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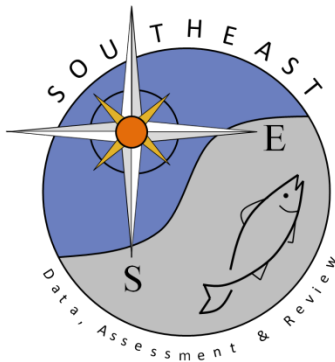 Standardized catch rates of blueline tilefish (*Caulolatilus microps*) in the South Atlantic and Gulf of Mexico waters of the U.S. from commercial logbook handline data

Sustainable Fisheries Branch – NMFS (contact: Eric Fitzpatrick)

SEDAR50-DW26

Submitted: 13 January 2017

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 commercial logbook handline data. SEDAR50-DW26. SEDAR, North Charleston, SC. 32 pp.

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

Sustainable Fisheries Branch[*]

February 2017

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

Commercial logbook data & Management issues to consider

1. A deep-water closure, primary habitat of blueline tilefish, at the end of January, 2011 prevented use of the 2011 data.
2. Snowy regulations in the area south of Cape Canaveral since the mid 2000s (these data were filtered in SEDAR 32)
3. Due to increase targeting of blueline tilefish north of Cape Hatteras, NC, these data were filtered similarly to SEDAR 32
4. Due to the IFQ in the Gulf tilefish fishery (implemented in January 2010), Gulf data during this time period is problematic and should be discussed

- see Gulf of Mexico 2015 Grouper-Tilefish IFQ Annual Report
- Figure 3, blueline tilefish catch pre-IFQ was approximately 80% then approximately 20% after IFQ
- Table 32, market conditions (adjusted ex-vessel price for golden tilefish) had impact on blueline tilefish landings

5. Additional data variables added to commercial logbook data since SEDAR 32

- Outlier_flag: outlier flag=1 if effort data (numgear, effort, away, crew) reported for gear fished during a trip were outside the 99.5 percentile of all gear-specific reports or were below the 0.5 percentile for length of longlines and gillnets or for longline effort (i.e., number of hooks per set) Avail_gag - avail_red_snapper: indicates the amount of IFQ (pounds) for each species available to the vessel and the end of each trip prior to unloading the catch
- Pounds: the total pounds landed of all species on the trip (used to estimate targeting)
- Pct_shk: the percent of sharks, by weight, of the total landings from the trip
- Pct_otr: the percent of other species (not shark, deep water grouper and tilefish, or shallow water grouper and snapper), by weight, of the total landings from the trip
- Pct_dw: the percent of deep water grouper and tilefish, by weight, of the total landings from the trip
- Pct_sw: the percent of shallow water grouper and snapper, by weight, of the total landings from the trip
- Target: the presumptive target species group for the trip; a species group was presumed to be targeted if landings of that species group accounted for 66.6% or more of the total landings from the trip; if no species group accounted for 66.6% or more, the target was labeled as mixed
- see Appendix for all variables

---

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

Effort

Commercial handline effort by gear (Figure 2) and proportion positive blueline tilefish trips (Figure 3). Additionally, 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 effort in these deepwater habitats.

Data considerations (decisions in bold)

IWG Decision: 5 alternative headboat indices were recommended. Two indices were developed at the coastwide level (with and without the Gulf of Mexico data) and 3 'regional' indices

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. Starting year

- **1993**

3. Terminal year

- **2007 for all indices in South Atlantic due to increased targeting after 2007 due to management regulations on other deepwater species (i.e., snowy grouper)**
- **2009 (Gulf only)**

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 handline index will exclude trips with the largest 0.5% values for catch in number (>) and cpue (> 7.28) for trips that caught blueline tilefish. The number of anglers on a trip can also influence cpue when calculated as fish/number of hooks*hours. Trips with the largest 0.5% values for reported anglers (> 5) were removed (Figure 5).

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 6.

Catch per unit effort= total pounds/number of lines*number of hooks*hours fished

Evaluation of explanatory variables

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

2. AREA

   - Area between North Carolina to Cape Canaveral (atl)
   - South Florida (sFL)
   - Gulf (gom)
   - Greater Atlantic Region ('VTR' data) 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.

3. SEASON

   - Winter (January - March)
   - Spring (April - June)
   - Summer (July - September)
   - Fall (October - December)

The patterns in the remaining positive blueline tilefish trips by month and region show that most trips occur in the Atlantic and Gulf region between May and August (Figure 7). The seasonal pattern in cpue across months seems consistent across areas (Figure 8 and Figure 9). Season was chosen as an explanatory variable.

4. Crew size (crewsize) -Crew size (crewsize) were pooled into two levels: one to two (1-2), and three or more (3plus) (Figure 10).

5. Days at sea - Days at sea were pooled into three levels: one or two days (1-2), three or four days (3-4) and more than five days (5plus) (Figure 11).

6. Proportion deepwater species - $< 50\%$ deepwater species or $>50\%$ deepwater species

Analytical decisions

1. Use Stephens and Maccall method

   - **model all trips that caught deepwater complex species**

2. Interaction terms

   - **no interaction terms**

3. CPUE units and transformation

   - **use catch/hook-hour for CPUE unit**

4. Weighting by region

   - **No weighting (unnecessary if no area*month interaction)**

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 variancewere 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 12 and Figure 14).

Results

The standardized index was similar to the nominal index with the exception of a few years associated with peaks in the catch rate (Figure 16).

Additional indices results are located in the appendix.

[1] "0 (total) records were removed by filter." [1] "0 positive records removed by filter."
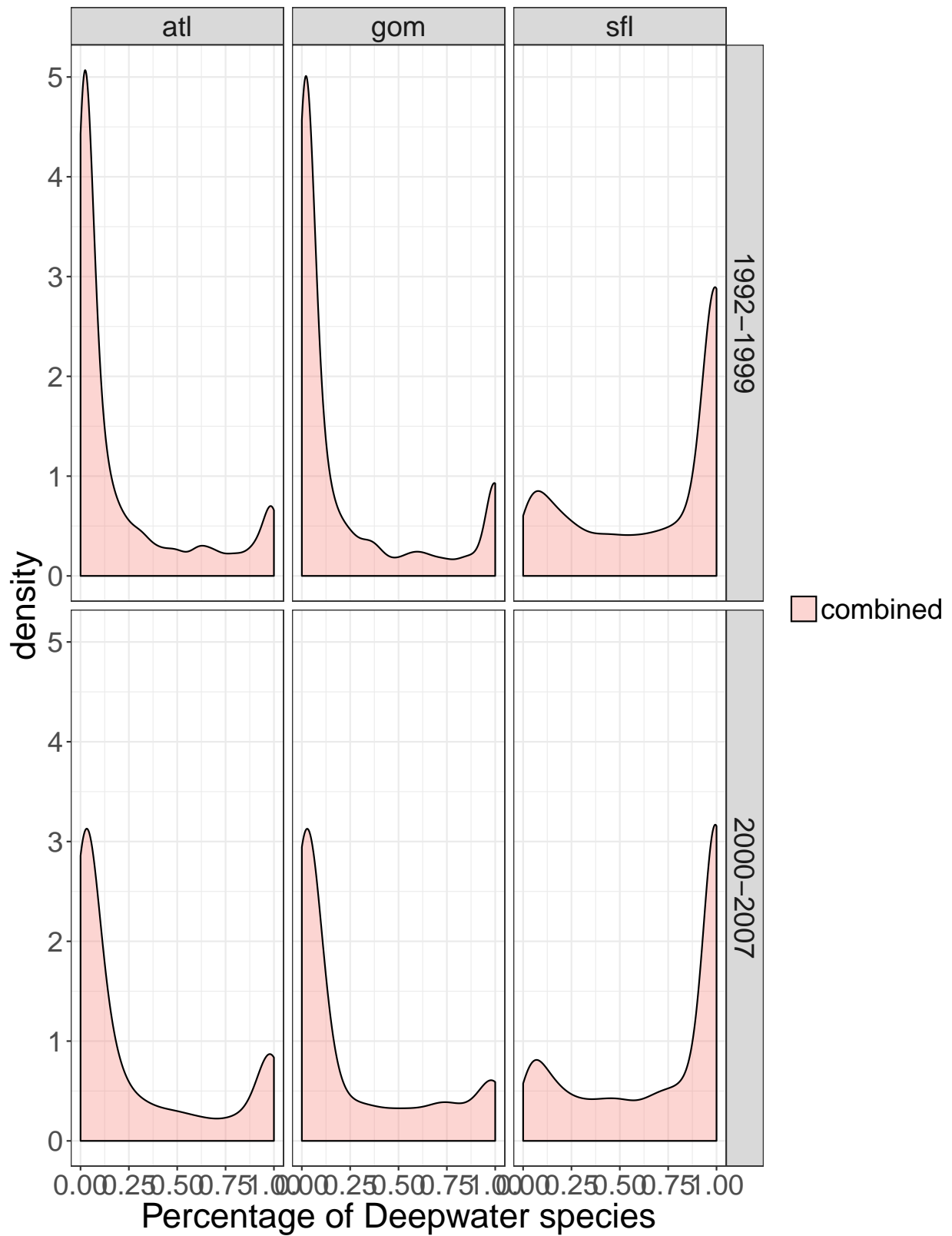
Figure 1: Density of percentage of deepwater species from commercial logbook by timeblock.
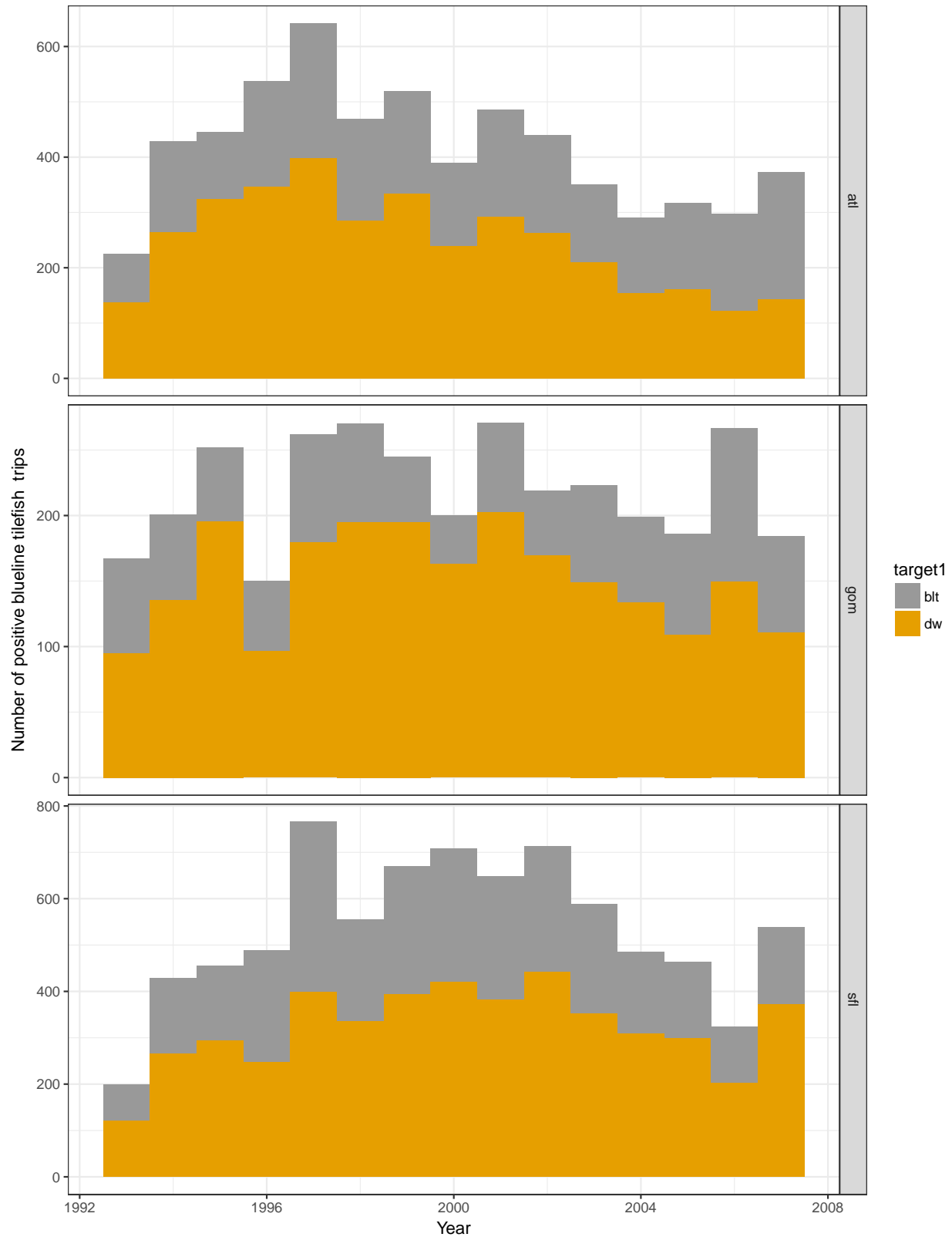
Figure 2: Number of positive blueline tilefish commercial logbook trips that submitted logbooks 1993-2015 for handline and bandit gear by region (atl-Atlantic (excluding south Florida and north of Hatteras) and gom - Gulf).
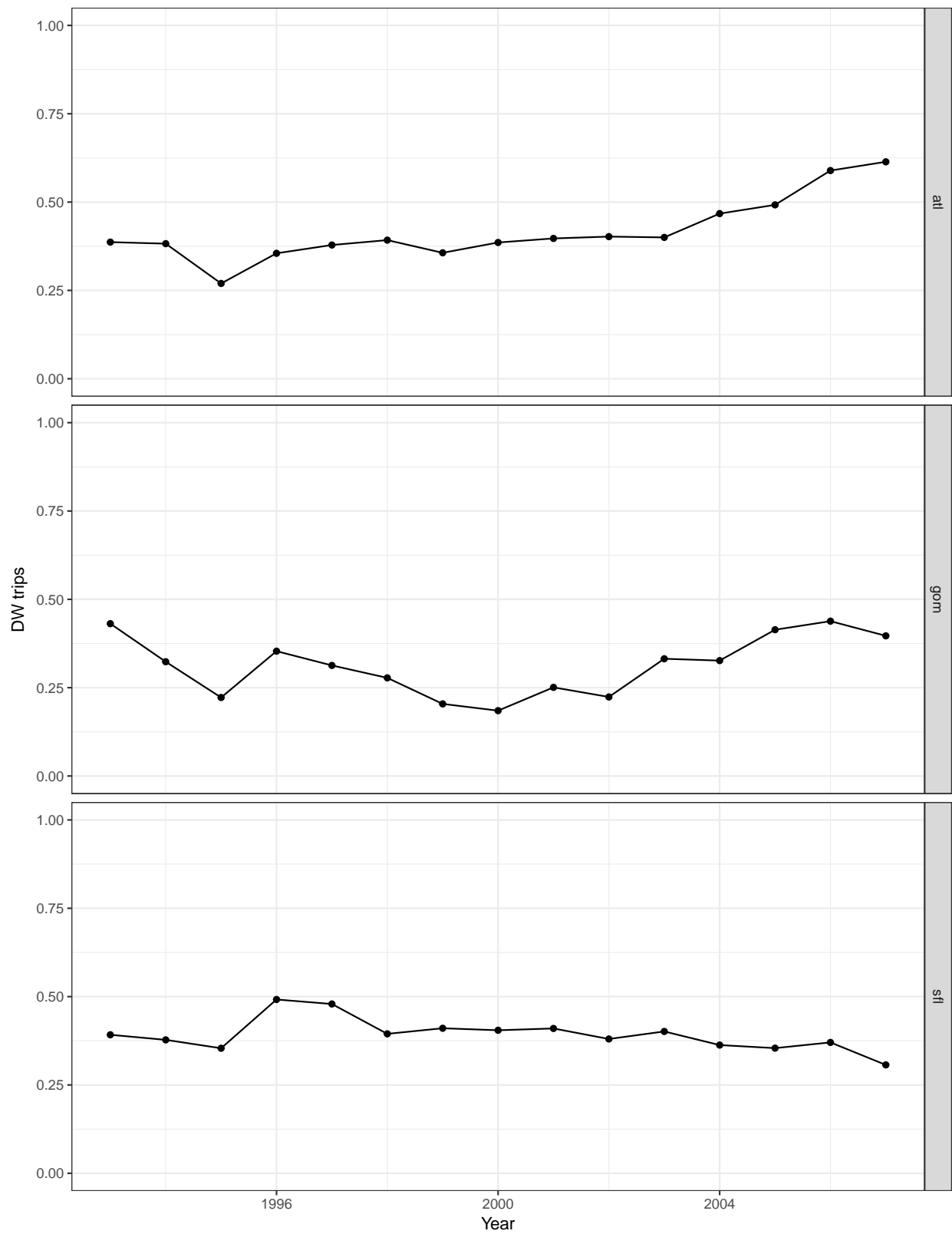
Figure 3: Proportion positive of (blueline trips/total DW trips) using deepwater complex species.
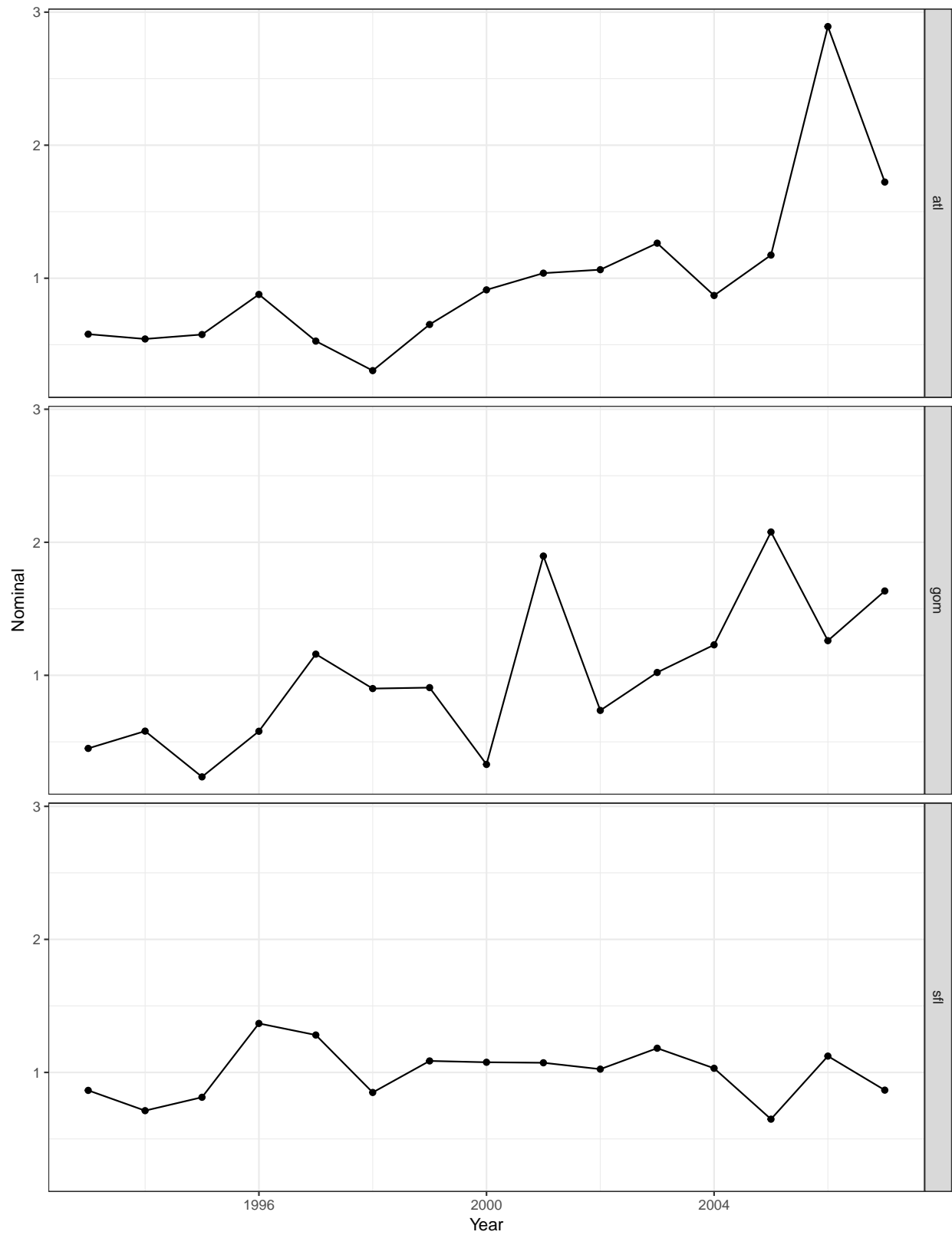
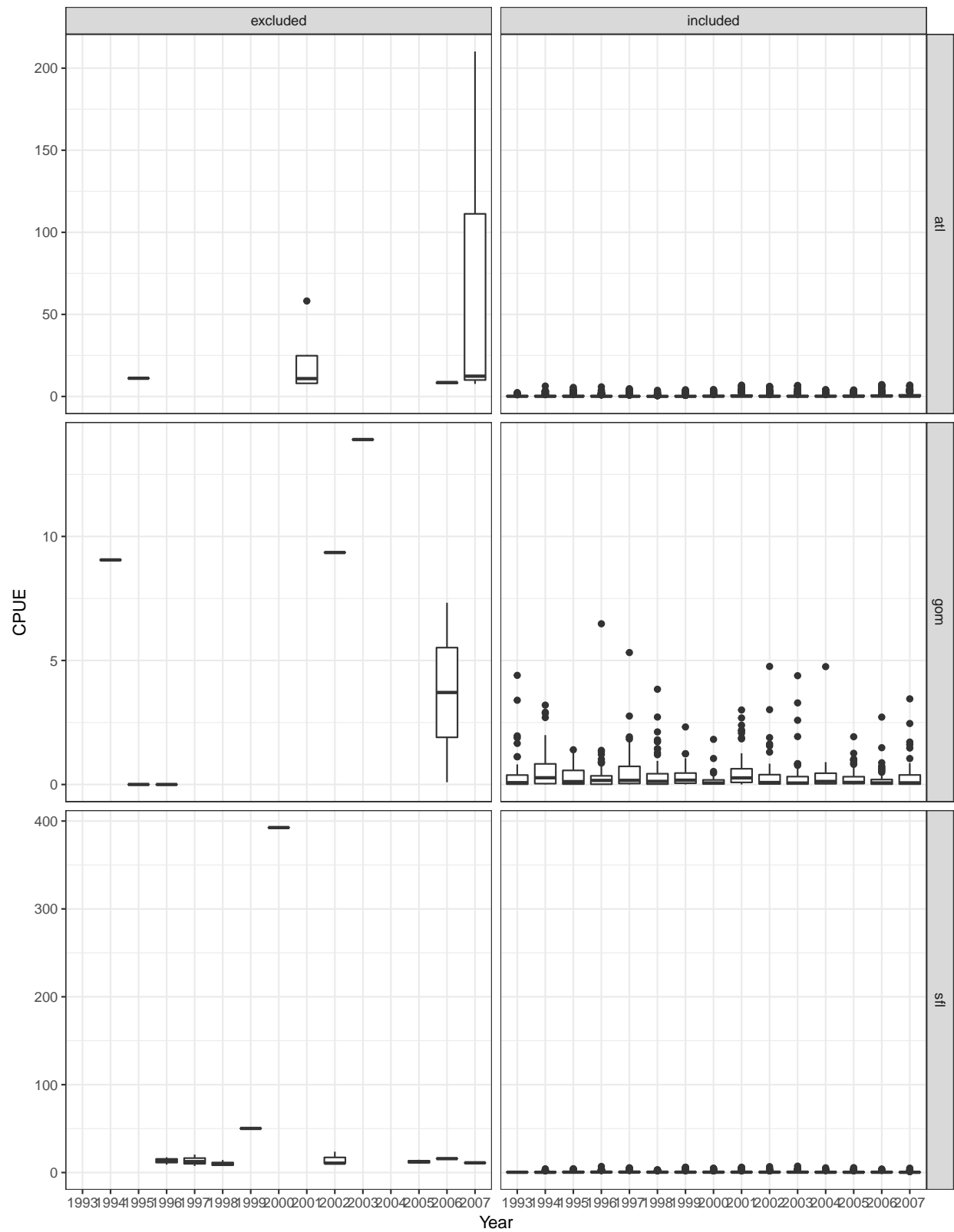Figure 4: Nominal blueline tilefish catch per unit effort.

Figure 5: Records determined as outliers (excluded) based on removal of values above the 99.5th percentile for crew, number of fish caught (totlbs), and cpue.
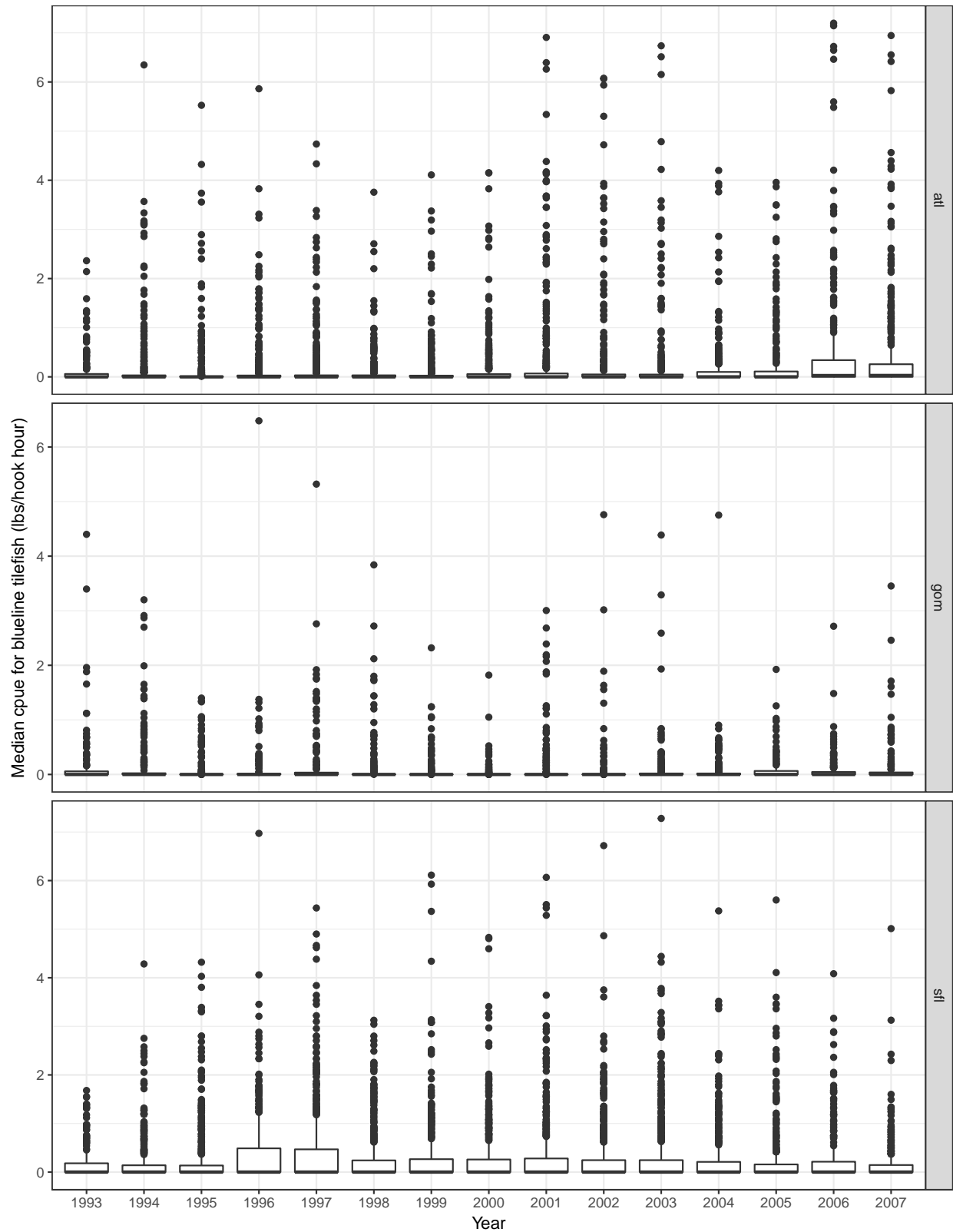
Figure 6: Blueline tilefish cpue by region (atl-Atlantic (excluding south Florida and north of Hatteras) and gom - Gulf) and season. Multi-day and 3/4-day trips are removed for all regions. Half-day trips are removed for CAR and GNF. Half-day trips and full-day trips are aggregated for SF. Years are limited to 1980-2015.
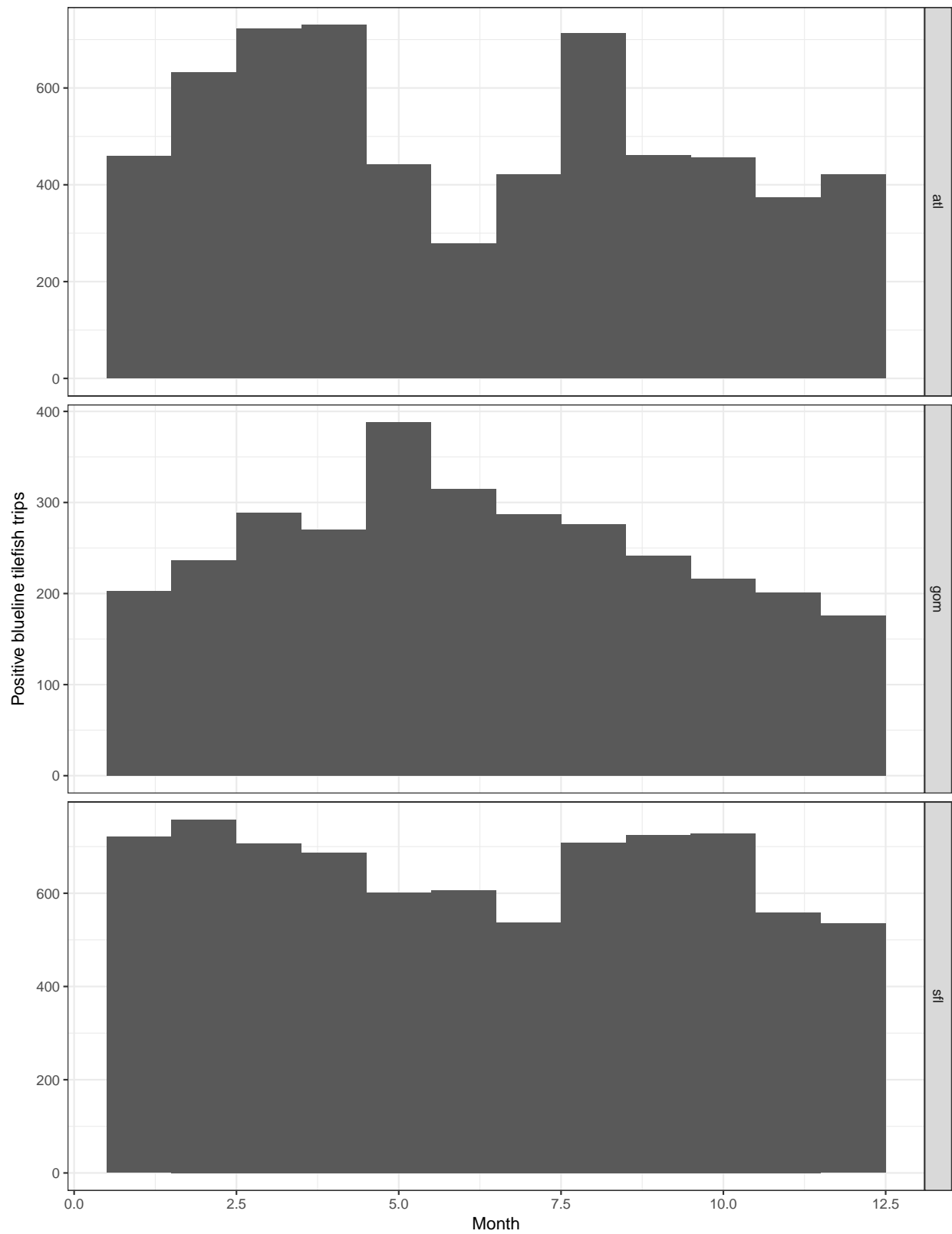
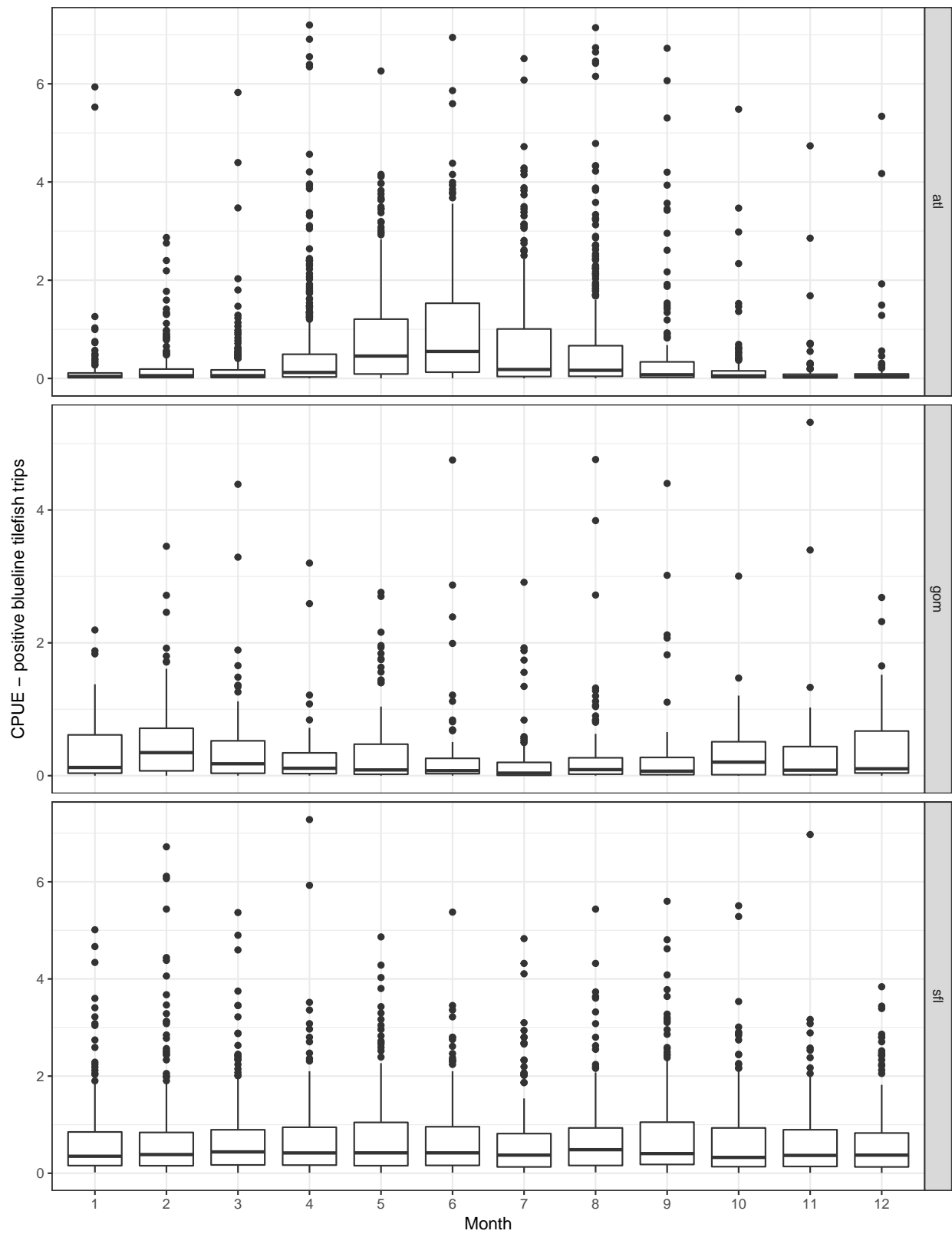Figure 7: Blueline tilefish trips by month and region.

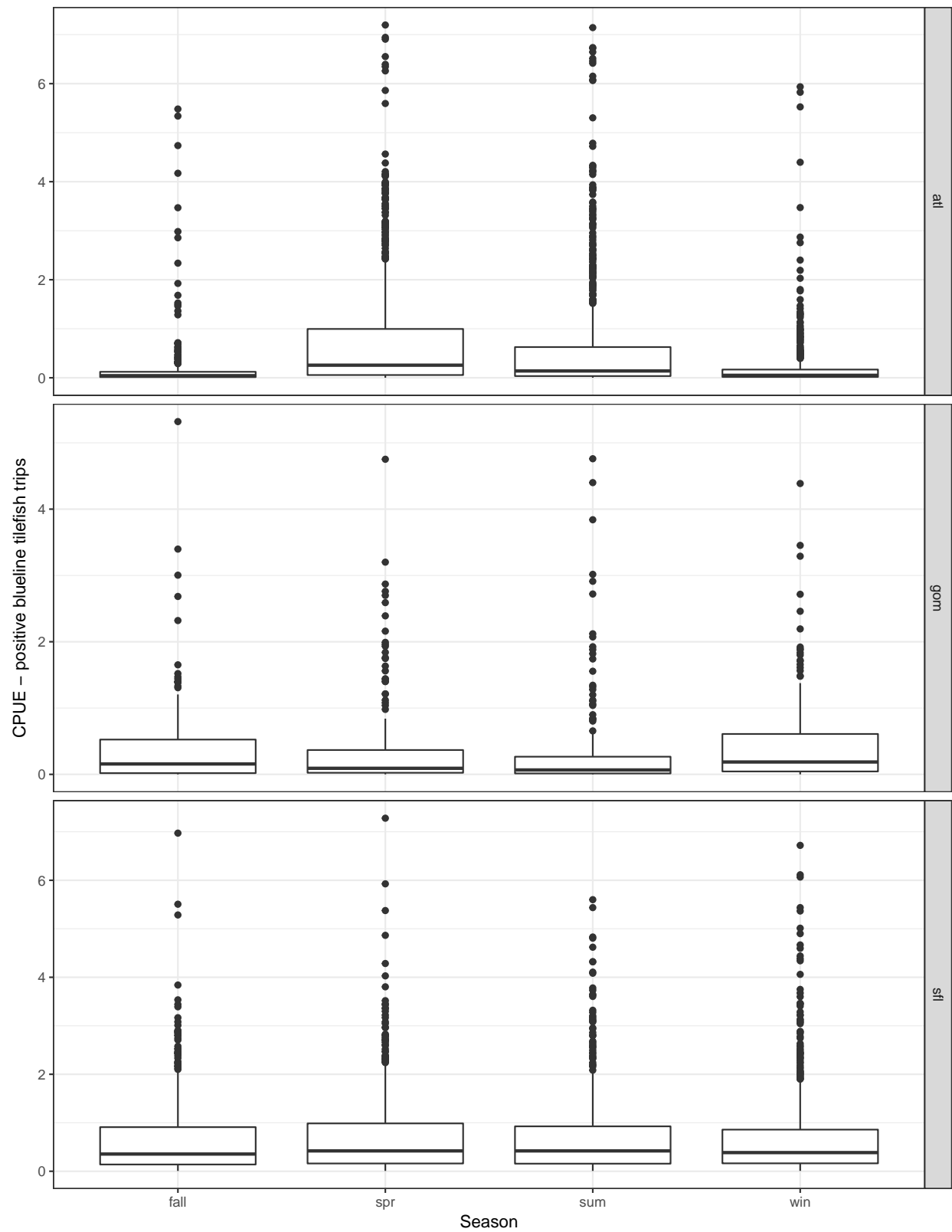Figure 8: Blueline tilefish cpue by month and region.

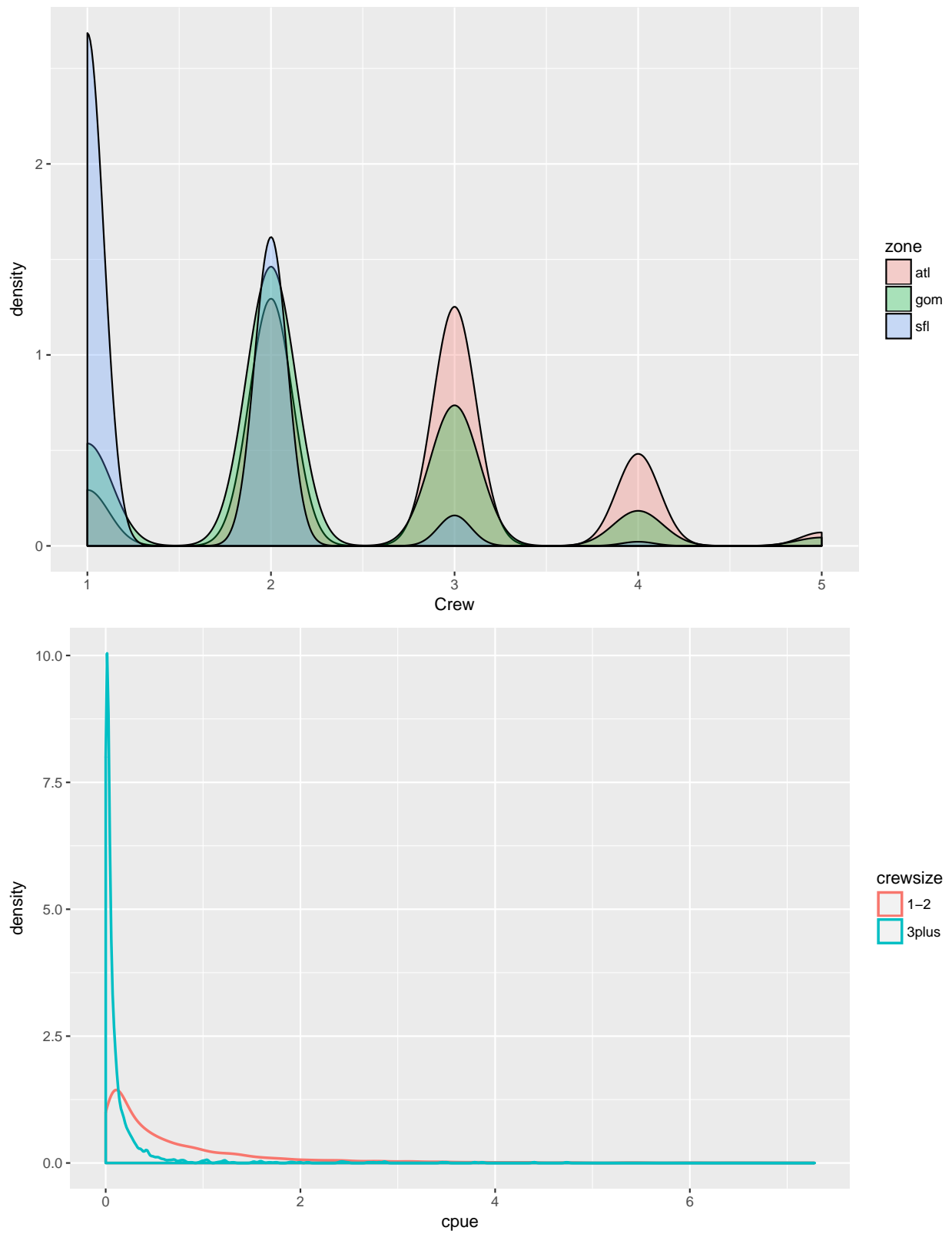Figure 9: Blueline tilefish cpue by season and region.

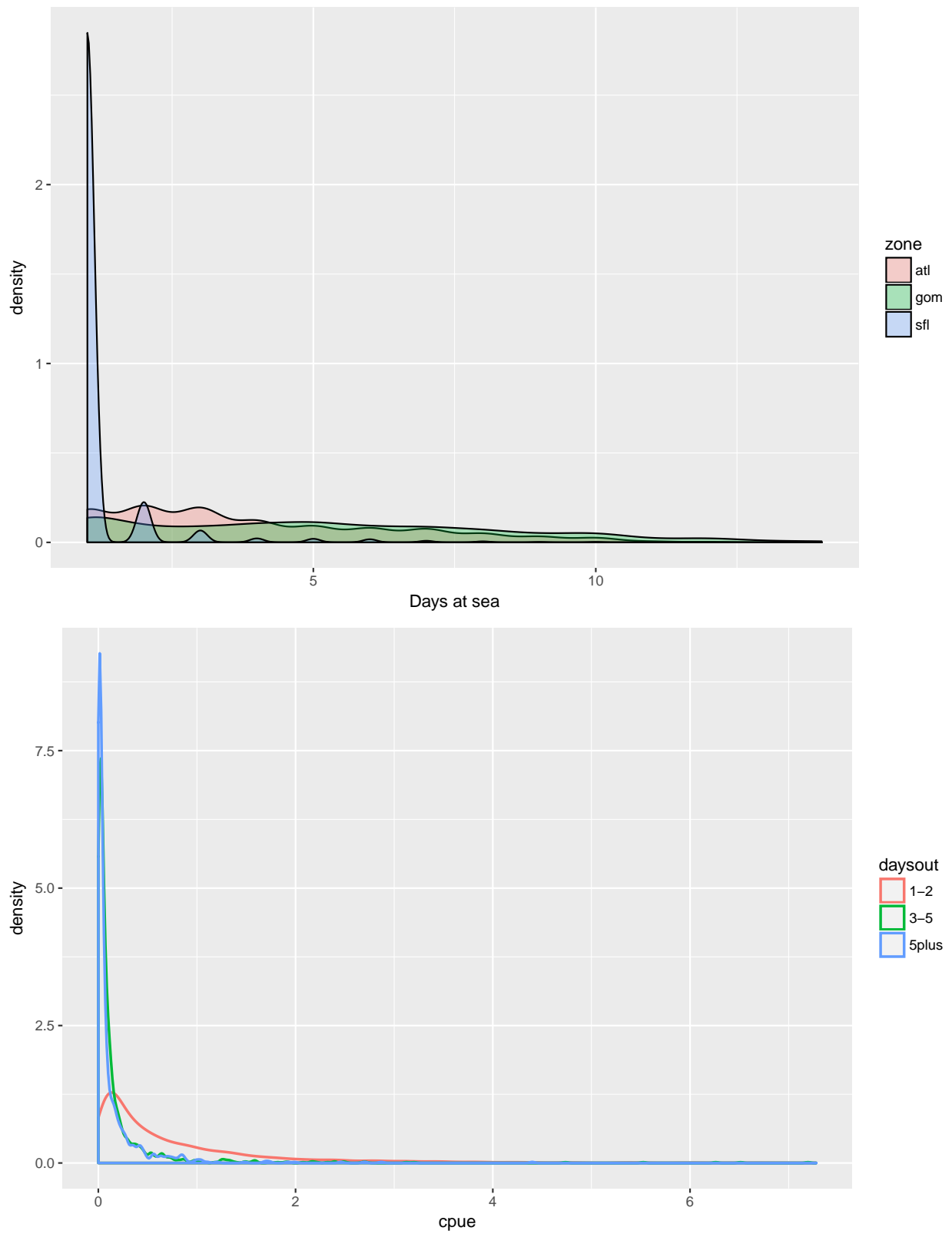Figure 10: Density of crewsize across areas and cpue.

Figure 11: Density of days at sea across areas and cpue.

Blueline tilefish pos handline CPUE

Raw residuals (pos CPUE)

Raw residuals (pos CPUE)

Figure 12: 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.

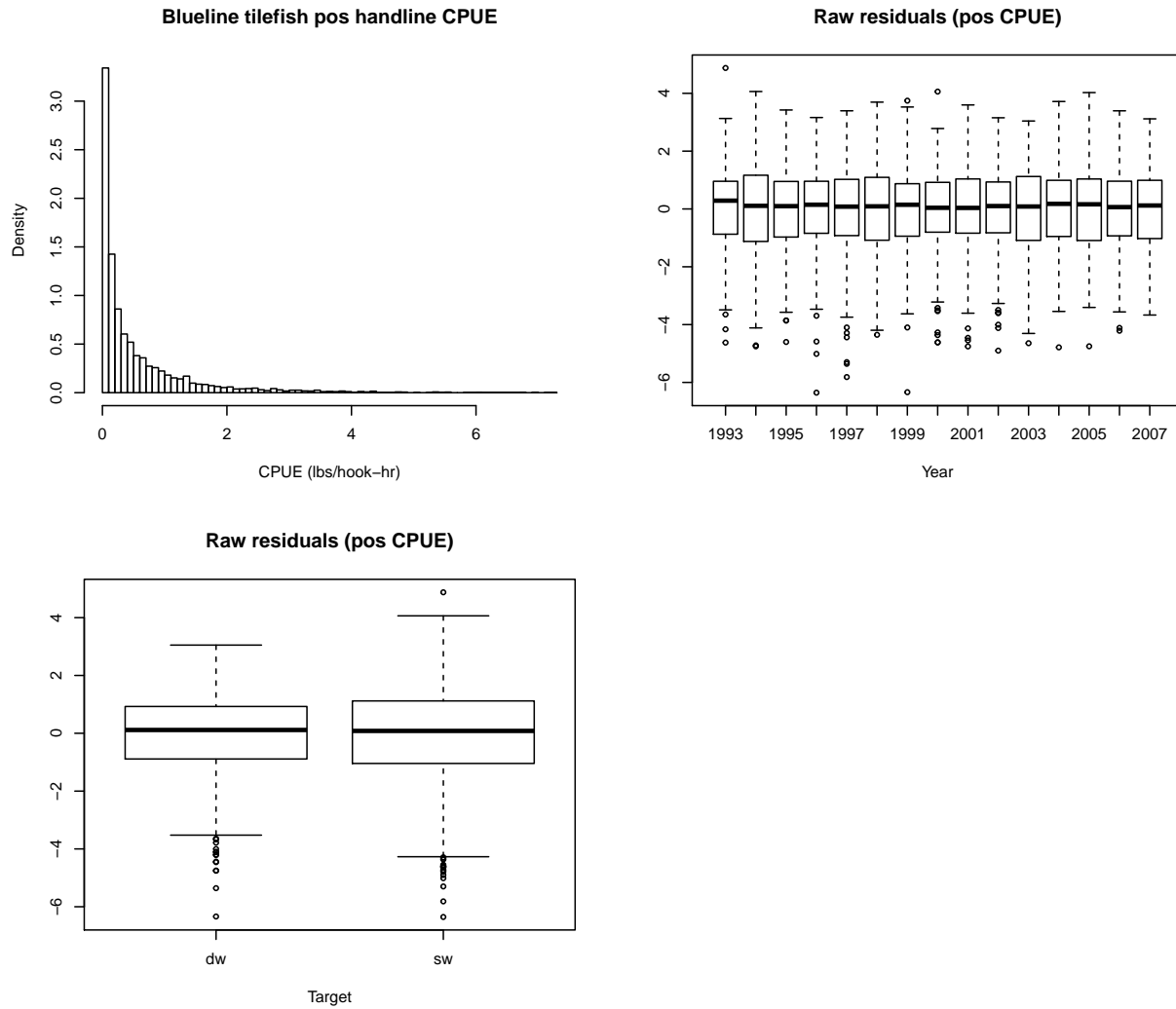Figure 13: Pearsons Diagnostics of lognormal submodel fits to positive CPUE data. Top left panel shows the distribution of positive cpue. Box and whisker plots give first, second (median) and third quartiles, as well as limbs that extend to approximately one interquartile range beyond the nearest quartile, and outliers (circles) beyond the limbs. Residuals are raw.
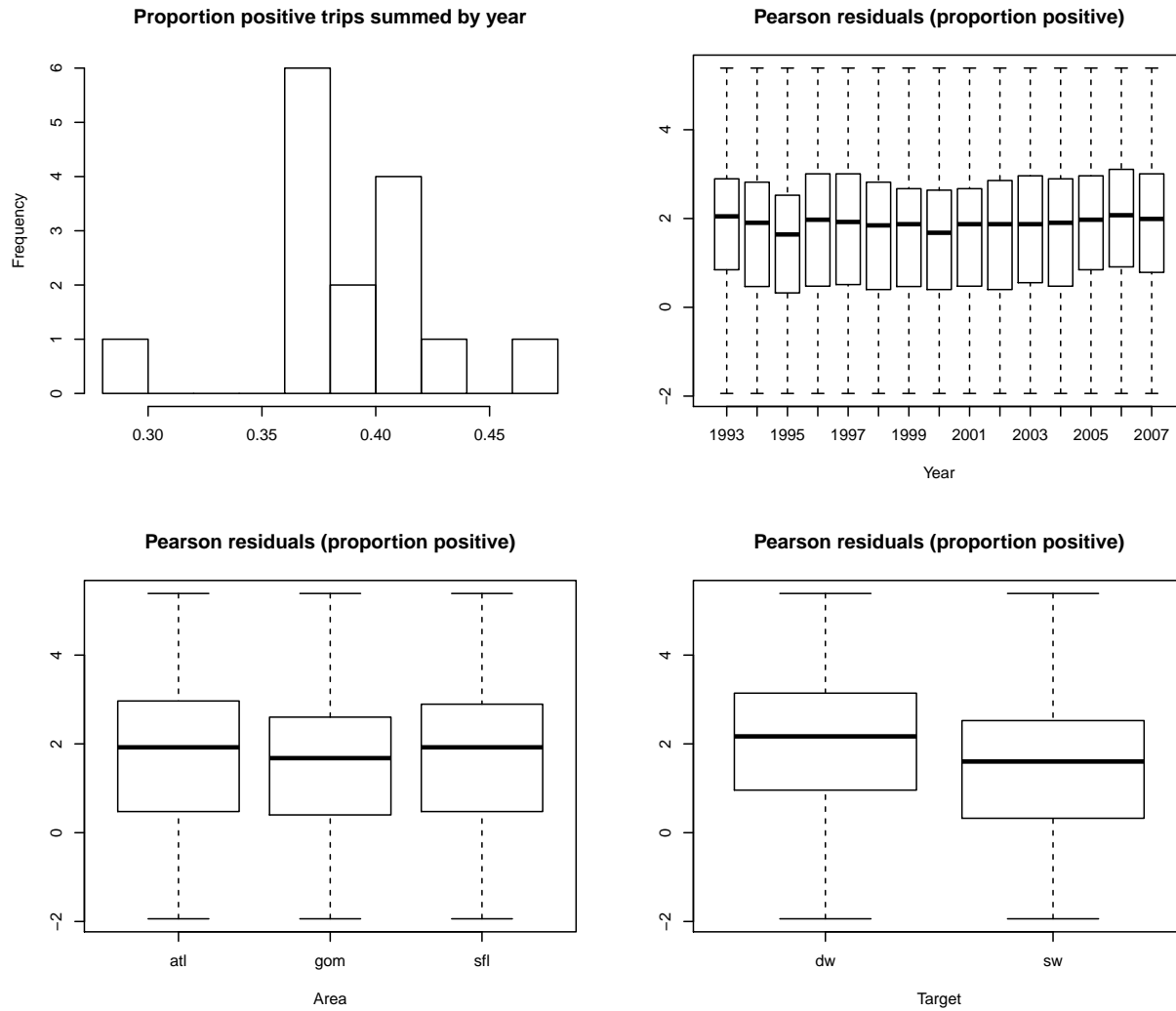
Blueline tilefish pos handline CPUE

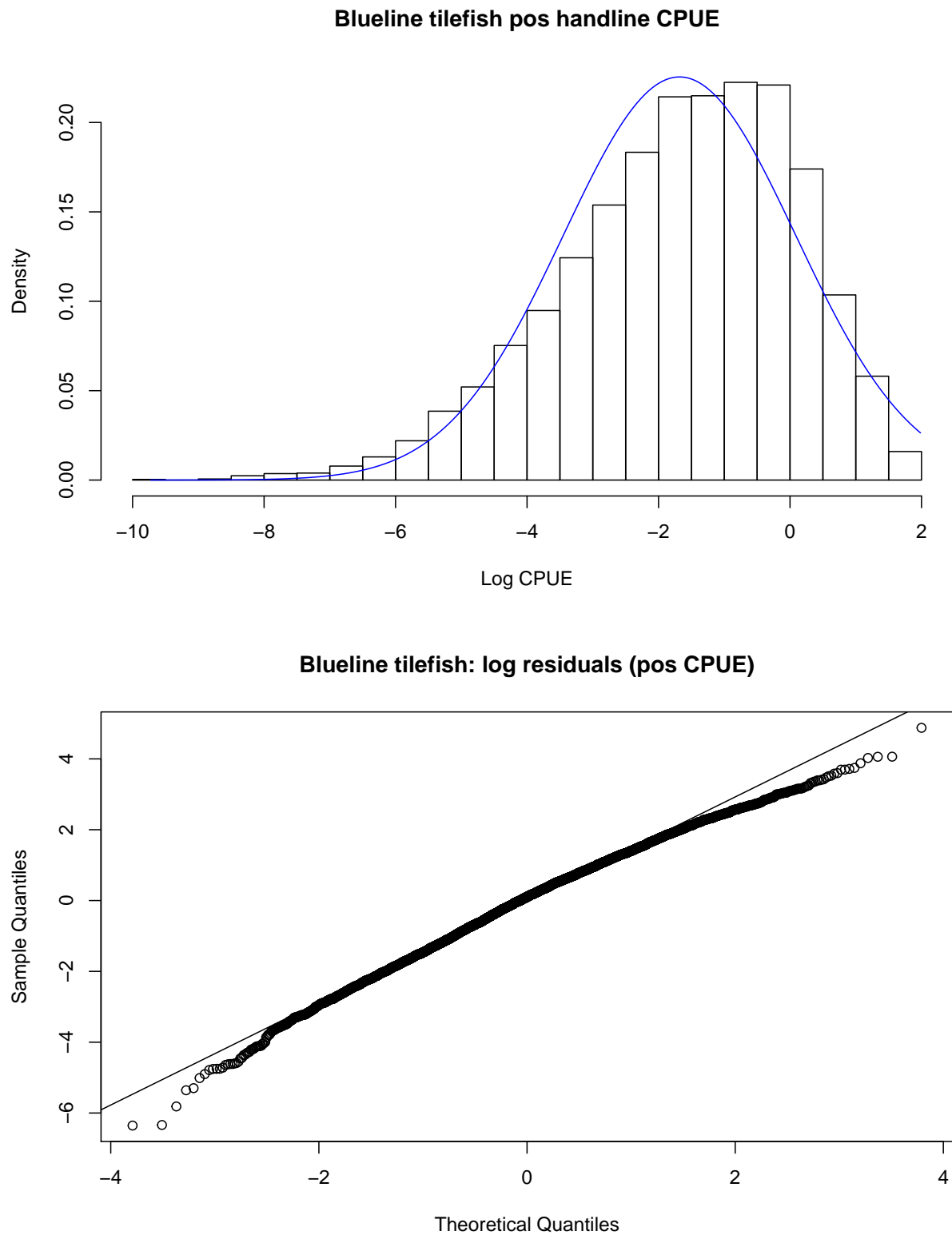**Blueline tilefish: log residuals (pos CPUE)**

Figure 14: 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 12:50:10 2017

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

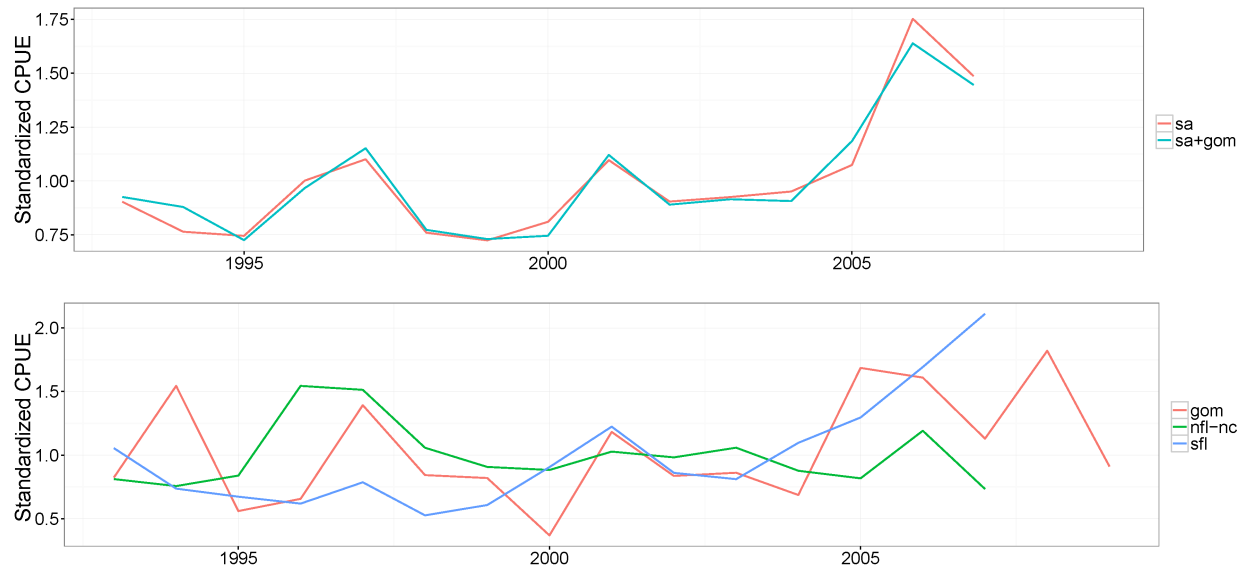|	year	atl	gom	sfl
1	1993.00	82	71	75
2	1994.00	164	61	160
3	1995.00	120	54	158
4	1996.00	189	51	237
5	1997.00	240	80	364
6	1998.00	181	72	218
7	1999.00	184	50	274
8	2000.00	150	35	286
9	2001.00	192	64	264
10	2002.00	177	49	271
11	2003.00	140	74	236
12	2004.00	136	63	176
13	2005.00	156	75	161
14	2006.00	173	111	117
15	2007.00	219	71	163

Figure 15: Various handline indices iterations including south Atlantic only, south Atlantic and Gulf of Mexico and regional indices.
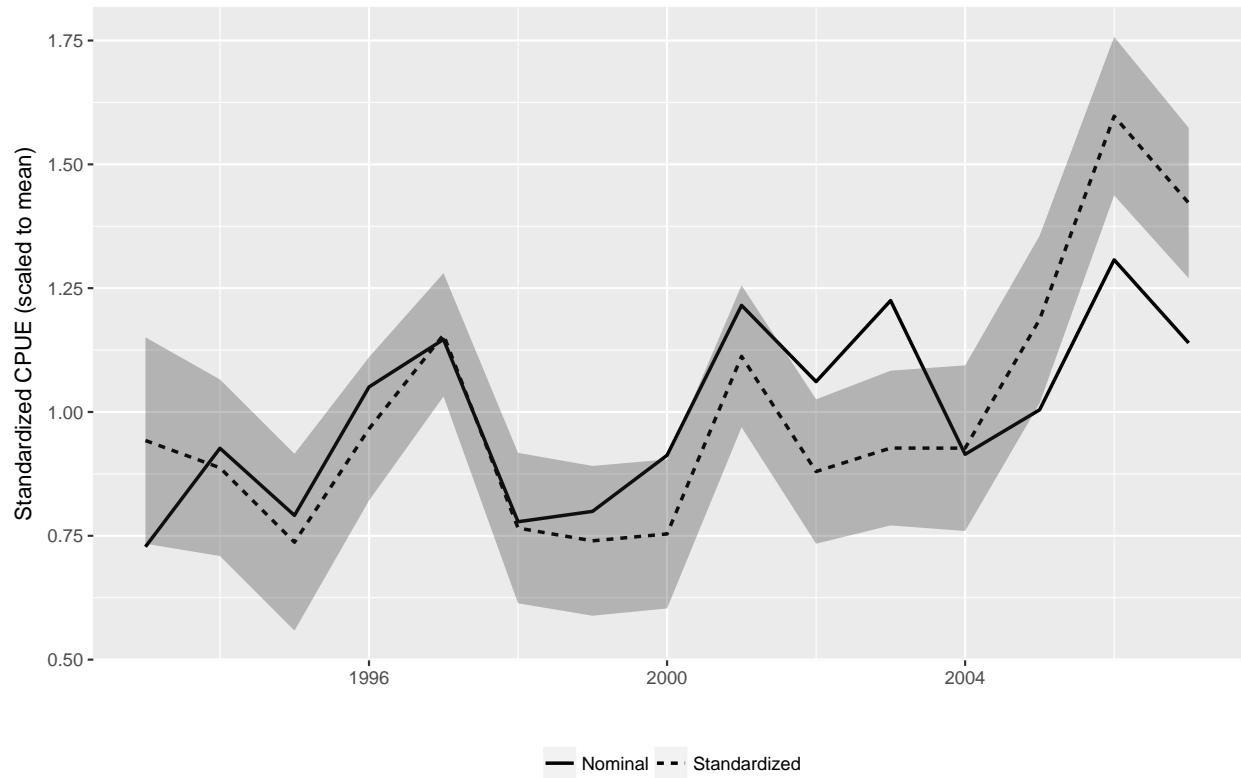
Figure 16: Standardized (dashed) with 95% confidence interval (shaded) and nominal index (solid) blueline tilefish catch rate from handline logbooks.

Appendix

Regional and South Atlantic indices
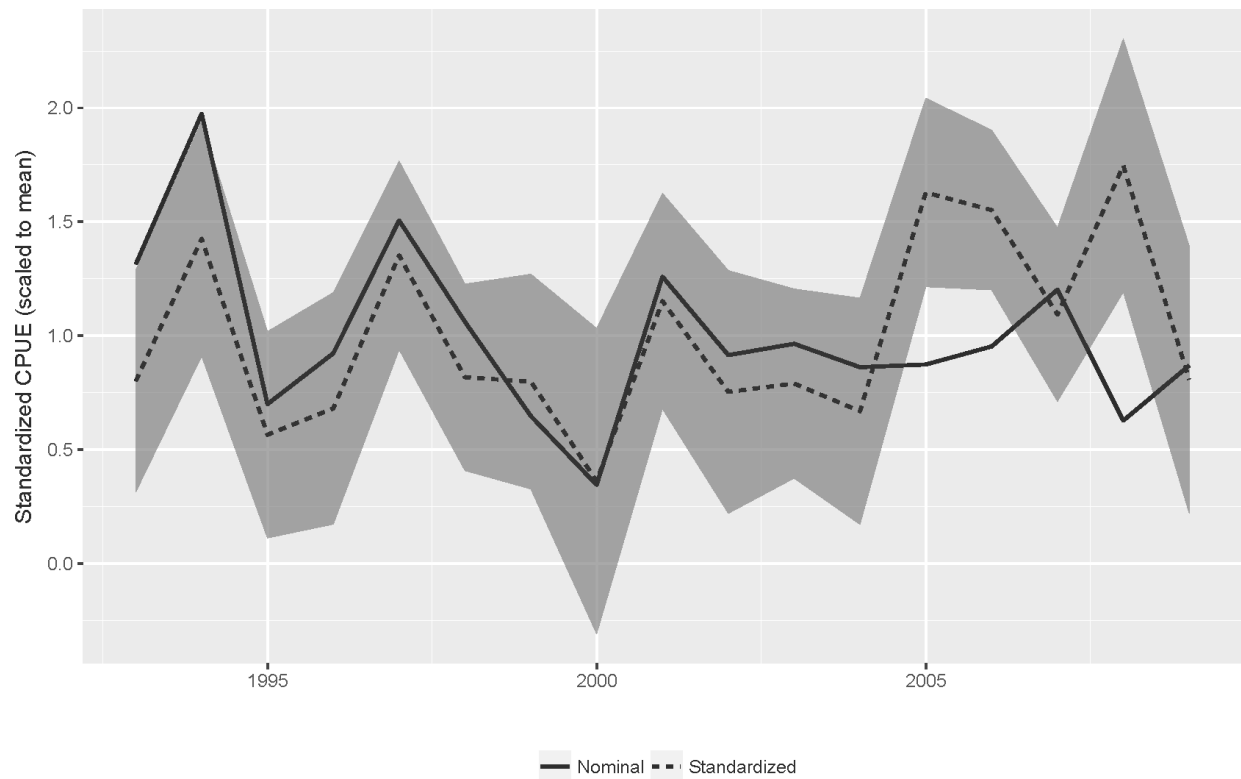
Figure 17: Standardized index of abundance from commercial handline (Eastern Gulf of Mexico) for blueline tilefish.

Figure 18: Standardized index of abundance from commercial handline (North Carolina to Cape Canaveral) for blueline tilefish.
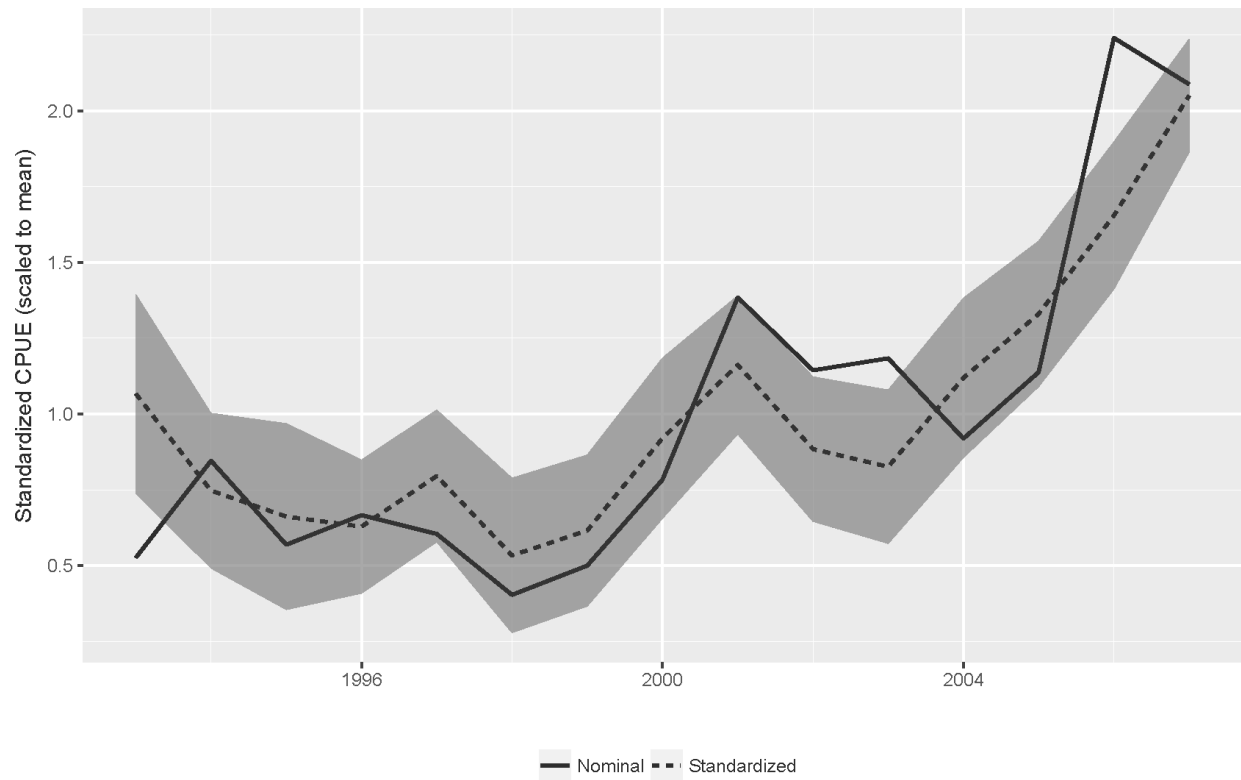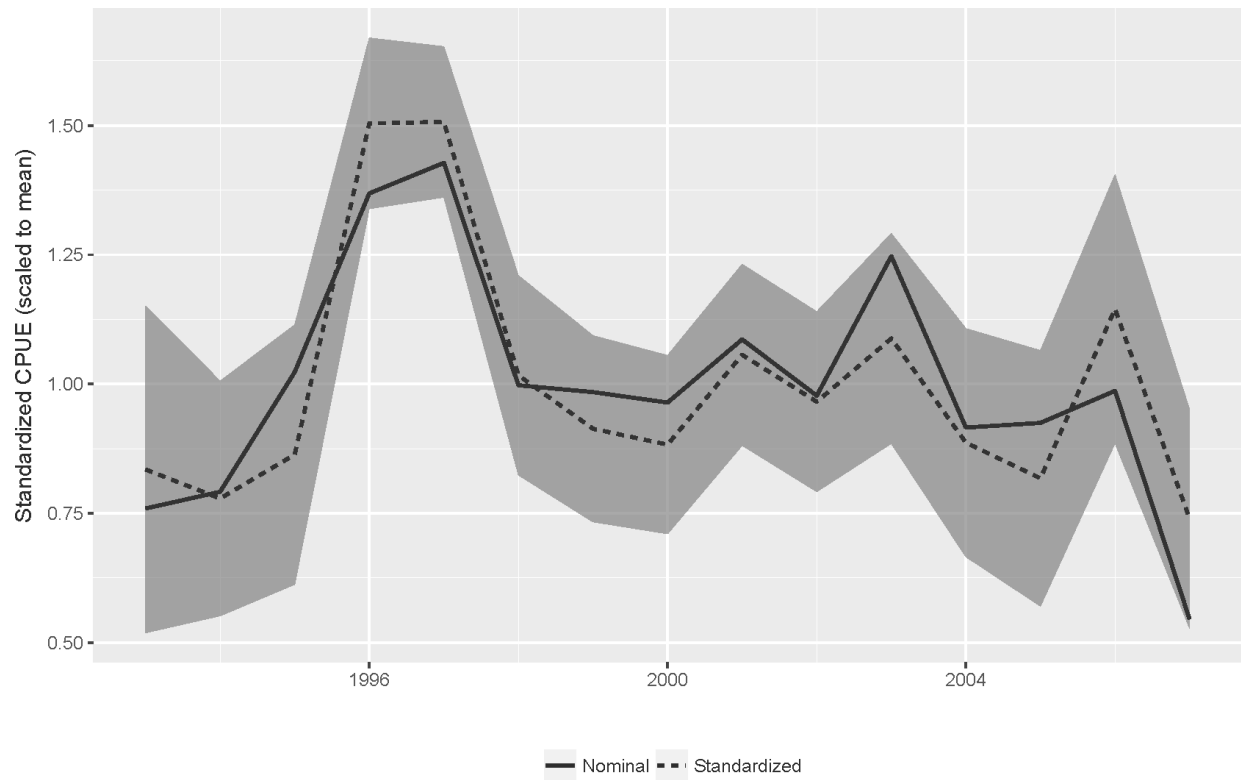
Figure 19: Standardized index of abundance from commercial handline (Cape Canaveral to Floriday Keys) for blueline tilefish.
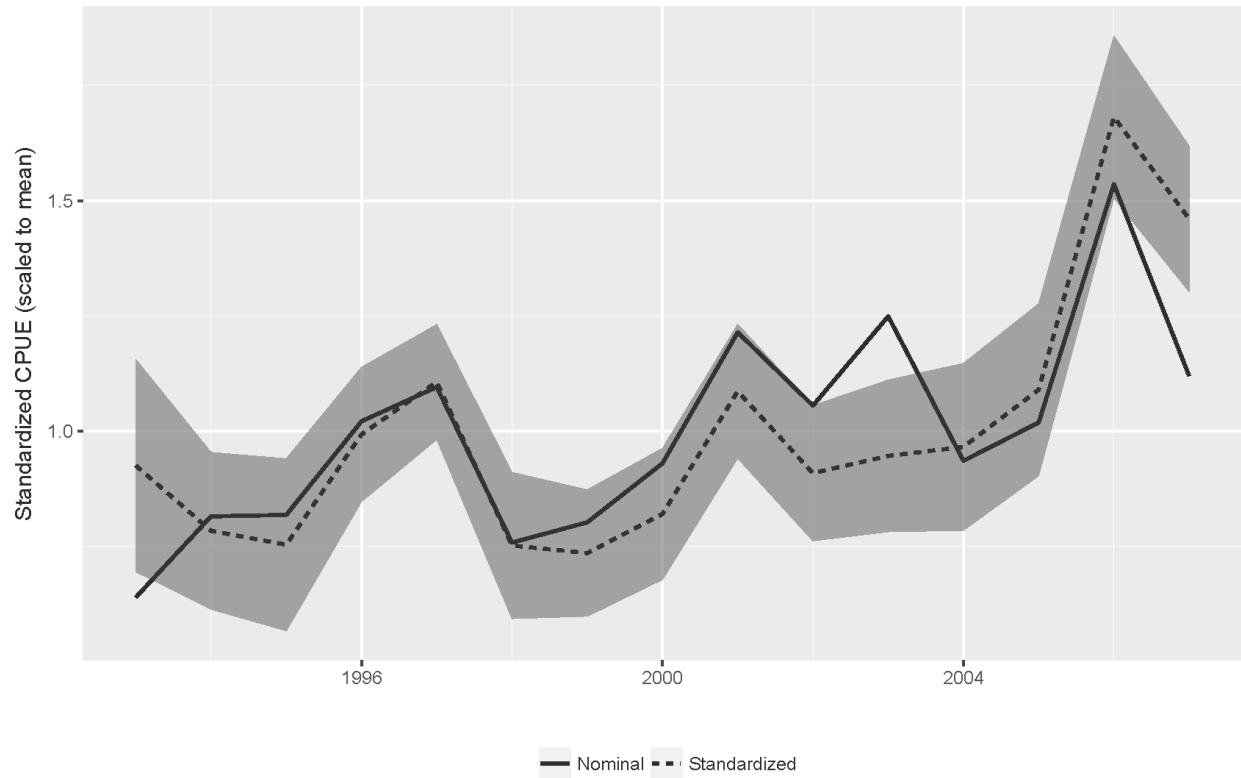
Figure 20: Standardized index of abundance from commercial handline (North Carolina to Florida Keys) to for blueline tilefish.

1. Additional data variables added to commercial logbook data following SEDAR 32

- Schedule: a unique trip identifier; each trip may have several records, one for each species landed
- Species: NMFS numeric species code
- Common and Scientific: species common and scientific names
- Gear: H=handline, E=electric reel (bandit gear), L=longline, GN=gill net, T=trap, TR=trolling, O=other, S=spear gun, P=powerheads/bangsticks
- Area: Gulf of Mexico shrimp grid or latitude/longitude where the animal was caught, see figure below
- Conversion: conversion factor for converting gutted weight to whole weight
- Depth: depth in feet where the majority of fish were caught, collection of these data began in 2005. Some fishers may have reported these data in fathoms. Reporting in fathoms may be more common for longline gear.
- Gutted: reported gutted weight in pounds
- Whole: reported whole weight in pounds
- Totlbs: calculated total (whole) weight landed (gutted x conversion + whole)
- Gear_type: additional description of gear fished
- Length: length of the longline in miles or length of GN in yards, GN length data were not collected prior to 1998
- Mesh1-Mesh4: mesh size on traps in inches or GN mesh in inches
- Set_time:

- Numgear: depends upon the gear used, for H, E, and TR=number lines fished L=number of sets T=number of traps used S, P=number of divers GN=number of nets fished O=does not apply
- Fished: hours fished, for L=total time fished for all sets combined. These data are problematic in the logbook data set due to changes in the way the data were reported. Often it is unclear whether hours fished was reported as hours per set or total hours fished. GN=total hours fished, all sets combined. T=total hours traps were in the water, e.g. 30 traps in the water for 4 hours equals 120 hours fished. H, E, TR, O=total hours fished. S, P=total hours divers were in the water.
- Effort: depends upon the gear used, for H, E, and TR=hooks per line, L= hooks per line,T=number of hauls (10 traps pulled 3 times each=30 hauls),S, P= does not apply,GN=depth of net in yards, these data were not collected prior to 1998,O=does not apply
- Vesid: vessel hull identification number
- Started: date fishing trip began
- Landed: date vessel returned to port
- Away: days at sea
- Crew: number of crew on the vessel
- State: NMFS state code, see list below
- County: NMFS county code, see list below
- Year, month, day: year, month, and day trip occurred = to date landed
- Gear1: equal to gear except gear1=H includes gears H and E; gear1=D includes gears S and P
- Region: defines region fished where sa=South Atlantic (south and east of FL Keys to 37N), gom=Gulf of Mexico (north and west of FL Keys to TX-Mexico border), and otr=other area (may include north of 37N, Mexico, or Caribbean)
- Gear flag: gear_flag=1 if effort information for that trip/gear combination are missing or are illogical (e.g., fishing more than 24 hours/day
- Report_days: number of days from the unload date to the date the logbook report was received at SEFSC
- Report_flag: report_flag=1 if report days > 45 or if report days<0
- Report1990_flag: report1990_flag=1 if report days>45 during 1990 - there were many reporting delays in 1990 and a unique flag was assigned to that year
- Region_flag: region flag=1 if more than one region was reported for a single trip
- Region123_flag: same as region flag, but for the years trips reported areas fished as Area1, Area2, Area3 (i.e., area fished was not associated with a particular species landed)
- Multgear_flag: multgear flag=1 if more than one gear was reported for a trip except when H and E

were both reported; those gears were not reported separately during the initial years of the logbook program
- Outlier_flag: outlier flag=1 if effort data (numgear, effort, away, crew) reported for gear fished during a trip were outside the 99.5 percentile of all gear-specific reports or were below the 0.5 percentile for length of longlines and gillnets or for longline effort (i.e., number of hooks per set) Avail_gag - avail_red_snapper: indicates the amount of IFQ (pounds) for each species available to the vessel and the end of each trip prior to unloading the catch
- Pounds: the total pounds landed of all species on the trip (used to estimate targeting)
- Pct_shk: the percent of sharks, by weight, of the total landings from the trip
- Pct_otr: the percent of other species (not shark, deep water grouper and tilefish, or shallow water grouper and snapper), by weight, of the total landings from the trip
- Pct_dw: the percent of deep water grouper and tilefish, by weight, of the total landings from the trip
- Pct_sw: the percent of shallow water grouper and snapper, by weight, of the total landings from the trip
- Target: the presumptive target species group for the trip; a species group was presumed to be targeted if landings of that species group accounted for 66.6% or more of the total landings from the trip; if no species group accounted for 66.6% or more, the target was labeled as mixed

Results of lognormal glm to determine factors.

```
## 
## Call:  glm(formula = log(cpue) ~ year + target + zone, family = gaussian(link = "identity"),
##     data = dat.pos)
## 
## Coefficients:
## (Intercept)      year1994      year1995      year1996      year1997
##    -1.16658       0.02134       0.05492       0.02380       0.17593
##     year1998      year1999      year2000      year2001      year2002
##    -0.12349      -0.06025      -0.06513       0.25704       0.06639
##     year2003      year2004      year2005      year2006      year2007
##     0.03192       0.05803       0.23973       0.39483       0.41565
##     targetsw       zonegom       zonesfl
##    -1.84528      -0.38317       0.47000
## 
## Degrees of Freedom: 6643 Total (i.e. Null);  6626 Residual
## Null Deviance:        20800
## Residual Deviance: 13330      AIC: 23520
```

Results of gamma glm to determine factors.

```
## 
## Call:  glm(formula = cpue ~ year + target + zone, family = Gamma(link = "log"),
##     data = dat.pos)
## 
## Coefficients:
## (Intercept)      year1994      year1995      year1996      year1997
##    -0.34315       0.10620       0.06338      -0.08313       0.14873
##     year1998      year1999      year2000      year2001      year2002
##    -0.09635      -0.12086      -0.15639       0.25130       0.04179
##     year2003      year2004      year2005      year2006      year2007
##     0.08439       0.04881       0.24247       0.27702       0.25521
##     targetsw       zonegom       zonesfl
##    -1.54415      -0.26300       0.27630
```

```
##
## Degrees of Freedom: 6643 Total (i.e. Null);  6626 Residual
## Null Deviance:        14910
## Residual Deviance: 11230      AIC: 1805
```

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 + target + zone"
##
## $positive.formula
## [1] "Formula for gaussian GLM: log(cpue) ~ year + target + zone"
##
## $deltaGLM.index
##          index jackknife
## 1993 0.1509857       NA
## 1994 0.1421719       NA
## 1995 0.1180804       NA
## 1996 0.1547456       NA
## 1997 0.1851850       NA
## 1998 0.1226615       NA
## 1999 0.1185323       NA
## 2000 0.1207623       NA
## 2001 0.1782205       NA
## 2002 0.1409530       NA
## 2003 0.1485418       NA
## 2004 0.1485066       NA
## 2005 0.1901595       NA
## 2006 0.2559200       NA
## 2007 0.2278165       NA
##
## $pos.effects
## $pos.effects[[1]]
##        dw        sw
## 0.9683678 0.1529836
##
## $pos.effects[[2]]
##       atl       gom       sfl
## 0.3739156 0.2548967 0.5982603
##
##
## $bin.effects
## $bin.effects[[1]]
##        dw        sw
## 0.5573432 0.2714126
##
## $bin.effects[[2]]
##       atl       gom       sfl
## 0.4950857 0.3856022 0.3429608
##
##
```

```
## $data.filter
## [1] "Data filter threshold set at 2 positive observations."
##
## $levels.deleted.by.filter
## $levels.deleted.by.filter$year
## [1] NA
##
## $levels.deleted.by.filter$target
## [1] NA
##
## $levels.deleted.by.filter$zone
## [1] NA
##
##
## $aic
##                        [,1]
## AIC.binomial   21607.264205
## AIC.lognormal   1187.008073
## sigma.mle          1.416611
```

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 + target + zone"
##
## $positive.formula
## [1] "Formula for Gamma GLM: cpue ~ year + target + zone"
##
## $deltaGLM.index
##          index jackknife
## 1993 0.3292965        NA
## 1994 0.3661905        NA
## 1995 0.3508429        NA
## 1996 0.3030276        NA
## 1997 0.3821029        NA
## 1998 0.2990493        NA
## 1999 0.2918081        NA
## 2000 0.2816223        NA
## 2001 0.4233766        NA
## 2002 0.3433503        NA
## 2003 0.3582904        NA
## 2004 0.3457687        NA
## 2005 0.4196548        NA
## 2006 0.4344050        NA
## 2007 0.4250340        NA
##
## $pos.effects
## $pos.effects[[1]]
##        dw        sw
## 0.7649966 0.1633215
##
```

```
## $pos.effects[[2]]
##       atl       gom       sfl
## 0.3519057 0.2705239 0.4638980
##
##
## $bin.effects
## $bin.effects[[1]]
## dw sw
##  1  1
##
## $bin.effects[[2]]
## atl gom sfl
##   1   1   1
##
##
## $data.filter
## [1] "Data filter threshold set at 2 positive observations."
##
## $levels.deleted.by.filter
## $levels.deleted.by.filter$year
## [1] NA
##
## $levels.deleted.by.filter$target
## [1] NA
##
## $levels.deleted.by.filter$zone
## [1] NA
##
##
## $aic
##                      [,1]
## AIC.binomial   36.0000000
## AIC.gamma    1652.9227387
## shape.mle       0.7129632
```

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.