Evaluation of MRFSS Intercept Data for Developing Gray Triggerfish and Blueline Tilefish Abundance Indices

Ed Martino, Julie Defilippi and Katie Drew

SEDAR32-DW-06

Submitted: 30 January 2013
Addendum: 4 March 2013*

***Addendum added to reflect changes made during the data workshop.**
Final index is found in the addendum.

Please cite this document as:

Martino, E., J. Defilippi, and K. Drew. 2013. Evaluation of MRFSS Intercept Data for Developing Gray Triggerfish and Blueline Tilefish Abundance Indices. SEDAR32-DW06. SEDAR, North Charleston, SC. 20 pp.

Evaluation of MRFSS Intercept Data for Developing Gray Triggerfish and Blueline Tilefish Abundance Indices

Ed Martino and Julie Defilippi, Atlantic Coast Cooperative Statistics Program

Abstract

We evaluated the feasibility of using MRFSS intercept survey data to develop gray triggerfish and blueline tilefish abundance indices. The MRFSS intercept database was subset to trips that either targeted or caught the species of interest. For both gray triggerfish and blueline tilefish, we included all observations within the MRFSS south Atlantic. mid-Atlantic, and northeast regions. We also included all waves, modes, areas, and gears for both species. Each set of grouped anglers in the intercept database was assumed to represent a single vessel-trip. The number of gray triggerfish positive trips was very similar to the total number of trips targeting or catching this species because the species was rarely targeted. There were 6,610 trips targeting or catching gray triggerfish and 6,485 trips (98%) reporting gray triggerfish catches (catch >0). Similarly, blueline tilefish was caught or targeted in 246 trips with 242 positive trips (98%) reporting catches of this species. There was insufficient data to develop a MRFSS index for blueline tilefish due to the very low number of observations almost exclusively comprised of positive tows. We used a log-normal generalized linear model (GLM) on the positive trip data to standardize the Gray Triggerfish index. Factors were selected for inclusion in the log-normal positive trip GLM using forward selection based on reductions in deviance for each component. The final model included the factors year, area fished, and mode (Residual deviance: 5894.7, 6446 df). Standardized model fit to the nominal CPUE time series appeared reasonably good although 95% confidence intervals were large due to both high variability in CPUE and modest sample sizes within factor levels. We suggest using alternative trip selection methodologies, as opposed to the MRFSS-targeted approach used here, to increase the number of observations and more importantly to reduce positive bias resulting from a lack of zero records. We will attempt to run these alternative selection methods and share results with the indices group at the Data Workshop.

Introduction

The MRFSS access-point angler intercept survey is conducted at public marine fishing access points to collect data on the individual catch of fishers, including species identification, total number and disposition of each species, and length and weight measurements of retained fish, as well as information about the fishing trip and the angler's fishing behavior. For more information on the methodology and variables collected, see the MRFSS Data User's Manual (available at http://www.st.nmfs.noaa.gov/st1/recreational/pubs/data_users/index.html).

In the south Atlantic, mid-Atlantic, and northeast survey regions there were a total of 6,610 interviews conducted for trips either catching or targeting Gray Triggerfish for the years 1982-2011 (Table 1 and Figure 1). For the same regions and years, a total of only 246 trips were intercepted that caught or targeted blueline tilefish (Table 2 and Figure 2).

Methods

Data from 1982 – 2011 were used in this analysis because the assessment is not using data from 2012, and wave 1 was not sampled in 1981.

The unit of effort used was directed angler-trip hours. The MRFSS intercept database was subset to trips that either targeted or caught (regardless of disposition) the species of interest. Each set of grouped anglers in the intercept database was assumed to represent a single vessel-trip; anglers with no follower records were also assumed to represent a single vessel-trip. For the nominal index, total available catch (Type A catch) was divided by the product of number of anglers contributing to that catch multiplied by hours fished to obtain Type A catch-per-angler-hours. The number of unavailable fish (Type B1 + B2 catch) was summed over all Type B records and divided by the product of number of anglers contributing to that catch multiplied by hours fished to obtain Type B catch-per-angler-hours. The Type A and Type B catch per angler-trip estimates were added together to get total catch per angler-trip.

The MRFSS intercept survey only counts anglers who contribute to the total catch, thus estimates of total catch per angler-trip-hour may be biased high in cases where anglers in the group fished but did not catch anything. In addition, the directed trips designation may not adequately identify zero trips. Anglers targeting other species or who do not report a target species may still have taken a trip that could have caught the species of interest, and that zero trip would not be been included in the directed trips subset. This issue is especially problematic for gray triggerfish and blueline tilefish because these species are not typically targeted by anglers.

For both gray triggerfish and blueline tilefish, we included all observations within the MRFSS south Atlantic. mid-Atlantic, and northeast regions. We also included all waves, modes, areas, and gears for both species. The number of gray triggerfish positive trips was very similar to the total number of trips targeting or catching this species because the species was rarely targeted. There were 6,610 trips targeting or catching gray triggerfish and 6,485 trips (98%) reporting gray triggerfish catches (catch >0) (Table 1 and Figure 3). Similarly, blueline tilefish was caught or targeted in 246 trips with 242 positive trips (98%) reporting catches of this species. There is insufficient data to develop a MRFSS index for blueline tilefish due to the very low number of observations almost exclusively comprised of positive tows (Table 2)

We used a log-normal generalized linear model (GLM) on the positive trip data to standardize the Gray Triggerfish index. The pdf of ln(nominal CPUE) appeared normal and this assumption was reinforced by the normal-quantile plot (Figure 4). We initially considered using a delta-lognormal approach (Lo *et al.*, 1992) to standardize this index because the delta-lognormal method utilizes information from both catch numbers from positive trips and the percentage of trips catching the species of interest. However, we were unable to use the delta-lognormal standardization because the approach is not recommended when positive trips exceed 80% of total trips.

Factors were selected for inclusion in the log-normal positive trip GLM using forward selection based on reductions in deviance for each component. The factors considered included region, state, wave, area fished, mode, and gear. A factor was included in the model if it reduced the deviance by 5% or more. The final model includes the factors year, area fished, and mode (Residual deviance: 5894.7, 6446 df; see Table 3 and Table 4). For comparison and to demonstrate effects of sub-setting this data set, we also show model results based on one possible subset including only the south Atlantic and predominate fishing modes (see Figure 5 for specific factor levels included).

Conclusions

Standardized model fit to the nominal CPUE time series appeared reasonably good although 95% confidence intervals were large due to both variability in CPUE and modest sample sizes within factor levels (Figure 5 and Figure 6). Given the difficulty of selecting an appropriate subset of trips and interpreting catch per trip from the MRFSS dataset, which was not designed to produce a CPUE index, as well as our reliance on models based exclusively on non-zero data, caution should be used when interpreting and using this index. We suggest using alternative trip selection methodologies (Stephens and MacCall, 2004), beyond the traditional MRFSS approach used here, to increase the number of observations and more importantly to reduce positive bias resulting from a lack of zero records. We will attempt to run these alternative selection methods and share results with the indices group at or before the Data Workshop.

Literature Cited

Lo, N.C., L.D. Jacobson, and J. L. Squire. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Canadian Journal of Fisheries and Aquatic Sciences 49: 2515-2526.

Stephens, A., and A. MacCall. 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fisheries Research (Amsterdam) 70(2-3):299–310.

Table 1. Gray triggerfish - distribution of observations (all trips on top and positive trips on bottom for each factor)

Region

	N.Atl	Mid.Atl	S.Atl
num	52.000	989.000	5569.000
prcnt	0.008	0.150	0.843
num	52.000	919.000	5514.000
prcnt	0.008	0.142	0.850

State

	CT	DE	FL	GA	MD	MA	NJ	NY	NC	RI	SC	VA
num	4.000	218.000	3245.000	409.000	137.000	14.000	235.000	175.000	1452.00	34.000	463.000	224.000
prcnt	0.001	0.033	0.491	0.062	0.021	0.002	0.036	0.026	0.220	0.005	0.070	0.034
num	4.000	202.000	3230.000	409.000	126.000	14.000	221.000	160.000	1412.000	34.000	463.000	210.000
prcnt	0.001	0.031	0.498	0.063	0.019	0.002	0.034	0.025	0.218	0.005	0.071	0.032

Wave

	1.000	2.000	3.000	4.000	5.000	6.000
num	25.000	43.000	52.000	70.000	36.000	20.000
prcnt	0.102	0.175	0.211	0.285	0.146	0.081
num	22.000	43.000	52.000	70.000	35.000	20.000
prcnt	0.091	0.178	0.215	0.289	0.145	0.083

Mode

	Man-made shore	Beach/bank	Shore	Party	Charter	Party/Charter	Private/Rental	Uknown
num	189.000	59.000	34.000	20.000	32.000	1295.000	1839.000	3142.000
prcnt	0.029	0.009	0.005	0.003	0.005	0.196	0.278	0.475
num	176.000	48.000	33.000	18.000	27.000	1271.000	1832.000	3080.000
prcnt	0.027	0.007	0.005	0.003	0.004	0.196	0.282	0.475

Gear

	Hook& Line	Dip A/Frame Net	Gill Net	Seine	Trap	Spear	Hand	Unknown
num	5713.000	3.000	1.000	1.000	2.000	57.000	1.000	832.000
prcnt	0.864	0.000	0.000	0.000	0.000	0.009	0.000	0.126
num	5592.000	3.000	1.000	1.000	2.000	53.000	1.000	832.000
prcnt	0.862	0.000	0.000	0.000	0.000	0.008	0.000	0.128

Area

	Ocean <3 mi	Ocean >3 mi	Inland
num	2247.000	3794.000	569.000
prcnt	0.340	0.574	0.086
num	2202.000	3751.000	532.000
prcnt	0.340	0.578	0.082

Table 2. Blueline tilefish - distribution of observations (all trips on top, positive trips on bottom for each factor)

Region

	Mid.Atl	S.Atl
Num	13	233
prcnt	0.05284553	0.9471545
Num	13	229
prcnt	0.05371901	0.946281

State

	DE	FL	NJ	NC	SC	VA
Num	1	29	9	203	1	3
prcnt	0.004065041	0.1178862	0.0365854	0.8252033	0.00406504	0.0121951
Num	1	28	9	200	1	3
prcnt	0.004132231	0.1157025	0.0371901	0.8264463	0.00413223	0.0123

Wave

	1	2	3	4	5	6
Num	25	43	52	70	36	20
prcnt	0.10162602	0.1747967	0.2113821	0.2845528	0.1463415	0.0813008
Num	22	43	52	70	35	20
prcnt	0.09090909	0.177686	0.214876	0.2892562	0.1446281	0.0826446

Mode

	Party/Charter	Private/Rental	Uknown
Num	9	200	37
prcnt	0.03658537	0.8130081	0.1504065
Num	9	200	33
prcnt	0.03719008	0.8264463	0.136363

Gear

	Hook& Line
Num	246
prcnt	1
Num	242
prcnt	1

Area

	1	2
Num	24	222
prcnt	0.09756098	0.902439
num	23	219
prcnt	0.09504132	0.9049587

Figure 1. Percent distribution of gray triggerfish trips by factor and factor level for all data and only positive trips

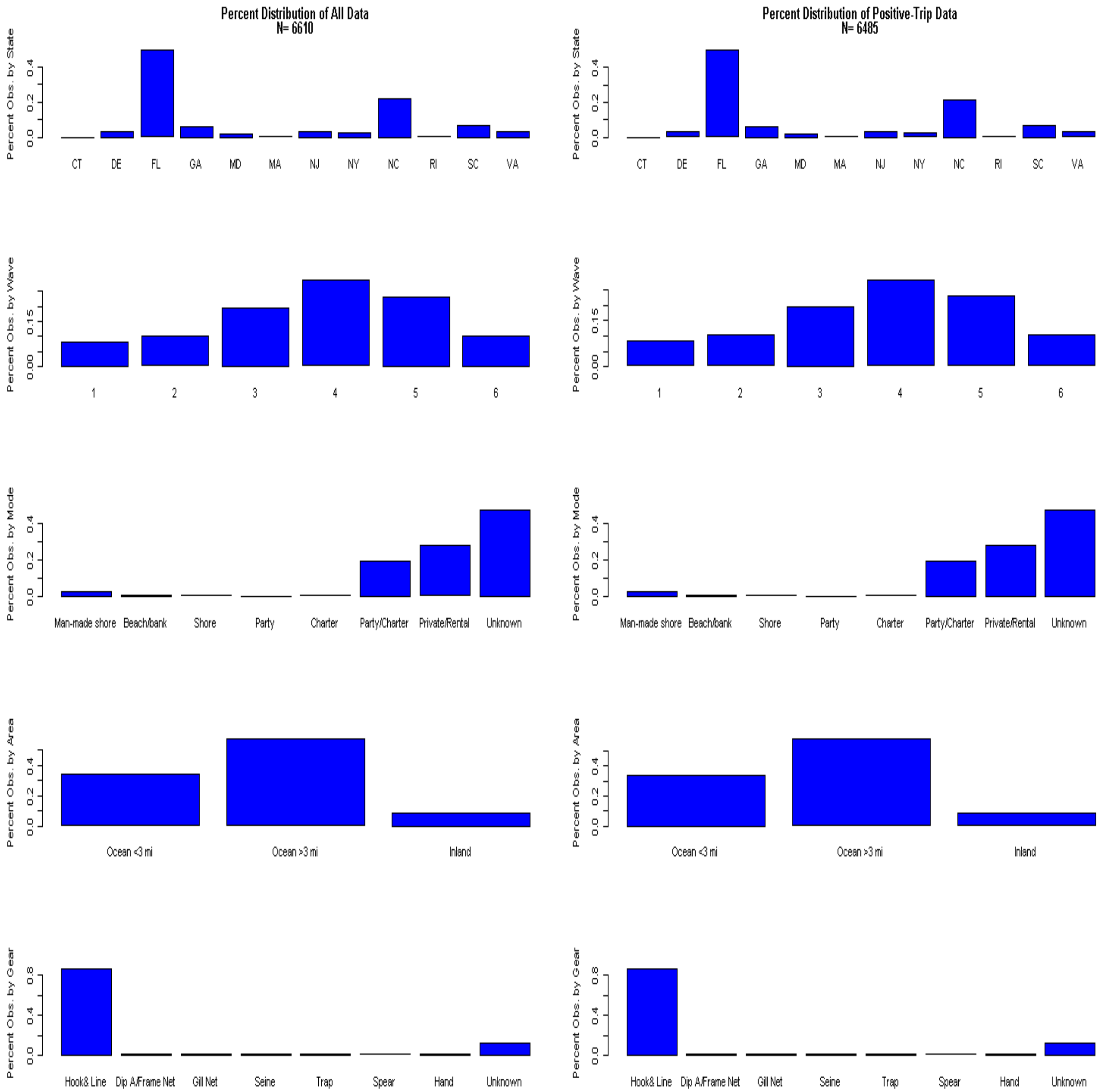


Figure 2. Percent distribution of blueline tilefish trips by factor and factor level for all data and only positive trips

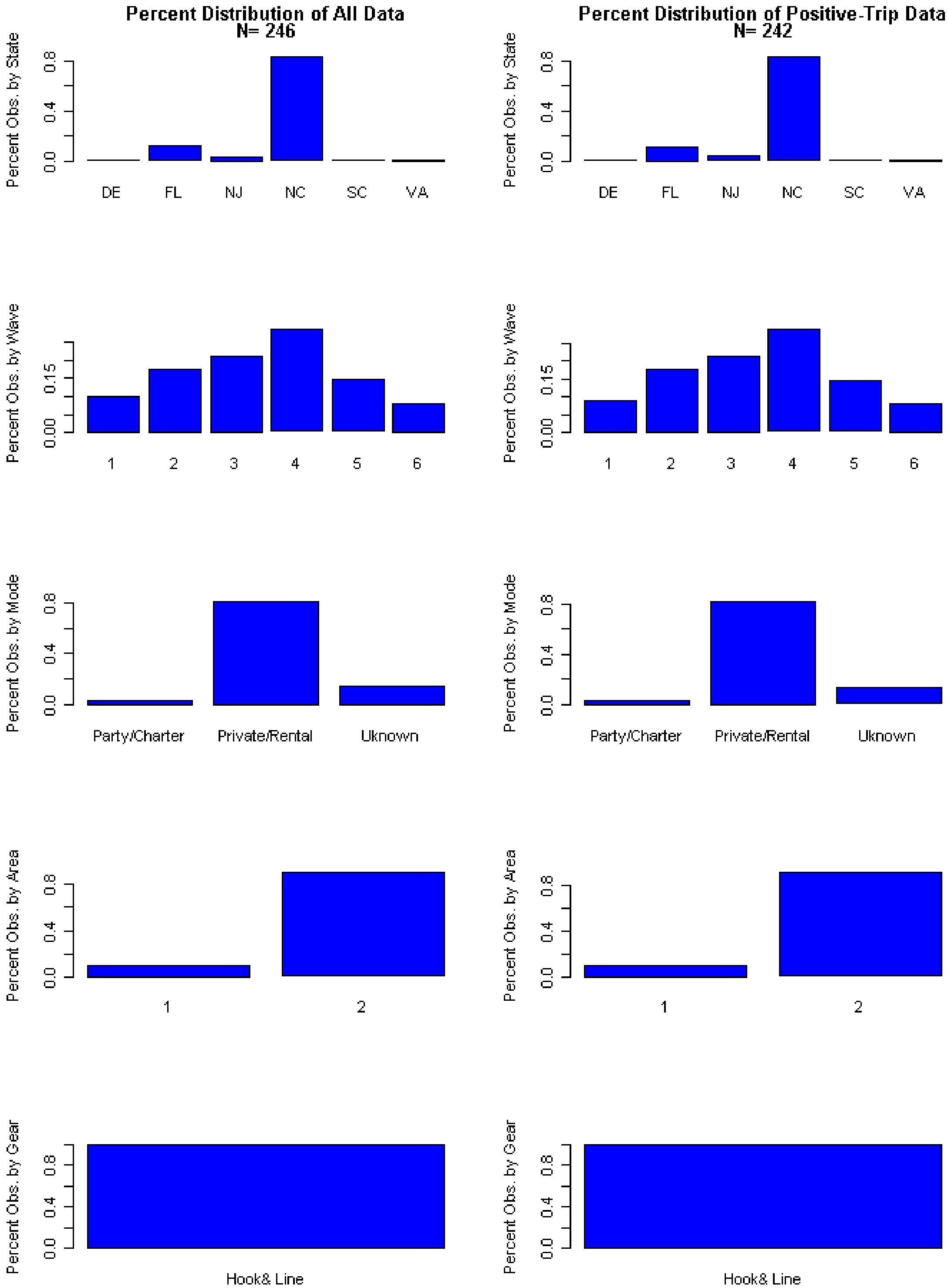


Figure 3. Distribution of gray triggerfish CPUE for all data and only positive trip data

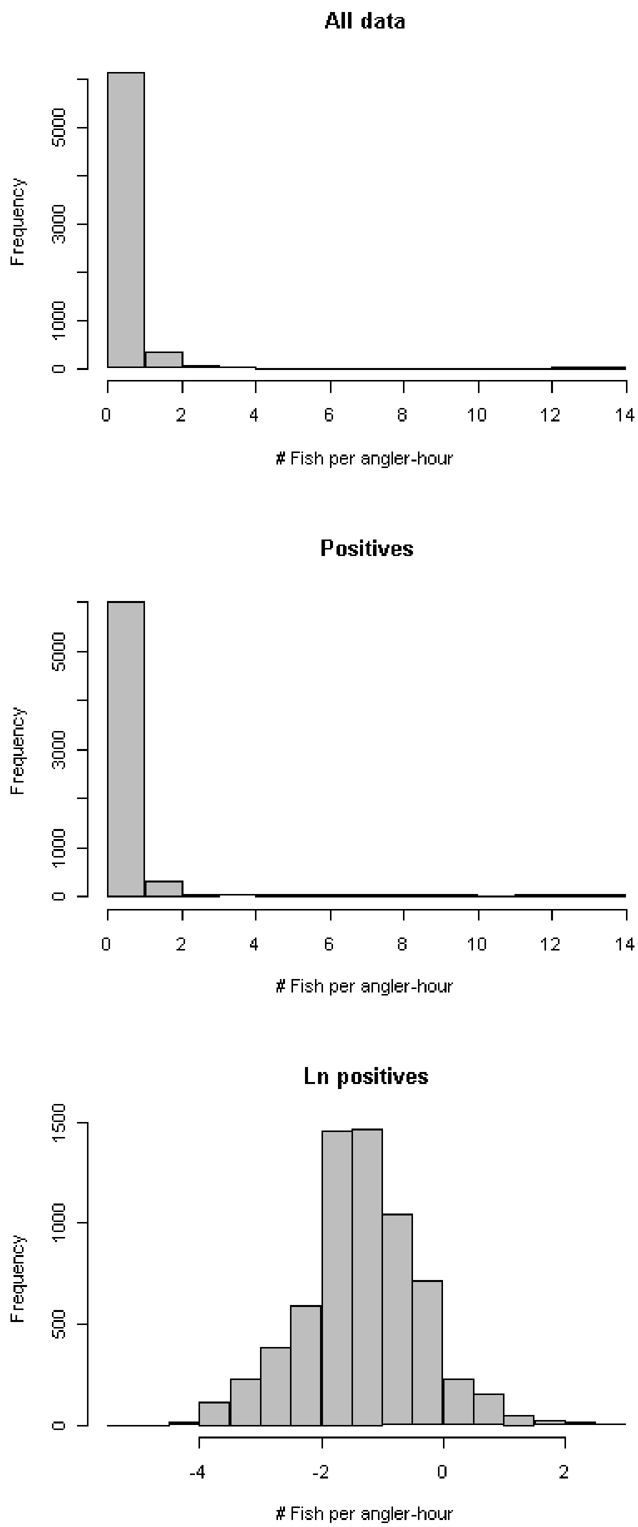


Figure 4. Pdf and normal-quantile plots to evaluate the distribution and normality of gray triggerfish ln(CPUE)

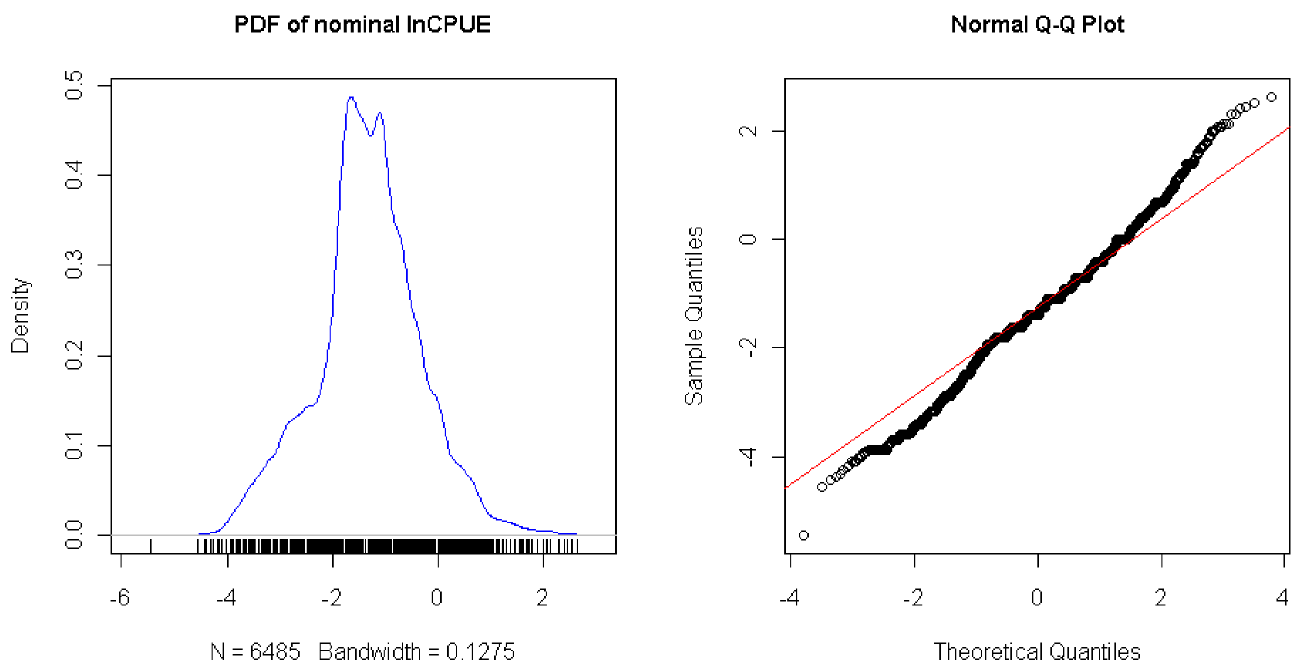
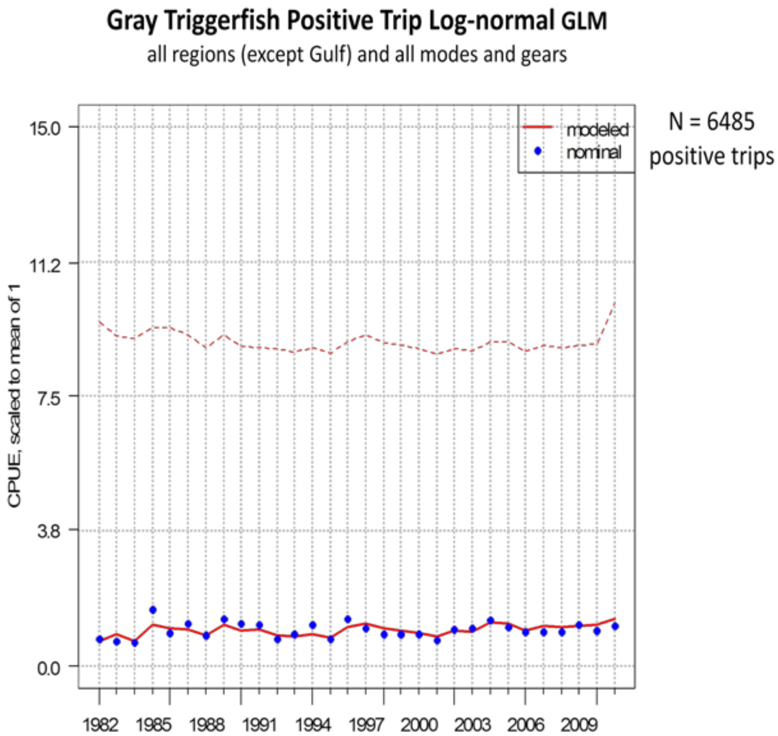


Figure 5. Gray triggerfish log-normal GLM on positive trip data for both all factor levels (top) and a subset of data including only the south Atlantic and predominant fishing modes (bottom)



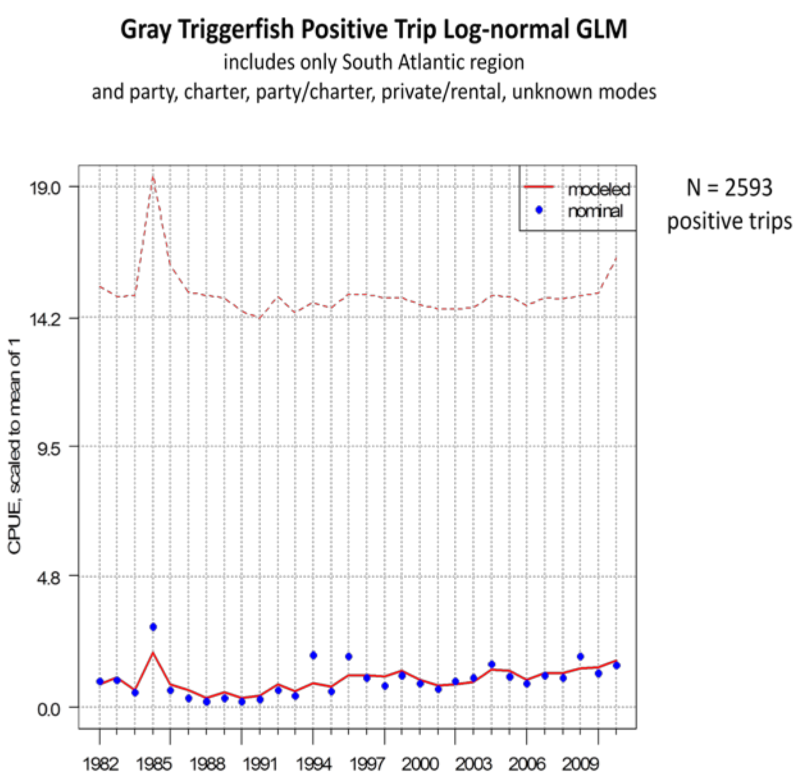


Table 3. Deviance table for factors included in final log-normal positive trip model (top model on previous page)

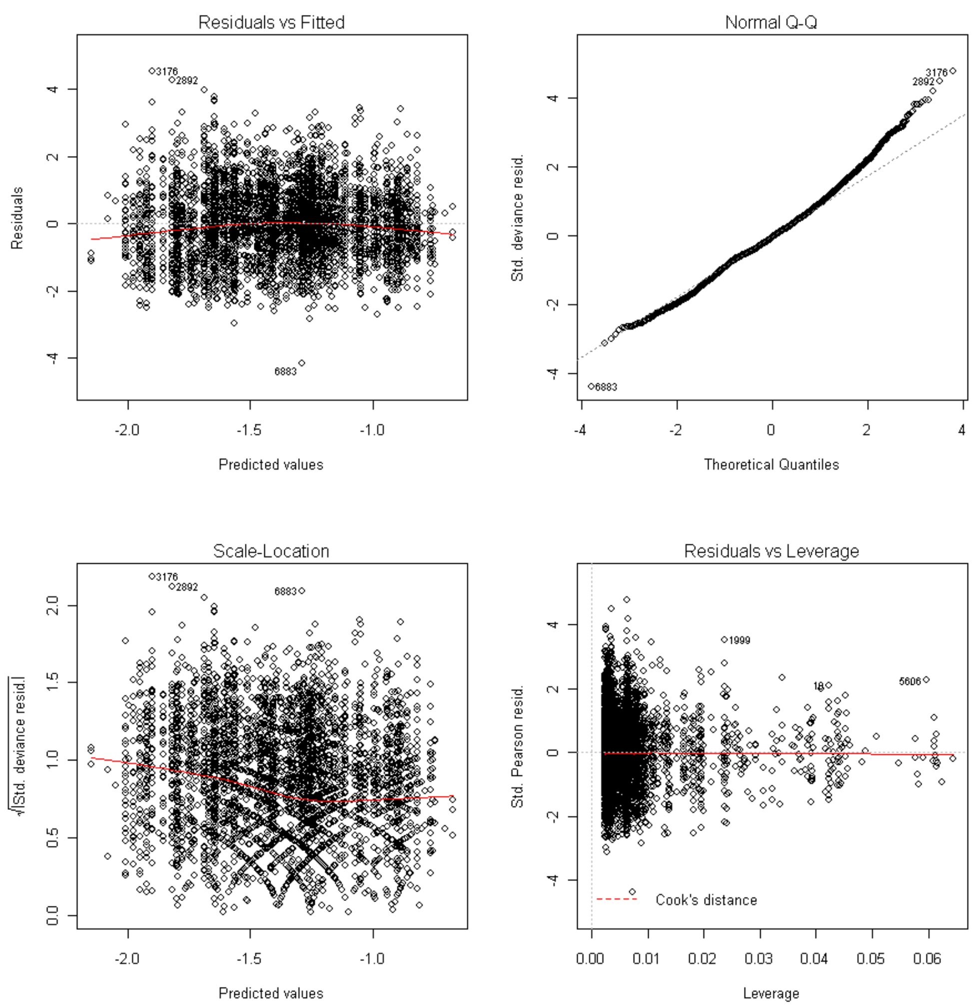
	Factor	Df	Deviance	Resid. Df	Resid. Dev	P(>|Chi|)	PercDevExplained
1	NULL	NA	NA	6484	6455.224	NA	NA
2	YEAR	29	113.8948	6455	6341.330	8.285385e-14	20.31829
3	AREA_F	2	262.9639	6453	6078.366	3.608880e-63	46.91153
4	MODE	7	183.6942	6446	5894.672	7.490673e-40	32.77018

Table 4. Convergence statistics for gray triggerfish positive trips model.

Final model lnCPUE ~ Year + AreaFished + Mode

Dispersion parameter for gaussian family 0.9144697
Null deviance: 6455.2 on 6484 degrees of freedom
Residual deviance: 5894.7 on 6446 degrees of freedom
AIC: 17865

Figure 6. Diagnostic residual plots for final gray triggerfish log-normal positive trips model including all factors levels (top model on previous page)



Addendum to the SEDAR 32 white paper entitled ***Evaluation of MRFSS intercept data for developing gray triggerfish and blueline tilefish abundance indices***

Ed Martino and Julie Defilippi, Atlantic Coast Cooperative Statistics Program
Katie Drew, Atlantic States Marine Fisheries Commission

Introduction

This document describes analyses completed by ACCSP during the SEDAR 32 data workshop. The results and methods described here are based on requested to changes to analyses reported in the previously submitted working paper. For a more detailed description of methods and preliminary results, see the SEDAR32 working paper entitled *Evaluation of MRFSS intercept data for developing gray triggerfish and blueline tilefish abundance indices.*

Methods

Gray triggerfish is not typically targeted by recreational fishers, and as a result the standard trip selection methodology employed by MRFSS (i.e. prim1 and prim2) to identify gray triggerfish trips results in data containing almost no zero trips. Preliminary analyses reported in the working paper were based on a log-normal GLM approach on positive trip data. These models and the resulting standardized indices were not very informative.

At the data workshop, the IWG agreed with ACCSP staff that the Jaccard method was suitable for gray triggerfish trip selection. The Jaccard method was used to identify additional gray triggerfish trips, and in particular to increase the number of zero trips. The Jaccard approach adds trips that report catches for other species identified as being highly associated with gray triggerfish. The approach calculates Jaccard coefficients to measure the degree of association between gray triggerfish and other species caught by fishers. Coefficients are calculated on a species presence-absence matrix and in this sense is similar to the Stephens and MacCall (2004) trip-selection approach. However, the Jaccard approach is appealing because it retains all trips that caught the focal species. The formula for the Jaccard coefficient used to measure the strength of species associations is shown below.

$$S_j = \frac{a}{a+b+c}$$

Where: a = # trips where triggerfish AND species j were caught
b = # trips where triggerfish was caught but NOT species j
c = # trips where species j was caught but NOT triggerfish

The indices group also recommended sub-setting of the gray triggerfish MRFSS data to include only the private/rental and charter fishing modes, ocean <3 and ocean >3 fishing areas, hook-and-line gear type. The six MRFSS wave levels were binned into three aggregate "seasonal" levels. Only data for the years 1993-2011 contained sufficient trip data for these factors and factor levels. In addition, the data was subset to include only Florida (east coast), Georgia, and North Carolina because non-zero gray triggerfish catches were too uncommon in South Carolina and more northern states. Trip data from Georgia and North Carolina was aggregated to increase the number of positive trips in this state group.

Jaccard-based trip selection was conducted on Florida and the Georgia-North Carolina groups separately. Thus, trip selection was based on species associations within each state group. Species identified as being highly associated with gray triggerfish in Florida and Georgia-North Carolina are shown in Figure 1 and Figure 2, respectively.

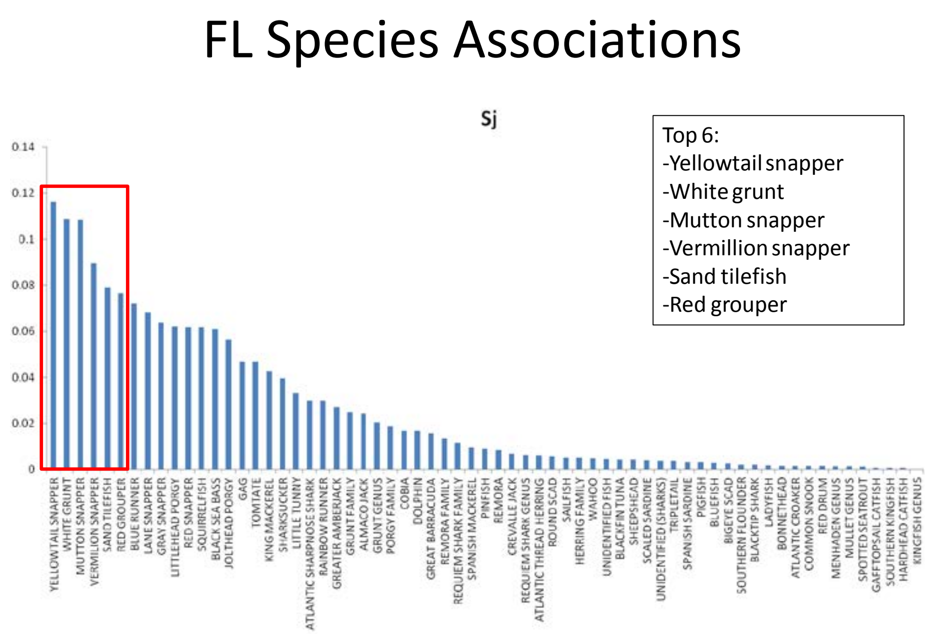


Figure 1. Jaccard coefficients showing association between gray triggerfish and other species caught in Florida. The top 6 most highly associated species were used to select additional trips to balance the degree of species association with the number of zero trips added.

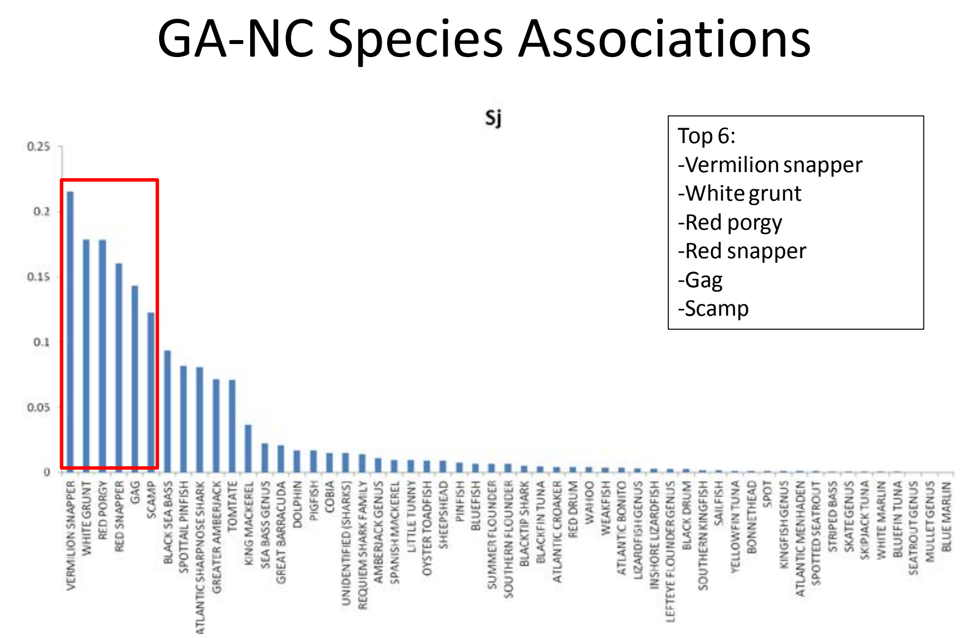


Figure 2. Jaccard coefficients showing association between gray triggerfish and other species caught in the Georgia-North Carolina aggregate group. The top 6 most highly associated species were used to select additional trips to balance the degree of species association with the number of zero trips added.

All trips were subsequently combined to represent a singly "coastal" trip dataset. The number of trips in the combined dataset was approximately 2-fold higher than the number of trips in the dataset based on the standard MRFSS trip-selection methodology (Figure 3, and see original white paper). The proportion of positive trips among factors and factor levels varied between 25 and 35 percent (Figure 4).

The increased number of trips, and in particular the number of zero trips, allowed us to calculate a standardized recreational index using the delta log-normal approach. The delta-lognormal approach was used because it incorporates both catch numbers from positive trips and the percentage of trips catching the species of interest (Lo *et al.*, 1992). The calculation of nominal CPUE is described in the original working paper. The distribution of gray trigger ln(CPUE) data was normal (Figure 5). Factors were selected for inclusion in the log-normal positive trip GLM using forward selection based on reductions in deviance for each component. The factors considered included state, wave, area fished, and mode. A factor was included in the model if it reduced residual deviance by 5% or more. The year factor was included in all model runs. The final positive trips model included year, wave, area, and mode, and the final proportion-positive model included year, wave, and state. Note that the wave and state factors used in these models reflect the aggregated groups described above.

Revised Results and Conclusion

The modeled standardized index provided a reasonably good fit to the nominal CPUE time series (3916.3 and 15266.8 residual dev., Table 1), and suggests a slight positive trend over the 1993-2011 time period (Figure 6 and Figure 7). Caution should be used when interpreting and using this index due to difficulties of selecting appropriate trip-data subsets, interpreting catch per trip from the MRFSS survey which was not designed for index development, and the reliance on alternative trip-selection methods to increase zero-trip numbers.

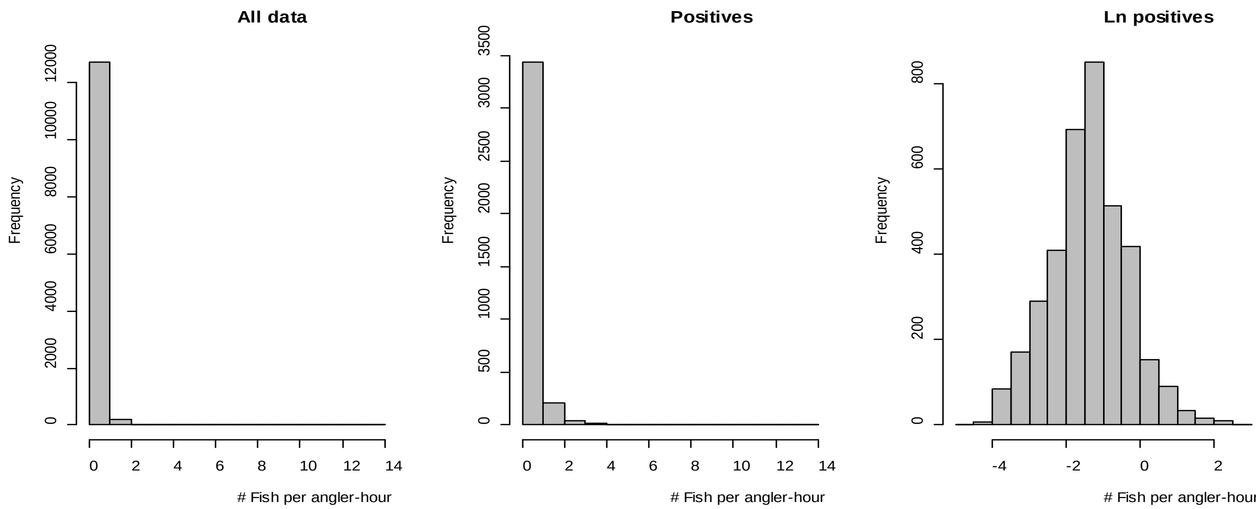


Figure 3. Distribution of gray triggerfish CPUE for all data and only positive trip data.

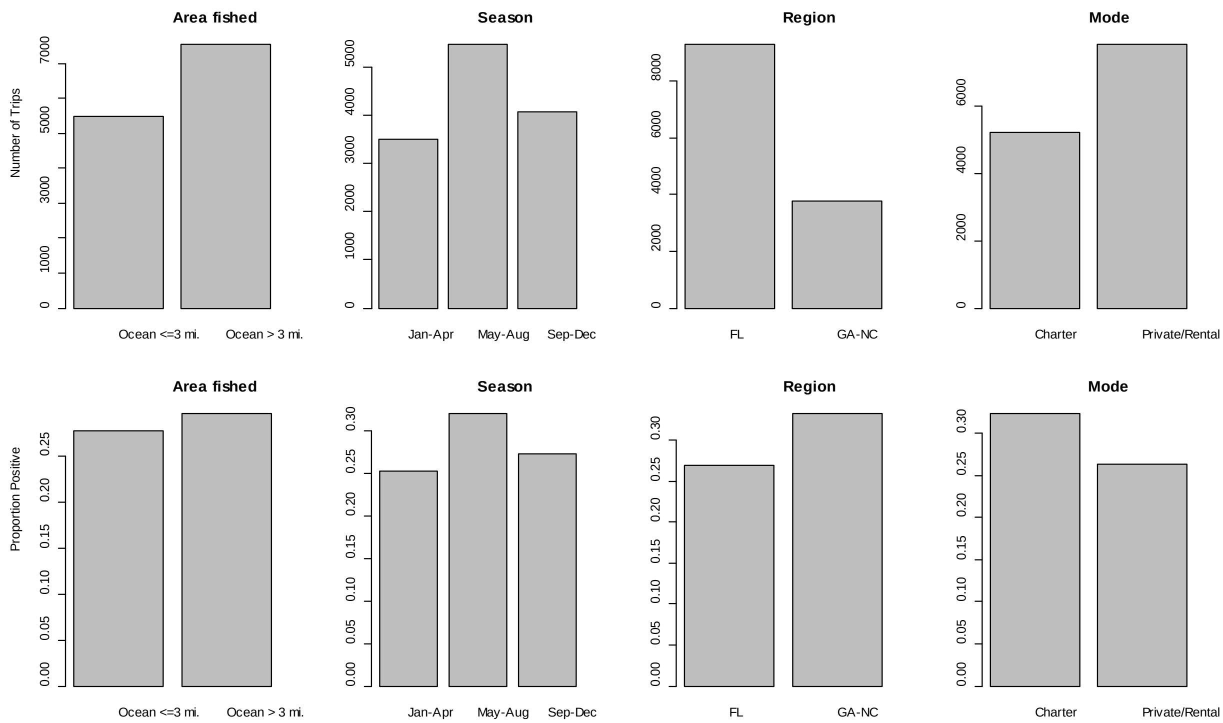


Figure 4. Percent distribution of blueline tilefish trips by factor and factor level for all data and only positive trips

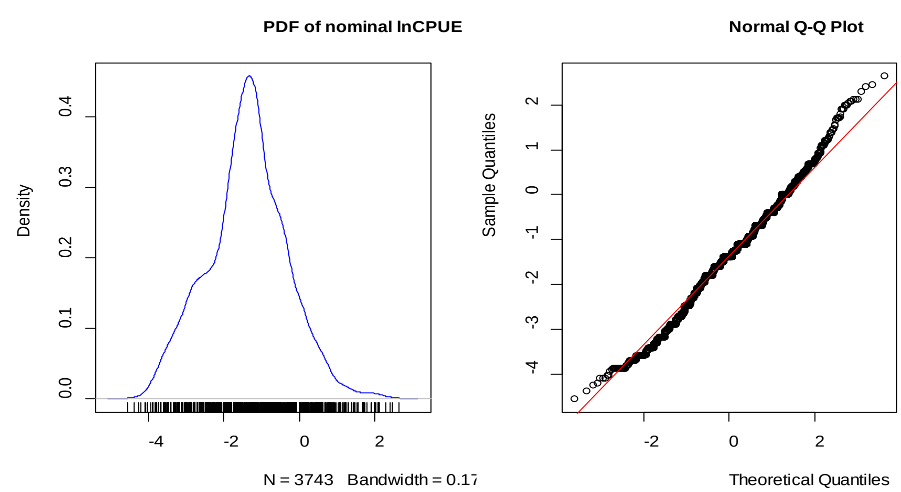


Figure 5. Pdf and normal-quantile plots to evaluate the distribution and normality of gray triggerfish ln(CPUE)

- **Positive Trips**

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)	PercDev Explained
NULL	.	.	3742	4237.3 .		.
YEAR	18	41.4	3724	4195.9	0.003	12.9
AREA_F	1	162.4	3723	4033.5	0.000	50.6
WAVE	2	17.4	3721	4016.2	0.000	5.4
STATE	1	1.5	3720	4014.7	0.240	0.5
MODE	1	98.4	3719	3916.3	0.000	30.7

Choose factors that reduce the deviance by 5% or more

- **Proportion Positive**

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)	PercDev Explained
NULL	.	.	13024	15624.3 .		.
YEAR	18	268.0	13006	15356.3	2E-46	71.5
AREA_F	1	15.6	13005	15340.7	8E-05	4.2
WAVE	2	43.0	13003	15297.8	5E-10	11.5
STATE	1	30.9	13002	15266.8	3E-08	8.3
MODE	1	17.1	13001	15249.8	4E-05	4.6
HOURS	1	0.1	13000	15249.7	8E-01	0.0

Table 1. Deviance tables used to select factors in positive trip and proportion-positive models. Highlighted factors were included in final models.

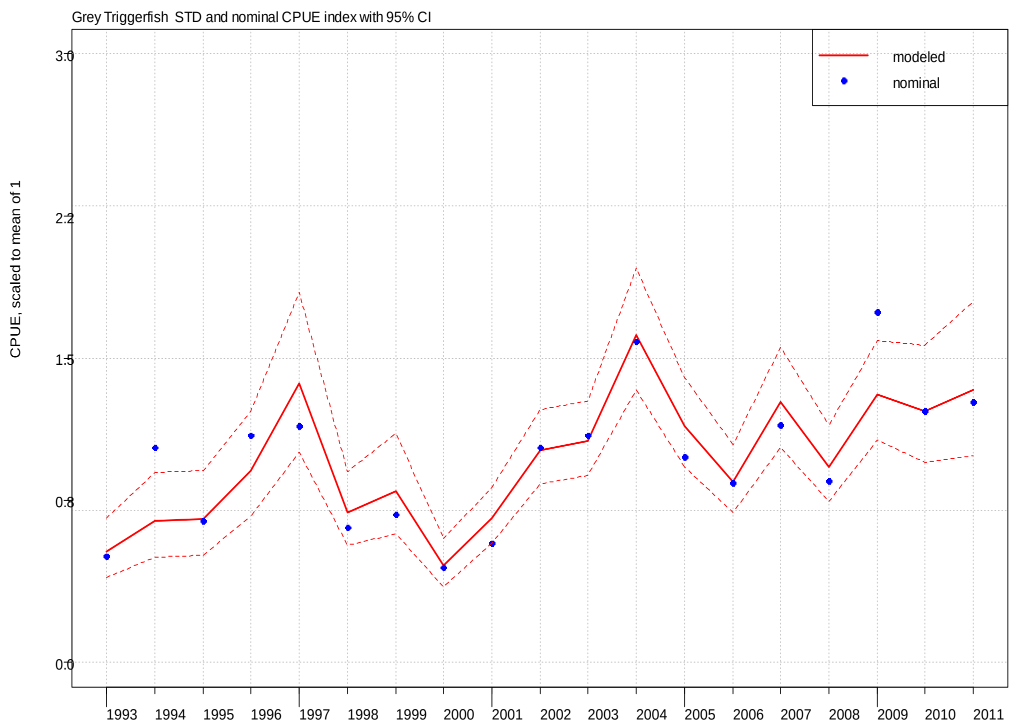


Figure 6. Nominal and standardized gray triggerfish indices based on delta log-normal model.

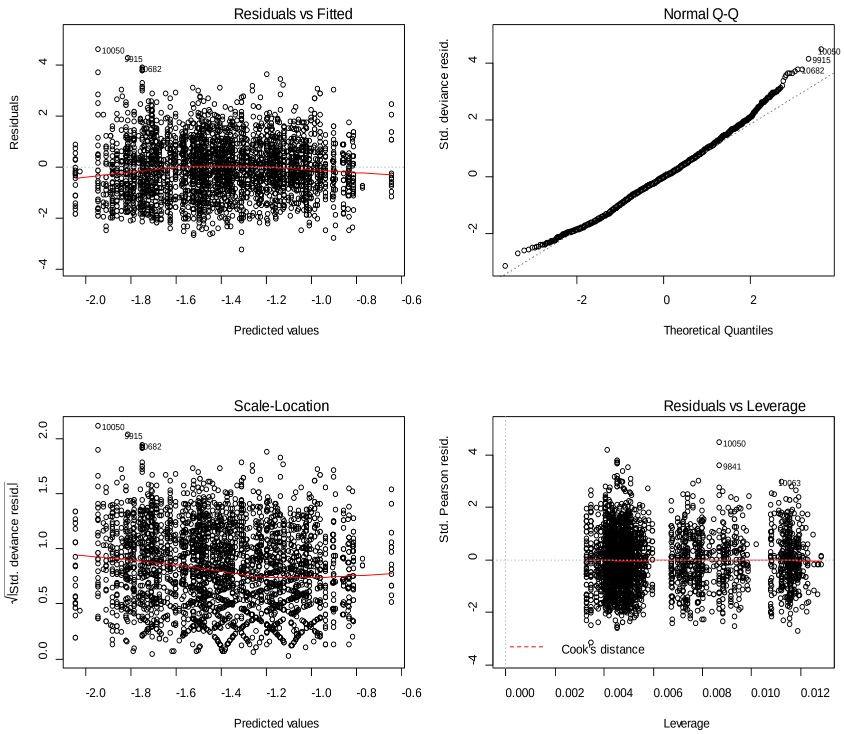


Figure 7. Diagnostic residual plots from the positive trips model.

References cited

Lo, N.C., L.D. Jacobson, and J. L. Squire. 1992. Indices of relative abundance from fish spotter data based on delta-lognormal models. Canadian Journal of Fisheries and Aquatic Sciences 49: 2515-2526.

Stephens, A., and A. MacCall. 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fisheries Research (Amsterdam) 70(2-3):299–310.