Standardized catch rates of blueline tilefish (*Caulolatilus microps*) in the South Atlantic and Gulf of Mexico waters of the U.S. from recreational headboat logbook data

Sustainable Fisheries Branch – NMFS (contact: Eric Fitzpatrick)

SEDAR50-DW25

Submitted: 13 January 2017

This information is distributed solely for the purpose of pre-dissemination peer review. It does not represent and should not be construed to represent any agency determination or policy.

Please cite this document as:

SFB-NMFS. 2017. Standardized catch rates of blueline tilefish (*Caulolatilus microps*) in the South Atlantic and Gulf of Mexico waters of the U.S. from recreational headboat logbook data. SEDAR50-DW25. SEDAR, North Charleston, SC. 29 pp.

Standardized catch rates of blueline tilefish (*Caulolatilus microps*) in the South Atlantic and Gulf of Mexico waters of the U.S. from recreational headboat logbook data

Sustainable Fisheries Branch[*]

February 2017

The headboat fishery includes for-hire vessels using hook and line gear that generally target hard bottom reefs including multiple species in the snapper-grouper complex.

Variable include: (year, month, day, area, location, trip type, number of anglers, species, catch, and vessel id.

This document describes the data and standardization of the SEDAR 50 headboat indices for blueline tilefish. Protocols were developed for SEDAR 32 that are reconsidered here due to changes in management and improvements in index development.

Headboat data from the mid-Atlantic Vessel Trip Report (VTR) database were initially included in this analysis but issues exist that required discussion at the data workshop (DW).

VTR headboat data issues

1. Potential under reporting in early years
2. Potential instances when number caught was reported in weight, severity and magnitude unknown
3. Limited time series relative to headboat data in South Atlantic and Gulf (approximately 10 years compared to 28 years)
4. Limited number of headboats and inconsistency in reporting by a majority of the fleet

Workshop Decision: These data were examined thouroughly at the data workshop and the Index Working Group (IWG) decided unanimously to exclude these data from the Southeast Regional Headboat Survey data.

Southeast Regional Headboat Survey data issues

1. Sampling of Area 1 vessels (Cape Hatteras to Virginia line) was inconsistent throughout the time series (sporadic in the early years and most recently sampling starting in 2009). For this reason, Area 1 was filtered from SEDAR 32 because these vessels were in operation but the effort was unknown.
2. Inconsistent effort in all all areas caused by:

- reduction in vessels going to deep water (Carolinas in the 1980s)
- lower compliance in GA/FL in the 1990s
- Potential targeting north of hatteras

[*]National Marine Fisheries Service, Southeast Fisheries Science Center, 101 Pivers Island Rd, Beaufort, NC 28516

Headboat effort

The proportion of the deepwater complex species (e.g., snowy grouper, blueline tilefish, tilefish, warsaw grouper, speckled hind, yellowedge grouper, etc) relative to all species was calculated at the trip level (>1% deepwater species) to investigate the occurance of these deepwater species in order to determine effective headboat effort in these deepwater habitats. Five time blocks and four spatial areas were considered. This figure suggests that deepwater species are infrequently targeted in the headboat fishery while differences exist spatially and temporally throughout the range.

For the standardization, headboat effort was defined as any headboat trip that caught at least 1 deepwater complex species. Headboat effort (positive blueline tilefish trips) in the Carolinas (CAR) dropped in the early 1990s with most trips being full-day. The majority of the trips in Georgia/East Florida(GA/FLE) region were half-day trips.

Data considerations (workshop decisions in bold)

1. Outlier removal

- **remove highest 0.5% of values for catch, anglers, cpue**
- Rationale: Extreme values occur more frequently in self-reported data because there are limited methods for validating data. Removing a small percentage of the trips with the hightest values for variables used to calculate CPUE is an unbiased method to correct for potential errors without influencing the results.

2. Cutoff for number of trips/vessel and anglers

- **retain all trips**
- remove vessels with less than 30 trips over all years, remove trips with 6 or fewer angler

3. Starting year (IWG Decision: 5 alternative headboat indices were recommended)

- **1986, first year with Gulf data (South Atlantic and Gulf of Mexico)**
- **1986, (Gulf of Mexico only)**
- **1973, (North Carolina to Cape Canaveral)**
- **1980, (Cape Canaveral to Florida Keys)**
- **1980, (North Carolina to Florida Keys)**

4. Terminal year (IWG Rationale: Increase in targeting in the late 2000s and management regulations (eg., snowy regulations and 2011 deepwater closure))

- **2005**

5. Trip types

- combine half and 3/4-day trips, combine full- and multi-day trips as in SEDAR 19
- Exclude 3/4-day and multi-day trips
- Exclude half-day trips for CAR and GNF
- include as factor
- **include all data due to limited sample size, (caveat: effective effort for multiday trips may not correspond to calculated effort (angler-hrs))**

6. Use month or season as factor in model

- use month as in SEDAR 19
- use season (May-Aug, Sep-Dec)
- use 4 season

Data Exclusions

Extreme values are common in self-reported data. Recent SEDAR stock assessments have removed values at the extreme tails of distribution for values associated with cpue values. For these reasons,the SEDAR 50 headboat index will exclude trips with the largest 0.5% values for catch in number (> 320) and cpue (> 2.25) for trips that caught blueline tilefish. 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 (> 109) were removed (Figure 6).

Nominal catch rates

Nominal catch rates of positive blueline tilefish trips by year and region from the data as filtered for input are shown in Figure 7.

Evaluation of explanatory variables

1. YEAR - Year was necessarily included, as standardized catch rates by year are the desired outcome.
2. AREA

- North Carolina to Cape Canaveral (CanaveralN)
- Cape Canaveral to Florida Keys (CanaveralS)
- Gulf (GULF)

3. Party size (number of anglers)

- **Greater than 30 anglers, Less than 30 anglers**

The patterns in the remaining positive blueline tilefish trips by month and region show few trips in the Carolinas for Jan and February. However, Jan and Feb have the most positive blueline tilefish trips for Georgia and East Florida. The seasonal pattern in cpue across months seems consistent across areas with slightly higher values for May to October. Season was not included as a factor.

4. VESSEL SIZE (vsize) - For SEDAR 32, a factor was developed for the number of anglers using the quartiles of the number of anglers across all trips as breaks for the factors. Given the large range of vessel sizes, a trip with 20 anglers could be either almost full or almost empty. Here we develop a factor for vessel size and crowding separately using the number of anglers. The proxy for vessel size is the maximum anglers reported over all trips for a vessel. This was then divided into four factors based on visual inspection of the density plots into: 1. less than 30 maximum anglers (a.lt30), 2. 30-59 maximum anglers (b.30-59), 3. 60-89 maximum anglers (c.60-89), and 4. 90 or more maximum anglers (d.ge90). IWG Decision: Vessel size was not included as a factor.
5. PERCENT FULL (pctfact) 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. This was then divided into 4 equally spaced factors; 1. less than 25% full (a.lt25), 25-49% full (b.25-49), 50-74% full (c.50-74), and 75% or more full(d.ge75). Percent full was not included as a factor.

Analytical decisions

1. Use Stephens and Maccall method

- experience from SEDAR 32 showed that Stephens and Maccall removed more than half the years from dataset due to low sample sizes
- model positive trips only
- **model all trips that caught at least one deepwater complex species**

2. Interaction terms

- **no interaction terms**

3. CPUE units and transformation

- **use catch/angler-hr for CPUE unit**
- use catch/angler with trip type as factor
- use 1/CPUE as dependent variable as in SEDAR 19

Standardization

CPUE was modeled using the GLM approach (Lo, L., and J. 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). Estimates of variance were based on 1000 bootstrap runs where trips were chosen randomly with replacement (Efron and Tibshirani 1993). All analyses were programmed in R, with much of the code adapted from Dick (2004).

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 blueline tilefish 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. Diagnostics, based on Pearson residuals, suggested reasonable fits of the Bernoulli submodel.

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. The two distributions, each with their best set of explanatory variables (all of them), were compared using AIC. Lognormal outperformed gamma, and was therefore applied in the final GLM. Diagnostics suggested reasonable fits of the lognormal submodel (Figure 8 and Figure 9).

Results - Coastwide (South Atlantic and Gulf of Mexico)

The standardized index was similar to the nominal index with the exception of a few years associated with peaks in the catch rate (Figure 11). Assessment methods that account for changes in catchability could be implemented over time periods where effort may have been influenced by management measures.

Additional indices are located in the appendix.

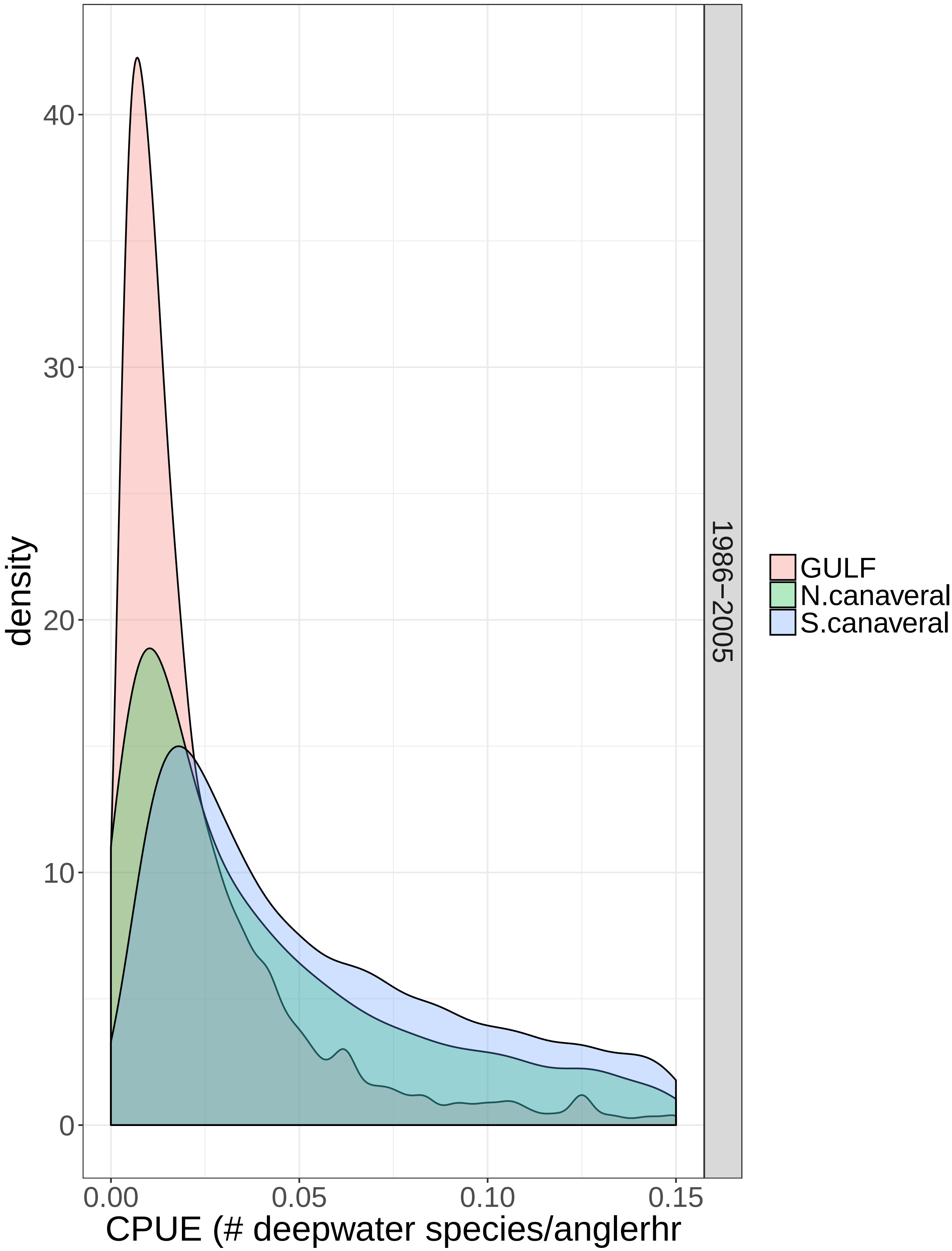


Figure 1: Density of percentage of deepwater species (ex: SG, BLT, TF, WSG, YG, SpkHD) for each headboat trip catching more than 1% of deepwater species.

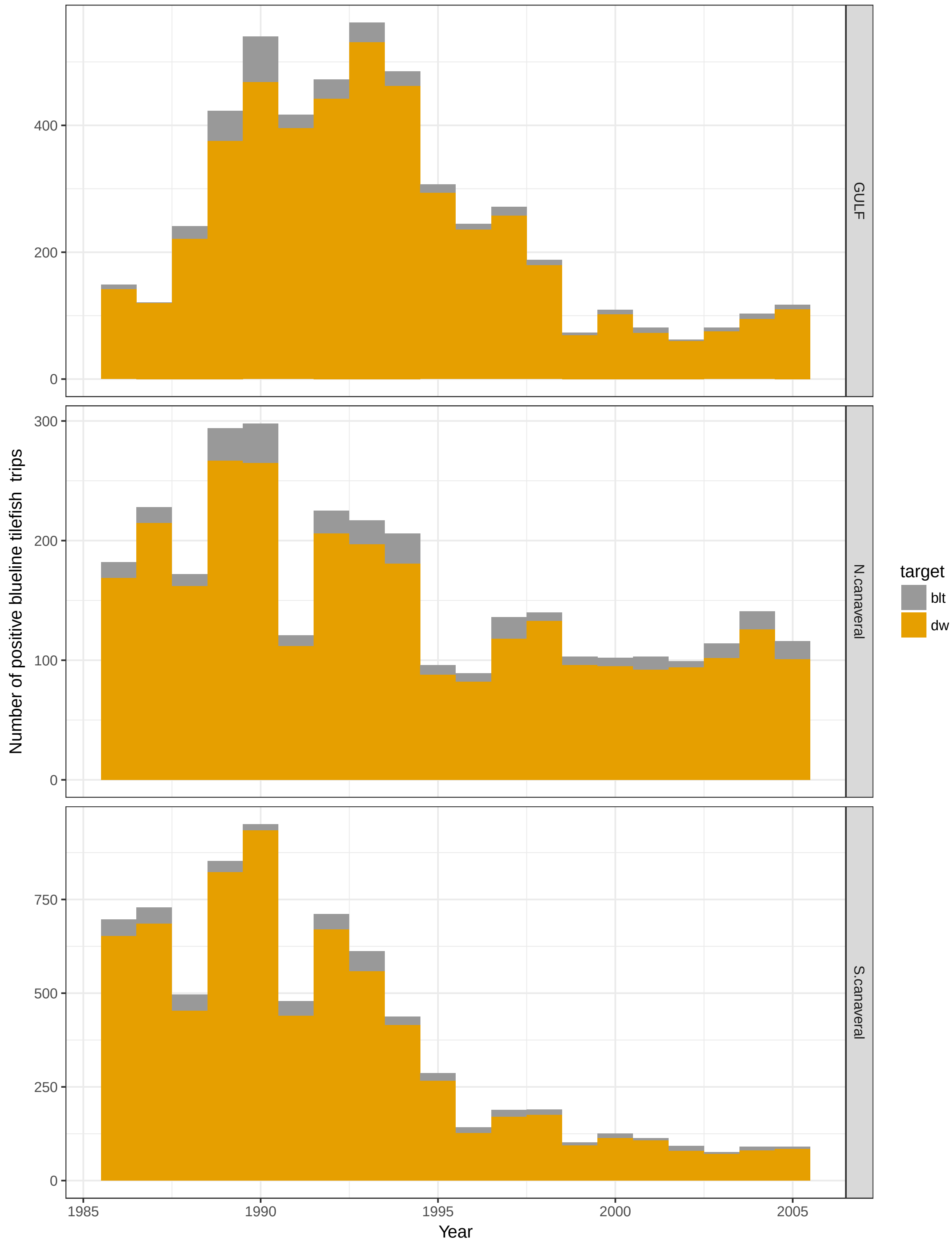


Figure 2: Number of deepwater and positive blueline tilefish headboat trips that submitted logbooks by region.

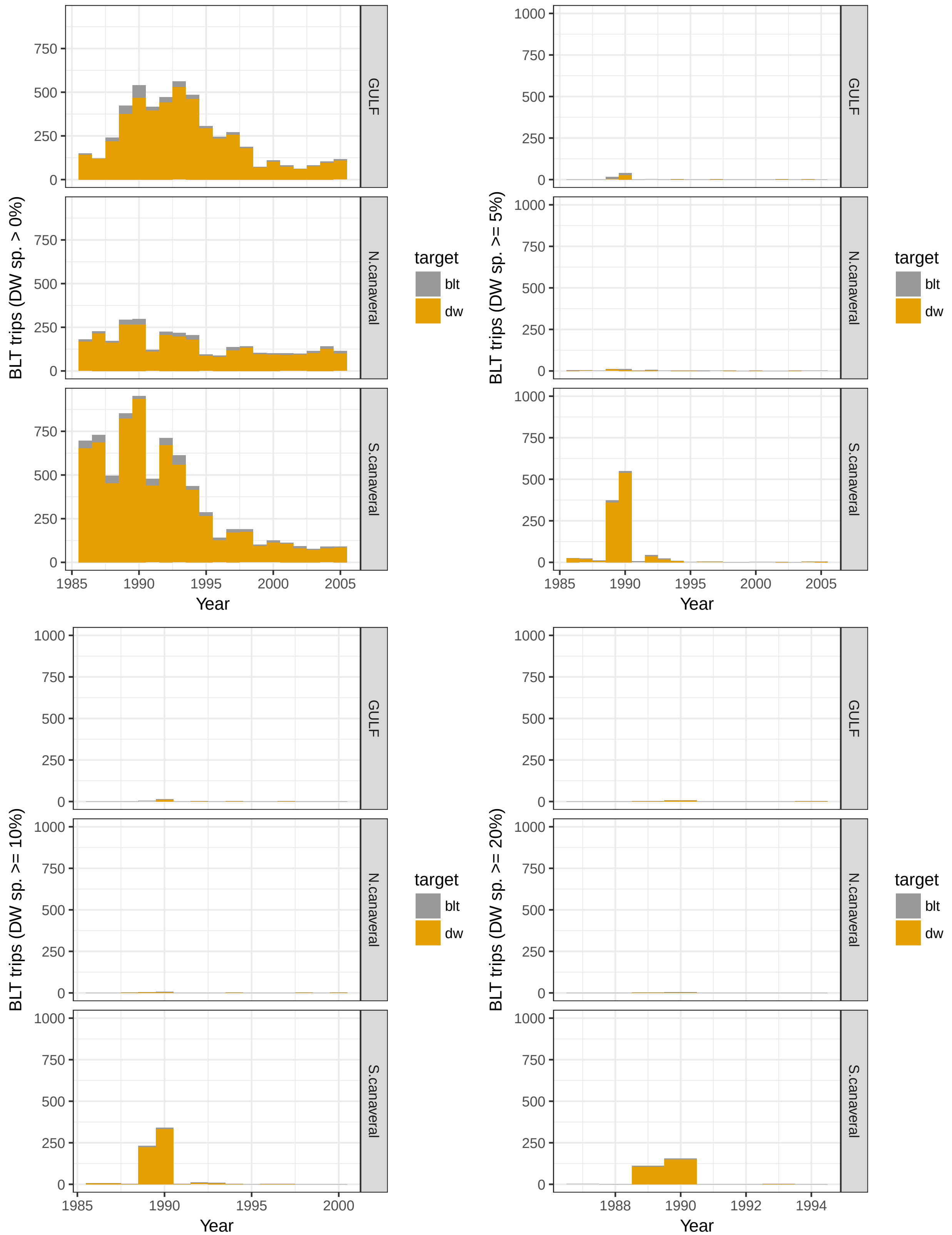


Figure 3: Number of deepwater and positive blueline tilefish headboat trips for various cutoff scenerios (>0% of catch was deepwater species (DW), >5% DW, >10% DW, >20% DW).

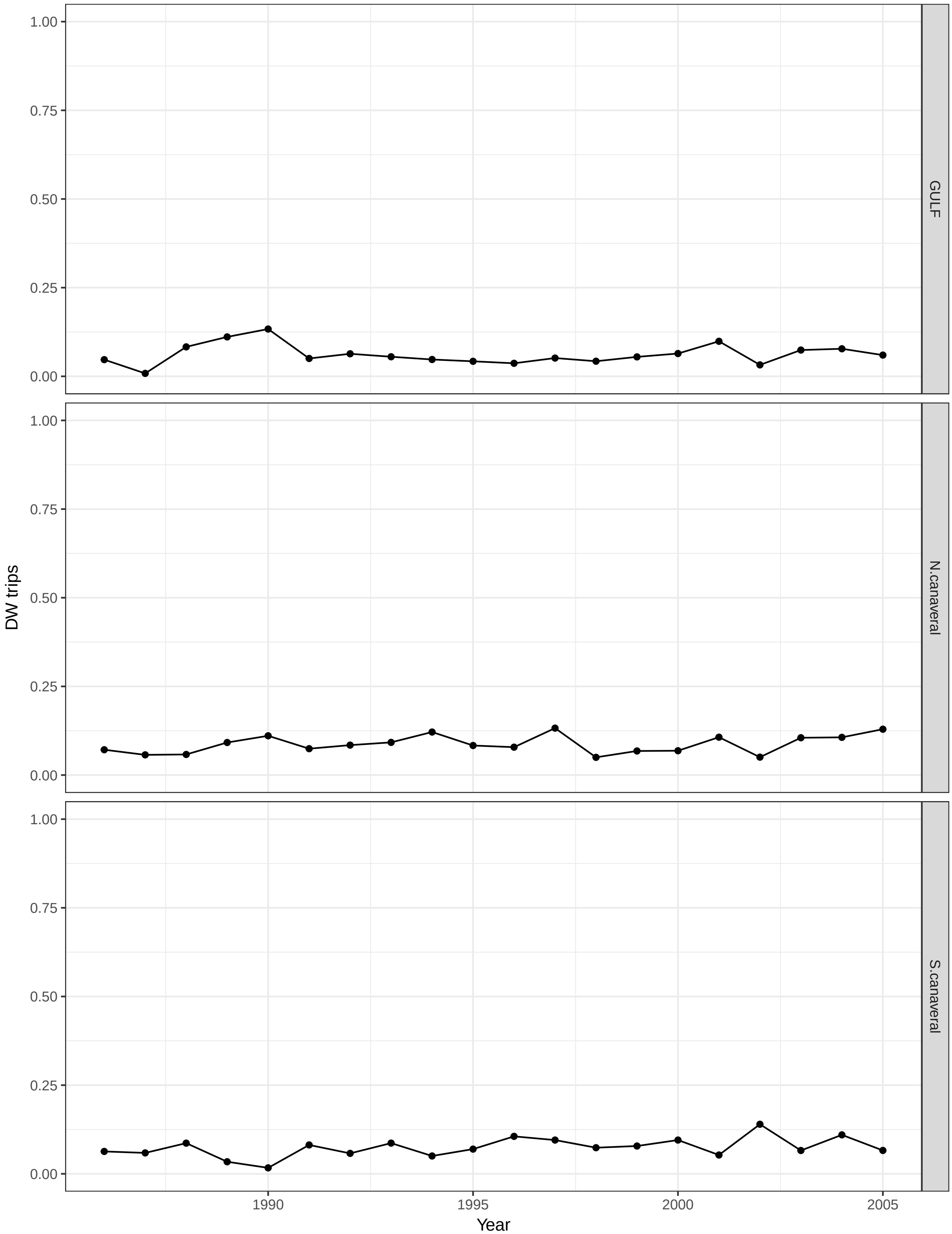


Figure 4: Proportion positive of (blueline trips/total DW trips) using deepwater complex species.

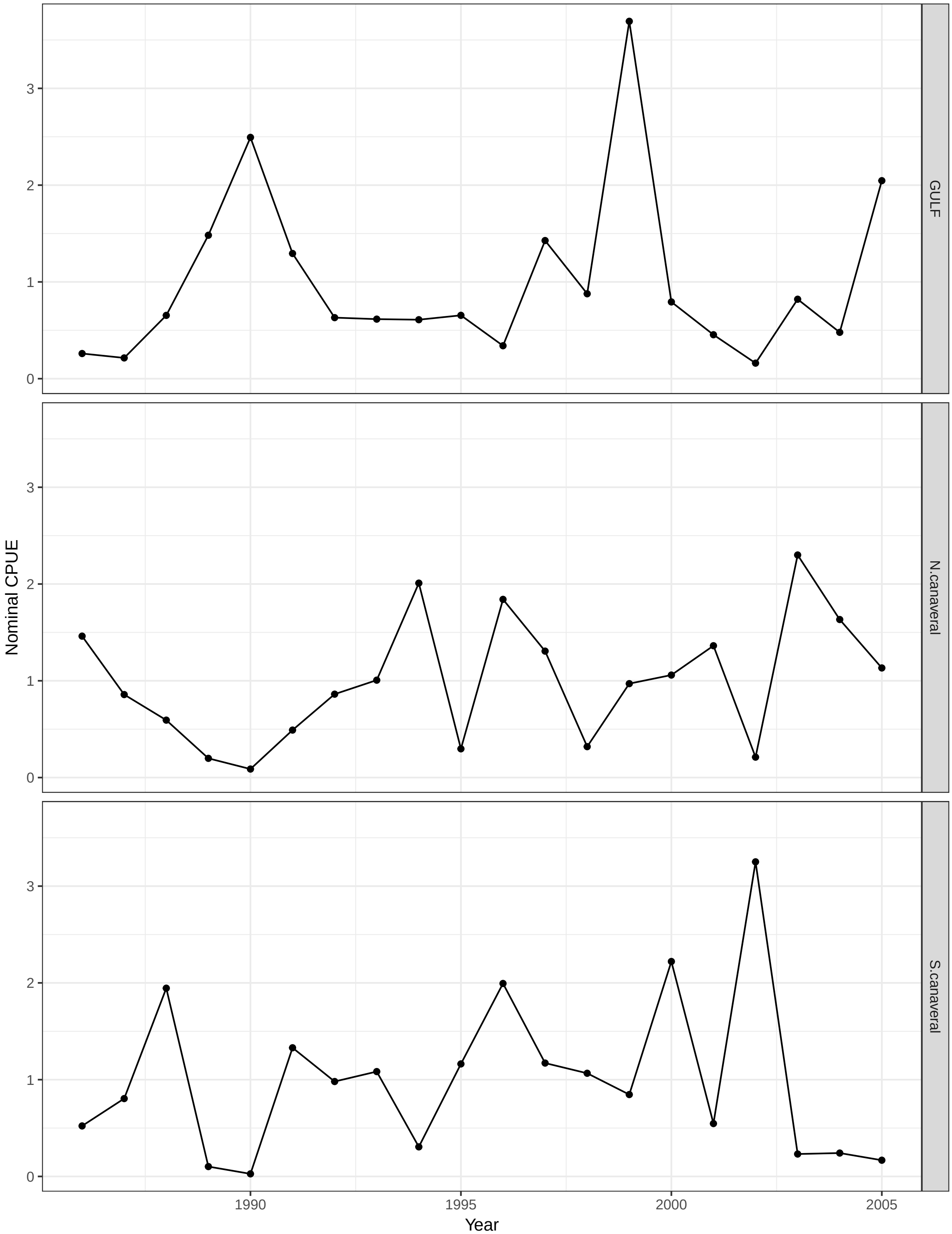


Figure 5: Nominal blueline tilefish catch per unit effort.

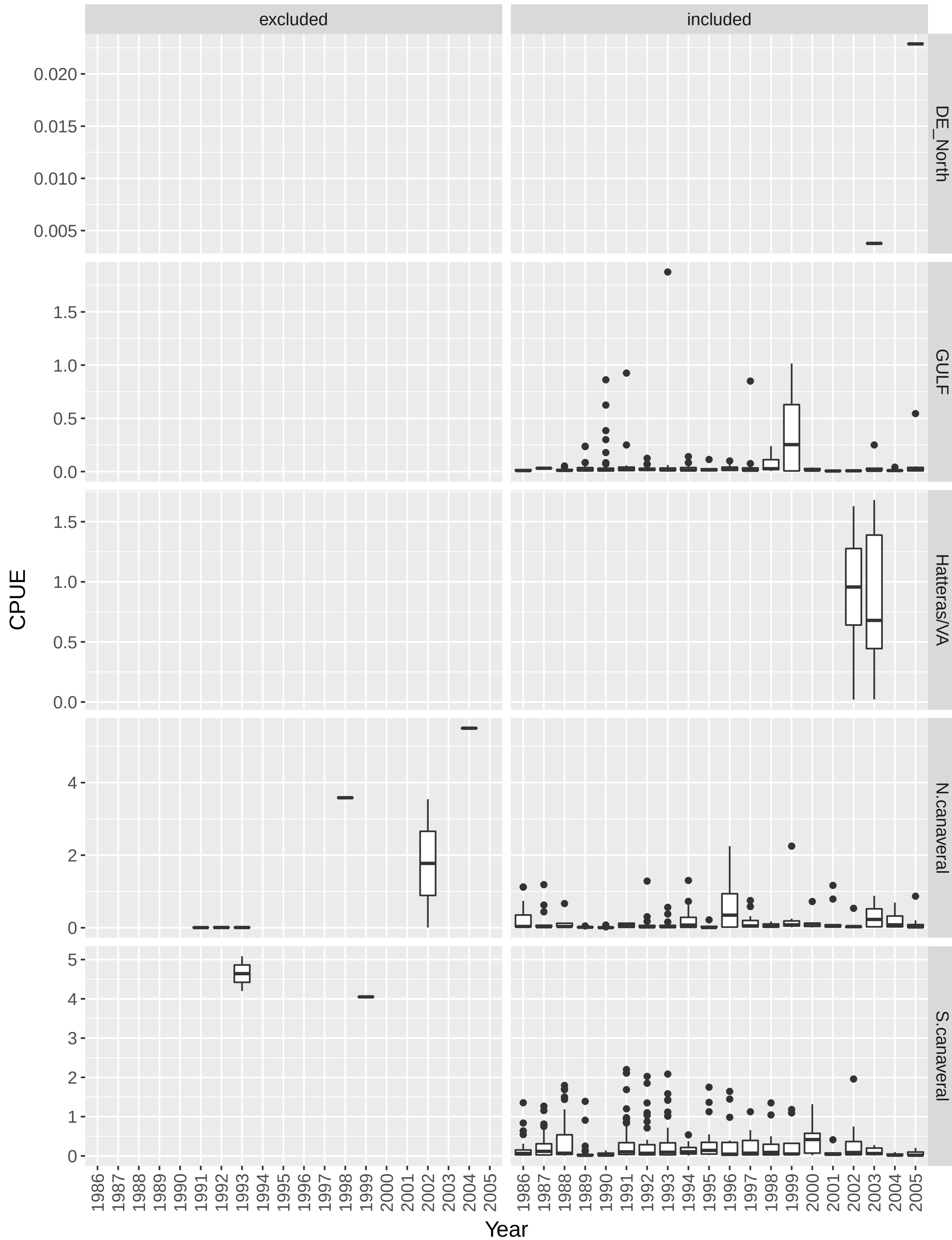


Figure 6: Records determined as outliers (excluded) based on removal of values above the 99.5th percentile for anglers, number of fish caught, and cpue.

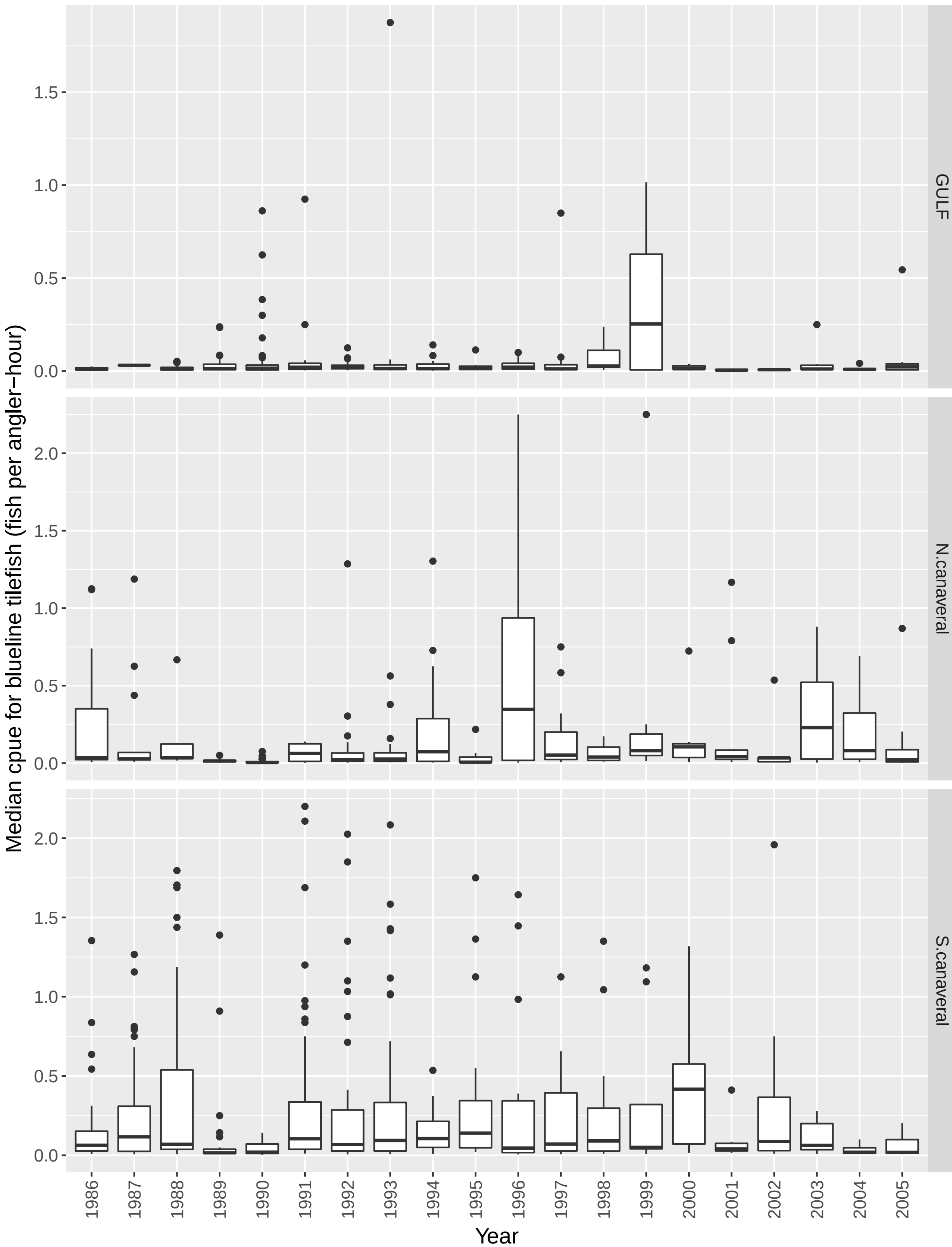
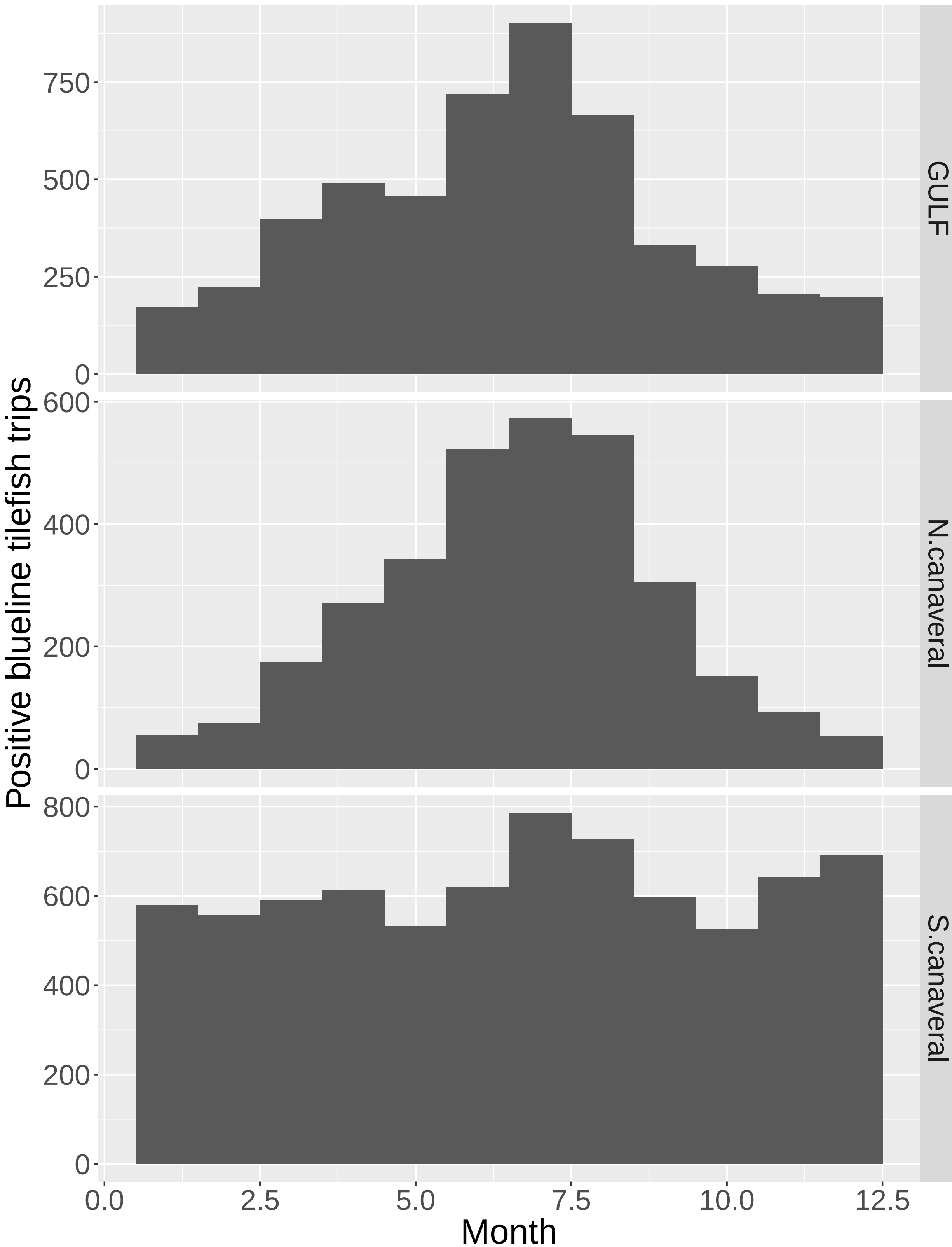
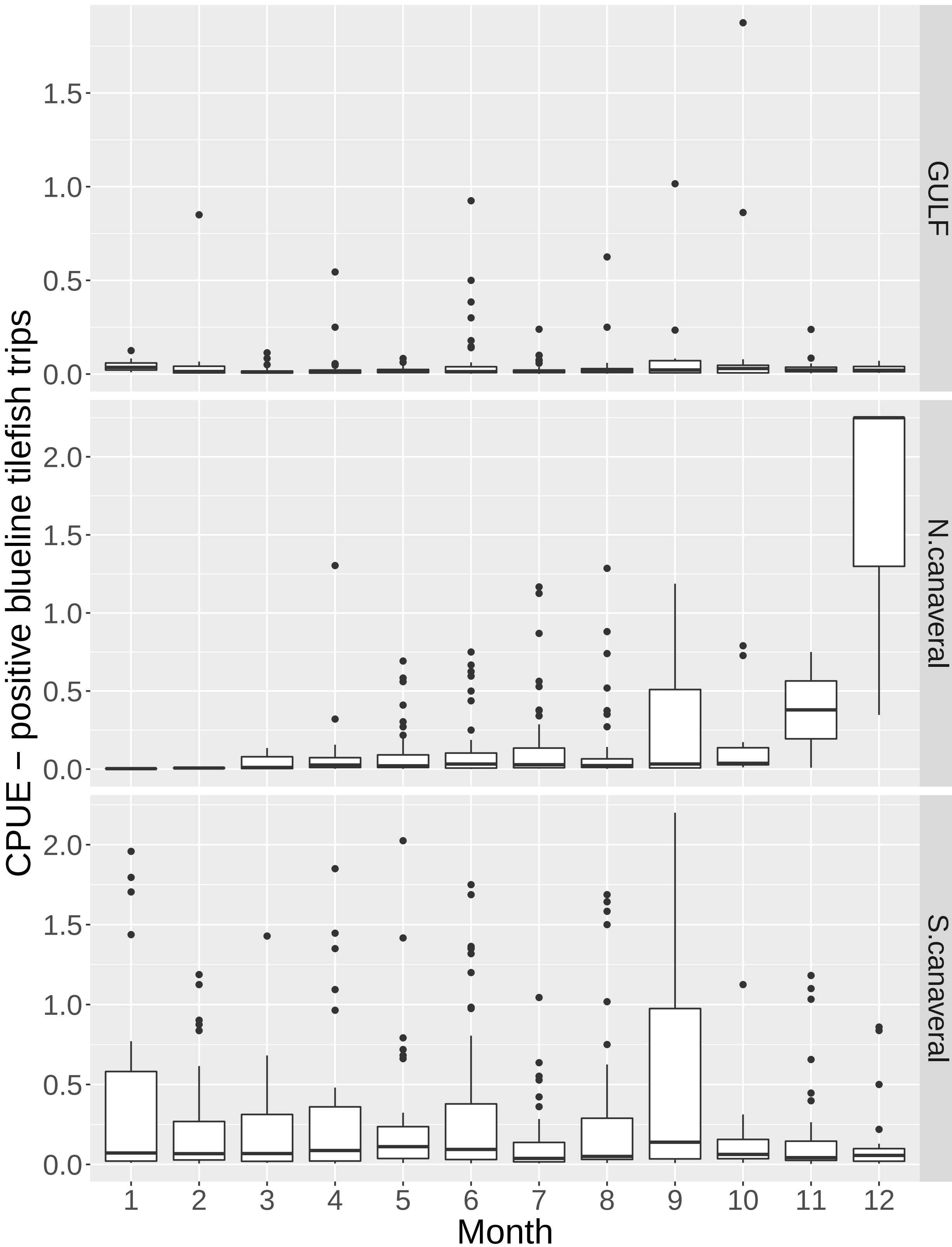


Figure 7: Blueline tilefish cpue by region and season.


750
500
250
0
GULF
600
400
200
0
N.canaveral
800
600
400
200
0
S.canaveral
Positive blueline tilefish trips
0.0
2.5
5.0
7.5
10.0
12.5
Month



[1] "0 (total) records were removed by filter." [1] "0 positive records removed by filter."

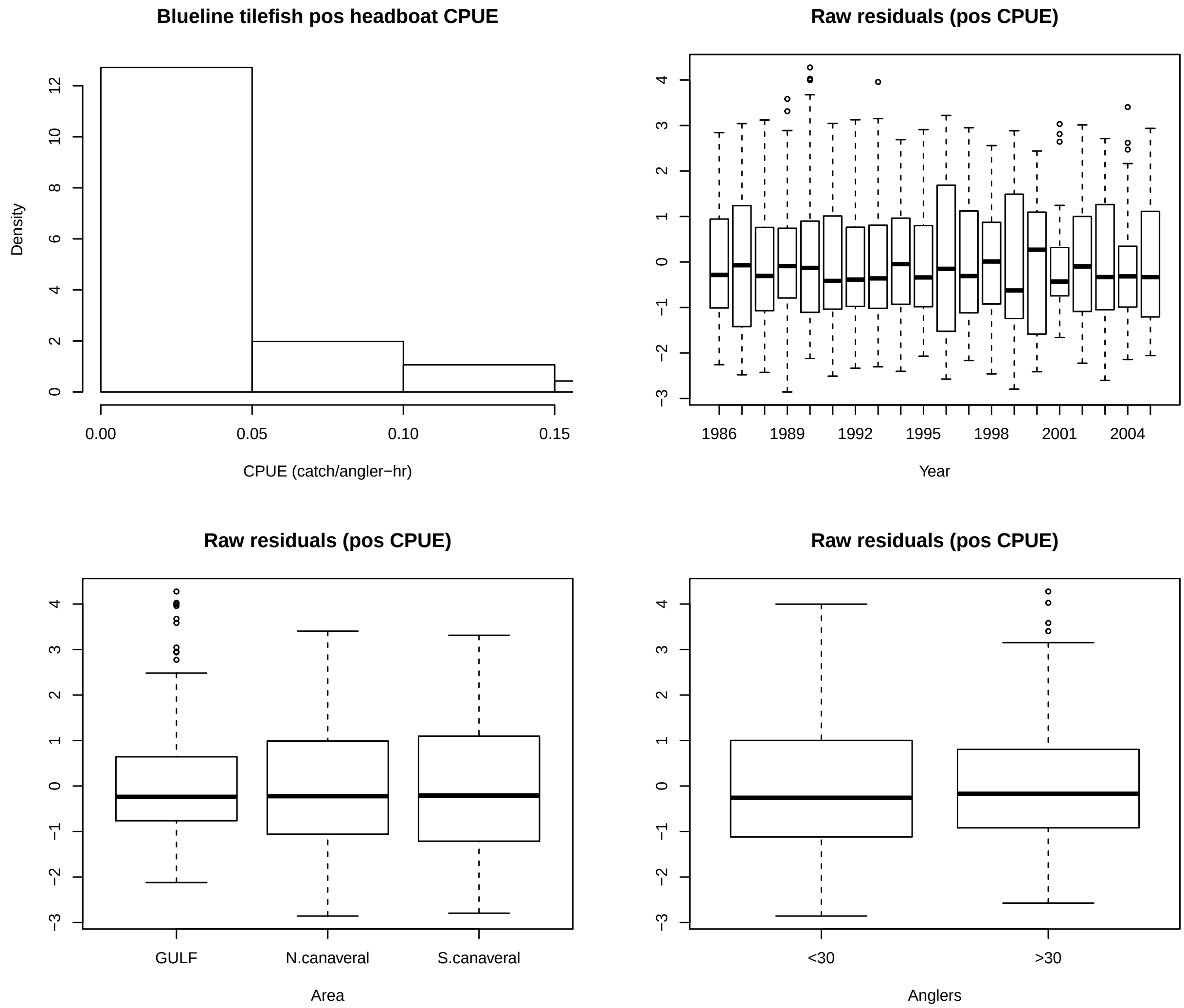


Figure 8: Diagnostics of lognormal submodel fits to positive CPUE data. Top left panel shows the distribution of positive cpue. Box and whisker plots give first, second (median) and third quartiles, as well as limbs that extend to approximately one interquartile range beyond the nearest quartile, and outliers (circles) beyond the limbs. Residuals are raw.

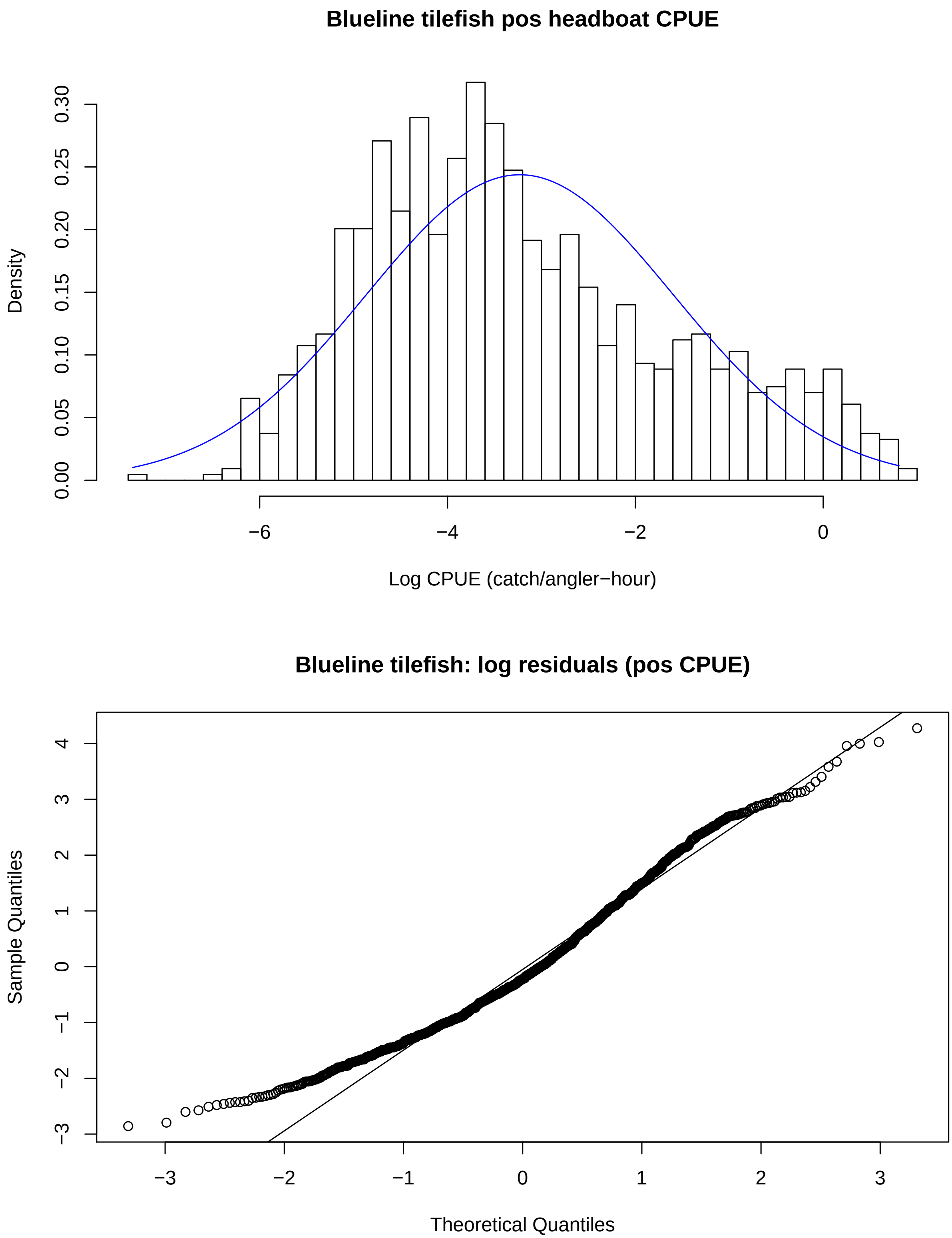


Figure 9: Histogram of empirical log CPUE, with the normal distribution (empirical mean and variance) overlaid. Quantile-quantile plot of residuals from the fitted lognormal submodel to the positive cpue cata.

% latex table generated in R 3.3.2 by xtable 1.8-2 package % Mon Feb 13 13:01:56 2017

Table 1: Number of positive headboat trips by region for blueline tilefish

	year	GULF	N.canaveral	S.canaveral
1	1986.00	7	13	43
2	1987.00	1	13	43
3	1988.00	20	8	43
4	1989.00	47	27	29
5	1990.00	70	33	16
6	1991.00	21	9	39
7	1992.00	30	19	41
8	1993.00	31	20	53
9	1994.00	23	25	22
10	1995.00	13	8	20
11	1996.00	9	7	15
12	1997.00	14	18	18
13	1998.00	8	7	14
14	1999.00	4	7	8
15	2000.00	7	7	12
16	2001.00	8	11	6
17	2002.00	2	5	13
18	2003.00	6	12	5
19	2004.00	8	15	10
20	2005.00	7	15	6

Appendix

Regional and South Atlantic indices

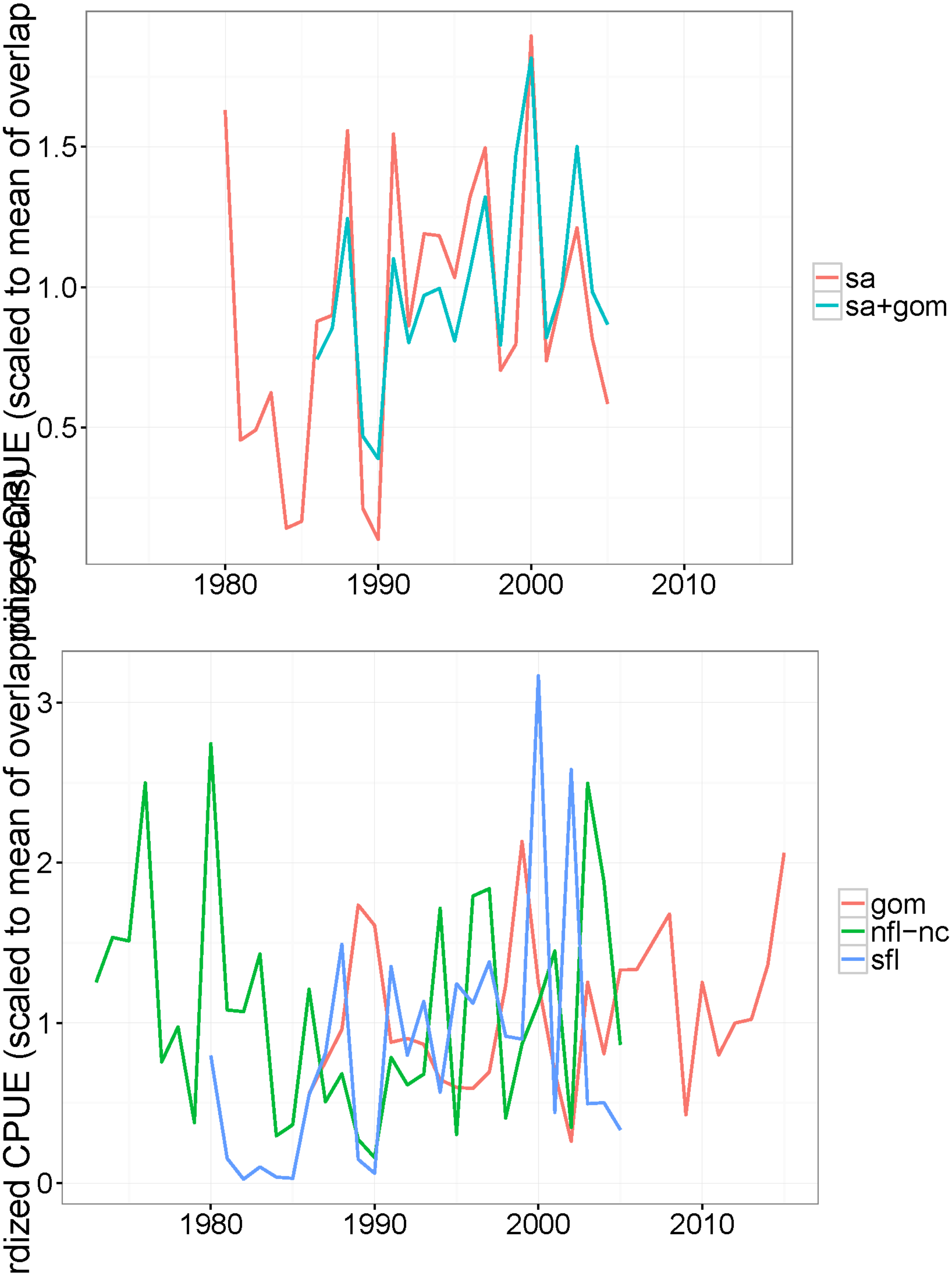


Figure 10: Various headboat indices iterations including south Atlantic only, south Atlantic and Gulf of Mexico and regional indices.

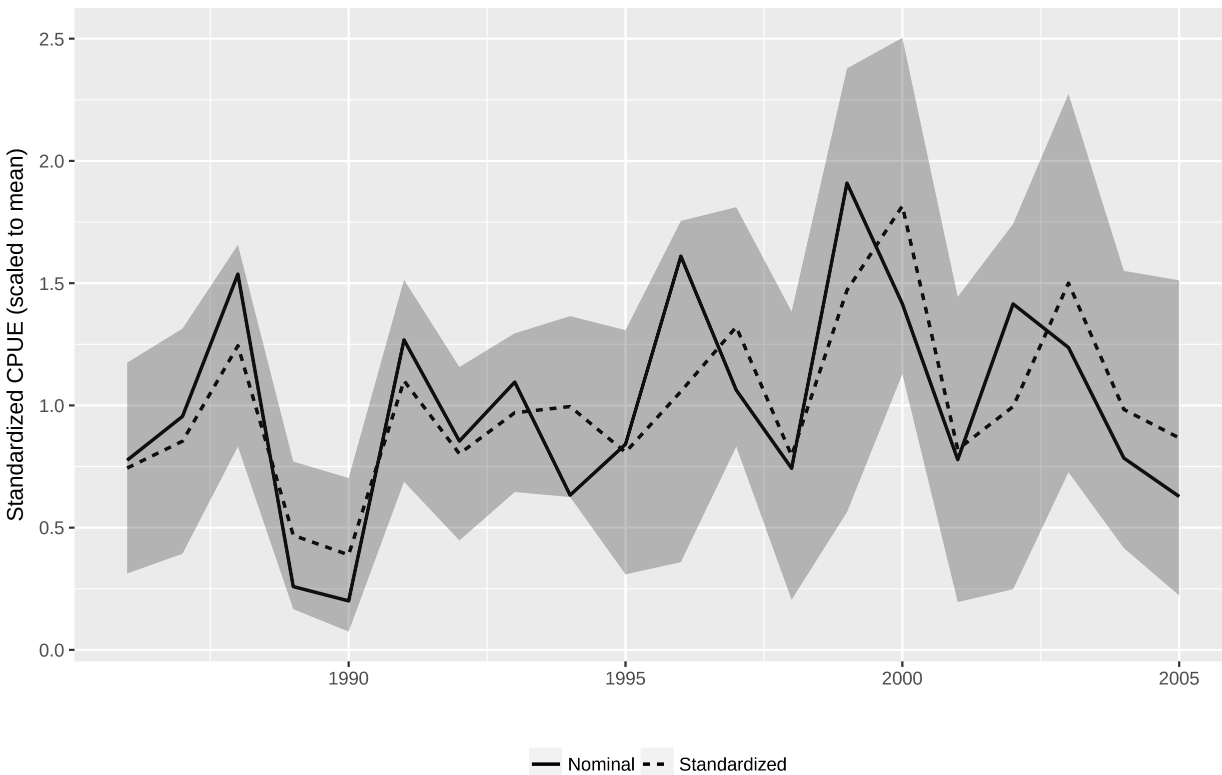


Figure 11: Standardized (dashed) with 95% confidence interval (shaded) and nominal index (solid) blueline tilefish catch rate from headboat logbooks.

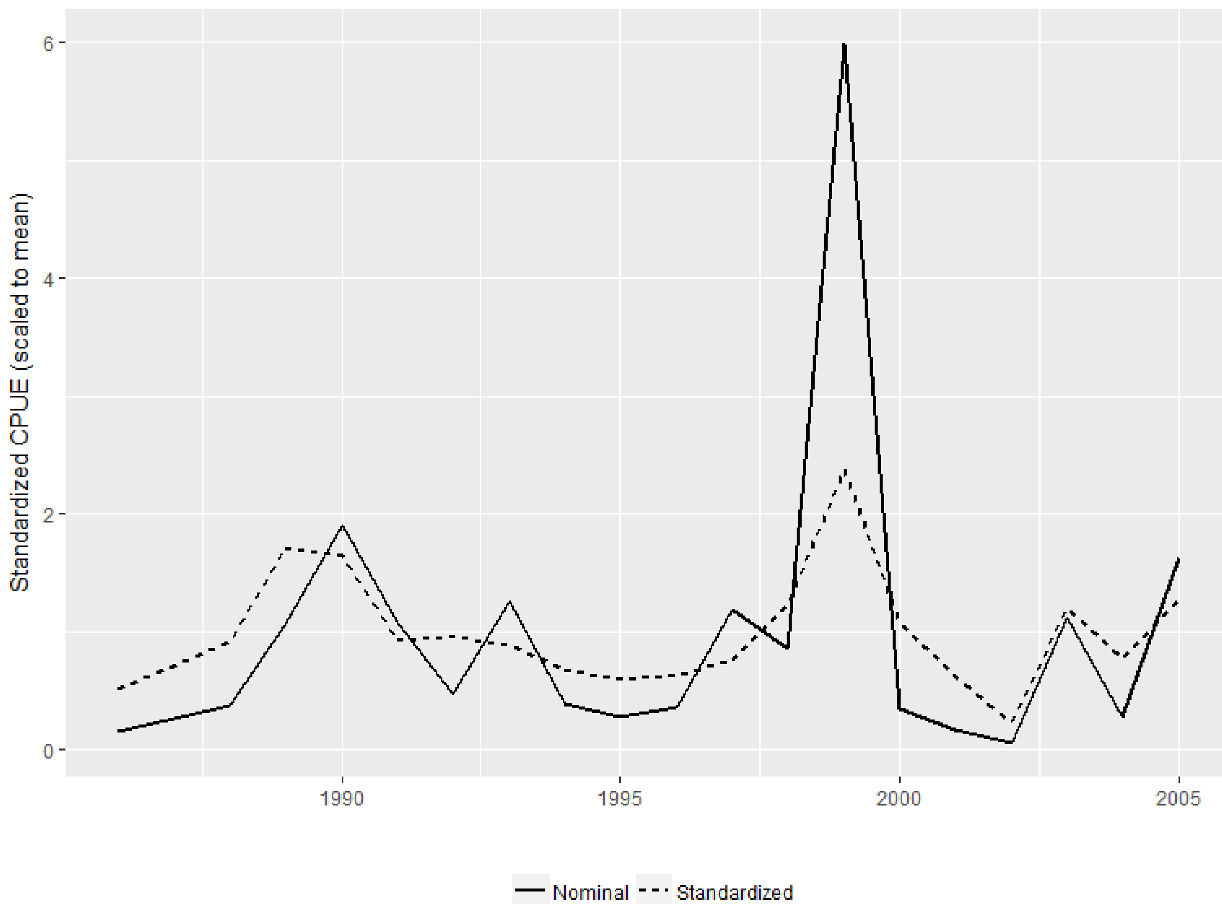


Figure 12: Standardized index of abundance from recreational headboat fleet (Eastern Gulf of Mexico) for blueline tilefish.

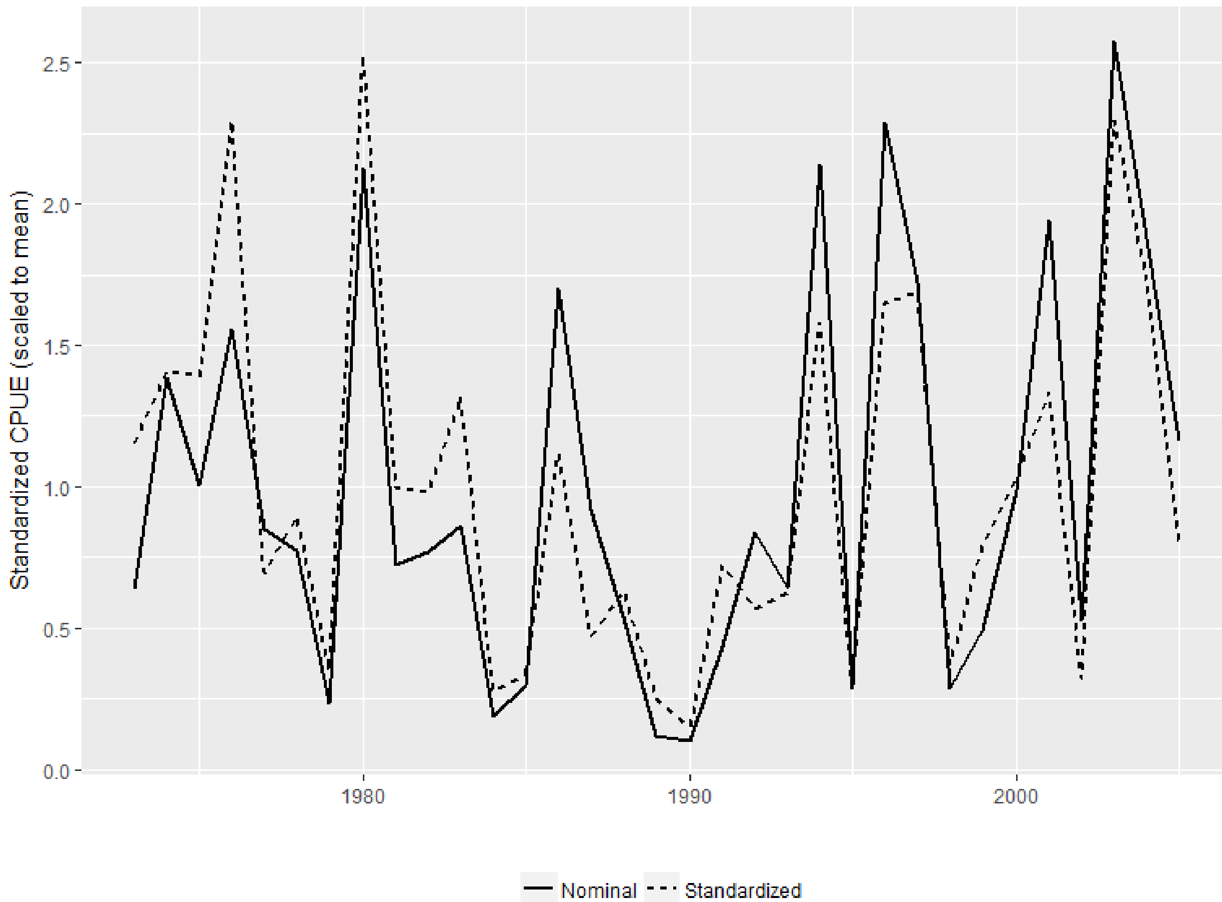


Figure 13: Standardized index of abundance from headboat fleet (North Carolina to Cape Canaveral) for blueline tilefish.

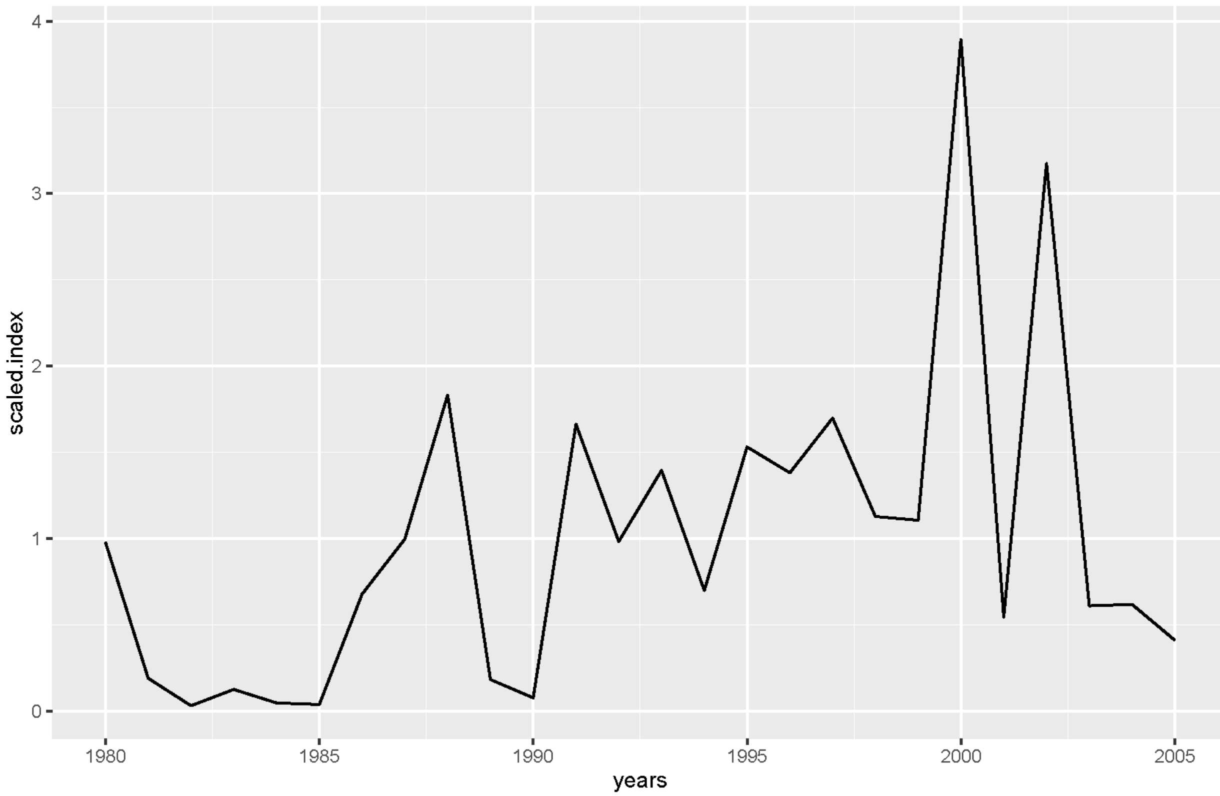


Figure 14: Standardized index of abundance from headboat fleet (Cape Canaveral to Floriday Keys) for blueline tilefish.

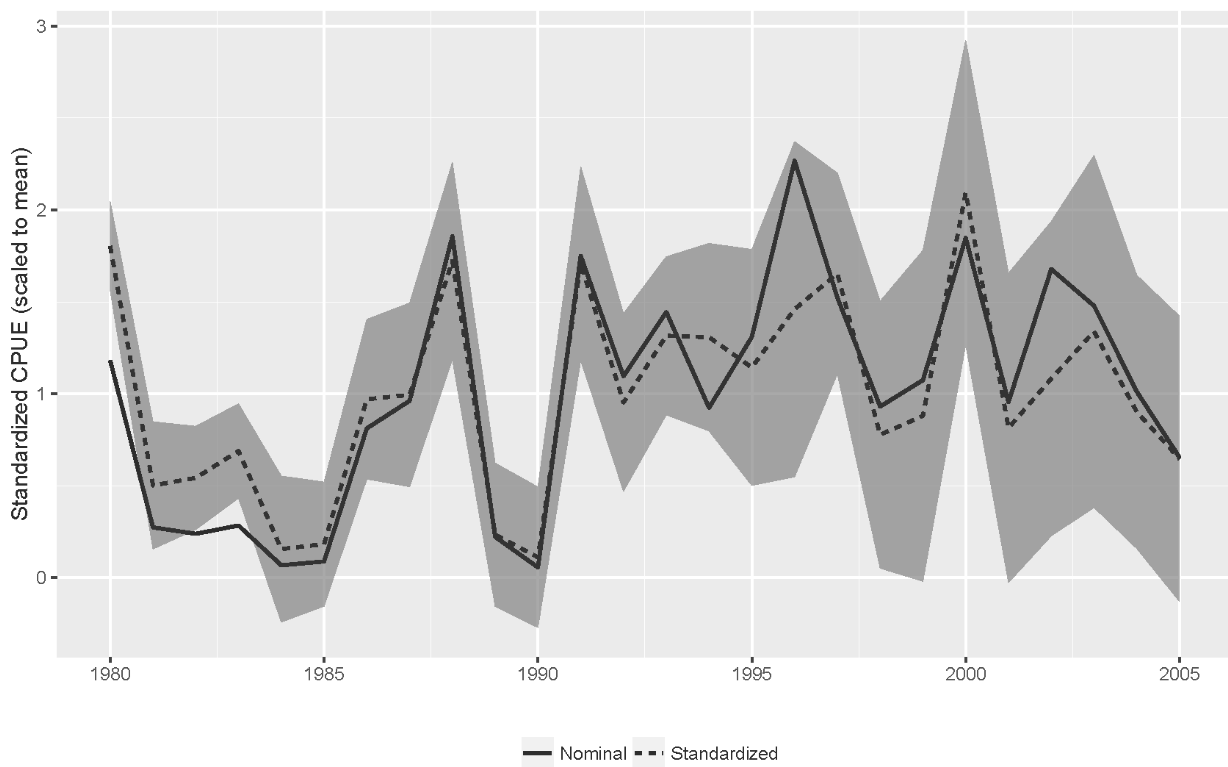


Figure 15: Standardized index of abundance from headboat fleet (North Carolina to Florida Keys) to for blueline tilefish.

Results of lognormal glm to determine factors.

```
##
## Call:  glm(formula = log(cpue) ~ year + zone1 + party, family = gaussian(link = "identity"),
##     data = dat.pos)
##
## Coefficients:
##      (Intercept)          year1987          year1988          year1989
##         -3.41059           0.29726           0.28409          -0.52293
##         year1990          year1991          year1992          year1993
##         -0.73540           0.28829           0.04229           0.08230
##         year1994          year1995          year1996          year1997
##          0.30029           0.11229           0.31530           0.29633
##         year1998          year1999          year2000          year2001
##          0.20359           0.65135           0.71486          -0.15383
##         year2002          year2003          year2004          year2005
##          0.12172           0.45920          -0.11282          -0.13385
## zone1N.canaveral  zone1S.canaveral          party>30
##          0.68606           0.94889          -1.08496
##
## Degrees of Freedom: 1070 Total (i.e. Null);  1048 Residual
## Null Deviance:      2866
## Residual Deviance: 1982  AIC: 3746
```

Results of gamma glm to determine factors.

```
##
## Call:  glm(formula = cpue ~ zone1 + party, family = Gamma(link = "log"),
##     data = dat.pos)
##
## Coefficients:
##      (Intercept)  zone1N.canaveral  zone1S.canaveral          party>30
##           -2.581             1.189             1.376            -1.284
##
## Degrees of Freedom: 1070 Total (i.e. Null);  1067 Residual
## Null Deviance:      3036
## Residual Deviance: 2250  AIC: -2733
```

Results of lognormal delta glm to compare models.

```
## $error.distribution
## [1] "Lognormal distribution assumed for positive observations."
##
## $binomial.formula
## [1] "Formula for binomial GLM: cpue ~ year + zone1 + party"
##
## $positive.formula
## [1] "Formula for gaussian GLM: log(cpue) ~ year + zone1 + party"
##
## $deltaGLM.index
##            index jackknife
## 1986 0.005501340        NA
## 1987 0.006315986        NA
```

```
## 1988 0.009204663        NA
## 1989 0.003466430        NA
## 1990 0.002875600        NA
## 1991 0.008145064        NA
## 1992 0.005934836        NA
## 1993 0.007177637        NA
## 1994 0.007363160        NA
## 1995 0.005979854        NA
## 1996 0.007819562        NA
## 1997 0.009772802        NA
## 1998 0.005876659        NA
## 1999 0.010882459        NA
## 2000 0.013433821        NA
## 2001 0.006066000        NA
## 2002 0.007362572        NA
## 2003 0.011098098        NA
## 2004 0.007275391        NA
## 2005 0.006413755        NA
##
## $pos.effects
## $pos.effects[[1]]
##        GULF N.canaveral S.canaveral
##  0.05601931  0.11124787  0.14468833
##
## $pos.effects[[2]]
##        <30        >30
## 0.16619471 0.05615969
##
##
## $bin.effects
## $bin.effects[[1]]
##        GULF N.canaveral S.canaveral
##  0.06777198  0.08945396  0.06199860
##
## $bin.effects[[2]]
##        <30        >30
## 0.07505834 0.06951732
##
##
## $data.filter
## [1] "Data filter threshold set at 2 positive observations."
##
## $levels.deleted.by.filter
## $levels.deleted.by.filter$year
## [1] NA
##
## $levels.deleted.by.filter$zone1
## [1] NA
##
## $levels.deleted.by.filter$party
## [1] NA
##
##
## $aic
```

```
##                        [,1]
## AIC.binomial   7815.832855
## AIC.lognormal -3173.489818
## sigma.mle         1.360288
```

Results of gamma delta glm to compare models.

```
## $error.distribution
## [1] "Gamma distribution assumed for positive observations."
##
## $binomial.formula
## [1] "Formula for binomial GLM: cpue ~ year + zone1 + party"
##
## $positive.formula
## [1] "Formula for Gamma GLM: cpue ~ year + zone1 + party"
##
## $deltaGLM.index
##           index jackknife
## 1986 0.07930796        NA
## 1987 0.10910939        NA
## 1988 0.13108456        NA
## 1989 0.04933092        NA
## 1990 0.06126320        NA
## 1991 0.11531449        NA
## 1992 0.08521518        NA
## 1993 0.10827124        NA
## 1994 0.08989473        NA
## 1995 0.08738168        NA
## 1996 0.15391193        NA
## 1997 0.10456545        NA
## 1998 0.08175282        NA
## 1999 0.21497125        NA
## 2000 0.14142657        NA
## 2001 0.07362923        NA
## 2002 0.09195037        NA
## 2003 0.15533973        NA
## 2004 0.08395379        NA
## 2005 0.08428366        NA
##
## $pos.effects
## $pos.effects[[1]]
##        GULF N.canaveral S.canaveral
##  0.04934387  0.12709582  0.15569151
##
## $pos.effects[[2]]
##        <30        >30
## 0.18196416 0.05408794
##
##
## $bin.effects
## $bin.effects[[1]]
##        GULF N.canaveral S.canaveral
##           1           1           1
##
```

```
## $bin.effects[[2]]
## <30 >30
##   1   1
##
##
## $data.filter
## [1] "Data filter threshold set at 2 positive observations."
##
## $levels.deleted.by.filter
## $levels.deleted.by.filter$year
## [1] NA
##
## $levels.deleted.by.filter$zone1
## [1] NA
##
## $levels.deleted.by.filter$party
## [1] NA
##
##
## $aic
##                       [,1]
## AIC.binomial    46.0000000
## AIC.gamma    -2787.3059687
## shape.mle        0.6134443
```

References

Dick, E.J. 2004. "Beyond Lognormal Versus Gamma: Discrimination Among Error Distributions for Generalized Linear Models." *Fisheries Research* 70: 351–66.

Efron, B., and R. Tibshirani. 1993. *An Introduction to the Bootstrap*. London: Chapman; Hall.

Lo, N., Jacobson L., and Squire J. 1992. "Indices of Relative Abundance from Fish Spotter Data Based on Delta-Lognormal Models." *Canadian Journal of Fisheries and Aquatic Sciences* 49: 2515–26.

Maunder, M., and A. Punt. 2004. "Standardizing Catch and Effort Data: A Review of Recent Approaches." *Fisheries Research* 70: 141–59.

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