Standardized catch rates of greater amberjack (*Seriola dumerili*) in the southeast U.S. from headboat at-sea observer data

Sustainable Fisheries Branch – NMFS Beaufort Lab
Contact: Eric Fitzpatrick

SEDAR59-WP06

Submitted: 6 December 2018

Please cite this document as:

SFB-NMFS. 2018. Standardized catch rates of greater amberjack (*Seriola dumerili*) in the southeast U.S. from headboat at-sea observer data. SEDAR59-WP06. SEDAR, North Charleston, SC. 13 pp.

Standardized catch rates of greater amberjack (Seriola dumerili) in the southeast U.S. from headboat at-sea observer data

Sustainable Fisheries Branch[*]

November 2018

This document describes the the development of the SEDAR 59 headboat at-sea index for greater amberjack.

Abstract

Standardized catch rates were generated from the Southeast headboat at-sea-observer program for 2005-2017. The analysis included areas from central North Carolina through south Florida. The index is meant to describe population trends of fish in the size/age range of fish landed and discarded by headboat vessels.

Trip-level information included state, county, Florida region, year, month, day, dock to dock hours (total trip hours), the number of hours fished (to the nearest half hour), the total number of anglers on the boat, the number of anglers observed on a trip, the number of greater amberjack discarded, minimum depth of the fishing trip, and maximum depth of the fishing trip. Depth information was not collected for South Carolina, North Carolina, and Georgia; therefore, it was not used in this analysis. SEDAR 59 terms of reference request that these data be evaluated and considered for inclusion in the standard assessment of greater amberjack. Data for this index were provided with the intention to explore greater amberjack discards as a possible index of juvenile abundance while avoiding harvested fish since these data would overlap from the headboat logbook index. Very few trips resulting in low proportion positives (4%) provide little data to pursue a standarized index (Figures 1 and 2). Due to these low sample sizes harvest and discards were combined in order to obtain enough sample sizes to pursue an index.

Response and explanatory variables

CPUE - Catch per unit effort (CPUE) is defined as units of fish/ angler interviewed*hours fished and was calculated as the number greater amberjack harvested and discarded divided by the number of anglers interviewed multiplied by hours fished.

YEAR - A summary of the total number of trips with greater amberjack effort per year is provided in (Figure 2.

AREA -Area was defined as North Carolina, South Carolina and Georgia (NCGA) and Florida (FL).

SEASON - The seasons were defined as winter (January, February, March), spring (April, May, June), summer (July, August, September) and fall (October, November, December) were considered and removed in the standarization process.

PARTY - Two categories (quantiles) for the number of anglers on the vessel were considered and removed in the standardization process.

HRSF- Two categories (quantiles) for the number of hours fished were considered and removed in the standardization process.

---

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

Standardization

CPUE was modeled using the delta-GLM approach (Lo, Jacobson, and Squire 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 red snapper on any given trip. Initially, all explanatory variables were included in the model as main eiiects, 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 removed all explanatory variables exept year and area. Diagnostics, based on standardized (quantile) residuals, suggested reasonable fits of the Bernoulli submodel considering the low sample sizes and low proportion positives (Figure 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 year and area as 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 delta-GLM. Diagnostics suggested reasonable fits of the lognormal submodel considering the low sample sizes and infrequency of greater amberjack (Figures 4 and 5).

Results

The standardized index was similar to the nominal index as well as the nominal discard index with the exception of a few years associated with peaks in the catch rate (Figure 6). The increase in the error since 2010 may be due to the lower proportion of positive catches. Interestingly, the index was also similar to the headboat logbook index for the overlapping years (correlation value= 0.60).

Recommendation

Due to limited samples and the overlapping data from the headboat logbook index, this headboat at-sea observer index is not recommended for use.

Table 1: Nominal and standardized CPUE for greater amberjack 2005-2017 with CVs for stardardized index of abundance.

Year	N	Nominal.CPUE	Relative.nominal	Standardized.CPUE	CV
2005	321	0.00	0.48	0.58	0.20
2006	289	0.00	1.06	0.75	0.31
2007	302	0.00	0.62	0.64	0.25
2008	239	0.00	0.65	0.66	0.26
2009	233	0.00	0.55	0.54	0.34
2010	250	0.00	0.33	0.37	0.35
2011	230	0.00	0.44	0.63	0.34
2012	261	0.00	0.54	0.78	0.26
2013	252	0.01	1.56	1.85	0.38
2014	255	0.01	1.56	1.35	0.31
2015	222	0.01	2.26	2.01	0.30
2016	271	0.01	1.41	1.73	0.24
2017	274	0.01	1.54	1.11	0.37

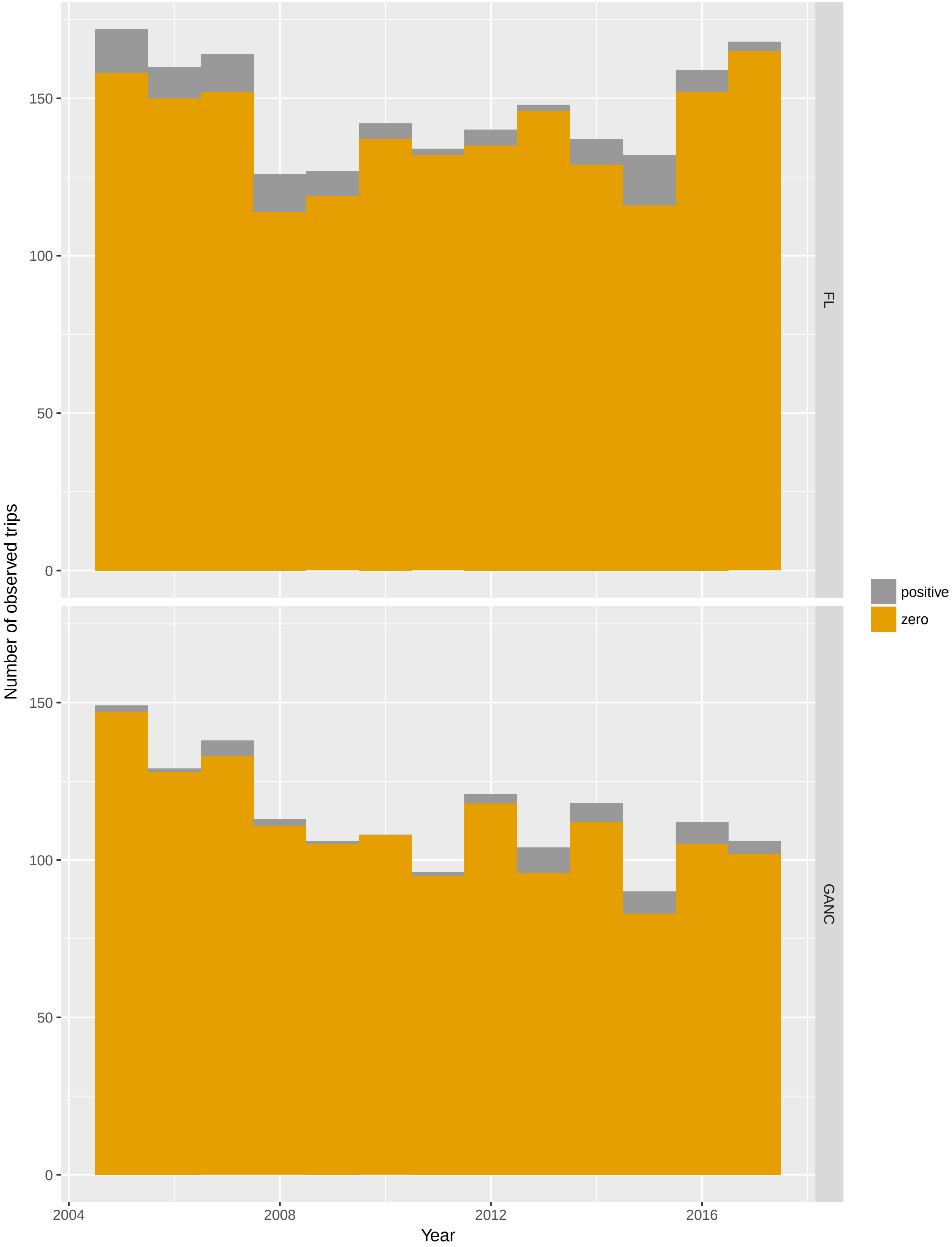


Figure 1: Number of observered headboat trips by region.

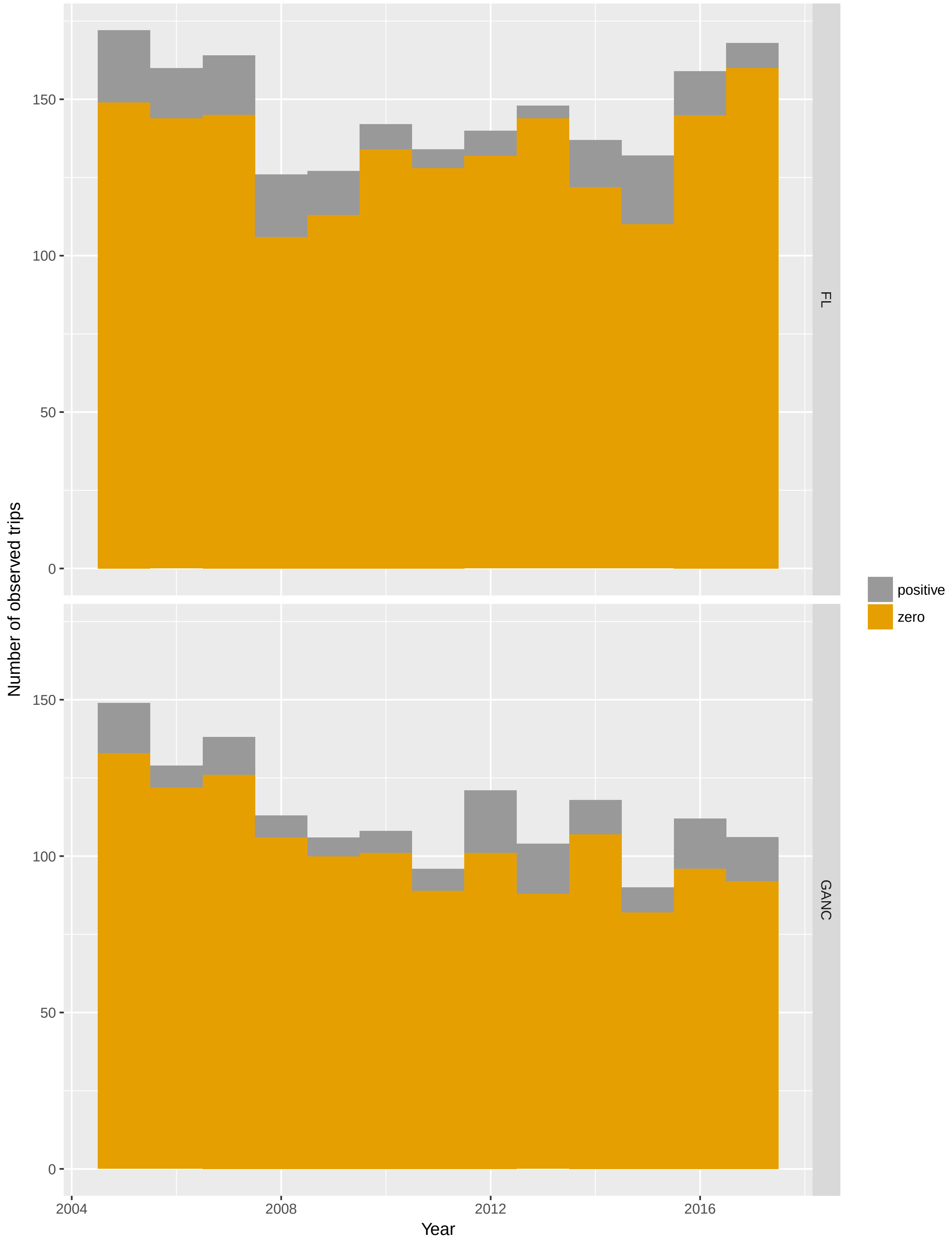


Figure 2: Number of observered headboat trips by region.

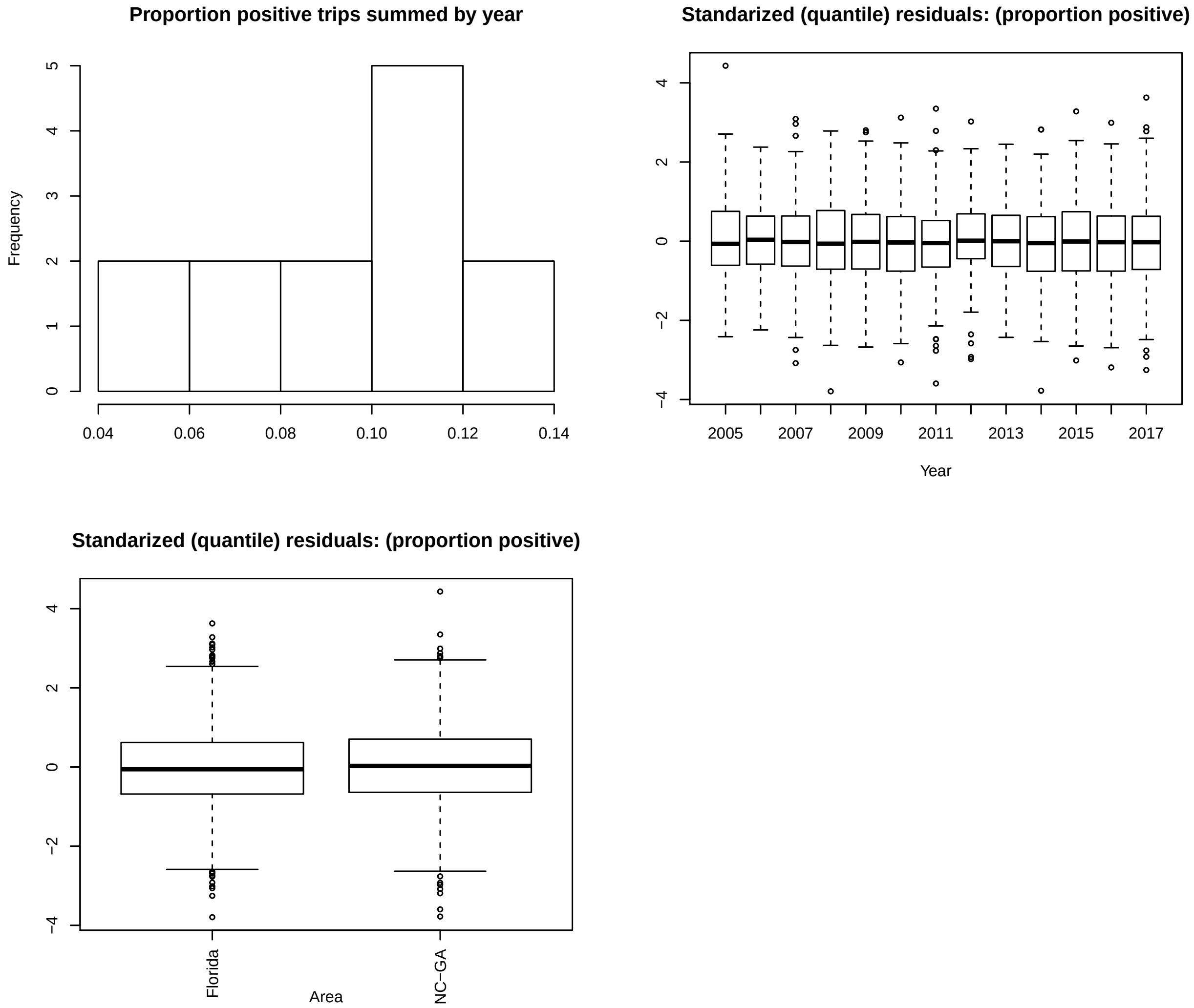
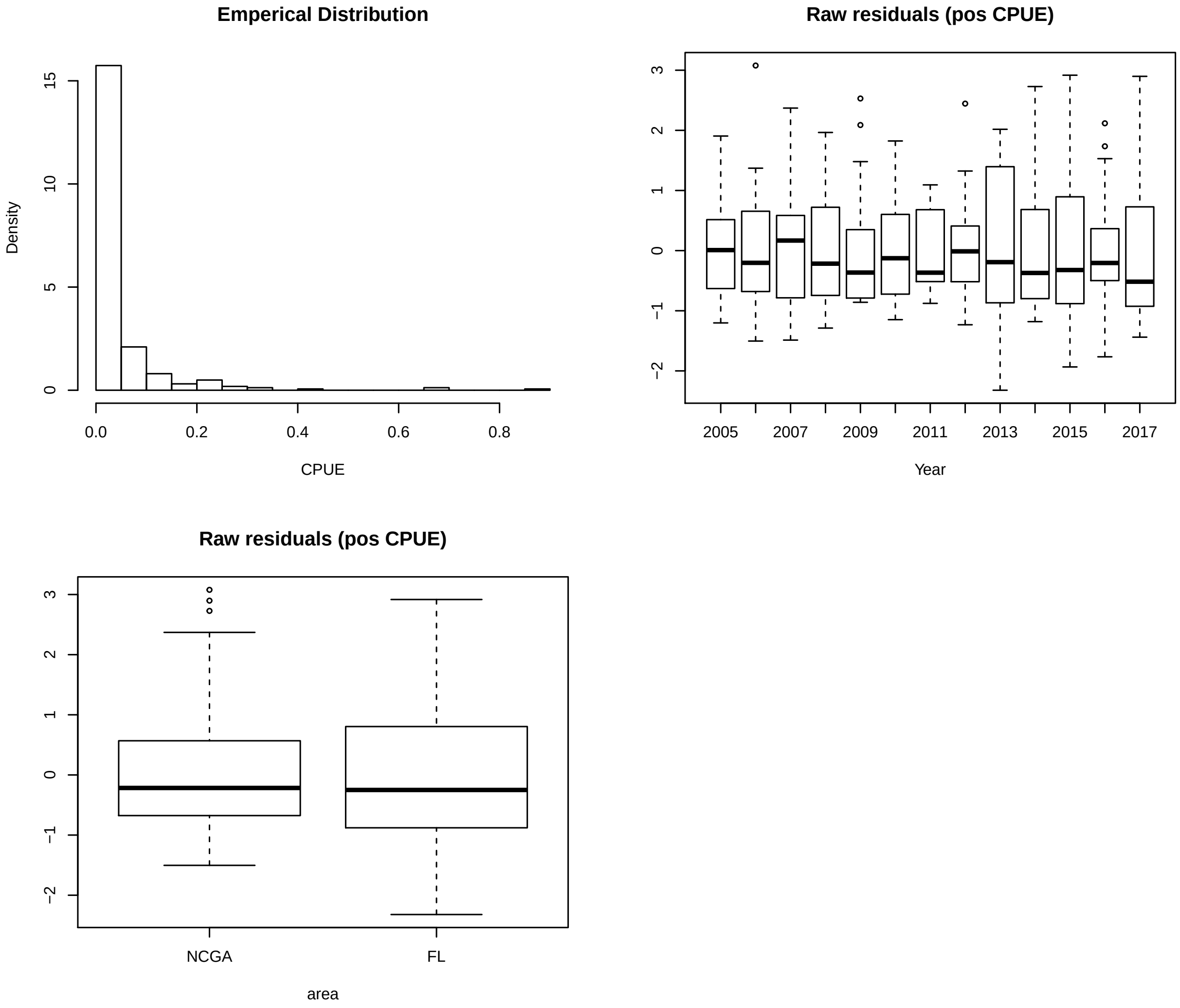


Figure 3: Diagnostics of Bernoulli submodel fits to positive versus zero CPUE data. Box and whisker plots give first, second (median) and third quartiles, as well as limbs that extend to approximately one interquartile range beyond the nearest quartile, and outliers (circles) beyond the limbs. Residuals are standardized (quantile) residuals.



Emperical Distribution

Raw residuals (pos CPUE)

Raw residuals (pos CPUE)

Figure 4: 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.

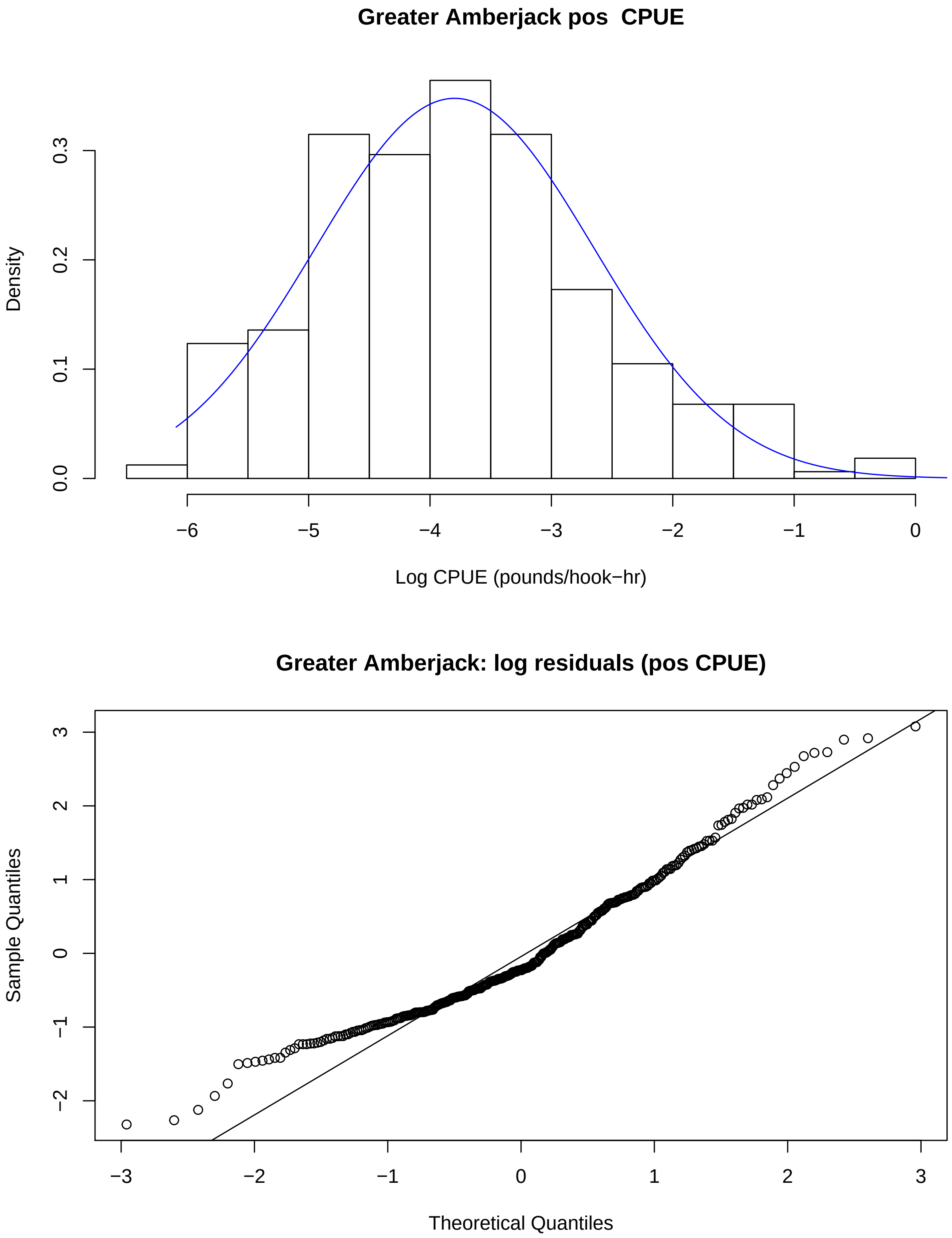


Figure 5: Histogram of empirical log CPUE, with the normal distribution (empirical mean and variance) overlaid. Quantile-quantile plot of residuals from the fitted lognormal submodel to the positive cpue data.

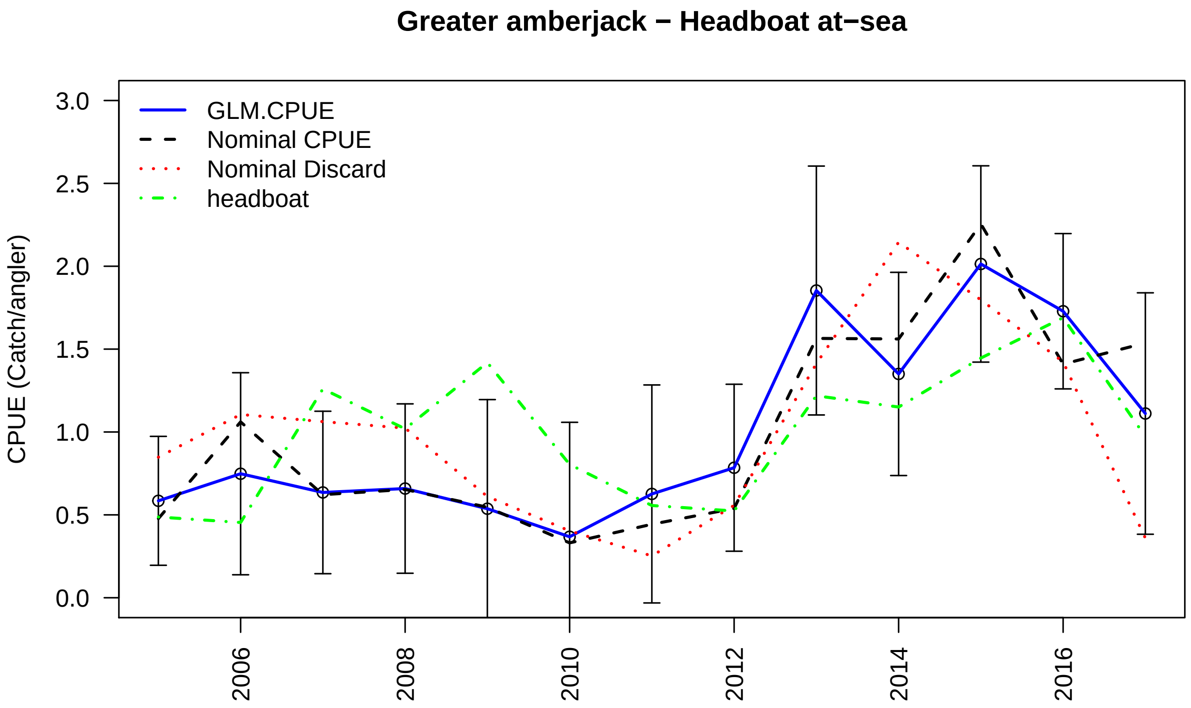


Figure 6: Standardized headboat at-sea observer greater amberjack catch rate (solid blue) with 95% confidence intervals and nominal catch rate (black dashed), nominal discard (red dashed) and standardized headboat logbook index (green dashed).

Appendix

Results of lognormal glm to determine factors.

```
##
## Call:  glm(formula = cpue ~ yr + area, family = gaussian(link = "identity"),
##     data = pos.dat)
##
## Coefficients:
## (Intercept)       yr2006       yr2007       yr2008       yr2009
##    0.025801     0.041651     0.009451     0.005803     0.009250
##      yr2010       yr2011       yr2012       yr2013       yr2014
##    0.008405     0.020513     0.010683     0.080421     0.053038
##      yr2015       yr2016       yr2017     areaGANC
##    0.056542     0.043111     0.074744    -0.018462
##
## Degrees of Freedom: 323 Total (i.e. Null);  310 Residual
## Null Deviance:       2.669
## Residual Deviance: 2.43  AIC: -635.8
```

Results of gamma glm to determine factors.

```
##
## Call:  glm(formula = cpue ~ yr + area, family = Gamma(link = "log"),
##     data = pos.dat)
##
## Coefficients:
## (Intercept)       yr2006       yr2007       yr2008       yr2009
##     -3.8430       1.1190       0.3983       0.3112       0.5902
##      yr2010       yr2011       yr2012       yr2013       yr2014
##      0.3236       0.7428       0.4694       2.0054       1.4556
##      yr2015       yr2016       yr2017     areaGANC
##      1.5548       1.3032       1.7006      -0.5930
##
## Degrees of Freedom: 323 Total (i.e. Null);  310 Residual
## Null Deviance:       492.7
## Residual Deviance: 370   AIC: -1408
```

Results of binomial glm to determine factors.

```
##
## Call:  glm(formula = cpue ~ 1, family = "binomial", data = bin.dat)
##
## Coefficients:
## (Intercept)
##       -2.25
##
## Degrees of Freedom: 3398 Total (i.e. Null);  3398 Residual
## Null Deviance:       2139
## Residual Deviance: 2139  AIC: 2141
```

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 ~ yr + area"
##
## $positive.formula
## [1] "Formula for gaussian GLM: log(cpue) ~ yr + area"
##
## $deltaGLM.index
##            index jackknife
## 2005 0.002317376        NA
## 2006 0.002964659        NA
## 2007 0.002517156        NA
## 2008 0.002610596        NA
## 2009 0.002128215        NA
## 2010 0.001457966        NA
## 2011 0.002481468        NA
## 2012 0.003108217        NA
## 2013 0.007345456        NA
## 2014 0.005350794        NA
## 2015 0.007980089        NA
## 2016 0.006850144        NA
## 2017 0.004404355        NA
##
## $pos.effects
## $pos.effects[[1]]
##         FL       GANC
## 0.05460356 0.02573361
##
##
## $bin.effects
## $bin.effects[[1]]
##         FL       GANC
## 0.09010239 0.09520251
##
##
## $data.filter
## [1] "Data filter threshold set at 2 positive observations."
##
## $levels.deleted.by.filter
## $levels.deleted.by.filter$yr
## [1] NA
##
## $levels.deleted.by.filter$area
## [1] NA
##
##
## $aic
##                         [,1]
## AIC.binomial   2147.2404694
## AIC.lognormal -1513.8198838
## sigma.mle         0.9956298
```

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 ~ yr + area"
##
## $positive.formula
## [1] "Formula for Gamma GLM: cpue ~ yr + area"
##
## $deltaGLM.index
##            index jackknife
## 2005 0.001938685        NA
## 2006 0.003892375        NA
## 2007 0.002440468        NA
## 2008 0.002459521        NA
## 2009 0.002472829        NA
## 2010 0.001325710        NA
## 2011 0.001900928        NA
## 2012 0.002737343        NA
## 2013 0.009435532        NA
## 2014 0.006975836        NA
## 2015 0.010240150        NA
## 2016 0.006520680        NA
## 2017 0.007048168        NA
##
## $pos.effects
## $pos.effects[[1]]
##         FL       GANC
## 0.05383247 0.02975135
##
##
## $bin.effects
## $bin.effects[[1]]
##         FL       GANC
## 0.09010239 0.09520251
##
##
## $data.filter
## [1] "Data filter threshold set at 2 positive observations."
##
## $levels.deleted.by.filter
## $levels.deleted.by.filter$yr
## [1] NA
##
## $levels.deleted.by.filter$area
## [1] NA
##
##
## $aic
##                      [,1]
## AIC.binomial  2147.240469
## AIC.gamma    -1412.499817
## shape.mle        1.009823
```

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., L. Jacobson, and J. Squire. 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.