Construction of a headboat index for south Atlantic red grouper (U.S.)

Paul B. Conn, National Marine Fisheries Service, Southeast Fisheries Science Center, 101 Pivers Island Rd, Beaufort, NC 28516

Background

The headboat fishery is sampled separately from other recreational fisheries, and includes an area ranging from North Carolina to the Florida keys. The headboat fishery comprises large, for-hire vessels that charge a fee per angler and typically accommodate 6–60 passengers. With simple hook & line gear, passengers on these vessels frequently target hard bottom reefs, sampling many members of the snapper-grouper complex. Headboat records were examined in detail, and catch-per-unit-effort (CPUE) standardization was employed in an effort to generate a fishery dependent index.

Possible confounding factors

Prior to analysis, data were examined in an effort to identify possible factors that might confound inferences about relative abundance.

BAG LIMIT CHANGES

For recreational fisheries in the south Atlantic, the relevant regulatory history proceeds as follows:

December 12, 1986 – State of Florida institutes a 5 grouper aggregate bag limit *January 1, 1992* – Federal regulations stipulate a 5 grouper aggregate bag limit *February 2, 1999* – Federal regulations dictate that no more than 2 members of the 5 grouper bag limit may be black or gag grouper

To investigate the potential for these regulations to constrain catch of red grouper, I examined the percentage of headboat trips where anglers caught 5 or more grouper, and contrasted this percentage before and after regulations were implemented. It was assumed that anglers would continue fishing until the boat's limit was reached (i.e., 5 groupers * number of anglers). Included in the list of grouper species were black grouper, gag, speckled hind, snowy grouper, golden tilefish, red grouper, scamp, blueline tilefish, Warsaw grouper, yellowfin grouper, and yellowmouth grouper. The analysis was conducted separately for North Carolina (NC; headboat sampling strata 2, 3, 9, and 10), South Carolina (SC; strata 4 & 5), north Florida & Georgia (NF; strata 6, 7, and 8), and south Florida (SF; headboat strata 11, 12, and 17). For reference, a map of headboat spatial coverage is provided in Figure 1. Headboat sampling strata 1 (northern NC) was censored due to sampling irregularities.

In all years and regions, the number of headboat trips meeting or exceeding 5 groupers per angler was less than 1% (Figure 2). There was no evidence that the proportion of trips meeting the limit decreased after bag limit regulations were implemented. This analysis suggests that CPUE of

grouper (including red grouper) is unlikely to be influenced by bag limit regulations in the south Atlantic.

SIZE LIMITS

The lead analyst suggested keeping the index time series intact in the face of increasing minimum size limits because decreases in CPUE associated with such changes can be accounted for with different selectivity curves. However, headboat length compositions should be examined prior to using headboat indices for simpler models (e.g., surplus production, stock reduction, etc.). In particular, if smaller fish are commonly retained prior to implementation of a size limit, one would expect CPUE to decrease after the size limit is employed. If this is the case, analysts should consider either not using or splitting the headboat index into multiple time series for analysis with simpler models.

Subsetting

Effective effort was based on those trips from areas where red grouper were available to be caught. Without fine-scale geographic information on fishing location, trips to be included in the analysis must be inferred. To do so, the method of Stephens and MacCall (2004) was applied. The method uses multiple logistic regression to estimate a probability for each trip that the focal species was caught, given other species caught on that trip. Because species composition differs markedly between south Florida (strata 11, 12 & 17) and areas north (Figure 1) the method was applied separately to data from regions north and south of Cape Canaveral (Shertzer and Williams 2008). To avoid numerical computation errors, we limited the number of species in each analysis to those occurring in 5% or more of trips (Table 1). I eliminated red porgy from this list because of strict harvest regulations since 1999, which creates the potential for erroneously removing trips likely to have caught red grouper in recent years. A backwards stepwise AIC procedure (Venables and Ripley 1997) was then used to perform further selection among possible species as predictor variables, where the most general model included all listed species as main effects. In this procedure, a generalized linear model with Bernoulli response was used to relate presence/absence of red grouper in headboat trips to presence/absence of other species. For the northern sampling area (NC, SC, NF), stepwise AIC eliminated spottail pinfish, sand perch, little tunny, and black sea bass; for the southern sampling area (SF), it eliminated dolphin and bluestriped grunt. Regression coefficients are presented for the northern (Figure 3) and southern (Figure 4) study areas.

A trip was then included if its associated probability of catching red grouper was higher than a threshold probability (Figure 5 & 6). The threshold was defined to be that which results in the same number of predicted and observed positive trips, as suggested by Stephens and MacCall (2004). After applying Stephens and MacCall (2004) and the constraints described above, the resulting subsetted data set contained 12,598 trips in the northern survey region (9.7% of trips), of which ~11% were positive. For the southern region, subsetted data contained 22,793 trips (13.6% of trips), of which ~21% were positive.

SEDAR 19-DW-03

Explanatory variables

TRIP TYPE – Trips were originally grouped according to whether they were half day, three quarters, full day, or multi-day trips. It was assumed that half day trips fished for 5 hours, three quarters day trips fished for 7 hours, full day trips fished for 9 hours, and multi-day trips fished for 12 hours/day. The proportion of three quarters and multi-day trips were relatively low but constant over time (Figure 7), while there appeared to be an increase in the proportion of full day trips starting in the early 1990s (at the expense of half day trips). Consistent with previous south Atlantic SEDARs (e.g. SEDAR 2008), multi-day trips were combined with full day trips and three quarters day trips were combined with half day trips as factor variables in the standardization process, while the original number of hours was retained for effort determinations.

Based on the subsetted data, there were n=25,251 half/three-quarters day red grouper trips, and n=10,761 full/multi-day trips. Of these, 15.0% of half day trips and 22.2% of full day trips caught at least one red grouper. Nominal CPUE was 0.0036 and 0.0024 red grouper per angler hour, repectively.

NUMBER OF ANGLERS

Based on subsetted data, most trips had fewer than 50 passengers (mean 26.4, median 21). Nominal CPUE appeared to decrease as a function of the number of anglers (Figure 8). As effort was summarized by angler-hours, the number of anglers was not independent of CPUE and thus I was reluctant to include it as a possible predictor variable in standardization. However, if headboat captain's behavior changes (e.g., fishing locations) as a function of the number of anglers (e.g., revenue to buy fuel, etc.), it may be an important variable to consider. As a compromise, I considered 4 discrete categories for the number of anglers as factors in the standardization process, corresponding to quartiles of the distribution of number of anglers. In particular, the following levels were considered: (1) 14 or fewer anglers, (2) 15-26 anglers, (3) 27-34 anglers, and (4) 35-133 anglers.

REGION

The total number of positive red grouper headboat trips by year and region is provided in Table 2. Most positive trips occurred in south Florida (64%), followed by north Florida/Geogia (19%), North Carolina (11%), and South Carolina (6%). When one looks at the subsetted data, North Carolina has the smallest number of selected trips (Figure 9), but many of these are positive. Trend in CPUE and percent of positive observations are only moderately correlated among regions (Table 3), suggesting that some procedure incorporating time by area interactions may be appropriate.

VESSEL

After subsetting data, there were 245 unique vessels that undertook trips likely to have caught red grouper. These ranged from 14 vessels that undertook only one trip to one vessel that

contributed 2,021trips. There was a large range of values among vessels for nominal CPUE and percent of positive trips (Figure 10), part of which is due to the large number of low sample sizes.

YEAR

A summary of the total number of positive red grouper trips is provided in Table 2. Following data subsetting, the number of records ranged from 195 in 1978 to 1793 in 1993. Nominal CPUE and the percent of positive observations largely mirrored each other, showing a minimum in 1991 and a maximum in 2004-2005 (Figure 11). The most recent two years (2007-2008) appear to be below average.

MONTH

The number of trips catching red grouper varied considerably by month and by region (Figure 12). For south Florida, the winter was the period of highest activity, while very few red grouper were caught off North and South Carolina during winter. This suggests the need to include month by area interactions in model structure.

Standardization

Prior to analysis, all vessels with fewer than 10 trips likely to have caught red grouper (that is, fewer than 10 subsetted trip) were eliminated to reduce dimensionality and better standardize effort. For the trips remaining, I modeled CPUE using the delta-glm approach (cf., Lo et al. 1992; Dick 2004; Maunder and Punt 2004). In particular, I compared fits of lognormal, gamma, and inverse-Gaussian models for positive CPUE, and examined which combination of predictor variables best explained CPUE patterns (both for positive CPUE and 0/1 CPUE). All analyses were performed in the R programming language, with much of the code adapted from Dick (2004).

POSITIVE CPUE SUBMODEL

Initial generalized linear model (GLM) fits to positive CPUE data using all predictor variables (but no interactions) indicated extreme lack of fit for large CPUE values (e.g., Figure 13). Successive trials with various transformations of CPUE and parametric families indicated that a model with CPUE^{-1.0} as the dependent variable and gamma as the parametric family better fit these observations and had only a few outliers (Figure 14). With CPUE^{-1.0} as the dependent variable, the gamma distribution highly outperformed lognormal and inverse-Gaussian distributions (Δ AIC>1000).

To determine predictor variables important for predicting positive CPUE, I started by fitting a model with all main effects, together with the interactions (year*area) and (month*area). Stepwise AIC (Venables and Ripley1997) with a backwards selection algorithm was then used to eliminate those that did not improve model fit. All predictor variables were modeled as fixed effects (and as factors rather than continuous variables).

Backwards model selection eliminated only the month by area interaction (Appendix 1); ostensibly the vessel variable contained enough redundancy to eliminate its possible contribution despite its importance if viewed marginally (i.e., Fig. 12). Standard model diagnostics (Figures 14-17) appeared reasonable, and were conducted with randomized quantile residuals (Dunn and Smyth 1996).

BERNOULLI SUBMODEL

The other component of the delta-GLM is a logistic regression model that attempts to explain the probability of either catching or not catching red grouper on a particular trip. The same approach was taken as for the positive GLM; that is, I first fit a model with all main effects, together with the interactions (year*area) and (month*area). Stepwise AIC (Venables and Ripley1997) with a backwards selection algorithm was then used to eliminate those that did not improve model fit. In this case, the stepwise AIC procedure did not remove any predictor variables (Appendix 1). There did not appear to be any recognizable patterns in randomized quantile residuals (Figures 18-20).

Extracting effects and interpreting the index

The presence of time*area interactions in both model components (positive and 0/1) presents a dilemma for interpretation of the index. For instance, the standardized CPUE trend will differ depending on what area is used to reference predictions; if a combination of areas is used, CPUE may be sensitive to the choice of weights. The SEDAR procedural workshop on indices (fall 2008; report yet unreleased) considered several alternatives, including (a) ignoring interactions time*area interactions (omitting them from model structure), (b) modeling them as random effects, and (c) supplying a weighting scheme based on knowledge of habitat availability in different areas or expert knowledge on historical (i.e., virgin) spatial distribution. Due to the software limitations in the face of a gamma error structure, alternative (b) is not possible in this case.

Two possible methods were considered in this working paper. First, time*area interactions were eliminated from model structure. Second, time*area interactions were retained and a spatial weighting scheme was employed. Two possible spatial weighting were considered: (i) equal weights (giving south Florida, north Florida/Georgia, and north/south Carolina equal weights), and (ii) giving NC/SC a weight of 0.5, with NF/GA and SF both getting weights of 0.25 (commercial landings are typically greater for NC/SC). Ultimately, these weights should reflect the percent of virgin biomass in each of the spatial areas. In both cases, population marginal means (Searle et al. 1980) were used for remaining variables. A bias correction was also employed to account for the transformation of the response variable from (CPUE^{-1.0}) to CPUE space as follows: (a) a large number of prediction in (CPUE^{-1.0}) space were generated for year y with a gamma error structure and a dispersion parameter obtained from the GLM, (b) these predictions were back-transformed into (CPUE) space, and (c) the mean value of the back transformed predictions was taken to be the index value for year y. These machinations correctly account for computation of the expected value of the prediction in CPUE space, where simple back-transformations of the mean trend in (CPUE^{-1.0}) space would not.

For the first method (interactions excluded), jackknife estimates of variance were available using the 'leave one out' estimator (Dick 2004), although requisite computing time was high (\approx 3 days). Computation time was even higher for the second approach (including interactions) and so I did not attempt to estimate accompanying variances.

Results

All indices are presented in Table 4, and visually in Figure 21.

LITERATURE CITED

- Dick, E.J. 2004. Beyond 'lognormal versus gamma': discrimination among error distributions for generalized linear models. Fish. Res. 70:351-366.
- Dunn, K. P. and Smyth, G. K. 1996. Randomized quantile residuals. J Comp. Graph. Stat. 5:1-10.
- Lo, N.C., Jacobson, L.D., 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.
- Maunder, M.N., Punt, A.E. 2004. Standardizing catch and effort data: a review of recent approaches. Fish. Res. 70:141-159.
- Searle, S.R., Speed, F.M., and G. A. Milliken. 1980. Population marginal means in the linear moel: An alternative to least squares means. Am. Stat. 34:216-221.
- Shertzer, K.W. and E.H. Williams. 2008. Fish assemblages and indicator species: reef fishes off the southeastern United States. Fish. Bull. 106:257-269.
- South Atlantic Fishery Management Council (SAFMC) 2007. Fishing Regulations for U.S. South Atlantic Federal Waters. Available from <u>www.safmc.net</u>.
- Stephens, A., and A. MacCall. 2004. A multispecies approach to subsetting logbook data for purposes of estimating CPUE. Fish. Res. 70:299-310.

Venables, W. N. and B. D. Ripley. 1997. Modern Applied Statistics with S-Plus, 2nd Edition. Springer-Verlag, New York. Table 1. Species (common name) included in logistic regressions for Stevens and MacCall (2004) method. Species were included if they appeared in the catch records of 5% or more of headboat trips (except for red porgy, which was subject to restrictive regulations in recent years).

North of Cape Caneveral

- Almaco jack
- Bank sea bass
- Black sea bass
- Cobia
- Dolphin
- Gag
- Gray snapper
- Gray triggerfish
- Greater amberjack
- Great barracuda
- King mackerel
- Knobbed porgy
- Lane snapper
- Little tunny
- Pinfish
- Pigfish
- Red snapper
- Sand perch
- Scamp
- Sharpnose shark
- Spottail pinfish
- Tomtate
- Vermilion snapper
- White grunt
- Whitebone porgy

South of Cape Caneveral

- Bigeye
- Black grouper
- Blue runner
- Bluestriped grunt
- Dolphin
- Gag
- Gray snapper
- Gray triggerfish
- Great barracuda
- Jolthead porgy
- King mackerel
- Knobbed porgy
- Lane snapper
- Little tunny
- Mutton snapper
- Sand tilefish
- Tomtate
- Vermilion snapper
- Yellowtail snapper
- White grunt

Table 2. Number of headboat trips reporting red grouper catch by year in the South Atlantic. Zones are south Florida (SF; headboat areas 11, 12, and 17), north Florida & Georgia (NF; areas 6, 7, and 8), South Carolina (SC; areas 4 & 5), and North Carolina (NC; areas 2, 3, 9, 10). Headboat sampling strata 1 (northern NC) was censored due to sampling irregularities.

Table of year by zone						
year						
Frequency	NC	NF	SC	SF	Total	
1978	97	396	10	83	586	
1979	142	288	15	685	1130	
1980	46	256	20	834	1156	
1981	51	128	14	1067	1260	
1982	89	105	10	804	1008	
1983	74	113	21	987	1195	
1984	28	125	3	930	1086	
1985	20	191	3	832	1046	
1986	17	169	4	1062	1252	
1987	31	135	23	1134	1323	
1988	58	502	70	632	1262	
1989	18	283	32	561	894	
1990	47	279	21	423	770	
1991	120	179	28	290	617	
1992	151	270	37	868	1326	
1993	119	194	37	1122	1472	
1994	211	162	40	927	1340	
1995	209	203	58	1011	1481	
1996	209	168	63	1134	1574	
1997	150	225	49	810	1234	
1998	339	536	210	1242	2327	
1999	297	394	234	617	1542	
2000	235	282	164	601	1282	
2001	181	271	103	596	1151	
2002	129	231	62	458	880	
2003	182	103	65	444	794	
2004	183	143	104	598	1028	
2005	161	310	79	851	1401	
2006	148	234	176	444	1002	
2007	118	166	203	420	907	
2008	119	70	107	398	694	
Total	3979	7111	2065	22865	36020	

Table 3. Pairwise Pearson correlation coefficients for time series of nominal CPUE and percentage of positive observations by region (NC: North Carolina; SC: South Carolina; NF: Northern Florida Georgia; SF: Southern Florida). Results are based on subsetted data (e.g., after method of Stephens and MacCall (2004) was used to select trips). An asterisk denotes a significant pairwise positive Pearson correlation at the α =0.05 level.

A. Nominal CPUE					
	NC	SC	NF	SF	
NC	1				
SC	0.72*	1			
NF	0.32*	0.34*	1		
SF	0.37*	0.27	0.25	1	
B. % Positive					
NC	1				
SC	0.46*	1			
NF	-0.22	-0.07	1		
SF	0.47*	0.45*	0.03	1	

Table 4. Standardized CPUE indices for red grouper from the headboat data. A CV was
calculated for the 'no interaction' index, but was not available for the two approaches not
accounting for interactions. The 'Equal wt' scenario gave equal weights to all spatial areas,
while 'NC High' gave more weight to the Carolinas (0.5) than north or south Florida (0.25 each).

	CPUE Index				
Year	No	CV	Equal wt	NC High	Nominal
	Interaction		1	e	
1978	2.04	19.7	1.92	1.54	0.81
1979	2.49	18.1	2.59	2.15	1.17
1980	1.45	20.3	1.05	0.71	0.77
1981	1.00	20.9	1.01	0.98	0.63
1982	0.89	21.7	0.67	0.59	1.02
1983	1.41	20.3	1.02	1.08	0.78
1984	1.38	20.3	0.69	0.58	0.93
1985	1.24	20.7	0.70	0.48	1.33
1986	1.06	21.1	0.62	0.47	0.90
1987	1.24	20.7	0.84	0.68	1.23
1988	0.69	21.9	0.68	0.46	0.77
1989	0.77	21.5	0.81	0.68	0.49
1990	0.46	22.8	0.65	0.67	0.24
1991	0.29	23.4	0.38	0.39	0.19
1992	0.45	22.5	0.43	0.37	0.60
1993	0.58	21.9	0.49	0.44	0.81
1994	0.59	21.8	0.57	0.58	1.03
1995	0.66	21.7	0.70	0.64	1.07
1996	0.89	21.2	0.67	0.67	1.50
1997	0.79	21.2	1.06	1.23	1.34
1998	1.30	20.0	1.86	2.19	1.40
1999	1.12	20.7	1.55	1.97	1.07
2000	0.63	21.9	0.90	1.09	1.19
2001	0.80	21.3	1.01	1.09	1.34
2002	0.66	21.8	0.74	0.81	0.68
2003	0.67	21.7	0.64	0.76	0.93
2004	1.36	20.2	1.45	1.48	2.53
2005	2.53	18.5	2.70	2.82	2.42
2006	0.70	21.7	1.05	1.25	0.72
2007	0.49	22.5	0.94	1.26	0.59
2008	0.34	23.1	0.59	0.89	0.52





Figure 2. Proportion of headboat trips resulting in five or more groupers per angler, by year and area. Definitions of areas are provided in the text.



Figure 3. Estimates of species-specific regression coefficients from Stephens and MacCall method applied to headboat data from areas in the northern region (excludes areas 11, 12, and 17), as used to estimate each trip's probability of catching the focal species.



Figure 4. Estimates of species-specific regression coefficients from Stephens and MacCall method applied to headboat data from areas in the southern region (includes areas 11, 12, and 17), as used to estimate each trip's probability of catching the focal species.



Regression coefficient

Figure 5. Absolute difference between observed and predicted number of positive trips from Stephens and MacCall method applied to headboat data from the northern region (excludes areas 11, 12, and 17). Left and right panels differ only in the range of probabilities shown.



Figure 6. Absolute difference between observed and predicted number of positive trips from Stephens and MacCall method applied to headboat data from the southern region (includes areas 11, 12, and 17). Left and right panels differ only in the range of probabilities shown.





Figure 7. Total number of headboat trips in the south Atlantic by trip type over time.

Year

Figure 8. Number of anglers per trip and nominal CPUE for headboat trips likely to have caught red grouper.





anglers





Figure 10. Violin plots giving the distribution of sample sizes (# trips), nominal CPUE, and % of positive observations for different vessels. The violin plot combines a traditional box-whisker plot with a kernel density estimate (essentially a smoothed histogram). Results are only summarized for trips selected as being likely to have caught red grouper using the method of Stephens and MacCall (2004).









Figure 12. Number of headboat trips catching red grouper by month and region.

A. zone = NC

B. zone = SC

SEDAR 19-DW-03

Figure 13. Standard diagnostic plots of predicted values versus residuals as well as a quantilequantile plot for a delta-lognormal model fit to positive red grouper CPUE data. A decreasing trend in residuals combined with deviation from the 1:1 quantile line indicate that larger CPUE values are being consistently underfit.



Figure 14. Standard diagnostics for the gamma GLM for positive red grouper headboat CPUE data. Residuals appear to have zero mean across predicted values; residual variance is also roughly constant across the range of fitted values but is difficult to tell visually because of differences in sample sizes.



Figure 15. A plot of randomized quantile residuals by predictor variable in the selected model for positive red grouper CPUE from the headboat survey. Box-and-whisker plots give median values instead of mean values.



4

NF

SC

area

SF





Figure 17. Observed (histogram) versus predicted (kernel density estimate; dark line) CPUE for the headboat survey for positive trips from the selected gamma GLM model. The GLM model was fit to 1/CPUE so this fit is also supplied for reference.



Figure 18. Standard diagnostics for the logistic regression model for red grouper headboat 0/1 CPUE data. Randomized quantile residuals appear to have zero mean across predicted values.



Fitted vs. Resid.

Normal Q-Q Plot

Figure 19. A plot of randomized quantile residuals by predictor variable in the selected model for 0/1 red grouper CPUE from the headboat survey. Box-and-whisker plots give median values instead of mean values.





Figure 20. A plot of randomized quantile residuals by vessel for the red grouper 0/1 CPUE model from the headboat survey.

Figure 21. A plot of potential headboat indices. Black circles and error bars represent values from the standardized indices, while gray dash-dot lines represent nominal CPUE. The top panel gives results for the approach not including interactions, while the middle panel gives results for an equal weighting scheme for each spatial area (NC/SC, GA/Northern FL, and Southern FL), and the bottom panel gives results when weighting NC/SC higher than the other spatial areas.



31

Appendix 1. Model selection results for delta-gamma model

A. Positive CPUE model

```
Start: AIC=66756.73
cpue.trans ~ pos.year + pos.type + pos.anglers + pos.vessel +
        pos.month + pos.area + pos.month * pos.area + pos.area *
        pos.year
                                             Df Deviance AIC
- pos.month:pos.area 19 1999 66747
                                                            1990 66757
<none>

      - pos.year:pos.area
      58
      2056
      66861

      - pos.type
      1
      2028
      66881

      - pos.vessel
      134
      2325
      67610

      - pos.anglers
      3
      2678
      69052

Step: AIC=66746.35
cpue.trans ~ pos.year + pos.type + pos.anglers + pos.vessel +
        pos.month + pos.area + pos.year:pos.area
                                           Df Deviance AIC
                                                        1999 66746
<none>

      - pos.month
      11
      2014
      66773

      - pos.year:pos.area
      58
      2066
      66855

      - pos.type
      1
      2036
      66868

      - pos.vessel
      134
      2334
      67597

      - pos.anglers
      3
      2690
      69044
```

B. Bernoulli CPUE model

		Df	Deviance	AIC
<1	ione>		24112	24712
-	bin.anglers	3	24154	24748
-	<pre>bin.month:bin.area</pre>	19	24188	24750
-	bin.year:bin.area	58	24426	24910
-	bin.type	1	24404	25002
-	bin.vessel	175	28947	29197
>				