# Standardized Catch Rate Indices for Gulf of Mexico Gray Triggerfish (*Balistes capriscus*) Landed During 1993-2013 by the Commercial Handline Fishery

Matthew W. Smith, Daniel Goethel, Adyan Rios, and Jeff Isley

## SEDAR43-WP-05

20 March 2015

Please cite this document as:

Smith, M.W., D. Goethel, A. Rios, and J. Isley. 2015. Standardized Catch Rate Indices for Gulf of Mexico Gray Triggerfish (*Balistes capriscus*) Landed During 1993-2013 by the Commercial Handline Fishery. SEDAR43-WP-05. SEDAR, North Charleston, SC. 16 pp.

# Standardized Catch Rate Indices for Gulf of Mexico Gray Triggerfish (*Balistes capriscus*) Landed During 1993-2013 by the Commercial Handline Fishery

by

Mathew W. Smith
Daniel Goethel
Adyan Rios
Jeff Isley

20 March 2015

NOAA Fisheries
Southeast Fisheries Science Center
75 Virginia Beach Drive
Miami, FL 33149

## INTRODUCTION

The National Marine Fisheries Service (NMFS) Gulf of Mexico Reef Fish Logbook data set was updated through 2013 and the procedures outlined in SEDAR 9 were implemented to provide standardized abundance indices. Only data from commercial handlines (including electric reels) were included in the analysis, because this is the primary fishery targeting gray triggerfish in the Gulf of Mexico. Catch-per-unit effort (CPUE) is derived from the commercial logbook data using total pounds of fish caught on a given trip divided by the amount of line-hours spent fishing, where:

$$\ln(CPUE) = \ln\left[\frac{Pounds\ of\ Fish}{Number\ of\ Lines*Hours\ Fished}\right]. \tag{1}$$

For the commercial handline data set, effort was estimated in line-hours where the number of hours a line soaked was multiplied by the number of lines set per trip. Two indices (east and west Gulf of Mexico) were calculated based on geographic area (east or west of the Mississippi delta) to better represent the variance and abundance trends in each zone, because effort can vary significantly from year-to-year between the two areas. The timeseries extends from 1993 (when all federally permitted vessels were required to submit a logbook) until 2013 (the terminal year of the stock assessment).

Trips were eliminated if they had missing values for any of the key factors, were in anyway incomplete, appeared to be misreported (e.g., reported zero hours fished) or represented multiple entries for a single trip. Similarly, because effort is reported on a per-trip basis, and not by area fished within a trip, all trips that reported multiple areas were removed from the analysis.

## METHODS
### Species Associations

The Stephens and MacCall (2004) approach was utilized to subset the number of trips based on species that co-occur in the catch using a logistic regression to determine commonly associated species. An indirect method was necessary to infer targeting behavior of fishermen, because no direct information was available. This statistical approach eliminates an overabundance of zero catches by dropping trips where gray triggerfish are unlikely to be located. In addition, it is reproducible and avoids *ad hoc* trip selection. However, the method relies on the occurrence of other species in the catch and trips that catch only the target species cannot be incorporated if they do not catch an associated species. The threshold probability determines a percentage of 'false positives' that will be removed from the data set by minimizing the difference between predicted and observed trips that reported catching the target species. By using only trip records where gray triggerfish are likely to occur (based on species associates), a more consistent and reliable set of trip records are used to develop the standardized CPUE.

### Index Standardization

A two-step delta-lognormal general linearized model (GLM; Lo et al. 1992) was used to standardize for variability and non-randomness in CPUE data collection methods not caused by the year effect (i.e., to factor out year to year variations in CPUE not due to changes in abundance). The combined approach first modeled the frequency with which trips caught the species of interest (i.e., proportion positive) using a logit regression assuming a binomial distribution of the response variable. In the second step, the logarithm of CPUE on successful trips (those that caught the target species) was used as the response variable assuming a normal distribution and an identity link function. The two models were then combined to provide the final standardized index of abundance.

A forward stepwise regression approach was utilized within the GENMOD procedure of SAS 9.2 (SAS Institute, 2008). In this procedure, potential factors were added to the base model one at a time based on the percent reduction in deviance per degree of freedom. With each run of the model the factor that caused the highest reduction in deviance was added to the base model (assuming the factor was significant based on a Chi-Square test with probability $\leq 0.05$) until no factor reduced the percent deviance by the prespecified level (i.e., 1%). Because the goal of the standardization was to model time trends in abundance, it was necessary to force the year effect as a factor even if it was not deemed significant. Two-way interaction terms were then investigated among each of the significant factors using the same stepwise approach. All higher order interactions were ignored.

The final delta-lognormal model was fit using the factors deemed significant in the GENMOD procedure using the SAS macro GLIMMIX (SAS Institute, 2008). Factors were modeled as fixed effects except for interaction terms involving year, which were modeled as random effects. Results of the binomial (proportion positive) and lognormal (mean CPUE on successful trips) were then multiplied to attain a single index of abundance based on the year effect.

## RESULTS

### Species Associations

There were 158,537 trip records available in the commercial logbook data base from the Gulf of Mexico with 36,735 encountering gray triggerfish. The Stephens and MacCall (2004) approach identified a threshold probability frequency of 45% (Figure 1A) for the western Gulf of Mexico, which resulted in a subset of 13,362 trips for use in the index standardization. The proportion of positive trips before the regression was applied was 0.38, which increased to 0.66 after the subset was taken. Coefficients of the logistic regression indicated species association and demonstrated that vermilion snapper and lane snapper were identified as the species most closely associated with gray triggerfish, while yellowedge grouper and king mackerel were most strongly negatively correlated (Table 1A).

For the eastern Gulf of Mexico, a threshold probability frequency of 39% (Figure 1B) was identified, which resulted in a subset of 23,492 trips for use in the index standardization. The proportion of positive trips before the regression was applied was 0.19, which increased to 0.57 after the subset was taken. Coefficients of the logistic regression indicated species association and demonstrated that vermilion snapper and red porgy were identified as the species most closely associated with gray triggerfish, while mutton snapper and yellowtail snapper were most strongly negatively correlated (Table 1B).

## Abundance Indices

A number of factors were investigated that could potentially influence yearly variations in catch rates including: Year, Season, Area, Red Snapper Season (i.e., open or closed), Red Snapper Permit (whether or not a vessel was permitted in the red snapper fishery), Days Away from Port, Crew Size, and Hook Hours Fished (only for the binomial model as this factor is confounded with effort for the CPUE response variable in the lognormal model). The levels and potential values for the various factors are provided in Table 2.

## Western Gulf of Mexico

For the binomial component of the western Gulf of Mexico index, the significant factors were Year, Area, Crew Size, and Hook Hours Fished. In the lognormal model Year, Crew Size, Area, and Days Away from Port were found to be significant along with the Year*Days Away from Port, Year*Area, and Crew Size*Area interaction terms. The final models were:

$$\begin{gathered} Proportion\ Positive = Year + Area + Crew + Hook\ Hours\ Fished \\ \ln(CPUE) = Year + Crew\ Size + Area + Days\ Away\ from\ Port + \\ Year * Days\ Away\ from\ Port + Year * Area + Crew\ Size * Area\ . \end{gathered} \qquad \textbf{(2)}$$

The final nominal and standardized CPUE (both provided as relative indices where each value is divided by the timeseries mean) along with confidence intervals and coefficients of variation ($CV$s) are given in Table 3.

Observed and predicted trends in the proportion of positive trips tend to overlap with values fluctuating between 50-70% for much of the timeseries (Figure 2A). In the last three years, the occurrence of gray triggerfish has tailed off to around 40%. Results from the lognormal model indicate a relatively strong fit with no obvious patterning in the residuals (Figure 3A), and adherence to distributional assumptions (Figure 3B).

Nominal and standardized CPUE (scaled to the timeseries mean) show similar trends for most of the timeseries with the nominal values within the 95% confidence intervals (Figure 4A). Both timeseries indicate a steady decline over time with the most recent values at timeseries lows. Standardized CPUE has been below the mean level since the early 2000s and is currently at around one quarter of the timeseries average.

## Continuity

Trends in the commercial west indices were similar between the original SEDAR 9 index, the update index, and the current SEDAR 43 index (Figure 5A). Although attempts were made to maintain modeling approaches, some alterations have been required over the last decade since the SEDAR 9 benchmark. Differences in the final models were mainly due to data updates causing changes in the significance of GLM factors, but this is also a result of new factors being tested between SEDAR 9 and the 2011 update. According to the SEDAR 9 analysis (Nowlis, 2005), Crew Size, Days Away from Port, and Hook Hours Fished were not factors that were tested. Not surprisingly, the change in factors that were tested has also led to changes in the final models chosen and resultant standardized CPUE values. For the binomial, the 2011 update and the current analysis indicated that Crew Size and Hook Hours Fished were significant, but that Red Snapper Permit was no longer significant. Similarly, in the lognormal model Red Snapper Permit was no longer found to be significant in the 2011 update and the current analysis, but Area, Crew Size, and Days Away from Port (along with a handful of interaction terms) were found to be significant. Given that each model shows the same general declining trend over the timeseries, the change in values is unlikely to be of importance since the relative changes through time are generally equivalent.

SEDAR 9 Model (Nowlis, 2005):

$$Proportion\ Positive = Year + Area + Red\ Snapper\ Permit$$
$$\ln(CPUE) = Year + Red\ Snapper\ Permit + Year * Red\ Snapper\ Permit\ . \tag{3}$$

2011 Update Model:

$$Proportion\ Positive = Year + Area + Crew\ Size + Hook\ Hours\ Fished \tag{4}$$
$$\ln(CPUE) = Year + Area + Crew\ Size + Days\ Away\ from\ Port + Area * Crew\ Size\ .$$

## Eastern Gulf of Mexico

For the binomial component of the eastern Gulf of Mexico index the significant factors were Year and Area along with the Year*Area interaction term. In the lognormal model Year, Area, Crew Size, and Days Away from Port were found to be significant along with the Year*Area, Year*Days Away from Port, Days Away from Port*Area, and Crew Size*Area interaction terms. The final models were:

$$Proportion\ Positive = Year + Area + Year * Area$$
$$\ln(CPUE) = Year + Area + Crew\ Size + Days\ Away\ from\ Port + Year * Area +$$
$$Year * Days\ Away\ from\ Port + Days\ Away\ from\ Port * Area + Crew\ Size * Area\ . \tag{5}$$

The final nominal and standardized CPUE (both provided as relative indices where each value is divided by the timeseries mean) along with confidence intervals and coefficients of variation ($CV$s) are given in Table 4.

Observed trends in the proportion of positive trips demonstrates a fairly consistent variation between 50-70% with a sharp decline in 2012, but was followed by a strong increase in 2013 (Figure 2B). Results from the lognormal model indicate a relatively strong fit with no obvious patterning in the residuals (Figure 3C), and little deviation from the expected lognormal error distribution (Figure 3D).

Nominal and standardized CPUE (scaled to the timeseries mean) show similar trends with the nominal values within the 95% confidence intervals for most of the timeseries (Figure 4B). The first five years are characterized by a sharp decline in both nominal and standardized CPUE followed by an increase until around 2003. CPUE has been generally declining since and the most recent estimates are near timeseries lows. Standardized CPUE has been below the mean level since 2006 and is currently at around one half of the timeseries average.

### Continuity

Generally the original SEDAR 9 indices, the 2011 updates, and the current estimates demonstrate consistent trends and similar values (Figure 5B). However, the SEDAR 9 values tend to be slightly lower than the other models, especially from 1996 onwards. Although attempts have been made to maintain modeling approaches, some alterations have been required over the last decade since the SEDAR 9 benchmark. Differences in the final models were mainly due to data updates causing changes in the significance of GLM factors, but this is also a result of new factors being tested between SEDAR 9 and the 2011 update. According to the SEDAR 9 analysis (Nowlis, 2005), Crew Size, Days Away from Port, and Hook Hours Fished were not factors that were tested. Not surprisingly, the change in factors that were tested has also led to changes in the final models chosen and resultant standardized CPUE values. For the binomial, the 2011 update and the current analysis indicated that Area was significant in addition to Year. In the lognormal model Red Snapper Season was no longer found to be significant in the 2011 update and the current analysis, but Area, Crew Size, and Days Away from Port (along with a handful of interaction terms) were found to be significant.

SEDAR 9 Model (Nowlis, 2005):

$$\begin{gathered} Proportion\ Positive = Year \\ \ln(CPUE) = Year + Red\ Snapper\ Season\ . \end{gathered} \tag{6}$$

2011 Update Model

$$\begin{gathered} Proportion\ Positive = Year + Area \\ \ln(CPUE) = Year + Area + Crew\ Size + Days\ Away\ from\ Port + Year * Area + \\ Year * Days\ Away\ from\ Port + Days\ Away\ from\ Port * Area + Crew\ Size * Area\ . \end{gathered} \tag{7}$$

The different factors tested resulted in different final models and are the reason for the difference in values between the SEDAR 9 results and those from the 2011 update and current analysis. However, it is difficult to determine what factor resulted in the biggest changes among models, but, given that the trends are very similar, the differences in standardized values are unlikely to be of concern.

## DISCUSSION

The results and model diagnostics indicate that the commercial handline data can be used to develop reliable indices of CPUE that can be used in the SEDAR 43 assessment of gray triggerfish. There are a few areas of concern that warrant future research, the most important of which is the impact of the 2008 regulatory changes (e.g., change in minimum size and forced implementation of circle hooks). Because management period would likely be confounded with the Year effect, it would be difficult to use a regulatory variable as an explanatory factor in the standardization. However, future investigations should consider splitting the series in 2008 as more years of data become available or dealing with these changes directly in the stock assessment model by allowing time-varying catchability or fishery selectivity. In addition, the impact of GLM modeling changes between SEDAR 9 and the 2011 update may need further investigation to determine which specific factors caused the change in standardized CPUE values.

## LITERATURE CITED


Lo, N.C., Jacobson, L.D., and 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.

Nowlis, J.S. 2005. Updated fishery-dependent indices of abundance for the Gulf of Mexico gray triggerfish (*Balistes capriscus*). SEDAR 9-AW07. SFD-2005-031.

SAS Institute Inc. 2008. *SAS/STAT 9.2 User's Guide*. Cary, NC: SAS Institute Inc.

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

# TABLES

**Table 1:** Coefficients of correlation used by the Stephens and MacCall (2004) species associates subsetting routine. Species with positive correlation coefficients tended to associate with gray triggerfish, while those with negative correlation coefficients did not. Panel A displays the results for the western Gulf of Mexico and Panel B displays the results from the eastern Gulf of Mexico

**A:**

| Species | Correlation Coefficient |
|---|---|
| Snapper, Vermilion | 1.286 |
| Snapper, Lane | 1.202 |
| Grouper, Black | 0.717 |
| Cobia | 0.609 |
| Jack, Almaco | 0.564 |
| Scamp | 0.562 |
| Grouper, Gag | 0.537 |
| Porgy, Red | 0.475 |
| Grouper, Warsaw | 0.467 |
| Snapper, Red | 0.465 |
| Blue Runner | 0.295 |
| Snapper, Mangrove | 0.178 |
| Amberjack, Greater | 0.075 |
| Sea Trout, White | -0.12 |
| Grouper, Yellowedge | -0.483 |
| King Mackerel | -0.643 |

**B:**

| Species | Correlation Coefficient |
|---|---|
| Snapper, Vermilion | 1.661 |
| Porgy, Red | 1.009 |
| Grunt, White | 0.907 |
| Snapper, Red | 0.567 |
| Snapper, Mangrove | 0.449 |
| Porgy, Whitebone | 0.351 |
| Scamp | 0.335 |
| Jack, Almaco | 0.288 |
| Grouper, Black | 0.278 |
| Grunt, Unclassified | 0.242 |
| Grouper, Gag | 0.174 |
| Snapper, Lane | 0.167 |
| Cobia | 0.109 |
| Amberjack, Greater | 0.103 |
| King Mackerel | 0.034 |
| Grouper, Red | -0.005 |
| Snapper, Mutton | -0.351 |
| Snapper, Yellowtail | -1.300 |

**Table 2:** Levels and values for the factors investigated for inclusion in the index standardization model.

Eastern Gulf of Mexico

| Factor | Levels | Values |
|---|---|---|
| Year | 21 | 1993 - 2013 |
| Season | 4 | Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec |
| Area | 6 | 6, 7, 8, 9, 10, 11 |
| Red Snapper Season | 2 | Closed, Open |
| Red Snapper Permit | 2 | Yes, No |
| Days away from port | 6 | 1, 2, 3, 4, 5, 6+ |
| Crew Size | 4 | 1, 2, 3, 4+ |
| Hook Hours | 5 | <100, 100-350, 350-1000, 1000-3000, 3000+ |

Western Gulf of Mexico

| Factor | Levels | Values |
|---|---|---|
| Year | 21 | 1993 - 2013 |
| Season | 4 | Jan-Mar, Apr-Jun, Jul-Sep, Oct-Dec |
| Area | 3 | 16, 17, 21 |
| Red Snapper Season | 2 | Closed, Open |
| Red Snapper Permit | 2 | Yes, No |
| Days away from port | 6 | 1, 2, 3, 4, 5, 6+ |
| Crew Size | 4 | 2, 3, 4, 5+ |
| Hook Hours | 5 | <500, 500-1000, 1000-2000, 2000-4000, 4000+ |

**Table 3:** Nominal CPUE and standardized index values (scaled to the timeseries mean) for the commercial handline fishery in the western Gulf of Mexico. Confidence intervals and CVs are for the standardized index.

| Western Gulf of Mexico | | | | | |
|---|---|---|---|---|---|
| | Index | | Confidence Limits (95%) | | |
| **Year** | **Standardized** | **Nominal** | **Lower** | **Upper** | **CV** |
| 1993 | 1.43 | 1.81 | 0.81 | 2.53 | 0.29 |
| 1994 | 2.30 | 2.43 | 1.35 | 3.94 | 0.27 |
| 1995 | 2.09 | 2.01 | 1.27 | 3.44 | 0.25 |
| 1996 | 1.54 | 1.70 | 0.93 | 2.55 | 0.26 |
| 1997 | 1.27 | 1.14 | 0.78 | 2.05 | 0.25 |
| 1998 | 1.49 | 1.37 | 0.92 | 2.41 | 0.25 |
| 1999 | 1.38 | 2.19 | 0.85 | 2.22 | 0.24 |
| 2000 | 1.06 | 0.93 | 0.65 | 1.73 | 0.25 |
| 2001 | 0.85 | 0.62 | 0.52 | 1.39 | 0.25 |
| 2002 | 0.98 | 1.13 | 0.61 | 1.59 | 0.24 |
| 2003 | 1.04 | 0.97 | 0.64 | 1.67 | 0.24 |
| 2004 | 0.92 | 0.87 | 0.57 | 1.48 | 0.24 |
| 2005 | 0.52 | 0.51 | 0.32 | 0.86 | 0.25 |
| 2006 | 0.52 | 0.69 | 0.32 | 0.85 | 0.25 |
| 2007 | 0.87 | 0.61 | 0.51 | 1.50 | 0.28 |
| 2008 | 0.83 | 0.69 | 0.46 | 1.47 | 0.30 |
| 2009 | 0.49 | 0.39 | 0.27 | 0.89 | 0.30 |
| 2010 | 0.41 | 0.29 | 0.22 | 0.76 | 0.31 |
| 2011 | 0.57 | 0.35 | 0.31 | 1.05 | 0.32 |
| 2012 | 0.23 | 0.18 | 0.12 | 0.44 | 0.35 |
| 2013 | 0.22 | 0.12 | 0.12 | 0.42 | 0.33 |

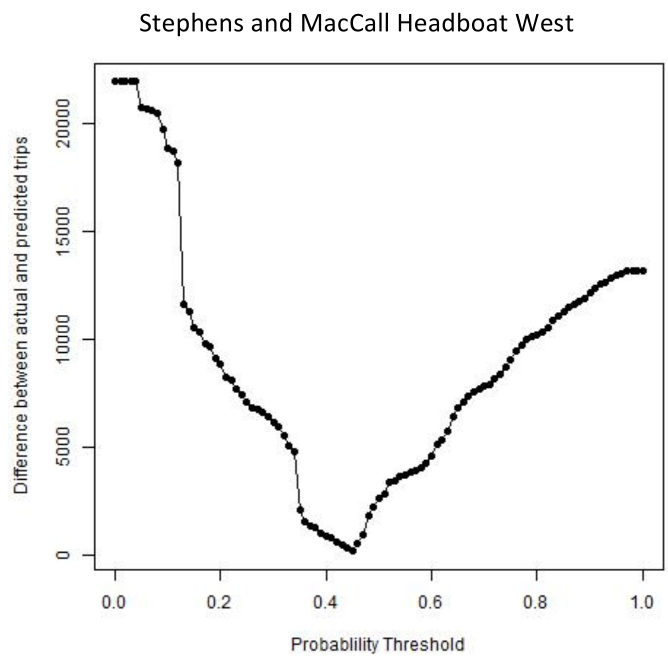
**Table 4:** Nominal CPUE and standardized index values (scaled to the timeseries mean) for the commercial handline fishery in the eastern Gulf of Mexico. Confidence intervals and CVs are for the standardized index.

| Eastern Gulf of Mexico | | | | | |
|---|---|---|---|---|---|
| | Index | | Confidence Limits (95%) | | |
| **Year** | **Standardized** | **Nominal** | **Lower** | **Upper** | **CV** |
| 1993 | 2.02 | 2.25 | 1.21 | 3.38 | 0.26 |
| 1994 | 2.00 | 2.74 | 1.22 | 3.28 | 0.25 |
| 1995 | 1.42 | 1.78 | 0.87 | 2.32 | 0.25 |
| 1996 | 0.95 | 0.98 | 0.57 | 1.60 | 0.26 |
| 1997 | 0.81 | 0.75 | 0.48 | 1.36 | 0.27 |
| 1998 | 0.90 | 0.74 | 0.53 | 1.52 | 0.27 |
| 1999 | 0.84 | 0.61 | 0.51 | 1.39 | 0.26 |
| 2000 | 0.59 | 0.52 | 0.35 | 1.00 | 0.27 |
| 2001 | 1.03 | 1.29 | 0.62 | 1.71 | 0.26 |
| 2002 | 1.47 | 1.83 | 0.90 | 2.38 | 0.25 |
| 2003 | 1.78 | 1.69 | 1.11 | 2.84 | 0.24 |
| 2004 | 1.28 | 1.40 | 0.78 | 2.09 | 0.25 |
| 2005 | 1.28 | 1.30 | 0.78 | 2.11 | 0.25 |
| 2006 | 0.70 | 0.68 | 0.42 | 1.18 | 0.26 |
| 2007 | 0.79 | 0.56 | 0.47 | 1.33 | 0.26 |
| 2008 | 0.57 | 0.38 | 0.33 | 0.96 | 0.27 |
| 2009 | 0.45 | 0.31 | 0.27 | 0.77 | 0.27 |
| 2010 | 0.50 | 0.40 | 0.30 | 0.84 | 0.26 |
| 2011 | 0.71 | 0.35 | 0.43 | 1.18 | 0.26 |
| 2012 | 0.39 | 0.17 | 0.22 | 0.71 | 0.31 |
| 2013 | 0.51 | 0.26 | 0.30 | 0.87 | 0.27 |

**FIGURES**

**Figure 1:** The difference between the number of records in which the target species are observed and those in which they are predicted to occur for each probability threshold using the Stephens and MacCall (2004) approach for the western (A) and eastern (B) Gulf of Mexico.

**A.**




**B.**

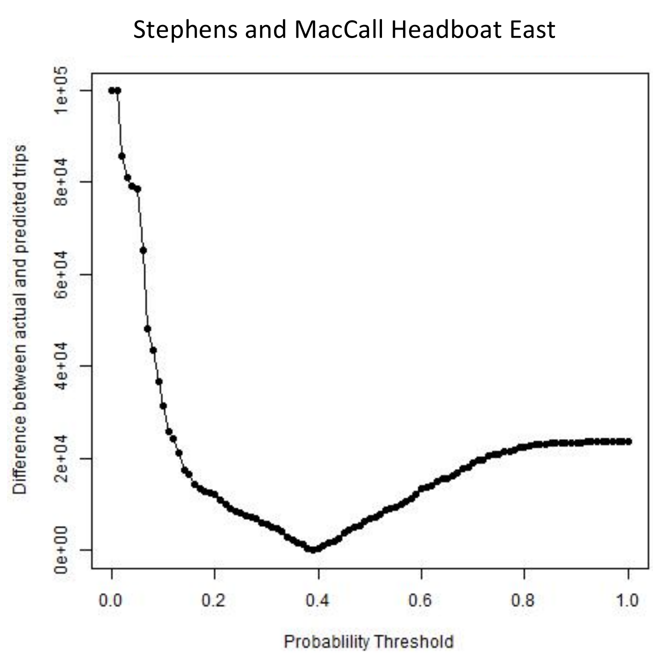
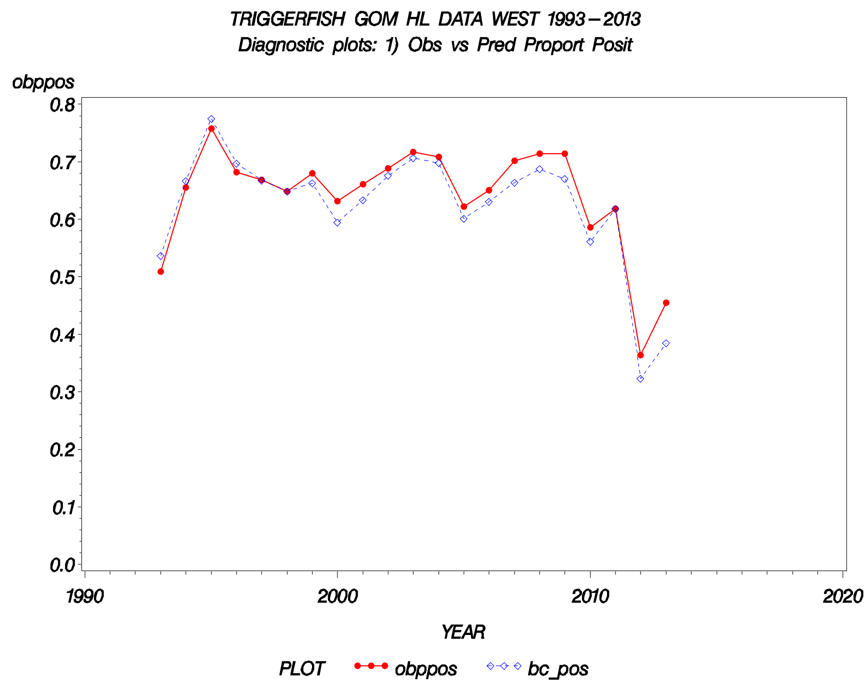


**Figure 2:** Diagnostic plots for the binomial model. Observed (red line) and predicted (blue line) proportion positive trips that caught the target species by year for the western Gulf of Mexico (A) and the eastern Gulf of Mexico (B).

**A.**




**B.**

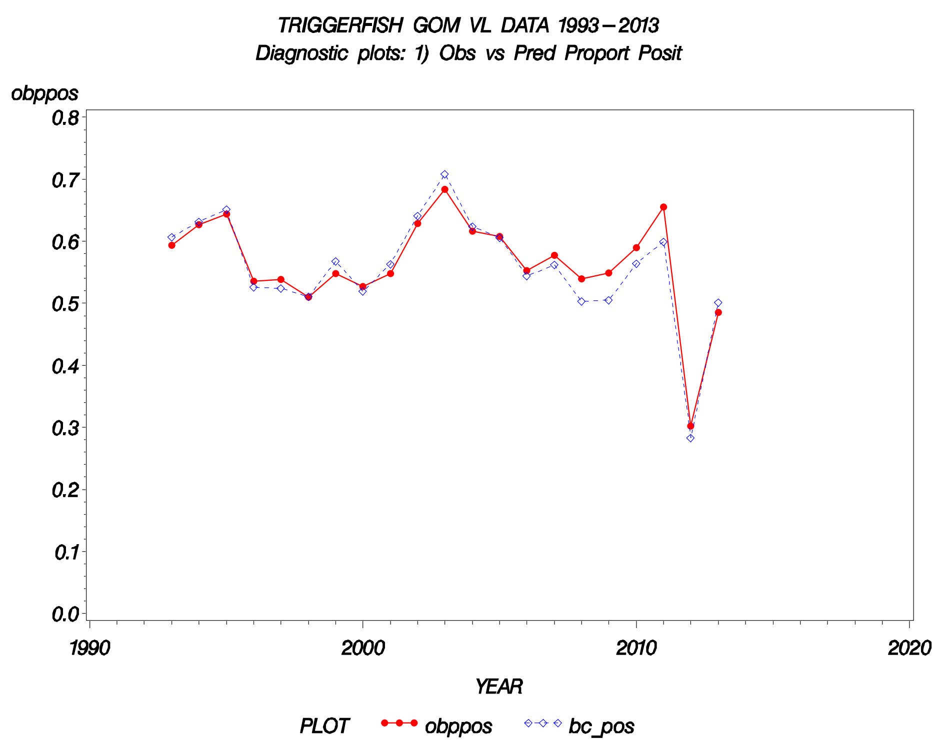
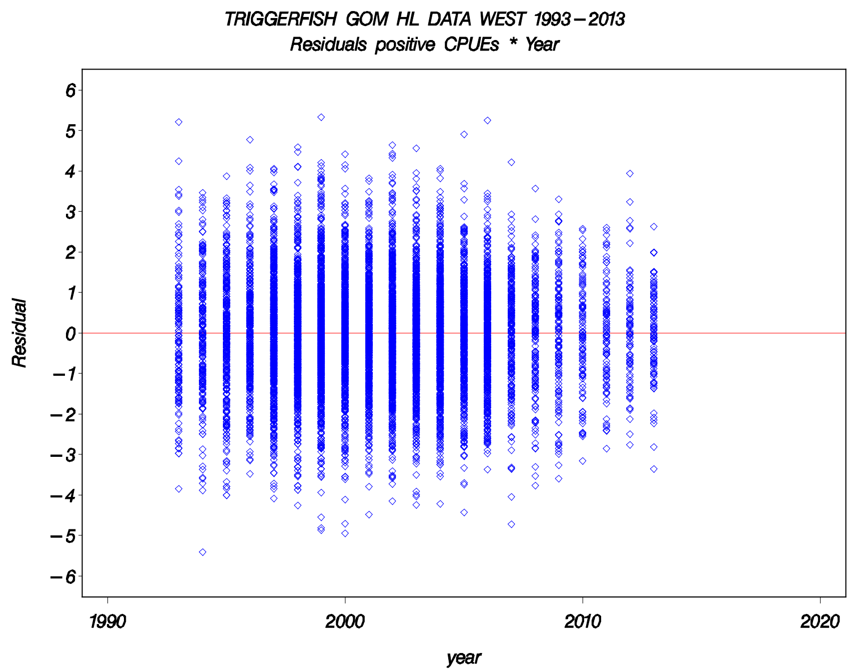
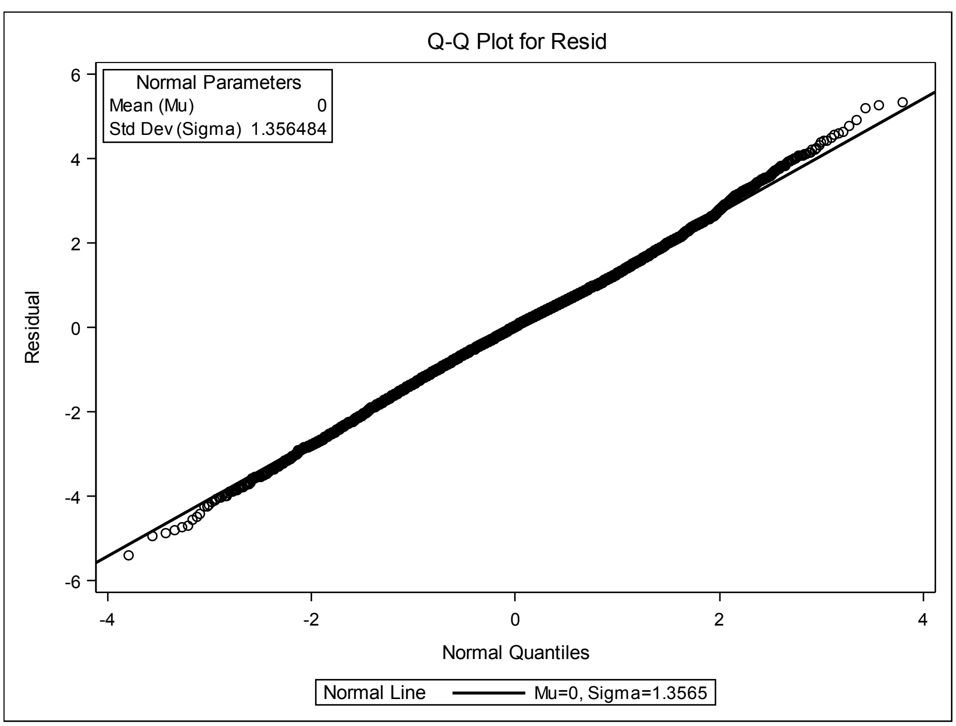


**Figure 3:** Diagnostic plots for the lognormal model: A) Residuals by year for the western Gulf of Mexico; B) Q-Q plot for the western Gulf of Mexico; C) Residuals by year for the eastern Gulf of Mexico; D) Q-Q plot for the eastern Gulf of Mexico.

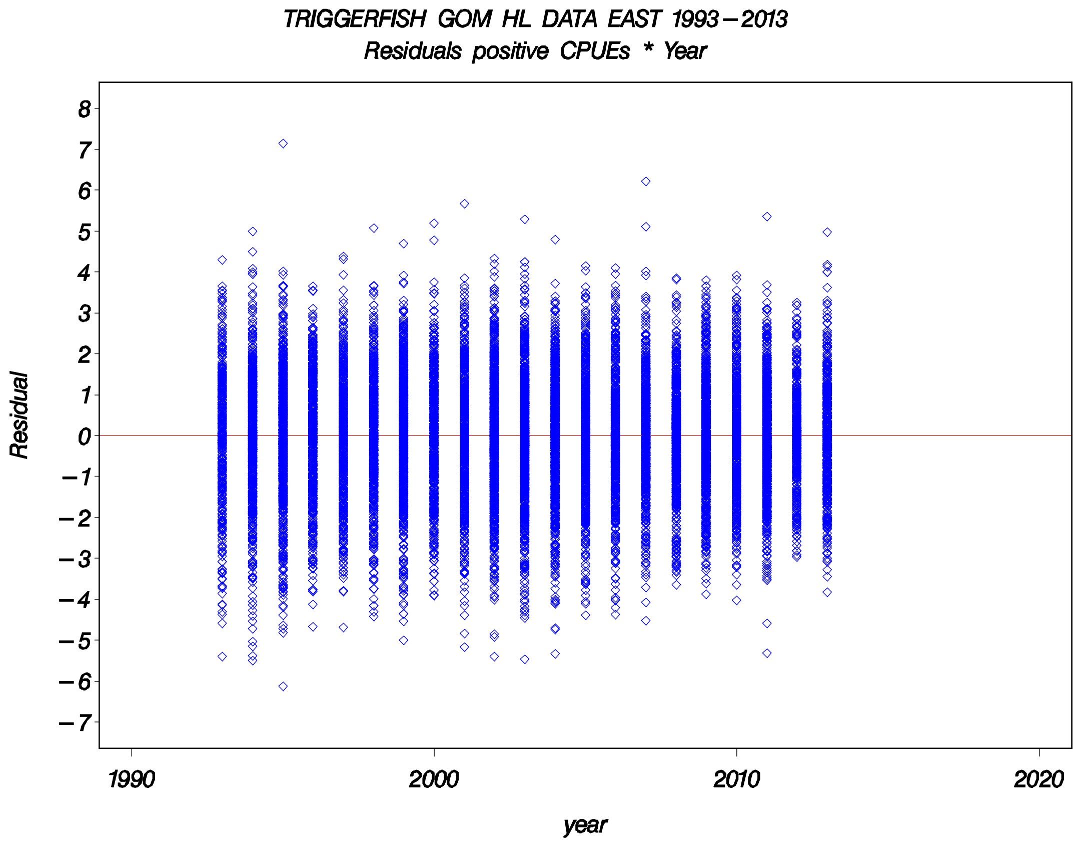
**A.**




**B.**



**C.**


TRIGGERFISH GOM HL DATA EAST 1993−2013
Residuals positive CPUEs * Year
Residual
8
7
6
5
4
3
2
1
0
−1
−2
−3
−4
−5
−6
−7
1990
2000
2010
2020
year

**D.**

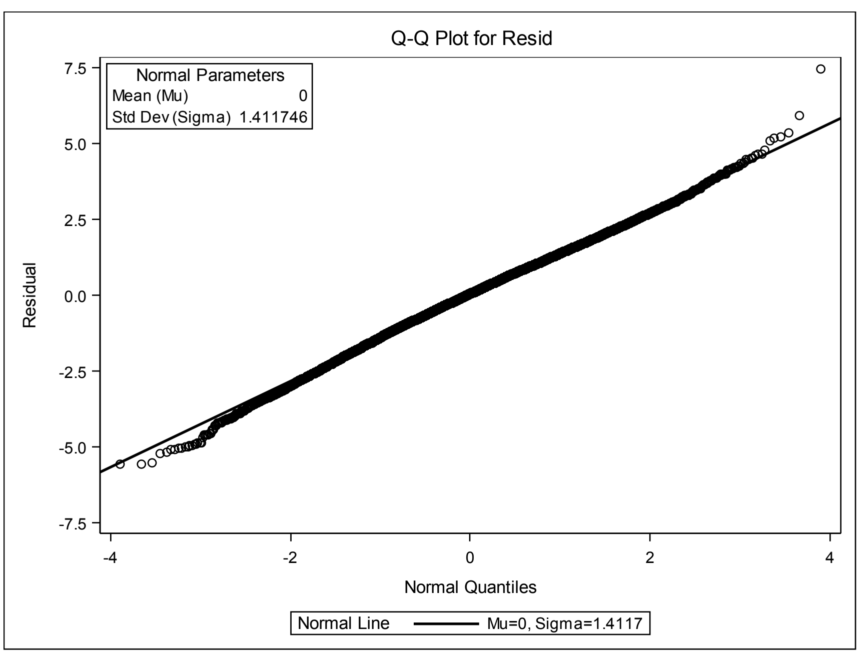
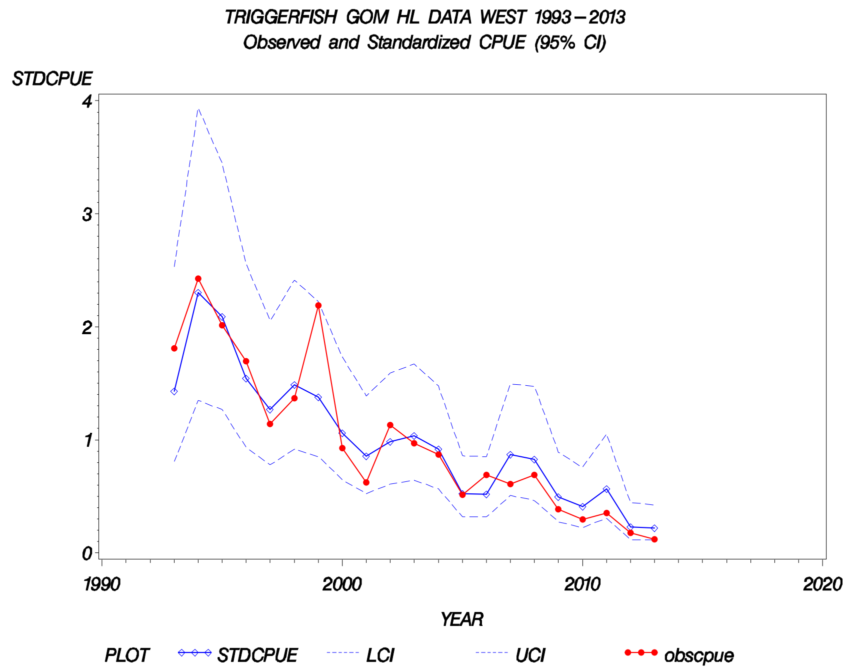

Q-Q Plot for Resid
Normal Parameters
Mean (Mu) 0
Std Dev (Sigma) 1.411746
Residual
7.5
5.0
2.5
0.0
-2.5
-5.0
-7.5
-4
-2
0
2
4
Normal Quantiles
Normal Line
Mu=0, Sigma=1.4117

**Figure 4:** Timeseries plots of nominal (red lines) and standardized (solid blue line) CPUE relative to the mean of the given timeseries for the western Gulf of Mexico (A) and the eastern Gulf of Mexico (B). 95% confidence intervals for the standardized CPUE are given by the dashed blue lines.

**A.**




**B.**

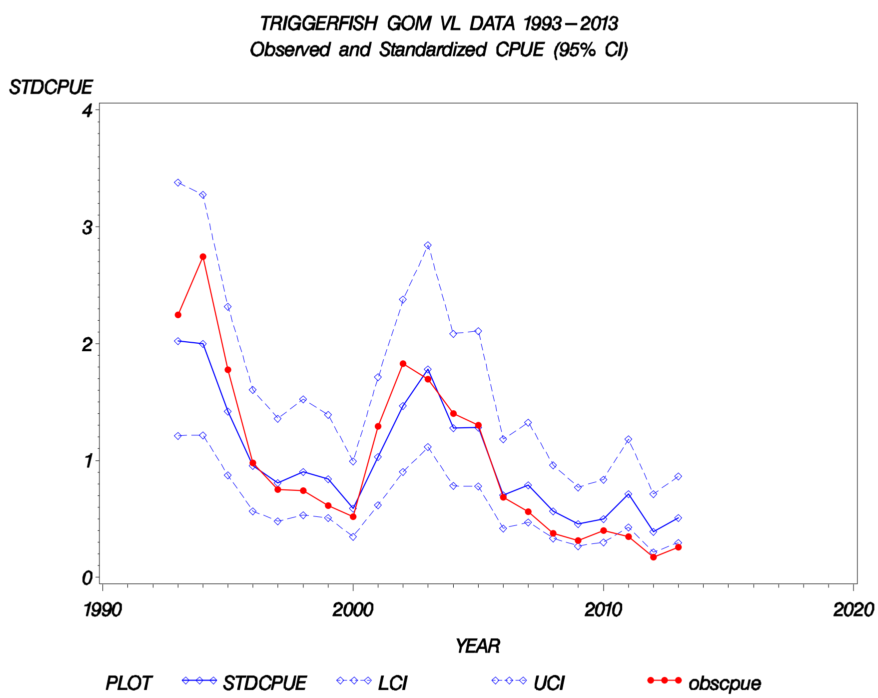
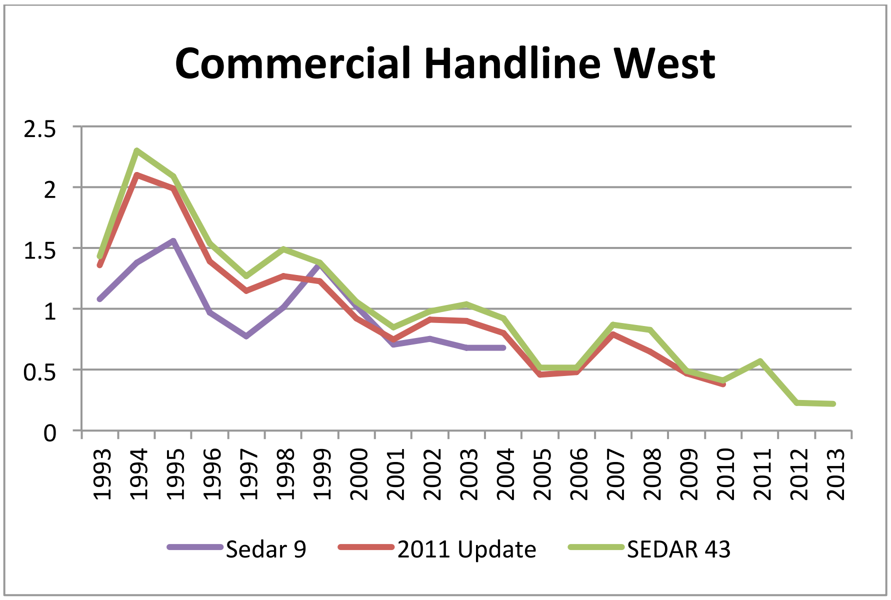


**Figure 5:** Timeseries plots of the SEDAR 9 (purple line), the 2011 update (red line), and the current SEDAR 43 (green line) standardized CPUE index for the western Gulf of Mexico (A) and eastern Gulf of Mexico (B).

**A.**




**B.**