# Center for Independent Experts (CIE) review report for the 2014 assessment of

# Gulf of Mexico Gag and Greater Amberjack

By

Anders Nielsen

April 13, 2014

# Executive Summary

This reviewer is very impressed with the assessment team. The two assessment reports were very complete and detailed. During the meeting the assessment team worked constantly and were impressively efficient at producing all the analysis requested by the reviewers. It was clear that they were able to answer any question posed by the panel, and that the fairly complex models were carefully constructed and fully understood by the assessment panel.

## *[GAG]*

For Gulf of Mexico Gag the data decisions and modeling approaches are generally scientifically sound and robust. The review meeting agreed to suggest the use of the combined gender reference points (and not the female only), as they are more conservative. With this convention in place the stock is overfished, or right on the border to be overfished, but not undergoing overfishing. A rapid recovery of Gag is seen in all model runs.

Confidence in the assessment and important output metrics could possibly be further improved by: 1) Investigating the use of female only versus combined reference points; 2) searching for more convincing stock recruitment relationship or simply borrowing a suitable steepness parameter from a similar stock; 3) restricting the model to fewer model parameters (especially variance parameters), such that the data weighting would become more objective; and 4) figuring out why the SEAMAP video index is inconsistent with the other data sources.

## *Greater Amberjack [GAJ]*

For Gulf of Mexico Greater Amberjack the data decisions are sound and robust. The two model approaches presented are scientifically sound, but not optimal for the data series. The first approach (ASPIC) is considered to be too simple, as it is unable to account for different selectivities between fleets and over time. The second approach (Stock Synthesis) is considered very promising, but the configuration presented had convergence problems, which made the important output metrics unreliable.

It is expected that the convergence issue can be resolved by relative minor adjustments to the model configuration, and that should be the main priority. A suggested approach is outlined below (TOR 2). Secondly it would be recommended to revisit the estimation of the steepness parameter (or to borrow a value from a similar but data rich stock) and to investigate a better way derive the effective sample size used for the composition data.

# Background

This report is prepared for the Center for Independent Experts (CIE). It contains an independent and impartial review of the assessments of Gulf of Mexico Gag (GAG) and Greater Amberjack. The SEDAR 33 Assessment Review was held February 24-27, 2014 in Miami, FL.

The assessment scientists presenting were the Lead Assessment Scientists: Jakob Tetzlaff, Meaghan Bryan, Nancie Cummings, and Adyan Rios. All are from the NMFS Southeast Fisheries Science Center in Miami, Florida.

Prior to the meeting the review panelists were given a link to an ftp site with background documents (Appendix 1). This reviewer's statement of work can be found in Appendix 2, and a list of review meeting participants in Appendix 3.

# Description of the reviewer’s role

This reviewer has independently read all documents deemed necessary in preparation for the review, traveled and participated actively in the review meeting, identified key issues in the assessment, contributed to the review panel’s summary report, and independently authored this review report.

# Findings for each term of reference

To ensure that all terms of reference are covered, and that comments are interpreted with reference to the correct terms, the terms are listed in gray boxes with corresponding reviewer comments following.

1. Evaluate the data used in the assessment, addressing the following:

– Are data decisions made by the Data and Assessment Workshops 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?

## *[GAG]*

The data decisions are clearly and precisely described in the assessment report. In most cases the decisions follow the data workshop and where they don't the report clearly describes the reason. Overall the decisions are sound and robust and uncertainties are acknowledged, reported and within normal levels.

This reviewer is confident that the data have been treated properly within the assessment model for the following reasons. The assessment team was able to answer all detailed questions posed by the review panel. The acknowledgments section mentions Stock Synthesis experts John F. Walter, Richard Methot and Mark Maunder for their advice on model configuration. The assessment team was able to match the previous model (CASAL).

Judging by model diagnostics and convergence the input data series are reliable and sufficient to support the assessment approach and findings.

This reviewer is concerned about the detailed description (Section 2.1 p. 14) of the different initial values for the von Bertalanffy growth parameters. Since those model parameters are estimated within the model, relatively small changes in the initial values should result in the same end result.

The sensitivity runs, runs starting from different starting values, and answers from the assessment panel assured the Review Panel that the model was converging properly. The growth curves are close.

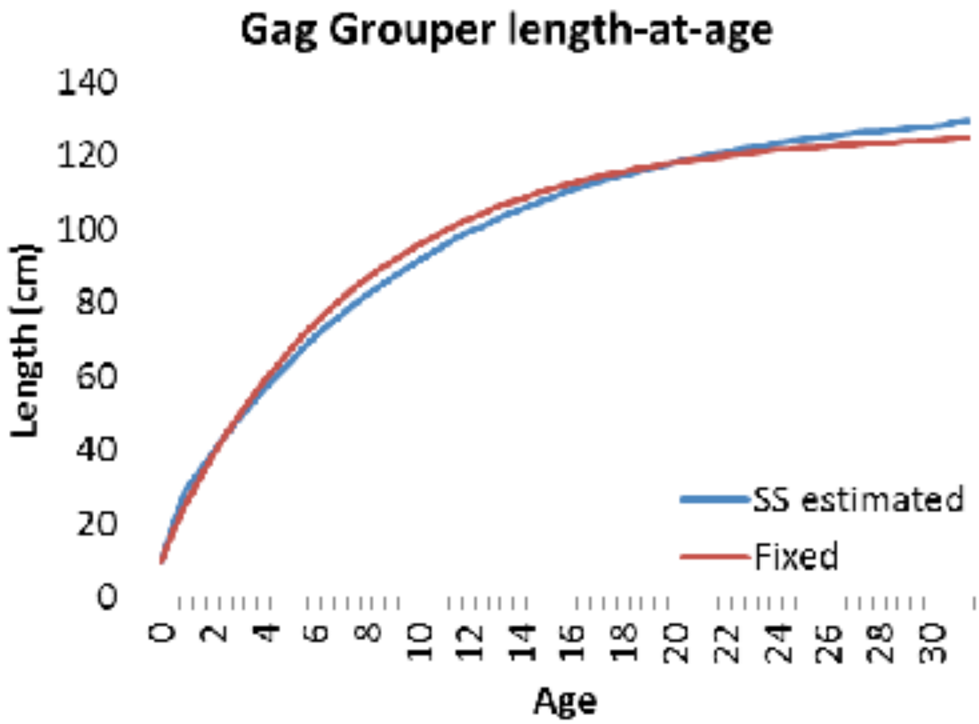


Figure 1: The estimated and fixed von Bertalanffy growth curves.

The panel suggested that the effective sample size to be used for the composition data should be based on the number of trips instead of the number of measured fish, because individual measured fish within a trip are likely correlated measurements.

The panel suggested that the recreational catches should not be treated as exact (very low CV), since they are estimated from angler surveys.

The GAG assessment has data from many sources. A wide array of sensitivities showed that the results and main conclusions were fairly robust to changes in individual sources, or different weightings of those sources (there were however differences in the first part of the estimated time series).

## *[GAJ]*

The data decisions for Greater Amberjack are overall sound and robust. The decisions mostly follow the data workshop, and where changes have been made they are clearly described in the assessment report. Changes include aggregating data sources into four identified fishing fleets, re-processing the discard numbers, and standardizing the catch per unit effort abundance indices.

With respect to uncertainties, the review panel was concerned with the composition sample sizes. Using the number of measured fish (although capped at 200) does not take into account that measurements taken from the same trip are likely correlated. The consequence of that is that the relative weight of the composition data could be too high. It proved difficult to obtain information on the number of trips sampled, which could be used as the effective sample size, so the panel instead suggested to simply (and subjectively) downweight the likelihood contribution of the composition data by multiplying the composition sample sizes by a constant smaller than one. This is a crude approach, but it serves to illustrate the sensitivity to the composition data.

The low sample sizes of the composition data raised the suggestion of simply leaving out the length compositions based on very few observations (e.g., n<50). This reviewer prefers not to leave out real observations, but ensure that they only enter with their appropriate weight.

The panel suggested that the recreational catches should not be treated as exact (very low CV), since they are estimated from angler surveys.

Some tables and figure have weights listed in pounds (lbs) and some in kilograms (kg). This is slightly confusing to the reader and it is recommended to only use one or the other.

The data are likely correctly treated in the assessment models (both in ASPIC and in Stock Synthesis). The assessment panelists were able to manipulate the models and data input to produce an additional run requested by the review panel, and no results were counterintuitive.

The input data series are not reliable and sufficient to support the current configuration of the assessment in Stock Synthesis. As detailed under TOR 2 the model has problems converging. This may sound like an academic issue, but it can have real consequences. The initial values are arbitrary values, and the model should be able to find exactly the same solution (minimum of the objective function) for a wide range of starting values. It can be sensible to choose initial values in the right ballpark to speed up computations, but the model should be sufficiently restricted that the solution is unique. For Greater Amberjack it is seen that the solution is sensitive to the initial value. In fact, when the initial values are changed in a range of 10%, the solution with respect to important output metrics (e.g., F[2012]/F[SPRTtgt]) also changes roughly in a range of 10%. After investigation at the review meeting this reviewer is optimistic with respect to solving this issue. It is expected that minor adjustments (e.g. fixing certain model parameters at reasonable values) to the model configuration will be enough to ensure proper convergence.

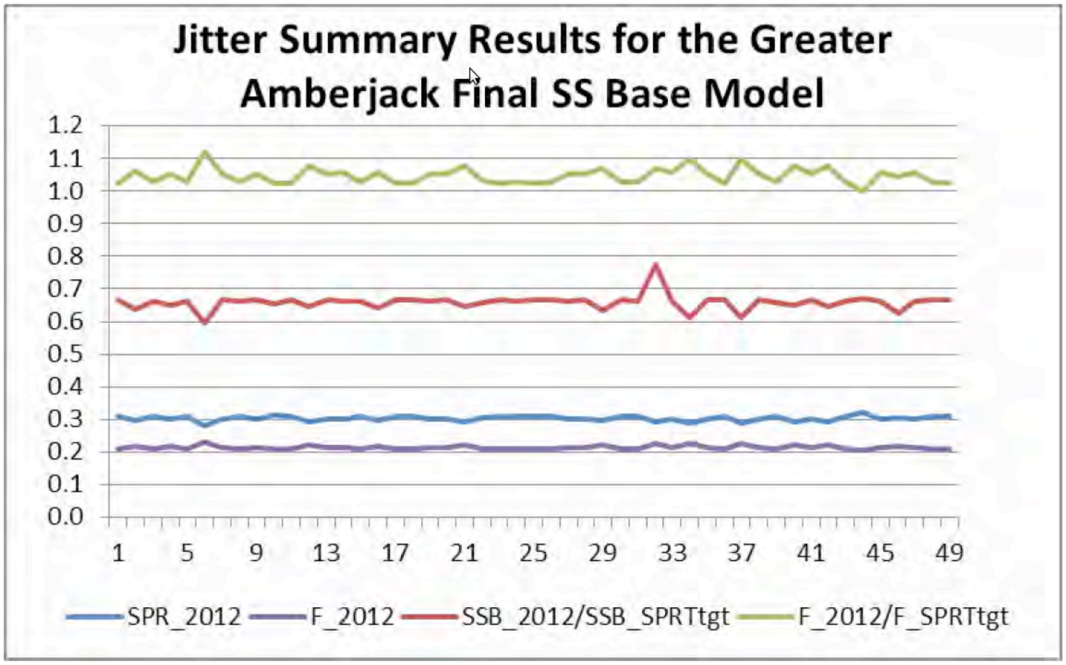


Figure 2: The influence of starting at different initial values on some important output metrics.

The data series are sufficient to inform the continuity ASPIC model, but changes in fishery selectivity have occurred, which the ASPIC model cannot take into account; so the ASPIC model is no longer the preferred method.

2. Evaluate the methods used to assess the stock, accounting for only the available data:

– Are the methods scientifically sound, robust, and appropriate for the available data?

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

The main method applied for both stocks is Stock Synthesis 3 (Methot 2013). Stock Synthesis is not a single model, but a modelling framework for full parametric stock assessments. It is well tested, as it has been applied to numerous thoroughly reviewed assessments. It can be configured to match almost any situation in terms of stock dynamics and observational likelihoods. In terms of data sources it can be configured to use many different data sources, from highly processed indices of abundances to fairly raw length and age data. An additional advantage of using such a widely used framework (in combination with graphics from the R-package r4ss) is that reviewers are familiar with it and know what to look for. Stock Synthesis is a sound and robust choice, which can be configured to be appropriate for the available data.

## *[GAG]*

For Gulf of Mexico Gag Grouper the model is likely to be properly configured and used consistent with standard practices.

The assessment team initiated their analysis by configuring a version of the GAG assessment in Stock Synthesis that closely mimics the previous accepted assessment, which was done in an independent framework CASAL (Bull et al. 2012). The assessment in Stock Synthesis was set to estimate the same number of parameters and the estimated values from CASAL were not used in the estimation process.

CASAL and Stock Synthesis are truly independent programs without any shared code base, as they are implemented via different software. The ability to get very similar, but not identical, results from both model configurations (Fig 3.1.5-3.1.8) strengthens confidence in both implementations and configurations.

The GAG assessment has a number of interesting non-standard features and the assessment team should be acknowledged for capturing these in the assessment model.

GAG are protogynous hermaphrodites (female at birth, then a proportion of the population transition into males). This feature was elegantly included in the model as a (fixed) logistic function.

A red tide event (an algal bloom, which had harmful effect on GAG) occurred in 2005, and is well supported by most indices. The assessment team investigated several ways to include this additional mortality in the model, and decided to include it as an additional fleet (a red tide fleet), which only has positive effort in 2005. In this approach the level of red tide associated mortality can be estimated and the same level can apply to all ages except age zero.

The fishing has changed over time due to the introduction and changing of bag-limits and size limits. This was accounted for in the model via different retention functions corresponding to the different periods, whereas the selectivity was kept constant.

A few issues did raise some concern, as described below.

The model allows for annual deviations from the stock recruitment relationship. For the first 10 years (1963-1972) the estimated logarithmic recruitment deviations are all estimated to be negative (Fig 3.2.69). This means that the time series of recruitment is artificially low for those years (Fig 3.2.68). According to the assessment team this is a technical matter where the model tries to adapt in that first period without composition data and those estimates should not be trusted. These low recruitment estimates may interfere with the model's ability to estimate the steepness of the stock-recruitment relationship. A number of sensitivity runs were performed (e.g. cutting off the first period and fixing steepness at different values) and this reviewer was satisfied that the key output metrics and all recent estimates were robust and not influenced by the initial low recruitment estimates. It would nevertheless be preferable if a configuration could be found where the issue could be avoided.

The recovery of GAG is predicted to be very strong and rapid. The reviewers noticed that this recovery is very much in synchrony with the Headboat index (Fig 3.2.13) and were concerned that the last points of this index alone was driving this recovery. Sensitivity runs were performed using down weighting and leaving out the Headboat index, and it was concluded that the recovery was robust to those changes.

Issues with fixing model parameters and ad hoc weighting of different data sources were raised (more under TOR 5), but these were certainly within standard practices.

## *[GAJ]*

For Gulf of Mexico Greater Amberjack it is this reviewer's impression that the optimal configuration of Stock Synthesis has not yet been found, which is not the same as rejecting the approach, as all models are imperfect; but it is this reviewer's opinion that addressing the issues identified below is needed before the assessment model is properly configured and used consistent with standard practices.

In addition to the model with Stock Synthesis, a model with ASPIC (A Stock-Production model Incorporating Covariates) was configured for Greater Amberjack. ASPIC is part of the NOAA Fisheries toolbox, and the model is scientifically sound and robust. It is however not an obvious choice for the data available for greater Amberjack. ASPIC is a surplus production model, and as such, it cannot deal with selectivity. An ASPIC model would not use the composition data.

The ASPIC model is supplied mainly as a continuity run, as it used to be the preferred method, but it is now giving a different result than it did previously. This is, according to the assessment panel, mainly due to the changes in the calculation of the indices outlined under TOR 1.

For the analysis in Stock Synthesis the main concern is the so-called jitter analysis. In a jitter analysis it is the intention to verify model convergence by starting all the model parameters in numerous different initial values and then see that the end result, in terms of the of objective function, model parameters, and important output metrics, is unchanged.

For Greater Amberjack, 50 runs were presented where the starting points had been randomly shifted by 10%. The result of this was not as expected. All 50 runs were reported at converged by the model. Of the 50 runs only one (run 33) gave a very different total likelihood (Fig. 3.2.2.1b), but the scale of the figure made it difficult to judge if other runs were all the same, or noticeable different. Even if the total likelihood is the same, the following is problematic. For the individual likelihood contributions of catch (Fig. 3.2.2.1c), survey (Fig. 3.2.2.1d) and the length and age compositions (Fig. 3.2.2.1e,f) it was clear that the model did not converge to one unique solution. Looking further at the important output metrics over the 50 runs (Fig. 3.2.2.1g) revealed that they too were changing when the starting point was changed. For example, the important F[2012]/F[SPRTtgt] showed that it varied about 10% when nothing except the starting point was changed (also 10%).

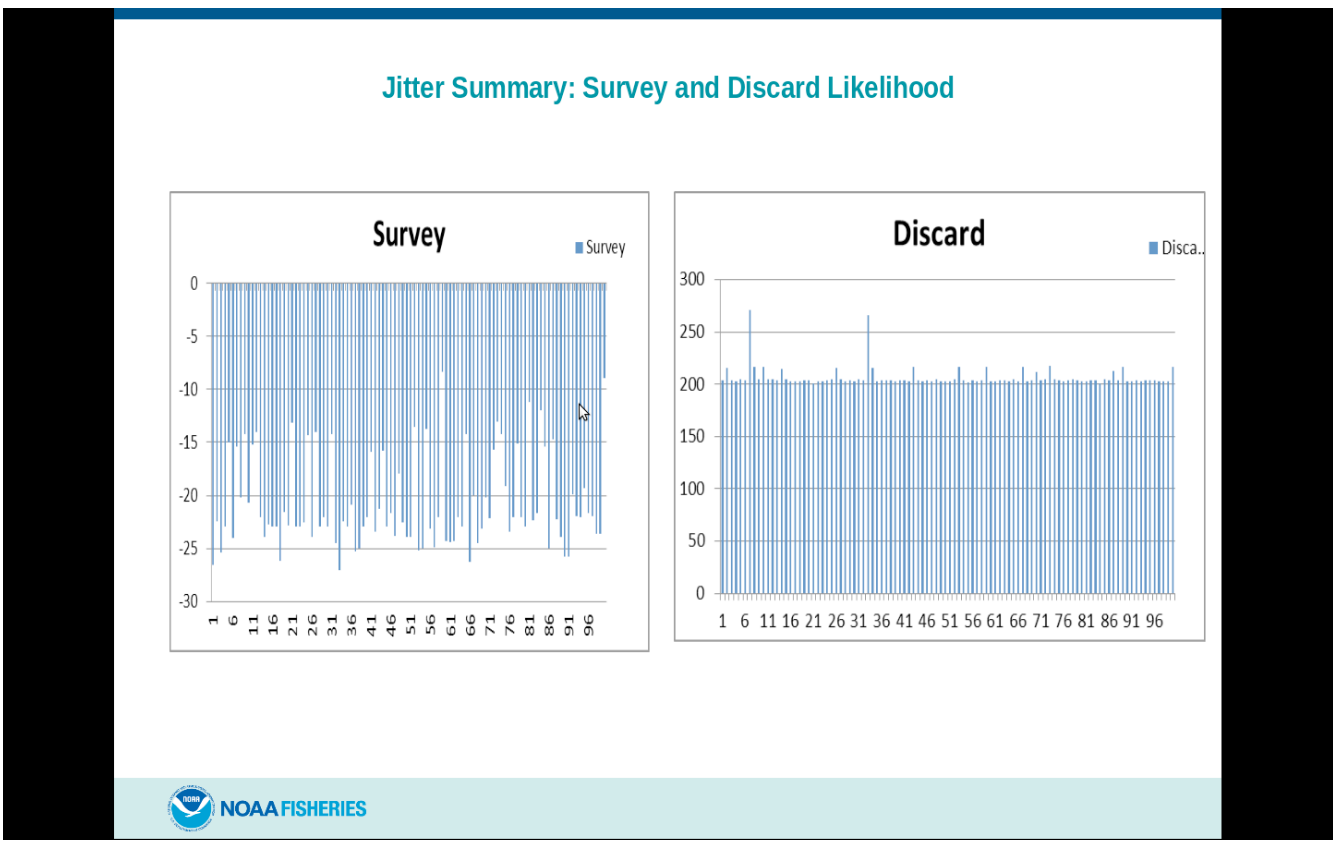


Figure 3: Result of 100 additional jitter runs from a slightly modified model run. All model runs are reported as converged, but the different likelihood components switch between runs.

Another place where the convergence problem is evident is in the profile likelihood with respect to the steepness parameter (Fig. 3.2.4.1), where sudden inexplicable high values occur in several places on otherwise nice convex curves.

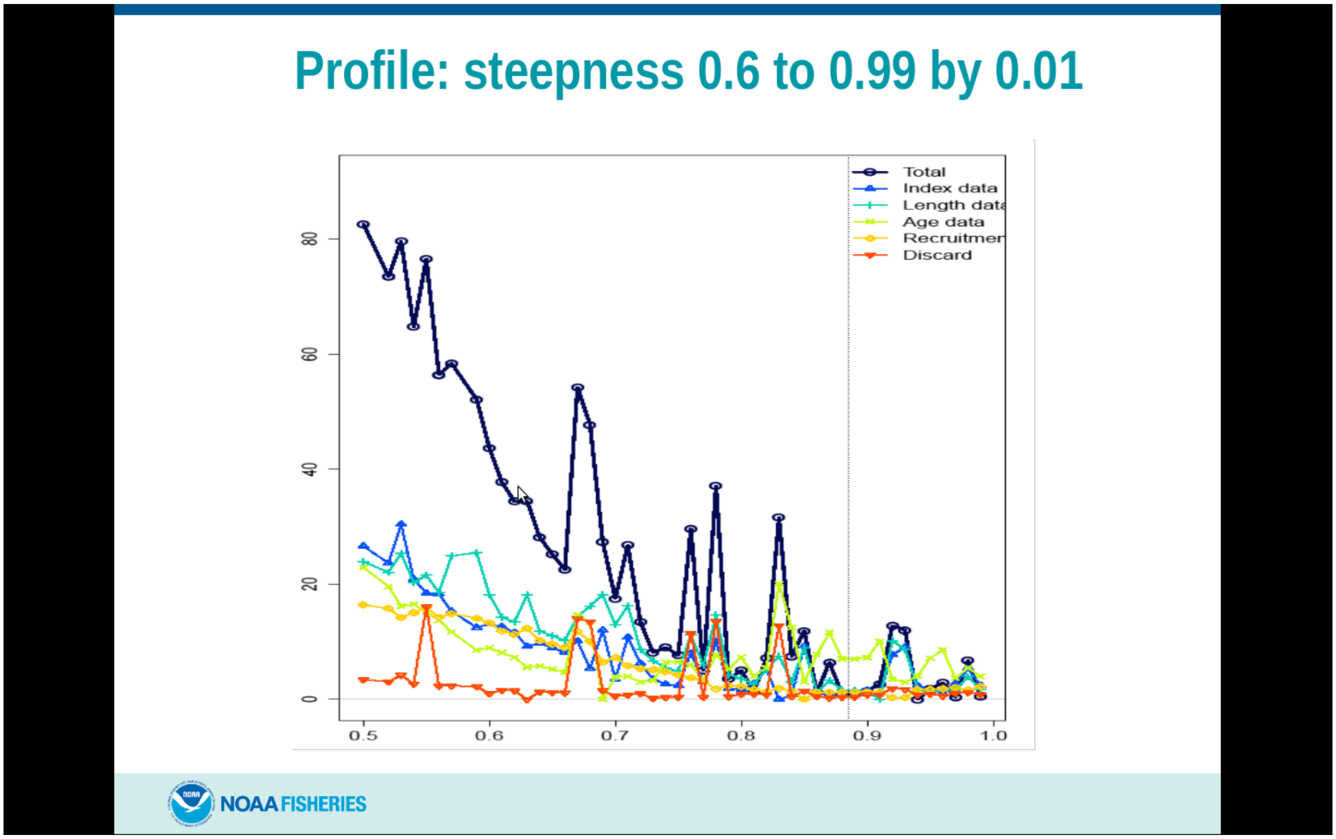


Figure 4: Profiling of the steepness parameter. Notice the unexpected peaks due to convergence problems.

It is important to note that the worry is not about lack of convergence, as that would be a lesser problem, which could solved in each individual case by choosing different starting points until convergence. The problem here is that Stock Synthesis (with its current configuration) reports convergence, but the point of convergence is highly dependent on the arbitrary starting values.

Stock Synthesis, in its current configuration, is not finding a unique minimum, which normally occurs if the model is non-identifiable (a.k.a. over-parameterized), which means that a change in some model parameters can compensate for a change in some other model parameters. To solve such an issue it is often necessary to fix some parameters, or assign priors to them. When looking for which model parameters to restrict it can be useful to look at correlations between model parameters, and to see if the standard deviations from a parametric bootstrap are similar to those derived from the inverse Hessian approximation. If they are very different it could be an indication of over-parameterization. Some of these methods were tried during the review meeting, and some results were improved, and some selectivity parameters were identified as the problematic ones, which means that there is reason to be optimistic with respect to solving this issue.

Another issue raised was the small sample sizes in the composition data. The concerned was that the composition data were given too much influence (as the sample sizes don't reflect independent samples). It was recommended to further down weight the composition data.

It also raised concern that the estimated fishing mortality of the commercial hand line fleet reached 2.0 around the year 1990 (Fig. 3.2.6.1), which is a very high fishing mortality, and thought to be unrealistic.

The results, with respect to the important output metrics, are not consistent between Stock Synthesis and ASPIC. This is not in itself surprising as Stock Synthesis uses more data (on age and length compositions) and is able to account for the different selectivities, but in this case it cannot simply be recommended to use the more advanced and detailed method, which uses the most data. This reviewer is of the impression that relative minor adjustments to the Stock Synthesis configuration will make it identifiable and suitable as basis for management of Greater Amberjack.

3. Evaluate the assessment findings with respect to the following:

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

– 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 quantitative estimates of 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?

## *[GAG]*

In the model for Gulf of Mexico Gag Grouper the ad hoc weighting of the different data sources, by assigning the level of precision to different data sources (e.g. CV for length-at-age were fixed at 0.13 and 0.01 for age 1 and age 31 respectively, sample sizes for composition data were capped at 200), makes it problematic to evaluate the individual abundance, exploitation, and biomass estimates, as the assigned weighting determines how closely the model fit will match the individual data sources. This was to some degree compensated for with a very thorough sensitivity analysis.

The fits, which are relevant to abundance estimates, are generally showing that the model is able to follow the main trends (Fig. 3.2.6-3.2.18). The residual patterns are certainly not perfect white noise, but the overall pattern is matched. The one important exception is the SEAMAP video (Fig. 3.2.16), where the estimated last year's strong recovery is not seen in the observations. This should be further investigated.

The estimates of the exploitation pattern are most visible in the fits to the composition data. The overall structure is captured. The misfits, seen at either the very low or high end, are not problematic, as they are based on very low observations/predicted numbers.

Sensitivity analysis showed that the biomass estimates are very reliable and consistent in the recent part of the time series, but more sensitive in the early part of the time series. The early part is especially sensitive to the steepness parameter of the stock-recruitment relationship, which is set to a value of almost one.

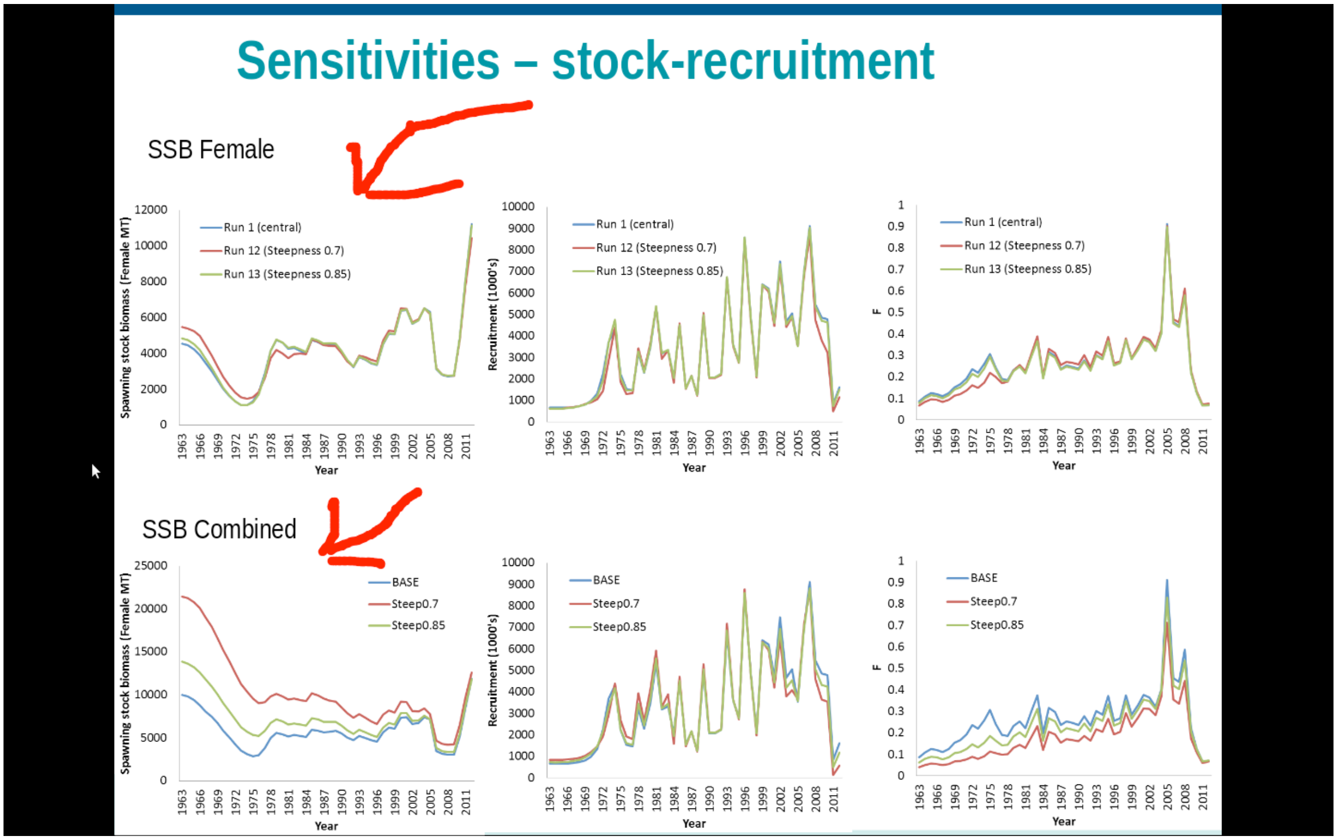


Figure 5: Sensitivity to the steepness parameters on the important estimated time series.

Notice that the female spawning stock biomass in the early part of the time series is less sensitive to the value of steepness than the combined spawning stock biomass.

The review meeting agreed to suggest the use of the combined reference points (and not the female only). With this convention in place the stock is overfished, or right on the border to be overfished (Fig. 2.3.108), but not undergoing overfishing (Fig. 2.3.110). However, across all sensitivity runs a rapid recovery of Gag is seen.

The stock-recruitment relationship is not very informative for Gag. Steepness is estimated by the base model to be approaching 1, which is a very high steepness. The panel recommended trying other steepness values (e.g. 0.7 and 0.85) to ensure that important management decisions are not based solely on the weak stock-recruitment relationship.

The stock-recruitment relationship is not very informative for this assessment, so the reference points, which depend on stock-recruitment, are also weakly defined. Using a fixed plausible value for the steepness parameter is a way to obtain better defined reference points (but also more subjective, but that can be a trade-off). The AW panel recommended use of SPR-based reference points, which can be calculated independently of the stock-recruitment relationship, which is appealing. The review panel recommended carrying forward the 6 combinations of steepness 0.99, 0.85, or 0.7, and spawning stock biomass, female only or combined. One thing which is well defined and consistent over a wide range of model options are the trends in spawning stock biomass and fishing mortality, and these can at least inform managers about trends and conditions.

## *[GAJ]*

The indices used in the ASPIC model have different selectivities, and their selectivities change over time e.g. due to introduced size limits. The ASPIC model cannot take these different selectivities into account, but applies the indices equally to the total biomass. The estimated time series from the ASPIC model (e.g. Fig. 3.4.3.1) are considered to be crude estimates, because of the selectivity issues, and only have limited usefulness to support inferences on stock status. Estimates from a correctly configured assessment model, which is able to take the different selectivities into account would be preferable.

The estimates of abundance, exploitation, and biomass from the Stock Synthesis model should be more reliable and useful, as Stock Synthesis is able to take the different selectivities into account. Unfortunately the convergence problem described under TOR 2 can affect these estimates too. From the figure showing the influence of starting at different initial values on some important output metrics (Fig 3.2.2.1g) it is evident that the estimates are affected. Neither the assessment report nor the slides contained a figure showing the estimated time series of abundance, exploitation, and biomass for the 50 different initial values, but the profiling of the steepness parameter (Fig 3.2.4.1) also has the convergence issue (seen as the unexpected peaks). For those profiling runs the individual time series are shown in the following figure (Fig. 6 taken from the slides). Seeing that these trajectories are parallel indicates that the relative trends from Stock Synthesis model are likely reliable in spite of the convergence problem. These relative estimates can be used to inform managers about stock trends and conditions, but the absolute values should not be used before the model configuration is improved to make the model identifiable. A suggested approach is outlined under TOR 2.

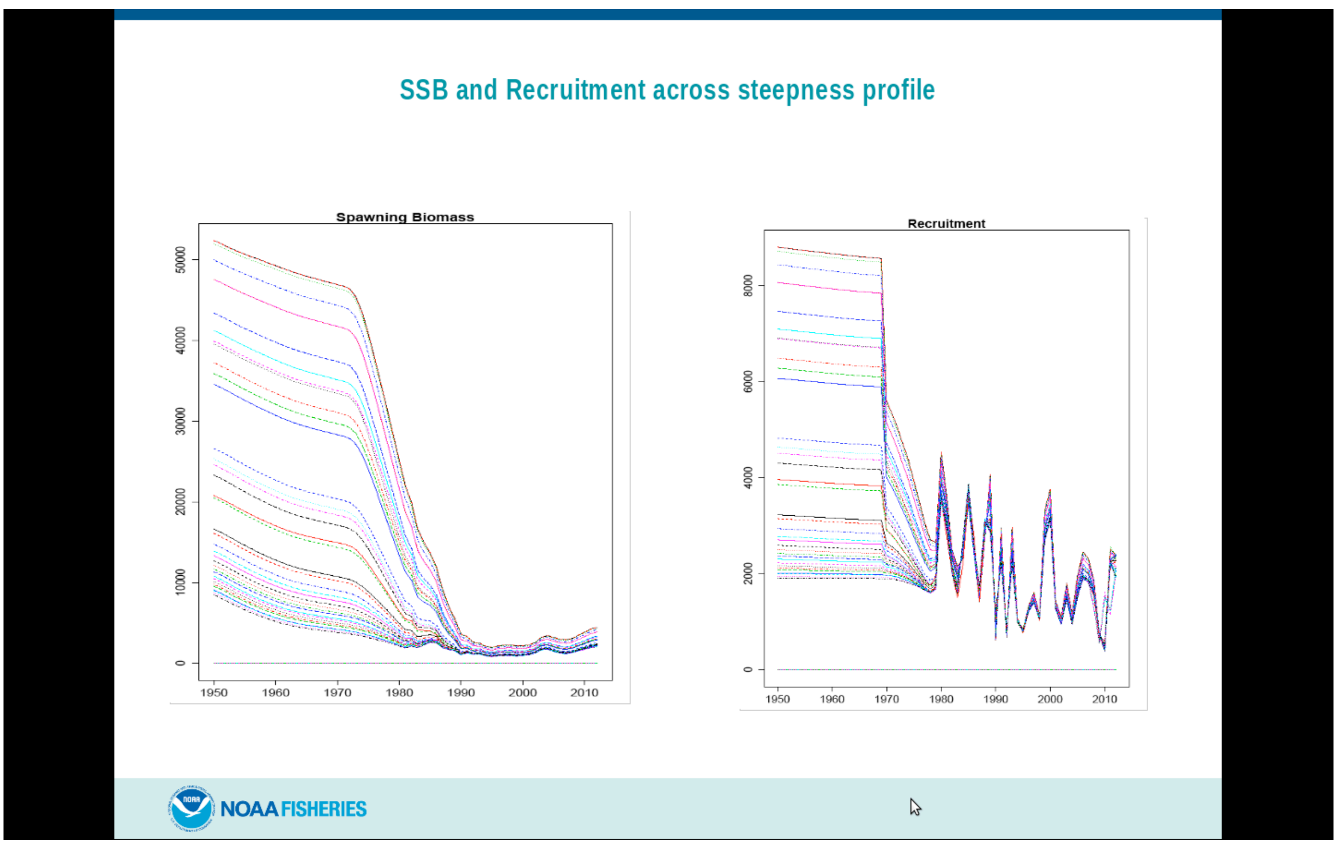


Figure 6: Time series of Spawning stock biomass and recruitment the individual steepness profiling runs.

Neither the ASPIC model nor the Stock Synthesis model (in its current configuration) is judged to give reliable estimates, so no definitive statements can be given about overfishing. Most indications from the ASPIC runs (Fig. 3.4.3.1) show that the stock has improved to be only borderline overfished, and that fishing mortality has decreased so the stock is no longer undergoing overfishing. However, indications from Stock Synthesis (Tab 3.2.8.1, Fig. 3.2.4.4b, Fig. 3.2.8.1b, and additional runs requested) suggest that the stock has increased in the last 5-10 years, but is slightly overfished, and that the stock is undergoing overfishing. Both results are borderline and given the convergence issue they will likely change once a more stable model configuration is used.

The stock-recruitment relationship is not very informative for this assessment, so the reference points, which depend on the stock-recruitment relationship, are also weakly defined. The configuration in the report estimated the steepness parameter of the stock-recruitment relationship, and the profile likelihood (vaguely) suggested that it was possible to do so (Fig. 3.2.4.1). Model improvements made during the review meeting however changed this, and the conclusion of the reviewers was that estimating steepness was very uncertain. Instead it was suggested to use values from a similar stock.

Neither the ASPIC model nor the Stock Synthesis model (in its current configuration) was identified as the final base run, as explained above, so it is not possible to evaluate the final estimates. An approach to obtain a stable configuration is suggested under TOR 2.

4. Evaluate the stock projections, addressing the following:

– Are the methods consistent with accepted practices and available data?

– Are the methods appropriate for the assessment model and outputs?

– Are results informative and robust, and useful to support inferences of probable future conditions?

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

The methods for the stock projections are implemented as part of Stock Synthesis, and as such have been frequently tested. They are in line with accepted practices, and appropriate for the assessment model and its outputs.

## *[GAG]*

The results are found to be informative and robust, and useful to support inferences of probable future conditions, and certainly within standard practices. Stochastic projections were not in the report, but some were presented at the meeting.

As mentioned the projections followed standard practices. Predicting the long term future is a difficult task, and these (and indeed most) projections must be used very carefully. It should be remembered these projections are based on a specific model with specific assumptions. The dynamics of the model are optimized to give good one year predictions, but here are being used to give long term projections. It is likely that events, which are not considered in the assessment model, will play a more important role in the long term. Examples could include decadal oceanic oscillations or temperature increases. After all, the models used to do daily or weekly weather forecasts are not the same models being used to predict long term climate changes.

### *[GAJ]*

For Greater Amberjack projections are presented both from the ASPIC model and for the Stock Synthesis model. Both sets of projections follow standard practices, but as neither of these can be considered a final base run, the projections are not considered useful in their present state. It is however expected that minor adjustments to the Stock Synthesis will solve the convergence issue, and then the projection procedure used in the report will give informative, robust, and useful projections.

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.

Uncertainties generally play an important role in assessment models. If a likelihood approach is applied, which it is for both stocks, the uncertainties determine the relative weighting of the different information sources entering the assessment. Furthermore it is important to correctly quantify the uncertainties on important output metrics to evaluate the risks of future fishing scenarios.

Both assessments supply standard deviations derived from the inverse hessian matrix of the objective function at its minimum. This is a standard output from most model fitting software, but it requires two things for these numbers to represent the uncertainty of our estimates: 1) the objective function should be well approximated by a quadratic function and 2) the model should correctly describe the observations including their observation uncertainties.

Item 1) is less of a concern, as standard approaches are available to circumvent this issue. For both assessments parametric bootstraps were provided. An alternative could be to use an MCMC approach. The review panel debated the difference, so a brief summary of the difference is described here. Parametric bootstrap simulates multiple independent data sets according to the assumptions in the model and the parameters estimated from the real observations. Estimation is then carried out for each data set. Parametric bootstrap is useful to obtain a simulation based, but otherwise exact, error propagation. It is also useful for revealing biases in the estimation procedure, as estimates can be compared to the assumed truth used when simulating the datasets. An MCMC approach simulates a Markov chain, such that its equilibrium distribution is the Bayesian posterior distribution of the model parameters (assuming flat priors where no prior is specified). MCMC approaches are useful for error propagation, but not for identifying biases, as no truth is assumed with which the estimates can be compared.

### *[GAG]*

Uncertainty about all model parameters are summarized by the estimated (Hessian based) standard deviations (Tab. 3.1.1) and supplemented by bootstrapped standard deviations and CV for selected quantities of interest based on 400 simulated data sets (Tab. 3.1.3 and Fig 3.2.73). Finally likelihood profiles were plotted for three important quantities: steepness, log of R0, and log of R1 (Fig. 3.2.65-3.2.67).

These are all good ways to represent uncertainties, but they are all based on the assumption that the model is describing the observations and their uncertainties correctly.

For the assessment of GAG it was decided to fix uncertainty parameters for different data sources at arbitrary values (e.g. CV for length-at-age were fixed at 0.13 and 0.01 for age 1 and age 31 resp., sample sizes for composition data were capped at 200). These arbitrary values are mainly based on experience and subjective judgment of the relative weighting between data sources, but they translate directly into scaling the estimated uncertainties.

The residual plots and plots of fitted lines and observations (Fig. 3.2.1-3.2.48) indicate that some fits are unrealistically close to the observations and others are too far off. Some of the plots of the fitted indices also showed auto-correlated errors (periods of only negative residuals followed by periods of positive residuals), which is in contrast with the assumed independent error structure.

Certain parameters are fixed in the model (e.g. natural mortality, some selection parameters and gender switching parameters (see table 3.1.1 for more)). This is necessary in these highly parameterized models, and some of these unacknowledged uncertainties would be picked up as larger observation uncertainties if the observation uncertainties were estimated.

These are all reasons to be skeptical about the estimated uncertainties, and suspect the real uncertainties to be larger. This is expected in such complex models, and the assessment team's choices are certainly within standard practices.

One additional source of uncertainty is the so-called model uncertainty, where the variations between different plausible model configurations are investigated. A wide array of sensitivity runs was presented in the assessment report, and additional runs were requested by the reviewers. These included different natural mortality, capping sample sizes by trip-number, and different data weightings. The overall impression from these sensitivity analyses were that the model results were very robust with respect to the recent period and important output metrics, but more sensitive in the first ca. 12 years of the data period.

To verify model convergence 50 runs were presented where the starting parameter values were randomly shifted 10%. Of the 50 runs 38 converged to a solution within 2 likelihood units of the base case. This is not optimal, but it was demonstrated that 42 of the runs provided similar key outputs.

The implications of uncertainty are clearly stated in the relevant graphs and tables.

## *[GAJ]*

For the ASPIC model for Gulf of Mexico Greater Amberjack a major part of the uncertainty originates from discarded fish. In recent years the discard is a large proportion (75%) of the total recreational catch (Fig 2.3.2.1). The amount of the discarded fish is based on the landings, but the size distribution of landed fish has changed due to regulations (Fig. 2.3.2.2). This raises uncertainty as to how this changed the size distribution of the discarded fish. Three methods were considered (details in section 2.3.2) "update", "low" (discard only undersized), and "high" (discard without respect to size). Results are reported for all three scenarios.

For the ASPIC model the reported uncertainties within each scenario are based on a 1000 bootstrap runs. This is a reasonable way to ensure that the uncertainties are correctly propagated in a non-linear model. These uncertainties are clearly stated w.r.t important output metrics in figure 3.4.3.1 and table 3.3.5.1.

In addition, sensitivities to the B1/K input ratio, discard mortality, and index weighting were conducted for the ASPIC results. These showed expected differences, but also that most conclusions about important output metrics were relatively stable (Fig. 3.4.6.1.2, 3.4.6.2.1).

The important output metrics from the ASPIC model are however sensitive to the new data compilation in SEDAR33 compared to the SEDAR9 update (Sec. 2.6.1). When 2009 is considered to be the final year and the new indices are used, B/BMSY doubles and F/FMSY halves (Fig. 3.4.7.1 and Tab. 3.4.7.1).

For the Stock Synthesis model the uncertainty about all model parameters are summarized by the estimated (Hessian based) standard deviations (Tab. 3.1.4.1). These standard deviations are however not likely to be reliable because the model with the present configuration is not identifiable (see description of this issue under TOR2). The quadratic approximation used when calculating these standard deviations is not appropriate if the objective function does not have a unique minimum.

Standard deviations based on a 1500 parametric bootstrap are also given (Tab. 3.2.2.1), and these have the advantage that they include all aspects of the implemented model when propagating uncertainties from observations to estimates of model parameters. This means that even if this model is not strictly identifiable (as some model parameters can compensate for others) this will, to some extent, be captured in the bootstrapped uncertainties. Hence, if the currently configured model were to be used as the basis for advice (which is not recommended) then it would be very important to use the bootstrapped uncertainties and not the Hessian based.

In addition to the main concern, which is that the model appears to be non-identifiable, a number of minor concerns were identified. For Greater Amberjack it was decided to fix uncertainty parameters for different data sources, to cap sample sizes, and to assign data weighting constants. These choices will directly influence the estimated uncertainties on derived model parameters, and therefore all uncertainty estimates are conditioned on those choices.

The residual plots and plots of fitted lines and observations (Fig. 3.2.1.3-3.2.1.6) indicate that some fits are close (compared to the data standard deviations) to the observations and others are far off. Many of the composition residuals systematically show positive residuals near the center of the distribution. Seeing residuals, which are less than perfect is however not uncommon in assessment models combining many different data sources.

Many sensitivity analyses were performed, some in the report and others requested by the review. The model was overall shown to be robust in the recent period, but more sensitive in the first period.

A final issue related to uncertainty was the steepness parameter. The configuration in the report estimated the steepness, and the profile likelihood (vaguely) suggested that it was possible to do so (Fig. 3.2.4.1). Model improvements made during the review meeting however changed this, and the conclusion of the reviewers was that estimating the steepness was very uncertain, and that it possibly would be more reliable to condition advice on a fixed chosen value.

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

The SEDAR procedure is, seen from this reviewer's perspective, very thorough, detailed, and well structured. In contrast to some assessment meetings this reviewer has attended, the SEDAR review meetings start out with a complete assessment report. This is extremely helpful to the reviewers and promotes a more in-depth review.

## *[GAG]*

The data workshop provides a long list of useful recommendations intended to improve the data quality. This reviewer supports those, but would like to add a recommendation to further study the SEAMAP video index. The strong recent recovery of Gag estimated by the assessment is not matched by this index, even though that should be expected. If the SEAMAP video index is for some reason not working as expected, then it should be known. If it is working as expected, then it leads to question the other data sources.

The assessment report lists five useful recommendations. Here this reviewer would like to emphasize number 3) Appropriate methods to deal with changing selectivity over time. Fully parametric assessment models such as Stock Synthesis express everything via model parameters to be estimated. This means that there is a trade-off between the number of model parameters that are tractable and the flexibility of the model. When a change in selectivity needs to happen, it leads to more model parameters. In contrast, time series models (a.k.a. State-space models) works by describing an underlying unobserved process via a few parameters, and then this process can adapt to changing selectivity, for example. These methods are gaining momentum in ICES, and they could possibly also be useful here. More information can be found at http://www.stockassessment.org.

In addition to the suggested recommendations, it is suggested to:

R1. Investigate if different stock-recruitment formulations (Ricker, hockey-stick, or even a non-parametric approach) will work better and avoid the problems with estimating steepness.

R2. Investigate if it is possible to estimate more parameters from the data and fix fewer. For instance, it could be of interest if the model could internally estimate the parameters in the logistic function which determines transition from female to male. This may however require additional data.

R3. Further investigation of the change from female to male. It is very important to know at what ages are the fish immature female, mature female, and mature male? Are these transitions constant in time? Do they adapt to the gender ratio in the population?

## *[GAJ]*

The data workshop lists recommendations regarding all aspects of the data, which are all useful. This reviewer would like to emphasize the continuation of the collection of composition data, which is greatly needed to inform integrated models such as Stock Synthesis.

The recommendations in the assessment are all useful. This reviewer would like to support number 3) to develop fisheries independent sampling programs for size and age compositions, and number 5) to develop program to increase sampling of discarded fish. Both of these would have helped inform the modeling approach for Greater Amberjack.

In addition it is suggested to:

R1. Investigate a better way derive the effective sample size used for the composition data. The number of measured fish may be too high due to within trip correlations, so a better proxy is needed to correctly weight this information source.

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

## *[GAG]*

The assessment of gag grouper is fairly complete, and the assessment team has done a great job of taking into account a number of non-standard issues. A few issues can be listed, which should be considered before the next assessment.

The difficulty in determining the steepness parameter within the assessment model has consequences for the reference points, and at present the best suggestion seems to be to use the "Robin Hood" approach of borrowing a value from a similar, but more data-rich stock.

There should be further investigation of the pros and cons of using female-only reference points versus using gender combined reference points.

There should be further study of the SEAMAP video index, as that data source is in conflict with the recent recovery seen in other data sources.

A lot of parameters are set to fixed values in this assessment, including many variance parameters. It would be interesting to see if some of these can be estimated instead.

Finally the sensitivity analysis focused on the point estimates, but an investigation of the effect and size of the fixed variance parameters (and priors) should focus on the estimated standard deviations of the important output metrics.

## *[GAJ]*

The most important issue to address for the Greater Amberjack assessment is the convergence issue of the configured Stock Synthesis model. To solve this issue it is recommended to fix some parameters, or assign priors to them. When looking for which model parameters to restrict it can be helpful to look at: 1) correlations between model parameters or 2) investigate if the standard deviations from a parametric bootstrap are similar to those derived from the inverse Hessian approximation. If they are very different, it can be an indication of over-parameterization. Some of these methods were tried during the review meeting, and some results were improved, and some selectivity parameters were identified as the problematic ones. To diagnose if the issue has been solved, repeat the jitter analysis and verify that in cases where the model reports convergence it is in fact converging to the same minimum, with the same important output metrics.

Once the convergence issue has been resolved it will also be useful to:

Revisit the estimation of the steepness parameter, or to borrow a value from a similar but data rich stock.

Investigate a better way derive the effective sample size used for the composition data. The current suggestion of simply down weighting the entire likelihood contribution of all compositions can only be considered a crude first approach.

8. Prepare a Peer Review Summary Report summarizing the Panel's evaluation of the stock assessment and addressing each Term of Reference. Develop a list of tasks to be completed following the workshop. Complete and submit the Peer Review Summary Report in accordance with the project guidelines.

The report was prepared in collaboration with the other reviewers. This reviewer especially helped shape the sections to evaluate the methods used to assess the stocks (TOR 2) and the sections considering the uncertainties in the assessments (TOR 5).

# Appendix 1: Bibliography of materials provided for review

SEDAR 33- Gulf of Mexico Gag and Greater Amberjack Document List

| Document Number | Species | Title | Authors |
|---|---|---|---|
| **Data Workshop Documents** | | | |
| SEDAR33-DW01 | Both | Greater Amberjack and Gag Grouper Catches from Mississippi Laboratories Fishery Independent Surveys | Pollack and Ingram |
| SEDAR33-DW02 | Gag | Protection of Grouper and Red Snapper Spawning in Shelf-Edge Marine Reserves of the Northeastern Gulf of Mexico: Demographics, Movements, Survival and Spillover Effects | Koenig and Coleman |
| SEDAR33-DW03 | Gag | Fishery-Independent Indices of Abundance for Gag (Mycteroperca microlepis) in the Northeastern Gulf of Mexico, with Intrinsic Habitat Quality Controlled and Contrasted | Lindberg, Christmas, and Marcinek |
| SEDAR33-DW04 | GAJ | Characterization of Greater Amberjack Discards in Recreational For-Hire Fisheries | Sauls and Cernak |
| SEDAR33-DW05 | Gag | Characterization of Gag Discards in Recreational For-Hire Fisheries | Sauls and Cernak |
| SEDAR33-DW06 | Gag | Condition and Relative Survival of Gag Mycteroperca microlepis Discards Observed Within a Recreational Hook-and-Line Fishery | Sauls |
| SEDAR33-DW07 | Gag | Natural Mortality of Gag Grouper from 1950 to 2009 Generated by an Ecosim Model | Chagaris and Mahmoudi |
| SEDAR33-DW08 | Gag | Satellite derived indices of red tide severity for input for Gulf of Mexico Gag grouper stock assessment | Walter, Christman, Landsberg, Linton, Steidinger, Stumpf, Tustison |
| SEDAR33-DW09 | Gag | Use of otolith microchemistry to improve fisheries-independent indices of recruitment for gag (Mycteroperca microlepis): linking estuarine nurseries to nearshore reefs in the eastern Gulf of Mexico | Jones, Switzer, Houston, and Peebles |
| SEDAR33-DW10 | Both | Incorporating various Gulf of Mexico Integrated Ecosystem Assessment products into the Stock Synthesis Integrated Assessment Model framework | Schirripa, Methot, et al. |

| | | | |
|---|---|---|---|
| SEDAR33-DW11 | Gag | Evaluation of natural mortality rates and diet composition for gag (Mycteroperca microlepis) in the West Florida Shelf ecosystem using the individual-based, multi-species model OSMOSE | Gruss, Schirripa, Chagaris, Drexler, Simons, Verley, Shin, Karnauskas, Penta, de Rada, and Ainsworth |
| SEDAR33-DW12 | GAJ | Seasonal movement and mixing rates of greater amberjack in the Gulf of Mexico and assessment of exchange with the South Atlantic spawning stock. | Murie, Parkyn, and Austin |
| SEDAR33-DW13 | Gag | Observer reported size distribution and discard characteristics of Gulf of Mexico Gag from the commercial vertical line and bottom longline fisheries | Johnson |
| SEDAR33-DW14 | GAJ | Observer reported size distribution and discard characteristics of Gulf of Mexico Greater Amberjack from the commercial vertical line and bottom longline fisheries | Johnson |
| SEDAR33-DW15 | Gag | SEAMAP Reef Fish Video Survey: Relative Indices of Abundance of Gag | Campbell, Rademacher, Felts, Noble, Felts, and Salisbury |
| SEDAR33-DW16 | GAJ | SEAMAP Reef Fish Video Survey: Relative Indices of Abundance of Greater Amberjack | Campbell, Rademacher, Felts, Noble, Felts, and Salisbury |
| SEDAR33-DW17 | Gag | Update concerning species misidentifications in the commercial landing data of gag groupers and black groupers in the Gulf of Mexico | Chih |
| SEDAR33-DW18 | Gag | Use of the Connectivity Modeling System to estimate movements of gag grouper (Mycteroperca microlepis) recruits in the northern Gulf of Mexico | Karnauskas, Paris, Zapfe, Gruss, Walter, and Schirripa |
| SEDAR33-DW19 | Both | A meta-data analysis of discard mortality estimates for gag grouper and greater amberjack | Lombardi, Campbell, Sauls, and McCarthy |
| SEDAR33-DW20 | Gag | Gag Life History Working Group Draft Working Document | Gag Life History Working Group |
| SEDAR33-DW21 | GAJ | Greater amberjack (Seriola dumerili) otolith ageing summary for Panama | Allman, Trowbridge, and |

| | | | |
|---|---|---|---|
| | | City laboratory (2009-2012) | Barnett |
| SEDAR33-DW22 | Gag | Age, length, and growth of gag (Mycteroperca microlepis) from the northeastern Gulf of Mexico: 1978-2012 | Lombardi, Fitzhugh, and Barnett |
| SEDAR33-DW23 | Gag | Catch and bycatch of gag grouper in the Gulf of Mexico shark and reeffish bottom longline fishery based on observer data | Gulak and Carlson |
| SEDAR33-DW24 | GAJ | Catch and bycatch of greater amberjack in the Gulf of Mexico shark and reeffish bottom longline fishery based on observer data | Gulak and Carlson |
| SEDAR33-DW25 | GAJ | Regional stock structure of greater amberjack in the southeastern United States using otolith shape analysis | Crandall, Parkyn, and Murie |
| SEDAR33-DW26 | Both | Relative abundance of gag grouper and greater amberjack based on observer data collected in the reeffish bottom longline fishery | Carlson, Gulak, Scott-Denton, and Pulver |
| SEDAR33-DW27 | GAJ | Non-lethal sex determination of greater amberjack with direct application to sex ratio analysis of the Gulf of Mexico stock | Smith, Murie, and Parkyn |
| SEDAR33-DW28 | Gag | Gag Mycteroperca microlepis Findings from the NMFS Panama City Laboratory Trap & Camera Fishery-Independent Survey – 2004-2012 | DeVries, Gardner, Raley, and Ingram |
| SEDAR33-DW29 | GAJ | Preliminary Release Mortality of Gulf of Mexico Greater Amberjack from Commercial and Recreational Hand-line Fisheries: Integration of Fishing Practices, Environmental Parameters, and Fish Physiological Attributes | Murie and Parkyn |
| SEDAR33-DW30 | GAJ | Greater Amberjack Index Report Cards: Adobe Portfolio | Index Working Group |
| SEDAR33-DW31 | Gag | Gag Index Report Cards: Adobe Portfolio | Index Working Group |
| | | | |
| **Assessment Workshop Documents** | | | |
| SEDAR33-AW01 | Both | Fisheries-independent data for gag and greater amberjack from reef-fish video surveys on the West Florida Shelf, 2008-2012 | Switzer, Keenan, McMichael, and Ingram |
| SEDAR33-AW02 | Gag | Length frequency distributions for gag groupers in the Gulf of Mexico from | Chih |

| | | | |
|---|---|---|---|
| | | 1984-2012 | |
| SEDAR33-AW03 | Gag | Age frequency distributions estimated with reweighting methods for gag groupers in the Gulf of Mexico from 1991 to 2012 | Chih |
| SEDAR33-AW04 | GAJ | Length frequency distributions and reweighted age frequency distributions for greater amberjacks in the Gulf of Mexico from 1984-2012 | Chih |
| SEDAR33-AW05 | GAJ | Greater Amberjack Seriola dumerili Findings from the NMFS Panama City Laboratory Trap & Camera Fishery-Independent Survey – 2004-2012 | DeVries, Raley, Gardner, and Ingram |
| SEDAR33-AW06 | Gag | Summary of fishery-independent surveys of juvenile gag grouper in the Gulf of Mexico | Ingram, Pollack, and McEachron |
| SEDAR33-AW07 | Gag | Standardized catch rate indices for gag grouper (Mycteroperca microlepis) landed by the commercial longline fishery in the U.S. Gulf of Mexico during 1990-2012 | Cass-Calay |
| SEDAR33-AW08 | Gag | Standardized catch rates for gag grouper from the United States Gulf of Mexico handline fishery during 1990-2009 | |
| SEDAR33-AW09 | Gag | Standardized catch rates for gag grouper from the Gulf of Mexico headboat fishery during 1986-2011 | |
| SEDAR33-AW10 | Gag | Standardized Catch Rates of Gulf of Mexico Gag Grouper from Recreational Inshore, Charterboat, and Private Boat Fisheries (MRFSS) 1986 to 2010 | |
| SEDAR33-AW11 | GAJ | Standardized Catch Rates for Greater Amberjack from the commercial longline and commercial handline fishery in the U.S. Gulf of Mexico | |
| SEDAR33-AW12 | GAJ | Standardized Catch Rates for Greater Amberjack from the Gulf of Mexico Headboat Fishery 1986-2011 | |
| SEDAR33-AW13 | GAJ | Standardized Catch Rates of Greater Amberjack from the Gulf of Mexico Recreational Charterboat and Private Boat Fisheries (MRFSS) 1986 to 2012 | |
| SEDAR33-AW14 | | | Calay |
| SEDAR33-AW15 | Gag | Standardized catch rates for gag | Bryan |

| | | | |
|---|---|---|---|
| | | grouper from the United States Gulf of Mexico handline fishery during 1990-2009 | |
| SEDAR33-AW16 | Gag | Standardized Catch Rates of Gulf of Mexico Gag Grouper from Recreational Inshore, Charterboat, and Private Boat Fisheries (MRFSS) 1986 to 2010 | Bryan |
| SEDAR33-AW17 | Gag | Standardized catch rates for gag grouper from the Gulf of Mexico headboat fishery during 1986-2010 | Bryan |
| SEDAR33-AW18 | GAJ | Commercial Indices of Abundance for Greater Amberjack in the Gulf of Mexico | Saul |
| SEDAR33-AW19 | GAJ | Standardized catch rates for greater amberjack from the Gulf of Mexico headboat fishery during 1986-2010 | Rios |
| SEDAR33-AW20 | GAJ | Standardized Catch Rates of Greater Amberjack from the Gulf of Mexico Recreational Charterboat and Private Boat Fisheries (MRFSS) 1986 to 2012 | Rios |
| SEDAR33-AW21 | Gag | Red tide mortality on gag grouper 1980-2009 | Gray, Ainsworth, Chagaris, and Mahmoudi |
| SEDAR33-AW22 | Both | Ageing error matrices for SEDAR33: gag grouper and greater amberjack | Lombardi |
| SEDAR33-AW23 | Gag | Meta-analysis of release mortality in the gag grouper fishery | Campbell, Lombardi, Sauls, and McCarthy |
| SEDAR33-AW24 | Gag | Natural mortality rates and diet patterns of gag grouper (Mycteroperca microlepis) in the West Florida Shelf ecosystem in the 2000s: Insights from the individual-based, multi-species model OSMOSE-WFS | Gruss, Schirripa, Chagaris, Drexler, Simons, Verley, Shin, Oliveros-Ramos, Karnauskas, and Ainsworth |
| | | | |
| **Review Workshop Documents** | | | |
| SEDAR33-RW01 | | | |
| | | | |
| **Reference Documents** | | | |
| SEDAR33-RD01 | GAJ | SEDAR 9: Gulf of Mexico Greater Amberjack Stock Assessment Report | SEDAR |
| SEDAR33-RD02 | GAJ | 2010 SEDAR 9 Update: Gulf of Mexico Greater Amberjack | SEDAR |
| SEDAR33-RD03 | Gag | SEDAR 10: Gulf of Mexico Gag Stock Assessment Report | SEDAR |

| | | | |
|---|---|---|---|
| SEDAR33-RD04 | Gag | 2009 SEDAR 10 Update: Gulf of Mexico Gag | SEDAR |
| SEDAR33-RD05 | GAJ | Gulf of Mexico Greater Amberjack Management History | GMFMC |
| SEDAR33-RD06 | Gag | Gulf of Mexico Gag Management History | GMFMC |
| SEDAR33-RD07 | Gag | Status of Gulf of Mexico Gag Grouper: Results and Projected Implications of the Revisions and Sensitivity Runs Suggested by the Grouper Review Panel | SEFSC |
| SEDAR33-RD08 | Gag | Final Model for Gulf of Mexico Gag Grouper as Recommended by the SEDAR Grouper Review Panel: Revised results and projections | SEFSC |
| SEDAR33-RD09 | Gag | Stock Assessment of Gag in the Gulf of Mexico: SEDAR Update Assessment Rerun | SEFSC |
| SEDAR33-RD10 | GAJ | Preliminary Analysis of Tag and Recapture Data of the Greater Amberjack, Seriola dumerili, in the Southeastern United States | McClellan and Cummings |
| SEDAR33-RD11 | GAJ | Trends in the Gulf of Mexico Greater Amberjack Fishery through 1998: Commercial Landings, Recreational Catches, Observed Length Frequencies, Estimates of Landed and Discarded Catch at Age, and Selectivity at Age | Cummings and McClellan |
| SEDAR33-RD12 | GAJ | Age, growth, and reproduction of greater amberjack, Seriola dumerili, in the southwestern north Atlantic | Harris |
| SEDAR33-RD13 | GAJ | Age, Growth and Sex Maturity of Greater Amberjack (Seriola dumerili) in the Gulf of Mexico | Murie and Parkyn |
| SEDAR33-RD14 | Gag | Annual Indices and Trends of Abundance for Gag (Mycteroperca microlepis) on the Shallow Continental Shelf in the Northeastern Gulf of Mexico | Lindberg, Christman, Marcinek, and Bohrmann |
| SEDAR33-RD15 | Gag | Stock Identification of Gag, Mycteroperca microlepis, Along the Southeast Coast of the United States | Chapman, Sedberry, Koenig, and Eleby |
| SEDAR33-RD16 | Gag | A Tag and Recapture Study of Gag, Mycteroperca microlepis, off the Southeastern U.S. | McGovern, Sedberry, Meister, Westendorff, Wyanski, and |

| | | | |
|---|---|---|---|
| | | | Harris |
| SEDAR33-RD17 | Gag | Empirical Use of Longevity Data to Estimate Mortality Rates | Hoenig |
| SEDAR33-RD18 | Gag | Discard composition and release fate in the snapper and grouper commercial hook-and-line fishery in North Carolina, USA | Rudershausen, Buckel, and Williams |
| SEDAR33-RD19 | Gag | Modeling Protogynous Hermaphrodite Fishes Workshop | Sheperd, Shertzer, Coakley, and Caldwell |
| SEDAR33-RD20 | GAJ | Field Based Non-Lethal Sex Determination and Effects of Sex Ratio on Population Dynamics of Greater Amberjack, Seriola dumerili | Smith |
| SEDAR33-RD21 | Neither | Release mortality in the red snapper fishery: a meta-analysis of three decades of research | Campbell, Driggers, Sauls, and Walter |
| SEDAR33-RD22 | | | |
| SEDAR33-RD23 | | | |
| SEDAR33-RD24 | | | |

# Appendix 2: Statement of Work

## Statement of Work

### External Independent Peer Review by the Center for Independent Experts

### SEDAR 33 Gulf of Mexico Gag and Greater Amberjack Assessment 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 Representative (COR), 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 33 will be a compilation of data, benchmark assessments of the stocks, and an assessment review conducted for Gulf of Mexico gag and greater amberjack. The review panel is ultimately responsible for ensuring that the best possible assessments are provided through the SEDAR process. The stocks assessed through SEDAR 33 are within the jurisdiction of the Gulf of Mexico Fishery Management Council and the state waters 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:** Three CIE reviewers shall have the necessary qualifications to complete an impartial and independent peer review in accordance with the tasks and ToRs described in the SoW herein. The CIE reviewers shall have expertise in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the tasks of the scientific peer-review described herein. 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 participate and conduct an independent peer review during the SEDAR 33 panel review meeting scheduled in Miami, Florida during February 24-27, 2014.

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

**Tasks prior to the meeting:** The contractor shall independently select qualified reviewers that do not have conflicts of interest to conduct an independent scientific peer review in accordance with the tasks and ToRs within the SoW. Upon completion of the independent reviewer selection by the contractor's technical team, the contractor shall provide the reviewer information (full name, title, affiliation, country, address, email, and FAX number) to the contractor officer's representative (COR), who will forward this information to the NMFS Project Contact no later than the date specified in the Schedule of Milestones and Deliverables. The contractor shall be responsible for providing the SoW and stock assessment ToRs to each reviewer. The NMFS Project Contact will be responsible for providing the reviewers with the background documents, reports, foreign national security clearance, and other information concerning pertinent meeting arrangements. The NMFS Project Contact will also be 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 COR prior to the commencement of the peer review.

Foreign National Security Clearance: The reviewers shall participate during a panel review meeting at a government facility, and the NMFS Project Contact will be responsible for obtaining the Foreign National Security Clearance approval for the reviewers who are non-US citizens. For this reason, the reviewers shall provide by FAX (not by email) the 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/.

Pre-review Background Documents: Approximately two weeks before the peer review, the NMFS Project Contact will send (by electronic mail or make available at an FTP site) to the COR the necessary background information and reports (i.e., working papers) for the reviewers to conduct the peer review, and the COR will forward these to the contractor. In the case where the documents need to be mailed, the NMFS Project Contact will consult with the COR on where to send documents. The reviewers are responsible only for the pre-review documents that are delivered to the contractor in accordance to the SoW scheduled deadlines specified herein. The reviewers shall read all documents deemed as necessary in preparation for the peer review.

**Tasks during the panel review meeting:** Each reviewer shall conduct the independent peer review in accordance with the SoW and stock assessment ToRs, and shall not serve in any other role unless specified herein. **Modifications to the SoW and ToRs shall not be made during the peer review, and any SoW or ToRs modifications prior to the peer review shall be approved by the COR and contractor.** Each 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 stock assessment ToRs as specified herein. The NMFS Project Contact will be responsible for any facility arrangements (e.g., conference room for panel review meetings or teleconference arrangements). The NMFS Project Contact will also be responsible for ensuring that the Chair understands the contractual role of the reviewers as specified herein. The contractor can contact the COR and NMFS Project Contact to confirm any peer review arrangements, including the meeting facility arrangements.

**Tasks after the panel review meeting:** Each reviewer shall prepare an independent peer review report, and the report shall be formatted as described in **Annex 1**. This report should explain whether each stock assessment ToR was or was not completed successfully during the SEDAR meeting. If any existing BRP or their proxies are considered inappropriate, each independent report shall include recommendations and justification for suitable alternatives. If such alternatives cannot be identified, then the report shall indicate that the existing BRPs are the best available at this time. Additional questions and pertinent information related to the assessment review addressed during the meetings that were not in the ToRs may be included in a separate section at the end of an independent peer review report.

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.

**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 at Miami, Florida from February 24-27, 2014
3) Conduct an independent peer review in accordance with the ToRs (**Annex 2**).
4) No later than March 14, 2013, each CIE reviewer shall 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 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**.

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

| | |
|---|---|
| 24 January 2014 | CIE sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact |
| 3 February 2014 | NMFS Project Contact sends the stock assessment report and background documents to the CIE reviewers. |
| 24-27 February 2014 | Each reviewer shall conduct an independent peer review during the panel review meeting in Miami, Florida |
| 14 March 2014 | CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator |
| 28 March 2014 | CIE submits CIE independent peer review reports to the COR |
| 4 April 2014 | The COR 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 COR within 10 working days after receipt of all required information of the decision on changes. The COR 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 COR 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 COR (William Michaels, via William.Michaels@noaa.gov).

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

**Distribution of Approved Deliverables:** Upon acceptance by the COR, the CIE Lead Coordinator shall send via e-mail the final CIE reports in *.PDF format to the COR. The COR will distribute the

CIE reports to the NMFS Project Contact and Center Director.

**Support Personnel:**

William Michaels, Program Manager, COR
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
Northern Taiga Ventures, Inc.
10600 SW 131st Court, Miami, FL 33186
shivlanim@bellsouth.net Phone: 305-383-4229

Roger W. Peretti, Executive Vice President
Northern Taiga Ventures, Inc. (NTVI)
22375 Broderick Drive, Suite 215, Sterling, VA 20166
RPerretti@ntvifederal.com Phone: 571-223-7717

**Key Personnel:**

NMFS Project Contact:

Ryan Rindone, SEDAR Coordinator
2203 N. Lois Avenue, Suite 1100
Tampa, Florida 33607
Ryan.Rindone@gulfcouncil.org Phone: 813-348-1630

**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. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed. The CIE independent report shall be an independent peer review of each ToRs.

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: Tentative Terms of Reference for the Peer Review

## SEDAR 33 Gulf of Mexico Gag and Greater Amberjack Assessment Review

1. Evaluate the data used in the assessment, addressing the following:
   a) Are data decisions made by the Assessment Workshop 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 the methods used to assess the stock, taking into account the available data.
   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 with respect to 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, addressing 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 Assessment workshop 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.
- Provide recommendations on possible ways to improve the SEDAR process.

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

## Annex 3: Tentative Agenda for

## SEDAR 33 Gulf of Mexico Gag and Greater Amberjack Assessment Review
## February 24-27, 2014
## Miami, FL USA

***Monday***

| | | |
|---|---|---|
| **1:00 p.m.** | **Convene** | |
| **1:00 – 1:30** | **Introductions and Opening Remarks** | **Rindone** |
| | *- Agenda Review, TOR, Task Assignments* | |
| **1:30 – 5:00** | **Assessment Presentations and Discussions** | **SEFSC** |
| **5:00 p.m. - 6:00 p.m.** | **Panel Work Session** | **Powers** |

***Tuesday***

| | | |
|---|---|---|
| **8:00 a.m. – 11:30 a.m.** | **Assessment Presentations and Discussions** | **SEFSC** |
| **11:30 a.m. – 1:00 p.m.** | **Lunch Break** | |
| **1:00 p.m. – 3:30 p.m.** | **Panel Discussion** | **Powers** |
| | **-** *Assessment Data & Methods* | |
| | *- Identify additional analyses, sensitivities, corrections* | |
| **3:30 p.m. – 3:45 p.m.** | **Break** | |
| **3:45 p.m. – 5:00 p.m.** | **Panel Discussion** | **Powers** |
| | *- Continue deliberations* | |
| | *- Review additional analyses* | |
| **5:00 p.m. - 6:00 p.m.** | **Panel Work Session** | **Powers** |

***Tuesday Goals*****:** Initial presentations completed, sensitivities and modifications identified.

***Wednesday***

| | | |
|---|---|---|
| **8:00 a.m. – 11:30 a.m.** | **Panel Discussion** | **Powers** |
| | *- Review additional analyses, sensitivities* | |
| | *- Consensus recommendations and comments* | |
| **11:30 a.m. – 1:00 p.m.** | **Lunch Break** | |
| **1:00 p.m. – 3:30 p.m.** | **Panel Discussion** | **Powers** |
| | *- Final sensitivities reviewed.* | |
| | *- Projections reviewed.* | |
| **3:30 p.m. – 3:45 p.m.** | **Break** | |
| **3:45 p.m. – 5:00 p.m.** | **Panel Discussion/Work Session** | **Powers** |
| | *- Review Consensus Reports* | |
| **5:00 p.m. - 6:00 p.m.** | **Panel Work Session** | **Powers** |

***Wednesday Goals:*** Final sensitivities identified, preferred models selected, projection approaches approved, final results made available. Summary report drafts begun.

***Thursday***

| | | |
|---|---|---|
| **8:00 a.m. – 12:00 p.m.** | **Panel Work Session** | **Powers** |

**12:00 p.m. ADJOURN**

# Appendix 3: List of review meeting participants

SEDAR 33 Attendance:

**Monday, February 24, 2014**

Mike Armstrong-CIE
Benjamin Blount-SSC
Meaghan Bryan- SEFSC
Shannon Cass-Calay-SEFSC
Nancie Cummings- SEFSC
John Froeschke- Council Staff
Jeff Iseley- SEFSC
Neil Klaeur-CIE
Michael Murphy-FWC
Anders Nielsen-CIE
Sean Powers-SSC
Ryan Rindone-Council Staff
Adyan Rios- SEFSC
Skyler Sagarese- SEFSC
Charlotte Schiaffo-Council Staff
Jessica Stephen- SEFSC
Greg Stunz-SSC
Jake Tetzlaff- SEFSC
Katyana Vert-Pre-UF

**Tuesday, February 25, 2014**

Mike Armstrong-CIE
Benjamin Blount-SSC
Meaghan Bryan- SEFSC
Shannon Cass-Calay-SEFSC
Nancie Cummings- SEFSC
John Froeschke- Council Staff
Jeff Iseley- SEFSC
Neil Klaeur-CIE
Michael Murphy-FWC
Anders Nielsen-CIE
Sean Powers-SSC
Ryan Rindone-Council Staff
Adyan Rios- SEFSC
Skyler Sagarese- SEFSC
Charlotte Schiaffo-Council Staff
Jessica Stephen- SEFSC
Greg Stunz-SSC
Jake Tetzlaff- SEFSC
Katyana Vert-Pre-UF

**Wednesday, February 26, 2014**

Mike Armstrong-CIE
Benjamin Blount-SSC
Meaghan Bryan- SEFSC
Shannon Cass-Calay-SEFSC
Nancie Cummings- SEFSC
John Froeschke -Council Staff
Justin Grubiche-PEW
Jeff Iseley- SEFSC
Neil Klaeur-CIE
Michael Murphy-FWC
Anders Nielsen-CIE
Sean Powers-SSC
Ryan Rindone-Council Staff
Adyan Rios- SEFSC
Skyler Sagarese- SEFSC
Charlotte Schiaffo-Council Staff
Jessica Stephen- SEFSC
Greg Stunz-SSC
Jake Tetzlaff- SEFSC
Katyana Vert-Pre-UF

**Thursday, February 27, 2014**

Mike Armstrong-CIE

Benjamin Blount-SSC
Meaghan Bryan- SEFSC
Shannon Cass-Calay-SEFSC
Nancie Cummings- SEFSC
John Froeschke -Council Staff
Neil Klaeur-CIE
Anders Nielsen-CIE
Sean Powers-SSC
Ryan Rindone-Council Staff
Adyan Rios- SEFSC
Skyler Sagarese- SEFSC
Charlotte Schiaffo-Council
Staff
Greg Stunz-SSC
Jake Tetzlaff- SEFSC