SEDAR10-DW20

Standardized catch rates of gag from the headboat fishery off the Southeastern United States

January 2006
Revised: February 2006 to include DW recommendations (Appendix)

Contacts: Rob Cheshire and Kyle Shertzer
Rob.Chesire@noaa.gov
Kyle.Shertzer@noaa.gov

Overview

Using catch-per-unit-effort (CPUE) data from the headboat fishery off the Southeastern U.S., we computed standardized catch rates of gag for possible use as an index of abundance. The time series spans 1973–2004. Trips used in the analysis were selected based on the probability of catching gag, computed according to the method of Stephens and MacCall (Stephens and MacCall, 2004, Fish. Res. 70:299-3210). Standardized catch rates were estimated using a generalized linear model assuming delta-lognormal error structure (Lo et al., 1992, Can. J. Fish. Aquat. Sci., 49:2515-2526). Explanatory variables were year, month, geographic area, and trip duration.

Data

A snapshot of the headboat catch record (CR) files was created on October 26, 2005. Years included in the analysis were 1973–2004. Before using the data, several issues had to be resolved. The primary issues were trip identification, duplicated data, black sea bass data entry errors, trips with anglers equal to 0, and trips with caught equal to 0.

We combined the annual catch record files, selecting trips that caught species in the snapper-grouper complex (SAS code in Appendix 1). Black sea bass were only recorded in weight and entered in the caught (numbers) field from 1973 to 1986 and in 1990. In 1992 the black sea bass weight was entered in either weight or caught and 0 was entered in the other. SAS code was written to move all weight data for black sea bass to the weight column for 1973-1986 and 1990. Likewise for 1992 the weight data were moved from the numbers column to the weight column if entered incorrectly. The caught (numbers) value was assigned as 1 for black seabass records where only weight was available to indicate this species was caught. A logical variable (1 for species caught) was created for use in determining species assemblages. The trip was defined as year, month, day, vessel, trip, period, anglers, and collection. A trip identification variable was created using these groupings. Collection is a primary key that was added in 1987. The data set was then exported to Microsoft Access as a table.

Duplicate values were present at two levels. Complete duplicates of all variables was believed to be erroneous and removed from the data prior to exporting from SAS to MS Access (1120 records with multiple species per trip). The fields duplicated were year, month, day, vessel, vessel type, trip, period, anglers, species, caught, weight, area, location, collection. Once in MS Access caught and weight fields were summed by trip for species recorded more than once per trip using a totals query. The query is grouped by the trip identifier and species. The first value of all duplicate species is used with the caught and weight field summed. Some of the summed data may be duplicate data. However, headboat personnel provided many reasons why species would be entered twice for a given trip;

1. Samplers were instructed to record numbers and weights of extremely large individuals of a species separately so as not to bias the mean weight estimates.
2. Misidentification of species was detected after records had been entered and reclassified as another species later.
3. Species with number caught greater than 999 were entered in twice because the database could only handle 3 digits for the caught field in early years.
4. Individuals of a species discovered after initial values entered and written in on the catch form.

A crosstab query was created using the totals query to create a matrix of trip by species with 1's in cells where a species was caught in a trip and blanks otherwise. This matrix was exported from MS Access as a comma separated file and opened in R. The blanks were replaced with zeros, to create a matrix of presence-absence data, where rows represented trips and columns represented species in the snapper-grouper complex. The workspace was saved as an .Rdata file.

Another query of the data was created to generate CPUE. Records with trips =0 or greater than 30 were removed as well as trips with anglers =0. Trip duration, a factor in the analysis, was assigned one of four possible levels: half-day, three-quarter day, full day, and multi-day (Table 1). Area, another factor, was assigned as in the data base, but with some areas combined (Table 2). Combined were areas 9 and 10 (central NC), 2 and 3 (southern NC), and 4 and 5 (SC). These three combinations were believed to better represent the resolution of our knowledge of area fished, as area is assigned to vessels rather than trips. CPUE was then calculated for each trip in units of gag/angler. (Units of gag/angler-hours was examined also) This data set was then exported as a comma separated file for use in R.

Methods

Standardized catch rates were estimated using a generalized linear model assuming delta-lognormal error structure (Lo et al., 1992, Can. J. Fish. Aquat. Sci., 49:2515-2526), in which the binomial distribution describes positive versus zero CPUE, and the normal distribution describes the log of positive CPUE. Explanatory variables were year, month, geographic area, and trip duration; and the response variable (CPUE) was in units gag per angler. Variability of estimates was estimated via empirical bootstrap (n=200).

Effective effort was based on those trips that caught gag (positive CPUE) and those that *could* have caught gag (zero catch, but positive effort). Positive catches are readily available from the data, but without information on targeting by fishermen, zero catches must be inferred. To do so, we applied the method of Stephens and MacCall (Stephens and MacCall, 2004, Fish. Res. 70:299-3210). In essence, the method uses multiple logistic regression to estimate a probability for each trip that gag was caught, given other species caught in that trip. Species used as factors in the regression were selected as those caught in at least 5% of trips. This cutoff simplifies the regression, by excluding rarely caught species; however, preliminary analyses indicated results were insensitive to the value of the cutoff (examined over a range of 0% to 10%). Trips were included if their associated probability was higher than a threshold probability. The threshold's value was defined as that which results in the same number of predicted and observed positive trips, as in Stephens and MacCall (2004). Our method differed slightly from that of Stephens and MacCall, in that we included all positive trips, not just those with probability higher than the threshold.

Results

Trips selected for the analysis (n=74,744 of a possible 264,1456 trips) included all positive trips (n=53,509) for gag and zero trips (n=21,235) identified via multiple logistic regression. The regression used 21 species as explanatory variables (Figure 2). A trip was identified as a zero trip if its probability of catching gag exceeded the threshold probability P=0.3409 (Figure 3).

Preliminary analyses using effort in units of angler-hours revealed a departure from lognormal error in positive CPUE (Figure 4A). With effort defined as number of anglers, however, that assumption appeared less violated (Figure 4B). Of course, hours fished would still be important in determining a trip's catch, an effect modeled by including trip duration as a factor.

The estimated index of abundance showed a general trend of decline (Figure 5, Table 3). Annual coefficients of variation, as estimated by empirical bootstrap, were generally about 5%.

Table 1. Trip duration from headboat sampling. For computing CPUE, trip duration was used as a factor, with levels as in column labeled “tripFACTOR.”

ID	ctrip	tripHRS	tripFACTOR	tripDAYS
1	1	5	half	0.5
2	21	5	half	0.5
3	9	5	half	0.5
4	29	5	half	0.5
5	3	7	threeQ	0.75
6	23	7	threeQ	0.75
7	2	10	full	1
8	25	18	fullplus	2
9	5	24	fullplus	2
10	6	36	fullplus	3
11	7	48	fullplus	4
12	8	60	fullplus	5
13	10	72	fullplus	6
14	11	84	fullplus	7

Table 2. Areas assigned to headboat vessels. For computing CPUE, several areas were combined: area 2 with 3, area 4 with 5, and area 9 with 10.

ID	area	area_n	def
1	1	NNCoff	NC/VA line to Ocracoke inlet offshore
2	2	SNCin	Topsail Island NC to NC/SC line inshore
3	3	SNCoff	Topsail Island NC to NC/SC line offshore
4	4	SCin	South Carolina inshore
5	5	SCoff	South Carolina offshore
6	6	GA	Georgia
7	7	NFL	North Florida
8	8	CFL	Central Florida
9	11	SFL	South Florida
10	12	KEYs	Florida Keys
11	17	TOR	Dry Tortugas
12	9	CNCin	Central North Carolina inshore
13	10	CNCoff	Central North Carolina offshore

Table 3

Year	CPUE (gag/angler)	standard deviation	CV
1973	0.25	0.01	0.06
1974	0.19	0.01	0.05
1975	0.12	0.01	0.05
1976	0.10	0.00	0.05
1977	0.10	0.00	0.05
1978	0.09	0.00	0.05
1979	0.14	0.01	0.04
1980	0.15	0.01	0.04
1981	0.12	0.01	0.04
1982	0.11	0.00	0.04
1983	0.12	0.00	0.04
1984	0.12	0.00	0.04
1985	0.11	0.00	0.04
1986	0.11	0.00	0.04
1987	0.12	0.00	0.04
1988	0.13	0.01	0.04
1989	0.13	0.01	0.04
1990	0.12	0.00	0.04
1991	0.12	0.01	0.05
1992	0.12	0.00	0.04
1993	0.12	0.00	0.04
1994	0.10	0.00	0.04
1995	0.11	0.00	0.04
1996	0.10	0.00	0.04
1997	0.10	0.00	0.04
1998	0.11	0.00	0.04
1999	0.08	0.00	0.04
2000	0.09	0.00	0.05
2001	0.08	0.00	0.04
2002	0.07	0.00	0.04
2003	0.07	0.00	0.06
2004	0.09	0.00	0.05

Figure 1. Area designations assigned to headboat vessels.

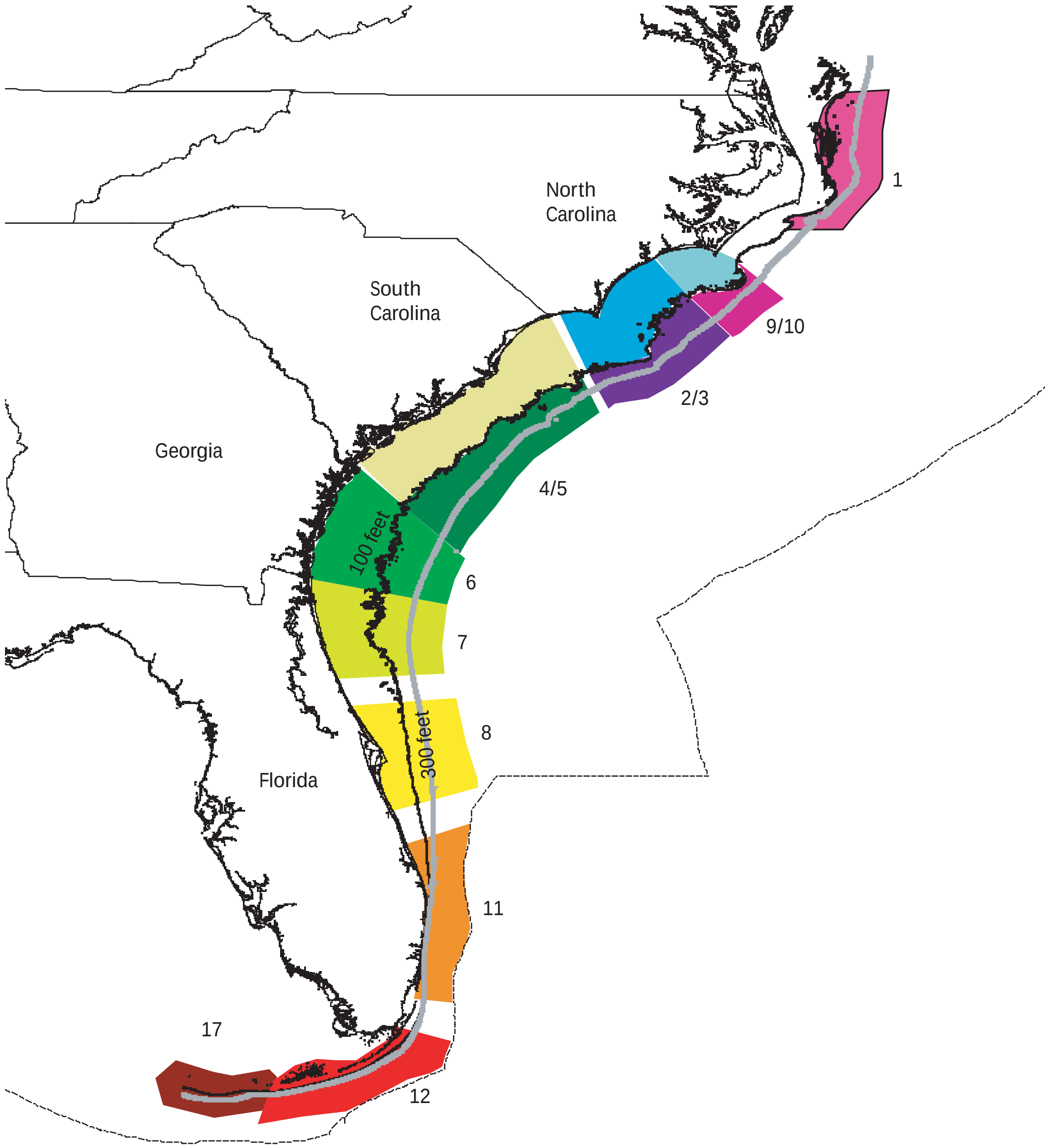


Figure 2. Estimates of species-specific regression coefficients used to estimate a trip's probability of catching gag.

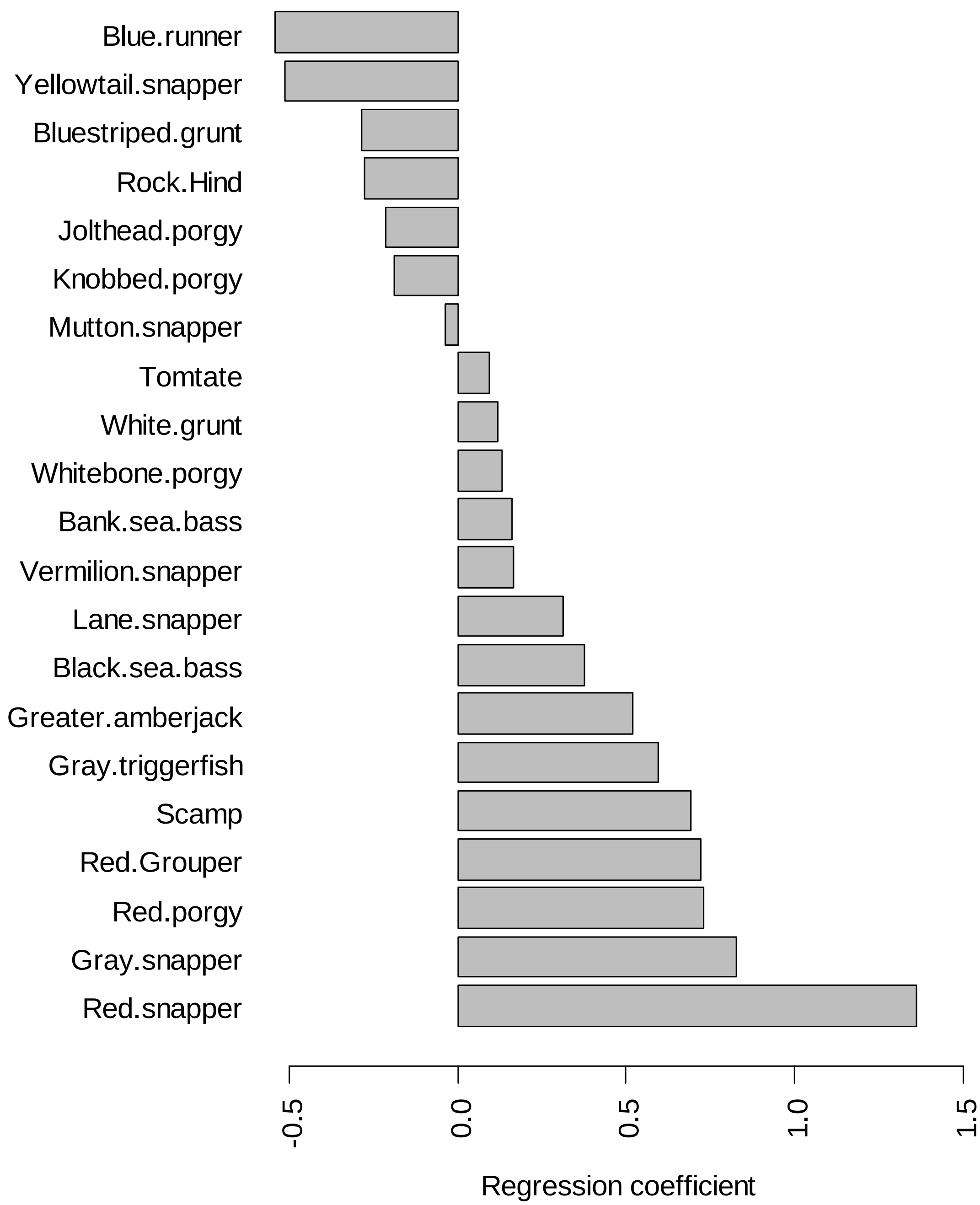


Figure 3. Absolute difference between observed and predicted number of positive gag trips. Top and bottom panels differ only in the range of probabilities shown.

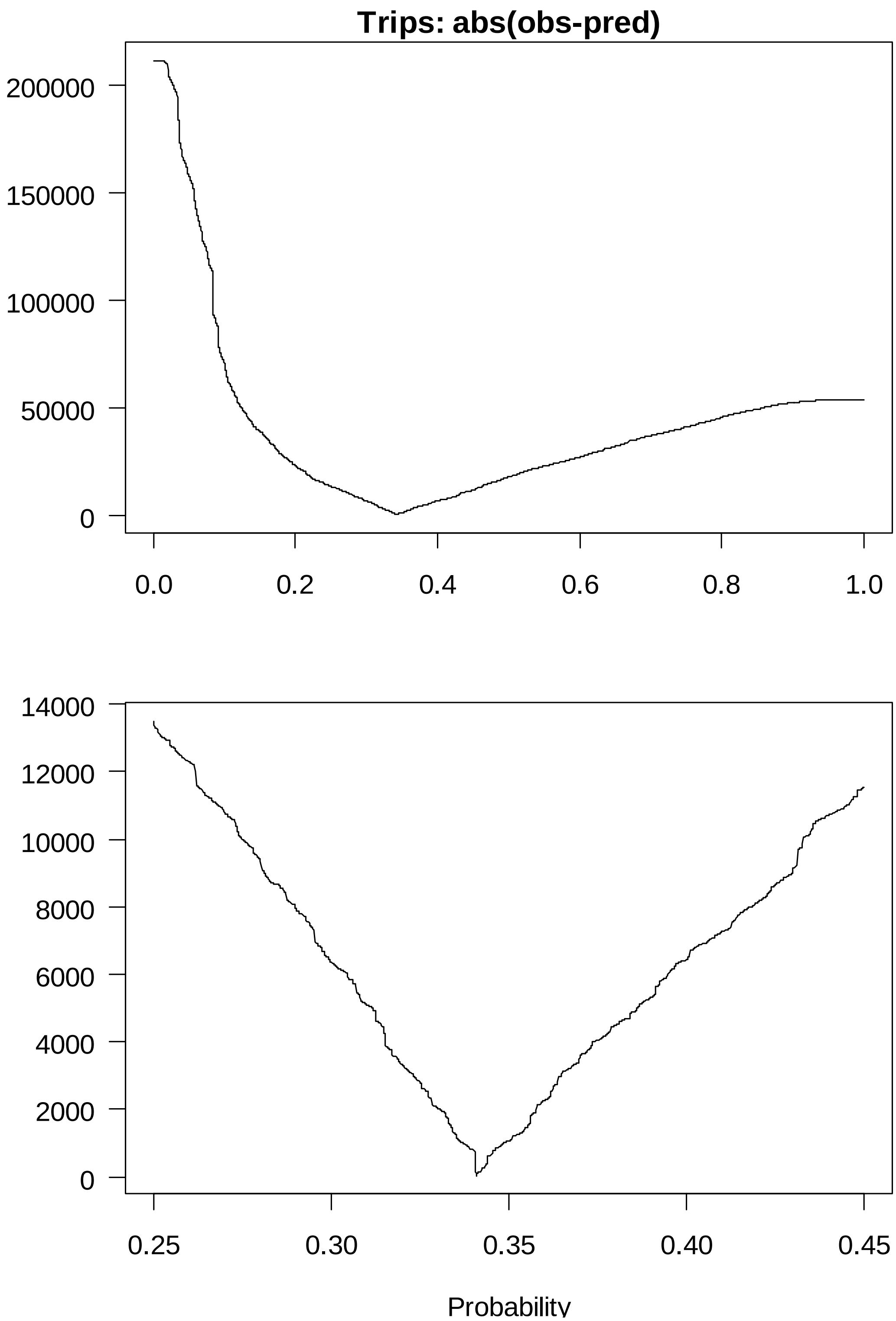


Figure 4. QQ plots of residuals from positive headboat trips. Top panel (A) from a model with effort defined as angler-hours. Bottom panel (B) from a model with effort defined as number of anglers.

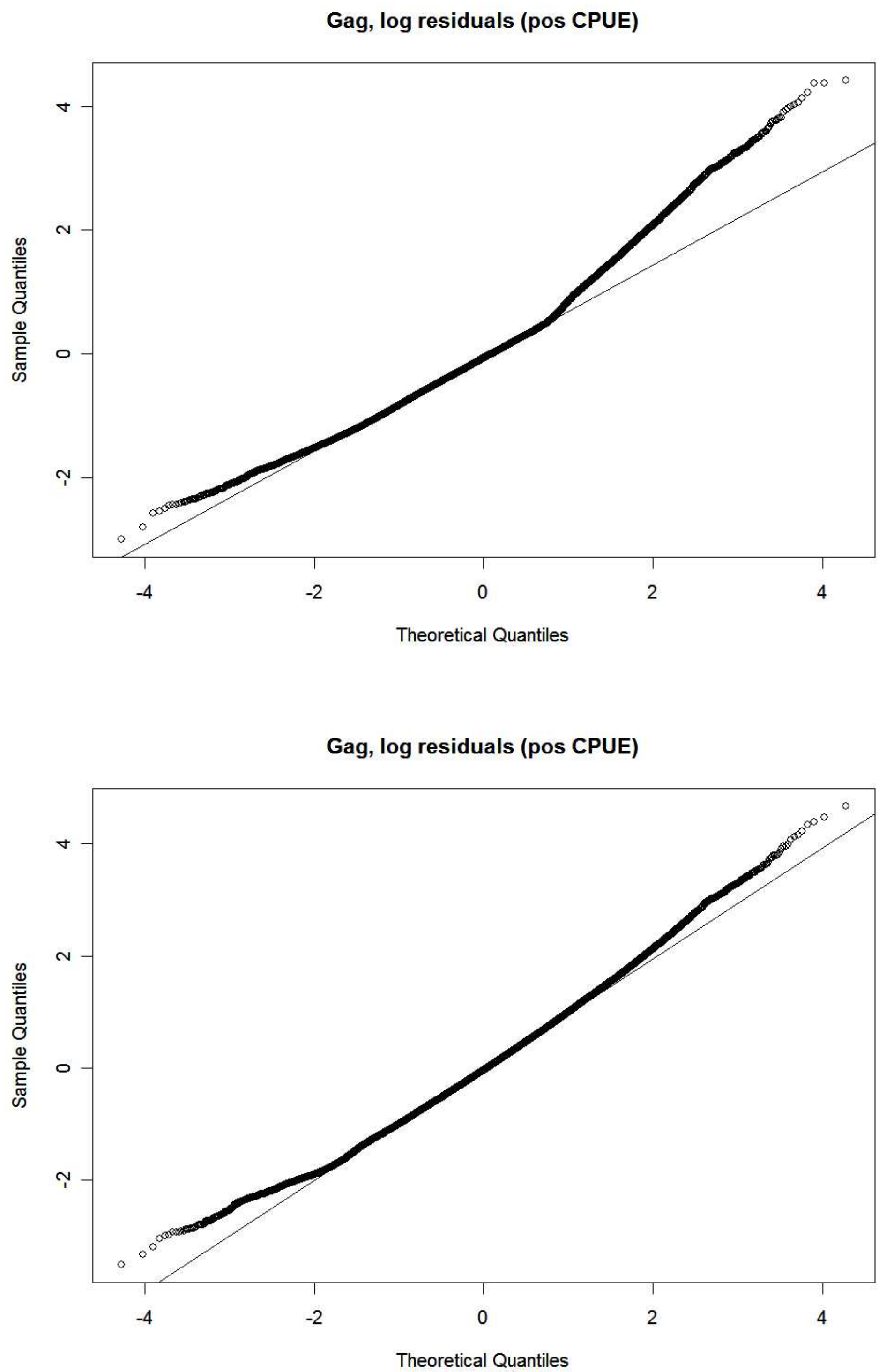
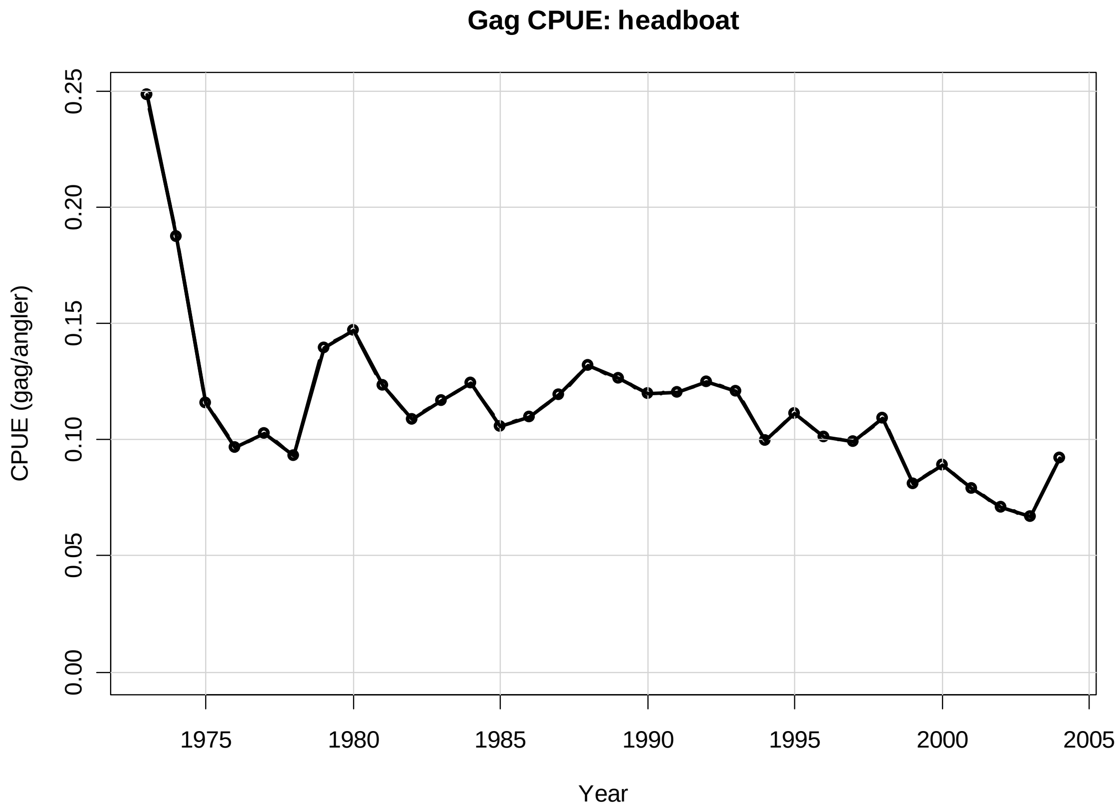


Figure 5. Index of abundance of gag from headboat data collected off the Southeastern U.S.



Appendix

Issues discussed at the DW:

Issue 1: Trip selection using method of Stephens and MacCall (2004)
Option 1: Include all positive trips and use Stephens and MacCall method to identify zero trips only.
Option 2: Include only those trips with associated probability of catching gag above the threshold probability, as in Stephens and MacCall (2004).
Decision: Option 2, to be consistent with the published method and to exclude trips with incidental catches of gag.

Issue 2: Interaction terms in the delta-GLM
Option 1: Include only main effects
Option 2: Investigate interaction terms
Decision: Option 2. Investigate interaction terms. The group decided not to include interactions with year effects, because such effects may be inseparable from changes in abundance.

Issue 3: Include/exclude years prior to full area coverage
Option 1: Exclude early years, because those years did not have full area coverage. All headboats from North Carolina and South Carolina were sampled starting in 1973. Headboats from Georgia and Florida were sampled starting in 1976, but without a complete census. All headboats in all states were sampled starting in 1978.
Option 2: Include early years, unless there is compelling empirical reason not to.
Decision: Option 2. The DW decided to include the early years, because the sampling covered a substantial proportion of the geographic area, and because the GLM accounts for area as a factor. Exploratory data analysis revealed nothing to suggest data in those early years were flawed.

Miscellaneous decisions
- Landings in 2004 from vessel #308 were apparently reported incorrectly. These landings were corrected for computing CPUE, as they were for computing headboat landings.

Updated analyses

A forward stepwise approach was used to construct each GLM (binomial and lognormal). First a GLM was fit on year. These results reflect the distribution of the nominal data. Next, each main effect (area, month, and trip type) was examined for its reduction in deviance per degree of freedom. The factor that caused the greatest reduction was added to the base model if it was significant based on a Chi-Square test ($\chi^2 \leq 0.05$) and if the reduction in deviance was greater than 1%. This model then became the base model. The process was repeated, adding main effects first and then two-way interaction terms, until no factor or interaction met the criteria for inclusion.

The iterative method above requires adequate sample sizes per year per factor, and an approximately balanced design. To achieve these requirements, three quarter day trips were combined with half day trips, as these trips are believed to fish approximately the same locations, and areas were combined as follows:
–Areas 2,3,9,10 (North Carolina)
–Areas 4,5 (South Carolina)
–Areas 6,7,8 (Georgia and northern Florida)
–Areas 11,2,13 (southern Florida)
Area 1 (north of Cape Hatteras) was removed from the analysis, because it contained a small number of headboats operating only in 1973 and 1974.

For the binomial GLM, the forward stepwise approach identified all main effects—area, month, and trip type—for inclusion in the analysis (Table A.1). For the lognormal GLM, it identified all main effects plus the area*trip type interaction for inclusion (Table A.2). Estimates of CPUE and CV are presented in Table A.3 and in Figure A.1. Diagnostics plots are in Figure A.2.

Table A.1 Linear regression statistics for the final GLM model on proportion positive trips.

Source	%RED DEV/DF	CHISQ	Pr>ChiSq
YEAR	NA	1369.25	<0.0001
AREA	3.08	2161.16	<0.0001
TYPE	2.44	1706.11	<0.0001
MONTH	1.24	883.61	<0.0001

Table A2.2 Linear regression statistics for the final GLM model on catch rates of positive trips.

Source	%RED DEV/DF	CHISQ	Pr>ChiSq
YEAR	NA	682.09	<0.0001
AREA*TYPE	7.12	2380.70	<0.0001
AREA	3.52	1154.78	<0.0001
MONTH	2.81	926.43	<0.0001
TYPE	1.45	470.11	<0.0001

Table A.3 Estimated CPUE (number/angler-hr) of gag off the Southeastern U.S., including lower (LCI) and upper (UCI) 95% confidence intervals and CV. Estimates based on data from the headboat fishery.

YEAR	CPUE (number/angler-hr)	Relative CPUE	LCI	UCI	CV
1973	0.027	2.486	1.452	4.256	0.274
1974	0.019	1.762	0.956	3.247	0.313
1975	0.010	0.925	0.397	2.154	0.442
1976	0.007	0.659	0.270	1.609	0.470
1977	0.007	0.678	0.280	1.642	0.465
1978	0.007	0.689	0.335	1.418	0.373
1979	0.011	1.037	0.589	1.826	0.289
1980	0.013	1.198	0.732	1.958	0.250
1981	0.011	1.064	0.607	1.866	0.287
1982	0.011	1.040	0.625	1.733	0.259
1983	0.012	1.150	0.723	1.829	0.235
1984	0.012	1.168	0.718	1.901	0.247
1985	0.011	0.985	0.601	1.613	0.251
1986	0.011	1.006	0.614	1.649	0.251
1987	0.012	1.084	0.690	1.705	0.229
1988	0.013	1.231	0.819	1.850	0.206
1989	0.012	1.166	0.705	1.928	0.256
1990	0.012	1.122	0.682	1.846	0.253
1991	0.012	1.098	0.664	1.818	0.256
1992	0.012	1.143	0.712	1.835	0.240
1993	0.011	1.050	0.615	1.793	0.273
1994	0.009	0.872	0.488	1.560	0.297
1995	0.010	0.914	0.515	1.624	0.293
1996	0.008	0.769	0.380	1.555	0.364
1997	0.009	0.821	0.379	1.780	0.402
1998	0.010	0.977	0.564	1.690	0.280
1999	0.007	0.670	0.320	1.402	0.383
2000	0.008	0.713	0.341	1.487	0.381
2001	0.007	0.658	0.306	1.414	0.397
2002	0.008	0.708	0.333	1.503	0.390
2003	0.006	0.522	0.190	1.429	0.538
2004	0.007	0.637	0.290	1.400	0.409

Figure A.1 Index of abundance for gag off the Southeastern U.S. Estimates based on data from the headboat fishery.

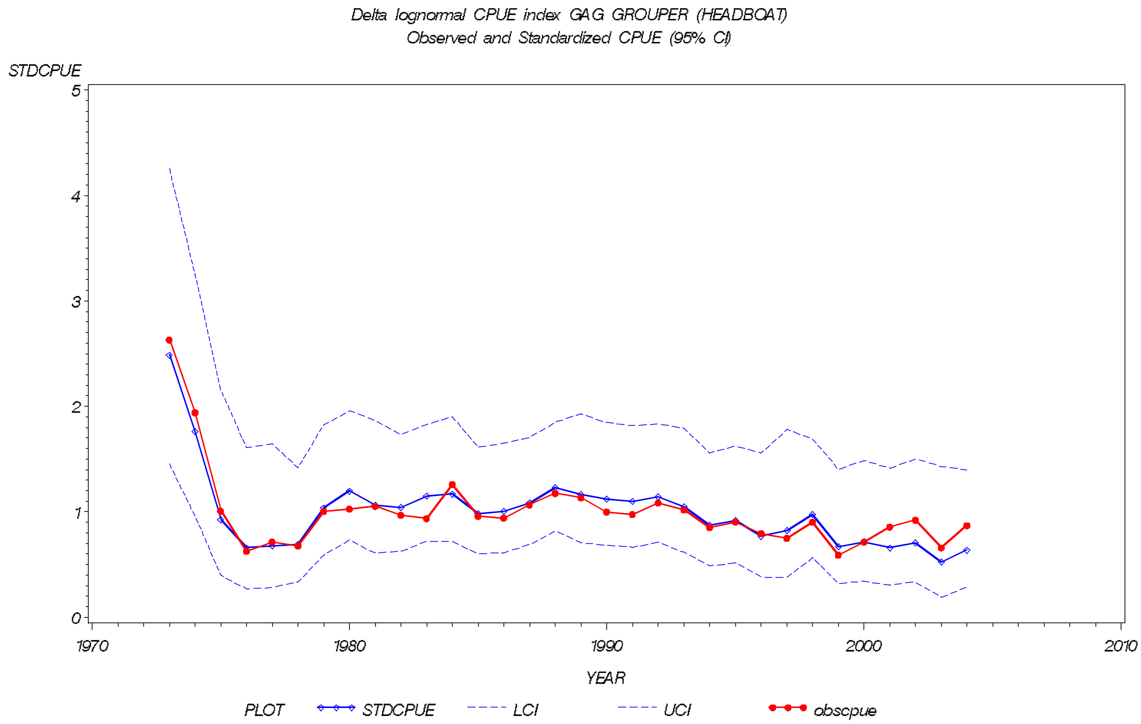
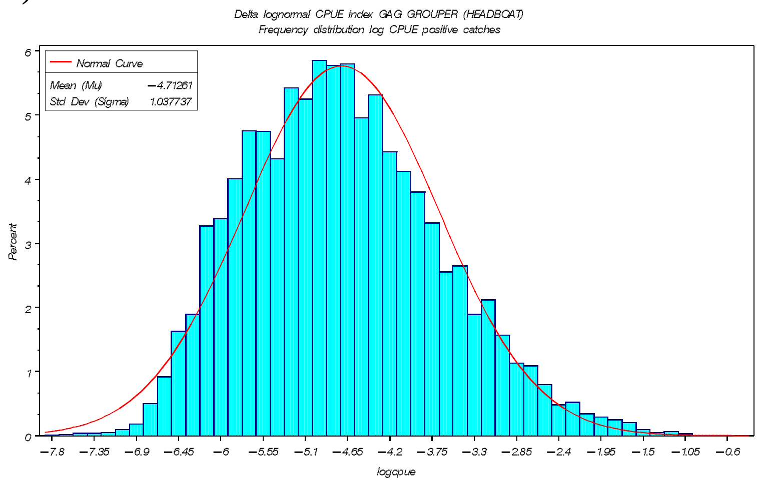
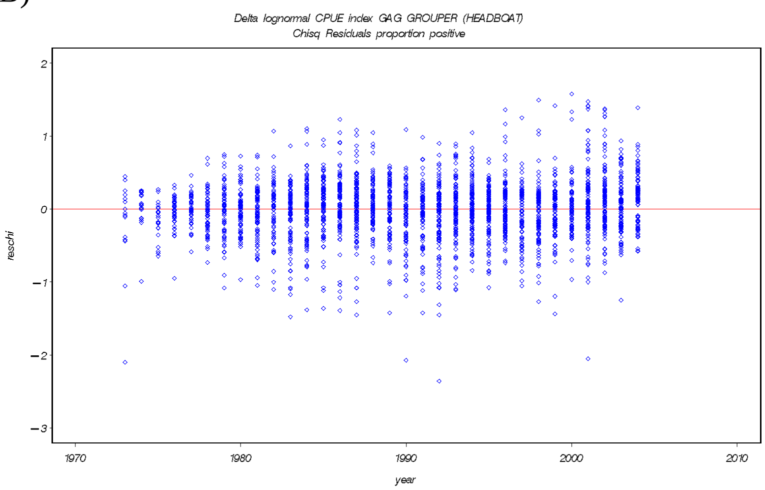


Figure A.2 Diagnostics of model fit.

A)



B)



C)

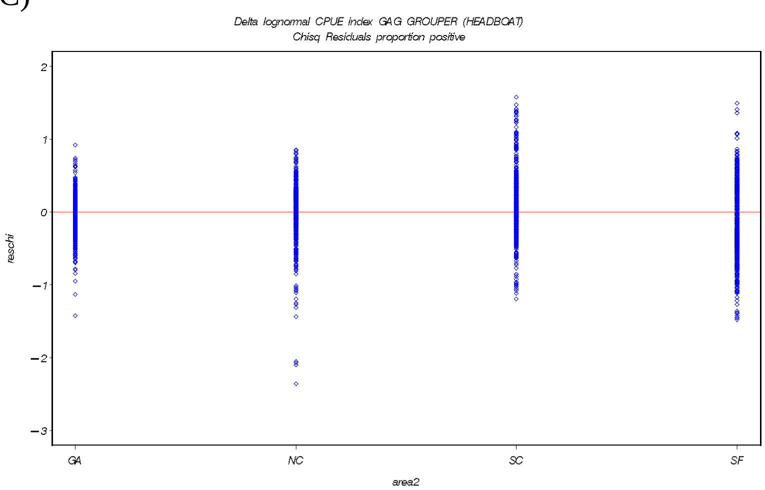
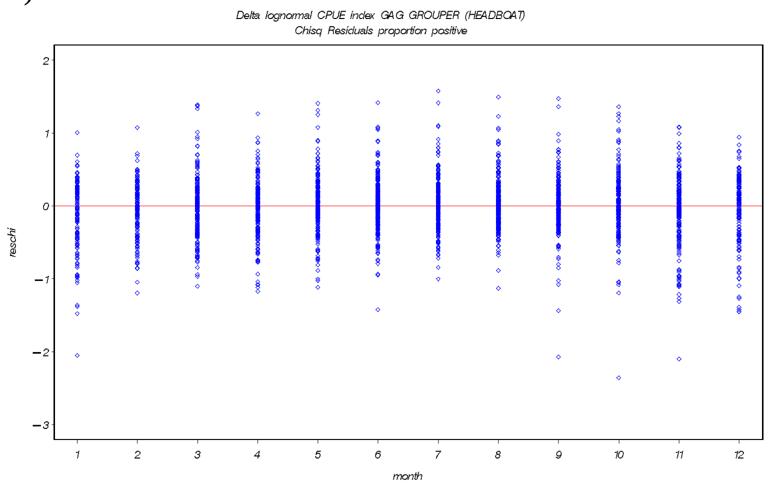
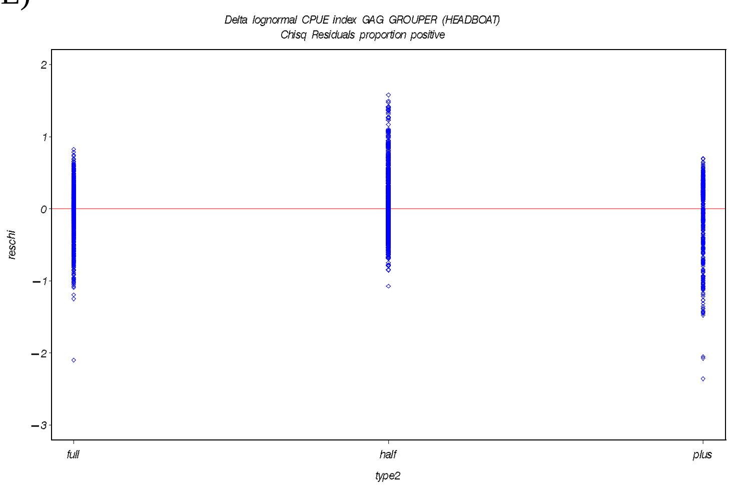


Figure A.2 (cont.)

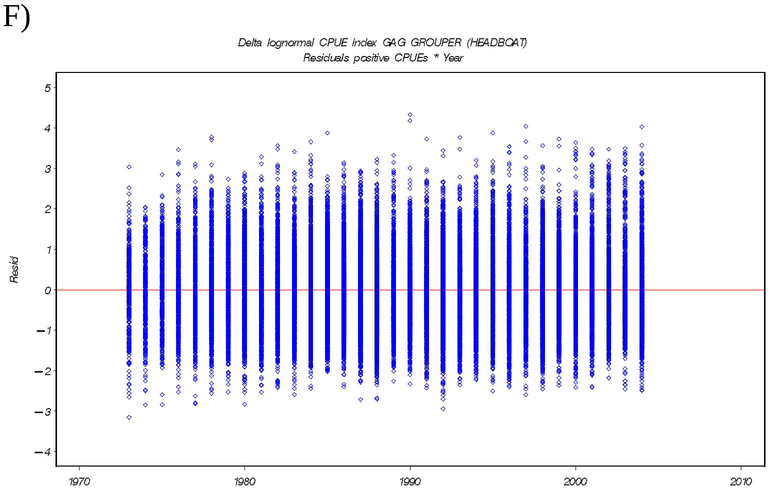
D)



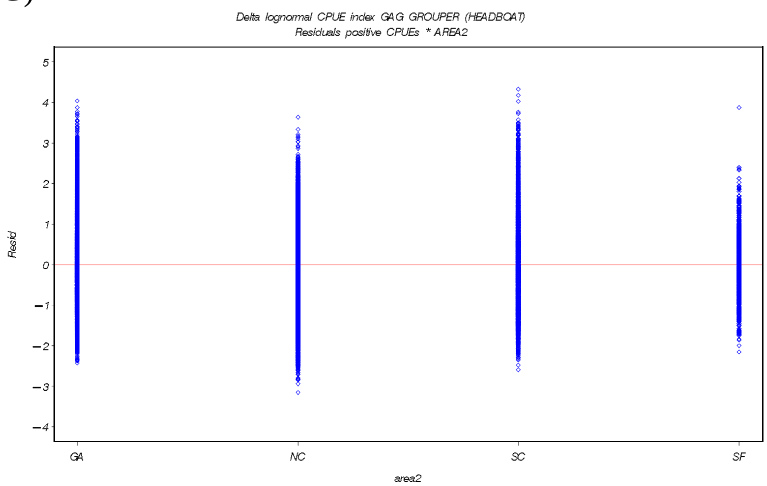
E)



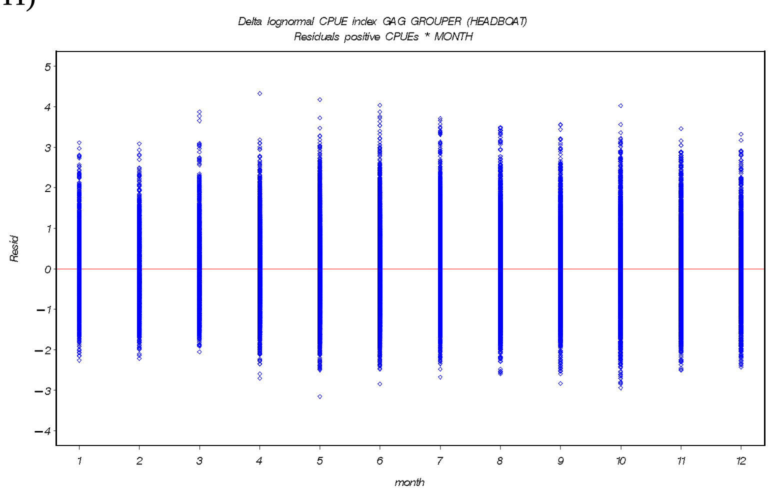
F)



G)


Delta lognormal CPUE index GAG GROUPER (HEADBOAT)
Residuals positive CPUEs * AREA2
Resid
5
4
3
2
1
0
-1
-2
-3
-4
GA
NC
SC
SF
area2

H)


Delta lognormal CPUE index GAG GROUPER (HEADBOAT)
Residuals positive CPUEs * MONTH
Resid
5
4
3
2
1
0
-1
-2
-3
-4
1
2
3
4
5
6
7
8
9
10
11
12
month

I)

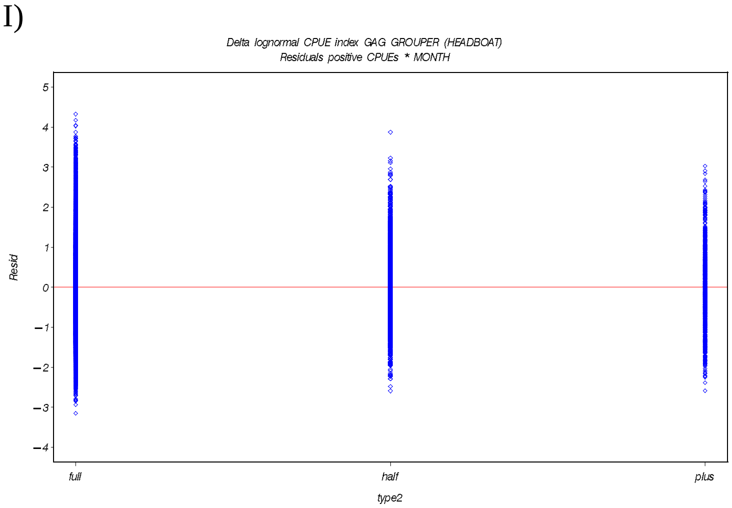

Delta lognormal CPUE index GAG GROUPER (HEADBOAT)
Residuals positive CPUEs * MONTH
Resid
5
4
3
2
1
0
-1
-2
-3
-4
full
half
plus
type2

Figure A.2 (cont)

J)