Center for Independent Experts (CIE) External Independent Peer Review

SEDAR 54 HMS Sandbar Shark Assessment Review.

December 2017

Dr. Noel Cadigan

Centre for Fisheries Ecosystems Research
Marine Institute of Memorial University of Newfoundland
St. John's, NL. Canada

Executive Summary

The science reviewed is the best scientific information available at present. The base model stock assessment indicated that Sandbar Shark in 2015 was overfished ($SSF_{2015}/SSF_{MSY} = 0.60$) but that overfishing was not occurring ($F_{2015}/F_{MSY} = 0.75$), with $Prob(F_{2015}<F_{MSY})=0.97$. However, sufficient evidence was not provided by SEDAR 54 for me to conclude that these status evaluations are reliable.

Background

The Southeast Data, Assessment, and Review (SEDAR) 54 was a compilation of data, a standard assessment of the stock, and a CIE assessment review conducted for Highly Migratory Species (HMS) sandbar sharks. This desk review provided an independent peer review of the sandbar stock assessment. The review was responsible for ensuring that the best possible assessment is provided through the SEDAR process and to provide guidance to the Southeast Fisheries Science Center to aid in their review and determination of best available science, and to HMS when determining if the assessment is useful for management.

CIE reviewers were tasked with conducting impartial and independent peer reviews in accordance with the SOW and Schedule of Milestones and Deliverables herein. The reviewers were required to have working knowledge and recent experience in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the primary task of providing peer-review advice in compliance with the workshop ToRs. Experience with elasmobranchs assessment methods was preferred. Each CIE reviewer's duties could not exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

The review was coordinated by Dr. Julie Neer, SEDAR Coordinator for the SEDAR 54 HMS Sandbar Shark assessment process.

Role of reviewer

I reviewed background materials and reports provided by Dr. Neer on October 18, 2017. These reports were made available via a website for the SEDAR 54 review. I also reviewed the previous sandbar shark assessment (SEDAR 21) which was provided at http://sedarweb.org/sedar-21-final-stock-assessment-report-hms-sandbar-shark. I reviewed the Final Stock Assessment Report for SEDAR 54 HMS Sandbar Shark which was provided October 23, 2017. These documents are listed in Appendix 1.

This report is structured according to my interpretation of the required format and content described in Annex 1 of Appendix 2.

Summary of findings

ToR 1. Prepare a Peer Review Report that summarizes the Reviewer's evaluation of the stock assessment and addresses each of the following Terms of Reference.

This report documents my evaluation of the stock assessment. The Final Stock Assessment Report for SEDAR 54 HMS Sandbar Shark that I was provided included descriptions of:

- List of data workshop working papers and reference documents,
- Statement addressing each term of Term of Reference,
- Data review and update,
- Replication and continuity data sets,
- New data sources considered (for new analyses): life history, catch, and indices of abundance,
- Stock assessment model and results for the replication analysis and the base case model and sensitivity analyses methods and results,
- Assessment research recommendations,
- Discussion.

ToR 2. Evaluate the data used in the assessment, addressing the following:

a) Are data decisions made by the assessment panel sound and robust?
b) Are data uncertainties acknowledged, reported, and within normal or expected levels?
c) Are data applied properly within the assessment model?
d) Are input data series reliable and sufficient to support the assessment approach and findings?

Landings of sandbar sharks were reconstructed back to 1960, and the assessment assumes negligible catches before 1960 and equilibrium unfished biomass in 1960. It is meritorious to have such a long time-series of landings, but the uncertainty in the landings has not been quantified which is a common problem in fish stock assessment. This is presumably more of a problem for the historic data. For example, the split of commercial landings into F1 and F2 components prior to 1991 was based on the average percent composition by region for the first five years with more reliable data (1991-1995). Hence, these early landings have additional uncertainty.

I was surprised by the research recommendation: "Continue work on reconstruction of historical catches, especially catches outside of the US EEZ". I did not find any discussion of the potential magnitude of catches outside the US EEZ so I am unsure how much of an assessment problem this may be.

However, there are important uncertainties with the recent catch data. The Mexican fishery catches were extended to 2015 using Conapesca fisheries statistics available online for 2009-

2013. Mexican catches for 2014 and 2015 were assumed equal to the mean of those in 2011-2013). This information from the Mexican fishery statistics should have been provided so I could assess the efficacy of the imputation procedure for unknown Mexican catches in 2014-15, and also to assess potentially how large (as a percentage of the total) the unknown 2014-15 Mexican catches could be. The F3 catches (Recreational+Mexican) are a substantial fraction of the total catch (i.e. ~47% in 2015) and the values of the Mexican catch in 2014 and 2015 could potentially have substantial impact on the assessment of stock status; however, there was insufficient information provided in the assessment documents for me to evaluate the potential impact on the assessment results.

For these reasons, I cannot conclude that the estimated catch of sandbar sharks are reliable and robust. Uncertainty in the size of commercial catches is a common problem in fish stock assessments. I recommend future assessments should use better methods to account for unknown catches and the associated uncertainty of the catches. For example, the censored catch approach (e.g., Van Beveren et al., 2017, Cadigan, 2015) could be used.

Van Beveren, E., Duplisea, D., Castonguay, M., Doniol-Valcroze, T., Plourde, S. and Cadigan, N., 2017. How catch underreporting can bias stock assessment of and advice for northwest Atlantic mackerel and a possible resolution using censored catch. Fisheries Research, 194, pp.146-154.

Cadigan, N.G., 2015. A state-space stock assessment model for northern cod, including under-reported catches and variable natural mortality rates. Canadian Journal of Fisheries and Aquatic Sciences.

New estimates of natural mortality (M) were produced for the SEDAR 54 assessment with the same indirect estimators used in SEDAR 21, but using updated life history estimates. These calculations should be documented so I could check for potential errors. There are some curious differences between M's presented in the SEDAR 21 and SEDAR 54 SAR's. For example, M's at ages 7 and 8 in SEDAR 21 were 0.154 and 0.153, while in the SEDAR 54 SAR they were 0.117 for both ages. This is a difference of 30% of the SEDAR 21 values. Differences at other ages were smaller than 30%, but they were still substantial and often greater than the plus/minus 10% range used in the alternative states of nature scenarios in SEDAR 54. Therefore, I cannot conclude that the M values used are reliable and sufficient to support the assessment approach and findings, nor has the stock assessment adequately evaluated the impacts of uncertainty in M. However, the assessment has demonstrated it is reasonably robust to plus/minus 10% changes in M, but I am uncertain if the range of alternative M's is large enough given the differences of M's estimated in SEDAR 21 and 54. A better approach would be to assign informative multivariate priors on M at age that cover a reasonable range (e.g., plus/minus 30% with 95% probability); however, a technical challenge will be to incorporate the high correlation that likely exists between M's at adjacent ages.

I am somewhat puzzled by the new (i.e. updated) growth data described in SEDAR54-WP-01. The Linf and k estimated from the updated data were not contained in the confidence intervals obtained from the updated + Hale and Baremore data, not for males, females, or combined sexes. The updated sample size was greater than the Hale and Baremore sample size. The rationale provided was that there was a lack of data at small sizes in the updated data. However, it seems that there are enough data at older ages for which growth is near Linf to estimate Linf really well, and I am puzzled by the differences in estimates of Linf. The two data series should have been shown using different plotting symbols (shape or color) in Figures 1 and 2 in SEDAR54-WP-01. These plots are not very helpful in diagnosing the adequacy of the Von Bertalanffy model for describing length at age. It would be useful to show the data means at each age in Figures 1 and 2, or even better a monotone increasing smoother of age (i.e. scam() in R). Nonetheless, the fits in these figures seem reasonable, and I conclude that the growth model used in the assessment is reliable and sufficient to support the assessment approach and findings.

Growth curves were estimated separately by sex. There seemed to be some evidence in the data of differences in growth rates. However, it is useful to evaluate the statistical significance of the differences in growth rates.

The proportions mature at age in the SEDAR 54 SAR were the same as in the SEDAR 21 SAR. However, updated estimates of proportion mature were provided in SEDAR54-WP-01, but these did not seem to be considered in the SEDAR 54 SAR. The estimated logistic maturity models in SEDAR54-WP-01 were length-based, and I used the estimated Von Bertalanffy growth models in SEDAR54-WP-01 and their maturity model parameter estimates to derive age-based maturity schedules by sex (see Fig. 1). There are substantial differences with the maturity ogive used in the assessment. For example, the A50 from the updated maturity model for females was 11.5 whereas it was 13 in the SEDAR 54 SAR. These results are presented for illustrative purposes only, and ideally one would want to account for the distribution of size at age when converting maturity at length to maturity at age. However, the potential seems to exist that the new maturity information presented in SEDAR54-WP-01 could warrant a change in the ogive used in the stock assessment. Rationale for not updating the maturity ogive should have been presented. Hence, I cannot conclude that this data decision made by the assessment panel is sound and robust.

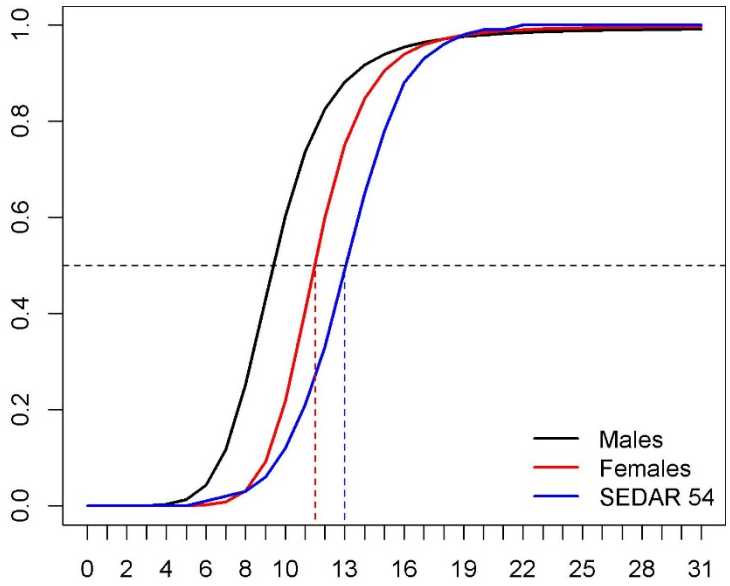

Figure 1. Maturity ogives by sex: (1) from the length-based logistic models as composite functions of the age-based Von Bertalanffy growth functions in SEDAR54-WP-01, and (2) from the SEDAR 54 SAR.

The SEDAR 21 SAR also presented estimates of proportion of maternal females. This information was not presented in SEDAR 54, and I assume not used in the stock assessment model. I was unclear how this information was used in the SEDAR 21 ASPM stock assessment model.

A variety of stock size indices were developed and used in this assessment:

Fisheries dependent

1. Large Pelagic Survey (S1_LPS) 1986-2015.
2. Bottom Longline Observer Program 1 (S2_BLLOP_1) 1994-2007.
3. Bottom Longline Observer Program 2 (S3_BLLOP_2) 2008-2015.
4. Southeast Pelagic longline observer program (S8_PLLOP) 1992-2015.

Fisheries independent

1. Virginia Longline Survey (S4_VA_LL) 1975, 1977, 1980, 1981, 1990-1993, 1995-2015.
2. NMFS Southeast Bottom Longline (S5_NMFS_LLSE) 1995-1997, 1999-2015.
3. Coastspan NE LL Survey (S6_CST_NE_LL) 2001-2015.
4. NMFS Longline Northeast Survey (S7_NMFS_NE) 1996, 1998, 2001, 2004, 2007, 2009, 2012, 2015.
5. Coastspan SE LL Survey (S9_COASTSPAN_SE_LL) 2000-2015.
6. South Carolina DNR red drum observer program (S10_SCDNR_RedDr) 1998-2006.
7. SEAMAP Longline SE Survey (S11_SEAMAP_LL_SE) 2007-2015.

Index comparison plots did not show much consistency between indices. There seems to be substantial and high-frequency sources of variation in the various indices that makes the interpretation about stock status difficult.

Many of the indices were based on fishing with longlines. An issue for longline surveys that I did not see addressed in the indices working papers (S54_WP_2-5) or the assessment document is hook saturation. It would be useful if some summary statistics were presented related to this: statistics on percent of hooks with no catch, and percent of hooks with no catch and no bait. The statistics could be percentiles by year and geographic region, and the percentiles could include 50, 25 and 10. Maybe hook saturation has been examined and found to not be an issue, and if so the documentation should indicate this.

The indices were standardized using delta-GLM models. There was insufficient information provided to review the index standardizations in detail. A different model selection method was used in WP-02 and WP-04 compared to WP-01, WP-03, and WP-05. I am not sure if this matters, but this could be addressed in a future methods review.

Effort was treated differently for the catch rate analysis of the Large Pelagic Survey (S1_LPS) compared to the Longline Observer Program. I prefer the approach in WP-03 for S1_LPS (log effort as an offset) because it naturally brings into question the appropriateness of the effort measurement. However, this is another technical issue that should be addressed in a future methods review. WP-03 indicates effort was only used in the positive observation (Poisson) submodel, and it was not clear to me why effort was not used in the presence/absence submodel. It seems likely that effort would affect whether at least one shark was caught.

The reliability of species identification is a concern that should somehow be addressed in future analyses. I was unsure of the potential magnitude of this problem.

The Large Pelagic Survey (S1_LPS) CPUE standardization analyses seemed to follow a simple update approach, without re-analysis of factors to standardize for in model selection. This may not be appropriate for a full assessment.

The S9 COASTSPAN SE indices were removed for 1988 and 1999 because they were uncertain. The report should give some indication why these indices were considered too uncertain to include. CVs were given in Table 2.6 of the SEDAR 54 SAR and, although they are high, there are many indices with higher values that were retained in the assessment model.

The hierarchical cluster analysis and cross-correlations of CPUE Indices described in SEDAR54-WP-06 was useful and provided a basis for formulating a major axis of assessment model uncertainty involving the different trends indicated by some indices. Robustness of cluster analyses is always a concern. To check for robustness, I replicated the correlation and cluster analyses and I also used Spearman's rho which is a more robust rank-based measure of association. The CPUE series are shown in Fig. 2. The correlation matrix in Fig. 3 is identical to

the one presented in SEDAR 54-WP-06; the only difference is that missing values are indicated by a ? symbol in Fig. 3 whereas they were indicated as zero's in SEDAR 54-WP-06. A cluster analysis dendrogram obtained using the hclust() function in R and with default settings (Fig. 4) demonstrates the two main cluster of CPUE indices that formed the basis of a SEDAR 54 sensitivity analysis. Note that for this analysis, missing correlations were replaced with zeros, although other approaches are possible. The left-hand cluster (S3, S4, S5, and S7) indicates the "POS_1" CPUE group; however, S2 was added to this group because the Assessment Panel felt that it helped extend the available time series to the period where the majority of the fishing effort occurred. Figures 5 and 6 illustrate the results from the same analysis but using Spearman's rank ordering correlations, which provide a more robust measure of association. The correlations in Figures 3 and 5 are reassuringly similar. However, the cluster analysis results in Fig. 6 suggest a different grouping of CPUE indices: group 1 – S1-S7, and group 2 S8-S11. This grouping may have provided a better basis for the sensitivity analyses.

I repeated the analyses using log CPUE indices. The Pearson results changed little, but the Spearman correlations were identical, as expected. This is because Spearman's correlations are invariant to one-to-one transformations of the indices.

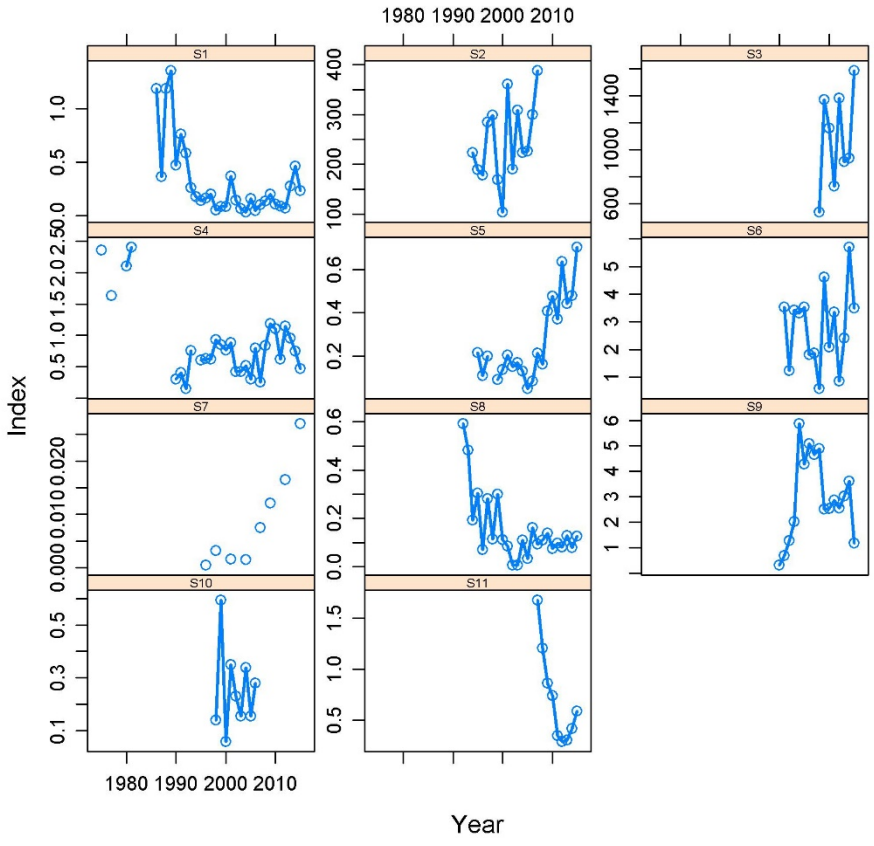

Figure 2. Stock size indices used in the SEDAR 54 stock assessment.

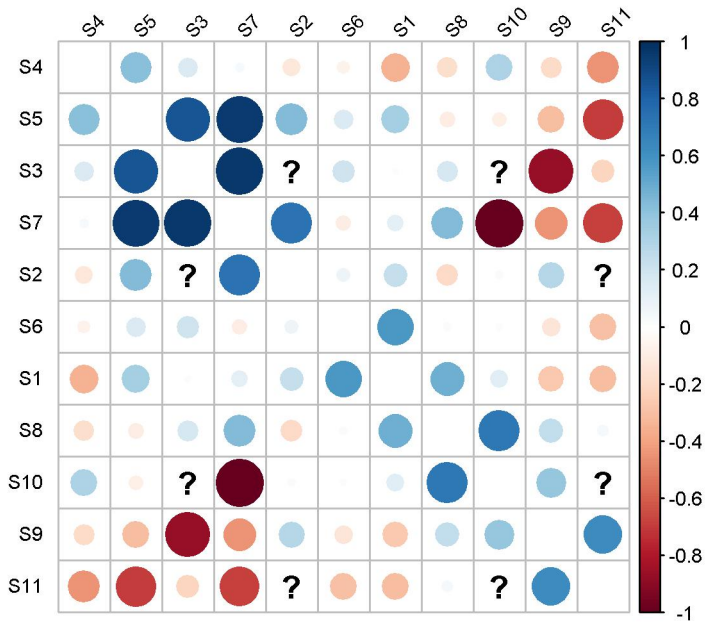

Figure 3. Pearson's correlation matrix for SEDAR 54 CPUE indices. Blue and red indicates positive and negative correlations, respectively. The ? indicates missing values due to non-overlapping series. The order of the indices is the same as in the SEDAR 54 correlation analyses.

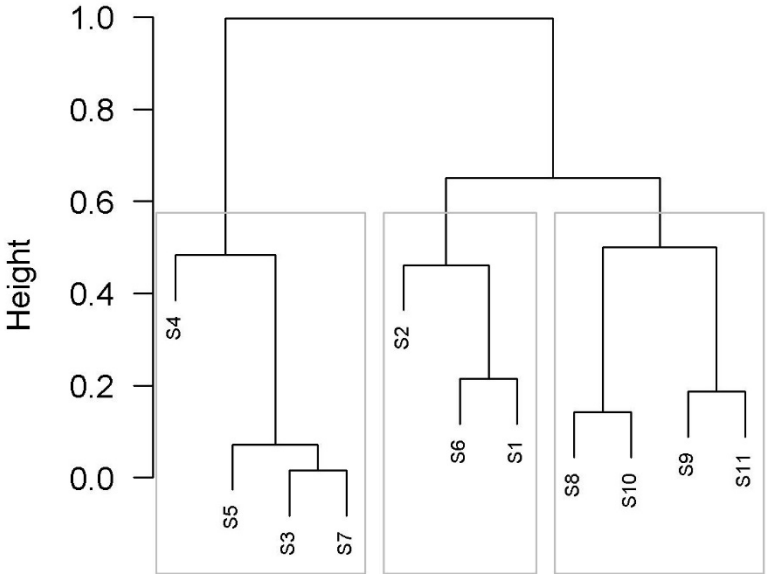

Figure 4. Cluster dendrogram based on the complete agglomeration method and the correlations indicated in Fig. 3. The three main clusters (1-left to 3-right) are indicated by the grey rectangles.

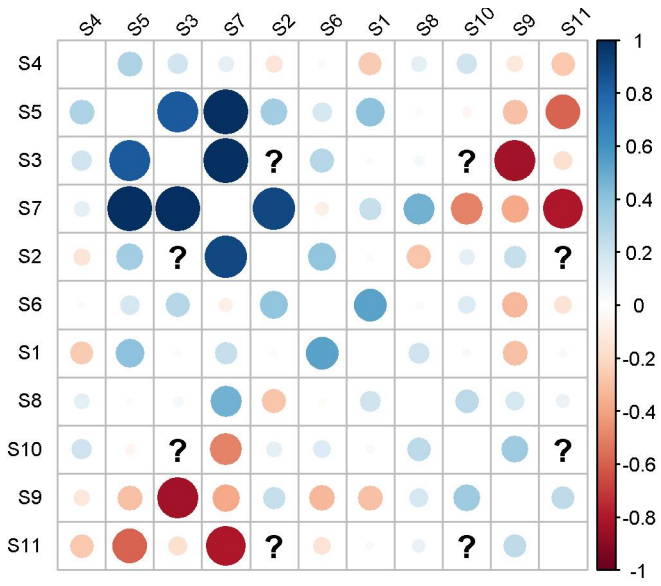

Figure 5. Spearman's correlation matrix for SEDAR 54 CPUE indices. Blue and red indicates positive and negative correlations, respectively. The ? indicates missing values due to non-overlapping series. The order of the indices is the same as in the SEDAR 54 correlation analyses.

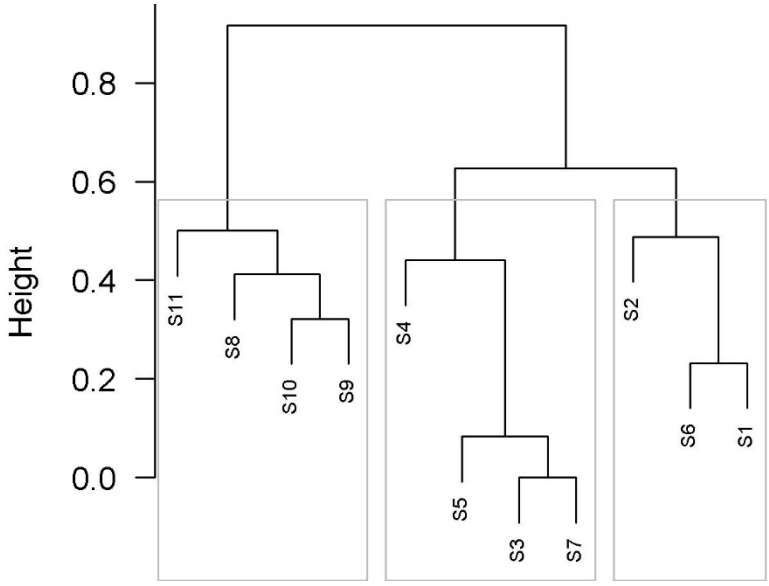

Figure 6. Cluster dendrogram based on the complete agglomeration method and the correlations indicated in Fig. 5. The three main clusters (1-left to 3-right) are indicated by the grey rectangles.

The assessment was a length-based age-structured statistical model. The SEDAR 54 SAR did not provide a description of the model, but did supply some references that were not provided to me as part of the CIE review package. I assume the model was age-based, but fit to length compositions by converting predicted fishery or survey catch at age to catch at length using a stochastic population growth curve. The SAR did indicate "The standard assumptions made concerning age and growth in the Stock Synthesis model are (i) the lengths-at-age are assumed to

be normally distributed for each age-class". In this case the variability of length at age is important; however, how this was quantified was not described in the SAR, and I could not determine this from the SS3 code presented in the Appendix. If these variances were estimated from growth curves then this should be described. If these variances were subjectively chosen, then this should also be described and robustness to the choice of values should be examined. I find that this aspect of data uncertainties was not acknowledged or reported.

Aggregate (over years) length compositions for various fishery and index fleets were presented, but no information was given about annual compositions. I gather that inter-annual variability was substantial.

There is a substantial amount of tagging data available for sandbar sharks and I was surprised that this data was not utilized more in the stock assessment. SEDAR21-DW-38 indicated that during 1962 through 2009 almost 31,000 Sandbar sharks were tagged. I appreciate concerns about "possibility of underreporting of recaptures" but I feel that more could be done with this data. A table of releases by year, and subsequent recaptures by year-at-liberty would be informative. Underreporting of recaptures is common. Even if the reporting rates are unknown, the rate of decline of recaptures as time-at-liberty increases can still provide information on mortality rates (see Myers et al., 1996). Adding a component to the stock assessment model for tagging data can result in the ability to estimate M and tagging data can contribute to improved precision of estimates of F (e.g., see Cadigan, 2015. Reference above).

Myers, R.A., Barrowman, N.J., Hoenig, J.M. and Qu, Z., 1996. The collapse of cod in Eastern Canada: the evidence from tagging data. ICES Journal of Marine Science, 53(3), pp.629-640.

ToR 3. Evaluate the methods used to assess the stock, taking into account the available data.

a) Are methods scientifically sound and robust?
b) Is the assessment model configured properly and used consistent with standard practices?
c) Are the methods appropriate for the available data?
d) Are differences between the current model and the previous model clearly documented and described?

The sandbar sharks were treated as a single coastwise stock. The SAR indicated that "this may be appropriate given that tagging results indicated a high amount of movement between the eastern US coast and the Gulf of Mexico"; however "there is little to no information concerning the degree of connectedness throughout the species southern range", which includes the western north Atlantic to the Gulf of Mexico, Caribbean and Brazil. The treatment of the stock seemed like a reasonable approach; however, the preservation of sub-stock structure should be a management concern, particularly in light of the importance of maintaining spawning components on overall stock productivity.

The stock assessment model is parameterized in terms of virgin biomass. This seems to create the need to reconstruct catch histories back to the year of virgin biomass. Many stock assessment models used internationally do not do this, and it would be useful if the assessment group could remind reviewers and others why they feel this approach is a good idea.

Fishery removals were divided among four fleets which seemed reasonable, although I was unsure why Mexican catches were combined with recreational catches.

The assessment model was sex-specific (i.e. assumed a 1:1 sex ratio) and fit to sex-specific data where available. Separate growth curves were estimated for females and males. The data suggest growth is somewhat sexually dimorphic and the modelling approach is appropriate.

The assessment was age-based, with a plus group at age 31 which seemed appropriate.

The natural mortality rate (M) was assumed to be the same for males and females but not estimated within the model, which is a good decision, because I very much doubt that the assessment modelling approach could estimate M reliably. However, if tagging data were included (see comment under ToR 2), then it is very possible that M could be estimated. Sexually dimorphic growth rates may suggest that M's will differ by age for males and females; however, the differences in size at age for the two sexes seem small enough that the assumption of the same M seems reasonable.

Recruitment dynamics modelled within SS3 were assumed to be governed by a Beverton-Holt stock-recruit function with a fixed steepness parameter. Spawning output in the stock-recruitment relationship was modelled as spawning stock fecundity (sum of female numbers at age times annual female pup production at age). This was appropriate.

Annual recruitment deviates from the recruitment relationship were estimated, but somehow constrained to reflect the limited scope for compensation given the estimates of fecundity. I was not sure what the constraint was. Variability in recruitment is expected to be low given the reproductive biology of this species. Although I am not a shark expert, I feel that it is reasonable to expect an almost linear relationship between recruitment and parental stock size for the range of stock sizes estimated by the assessment model. Hence, I conclude that the form of the recruitment model was appropriate in this assessment.

The value of stock-recruit steepness was fixed at h=0.3, which was changed slightly from the 0.29 value used in SEDAR 21. I anticipate steepness is low for this species.

Selectivity is fishery and index specific and was assumed to be time-invariant. A double normal and possibly domed-shaped functional form was assumed for the fishery selectivity curves F1 and F3, and logistic with asymptotic selectivity was used for F2 and F4. Selectivity curves for the CPUE series were almost all double normal, and none were fixed to be asymptotic. The selectivity was fixed (not estimated) for the CPUE series S2 and S3 as these CPUE series share

the length compositions with the fisheries F1 and F2, so as to not use the same data in the estimation phase more than once. I was unsure how these selectivity curves were fixed. My approach to stock assessment selectivity is to assume it is asymptotic unless there are good reasons to expect otherwise. Hence, I agree with assuming F2 selectivity is asymptotic; however, I would prefer to see arguments presented why other fisheries and CPUE indices could possibly have dome-shaped selectivity curves that are appropriately modelled using a double normal functional form. This issue is considered further under ToR 4.

Annual length composition sample sizes were weighted in the SS3 integrated likelihood using the Francis (2011) method. I unfortunately do not understand the details of this method. From what I understand, the weighting involves changing the input sample sizes in the multinomial likelihoods used by SS3 to fit compositional data. In this assessment the input sample sizes were taken to be the number of sets sampled for length compositions. This seemed reasonable.

The SAR indicated that weightings (i.e. input CVs) for relative abundance indices were rescaled so that the mean CV (over years) was the same as the CV for an independently fit smooth line. However, the SAR indicated in another section that SS3 estimated "the minimum average CV associated with the indices of abundance". This seems different than the first approach described. This should be clarified.

Replication analysis. I appreciate that it is difficult to exactly mimic the ASPM assessment structure using SS3. There are many differences, some subtle, between the two models. However, I conclude that the major differences between the current SS3 model and the previous ASPM model were clearly documented and described. The comparison of biomass trends from the replication analysis are very similar for the post-1985 period in which there are substantial stock size indices available. Historic differences are a concern and indicate to me the difficulty in estimating virgin biomass.

ToR 4. Evaluate the assessment findings with respect to the following:

a) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
b) Is the stock overfished? What information helps you reach this conclusion?
c) Is the stock undergoing overfishing? What information helps you reach this conclusion?
d) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
e) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?
f) Are base model runs, sensitivity runs, and alternate states of nature runs clearly described and reasonable?

The basic way I evaluate ToR4a is to assess the SEDAR 54 SS3 model fits to tuning indices and length compositions. Also, one can examine if model estimates, and in particular estimated fishery and CPUE index selectivity curves, are consistent with other knowledge on the gears used and locations fished relative to the spatial distribution of the stock and how this may change with the age and size of the fish. Retrospective analyses are another way to examine the reliability of model estimates. Examination of convergence diagnostics is also important and was part of the ToRs for the assessment process.

Fits to survey indices were sometimes poor with substantial and temporally systematic differences between indices and model predictions. However, the indices were not really consistent with each other anyway, and had short-term trends that seem to not have indicated trends in the sandbar shark stock. Hence, it seems reasonable that the assessment model could not fit these indices well. Nonetheless, it is difficult to conclude from fits to the CPUE indices that model estimates are reliable.

Fits to the various aggregate (over years) length compositions were OK, and did not provide strong evidence of model mis-specification. However, the estimated selectivity curves from the base case model were not provided, so I cannot comment on the plausibility of this aspect of the estimated assessment model. It is very unusual not to provide this information, and this is an important criticism of the SEDAR 54 SAR.

I found some of the length composition sample size multipliers in Table 3.2.2 of the SAR surprising – especially some of the values that were substantially greater than one. I realize that this is a difficult and somewhat controversial area in statistical stock assessment model fitting, and an area of continuing research (e.g., Francis, 2016). I feel that the assessment should have investigated if the results were sensitive to alternative and reasonable weighting of the length compositions.

Francis, R.C., 2017. Revisiting data weighting in fisheries stock assessment models. Fisheries Research, 192, pp.5-15.

The SEDAR 54 SAR did not provide results on the convergence of the SS3 (i.e. maximum absolute gradients, jitter analyses, phase alternation runs). Some of the estimated parameter values in Table 3.2.3 are surprisingly close to their initial values (e.g., SizeSel_2P_1_F2_COM_SA 93.63 vs 94.68; SizeSel_2P_2_F2_COM_SA 29.72 vs 31.03). Maybe this is because the initial values were derived from preliminary models runs and therefore close to their maximum likelihood values, but it also could represent poor convergence and a flat likelihood surface for these parameters. This is something that would be checked in a full review panel meeting. Hence, I cannot conclude that the SS3 model solution was reliable.

The base model had little retrospective pattern in stock depletion or in the size of virgin recruitment. This is good, although absence of a retrospective pattern does not necessarily

indicate reliable estimation. It would have been useful to provide retrospective estimates of Fcurrent/Fmsy and SSFcurrent/SSFmsy.

For the above reasons, I conclude that sufficient evidence has not been presented to support ToRa.

The base case model indicated that the stock is overfished (P(SSB2015<SSBMSY)=0.99) but there was no overfishing in 2015 (P(F2015<FMSY)=0.97). The overfished conclusion was different in all sensitivity analyses based on the POS CPUE set of indices, while the overfishing conclusion was different (i.e. overfishing was occurring) in the sensitivity analysis based on the 'NEG' set of CPUE indices. Some discussion of the potential reliability of various indices was presented, but it did not help me conclude that one of the "POS" or "NEG" sensitivity analyses were more reliable, or that the base model run was more reliable. Hence, I conclude that sufficient evidence has not been presented to conclude whether the stock is overfished or if overfishing occurred in 2015. I conclude that the quantitative estimates of the status determination criteria for this stock are not reliable. The potential exists that an assessment model that utilizes additional tagging data for this stock could produce more reliable status determinations.

The stock-recruit steepness parameter was fixed at 0.3, which was very close to the value used in SEDAR 21 (i.e. h=0.29). The new value was based on a recalculation of the parameter based on the updated life history inputs. This provides some evidence that the steepness value is reliable. For example, had the value changed substantially after recalculation then I would conclude it was unreliable. Hence, the assessment uses an informative stock recruitment relationship. I conclude that the stock recruitment curve is reliable and useful for evaluation of productivity and future stock conditions as long as projected SSF does not greatly exceed the bounds of estimated historic values.

I find that the base model runs, sensitivity runs, and alternate states of nature runs were clearly described and reasonable. However, more model outputs should have been provided – in particular, the estimated selectivity curves for the fishing fleets and abundance indices. As usual, additional sensitivity runs would have been informative (see ToR 6).

ToR 5. Evaluate the stock projections, rebuilding timeframes, and generation times, addressing the following:

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

e) If the results indicate a new rebuilding schedule is required, are the scientific/technical reasons for the new schedule clearly articulated and appropriate?

Projections were carried out using the forecast module internal to SS3, and therefore I find that the SEDAR 54 SAR fully satisfied ToRa,b.

The base model projection results were informative. They were used to determine elements of the current rebuilding plan; for example, the constant TAC strategies that would allow stock rebuilding by 2070 with a 50% and 70% probability, respectively, were 208 and 148 mt (whole weight). Since the assessment model results were not robust to the POS/NEG CPUE grouping analyses, I suspect the projections were not robust in this aspect as well; however, results comparing projection results from the base case and sensitivity runs were not presented. Hence, I conclude that sufficient evidence was not presented to demonstrate that stock projections were useful to support inferences of probable future conditions.

Only uncertainty associated with the parameter estimates calculated internally to SS3 were included in the projections. Recruitment variability was not included; however, the SAR indicated that "given the reproductive biology of this species, variability in recruitment is expected to be low". I accept this argument. Projections were only carried out for the base case productivity assumptions, because the influence of the high and low productivity scenarios had minimal effect on stock status in comparison to the CPUE groupings. However, the high and low productivity scenarios only involved a plus-minus 10% change in M, and for reasons outlined under ToR2, I feel that a larger change in M should have been considered. It is common in stock assessment to consider a range for the steepness parameter in sensitivity analyses. SEDAR 54 did not, although it did consider a range in the reproductive cycle (i.e. 2.5 years plus/minus 0.5 years). I suggest the assessment should have considered a range of steepness values for the sensitivity analyses, or at least presented rationale why this was not necessary. The assessment should investigate if results are sensitive to alternative and reasonable weighting of the length compositions.

The determination of stock status in SEDAR 54 was the same as in SEDAR 21 (i.e. overfished, but not experiencing overfishing); therefore, no new rebuilding schedule was warranted. However, in one sensitivity analysis (i.e. NEG CPUE) the stock was estimated to be overfished and a new rebuilding schedule had to be calculated.

ToR 6. 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
b) Ensure that the implications of uncertainty in technical conclusions are clearly stated.

The Hessian matrix computed at the mode of the posterior distribution was used to obtain estimates of the covariance matrix of parameter estimators. Variances of derived parameter estimators were then obtained using the delta method. Markov Chain Monte Carlo (MCMC) was used to carry the uncertainty in the parameter estimates forward to projection results. These methods are provided by the SS3 package and seem appropriate. Sensitivity analyses were also used to indicate the potential for additional uncertainty; however, this was not formally quantified using model averaging or some other procedure. Research on methods to do this is ongoing, and I appreciate why this was not done in SEDAR 54.

Some additional sensitivity analyses should have been provided. Low and high catch scenarios were considered in SEDAR 21, and the catch range was wide. It is common in stock assessment to do this. This was not considered in SEDAR 54 and no rationale was provided for this. However, I suspect that investigating catch ranges does not fully address uncertainty in catch – there are other ways to consider this issue (e.g., Van Beveren, 2017; reference under ToR 2). The assessment should have examined a range of steepness values that are plausible for this species (although see comment below for a better approach). A sensitivity analysis where the F1 and F3 fleets are also given asymptotic selectivity should be conducted, to examine if there are good reasons (i.e. data signals) not to do this. A robust cluster analysis indicated CPUE indices should be grouped S1-S7 and S8-S11. This grouping may have provided a better basis for the sensitivity analyses. I was unsure how variability in size at age was included in the assessment, or if the assessment was sensitive to this. This could be addressed in a sensitivity analysis.

Some potentially important sources of uncertainty were not accounted for in the assessment, such as the value of M and steepness. A better approach may be to assign informative priors (derived from life-history experts) to these parameters. If the data and assessment model are uninformative about these parameters, then the prior variances will be directly propagated to variances for stock assessment results. Similarly, there is unaccounted uncertainty in catches (see above comment).

Uncertainty in technical conclusions were clearly stated.

ToR 7. 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.
b) Provide recommendations on possible ways to improve the SEDAR process

I have provided suggestions for improvements under the previous ToRs, and summarized these below.

Documentation

1. Provide figures of estimated selectivity curves.
2. Clarify the potential magnitude of catches outside the US EEZ.
3. Provide more information about the Mexican fishery catch statistics.
4. Document the details of M estimation.
5. Provide residual plots for fitted growth curves, or show the data means at each age or, even better, a monotone increasing smoother of age.
6. Describe how variability in size at age was included in the assessment.
7. Provide improved rationale why the S9 COASTSPAN SE indices were removed for 1988 and 1999.

Input data and analytical methods

Short-term

1. Provide tables of tagging data: Number of releases by year, and subsequent recaptures by year-at-liberty.
2. Quantify uncertainty in the landings.
3. Evaluate the statistical significance of the differences in growth rates.
4. Evaluate via a sensitivity analysis if the assessment is sensitive to assumed values for variability in size at age.
5. Update the maturity ogives.
6. Provide some summary information about the potential for saturation effects to occur in longline surveys.
7. Consider using Spearman's rho in the correlation/cluster analysis. This approach is invariant to one-to-one transformations of the indices, which is a good thing.
8. Consider a priori if fisheries and CPUE indices could possibly have dome-shaped selectivity curves.
9. Provide retrospective estimates of Fcurrent/Fmsy and SSFcurrent/SSFmsy.

Long-term

1. Investigate better utilization of the substantial tagging data for the stock assessment of sandbar sharks.
2. Investigate assessment models that do not try to estimate virgin biomass.
3. Investigate better methods to account for unknown fishery catches and the associated uncertainty of the catches.
4. Use informative multivariate priors on M at age that cover a reasonable range (e.g., plus/minus 30% with 95% probability); however, a technical challenge will be to incorporate the high correlation that likely exists between M's at adjacent ages.
5. Include a prior for steepness to better account for uncertainty in this parameter.

6. CPUE index standardization procedures may not be working that well in removing nuisance effects from catch rates. It also seems possible that some of the stock signal is being removed from the data in the standardization. A variety of model selection procedures were used for SEDAR 54. Effort is treated differently in two of the CPUE models. These technical issues should be addressed in a future methods review meeting/process.
7. Evaluate reliability of species identification.
8. Investigate if there is evidence of within season depletion in local fishery CPUE that could be used to indicate local stock sizes.

Possible ways to improve the SEDAR process

The SEDAR 54 SAR did not address some elements of the CIE review ToRs. I wonder if these ToRs were available to the assessment team. The ToRs should be available and an emphasis should be placed on documenting results that reviewers will need to evaluate the review ToRs. This could go as far as to ask the assessment team to "self-review" their assessment, and then the CIE reviewers could focus on whether they agree with the "self-review" and to explain differences in review conclusions. In this situation it would be less likely that the assessment team would omit results by mistake, or that CIE reviewers would mis-interpret or not notice some important assessment results. However, I appreciate that this would add additional workload to assessment staff who are already very busy.

The desk-review process requires that all review ToRs be evaluated from the documented assessment results. I know from experience that it is difficult to document all aspects of an assessment model or to anticipate what results reviewers will want to evaluate. In a desk review there are no opportunities for reviewers to ask for clarifications or additional results. This makes it more difficult to determine if assessment results (e.g., status evaluations, projections) are reliable and robust. In a desk review it is more difficult to conclude that ToRs have been satisfactorily addressed because less assessment information is available to the reviewers.

ToR 8. Provide guidance on key improvements in data or modeling approaches that could be considered when scheduling the next assessment.

I speculate that key improvements will involve better modelling of CPUE indices and incorporation of tagging data into the stock assessment model. I agree with all of the SEDAR 54 SAR research recommendations, and especially the one involving investigations of the "distribution and movements of the stock relative to sampling coverage". It may be that a spatial model is required to reconcile difference in some CPUE indices. Of course tagging data will be very useful in a spatial model.

Conclusions and Recommendations

Recommendations are provided under ToR 7. Conclusions are listed below for each ToR 2-7.

ToR 2. Evaluate the data used in the assessment, addressing the following:

a) Are data decisions made by the assessment panel sound and robust?
b) Are data uncertainties acknowledged, reported, and within normal or expected levels?
c) Are data applied properly within the assessment model?
d) Are input data series reliable and sufficient to support the assessment approach and findings?

Uncertainty in fisheries landings has not been quantified. There could be catches outside the US EEZ but no information on their potential magnitude was provided. Insufficient information was provided in the assessment document for me to evaluate the potential impact of unknown Mexican catches in 2014-15 on the assessment results. For these reasons, I cannot conclude that the estimated catches of sandbar sharks are reliable and robust.

New estimates of natural mortality (M) were produced for SEDAR 54 and they differed as much as 30% of the SEDAR 21 values. I cannot conclude that the M values used are reliable and sufficient to support the assessment approach and findings, nor has the stock assessment adequately evaluated the impacts of uncertainty in M.

I conclude that the growth model used in the assessment is reliable and sufficient to support the assessment approach and findings.

Maturity ogives were not updated with recent data and this decision made by the assessment panel may not be sound and robust.

The procedures used to derive CPUE indices seemed sound overall. These procedures also seemed robust in that updated indices were reasonably similar with those calculated for SEDAR 21.

The hierarchical cluster analysis and cross-correlations of CPUE indices was a good and sound idea. This analysis was based on Pearson's correlation coefficient which is known to not be robust to outliers. However, I replicated the analyses with a robust correlation measure and the main results did not change much.

ToR 3. Evaluate the methods used to assess the stock, taking into account the available data.

a) Are methods scientifically sound and robust?
b) Is the assessment model configured properly and used consistent with standard practices?
c) Are the methods appropriate for the available data?

d) Are differences between the current model and the previous model clearly documented and described?

The assessment model was implemented in Stock Synthesis version 3.24f. The model is used widely in some jurisdictions and has been tested extensively. Hence I conclude that the software used was scientifically sound and robust.

In terms of model configuration, the sandbar sharks were treated as a single coastwise stock, which seemed appropriate. The assessment model was sex-specific which was appropriate because growth is somewhat sexually dimorphic. The assessment was age-based, with a plus group at age 31 which seemed appropriate. The way recruitment dynamics were modelled was appropriate for the species and the data available. The choice of the steepness parameter values seemed appropriate, although methods to account for uncertainty in this value were not appropriate.

Insufficient information was presented for me to make conclusions about whether the way fishery and CPUE selectivity curves were modelled was appropriate or not. I would prefer to see arguments presented why fishery and CPUE indices could possibly have dome-shaped selectivity curves.

The assessment was a length-based age-structured statistical model, which is appropriate for the available data. However, insufficient information was presented for me to make conclusions of the appropriateness of the stochastic growth model to infer the distribution of length at age.

Sandbar sharks are data poor in some aspects. However, available tagging data that is possibly relevant to understanding mortality rates was not used.

Major differences between the current SS3 model and the previous ASPM model were clearly documented and described.

ToR 4. Evaluate the assessment findings with respect to the following:

a) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
b) Is the stock overfished? What information helps you reach this conclusion?
c) Is the stock undergoing overfishing? What information helps you reach this conclusion?
d) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
e) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?
f) Are base model runs, sensitivity runs, and alternate states of nature runs clearly described and reasonable?

Sufficient evidence was not provided in the SEDAR 54 SAR for me to conclude that status evaluations are reliable. Hence,

a) I cannot conclude that abundance, exploitation, and biomass estimates are reliable and useful to support status inferences;
b) I cannot conclude the stock was overfished;
c) I cannot conclude that overfishing is occurring; and
d) I cannot conclude that the quantitative estimates of the status determination criteria for this stock are reliable.

It is important to note that I do not know that the results are unreliable, or that the stock is not overfished, etc. I am simply concluding that the reliability of assessment estimates and the overfished and overfishing status are unknown. The status determinations change depending on what CPUE indices are used to fit the model, and I have no basis to decide which indices are more informative. I also do not conclude that the base model formulation, which fits to the combined set of indices (i.e. finding the compromise fit), results in a more reliable assessment because adding unreliable indices does not improve the reliability of an assessment.

I conclude that the assessment uses an informative stock recruitment relationship that is reliable and useful for evaluation of productivity and future stock conditions as long as projected SSF does not greatly exceed the bounds of estimated historic values.

I find that the base model runs, sensitivity runs, and alternate states of nature runs were clearly described and reasonable. However, more model outputs should have been provided.

ToR 5. Evaluate the stock projections, rebuilding timeframes, and generation times, addressing the following:

a) Are the methods consistent with accepted practices and available data?
b) Are the methods appropriate for the assessment model and outputs?
c) Are the results informative and robust, and useful to support inferences of probable future conditions?
d) Are key uncertainties acknowledged, discussed, and reflected in the projection results?
e) If the results indicate a new rebuilding schedule is required, are the scientific/technical reasons for the new schedule clearly articulated and appropriate?

I conclude that the SEDAR 54 projection methods are consistent with accepted practices and available data and appropriate for the assessment model and outputs. However, projection results were not robust. Sufficient evidence was not presented to demonstrate that stock projections were useful to support inferences of probable future conditions.

Only uncertainty associated with the parameter estimates calculated internally to SS3 were included in the projections. Uncertainty related to recruitment variability (presumed to be low for this species) and M were not included in the projections.

The determination of stock status in the base model provided by SEDAR 54 was the same as in SEDAR 21 (i.e. overfished, but not experiencing overfishing); therefore, no new rebuilding schedule was warranted.

ToR 6. 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
b) Ensure that the implications of uncertainty in technical conclusions are clearly stated.

Variances of assessment estimators were obtained using the delta method and Hessian-based covariances of model parameter estimators. Markov Chain Monte Carlo (MCMC) was used to carry the uncertainty in the parameter estimates forward to projection results. These method are provided by the SS3 package and seem appropriate. Sensitivity analyses were also used to indicate the potential for additional uncertainty; however, this was not formally quantified using model averaging or some other procedure. Some potentially important sources of uncertainty were not accounted for in the assessment, such as the value of M, steepness, and catch values.

Uncertainty in technical conclusions were clearly stated.

ToR 7. 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.
b) Provide recommendations on possible ways to improve the SEDAR process.

Suggestions on improvement to assessment documentation and short- and long-term research recommendations to improve input data and analytical methods are provided above. My main recommendation is to continue the objective of implementing an integrated stock assessment model that accounts for all relevant information on the productivity of sandbar sharks. In particular, there is a potential that tagging data could provide new and useful information on mortality rates.

The CIE review ToRs should be available to assessment teams, and during the assessment process an emphasis should be placed on documenting results that reviewers will need to evaluate the review ToRs.

There were no opportunities during the CIE desk review of SEDAR 54 to ask for clarifications or additional results. This makes it more difficult to determine if assessment results (e.g., status evaluations, projections) are reliable and robust. I suggest the utility of the desk review process should be re-evaluated. The process could be improved if it were augmented with a webinar or some other process where reviewers could ask for clarifications or additional results.

Appendix 1: Bibliography of materials provided for review

SEDAR 54 HMS Sandbar Shark Document List

Document #	Title	Authors	Date Submitted
Documents Prepared for the Assessment			
SEDAR54-WP-01	Updated life history parameters for sandbar sharks, *Carcharhinus plumbeus*	William B. DriggersIII, Bryan S. Frazier, John K. Carlson, Bethany M. Deacy, Michael P. Enzenauer and Andrew N. Piercy	8 May 2017
SEDAR54-WP-02	Updated catch rates of sandbar sharks(*Carcharhinus plumbeus)* in the northwest Atlantic Ocean from the Shark Bottom Longline Observer Program, 1994-2015	John K. Carlson and Alyssa N. Mathers	3 May 2017
SEDAR54-WP-03	Standardized catch rates of sandbar sharks from the Large Pelagics Rod and ReelSurvey 1986-2015	John Walter and Craig A. Brown	7 April 2017
SEDAR54-WP-04	Sandbar Shark Abundance Indices from NMFS Bottom Longline Surveys in theNorthern Gulf of Mexico	Adam G. Pollack, David S. Hanisko and G. Walter Ingram, Jr.	19 May 2017
SEDAR54-WP-05	Standardized catch rates for sandbar sharks from the U.S. pelagic longlineobserver program using generalized linear mixed models	Enric Cortés and Xinsheng Zhang	15 May 2017
SEDAR54-WP-06	Example Implementation of aHierarchical Cluster Analysis and Cross-correlations of Selected CPUE Indices for the SEDAR 54 Assessment	Dean Courtney	20 Sept 2017
SEDAR54-WP-07			
Final Stock Assessment Reports			
SEDAR54-SAR1	HMS Sandbar Shark	SEDAR 54 Panel	

Appendix 2: CIE Statement of Work

Statement of Work
National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
Center for Independent Experts (CIE) Program
External Independent Peer Review

SEDAR 54 HMS Sandbar Shark Assessment Review

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf).

Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

Southeast Data, Assessment, and Review (SEDAR) 54 will be a compilation of data, a standard assessment of the stock, and CIE assessment review conducted for Highly Migratory Species (HMS) sandbar 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 Southeast Fisheries Science Center to aid in their review and determination of best available science, and to HMS when determining if the assessment is useful for management. The stock assessed through SEDAR 54 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

NMFS requires three reviewers to conduct an impartial and independent peer review in accordance with the Statement of Work (SOW), OMB Guidelines, and the ToRs below. The reviewers shall have working knowledge and recent experience in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the primary task of providing peer-review advice in compliance with the workshop ToRs. Experience with elasmobranches assessment methods would be preferred. Each CIE reviewer's duties shall not exceed a maximum of 10 days to complete all work tasks of the peer review described herein.

Tasks for reviewers

Each CIE reviewers shall complete the following tasks in accordance with the SOW and Schedule of Milestones and Deliverables herein.

Pre-review Background Documents: Review the following background materials and reports prior to the review:

Working Papers, Reference Documents, and the Assessment Report will be available no later than 23 October 2017. All materials will be available on the SEDAR website:

http://sedarweb.org/sedar-54-dataassessment-process

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 cannot be made during the peer review, and any SOW or ToRs modifications prior to the peer review shall be approved by the Contracting Officer's Representative (COR) and the CIE contractor.

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.

Place of Performance

Each CIE reviewer shall conduct an independent peer review as a desk review, therefore no travel is required.

Period of Performance

The period of performance shall be from the time of award through December 2017. Each reviewer's duties shall not exceed 10 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers
Within four weeks of award	Contractor provides the pre-review documents to the reviewers
October 2017	Each reviewer conducts an independent peer review as a desk review
Within two weeks after review	Contractor receives draft reports
Within two weeks of receiving draft reports	Contractor submits final reports to the Government
Within two weeks of Government receiving final reports	Government distributes final reports to Project Contact and SEDAR

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards:

(1) The reports shall be completed in accordance with the required formatting and content (2) The reports shall address each ToR as specified (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

Since this is a desk review travel is neither required nor authorized for this contract.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

Annex 1: Peer Review Report Requirements

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not 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.

3. The reviewer report shall include the following appendices:

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

Annex 2: Terms of Reference for the Peer Review

SEDAR 54 HMS Sandbar Shark Assessment Review

1. Prepare a Peer Review Report that summarizes the Reviewer's evaluation of the stock assessment and addresses each of the following Terms of Reference.

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

    a) Are data decisions made by the assessment panel sound and robust?
    b) Are data uncertainties acknowledged, reported, and within normal or expected levels?
    c) Are data applied properly within the assessment model?
    d) Are input data series reliable and sufficient to support the assessment approach and findings?

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

    a) Are methods scientifically sound and robust?
    b) Is the assessment model configured properly and used consistent with standard practices?
    c) Are the methods appropriate for the available data?
    d) Are differences between the current model and the previous model clearly documented and described?

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

a) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
b) Is the stock overfished? What information helps you reach this conclusion?
c) Is the stock undergoing overfishing? What information helps you reach this conclusion?
d) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
e) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?
f) Are base model runs, sensitivity runs, and alternate states of nature runs clearly described and reasonable?

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

a) Are the methods consistent with accepted practices and available data?
b) Are the methods appropriate for the assessment model and outputs?
c) Are the results informative and robust, and useful to support inferences of probable future conditions?
d) Are key uncertainties acknowledged, discussed, and reflected in the projection results?
e) If the results indicate a new rebuilding schedule is required, are the scientific/technical reasons for the new schedule clearly articulated and appropriate?

6. 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
b) Ensure that the implications of uncertainty in technical conclusions are clearly stated.

7. 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.
b) Provide recommendations on possible ways to improve the SEDAR process.

8. Provide guidance on key improvements in data or modeling approaches that could be considered when scheduling the next assessment.