Center for Independent Experts (CIE) Report on SEDAR 34 Assessment of Atlantic sharpnose and bonnethead shark

Robin Cook

October 2013

Prepared for The Center for Independent Experts

Contents

Executive Summary

1. *A desk review of the bonnethead and Atlantic sharpnose stock was conducted during September/October 2013. This was based on detailed assessment reports and supporting documentation made available on 17th September.*
2. *Data used in the assessments comprise catch data, abundance indices and biological parameters. I concurred with the assessment panel's decisions regarding the choice and selection of data to be used in the assessment.*
3. *In both assessments the abundance indices are highly variable with very little common signal evident between them. The hierarchical index offers a promising way to combine the indices and obtain measures of process and measurement error. The trend in the hierarchical index may be the best indication of recent stock trends.*
4. *Both assessments are dominated by the estimates of bycatch from the shrimp fishery which forms the largest single component of the total catch. The data were derived using mean CPUE values raised by effort to obtain catch. Because fixed values of CPUE will not reflect changes in stock biomass, this procedure will be prone to generating a biased time series of catch that will be translated into the estimates of stock biomass and fishing mortality in the assessment. The shrimp bycatch data are better regarded as an index of fishing effort.*
5. *The principal assessment method used was an age structure statistical model, SSASPM. This is a modern well used approach that conforms to current practices in fish stock assessment. The model is complex and requires detailed age structured data as well as careful choice of error distributions associated with the data. Since no age structured data exist for both the assessed species, age data had to be derived from a limited amount of length frequency data. This will inevitably add to the uncertainty in the model results.*
6. *I felt that the complexity of the model was not well suited to the very limited data available and meant that a very large number of simplifying assumptions were required. In turn this made interpreting the robustness of the model output challenging.*
7. *I would recommend that an assessment model that is more closely designed around the available data is developed. If, in reality, shrimp bycatch data cannot be estimated with any precision, then it might be better to treat these catches as unknown and use the effort data directly in the model.*

8. *A number of sensitivity scenarios were used to investigate the performance of the assessment model. These are useful and suggest that stock status judgments are robust to the scenarios tested. However, the effect of bias in the shrimp bycatch was not investigated which leaves an important area of uncertainty unexplored and might affect conclusions about overfishing and overfished thresholds.*
9. *Both assessments suggest that the stocks are not over-fished and that over-fishing is not occurring. While there is very little evidence to contrary, the absence of a test of the influence of bias in the shrimp bycatch on the assessment means that the uncertainty associated with stock status is higher than the sensitivity analysis implies. Without such a test it would be unwise to have confidence in the assessed stock status relative to reference points.*
10. *Projections are based on model output from the base run and the sensitivity scenarios under a range of fixed catch regimes. A standard approach has been used that considers uncertainty in the initial conditions. Given the uncertainty in the shrimp bycatch data, the probability levels associated with the projections should be regarded as highly imprecise.*
11. *The problems with the shrimp bycatch could be avoided if a simpler assessment method was used. The methods developed by Brooks et al (2010) appear to offer an alternative assessment method without using catch data. The hierarchical index (Conn, 2010) would be an obvious choice of abundance index to use with this method. I would suggest that this approach is implemented alongside the SSASPM method for comparison and as an additional sensitivity test.*
12. *The assessment report documents are very well written and presented and are a credit to the authors. As a desk review, I missed the opportunity to raise questions about the assessments with the assessment panel. It might be useful to organize a web conference at some stage during the review so that reviewers can seek clarification from the assessment panel to avoid mistakes and mis-understandings.*

Background

SEDAR 34 is a compilation of data, a standard assessment of the stock, and CIE assessment review conducted for HMS Bonnethead and Atlantic sharpnose sharks. The desk review reported here provides an independent peer review of these stock assessments. The review is intended to ensure that the best possible assessment is provided through the SEDAR process and will provide guidance to the SEFSC to aid in their review and determination of best available science, and when determining if the assessment is useful for management.

Description of the Individual Reviewer's Role

The reviewer accessed documents from the SEDAR ftp site on the 17th September and commenced reviewing the main assessment reports. After an initial examination of the reports, the reviewer contacted the SEDAR co-ordinator to seek clarification on catch and effort data used in the assessments. Timely responses were received from the assessment panel that facilitated the reviewer's work. During the review of the main assessments a number of he supporting working documents were consulted. A draft report was completed on the 7th October and sent to the CIE as required.

Summary of Findings

Bonnethead shark

1. Evaluate the data used in the assessment

a) *Are data decisions made by the assessment panel sound and robust?*

The assessment report provides a clear summary of the relevant documents and the reasons for taking the various decisions. Overall I would concur with the decisions made but make a number of observations.

Clearly there are reasons to split the stock into two assessment areas (Gulf and Atlantic) but it was decided to run the assessment on the combined stock. This appears to have been done for pragmatic reasons, to avoid delays in completing the assessment. Without direct knowledge of the circumstances I would support this course of action, but the question of stock identity will need to be more comprehensively explored at some future date as acknowledged in the report.

The decision to adopt "method 2" in the reconstruction of shrimp bycatch is fair enough given the choice between a raw raising procedure and one using logs. However, I was not convinced that the use of a single CPUE mean (over the years 2009-2011) represented the best way to convert effort data into catch data. I believe this to be a potentially very significant issue, and is discussed at various points in this review, because it may have the largest single impact on the results of the assessment.

Whilst I would agree with the decisions on the choice of discard mortality rates in the absence of highly informative data, I would like to have seen some discussion on the impact of errors associated with this choice on the estimate of the catches since small changes in this mortality in the recreational fishery has potentially large effects on the estimate of dead catch, with consequences for the assessment.

The decisions to include/exclude abundance indices were based on sensible and objective criteria, such as length of series, area coverage, reproducibility of indices etc. I felt that these choices were sound.

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

The assessment report does discuss many of the uncertainties in the data, but I felt much more could have been done to provide quantitative information on the uncertainty in the estimates of the derived catches and the abundance indices. I have been led to understand that the MRIP survey is now based on a proper statistical design and it should therefore be possible to provide estimates of the variances for at least the sampling error associated with the recreational catch. It should also be possible to do this for the derived shrimp bycatch by considering the variance associated with the mean CPUE used and the stratification of the effort data. These two catch components dominate the assessment estimates of fishing mortality and it is particularly important to consider the uncertainty in these data.

I was particularly concerned about the use of a single mean CPUE per stratum to raise the annual effort data to total catch for the shrimp fishery. This procedure effectively assumes that the stock from which the catch is derived is of fixed abundance over time. It means that the catch increases linearly with effort and implies increases in fishing effort have no effect on stock abundance. This is not a credible assumption. In effect the shrimp bycatch is an index of effort scaled to catch units. In reality you would expect an increase in effort to lead to a transient increase in catch followed by a decline as the stock adjusts to a new state. Similarly you would expect a decrease in effort to lead to a transient loss of catch followed by some recovery. These effects will not be reflected in the shrimp bycatch data, which are therefore likely to contain an unknown bias. Such biases will emerge as biased estimates of fishing mortality and biomass in the assessment. It is of some importance to establish whether these biases undermine the assessment results.

An interesting development reported is the derivation of a hierarchical abundance index. The analysis on which this is based should provide estimates of both process and measurement error associated with the surveys and could be presented in the report to illustrate the uncertainties in the indices associated with these two elements.

Are data applied properly within the assessment model?

Given the choice of SSASPM as the preferred assessment tool, the data have been applied in the appropriate manner. Length frequency data are converted to age, selectivity curves have been generated and error distributions defined for each data component.

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

This is an extremely challenging question to answer and goes to the heart of the assessment. On purely scientific grounds I would argue that the data do not contain sufficient information to justify the complexity of assessment model attempted here. An indication of this is the fact that some parameters (e.g. fleet selectivity) have had to be estimated externally from the model as they appear not to be estimable within it. However, it is frequently the case that stakeholders and managers expect far more detail, and want the inclusion of information of particular concern to them, to be included in an assessment. This often leads to inflation in the complexity of the assessment model, which may satisfy clients of the assessment but which adds to the uncertainty of the results due to the necessity to make numerous structural and simplifying assumptions, each with an associated error, which tend to accumulate.

Above I have already commented on the issues of uncertainty relating to the shrimp bycatch data. It is noteworthy that the abundance indices, while numerous, do not show a high degree of consistency and this is itself indicative of uncertainty in the assessment. The hierarchical index provides a useful summary of the indices to the extent that it suggests a smooth stock decline in the earlier period and an increase more recently. This is consistent with trends in the shrimp fishery effort, from which most of the catch appears to come, and provides some reassurance about overall stock trends.

Given my reservations about the catch data, I would be skeptical about the ability of the assessment to determine stock status relative to reference points.

2. Evaluate the methods used to assess the stock, taking into account the available data.

a) *Are methods scientifically sound and robust?*

The SSASPM model which lies at the core of the assessment is a fully age structured forward projection statistical model which conforms to current standards and practice for the assessment of stocks with appropriate data. The model uses a maximum likelihood approach to estimate the parameters and allows restrictions to be placed on these parameters through the incorporation of penalty functions in the likelihood. This allows prior knowledge about the parameters to be incorporated into the estimation process but it is necessary to test the sensitivity of the estimated parameters to these priors. I would like to have seen more exploration of these sensitivities to understand the robustness of the model. For example, how sensitive is the model to the choice of CV and autocorrelation in the recruitment series? Potentially this could affect the estimation of MSY reference points.

The model is highly complex and the underlying state-space formulation can in principle estimate both process and measurement error. Where data are informative the model is likely to be able to estimate these variances, but with poor or noisy data there is a risk that model estimates are unreliable. I felt that this assessment may fall into the latter category. Thus while the assessment model is in itself scientifically sound I would not regard it as robust, given the limitations in the data available, without a clearer exploration of the consequences of choices about the values used to configure the model.

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

The population model within SSASPM is age and fleet structured and this requires information on fleet selectivity by age. The assessment panel has made sensible and pragmatic use of available information to convert length frequency data to age using growth models and then to derive selectivity functions based on a visual interpretation of the age frequencies. This is a necessity given the model formulation but is far from ideal and the approximations will necessarily be crude and may well be biased. Furthermore, the availability of length frequency data appears to be very limited and it is difficult to tell whether the length-to-age-to-selectivity conversions are realistic.

The model has a complex error structure having both process and measurement components. In order for these to be estimable the assessment panel has had to make further assumptions about the relative errors in the different data components and the process error. While it is clear that great care has been taken in arriving at these assumptions, and they are consistent with what any reasonable practitioner would do, they add to the uncertainty in the assessment because they rely on expert judgment. It is indicative of problems with the data that these constraints were necessary.

c) *Are the methods appropriate for the available data?*

The assessment results show that it is possible to interpret a very limited amount of data in order to infer plausible stock trends. But this is conditioned on a very large number of external assumptions about the data whose consequences are not easily understood even with the sensitivity analysis provided. I would regard the methods as giving a credible insight into possible stock scenarios based on incomplete clues in the disparate data.

Whether the methods are appropriate for the available data is very much a question of preference. I felt the complexity of the assessment model was not consistent with the limited amount of data. For example, the shrimp bycatch data are not really an observation of catch but an index of effort, yet the data are treated as catch in the model and no effort index is included in the model. Furthermore, these data dominate the assessment. Similarly there are very limited data on length, yet these have not only been converted to age without aging information but also used to estimate fleet selectivity. This has been done because the SSASPM model requires it, not because the data demand it. I would have thought, at least for exploratory purposes, a simpler model that treats the data in its natural form would be desirable since this would make clearer where the limitations of the data lie and simplify the number of *ad hoc* assumptions that need to be made. It would not solve the problem of the amount of information in the data, nor necessarily lead to different conclusions about stock status, but it should make interpretation of the results more straight forward.

3. Evaluate the assessment findings

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

The assessment is divided into an historical period and a modern period (from 1972 onwards). The historical period is based almost entirely on assumptions about the fishery and the way catches may have developed. It is therefore largely speculative and I would regard trends in biomass and exploitation during this period as a tantalizing insight into what might have occurred rather than reliable estimates of how the stock and fishery developed.

For the more recent period the model has more data to inform population estimates and trends here may be more useful. However, the assessment is dominated by the shrimp bycatch data, which I have suggested cannot truly reflect abundance due to the way the estimates were derived. I would therefore regard the trends estimated by the hierarchical index as the best indicator of stock trends because it should correctly weight the abundance indices without introducing bias from the shrimp bycatch data. The trends will be more adequately estimated from about the mid1990s onwards when there is a more complete set of indices available.

With regard to exploitation rate, I have already expressed reservations about the shrimp bycatch data and I therefore fear that the estimated trends in F may be biased. In fact the bycatch data may well be the best indicator of trends in F since it is a direct measure of effort and this fishery dominates the catch.

The recent trend in the hierarchical index and the bycatch suggest the biomass is increasing and F is decreasing. The assessment also shows similar qualitative trends. Consequently it seems likely that these estimated trends are robust. It is far more difficult, however, to have confidence in the estimate of stock status relative to reference points.

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

A number of sensitivity runs were performed that examine aspects of the data (mainly the abundance indices), weighting regimes and the biology of the Atlantic/Gulf split. These runs tend to show most evaluations of stock status as being not over-fished, but many are close to the threshold. Given the uncertainty in the shrimp bycatch data and its potential impact on the assessment, I do not feel it is possible to be confident that the stock is not over-fished.

The only sensitivity run examining the catch data is to consider halving the shrimp bycatch. While useful, this does not address the problem that the trend in the shrimp bycatch data may well be biased. Sensitivity runs need to be performed looking at this likely bias to see how stock status evaluations may be affected.

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

In common with the over-fished evaluation above, most sensitivity runs suggest the stock is not experiencing overfishing. While the recent reduction in F is likely to have occurred, it is problematic to evaluate overfishing status without sensitivity runs that consider potential bias in the shrimp bycatch estimates.

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

I was not able to fully understand how recruitment was modeled in the assessment. Equation (2) (page 70) shows a Beverton-Holt relationship relating recruitment structurally to spawning stock fecundity. In addition the report refers to recruitment (as opposed to recruitment deviations) following a first order autoregressive process which has no structural connection to spawning stock fecundity (SSF). Typically the autoregressive process would refer to the multiplicative process error added to nominal log recruitment but the report indicates that the annual recruitment deviations were not estimated. Thus I could not see how these two descriptions of recruitment were reconciled. This issue needs clarification.

Results from the model fit give plausible estimates of the stock-recruitment model parameters, though it would be necessary to test the robustness of these values with sensitivity tests that looked at shrimp bycatch data and the possible biases therein.

I would suspect that the estimated stock – recruitment relationship is adequate for exploratory purposes but needs to be treated with some caution.

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?*

Criteria for status determination are the ratio of current stock size and F to respective MSY values. Such indicators will be more robust than point estimates since they avoid the problem of arbitrary scaling which is likely to affect this assessment. Their interpretation in this assessment, however, is affected by the uncertainties in the catch data.

4. Evaluate the stock projections, rebuilding timeframes, and generation times

a) *Are the methods consistent with accepted practices and available data?*

Projection methods adopted conform to standard procedures where initial population sizes mortality rates and recruitment are drawn from parametric error distributions estimated from the assessment model. Generation times are defined in terms of mean age of parents of offspring over the lifetime of a cohort. This is a commonly adopted definition.

The projections are run for fixed levels of catch and in a sense this implies controlling the catch leads to a control on fishing mortality. This is, of course, standard practice, especially where targeted fisheries are concerned. In this fishery most of the catch appears to come from the shrimp fishery as bycatch and effort reductions are likely to be the most important variable in limiting catch. It might therefore be worth considering projection models where future effort is considered as the control variable rather than catch. This would require deriving a relationship between shrimp fishery F and effort.

b) *Are the methods appropriate for the assessment model and outputs?*

Given the choice of SSASPM as the assessment tool the projection methods flow naturally from this and its underlying population model. Clearly the initial starting conditions are dependent on the quality of the assessment output, thus all future scenarios are conditioned on the assumption that the SSASPM output is a good representation of reality.

Forward projections using the same population model as the assessment are desirable because some errors in the assessment will cancel out, at least for short time horizons and hence reduce sensitivity to errors.

c) *Are the results informative and robust, and useful to support inferences of probable future conditions?*

The results from the projections are likely to adequately reflect stock development for a few years ahead. Any projections beyond a decade are fraught with uncertainties and should be regarded only as illustrative of what might happen. When the projection reaches a condition where all the population has been generated from within the model (as opposed to measurements taken in the base year) the result is entirely driven by model assumptions and are likely to be unreliable.

While I would support the view that the projections offer an insight into stock trajectories, I would be quite skeptical of the probabilities associated with the reference points. These are too heavily dependent on the model estimates of fishing mortality, which are likely to be highly uncertain.

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

Two principal sources of uncertainty are considered in the projections. These are the error distributions associated with the initial conditions and model uncertainty as expressed by the sensitivity scenarios. I believe the uncertainties are much larger than is reflected in the chosen sensitivity scenarios, especially in relation to the shrimp fishery bycatch, but there are also questions about the length-age conversions and resultant fleet selectivity functions. Over long time horizons these errors could accumulate and have a major impact on the stock trajectory.

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

a) *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*

Uncertainties are expressed mainly in relation to the SSASPM model output. These are the asymptotic variances, likelihood profiles or posterior distributions. In addition a large number of sensitivity scenarios are considered that examine uncertainty in the survey data, the catch data, the area of the stock, biological productivity and base year in the assessment. These are all important and useful analyses that help understand the model used conditioned on the data presented to it. I did not feel, however, that these analyses covered some critical areas of uncertainty relating to the input data and I would have liked to see an attempt to explore model uncertainty with a substantially different model that made alternative assumptions about population dynamics (e.g. length based, or simply biomass based).

The attentive reader will by now be aware of my concern over the catch data. At the very least minimal estimates of the measurement error associated with the MRIP and shrimp bycatch data need to be presented. It would also be useful to compare these error estimates with those that emerge from fitting the model. It is noteworthy that the model appears to fit the catch data very well, far better than might be expected given the fact that these data are derived from surveys or reconstructed from effort data. It shows that the model is heavily driven by the catch data and hence the inevitable errors in these data are translated directly into the estimated parameters. A sensitivity test that would be of particular interest would be to draw random vectors of catch from a multivariate lognormal distribution based on the measurement errors in the catch to see how the model responds to different catch trends.

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

As the assessment report is already complete, the reviewer can only comment on uncertainty in the review report.

6. Consider the research recommendations provided and make any additional recommendations or prioritizations warranted.

a) *Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments.*

Two research recommendations are made in the report which relate to improving the shrimp by-catch estimates and the integration of indices. I would strongly support both of these recommendations. In the case of the shrimp bycatch it is of particular importance to develop a method that is able to reflect the response of the catch to changes in stock abundance. The method used for this assessment uses only effort and is a severe weakness which is bound to result in bias in the estimates.

Assessment models are often a matter of taste where practitioners adopt the model that they prefer. I felt the model used in this assessment demanded far too much from the data and as a consequence a very large number of assumptions had to be made which can only be examined by a very complex set of sensitivity tests whose results themselves are not easy to digest or interpret. I would recommend that an assessment model that is more closely designed around the available data is developed. If, in reality, shrimp bycatch data cannot be estimated with any precision, then it might be better to treat these catches as unknown and use the effort data directly in the model. It would also be preferable to describe the population in terms of length rather than age.

b) *Provide recommendations on possible ways to improve the SEDAR process.*

The assessment report documents are very well written and presented. They were available on time and all this made the work of review much easier. I missed the opportunity to raise questions about the assessments with the assessment panel. It might be useful to try to organize a webinar at some stage during the review so that reviewers can seek clarification from the assessment panel to avoid mistakes and mis-understandings in their reports. I imagine that some of my comments may have arisen through incomplete understanding of what was done.

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

The problems with the shrimp bycatch could be avoided if a simpler assessment method was used. The methods developed by Brooks *et al* (2010) appear to offer a very promising alternative and have been used for this stock. One of the main questions is choosing an appropriate abundance index to use. The development of the hierarchical index (Conn, 2010) would be an obvious choice. The simplicity of the Brooks method means that it would make sensitivity analysis much simpler and easier to interpret. I would suggest that this approach is tried alongside the SSASPM method for comparison and as an additional sensitivity test.

Atlantic sharpnose shark

Introductory remarks

There are substantial similarities between the assessments of the bonnethead shark and Atlantic sharpnose stocks, both in the data used and the methodology applied. Indeed the two reports contain much the same text. Consequently much of the review commentary on the bonnethead shark assessment applies to the Atlantic sharpnose assessment and I have adopted a similar approach of reproducing bonnethead text below with relevant modifications where species differences arise.

One general point that should be made is that the bycatch data from the shrimp trawl have been derived from the same effort data as was used for bonnethead. As a result the catch trends for sharpnose are almost identical to those of bonnethead and this results in very similar estimated trends in fishing mortality and stock biomass. Figure 1 shows the trends in the shrimp bycatch for the two species illustrating the close correlation and inconsistent change of scale in the transition from the historical period to the modern period. I would have thought this close correlation is unusually high and indicates a problem with the data and is likely to be an artifact of the estimation procedure.

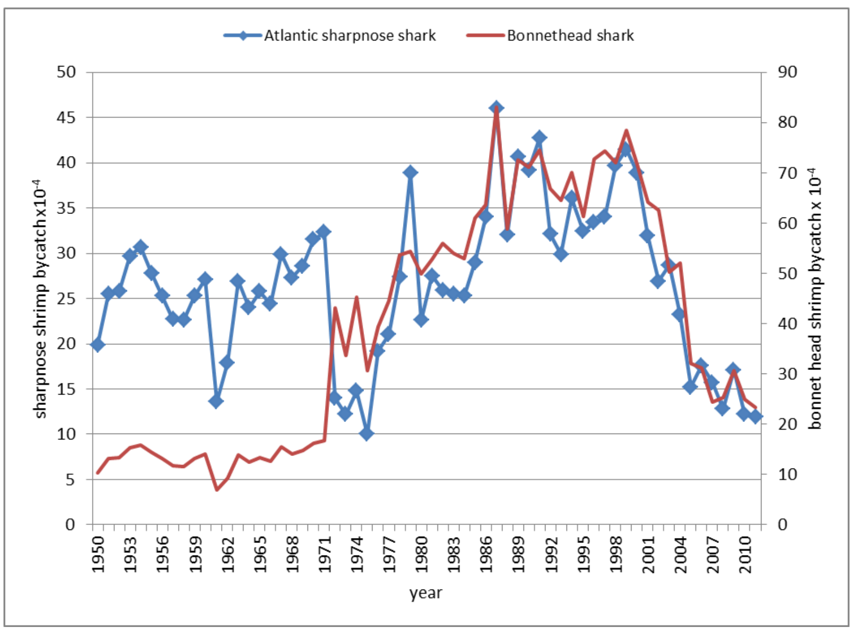

Figure 1. Trends in bycatches in the shrimp fishery of bonnethead and Atlantic sharpnose sharks.

1. Evaluate the data used in the assessment

a) Are data decisions made by the assessment panel sound and robust?

The assessment report provides a clear summary of the relevant documents and the reasons for taking the various decisions. Overall I would concur with the decisions made, but make a number of observations.

The decision to adopt "method 2" in the reconstruction of shrimp bycatch is fair enough given the choice between a raw raising procedure and one using logs. However, I was not convinced that the use of a single CPUE mean (over the years 2009-2011) represented the best way to convert effort data into catch data. I believe this to be a potentially very significant issue, and is discussed at various points in this review, because it may have the largest single impact on the results of the assessment.

Whilst I would agree with the decisions on the choice of discard mortality rates in the absence of highly informative data, I would like to have seen some discussion on the impact of errors associated with this choice on the estimate of the catches since small changes this mortality in the recreational fishery has potentially large effects on the estimate of dead catch with consequences for the assessment.

The decisions to include/exclude abundance indices were based on sensible and objective criteria, such as length of series, area coverage, reproducibility of indices etc. I felt that these choices were sound.

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

The assessment report does discuss many of the uncertainties in the data but I felt much more could have been done to provide quantitative information on the uncertainty in the estimates of the derived catches and the abundance indices. I have been led to understand that the MRIP survey is now based on a proper statistical design and it should therefore be possible to provide estimates of the variances for at least the sampling error associated with the recreational catch. It should also be possible to do this for the derived shrimp bycatch by considering the variance associated with the mean CPUE used and the stratification of the effort data. These two catch components dominate the assessment estimates of fishing mortality and it is particularly important to consider the uncertainty in these data.

I was particularly concerned about the use of a single mean CPUE per stratum to raise the annual effort data to total catch for the shrimp fishery. This procedure effectively assumes that the stock from which the catch is derived is of fixed abundance over time. It means that the catch increases linearly with effort and implies increases in fishing effort have no effect on stock abundance. This is not a credible assumption. In effect the shrimp bycatch is an index of effort scaled to catch units. In reality you would expect an increase in effort to lead to a transient increase in catch followed by a decline as the stock adjusts to a new state. Similarly you would expect a decrease in effort to lead to a transient loss of catch followed by some recovery. These effects will not be reflected in the shrimp bycatch data, which are therefore likely to contain an unknown bias. Such biases will emerge as biased estimates of fishing mortality and biomass in the assessment. It is of some importance to establish whether these biases undermine the assessment results.

An interesting development reported is the derivation of a hierarchical abundance index. The analysis on which this is based should provide estimates of both process and measurement error associated with the surveys and could be presented in the report to illustrate the uncertainties in the indices associated with these two elements.

c) *Are data applied properly within the assessment model?*

Given the choice of SSASPM as the preferred assessment tool, the data have been applied in the appropriate manner. Length frequency data are converted to age, selectivity curves have been generated and error distributions defined for each data component.

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

On purely scientific grounds I would argue that the data do not contain sufficient information to justify the complexity of assessment model attempted here. An indication of this is the fact that some parameters (e.g. fleet selectivity) have had to be estimated externally from the model as they appear not to be estimable within it. However, it is frequently the case that stakeholders and managers expect far more detail, and want the information of particular concern to them to be included in an assessment. This often leads to inflation in the complexity of the assessment model, which may satisfy clients of the assessment but which adds to the uncertainty of the results due to the necessity to make numerous structural and simplifying assumptions, each with associated errors, which tend to accumulate.

Above I have already commented on the issues of uncertainty relating to the shrimp bycatch data. It is noteworthy that the abundance indices, while numerous, do not show a high degree of consistency and this is itself indicative of uncertainty in the assessment. The hierarchical index provides a useful summary of the indices to the extent that it suggests some variability around a stationary mean in the earlier period and a gradual increase more recently. This shows some similarity with trends in the shrimp fishery effort, from which most of the catch appears to come, and provides some reassurance about overall stock trends.

Given my reservations about the catch data, I would be skeptical about the ability of the assessment to determine stock status relative to reference points.

2. Evaluate the methods used to assess the stock, taking into account the available data.

a) *Are methods scientifically sound and robust?*

The SSASPM model that lies at the core of the assessment is a fully age structured forward projection statistical model that conforms to current standards and practice for the assessment of stocks with appropriate data. The model uses a maximum likelihood approach to estimate the parameters and allows restrictions to be placed on these parameters through the incorporation of penalty functions in the likelihood. This allows prior knowledge about the parameters to be incorporated into the estimation process, but it is necessary to test the sensitivity of the estimated parameters to these priors. I would like to have seen more exploration of these sensitivities to understand the robustness of the model. For example, how sensitive is the model to the choice of CV and autocorrelation in the recruitment series? Potentially this could affect the estimation of MSY reference points.

The model is highly complex and the underlying state-space formulation can in principle estimate both process and measurement error. Where data are informative the model is likely to be able to estimate these variances, but with poor or noisy data there is a risk that model estimates are unreliable. I felt that this assessment may fall into the latter category. Thus, while the assessment model is in itself scientifically sound, I would not regard it as robust given the limitations in the data available without a clearer exploration of the consequences of choices about the values used to configure the model.

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

The population model within SSASPM is age and fleet structured and this requires information on fleet selectivity by age. The assessment panel has made sensible and pragmatic use of available information to convert length frequency data to age using growth models and then to derive selectivity functions based on a visual interpretation of the age frequencies. This is a necessity given the model formulation, but is far from ideal and the approximations will necessarily be crude and may well be biased. Furthermore, the availability of length frequency data appears to be very limited and it is difficult to tell whether the length-to-age-to-selectivity conversions are realistic.

The model has a complex error structure having both process and measurement components. In order for these to be estimable the assessment panel has had to make further assumptions about the relative errors in the different data components and the process error. While it is clear that great care has been taken in arriving at these assumptions, and they are consistent with what any reasonable practitioner would do, they add to the uncertainty in the assessment because they rely on expert judgment. It is indicative of problems with the data that these constraints were necessary.

c) *Are the methods appropriate for the available data?*

The assessment results show that it is possible to interpret a very limited amount of data in order to infer plausible stock trends. But this is conditioned on a very large number of external assumptions about the data whose consequences are not easily understood even with the sensitivity analysis provided. I would regard the methods as giving a credible insight into possible stock scenarios based on incomplete clues in the disparate data.

Whether the methods are appropriate for the available data is very much a question of preference. I felt the complexity of the assessment model was not consistent with the limited amount of data. For example, the shrimp bycatch data are not really an observation of catch but an index of effort, yet the data are treated as catch in model and no effort index is included in the model. Furthermore, these data dominate the assessment. Similarly there are very limited data on length, yet these have not only been converted to age without aging information but also used to estimate fleet selectivity. This has been done because the SSASPM model requires it, not because the data demand it. I would have thought, at least for exploratory purposes, a simpler model that treats the data in its natural form would be desirable since this would make clearer where the limitations of the data lie and simplify the number of *ad hoc* assumptions that need to be made. It would not solve the problem of the amount of information in the data, nor necessarily lead to different conclusions about stock status, but it should make interpretation of the results more straight forward.

3. Evaluate the assessment findings

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

The assessment is divided into an historical period and a modern period (from 1972 onwards). The historical period is based almost entirely on assumptions about the fishery and the way catches may have developed. It is therefore largely speculative and I would regard trends in biomass and exploitation during this period as a tantalizing insight into what might have occurred rather than reliable estimates of how the stock and fishery developed.

For the more recent period the model has more data to inform population estimates, and trends here may be more useful. However, the assessment is dominated by the shrimp bycatch data, which I have suggested cannot truly reflect abundance due to the way the estimates were derived. I would therefore regard the trends estimated by the hierarchical index as the best indicator of stock trends because it should correctly weight the abundance indices without introducing bias from the shrimp bycatch data. The trends will be more adequately estimated from about the mid1990s onwards when there is a more complete set of indices available.

With regard to exploitation rate, I have already expressed reservations about the shrimp bycatch data and I therefore fear that the estimated trends in F may be biased. In fact the bycatch data may well be the best indicator of trends in F since it is a direct measure of effort and this fishery dominates the catch.

The recent trend in the hierarchical index and the bycatch suggest the biomass is increasing and F is decreasing. The assessment also shows similar qualitative trends. Consequently it seems likely that these estimated trends are robust. It is far more difficult, however, to have confidence in the estimate of stock status relative to reference points.

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

A number of sensitivity runs were performed that examine aspects of the data (mainly the abundance indices) and weighting regimes. These runs tend to show most evaluations of stock status as being not over-fished. Given the uncertainty in the shrimp bycatch data and its potential impact on the assessment, I do not feel it is possible to be confident that the stock is not over-fished.

The only sensitivity run examining the catch data is to consider halving the shrimp bycatch. While useful, this does not address the problem that the trend in the shrimp bycatch data may well be biased. Sensitivity runs need to be performed looking at this likely bias to see how stock status evaluations may be affected.

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

In common with the over-fished evaluation above, most sensitivity runs suggest the stock is not experiencing overfishing. While the recent reduction in F is likely to have occurred, it is problematic to evaluate overfishing status without sensitivity runs that consider potential bias in the shrimp bycatch estimates.

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

I was not able to fully understand how recruitment was modeled in the assessment. Equation (2) (page 72) shows a Beverton-Holt relationship relating recruitment structurally to spawning stock fecundity (SSF). In addition the report refers to recruitment (as opposed to recruitment deviations) following a first order autoregressive process that has no structural connection to SSF. Typically the autoregressive process would refer to the multiplicative process error added to nominal log recruitment, but the report indicates that the annual recruitment deviations were not estimated. Thus I could not see how these two descriptions of recruitment were reconciled. This issue needs clarification.

Results from the model fit give plausible estimates of the stock-recruitment model parameters, though it would be necessary to test the robustness of these values with sensitivity tests that looked at shrimp bycatch data and the possible biases therein.

I would suspect that the estimated stock – recruitment relationship is adequate for exploratory purposes but needs to be treated with some caution.

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?

Criteria for status determination are the ratio of current stock size and F to respective MSY values. Such indicators will be more robust than point estimates since they avoid the problem of arbitrary scaling which is likely to affect this assessment. Their interpretation in this assessment, however, is affected by the uncertainties in the catch data.

4. Evaluate the stock projections, rebuilding timeframes, and generation times, addressing the following:

a) Are the methods consistent with accepted practices and available data?

Projection methods adopted conform to standard procedures where initial population sizes mortality rates and recruitment are drawn from parametric error distributions estimated from the assessment model. Generation times are defined in terms of mean age of parents of offspring over the lifetime of a cohort. This is a commonly adopted definition.

The projections are run for fixed levels of catch and in a sense this implies controlling the catch leads to a control on fishing mortality. This is, of course, standard practice, especially where targeted fisheries are concerned. In this fishery most of the catch appears to come from the shrimp fishery as bycatch, and effort reductions are likely to be the most important variable in limiting catch. It might therefore be worth considering projection models where future effort is considered as the control variable rather than catch. This would require deriving a relationship between shrimp fishery F and effort.

b) *Are the methods appropriate for the assessment model and outputs?*

Given the choice of SSAASPM as the assessment tool, the projection methods flow naturally from this and its underlying population model. Clearly the initial starting conditions are dependent on the quality of the assessment output, thus all future scenarios are conditioned on the assumption that the SSASPM output is a good representation of reality.

Forward projections using the same population model as the assessment are desirable because some errors in the assessment will cancel out, at least for short time horizons and hence reduce sensitivity to errors.

c) *Are the results informative and robust, and useful to support inferences of probable future conditions?*

The results from the projections are likely to adequately reflect stock development for a few years ahead. Any projections beyond a decade are fraught with uncertainties and should be regarded only as illustrative of what might happen. When the projection reaches a condition where all the population has been generated from within the model (as opposed to measurements taken in the base year), the result is entirely driven by model assumptions and are likely to be unreliable.

While I would support the view that the projections offer an insight into stock trajectories, I would be quite skeptical of the probabilities associated with the reference points. These are too heavily dependent on the model estimates of fishing mortality, which are likely to be highly uncertain.

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

Two principal sources of uncertainty are considered in the projections. These are the error distributions associated with the initial conditions and model uncertainty as expressed by the sensitivity scenarios. I believe the uncertainties are much larger than is reflected in the chosen sensitivity scenarios, especially in relation to the shrimp fishery bycatch, but there are also questions about the length-age conversions and resultant fleet selectivity functions. Over long time horizons these errors could accumulate and have a major impact on the stock trajectory.

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

a) *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*

Uncertainties are expressed mainly in relation to the SSASPM model output. These are the asymptotic variances, likelihood profiles or posterior distributions. In addition a large number of sensitivity scenarios are considered that examine uncertainty in the survey data, the catch data, the area of the stock, biological productivity and base year in the assessment. These are all important and useful analyses that help understand the model used conditioned on the data presented to it. I did not feel, however, that these analyses covered some critical areas of uncertainty relating to the input data, and I would have like to see an attempt to explore model uncertainty with a substantially different model that made alternative assumptions about population dynamics (e.g. length based, or simply biomass based).

I have frequently referred to concerns over the catch data. At the very least, minimal estimates of the measurement error associated with the MRIP and shrimp bycatch data need to be presented. It would also be useful to compare these error estimates with those that emerge from fitting the model. It is noteworthy that the model appears to fit the catch data very well, far better than might be expected given the fact that these data are derived from surveys or reconstructed from effort data. It shows that the model is heavily driven by the catch data and hence the inevitable errors in these data are translated directly into the estimated parameters. A sensitivity test that would be of particular interest would be to draw random vectors of catch from a multivariate lognormal distribution based on the measurement errors in the catch to see how the model responds to different catch trends.

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

As the assessment report is already complete, the reviewer can only comment on uncertainty in the review report.

6. Consider the research recommendations provided and make any additional recommendations or prioritizations warranted.

a) *Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments.*

Two research recommendations are made in the report which relate to improving the shrimp by-catch estimates and the integration of indices. I would strongly support both of these recommendations. In the case of the shrimp bycatch it is of particular importance to develop a method that is able to reflect the response of the catch to changes in stock abundance. The method used for this assessment uses only effort and is a severe weakness that is bound to result in bias in the estimates.

Assessment models are often a matter of taste, where practitioners adopt the model that they prefer. I felt the model used in this assessment demanded far too much from the data and as a consequence a very large number of assumptions had to be made which can only be examined by a very complex set of sensitivity tests whose results themselves are not easy to digest or interpret. I would recommend that an assessment model that is more closely designed around the available data is developed. If, in reality, shrimp bycatch data cannot be estimated with any precision, then it might be better to treat these catches as unknown and use the effort data directly in the model. It would also be preferable to describe the population in terms of length rather than age.

b) *Provide recommendations on possible ways to improve the SEDAR process.*

The assessment report documents are very well written and presented. They were available on time and all this made the work of review much easier. I missed the opportunity to raise questions about the assessments with the assessment panel. It might be useful to try to organize a webinar at some stage during the review so that reviewers can seek clarification from the assessment panel to avoid mistakes and mis-understandings in their reports. I imagine that some of my comments may have arisen through incomplete understanding of what was done.

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

The problems with the shrimp bycatch could be avoided if a simpler assessment method was used. The methods developed by Brooks *et al* (2010) appear to offer a very promising alternative and have been used for this stock. One of the main questions is choosing an appropriate abundance index to use. The development of the hierarchical index (Conn, 2010) would be an obvious choice. The simplicity of the Brooks method means that it would make sensitivity analysis much simpler and easier to interpret. I would suggest that this approach is tried alongside the SSPASM method for comparison and as an additional sensitivity test.

References

Brooks, E.N., J.E. Powers, and E. Cortés. 2010. Analytic reference points for age-structured models: application to data-poor fisheries. ICES J. Mar. Sci. 67:165-175.

Conn, P.B. 2010. Hierarchical analysis of multiple noisy abundance indices. Can. J. Fish. Aquat. Sci. 67:108-120.

Appendix 1: Bibliography of materials provided for review

Documents Prepared for the Assessment Process		
SEDAR34-WP-01	Standardized catch rates of Atlantic sharpnose sharks (*Rhizoprionodon terraenovae*) in the U.S. Gulf of Mexico from the Shark Bottom Longline Observer Program, 1994-2011	John Carlson and Simon Gulak
SEDAR34-WP-02	Standardized catch rates of bonnetheads from the Everglades National Park Creel Survey	John K. Carlson and Jason Osborne
SEDAR34-WP-03	Standardized Catch Rates of Bonnethead and Atlantic Sharpnose Shark from the Southeast Shark Drift Gillnet Fishery: 1993-2011	John Carlson, Alyssa Mathers and Michelle Passerotti
SEDAR34-WP-04	Tag and recapture data for Atlantic sharpnose, *Rhizoprionodon terraenovae*, and bonnethead shark, *Sphyrna tiburo*, in the Gulf of Mexico: 1999-2011	Dana Bethea and Mark Grace
SEDAR34-WP-05	Relative abundance of bonnethead and Atlantic sharpnose sharks based on a fishery-independent gillnet survey off Texas	Walter Bubley and John Carlson
SEDAR34-WP-06	Update to maximum observed age of Atlantic sharpnose sharks (*Rhizoprionodon terraenovae*) in the western North Atlantic Ocean based on a direct age estimate of a long term recapture	Bryan S. Frazier and Joshua K. Loefer
SEDAR34-WP-07	Validated age and growth of the bonnethead (*Sphyrna tiburo*) in the western North Atlantic Ocean	Bryan S. Frazier, Douglas H. Adams, William B. Driggers III, Christian M. Jones, Joshua K. Loefer, Linda A. Lombardi
SEDAR34-WP-08	A preliminary review of post-release live-discard mortality rate estimates in sharks for use in SEDAR 34	Dean Courtney
SEDAR34-WP-09	Standardized catch rates of Atlantic sharpnose (*Rhizoprionodon terraenovae*) and bonnethead (*Sphyrna tiburo*) sharks collected during a gillnet survey in Mississippi coastal waters, 1998-2011	Eric R. Hoffmayer, Glenn R. Parsons, Jill M. Hendon, Adam G. Pollack, and G. Walter Ingram, Jr.
SEDAR34-WP-10	Standardized catch rates of Atlantic sharpnose sharks (*Rhizoprionodon terraenovae*) collected during a bottom longline survey in Mississippi coastal waters, 2004-2011	Eric R. Hoffmayer, Jill M. Hendon, and Adam G. Pollack
SEDAR34-WP-11	Standardized catch rates of Atlantic sharpnose sharks (Rhizoprionodon terraenovae) collected during bottom longline surveys in Mississippi, Louisiana, Alabama, and Texas coastal waters, 2004-2011	Eric Hoffmayer, Adam Pollack, Jill Hendon, Marcus Drymon, and Mark Grace
SEDAR34-WP-12	Atlantic Sharpnose Shark: Standardized index of relative abundance using boosted regression trees and generalized linear models	John Froeschke and J. Marcus Drymon

SEDAR34-WP-13	Atlantic Sharpnose Abundance Indices from SEAMAP Groundfish Surveys in the Northern Gulf of Mexico	Adam G. Pollack and G. Walter Ingram, Jr.
SEDAR34-WP-14	Bonnethead Abundance Indices from SEAMAP Groundfish Surveys in the Northern Gulf of Mexico	Adam G. Pollack and G. Walter Ingram, Jr.
SEDAR34-WP-15	Atlantic Sharpnose and Bonnethead Abundance Indices from NMFS Bottom Longline Surveys in the Western North Atlantic and Northern Gulf of Mexico	Adam G. Pollack and G. Walter Ingram, Jr.
SEDAR34-WP-16	Continuity Runs for Atlantic Sharpnose and Bonnethead SEAMAP Groundfish Surveys and NMFS Bottom Longline Surveys	Adam G. Pollack and G. Walter Ingram, Jr.
SEDAR34-WP-17	Variability in the Reproductive Biology of the Atlantic Sharpnose Shark in the Gulf of Mexico	Eric R. Hoffmayer, Jill M. Hendon, William B. Driggers III, Lisa M. Jones, and James A. Sulikowski
SEDAR34-WP-18	Shrimp Fishery Bycatch Estimates for Atlantic Sharpnose and Bonnethead Sharks in the Gulf of Mexico, 1972-2011	Xinsheng Zhang, Brian Linton, Enric Cortés and Dean Courtney
SEDAR34-WP-19	Standardized catch rates of Atlantic sharpnose and bonnethead sharks from the SEAMAP-South Atlantic Shallow Water Trawl Survey	Enric Cortés and J. Boylan
SEDAR34-WP-20	Updated catches of Atlantic sharpnose and bonnethead sharks	Enric Cortés and Ivy Baremore
SEDAR34-WP-21	Dead discards of Atlantic sharpnose sharks in the shark bottom longline fishery	John Carlson, Kevin J. McCarthy and Simon J.B. Gulak
SEDAR34-WP-22	Preliminary data on the reproductive biology of the bonnethead (*Sphyrna tiburo*) from the southeast U.S. Atlantic coast	Bryan Frazier, Jim Gelsleichter, and Melissa Gonzalez De Acevedo
SEDAR34-WP-23	Interannual site fidelity of bonnetheads (*Sphyrna tiburo*) to two coastal ecosystems in the western North Atlantic Ocean	William B. Driggers III, Bryan S. Frazier, Douglas H. Adams, Glenn F. Ulrich and Eric R. Hoffmayer
SEDAR34-WP-24	Size composition and indices of relative abundance of the Atlantic sharpnose shark (*Rhizoprionodon terraenovae*) in coastal Virginia waters	Robert J. Latour, Christopher F. Bonzek, and J. Gartland
SEDAR34-WP-25	Mark/Recapture Data for the Atlantic Sharpnose Shark (*Rhizoprionodon terranovae)*, in the Western North Atlantic from the NEFSC Cooperative Shark Tagging Program	Nancy E. Kohler, Danielle Bailey, Patricia A. Turner, and Camilla McCandless
SEDAR34-WP-26	Mark/Recapture Data for the Bonnethead (*Sphyrna tiburo)*, in the Western North Atlantic from the NEFSC Cooperative Shark Tagging Program	Nancy E. Kohler, Elizabeth Sawicki, Patricia A. Turner, and Camilla McCandless
SEDAR34-WP-27	Preliminary mtDNA assessment of genetic stock structure of the bonnethead, *Sphyrna tiburo*, in the eastern Gulf of Mexico and northwestern Atlantic	Píndaro Díaz-Jaimes, Douglas H. Adams, Nadia S. Laurrabaquio-Alvarado, Elena Escatel-Luna
SEDAR34-WP-28	Standardized Catch Rates of Bonnethead and	John Carlson, Alyssa

	Atlantic Sharpnose Shark from the Southeast Sink Gillnet Fishery: 2005-2011	Mathers and Michelle Passerotti
SEDAR34-WP-29	Relative abundance of Atlantic sharpnose and bonnethead shark from the northeastern Gulf of Mexico	John K. Carlson, Dana M. Bethea, Eric Hoffmayer, John Tyminski, Robert Hueter, R. Dean Grubbs, Matthew J. Ajemian, and George H. Burgess
SEDAR34-WP-30	Reproductive parameters for Atlantic sharpnose sharks (*Rhizoprionodon terraenovae*) from the western North Atlantic Ocean	William B. Driggers III, Eric R. Hoffmayer, John K. Carlson and Joshua Loefer
SEDAR34-WP-31	Tag-recapture results of bonnethead (*Sphyrna tiburo*) and Atlantic sharpnose (*Rhizoprionodon terraenovae*) sharks in the Gulf of Mexico and Florida Coastal Waters	John P. Tyminski, Robert E. Hueter, John Morris
SEDAR34-WP-32	Standardized catch rates of bonnethead (*Sphyrna tiburo*) from the South Carolina Department of Natural Resources trammel net survey	Bryan S. Frazier and Camilla T. McCandless
SEDAR34-WP-33	Tag and recapture data for Atlantic sharpnose, *Rhizoprionodon terraenovae*, and bonnethead, *Sphyrna tiburo*, sharks caught in the northern Gulf of Mexico from 1998-2011	Jill M. Hendon, Eric R. Hoffmayer, and Glenn R. Parsons
SEDAR34-WP-34	Standardized indices of abundance for Atlantic sharpnose sharks from the Georgia Department of Natural Resources red drum longline survey	C.T. McCandless, C.N. Belcher
SEDAR34-WP-35	Standardized indices of abundance for bonnethead and Atlantic sharpnose sharks from the Georgia Department of Natural Resources ecological monitoring trawl surveys	C.T. McCandless, J.Page, C.N. Belcher
SEDAR34-WP-36	Standardized indices of abundance for bonnethead and Atlantic sharpnose sharks caught during the South Carolina Department of Natural Resources red drum longline and Cooperative Atlantic States Shark Pupping and Nursery gillnet surveys	C.T. McCandless, B.S. Frazier
SEDAR34-WP-37	Standardized indices of abundance for bonnethead and Atlantic sharpnose sharks caught during the Cooperative Atlantic States Shark Pupping and Nursery longline surveys from South Carolina to northern Florida	C.T. McCandless, C.N. Belcher, B.S. Frazier, M. McCallister, R. Ford, J. Gelsleichter
SEDAR34-WP-38	Standardized indices of abundance for Atlantic sharpnose sharks from the University of North Carolina bottom longline survey	Frank Schwartz, Camilla McCandless, and John Hoey
SEDAR34-WP-39	A Summary of Evaluation Worksheets of abundance indices for Atlantic sharpnose shark and bonnethead shark	SEDAR 34 Panel
Final Stock Assessment Reports		
SEDAR34-SAR	Atlantic Sharpnose Sharks	SEDAR 34 Panel
SEDAR34-SAR	Bonnethead Sharks	SEDAR 34 Panel

Reference Documents		
SEDAR29-RD01	SEDAR 13 (SCS) Final Stock Assessment Report	SEDAR 13 Panels
SEDAR29-RD02	Abundance Indices Workshop: Developing protocols for submission of abundance indices to the SEDAR process	SEDAR Procedural Workshop I
SEDAR29-RD03	Characterization of the U.S. Gulf of Mexico and South Atlantic Penaeid and Rock Shrimp Fisheries Based on Observer Data	ELIZABETH SCOTT-DENTON, PAT F. CRYER, MATT R. DUFFY, JUDITH P. GOCKE, MIKE R. HARRELSON, DONNA L. KINSELLA, JAMES M. NANCE, JEFF R. PULVER, REBECCA C. SMITH, and JO A. WILLIAMS
SEDAR29-RD04	Effects of Turtle Excluder Devices (TEDs) on the Bycatch of Three Small Coastal Sharks in the Gulf of Mexico Penaeid Shrimp Fishery	Scott W. Raborn, Benny J. Gallaway, John G. Cole, William J. Gazey & Kate I. Andrews

Appendix 2: CIE Statement of Work

Statement of Work for Dr. Robin Cook

External Independent Peer Review by the Center for Independent Experts

SEDAR 34: Highly Migratory Species Bonnethead Shark and Atlantic Sharpnose Shark Assessment Desk 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 34 will be a compilation of data, a standard assessment of the stock, and CIE assessment review conducted for HMS Bonnethead and Atlantic sharpnose sharks. The desk review provides an independent peer review of SEDAR stock assessments. The review is responsible for ensuring that the best possible assessment is provided through the SEDAR process and will provide guidance to the SEFSC to aid in their review and determination of best available science, and to HMS when determining if the assessment is useful for management. The stocks assessed through SEDAR 34 are within the jurisdiction of the Highly Migratory Species Division of NOAA Fisheries and the states of Texas, Louisiana, Mississippi, Alabama, Florida, Georgia, South Carolina, North Carolina, Virginia, Maryland, Delaware, Pennsylvania, New Jersey, New York, Connecticut, Rhode Island, Massachusetts, New Hampshire, and Maine. 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 statement of work (SoW) tasks and terms of reference (ToRs) specified herein. The CIE reviewers shall have expertise in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the tasks of the peer-review described herein. Each CIE reviewer's duties shall not exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

Location of Peer Review: Each CIE reviewer shall participate and conduct an independent peer review as a desk review, therefore travel will not be required.

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

Prior to the Peer Review: Upon completion of the CIE reviewer selection by the CIE Steering Committee, the CIE shall provide the CIE reviewer contact information to the COR, 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 assessment and other pertinent background documents for the peer review. Any changes to the SoW or ToRs must be made through the COR 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.

Desk Review: 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 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 CIE Lead Coordinator. The CIE Lead Coordinator can contact the Project Contact to confirm any peer review 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**.

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) Conduct an impartial and independent peer review in accordance with the tasks and ToRs specified herein, and each ToRs must be addressed (**Annex 2**).
3) No later than October 7, 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 CIE Regional Coordinator, via email to Dr. 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.

10 September 2013	CIE sends reviewer contact information to the COR, who then sends this to the NMFS Project Contact
17 September 2013	NMFS Project Contact sends the CIE Reviewers the assessment report and background documents
18 September through 02 October 2013	Each reviewer conducts an independent peer review as a desk review
07 October 2013	CIE reviewers submit draft CIE independent peer review reports to the CIE Lead Coordinator and CIE Regional Coordinator
21 October 2013	CIE submits CIE independent peer review reports to the COR
28 October 2013	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:

Julie 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.

   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

Annex 2 – Terms of Reference

SEDAR 34: Atlantic Sharpnose and Bonnethead Shark Assessment Review

1. Evaluate the data used in the assessment, addressing the following:
    e) Are data decisions made by the assessment panel sound and robust?
    f) Are data uncertainties acknowledged, reported, and within normal or expected levels?
    g) Are data applied properly within the assessment model?
    h) 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.
    d) Are methods scientifically sound and robust?
    e) Are assessment models configured properly and used consistent with standard practices?
    f) Are the methods appropriate for the available data?
3. Evaluate the assessment findings with respect to the following:
    f) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
    g) Is the stock overfished? What information helps you reach this conclusion?
    h) Is the stock undergoing overfishing? What information helps you reach this conclusion?
    i) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
    j) 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, rebuilding timeframes, and generation times, addressing the following:
    e) Are the methods consistent with accepted practices and available data?
    f) Are the methods appropriate for the assessment model and outputs?
    g) Are the results informative and robust, and useful to support inferences of probable future conditions?
    h) 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 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.
8. Prepare a Peer Review Report summarizing the Reviewer's evaluation of the stock assessment and addressing each Term of Reference.