Independent peer reviewer's report on the SEDAR 47 Benchmark stock assessment for Goliath Grouper

Desmond M. Kahn

Prepared for the Center for Independent Experts (CIE)

Table of Contents

Table of Figures

Executive Summary

The SEDAR 47 review took place in St. Petersburg, Florida, from May 17-19, 2016 to evaluate the latest stock assessment of Goliath grouper fishery. The assessment team presented data from the Goliath grouper fishery in Florida, including commercial landings prior to a moratorium imposed in 1990 through the present. Commercial landings were subjected to a major adjustment for almost 20 years, which creates uncertainty. Recreational landings estimates are available from 1981 through 1989; during the moratorium, estimates of the number of recreational Goliath discards per year are available. Since 1990, the only "catch" is an assumed 5% mortality of recreational discards. These numbers were converted to weight, based on little data, and could have misestimated the biomass killed by recreational discards. There is no age-structure to the catch data, although some research projects have provided estimates of age-structure of both the estuarine nursery area and a smattering of age data from the "offshore" component. Four indices of relative abundance were presented, three based on recreational catch-per-trip, and a fourth based on visual observation by SCUBA divers.

The assessment included two models, a Stochastic Stock Reduction Analysis model and a Catch-Free model; both the models were variants of Age Structured Production Models. The Review Panel decided the model results were not reliable for management purposes. The handling of indices raised questions, including a truncation of sixteen years from two indices due to sparse data; however, elimination of these years deprived the models of information that the stock was at low abundance during that time (sparseness was due to low catches). The indices indicate a surprising spike of exponential growth in the early-mid 2000s followed by an equally abrupt crash, reported to be caused by a red tide and two severe cold snaps which devastated the stock, particularly the younger, immature inshore component. The indices do not yet show recovery from the crash. These incidents mean natural mortality (M) has been extremely non-constant; the models are outfitted with constant M, which means they cannot fit the data just before and during the crash. They appear to misinterpret the increased M as an increase in fishing mortality (F). Likely overestimation of the weight of recreational discards may also contribute to overestimation of fishing mortality for the SSRA. The model results for exploitation were untenable, with the Catch-Free model finding that an assumed 5% mortality rate of recreational discards had exceeded the overfishing threshold. Both models portray increases in biomass, including spawning stock biomass, in the last few years, but at least three of the indices do not support this increase.

Model configurations, including Beverton-Holt stock recruit functions, flat-topped selectivity for the SSRA and constant natural mortality, should be revisited. Models have to be informed of high M during cold-kill years in order to fit these unusual data. All available years of the recreational catch-per-trip indices should be included in the inputs. Conversion of recreational discards to weight should be revisited for the SSRA. A series of sensitivity runs, assuming these models can be reconfigured, could be helpful. A fruitful undertaking may be to employ some more basic assessment approaches, such as an index-based analysis and surplus production models, possibly including Steele-Henderson modeling to incorporate cold-kill events.

Background

Two previous SEDARs have presented stock assessments of Goliath grouper. SEDAR 3 (2004) assembled data, and SEDAR 6 (2006) presented an assessment based on the Catch-Free model. The assessment concluded that, prior to the 1990 moratorium on landings, there had been overfishing and the stock had been overfished. The assessment also concluded that a significant reduction in fishing mortality had occurred, and the stock could be on a trajectory to recover to the management proxies for Optimal Yield, which were $F50_{50\%SPR}$ and $SSB_{F\ at\ 50\%SPR}$.

SEDAR 23 (2010) revisited the status of Goliath grouper, using indices through 2009 and again employing the Catch-Free model while exploring sensitivity to variation in priors for reduction in F post-moratorium and for natural mortality, and found that some scenarios showed SSB had recovered to the OY proxies, depending on the assumed estimate of longevity.

An update (O'Hop et al. 2015) for the Florida Fish and Wildlife Conservation Commission (FWC) that employed variations in the indices also found evidence for recovery, but in this update, the impact of the cold kills of 2008 and 2010 came to the fore.

On May 17, 18 and 19, 2016, SEDAR 47 convened in St. Petersburg, FL with a Review Panel consisting of representatives of the Scientific and Statistical Committees of the South Atlantic and Gulf of Mexico Fishery Management Councils, including Dr. Carolyn Belcher, Dr. Bob Ellis, and Dr. Mary Chrisman and three representatives of the Center for Independent Expert: Dr. Robin Cook, Dr. Joel Rice, and me. The Panel was chaired by Dr. Marcel Reichert, South Carolina Marine Resources Research Institute. The Assessment team presenting consisted of Joe O'Hop and Joseph Munyandorero, Florida Fish and Wildlife Research Institute. The meeting was assisted by Dr. Julie Neer, SEDAR coordinator for the South Atlantic Fishery Management Council. Dr. Manoj Shivlani and Roberto Koeneke coordinated the travel and housing arrangements.

Description of Individual Reviewer's Role in the Review Activities

The review consisted of four sequential tasks: 1) A reading of the assessment report and some references cited therein; 2) Participation in a three-day Review Panel Meeting in which the assessment was presented in detail and discussed in detail; 3) Participation in development of the Panel's Summary Report; and 4) Completion of this individual report.

The assessment report, by Joe o'Hop and Joseph Munyandorero of the Florida Fish and Wildlife Commission's Fish and Wildlife Research Institute, was supplied to the Review Panel members two weeks prior to the meeting. Upon studying the report, I developed some questions and contacted the assessment team. They were very cooperative and were able to supply me with answers in a very helpful way.

On the 17th, the assessment team presented the assessment. One section focused extensively on the assessment data, another section presented the Stochastic Stock Reduction Analysis (SSRA) and a third section presented the Catch-Free (CF) model. After extended discussion, the panel requested some additional model runs to be presented on the 18th. On that day, these new runs were presented and discussed. On the afternoon of the 18th, the Panel began writing its Summary Report. Discussion of various issues perceived in the assessment continued that afternoon and into the 19th, when work continued on the Summary Report.

Following the meeting, I continued to contribute to the Summary report and wrote my individual report.

Review of the Terms of Reference

TOR 1 - Evaluate the data used in the assessment, including discussion of the strength and weakness of data sources and decisions, and consider the following:

a). Are data decisions made by the data providers and assessment analysts sound and robust?

Commercial landings

The assessment gave a helpful summary of early literature reports of commercial landings as far back as Jordan (1884). Some reports indicated catches of vessels from the Keys were landed in Havana. That pattern of foreign landings of Goliaths caught in U.S. waters creates a serious problem for developing reliable landings records. The assessment cites numerous sources substantiating a significant commercial fishery from at least the late nineteenth century into the early decades of the twentieth century. One report stated the fish were usually 100-250 pounds and up to 400-500 pounds. This information shows clearly that the stock was not virgin in 1950, but had been subjected to almost a century of fishing, including significant landings.

Although landings records include landings in other states, the assessment restricted landings to Florida records. There are significant questions, as reported in the assessment, as to the extent of unit stocks. Are Florida landings all from one stock? Do east and west coast Goliaths belong to separate stocks? Are fish landed in Alabama from the same stock as those landed in Florida? These questions are under study currently and clear answers are not yet available. The analysts' approach to focus the current assessment on Florida data is sound.

The first annual federal estimate of Goliath landings was produced in 1918. Figure 3.3.1 is a chart of Florida landings reports from 1915 through 2010, and these estimates were sporadic and erratic through the 1940s, varying by a factor of four in magnitude, ranging up to 475,000 pounds. There are two periods where federal landings were supplemented by landings estimates collected by the state of Florida. The second of these was actually a correction for suspected over-reporting by one particular dealer. The correction was significant, dropping landings by almost 50% for almost a twenty-year period from about 1965 through 1984. The rationale for the correction is sensible, and it seems advisable to make a correction, but the resulting estimate has high uncertainty, as the report states. It would be advisable to analyze the impact of this correction with at least one sensitivity run.

Logbooks for commercial vertical line and longline vessels report Goliath bycatch on up to 14% of trips during the apparent stock build-up in the early 2000s, but the percentage declined to just a few per cent by 2012, as the stock appeared to decline, based on the indices of relative abundance. At-sea observer reports indicated that from 7.5% to 11% of longline trips caught Goliaths from 2006-2015. These reports raise the possibility that commercial bycatch could be a factor affecting stock dynamics, but the assessment did not include commercial bycatch. The frequency of such commercial trips would determine the potential impact of commercial bycatch. Without the ability to devote some time to this question, it is reasonable to leave such bycatch out of the removals at present, but a data workshop in future could take an in-depth look at the issue.

Recreational landings and catch

The recreational catch estimates were from the Marine Recreational Fishery Statistics Survey (MRFSS) / Marine Recreational Information Program (MRIP) program conducted by the National Marine Fisheries Service (NMFS). This is the only game in town for recreational catch data. Since the moratorium was imposed in 1990, there have been no legal landings, but the MRIP estimate includes discards. Based on what seems a very rough guesstimate, the dead discards are estimated as 5% of the released fish. That constitutes the "catch" for the moratorium period. However, there are some questions about these data. Figure 3.4.1 presents the total of recreational plus commercial landings. From 1990 through about 2015, the "landings" are the 5% of recreational discards assumed killed, converted to estimated weight in kg. The peak kill of discards in this graph occurs in 2007, when it was about 90,000 kg, or 90 tons. If this is 5% of the fish caught and released, it implies that a total of 1800 mt were caught and released in that year. This number is surprisingly high. Since the MRIP estimated number of recreational discards for that year was 138,000, the average weight in the conversion from numbers discarded to kilograms discarded is 13 kg per fish. The offshore MRIP discards were assumed to weigh 59.4 kg on average, based on a research study, which could be high, especially since the average weight from data collected by MRFSS prior to the moratorium from offshore recreational landings was only 6.6 kg. The assessment team stated that recovery of stock size structure during the moratorium could explain this discrepancy, but the research team was targeting Goliath and surely used heavy tackle capable of catching large specimens. For estuarine catches, the assumed average weight was 5.3 kg, based on data collected as part of the Everglades National Park survey.

The assessment team states that the age frequency distribution from a research catch of 22 fish offshore was used to construct what they term a "vulnerability curve". In the Catch Free (CF) model, such curves are used to interpret the indices, showing, for example, that the estuarine indices apply to younger fish than the offshore MRIP index. Questions were raised by the Panel about use of the term selectivity and how this and other such curves were applied. If they were applied to estimate the catch-at-age of recreational catch, they seem fine. But use of the term vulnerability implies that the curves estimate fishery selectivity at age. If that is how they were used, there may be some question. The Panel commented on this issue at some length.

Indices of relative abundance

Three indices were developed from recreational catch-per-trip (CPT) data (Everglades National Park (ENP) Index, MRIP estuarine CPT and MRIP offshore CPT) and one was developed from a visual survey by volunteer divers (the REEF index). The ENP index extends back to 1974, pre-moratorium, while the assessment employed the two MRIP surveys only from 1997 through 2014.

As of this writing, the summary report stated that the MRFSS/MRIP indices were "highly variable, ad-hoc corrections for over-reporting were conducted and estimates of mean weight per fish estimated by the assessment team varied among indices by more than an order of magnitude." I disagree with these statements. While they may be changed in the final summary report, I would point out that the "ad-hoc corrections for over-reporting" occurred in the commercial landings, not the MRIP survey. I don't find the indices highly variable; indeed they agree in trend with the ENP index and seem quite coherent. The comment about mean weight variation misconstrues the MRIP survey. It is designed to collect data on catch in numbers, not in weight. The survey website explicitly warns users that weight data are not reliable, since interviewers are often not able to obtain weights. However, weight data present in the survey are highly useful at times.

In fact, the MRFSS survey began in 1981 and catch and effort data are available from that year through 2015. The assessment team (AT) subjected the data to generalized linear modeling using the delta distribution, but stated that the data were so sparse when individual trip data were examined, that only data after 1996 were acceptable for their purposes. The reason for sparse data is low abundance, meaning catch rates are low. By truncating the MRFSS data prior to 1997, the assessment team lost the ability to feed the model the signal that abundance was low for those years. The strong contrast between the low abundance from 1981 through 1996 and the peak abundance in the mid-2000s could be highly beneficial to the modeling process.

Part of the cause of sparse data could be the split of the MRFSS/ MRIP data into areas – estuarine vs all other areas ("offshore"). Splitting the data like this will reduce precision in general, although I'm not sure exactly how it would affect the raw data as the assessment team examined it. In order to get a feel for the basic data on this stock, I downloaded the MRIP data on fishing effort (trips) and Goliath grouper catch for Florida via a data query at the MRIP website. I calculated total catch per trip by dividing estimated catch for a year by estimated fishing trips for a year (Figure 1). Later, I was able to construct confidence intervals using the CV (proportional standard error) provide for each parameter by MRIP, as advised by Mary Christman (a statistician on the Review Panel). Clearly, this index for the period 1981-1996 gives a strong signal of low abundance.

An alternative approach to the one the AT employed is to use all the data as total catch per trip. This can be broken down by area, but precision could be reduced. Possibly the ENP index could be retained as a recruitment index and the pooled MRIP index used as a second index. While generalized linear modeling such as the AT employed is a standard approach to employing CPT data, the analysis frequently makes very little difference in the final index. Here, it would be helpful to compare the modeled indices with the nominal catch-per-trip to inform the Review Panel about the difference the modeling made.

Several questions arose about the REEF index, which showed a quite optimistic trend in Goliath abundance, except for a drop at the end, which is in concert with the drop seen in the CPT surveys. One issue is the lack of a randomized structure, leading to the fact that divers choose which reef to visit, which can induce a bias in favor of reefs with higher abundance. A diver index, though, may be valuable because Goliaths will have higher "catchability", since they only need to be seen, whereas a hook-and-line-based index, like the MRFSS-MRIP-ENP, requires Goliath to be brought to the boat. Larger Goliath can more easily break tackle and may never be counted by recreational anglers. Consequently, this index has the potential to present a more accurate trend than CPT indices.

b). Are data uncertainties acknowledged, reported, and within normal or expected levels?

The report does not present the uncertainty around the estimates of recreational catch from MRFSS and MRIP of the National Marine Fisheries Service, which are provided by MRFSS-MRIP in the form of proportional standard errors (pse, coefficients of variation). Although they were not presented in the recreational landings table 3.3.1 in the report, I have downloaded these estimates and, particularly prior to 1994 during the period of low abundance, the pse values exceed 50% particularly prior to 1994, even when catch is pooled over all areas and modes of fishing. When catch estimates are partitioned out by area, as in the assessment, the pse values will be higher. The fact that a confidence interval exists around these recreational catch estimates is unavoidable due to the fact that the values are estimated from a complex sampling design and are not a census, as commercial landings are thought to be. The low catches of Goliaths during the period from 1981 through 1993, at least, are probably the reason that precision is low. This problem is unavoidable, but the pse values should be presented in the report. These higher pse levels are normal for very low catches in MRFSS/MRIP data in my experience. The uncertainty around the recreational catch violates the SSRA assumption that catch is known without error. Implications of this violation for model results are unknown, however.

The second area of uncertainty in recreational landings is the assumption that 5% of the discarded grouper die when released. This estimate, while a reasonable first cut, hopefully will be verified or corrected by a holding study of hook-and-line caught fish.

As discussed above, a third area of uncertainty is conversion of discard numbers to weight. This seems to overestimate kill in biomass to a significant degree, unless knowledgeable people consider it a reasonable proposition that 1800 mt was caught and released in 2006. This overestimate could be due to the average weight of 59.4 kg applied to offshore recreational discards to convert numbers to weight, based on a research study. If these discards in weight are significantly overestimated, as seems likely, it could bias the SSRA results in unknown, but significant ways, possibly including overestimating fishing mortality.

The commercial landings uncertainty was clearly outlined in the report, and seems substantial, exceeding normal limits, due to the expressed need to make a major adjustment to reduce landings by close to 50% over 19 years to correct for suspected over-reporting by one seafood dealer. This uncertainty seems unresolvable, and can be explored by sensitivity runs with trial adjustments.

c) Are data applied properly within the assessment models?

Some comments by the Review Panel, which at this writing were in the summary report, were made that a potential problem existed because the MRIP catch data were also employed in the CPT indices, so the index and the catch could be correlated. I don't see this as a problem. For one thing, the index has the catch standardized by effort. In some assessments, commercial landings are standardized by effort into CPUE estimates that are used to tune the model; this present case is not really different. Since

$$\text{Catch} = \text{effort} * \text{catchability} * N,$$

we would expect catch to be correlated with abundance if effort and catchability are roughly constant.

d) Are input data series reliable and sufficient to support the assessment approach and findings?

<u>Catch/losses data</u>

The management periods affect the availability of catch data in this assessment, primarily through the imposition of the moratorium. Catch data are available prior to the moratorium, including commercial landing data collected primarily by the State of Florida, with recreational catch estimates for 1981-1989 from the Marine Recreational Fishery Statistics Survey (MRFSS) of the National Marine Fisheries Service. However, the use of significant adjustments to the commercial landings for two decades, 1965 through 1984, leaves uncertain what the correct landings were, raising questions as to the reliability of this input. The adjustment seems to have been called for, but the question is whether it was the correct adjustment. Sensitivity runs seem the only way to deal with the effect of this uncertainty. The question is what would be the impact on the assessment results of different adjustments.

Both the uncertainty about the true commercial landings and the uncertainty around the recreational landings mentioned above, strictly speaking, violate the SSRA model assumption that landings are known without error. The effect of this violation is not clear, however. The issue of using landings estimates with a confidence interval as input into a model that assumes landings are known without error should not be considered something that disqualifies a modeling effort, but this violation of assumptions should perhaps be kept in mind in evaluating model results. Often, totally accurate landings are not available, if they ever are.

Indices of relative abundance

The ENP and MRIP indices are reliable, but as mentioned, the MRFSS/MRIP indices should be extended back to 1981. The fact that both the MRIP Estuarine and the ENP indices showed similar trends supports each of them. The REEF index seems valuable, but possibly could or should be modeled or processed differently, due to the nonrandom sampling. Could the survey data be sampled in a random manner? Panel members also thought a different statistical distribution should be employed in processing the data.

TOR 2 – Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account the available data, and considering the following:

At the outset, if modeling is a scientific activity, it must be subject to testing, since models are akin to hypotheses (see "Conjectures and Refutations", K. Popper). If a model cannot be tested, using the available data, then models are not scientific endeavors. I therefore try to obtain basic data to compare to model output, including an index of the trends in abundance, the catch over time, and the trend in fishing mortality. The latter is calculated according to a relative version of the catch equation formulation, F = total catch/ (mean stock size). The relative version is relative F = catch/(mean index of relative abundance). Two highly valuable attributes of relative F are that it has a sound theoretical basis, and that it is totally independent of any need to estimate M, which we generally know little about for any particular year, but on which we condition most of our estimates of fishing mortality. I then use these data sets to evaluate the outputs of stock assessment models. In Fig. 1, I portrayed the trend in abundance in numbers mentioned above from 1981 through 2015; here the estimated total recreational catch, including released fish (Type A + B1 + B2) for a year was divided by the total estimated recreational fishing trips for the year to estimate mean catch per trip for the year. The catch in kilograms (Figure 2) was taken from the assessment report, including the estimated recreational discard mortality. When the catch in kg is divided by the relative abundance value for the year, we see relative F (Figure 3), portraying the trend in total fishing mortality for the year. Note that the plot of relative F shows no increase in the period 2004-2014, indicating that the dramatic changes in "catch" (discard mortality) were exactly in proportion to the changes in the index of relative abundance, which is what we would expect.

a) Are methods scientifically sound and robust?

In theory, yes. The issues are whether the assumptions of the models are violated sufficiently to cause a biased outcome, and whether the data are biased enough to cause a biased outcome. The fact that two different models were employed is certainly a plus, as opposed to the serious limitations of the "one preferred model" approach that is often employed.

Both models, the Stochastic Stock Reduction Analysis (SSRA) originated by Martell et al. (2008) and the Catch-Free (CF) model (Porch et al. 2006), are described as variants on the group of models known as Age-Structured Production Models. However, this category is so broad (Butterworth and Rademeyer 2008) that it verges on being meaningless. These two models have significant differences, but they still differ from some other modeling approaches, such as the simpler approaches of Index-Based Models and Surplus Production or Biomass Dynamic models.

In general, both models have requirements and assumptions that are not always met in the current application, leading to unreliable results, as the Review Panel has agreed. Consequently, as detailed below, although the models themselves are sound and robust, their application to Goliath grouper is neither.

The common features of the two models, as described in the assessment report, include:

- an estimation of stock-recruit functions instead of production parameters such as carrying capacity and intrinsic population growth rate (these two sets of parameters are mathematically related),
- they incorporate the age structure, through inputs and simulations, although they do not require highly age-structured catch data,
- they project the population forward while incorporating more biological complexity compared to simpler surplus production models, such as time lags, fishery selectivity, and several parameters that are age-structured, such as natural mortality, maturity, fecundity and size
- They can be tuned with indices, which can have their own age-selectivity, although the indices can be aggregated. Fishery selectivity patterns "must be specified by the user."

Some differences between the models are, first, that the well-named Catch-Free model does not employ catch, but rather indices and life history parameters, including an index of fishing effort and starting values for fishing mortality. The report states that this model requires "indices of abundance and associated vulnerabilities-at-age", which prompted questions from the Panel. The report states that "fishing mortality rates must be estimated using from (sic) the catchabilities and indices, and finding the best solution for trends in the indices in comparison to the reconstructed population biomass-at-age levels." This approach to estimation of F seems unique in my experience, and I infer that it puts heavy reliance on the model estimate of age-specific M, which is influenced by priors. If the estimated M is biased for certain years, in the case of non-constant M, then the output F will be biased for at least those years, and we shall see below that there is strong evidence supplied to the Panel that M was subject to one or more strong spikes in recent years.

In contrast to the Catch-Free model, the Stochastic Stock Reduction Analysis did require catch input. This model appears to puts heavy reliance on a stock-recruit (SR) function for its principal output of the management parameters of Fmsy and MSY on a stock-recruit function. A standard Beverton-Holt (BH) form was chosen for the SR function. In such a model, the choice of SR function, i.e. BH versus Ricker, may have major implications for the output parameters. The model incorporates a parameter, κ, related to Stock-recruit steepness. Recruitment is described as "a function of spawning stock egg production (Eq. A27) via a stock-recruit model expressed by Eq. A28", which sounds deterministic, yet Figure 6.9.6 shows estimated recruitment values from the model that were obviously influenced by the large spike in the two pre-recruit indices in the mid-2000s. The summary report states that "Recruitment deviations are treated as random effects and characterize relative year class strength". This model can incorporate proportions-at-age for the catch. This model does not estimate M. Annual fishing mortality rate (Ft) values are solved by iteration.

Markov Chain Monte Carlo chains were employed to evaluate uncertainty in parameters. A retrospective analysis was run, but the results indicate that retrospective bias was not an issue with the SSRA. The determination of stock status from the SSRA was based on a combination of the model output and Spawning Potential Ratio calculations. The matrix of F at age calculated by the SSRA was employed, for example.

b) Are assessment models configured properly and used consistent with standard practices?

At the outset of this topic, I would state that "standard practice" does not necessarily imply that a practice is desirable or helpful. The main points of model configuration I will discuss here are the choice of stock-recruit function, the characterization of natural mortality and the aforementioned selectivity configuration.

In 2a) I mentioned that the Catch-Free model employs a Beverton-Holt (BH) SR function, as does the SSRA. While this is the most commonly employed SR function, it has implications that may not be correct for Goliath grouper. This function has some degree of density-dependence into it, in that at larger stock sizes, survival to recruitment of eggs produced declines, such that recruitment fails to keep pace with an increase in fecundity, i.e. spawning stock biomass. However, this density-dependence is limited, in that no matter how large the spawning stock becomes, the function predicts a constant level of recruitment. This causes the projection that biomass can become huge, in terms of an estimation of virgin biomass or carrying capacity. That projection is biologically unrealistic. The science of ecology has produced reports that groupers, predators that they are, are or can be food-limited. The result of applying the BH function can be an inflated estimate of virgin biomass for grouper that may be unsustainable. No more prey, no more additional grouper. The data, and their effects on the recruitment estimates from the SSRA in the stock-recruit plot in Figure 6.9.6, show some features more consistent with a Ricker function, such as highest recruitment in medium values of SSB and some limited decline in recruitment at high SSB. A configuration with a Ricker function, which produces reductions in recruitment at higher SSB, may be more accurate for grouper and would almost certainly produce more moderate, possibly more accurate estimates of virgin biomass, which affects estimated MSY for the SSRA and relative biomass for the Catch-Free model. This in turn would affect the F reference point outputs.

The models are configured with a constant set of age-structured M estimates, although the Catch-Free model estimates this based partly on priors. However, the data and information supplied in the report indicate two or three instances of large spikes in M that had a drastic effect on stock abundance, to judge by the steep drop in CPT in the three CPT indices and also in the REEF index. A photo of a Goliath killed by a severe cold-snap indicates that mortality inshore was documented. In addition to two cold-kill events in 2008 and 2010, a widespread red tide may have caused mortality. How frequent such incidents will be is certainly unclear, but a model algorithm configured with constant M cannot compute these events. It has clearly caused a poor fit to the spike and then the steep drop in the indices in the 2000s and 2010s evidenced in the SSRA in Figs. 6.9.1. Included in the poor fit is the very poor residual pattern in Figure 6.9.2. This inability to fit the data is caused both because the model may find it difficult to fit the astounding and unprecedented spike in recruitment beginning about 2004 (See Figure 1), but also because it cannot fit the ensuing steep decline, which greatly exceeds the low constant M the models are equipped with. Of course, since the SSRA discounts the increase in abundance evident in the data, based on the increased catch-per-trip, it interprets the increase in catch as an increase in F, and in Figure 6.9.4 a), shows that the 5% discard mortality has exceeded the overfishing threshold for several years! That result is hardly tenable, and is due to the underestimation of M during this period, particularly when compared to the trend in relative F (Figure 3 below).

The Catch-Free model does a better job of fitting the indices in Figure 7.5.2. The fit attains 70% of the peak in the two inshore indices, but only about half of the peak in the offshore index and seriously misses the steep offshore drop, seemingly more heavily influenced by the REEF index, which it fits until the sharp drop in the last year of that index. The residuals in Figure 7.5.3 bear these patterns out, except they show a very good pattern of fit to the MRIP estuarine index. Conversely, the residuals for the peak period in the ENP index are poor and the MRIP offshore fit is not close. The Catch-Free estimate of F during the last decade and a half exceeds the threshold, which is an extremely conservative F at 50%SPR, meaning even a catch and release fishery exceeds the F threshold; this finding is probably, as Joe O'Hop stated in the workshop, due to the increased M which the model translates into increased F.

The selectivity assumption for the SSRA model was flat-topped, meaning older fish were assumed to be fully selected by the fishery (Figure 6.3.1). Yet we were informed that larger, older Goliath are abler to break tackle and evade being caught. That means a dome-shaped selectivity would be more accurate. This difference in selectivity can have major implications for model results, since a dome-shaped selectivity means there are more older fish with higher fecundity in the population, and the SSB is higher. These choices should be explored with sensitivity runs, at a minimum.

However, for the Catch-Free model, Figure 7.5.5 portrays cumulative proportions of ages in the catch, as opposed to actual selectivity at age. It is not clear exactly how the selectivity was finally configured, including the fact that the model fit a cumulative selectivity function.

Selectivity of the indices was also fit by the model, but again, the question was raised by the panel whether this is actually selectivity or just the age-frequency of the indices. IF we step back from the fishery operating in the inshore area and look at the whole stock, the estuarine indices can be correctly described as having a selectivity for the younger ages, while the offshore index has a selectivity for somewhat older ages.

c) Are the methods appropriate for the available data?

These two models are not inappropriate, especially the Catch-Free model, which is totally unique to my knowledge as a stock assessment model which does not employ catch data, and was apparently designed specifically for the Goliath grouper case where landings have high uncertainty. However, it should be pointed out that these two models both have fairly high data requirements and also rely heavily on internal simulations of age-structure, employing many other parameter estimates, either as inputs (SSRA) or self-estimated (CF). They must be modified, at least in the input M values for the SSRA, and in some way for the CF model to recognize the episodes of high natural mortality that devastated this stock in recent years. There appear to be still problems with the input data, described above; if these were corrected or else explored via sensitivity analysis, these models could probably do a reasonable job.

That said, I believe it would be highly informative to employ simpler models, including what is not really a model at all, but an index-based approach, as I have looked at in my Figures 1-3. The National Fisheries Toolbox has software for this approach. Surplus production modeling is another promising avenue, especially if extended with a possible Steele-Henderson approach incorporating a categorical covariate for cold-kill events.

In any case, no model that incorporates constant natural mortality will be able to cope effectively with this data set, as described above. Models will have to be modified to incorporate the extreme episodic spikes of natural mortality that have severely affected this stock.

TOR 3 - Evaluate the assessment findings and consider the following:

a) Are abundance, exploitation and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inference?

SSRA

The model's trend in abundance is somewhat similar to the indices of relative abundance, except that the model results greatly underestimate the surprising peak attained in 2007, when the index of total abundance increased about 16 fold above its low average during 1983-1997 (Total CPT, My Figure 1). This underestimation can be clearly seen in Figure 6.9.1 and 6.9.2. Since the model underestimates abundance, it overestimates fishing mortality (Figure 6.9.4 a), where it indicates that the 5% recreational discard mortality exceeds the $F_{50\%SPR}$ reference point, averaging 0.11-0.17 from 2003-2008. Those estimates are not tenable, as I tried to explain above.

The estimate of total biomass (Fig. 6.9.4 c) does not jive with the implied estimate of total recreational discards. When the 5% mortality is converted to weight with an average of 13 kg per fish, the resulting 90,000 kg for 1993 implies a total of 1800 mt were discarded by the recreational fishery in 2007. Yet the model estimates total stock biomass was only about 1,000 mt. This would only add up if almost all the fish in the stock were caught and discarded twice. Yet the model estimates the total stock size in 2007 as about 1,200,000 fish (Figure 6.9.4 b). This means the average weight was only 0.8 kg, which could be true if the overwhelming majority were 0 – 1 year of age. Figure 6.9.4 (d) depicts "vulnerable biomass", which seems not clearly defined, although Table 6.9.3 indicates it is significantly lower than the SSB.

In summary, the estimated exploitation is too high, partly because the estimated stock size seems too low, based on failing to fit the indices at peak values, almost certainly because the model cannot fit the steep decline due to density-independent spikes of M in the cold-kills. Total stock biomass is not consistent with estimated weight of recreational discards, indicating that either the stock weight estimate is too low, or the estimated weight of discards is too high or both, and average stock weight seems too low.

The estimated value of 50%SPR fishing mortality=0.08 is extremely low. We are dealing with a teleost with extremely high fecundity, not a shark with low fecundity. Conversely, the Minimum Stock Size Threshold, at 730,216 kg (presumably this refers to spawning stock biomass), is too high, since the peak recruitment occurred at estimated SSB well below this level (Figure 6.9.6). While model results indicate it has been achieved in recent years (Figure 6.9.4 (e)), the large increase in biomass seen in that Figure is not consistent with the MRIP offshore index, which peaked in 2008, followed by a drop (Figure 4.3.2). The REEF index also dropped sharply in 2014, although that is only one year. Still, there is no support in the indices for the steep climb in SSB in Figure 6.9.4 (e), although granted, the indices are total abundance, not spawning biomass abundance.

The model output or associated analysis also reported results of equilibrium Spawning Potential Ratio and Yield-Per-Recruit models. As discussed above, dynamics of this stock have been strongly non-equilibrium, so these analyses, based on constant natural mortality, can be very misleading.

CF model

Exploitation estimates are untenable for recent years, when the catch is only 5% of recreational discards. Figure 7.5.7 shows F has exceeded the $F_{50\%SPR}$ every year since 1990, as do most values of MCMC samples in Figure 7.5.10. On the positive side, most MCMC samples show SSB has exceeded the biomass threshold in the latter figure. Here, the same questions arise as for the SSRA above. Why does the CF model show SSB is above the threshold when most indices have declined in the most recent years? This question does not automatically indicate that the model SSB estimate is inaccurate, but still, how could the optimistic SSB estimates be correct in light of the data?

b) Is the stock overfished? What information helps you reach this conclusion?

c) Is the stock undergoing overfishing? What information helps you reach this conclusion?

Because the catch since 1990 was only assumed to be 5% of the recreational discards (discard mortality), exploitation must be less than 5%, since presumably not all fish are caught by the recreational fishery. The proportion caught is unknown, but it is most likely less than 50% in a widely dispersed stock such as this. Consequently, exploitation is well below 5%, probably a few percent at most, so F will be in that range, assuming commercial bycatch is negligible. Therefore, since overfishing cannot have occurred during the last 25 years, the stock cannot be overfished.

d) Is there an informative stock-recruit relationship? Is the stock-recruit curve reliable and useful for evaluation of productivity and future stock conditions?

A stock-recruit plot is in Figure 6.9.6. The ascending limb shows an impressive fit. Basically, though, recruitment in other years was flat except for a striking spike in the mid-2000s, which depicts aberrantly high recruitment for about 5 years, all at the same SSB level in the middle of the SSB range. Recruitments at higher SSB are slightly lower. As discussed above, this seems Ricker-like, in that highest recruitment was at moderate SSB, and higher SSB produced lower recruitment. The run of high recruitment in 2003-2006, which appeared in the estuarine indices as depicted, is inexplicable. Presumably, environmental conditions were excellent for larval survival in that set of years, but the only partial explanation is that the stock was high enough for high fecundity and fertility, but not so high that recruitment suffered density-dependent negative feedback; in other words, a Ricker function. As mentioned above, the very high estimates of SSB in 2012-2014 do not seem consistent with the data.

The SR relationship does seem to reliably indicate spawning stock below a certain point, here depicted as about 100,000 kg, produces reduced recruitment. According to model output, this low-recruitment floor was breached from about 1987-1992. The peak recruitment occurred at SSB values depicted here as about 500,000 to 600,000 kg. The question is whether the SSRA is accurately scaling the SSB to the correct values, and whether the peak recruitment occurred because the SSB was at that level, or for other reasons unrelated to the SSB level.

e) **Are the quantitative estimates of the stock status determination for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?**

I am not sure if this question refers to the values estimated for the target reference points, based on SPR50% values, or if it refers to the quantitative estimates of the recent F values and the SSB levels. As discussed above, I find the estimates of fishing mortality seriously overstated from both models, so they are not reliable (see my Figure 3). They are severely biased by the underestimation of natural mortality during the cold-kill events, and apparently by overestimating the weight of the 5% discards assumed to die from being caught.

As mentioned above, the SSB estimates from both models do not jive with the indices data, which make me question the validity of these estimates.

There are other theoretically sound indicators for stock trends and trends in fishing mortality relying on few assumptions and not affected by highly uncertain estimates of natural mortality, based on the indices of relative abundance and total catch. Trends in one or more of said indices are shown in my Figure 1, and the trend in relative F in my Figure 3.

TOR 4 – Evaluate the stock projections, including discussing strength and weaknesses, and consider the following:
a) Are the methods consistent with accepted practices and available data?
b) Are the methods appropriate for the assessment model and outputs?
c) Are the results informative and robust, and useful to support inferences of probable future conditions?
d) Are key uncertainties acknowledged, discussed, and reflected in the projection results?

Stock projections were deemed not reliable by the Panel, since the model outputs were deemed not reliable.

TOR 5 – Consider how uncertainties in the assessment, and their potential consequences, are addressed.
a) Comment on the degree to which methods used to evaluate the uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods.
b) Ensure the implications of uncertainty in technical conclusions are clearly stated.

First, the fact that the team presented two models goes some distance towards addressing model uncertainty. These two models, however, were two variations of age-structured production models, as opposed to two models with major differences in assumptions, etc., so model uncertainty remains as a major source of uncertainty in the assessment.

Sensitivity runs were conducted at the request of the panel, including truncation of the modeled period from 1950-2014 to 1975-2014, and elimination of various indices. At this writing, the summary report says that effects of index elimination lead to ”further changes in the model fit and predictions.” The report goes on to describe these results as “describing good indications of the degree of uncertainty in model results.” I would like to say that I do not expect a model to give unchanged results when indices are omitted, because the model, hopefully, is fitting these indices. I would expect the output to change if important inputs are omitted. There was some important change when the period was truncated, and the cause of this was not clear.

SSRA

The plots of observed indices vs model fits were good practice and quite helpful, as were the residual plots of fits to indices. The use of MCMC mode enabled graphical portrayal of potential uncertainty and was quite helpful to the Panel. The retrospective analysis was quite helpful and showed no retrospective issue.

That said, however, significant uncertainties described above, including especially the realized values of natural mortality, and commercial catch and conversion of numbers of fish discarded by the recreational fishery during the moratorium to weight are not depicted by these methods, nor should we expect them to be. Some of these questions could be addressed by further sensitivity runs, as mentioned above.

CF model

Use of prior distributions for some parameters and use of MCMC chains were good measures to include uncertainty and portray uncertainty in these parameters and model outputs. Phase plots were helpful portrayals of the variability and uncertainty around management parameters.

TOR 6 - Consider the research recommendations provided and make any additional recommendations or prioritizations warranted.

a) Clearly denote research and monitoring that could improve future assessments

The following represent research and monitoring activities in support of future assessments:

Attempt to evaluate accurate values of natural mortality experienced by the stock from cold-kills and red tide events. One possible approach is to use the rate of decline in the indices following these events as a measure of survival. Once values are estimated, attempt to input them in models like SSRA in appropriate years.

One question is inspired by the delay in recovery of recruitment from the last cold-kill in 2010, seen in the ENP and MRIP estuarine indices. SSB, which is described as mostly offshore would not be expected to have declined dramatically as a result of the cold-kills, although this is not certainly known. The offshore and REEF indices had a decline that was delayed after 2010, which could have been caused by reduced recruitment from inshore. This long period of reduced recruitment raises the possibility of a predator pit, where recruits were reduced to the point that increased predation could have kept abundance low. A Google search for predators on Goliath groupers returned the following, among several other replies.

> Before the Atlantic goliath grouper reaches full size it is susceptible to the attack of **barracuda**, king mackerel and **moray eels**, as well as **sandbar sharks** and **hammerhead sharks**. Once it is fully grown, humans and large **sharks** are the Atlantic goliath grouper's only predators (Atlantic goliath grouper videos, photos and facts - Epinephelus itajara...*www.arkive.org/atlantic-**goliath-grouper**/epinephelus-itajara/)*.

Is there some possibility that one or more of the predators listed above have increased in abundance recently? To pursue an ecosystem approach, which is desirable, this question should be pursued. Impacts of predators have recently been modeled in several cases with use of Steele-Henderson (SH) modeling, which is a surplus production model with a covariate for one or more predators.

Explore SH modeling with a categorical variable to include or exclude cold-kill mortality in appropriate years.

Analysis of the extent of commercial discards to determine the level of such discards, also anecdotal reports of diver killing of Goliaths, should be investigated. Education programs could be developed to reduce this source of mortality, if significant.

Of the recommendations listed in the assessment report, some that seem especially relevant to issues discussed in my report include:

- Although costly, a research survey targeting Young of Year or one-year old grouper could shed helpful light on recruitment and survival.
- Age-structure sampling on a research basis.
- Developing a volunteer angler program to obtain lengths/weights of releases; also, headboat/charter measurements could be helpful.

TOR 7 - Consider whether the stock assessment constitutes the best scientific information available using the following criteria as appropriate: relevance, inclusiveness, objectivity, transparency, timeliness, verification, validation, and peer review of fishery management information.

I take this TOR as asking if the assessment report is acceptable as good, sound scientific information suitable for management use as a basis for management action. The assessment report, while informative and a good source of data and information on Goliath grouper status, does not meet the criteria of sound enough to be suitable for management action. The peer review panel was clear about this and I agree with that judgement.

While the assessment is certainly relevant to the Goliath stock, it falls short of inclusiveness; for example, the report documents winterkill in 2008 and 2010, but the modeling approaches did not include these devastating events in any meaningful way, as I discussed above at some length. Instead, the models essentially employed equilibrium approaches in a decidedly non-equilibrium situation. Particularly if winterkill and red tide events continue to occur, a serious effort must be made to model them; failing that, conventional assessment approaches will not be up to the challenge raised by this stock due to lack of inclusiveness of these important events. The assessment also did not include the MRIP/MRFSS catch per trip data prior to 1997, based on concerns about scanty data. This scanty data is what is produced when a stock is at low abundance, and the model was denied the signal of very low abundance from 1981 through the mid-1990s.

Objectivity does not seem to be an issue, but there could have been more transparency in the report, specifically as the Review panel mentioned, in portraying the effect and fit of the modeling of the data from MRIP used in the two indices – estuarine and offshore.

I would have preferred to see some additional attempts at validation and verification of model estimates of management parameters, such as recent fishing mortality levels and recent abundance levels. I discussed an approach to checking model outputs against available data, such as indices, catch and relative F. The model results in estimates of fishing mortality and peak abundance in the mid-2000s were not verified by these comparisons. Some model results such as estimates of carrying capacity and relative stock status may be more difficult to verify or validate, but trends in abundance and fishing mortality can be more readily verified.

In addition, under TOR 6, I suggested some additional research areas that could possibly increase the value and accuracy of the scientific information in a Goliath grouper assessment"

TOR 8 - Provide guidance on key improvements in data or modeling approaches that should be considered when scheduling the next assessment.

<u>Data improvements</u>

- Conduct experimental studies to confirm mortality rate from recreational discards.
- Develop data on the size structure of recreational discards, possibly from a volunteer angler program.
- Employ all eighteen years of the MRFSS/MRIP indices. If necessary, employ raw or nominal catch-per-trip indices to supply models with the earlier years of low abundance.

- Develop data on the age-structure of the offshore stock component. Consider annual monitoring of age-structure of estuarine component to better understand recruitment variation.

Modeling improvements:

- Modify inputs or employ alternative modeling approaches, such as possibly Steele-Henderson modeling to account for the large increases in natural mortality during cold-kill events.
- Conduct index-based analyses, similar to my Figures 1-3. Explore the AIM (An Index-based Model) in the National Fisheries Toolbox.
- Conduct surplus production modeling.
- If important predators of juvenile grouper have shown trends in abundance, explore Steele-Henderson or other models that incorporate predation, particularly as potentially influencing recruitment.
- Employ the Ricker stock-recruit function instead of a Beverton-Holt function.
- Employ dome-shaped fishery selectivity instead of the flat-topped selectivity currently employed in the SSRA.

References

Butterworth, S.D. and Rademeyer, R.A. 2008. Statistical catch-at-age analysis vs. ADAPT_VPA: the case of Gulf of Maine cod. ICES J. Mar. Sci. 65:1717-1732.

Jordan, D.S. 1884. The Fishes of the Florida Keys. In: Bulletin of the U.S. Fish Commission, Vol. IV, for 1884. U.S. Government Printing Office, Washington, D.C.

Popper, K. 1963. Conjectures and Refutations: The Growth of Scientific Knowledge. Routledge. New York.

Porch, C., A.-M. Eklund and G.P. Scott. 2006. A catch-free stock assessment model with application to goliath grouper (*Epinephelus itajara*) off southern Florida. Fish. Bull. 104: 89-101.

Figures

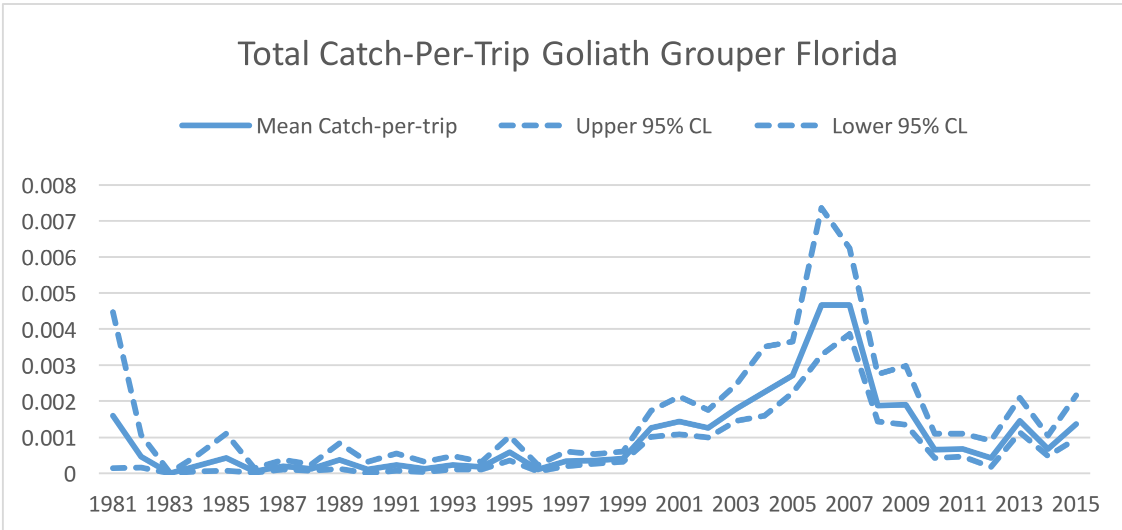

Figure 1: Index of relative abundance based on MRIP total catch of Goliath grouper for Florida and MRIP estimate of total recreational fishing trips in Florida for 1981 through 2015. Personal communication from the National Marine Fisheries Service, Fisheries Statistics Division, May 10, 2016

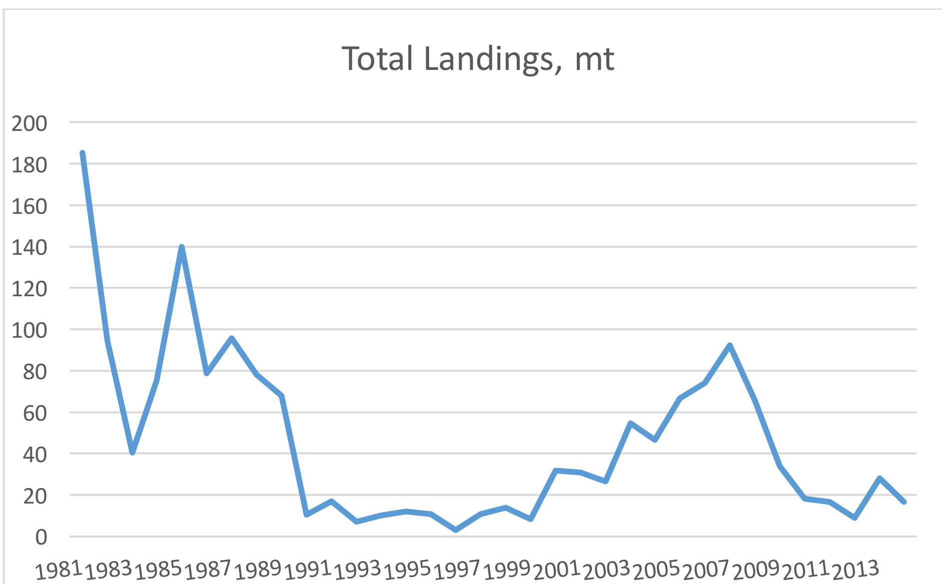

Figure 2: Total landings, Florida Goliath grouper, 1981-2014. Source: Tables 3.1.1 and 3.3.1, O'Hop and Munyandorero. 2016. SEDAR 47 Stock Assessment Report for Goliath Grouper of the South Atlantic and Gulf of Mexico.

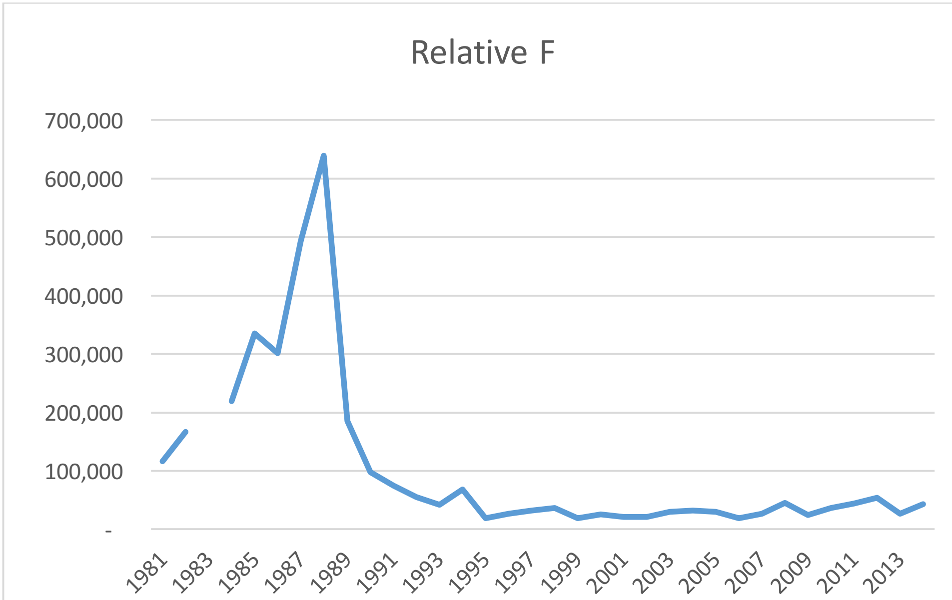

Figure 3: Relative F, Florida Goliath grouper, for the period 1981 through 2014. Value for 1983 deleted to adjust scale. Relative F = total catch/ (index of relative abundance).

Appendix 1. Bibliography of materials provided for review

O'Hop, J. and Munyandorero, J. 2016. SEDAR 47 Stock Assessment Report for Goliath Grouper of the South Atlantic and Gulf of Mexico. Fish and Wildlife Research Institute, St. Petersburg, FL.

Gilmore, R.G., L.H. Bullock and F.H. Berry. 1978. Hypothermal mortality in marine fishes of south-central Florida January 1977. Northeast Gulf Science 2 (2): 77-97.

Appendix 2. Copy of the Statement of Work

Statement of Work
External Independent Peer Review by the Center for Independent Experts
SEDAR 47 Southeastern Goliath Grouper Assessment Review Workshop

Scope of Work and CIE Process: The National Marine Fisheries Service's (NMFS) Office of Science and Technology coordinates and manages a contract providing external expertise through the Center for Independent Experts (CIE) to conduct independent peer reviews of NMFS scientific projects. The Statement of Work (SoW) described herein was established by the NMFS Project Contact and Contracting Officer's Technical Representative (COTR), and reviewed by CIE for compliance with their policy for providing independent expertise that can provide impartial and independent peer review without conflicts of interest. CIE reviewers are selected by the CIE Steering Committee and CIE Coordination Team to conduct the independent peer review of NMFS science in compliance the predetermined Terms of Reference (ToRs) of the peer review. Each CIE reviewer is contracted to deliver an independent peer review report to be approved by the CIE Steering Committee and the report is to be formatted with content requirements as specified in **Annex 1**. This SoW describes the work tasks and deliverables of the CIE reviewer for conducting an independent peer review of the following NMFS project. Further information on the CIE process can be obtained from www.ciereviews.org.

Project Description: SEDAR 47 will be a compilation of data, an assessment of the stock, and CIE assessment review conducted for Southeastern Goliath Grouper. The review workshop provides an independent peer review of SEDAR stock assessments. The term review is applied broadly, as the review panel may request additional analyses, error corrections and sensitivity runs of the assessment models provided by the assessment panel. The review panel is ultimately responsible for ensuring that the best possible assessment is provided through the SEDAR process. The stocks assessed through SEDAR 47 are within the jurisdiction of the South Atlantic and Gulf of Mexico Fisheries Management Council and the states of Florida, Georgia, South Carolina, and North Carolina, Mississippi, Alabama, Louisiana, and Texas. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

Requirements for CIE Reviewers: Three CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have working knowledge expertise in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the primary task of providing peer-review advice in compliance with the workshop Terms of Reference. Experience with data-limited or catch-free assessment methods would be preferred. Each CIE reviewer's duties shall not exceed a maximum of 14 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review during the panel review meeting **scheduled in St. Petersburg, FL during May 17-19, 2016**.

Statement of Tasks: Each CIE reviewers shall complete the following tasks in accordance with the SoW and Schedule of Milestones and Deliverables herein.

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

Foreign National Security Clearance: When CIE reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for CIE reviewers who are non-US citizens. For this reason, the CIE reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/
http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html

Pre-review Background Documents: Two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Panel Review Meeting: Each CIE reviewer shall conduct the independent peer review in accordance with the SoW and ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs can not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COTR and CIE Lead Coordinator.** Each CIE reviewer shall actively participate in a professional and respectful manner as a member of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NMFS Project Contact is responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements).

The NMFS Project Contact is responsible for ensuring that the Chair understands the contractual role of the CIE reviewers as specified herein. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

CIE reviewers shall conduct an impartial and independent peer review of the assessment in accordance with the SoW and ToRs herein.

A description of the SEDAR Review process can be found in the SEDAR Policies and Procedures document:

http://sedarweb.org/docs/page/A6-SEDARPoliciesandProcedures_June2014_0.pdf
The CIE reviewers may contribute to a Summary Report of the Review Workshop produced by the Workshop Panel.

Contract Deliverables - Independent CIE Peer Review Reports: Each CIE reviewer shall complete an independent peer review report in accordance with the SoW. Each CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. Each CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.
Other Tasks – Contribution to Summary Report: Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of the reviewer's views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

Specific Tasks for CIE Reviewers: The following chronological list of tasks shall be completed by each CIE reviewer in a timely manner as specified in the **Schedule of Milestones and Deliverables**.

1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
2) Participate during the panel review meeting **tentatively scheduled in St. Petersburg, FL during May17-19, 2016.**
3) **Tentatively in St. Petersburg, FL during May 17-19, 2016** as specified herein, and conducts an independent peer review in accordance with the ToRs (**Annex 2**).
4) No later than June 9, 2016, each CIE reviewer shall submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Dr. Manoj Shivlani, CIE Lead Coordinator, via email to mshivlani@ntvifederal.com, and Dr. David Sampson, CIE Regional Coordinator, via email to david.sampson@oregonstate.edu. Each CIE report shall be

written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

March 29, 2016	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
April 29, 2016	NMFS Project Contact sends the CIE Reviewers the pre-review documents
May 17-19, 2016	Each reviewer participates and conducts an independent peer review during the panel review meeting
June 9, 2016	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
June 23, 2016	CIE submits CIE independent peer review reports to the COTR
June 30, 2016	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: This 'Time and Materials' task order may require an update or modification due to possible changes to the terms of reference or schedule of milestones resulting from the fishery management decision process of the NOAA Leadership, Fishery Management Council, and Council's SSC advisory committee. A request to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent changes. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on changes. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

Acceptance of Deliverables: Upon review and acceptance of the CIE independent peer review reports by the CIE Lead Coordinator, Regional Coordinator, and Steering Committee, these reports shall be sent to the COTR for final approval as contract deliverables based on compliance with the SoW and ToRs. As specified in the Schedule of Milestones and Deliverables, the CIE shall send via e-mail the contract deliverables (CIE independent peer review reports) to the COTR (William Michaels, via William.Michaels@noaa.gov).

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:

(1) The CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) The CIE report shall address each ToR as specified in **Annex 2**,
(3) The CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

Distribution of Approved Deliverables: Upon acceptance by the COTR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COTR. The COTR will distribute the CIE reports to the NMFS Project Contact and Center Director.

Support Personnel:
Allen Shimada
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
Allen Shimada@noaa.gov Phone: 301-427-8174

Manoj Shivlani, CIE Lead Coordinator
NTVI Communications, Inc.
10600 SW 131st Court, Miami, FL 33186
mshivlani@ntvifederal.com Phone: 305-968-7136

Key Personnel:
NMFS Project Contact:
Julie A Neer
SEDAR Coordinator
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405
(843) 571-4366
julie.neer@safmc.net

Annex 1: Format and Contents of CIE Independent Peer Review Report

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.

2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.

   a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.

   b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.

   c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.

   d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.

   e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review

   Appendix 2: A copy of the CIE Statement of Work

   Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference

SEDAR 47 Southeastern Goliath Grouper Assessment Review Workshop

1. Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions, and consider the following:
    a) Are data decisions made by the data providers and assessment analysts sound and robust?
    b) Are data uncertainties acknowledged, reported, and within normal or expected levels?
    c) Are data applied properly within the assessment model?
    d) Are input data series reliable and sufficient to support the assessment approach and findings?
2. Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account the available data, and considering the following:
    a) Are methods scientifically sound and robust?
    b) Are assessment models configured properly and used consistent with standard practices?
    c) Are the methods appropriate for the available data?
3. Evaluate the assessment findings and consider the following:
    a) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
    b) Is the stock overfished? What information helps you reach this conclusion?
    c) Is the stock undergoing overfishing? What information helps you reach this conclusion?
    d) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
    e) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

4. Evaluate the stock projections, including discussing strengths and weaknesses, and consider the following:
   a) Are the methods consistent with accepted practices and available data?
   b) Are the methods appropriate for the assessment model and outputs?
   c) Are the results informative and robust, and useful to support inferences of probable future conditions?
   d) Are key uncertainties acknowledged, discussed, and reflected in the projection results?
5. Consider how uncertainties in the assessment, and their potential consequences, are addressed.
   - Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods.
   - Ensure that the implications of uncertainty in technical conclusions are clearly stated.
6. Consider the research recommendations provided and make any additional recommendations or prioritizations warranted.
   - Clearly denote research and monitoring that could improve future assessments
7. Consider whether the stock assessment constitutes the best scientific information available using the following criteria as appropriate: relevance, inclusiveness, objectivity, transparency, timeliness, verification, validation, and peer review of fishery management information.
8. Provide guidance on key improvements in data or modeling approaches that should be considered when scheduling the next assessment.
9. CIE Reviews may contribute to Peer Review Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference.

Annex 3: Agenda

SEDAR 47 Southeastern Goliath Grouper Review Workshop

Saint Petersburg, Florida
17-19 May 2016

Tuesday

9:00 a.m.	**Introductions and Opening Remarks**	**Coordinator**
	- Agenda Review, TOR, Task Assignments	
9:30 a.m. – 11:30 a.m.	**Assessment Presentations**	**Analytic Team**
	- *Assessment Data & Methods*	
	- Identify additional analyses, sensitivities, corrections	
11:30 a.m. – 1:00 p.m.	**Lunch Break**	
1:00 p.m. – 6:00 p.m.	**Assessment Presentations (continued)**	**Analytic Team**
	- *Assessment Data & Methods*	
	- Identify additional analyses, sensitivities, corrections	
6:00 p.m. – 6:30 p.m.	**Public comment**	**Chair**

Tuesday Goals: Initial presentations completed, sensitivity and base model discussion begun

Wednesday

8:00 a.m. – 11:30 a.m.	**Panel Discussion**	**Chair**
	- *Assessment Data & Methods*	
	- Identify additional analyses, sensitivities, corrections	
11:30 a.m. – 1:00 p.m.	**Lunch Break**	
1:00 p.m. – 6:00 p.m.	**Panel Discussion/Panel Work Session**	**Chair**
	- Continue deliberations	
	- Review additional analyses	
	- Recommendations and comments	

6:00 p.m. – 6:30 p.m.	**Public comment**	**Chair**

Wednesday Goals**:** sensitivities and modifications identified, preferred models selected, projection approaches approved, Report drafts begun

Thursday

8:00 a.m. – 11:30 a.m.	**Panel Discussion**	**Chair**
	- Final sensitivities reviewed.	
	- Projections reviewed.	**Chair**
11:30 a.m. – 1:00 p.m.	**Lunch Break**	
1:00 p.m. – 5:30 p.m.	**Panel Discussion or Work Session**	**Chair**
	- *Review Reports*	
5:30 p.m. – 6:00 p.m.	**Public comment**	**Chair**
6:00 p.m.	**ADJOURN**	

Thursday Goals: Complete assessment work and discussions, final results available. Draft Reports reviewed.

Appendix 3: List of Participants

Carolyn Belcher, DNR, GA

Mary Christman, University of Florida

Robin Cook, CIE

Bob Ellis, FWRI, FL

Desmond Kahn, CIE

Marcel Reichert (Chair), DNR-Marine Resources Division, SC

Joel Rice, CIE