

SEDAR

Southeast Data, Assessment, and Review

SEDAR Grouper Assessment Review

Evaluation Panel Report

SEDAR
4055 Faber Place
Suite 201
North Charleston, SC 29405
(843) 571-4366

Table of Contents

1.	Introduction.....	1
1.1.	SEDAR Overview.....	1
1.2.	SEDAR Grouper Assessment Review Overview and Need.....	1
1.3.	Workshop Time and Place.....	3
1.4.	Terms of Reference.....	3
1.5.	List of Participants.....	4
1.6.	Supporting Documents.....	4
2.	Responses to Terms of Reference.....	5
2.1.	TOR 1.....	5
2.2.	TOR 2.....	14
2.3.	TOR 3.....	16
2.4.	TOR 4.....	18

1. Introduction

1.1. SEDAR Overview

SEDAR (Southeast Data, Assessment and Review) was initially developed by the Southeast Fisheries Science Center and the South Atlantic Fishery Management Council to improve the quality and reliability of stock assessments and to ensure a robust and independent peer review of stock assessment products. SEDAR was expanded in 2003 to address the assessment needs of all three Fishery Management Council in the Southeast Region (South Atlantic, Gulf of Mexico, and Caribbean) and to provide a platform for reviewing assessments developed through the Atlantic and Gulf States Marine Fisheries Commissions and state agencies within the southeast.

SEDAR strives to improve the quality of assessment advice provided for managing fisheries resources in the Southeast US by increasing and expanding participation in the assessment process, ensuring the assessment process is transparent and open, and providing a robust and independent review of assessment products. SEDAR is overseen by a Steering Committee composed of NOAA Fisheries representatives: Southeast Fisheries Science Center Director and the Southeast Regional Administrator; Regional Council representatives: the Executive Directors and Chairs of the South Atlantic, Gulf of Mexico, and Caribbean Fishery Management Councils; and Interstate Commissions: the Executive Directors of the Atlantic States and Gulf States Marine Fisheries Commissions.

SEDAR is organized around three workshops. First is the Data Workshop, during which fisheries, monitoring, and life history data are reviewed and compiled. Second is the Assessment workshop, during which assessment models are developed and population parameters are estimated using the information provided from the Data Workshop. Third and final is the Review Workshop, during which independent experts review the input data, assessment methods, and assessment products.

SEDAR workshops are organized by SEDAR staff and the lead Council. Data and Assessment Workshops are chaired by the SEDAR coordinator. Participants are drawn from state and federal agencies, non-government organizations, Council members, Council advisors, and the fishing industry with a goal of including a broad range of disciplines and perspectives. All participants are expected to contribute to the process by preparing working papers, contributing, providing assessment analyses, and completing the workshop report.

SEDAR Review Workshop Panels consist of a chair and 3 reviewers appointed by the Center for Independent Experts (CIE), an independent organization that provides independent, expert reviews of stock assessments and related work. The Review Workshop Chair is appointed by the SEFSC director and is usually selected from a NOAA Fisheries regional science center. Participating councils may appoint representatives of their SSC, Advisory, and other panels as observers to the review workshop.

1.2. SEDAR Grouper Assessment Review Overview and Need

The SEDAR Steering Committee determined that additional scrutiny should be devoted to recent grouper assessments.

A basic tenet of each SEDAR assessment is that all previous assessment decisions and assumptions are up for debate, and that each decision and assumption included in the current assessment is to be specifically evaluated and judged on scientific merit. Participants are clearly instructed that any decisions made in previous assessments are to be thoroughly evaluated in light of current knowledge. There is no requirement or expectation that decisions made regarding one assessment should be consistent with those in prior assessments, and, in fact, justifications based solely on past decisions are explicitly discouraged. As a result, SEDAR participants are compelled to continually improve assessment quality and it is acknowledged within the Southeast fisheries management community that SEDAR has improved assessment methods, data evaluation techniques, and awareness of critical data collection program characteristics.

One consequence of continually evaluating all prior decisions and striving to improve methods is that current assessments may develop solutions to data deficiencies and analytical challenges that differ from solutions applied in previous assessments. Previous SEDAR assessments have faced post-approval criticism brought on by technological advancements and improved understanding of data sources stemming from later assessments, and the accepted solution has been to apply the most up to date methods to each problem at the next available opportunity. For example, updates to SEDAR 1 and 2 assessments included model configurations and data treatments developed through subsequent assessments.

A similar situation arose recently when the findings of assessments for Gulf of Mexico and South Atlantic gag grouper were compared with those for Gulf of Mexico red grouper. Although many of the same datasets were included in the assessments for gag and red grouper in the Gulf of Mexico, the two species are exploited by similar fisheries, and there is potential overlap in the species range, the SEDAR 10 (gag) and SEDAR 12 (red grouper) assessments differed in key areas including data time series, discard mortality rates, estimation of natural mortality, and analysis of fishery-dependent catchability. Similar differences are also noted within the SEDAR 10 assessments for South Atlantic gag grouper and Gulf of Mexico gag grouper. It should be noted that the assessments prepared during SEDAR 10 and SEDAR 12 were judged separately on their individual merits and found adequate and acceptable by independent scientific review panels. In addition, the SEDAR 10 assessment of Gulf of Mexico gag grouper was also reviewed and deemed acceptable by the Gulf of Mexico Fishery Management Science and Statistics Committee. Nonetheless, the SEDAR Steering Committee determined that additional scrutiny should be devoted to recent grouper assessments.

This special review project is convened by request of the SEDAR Steering Committee to evaluate key decisions of the SEDAR 10 and SEDAR 12 stock assessments. The Steering Committee determined that additional evaluation should be devoted to these issues to ensure confidence in both the assessment process and assessment findings. The Steering Committee recognizes the inherent challenge in balancing demands to scientists to prepare each assessment with the best available data and most up to date methods with constituents' expectations that similar fisheries should receive similar analytical treatments. By initiating this project, the SEDAR Steering Committee intends to ensure every effort is made to verify that all decisions are scientifically sound and adequately scrutinized within the assessment process.

The SEDAR Grouper Assessment Review will be carried out in two steps. First, an evaluation panel of experts knowledgeable in the fisheries and the SEDAR 10 and 12 stock assessments will be convened to review the assessment findings and recommendations in light of current knowledge. This group will prepare a report including recommendations for subsequent

assessment analyses if justified. Second, a SEDAR review panel will be convened to independently review the findings and recommendations of the evaluation panel as well as any additional assessment analyses prepared as a result of the evaluation panel findings.

1.3. Workshop Time and Place

The SEDAR Grouper Assessment Review Evaluation Workshop was held March 19 - 22, 2007, in Miami, FL.

1.4. Terms of Reference

1. Review SEDAR 10 and SEDAR 12 assessment reports, relevant supporting documents, and recommendations, along with any additional research available since the SEDAR assessments, regarding the following specific topics:
 - A. The length of the time series to be used for the base cases in each assessment (Gulf gag, Atlantic gag and Gulf red grouper).
 - B. The treatment of the catchability coefficient for fishery-dependent indices of abundance in each assessment.
 - C. The estimation of the number and size composition of discarded fish, as well as the fraction of the discards that die in each assessment.
 - D. The treatment of the natural mortality rate and, in particular, the method used to scale the Lorenzen curve in each assessment.
 - E. Recommended reference points (minimum stock size threshold, maximum fishing mortality threshold and optimal yield) and whether those choices are consistent with the goals of the respective Fishery Management Plans and the Magnuson-Stevens Reauthorization Act.
2. Discuss how consistency in methodology should be balanced against the need to address differences in the data, fisheries and biology of the three stocks in question. Include in this discussion the significance of using different stock assessment algorithms for each stock.
3. Formulate recommendations for any additional analyses, sensitivity runs, or changes to the base cases that need to be made to the Gulf gag, Atlantic gag, and Gulf red grouper assessments based on the reviews of the specific issues addressed in TOR #1 and given the conclusions reached during the discussion of TOR #2.
4. Prepare a consensus report documenting committee discussions and recommendations. The report should be drafted during the workshop and finalized within one week of workshop conclusion.

1.5. List of Participants

<u>NAME</u>	<u>Affiliation</u>
-------------	--------------------

Workshop Panel

John Carmichael.....	SEDAR, Chair
Andrew Cooper.....	SAFMC SSC
Behzad Mahmoudi.....	FL FWC/GMFMC FSAP
Bob Muller.....	FL FWC/SAFMC SSC & GMFMC FSAP
Clay Porch.....	SEFSC
Erik Williams.....	SEFSC/SAFMC SSC

Council Appointed Observers

Dennis O'Hern.....	GMFMC AP
Bill Tucker.....	GMFMC AP
Bob Zales II.....	GMFMC AP

General Observers

Theo Brainerd.....	SEFSC
Craig Brown.....	SEFSC
Shannon Calay.....	SEFSC
Alex Chester.....	SEFSC
Guillermo Diaz.....	SEFSC
Dennis Heinemann.....	The Ocean Conservancy
Larry Massey.....	SEFSC

Staff

Tyree Davis.....	SEFSC
Rachael Lindsay.....	SEDAR
Larry Massey.....	SEFSC

1.6. Supporting Documents

- SEDAR. 2006. Stock Assessment Report for South Atlantic Gag Grouper. SEDAR10-SAR1. SEDAR, Charleston SC.
- SEDAR. 2006. Stock Assessment Report for Gulf of Mexico Gag Grouper. SEDAR10-SAR2. SEDAR, Charleston SC.
- SEDAR. 2007. Stock Assessment Report for Gulf of Mexico Red Grouper. SEDAR12-SAR1. SEDAR, Charleston SC.

2. Responses to Terms of Reference

2.1. TOR 1.

1. Review the SEDAR 10 and SEDAR 12 assessment reports, relevant supporting documents and recommendations, along with any additional research available since the SEDAR assessments, regarding the following specific topics:

- A. The length of the time series to be used for the base cases in each assessment (Gulf gag, Atlantic gag and Gulf red grouper).

The assessment evaluation panel (AEP) discussed pros and cons of different catch time series used for the base cases of each assessment. The catch time series started with 1962 for South Atlantic gag, 1963 for Gulf of Mexico gag, and 1986 for the Gulf of Mexico red grouper. The main focus of discussion was why shorter time series for red grouper given similarities among these fisheries. The AEP agreed that in order to provide best possible long-term perspective on stock status (virgin or near virgin biomass and recruitment levels) assessments must use the longest possible catch time series. Short catch time series, even with detailed composition data, can give misleading estimate of current stock status relative to un-fished stock level. However, there are difficulties with constructing long time series. These include questions concerning the quality of catch statistics, availability of reliable age composition and abundance indices data, and changes in fleet selectivity and fishing practices. These issues were discussed and evaluated by AEP for each of the three catch time series.

For all three assessments, reliable information on age composition, life history parameters, and indices were generally available from mid 1980s (except for South Atlantic gag headboat index which started in early 1970s?). Age composition data with high sampling resolutions were available after 1990. The key differences between the three catch time series were the quality of catch statistics, waterbody information, and whether the historical catch trends reflected virgin or near virgin stock status. Information on waterbody of landings is available beginning in 1962 for the South Atlantic gag and 1963 for Gulf of Mexico gag. For South Atlantic gag, exploitation prior to 1962 is believed to be slight, so the model can begin with a near virgin stock. For Gulf of Mexico gag, the landings prior to 1963 were also relatively low in comparison to recent years (although they were not considered low enough to justify assuming the stock was near virgin levels). For Gulf of Mexico red grouper, however, the landings prior to the mid 1970s and perhaps as far back as the 1930s were likely quite high owing to a large Cuban fleet that fished for red grouper off Florida and Mexico. Unfortunately, the Cuban landings are not well documented in many years and it was not possible to construct a reliable time series of landings prior to the 1980s. The AEP suggested that research effort should continue with constructing the historical catch time series for future Gulf of Mexico Gag and red grouper assessments

The panel agreed that the catch time series selected for base runs of the three assessments were appropriate and noted agreements among base runs and sensitivity runs with various catch time series and alternative methods (i.e., SRA). For example, results from the Gulf of

Mexico gag runs with catch time series starting in 1963 were consistent with sensitivity runs using catch time series starting in 1880. The stochastic SRA applied to the Gulf of Mexico gag and Gulf of Mexico red grouper catch time series starting in 1880 generated similar results as those generated from the base runs with catch time series starting in 1963 for Gulf of Mexico gag and in 1986 for the Gulf of Mexico red grouper. Due to the assumptions required to develop the 1880 time series, the evaluation panel agrees with the SEDAR 10 and 12 review panel decisions to develop recommendations based on models including the shorter time series.

- B. The treatment of the catchability coefficient for fishery-dependent indices of abundance in each assessment.

Review of SEDAR 10 and SEDAR 12 decisions

The assessments of South Atlantic gag grouper, Gulf gag grouper, and Gulf red grouper each explored models that:

- 1) held the catchability coefficients constant over time for fishery dependent indices of abundance,
- and
- 2) included an annual 2% increase in the catchability coefficients for fishery dependent indices.

Considerable discussion occurred during the SEDAR 10 Assessment Workshop concerning the use of time-varying or constant catchability coefficients for the fishery dependent abundance indices. Panelists agreed that catchability has likely increased over the last few decades, but disagreed on how much. As stated on page 8 of the SEDAR 10 Assessment Workshop report for South Atlantic gag grouper:

The group recognized that technology improvements over time, in particular better electronics, have made fishermen more effective and efficient at catching fish, although there was no firm conclusion about details. This issue is important for the present stock assessment because the assessments rely heavily on fishery-dependent catch rate abundance indices. Such indices divide catch by effort. When a unit of effort becomes more efficient at catching fish, the resulting abundance index becomes biased, making fish appear relatively more abundant.

In response, a proposal was discussed to assume an increased catchability of 2% per year (non-compounding), beginning in 1980 and continuing to the present. The value of 2% reflects findings of a recent published paper (Robins et al., 1996) and an ICES paper (Skjold et al., 1996), which examined other fisheries. The starting data reflects increased availability of better electronics.....Participants believe that some increase in catchability has occurred, but that estimating its magnitude is too difficult to be done at this assessment workshop. Workshop participants agreed to send to the Review Workshop runs made under both assumptions (constant or increasing catchability), without labeling either one the “base run” to the exclusion of the other.

The SEDAR 10 Gulf and SA gag grouper Assessment Workshop followed a similar path, sending runs made under both assumptions to the Review Workshop without labeling either one as the “base run”. The S10 Review Workshop for both South Atlantic and Gulf gag grouper, agreed that catchability has likely increased over time but did not believe that a constant 2% increase adequately describes the complex ways in which it has changed. As such, the SEDAR 10 Review Workshop rejected the assumption of a 2% annual increase in catchability for fishery dependent indices in favor of an assumption of constant catchability.

Citing the discussions during SEDAR 10, the SEDAR 12 Gulf Red Grouper Assessment Panel included both the assumption of constant catchability and a 2% annual increase in catchability for fishery dependent indices in their analyses. The “base case” sent to the Review Workshop assumed constant catchability, and a 2% annual increase was included as a sensitivity run. However, the label of “base run” versus “sensitivity” run was not intended to imply preference, but rather as a way to structure the presentation of the models (e.g., a baseline model and then a list of ways the baseline could be modified) (Behzad Mahmoudi and Shannon Cass-Calay, personal communication)

The SEDAR 12 Gulf Red Grouper Review Panel chose the assumption of a 2% increase in the catchability coefficients for the fishery dependent indices as the preferred model. On page 9 of the Review Panel Consensus Summary of the SEDAR 12 Review Workshop, it states:

The panel agreed that it would be unrealistic to assume constant fishery catchability over 20 years and requested that an annual 2% increase in catchability be incorporated in the base run to reflect increased fishing power (efficiency) principally due to technology innovations (GPS, GIS, cell phone communication, etc.) that cannot be quantitatively included in the standardization. This means that over a 15-year period, a 35% increase in observed fishery CPUE would be expected from a stock that was level in its abundance. The representatives of the fishing industry attending the meeting agreed that 2% per year was within a likely range.

SEDAR Grouper Assessment ad hoc Evaluation Panel Discussion

The Evaluation Panel spent considerable time discussing the history of the decisions outlined above along with the various reasons why catchability has likely increased for fishery dependent indices. The Evaluation Panel agrees with both SEDAR 10 and SEDAR 12 Assessment Panels and Review Panels in that catchability has likely increased. The discrepancy between the SEDAR 10 and SEDAR 12 decisions as to whether a constant 2% increase in catchability should be included in the preferred model arose as a result of different groups of people arriving at different conclusions when presented with imperfect information. In particular, neither panel was aware of any information quantifying the dynamics of the change in efficiency of the snapper-grouper fishery. The SEDAR 10 Review Panel was simply not comfortable with making a simplistic assumption of a constant 2% increase whereas the SEDAR 12 Review Panel decided that a simplistic assumption of a

constant 2% increase was more realistic than assuming a constant catchability. Such discrepancies should not be considered unusual or cause for concern.

The Evaluation Panel does, however, feel that future assessments of gag and red grouper should be consistent in their assumption of increasing catchability, but the Panel is not in a position to comment on whether one specific value or method for modeling this increase is more appropriate than another. The assumption of constant catchability for fishery dependent abundance indices in the snapper-grouper complex, however, should no longer be the default assumption for future assessments. Future assessments should examine the sensitivity of their output to the assumed level of increasing catchability. The Evaluation panel agrees with the recommendation from the SEDAR 10 Assessment Panel and Review Panel that a special workshop be convened to estimate and quantify changes in catchability over the last 25 to 30 years.

C. The estimation of the number and size composition of discarded fish, as well as the fraction of the discards that die in each assessment.

The response to this issue is addressed in two sections. First, approaches to inferring the size composition of discarded fish is addressed. Second, methods of determining the mortality rate of fish that are discarded is discussed. Greater detail is included in these discussion because these issues were not always addressed clearly and in detail in the original SEDAR 10 and 12 reports. The greater detail herein is the result of considerable effort, including examination of spreadsheets and computer programs prepared during the various SEDAR workshops, evaluation of workshop notes and discussions with workshop participants, and discussions with those who prepared various supporting analyses.

Discard size composition

No information on the size of discarded gag grouper was available for the Atlantic population. Anecdotal accounts suggested that most of the discarded gag were below the size limit, but legal-sized gag grouper were occasionally discarded by recreational fishers owing to the bag limit. To model this, an ad hoc approach was developed where the selectivity curve on discarded fish was assumed to be equal to the difference between the selectivity curve estimated for the landings and a curve with the same shape parameters except shifted to the left by two age classes. The choice to shift by two age classes was not based directly on data, however the length composition of the discards estimated by the assessment model using this approach was qualitatively consistent with the view that most of the discards were discarded because of the size limit. The evaluation panel noted that the Goodyear approach used for red grouper (see below) could have been applied to Atlantic gag, but would still require making ad hoc assumptions about the selectivity on young fish. The evaluation panel found little basis for recommending one approach over the other and noted that it would be difficult to fully evaluate the effect of employing the Goodyear approach in time for the upcoming review in May. Accordingly, no additional runs were recommended.

Information on the size of discarded gag in the Gulf of Mexico recreational fishery was limited to a survey out of the Mote Marine laboratory (covering mostly the region offshore of Tampa) and larger samples from GULFIN and TIP (which included landed fish). The assessment models examined by SEDAR 10 assumed that the size composition of recreational discards was similar to the size composition from these combined data sets, which effectively implies that most recreationally-caught gag were discarded because of the bag limit and a smaller fraction were discarded because they were below the legal size. Anecdotal testimony from several recreational fishers suggested that the majority of discarded fish were below the size limit and that the bag limit was seldom reached. It was pointed out that recreational discards represent a very large fraction of the total estimated removals by the fishery, therefore any substantial changes in the assumptions regarding those discards would likely have important implications for the assessment. In the case of commercial discards, the size composition was assumed to be equal to the difference between the observed distributions before and after the size limit regulations went into effect. It was pointed out that this approach cannot account for variations in discard size composition attributable to variations in year-class strength and mortality. However, it was also noted that the discard size composition was given relatively little weight in the model and that the magnitude of the commercial discards was estimated to be only a small fraction of the total removals, therefore changes in assumptions regarding the commercial discard size composition would be unlikely to have important implications for the assessment. The Evaluation panel recommends two analyses be prepared to be submitted to the upcoming Review Panel: (1) catch per trip analysis to determine how frequently recreational fishers are limited by the current aggregate-grouper bag limit; and (2) an additional assessment run that replaces the current length frequency distributions assumed for discarded fish with the same data truncated at the minimum size limit.

The sources of information on the size of discarded red grouper in the Gulf of Mexico were essentially the same as for Gulf of Mexico gag and the treatment of total discard levels was similar between the two species. The size composition of the discards, however, was not based on data observed prior to the implementation of the minimum size limits (as for commercial gag discards) or on the Mote/GULFIN observations (as for recreational gag discards). Instead, the Goodyear probabilistic approach (SEDAR7-DW56, SEDAR12-AW01) was employed whereby the length frequency of the discards was inferred from the expected distribution of lengths at age below the size limit (based on extensive growth analyses) and assumptions of the relative fraction of younger age classes available to the fishery. This approach is consistent with an assumption that discarding is due to the size limit. It was noted that the approaches used for gag could also have been applied for red grouper. However, sensitivity runs conducted during SEDAR 12 demonstrated that the assessment results were relatively insensitive to the treatment of discard size composition. Accordingly, further exploration of the treatment of red grouper discards does not seem warranted for this review.

Discard mortality rates

Physiological and experimental studies have established a clear relationship between the mortality rate of released gag and the depth at which they were caught.

However, in the case of Atlantic gag there were no data relating the distribution of catches to depth. The discard mortality estimates that were used were based on consultations with fishers relating mortality to the average depth of capture and handling procedures for the various fisheries.

In the case of Gulf gag, the TIP samples included a large number of observations on the distribution of catch at size by depth for the various commercial fisheries and 371 observations of catch at size by depth for the recreational fishery (mostly from charter fishermen). This information was used in SEDAR 10 to construct fishery-specific matrices of discard mortality by size that were multiplied by the fishery-specific estimates of total discards at size to obtain the number of dead discards by size for each fishery. Similar size at depth data were available for Gulf red grouper, but the relationship between discard mortality and depth was unclear because the available studies did not treat the study subjects in the same way for all depths. The SEDAR 12 assessment panel did note that about 20% of the red grouper caught by long liners were discarded already dead and at least 25% of those that were discarded alive would be expected to die based on the results from caging studies conducted at depths typically fished by long liners. Caging studies conducted at depths typically fished by commercial handline and recreational anglers (less than 40m) suggested that fewer than 10% of the animals die directly because of decompression, but testimony during the workshops suggested a few percent may also die owing to depredation by dolphin, barracuda and sharks.

The Evaluation panel did not find additional information that would suggest the approaches to estimating discard mortality should have been more similar across stocks. The panel noted that the effective discard mortality for recreationally caught Gulf gag may decrease with the recommended change in size composition (smaller fish tend to be caught at shallower depths with lower discard mortality rates). The panel believes that the solution to such problems is simple: data are needed to allow determination of discard mortality rates and allocation of discarded fish into size and age classes to eliminate the need for elaborate assumptions.

D. The treatment of the natural mortality rate and, in particular, the method used to scale the Lorenzen curve in each assessment.

Fish are not immortal but quantifying mortality in the absence of fishing is very difficult especially in the S.E. United States where many species have been exploited for more than a century. In previous assessments for these two species, natural mortality was treated as constant across all ages. Based on maximum ages, the natural mortality rate was 0.15 per year in both gag assessments and 0.20 per year for red grouper. However in the recent assessments, analysts thought that it was unlikely that an age-1 grouper had the same natural mortality as did an age-10 grouper and so they developed age-specific natural mortality rates based on a method proposed by Lorenzen (1996). Lorenzen's method assumes that natural mortality is inversely related to length such that older fish have progressively lower natural mortality rates. The Lorenzen age-specific rates were scaled to the cumulative survival to the maximum observed age.

The DW for gag recommended constant natural mortality rates of $M = 0.15$ per year for both coasts (maximum age 31 years in Gulf of Mexico gag, 30 years in South Atlantic gag); however, the AW used $M = 0.14$ to maintain consistency with the Hoenig's regression (1983). The red grouper DW members thought that $M = 0.20$ per year was too high and, because the oldest age in the dataset was 29 years, they recommended using an equivalent of $M = 0.14$ per year also based on Hoenig's regression. The AW stated that the maximum age of 29 was tenuous because it was based on a single fish that had been aged as young as 18 and as old as thirty by different readers, and instead recommended assuming a maximum age of 25 years with an associated $M = 0.167$ per year. For all three assessments, the AW's scaled the Lorenzen curve such that the cumulative mortality from age-0 to the respective maximum ages was equivalent to that for the constant value estimated from Hoenig's regression equation (this approach was also used in SEDAR 04 for snowy grouper and tilefish). The SEDAR 10 CIE reviewers recommended evaluating natural mortality with mark-recapture models but accepted the age-specific natural mortality rates using all ages to scale both the South Atlantic and the Gulf of Mexico gag. However, the SEDAR 12 CIE reviewers for red grouper recommended two modifications to natural mortality: (1) using the original maximum age of 29 years ($M = 0.14$ per year) instead of 25 years ($M = 0.167$ per year), because several red grouper had been estimated at more than 25 years, and (2) using ages 5 and older to scale the curve because the reviewers thought that "natural mortality was underestimated for the older fish".

Scaling the curve with older ages had an impact on the determination of stock status of red grouper. When the Lorenzen curve was scaled with all ages (ages 0-29), the F_{2005}/F_{MSY} was 0.91 implying that red grouper were not undergoing overfishing and SS_{2005}/SS_{MSY} was 1.04 implying that the red grouper stock was just at being rebuilt. When the Lorenzen curve was scaled with ages 5-29, the F_{2005}/F_{MSY} was 0.59 implying that red grouper were not undergoing overfishing and that current fishing mortality was less than OY ($F_{2005}/F_{OY} = 1.17$); SS_{2005}/SS_{MSY} was 1.40 implying that the red grouper stock was rebuilt. It would have been very fruitful for the reviewers to have included the rationale for their choice of ages to include in the scaling of natural mortality.

Since the basis of the natural mortality rates was Hoenig's regression of total mortality on maximum ages, we contacted Hoenig about which ages he included in calculating the catch curves that went into developing his regression. He responded that he included the fully recruited ages so as to avoid the confounding of selectivity with mortality in the younger ages. Determining the specific ages to use in the scaling is problematic for many reasons, not the least of which being that estimated selectivity patterns may take many forms (e.g. dome, flat, reverse logistic) and the age of full selectivity may vary over time. Another complication is that, because Hoenig used unexploited or lightly exploited stocks, his regression ignores that the maximum observed age in an exploited stock depends upon the level of exploitation.

The group supports the SEDAR 12 reviewers recommendation that a technical paper be developed regarding the application of Lorenzen's method to convert conventional constant M to age-dependent M . This exercise does not address the larger question that natural mortality is poorly known. Hence, methods and data need to be developed to allow the

empirical estimation of natural mortality. A possibility noted at SEDAR 12 was to explore using marine protected areas such as the Dry Tortugas Marine Reserve to estimate natural mortality from tagging and catch curves.

E. Recommended reference points (minimum stock size threshold, maximum fishing mortality threshold and optimal yield) and whether those choices are consistent with the goals of the respective Fishery Management Plans and the Magnuson-Stevens Reauthorization Act.

- The evaluation panel considers reference points recommended by the SEDAR 12 review panel for Gulf of Mexico red grouper consistent with the management requirements as stated in the SEDAR 12 Stock Assessment report.
- The evaluation panel considers the SEDAR 10 review panel recommendations for South Atlantic gag grouper references based on MSY concepts consistent with the management requirements as stated in the SEDAR 10 Stock Assessment report.
- The evaluation panel questioned whether the SEDAR 10 review panel provided recommendations for Gulf of Mexico gag grouper management criteria that are consistent with the management requirements as stated in the SEDAR 10 Stock Assessment report. Specifically, the SEDAR 10 review panel did not provide clear recommendations for MFMT and MSST.

The concerns raised by this evaluation panel regarding the SEDAR 10 review panel's recommendations for management parameters for Gulf of Mexico gag grouper center around providing guidance for appropriate advice for MFMT, the reliability of biomass levels chosen from visual examination of stock-recruit plots, and the differences in the stock recruit relationships for Gulf and South Atlantic gag grouper. Although the SEDAR 10 Review panel provided values for exploitation, yield, and spawning stock biomass corresponding to variety of potential management criteria including F_{MAX} , $F_{0.1}$, $F_{20\%SPR}$, and $F_{30\%SPR}$, the panel did not identify any of these as suitable values for MFMT and MSST. The panel suggested that a 'working definition' of MSST could be derived from an empirical evaluation of the stock recruitment plot, but declined to clearly recommend a particular MSST.

Given the lack of clear recommendations for management parameters MFMT and MSST in particular, the evaluation panel agreed to provide additional guidance for consideration.

MFMT

Twenty-year average exploitation of $F=0.36$ was cited by the SEDAR 10 Review Panel as evidence that the current F_{MSY} proxy of $F_{30\%SPR} = 0.25$ is unreasonably stringent:

For the Gulf of Mexico, a MFMT of 0.25 (current value of $F_{30\%SPR}$) is not consistent with the recent dynamics of gag grouper: fishing mortality has been fluctuating around $F = 0.36$ for more than twenty years (1985-2004) and the stock biomass is near its historical maximum. The Review Panel could not provide advice

on target F and biomass reference points, but noted that the stock has apparently increased as a result of good recruitment under estimated fishing mortality rates that have fluctuated around an average value of $F = 0.36$ since the early 1980s. The Review Panel advised that it would be prudent to reduce fishing mortality below $F = 0.36$.

This evaluation panel considers F_{MAX} a reasonable proxy for F_{MSY} and useful for deriving MFMT. F_{MAX} falls below the $F=0.36$ boundary recommended by the SEDAR 10 Review panel. The evaluation panel notes that F_{msy} seldom exceeds F_{MAX} in instances when both parameters are estimated, and further cites the estimates provided in SEDAR 10 for South Atlantic gag that suggested $F_{MAX} = 0.26$ and $F_{MSY} = 0.24$. Finally, the estimate of F_{MAX} is similar to that of $F_{30\% SPR}$, which is the recommended F_{MSY} proxy in the GMFMC generic SFA amendment.

MSST

Although the SEDAR 10 Review panel left MSST blank in the table of proposed management criteria, a ‘working definition’ for MSST of 20 million pounds was suggested based on visual examination of the stock-recruit plot. The stock recruit relationships for both stocks were considered ‘equally uncertain’, though MSY-based reference points were accepted for South Atlantic gag because the stock-recruit model suggested equilibrium recruitment was within the range of observed stock status, whereas the ad hoc stock-recruit model fitted externally from the assessment model for Gulf of Mexico gag suggested equilibrium recruitment was several times greater than that observed over the assessment period.

The panel discussed whether deriving biomass reference points from visual interpretation of stock-recruitment plots, as proposed for Gulf gag grouper, is a valid and objective approach. It was agreed that applying a consistent objective standard is difficult through visual evaluation, and that if a clear visual pattern emerged then it is likely that a stock-recruitment model would fit the data. The evaluation panel agrees that any biomass reference points derived without use of an accepted stock recruitment relationship will be subjective and may just as likely reflect the particular circumstances evaluated as the overall population productivity. The evaluation panel believes that the method used by the SEDAR 10 Review Panel to suggest a level of MSST for gulf gag is not consistent with methods used for other stocks with similar levels of uncertainty.

The evaluation panel considers information provided by estimates of recruitment and the Yield-per-Recruit based references value useful for deriving management parameters. The Panel recommends applying an approach that is consistent with previous solutions to this problem, and estimating a proxy for the biomass at MSY by scaling the expected SSB per recruit at the F_{msy} proxy by the average recent recruitment.

The panel discussed this Term of Reference along broader lines, considering the composition of SEDAR review panels, appropriate expectations of SEDAR review panels, and the products Assessment Workshop panels should provide future Review

Panels. SEDAR Review Panels are composed of independent experts selected for their knowledge in stock assessment. They are intentionally independent of both the assessments being considered and NOAA. Therefore, it is highly unlikely that they are also well-versed in US Federal fisheries management requirements. Furthermore, the charge to the Review Panel is to provide scientifically sound advice and recommendations, and not to consider other ramifications of their recommendations. They are intentionally prohibited from providing specific management advice, and are advised that such recommendations will be provided by the appropriate Council committees that will receive the SEDAR reports. For these reasons, the evaluation panel believes that ensuring compatibility with MS-SFA requirements is beyond the current charge to the Review panels, expectations for such consistency are unreasonable given the restrictions placed upon reviewer selection, and council processes exists to make the transition from review panel recommendations to FMP and MSA compatible regulations.

The Evaluation Panel recommends that such issues can best be resolved by ensuring SEDAR Assessment panels provide ample reference point and stock recruit relationship options for consideration by review panels, and that those proposed references are addressed in the Advisory Report. Those appointed to SEDAR assessment workshop panels are better versed in US management requirements than those appointed to SEDAR review panels, and are more aware of the needs and expectation of the Councils.

2.2. TOR 2.

2. Discuss how consistency in methodology should be balanced against the need to address differences in the data, fisheries and biology of the three stocks in question. Include in this discussion the significance of using different stock assessment algorithms for each stock.

There is a fine balance between model consistency and the need to address differences in data, fisheries, and biology. The frequency of benchmark assessments often occurs over a long enough time period that data collection, fishery prosecution, fishery regulations, biological studies, modeling techniques, etc. have changed. Ignoring these changes may erode confidence in the results from any modeling exercise. Incorporating these changes in new benchmark stock assessments should increase our confidence in the results.

However, consistency can be important for public perception and understanding. It is incumbent that stock assessment scientists avoid the perception that the methods being applied are constantly changing with little justification. In almost every case, the decision to change modeling methodology is done to improve our confidence in the results. The justification for a change in modeling methods needs to be documented and described in more detail in the future. Guidance from the SEDAR process suggests that consistency should not be a major force in determining modeling methodology, but instead using the best model for the situation.

In fisheries we deal with diverse species, fisheries, and stock assessment scientists; this necessitates the need for many different stock assessment models. These models include

catchy names such as SPASM, CASAL, CATCHEM, ASAP, etc. However, at the core of these models, the algorithms employ the same equations that have been used for many years.

There are some fundamental differences between general categories of models. These class of models include productions models, delay-difference models, virtual population analysis (VPA), stock-reduction analysis, statistical catch-at-age models, and length-based models. In SEDAR assessments we have seen several classes of models applied to the same species being assessed. This is often done for continuity reasons or to examine results using different model classes, which have different underlying assumptions.

For South Atlantic gag, Gulf of Mexico gag, and Gulf of Mexico red grouper, several model classes were employed. In all cases, one of the secondary model classes applied was the ASPIC production modeling software. Additionally, the Gulf gag assessment applied a stochastic stock reduction analysis and a VPA model. The VPA model was run for continuity purposes to compare results from the previous stock assessment. The Gulf red grouper assessment applied an additional stochastic stock reduction model, as well.

The secondary models used in the South Atlantic gag, Gulf of Mexico gag, and Gulf of Mexico red grouper assessments all have some limitation or aspect that could be consider “incomplete”. For example, the ASPIC production model does not use any age-structure and therefore does not take advantage of age/length data. The stochastic stock reduction analysis cannot capture age-specific selectivity or distinguish between age/size of landed and discarded fish. The VPA method assumes catches are known without error, an assumption that is rarely met with many snapper-grouper fisheries.

In general, the state-of-the-art model class is the statistical catch-at-age models. For South Atlantic gag, Gulf of Mexico gag, and Gulf of Mexico red grouper, the base model recommended for use in management decisions is a statistical catch-at-age model. The South Atlantic gag model is a statistical catch-at-age model which has been used for almost every snapper-grouper assessment in the South Atlantic. The Gulf of Mexico gag assessment employed the statistical catch-at-age modeling software CASAL, developed in New Zealand and used in many of the international forums. The Gulf of Mexico red grouper assessment applied the statistical catch-at-age modeling software ASAP, available in the NMFS toolbox and which was also used in the previous benchmark assessment. All these models are widely used, and, at their core, these models are almost identical.

The statistical catch-at-age models mentioned above employ the Baranov catch equation, which has been in use since the early 1900’s. These models also share identical methods for estimating annual recruitment values, which are loosely conditioned to follow a stock recruit curve. Another similarity in these models includes the estimates of annual fishing mortality parameters and age-specific selectivity functions. The models are all optimized using Newton-based gradient algorithms and utilize maximum-likelihood objective functions. Given the same data and assumptions, any differences between these models are purely interface differences (i.e. methods of inputting data and outputting results). The choice of which software to use is then an analyst’s preference, having no bearing on the results or conclusions about the stock.

Some stock assessment modeling software packages are more flexible than others. As is often the case with fisheries data, we deal with data types and limitations which require tailoring the model specifications to accommodate the data. Some of the software packages are better at accommodating these unique situations than others.

For the specific cases of South Atlantic gag, Gulf of Mexico gag, and Gulf of Mexico red grouper, the differences in the statistical catch-at-age models are primarily due to differences in available data. Some of the notable differences are with respect to the amount of length, age, and depth data. In the case of South Atlantic gag, the length, age, and depth data are relatively limited compared to Gulf of Mexico gag and Gulf of Mexico red grouper. The South Atlantic gag had no information on depth of fishing from the fisheries, no information on the size/age of discards, and limited annual age sample sizes. These data limitations translated directly into modeling limitations for discards and their mortality rate, and prohibited the use of annual age-length keys. In contrast, the Gulf of Mexico gag had sufficient data to model a depth specific discard mortality rate, model the size of discards, and compute annual age-length keys.

The bottom line is that modeling methods differ primarily depending on the class of model the fall under (see classes mentioned above). Within these classes there are often several software packages for applying the model. In general, these software packages differ only in their user interfaces and their ability to handle minor details, usually resulting from species/fishery specific data limitations.

2.3. TOR 3.

3. Formulate recommendations for any additional analyses, sensitivity runs, or changes to the base cases that need to be made to the Gulf gag, Atlantic gag, and Gulf red grouper assessments based on the reviews of the specific issues addressed in TOR #1 and given the conclusions reached during the discussion of TOR #2.

Recommendations were made regarding the following specific topics identified in TOR 1 and 2. The panel identified both analyses that should be prepared and offered to the Grouper Assessment Review Panel for consideration as well as recommendations for future SEDAR assessments.

1) A. Length of analytical time series:

The AEP was satisfied with the length of time series chosen in all three assessments and made no recommendation for additional runs for the upcoming review workshop.

The Panel suggested future assessments pursue more accurate information on the Cuban catch of red grouper.

1) B. The treatment of the catchability coefficient for fishery-dependent surveys:

The AEP made no recommendations for additional runs for the upcoming review workshop. Runs are available for all 3 assessments based on constant and changing catchability.

The Panel suggested that future SEDAR assessments acknowledge and address catchability changes and examine the sensitivity of their output to the assumed trend in catchability.

The Evaluation panel agrees with the recommendation from the SEDAR 10 Assessment Panel and Review Panel that a special workshop be convened to estimate and quantify changes in catchability over the last 25 to 30 years.

1) C. The estimation of number and size composition of discard fish:

The panel offers no recommendations for changes in the discard assumptions for South Atlantic gag grouper and Gulf of Mexico red grouper.

The panel recommends constructing an analysis of Gulf of Mexico gag grouper based on assigning discarded fish to sizes below the minimum limit.

The panel requests frequency distributions from MRFSS (and other recreational sampling sources) of catch per angler for Gulf of Mexico grouper by year. This should be constructed to include all grouper species that are subject to the 5 fish bag limit.

The panel requests tabulation of the number of samples available for determining depth of recreational discards for Gulf of Mexico gag grouper.

1) D. The treatment of natural mortality:

The evaluation panel does not recommend development of assessment runs based on alternative natural mortality assumptions for evaluation by the Review Panel.

The panel recommends developing a consistent method of estimating age specific natural mortality for use in future SEDAR assessments. The panel requests that the Review panel provide guidance to facilitate such efforts, focusing on determining the appropriate ages to use in scaling and commenting on how such decisions may be affected by changes in selectivity at age over time and differing selectivity patterns.

The panel recommends studies and monitoring that will allow estimation of natural mortality

1) E. Reference Points:

The panel recommends developing reference points for Gulf of Mexico gag grouper based on yield-per-recruit concepts for consideration by the Review Panel. Expected yield and spawning stock biomass should be provided for a range of fishing mortality values for use in determining target and limit references.

For future assessments, the panel recommends that the SEDAR Assessment workshop terms of reference be modified to require that assessment workshops provide management benchmarks that meet FMP and MS-SFA requirements for consideration by subsequent review panels.

The panel recommends clarifying SEDAR Review Panel Terms of Reference to ensure that all reference points submitted by the assessment workshop are addressed and evaluated, and that the Review Panel clearly justify any deviations from Assessment workshop recommendations.

2. Consistency in methods

Due to the similarity of the base models used to assess the stocks in question, the evaluation panel does not recommend any additional analyses for consideration by the Review panel.

The panel recommends development of a glossary of common management and assessment terms and model descriptions for inclusion in all SEDAR assessment reports -- be prepared for consideration by the Review Panel.

The panel recommends that future SEDAR Assessment workshops reports include clear and explicit justification for model changes.

2.4. TOR 4.

This report addressing the recommendations of the evaluation panel was drafted during the SEDAR Grouper Review Evaluation Workshop.