Notice on SEDAR Working Papers

Standardized catch rates of U.S. Atlantic red snapper (*Lutjanus campechanus*) from commercial logbook data

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

3 May 2010

1. Introduction

Landings and fishing effort of commercial vessels operating in the southeast U.S. Atlantic have been monitored by the NMFS Southeast Fisheries Science Center through the Coastal Fisheries Logbook Program (CFLP). The program collects information about each fishing trip from all vessels holding federal permits to fish in waters managed by the Gulf of Mexico and South Atlantic Fishery Management Councils. Initiated in the Gulf in 1990, the CFLP began collecting logbooks from Atlantic commercial fishers in 1992, when 20% of Florida vessels were targeted. Beginning in 1993, sampling in Florida was increased to require reports from all vessels permitted in coastal fisheries, and since then has maintained the objective of a complete census of federally permitted vessels in the southeast U.S.

Catch per unit effort (CPUE) from the logbooks was used to develop an index of abundance for red snapper landed with vertical lines (manual handline and electric reel), the dominant gear for this red snapper stock. Thus, the size and age range of fish included in the index is the same as that of landings from this same fleet. The time series used for construction of the index spanned 1993−2009, when all vessels with federal snapper-grouper permits were required to submit logbooks on each fishing trip.

2. Data and treatment

2.1 Available Data

For each fishing trip, the CFLP database included a unique trip identifier, the landing date, fishing gear deployed, areas fished, number of days at sea, number of crew, gear-specific fishing effort, species caught, and weight of the landings (reported fields described in Appendix 1). Fishing effort data available for vertical line gear included number of lines fished, hours fished, and number of hooks per line. For this southeast U.S. Atlantic stock, areas used in analysis were those between 24 and 36 degrees latitude, inclusive of the boundaries (Figure 1).

Data were restricted to include only those trips with landings and effort data reported within 45 days of the completion of the trip (some reporting delays were longer than one year). Reporting delays beyond 45 days likely resulted in less reliable effort data (landings data may be reliable even with lengthy reporting delays if trip ticket reports were referenced by the reporting fisher). This restriction excluded ~22% of the full data set (i.e., the data set with all gears and all areas, including Gulf of Mexico). Also excluded were records reporting multiple areas or gears fished, which prevents designating catch and effort to specific locations or gears. Therefore, only trips which reported one area and one gear fished were included in these analyses.

Clear outliers in the data, e.g. values falling outside the 99.5 percentile of the data, were also excluded from the analyses. These outliers were identified for manual handlines as records reporting more than 8 lines fished, 8 hooks per line fished, 8 days at sea, or 4 crew members, and they were identified for electric reels as records reporting more than 5 lines fished, 8 hooks per line fished, 11 days at sea, or 5 crew members. Also omitted were any vertical line records reporting more than 24 hours fishing per day at sea. Together, these restrictions removed ~4.7% of records.

2.2 Subsetting

Effective effort was based on those trips from areas where red snapper were available to be caught. Without fine-scale geographic information on fishing location, trips to be included in the analysis must be inferred, which was done here using the method of Stephens and MacCall (2004). The method uses multiple logistic regression to estimate a probability for each trip that the focal species was caught, given other species caught on that trip. Because a zoogeographic boundary is apparent near Cape Canaveral (Shertzer et al., 2009), the method was applied separately to data from regions north and south of 28 degrees latitude (near Cape Canaveral). To avoid undue influence of rare species on regression estimates, species included in each analysis were limited to those occurring in 1% or more of trips (Table 1). Red porgy was also omitted because of strict harvest regulations since 1999 (including a temporary moratorium), which creates the potential for erroneously removing trips likely to have caught red snapper during years of red porgy restrictions. A backwards stepwise AIC procedure (Venables and Ripley, 1997) was then used to perform further selection among possible species as predictor variables, where the most general model included all listed species as main effects. In this procedure, a generalized linear model with Bernoulli response was used to relate presence/absence of red snapper in each trip to presence/absence of other species. For the northern sampling area (NC, SC, GA, north FL), stepwise AIC eliminated mutton snapper and sand tilefish; for the southern sampling area (south FL), it eliminated black grouper and almaco jack. Regression coefficients of included species for the northern sampling areas are shown in Figure 2, and for the southern areas in Figure 3.

A trip was then included if its associated probability of catching red grouper was higher than a threshold probability (Figures 4, 5). The threshold was defined to be that which resulted in the same number of predicted and observed positive trips, as suggested by Stephens and MacCall (2004). After applying the Stephens and MacCall method, and the constraints described above, the resulting subsetted data set contained 17,692 trips in the northern sampling areas, of which ~63% were positive, and 2,603 trips from the southern sampling area, of which ~35% were positive.

3. Standardization

CPUE was modeled using the delta-GLM approach (Lo et al., 1992; Dick, 2004; Maunder and Punt, 2004). This approach combines two separate generalized linear models (GLMs), one to describe presence/absence of the focal species, and one to describe catch rates of successful trips (trips that caught the focal species). The response variable, CPUE, was calculated for each trip as,

CPUE = pounds of red snapper/hook-hours

where hook-hours is the product of number of lines fished, number of hooks per line, and total hours fished. Explanatory variables, all categorical, are described below. Estimates of variance were based on the jackknife "leave one out" estimator. All analyses were programmed in R, with much of the code adapted from Dick (2004).

3.1 Explanatory variables considered

YEAR — Year was necessarily included, as standardized catch rates by year are the desired outcome. Years modeled were 1993−2009. The total number of red snapper trips by year, as well as proportion positive by year, is provided in Table 1.

SEASON — Season included four levels: winter (January−March), spring (April−June), summer (July−September), and fall (October−December). The number of trips per year by season, as well as proportion positive, is shown in Figure 6.

AREA — Areas reported in the logbook (Figure 1) were pooled into the broader geographic levels: North Carolina (NC), South Carolina (SC), Georgia (GA), north Florida (NF), and south Florida (SF). The break between north and south Florida occurred at 28 degrees latitude, near Cape Canaveral, which has been identified as a zoogeographical boundary (Shertzer et al., 2009). The number of trips per year by area, as well as proportion positive, is shown in Figure 7.

DAYS AT SEA — Days at sea (sea days) were pooled into three levels: one day (one), two to four days (twotofour), and five or more days (fiveplus). The number of trips per year by sea days, as well as proportion positive, is shown in Figure 8.

CREW SIZE — Crew size (crew) was pooled into three levels: one (one), two (two), and three or more days (threeplus). The number of trips per year by crew, as well as proportion positive, is shown in Figure 9.

3.2 Bernoulli submodel

The bernoulli component of the delta-GLM is a logistic regression model that attempts to explain the probability of either catching or not catching red snapper on any given trip. Initially, all explanatory variables were included in the model as main effects, and then stepwise AIC (Venables and Ripley, 1997) with a backwards selection algorithm was used to eliminate those variables that did not improve model fit. In this case, the stepwise AIC procedure did not remove any explanatory variables (Table 2). Diagnostics, based on randomized quantile residuals (Dunn and Smyth, 1996), suggested reasonable fits of the Bernoulli submodel (Figure 10).

3.3 Positive CPUE submodel
Two parametric distributions were considered for modeling positive values of CPUE, lognormal and gamma. For both distributions, all explanatory variables were initially included as main effects, and then stepwise AIC (Venables and Ripley, 1997) with a backwards selection algorithm was used to eliminate those variables that did not improve model fit. For both lognormal and gamma distributions, the best model fit included all explanatory variables (lognormal shown in Table 2). The two distributions, each with their best set of explanatory variables (all of them), were compared using AIC: lognormal highly outperformed gamma (ΔAIC>1000), and was therefore applied in the final delta-GLM. Diagnostics suggested reasonable fits of the lognormal submodel (Figures 11, 12).

Results

Nominal CPUE averaged across areas tracked more closely the nominal CPUEs of GA, north FL, and south FL, than it did the nominal CPUEs of NC and SC (Figure 13). The standardized index has fluctuated substantially throughout the time series, with a maximum value equal to 3.4 times the minimum (Figure 14, Table 1). Over the past four years, the pattern has been one of strict increase, culminating in the highest expected value of the full series.

Literature cited

Dick, E.J. 2004. Beyond 'lognormal versus gamma': discrimination among error distributions for generalized linear models. Fish. Res. 70:351−366.

Dunn, K. P. and Smyth, G. K. 1996. Randomized quantile residuals. J Comp. Graph. Stat. 5:1−10.

Lo, N.C., Jacobson, L.D., Squire, J.L. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Can. J. Fish. Aquat. Sci. 49:2515−2526.

Maunder, M.N., Punt, A.E. 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70:141−159.

Shertzer, K.W., E.H. Williams, and J.C. Taylor. 2009. Spatial structure and temporal patterns in a large marine ecosystem: Exploited reef fishes of the southeast United States. Fish. Res. 100:126−133.

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

Venables, W. N. and B. D. Ripley. 1997. Modern Applied Statistics with S-Plus, 2nd Edition. Springer-Verlag, New York.

Table 1. Standardized index of red snapper from commercial logbook data.

Year	Relative Nominal CPUE	N	Proportion N Positive	Relative Standardized Index	CV (Index)
1993	0.885	843	0.708	1.137	0.060
1994	0.764	1357	0.704	0.914	0.048
1995	0.810	1528	0.656	0.922	0.047
1996	0.496	1240	0.582	0.573	0.056
1997	0.484	1479	0.546	0.567	0.059
1998	0.508	1365	0.495	0.632	0.058
1999	0.644	1172	0.520	0.756	0.062
2000	0.774	1160	0.521	0.745	0.060
2001	1.197	1381	0.663	1.218	0.050
2002	1.090	1430	0.706	1.365	0.047
2003	1.036	1178	0.626	1.111	0.054
2004	1.334	1059	0.630	1.440	0.053
2005	1.105	1068	0.582	1.228	0.060
2006	0.543	950	0.483	0.608	0.068
2007	0.665	1123	0.477	0.664	0.066
2008	1.976	1013	0.560	1.201	0.068
2009	2.688	948	0.631	1.918	0.073

Table 2. Model selection results from delta-lognormal model

A. Bernoulli submodel

Removed	Df	Deviance	AIC
<none>		23649	23705
- crew	2	23680	23732
- season	3	23753	23803
- days	2	24104	24156
- year	16	24148	24172
- area	4	25165	25213

B. Lognormal submodel

Removed	Df	Deviance	AIC
<none>		19111	39895
- season	3	19318	40019
- crew	2	19773	40302
- year	16	20026	40428
- area	4	20993	41022
- days	2	21470	41297

Figure 1. Areas reported in commercial logbooks. First two digits signify degrees latitude, second two degrees longitude. Areas were excluded from the analysis if north of 36 degrees latitude or if in the Gulf of Mexico. In analyses, south Florida was treated separately from north Florida, with the boundary occurring at 28 degrees latitude (break near Cape Canaveral; boundary included in the south).

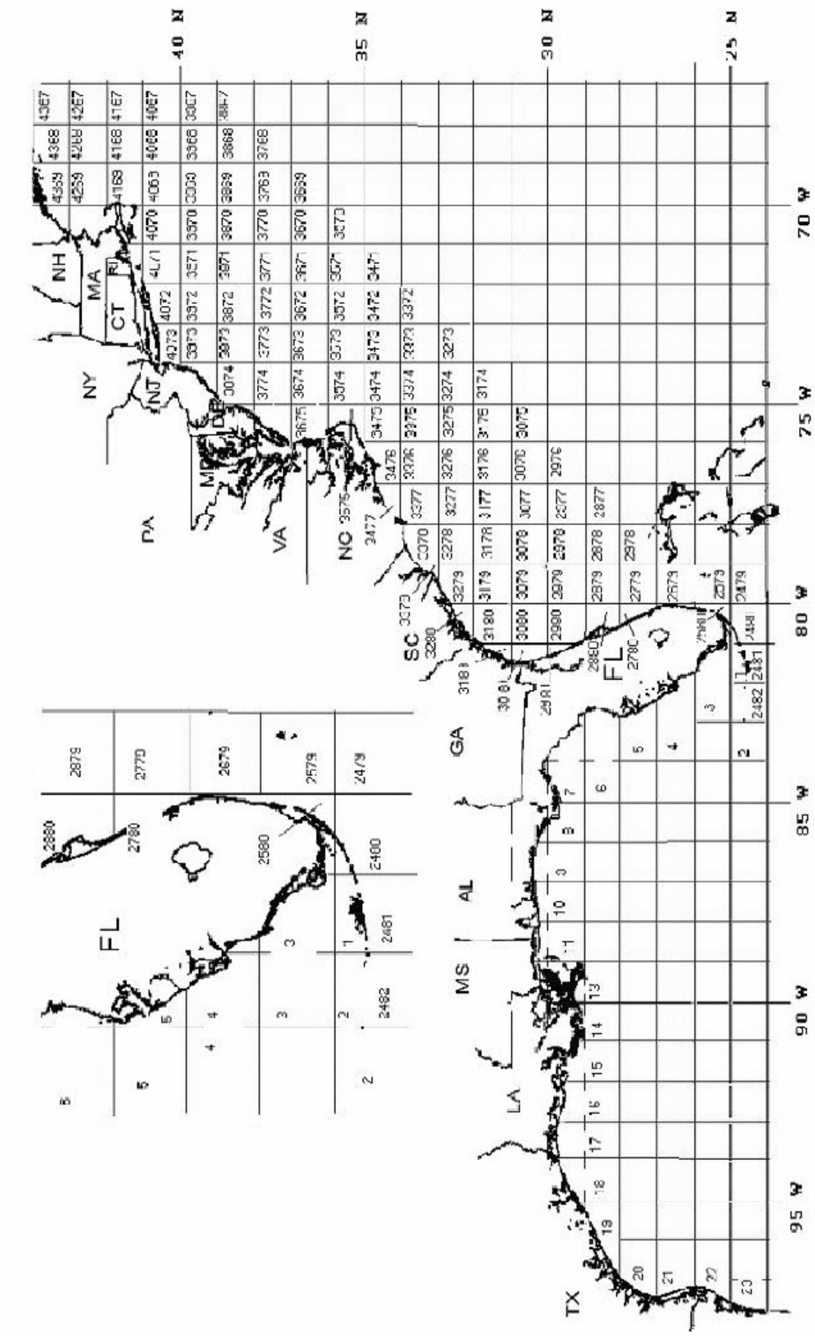


Figure 2. Estimates of species-specific regression coefficients from Stephens and MacCall method applied to logbook data from areas in the northern region (NC, SC, GA, north FL), as used to estimate each trip's probability of catching the focal species.

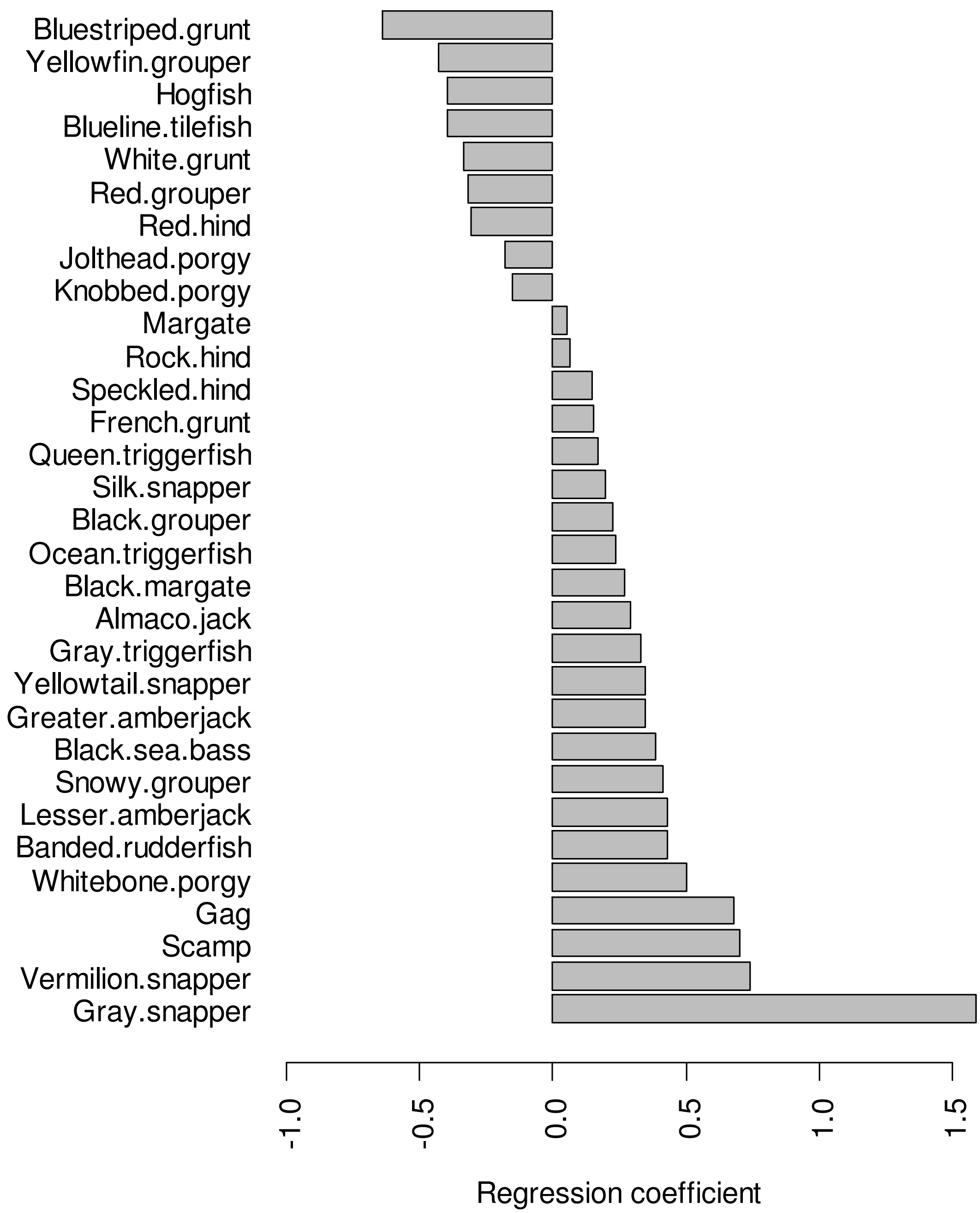


Figure 3. Estimates of species-specific regression coefficients from Stephens and MacCall method applied to logbook data from areas in the southern region (south FL), as used to estimate each trip's probability of catching the focal species.

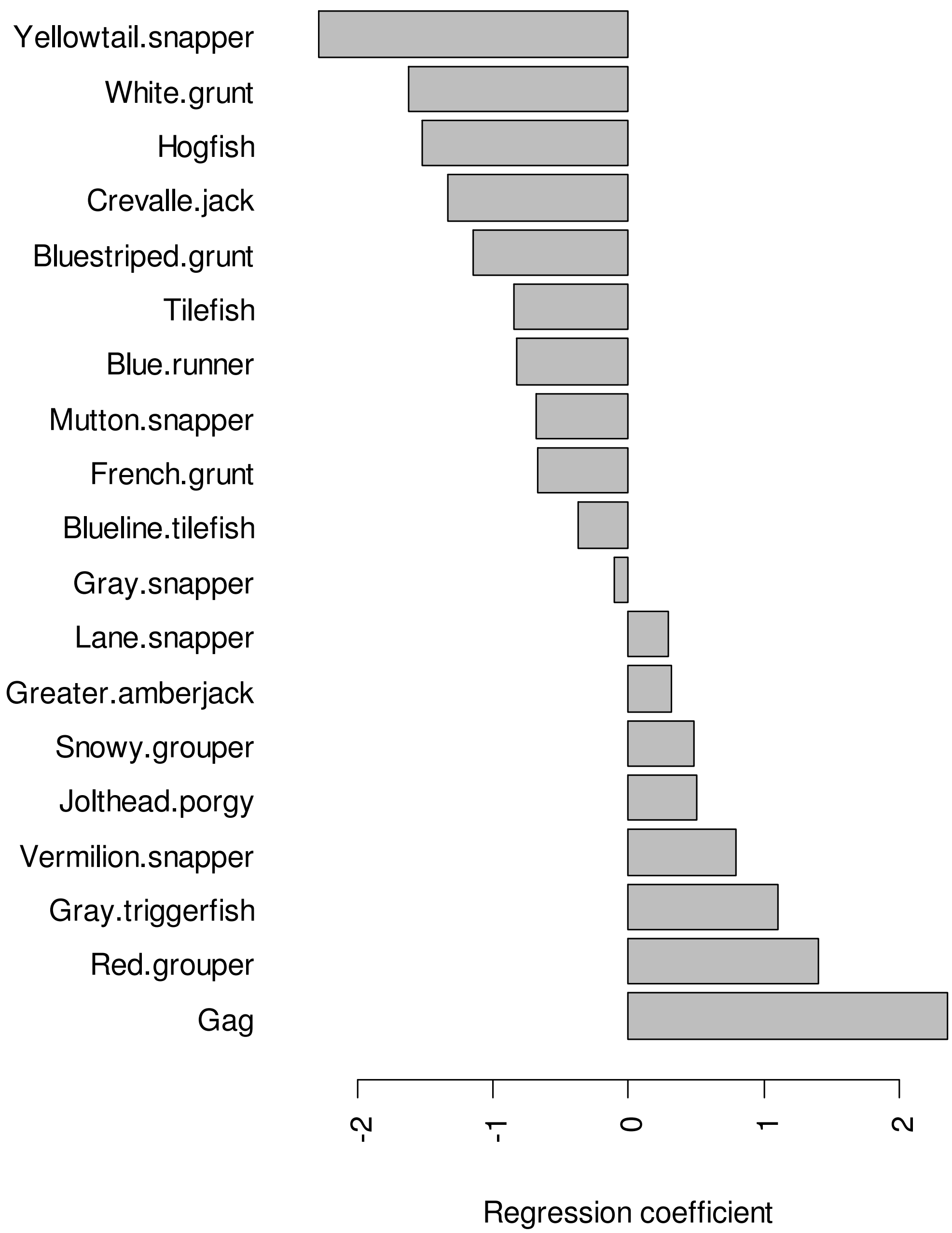


Figure 4. Absolute difference between observed and predicted number of positive trips from Stephens and MacCall method applied to logbook data from the northern region (NC, SC, GA, north FL). Left and right panels differ only in the range of probabilities shown.

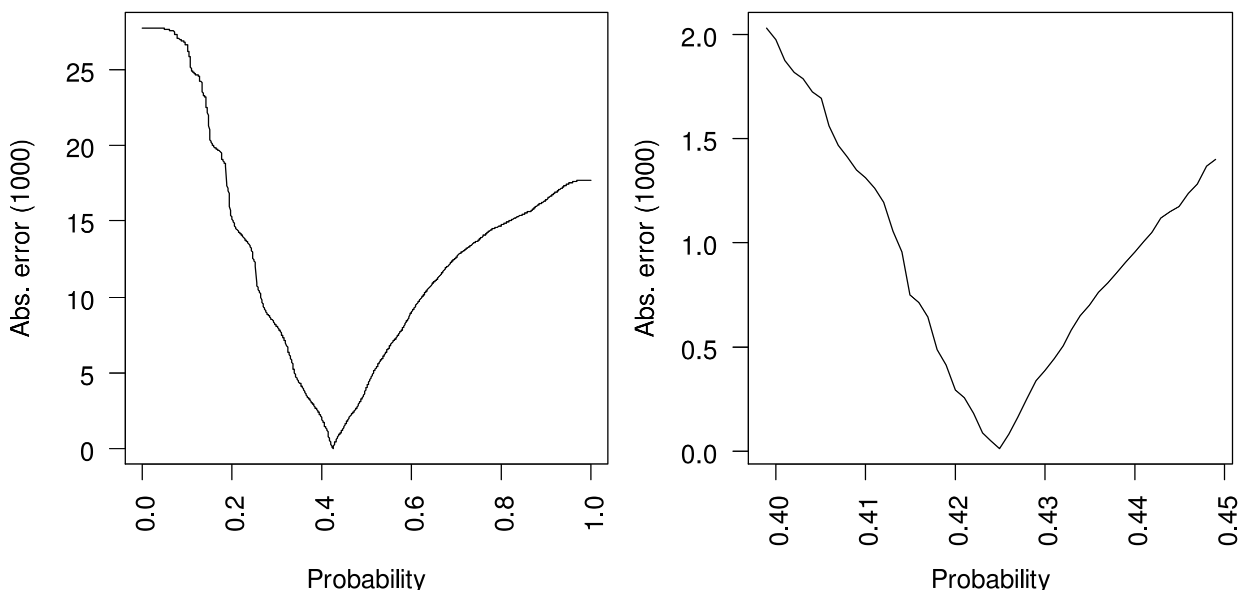


Figure 5. Absolute difference between observed and predicted number of positive trips from Stephens and MacCall method applied to headboat data from the southern region (south FL). Left and right panels differ only in the range of probabilities shown.

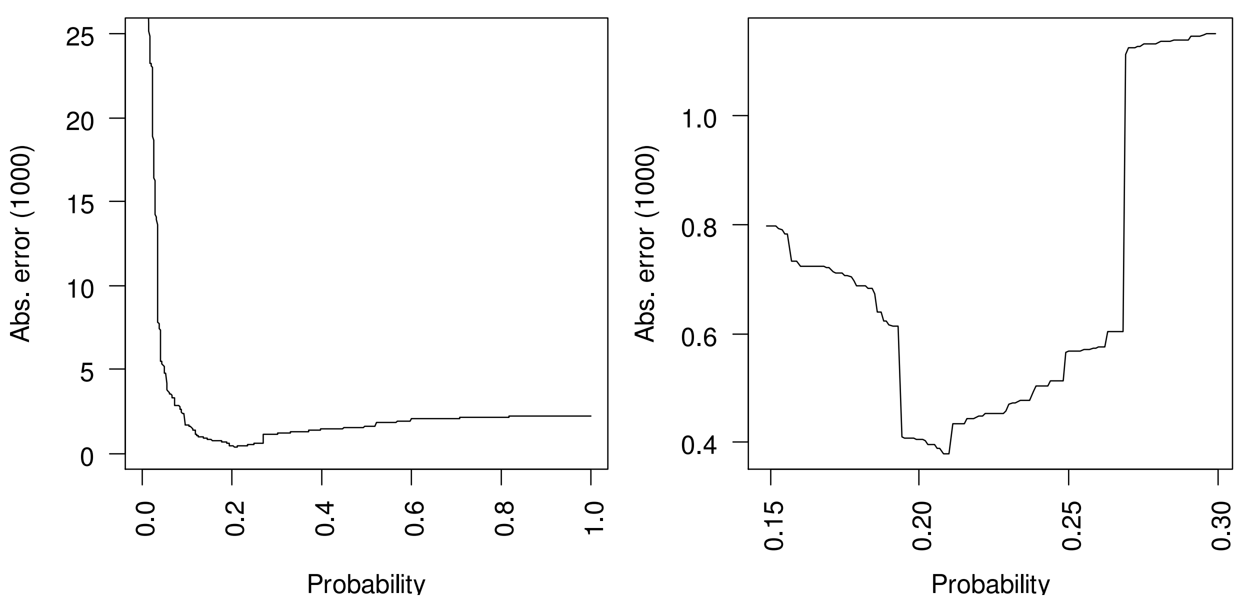


Figure 6. Total number of trips and proportion positive over time by season, in the subsetted data set.

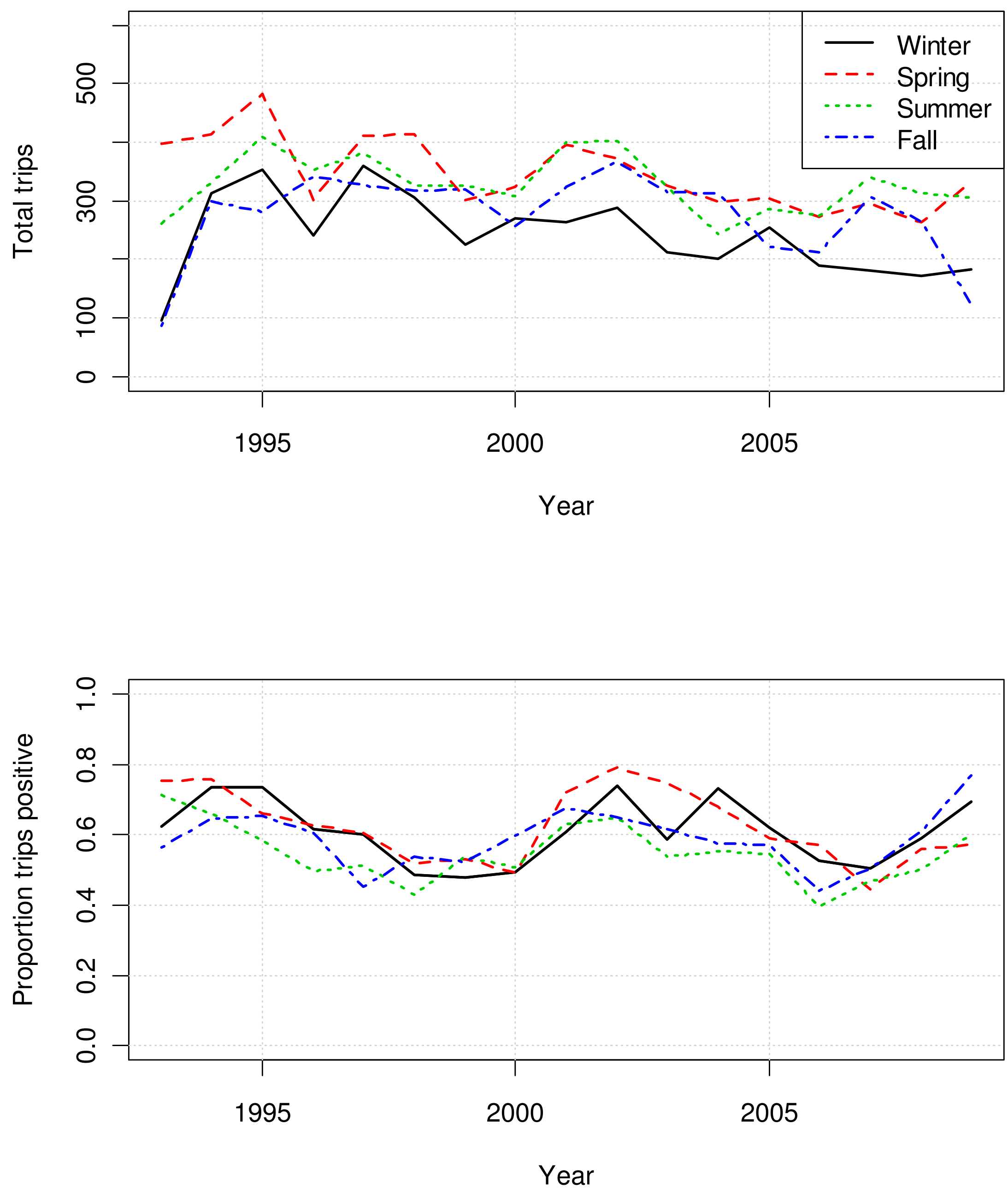
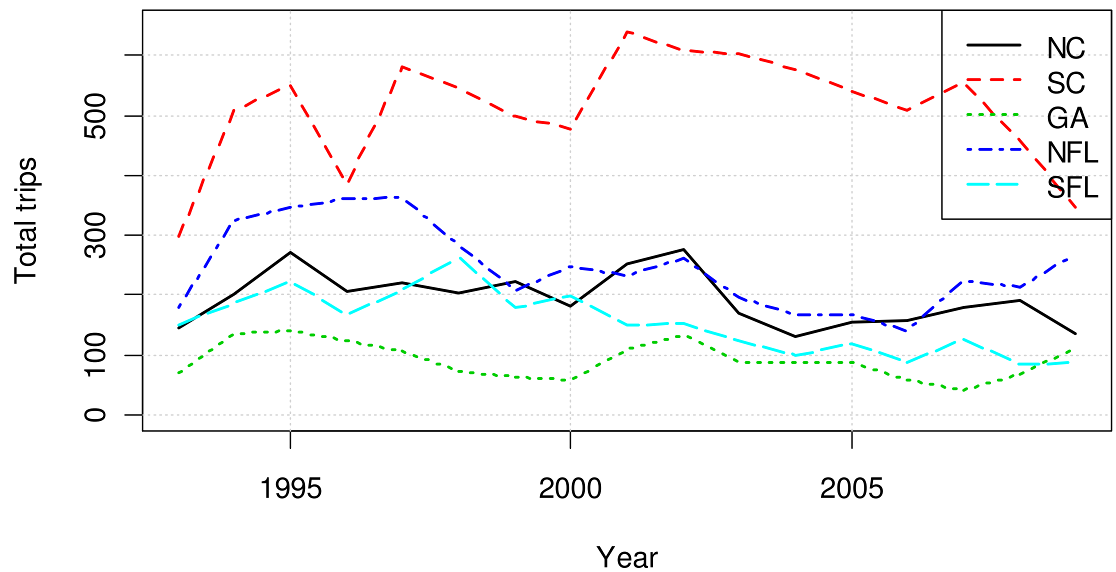


Figure 7. Total number of trips and proportion positive over time by area, in the subsetted data set.



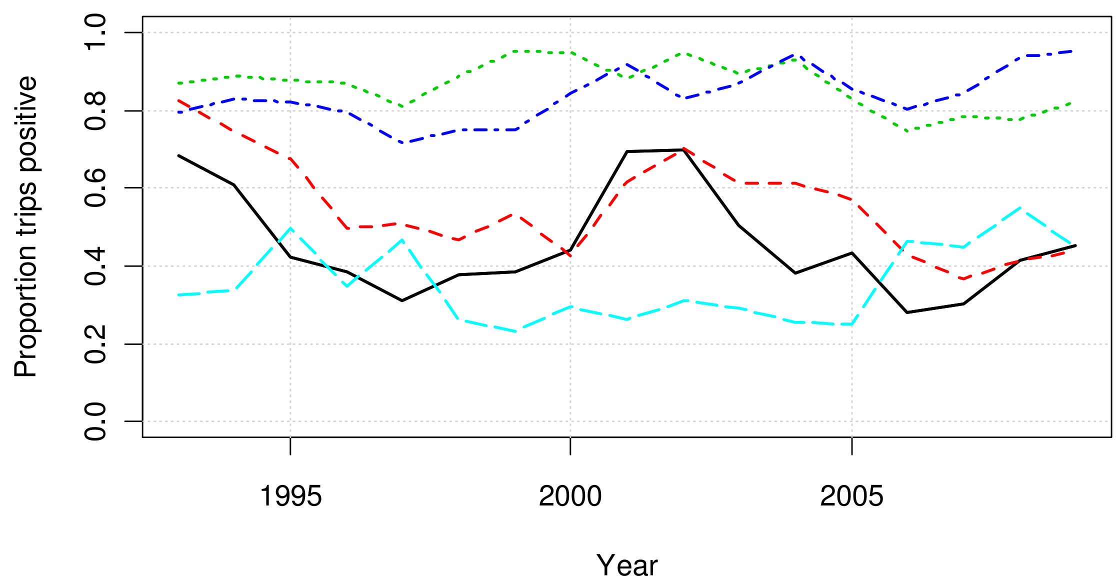


Figure 8. Total number of trips and proportion positive over time by sea days, in the subsetted data set.

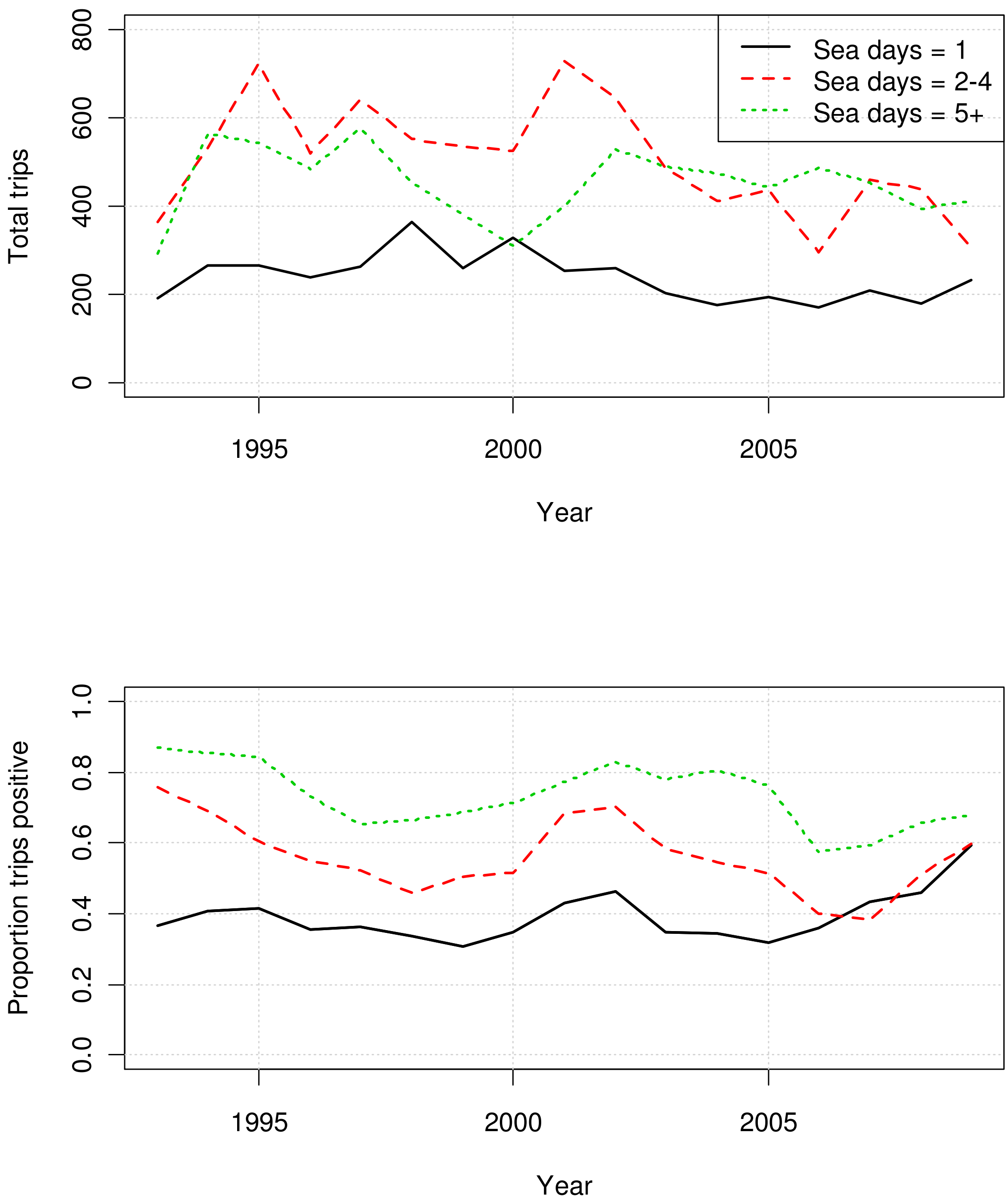


Figure 9. Total number of trips and proportion positive over time by crew, in the subsetted data set.

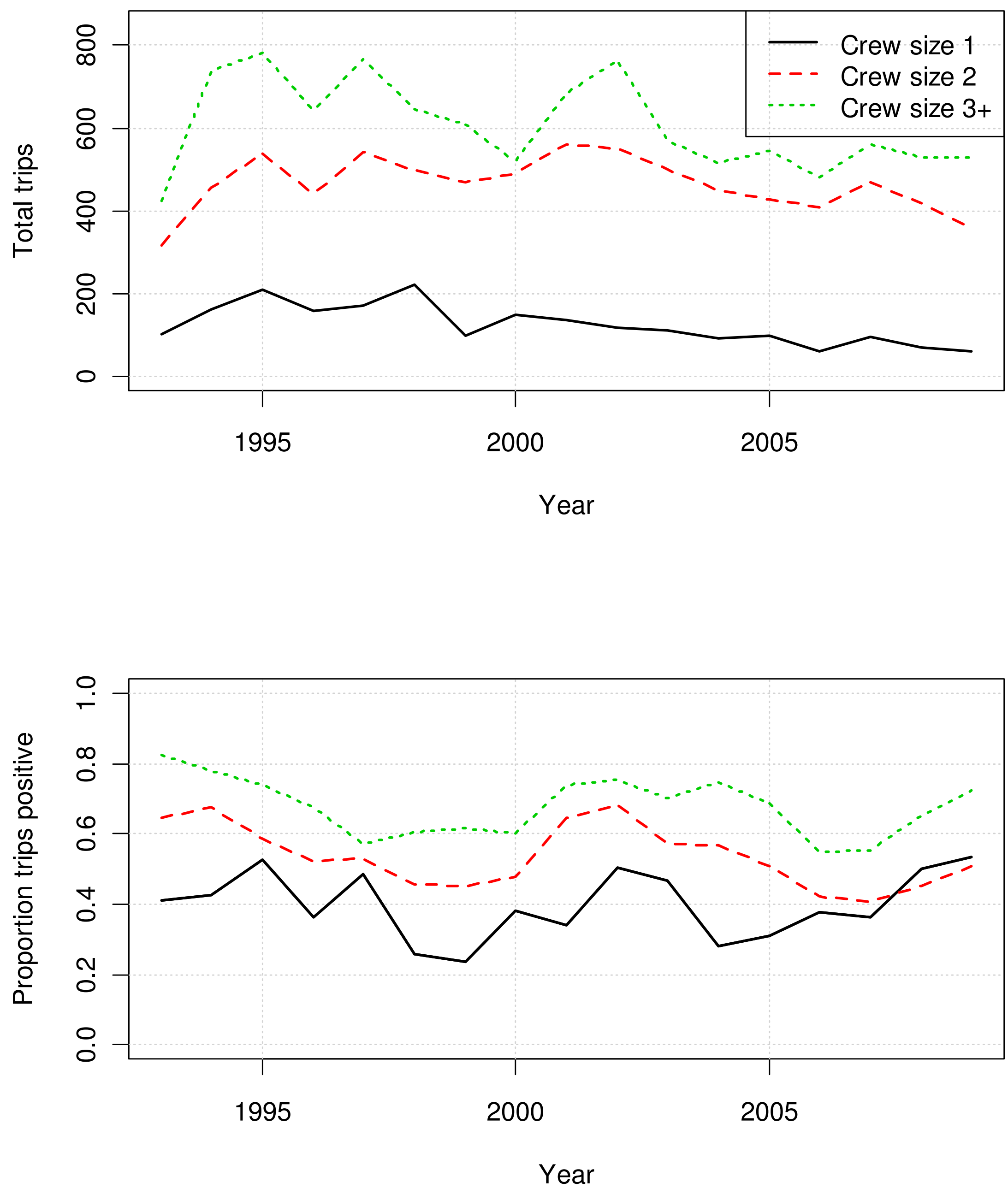


Figure 10. Diagnostics of Bernoulli submodel fits to positive versus zero CPUE data. Box-and-whisker plots give first, second (median), and third quartiles, as well as limbs that extend approximately one interquartile range beyond the nearest quartile, and outliers (circles) beyond the limbs. Residuals are randomized quantile residuals.

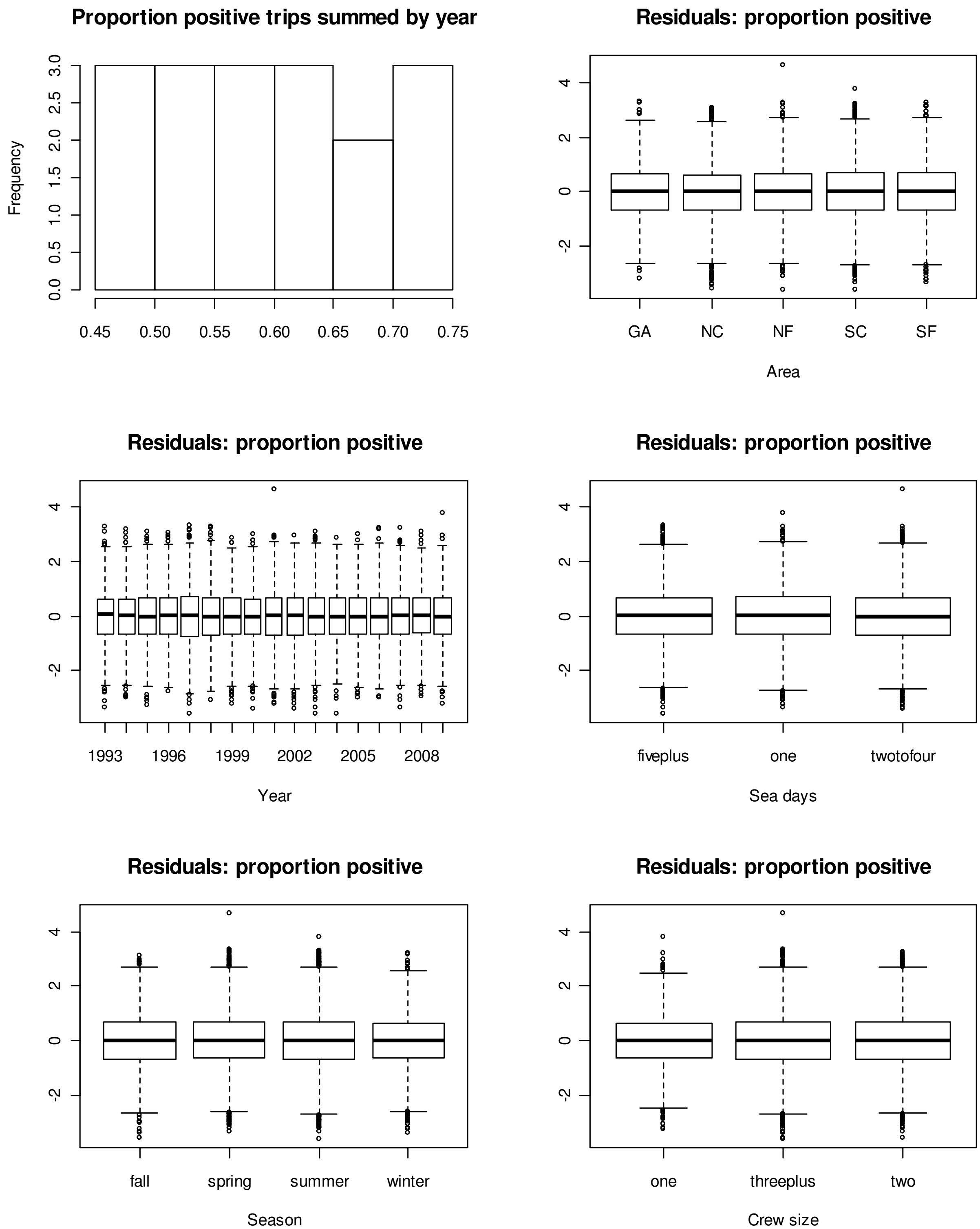
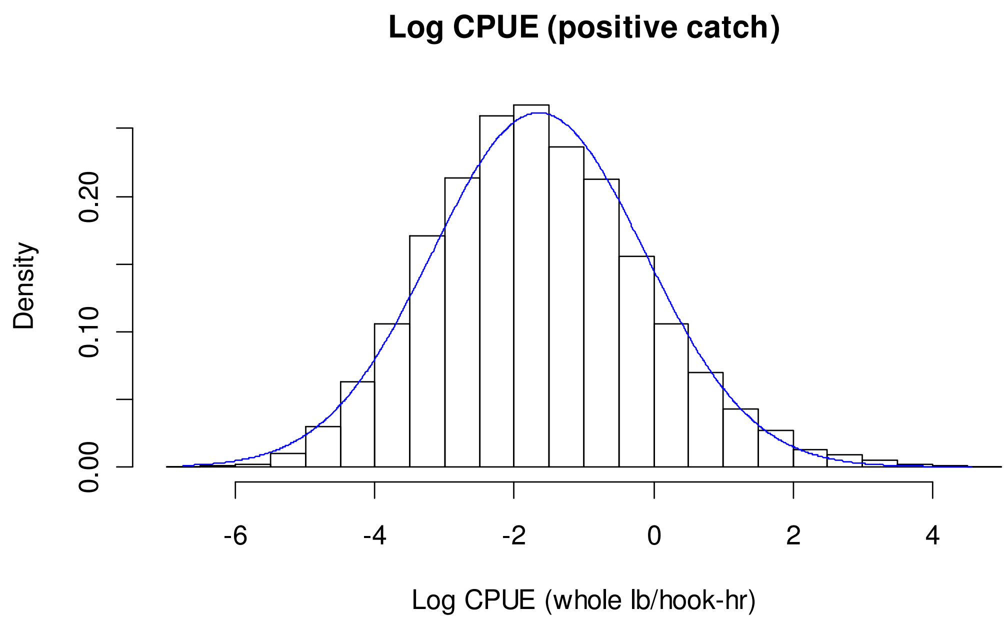


Figure 11. Diagnostics of lognormal submodel fits to positive CPUE data. Top panel shows the histogram of empirical log CPUE, with the normal distribution (empirical mean and variance) overlaid. Bottom panel shows the quantile-quantile plot of residuals from the fitted model.



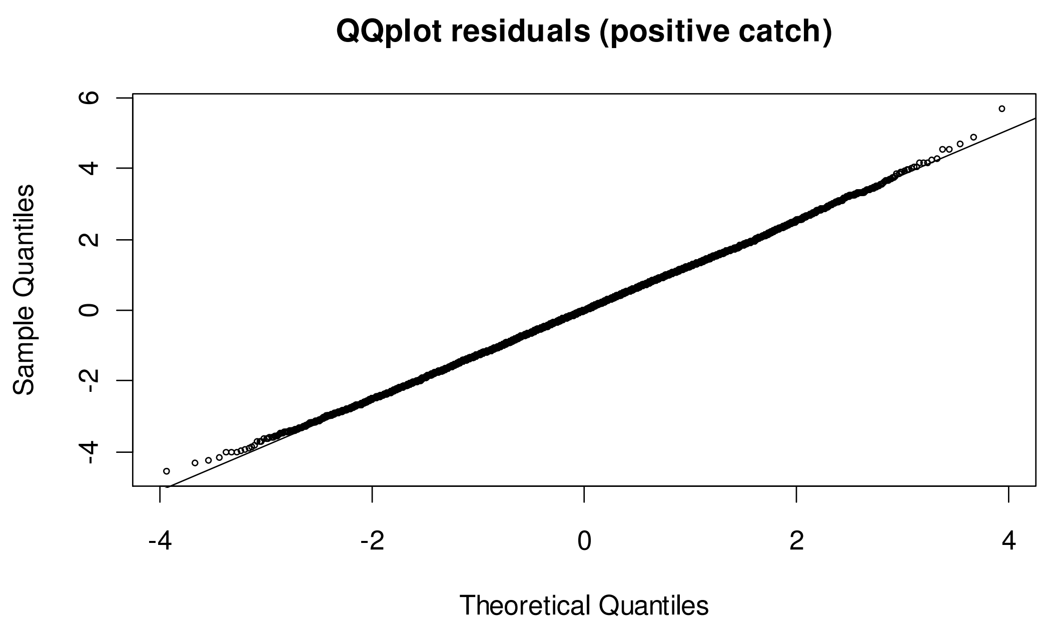


Figure 12. Diagnostics of lognormal submodel fits to positive CPUE data. Box-and-whisker plots give first, second (median), and third quartiles, as well as limbs that extend approximately one interquartile range beyond the nearest quartile, and outliers (circles) beyond the limbs. Residuals are raw.

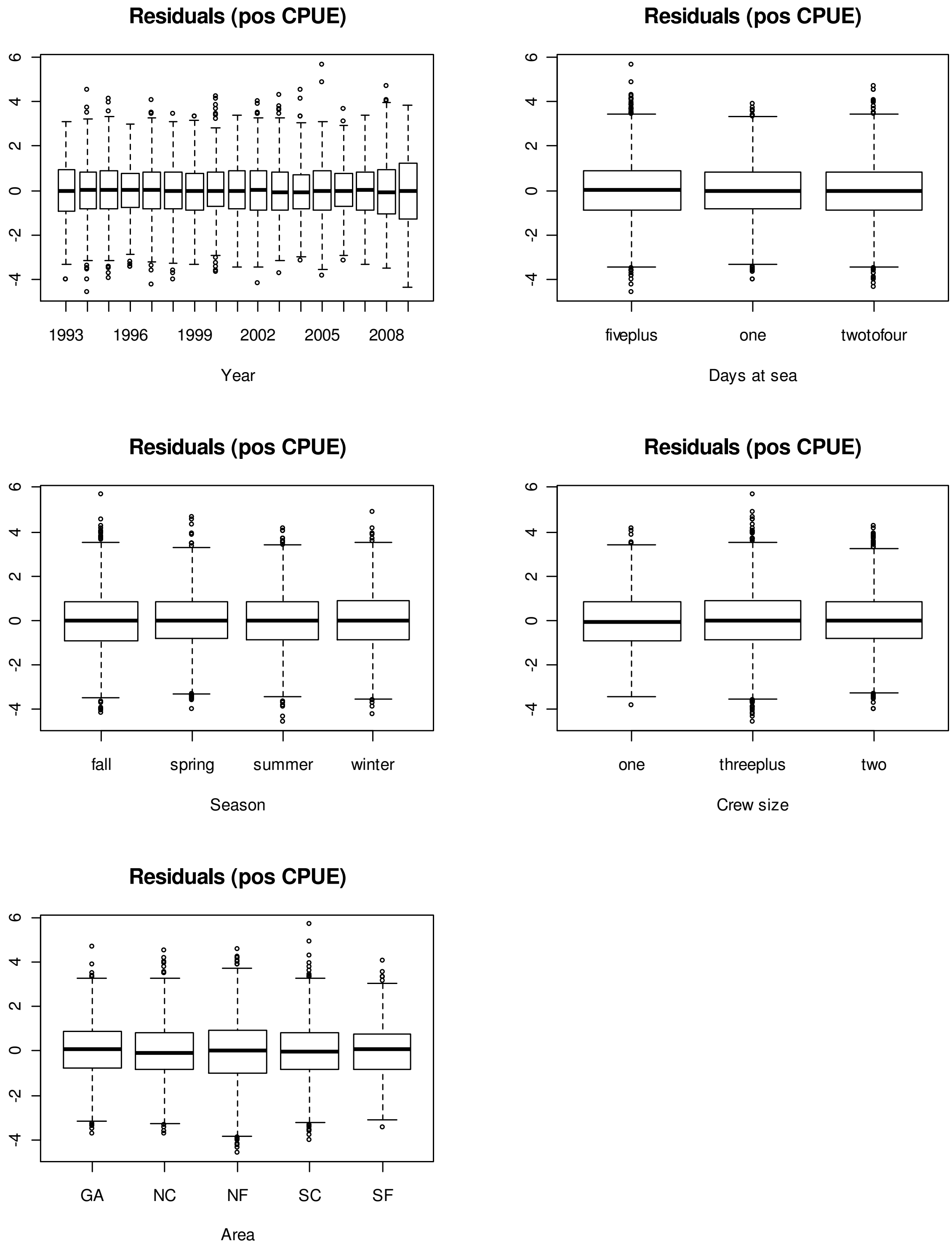


Figure 13. Nominal CPUE by area and across all areas.

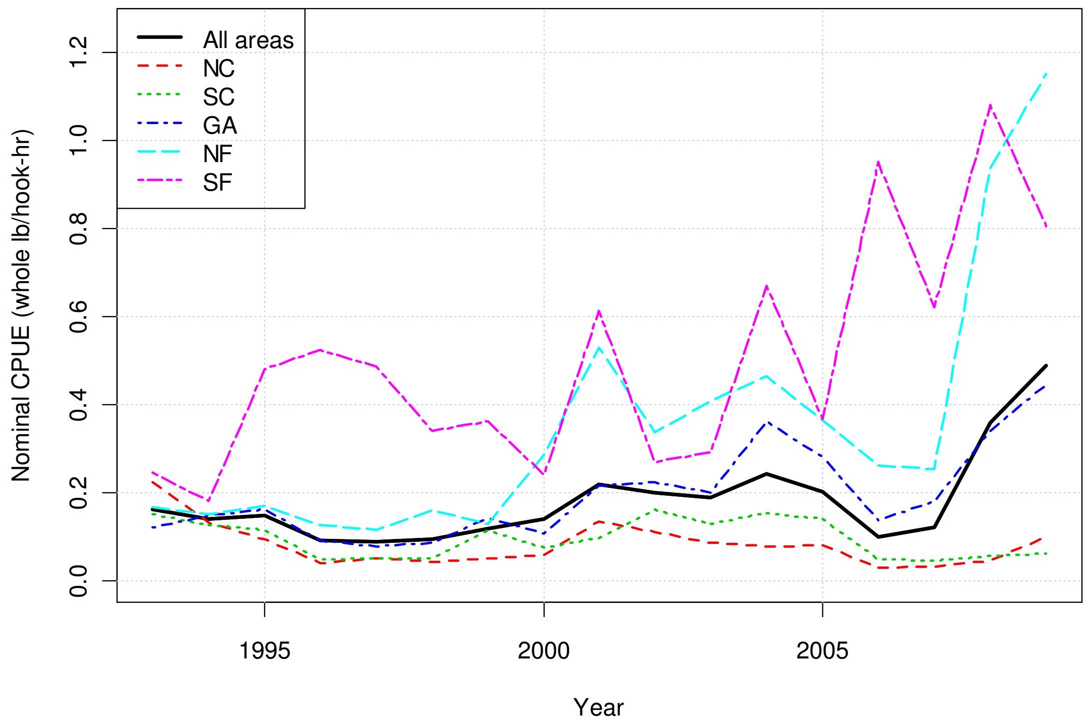
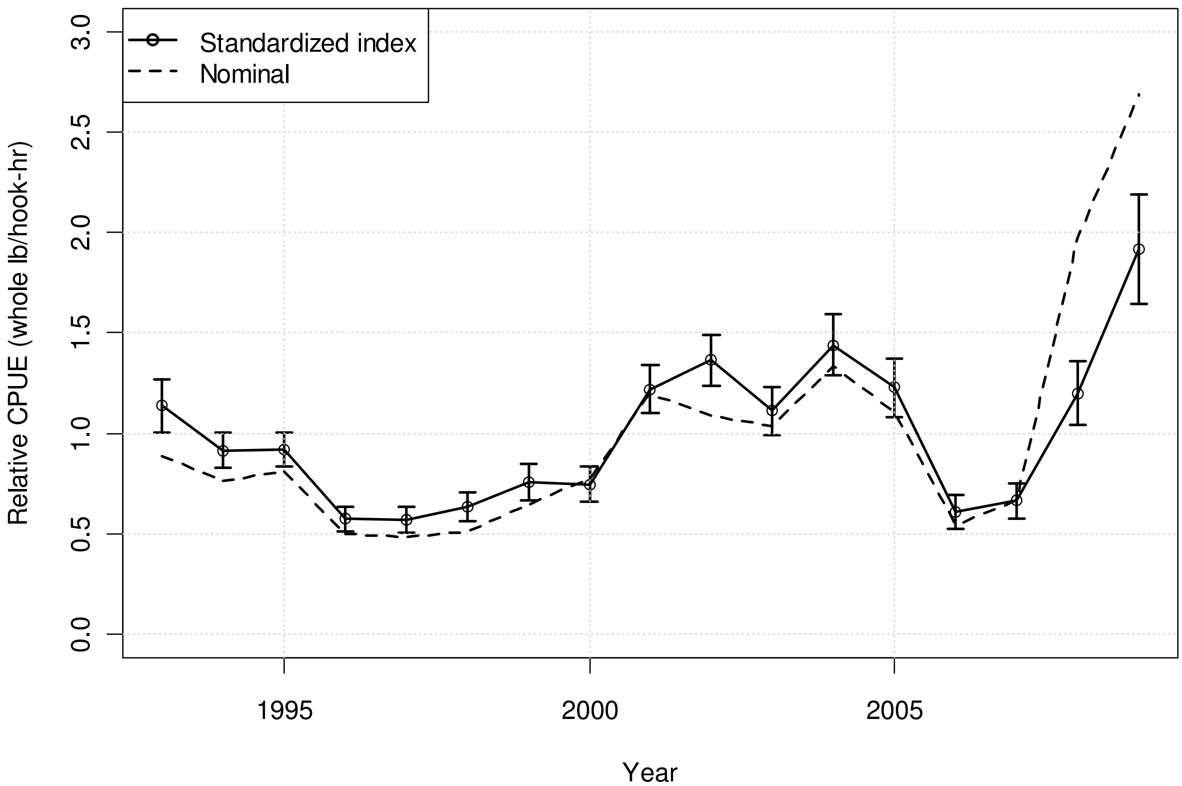


Figure 14. Relative standardized index (solid line, black circles, 95% error bars) and relative nominal index (dashed).



Appendix 1. The commercial logbook data set contains the following variables (all are numeric unless otherwise noted):

schedule: this is a unique identifier for each fishing trip and is a character variable
species: a character variable to identify species caught.
gear: a character variable, the gear type, multiple gear types may be used in a single trip, L = longline, H = handline, E = electric reels, B = buoy gear, GN = gill net, P = diver using power head gear, S = diver using spear gun, T = trap, TR = trolling
area: area fished, in the south Atlantic these codes have four digits- the first two are degrees of latitude and the second two are the degrees of longitude
conversion: conversion factor for calculating total pounds (totlbs) from gutted weight
gutted: gutted weight of catch for a particular species, trip, gear, and area
whole: whole weight of catch for a particular species, trip, gear, and area
totlbs: a derived variable that sums the gutted (with conversion factor) and whole weights, this is the total weight in pounds of the catch for a particular species, trip, gear, and area
length: length of longline (in miles) or gill net (in yards)
mesh1 – mesh4: mesh size of traps or nets
numgear: the amount of a gear used, number of lines (handlines, electric reels), number of sets (longlines), number of divers, number of traps, number of gill nets
fished: hours fished on a trip, this is problematic for longline data as discussed later
effort: like numgear, the data contained in this field depends upon gear type; number of hooks/line for handlines, electric reels, and trolling; number of hooks per longline for longlines; number of traps pulled for traps; depth of the net for gill nets, this field is blank for divers
source: a character variable, this identifies the database that the record was extracted from, sg = snapper grouper, grf = gulf reef fish, all records should have this source code
tif_no: a character variable, trip identifier, not all records will have a tif_no
vesid: a character variable, a unique identifier for each vessel
started: numeric (mmddyy8) variable, date the trip started
landed: numeric (mmddyy8) variable, date the vessel returned to port
unload: numeric (mmddyy8) variable, date the catch was unloaded
received: numeric (mmddyy8) variable, date the logbook form was received from the fisherman
opened: numeric (mmddyy8) variable, date the logbook form was opened and given a schedule number
away: number of days at sea, this value should equal (landed-started+1)
crew: number of crew members, including the captain
dealer: character variable, identifier for the dealer who bought the catch, in some cases there may be multiple dealers for a trip
state: character variable, the state in which the catch was sold
county: character variable, the county in which the catch was sold
area1 – area3: areas fished, if the trip included catch from multiple areas, those areas will be listed here
trip_ticke: character variable, trip ticket number, a unique identifier for each trip not all trips have this identifier.