External Independent Peer Review for the Center for Independent Experts

SEDAR 23 Gulf of Mexico and South Atlantic Goliath Grouper Review Workshop

15–17 November 2010
Key West, Florida

Sven Kupschus

Summary:

The review panel was asked to assess the available information on the status of the US goliath grouper stock in relation to stock benchmarks and management parameters to advise management, based on the work provided by the data and assessment workshops of the SEDAR 23 process. The work conducted by these two workshops had been thorough, but the output was highly uncertain owing to the unusual circumstances surrounding the data sources, especially in terms of the availability of age and growth and catch information since the implementation of the moratorium on fishing in 1990, as well as the general data paucity and poor knowledge of the basic biology and stock structure of the species. Despite this, a lot of information addressing the respective terms of reference had been collated and included in a catch-free assessment model developed for the species and used in SEDAR 6.

Two sets of output from the model, a continuity run consistent with the setting developed in SEDAR 6 and a new base model differing in the choice of indices and implementing an age-dependent rate of natural mortality in place of the previously used constant M were presented. However, the assessment was deemed to be inappropriate for management purposes mainly because it produced unrealistic trajectories of the stock dynamics necessary to inform management.

Fishing mortality trends indicated a slowly increasing trend from the early part of the time-series until 1980, after which it spiked dramatically to peak in 1985 before dropping rapidly down to earlier levels in 1989 and then plummeting to near zero levels as the moratorium was implemented. Although it was not possible to examine the cause of this unusual trend experimentally, there were a number of issues in the specification of the model and input data heavily constraining the model with respect to assessing trends in F.

Trends in spawning stock biomass indicated sharp declines in historic values until the moratorium, with recent exponential growth forming the basis of some recovery. The latter is driven by the recent very sharp increases in the indices, but little or no information on the early part of the historic decline is provided by the data, so the assessment response is surprising, if not worrying. More important though is that the recruitment dynamics specified in the model are unable to follow the recent increase in the indices and also that the stock–recruit dynamics modelled in the assessment are in contrast to the assessment results.

In summary, therefore, the assessment was found not to be adequately reflecting stock dynamics, with the concerns being sufficiently severe for the assessment to be deemed inappropriate for providing management advice. It was therefore not possible to assess the status of the stock in relation to potential benchmarks, or to characterize the sensitivity. Although the model is unable to provide management advice, it is not necessarily the modelling approach or the effort at the date workshop and assessment workshop that is at fault, but more likely the quality and availability of data that was contributing to this regrettable outcome.

Workshop results by term of reference:

TOR1: Evaluate the adequacy, appropriateness, and application of data used in the assessment.

It is not possible to evaluate this TOR entirely in isolation of the stock assessment procedure because the appropriateness of data is conditional on the assumptions under which they are used. Consequently I interpret this TOR with respect to the more generally available information for goliath grouper in terms of its biology and stock structure, dealing with more assessment-specific comments such as the utility of the selectivity information and the appropriateness of indices under TOR 2, despite this information strictly speaking being provided by the data workshop. Figures to illustrate the points in this report were taken from the assessment report or provided by Joe O'hop at the meeting.

Stock structure: This is a frequently overlooked, but essential, component to understand in any management procedure. In this case information was presented to the panel suggesting that there was little evidence of a significant substructure to the goliath grouper population encountered in US waters on the basis of genetic information. Consequently, it was suggested that a single assessment covering the whole stock with its centre of density located in Florida's offshore waters would be appropriate.

However, the level of mixing required for genetic homogeneity at the gene level is much less than that required to fulfil the assumptions in most stock assessment methodologies. In addition, there may be ontogenetic stages such as the larva where mixing is great, but there can still be different stocks if these are effectively isolated at the time of settlement. The biological information available to the panel suggests that the species exhibits a high degree of site fidelity, with tagged fish frequently being encountered on subsequent dives in virtually identical locations, suggesting that the appropriate unit of management may well be the reef. In stark contrast with this is the evidence from Collins and Barbieri that indicates that some individuals travel significant distances (175 km) over a short time-frame, implying that full mixing of the populations is at least possible.

In reality the large distances apparently travelled likely imply spawning migrations of an otherwise mostly sessile fish species. Therefore, whether a single population or a reef-based unit are appropriate for the assessment depends upon whether the fishery targets spawning aggregations or exploits the fish when movement is minimal, leading potentially to significant local depletion, especially if individual fish return to the same reef post-spawning (which is less clear from the evidence presented).

In any case, the spatial disaggregation of the stock is likely to be much greater than the resolution of the available data, so few options exist other than to attempt a single stock unit assessment. This practice is common, but it does require a very careful examination of the results shown that mixing is incomplete and that not all available information is representative of the population as a whole. If carried out appropriately, it can, however, provide suitable management information.

Reproduction: Reproductive studies on the species are still inconclusive in terms of the assumed protogynous hermaphrodite status, and particularly confusing is the fact that Bullock *et*

al. suggest that males of the species mature younger than females. In addition to this uncertainty, the data workshop provided proxy information based on a meta-analysis of similar species. This uncertainty, therefore, coupled with the fact that the assessment represents a single-sex assessment, indicates that calculations of biological parameters such as steepness that are important in the considerations of stock productivity and management benchmarks are unlikely to be appropriately estimated even if the model had been able to estimate the underlying age structure of the population accurately.

Although the panel agreed in general that the information was lacking or poor, the degree of concern differed among individual members of the panel. In particular, the summary report indicates concerns about metadata based on “early studies or from proxy species”. I share the concern that these values are likely to be inaccurate, but feel personally that reproduction is no more likely to have changed over time or to differ between species than natural mortality, so find the acceptance in principle of using a meta-analysis estimate of one parameter and rejection of another unacceptable, especially given the fact that at least some of the reproductive parameters are estimated within the model, so would be somewhat corrected by the evidence in the data.

Some concerns were raised in discussion over the possible presence of an Allee effect with regards to reproduction, because it was suggested that goliath grouper needed a minimum spawning aggregation density to reproduce effectively, and that this was not taken into account in the assessment. However, this opinion is based largely on conjecture because little is known about the behaviour of the species at low density, and it may simply be that at low density the remaining fish aggregate in the same densities at fewer sites. In any case the current rate of population increase suggested by the indices implies a much faster rate of reproduction than the current parameter set in the assessment is able to replicate, so one has to assume that biomass is currently greater than the level at which an Allee effect might be observed.

Natural mortality: The assessment workshop spent some time developing an age-specific natural mortality vector as an improvement to the constant M=0.12 used at SEDAR 6. Benchmark values in absolute terms are likely to be highly sensitive to the choice of M and will strongly reflect the estimates of catch that can sustainably be taken from the fishery. However, no data were available to estimate M for the stock, so M was based on a meta-analysis of related species using longevity (Tmax) as the basis of correlation. The oldest goliath grouper aged by Bullock *et al*. was 37 years, which provided a sensible starting point for the analysis. However, the age data relate to a period when the stock was already heavily exploited so that the true Tmax is likely to be higher. A number of other values ranging up to 80 years were used to evaluate the sensitivity of the analysis. This had little effect on the equilibrium values of SSB/SSB(spr50%) in the assessment, mainly because fishing mortality was estimated close to 0 since 1990, but it did influence the rate of recovery to the long-term equilibrium and hence the current estimate of stock status.

Significant time was spent at the review workshop trying to follow the methodology of how the M vector was derived, and there seemed to be some inconsistencies in the way that the Hoenig method had been applied here and in other grouper stocks, which is something that should certainly be examined prior to a future assessment. However, my opinion is that some of the other shortcomings in the data and assessment methodology provide far greater uncertainty than the effect that a Tmax within a reasonable range (that examined by the assessment

workshop) might have on estimates of stock status and benchmarks mainly because the estimates of F are unreasonably low in the assessment as a whole.

Growth: The available data on growth was sparse. Although the data spanned a good number of ages, the years in which it was collected and the potential bias in the data from recent times (taken from individuals suffering incidental mortality in cold kill events), along with the unavailability of data for many of the specimens collected then precluded any analysis of interannual or inter-cohort differences in size-at-age, so growth was estimated to be constant across all cohorts. For age-based assessments in general, this presents at worst a small problem, but the catch-free model, although formulated in an underlying age structure, estimates the abundance driven by the choice of selectivities of the indices, which in this case have been derived from the available length information ignoring differences in the variability of size-at-age over time. In addition and probably more importantly, selectivity estimates use data that suggest extensive overlap between the size at different ages, so there is little contrast in the available information to reflect variation in cohort strength accurately within the model, as suggested by the unrealistically smooth estimates of recruitment derived from the model (Figure 1).

Selectivities: Growth and mortality estimates were used to determine the selectivities of the individual indices used in the assessment. This work is usually conducted within the assessment process, and although not very clear from the information provided, it appeared to be treated as an input parameter in the model.

Two methods of estimating selectivity (referred to as vulnerabilities) for the index information were examined. The first was based on a purely probabilistic age/length key, but because this did not take reasonable account of the different sizes-at-age was rejected as unrealistic and replaced with an iterative method accounting for the likely differences in abundance-at-age using a constrained iterative method that minimized the predicted and observed abundances-at-age for each index in turn. Although theoretically appropriate to determine the selectivity of a gear, such a procedure in many ways defeats the object of an assessment model by trying to come up with a unified view of stock dynamics, so this process should be carried out within the assessment model, not external to it. The effects of this on the model are discussed in detail under the section relating to TOR2.

Indices: All indices show a general increase in the abundance of goliath grouper recently and although in general more information could have been presented on spatio-temporal aspects and a better description of the background to the indices provided, the panel took the view that these had been examined satisfactorily for consistency and robustness. Where additional information on the development and the background for each index was really missing was in the information that could have suggested what the indices were representative of, such as

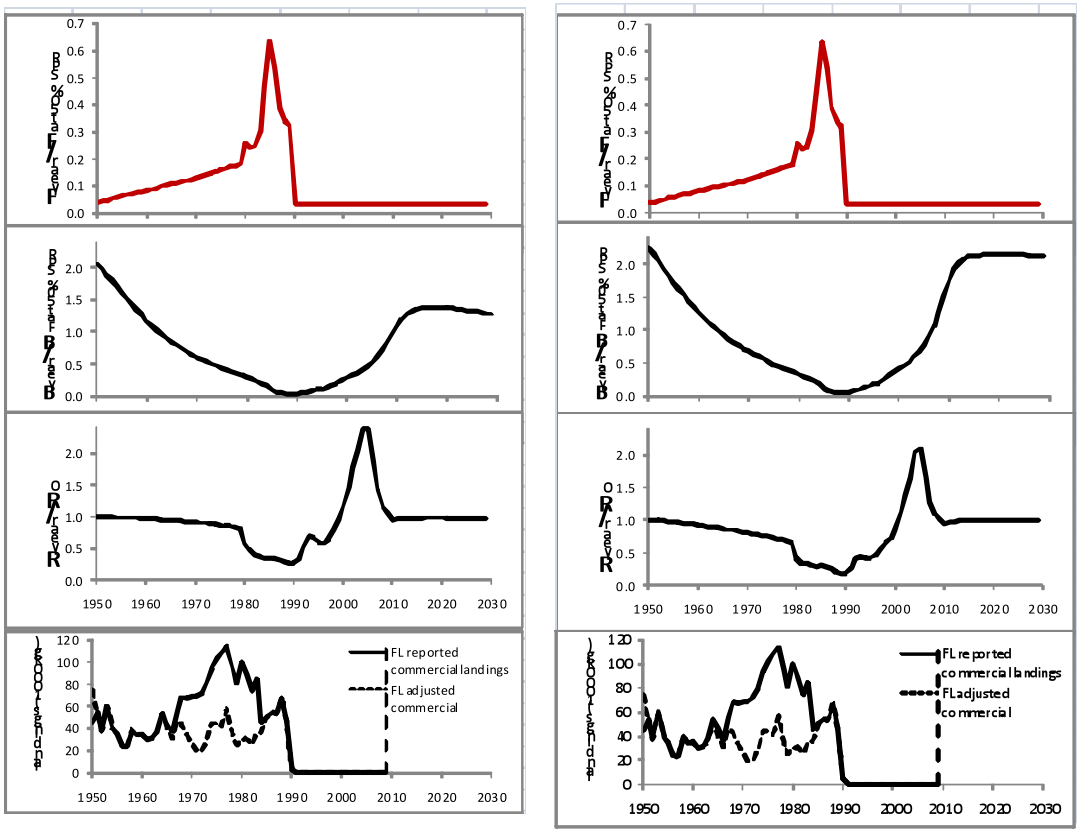

Figure 1: Summary stock output for the continuity and proposed base model assessment (base run left, continuity run right) showing F , SSB, recruitment and landings and corrected landings (top to bottom).

Specific age ranges/habitats/subpopulations and which possible sources of bias had been accounted for. This would have been important because the indices, although all suggesting a recent rise, disagreed substantially on the rate of this rise even for indices using the same methodology (i.e. REEF east and west), and also because the historic trends, where available, diverged significantly.

The assessment ultimately interprets the index as viewed through the selectivity 'filter'. If this is incorrectly specified or estimated in the model, then the index may well be appropriate, but inappropriately used, in the model. Without the ancillary information it is not possible to at least judge the appropriateness of the indices and selectivity information qualitatively.

ENP creel survey**:** This index is based on recreational angler intercept surveys and claims to be covering mainly ages 2–8, the basis of which is not entirely clear to me because it seems to provide no age information. The basis may be size, but then it is very dependent on the appropriateness of the general age information or may simply be on the basis that it is derived from juvenile habitat, but then I would assume that age 8 is rather high because the majority of fish are mature by that stage. It is almost certain that there has been some retargeting of the recreational fishery over time, not necessarily specifically with respect to goliath grouper, but any retargeting of recreational catches is likely to influence the encounter rate with that species, so it has to be assumed that there is some bias in the trends derived from the data.

MRFSS: Like the ENP index, this index is based on angler interviews corrected for differences in a number of covariables, but unlike the former index it is based on a proportion of positive encounters mainly because insufficient data were available to estimate the average size of fish caught correctly and hence to develop a biomass index. The nature of the index is therefore binomial, so its ability to guide the model in terms of the relative trends in abundance is limited at the extremes of the range of abundances. Although technically correct as an index, the way the data are treated in the model is therefore inappropriate from a likelihood perspective.

One of the major benefits of including these data in the model, however, is that it is a fishery-based index, which could in theory provide some information on the likely selectivity of any future recreational fishery, while still providing abundance information during the period of the moratorium as catch-and-release fish are included. The applicability of this information is somewhat determined by the accuracy of the data, and given the pressure to reopen the fishery by recreational fishers in conjunction with the protected species classification of goliath grouper, there may well be some biases in the data of an unquantifiable size or even in terms of a qualitative estimate of its direction.

REEF: This index is based on a visual census index from a volunteer diver program, but the data are available only for very rough classes, i.e. abundances of 0, 1, 2–10 and >10 fish, which appears to have been frequently used as an explanation as to why the index was not considered reliable, as also suggested by the review workshop summary report. However, SEDAR-AW-01 explains how this is standardized into a relative abundance index (formula 4) which should make it appropriate for use in stock assessment. However, for this to work as intended, this should be done within the assessment process rather than deriving a maximum likelihood of abundance external to the process and then providing it as a log abundance index. From a variance and likelihood perspective, the two are not the same thing; for example, once all abundance values fall within one of the multiple frequency categories, there is no longer any information in the data to suggest which way the index is moving, until at least some observations fall outside the category. There will still be a maximum likelihood estimate from the standardized index, which will be the same as in previous years. In reality there is no evidence to suggest that abundance has changed, so if other indices were to indicate increases in abundance, then using the index as provided by SEDAR-AW-01 information would be contradictory and would provide incorrect parameter estimates, but not so if the likelihood estimator was applied within the model (i.e. given some abundance N, what is the chance of obtaining the observed index).

The REEF index has been split into two separate components, mainly because the time-series available for the eastern and western portions differ in length, and it was not immediately clear how one might combine the two sets of information into a single index. Both indices indicate an increase, but they provide different information as to the likely rate of increase. Given the similarity of the sampling protocol, adding such conflicting data sources to an assessment should be considered carefully. The differences in trend may reflect differences in the distribution as the increasing population expands spatially. However, the model is unable to account for such spatial variations, so the inclusion of both indices in an assessment may present significant problems.

DeMaria: This index is based on the observations of a commercial fisher. The original steep decline in the number of fish aggregating at four wrecks involved in what appeared to be

spawning behavior ultimately represent the evidence base for the moratorium. However, the basis of the index calculation and whether it could be determined to be representative of stock size is unclear from the evidence provided to the panel. Both the fact that the initial two years, the period of the steepest decline, have been removed for assessment purposes and that the data were never originally designed to be used as an index left a number of panel questions about the appropriateness of the index unanswered.

Don DeMaria was present at the meeting and was able to shed some light on his interpretation of the information. One point he made seemed to imply that people were going farther and farther afield to exploit goliath grouper, which in conjunction with the limited movement of the species may explain the extremely rapid decline at the index sites, at least in part as an effect of local depletion, so care needs to be taken in the interpretation of this information in the assessment, particularly in the light of the constant or slightly increasing landings information available. The index crucially provides some information regarding the decline of the adult population during the period where the model is not constrained in other ways, and F is actually relatively free to fluctuate. This index is therefore important, but it should not be allowed to dominate the assessment in the way it is currently doing, especially given the above-mentioned concerns regarding spatial issues and survey design.

Catches: Catches are unimportant to the catch-free model, but the data workshop provided such information as additional information and to advance the option of utilizing other models in future. A catch history was presented at the review workshop, but there was considerable uncertainty in the estimates of landings prior to the moratorium, apparently driven largely by the efforts of a single dealer determined to create the impression that Lee County was the “goliath grouper capital of the world”. Corrected catch estimates excluding this dealer changed the impression of rising catches in the 1980s to a much more level trend which, given that there was little indication of a great increase in effort then, would be contrary to the assertion that the stock was collapsing during the 1980s. This hyperstability in catches may be explained by spatial expansion of the fishery, or by the aggregating behavior of the species at spawning time. The SRA model would be immune to such effects, provided the indices are insensitive, but the sharp decline in the DeMaria index certainly suggests that it may be susceptible to such assumptions.

The uncertainty surrounding the true catches during the 1980s is likely to have a great impact on both management benchmark estimates and current status indicators, so a catch-based analysis should only be considered if this uncertainty is resolved.

Discarding: No information on discarding was provided by the data workshop, so this remains unknown for the historic period, as does current catch-and-release mortality. Recently, recreational hook-and-line fishers are likely to have the greatest impact because commercial longliners are excluded from a large percentage of the area occupied by the species. Consequently, mortality is likely to be low except in cases where the fish are brought rapidly to the surface from depth. The catch-free model is unaffected by the lack of such estimates, but the stock-reduction analysis presented as additional information at the review workshop uses a constant catch assumption of 2 tonnes. Given the rise in abundance of the stock this will likely lead to an underestimate of F, because the 2 tonnes represent continually decreasing portions of the biomass. Realistically, the rate of mortality should be proportional to the encounter rate, which in turn is driven by near constant effort and a sharply increasing abundance.

Summary: The data workshop provided a great breadth of information from which it should theoretically be possible to determine stock history. The problem with the information in most cases is not its variability, but its potential for bias, because much of the biological work relied on meta-analysis, the age information is based on historic growth information, and the indices, some of which imply implausible rates of increase recently, have generally been constructed *post hoc* from work conducted for other purposes or other species. Nevertheless, if such disparate information could be corroborated in a single model it could provide a sufficiently robust basis for management for the stock.

TOR2: Evaluate the adequacy, appropriateness, and application of methods used to assess the stock.

The review panel was presented with two runs of the catch-free assessment model, the first being the SEDAR 6 continuity run, and the second the SEDAR 23 base run. The model setups differed mainly in their choice of indices, the latter including the new MRFSS index and the REEF south-east index rather than the two-point fisher interview information. In addition, M was treated as age-specific rather than constant.

Summary output from the two runs suggests similar trends in stock dynamics, with a historically low fishing mortality rising slowly up to the mid 1980s, whereupon it spikes sharply with an up to fivefold increase in F, then dropping to very low levels with the introduction of the moratorium in 1990. Recruitment estimates show unrealistically low interannual variation, starting historically at moderate levels, before dropping to low levels during the period consistent with the spike in F. Subsequent recruitments were initially very high, but declined markedly to levels close to historic levels as biomass appeared to increase.

The overall model fits were described by the assessment workshop as “reasonable” on the basis that the models tended to follow the information contained in the indices and the selectivities followed the perceived pattern described for the indices. The latter should not be seen as an appropriate measure for model fit, however, because the selectivities were determined external to the model and their functional form was determined *a priori* on the basis of expert opinion, so cannot be independent characteristics against which to judge model fit. With respect to the fits to indices, I consider there to be significantly autocorrelated residuals present in all indices (Figure 2), suggesting that the model is unable to address the differences in the information from the indices on the basis of the parameters left unconstrained in the model.

In this case I characterize model fit as rather poor, particularly with the summary output suggesting some rather unrealistic stock dynamics in terms of trends in F, SSB and recruitment.

Trends in fishing mortality: The assessment suggests dramatic increases in F (3–5 times, dependent on the run) followed by a decrease of similar size over a very short period, prior to dropping to near zero since the moratorium. Although the relative magnitude of this change in F differs between the base and continuity run because of the different assumptions regarding natural mortality, the pattern is virtually identical in terms of F.

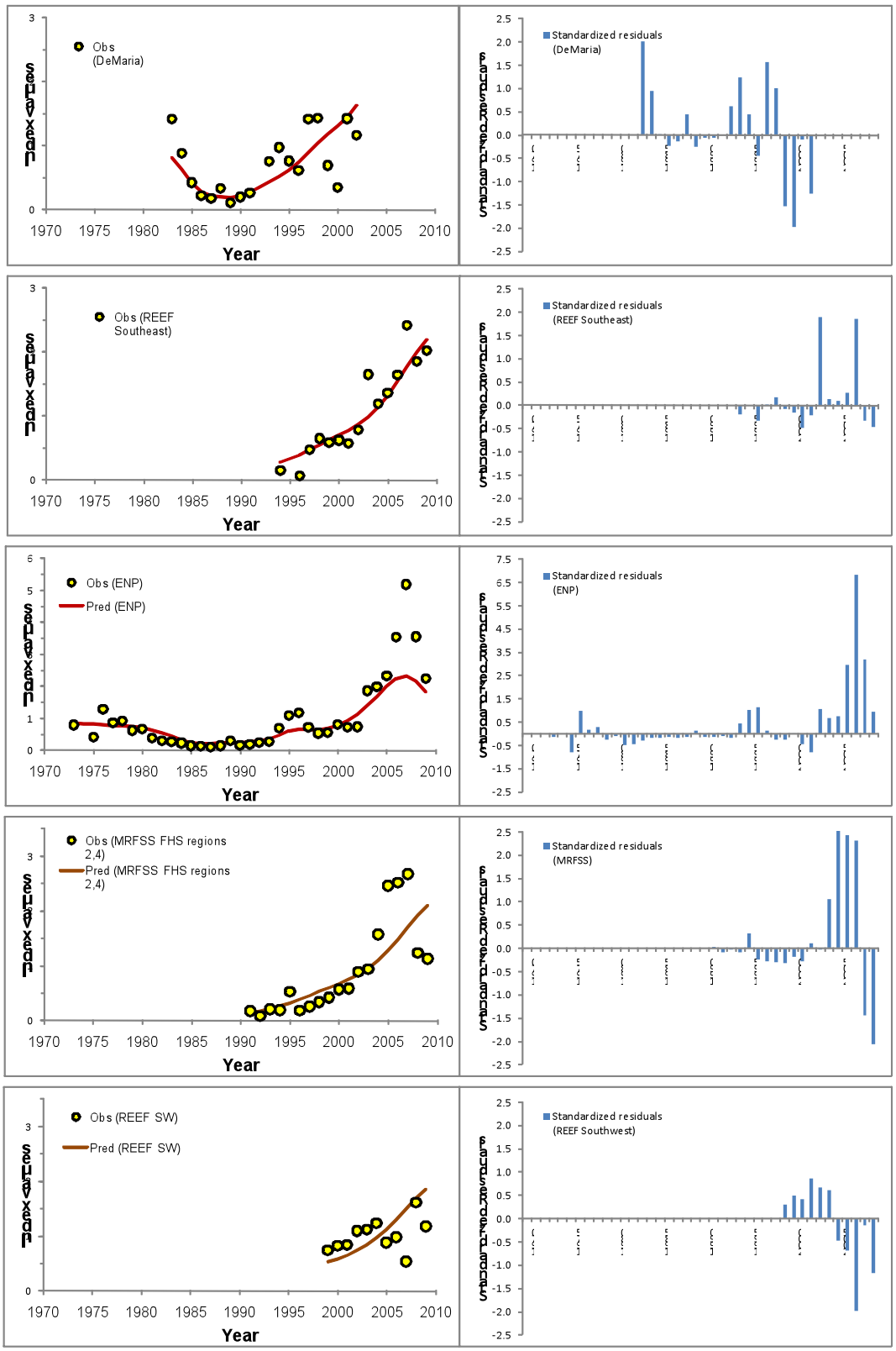

Figure 2: Index information and predicted assessment trends (left) and residual pattern indicating strong temporal autocorrelation patterns particularly for the last three indices in the most recent period.

There appears to be no reasonable basis for such a sharp spike in F. SCUBA gear, one of the main methods of exploiting goliath grouper, was already well established in 1980, as were recreational and commercial fisheries, and although there is little doubt that all these gears expanded in terms of effort during the 1980s, there is no qualitative or quantitative independent information available, such as fisher reports or catch information that would appear to be consistent with such dramatic changes in F.

On closer examination of the model set-up it becomes apparent that the pattern in F has been highly constrained in the model in three separate periods and that it is likely the discontinuity between the differing constraints is driving the spike in F.

1) Trends in historic fishing mortality in the periods prior to index availability are constrained by an assumed increase in effort proportional to the expanding population in Florida. Although this is a relative constraint, it does fix the direction of the trend and does not allow for any interannual fluctuations in F in the way that it has been fitted. In this model the constraint is necessary in order to be able to set benchmarks, but it ignores the fact that it is a very crude constraint and likely highly influential on the benchmarks and estimate of stock status, although this was not investigated by the assessment workshop.
2) Estimation of the efficiency of the moratorium set the relative level of F in the years 1991–2009 proportional to that in 1990. This is a relative measure, and the review workshop was provided with a sensitivity analysis using an uninformative prior of the measure. It turns out that the data suggest that moratorium effectiveness is 1, which effectively fixes F at 0, which we know not to be the case, so the level of constraint implemented in the current assessment is incorrect. The cause of this relates to the recruitment dynamics and will be discussed under that section, later.
3) The period in between is theoretically less constrained in terms of its scaling, but in many ways it is constrained by the selectivities (vulnerabilities) derived external to the model.

In general, to determine selectivity at age it is necessary to know the total mortality rate and a relative measure of cohort size. Here, equilibrium conditions are assumed in terms of the estimate of cohort strength, and an assumed minimum F is used in a minimization. The assumption about equilibrium condition is certainly inappropriate for the REEF, MRFSS and ENP indices, which show recent substantial increases in abundance. Determining selectivities external to the model with an increasing population as conducted here is almost certain to underestimate F at the start of the series while overestimating it at the end, the effect being particularly severe for short time-series.

The estimates of F and selection at age are correlated both in the selectivity estimation (and hence the need to provide a minimum F value) and in the assessment, leading to incorrect assessment values of F if the selections have been incorrectly specified. The correlation may be reduced by a gamma or logistic function, but the premise still holds. The estimation of selectivities needs to be carried out internal to the modelling process. Only in that way can selection be estimated consistently with F, and I cannot understand why this is not implemented within the model. If there is insufficient information in the data for the model to converge under these circumstances, then clearly that information is also insufficient to determine selectivities outside the model, so the estimation of F will be very susceptible to the minimum F chosen in the selectivity analysis.

To my mind then, there is a worrying degree of circularity in the determination of F within this assessment in the sense that one of the major questions to be answered is exploitation rate, yet very little of this parameter is actually estimated within the assessment. Although current F is likely to be small and F values are unlikely to be important in determining stock status with respect to “overfishing”, these estimates will still have a major impact on the estimates of abundance and therefore strongly drive the estimation of the rate of stock recovery and its potential productivity. Future assessments need to address these issues if they are to be used in management.

Although in general there are improvements possible to the assessment methodology (set-up and data use) in many instances, my personal feeling here is that even with much additional modelling work, the paucity and uncertainty of the data may preclude appropriate elucidation of cohort strength, so questions relating to specific cohorts of years such as potential ABCs and stock status advice, or questions as to whether the 2008 cold snap had an effect on recruitment, are unlikely to be answered satisfactorily without substantial additional data collection. This does not mean that it is not possible to provide useful management advice, but it does suggest that the questions may need to be posed differently from how they are posed for stocks where better and more data are available. General productivity principles and relative abundance information in an appropriate set-up of the model should be able to address whether reopening of the fishery in an adaptive management framework might be appropriate or whether further protection measures may be necessary to protect goliath grouper. However, uncertainty estimates surrounding such advice will likely never reach the levels of precision available in cases where substantially more data are available and uncertainties regarding the quality of the data are fewer.

One more interesting point regarding estimates of F is that the base assessment suggests that under the currently low levels of exploitation, the stock will increase to an equilibrium level of around 50% SPR, which given the stock–recruit parameters implies an SSB of just under 50% of virgin SSB. In other words, we are close to a theoretical F_{msy}, with just incidental fishing mortality under current conditions. I find this situation implausible, because it would mean that the population is not at all buffered against variation in mortality which, given the history of cold kills, would make it very unlikely that the population would persist in Florida.

Spawning Stock Biomass trends: Trends in SSB describe a steep decline from 1950, originally thought to be driven largely by the sharp decline in the DeMaria index, but a sensitivity analysis excluding one index at a time suggested that the latter index only affected the rate and timing of the decline, not the total change in abundance over the period to the moratorium. It seems that the model views the population as virgin in 1950 and at extremely low levels during the late 1980s. The cause of the latter is likely the high estimates of F during this period in conjunction with the ENP index suggesting a virtual absence of these fish in the nursery areas.

Historic trends in SSB, although highly uncertain given that most of the information is available only for the recent period, appear to be inappropriately stable in the various sensitivity runs. In general, the parameter declines from twice SSB at 50% SPR to virtually zero. Not only does this suggest that the relative assessment becomes near absolute at SSB ~0, but the high estimate of virtually virgin SSB in 1950 irrespective of the assessment results from a scalar constraint implemented in the proposed base run as no data on abundance exists prior to 1973 (ENP index), so there is no quantitative information available to the model to determine earlier SSB

trends. However this period of the assessment seems to drive the benchmark characteristics and the stock recruit relationship on which productivity estimates are based.

SSB trends since the moratorium imply steep rises, suggesting that the moratorium has been very effective in improving SSB by lowering F. In fact, the rate of increase suggested by the indices is much faster than numerically possible in the model on the basis of the stock–recruit relationship, so a number of strong positive deviates in recruitment are necessary immediately following the moratorium to explain the fast increase as well as suggesting that steepness in the stock recruit relationship is near 1, i.e. independent of stock size at very low stock levels.

For all runs, including the sensitivity runs, SSB simply declines from 1950 to 1990, whereupon it rises exponentially until today. The rate of this increase, the relative SSB-SSBspr50% in 2009, and the level at which the stock is projected to reach equilibrium is very dependent on the sensitivities within the model regarding mortality and the effectiveness of the moratorium, as well as any assumptions that constrain the stock–recruit relationship estimated within the model.

Recruitment: Estimates of recruitment were not provided by the assessment workshop, but were extracted from the model output during the review workshop. The implied recruitment trends show virtually no interannual variation, but an overly smoothed trend. Moreover, the trend is almost entirely inconsistent with the presumed stock–recruit relationship that recruitment now appears to be similar to historic levels, when SSB was estimated to have been very much higher and is estimated to have been greatest at the time immediately following the moratorium when stock size was at its minimum.

Neither effect is likely, but if the model completely ignores the stock–recruit function, then either the recruitment dynamics or the mortality trends are incorrectly specified. Even if some *a priori* knowledge exists that allows for an assessment of the likely cause, the output of a model with such a strong process error should not be used to inform management.

Given the problems with the model presented above, it was not possible to determine appropriate estimates of stock abundance, biomass, and exploitation. It seems likely, however, that SSB has increased markedly since the implementation of the moratorium because all indices suggest an improving trend, but the rate of increase or its persistence is much less clear especially in light of the slight decline in abundance in the most recent year for some of the indices.

Statistical Stock Reduction Analysis: An alternative assessment based on the stochastic stock reduction analysis (SRA) was also presented to RW, but only as additional information since the assessment had not been previously approved by the AW. The modeling approach offers some promising aspects for future management evaluation. In particular it should allow the determination of absolute estimates of SSB which will make model output more appropriate for current management practice that the attempted catch-free model, although estimating likely patterns of selectivity for future exploitation would still prove difficult.

Despite this promise though there are some serious concerns regarding the historic catch information and the way it has been smoothed. As this is one of the cornerstones of this type of assessment catch information needs to be both precise and accurate and it appears that we are still some way off from attaining a consensus as to what the most likely time series of catches

might have been. Some members of the panel felt that pursuit of this approach could quite quickly lead to an assessment appropriate for management advice, but my feeling is that the poor age and length information will still not allow the SRA to accurately reflect the stock dynamics either, especially as it uses many of the same biological and selectivity inputs.

One of the two runs presented shows a very similar picture of the F trend as that produced by the catch-free model, but with this very similar F pattern, it estimated SSB 2009 to be equal to historic levels SSB 1918, unlike the catch-free model which still estimated SSB to be a fraction of the SSB in 1950. The major difference between this and the other run presented was merely the degree of permitted recruitment deviates. Constraining recruitment variation provided the impression of a stock that had been very close to complete collapse for a period of 40 years under low levels of F and had recovered unexpectedly quickly in response to the moratorium.

Examination of the cohort strength estimates as in the catch-free model indicated that the model had been unable to resolve the underlying relative cohort signal. Recruitment estimates were strongly auto-correlated with little inter-annual variability. Releasing the recruitment variation in the first run merely led to the model dramatically increasing recruitment during the period following the moratorium (as had the catch-free model) rather than the intended purpose of creating a more usual recruitment pattern. Although changes in the catch information may alter these trends, I cannot see it increasing the data contrast to the point where it may overcome the poor age information in this assessment.

TOR 3 & 4: Recommend appropriate estimates of stock abundance, biomass, and exploitation. + Evaluate the methods used to estimate population benchmarks and management parameters (*e.g., MSY, F_{msy}, B*msy*, MSST, MFMT, or their proxies*); recommend appropriate management benchmarks and provide estimated values for management benchmarks, and declarations of stock status.

The assessment workshop felt uncomfortable with providing more than the bare essentials with respect to benchmark and management parameters, presenting only F50%spr and SSB2009 because a ratio of SSB50%spr appears to be calculated in an appropriate manner; but given that the assessment on which they are based was rejected, little time was spent on these discussions. My own concerns fundamentally mirror those of the AW with respect to the absolute estimates of abundance, i.e. they are of little value for management given the relative nature of the assessment method used. However my concerns go further as I do not believe these estimates to be suitable either as a relative measure or even qualitatively on the basis of the concerns regarding the stock dynamics mentioned in TOR 2 and the uncertainty underlined in TOR 6. Potential productivity from this stock and hence any estimate of equilibrium benchmarks are inappropriate given that the stock recruit relationship as models is diametrically opposed to the available data information, and estimates of natural mortality appear to have been calculated inconsistently with the ecological theory underlying the meta-estimates (TOR 1). These terms of reference require a sound stock assessment at the very least reasonably replicating some plausible stock dynamics and illustrating coherence with the available data before such evaluations can be carried out. Therefore, recommendations regarding benchmark and management parameters appropriateness with respect to goliath groupers will have to await a new assessment process.

TOR 5: Evaluate the adequacy, appropriateness, and application of the methods used to project future population status; recommend appropriate estimates of future stock condition (e.g., exploitation, abundance, and biomass).

As for the TOR above, the projections provided by the assessment workshop seemed of little use to management given their numerical basis in a flawed assessment. However, the procedure for calculating future stock trends appeared to be reasonable. Stock projection carried forward to 2100 at the review workshop as part of one of the sensitivity analysis suggested that there may be some problems with the implementation of the stock projections if taken too far into the future (Figure 3). It is likely, however, that this is a bug in the program where it runs out of array space, but the cause of this should be investigated further to clarify the cause of this unexpected behaviour.

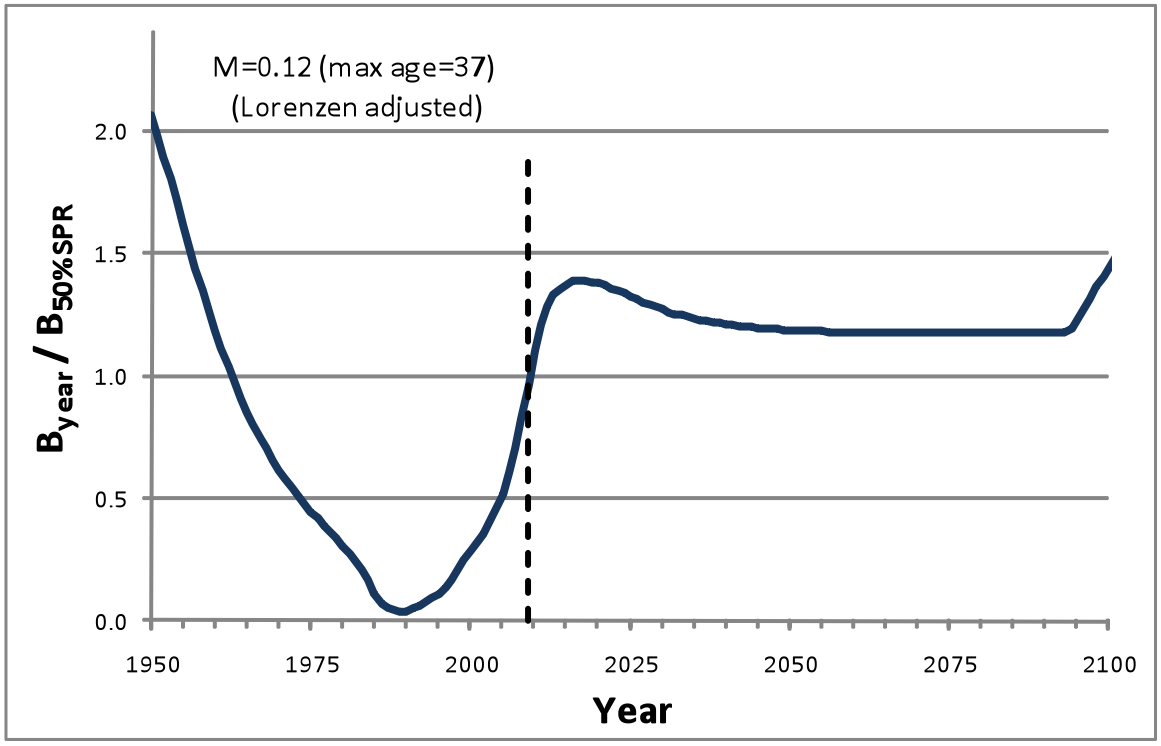

Figure 3: showing that he stock under currently estimated recruitment and F conditions will reach equilibrium values at around 1.2 times B50%SPR, suggesting that recent recruitment has been unusually high. Also noticeable is a sharp increase in SSB in 2090 likely a programming artefact as the model was not designed to project this far.

TOR 6: Evaluate the adequacy, appropriateness, and application of methods used to characterize uncertainty in estimated parameters. Provide measures of uncertainty for estimated parameters. Ensure that the implications of uncertainty in technical conclusions are clearly stated.

In general I believe that if the mean cannot be estimated, then the uncertainty cannot either. Given that the assessment is sufficiently questionable, I feel that none of the quantitative conclusions about the uncertainties or sensitivities of the assessment are valid. This pertains to

both factors that have been found to alter the outcome of the assessment significantly as well as those that have little effect. I provide one example of each.

The assessment workshop analysis examining the sensitivity of the assessments to the assumption of moratorium effectiveness suggests that the model is not overly sensitive to the prior, with a near 90% effectiveness implied by a number of different settings examined. Using an uninformative prior at the review workshop, however, it was postulated that the model would be unlikely to converge in the absence of information on current F in relation to earlier values of F. Surprisingly, the model did converge, but it did so at an unrealistic 100% effectiveness rating. The convergence in the absence of absolute information is concerning. The most likely explanation for this is that SSB has been increasing exponentially since the moratorium and, given the recruitment parameterization estimate, the model is unable to follow the trend. As a consequence, any additional mortality would increase the size of the residuals so that a fully effective moratorium would appear to be the most parsimonious outcome. It is unlikely that a properly parameterized model would be equally insensitive to the uncertainty of such an important prior.

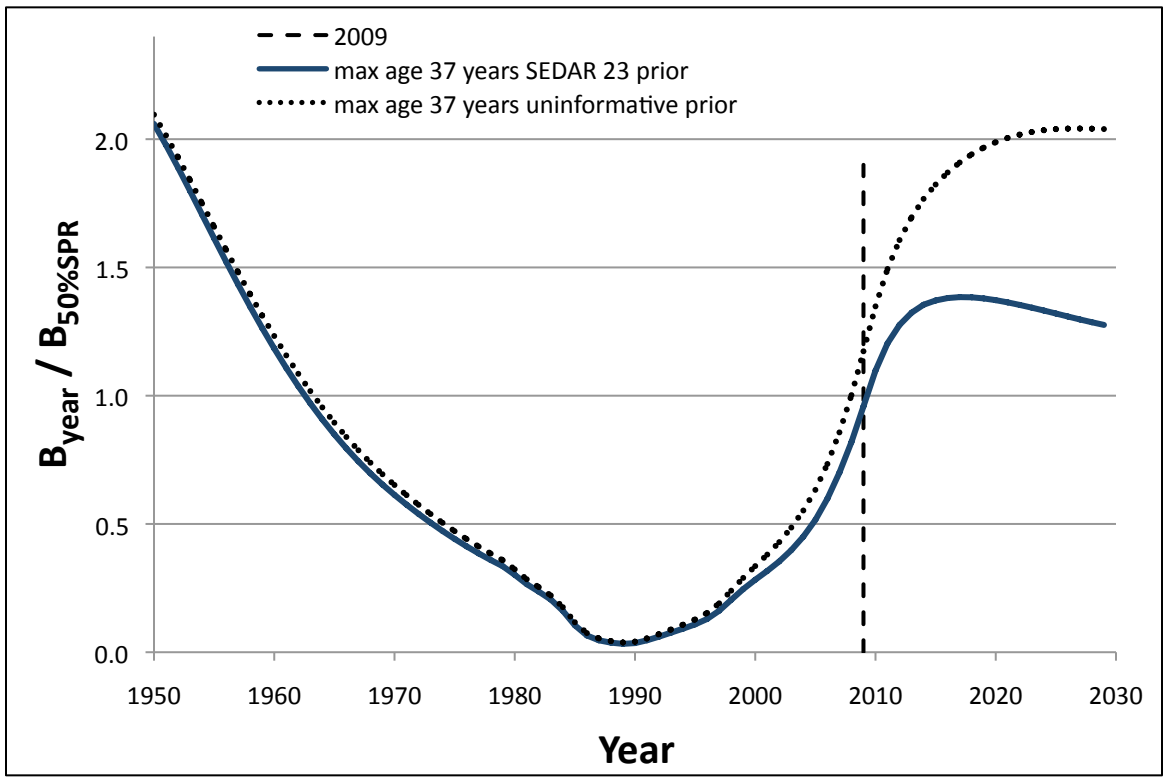

Figure 4: showing that using an uninformative prior there is considerable uncertainty in stock status estimates based on the uncertainty of the moratorium effectiveness parameter, and the apparent insensitivity in the MCMC analysis is derived from the fact that in the absence of information from the data the model converges to the expert determined prior.

Other sensitivity work performed at the assessment workshop concerned itself with the estimation of natural mortality, specifically the maximum age estimate. The uncertainty in this case ranged from the fact that the population is expected to have been strongly depleted prior to

the collection of age information, implying that the maximum observed age in the catch (age 37) may well represent an underestimate of the maximum age in the natural population. This fact was illustrated at the review workshop by a meta-analysis between growth and maximum age within the grouper/snapper complex. Different values of t_{max} were examined, and the assessment was found to be highly sensitive in terms of stock status estimation and the time to recovery. The degree of sensitivity was rather surprising, because the number of fish at older ages in any population make up a relatively small number/proportion of the total. This is because of the effect of cumulative mortality over time, especially in this case where the population at least within the assessment was driven down to such an extent that there were very few old fish in the population and SSB should then have been largely dependent on the number of recent cohorts, not on the mortality at older age.

One of the reasons for this sensitivity may have been the way that natural mortality at age was calculated. According to Hoenig it should not have been rescaled using the Lorenzen method. Instead Lorenzen M should be used for the younger ages and Hoenig M for the older ages. In this case it is unlikely that there would be much of a difference in current stock status because according to the assessment there are few if any fish of that age out there. If fish live longer, there will be a difference in the equilibrium biomass though so that some sensitivity will remain. The extent of the sensitivity is also over emphasised by the fact that through most of the assessed period F is very low, probably unrealistically so. At higher Fs stock status would be very much less affected than indicated in figure 5, because far fewer of the fish would survive to the ages of uncertainty.

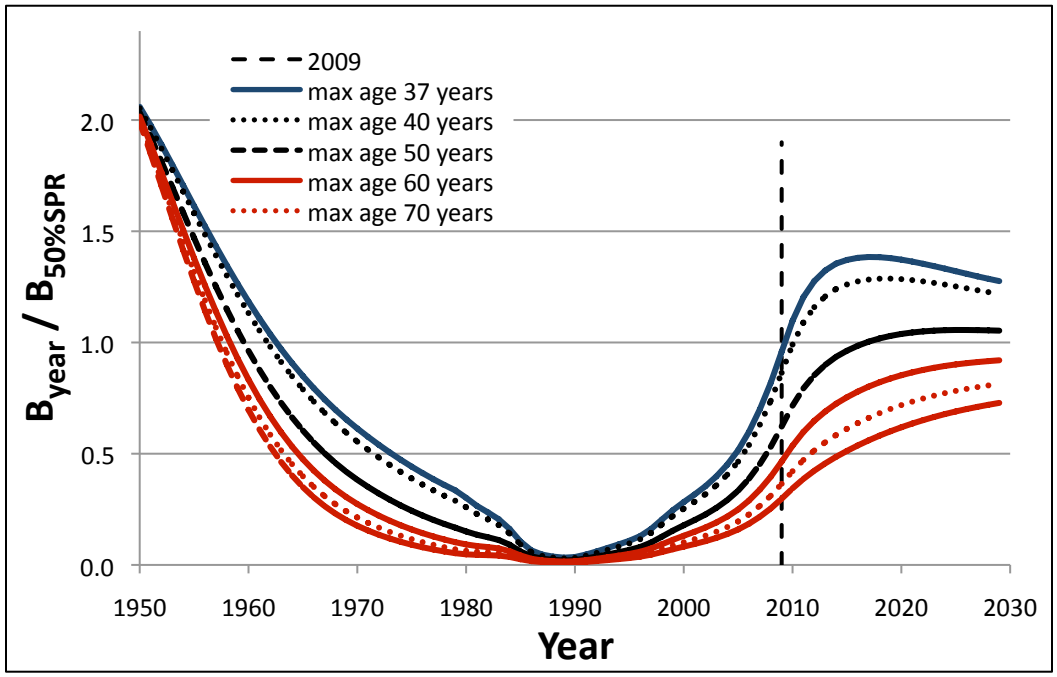

Figure 5: SSB stock trajectories given different maximum ages for age specific M

In principle, the Markov Chain Monte Carlo (MCMC) methodology employed here is suitable for assessing uncertainty and the criteria for choosing burn-in and acceptance rate are fundamentally reasonable. However, applying these criteria to a model that falls apart under

critical review is inappropriate. I feel therefore that too much effort may have been devoted to the time-consuming MCMC analysis when a single run would have clearly identified the inappropriateness of the model. Of particular concern is the fact that the MCMC output was interpreted despite the model implying strongly autocorrelated residuals. Under such conditions uncertainty is generally underestimated, because this will ignore the process error clearly present in this assessment.

Undoubtedly the reason so much time was designated to determining the uncertainty are the very generic terms of reference provided for this stock. In my mind it is questionable whether there is any point in assessing uncertainty in a flawed assessment, and more time could have been spent investigating the reasons for the poor results. Although this might not have provided the necessary answers either given paucity of the data, it might have however provided better information as to the process error underlying the difficulties and hence suggested ways forward.

Some more specific terms of reference might also have allowed the provision of more qualitative expert advice. The current situation is that any advice is entirely dependent on an acceptable assessment so that without such an assessment no advice can be given. In reality there is still a lot of information in the data though and some questions of interest could still have been addressed.

TOR 7: Ensure that stock assessment results are clearly and accurately presented in the Stock Assessment Report and that reported results are consistent with Review Panel recommendations.

Stock assessment results were presented by the assessment workshop, and most of the information necessary to draw the conclusions reached from the assessment was available and the balance was produced on request of the review panel. However, had the assessment been accepted, a lot of reordering and clarification would have been required to convey the information and its basis to managers. I would have preferred to have seen more information comparing different runs in a single plot, rather than seeing different parameters plotted in the same graph for a single run.

TOR 8: Evaluate the SEDAR process as applied to the reviewed assessments and identify any Terms of Reference that were inadequately addressed by the Data or Assessment Workshops.

The data and assessment workshops provided a lot of information considering the sparse amount of data collected on this specific stock. The terms of reference provided were well addressed (though see my comment above), although there still seemed to be some room for improvement in the way the information was passed between the two meetings. A number of uncertainties in the interpretation remained, for example with respect to the natural mortality calculations that resulted in some uncertainty at the review workshop. In addition the background information on index sampling and derivation should be presented more clearly. In particular, spatio-temporal information on the development of the indices would have been helpful in assessing their appropriateness for use in the assessment, especially for the new indices developed specifically for this assessment. At the review workshop, a standard index presentation procedure for the SEDAR process was provided, and it would have been helpful to reviewers if the index information would have been presented in this format.

ToR 4 regarding the population benchmarks and management parameters was minimally addressed, with only basic plots of F and SSB trends, and not even a comparison of the continuity and base runs. However, it became clear during the review workshop that the assessment workshop had felt that the information provided by these assessments was inappropriate for the evaluation of such benchmarks and hence had provided only minimal information. The review panel agreed fully with both the determination of the utility of the assessment and the fact that under such circumstances such information should not be presented as substantive.

Given the difficulties regarding TOR 2, the subsequent terms of reference appeared to be of little relevance to the assessment process, yet a significant amount of time was still devoted to these terms of reference during the assessment and review workshops, particularly regarding the uncertainties in the assessment. I feel that additional time spent on determining the underlying causes of process error in the model might have yielded further insight as to how one might proceed in future and would have been a more valuable investment of time by the assessment workshop instead of developing a full MCMC approach for a model that clearly had problems. Certainly under such circumstances, the review workshop might have been in a better position to address their TOR 6. The generic nature of the TORs for the SEDAR process and the apparent pressure to address all of these may in some instances reduce the value of the advice provided, and additional guidelines or more stock-specific ToRs may be helpful to improve the process in future (in my opinion, this is a very important point, which is why I keep emphasizing it).

TOR 9: Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted. Clearly denote research and monitoring needs that could improve the reliability of future assessments. Recommend an appropriate interval for the next assessment, and whether a benchmark or update assessment is warranted.

The review workshop unanimously determined the assessment to be unsuitable for management advice. In that sense the discussions by the panel were very constructive in establishing the problems with the assessment. However, the question as to how the situation could be improved highlighted the different areas of expertise and experiences of the members of the panel, so opinions differed significantly in terms of the prioritization of research needs. The panel did present a summary of proposals, but also agreed that a better understanding of what the management aims for the stock currently under moratorium would be necessary to provide advice on which research should be supported to provide better advice in future.

A better understanding of stock structure is vital for any assessment and management advice. Genetic information does not appear able to resolve these issues at sufficiently fine resolution, and tagging information will be required to determine the extent of mixing in the population and the recruitment process from the nursery grounds, particularly the major center of these located within the Everglades National Park.

If general questions regarding potential productivity of the stock and likely direction of stock trends require some clarity in terms of reproductive biology, particularly the question of whether goliath grouper are protogynous and the factors that drive sex change could be highly influential

in assessing the resilience of the stock and may serve to explain some of the difficulties encountered in the assessment with regards to the stock–recruitment relationship.

If more-specific management questions such as setting ABC's or determining the stock status in relation to reopening a fishery are required, then I cannot see a way around obtaining better information on cohort strength in the form of age and length information. Both assessment methodologies presented to the review workshop indicated that the inability of the models to resolve cohort signals was because of the sparse length information with an age/length key cumbersomely reconstructed via a length probabilistic distribution derived from time-invariant growth based on even sparser age information. There are a number of ways discussed in this review in which the process could be improved, but simply the contrast in the data is insufficient given the complexity of the assessment model used currently, and I would expect that although refined assessment attempts may provide more accurate information on stock dynamics, the pressing question of specific cohort strength is likely to remain unresolved in the absence of a substantial increase in age sampling.

Given the deficiencies of the assessment and the uncertainty of the management requirements, it is very difficult to suggest a time-frame for re-evaluation. Certainly a full assessment process is expensive and additional work in the interim could address this question much more appropriately. Certainly the collection of age information is a drawn-out long-term approach and cannot be assumed to yield significant improvements to the assessment in terms of resolving cohort strength issues in the near future. Similarly, tagging information (data storage tags or traditional tags) to resolve stock definition, migration and mixing issues is likely to require substantial time until sufficient information becomes available, so I cannot see much benefit in trying to reassess this stock in the next 5–10 years unless other information becomes available in the interim, which seems unlikely.

This does not mean that additional assessment work in the interim could not provide significant improvements on what has been produced now, and indeed it may guide or even refine the research needs better so that the most important effort be continued. In particular it may be useful to explore other assessment techniques that can provide absolute estimates of stock size as opposed to the relative measures provided by the catch-free model.

TOR 10: Prepare a Peer Review Summary 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 Summary Report within 3 weeks of workshop conclusion.

The review panel compiled a peer review summary describing the pertinent conclusions regarding the SEDAR 23 process for each of the terms of reference. Detail of my own findings is found in the text above.

Appendix 1: SEDAR 23 South Atlantic and Gulf of Mexico Goliath Grouper
Workshop Document List

Document #	Title	Authors
SEDAR23-DW-01	Bottom longline fishery bycatch of Goliath Grouper (*Epinephelus itajara*) from observer data	Loraine Hale
SEDAR23-DW-02	Monitoring changes in the catch rates and abundance of juvenile goliath grouper using the ENP creel survey, 1973-2009	Shannon L. Cass-Calay
SEDAR23-DW-03	Goliath grouper surveys and samples: A summary of recent work by the Fish and Wildlife Research Institute (2006 -2010)	Angela Collins & Luiz Barbieri
SEDAR23-DW-04	Calculated Goliath grouper discards from commercial vertical line and longline fishing vessels in the Gulf of Mexico and US South Atlantic	Kevin McCarthy
Documents Prepared for the Assessment Workshop		
SEDAR23-AW-01	Standardized visual counts of goliath grouper off south Florida	Clay Porch
SEDAR23-AW-02	Analysis of Headboat Data for Goliath Grouper	Walter Ingram
SEDAR23-AW-03	Standardized proportion of private vessel trips with catches of goliath grouper from the Marine Recreational Fisheries Statistics Survey in south Florida, 1991-2009	Joe O'Hop

Documents Prepared for the Review Workshop		
SEDAR23-RW-01	Application of Stock Reduction Analysis to goliath grouper (*Epinephelus itajara*) off southeastern U.S.A, 1918 – 2009	Joseph Munyandorero
Final Stock Assessment Reports		
SEDAR23-SAR	Goliath Grouper	
Reference Documents		
SEDAR23-RD01	Age, growth, and reproduction of jewfish, *Epinephelus itajara* in the eastern Gulf of Mexico	L.H. Bullock, M.D. Murphy, M.F. Godcharies, and M.E. Mitchell
SEDAR23-RD02	Monitoring changes in the catch rates and abundance of juvenile goliath grouper using the ENP creel survey, 1973-2006	Shannon L. Cass-Calay and Thomas W. Schmidt
SEDAR23-RD03	How many species of goliath grouper are there? Cryptic genetic divergence in a threatened marine fish and the resurrection of a geopolitical species	M. T. Craig, R. T. Graham, R. A. Torres, J. R. Hyde, M. O. Freitas, B. P. Ferreira, M. Hostim-Silva, L. C. Gerhardinger, A. A. Bertoncini, D. R. Robertson10
SEDAR23-RD04	Habitat affinities of juvenile goliath grouper to assess estuarine conditions	Anne-Marie Eklund
SEDAR23-RD05	A stepwise approach to investigating the movement patterns and habitat utilization of goliath grouper, *Epinephelus itajara*, using conventional tagging, acoustic telemetry and satellite tracking	Anne-Marie Eklund and Jennifer Schull
SEDAR23-RD06	Activity patterns of three juvenile goliath grouper, *Epinephelus itajara*, in a mangrove nursery	Sarah Frias-Torres, Pedro Barroso, Anne-Marie Eklund, Jennifer Schull, and Joseph E. Serafy

SEDAR23-RD07	Mangroves as essential nursery habitat for goliath grouper (*Epinephelus itajara)*	Christopher C. Koenig, Felicia C. Coleman, Anne-Marie Eklund, Jennifer Schull, and Jeffrey Ueland
SEDAR23-RD08	Early life history stages of goliath grouper *Epinephelus itajara* (Pisces: Serranidae) from Ten Thousand Islands, Florida	Monica R. Lara, Jennifer Schull, David L. Jones, Robert Allman
SEDAR23-RD09	Goliath grouper *Epinephelus itajara* sound production and movement patterns on aggregation sites	David A. Mann, James V. Locascio, Felicia C. Coleman, Christopher C. Koenig
SEDAR23-RD10	Documenting Loss of Large Trophy Fish from the Florida Keys with Historical Photographs	Loren McClenachan
SEDAR23-RD11	Status report on the continental United States distinct population segment of the goliath grouper (*Epinephelus itajara*)	NMFS
SEDAR23-RD12	A catch-free stock assessment model with application to goliath grouper (Epinephelus itajara) off southern Florida	Clay E. Porch, Anne-Marie Eklund, and Gerald P. Scott
SEDAR23-RD13	A Preliminary Discussion of Acceptable Harvest Levels for Scientific Sampling of Goliath Grouper in the U.S. South Atlantic and Gulf of Mexico	Clay E. Porch and Luiz R. Barbieri
SEDAR23-RD14	Range-wide status and conservation of the goliath grouper *Epinephelus itajara*: Introduction	Kevin L. Rhodes and Rachel T. Graham
SEDAR23-RD15	Synopsis of biological data on the Nassau grouper, *Epinephelus striatus* (Bloch, 1792), and the jewfish, *E. itajara* (Lichtenstein, 1822)	Yvonne Sadovy and Anne-Marie Eklund
SEDAR23-RD16	Complete Stock Assessment Report of SEDAR 6 - Goliath Grouper	SEDAR 3 DW participants/ SEDAR 6 RW participants
SEDAR23-RD17	Habitat use of juvenile goliath grouper *Epinephelus itajara* in the Florida Keys, USA	Sarah Frias-Torres

SEDAR23-RD18	Standardized visual counts of goliath grouper off south Florida and their possible use as indices of abundance	Clay E. Porch and Anne-Marie Eklund
SEDAR23-RD19	Population density, demographics, and predation effects of adult goliath grouper	Christopher C. Koenig and Felicia C. Coleman
SEDAR23-RD20	The role of dispersal and demography in determining the efficacy of marine reserves	Gerber LR, Heppell SS, Ballantyne F, Sala E.
SEDAR23-RD21	Spawning aggregations and reproductive behavior of reef fishes in the Gulf of California	Sala E, Aburto-Oropeza O, Paredes G, Thompson G.
SEDAR23-RD22	American Fisheries Society Position Statement. Long-lived reef fishes: the grouper-snapper complex	Coleman, F.C., C.C. Koenig, G.R. Huntsman, J.A. Musick, A.M. Eklund, J.C. McGovern, R.W. Chapman, G.R. Sedberry, and C.B. Grimes
SEDAR23-RD23	Preliminary Investigations of Reproductive Activity of the Jewfish, Epinephelus itajara (Pisces: Serranidae)	Colin, P.L.
SEDAR23-RD24	Grouper Stocks of the Western Central Atlantic: The Need for Management and Management Needs	Sadovy, Y.
SEDAR23-RD25	Hypothermal mortality in marine fishes of southcentral Florida	Gilmore RG, Bullock LH, Berry FH
SEDAR23-RD26	Evaluation of finrays as a non-lethal ageing method for protected goliath grouper *Epinephelus itajara*	Murie DJ, Parkyn DC, Koenig CC, Coleman FC, Schull J, Frias-Torres S.
SEDAR23-RD27	Mercury concentrations in the goliath grouper of Belize: an anthropogenic stressor of concern	Evers DC, Graham RT, Perkins CR, Michener R, Divoll T.
SEDAR23-RD28	Behavior, Habitat, and Abundance of the Goliath Grouper, *Epinephelus itajara*, in the Central Eastern Gulf of Mexico	Angela B. Collins and Luiz R. Barbieri

Appendix 2: Statement of Work

External Independent Peer Review by the Center for Independent Experts

SEDAR 23 Gulf of Mexico and South Atlantic Goliath Grouper 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 23 will be a compilation of data, a benchmark assessment of the stock, and an assessment review conducted for Gulf of Mexico and South Atlantic Goliath Grouper. The review workshop provides an independent peer review of SEDAR stock assessments. The review panel is ultimately responsible for ensuring that the best possible assessment is provided through the SEDAR process. The stocks assessed through SEDAR 23 are within the jurisdiction of the Gulf of Mexico and South Atlantic Fishery Management Councils and the states of Texas, Louisiana, Mississippi, Alabama, Florida, Georgia, South Carolina, and North Carolina. 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 shall have expertise, working knowledge, and recent experience in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the primary task of reviewing the technical details of the methods used for the assessment. Each CIE reviewer's duties shall not exceed a maximum of 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 Key West, Florida during 15-17 November 2010.

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.

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 can not 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 Key West, Florida during 15-17 November 2010.
3) During 15-17 November in Key West, Florida as specified herein, conduct an independent peer review in accordance with the ToRs (**Annex 2**).
4) No later than 1 December 2010, 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 CIE Regional Coordinator, via email to David Sampson 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.

11 October 2010	CIE sends reviewer contact information to the COTR, who then sends this to the NMFS Project Contact
1 November 2010	NMFS Project Contact sends the CIE Reviewers the pre-review documents
15-17 November 2010	Each reviewer participates and conducts an independent peer review during the panel review meeting
1 December 2010	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
15 December 2010	CIE submits CIE independent peer review reports to the COTR
22 December 2010	The COTR distributes the final CIE reports to the NMFS Project Contact and regional Center Director

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

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

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

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

Support Personnel:

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

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

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:

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

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

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

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

   a. Reviewers should describe in their own words the review activities completed 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 23 Gulf of Mexico and South Atlantic Goliath Grouper Review Workshop

1. Evaluate the adequacy, appropriateness, and application of data used in the assessment.
2. Evaluate the adequacy, appropriateness, and application of methods used to assess the stock.
3. Recommend appropriate estimates of stock abundance, biomass, and exploitation.
4. Evaluate the methods used to estimate population benchmarks and management parameters (*e.g., MSY, Fmsy, Bmsy, MSST, MFMT, or their proxies*); recommend appropriate management benchmarks and provide estimated values for management benchmarks, and declarations of stock status.
5. Evaluate the adequacy, appropriateness, and application of the methods used to project future population status; recommend appropriate estimates of future stock condition (e.g., exploitation, abundance, biomass).
6. Evaluate the adequacy, appropriateness, and application of methods used to characterize uncertainty in estimated parameters. Provide measures of uncertainty for estimated parameters. Ensure that the implications of uncertainty in technical conclusions are clearly stated.
7. Ensure that stock assessment results are clearly and accurately presented in the Stock Assessment Report and that reported results are consistent with Review Panel recommendations.
8. Evaluate the SEDAR Process as applied to the reviewed assessments and identify any Terms of Reference which were inadequately addressed by the Data or Assessment Workshops.
9. Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted. Clearly denote research and monitoring needs that could improve the reliability of future assessments. Recommend an appropriate interval for the next assessment, and whether a benchmark or update assessment is warranted.
10. Prepare a Peer Review Summary 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 Summary Report within 3 weeks of workshop conclusion.

The review panel may request additional sensitivity analyses, evaluation of alternative assumptions, and correction of errors identified in the assessments provided by the assessment workshop panel; the review panel may not request a new assessment. Additional details regarding the latitude given the review panel to deviate from assessments provided by the assessment workshop panel are provided in the *SEDAR Guidelines* and the *SEDAR Review Panel Overview and Instructions*.

** The panel shall ensure that corrected estimates are provided by addenda to the assessment report in the event corrections are made in the assessment, alternative model configurations are recommended, or additional analyses are prepared as a result of review panel findings regarding the TORs above.**

Annex 3: Tentative Agenda
SEDAR 23 Gulf of Mexico and South Atlantic Goliath Grouper Review Workshop
Key West, Florida during 15-17 November 2010

Monday

10:00 a.m.	**Convene**	
10:00 – 10:30	**Introductions and Opening Remarks**	**Coordinator**
	- Agenda Review, TOR, Task Assignments	
10:30 – 11:30	**Assessment Presentation**	**TBD**
11:30 a.m. – 1:00 p.m.	**Lunch Break**	
1:00 p.m. – 3:30 pm	**Continue Presentation/Discussion**	**Chair**
3:30 – 4:00	**Break**	
4:00 – 6:00	**Continue Presentation/Discussion**	**Chair**

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

Tuesday

8:30 a.m. – 11:30 a.m.	**Panel Discussion**	**Chair**
	- Assessment Data & Methods	
	- Identify additional analyses, sensitivities, corrections	
11:30 a.m. – 1:00 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Panel Discussion**	**Chair**
	- Continue deliberations	
	- Review additional analyses	
3:30 p.m. – 4:00 p.m.	**Break**	
4:00 p.m. – 6:00 p.m.	**Panel Discussion/Panel Work Session**	**Chair**
	- Recommendations and comments	

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

Wednesday

8:30 a.m. – 11:30 a.m.	**Panel Discussion**	**Chair**
	- Final sensitivities reviewed.	
	- Projections reviewed.	**Chair**
11:30 a.m. – 1:00 p.m.	**Lunch Break**	
1:30 p.m. – 3:30 p.m.	**Panel Discussion or Work Session**	**Chair**
	- Review Reports	
4:00 p.m.	**ADJOURN**	

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

Appendix 3: List of panel members

Dr. Luis Barbieri (Chair)

Dr. Julie Neer

Dr. John Hoenig

Dr. Shannon Calay

Dr. Jamie Gibson

Dr. Kevin Stokes

Dr. Barbara Dorf

Sven Kupschus