Standardized catch rates of yellowtail snapper (*Ocyurus chrysurus*) from the Marine Recreational Fisheries Statistics Survey in south Florida, 1981-2010

Introduction

Yellowtail snapper are caught by recreational anglers primarily in south Florida from Palm Beach County on the east coast to Monroe County on the west coast. The Marine Recreational Fisheries Statistics Survey (MRFSS) was initiated in 1981 to collect catch, effort, and participation estimates from the recreational sector. MRFSS consists of a telephone survey of fishing effort and an access-site intercept survey of angler catch. Intercepts are conducted at public marine fishing access points (boat ramps, piers, beaches, marinas, etc.) to collect individual catch data including number of each species caught, number discarded, length, and weight. Access points are selected by a proportional random selection process in order to sample high activity sites most often. From these intercept data, catch per trip estimates can be made for each species encountered. These catch rates can provide an indication of population trends over time and are combined with the effort estimates from the telephone survey to produce total catch and harvest estimates. In 1991, MRFSS made several improvements to the survey and one of which was the linking together of separate intercepts of anglers that fished on the same trip and recording the total number of anglers in the party.

Methods

Data Preparation

To generate a standardized MRFSS catch rate, I used intercepts made during wave 2, 1981 through wave 6, 2010 on both coasts of Florida where the fishing mode was either private/rental or charter boat. Prior to 1986, headboat and charterboat intercepts were combined into a single fishing mode, however we excluded the headboat interviews by using the mode_f variable. Type 4 records were used to collapse ancillary intercepts into the appropriate type 1 fishing trip for years prior to 1991. All inland trips were excluded as well as any species that were observed in at least 1% of total trips. These constraints resulted in 124,998 recreational fishing trips, of which 7,159 caught yellowtail snapper.

Data Subsetting

Of the 124,998 intercepts made from 1981 through 2010, some of these did not occur in areas where yellowtail snapper occur and should therefore be removed from the standardization process. To subset for trips likely to have caught yellowtail snapper, I used the clustering technique of Shertzer and Williams (2008) to identify species that co-occur with yellowtail snapper. To do this, I performed hierarchical cluster analysis with average linkage on the Bray-Curtis similarity measure calculated on the square-root transformed catch data (total number of fish). Because choosing the appropriate number of clusters can be somewhat subjective, I plotted the number of clusters against the average distance between clusters (height) and used a piecewise regression with one breakpoint to determine the inflection point of the plot (Figure 1). The inflection point was chosen as that with the lowest residual mean square error, and at this point little information is gained from adding additional clusters. Eleven clusters were identified using this method with yellowtail snapper clustering with mutton snapper, blue runner, lane snapper, gray snapper, cero mackerel, black grouper, and grunts (Figure 2). This cluster can be considered a southern assemblage and is very similar to that revealed by Shertzer and Williams

(2008). After removing all trips that did not capture one of these species, 29,485 intercepts remained.

Response and Explanatory Variables

After subsetting and prior to standardization, only trips occurring in either Monroe, Dade, Broward, or Palm Beach counties were retained as these counties account for nearly 98% of the catch and is considered to be encompass the center of yellowtail snapper's distribution. Records where hours fished, number of contributors, or days fished in previous wave (avidity) were not available were also removed. Finally, I removed any records with high leverage, calculated as that which is above a critical value determined by 2*p/n where p is the number of potential parameters in the model. The filtered dataset used in standardization consisted of 15,026 trips, of which 6,574 captured yellowtail snapper.

CPUE - The response variable is the total number of yellowtail snapper caught on a trip (A + B1 +B2). There were many trips which captured no yellowtail snapper and those that did usually caught less than 10 (Figure 3).

AREA - The MRFSS variable area_x was combined into two categories, inshore and offshore. There were 8,096 trip in the inshore area and 6,930 offshore trips. Mean CPUE was similar for both areas up until 1999 when CPUE for inshore trips became consistently lower than offshore trips indicating a potential year*area interaction (Figure 4a).

HOURS FISHED - Hours fished was combined into 6 levels (lower number is inclusive in group): 0-2, 2-4, 4-6, 6-8, 8-10, 10-12, and 12+. Most trips (5,377) fished for 4-6 hrs followed by 2-4 hrs (4,660) and 6-8 hrs (3,286). Mean CPUE is fairly constant across all categories of hours fished and there appears to be no interaction with year (Figure 4b).

CONTRIBUTORS - The number of contributors to an intercept was combined into 7 groups, 1-7+. The number of intercepts declined with the number of contributors. As expected, CPUE appears to increase with the number of contributors and there is no apparent year*contributors interaction (Figure 4c).

AVIDITY - Avidity was created as the number of days fished within last 2 months as a factor with 7 levels (lower number is inclusive in group): 0-5, 5-10, 10-15, 15-20, 20-25, 25-30, 30+. The number of intercepts generally was less with increasing avidity. CPUE is similar across all avidity levels and years (Figure 4d).

COUNTY – As mentioned above, four counties were used in the analysis: Monroe, Dade, Broward, and Palm Beach. Monroe County had the most number of intercepts (7,890) followed by Palm Beach (4,003), Dade (2078), and Broward counties (1,055). Mean CPUE is higher in Monroe county than other counties, with mean CPUE in the other counties relatively equal but the trend in CPUE appears to be similar for all counties over time (Figure 4e).

MODE – Fishing mode was either charter boat or private/rental boat. There were more intercepts of the private/rental mode (8,939) than charter boat mode. CPUE was similar in both fishing modes up until 1997 when CPUE began to trend higher for charter boats than private/rental modes (Figure 4f).

WAVE – Two month wave were included in the analysis and the number of intercepts was fairly uniform across waves. Mean CPUE appears to similar over time across all waves (Figure 4g).

Standardization

CPUE was modeled using the delta-glm approach (Dick 2004; Lo et al. 1992; Maunder and Punt 2004) with R code provided by the SEFSC. This approach calculates an index as the product of the indices from binomial (presence/absence) and positive submodels. In this particular program, the response variable in the positive submodel can be defined by either lognormal or gamma distribution. To determine which distribution best described the data, I used the 'fitdistr' function of the MASS package in R to fit CPUE to the lognormal and gamma distributions (Figure 5). Positive CPUE of yellowtail snapper was described better by the lognormal distribution based on an improvement in AIC of 724 units. For both the positive and binomial submodels, explanatory variables were selected using stepwise forward selection based on AIC where k = log(n) rather than 2. This is thought to be more conservative than genuine AIC, especially when n is large, and usually retains only the variables that explain a large portion of the deviance. The final positive submodel included the terms year, county, and contributors (Table 1Table 2) while the binomial submodel included year, contributors, county, wave, mode, and area (Table 3Table 4). The least squared means for the year factor from each model were multiplied together with a bias correction applied to the positive CPUE to account for transformation of the response variable from log space back to CPUE.

Results

To evaluate residuals of the binomial model randomization was introduced to produce continuous normal residuals using the 'qres.binom' function of the 'statmod' package in R. Randomized quantile residuals for the binomial submodel were normally distributed and showed no pattern across predictor variables (Figure 6). Residuals from the positive submodel were also normal with no pattern across predictor variables (Figure 7). Diagnostic plots of the positive submodel indicate that residuals are normally distribution and exhibit no pattern, variance is homoscedastic, and there are no influential outliers in the dataset (Figure 8). The observed annual mean CPUE, modeled CPUE, and proportion of trips positive is provided in Table 5 and plotted in Figures 9 - 10.

References

Shertzer, K. W., and E. H. Williams. 2008. Fish assemblages and indicator species: reef fishes off the southeastern United States. Fishery Bulletin 106:257-269.

Table 1. Model selection steps for the positive submodel. The final model consists of year, county, and contributors.

Variable	Df	Deviance	AIC
log(response) ~ year			**19841**
cnty	3	7286	19262
cntrbtrs2	6	7263	19272
mode_fx	1	7660	19563
area2	1	7850	19719
none>		8014	19841
wave	5	7957	19844
hrsf2	6	7947	19846
avidity	6	7999	19887
log(response) ~ year + cnty			**19262**
cntrbtrs2	6	6770	18852
mode_fx	1	7210	19206
area2	1	7269	19257
none>		7286	19262
hrsf2	6	7244	19284
wave	5	7275	19302
avidity	6	7270	19307
log(response) ~ year + cnty + cntrbtrs2			**18852**
<none>		6770	18852
area2	1	6768	18860
mode_fx	1	6769	18861
avidity	6	6738	18880
hrsf2	6	6745	18887
wave	5	6758	18889

Table 2. Deviance table for final positive submodel.

Var	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)	PercDevExp
NULL	NA	NA	6374	8282	NA	NA
year	29	269	6345	8014	2.82E-37	17.75
cnty	3	728	6342	7286	2.16E-147	48.13
cntrbtrs2	6	516	6336	6769	3.73E-101	34.12

Table 3. Model selection steps for the binomial submodel. The final model consists of year, county, and contributors.

Variable	Df	Deviance	AIC
log(response) ~ year			**20601**
cntrbtrs2	6	19714	20061
cnty	3	20006	20324
mode_fx	1	20206	20504
area2	1	20230	20529
wave	5	20251	20588
hrsf2	6	20242	20589
none>		20312	20601
avidity	6	20301	20648
response ~ year + cntrbtrs2			**20061**
cnty	3	19569	19945
area2	1	19690	20046
wave	5	19655	20050
<none>		19714	20061
mode_fx	1	19713	20070
hrsf2	6	19681	20085
avidity	6	19706	20110
response ~ year + cntrbtrs2 + cnty			**19945**
wave	5	19500	19924
mode_fx	1	19546	19931
<none>		19569	19945
area2	1	19563	19948
hrsf2	6	19534	19968
avidity	6	19551	19984
response ~ year + cntrbtrs2 + cnty + wave			**19924**
mode_fx	1	19476	19910
<none>		19500	19924
area2	1	19494	19928
hrsf2	6	19465	19947
avidity	6	19482	19964
response ~ year + cntrbtrs2 + cnty + wave + mode_fx			**19910**
area2	1	19463	19906
<none>		19476	19910
hrsf2	6	19439	19930
avidity	6	19466	19957
response ~ year + cntrbtrs2 + cnty + wave + mode_fx + area2			**19906**
<none>		19463	19906
hrsf2	6	19429	19930
avidity	6	19453	19954

Table 4. Deviance table for final binomial submodel.

Variable	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)	PercDevExp
NULL	NA	NA	15025	20484	NA	NA
year	29	173	14996	20312	2.27E-22	16.90
cntrbtrs2	6	597	14990	19714	8.19E-126	58.49
cnty	3	145	14987	19569	3.21E-31	14.19
wave	5	69	14982	19500	1.33E-13	6.80
mode_fx	1	23	14981	19476	1.28E-06	2.30
area2	1	14	14980	19463	2.31E-04	1.33

Table 5. Nominal mean CPUE and final modeled index.

Year	**Nominal CPUE**	**N**	**Proportion Positive**	**Index**	**Index CV**
1981	4.109	92	0.500	3.901	0.150
1982	2.722	194	0.402	3.675	0.196
1983	1.771	131	0.328	2.960	0.161
1984	1.910	189	0.402	3.307	0.199
1985	1.422	135	0.281	2.627	0.159
1986	1.788	231	0.346	3.525	0.132
1987	2.067	372	0.336	2.786	0.142
1988	2.328	293	0.392	3.809	0.145
1989	2.154	267	0.318	3.787	0.131
1990	2.227	256	0.551	4.587	0.112
1991	4.892	278	0.518	7.183	0.094
1992	3.264	550	0.522	6.113	0.097
1993	3.210	499	0.473	4.819	0.106
1994	3.265	423	0.487	4.578	0.121
1995	3.572	339	0.407	5.179	0.117
1996	3.332	404	0.455	4.048	0.120
1997	4.085	424	0.394	4.408	0.118
1998	4.203	536	0.384	4.066	0.105
1999	3.477	792	0.365	4.397	0.106
2000	3.308	672	0.408	4.016	0.108
2001	3.977	665	0.350	4.168	0.104
2002	3.568	911	0.393	3.710	0.100
2003	3.956	895	0.397	4.407	0.099
2004	4.048	831	0.432	5.125	0.096
2005	4.392	738	0.466	5.325	0.094
2006	4.010	762	0.454	5.296	0.090
2007	4.386	888	0.453	5.436	0.090
2008	4.296	948	0.465	4.797	0.100
2009	3.164	623	0.424	4.284	0.099
2010	4.792	688	0.458	5.140	0.195

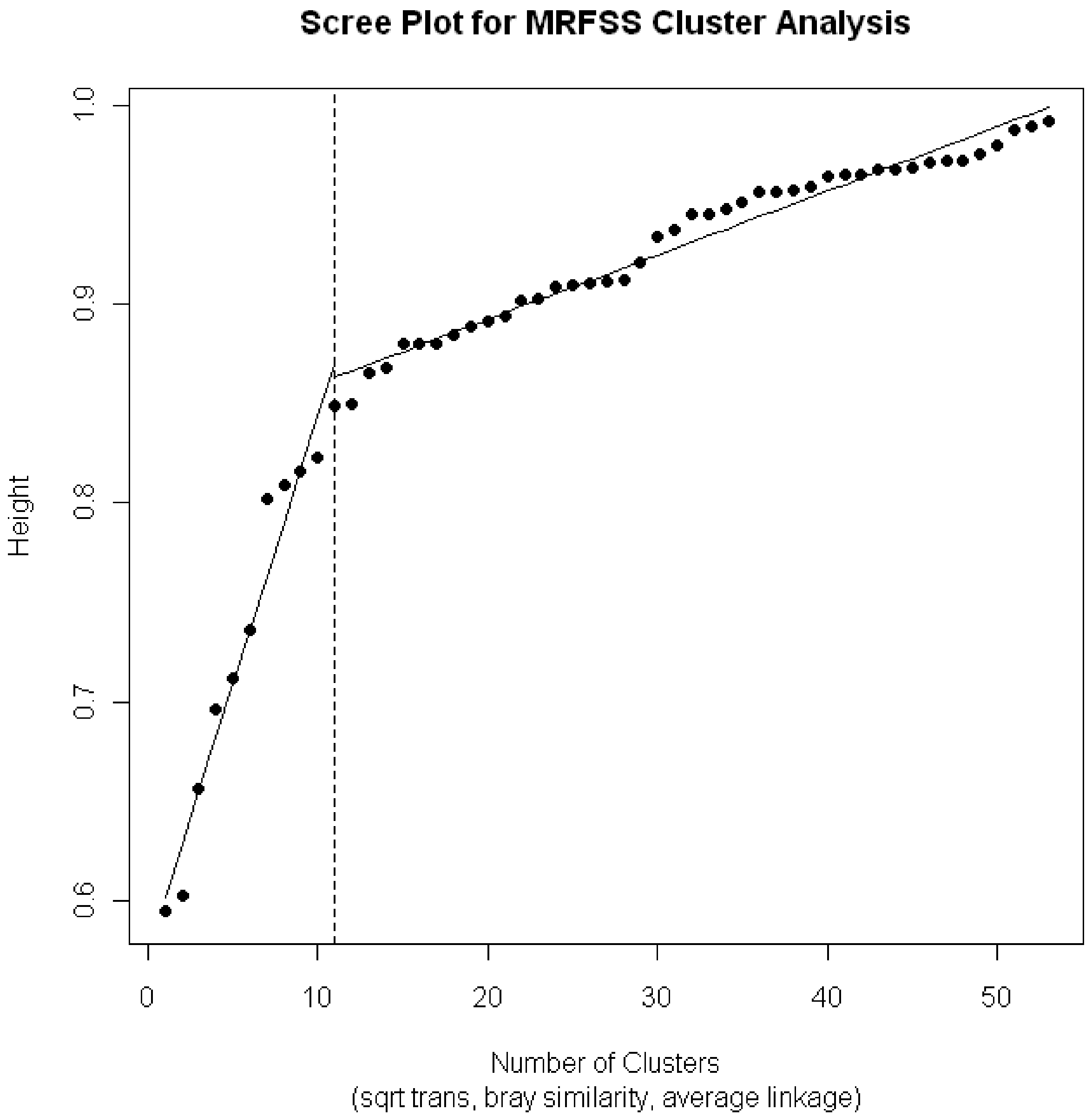

Figure 1. Plot of number of clusters against height from the hierarchical cluster analysis, where height is the average dissimilarity among species in a cluster with 1 being most similar. Included are the piecewise regression lines (solid line) using a breakpoint (dashed vertical line) that minimized the residual mean square error.

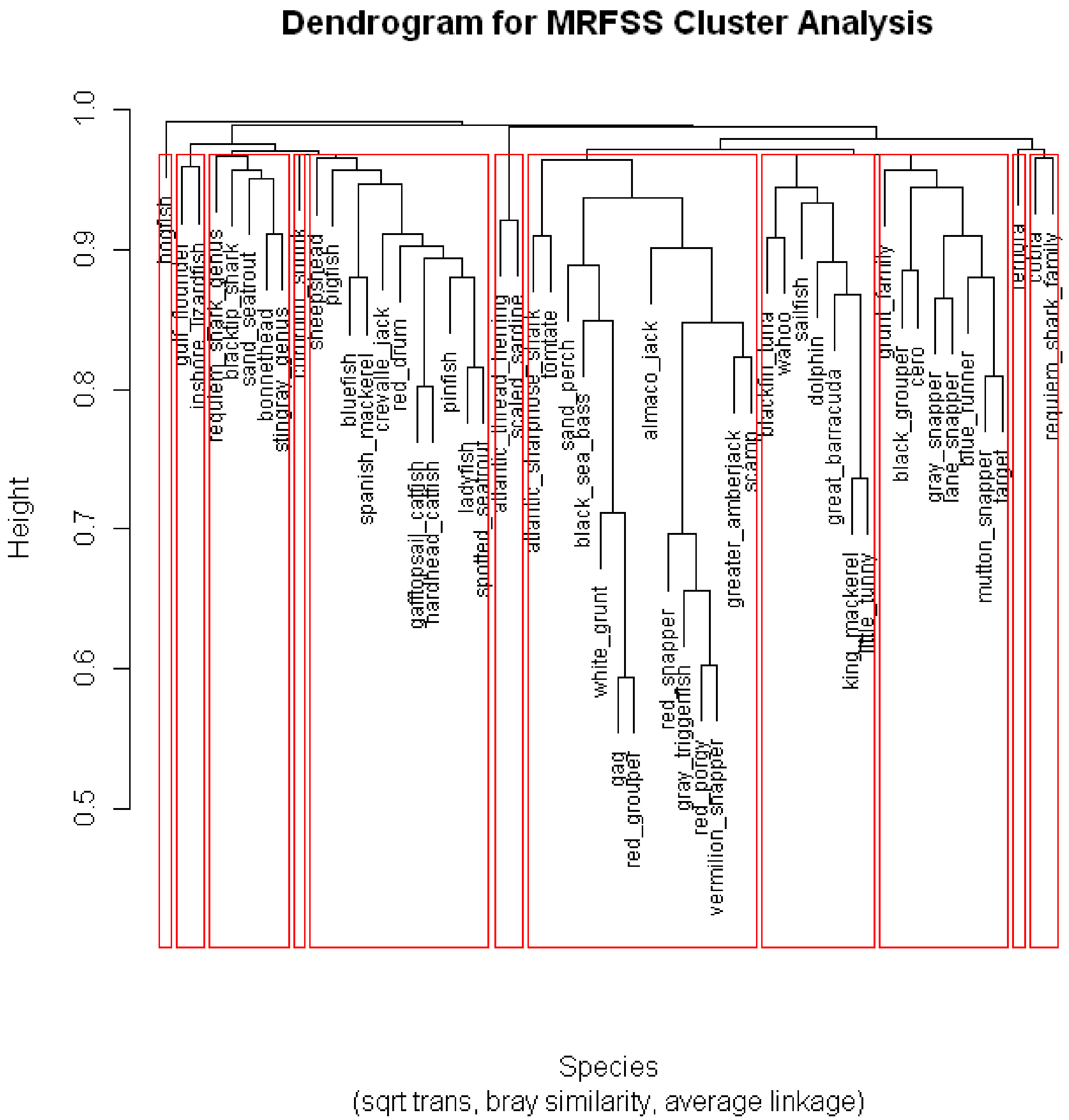

Figure 2. Dendrogram from hierarchical cluster analysis of species in the MRFSS data. Height measures the average dissimilarity among species within a branch with a value of 1 being most similar. Yellowtail snapper is represented as ‘target’ in this plot.

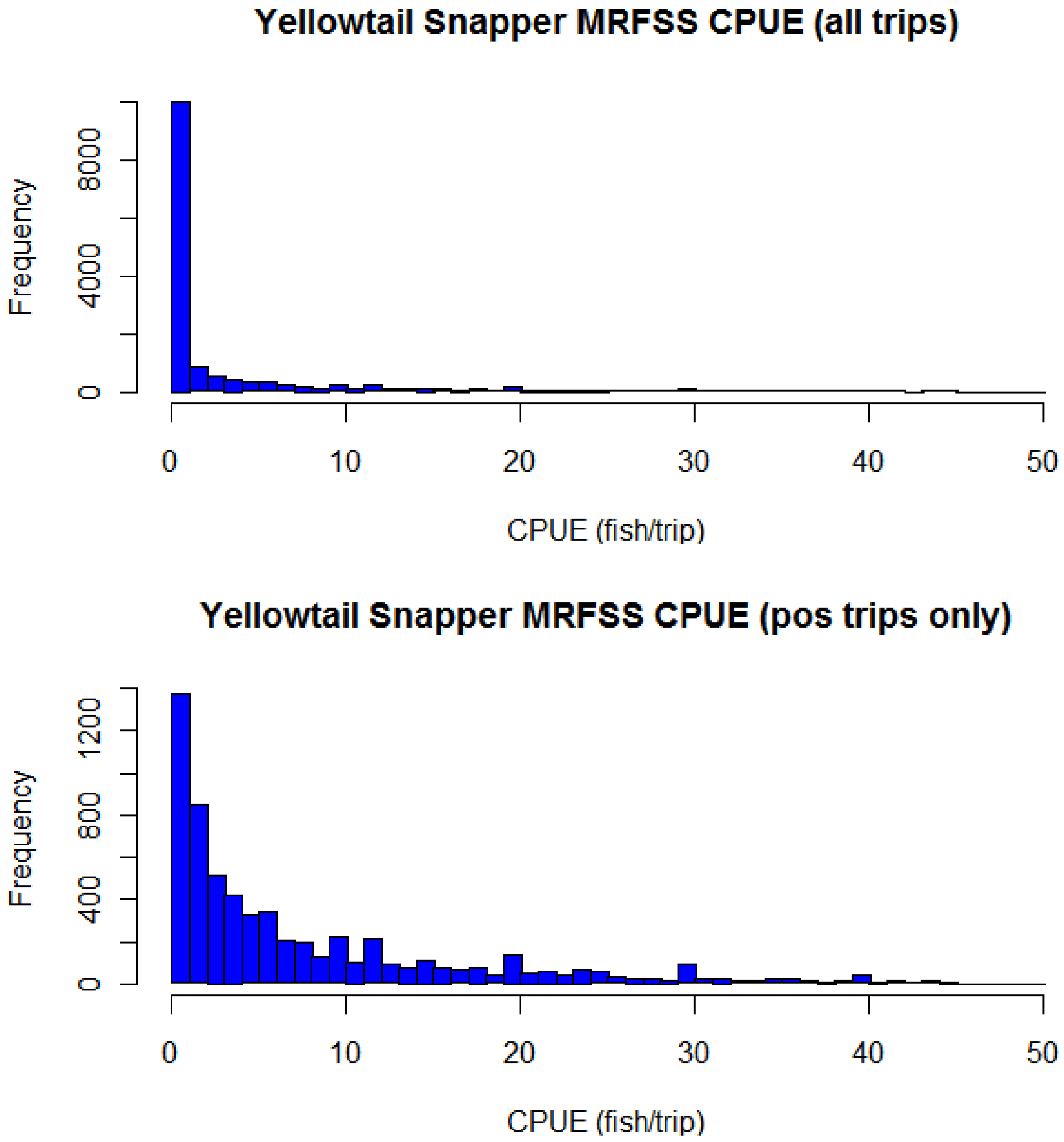

Figure 3. Distribution of CPUE of yellowtail snapper from MRFSS for all trips and positive trips only after applying filters and subsettings.

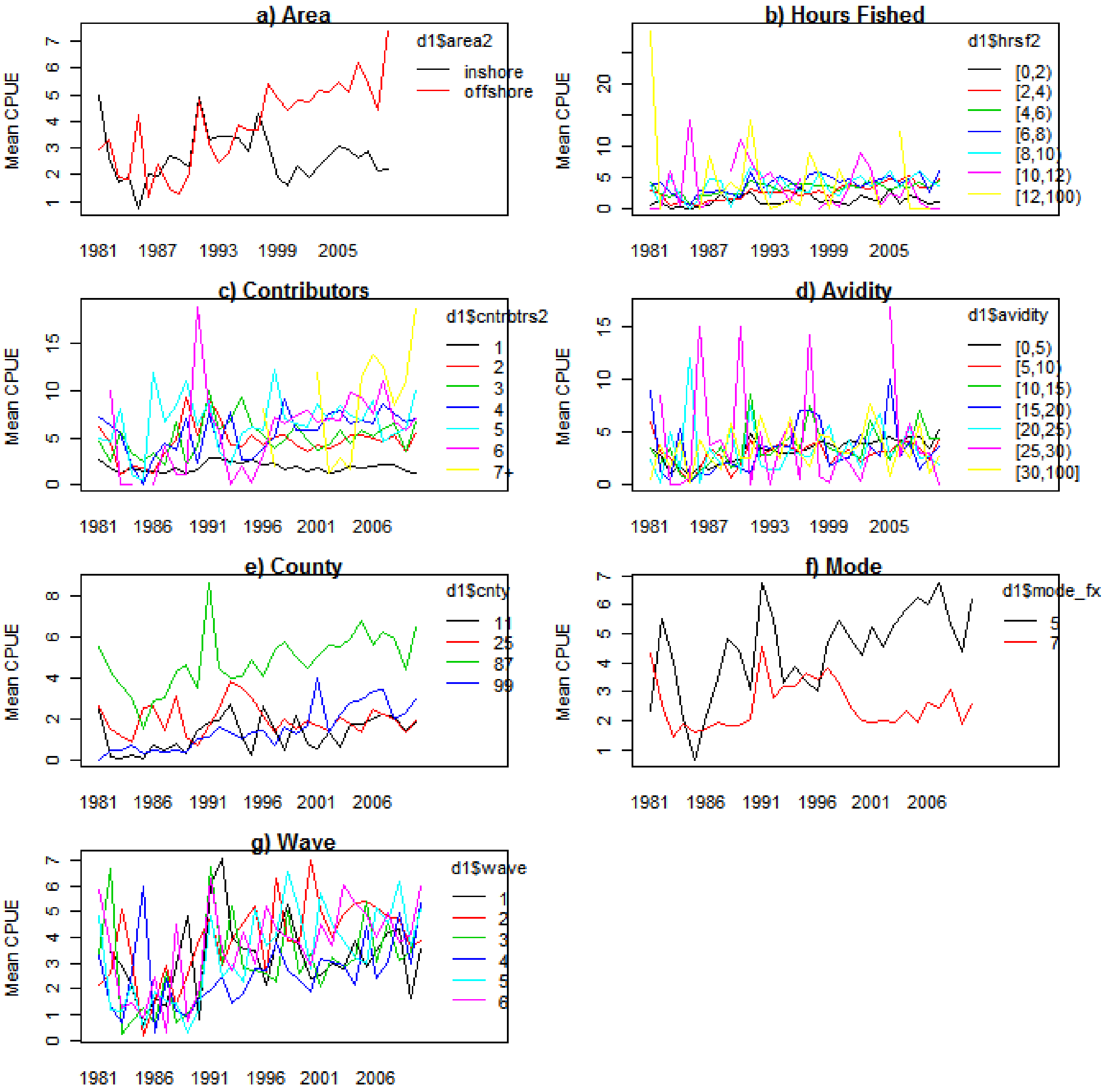

Figure 4. Interaction plots of year and each predictor variable on CPUE.

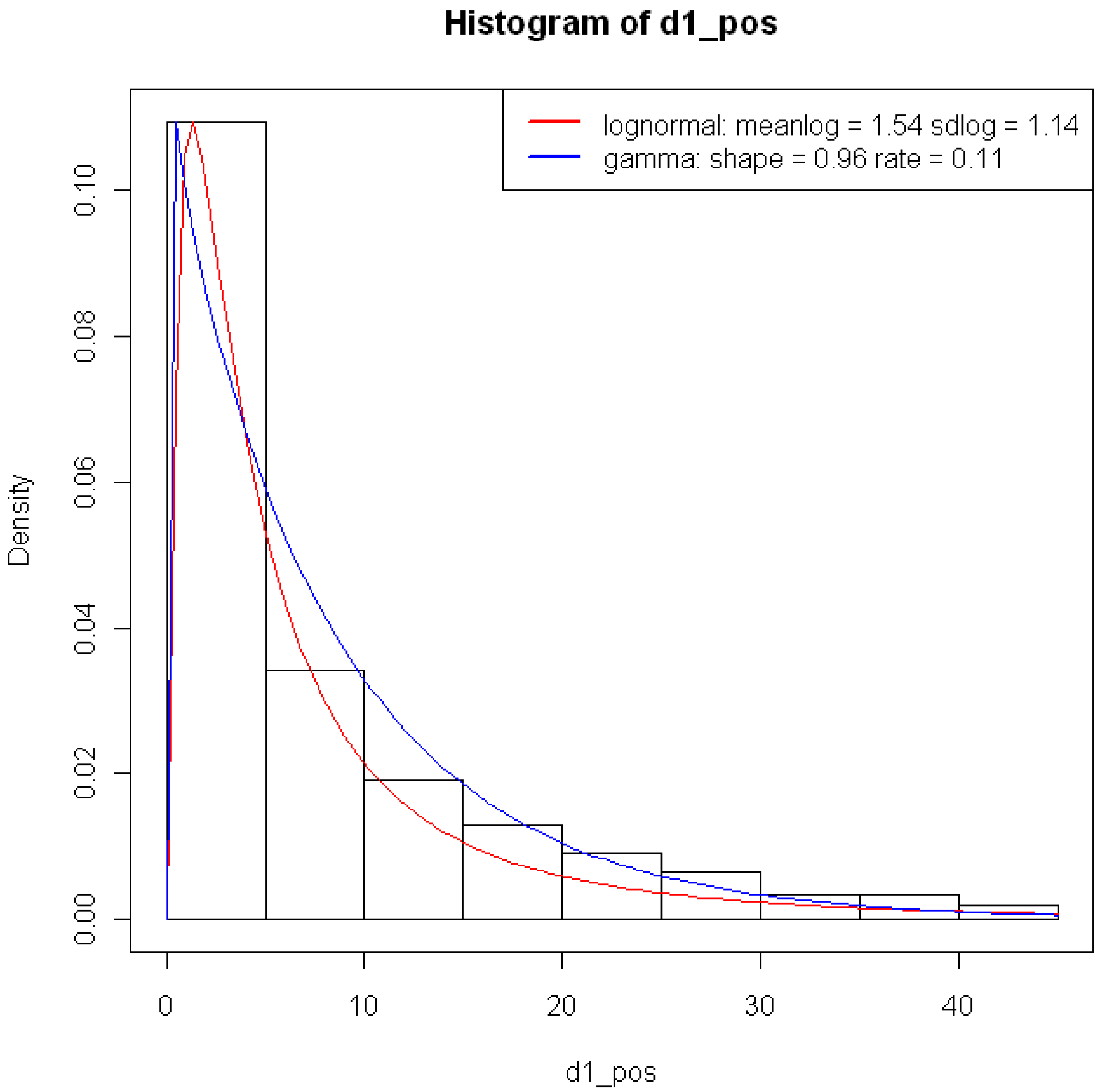

Figure 5. Fit of CPUE to the gamma and lognormal distributions.

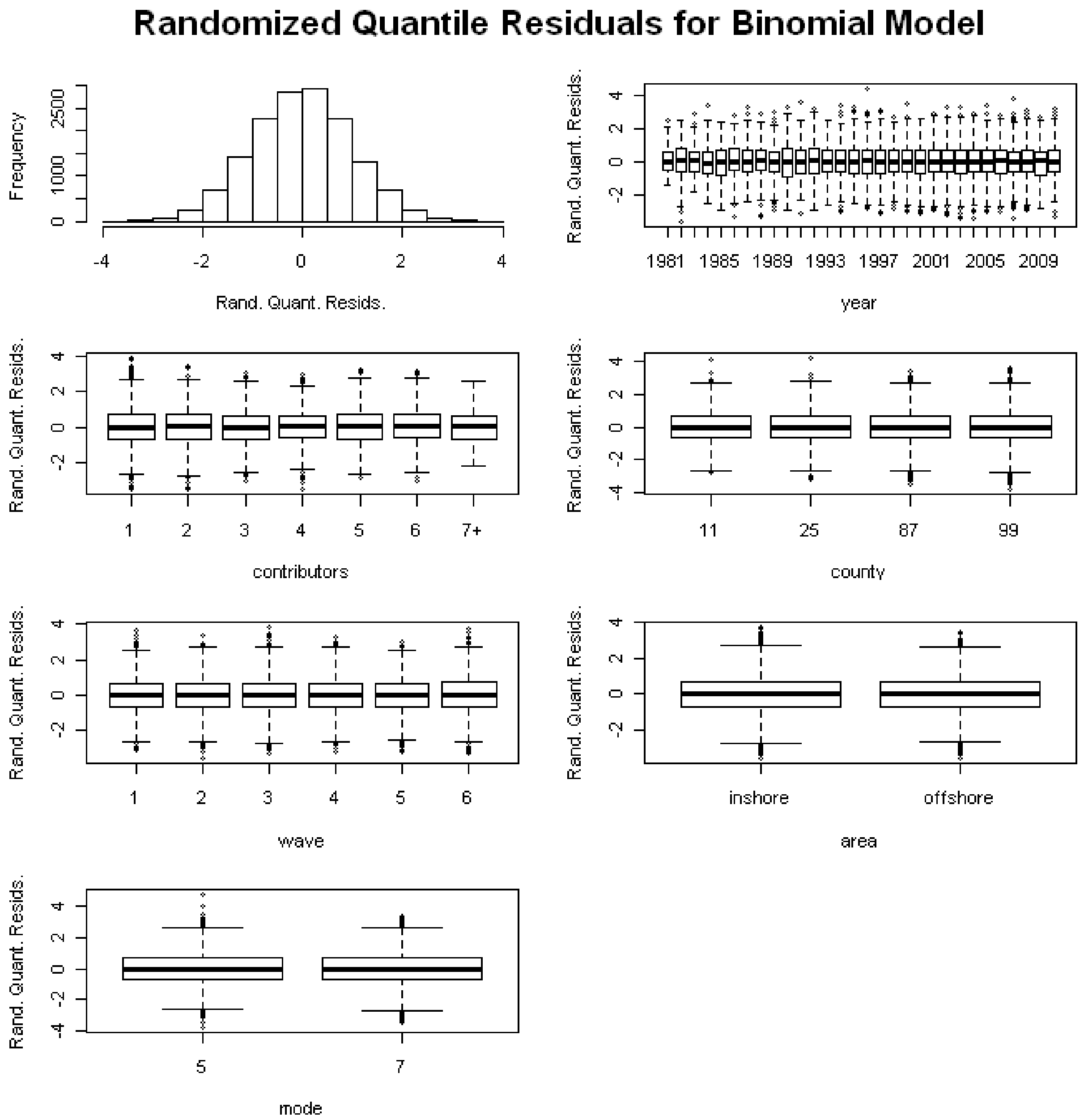

Figure 6. Standardized residuals for the binomial submodel.

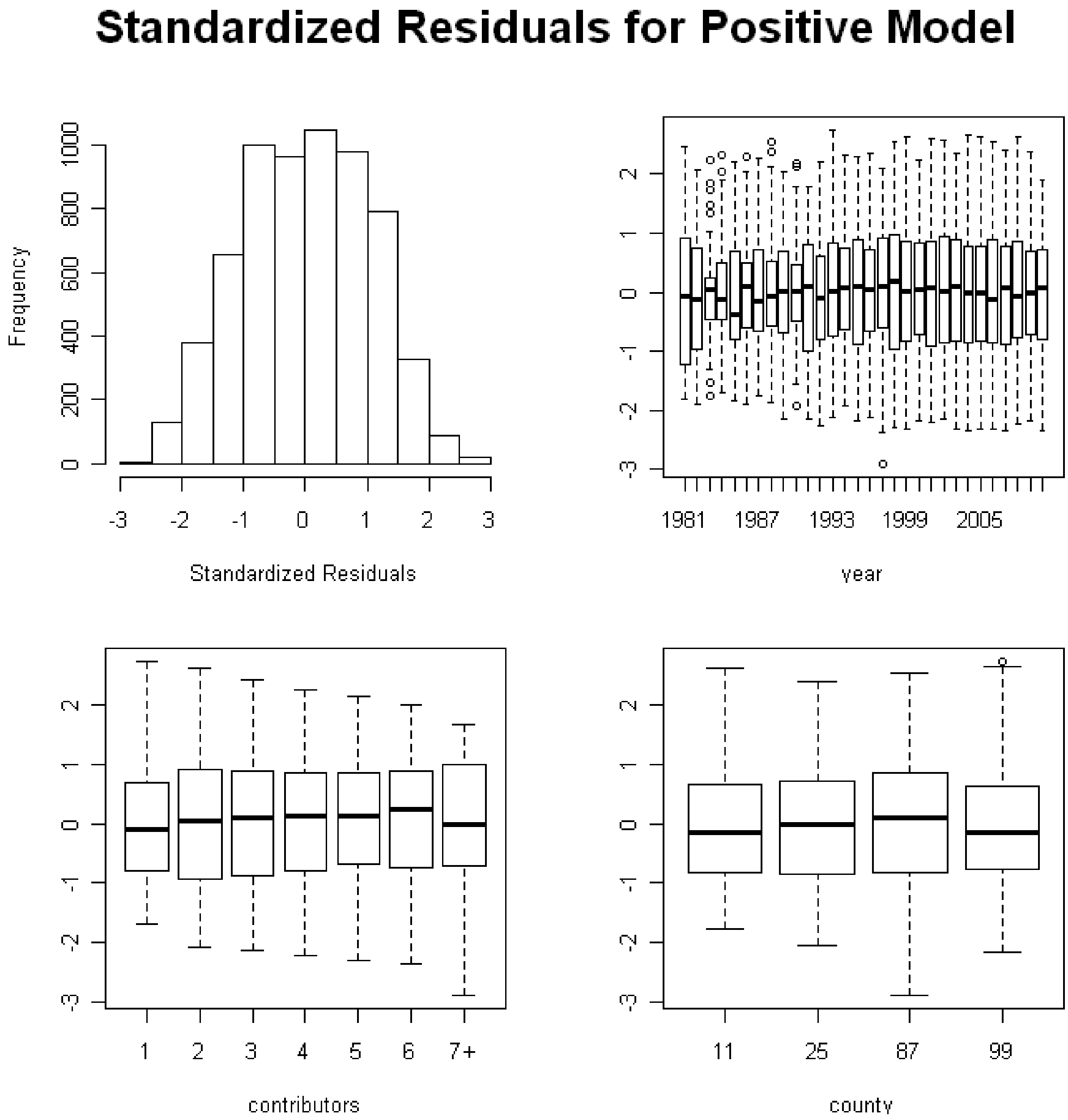

Figure 7. Standardized residuals for the positive submodel.

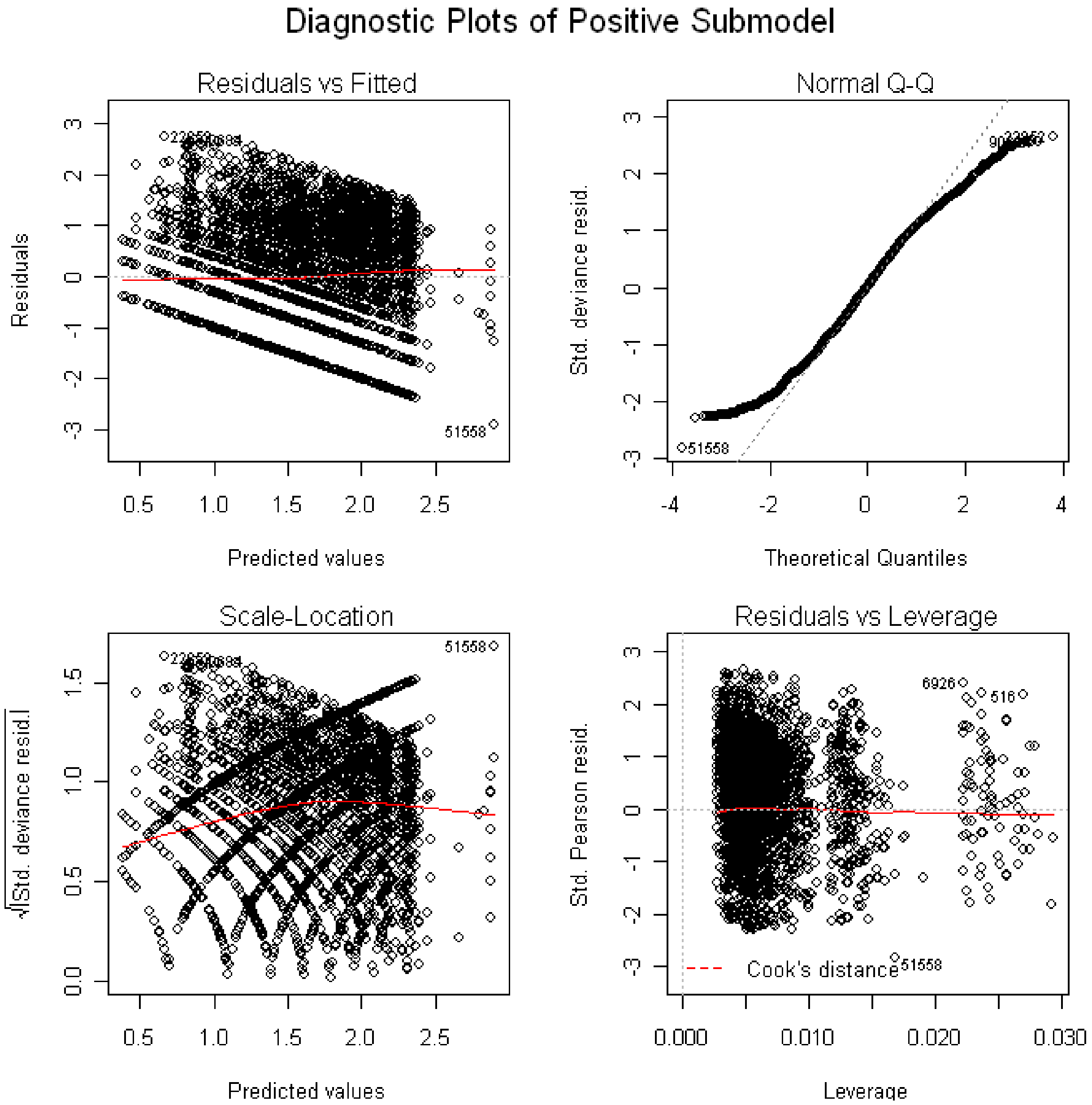

Figure 8. Diagnostic plots for the positive submodel.

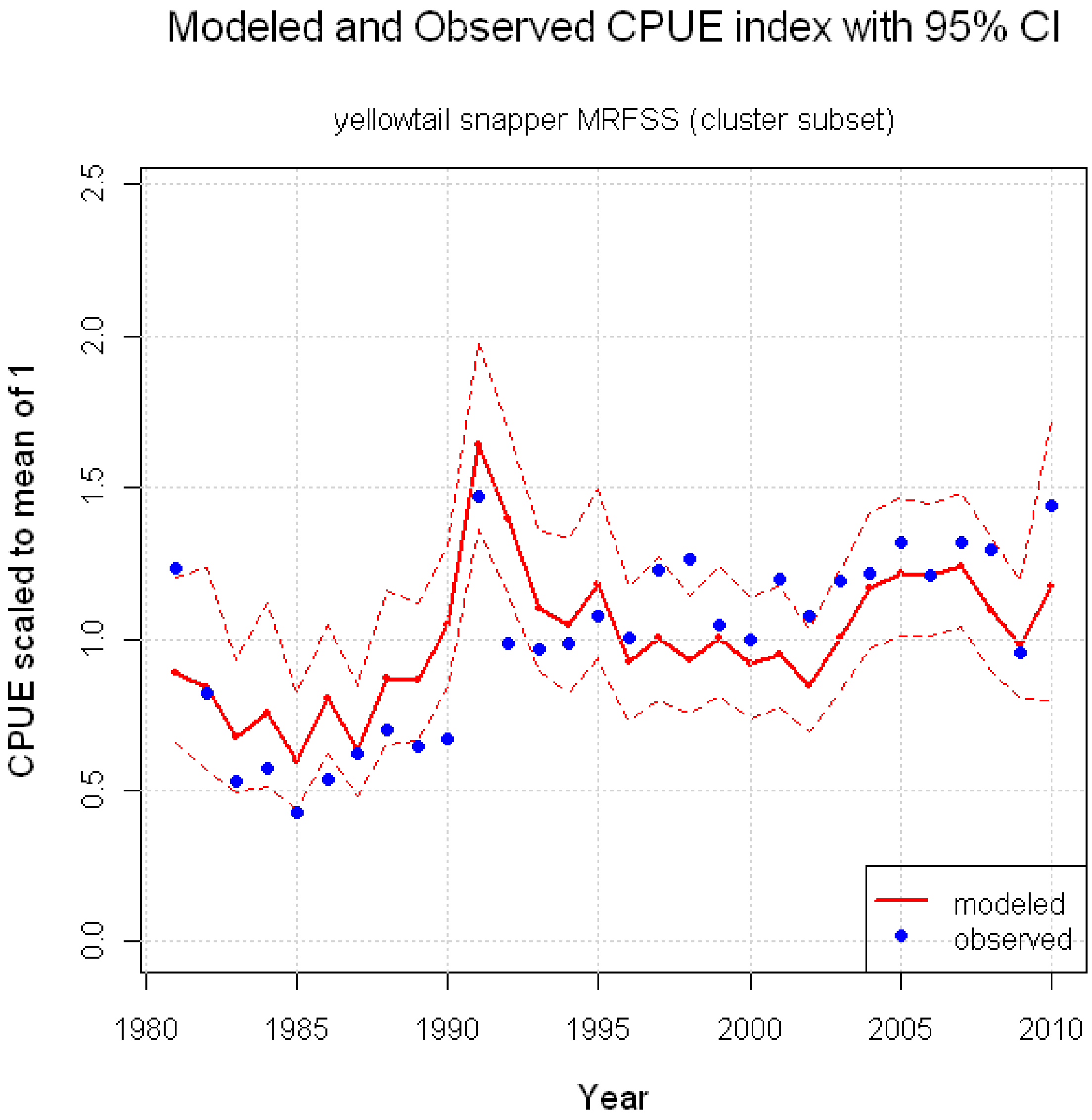

Figure 9. Modeled and observed CPUE, scaled to mean = 1, of yellowtail snapper from MRFSS data.

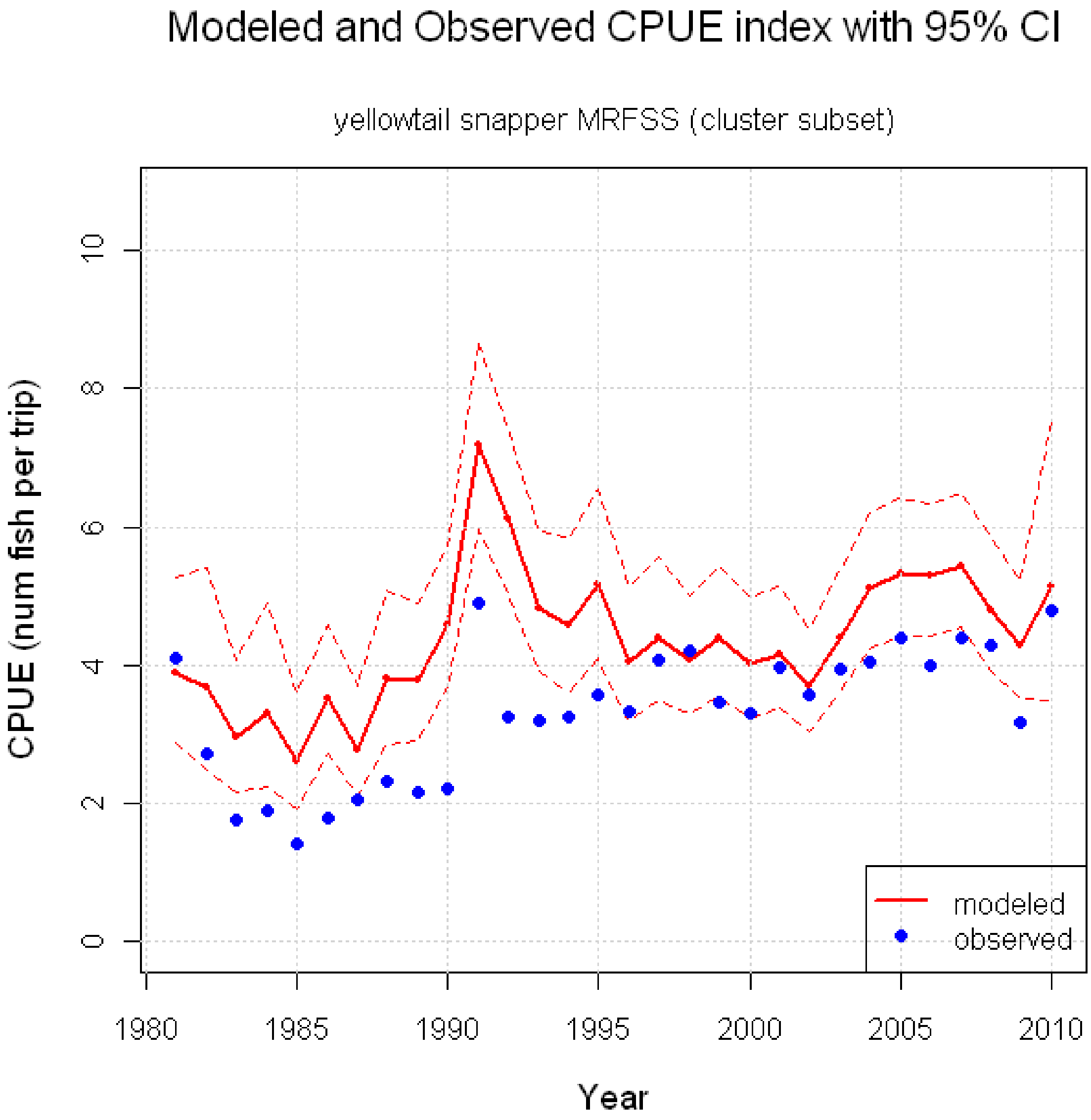

Figure 10. Modeled and observed CPUE (number of fish per trip) of yellowtail snapper from the MRFSS data.