Individual CIE Report

Atlantic red drum stock assessment review

Prepared for the Centre for Independent Experts

by

Dr. Geoff Tingley

Gingerfish Ltd, Wellington, New Zealand
email: gingerfish.ltd@gmail.com

October 2024

Table of Contents

Executive Summary

- This document is the individual CIE Reviewer Report of the review of the 2024 Atlantic red drum benchmark stock assessment. This was an in-person review conducted during August 2024. This report represents the sole views of the independent CIE reviewer, Dr Geoff Tingley.
- Documents describing the 2022 Simulation Assessment and the 2024 stock assessments of the northern and southern red drum stocks were provided together with a few relevant background documents. These materials were provided in a downloadable format posted on a dedicated password protected ASMFC ShareFile drive prior to the day of the prearranged webinar briefing call between the Panel Chair, the three CIE reviewers and key NOAA analytical staff. The documents were also posted on the SEDAR website. All documents are listed in Appendix 1.
- All Terms of Reference of the review were addressed and recommendations made where improvements were likely to be possible. The recommendations were developed and are organized by the identified ToR elements.
- For the southern stock of Atlantic red drum, the Stock Synthesis stock assessment represents the best available science and is fit for use in providing input to management. The associated TLA is also available to provide additional insight into stock status.
- For the northern stock of Atlantic red drum, no accepted Stock Synthesis stock assessment exists. The associated TLA is available to provide quantitative insight into stock status, provided that there is an appropriate understanding of uncertainty.
- Recommendations have been developed for each ToR where considered appropriate.
- One key recommendation is that, for the southern stock, the Stock Synthesis assessment should be updated in 2025, to address some, mostly data, issues identified during this review.
- It was also recommended to continue to develop a Stock Synthesis model for the northern stock.
- A further key recommendation was for an improved process to determine whether abundance time series should be considered indices of abundance of the stock as a whole, including capturing the full justification for retaining or excluding particular abundance data in assessments.

Background

The main fisheries for Atlantic red drum operate off multiple states from Florida to North Carolina but with some smaller catches further north. Most catches have been made from Florida, Georgia, South Carolina and North Carolina and Virginia. The fishery is principally a recreational fishery, with a small commercial catch mostly taken as bycatch in target fisheries for other species. The commercial fishery in Florida ended in 1988. The core assessment data available included catch data estimated from recreational fisher surveys and standard commercial reporting, abundance index data from multiple fishery-independent surveys and from fishery dependent CPUE estimates, biological data from observer carcass collection programs, and tagging data from a number of different tag release and recapture programs. Based on a variety of evidence, the Atlantic red drum are managed as two stocks, a northern stock and a southern stock, with the stock boundary at the NC-SC state line, although there remains uncertainty about whether this is the most appropriate stock boundary (ASFMC, 2024).

In preparation for this benchmark stock assessment, an extensive simulation assessment was developed in 2022 (ASFMC, 2022). The current benchmark stock assessment was completed in 2024 using complete data up to the end of 2022 and some 2023 data (ASFMC, 2024). These activities were coordinated by the Atlantic States Marine Fisheries Commission (ASMFC), with this review coordinated through the SouthEast Data, Assessment, and Review (SEDAR).

This review of the 2024 benchmark assessment of the two Atlantic red drum stocks was conducted as part of an independent review for the Center for Independent Experts (CIE).

All views expressed in this report are solely those of the named, independent CIE reviewer.

Description of Review Activities

This review and reporting was undertaken by Dr Geoff Tingley (Gingerfish Ltd) during August and early September 2024.

The supporting documents for the review were initially provided to the reviewers in downloadable electronic formats from a password protected ShareFile site, and subsequently from the SEDAR website (https://sedarweb.org/assessments/sedar-93-atlantic-red-drum/). These resources were provided in advance of the initial review online meeting The documents provided included the Agenda, Terms of Reference (ToR), inputs data descriptions, programming code and some specific scientific background documents. All documents (but not code files) provided and used are listed in the Bibliography.

A half hour pre-meeting GoToWebinar call was held at 06.00 on 6 August 2024 (NZDT) (14.00 5 August 2024 ET). This call was to address any early questions the CIE reviewers may have had, and to identify any specific requests for additional work that could be progressed prior

to the in-person review meeting. The call included the Panel Chair, the three CIE reviewers, assessment scientists, and SEDAR coordinators.

The key documents describing and reporting the simulation work, and the northern and southern red drum stock assessments were clearly written and informative.

Information relevant to this review is presented in the appendices to this report, as required by the ToR provided by the CIE. These are, Appendix 1 Bibliography of documents, Appendix 2 the CIE Performance Work Statement (which includes its own annexes describing (1) the Peer review reporting requirements, and (2) the ToR for the peer review, and Appendix 3 Review Panel.

The three CIE reviewers worked together, with the Chair, as a Panel during the review meetings and preparation of the panel report. This reviewer then worked independently to prepare this individual CIE report. Questions about the assessment the Panel members had were fully addressed by the assessment team in a timely manner.

The panel review was prepared as a Google Doc, managed by the Panel Chair, with the full engagement and cooperation of the three CIE reviewers and the Panel Chair.

This reviewer greatly appreciated the full engagement and openness of the assessment team members during the review.

Summary of Findings

The review provided an adequately detailed understanding of the fisheries that catch Atlantic red drum that have operated off the southeast coast of the USA, and an in-depth understanding of the simulation and the benchmark assessment for the two stocks of Atlantic red drum. Based on the documentation, presentations and discussions, the assessment team are clearly a knowledgeable and highly competent stock assessment group.

This assessment includes a complex array of different fisheries, as well as spatial and jurisdictional complexity. It is, therefore, not surprising that the available input data are highly complex or that this generates a number of data quality issues within the assessment. It is to the assessment team's credit that they managed this complexity so well and produced several informative models applied to the two red drum stocks.

The review did identify a number of issues that would benefit from further work. These issues were not sufficiently important so as to suggest rejection of the assessment but some of them do need to be considered with a degree of urgency. The details of these issues, most of which the Review Panel showed a high degree of agreement with, are provided in the text of this report.

Overall, the assessment produced information suitable for supporting management but with higher than desirable uncertainty in some areas. Suggestions to reduce this uncertainty in the short-term form the basis of key recommendations.

A number of areas for improvement or development have been identified and recommendations have been developed for all such areas.

This reviewer was impressed by the approach of the assessment team to the review and thanks them for their openness and their full participation.

Addressing the Terms of Reference (ToR) for the Peer Review

Detailed findings and recommendations are presented below, as required by the ToR for the review. Text in italics is taken directly from the requirements of the review defined in the ToR.

Review of the 2024 Atlantic red drum stock assessment.

1. Evaluate responses to Simulation Assessment Peer Review Panel recommendations.

The Simulation Assessment Peer Review Panel made a number of recommendations to the Red Drum Stock Assessment Subcommittee (SAS).

Principle of these recommendations was that the SAS should use the statistical catch-at-age model developed in the Stock Synthesis framework to develop assessments for both the northern and southern stocks of Atlantic red drum, and that this approach should be supported by application of the traffic light analysis (TLA) model as an accessary tool. The SAS followed this recommendation in how the assessment was developed.

Other recommendations were also addressed, including:

(i) using a revised grid search for deriving reference points for the TLA to only include data that would be available to a TLA model when it was applied in practice (i.e. pre-2023), with the revised grid search outputs used as the basis for defining the optimized TLA reference points presented in the assessment report, and

(ii) seeking to demonstrate that the southern Stock Synthesis estimation model could produce unbiased estimates when using data with no observation error. The SAS showed the review panel how this was addressed with versions of the estimation model fitted to otherwise identical models using data generated from the observation model with no observation error. In these runs, the southern stock estimation model runs showed less relative error than in the original simulations, with the relative error approaching zero as the degree of mis-specification decreased. This was a positive result, although it was only conducted using a single iteration. Ideally this should have been run over more scenarios to ensure that this understanding was robust to key parameter differences, such as in the recruitment time series.

This ToR was considered to have been met.

2. Evaluate the thoroughness of data collection and the presentation and treatment of fishery-dependent and fishery-independent data in the assessment, including the following but not limited to:

a. Presentation of data source variance (e.g., standard errors).

The presentation of variance in the various data sources in the documents provided was generally good, with standard error or confidence intervals provided for plotted and tabulated data where understanding variance or uncertainty were important. Some of the plots in the presentations at the meeting did not include variance but this was clarified during the meeting as necessary.

b. Justification for inclusion or elimination of available data sources.

Given the geographical and jurisdictional range of the fisheries under consideration, there were a large number of different data sets to consider as potential input sources for the northern and southern assessments, and also for the TLA and Skate analyses.

Generally, the justification for excluding an individual dataset was clear and well argued. The justification for retaining datasets was mostly, similarly well developed, but two areas were considered lacking in some respects.

Of particular note was an apparent acceptance that some of the time series of abundance information were genuine indices of abundance without always being thoroughly tested. For example, clear analyses were not available that demonstrated that the time series included in the assessment models were all indexing stock abundance and that there were no conflicts between these time series. This issue was explored as far as possible during the review meeting by examining different data sets and by repeating model runs, each time dropping out some of the abundance datasets, which proved informative for some datasets. Given this, the reviewer considers that improvements to the process of deciding on the inclusion or exclusion potential abundance indices in these assessments may be made, and that these may enable improvements in future assessments. This is particularly important as all three methods considered in this assessment (Stock Synthesis, Traffic Light Analysis, and Skate), are fairly or almost entirely reliant on high quality abundance indices.

The historical longline data were excluded due to insufficient coverage and their lack of representativeness for the populations in both the northern and southern models. This was based in part on sensitivity analysis related to the surveys, and for the northern model, the exclusion of the contemporary longline data also had little impact, leading to a recommendation to remove those also.

Collectively, the Review Panel also agreed that some greater consideration of the interaction of different data components may help, for example, ensuring that data were included that inform on all of the different life stages of each stock (e.g., recruitment index, subadult index, adult index, age and growth for older fish).

The Review Panel also collectively identified other areas, including,

(i) The MRIP CPUE time series was recommended to be excluded from the assessment due to a relatively high likelihood of hyperstability.

(ii) The geographical coverage of the stopnet survey was very limited, with data collected from only one site. This raised concern about the representativeness of this time series as a potential index for the stock as a whole. This logic provided a fairly strong basis for recommending exclusion, and to more thoroughly evaluate its consistency with other indices of abundance, or at least use this time series in a sensitivity run.

(iii) The justification provided for excluding all scale-based age data was inadequate. While there is increasing bias in scale ages as fish get older, these data appear to be excellent

for the youngest cohorts, say 0 to 3 year-olds. This source of data also fills what would otherwise be a gap in the age data. The original justification to exclude also used monthly comparisons but there was insufficient evidence presented regarding the overall time series, shorter time periods, and the spatial coverage or the samples.

(iv) Length compositions for the recreational catch were estimated from angler tag release discard data. This process was thoroughly explained and clear. Incorporating these data into the assessment filled an information gap that would otherwise not have been filled.

c. Consideration of data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivities, ageing accuracy, sample size).

The stock assessment team clearly worked diligently to understand and describe the many different datasets available to the assessment. These were well documented in the assessment report and in the various presentations shown at the review meeting. The way this was done greatly assisted the review process.

Many of the datasets had limited geographical coverage, which raised issues of how to evaluate whether they were reflecting stock-wide or just local patterns, especially important for potential abundance indices. Most of these issues were well considered in the assessment report.

It was of note that some surveys showed changes in sampling intensity and distribution over their time series. Where issues related to this were not entirely clear in the assessment report, such as sampling heterogeneity for adult and subadult surveys, clarification was provided during the review meeting.

The sample sizes for length composition data were described (and input to assessment models as initial effective sample size) as the number of sets.

The tagging data available to the assessment were collected over a long period of time and sourced from three different sampling programs (electrofishing, longline and trammel net). The different methods had markedly different selectivities by length (Figure 1),

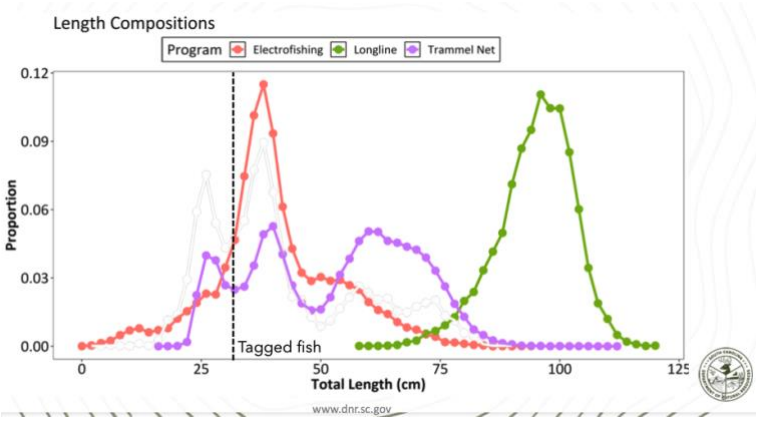

Figure 1: Length compositions of tagged fish from three different sampling sources.

and were also likely to have different post-release mortalities, and the details of the data collection (e.g., when the different gears were sampled over the time period) was not sufficiently clear to be sure that the tagging data represented a single dataset or were unbiased.

d. Calculation and/or standardization of abundance indices.

There were some issues with the standardization of some of the fishery-dependent CPUE time series used as indices of abundance, as indicated by patterns in the residuals. Re-standardization during the review reduced or eliminated patterns in the residuals for several CPUE time series considered, as was seen following repeated sensitivity runs.

3. Evaluate the methods and models used to estimate population parameters (e.g., F, abundance) and reference points, including but not limited to:

a. If modeling approaches differ from those recommended during the Simulation Assessment, were these differences warranted and appropriate?

The most obvious parameterization difference between the simulation assessment and the SS models of the northern and southern red drum stocks was the choice of the value for *h*, the Beverton-Holt steepness parameter. In the simulation assessment operating model, *h* was set at 0.84 based on published evidence, *h* was estimated in the estimation model, which gave unlikely low estimates. For the SS assessments, *h* was fixed at 0.99, a recommendation of the simulation peer review. However, as noted in the assessment report, this effectively made the stock-recruit relationship a constant. While there was good reason to not try and estimate *h* in the SS assessments (too little relevant information in the available data), fixing *h* at 0.99 is viewed by the reviewer as a very strong assumption that is unsupported by evidence, especially when there was an alternative and likely more appropriate value based on evidence and that was used in the simulation work.

b. Evaluate the choice and justification of the preferred model(s). Was the most appropriate model (or model averaging approach) chosen given available data and life history of red drum?

The choice of SS as the modelling framework was appropriate given SS's ability to handle multiple and complex datasets.

The choice of the SAS preferred model structure for the southern stock was well founded and justified and the selection of input data and assumptions were generally well considered and appropriate. The review did find some options for both excluding and including datasets and that differed from the choices in the SAS preferred model. Where there were differences, these were fully discussed during the meeting and explored through some additional model runs. Potentially important differences are

identified in the reviewer responses to the relevant ToR, as for example the recommended exclusion of the longline CPUE time series and the inclusion of the scale-based age data for young fish (2.b above).

As no Stock Synthesis model was accepted for the northern stock, the choice of available model for this stock was much reduced, essentially leaving the TLA approach.

The Traffic Light Analysis (TLA), a data limited, indicator-based approach able to inform across a range of stocks and fisheries either with or without an accepted SS assessment, did provide a basic approach to understanding status. The TLA was simulation tested. The TLA provides a backup to the fully quantitative approach (SS), should the SS model development be unsuccessful (as was the case for the northern stock this time). The review did note that TLA reference periods appeared somewhat long, and that this choice was not well justified (1996-2013 for the northern stock and 1991-2013 for the southern stock). Having an overly-long reference period increases the risk of including times when the stock and fishery were not in a good state; good state being a principal criterion for choosing a reference period. The reference period was also based on the outcomes of previous assessments rather than those of the current assessment where available. The basis for determining the adult abundance indicator was clear but the scaling adjustment factor (0.5) was not well justified. Provided that appropriate reference periods are determined, the application of TLAs, as a support to a stock assessment or to provide limited status guidance in the absence of a stock assessment, is appropriate and thus the choice to use this approach by the SAS team in this regard appears well founded. This is especially so as the performance of the TLA can be compared with (southern assessment) and without a SS model.

The Skate method is an alternative data limited approach but only provides guidance on the scale of potential management response needed, not on when such a response should be made. The review spent limited time considering this approach but noted that the model parameterization appeared to be appropriate. The Skate method was not subject to simulation testing but probably should have been. The assessment team ensured that both the abundance index and the catch time series targeted the same, younger age groups, ensuring consistency within the evaluation and linked to the focus of the fishing mortality. Wide inter-annual fluctuations were minimised by using three-year moving averages for both the abundance index and the catch, and a fixed reference point was used, both of which were appropriate. The review did have some concerns over the basis of the reference point selection, which was based on the median of the catch-to-index ratio for the time series. This choice was not well justified and could potentially be influential in some management decisions, such as the setting of catch limits. Some other weaknesses of this data-limited approach were also noted, specifically including a sensitivity to changes in the size of the year classes potentially causing over- or underestimation of available catch quantities.

Tagging data were available and the Cormack-Jolly-Seber (CJS) model was selected for the analysis. While a number of alternative models could also have been used, this choice seems entirely appropriate for the available data and intended use. Some

necessary information about the tagging programs was not included in the report but was provided during the review meeting, including information on tag loss and gear-dependent selectivity. More importantly, there was an unresolved question about possible differences in gear-specific post-release mortality, with different gear types providing fish for tagging over different periods of the overall tagging program, with the potential for undefined bias in the results. The presentation of the annual apparent survival from the CJS analysis was also considered to be over-smoothed.

c. Evaluate model parameterization and specification (e.g., choice of CVs, effective sample sizes, likelihood weighting schemes, calculation/specification of M, stock-recruitment relationship, choice of time-varying parameters, plus group treatment).

Some time was spent during the review understanding in detail the way in which effective sample sizes were determined, with no suggestions for alternatives forthcoming. Specification of other parameters, such as *M*, were discussed and although different ways of dealing with sex or age were discussed. no recommendations were made. There was a different approach proposed for parameterising the Beverton-Holt stock–recruit relationship, which is discussed in more detail at 3a (above).

4. Evaluate the diagnostic analyses performed, including but not limited to:

Comprehensive model diagnostic analyses were carried out and fully reported for the Stock Synthesis (SS) models for both the southern and northern stocks. These analyses included evidence of adequate convergence, goodness-of-fit, and a range of sensitivities to key assumptions or uncertainties.

For the assessment of the southern stock, the assessment team adequately demonstrated model convergence, with no parameters at bounds. Goodness-of-fit to abundance indices, and compositions data was tested using residual analyses. These were generally good but there were some indications of pattern indicative of either issues in the data or some model misspecification. Some of these issues were resolved during the review, particularly by applying different standardization models to some of the abundance indices and running additional sensitivities showing improved diagnostics.

a. Sensitivity analyses to determine model stability and potential consequences of major model assumptions.

The sensitivity of the southern stock SS model results to specific inclusion and exclusion of data was explored as the model was developed. No such sensitivities were conducted on the base case model, but some should certainly have been done for key assumptions, such as the Beverton-Holt steepness parameter, *h*. No developmental bridging runs from

the previous assessment were presented, which was considered acceptable given the substantive changes in current model specification from the previous one.

While there was no accepted SS model for the northern stock, the developmental process was clearly similar to that used for the southern stock, and any comments and recommendations should also be applicable to future attempts to deliver an assessment for the northern stock.

The TLA results for both stocks were tested for sensitivity to alternative reference periods, with 11 different reference periods evaluated but not fully reported. These sensitivity runs did not find any substantive issues with the TLA, with most results broadly aligned. A number of other potential sensitivities were discussed but not progressed during the review including, using the updated (improved residual pattern) abundance indices and changes in TLA thresholds due to changes in the assumed steepness.

For the Skate model (both stocks), sensitivity to the selection of terminal year was explored and reported. Some other potential sensitivities were discussed but not progressed during the review including, using the updated (improved residual pattern) abundance indices, the *F* reference period selected (i.e. not the whole time series), and the number of years in the moving average.

b. Retrospective analysis.

Retrospective analysis of the assessment for the southern stock was well conducted and fully reported. There was a small retrospective pattern for which the assessment team had a plausible, data-driven cause due to low abundance estimates in 2019. This issue should be reconsidered following any re-standardization of the indices (as recommended elsewhere), which may change the outcome. Historical retrospective analysis was used to compare the current model to the performance of earlier assessments. There was nothing indicated in the retrospective analyses that should give managers cause for concern with regard to the setting of catch limits (e.g., too high or too low).

For the TLA, there remain some opportunities for a historical retrospective analysis but this may well be limited by the length of the time series of key datasets.

5. Evaluate the methods used to characterize uncertainty in estimated parameters. Ensure that the implications of uncertainty in technical conclusions are clearly stated.

Uncertainty in the SS models was presented through detailed sensitivity analyses, likelihood profiles for key parameters, and asymptotic standard errors. This approach was informative and appropriate.

The choice of sensitivity runs for the SS models was informed by previous work and review findings and these choices were largely appropriate. Two areas where the approach diverged from best practice were in not carrying through sensitivity analysis to the base

case model (southern stock), and not doing so for the strong assumption made for steepness of 0.99. A steepness of 0.99 implies that the stock-recruit relationship is a horizontal line, which is not supported by the available information. When a lower value steepness was tested in sensitivity runs requested during the review (0.84 from the simulation assessment), this led to a drop in both biomass and stock status. There was no evidence that a value of *h* lower than 0.99 was inappropriate based on the model performance and diagnostics. Other sensitivities were also requested by the review Panel (using an updated standardized South Carolina Trammel index, removing the age 0 data from the early years of the Florida haul index, removal of sub-adult length data, and adjusting the MRIP catch estimates), and while all results fell withing the 95% CI of the base case median, some of the other changes suggested there remained some unresolved issues.

Log-likelihood profiles, a standard SS output, were used to good effect, and demonstrated that the model was principally informed by recruitment, length, and discard information. There was, however, some evidence of potential conflicts between the index and composition data, possibly related to the index standardization issue discussed above.

For the TLA (both stocks), uncertainty in the results was driven by changes to the reference period. The selected reference period was relatively long and may therefore include years when the fishery was not performing optimally, which will tend to increase uncertainty. Given this, a shorter reference period may have been more appropriate. Some sensitivity testing was conducted on the length of reference period, which demonstrated the different reference periods tested had little or no effect on the stock status for the southern stock (when compared with the SS output).

In proposing effective management reference points for the TLA, the periods of sequential outcome status for the TLA (e.g., the number of occurrences of an indicator being below a specific threshold, i.e., red) was used to infer management reference points. While this was conceptually helpful, it was not well tested.

The application of the Skate model in the context of the assessment and development of management advice for red drum has value. Using the 3-year moving average dampens the noise in the available indices and likely helps in the robustness of this approach. While no reference points exist for this method, Panel members recognized the value of this approach in providing some quantitative advice, especially if supported by an MSE of simulation testing.

6. If a minority report has been filed, review minority opinion and any associated analyses. If possible, make recommendation on current or future use of alternative assessment approach presented in minority report.

As no minority report was filed; this ToR was considered met.

7. ***Recommend best estimates of stock biomass, abundance, and exploitation from the assessment for use in management, if possible, or specify alternative estimation methods.***

Southern stock:

For the southern stock, the Stock Synthesis (SS) model is clearly the best and most appropriate source of information about the status of the stock (stock biomass, abundance, and exploitation).

The review did identify some issues with the proposed base case that warrant further examination and have been discussed under different ToRs above. In summary, these issues are with the approach to index standardisation, using the most recent catch and effort data, and concern about whether the longline survey is indexing the stock.

The TLA and Skate model can be used to provide additional quantitative information to support advice on the southern stock but can also be used as bridge to better understand the status of the northern stock by comparison with the southern stock.

Northern stock:

For the northern stock, there is no accepted Stock Synthesis (SS) model and no usable model outputs.

The TLA approach offers a backstop option but relying solely on the TLA implies accepting some degree of higher management risk. However, other evidence, including directly from the abundance indices and on-going recruitment, provide no evidence for sustainability issues in the northern stock. It would be important for the TLA to be demonstrably robust and responsive to the state of the stock. Therefore, associated with the use of the TLA, there should be some robustness testing, specifically including for the choice of the reference period and aspects of index reliability. Options for robustness testing exist and include both a simulation-based evaluation or an MSE.

8. ***Evaluate the choice of reference points and the methods used to estimate them. Recommend stock status determination from the assessment, or, if appropriate, specify alternative methods/measures.***

For both northern and southern stocks, there are existing reference points available and used in management. These include the thresholds $F_{30\%}$ and $SPR_{30\%}$ and targets $F_{40\%}$ and $SPR_{40\%}$ and appear to have been derived appropriately. There are also new spawning stock biomass (SSB) reference points but these have not yet been specifically evaluated or accepted.

For the northern stock SS model reference points should be part of the next iteration of development of this assessment and should be informed by what happens in the development of the southern stock assessment and its associated management parameters.

Reference points for the TLA were also presented, based on expert judgement and intended to be precautionary but this has not yet been evaluated.

The Skate approach has no reference points at present. Such quantities could be determined using simulations or MSE but the effort to do this would need to be weighed against the other priorities and the balance of whether having reference points would materially improve the usability of the Skate methodology.

9. ***Review the research, data collection, and assessment methodology recommendations provided by the TC and make any additional recommendations warranted. Clearly prioritize the activities needed to inform and maintain the current assessment, and provide recommendations to improve the reliability of future assessments.***

   The reviewer broadly agrees with the list of research recommendations and their prioritization as presented in the assessment report (ASFMC, 2024). However, some of the recommendations arising from this review take the form of research recommendations and some of these should rate high priority and have short- to medium-term delivery timeframes.

   Four areas where additional research is considered necessary and useful are identified, with a recommendation developed for each. These are :

   (i) CPUE standardization

   (ii) Robustness testing using either simulation analyses or management strategy evaluation (MSE) techniques

   (iii) Tagging data

   (iv) Spatially and temporally patchy data sets

10. ***Review the recommended timeframe for future assessments provided by the TC and recommend any necessary changes.***

    The recommended five-year timeframe for the next Benchmark Assessment appears appropriate from both a resources and a management need perspective. Moreover, having a reasonable period before trying another northern assessment will allow considerably more data to accumulate, especially abundance index data critical to the success of any such modelling attempt.

    Noting the various issues in the proposed base case for the SS assessment of the southern red drum stock, a shorter time frame for an update assessment would be appropriate.

    For the northern red drum stock, an earlier update to the TLA is also advisable in order to be using the most recent data to develop advice.

    In general, where the TLA is providing the core of the advice (as for the northern stock), an update frequency of 2-years is probably appropriate (as advised by the SAS). In order to retain the ability to compare the TLA for the northern and southern stocks and the southern SS model, the TLA for both areas should be updated at the same time.

11. ***Prepare a peer review panel terms of reference and advisory report summarizing the panel's evaluation of the stock assessment and addressing each peer review term of***

reference. Develop a list of tasks to be completed following the workshop. Complete and submit the report within 4 weeks of workshop conclusion.

An agreed Panel Report was prepared, led by the Panel Chair with contributions from all four members of the Panel. This ToR was completed.

Recommendations

The following are CIE reviewer recommendations arising from the review of the 2024 Atlantic red drum benchmark stock assessment.

Recommendations against each ToR

1. ***Evaluate responses to Simulation Assessment Peer Review Panel recommendations.***

None.

2. ***Evaluate the thoroughness of data collection and the presentation and treatment of fishery-dependent and fishery-independent data in the assessment, including the following but not limited to:***

   a. ***Presentation of data source variance (e.g., standard errors).***

      None

   b. ***Justification for inclusion or elimination of available data sources.***

      (i) Develop guidance or a process of analyses such that it is demonstrated that for each time series considered for inclusion in an assessment, the time series does index the overall stock abundance and that there are no conflicts with other time series. This will then provide the best set of abundance indices for use and clear justification for either inclusion or exclusion of particular time series.

      ii) Include scale age data for those ages where there is little or no bias.

   c. ***Consideration of data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivities, ageing accuracy, sample size).***

      (i) A review of the statistical approach to standardizing CPUE data is probably warranted to reduce or eliminate problematic residual patterns from all such time series prior to their consideration as indices of abundance.

      ii) Ensure that there are no biases in the tagging data accruing from different sampling methods applied at different time periods due to differing selectivities or post-release mortality.

      iii) For those datasets, including some surveys, that were constrained and/or excluded from the assessment due to patchy coverage in time and/or space, reanalysis using modern spatio-temporal assessment techniques (such as Integrated Nested Laplace

Approximation (INLA) or Vector Autoregressive Spatio-Temporal (VAST) modelling) may enable some of these datasets to become informative in future assessments.

d. Calculation and/or standardization of abundance indices.

(i) A review of the statistical approach to standardizing CPUE data is probably warranted so as to reduce or eliminate problematic residual patterns from all such time series prior to their consideration as indices of abundance.

3. Evaluate the methods and models used to estimate population parameters (e.g., F, abundance) and reference points, including but not limited to:

a. If modeling approaches differ from those recommended during the Simulation Assessment, were these differences warranted and appropriate?

(i) It is recommended that a sensitivity analysis of the southern stock SS model to the choice of the fixed value of the Beverton-Holt steepness parameter, *h*, should be conducted and based on available evidence.

b. Evaluate the choice and justification of the preferred model(s). Was the most appropriate model (or model averaging approach) chosen given available data and life history of red drum?

(i) Based on the findings during this review, it is recommended that the base case SS model for the southern stock be reconsidered specifically with respect to (a) the standardization of the abundance time series, and (b) which abundance time series are retained as abundance indices based on model performance and coherence between the different indices.

(ii) It is recommended that alternative, shorter reference periods be explored and tested for the TLA to ensure the robustness of this methodology.

(iii) It is recommended that, if the Skate method is to be used to inform management, it should undergo some form of simulation or MSE robustness testing, including to identified potential weaknesses such as sensitivity to year class strength variability.

(iv) It is recommended that the basis of the reference point selection for the Skate methodology be revisited and fully justified.

c. Evaluate model parameterization and specification (e.g., choice of CVs, effective sample sizes, likelihood weighting schemes, calculation/specification of M, stock-recruitment relationship, choice of time-varying parameters, plus group treatment).

None.

4. Evaluate the diagnostic analyses performed, including but not limited to:

a. Sensitivity analyses to determine model stability and potential consequences of major model assumptions.

(i) It is recommended that the sensitivity of the southern stock model to any updated standardized abundance indices be tested to ensure that the final base model is robust.

ii) Revisit the sensitivity testing of the TLA to explore how any updated abundance indices influence the performance of the TLA. It is also recommended to consider other sensitivity analyses, including but not limited to the fixed value of the Beverton and Holt steepness parameter, *h*.

iii) Revisit the sensitivity testing of the Skate analysis to explore how any updated abundance indices influence performance. It is also recommended to consider other sensitivity analyses such as the number of years in the moving average and the *F* reference period selected (i.e., less than the whole time series).

b. Retrospective analysis.

(i) Given the potentially important role the TLA may have in providing management advice, especially for the northern stock, a more thorough historical retrospective analysis should be conducted to ensure both a fuller understanding of the TLA performance but also its robustness.

5. Evaluate the methods used to characterize uncertainty in estimated parameters. Ensure that the implications of uncertainty in technical conclusions are clearly stated.

(i) It is recommended that the additional sensitivity runs delivered during the review meeting be used to better define the suit of sensitivities used in testing future red drum assessment models.

(ii) It is recommended that, for the TLA, the process of inferring management reference points should be the subject of a formal management strategy evaluation (MSE).

(ii) It is recommended that to support the application of the Skate method to be useful in providing quantitative advice to managers, appropriate robustness testing using, for example, MSE or a simulation approach should be explored.

6. If a minority report has been filed, review minority opinion and any associated analyses. If possible, make recommendation on current or future use of alternative assessment approach presented in minority report.

None.

7. ***Recommend best estimates of stock biomass, abundance, and exploitation from the assessment for use in management, if possible, or specify alternative estimation methods.***

   (i) In order to produce the most informative and robust advice on the southern stock of red drum, it is recommended that the index standardization, recent catch history, and exclusion of the longline time series be updated for the base case SS model.

   (ii) If the management advice for the norther red drum stock is going to principally rely on the TLA model, then urgent, formal robustness testing of the TLA is recommended, probably using either simulation or MSE.

8. ***Evaluate the choice of reference points and the methods used to estimate them. Recommend stock status determination from the assessment, or, if appropriate, specify alternative methods/measures.***

   (i) The proposed SSB reference points (RPs) for the southern stock should be evaluated and adopted if appropriate.

   (ii) The proposed SSB reference points for the northern stock should be subject to further review as part of the next iteration of development of this assessment. The outcomes of the development of the southern stock RPs should also be considered.

   (iii) The proposed reference points for the TLA should be evaluated and adopted if appropriate.

   (iv) It is recommended that the development of RPs for the Skate methodology should be considered, commensurate with the likelihood of the application of this methodology in providing management advice.

9. ***Review the research, data collection, and assessment methodology recommendations provided by the TC and make any additional recommendations warranted. Clearly prioritize the activities needed to inform and maintain the current assessment, and provide recommendations to improve the reliability of future assessments.***

   In addition to the research recommendations in the assessment report, four additional ones are included here:

   (i) CPUE standardization: the approach to standardizing the CPUE indices that are an important component of this assessment needs to be reviewed and improved, using the most up-to-date approaches. Probably all of these times series should be reviewed, with the quality of the standardization tested though sensitivity analysis for lack of problematic residual patterns. The implications of all such changes need to be followed through to the other model frameworks to ensure that the selection, running, and outputs of the TLA and Skate models remain appropriate. This is a substantial amount of work but is clearly both urgent and a high priority.

(ii) There are a number of recommendations concerning robustness testing using either simulation analyses or management strategy evaluation (MSE) techniques. These include issues of reference points for the SS, TLA and Skate models, reference periods for the TLA, etc. These various strands of work should be considered as a whole and progressed in a resource and cost-effective way to deliver the required science quality to support the advice (high priority).

(iii) Tagging data: in order to make full use of the historic and current tagging data, uncertainties related to potential gear-related differences in post-release mortality and in tag loss need to be fully quantified (medium priority).

(iv) The spatially and temporally patchy data sets, particularly the surveys, should be subject to spatio-temporal modelling research to try and establish usable data from these time series (medium to low priority).

10. Review the recommended timeframe for future assessments provided by the TC and recommend any necessary changes.

(i) It is recommended that the SS assessment for the southern stock be updated in 2025 to address the issues identified during this review, and should include:

(a) the most recent data available (including catch, biological and abundance indices data);

(b) updating the model according to the recommendations in the Panel and individual CIE review reports, specifically including the approach to standardization of abundance indices and in the testing and selection of retained abundance indices; and

(c) incorporating the expected changes in the catches derived from the MRIP (if available).

(ii) It is recommended that the TLA be updated every second year for both stocks to enable cross comparison between the two and strengthen the advice for the northern stock.

11. Prepare a peer review panel terms of reference and advisory report summarizing the panel's evaluation of the stock assessment and addressing each peer review term of reference. Develop a list of tasks to be completed following the workshop. Complete and submit the report within 4 weeks of workshop conclusion.

No recommendations.

Appendix 1: Bibliography

Atlantic red drum Review Documents

Background Documents

Arnott1, S.A., Roumillat, W.A., Archambault, J.A., Wenner, C.A., Gerhard, J.I., Darden, T.L. and Denson, M. R. (2010). Spatial synchrony and temporal dynamics of juvenile red drum *Sciaenops ocellatus* populations in South Carolina, USA. Mar. Ecol. Prog. Ser. 415, 221-236.

ASMFC (2022). Red Drum Simulation Assessment and Peer Review Report. 567pp.

ASMFC (2024). 2024 Red Drum Benchmark Stock Assessment Report. 189pp.

Troha, L. U. (2023). Estimating the tag-reporting rate and length-based selectivity of red drum (S*ciaenops ocellatus*) in South Carolina using a long-term tag-recapture study. MSC Thesis, Univ. Charleston SC. 89 pp.

Other Material Provided or used

Full code for the northern and southern assessments, and for the traffic light analysis (TLA) model were provided.

Thirteen presentations covering all aspects of the assessments were also made available during the meeting.

Appendix 2: Performance Work Statement

Performance Work Statement (PWS)
National Oceanic and Atmospheric Administration (NOAA)
NOAA Fisheries
Center for Independent Experts (CIE) Program
External Independent Peer Review
Southeast Data, Assessment, and Review (SEDAR) 93 Atlantic Red Drum Assessment Review

Background

The NOAA Fisheries 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). NOAA Fisheries 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 (1) 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. Specifically, science products that the agency can reasonably determine that will have, when disseminated, *"a clear and substantial impact on important public policies or private sector decisions*." Additionally, peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards[1].

Scope

The SEDAR is the cooperative process by which stock assessment projects are conducted in NOAA Fisheries Southeast Region. SEDAR was initiated to improve planning and coordination of stock assessment activities and to improve the quality and reliability of assessments.
SEDAR 93 will be a CIE assessment review conducted for Atlantic Red Drum. There are two (2) models to be reviewed: Southern and Northern Stocks. The review workshop provides an independent peer review of SEDAR stock assessments. The term review is applied broadly, as the review panel may request additional analyses, error corrections and

[1] https://www.whitehouse.gov/wp-content/uploads/legacy_drupal_files/omb/memoranda/2005/m05-03.pdf

sensitivity runs of the assessment models provided by the assessment panel. The review panel is ultimately responsible for ensuring the scientific basis of the assessment through the SEDAR process. The specified format and contents of the individual peer review reports are found in **Annex 1**. The Terms of Reference (ToR) of the peer review are listed in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3.**

Requirements

NOAA Fisheries requires three (3) reviewers and a chairperson to conduct an impartial and independent peer review in accordance with the PWS, OMB guidelines, and the ToR below. The reviewers shall have a working knowledge in stock assessment, statistics, fisheries science, and marine biology sufficient to complete the primary task of providing peer-review advice in compliance with the workshop Terms of Reference for the stock assessment. The chair, who is in addition to the three (3) reviewers, will not be provided by the CIE. Although the chair will be participating in this review, the chair's participation (e.g., labor and travel) is not covered by this contract.

Tasks for Reviewers

1. Two (2) weeks before the peer review, the Project Contacts will send (by electronic mail) the necessary background information to the CIE reviewers and reports for the peer review. In the case where the documents need to be mailed, the Project Contacts will consult with the contractor on where to send documents. CIE reviewers are responsible only for the pre-review documents that are delivered to the reviewer in accordance with the PWS scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.
2. Attend and participate in an in-person review meeting. The meeting will consist of presentations by NOAA and other scientists, stock assessment authors and others to facilitate the review, to answer any questions from the reviewers, and to provide any additional information required by the reviewers.
3. After the review meeting, reviewers shall conduct an independent peer review report in accordance with the requirements specified in this PWS, OMB guidelines, and ToR, in adherence with the required formatting and content guidelines. Reviewers are not required to reach a consensus.
4. Each reviewer shall assist the Chair of the meeting with contributions to the summary report.
5. Deliver their reports to the Government according to the specified milestones dates.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NOAA Fisheries Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NOAA Fisheries Project Contact for

the purpose of their security clearance, and this information shall be submitted at least 30 days in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Foreign National Guest website. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance
The places of performance shall be in Charleston, SC.

Period of Performance
The period of performance shall be from the time of award through **September 2024**. Each CIE reviewer's duties shall not exceed 14 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 (2) weeks of award	Contractor selects and confirms reviewers
Two (2) weeks prior to the panel review	Contractor provides the pre-review documents to the reviewers
August 13 – 16, 2024	Panel review meeting
Approximately three (3) weeks later	Contractor receives draft reports
Within two (2) weeks of receiving draft reports	Contractor submits final reports to the Government

Applicable Performance Standards
The acceptance of the contract deliverables shall be based on three (3) 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; and (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel
All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $15,000.

Restricted or Limited Use of Data
The contractors may be required to sign and adhere to a non-disclosure agreement.

Project Contacts:
Larry Massey – NOAA Project Contact
150 Du Rhu Drive, Mobile, AL 36608
(386) 561-7080
larry.massey@noaa.gov

Julie Neer - SEDAR Program Manager
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405
Julie.Neer@safmc.net

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 the science reviewed is adequate.

2. The report must contain a background section, description of the individual reviewers' roles 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 ToR.

a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.

b. Reviewers should discuss their independent views on each ToR even if these were consistent with those of other panelists, but especially where there were divergent views.

c. Reviewers should elaborate on any points raised in the summary report they believe might require further clarification.

d. Reviewers shall provide a critique of the NOAA Fisheries review process, including suggestions for improvements of both process and products.

e. The 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 report shall represent the peer review of each ToR, and shall not simply repeat the contents of the summary report.

3. The report shall include the following appendices:

Appendix 1: Bibliography of materials provided for review
Appendix 2: A copy of this Performance Work Statement
Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review
SEDAR 93 Atlantic Red Drum Assessment
Review Workshop Terms of Reference

1. Evaluate responses to Simulation Assessment Peer Review Panel recommendations.
2. Evaluate the thoroughness of data collection and the presentation and treatment of fishery-dependent and fishery-independent data in the assessment, including the following but not limited to:
   a. Presentation of data source variance (e.g., standard errors).
   b. Justification for inclusion or elimination of available data sources.
   c. Consideration of data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivities, ageing accuracy, sample size).
   d. Calculation and/or standardization of abundance indices.
3. Evaluate the methods and models used to estimate population parameters (e.g., F, abundance) and reference points, including but not limited to:
   a. If modeling approaches differ from those recommended during the Simulation Assessment, were these differences warranted and appropriate?
   b. Evaluate the choice and justification of the preferred model(s). Was the most appropriate model (or model averaging approach) chosen given available data and life history of red drum?
   c. Evaluate model parameterization and specification (e.g., choice of CVs, effective sample sizes, likelihood weighting schemes, calculation/specification of M, stock-recruitment relationship, choice of time-varying parameters, plus group treatment).
4. Evaluate the diagnostic analyses performed, including but not limited to:
   a. Sensitivity analyses to determine model stability and potential consequences of major model assumptions.
   b. Retrospective analysis.
5. Evaluate the methods used to characterize uncertainty in estimated parameters. Ensure that the implications of uncertainty in technical conclusions are clearly stated.
6. If a minority report has been filed, review minority opinion and any associated analyses. If possible, make recommendation on current or future use of alternative assessment approach presented in minority report.
7. Recommend best estimates of stock biomass, abundance, and exploitation from the assessment for use in management, if possible, or specify alternative estimation methods.
8. Evaluate the choice of reference points and the methods used to estimate them. Recommend stock status determination from the assessment, or, if appropriate, specify alternative methods/measures.
9. Review the research, data collection, and assessment methodology recommendations provided by the TC and make any additional recommendations warranted. Clearly prioritize the activities needed to inform and maintain the current assessment, and provide recommendations to improve the reliability of future assessments.
10. Review the recommended timeframe for future assessments provided by the TC and recommend any necessary changes.
11. Prepare a peer review panel terms of reference and advisory report summarizing the panel's evaluation of the stock assessment and addressing each peer review term of reference. Develop a list of tasks to be completed following the workshop. Complete and submit the report within 4 weeks of workshop conclusion.

Annex 3: Tentative Agenda - SEDAR 93 Atlantic Red Drum Assessment Review
August 13 – 16, 2024

Tuesday

8:30 am – 9:00 am **Introductions and Opening Remarks** **Coordinator**
- Agenda Review, TOR, Task Assignments
9:00 am – 12:00 pm **Assessment Presentations** **Analytic Team**
- *Background*
- *Assessment Data & Methods*
12:00 pm – 1:30 pm **Lunch Break**
1:30 pm – 4:30 pm **Assessment Presentations (continued)** **Analytic Team**
- *Assessment Data & Methods*
- Identify additional analyses, sensitivities, corrections
4:30 pm – 5:00 pm **Wrap Up/Public Comment** **Chair**
5:00 pm - 6:00 pm **Panel Work Session** **Chair**
***Tuesday Goals*:** Initial assessment presentations completed, sensitivities and modifications identified.

Wednesday

8:30 a.m. – 11:30 pm **Assessment Presentations (continued)** **Analytic Team**
- *Assessment Methods*
- Identify additional analyses, sensitivities, corrections
11:30 a.m. – 1:00 pm **Lunch Break**
1:00 pm – 5:30 pm **Panel Discussion** **Chair**
- Review additional analyses, sensitivities
- Recommendations and comments
5:30 pm - 6:00 pm **Public Comment** **Chair**
Wednesday Goals: Presentations completed, additional sensitivities identified, preferred models selected, projection approaches approved, Summary report drafts begun

Thursday

8:30 a.m. – 11:30 pm **Panel Discussion** **Chair**
- Review additional analyses, sensitivities
- Recommendations and comments
11:30 a.m. – 1:00 pm **Lunch Break**
1:00 pm – 5:30 pm **Panel Discussion** **Chair**
- Final sensitivities reviewed.
- Projections reviewed.
5:30 pm - 6:00 pm **Public Comment** **Chair**
Thursday Goals: Review final sensitivities, complete assessment work, and finalize discussions.

Friday

8:30 a.m. – 12:00 pm **Panel Discussion or Work Session** **Chair**
- Review Summary Reports
Friday Goals: Final results available. Draft Summary Report reviewed.

Appendix 3: Panel Membership

Review Panel

Gavin Fay (Chair) .....
Kotaro Ono ..... CIE Reviewer
Geoff Tingley ..... CIE Reviewer
Katyana Vert-Pre ..... CIE Reviewer

Analytic Team

Joey Ballenger ..... SCDNR
Tracey Bauer ..... ASMFC
Jared Flowers ..... GADNR
Angela Giuliano ..... MADNR
Jimmy Kilfoil ..... SCDNR
Jeff Kipp ..... ASMFC
CJ Schlick ..... SCDNR

Staff

Julie A Neer ..... SEDAR
Emily Ott ..... SEDAR
Rachael Silvas ..... SAFMC Staff

Workshop Observers

Chip Collier ..... SAFMC Staff

Workshop Observers via Webinar

Alan Bianchi ..... NCDNR
Pat Campfield ..... ASMFC
Manuel Coffill-Rivera .....
Dawn Franco ..... GADNR
Ryan Harrell ..... GADNR
Matthew Jargowsky ..... MADNR
Chris Kalinowsky ..... GADNR
Cara Kowalchyk ..... NCDNR
Laura Lee ..... NCFWS
Rebecca Scott ..... FWC