SEDAR

Southeast Data, Assessment, and Review

SEDAR 47

U.S. Southeastern Goliath Grouper

SECTION IV: Review Workshop Report

June 2016

SEDAR
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405

Table of Contents

Table of Contents 2
1. INTRODUCTION 2
1.1 WORKSHOP TIME AND PLACE 2
1.2 TERMS OF REFERENCE 2
1.3 LIST OF PARTICIPANTS 3
1.4 LIST OF REVIEW WORKSHOP WORKING PAPERS AND DOCUMENTS 4
2. REVIEW PANEL REPORT 4

1. INTRODUCTION

1.1 WORKSHOP TIME AND PLACE

The SEDAR 47 Review Workshop was held May 17-19, 2016 in St. Petersburg, Florida.

1.2 TERMS OF REFERENCE

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. Prepare a Peer Review Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference.

1.3 LIST OF PARTICIPANTS

Workshop Panel

Marcel Reichert, Chair ..... Chair, SSC
Carolyn Belcher ..... SSC
Mary Christman ..... SSC

Robin Cook ..... CIE Reviewer
Bob Ellis ..... SSC
Desmond Kahn ..... CIE Reviewer
Joel Rice ..... CIE Reviewer

Analytic Representation
Joe O'Hop ..... FWRI, St. Petersburg
Joseph Munyandorero ..... FWRI, St. Petersburg

Observers
Dustin Addis ..... FWRI, St. Petersburg
Steven Atran ..... GMFMC
Shanae Allen ..... FWRI, St. Petersburg
Michael Drexler ..... Ocean Conservancy
Doug Gregory ..... GMFMC
Elizabeth Herdter ..... FWRI, St. Petersburg
Michelle Masi ..... FWRI, St. Petersburg
Bob Muller ..... FWRI, St. Petersburg
Mike Murphy ..... FWRI, St. Petersburg
Brian Schoonard ..... GMFMC

Staff
Julie Neer ..... SEDAR
Ryan Rindone ..... GMFMC Staff
Charlotte Schiaffo ..... GMFMC Staff

1.4 LIST OF REVIEW WORKSHOP WORKING PAPERS AND DOCUMENTS

Documents Prepared for the Review Workshop			
SEDAR47-RW-01	The tpl file, data file, and control file for a Stochastic Stock Reduction Analysis (SSRA) program	Joseph Munyandorero	17 May 2016 Updated 24 May 2016

2. REVIEW PANEL REPORT

Review panel Report SEDAR 47 – Goliath Grouper Benchmark assessment

2. Executive Summary

The SEDAR 47 Goliath Grouper stock assessment presented to the Review Panel (RP) included two assessments models that were similar in terms of structure, but differed in the parameterization and use of the data. The assessment team acknowledged, and the RP agreed, that Goliath Grouper currently is a data poor species, and the approaches were selected accordingly. The two models were a "Stochastic Stock Reduction Analysis (SSRA)" and a "Catch Free" model. The RP concluded that the assessment does not constitute the best available scientific information and there is a significant level of uncertainty that has not been explored. Main areas of concern were the available data, treatment, and high uncertainty of the landings (catch) and the indices of relative abundance, and the structure of the chosen assessment models. As a result, the RP recommended that this assessment was not adequate to support status inferences, and as a result should not be used as a basis for management advice. However, a general increase in abundance since moratorium appears to be a reliable signal.

2.1 Terms of Reference

2.1.1. 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?

There was no data workshop so this was difficult to evaluate; the analysts provided some detail, but the RP concluded that there are numerous issues with the data and its treatments, which are outlined below. The RP felt that this assessment could have benefitted from a data workshop (or webinars) to discuss important issues related to the data.

<u>Catch / Removals</u>

Observed fishery removals consisted of commercial landings (1950–1989), reported recreational landings (1981–1989) and recreational dead discards (1990- 2014) with an assumed release mortality of 5%; they were considered to be known without error. There are significant concerns with the removals data. Historical catch data is highly variable and discontinuous pre 1950. The assessment team also considered helpful accounts of landings back to 1884. These accounts mention a pattern of catches from U.S. waters being landed in Havana, which could be problematic for landings records. The account presents evidence of a significant commercial fishery for Goliaths averaging over 100 lbs. in the early decades of the twentieth century back to the late 19th century (see also McClenahan et al. 2009). Acknowledging the high uncertainty of historic landings, the RP suggested to assess the sensitivity of assumption of an earlier virgin stock (some time prior to 1950) to the assessment. Landings after that time period were mainly restricted to the state of Florida. These data were adjusted for the period of 1965-1984, based on suspicion of over-reporting by a single dealer in Lee County (FL); this suspicion was supported by visits from biologists to this dealer, who did not observe the amount of Goliath Grouper reported by the dealer. The reasoning behind the exact adjustment factor was explained; however, no sensitivities were given nor were alternative catch histories given. The specific percentages as chosen by the analysts were not examined in detail and subsequent analysis could result in different percentage reductions.

Commercial discards were not estimated. This is another large area of data uncertainty, with the report showing an approximate 7.5%-11% observed occurrence of Goliath Grouper in the vertical and bottom

longline fisheries in recent years. Recreational removals varied significantly in the average weight per fish and overall catch size by year, and was a source of considerable variability in the data.
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 (or CV: coefficients of variation). Although they were not presented in the recreational landings in the report (Table 3.3.1), the CV 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 CV values will probably be higher due to the partitioning of the sample size among areas. The low catches of Goliaths during the period from 1981 through 1993, may have a high uncertainty and it would have been helpful if CVs were presented in the report. The (recognized) uncertainty around the recreational catch violates the model assumption that catch is known without error (see also discussions below). The RP recommends that re-examining the Assessment Teams methods of constructing historical removals should be a research priority. This recommendation is not to reinvent the catch data using new methods, rather to try to understand the influence of plausible alternative catch streams on the assessment.

Indices of Abundance

Four indices of abundance were used in the assessments. These are addressed individually below. The details concerning general index model selection, development, and diagnostics (apart from the final deviance table) were missing from the report.

The Everglades National Park (ENP) index is a fishery index that is conducted annually by the National Park Service biologists who sample (interview) recreational anglers in the Park. The index covers important juvenile habitat, thought to be the core habitat at the beginning of the moratorium.
The raw data from the index were subjected to generalized linear modeling based on the delta distribution, which has become standard practice, and attempts to remove variation attributable to factors other than abundance. This index is valuable as a general recruitment index of immature fish and because it extends back to 1974. The fact that this index and the MRIP estuarine index show similar trends is reassuring, and suggests that both may be reliable. The ENP index is largely unable to take account of changes in the fishers behavior over time (e.g. due to the moratorium; “effort creep”), which may or may not be significant, but would relevant to explore. An important feature of this index is that it covers the period before and after the moratorium.

The REEF Dive Index is an index developed from reported sightings by volunteer divers which have gone through a training program in fish identification and survey techniques taught by the REEF organization. This index has no rigid experimental design, includes numerous reports (of sightings and non-sightings), and is generally not oriented at observing Goliath Groupers. Arbitrary criteria intended to balance the need for spatial coverage by the Assessment Team was developed to require a dive site to have at least 10 reports over the last 20 years, and at least one positive sighting of a Goliath Grouper. Data are reported as categorical variables (0, 1, 2-10, 11-100, 100+). The RP concluded that the standardization of these data which was done with a Poisson generalized linear model, is inappropriate for categorical ordinal data. The REEF data was supplemented by a targeted survey from the Great Goliath Grouper Count (GGGC) data, a targeted dive survey that is similar in method to the REEF survey conducted annually by the Florida Wildlife Commission and Florida Sea Grant from 2010 to present. The RP felt that the combination of the non-target REEF data and the targeted GGGC data was problematic for the appropriate interpretation of the index.
Another issue discussed by the RP was that divers choose which reef to visit, which could introduce bias in favor of reefs with higher abundance of Goliath Grouper.

However, overall, a diver index may have added value because Goliath Grouper will have higher "catchability", since they only need to be seen, whereas a hook-and-line-based index (e.g. MRFSS, MRIP, ENP) requires Goliath Grouper to be brought to the boat (for identification). Large Goliath Grouper can more easily break tackle and as a result, may never be reported by recreational anglers.

The MRFSS/MRIP Indices are angler intercept surveys conducted throughout Florida waters. This data set was separated to create an inshore and an offshore index. Both 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 an order of magnitude in the same year. As discussed above and further detailed under section 1.c., these surveys were also the data source of recreational catch estimates, and there is a potential confounding problem with recreational catch estimates and catch-per-trip indices originating from the same survey data.

Stock Structure

A single stock within US waters was assumed for the assessment and the RP considered this to be reasonable assumption. Goliath Grouper are distributed throughout the tropics, subtropics, and warm temperate coastal waters of the Atlantic Ocean. Genetic data indicate that the stocks on the western and eastern shores of Florida are related. The RP was content with the assumption of a "US only" stock in the absence of compelling data to the contrary. However, within the United States the resolution of the data is not detailed enough to determine spatial structure of the stock. Tagging data demonstrate site fidelity, but also long distance movements (~400 km), in particular related to spawning. The vast majority of the data used in the assessment originate from Florida. However, historical landings indicate a range that in the US, Goliath Grouper occurs through southern Texas in the GOM, and up to North Carolina in the South Atlantic.

Life History Information

Key decisions for the life history information were based on small sample sizes (e.g. the fecundity estimate was based on 2 fish). The maturity is assumed to be knife edge, but there is a considerable amount of uncertainty with respect to the maturity. Although current data are inconclusive, there was some discussion of unpublished report about the reproductive strategy of Goliath Grouper (diandric protogynous hermaphrodite vrs. gonochorist).

Age data

The chosen assessment models were, in essence age-based models. However, the age data for Goliath Grouper are relatively sparse and highly uncertain.

Selectivity

The selectivity of the fisheries was separated into two blocks in the assessment, before and after the moratorium. The RP agreed that, in general, this was a sound decision by the analytic team, however because commercial discards are ignored in the assessment, the assumed selectivity of sub-adult and adult age class is likely mis-specified. Furthermore it is questionable as to whether the largest fish (>2 meters) would be retained by the majority of the fishing gear.

The assumed hook-and-line selectivity for the SSRA model (report figure 6.3.1) is flat-topped, meaning that the fishery catches the oldest, largest grouper as readily as it catches those just attaining full selectivity (ages 14+). However, remarks made at the Review Workshop indicates that the larger Goliath grouper often break lines, and may also break commercial gear. This may result in a dome shaped selectivity for hook-and-line gear. As no selectivity sensitivity runs were presented, the RP

recommended that the sensitivity of model output to selectivity assumptions (flat-topped vrs. dome) should be investigated.
The selectivity associated with the index of abundance for the MRFSS/MRIP and ENP estuarine indices is representative of the frequency distribution of the age of fish in the estuarine catch, rather than the proportion of fish selected by the gear from the population. Because the younger fish are more abundant in the population this "selectivity curve" will overestimate the selectivity of young fish, and underestimate the selectivity of older fish.
In general, the RP concluded that the selectivity choices and development were poorly described in the assessment report, and no sensitivity runs were presented. This hampered a thorough review of consequences of these choices for the assessment by the RP.

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

The analysts reported that this was a "data poor" assessment, and acknowledged that the uncertainties are likely at the high end of the expected level, especially those related to the catch. The CV's for the indices of abundance were highly variable, but it was unclear to the RP what the various CV estimates represented. These CVs were used to set priors in the assessment model. The RP suggested that it would be helpful to include model diagnostics, such as plots of residuals or other portrayals of uncertainty measures from linear modeling of the indices. Also, it would be helpful to compare the indices before fitting the model (raw or nominal indices) with the results of modeling to explore how influential a-priori choices were to the model outcome.
Uncertainty in commercial landings prior to the moratorium due to the suspected over-reporting by one dealer was thoroughly discussed. The RP felt that the assessment team made a reasonable and well-explained correction, reducing the landings by almost 50% from the reported landings. As this is a major correction, the RP felt that it would be helpful to explore a model run using the uncorrected reported landings to investigate the effect of this correction.

c) Are data applied properly within the assessment model?

MRFSS/MRIP data was used to develop an index of relative abundance, and also to derive part of the catch. The way this information is used in the model resulted in the fact that the estimated errors in the catch may be correlated to those in the index. However, the RP acknowledged that this is not unusual in assessments that use the MRIP data, but the correlation should be made explicit. The RP recognized that the MRIP CPUE data was used as an input in the index of relative abundance, which is appropriate. The reported B_2 catch (recreational discards) was the only catch allowed during the moratorium beginning in 1990, and as a result was the basis for the CPUE index. This was converted into the input "catch" from the fishery, based on 5% mortality of discards. The RP raised a concern that the indices and the discard estimate were derived from the same data source, and the resulting "catch" trend is identical, or at least (highly) correlated with the trend in the indices they were derived from. The RP recognized that this approach is not unusual, and that the impact on the model may be somewhat limited because the "catch" (discard mortality) was derived as the sum of estimated discards from the two surveys: MRIP estuarine and MRIP offshore, and catch was standardized by effort (catch-per-trip) to construct the index. Also, the index model provided a measure of uncertainty around index values portrayed in the confidence intervals (see report figures 4.3.1 and 4.3.2). Conversely, if abundance rises and effort remains roughly similar, it is reasonable to expect higher recreational catches, and thus more discards. As a result, the RP agreed that it is reasonable to expect some correlation between abundance and

discards. However, the RP felt that the documentation and information provided did not allow a thorough evaluation of the indices, their use to construct catch, and the impact on the model.

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

The assessment approach was twofold by exploring a SSRA model and a Catch Free model. The SSRA model used catch information that was not considered reliable, and the MRIP data was used to construct an index as well as a measure of effort to get catch (see discussion above). The choice of a catch free approach seems to be supported by several indices and other information.
The MRFSS/MRIP recreational catch per trip data is available from 1981 through 2014. The analysts conducted analysis of the raw data to attempt to increase its accuracy and precision, but concluded that only data from 1997 through 2014 was adequate. This is, in part, because Goliath Grouper occurrences in the recreational catch data were sparse prior to 1997. The RP felt that it would seem helpful to the modeling process to include survey data during the period of lower abundance (from 1981 forward). A possible alternative approach could be to divide the total catch estimates by the total trips estimate for at least for the earlier period, when more refined analyses may be impractical due to sparseness or data deficiencies. Other (statistical) approaches could also be used for the earlier part of the time series. A possible sensitivity analyses extending the indices back to 1981 could be helpful for assessing model adequacy and the impact of data from time periods with lower abundances.
The analysts divided the data into two geographical area based on habitat: estuarine and off-shore. The estuarine was used as information about younger, immature fish, which is appropriate. However, employing the information by area resulted in two smaller data "components", which increased uncertainty, especially around estimates of trips and catch. However, the RP acknowledged that, the advantage of using two areas was that the estuarine area functions as a nursery area for immature fish, which tend to move offshore as they age. As a result, the MRFSS/MRIP and ENP estuarine indices can be fortuitously employed as an index of recruitment. Conversely, the offshore area data can be employed as an index of the mature portion of the stock. As a result the RP supported that decision, in spite of it potential for increasing uncertainty.

2.1.2. 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?

Two principal assessment models were presented and discussed in the assessment report. These were the "Stochastic Stock Reduction Analysis" (SSRA) (Martell et al 2009) and the "Catch Free" model (Porch et al, 2006). Both models have been published in the peer reviewed literature. The RP noted that it is important to recognise that the assessment team indicated SSRA model has been modified by FWC to allow the inclusion of multiple survey indices, and that the Catch Free model configuration was modified also (see page 33 of the assessment report). These modifications do not appear to have been reviewed externally and the RP was unable to fully evaluate the impact of these modifications for either model.
The models share some important similarities which include:

- The underlying population is model age structured
- A Beverton-Holt stock recruitment is assumed
- Recruitment deviations are treated as random effects and characterise relative year class strength

- Fishing mortality is modelled as the product of an age and year effect
- Survey indices are treated as proportional to biomass or numbers conditioned on age specific selectivity.
- Parameters are estimated by maximizing a likelihood function
- Penalty functions are used to constrain some of the model parameters. These are referred to as "priors" but are not true Bayesian priors and may result in improper posterior distributions.

Important differences between the models are:

- SSRA uses an estimate of total fishery removals (dead catch) and these are treated as error free. They do not contribute to the likelihood.
- Unlike SSRA, the Catch Free model treats selectivity, natural mortality, growth parameters and fishing mortality as parameters to be estimated.
- SSRA parameterises the stock recruitment function in terms of F_{msy} and MSY, and these are the main ("lead") parameters to be estimated. An important consequence of this is that the stock recruitment parameters are conditioned on the assumption of selectivity and will change if the selectivity assumption is changed.

The RP noted that, while not a feature of either model, the analysts assumed different relationships between natural mortality and weight for the two models, which reduces the comparability between the models, and affects the calculation of MSY reference points.
The models are well known variants of age-structured production models and can be regarded as scientifically sound. Whether they are considered "robust" depends heavily on the data used. Here "data" may include constants, such as age at maturity, selectivity, M, etc. These do not enter the likelihood unlike observations, such as survey indices. Where data enter the model as constants, the accuracy is particularly important to avoid cumulative errors. For example, fishery removals and selectivity estimates can be critical in determining the model outcome, yet there is considerable uncertainty surrounding the values used in these assessments. The RP felt that it was not possible to conclude that the methods used are robust. This decision was largely based on various analyses reported in the assessment document (i.e. the MCMC runs for the Catch Free model) and additional model runs that were performed during the meeting (i.e. the "leave-one-index-out" survey analysis for SSR, see post RW addendum to the Assessment Report). Where priors are used, as was the case in these assessments, it is particularly important to examine whether these are updated by the observations, and to examine the sensitivity of model estimates to the priors. These diagnostics were not done, which prevented the RP from fully assessing robustness of the model.

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

For each model, only one or a very few model configurations were presented. The RP concluded that although these configurations were plausible, they may not necessarily represent the optimal model configuration. As mentioned under section 2.1.2-a, considerably more analysis of the prior assumptions is needed for a full evaluation of the models. More consideration should be given to the choice of indices are included. The RP expressed concern that, given the uncertainty in the catch data, these are included in the model as error-free constants. It was suggested that this could be addressed by treating the catches as observations that enter the likelihood, and would allow errors to be estimated.

c) Are the methods appropriate for the available data?

The RP concluded that the SSRA and Catch Free models are appropriate tools, but given the available data, should be used as part of a suite of alternative models, if only to better characterize model uncertainty. The available data could potentially be analysed using a variety of models, including surplus production models and other data-poor approaches. The RP recommended that considerably more thought needs to be given to the implication of handling the fishery removals as known constants and developing changes to the SSRA/Catch Free models so that this issue can be explored.
Given the significant difficulty of trying to construct abundance indices and fishery removals for Goliath Grouper, it is important to review what can be realistically derived in terms of useful reference points. Both models need age information in the data to estimate parameters, however, there is a lack of such information in the data. As a result, many assumptions have to be made to estimate age dependent parameters. Also, the VonBertalanffy (VB) growth parameters are correlated, yet in the models they can vary independently. The RP recommended that if one parameter is chosen, the other parameters should be fixed based on the age-at-length analysis (e.g. the VB model).
The RP was unable to fully evaluate how influential these assumptions were to the outcome of the assessment models. Simpler methods that consider only stock trends may be (more) useful than trying to reconstruct a fully age-structured population model. The RP felt that this assessment could have benefitted from an assessment workshop (or webinars) to discuss important issues related to the model and model parameter choices.
The RP noted that the Goliath Grouper stock attained an exponential increase, seemingly due to one or more very strong recruitment events in the 2000s. This dramatic increase was followed by a steep decline of least the estuarine pre-recruits, possibly due to an episodic natural mortality caused by some combination of red tides and severe cold snaps in 2008-2010. Since the model is informed that natural mortality is low and constant, it may not be able to appropriately account for episodes of high M. Indeed, both the SSRA and the Catch Free model show a poor fit of the indices' increase, as is illuminated by the residual pattern (see assessment report figures 6.9.1, 6.9.2, 7.5.2 and 7.5.3). This also affects model estimates of F during this period.

2.1.3. 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?**

The RP concluded that neither the SSRA nor the Catch Free model outcomes support inferences on stock status. For the SSRA, this conclusion was largely based on the uncertainties in the data and the sensitivity to, among other things, the choice of the beginning of the time series and the REEF index. For the Catch Free models this was largely based on the uncertainties in the data and poor convergence results of the MCMC, which suggest multiple local maxima in the likelihood.
The RP was unable to fully evaluate the abundance, exploitation, and biomass estimates from both the SSRA and Catch Free models. Both models were valuable in illustrating plausible stock and exploitation trends, but neither was sufficient to support status inferences, and thus not adequate for management advice based on MSY reference points.
By definition, the Catch Free model can only provide estimates of relative abundance, exploitation, and biomass, so there is no information provided to the model to allow scaling to absolute values. The Catch Free model was previously used in SEDAR 6 and SEDAR 23, where it were adopted to provide relative estimates and guidance on the possible recovery time of Goliath Grouper. In SEDAR 23, the Catch Free model was employed again, but the context was changed with the management need to provide OFL and ABC recommendations. However, the Catch Free model cannot provide this information as it does

not use data on removals to scale necessary estimates. Thus, the RP concluded that for SEDAR 47 this model is, again, not appropriate for stock status determination.
During the SEDAR 23 RW, an SSRA model was presented for exploratory purposes, but the review panel did not use it to make inferences about stock status as it had not been previously considered by the SEDAR 23 Assessment Workshop. In principle, with better quantification of removals, and with conducting various sensitivity runs, the SSRA could be used to provide more relevant information for management. However, the SEDAR 47 RP concluded that the SSRA model critically depends on credible inputs of removals, which were deemed too uncertain in the current assessment. Thus, for SEDAR 47 the RP does not consider the SSRA model appropriate for stock status determination.

(a) Is the stock overfished? What information helps you reach this conclusion?
(b) Is the stock undergoing overfishing? What information helps you reach this conclusion?

The RP did not accept either model as sufficient to infer stock status and support management decisions, thus it cannot determine if the stock is overfished or undergoing overfishing.

(c) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?

The stock recruitment curve used was estimated internal to the models and assumed to follow a Beverton-Holt relationship. The robustness of the chosen stock recruitment relationship was not explored, therefore it is not known how informative the presented stock recruitment relationship is. The estimated stock recruitment plot from the SSRA model (Figure 6.9.6) suggests that a Ricker model may be appropriate, because the highest recruitment estimates occurred at moderate values of SSB, and recruitment declines somewhat at higher SSB. In addition, Goliath Grouper biology and ecology may support a Ricker stock-recruit model choice.

(d) Are the quantitative estimates of the stock 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?

The Review Panel felt that the quantitative estimates produced by both the SSRA and Catch Free models were not reliable. During the RW the Panel requested various sensitivity runs be produced from the SSRA, including starting the assessment at 1975, dropping the REEF diver data index, and including each index in isolation. The stock status determinations produced from these various sensitivity runs varied greatly and contributed to the lack of confidence that the RP had in the model's ability to reliably estimate stock status for the Goliath Grouper population (see also section 2.1.5).
One of the main sources of available data to determine the Goliath Grouper stock trends comes from diver observations collected by the REEF Foundation and, more recently, from the GGGC conducted annually by the FWC and Florida Sea Grant from 2010-2014. Positive aspects of these surveys are that they are not fishery based and have broad spatial coverage. Despite numerous concerns from the RP regarding the treatment of the REEF diver index in both models (e.g. not a random survey, variability in spatial coverage over time, and see notes above), these data might be useful as population indicators of trends in relative abundance and/or spatial distribution if standardized appropriately. At this point, the REEF foundation data extend back to 1993, and thus covers nearly the entire duration of the moratorium and subsequent population recovery. While these data may not be appropriate for use in the models presented at the RW, the Panel recommends further exploration of methods of index

formation. Spatial analyses of these data may be informative in terms of the spatial extent of population trends.

2.1.4. 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?**

No projections were presented for SSRA. Projections from the Catch Free model give an indication of possible future biomass trends, but cannot indicate where biomass lies in relation to reference points.

2.1.5. 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.**

The Assessment Team indicated that in addition to the SSRA and the Catch Free models, other models were considered and rejected. However, the Assessment Team did not provide any details of the issues or sources of problems with these alternative models. Hence, the RP was unable to evaluate these efforts as potentially useful for addressing model uncertainty and assessing stock status.

Sensitivity Runs

The number of sensitivity runs provided in the assessment report was very limited, which significantly impaired the RP's ability to fully evaluate the models. Additional sensitivity analyses were performed for both models during the Review Workshop at the request of the reviewers. The SSRA model was run with and without various indices of abundance and for different time periods (1975 - 2014 vs. 1950 - 2014). The results indicated that biomass trends were strongly impacted by the changes in start date. In addition, the sensitivity runs with different indices led to further changes in the model fit and predictions. Together these provide good indications of the high degree of uncertainty in model results.
For the Catch Free model, sensitivity analyses included: 1) putting prior distributions around the two levels of natural mortality (M = 0.12 and 0.18); and 2) either fixing selectivity curves or using priors on selectivity parameters. Model outputs under the different assumptions in (1) were provided as a means of assessing consistency of the conclusions. For (2), comparison of approximate catch to predicted catch from the Catch Free model was provided.
The RP was unable to fully evaluate the requested sensitivity runs (with a start year of 1975) because the models assumed a virgin stock biomass in 1975, which is likely unrealistic. As a consequence, the RP was unable to fully compare the "base" model with these sensitivity runs.
None of the sensitivity runs for either model tested whether the technical conclusions concerning overfishing were similar, regardless of model inputs and decisions. This is a serious omission making it difficult to judge the robustness of the model results.

Sources of uncertainty in the life history data sources

I) Catch Free model

Although full details for every parameter were not provided (see Table 7.2.1 in the Assessment Report for provided details), it appears that uncertainty in the life history parameters was allowed for by drawing them from a prior distribution. The RP agreed that this is a good approach to incorporating and propagating uncertainty in the model, as is the use of MCMC to obtain posterior distributions that include variability in the outputs. The use of multiple chains was helpful for checking convergence of the model and revealing the uncertainty in the model fit.

The RP was not convinced that the parameters and methods were appropriately chosen. A specific example is the use of independent prior distributions for VB model growth parameters, which should be treated as either a bivariate or tri-variate joint distribution with correlation between parameters. Use of the independent distributions leads to parameters selections that may or may not be within the valid sample space of the joint parameters. This could be part of the problem with non-convergence of the chains in the MCMC runs of the model.

The use of phase plots was a good portrayal of the uncertainties or variance around the model optimization solutions. In addition, the incorporation of an overall variance parameter to include process/unexplained error as described in Porch et al. (2006) was useful for capturing sources of variability not optimally described by CVs or variances for data inputs.

II) SSRA Model

Unlike the Catch Free model, life history parameters were assumed to be fixed (Appendix A of the Assessment Report). Uniform bounded priors were placed on F_{msy} and MSY, and a prior was placed on the compensation ratio (κ). Like the Catch Free model, MCMC was used to explore uncertainty. All other life history parameters, e.g. weight-length relationships were assumed to be fixed and known. As a result, these sources of uncertainty were not included in the model outcomes variability.

The retrospective analyses, the plots comparing observed and predicted indices, the residual plots, and the MCMC simulations were all well-done and helpful to the RP for addressing model validity and assessing uncertainty.

Sources of uncertainty in the data sources

The SSRA model did not include measures of uncertainty for the catch. The analysts appropriately raised the strong possibility of over-reporting, and used an approach to correct it. Although worthy, the adjustment to the catches may not be accurate and may influence the model outcomes. It would be helpful to run sensitivity analyses in the SSRA model to explore the impact of the range of possible corrections.

Uncertainty for the other indices (see above) were provided, but the RP was concerned about the validity of the value of the CVs. The CVs were based on only the probability distributions assumed for the data index and sample sizes, and as such may be a poor representation of the true variability. However, the RP realized that to account for this would have required adjustments to the standardization of the indices. The assessment report provided insufficient information for a full evaluation by the RP.

2.1.6. Consider the research recommendations provided and make any additional recommendations or prioritizations warranted.

a) Cleary denote research and monitoring that could improve future assessments.

The data used in the models mostly originated from Florida. Sparse data from elsewhere in the species (historical) range may be indicative of low population size (either as a function of natural distribution patterns or constriction of the population due to heavy fishing pressure), or poor sampling (including landings). This issue needs to be further explored as it has bearing on the geographical validity and usefulness of the assessment for regional management.
There was some concern by the RP about the method the Assessment Team used for combining the GGGC and the REEF survey as there is a potential for bias (e.g. potential for targeting sites with known high abundance of Goliath Grouper in the GGGC survey). How influential the inclusion of the GGGC data was to the outcome of the model should be explored.
Many of the research recommendations provided in the Assessment Report include research that would not necessarily improve future assessments for this species. The SEDAR 23 RW concluded that "The next benchmark assessment cannot be successfully completed without data from the research recommended by the Data, Assessment, and Review Panels." The outcome of the SEDAR 47 benchmark assessment process indicates that much of this information is still needed in order to successfully complete an assessment for Goliath Grouper.
Specifically, research and monitoring efforts that could improve future assessments for Goliath Grouper include:

<u>Life history information</u>
Basic reproductive data is lacking throughout the species distribution. This includes size and age at maturity for each sex, sexual sequence with size and age for each sex, and fecundity. In the SEDAR 47 assessment, the reproduction functions used in the models made some strong assumptions about the maturity schedule and fecundity rates that were based on insufficient data. Greater resolution of data, especially maturity at size or age, would alleviate the impact of these assumptions for future assessments.
A limited research harvest should be considered to fill the remaining gaps in life history information for Goliath Grouper. Such a harvest should incorporate individuals from across the size spectrum, but should focus on larger individuals as they may be beneficial to ground truth the fin-ray aging techniques used for the offshore age composition, and to develop fecundity schedules.
Additional research on the age structure of the catch, especially in the offshore recreational fishery, is needed. The SEDAR 47 assessment used age composition of only 22 adult individuals that were caught by a research fishery and aged with fin rays (Koenig et al. 2013). This age composition was used for multiple parts of the assessment and may provide a large source of the assessment uncertainty. Cooperative research efforts with the recreational charter and headboat fisheries could be informative towards generating better information on the offshore recreational age composition.
Discard mortality estimates are needed across the species distribution. For the SEDAR 47 assessment, a fixed discard mortality estimate was applied to the post-moratorium harvest. However, the uncertainty around this estimate is unknown and may be substantial.

<u>Stock definition</u>
SEDAR 23 recommended that Goliath Grouper should be genetically sampled from areas across the stock range in the South Atlantic and Gulf of Mexico to allow for a more thorough examination of the current single stock definition. The SEDAR 47 RW was presented with a brief summary of these efforts,

which seem to support that single stock definition. Like many other sources of information informing the SEDAR 47 assessment, this information remains in progress or is incomplete and has not yet been vetted by peer review.
Examination of spawning aggregations over the entire distribution range should include seasonality, sex ratios, and individual fidelity.

Fishery independent sources of information are lacking or uncertain
The SEDAR 47 AT indicated that a specifically designed pre-fishery recruit survey (e.g. mangrove habitat) would help guide recruitment in the assessment model.

Develop and/or explore methods to take into account episodic mortality events.
One issue with the SEDAR 47 assessment was the use of a fixed value for natural mortality at age, despite evidence that episodic mortality events (i.e. cold-kills) have affected the Goliath Grouper population. Options to account for this mortality should be explored for future assessments. Methods used in other assessments (e.g. to address red tide events affecting red and gag grouper in the GOM) include incorporating episodic mortality events as a separate removal fleet. These methods may be appropriate for Goliath Grouper and could reduce some of the uncertainty in the estimates of natural mortality.

Reexamine methods of constructing historical removals
The use of length data from MRFSS/MRIP recreational Goliath Grouper removals need to be further examined. In SEDAR 47, the methods used to apply mean length of catch was inconsistent between years when there was missing and/or suspect data, and years with an estimate from the MRFSS/MRIP database. This introduced a significant amount of uncertainty to the harvest estimates.

Incorporate Data from Low Abundance Years into Indices
The Assessment Team discarded some of the data from index development due to very low catch rates in years adjacent to the moratorium. As a result, low abundance indices are removed from the assessment. Methods for incorporating these data into appropriate statistical models for standardization and development of indices should be explored.

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

The model was appropriately configured, but the nature of the data, data choices, and model choices not provide results that can be considered BSIA. Details are provided under various TORs above.

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

Inclusion of a complete and comprehensive data report, as well as a complete assessment report would have been helpful to the RP in evaluating this assessment, specifically how the data were handled. The analysts indicated other modeling approaches were investigated, but uninformative results led to their exclusion from the discussion. The RP indicated that information associated with those runs would have helped evaluate model uncertainty.

- Given the nature of the (limited) data and resulting modeling approaches, a Data Workshop, and possibly an Assessment Workshop (or webinars) should be considered when scheduling a next assessment.
- Consideration of other (data poor) model approaches. It will be useful for the RP if explored models and outcomes of those explorations are included in the assessment report.
- Explore methods that allow for a varying M (e.g. as result of cold kills).
- Considerably more age data is required to inform the model.
- Explore other stock/recruit relationships.
- More complete sampling of the catch to provide lengths and weights of all individual fish.
- The REEF "abundance" was a ranked abundance scale, not a true abundance scale. Other ways of estimating abundance, and the effect of the choice on the model configuration and outcome should be explored.
- Spatial analyses of the REEF data may be informative in terms of the spatial extent of population trends.
- Improved estimate of bycatch mortality, e.g. by using experiment studies. Also, commercial bycatch mortality was not included in the model. Acknowledging the paucity of data, estimates of this source of mortality may improve the model outcome.

2.2. Summary of Results of Analytical Requests

The analytical team provided several additional analyses and clarification of model structure and results. All are summarized and discussed in the previous sections of this report.

2.3. Additional Comments.

The RP had no additional comments.

3. Submitted Comment.

There were no additional submitted comments.