Individual CIE Report

SEDAR 22 Gulf of Mexico Yellowedge Grouper and Gulf of Mexico Tilefish.

Prepared for the Center for Independent Experts

By

Dr Geoff Tingley

December 2010

Cefas Contract
C5192

Table of contents

Executive Summary

- This document is the individual CIE Reviewer report of the SEDAR 22 Gulf of Mexico yellowedge grouper (*Epinephelus flavolimbatus*) and Gulf of Mexico tilefish (*Lopholatilus chamaeleonticeps*) desk-based review.

- This report solely represents the views of the independent reviewer (Dr Geoff Tingley).

- A principal finding is that the technical assessment teams for yellowedge grouper and for tilefish met all of their review terms of reference.

- The assessment for yellowedge grouper is a significant improvement on the previous assessment conducted in 2002. This has been largely as a result of considerable efforts to provide fisheries landings data back to 1975.

- Yellowedge grouper model outputs were suitable for providing basic management advice but there remain concerns about the robustness of this approach with the available data and this needs further development to provide robust management advice.

- The assessment outputs for tilefish were less robust and less useful. This was largely due to the lack of data and uncertainty in the data that were available.

- The reviewer considers that the science basis for the assessments of both yellowedge grouper and tilefish are the best available with some minor adjustments as per the recommendations made.

- The assessment teams clearly recognized the shortcomings of the analyses for both species and made some appropriate recommendations aimed at improving the current approach to the tilefish and tilefish stock assessments through additional research.

- Additional recommendations aimed at improving the current approach to both tilefish and tilefish stock assessments through additional research are made by the reviewer.

Introduction

This desk-based review of the assessments of two species, yellowedge grouper (*Epinephelus flavolimbatus*) and tilefish (*Lopholatilus chamaeleonticeps*), was conducted as part of an independent review of the overall assessment process under the Center for Independent Experts (CIE). This review was delayed for a number of weeks due the unavailability of the final assessment documents. The final submission date for this review report was also changed to reflect the change in the review dates.

All documents have been clearly presented and contain few omissions or typographical errors. One omission was that there was no list of acronyms used in either assessment report which, given the number of acronyms used would have been very useful and would have expedited the work of the reviewer (see for example http://sero.nmfs.noaa.gov/sf/pdfs/Glossary%20of%20Fishery%20Terms.pdf).

The separate section approach taken in the preparation of the assessments reports made reviewing the documents considerably more difficult as re-finding key text or figures was a chore. A report structure with fully consecutive page numbering and no duplication in table or figure numbers should be norm.

The SEDAR staff supporting the review process was fully helpful.

All views expressed in this report are those of this reviewer.

This assessment for yellowedge grouper follows an initial assessment in 2002 using an age-structured production model. No interim assessments have been reported but considerable improvements in input data have been achieved over the intervening time.

This review also addresses the first tilefish assessment reported.

Description of review activities

This review was undertaken by Dr Geoff Tingley between the 23rd November and 10th December 2010 as part of the SEDAR 22 review of the assessments of Gulf of Mexico yellowedge grouper and Gulf of Mexico tilefish.

The supporting documents (see Bibliography, Appendix 1) for the assessment were provided to the reviewer in electronic format adequately in advance of the original review date. The two main assessment documents were provided to the reviewer in electronic format at the start of the rescheduled review date.

The reviewer reviewed the material against the specific terms of reference (ToR) provided by the CIE (see Appendix 2). No additional material was considered necessary by the reviewer and none was used.

Background information relevant to this review is presented in appendices to this review report. These are Appendix 1: a Bibliography of documents and Appendix 2: the CIE Statement of Work (which includes background information and Annexes to cover (i) the

Format and Contents of the CIE Peer Review Report, (ii) Terms of Reference for the Peer Review, for SEDAR 22 Gulf of Mexico yellowedge grouper and Gulf of Mexico tilefish Review.

Comments are provided against the specific terms of reference (ToR) given in Annex 2 of Appendix 2 and are solely those of the reviewer.

Summary of findings

The yellowedge grouper and tilefish assessment teams should be commended for their thorough and professional approach to processing the basic data and developing and applying the models to provide advice to managers. A summary of findings and recommendations from this reviewer for each species are presented below.

The findings of this reviewer are reported within relevant sections, addressing each of the Terms of Reference (ToR) as set out in Annex 2 of Appendix 2.

Overall findings

The principal finding is that all key terms of reference were met by both assessment teams. The approaches to modeling showed some demonstrable areas of weakness and were only able to provide partial information to managers. The choice of sophisticated data hungry models for application in what are essentially data poor fisheries is questioned.

1. Yellowedge grouper

Summary

- The results from the yellowedge grouper assessments using two very different models were fairly consistent but there were a number of concerns about robustness. The model results and the status determination are, however, probably adequate for providing basic management advice.
- The status of the yellowedge grouper in the Gulf of Mexico derived from the results of the Stock Reduction Analysis under the 40:10 rule from the Pacific Fishery Management Council (which is not directly comparable to GMFMC benchmarks) predicts yellowedge grouper in the Gulf of Mexico to be experiencing overfishing (p. overfishing: east 68%, west 94%) and to be overfished (p. overfished: east 87%, west 99%).
- The more complex SS3-based assessment provides guidance on stock status but is sensitive to the proxy for MSY used (SPR 30% or SPR 40%). Thus some base runs yielded outputs suggesting the stock was both overfished and overfishing was occurring through to not overfished with some overfishing.
- The outputs of the two different approaches to modelling are not directly comparable. Thus, there is some uncertainty as to the exact status of the stock but the assessment does provide sufficient information relating to the overfishing status to inform management.
- Further development work on some aspects of the data inputs may yield improved model fits and there are some areas where additional focus on uncertainty may also be helpful.
- The assessment team did a thorough job of preparing the various data sets and in applying the models.

Comments on Individual Terms of Reference: yellowedge grouper

1. *Evaluate the adequacy, appropriateness, and application of data used in the assessment.*

Appropriate biological data for yellowedge grouper were considered, including catch distribution (spatial and depth), stock definition, population genetics, age determination, maturity (age and size), sex transition (age and size), growth and natural mortality (M). An outstanding issue was the possible use of otolith weight as a proxy for age in those individuals not aged. The approaches to separating landings of mixed species helped to reduce uncertainties but did not remove them. Also, the magnitude of discards is assumed to be low but this assumption is based on very little evidence, and some information suggests that discarding is likely to be higher than previously believed (but still unquantified).

Information on catches (i.e. landings and discards), including separation of landings by species, is fairly comprehensive. It highlights the uncertainty of landings in the earlier years of the fishery and considers factors such as misreporting of species and separation of landings by species from collective 'grouper' codes.

While the approaches to separating mixed catches into species and addressing discards of yellowedge grouper from very limited data have been done in a clear and scientifically rigorous way, there remains considerable uncertainty about both the scale of the early landings, and the proportion of overall grouper landings that were yellowedge grouper. It is important that the potential magnitude of these uncertainties be fully tested in assessments.

The yellowedge grouper has been assessed based on an assumed single stock, split into two areas based on fishery distinctions. There is no available genetic data to support or refute this assumption. Morphometric data suggest that there may be some measure of population differentiation. No tagging (migration, movement) studies have been conducted. However, it seems unlikely that there is only one stock over the whole of the range of this species (North Carolina to Brazil) and that understanding the stock structure in US waters may require a wider perspective than US waters alone. The concern of local depletion was considered but was not addressable due to data constraints.

Data collection and analysis of basic biological data (length, age, maturity, sex change, etc.) appear to be well managed.

The temporal and spatial patchiness of age data (e.g. most coming from the last few years and most from Florida waters) needs to be addressed for future assessments.

Aging using otoliths has been validated using radiocarbon (^{14}C) from nuclear contamination provides a greater degree of confidence in the age data than in many other species. The error on aging from otolith reading is, however, quite high and attempts to improve this could be made.

In the absence of discard mortality data, an assumed discard mortality of 100% is not unreasonable, but better understanding of this is possible.

The development of a number of fishery-dependent and fishery-independent indices is a strength of this assessment. Moreover, the use of commercial data to develop abundance indices is to be commended. A minor criticism, and one that should be easily addressed, is that it is difficult to get a relative feel for the two different measures of commercial CPUE presented (i.e. number of hooks for longline vessels and hook hours for the vertical line fishery). It would be helpful to include a comparison of these measures.

The commercial and research longline CPUE indices all show a consistent trend, with an early years' sharp decline followed by a long-term steady improvement.

There are some minor errors that detract from the assessment report but these also affect the confidence of the reader. For example, on page 15, text describes the range of lengths of yellowedge grouper as 100-1,288 mm and the following sentence asserts that the majority of fish were 90-929 mm in length: these statements are clearly not compatible. Moreover, the relevant figure shows that there were practically no fish less than 400 mm (Figure 6a).

A number of statements were made concerning the data that were not really supported, such as on page16 concerning differences in mean length and mean age. These stated differences were unsupported by statistical analysis and, given the sample sizes, subject to significant variation by adding or removing the measurements from single fish (something I tried).

Overall, however, given a fairly difficult set of data to work with, the assessment team have done a professional job of preparing the data for assessment.

2. *Evaluate the adequacy, appropriateness, and application of methods used to assess the stock.*

Model 1 is Stochastic Stock Reduction (SRA) a deterministic age-structured population model with a Beverton-Holt stock-recruitment function. The model produces a single trajectory of biomass over time given MSY and exploitation at MSY. This model does not provide measures of overall model fit or some key parameters (e.g. SSB). There are also a number of inconsistencies in the model set up that makes comparison with Model 2 difficult. For example it uses the Pacific Fisheries Pacific Fisheries Council (PFMC) 40:10 rule for overfishing that has no direct counterpart in the Gulf of Mexico.

Model 2 is Stock Synthesis 3 (SS3), a well tested application and not necessarily inappropriate for developing an assessment for this species. Of particular relevance, SS3 can cope with temporal variation on quantity of data, in this case, data poor periods in the early years with increasing data complexity in later years. SS3 can also cope with missing data without the need for the missing data to be 'artificially created' (e.g. survey data in 2005 affected by hurricane Katrina).

SS3 does assume that landings are precisely known, and where there is uncertainty about landings, as is the case in this fishery, adequate sensitivity runs to the assumptions about landings need to be done.

It is interesting that SS3 failed to adequately fit the three area–based approach to providing spatial scale. It is noted that this is almost certainly due to problems in the data, associated with, for example, differences in the spatial and temporal distributions of the catches and the length and age data.

The reviewer has concerns over how the selection process for some of the basic input parameters (sigmaR, reference age, steepness), not uncommon to SS3 models, appears rather arbitrary and done more to generate a fit rather than being based on any data or scientific methodological source. Also, given the limitations of the quantity and quality of the data, the requirement of SS3 to estimate 73 parameters seems challenging.

The base run of SS3 did not show a good fit to the CPUE indices but provided a better fit to the length and age composition data. The survey indices were also poorly fitted.

The Assessment Team reported that the Beverton and Holt stock-recruit relationship fitted by SS3 did not show a particularly good fit to the data but no alternative to Beverton and Holt appears to have been tried.

Collectively these models appear to be able to generate information on the stock status in relation to biomass and overfishing status.

3. *Evaluate the methods used to estimate population benchmarks and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT, or their proxies); comment on the reliability of the estimated benchmarks.*

The use of the two models to produce population benchmarks and management information is appropriate. However, the applicability of the output values relies on the robustness of the two models, about which there are some questions mostly focussed on the uncertainties.

The SRA results predict yellowedge grouper in the Gulf of Mexico to be experiencing overfishing and there are overfished conditions in the both regions, all with fairy high levels of probability.

The SS3 model outputs are unfortunately difficult to compare with those derived using the SRA model. However, both models pitch the stock one side or the other of the boundary of overfishing/overfished and so provide useful inputs to management.

The reliability of the SS3 model runs has been influenced by uncertainties in the data. These uncertainties, from a number of different sources, are described in the assessment report and listed below. The key uncertainties are those with potentially large impacts and for which there is no or only limited ways to define or reduce the scale of the problem. The principal issue for this species is the magnitude of the fishery, especially in the early years and the inability to accurately define the catches of yellowedge grouper. The efforts expended to allocate landings reported as mixed species have been commendable but uncertainty remains in this area and this flows through to the assessment results. It is also of note that SS3 found it difficult to adequately fit the abundance indices.

4. *Evaluate the adequacy, appropriateness, and application of the methods used to project future population status. Evaluate the adequacy, appropriateness, and application of methods used to characterize uncertainty in estimated parameters. Comment on the degree to which uncertainties are identified and evaluated, and implications of uncertainties stated. Identify any Terms of Reference which are inadequately addressed by the Data or Assessment Workshops*.

Some of the parameters applied to the SRA rendered projections unrealistic. For example, projections were based on an unsustainably high fishing mortality, resulting in stock declines and would not be of use.

SS3 provided a number of reasonably good fits to the data, including length and age composition data.

The data available created the need for a series of decisions about how to handle data inputs. While the decisions made all appeared to be scientifically sound and well justified, there does appear to be something missing in addressing how some of these decisions flow through the assessment and affect the outcome.

On the whole, the approaches to data management and data processing minimised, as far as was reasonably possible, the levels of uncertainty in the input data. However, considerable uncertainty remains in elements of the input data. For example, the magnitude of the early landings, hind-casting how combined landings were split into different species, etc. Different approaches to handling misidentification in statistical areas 6 and 7 (Table on page 11) show that changes in approach can change the input data and results.

Specific examples explored by the sensitivity analyses include:

(i) Assumptions about whether the very large landings in early years were real or not;
(ii) The impact of the reliability of the split of combined reported grouper species into their component species.

The major areas of uncertainty have been addressed by the Assessment Team. Other areas remain to be addressed in future assessments including, for example, ageing errors, and assumptions about discards: survival (assumed small but probably under reported and including mis-reporting), magnitude in the longline fishery, stability over time.
The Assessment Team indicated that some of the MCMC investigations of uncertainty were incomplete and that further work remains to be done.

5. *Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted. Clearly denote research and monitoring needs that could improve the reliability of future assessments*.

The research recommendations in the yellowedge grouper assessment report were all identifying appropriate areas for further investigation but a number of them were rather short on proposed investigative methodologies.

One proposal was to look at genetics. The application of genetics to fisheries management has had mixed success but here is a need to address stock structure and a regional genetics program may be able to address this issue, not only for this species but for others in the same position.

The fishery dependent research recommendations were both good but it is probably worth defining how much observer coverage would be required to provide adequate data from which to construct alternative indices. The additional fishery information obtained from an expanded observer program (on such things as discards) would, however, also be very welcome.

Direct aging of the Johnson otoliths from 1982 and 1983 is a low cost and worthwhile study that will directly feed into future assessments and specifically help to correct the paucity of data in the earlier years of the fishery.

Additional research recommendations have been identified by the reviewer and are presented below in priority order.

Yellowedge grouper Reviewer Recommendations

- In a fishery with multiple data deficiencies, one of the objects of modeling is to identify those data sets that, by their inadequacy or absence, have a disproportionate impact on the outcome of the assessment. This then provides an independent assessment of the prioritization of future research effort aimed at improving the assessment most effectively. More could probably be made of this in defining immediate future research focus.

- Analyze existing data, or collect and analyze new data to confirm that the yellowedge grouper is composed of only a single stock. This could focus on a genetics program aimed at a number of species in the region, as this appears to be a shared problem amongst a number of species.

- Selection bias has occurred in yellowedge grouper age samples, with many more samples in recent years and more from some fishery areas than others (e.g. Florida). Some attempts to obtain a balance of samples from the different areas of (i) the fishery and (ii) the wider stock distribution should be developed and implemented

- While the recreational landings represent a small proportion of the landings it could be worth reviewing the biological data available as recreational fisheries often either target or catch different age or length components of the stock compared to other fisheries. This can be seen in differences between the handline and longline fisheries here. If this is the case then this small part of the fishery may contain useful information about length or age. A basic analysis of length and possibly otolith weight (as a proxy for age) would advise whether this merits further consideration.

- The core input data are in imperial units (lbs) while model processed data (e.g. weight at length or age) are presented in metric units. More importantly the landings/catch data are in lbs and model outputs are in kgs making comparison somewhat difficult. Input and output data should be presented in consistent units.

2. Tilefish

Summary

- The assessment team looking at tilefish used the same two model approach used for yellowedge grouper, a Stock Reduction Analysis (SRA) and a Stock Synthesis 3 (SS3) model.
- Tilefish data were sparser than those for yellowedge grouper but with equal issues concerning periods of inadequate data, catch definition, etc.
- Results from the tilefish assessments using SS3 gave poor fits and unreliable information for management purposes.
- The SRA approach did provide information that is useful in formulating management advice but the lack of a crosscheck with the SS3 results does weaken this compared to the yellowedge grouper assessment.
- Further development work on key aspects of the data inputs may yield improved model fits.
- The assessment team did a thorough job of preparing the various data sets and in applying the model.
- That this fishery is clearly data poor (Section III, page 55) does raise questions as to the applicability of selecting a multi-parameter, a data hungry assessment framework such as SS3.

Comments on Individual Terms of Reference: tilefish

1. *Evaluate the adequacy, appropriateness, and application of data used in the assessment.*

Appropriate biological data for tilefish were considered, including catch distribution (spatial and depth), stock definition, population genetics, age determination, maturity (age and size), sex transition (age and size), growth and natural mortality (M).

Information on catches (i.e. landings and discards) including separation of landings by species are fairly comprehensive. They highlight the uncertainty of landings in the earlier years of the fishery and consider factors such as misreporting of species and separation of landings by species from collective 'grouper' codes. The approaches to separating landings of mixed species helped to reduce uncertainties but did not remove then. Also, the magnitude of discards is assumed to be low but this assumption is based on very little evidence and some information suggests that discarding is likely to be higher than previously believed (but still unquantified).

While the approaches to separating mixed catches into species and addressing discards of tilefish from very limited data has been done in a clear and scientifically rigorous way, there remains considerable uncertainty about both the scale of the early landings, and the proportion of overall grouper landings that were tilefish. It is important that the potential magnitude of these uncertainties be fully tested in assessments.

One element that could influence how a fish is reported is the amount of quota remaining. Thus, as quota is progressively used up there may be a bias in what quota group a fish is

reported. While this may not be likely, eliminating it as a possible bias in the input data might be sensible. It is not clear whether this has been attempted or not.

The tilefish has been assessed based on an assumed single stock, originally intended to be split into three sub-areas but due the outcome of reallocation of landings to species, was assessed as two areas only. These areas were the eastern and western Gulf. There are some genetics data to support a single stock basis for this species, but there is only weak evidence at best. Limited tagging (migration, movement) studies have also been conducted, and show the species to be sedentary. There remains uncertainty as to whether there is only one stock over the whole of the range of this species. The concern that local depletions may be occurring was considered but was not addressable due to data constraints.

Data collection and analysis of basic biological data (length, age, maturity, sex change, etc.) appears to be well managed.

Aging using otoliths, validated using radiocarbon (^{14}C) from nuclear contamination, provides a greater degree of confidence in the age data than in many other species but problems remain, especially for older fish. The error on aging from otolith reading is quite high and attempts to reduce age errors could be made.

In the absence of discard mortality data, an assumed discard mortality of 100% is not unreasonable, but better understanding of this is also possible.

As with the yellowedge grouper assessment, the development of a number of fishery-dependent and fishery-independent indices is a strength of this assessment of tilefish. Moreover, the use of commercial data to develop abundance indices is to be commended. A minor criticism, and one that should be easily addressed, is that it is difficult to get a relative feel for the two different measures of commercial CPUE presented (i.e. number of hooks for longline vessels and hook hours for the vertical line fishery). It would be helpful to include a comparison of these measures.

The commercial and research longline CPUE indices all show a shallow but consistent trend over the time period.

Given a fairly sparse set of data to work with, the assessment team have done a professional job of preparing data for assessment. One focus for future assessments is the discard estimation about which little is known and the hind-casting has no variation applied to it.

2. *Evaluate the adequacy, appropriateness, and application of methods used to assess the stock.*

Model 1 is Stochastic Stock Reduction (SRA), a deterministic age structured population model with a Beverton-Holt stock-recruitment function. The model produces a single trajectory of biomass over time given MSY and exploitation at MSY. This model does not provide measures of overall model fit or some key parameters (e.g. SSB). There are also a number of inconsistencies in the model set up that makes comparison with SS3 outputs difficult. For example it uses the PFMC 40:10 rule for overfishing that has no direct counterpart in the Gulf of Mexico.

Model 2 uses Stock Synthesis 3 (SS3), which is a well tested application and not necessarily inappropriate for developing an assessment for this species. Of particular relevance, SS3 can cope with temporal variation on quantity of data, in this case data poor periods in the early years with increasing data complexity in later years. SS3 can also cope with missing data without the need for the missing data to be 'artificially created' (e.g. survey data affected by hurricane Katrina in 2005). However, SS3 is a sophisticated, data-hungry approach and this is clearly a fishery that could be described as data poor, and thus the appropriateness of selecting SS3 as the assessment vehicle must be raised.

SS3 does assume that landings are precisely known, and where there is uncertainty about landing, as is the case in this fishery, adequate sensitivity runs to the assumptions about landings need to be done. This does not sit well with what we believe we know about the landings data.

Spatial patterns of variable biology were taken into account in the input data (e.g. two growth models for the two regions). It is interesting that SS3 failed to adequately fit the two area, two growth model-based approach to providing spatial scale. It is noted that this is almost certainly due to problems in the data associated with, for example, differences in the spatial and temporal distributions of the catches and the length and age data.

The reviewer has concerns over how the selection process for some of the basic input parameters (sigmaR, reference age, steepness), not uncommon to SS3 models, appears rather arbitrary and done more to generate a fit rather than being based on any data or scientific methodological source. Also, given the limitations of the quantity and quality of the data, the requirement of SS3 to estimate 73 parameters seems challenging.

The base run of SS3 did not show a good fit to the CPUE indices, the length composition data or the age composition data. The survey indices were also poorly fitted. This overall poorness of fit must raise some doubts as to what the model is doing and its suitability.

The Assessment Team reported that the Beverton and Holt S-R relationship fitted by SS3 did not show a particularly good fit, but no alternative to BH appears to have been tried.

Collectively these models appear to be appropriate and able to generate information on the stock status in relation to biomass and overfishing status but may not be robust.

3. *Evaluate the methods used to estimate population benchmarks and management parameters (e.g., MSY, Fmsy, Bmsy, MSST, MFMT, or their proxies); comment on the reliability of the estimated benchmarks.*

The use of the two models to produce population benchmarks and management information is appropriate. However, the applicability of the output values relies on the robustness of the two models, about which there are some questions, mostly focussed on the uncertainties and poorness of fit to the data.

The SRA results predict tilefish in the Gulf of Mexico to be experiencing overfishing in one region and overfished conditions in the other region, but with low probability.

Specification issues prevented the Assessment Team from having the confidence to use the SS3 assessments to generate benchmarks. Specifications of the reference age for M, and M at the reference age were highlighted by the Assessment Team (This refers to a scaling factor derived following Lorenzen which scales M according to the growth curve; so the actual scaling of M varies between males and females and according to the growth rates in the different regions and is in input into SS3).

The reliability of the SS3 model runs has been influenced by uncertainties in the data. These uncertainties, from a number of different sources, are described in the assessment report and listed below. The key uncertainties are those with potentially large impacts and for which there is no or only limited ways to define or reduce the scale of the problem. The principal issue for this species is the magnitude of the fishery, especially in the early years and the inability to accurately define the catches of tilefish. The efforts expended to allocate landings reported as mixed species have been commendable but uncertainty remains in this area and this flows through to the assessment results. It is also of note that SS3 found it difficult to adequately fit the abundance indices.

The sensitivity runs were used to assess stock status and at SPR 30% all runs indicated that the stock was undergoing overfishing but was not overfished.

4. *Evaluate the adequacy, appropriateness, and application of the methods used to project future population status. Evaluate the adequacy, appropriateness, and application of methods used to characterize uncertainty in estimated parameters. Comment on the degree to which uncertainties are identified and evaluated, and implications of uncertainties stated. Identify any Terms of Reference which are inadequately addressed by the Data or Assessment Workshops*.

Some of the parameters applied for the SRA rendered projections unrealistic. For example, projections were based on an unsustainably high fishing mortality, resulting in stock declines and would not be of use.

SS3 provided only relatively poor fits to the data, but in a consistent manner on a number of different runs.

The data available created the need for a series of decisions about how to handle data inputs. While the decisions made all appeared to be scientifically sound and well justified, there does appear to be something lacking in addressing how some of these decisions feed through to the assessment outputs.

On the whole, the approaches to data management and data processing minimised, as far as was reasonable possible, the levels of uncertainty in the input data. However, considerable uncertainty remains in elements of the input data, for example, the magnitude of the early landings, hind-casting how combined landings were split into different species, etc. Different approaches to handling misidentification in statistical areas 6 and 7 (Table on page 11) show that changes in approach can change the input data and results.

Specific examples explored by the sensitivity analyses include:
(i) Assumptions about whether the very large early years landings were real or not;

(ii) The impact of the reliability of the split of combined reported tilefish species into their component species.

The major areas of uncertainty have been addressed by the Assessment Team. Other areas remain to be addressed in future assessments including, for example, ageing errors, and assumptions about discards (assumed small but probably under reported), discard survival, magnitude of the longline fishery, stability over time, etc.

5. *Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted. Clearly denote research and monitoring needs that could improve the reliability of future assessments*.

The research recommendations in the tilefish assessment report were all identifying appropriate areas for further investigation but a number of them were rather short on proposed investigative methodologies.

One proposal to look at stock structure should be part of a larger, probably genetics program to look at regional stock structure in a number of similar species.

The fishery dependent research recommendations were both good but it is probably worth defining how much observer coverage would be required to provide adequate data from which to construct alternative indices. The additional fishery information obtained from an expanded observer program (on such things as discards) would, however, also be welcome.

Additional research recommendations have been identified by the reviewer and are presented below in priority order.

Tilefish Reviewer Recommendations

- In a fishery with multiple data deficiencies, one of the objects of modeling is to identify those data sets that, by their inadequacy or absence, have a disproportionate impact on the outcome of the assessment. This then provides an independent assessment of the prioritization of future research effort aimed at improving the assessment most effectively. More could probably be made of this in defining immediate future research focus.

- Analyze existing data, or collect and analyze new data to confirm that the tilefish is composed of only a single stock. This could focus on a genetics program aimed at a number of species in the region, as this appears to be a shared problem amongst a number of species.

- Review the information about distribution of tilefish age in time and geographical area with a view to obtain better quality data going forward (i.e. attempts to obtain a balance of samples from the different areas of (i) the fishery and (ii) the wider stock distribution should be developed and implemented).

- Evaluating whether the amount of remaining quota influences how landings are reported by species should be considered.

- While the recreational landings represent a small proportion of the landings it could be worth reviewing the biological data available as recreational fisheries often either target or catch different age or length components of the stock compared to other fisheries. If this is the case then this small part of the fishery may contain useful information about length or age. A basic analysis of length and possibly otolith weight (as a proxy for age) would advise whether this merits further consideration.

- The core input data are in imperial units (lbs) while processes data (e.g. weight at length or age) are presented in metric units. More importantly the landings/catch data are in lbs and model outputs are in kgs making comparison somewhat difficult. Input and output data should be presented in consistent units.

Conclusions

Having reviewed all the material, the reviewer is confident in the robustness of the outcome of the review. Assessments for both species are based on high quality science and draw on most of the data that are available. While some management advice can be generated, the degree to which the assessments were fit for purpose is debatable. The application of data hungry models such as SS3 in data poor fisheries needs further justification. Future research effort might be better spent using existing and developing specific data poor approaches for these fisheries.

Despite criticisms in this review as to the choice of models used, the assessment teams are to be congratulated on the very thorough approaches taken to address serious shortcomings in key datasets and in the application of the selected models.

Where improvements can be made these have been addressed through making appropriate recommendations.

Appendix 1: Bibliography

SEDAR 22
Gulf of Mexico Yellowedge Grouper and Tilefish
Workshop Document List

Document #	Title	Authors	Working Group
Documents Prepared for the Data Workshop			
SEDAR22-DW-01	Golden tilefish (*Lopholatilus chamaeleonticeps*) age, growth, and reproduction from the northeastern Gulf of Mexico: 1985,1997-2009	Linda Lombardi, Gary Fitzhugh, Hope Lyon	Life History
SEDAR22-DW-02	Commercial longline vessel standardized catch rates of yellowedge grouper in the Gulf of Mexico	Neil Baertlein and Kevin McCarthy	Indices
SEDAR22-DW-03	Golden tilefish and blueline tilefish standardized catch rates from commercial longline vessels in the Gulf of Mexico	Kevin McCarthy	Indices
SEDAR22-DW-04	Discards of yellowedge grouper, golden tilefish, and blueline tilefish from commercial fishing vessels in the Gulf of Mexico	Kevin McCarthy	Catch Statistics
SEDAR22-DW-05	Explorations of habitat associations of yellowedge grouper and golden tilefish	John F Walter, Melissa Cook, Brian Linton, Linda Lombardi, and John A. Quinlan	Life History
SEDAR22-DW-06	Abundance Indices of subadult Yellowedge Grouper, *Epinephelus flavolimbatus*, Collected in Summer and Fall Groundfish Surveys in the northern Gulf of Mexico	Adam G. Pollack and G. Walter Ingram, Jr.	Indices
SEDAR22-DW-07	Abundance Indices of Yellowedge Grouper and Golden Tilefish Collected in NMFS Bottom Longline Surveys in the northern Gulf of Mexico	G. Walter Ingram, Jr. and Adam G. Pollack	Indices
SEDAR22-DW-08	Yellowedge grouper (*Epinephelus flavolimbatus*) age, growth and reproduction from the northern Gulf of Mexico	Melissa Cook and Michael Hendon	Life History
SEDAR22-DW-09	Observed Length frequency distributions and otolith sampling issues for yellowedge groupers caught in the Gulf of Mexico from 1984 to 2009.	Ching-Ping Chih	Life History/ Catch Statistics
SEDAR22-DW-10	Observed Length frequency distributions and otolith sampling issues for tile fish caught in the Gulf of Mexico from 1984 to 2009	Ching-Ping Chih	Life History/ Catch Statistics
SEDAR22-DW-11	Length frequency distributions for blue line tile fish caught in the Gulf of Mexico from 1984 to 2009	Ching-Ping Chih	Life History/ Catch Statistics
SEDAR22-DW-12	Estimation of species misidentification in the commercial landing data of tile fish in the Gulf of Mexico from 1984 to 2009	Ching-Ping Chih	Catch Statistics
SEDAR22-DW-13	Estimation of species misidentification in the commercial landing data of yellowedge groupers in the Gulf of Mexico from 1984 to	Ching-Ping Chih	Catch Statistics

	2009		
SEDAR22-DW-14	Evidence of hermaphroditism in Golden Tilefish (*Lopholatilus chamaeleonticeps*) in the Gulf of Mexico	Hope Lyon	Life History
SEDAR22-DW-15	Recreational Survey Data for Yellowedge Grouper, Tilefish (golden), and Blueline Tilefish in the Gulf of Mexico	Vivian M. Matter	Catch Statistics
SEDAR22-DW-16	Estimated Recreational Catch in Weight: Method for Filling in Missing Weight Estimates from the Recreational Surveys	Vivian M. Matter	Catch Statistics
SEDAR22-DW-17	Commercial Landings of Yellowedge Grouper, Golden Tilefish, and Blueline Tilefish from the Gulf of Mexico region	Refik Orhun	Catch Statistics

Document #	Title	Authors
Documents Prepared for the Assessment Process		
SEDAR22-AP-01	United States Commercial Longline Vessel Standardized Catch Rates of Golden and Blueline Tilefish in the Gulf of Mexico, 1992-2009: Revised	Kevin McCarthy
SEDAR22-AP-02	United States Commercial Longline Vessel Standardized Catch Rates of Yellowedge Grouper (*Epinephelus flavolimbatu*s) for Three Regions in the Gulf of Mexico, 1991-2009	Neil Baertlein and Kevin McCarthy
Documents Prepared for the Review Workshop		
SEDAR22-RW-01		
Final Stock Assessment Reports		
SEDAR19-SAR1	Yellowedge Grouper	
SEDAR19-SAR2	Golden Tilefish	
Reference Documents		
SEDAR22-RD01	Lead-radium dating of golden tilefish (*Lopholatilus chamaeleonticeps*)	Allen Andrew
SEDAR22-RD02	Status of the yellowedge grouper fishery in the Gulf of Mexico	Shannon L. Cass-Calay and Melissa Bahnick
SEDAR22-RD03	Yellowedge grouper (*Epinephelus flavolimbatus*) and golden tilefish (*Lopholatilus chamaeleonticeps*) distributions, habitat preferences and available biological samples	Melissa Cook and Linda Lombardi-Carlson
SEDAR22-RD04	Validation of yellowedge grouper, *Epinephelus flavolimbatus*, age using nuclear bomb-produced radiocarbon	Melissa Cook & Gary R. Fitzhugh & James S. Franks
SEDAR22-RD05	Population dynamics structure, and per –recruit analyses of yellowedge grouper, *Epinephelus flavolimbatus* from the northern Gulf of Mexico	Melissa Cook
SEDAR22-RD06	Reproduction of yellowedge grouper *Epinephelus flavolimbatus*, from the eastern Gulf of Mexico	Bullock, L. H., M. F. Godcharles and R. E. Crabtree
SEDAR22-RD07	Burrow utilization by yellowedge grouper, *Epinephelus flavolimbatus*, in the northwestern Gulf of Mexico	Jones, R. S., E. J. Gutherz, W. R. Nelson and G. C. Matlock

SEDAR22-RD08	Age and growth of the yellowedge grouper, *Epinephelus flavolimbatus*, and the yellowmouth grouper, *Mycteroperca interstitialis*, off Trinidad and Tobago	Manickchand-Heileman, S. C. and D. A. T. Phillip
SEDAR22-RD09	A descriptive survey of the bottom longline fishery in the Gulf of Mexico	Prytherch, H. F.
SEDAR22-RD10	Comparison of Two Techniques for Estimating Tilefish, Yellowedge Grouper, and Other Deepwater Fish Populations	Matlock, Gary C., Walter R. Nelson, Robert S. Jones, Albert W. Green, Terry J. Cody, Elmer Gutherz, and Jeff Doerzbacher
SEDAR22-RD11	Deep-water sinkholes and biotherms of South Florida and the Pourtales Terrace – Habitat and Fauna	John K. Reed, Shirley A. Pomponi, Doug Weaver, Charles K. Paull, and Amy E. Wright
SEDAR22-RD12	Tilefishes of the genus *Caulolatilus* construct burrows in the sea floor	K.W. Able, D.C. Twichell, C.B. Grimes, and R.S. Jones
SEDAR22-RD13	Spawning Locations for Atlantic Reef Fishes off the Southeastern U.S.	George R. Sedberry, O. Pashuk, D.M. Wyanski, J.A. Stephen, and P. Weinbach
SEDAR22-RD14	Trends in tilefish distribution and relative abundance off South Carolina and Georgia	Charles A. Barnes and Bruce W. Stender
SEDAR22-RD15	Age, growth, and reproductive biology of blueline tilefish along the Southeastern coast of the United States, 1982-1999	Patrick J. Harris, David M. Wyanski, and Paulette T. Powers Mikell
SEDAR22-RD16	Temporal and spatial variation in habitat characteristics of tilefish (*Lopholatilus chamaeleonticeps*) off the east coast of Florida	Kenneth W. Able, Churchill B. Grimes, Robert S. Jones and David C. Twichell
SEDAR22-RD17	The Complex Life History of Tilefish *Lopholatilus chamaeleonticeps* and Vulnerability to Exploitation	Churchill B. Grimes and Stephen C. Turner
SEDAR22-RD18	The fishery for tilefish, *Lopholatilus chamaeleonticeps*, off South Carolina and Georgia	Bob Low, Glenn Ulrich, and Frank Blum
SEDAR22-RD19	Tilefish off South Carolina and Georgia	R.A. Low, Jr., G.F. Ulrich, and F. Blum
SEDAR22-RD20	Spawner-recruit relationships of demersal marine fishes: Prior distribution of steepness for possible use in SEDAR stock assessments	SEDAR 24−AW−06 - Sustainable Fisheries Branch

Appendix 2: Statement of Work

Attachment A: Statement of Work for Dr. Geoff Tingley (CEFAS)

External Independent Peer Review by the Center for Independent Experts

SEDAR 22
Gulf of Mexico yellowedge grouper and tilefish pre-review workshop review

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

Project Description: SEDAR 22 will be a compilation of data, a benchmark assessment of the stock, and an assessment review for conducted for Gulf of Mexico yellowedge grouper and tilefish. The desk review will provide an independent peer review of SEDAR stock assessments prior to the panel Review Workshop. The term review is applied broadly, as the reviewer may suggest additional analyses, error corrections and sensitivity runs of the assessment models provided by the assessment workshop panel. The stocks assessed through SEDAR 22 are within the jurisdiction of the Gulf of Mexico Fishery Management Council and the states of Texas, Louisiana, Mississippi, Alabama, and Florida. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**.

Requirements for CIE Reviewers: One CIE reviewers shall conduct an impartial and independent peer review in accordance with the SoW and ToRs herein. CIE reviewers shall have working knowledge and recent experience in the application of stock assessment, statistics, fisheries science, and marine biology sufficient to complete the primary task of reviewing the technical details of the methods used for the assessment. Each CIE reviewer's duties shall not exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Desk Review (no travel needed)

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

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

Pre-review Background Documents: Two weeks before the peer review period, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the CIE reviewer the necessary background information and reports for the peer review. The actual report for review will be provided no later than the date the review period is scheduled to begin in accordance with the schedule. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the CIE Lead Coordinator on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance to the SoW scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.

Contract Deliverables - Independent CIE Peer Review Report: The CIE reviewer shall complete an independent peer review report in accordance with the SoW. The CIE reviewer shall complete the independent peer review according to required format and content as described in Annex 1. The CIE reviewer shall complete the independent peer review addressing each ToR as described in Annex 2.

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

1) Review background materials and supporting documentation for the assessment and assessment report.
2) Conduct an independent peer review of the assessment report, in accordance with the ToRs (**Annex 2**), between 23 November 2010 and 3 December 2010.
3) No later than 3 December 2010, submit an independent peer review report addressed to the "Center for Independent Experts," and sent to Mr. Manoj Shivlani, CIE Lead Coordinator, via email to shivlanim@bellsouth.net, and David Sampson CIE Regional Coordinator, via email to david.sampson@oregonstate.edu. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

Schedule of Milestones and Deliverables: CIE shall complete the tasks and deliverables described in this SoW in accordance with the following schedule.

15 August 2010	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
23 November 2010	NMFS Project Contact sends the CIE Reviewer the pre-review documents
23 November - 3 December 2010	Reviewer conducts an independent peer review
3 December 2010	CIE reviewer submits draft CIE independent peer review report to the CIE Lead Coordinator and CIE Regional Coordinator
15 December 2010	CIE submits CIE independent peer review reports to the COTR
17 December 2010	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

Modifications to the Statement of Work: Requests to modify this SoW must be approved by the Contracting Officer at least 15 working days prior to making any permanent substitutions. The Contracting Officer will notify the COTR within 10 working days after receipt of all required information of the decision on substitutions. The COTR can approve changes to the milestone dates, list of pre-review documents, and ToRs within the SoW as long as the role and ability of the CIE reviewers to complete the deliverable in accordance with the SoW is not adversely impacted. The SoW and ToRs shall not be changed once the peer review has begun.

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

Applicable Performance Standards: The contract is successfully completed when the COTR provides final approval of the contract deliverables. The acceptance of the contract deliverables shall be based on three performance standards:
(1) Each CIE report shall completed with the format and content in accordance with **Annex 1**,
(2) Each CIE report shall address each ToR as specified in **Annex 2**,
(3) The CIE reports shall be delivered in a timely manner as specified in the schedule of milestones and deliverables.

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

Key Personnel:

William Michaels, Contracting Officer's Technical Representative (COTR)
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-713-2363 ext 136

Manoj Shivlani, CIE Lead Coordinator
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Julie Neer, SEDAR Coordinator, NMFS Project Contact
South Atlantic Fishery Management Council
4055 Faber Place, Suite 201
North Charleston, SC 29405
Julie.Neer@SAFMC.net Phone: 843-571-4366

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

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

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

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

   b. Reviewers should discuss their independent peer review on each ToR.

   c. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed,.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE Statement of Work.

Annex 2: Terms of Reference for the Peer Review

SEDAR 22 Gulf of Mexico Yellowedge grouper and Tilefish Pre-Review Workshop review

1. Evaluate the adequacy, appropriateness, and application of data used in the assessment.
2. Evaluate the adequacy, appropriateness, and application of methods used to assess the stock.
3. Evaluate the methods used to estimate population benchmarks and management parameters (*e.g., MSY, Fmsy, Bmsy, MSST, MFMT, or their proxies*); comment on the reliability of the estimated benchmarks.
4. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status; Evaluate the adequacy, appropriateness, and application of methods used to characterize uncertainty in estimated parameters. Comment on the degree to which uncertainties are identified and evaluated, and implications of uncertainties stated. Identify any Terms of Reference which are inadequately addressed by the Data or Assessment Workshops.
5. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted. Clearly denote research and monitoring needs that could improve the reliability of future assessments.
6. Prepare a Peer Review Report documenting findings pertaining to these Terms of Reference.