

**Center for Independent Experts (CIE) Independent Peer Review of SEDAR 42
Gulf of Mexico Red Grouper**

Prepared for the Center for Independent Experts

Matthew. D. Cieri

August 2015

Executive Summary

The Review Workshop for SEDAR 42 for Red Grouper was held ^{from} July 13-16, 2015, in Miami, Florida. At this Workshop, the Assessment Team made many presentations on the materials provided two weeks prior to the Review Panel. This was a new assessment using a new model type (SS3) which was applied to a data stream after much care and consideration.

After presentations the first day, many deficiencies were found. These included base model fits to discards and indices. Because of the poor fits, the Panel made a number of suggestions for alternate runs, and discussed the implications of the base and alternate runs at length. In the end, the model was simply not fitting the data well. Attempts to force fit, or any work around suggested during the workshop, resulted in other unexplained degradation of the diagnostics. On the final day of the Workshop, the Panel made a few more requests, with the assumption that the results and diagnostics would be forthcoming in the next week.

After the results were emailed, it was clear to me that the base and alternate runs still had many difficulties. After careful consideration and reflection, I decided that the current formulation, and any alternate runs suggested, simply were not feasible for providing management advice. As such, projections and reference points were also deemed useable.

I've made a number of recommendations and suggestions to help both the modeling as well as the process. My hope is that some of these will be useful in the future.

Background

The Review Workshop for SEDAR 42 was held in Miami from July 13-16, 2015. At this meeting a Panel consisting of SSC members and CIE reviewers were tasked with reviewing and providing recommendations to the assessment process documentation. This Workshop was a culmination of months of preparation and analysis by the Assessment Team and the Assessment Workshop. The end goal was an analysis that could be used to provide management advice on ABC/OY for future specifications for this and removal levels for this stock.

Description of Role

In keeping with the Terms of Reference (or TORs), the reviewer's role is to examine the methods used to assess the Gulf of Mexico Red Grouper stock including data inputs, application of methods, and assumptions. After doing so, the reviewer should critique the application of these methods and analysis to each of the stocks under consideration. Where applicable the reviewer should make comments, ask appropriate questions, and where deficiencies are found, highlight those, make recommendations and provide conclusions.

Review Workshop Technical Reviewer Responsibilities

1. Approximately two weeks prior to the meeting, the reviewers shall be provided with the stock assessment reports, associated supporting documents, and review workshop instructions including the Terms of Reference. Reviewers shall read these documents to gain an in-depth understanding of the

stock assessment, the resources and information considered in the assessment, and their responsibilities as reviewers.

2. During the Review Workshop, reviewers shall participate in panel discussions on assessment methods, data, validity, results, recommendations, and conclusions as guided by the Terms of Reference. The reviewers shall develop a Review Panel Report for each assessment reviewed. Reviewers may be asked to serve as an assessment leader during the review to facilitate preparing first drafts of review reports.

3. Following the Review Workshop, reviewers shall work with the chair to complete and review the Review Workshop Reports. Reports shall be completed, reviewed by all panelists, and comments submitted to the Chair within two weeks of the conclusion of the workshop.

4. Additional obligation of CIE-appointed reviewers: Following the Review Workshop, each reviewer appointed by the CIE shall prepare an individual CIE Reviewer Report and submit it in accordance with specifications provided in the Statement of Work.

Summary of Findings by TOR

TOR 1: Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions, and consider the following:

- Are data decisions made by the DW and AW sound and robust?
- Are data uncertainties acknowledged, reported, and within normal or expected levels?
- Are data applied properly within the assessment model?
- Are input data series reliable and sufficient to support the assessment approach and findings?

General Comments

Overall, key descriptions on the data elements were not in the final report prepared for the review, despite their importance in how the base model was set up. While many of these can be found in the data workshop report, it was thought that inclusion of these in the review report would be clearer. Further, there was little discussion or justification in the review report as to why the Assessment Working group chose the current model, fleet structure, start year, and other key elements. These need to be included in the review report so that the train of logic can be followed from Data Workshop, through Assessment Workshop, to Review Workshop.

A wide array of input data were used in this assessment. These included:

Life History

- Age and growth
- Fecundity
- Natural mortality
- Maturity
- Sex transition

Landings

- Handline: 1986-2013
- Longline: 1986-2013

- Trap: 1986-2006
- Charter /Private: 1986-2013
- Headboat: 1986-2013

Discards

- Handline: 2007-2013 (observer program)*
- Longline: 2007-2013 (observer program)*
- Trap: Charter/Headboat/Private: 1986-2013

Length composition of discards by fleet

- Handline: 2006-2013 (observer program)
- Longline: 2006-2013 (observer program)
- Charter: 2010-2013 (observer program)
- Headboat: 2005-2007, 2009-2013 (observer program)

Age composition of retained catch

- Handline: 1991-2013
- Longline: 1991-2013
- Charter/Headboat/Private: 1991-2013

Abundance indices

- Handline: 1993-2009
- Longline: 1993-2009
- Charter/Private: 1986-2013
- Headboat: 1986-2013
- Combined video survey: 1993-1997, 2002, 2004-2013
- SEAMAP summer groundfish survey: 2009-2013
- NMFS bottom longline survey: 2001, 2003-2013

The Panel examined the available information on Life History. I noted that the maturity schedule, the transition from females to males, fecundity, and growth/size at age were all static across years. While there is little data to support alternatives, it is likely that Life History traits change based on environmental factors and/or may have density dependence. In particular, the ratio of males to females, the transition from females to males, and the fecundity could have large impacts on reference points and advice on sustainable harvest levels.

I had a number of concerns with the landings and discard data series, as well as the fishery dependent sampling. While commercial landings were well known, recreational landings were not as precise given the change from MRFSS to MRIP. More importantly, the level of removals as dead discards in this fishery is high (Figure 1). Similarly, I in particular, had concerns with the length and age information because of low sampling for the discarded and landed fractions of removals, particularly when parsed by fleets. Given the low sampling at-sea, we recommended down-weighting the discard length compositions by increasing the CV. However, to get the model to perform, required increasing the weight of total discards.

We also examined the available fishery dependent and independent abundance indices. I noted that these were restricted geographically to the central part of this stock, and as such would not be as sensitive to changing stock levels if density remained constant, but would be sensitive if the range expanded or contracted in response to stock growth or decline. I suggested that fishery dependent estimates (as CPUE) were prone to hyper-stability, as well as large uncertainties in standardization given changing management measures.

Are data decisions made by the DW and AW sound and robust?

Despite the above concerns, I believe both Data Workshop and Assessment Workshop decisions were sound and robust given the lack of sampling and other uncertainties. However, we as a Panel suggested a number of alternatives and sensitivity runs to the base model to explore how sensitive the model outputs were to input data uncertainties (see below). Because of model fit issues, we decided to go with a different configuration as the base model run.

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

Overall, the Panel concluded that the data uncertainties were acknowledged, reported, and within expected levels given low fishery dependent sampling.

Are data applied properly within the assessment model?

Overall, we as a group found that the application of the data within the model was properly done. However, we also did recommend changes to the model structure as well as sensitivity analysis to examine model behavior in light of data variability and uncertainty (below).

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

The Panel noted the lack of fishery dependent sampling, the restricted geographic range of the surveys, the high discard and low at-sea sampling, and the socio-biology of the species spawning. This led the Panel to recommend a number of model runs to test the sensitivity of the model results, and to examine model behavior. These runs included: modeling the recreational fleet as one fleet, as was done in the previous benchmark, starting the model at 1993 to better encapsulate the start of both the fishery dependent and independent sampling, relaxing the fit to the landings in light of the transition from MRFSS to MRIP, removing the fishery dependent abundance indices given issues with hyper-stability, increasing and decreasing the fit to the discards and the length distributions of discards given low sampling. For these runs resulting residuals and selectivity's were examined to observe model behavior.

After examination, we decided to recommend changes to the base model. These included (RW1): starting the model in 1993 as opposed to 1986, using one recreational fleet, setting steepness to 0.99 and using median recruitment in projections, relaxing the fit to the landings, increasing the fit on the survey indices, increasing the CV on the discards, and using batch fecundity as number of eggs. An alternate run (RW2) was also proposed that was the same as the above run, but with only the fishery independent indices used. Because these runs could not be delivered during the course of the workshop, the results were made available later and are outlined elsewhere in the report.

Conclusions and Recommendations

I have a number of concerns with the data stream for this assessment. Chiefly the uncertainty surrounding both the magnitude of discards and the length frequency was not handled very well in the base model or any of its subsequent configurations. As outlined above, this is a major source of

removals from the population, yet actual data are only available from 2007 onwards, and sampling is very low at-sea.

I also had concerns about the survey indices being constricted to the central portion of the range of the stock. Given the biology of the species, where males build nests and have harems of females, I wondered if this restriction could lead to hyper stability of the fishery independent and dependent indices. One might expect that changes to population size might be reflected in spatial growth of the population, rather than in increase in density within the area surveyed.

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

- Are methods scientifically sound and robust?
- Are assessment models configured properly and used consistent with standard practices?
- Are the methods appropriate for the available data?

General Comments

During the first day of the Review Workshop, the assessment team gave a wonderful presentation bridging the gap between the previous benchmark assessment using ASAP, and the current SS3 modeling approach. This was a highly informative presentation and should be included in the Assessment Report. It is important for the review, as well as the management process, to discuss changes in the model, how they compare to the older approach, and the rationale behind model changes.

The following is a recurring theme in the review. There was not, in my opinion, adequate discussion or justification in the review report as to why certain decisions were made over others. Why was SS3 used vs. the ASAP approach? Why was growth modeled outside SS3 when SS is capable of doing so internally? Why was the fleet structure set up the way it was? Why was the base model not fitting the discards? While some of this information can be found elsewhere or was answered during the review, it is unclear in the Assessment Workshop report.

After examination of the base run diagnostics, it was clear that there were two main difficulties; that the model as configured was fitting neither the indices (Figures 3.2.11-18 in Assessment Report), nor the discards (Figures 3.2.6-10 in Assessment Report). A number of sensitivity analyses were suggested to examine model behavior as outlined in the table below.

Sensitivity:	Level of difficulty:	Priority
Rec. Fleet Structure (one recreational fleet)	Time consuming	
Start Year - 1993	Simple	1
Relax Fit to Landings*	Moderate	4
Fishery-independent indices only run	Simple	2
Fishery-dependent indices only	Simple	3
Explore fit to discards	Moderate	5

These plus other sensitivity analyses requested showed that the model was overfitting the landings and not fitting the indices or the discards, despite suggested changes. After much discussion and examination, the Panel and I suggested side by side examination of three different configurations:

Model Base_orig:

- Start year = 1986
- Uncertainty in landings: CV = 0.05 for all fleets
- Commercial discard $CV_{1986-2006} = 0.9$ and $CV_{2007-2013} = 0.5$
- Fixed retention 1990 – 2008 for commercial handline and longline
- Estimated retention 2009 – 2013 for commercial handline and longline
- Fixed retention for commercial trap
- Estimated retention 1990-2013 for charter/private and headboat
- NMFS bottom longline selectivity was estimated

Model RW1:

- Start year = 1993
- Uncertainty in landings: CV=0.3 for the recreational landings, 0.15 for the commercial landings
- Headboat has been aggregated with charter boat and private to create a single recreational fleet
- Index lambda = 5 (upweighting the indices)
- Discard CV = 0.3
- Retention was estimated, except commercial trap
- NMFS bottom longline selectivity was fixed to be flat topped

Model RW2:

- Same as RW1, but the fishery-dependent indices were removed from the assessment

Are methods scientifically sound and robust?

While SS3 as a methodology may be appropriate for the assessment, the current model diagnostics and results of sensitivity analysis suggest that none of the models formulations considered are sound and robust.

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

Given the below mentioned difficulties, the problem may lie in a simple misconfiguration of the fleet structure. As outlined above, changing fleet structure to include a discard fleet maybe a simple fix. But as it stands currently, the model and the alternate configurations examined are not properly working.

Are the methods appropriate for the available data?

At present this is difficult to say. While it may be a configuration problem with fleet structure, it could also be possible that SS3 is not the right model to use. Fixing growth outside the model, the inability of SS3 to handle both length and age information from the same fleet simultaneously, and the complexity in changing the model to attempt different sensitivity analyses could point to a more simplistic model as a better choice.

Conclusions and Recommendations

Because of the complexity of SS3 as well as the short time during the workshop, the results of comparisons among Base, RW1 and RW2 (above) were not available until well after the workshop had been concluded.

Examination of diagnostics for RW1 showed better fits to indices (Figure 2) and discards (Figure 3), but at the expense of unreasonable changes to the selectivity, as well as inducing a retrospective uncertainty (Figures 4-6). RW2 showed similar fits to discards, resulted in a poorer fit to the indices, produced moderate retrospective uncertainty, and had an unreasonable selectivity pattern (particularly for the combined recreational fleet) similar to the base model. In Both RW1 and RW2, the model sets a much higher R_0 than the base, suggesting very different reference points for biomass and different stock status over the last decade. All runs agree, however, that the most recent 3 year status is not overfished and overfishing is not occurring.

Model	SPR30%	FSPR30%
Base_orig	1203500.00	0.204
RW1	2444260.00	0.212
RW2	2265390.00	0.214

Additionally, the stock recruitment relationship was not very informative. There was almost no correlation between R and S. A steepness (h) of 0.8 used in the base run was high compared to other comparable fish stocks. As such it was suggested fixing it at 0.99. Fixing $h = 0.99$ should not be interpreted as a measure of very high stock productivity of the stock at very low stock sizes, but is merely a method for implementing a forecast going forward with random recruitment. One possible explanation for the poor S-R is that the base model is seeing the stock at high biomass over the time series; that there is little contrast to inform the relationship.

After much thought, reflection, and consideration, my opinion is that none of the final configurations are acceptable for use in management. Clearly this model is not handling discards. As discards are a major source of removals this suggests an unacceptable level of uncertainty for which to make management decisions. While usually discard information is lacking in many assessments, here the uninformative nature of the indices and the variability of the age comps cannot “fill in the gaps”. I’m unsure any modeling approach other than having discards as a simple scalar would be appropriate; but there are certainly approaches to try before making a final decision to that regard.

One approach to attempt is to simply model discards as its own separate fleet in SS3, as recommended in the Panel report. Because of the time constraints in the workshop, and because of the complexity of SS3, this could not be accomplished during the workshop or by the time this report was due. While this is a “low hanging fruit” that the assessment team could conceivably get done relatively soon, the diagnostics and model performance cannot be evaluated by this Panel, and as such should be independently reviewed prior to making management recommendations or providing advice on catch.

Despite the difficulties with the modeling approach, it should be clearly stated that the status of this stock is good, none of the configurations examined suggested that the stock was below its biomass or above its fishing mortality reference points, as outlined elsewhere.

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

- Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
- Is the stock overfished? What information helps you reach this conclusion?
- Is the stock undergoing overfishing? What information helps you reach this conclusion?

- Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
- Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

General Comments

Give the findings in TOR 2, it is difficult to address TOR 3. Overall, the model results, no matter which final configuration was examined, show the stock in a favorable condition. Spawning output seems to be higher than historical, while exploitation appears low. Further, most indices of abundance have shown increases, the fishery is catching less, and the age/size structure has filled out so that older/larger individuals are now present in the population. All indicators are positive.

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

In my option, the results of all assessment configurations suggest a positive status, despite their difficulties. Qualitatively, the results of the configurations are consistent with commercial catch sampling as well as the fishery dependent and independent indices. However, quantitative estimates from the base, RW1 or RW2 runs are not reliable for management purposes. It is hoped that some further analysis and re-working could produce a usable analysis in the short term.

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

Because the current assessment is not usable, I am unable to say that the stock is not overfished. However, the results of the final configurations paired with the information from the catch sampling and indices suggest that it is unlikely that this stock is overfished.

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

Because the current assessment is not usable, I am unable to say if the stock is experiencing overfishing. However, the results of the final configurations paired with the information from the catch sampling and indices suggest that it is unlikely that overfishing is occurring.

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

As stated under TOR 2, the SR for this assessment was un-informative. Because of this, the Panel decided to use a fixed steepness, and median recruitment for projections. Because of this, that steepness value should not be considered reliable for measuring stock productivity.

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

As stated elsewhere, the current base as well the alternative runs proposed during the Workshop are not usable for management purposes. However, clearly all the sensitivity analyses suggest that the stock is in good shape relative to historical levels. All of the fishery independent and dependent indices show positive to neutral signs for this stock in the last few years. Additionally, commercial sampling also shows larger and older fish in the population, suggesting a filling out of the age structure. All of these are positive indications on stock health.

Conclusions and Recommendations

During the Review Workshop, the Panel examined the proposed MSY based reference points recommended by the assessment team and the Assessment Workshop. Given the uninformative SR relationship and the decision to move to a fixed steepness, I and other Panel members suggested that MSY-based reference points are probably not very appropriate, even if the model was acceptable for management use. As an alternative, SPR based reference points, particularly SPR 30%, were suggested.

While I concur with the rest of the Panel that SPR 30% is an option, I have difficulty making it a recommendation despite the popularity of SPR reference points, which are on a per recruit basis. As such, they do not, even when translated into a biomass, give a true minimum stock size and reflect fishing mortality rather than a level below which fishing should be curtailed. Moreover, given the reproductive biology of this species, I'm unsure if SPR based reference points alone are sufficient.

As such I recommend two alternatives or additions: 1) that historical based performance reference points be explored, and 2) that an additional management trigger be evaluated based on sex ratios. For the former, historical based reference points may be more instructive given the different R_0 values among the final runs of this assessment. Likewise, basing biomass reference points on, for example, the lowest level the stock has recovered from, maybe a more common sense approach. Also, additional management triggers based on some historical sex ratio could go a long way in ensuring overall stock reproductive health by accounting for the reproductive socio-biology of the species. To be direct, reproductive health at certain stock sizes maybe tied to the number of males and burrows, rather than the number of females in the harem.

TOR 4: Evaluate the stock projections, including discussing strengths and weaknesses, and consider the following:

- Are the methods consistent with accepted practices and available data?
- Are the methods appropriate for the assessment model and outputs?
- Are the results informative and robust, and useful to support inferences of probable future conditions?
- Are key uncertainties acknowledged, discussed, and reflected in the projection results?

General Comments

Given the findings in TOR 2 it is difficult to address TOR 4. Also, the Panel spent very little time discussing the projections, as there were concerns about the model which occupied much of the meeting. Because the underlying model and alternative runs, in my option, are not suitable for management, the projections as analyzed are also not useful.

Are the methods consistent with accepted practices and available data?

As proposed, the methods expressed in the report seem to be an excellent way of projecting this stock's future under various scenarios. However, as the underlying assessment has some fatal flaws, the projection results are not useable.

Are the methods appropriate for the assessment model and outputs?

The methods proposed seem reasonable given the structure of SS3. However, as the underlying assessment has some fatal flaws, the projection results are not useable.

Are key uncertainties acknowledged, discussed, and reflected in the projection results?

Given that the underlying model is not usable for management purposes, this is not relevant.

Conclusions and Recommendations

The methods proposed in the report seem to be a good way of providing projections. However, as the underlying model is not usable for management purposes, it impossible to say if the methods are practical.

TOR 5: Consider how uncertainties in the assessment, and their potential consequences, are addressed.

- Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods.
- Ensure that the implications of uncertainty in technical conclusions are clearly stated.

General Comments

Given the results of TOR 2; it is difficult to address this TOR. The Panel spent very little time discussing assessment uncertainty. During the presentation of the base model, a number of different sensitivity analyses were presented including: Steepness, Natural Mortality, Discards, Selectivity pattern of the longline survey, and sequentially removing survey indices. For within model sensitivity, the approach was to use the inverted Hessian Matrix to calculate asymptotic standard errors. In addition, both likelihood profiles on key parameters and a retrospective analysis were provided.

Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods.

The methods used were very good at revealing the uncertainties of the base model, particularly the discards. What was not made apparent in the report was how important those discards were to the overall estimates. Discards are a very large fraction of the total removals, and as such there was not a lot of discussion of why the discards were not being fitted as well as sensitivity analysis beforehand to show the trade-offs by forcing fit to these. In the end, the uncertainties in the discards, their fit, and the lack of fit to the indices produced a model which simply was not fitting key components of the input data. Further, no attempts were apparent to explain the lack of fit, nor to run sensitivity analyses around those issues.

Ensure that the implications of uncertainty in technical conclusions are clearly stated.

Given the findings in TOR 2, this is not relevant.

Conclusions and Recommendations

Discussion of uncertainty for a model that is not very useful for management seems to be a moot point. However, there were key uncertainties which were described but not explained within the assessment document. Further once these uncertainties were revealed, there was no follow up work to either explain, nor to revamp the model to correct these deficiencies. In retrospect, this would have been a tall order, as the lack of fit and a retrospective uncertainty induced by forcing a fit was not predictable.

TOR 6: Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.

- Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments with particular emphasis on the Deepwater Horizon oil spill.
- Provide recommendations on possible ways to improve the SEDAR process.

General Comments

Data Workshop recommendations included:

- Time varying natural mortality.
- Index of red tide mortality derived from Ecopath with Ecosim.
- Elucidating the response of red grouper to red tide events.
- Modifications to the CMS modeling framework.
- Enhance fish kill reporting, particularly in offshore regions.

Assessment Workshop Recommendations included:

- Evaluate existing methods for deriving historical discard numbers and discard rates and improve methods as appropriate.
- Develop/evaluate methods to maintain continuity of fishery-dependent indices in light of management regulations and IFQs.
- Considering red tide as an unpredictable event, but can be a significant source of mortality, a response protocol should be developed for data collection and incorporation of the information into updated assessments.
- The start of this assessment is in 1986. Future assessments should investigate extending the assessment model further back in time.
- Develop a protocol for reliable estimation of fishery discards.

Review Workshop Recommendations

- Not finalized at the time of this report's due date.

Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments with particular emphasis on the Deepwater Horizon oil spill.

Please see below. Note: the Deepwater Horizon issue was not discussed and does not appear to be pertinent.

Provide recommendations on possible ways to improve the SEDAR process.

There are many. Please see the overall Conclusions and Recommendations section below.

Conclusions and Recommendations

While the Data and Assessment Workshop recommendations might be useful, my focus is on improving the model. I found many of the above research recommendations more for management than for improving the assessment. As such, I'm going to recommend a different set. Some of these will match the ones above, but others will not. In order of priority, under each section, they are:

Data

- 1) Better discard information is vital to the future success of any assessment for this stock.
- 2) Consider your indices carefully; re-analyze and drop, or up-weight ones based on diagnostics.
- 3) Look for some index (if possible) that is not in the central portion of the range.
- 4) Develop an error distribution around the transition between sexes.
- 5) Develop size at age by year or blocks of years.
- 6) Better elucidate fecundity.

Assessment

- 1) Reconfigure the current SS3 model to incorporate discards as a separate fleet.
- 2) Re-analyze the discard data and incorporate as age rather than length.
- 3) Re-run the model as ASAP or other age-structured model, check resulting selectivity's and diagnostics against base and both RW alternates.
- 4) As a last resort try a biomass model or a VPA to be certain the overall trends are the same.
- 5) Later, try incorporating growth within SS3, as well as modeling the transitions between sexes.

The "Low hanging fruit" here is the input of discards as a separate fleet. But given the appearance of retrospective uncertainty in the RW alternates, diagnostics should be carefully checked.

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

Conclusions and Recommendations

As outlined in TOR 2, the base as well as the Review Workshop alternates are not, in my opinion, useable for management purposes. This includes setting OY/ABC, status determination, reference points as well as projections.

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

Conclusions and Recommendations

Please see Overall Conclusions and Recommendations section (below) for specific recommendations on Report writing, Process, and Modeling.

Overall Conclusions and Recommendations

This was one of the more difficult and complex assessments I have reviewed for the CIE. A part of this is simply the nature of the assessment and the stock/data involved. But there are other factors that increased its difficulty and complexity. There were many challenges in report writing, process, and post workshop follow-up.

Report writing

Clearly there is a need within the document, as presented, to provide more pertinent information to the reviewers, public, managers, and stakeholders. Key pieces seemed to be missing. These included the information on the “bridge” between ASAP and SS3, why data were handled the way they were, why certain assumptions were made, explanations as to why the model wasn’t fitting the data, and others. Sometimes, the reader was directed back to the Data Workshop Report for key pieces of information. While acceptable, my thought is that the Assessment Report should stand alone. Additionally, there was information that was not needed within the document. For example, there was little discussion on model choice, the previous model, the reference points chosen, or why the model wasn’t fitting the data. Yet there was a lengthy section (relatively) going over workshop participation, TOR, and other information that I assume was required by format. In the future, it would be best to provide clear justification and rationale for assumptions/choices, discuss key issues with the model and their implications, explore alternate ideas and link them to the research recommendations.

Process

There must have been some difficulties with the Assessment Workshop process. It is difficult to think that an assessment that was not fitting the indices or the discard data was able to pass through the review of a committee with little to no comment or efforts to explore or repair. In my view, the Assessment workshop is vital for picking up on these issues, for making suggestions to the Assessment Team, and helping them explore the model’s behavior.

After some digging, I found the actual workshop proceedings (which were not given as documentation for the review). The last webinar (<http://gulfcouncil.org/resources/SEDAR/SEDAR%2042%20DW/S42%20AW%20Webinar%204%20Summary.pdf>) held on June 4, 2014 (less than a month from the Review Workshop document deadline) stated:

“Fits to discard data are poor for the commercial indices (underestimation), and marginal for recreational indices. The AW Panel recognized the poor commercial fits, and will recommend possible solutions to the review panel.”

Those recommendations, and the exploration of this behavior, were not available in the document. Further, there were four webinars, each 3-4 hours long. In my experience it takes longer than 12 to 16 hours to explore model behavior in a committee setting when problems arise. Adding in the fact that these were not in-person meetings, the complexities of SS3, the issues with the discards and indices, and it’s clear that there simply was not enough time. Assessment workshop participants, the Assessment team, and others are very busy with this and other assessments. As such, this probably is a case of “too much to do, with too little time and resources to do them in”. For example, the Data Workshop was held eight months prior to the Review Workshop, and produced a report only 4-5 months prior. This leaves very little time for the Assessment Workshop to complete their work with only 3-4 hour monthly webinars in March-June, especially when there is a change in models to one of higher complexity.

In my opinion, the Assessment Team simply did not have enough guidance from the Assessment Workshop to produce a viable product. This was due to both the reliance on webinars and the short turnaround time. While the time and other resource savings of webinars is huge, in the future Assessment Workshops should have at least one in-person meeting. This is particularly true as one starts using a very complex model that is relatively new to the region.

The Review Workshop suffered from similar issues. While the preparatory documentation was on time, the Review Workshop itself was rather short (three full days). Generally, these are 4-5 days long rather than three. Also, while the Chair attempted to hold a conference call prior to the Workshop, given the short notice, it was simply not possible. Because of the issues in the assessment, and the complexities associated with SS3, model suggestions by the Panel took a long time for the Assessment Team to complete during the review. Some of the issues identified in my review could have been avoided had a conference call been held prior to our in-face meeting. At least a good jump on the request could have been made by the assessment team, leaving more time for discussions by the Panel. Alternatively a longer meeting might have produced a similar result. In any case, very little Workshop time was spent discussing vital aspects, like uncertainty, reference points, projections, and other issues. Partly because of the hang ups with the model fitting, and partly because of long wait times for alternate runs.

The “spill-over” meant that the post-Workshop process also suffered. Because of late requests by the Panel on the last day, the Assessment Team took well over a week to complete alternate runs (RW1 and RW2) and associated diagnostics. This delayed the Panel’s report; which by my review’s due date has still not been completed, nearly three weeks later. This in turn resulted in the request by the CIE reviewers for an extension of our contract by a full week. While I didn’t make the request for the extension, I was happy to see it. I was very concerned that my report would not reflect the most recent runs requested on the last day of the workshop. Even by this report’s writing, I am still unsure how my views overlap those on the Panel, as the Panel’s report is still unfinished.

An additional, and more minor concern, was how the documents were distributed during the meeting. SEDAR uses an FTP-type protocol with proprietary software to distribute files during the Review Workshop. This is rather archaic, requiring loading of new software and learning how the new procedure works; including a stand-alone document. A switch to a more commonly used procedure, like Drop Box or Google Drive, would make things much easier

Given all of this, my explicit recommendations are established to hopefully prevent these issues from happening again during the Assessment/Review process. Some of these have been hinted at earlier, while others have not. They include the following:

- 1) A recommendation that the Assessment workshop include one person-to-person meeting.
- 2) That extra time be built-in when there is a change in model or when issues in the base run are detected during the Assessment Workshop.
- 3) That the Chair ask the Panel if there is a need for a pre-Workshop Webinar prior to the Review Workshop, and that this webinar be standard procedure if there is a change in model.
- 4) That this Webinar be held shortly after getting the documents, allowing for more lead-time for the Assessment Team to conduct sensitivity analyses.
- 5) That for a complex assessment more time at the Review Workshop be given for their review.
- 6) That Panels be prevented from making requests on the final day that will take over a week to accomplish.

- 7) That in the case where final determination hinges on results which are to be available after the Workshop concludes, that a final webinar be scheduled at a later date (within a week or two) while the Panel is in the room on the last day of the workshop.
- 8) That in cases where there is a complex model, or where significant time is involved after the conclusion of the Workshop, that CIE reports be delayed accordingly.

Modeling

As state previously, the current model and the requested alternate runs are not usable for management purposes, in my opinion. However, I struggled and reflected for a long time to arrive at that conclusion. Without a doubt this model is close to being useful. During my contemplation, I was heartened that the model, even with the alternate runs, was giving the same answer; and that this answer was in line with the data and the overall perceptions of the stock. But just because you seem to be getting the right answer doesn't mean that the tool itself is appropriate. These models are used well past this Review and for far longer than most would like in setting specifications. This plus the accessory use of providing management advice on min sizes, bag limits, and other measures made me hesitant to come to any other conclusion.

My recommendation for this is that in the short term the Assessment Team explore setting discards as a separate fleet, that they carefully select only those indices that are useful, and that they reconsider the retention/selectivity issue. My hope is that one of these, most likely the discards, will resolve the issues quickly. But I cannot say if the result will be useful at this point.

I'm sure that the SS3 framework is the way to go for this and other assessments in the region. Its flexibility and its adaptability will be very useful as the region struggles with different and data poor situations. However, there were times during the Workshop when the Assessment Team was hampered by the model and they spent a lot of time trying to figure out how to get SS3 to do what they wanted. While we can all struggle with this, sometimes when a model becomes too difficult to work with, it's time to build your own! I have no doubt that they and their colleagues could build their own regional model, much like ASAP, SS3, BAM and others were built to meet the needs specific of their regions.

Final Thoughts

It is always a real pleasure to work in the SEDAR process for a number of reasons. The Chair was excellent and is always fun to be around. The support staff is accommodating and professional as always. My colleagues from CIE and the SSC were very knowledgeable and I relied on them extensively to provide their views and guidance during the process. This was particularly true for the SSC Panel members who added a lot and reined me in with the realities of the process from time to time. Meeting and working with both Geoff and Henrik was a blast. I have read their work and it was finally nice to meet them in person. I found that the Panel's composition was an excellent balance of local experience, social/economic views, as well as technical expertise.

Miami in July is a little hot for someone from Maine, but the accommodations were comfortable and the choice of venue well above satisfactory.

What I was most impressed with was the Assessment Team. While relatively inexperienced as leads and being new to SS3, they performed admirably with a flawed assessment and data. They were very accommodating, highly professional, and very capable. This was a stressful assessment, and they took

most of the criticisms in stride and were diligent in both requests and in conducting alternate runs. They are a significant credit to the Miami laboratory (Southeast Fisheries Science Center).

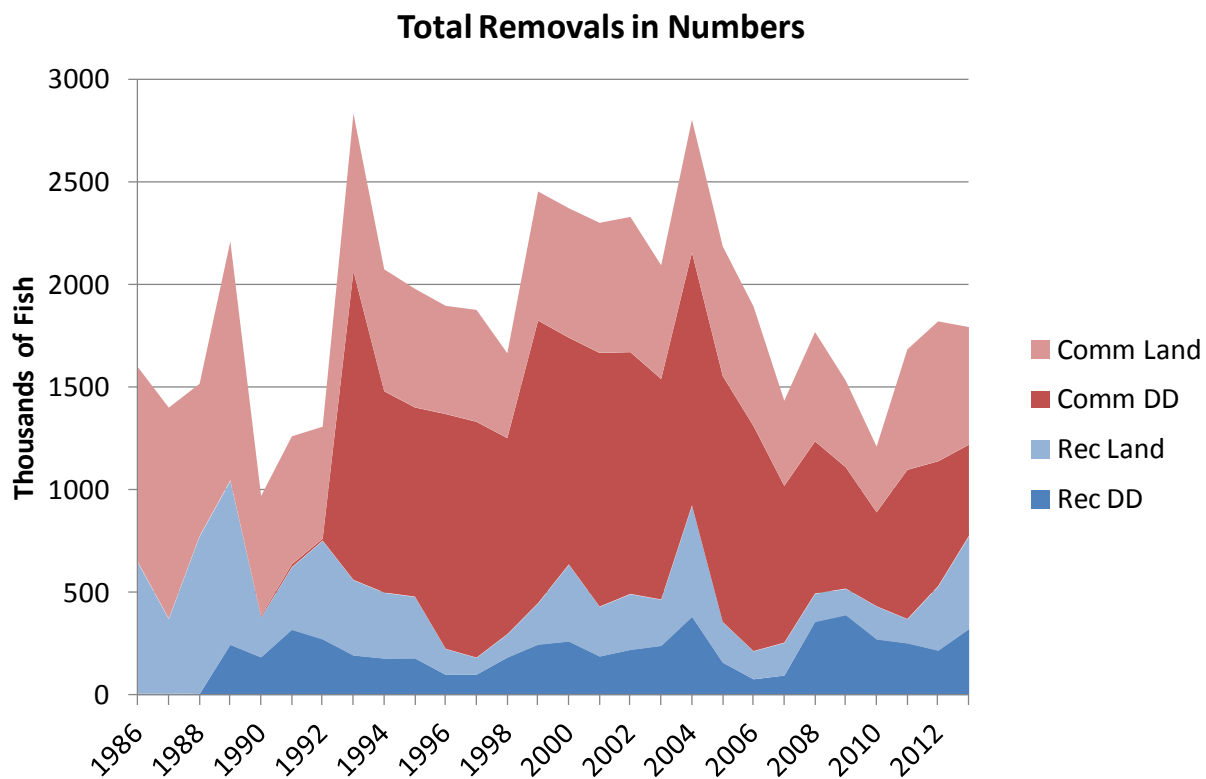


Figure 1. Removal by landings or discards by fleet.

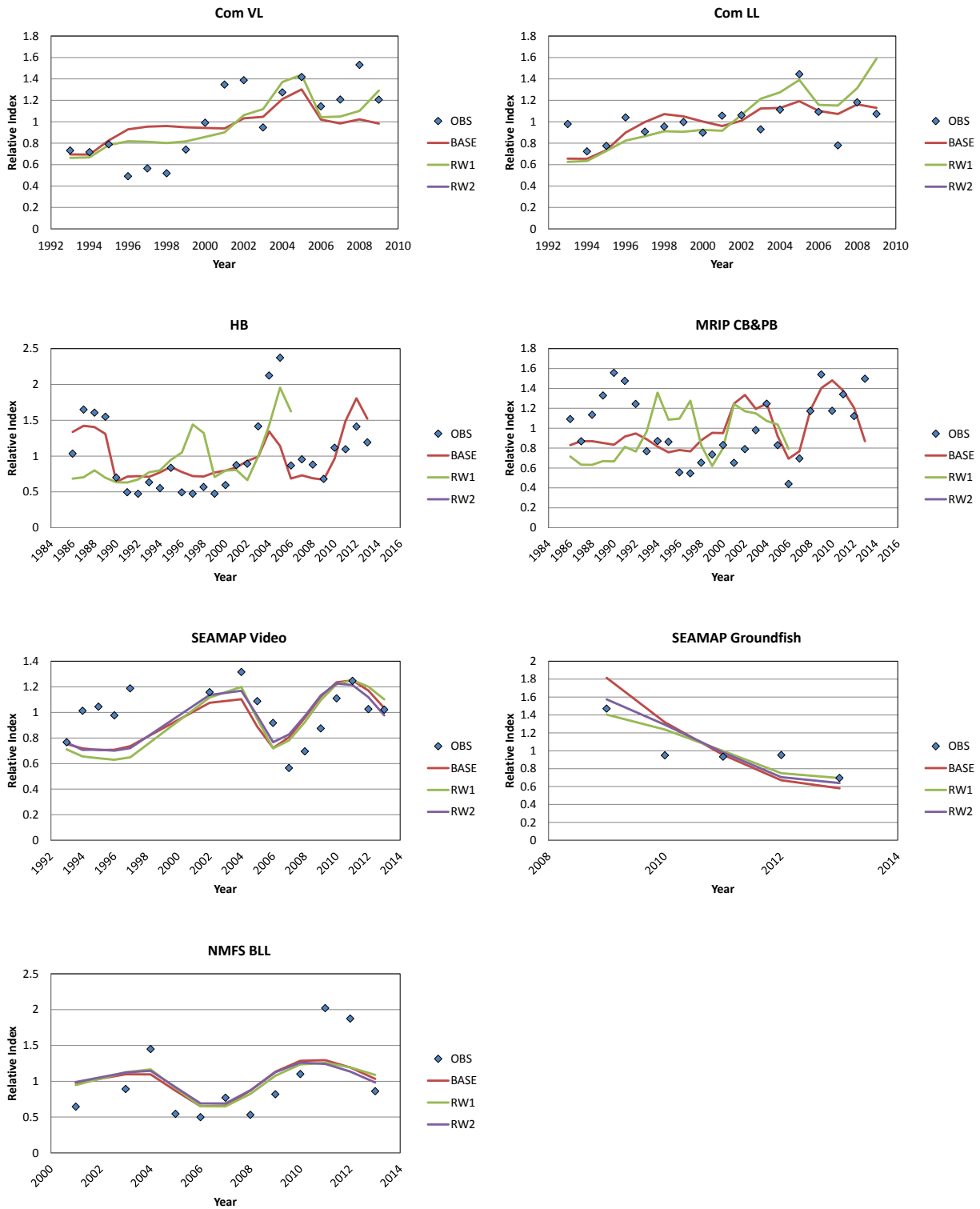


Figure 2. Comparison of fits to indices for Base, RW1 and RW2

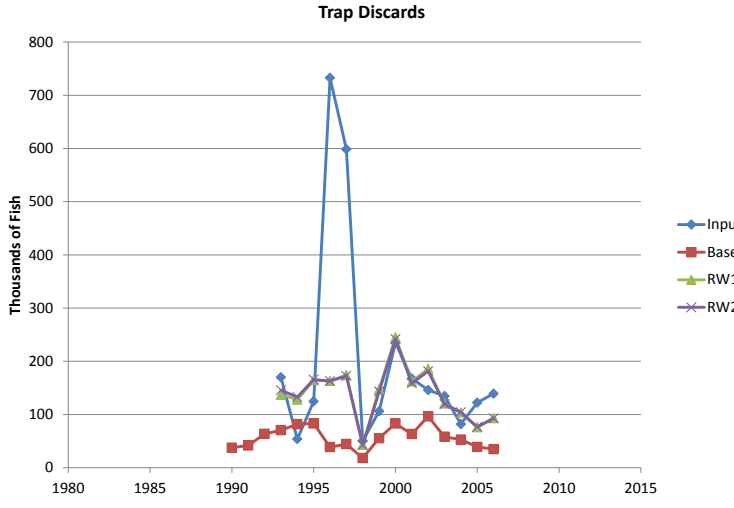
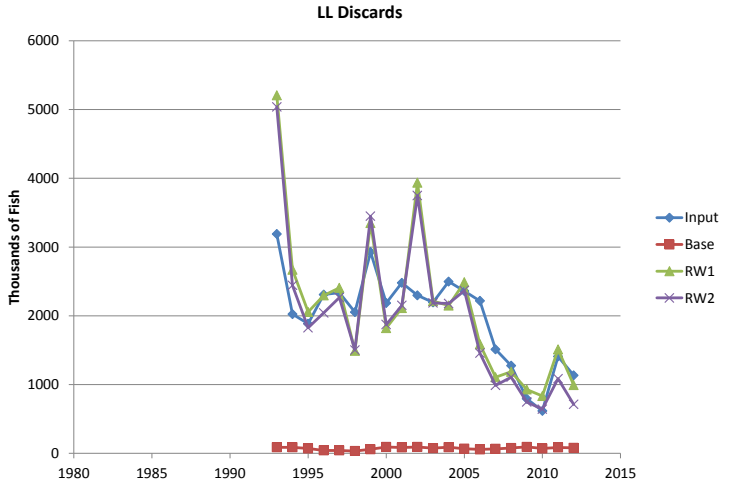
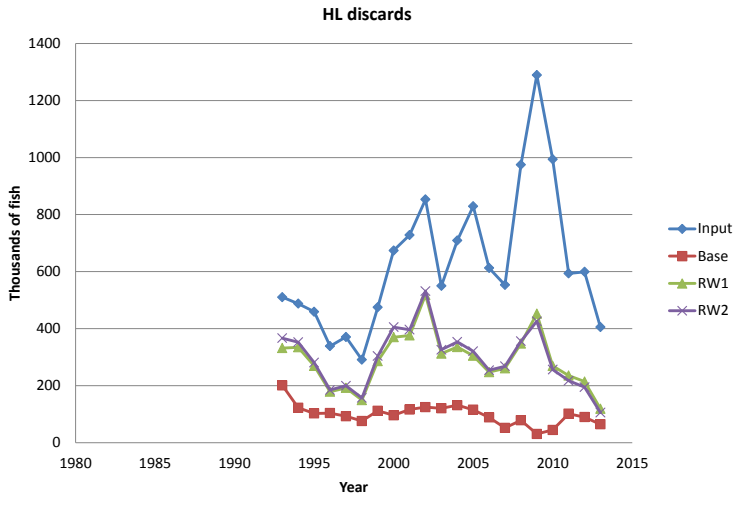


Figure 3. Comparison of fit to discard estimates for Base, RW1 and RW2.

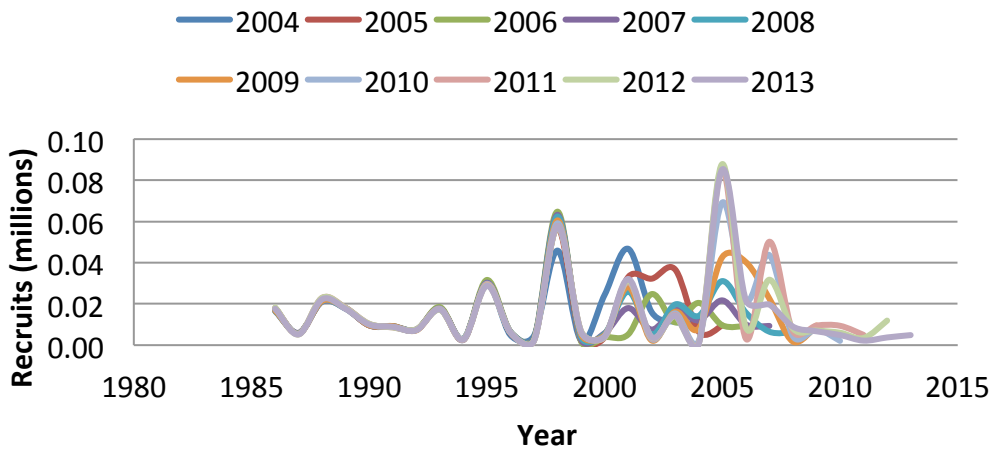
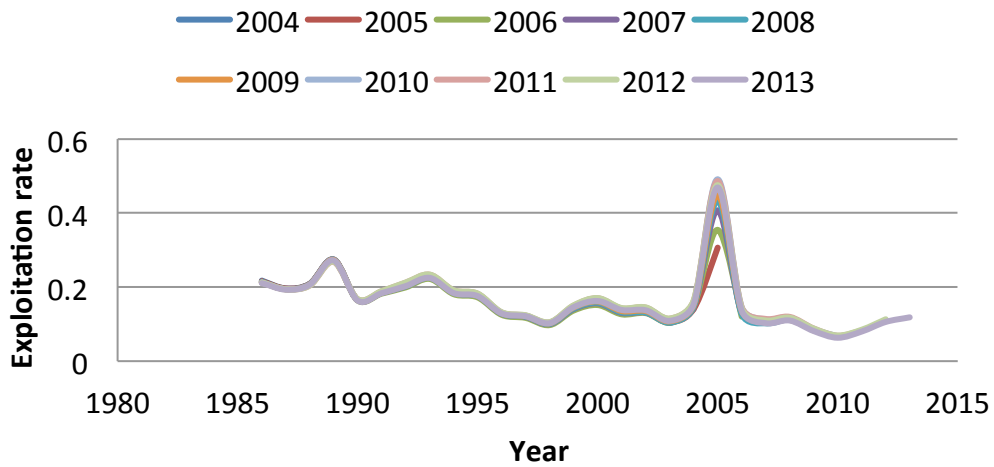
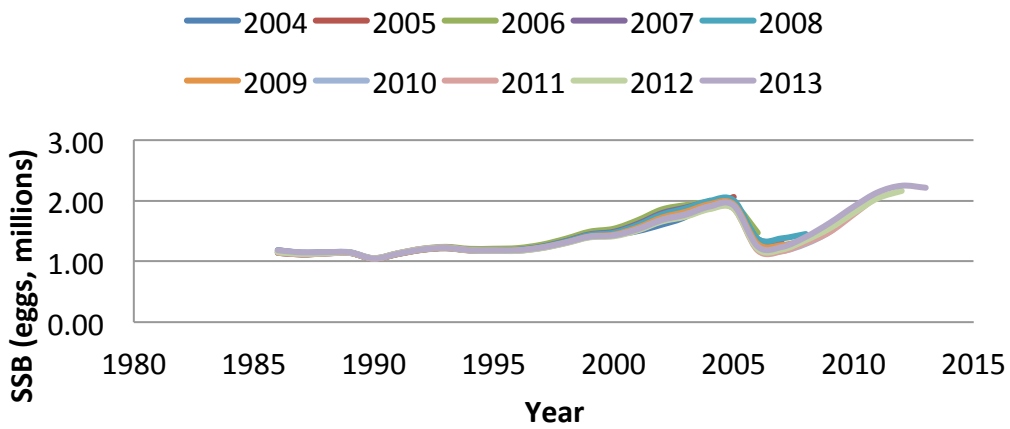


Figure 4. Retrospectives results for model Base_orig. The proxy for SSB is eggs (as shown in the top panel), exploitation rate represents that total catch over total biomass (middle panel), and age-0 recruits (bottom panel).

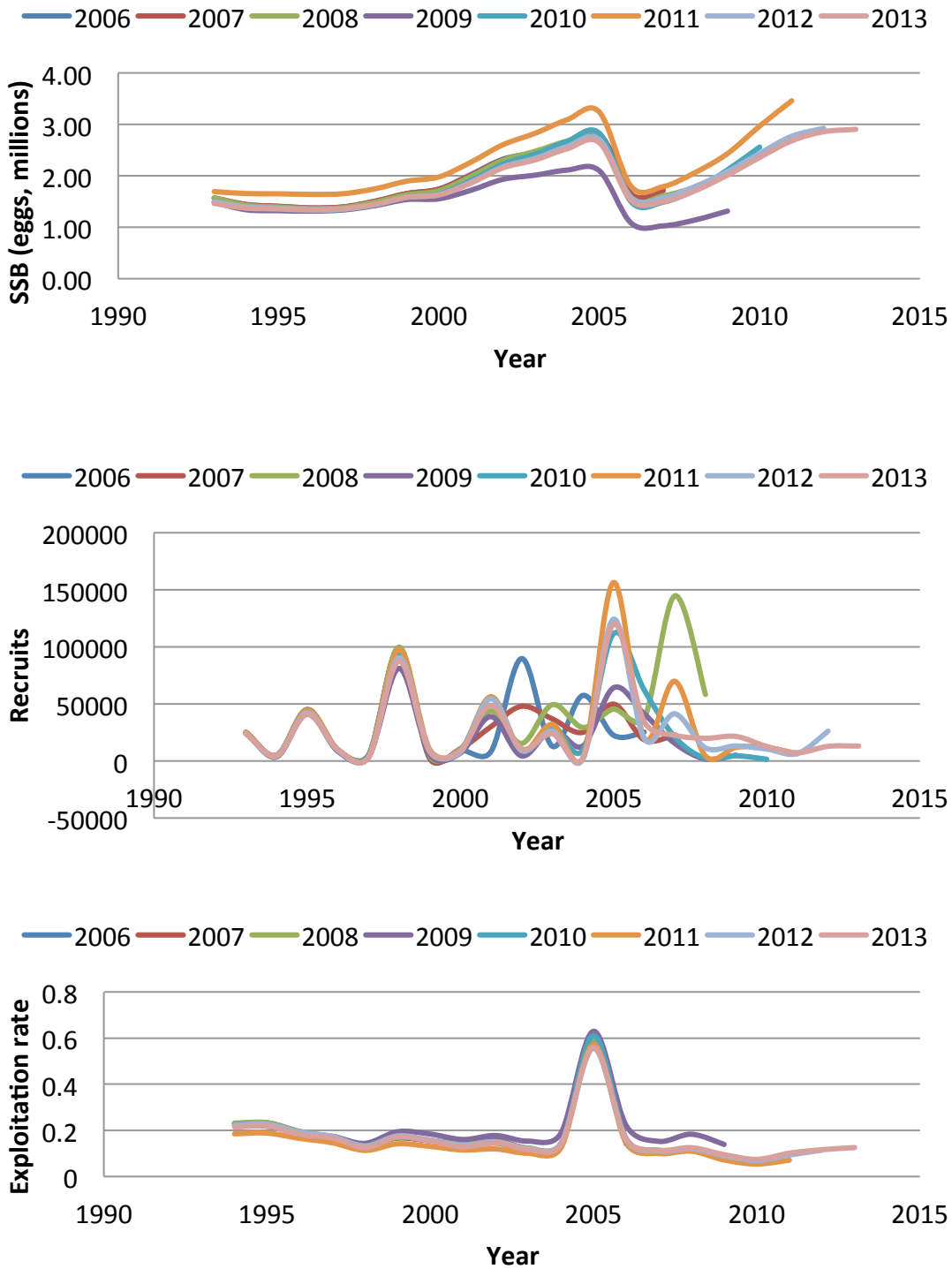


Figure 5. Retrospectives results for model RW1. The proxy for SSB is eggs (as shown in the top panel), exploitation rate represents that total catch over total biomass (middle panel), and age-0 recruits (bottom panel).

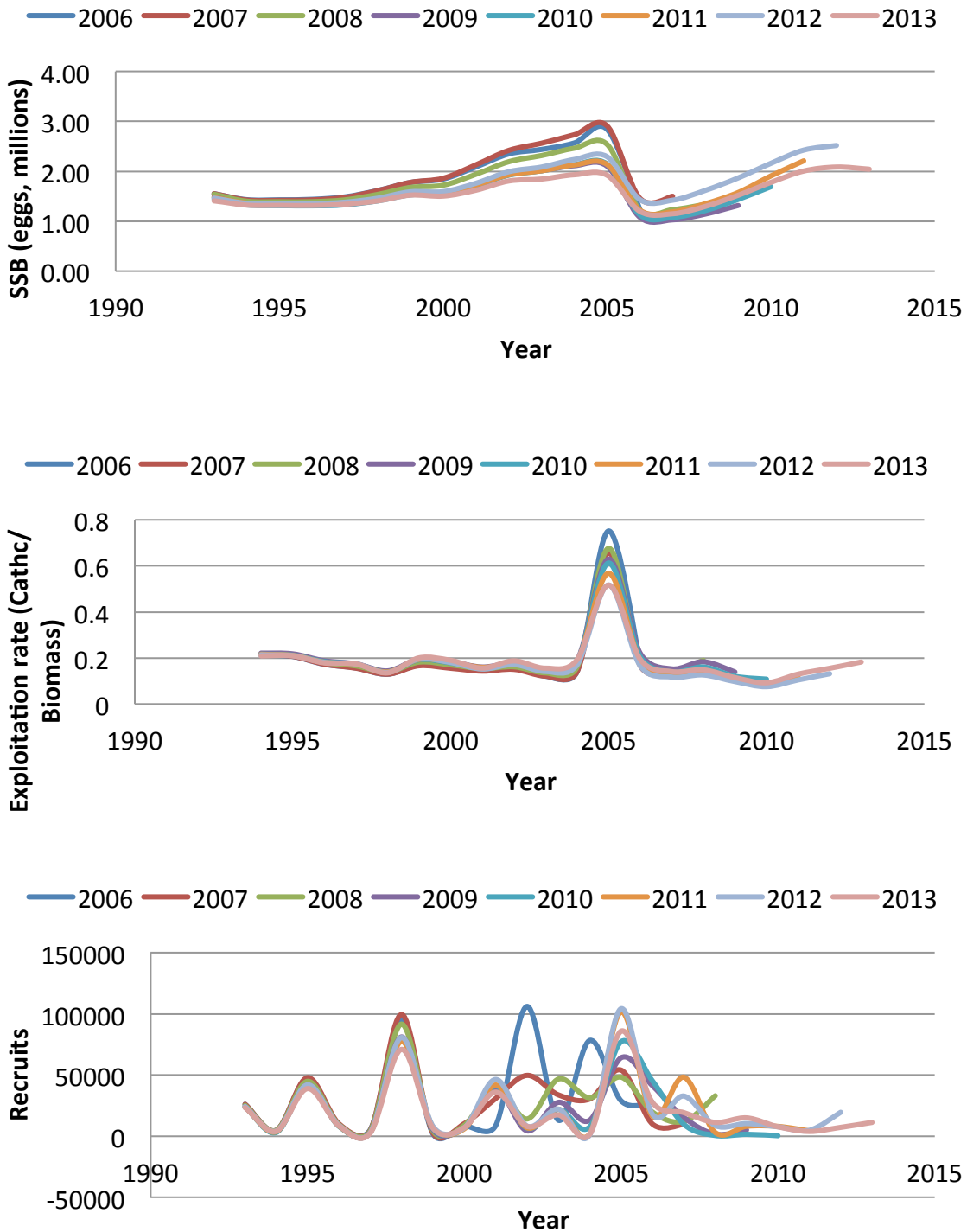


Figure 6. Retrospectives results for model RW2. The proxy for SSB is eggs (as shown in the top panel), exploitation rate represents that total catch over total biomass (middle panel), and age-0 recruits (bottom panel).

Appendix 1: Bibliography of materials provided for review

**SEDAR 42
Gulf of Mexico Red Grouper
Document List**

Document #	Title	Authors	Date Submitted
Documents Prepared for the Data Workshop			
SEDAR42-DW-01	Summary of commercial red grouper (<i>Epinephelus morio</i>) catch data based on fishery observer coverage of the Gulf of Mexico reef fish fishery	Jeffrey R. Pulver, Linda Lombardi, and Elizabeth Scott-Denton	27 Oct 2014
SEDAR42-DW-02	Evaluation of the natural mortality rates of red grouper (<i>Epinephelus morio</i>) in the West Florida Shelf ecosystem using the individual-based, multi-species model OSMOSE-WFS	A. Grüss, M. J. Schirripa, D. Chagaris, P. Verley, Y.-J. Shin, L. Velez, C. H. Ainsworth, S. R. Sagarese, and M. Karnauskas ²	1 Nov 2014 Updated: 11 March 2015
SEDAR42-DW-03	Use of the Connectivity Modeling System to estimate the larval dispersal, settlement patterns and annual recruitment anomalies due to oceanographic factors of red grouper (<i>Epinephelus morio</i>) on the West Florida Shelf	A. Grüss, M. Karnauskas, S. R. Sagarese, C.B. Paris, G. Zapfe, J.F. Walter III, W. Ingram, and M. J. Schirripa	2 Nov 2014 Updated: 14 Nov 2014
SEDAR42-DW-04	Ontogenetic spatial distributions of red grouper (<i>Epinephelus morio</i>) within the northeastern Gulf of Mexico and spatio-temporal overlap with red tide events	S. R. Sagarese, A. Grüss, M. Karnauskas, J.F. Walter III	3 Nov 2014
SEDAR42-DW-05	Red Grouper Abundance Indices from SEAMAP Groundfish Surveys in the Northern Gulf of Mexico	Adam G. Pollack and G. Walter Ingram, Jr.	7 Nov 2014 Updated: 26 Nov 2014
SEDAR42-DW-06	Red Grouper Abundance Indices from NMFS Bottom Longline Surveys in the Northern Gulf of Mexico	Adam G. Pollack and G. Walter Ingram, Jr.	19 Nov 2014
SEDAR42-DW-07	Maturity, sexual transition, and spawning seasonality in the protogynous red grouper on the West Florida Shelf	Susan Lowerre-Barbieri, Laura Crabtree, Theodore S. Switzer, and Robert H. McMichael, Jr.	17 Nov 2014 Updated: 21 Nov 2014
SEDAR42-DW-08	Indices of abundance for Red Grouper (<i>Epinephelus morio</i>) from the Florida Fish and Wildlife Research Institute (FWRI) video survey on the West Florida Shelf	Cameron B. Guenther, Theodore S. Switzer, Sean F. Keenan, and Robert H. McMichael, Jr.	12 Nov 2014
SEDAR42-DW-09	Indices of abundance for Red Grouper (<i>Epinephelus morio</i>) from the Florida Fish and Wildlife Research Institute (FWRI) chevron trap survey on the West Florida Shelf	Cameron B. Guenther, Theodore S. Switzer, Sean F. Keenan, and Robert H. McMichael, Jr.	12 Nov 2014
SEDAR42-DW-10	An age and growth description of Red	Linda Lombardi-	13 Nov 2014

SEDAR 42 Document List

30-Jul-15

SEDAR 42
Gulf of Mexico Red Grouper
Document List

	Grouper (<i>Epinephelus morio</i>) from the northeastern Gulf of Mexico: 1978-2013 for SEDAR42	Carlson	Updated: 10 Dec 2014
SEDAR42-DW-11	SEAMAP Reef Fish Video Survey: Relative Indices of Abundance of Red Grouper	Matthew D. Campbell, Kevin R. Rademacher, Michael Hendon, Paul Felts, Brandi Noble, Michael Felts, Joseph Salisbury, and John Moser	13 Nov 2014
SEDAR42-DW-12	Variations in length frequency distributions and age length keys for red groupers collected in the Gulf of Mexico	Ching-Ping Chih	14 Nov 2014
SEDAR42-DW-13	The use of Otolith Reference Collections to Determine Ageing Precision of Red Grouper (<i>Epinephelus morio</i>) Between Fisheries Laboratories	Palmer, C.L., L. Lombardi, J. Carroll, and E. Crow	18 Nov 2014 Updated: 16 Dec 2014
SEDAR42-DW-14	Size Distribution of Red Grouper Observed in For-Hire Recreational Fisheries in the Gulf of Mexico	Alisha Gray and Beverly Sauls	20 Nov 2014 Updated: 15 Dec 2014
SEDAR42-DW-15	Red Grouper <i>Epinephelus morio</i> Findings from the NMFS Panama City Laboratory Trap & Camera Fishery-Independent Survey – 2004-2014	D.A. DeVries, C.L. Gardner, P. Raley, and W. Ingram	5 Dec 2014
SEDAR42-DW-16	Estimates of Historical Private/Charterboat and Headboat Fishery Red Grouper Angler Catch in the Gulf of Mexico 19xx-1980	Jeff Isely, Nancie Cummings and Adyan Rios	9 Dec 2014
SEDAR42-DW-17	Discards of red grouper (<i>Epinephelus morio</i>) for the headboat fishery in the US Gulf of Mexico	Fisheries Ecosystems Branch, Beaufort, NC	21 Nov 2014 Updated: 10 Dec 2014
SEDAR42-DW-18	Length and age frequency distributions for red groupers collected in the Gulf of Mexico from 1984 to 2013	Ching-Ping Chih	11 Dec 2014
SEDAR42-DW-19	Index report cards	Indices Working Group	17 Dec 2014
Documents Prepared for the Assessment Process			
SEDAR42-AW-01	Red tide mortality on red grouper (<i>Epinephelus morio</i>) between 1980 and 2009 on the West Florida Shelf	Skyler R. Sagarese, Alisha M. Gray, Cameron H. Ainsworth, David D. Chagaris, Behzad Mahmoudi	5 Feb 2015

SEDAR 42
Gulf of Mexico Red Grouper
Document List

SEDAR42-AW-02	Standardized catch rates for red grouper from the United States Gulf of Mexico vertical line and longline fisheries	Meaghan D. Bryan and Kevin McCarthy	10 March 2015
SEDAR42-AW-03	Standardized Catch Rates of Red Grouper (<i>Epinephelus morio</i>) from the U.S. Headboat Fishery in the Gulf of Mexico, 1986-2013	Adyan Rios	13 March 2015
SEDAR42-AW-04	Standardized Catch Rates of Red Grouper (<i>Epinephelus morio</i>) from the Gulf of Mexico Recreational Charterboat and Private Boat Fisheries (MRFSS) 1986-2013	Adyan Rios	13 March 2015
SEDAR 42-AW-05	Estimating age- and size-specific natural mortality rates for Gulf of Mexico red grouper (<i>Epinephelus morio</i>) using the ecosystem model OSMOSE-WFS	A. Grüss, M. J. Schirripa, D. Chagaris, P. Verley, Y.-J. Shin, L. Velez, C. H. Ainsworth, S. R. Sagarese, and L. Lombardi-Carlson	11 March 2015
Documents Prepared for the Review Workshop			
SEDAR42-RW-01	Incorporating ecosystem considerations within the Stock Synthesis integrated assessment model for Gulf of Mexico Red Grouper (<i>Epinephelus morio</i>)	Skyler R. Sagarese, Meaghan D. Bryan, John F. Walter, Michael Schirripa, Arnaud Grüss, Mandy Karnauskas	29 June 2015
SEDAR42-RW-02	Assessing the impact of the 2014 red tide event on red grouper (<i>Epinephelus morio</i>) in the Northeastern Gulf of Mexico	John F. Walter III, Skyler R. Sagarese, William J. Harford, Arnaud Grüss, Richard P. Stumpf, Mary C. Christman	14 July 2015 Updated: 20 July 2015
Final Stock Assessment Reports			
SEDAR42-SAR1	Gulf of Mexico Red Grouper	SEDAR 42 Panels	
Reference Documents			
SEDAR42-RD01	A Directed Study of the Recreational Red Snapper Fisheries in the Gulf of Mexico along the West Florida Shelf – Final Project Report	Beverly Sauls, Oscar Ayala, and Richard Cody	
SEDAR42-RD02	SEDAR 12-DW-11: Quantitative Historical Analysis of the United States and Cuban Gulf of Mexico Red Grouper Commercial Fishery	Steve Saul	
SEDAR42-RD03	Age, growth, and mortality of red grouper,	Todd C. Stiles and Michael L. Burton	

SEDAR 42
Gulf of Mexico Red Grouper
Document List

	<i>Epinephelus morio</i> , from the Southeastern U.S.	
SEDAR42-RD04	Life history of red grouper, (<i>Epinephelus morio</i>) off the coasts of North Carolina and South Carolina	J.M. Burgos, G.R. Sedberry, D.M. Wyanski, and P.J. Harris

Appendix 2: A copy of the CIE Statement of Work

SEDAR 42 Gulf of Mexico Red Grouper Assessment Review Workshop

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

Project Description: SEDAR 42 will be a compilation of data, an assessment of the stock, and CIE assessment review conducted on Gulf of Mexico Red Grouper. The review workshop provides an independent peer review of SEDAR stock assessments. The term review is applied broadly, as the review panel may request additional analyses, error corrections and sensitivity runs of the assessment models provided by the assessment panel. The review panel is ultimately responsible for ensuring that the best possible assessment is provided through the SEDAR process. The Terms of Reference (ToRs) of the peer review are attached in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

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

Location of Peer Review: Each CIE reviewer shall conduct an independent peer review during the panel review meeting scheduled in **Miami, Florida** during **July 14-16, 2015**.

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

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer information (full name, title, affiliation, country, address, email) to the COTR, who forwards this information to the NMFS Project Contact no later the date specified in the Schedule of Milestones and Deliverables. The CIE is responsible for providing the SoW and ToRs to the CIE reviewers. The NMFS Project Contact is responsible for providing the CIE reviewers with the

background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact is also responsible for providing the Chair a copy of the SoW in advance of the panel review meeting. Any changes to the SoW or ToRs must be made through the COTR prior to the commencement of the peer review.

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

<http://deemedexports.noaa.gov/>

http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html

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

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

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

Other Tasks – Contribution to Summary Report: Each CIE reviewer may assist the Chair of the panel review meeting with contributions to the Summary Report, based on the terms of reference of the review. Each CIE reviewer is not required to reach a consensus, and should provide a brief summary of

the reviewer’s views on the summary of findings and conclusions reached by the review panel in accordance with the ToRs.

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

- 1) Conduct necessary pre-review preparations, including the review of background material and reports provided by the NMFS Project Contact in advance of the peer review.
- 2) Participate during the panel review meeting in Miami, Florida during July 14-16, 2015.
- 3) Conduct an independent peer review, as specified herein, in Miami, Florida during July 14-16, 2015, in accordance with the ToRs (Annex 2).
- 4) No later than August 3, 2015, each CIE reviewer shall submit an independent peer review report addressed to the “Center for Independent Experts,” and sent to Dr. Manoj Shivlani, CIE Lead Coordinator, via email to *MShivlani@ntvifederal.com*, and Dr. David Sampson, CIE Regional Coordinator, via email to *david.sampson@oregonstate.edu*. Each CIE report shall be written using the format and content requirements specified in Annex 1, and address each ToR in **Annex 2**.

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

<i>June 15, 2015</i>	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
<i>June 29, 2015</i>	NMFS Project Contact sends the CIE Reviewers the pre-review documents
<i>July 14-16, 2015</i>	Each reviewer participates and conducts an independent peer review during the panel review meeting
<i>July 30, 2015</i>	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
<i>August 10, 2015</i>	CIE submits CIE independent peer review reports to the COTR
<i>August 17, 2015</i>	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

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

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

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

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

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

Support Personnel:

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

William Michaels
NMFS Office of Science and Technology
1315 East West Hwy, SSMC3, F/ST4, Silver Spring, MD 20910
William.Michaels@noaa.gov Phone: 301-427-8155

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

Key Personnel:

NMFS Project Contact:

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

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

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether the science reviewed is the best scientific information available.
2. The main body of the reviewer report shall consist of a Background, Description of the Individual Reviewer's Role in the Review Activities, Summary of Findings for each ToR in which the weaknesses and strengths are described, and Conclusions and Recommendations in accordance with the ToRs.
 - a. Reviewers should describe in their own words the review activities completed during the panel review meeting, including providing a brief summary of findings, of the science, conclusions, and recommendations.
 - b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, and especially where there were divergent views.
 - c. Reviewers should elaborate on any points raised in the Summary Report that they feel might require further clarification.
 - d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.
 - e. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The CIE independent report shall be an independent peer review of each ToRs, and shall not simply repeat the contents of the summary report.
3. The reviewer report shall include the following appendices:
 - Appendix 1: Bibliography of materials provided for review
 - Appendix 2: A copy of the CIE Statement of Work
 - Appendix 3: Panel Membership or other pertinent information from the panel review meeting.

Annex 2: Tentative Terms of Reference for the Peer Review

SEDAR 42 Gulf of Mexico Red Grouper Assessment Review Workshop

1. Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions, and consider the following:
 - a) Are data decisions made by the DW and AW sound and robust?
 - b) Are data uncertainties acknowledged, reported, and within normal or expected levels?
 - c) Are data applied properly within the assessment model?
 - d) Are input data series reliable and sufficient to support the assessment approach and findings?
2. Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account the available data, and considering the following:
 - a) Are methods scientifically sound and robust?
 - b) Are assessment models configured properly and used consistent with standard practices?
 - c) Are the methods appropriate for the available data?
3. Evaluate the assessment findings and consider the following:
 - a) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
 - b) Is the stock overfished? What information helps you reach this conclusion?
 - c) Is the stock undergoing overfishing? What information helps you reach this conclusion?
 - d) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
 - e) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?
4. Evaluate the stock projections, including discussing strengths and weaknesses, and consider the following:
 - a) Are the methods consistent with accepted practices and available data?
 - b) Are the methods appropriate for the assessment model and outputs?
 - c) Are the results informative and robust, and useful to support inferences of probable future conditions?
 - d) Are key uncertainties acknowledged, discussed, and reflected in the projection results?
5. Consider how uncertainties in the assessment, and their potential consequences, are addressed.
 - Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods.
 - Ensure that the implications of uncertainty in technical conclusions are clearly stated.

6. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.
 - Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments with particular emphasis on the Deepwater Horizon oil spill.
 - Provide recommendations on possible ways to improve the SEDAR process.
7. Consider whether the stock assessment constitutes the best scientific information available using the following criteria as appropriate: relevance, inclusiveness, objectivity, transparency, timeliness, verification, validation, and peer review of fishery management information.
8. Provide guidance on key improvements in data or modeling approaches which should be considered when scheduling the next assessment.
9. CIE Reviewer may contribute to a Peer Review Summary summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference.

Annex 3: Tentative Agenda
SEDAR 42 Gulf of Mexico Red Grouper Review Workshop

Miami, Florida
14-16 July 2015

Tuesday

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

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

Wednesday

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

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

Thursday

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

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

Appendix 3: Panel Membership and Participant list

Workshop Panel

Luiz Barbieri, Chair	Chair, SSC
Ben Blount	SSC
Matt Cieri	CIE Reviewer
Kai Lorenzen	SSC
Henrik Sparholt	CIE Reviewer
Geoff Tingley	CIE Reviewer

Analytic Representation

Meaghan Bryan	SEFSC, Miami
Adyan Rios	SEFSC, Miami

Appointed Observers

Bo Gorham	Recreational Fisherman
Ed Walker	Industry Rep

Observers

Shannon Cass-Calay	SEFSC, Miami
Nancie Cummings	SEFSC, Miami
Michael Drexler	Ocean Conservancy
Chad Hanson	Pew
Bill Harford	RSMAS/SEFSC
Michael Larkin	SERO
Patrick Lynch	NMFS/ST
Clay Porch	SEFSC, Miami
Skyler Sagarese	SEFSC/UM

Staff

Julie Neer	SEDAR
John Froeschke	GMFMC Staff
Ryan Rindone	GMFMC Staff
Charlotte Schiaffo	GMFMC Staff