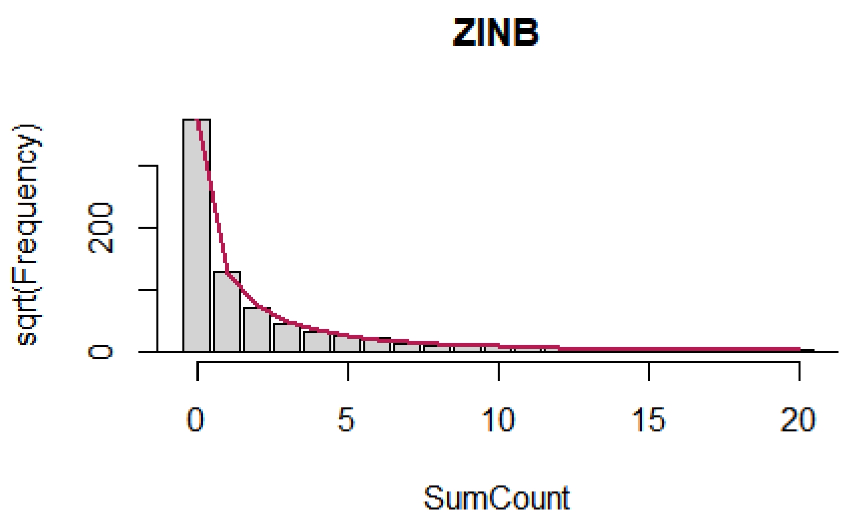


Figure 16. Model diagnostic plots of fitted model values (red line) against the original data distribution for the preferred model.

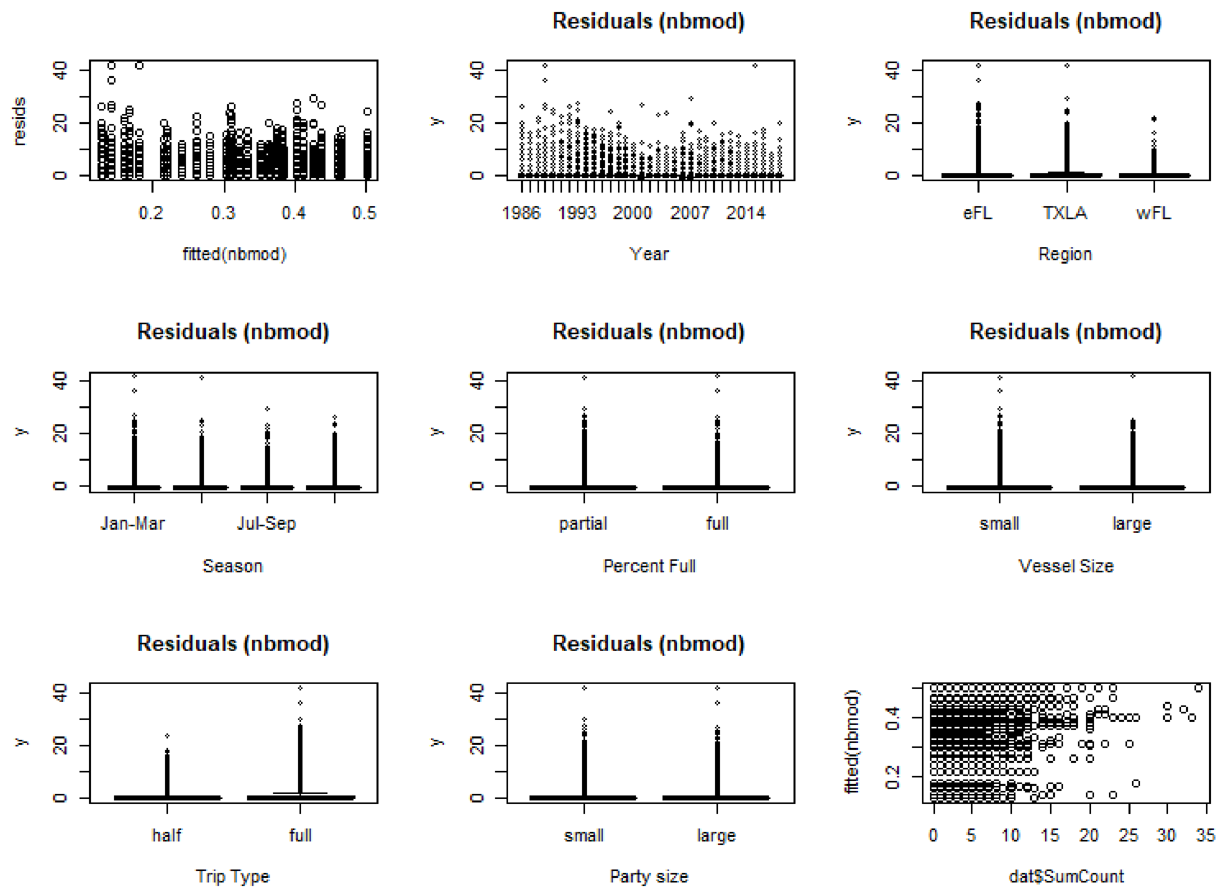


Figure 17. Model diagnostic plot showing residuals from final model (ZINB).

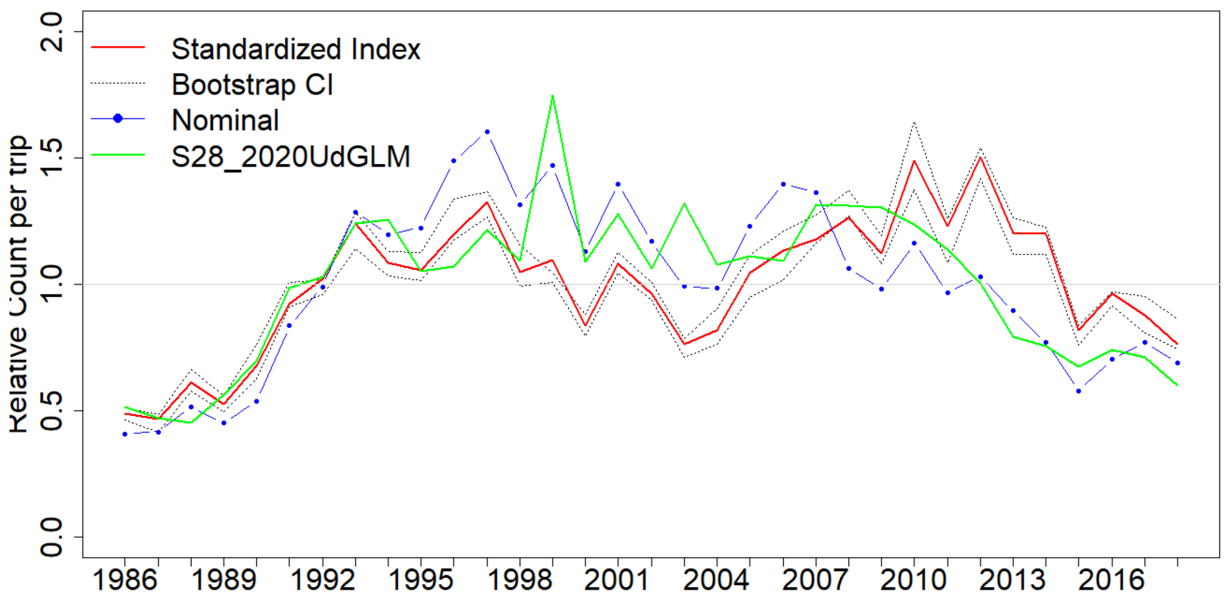


Figure 18. Relative standardized index (solid red line) with 2.5% and 97.5% confidence intervals (dashed lines) and the relative nominal index (blue) for Gulf of Mexico cobia in the SRHS headboat logbook data. The green line represents the most recent SEDAR 28 updated index using the original delta-GLM approach.