Center for Independent Experts (CIE)

Independent review report

SEDAR 100 Gulf of America Gray Triggerfish Stock Assessment

Nicholas Fisch

September 2026

Executive Summary

The SEDAR 100 Gulf of America Gray Triggerfish stock assessment represents a substantial improvement over the previous assessment conducted under SEDAR 43. The assessment benefits from substantially revised and reanalyzed removals data, updated life-history information, improved age readings, updated abundance indices and composition datasets, and a generally appropriate integrated age-structured modeling framework. Overall, the data and analytical procedures are largely sound and consistent with regional stock-assessment practices, and the assessment provides a substantially improved basis for understanding the historical dynamics of Gulf Gray Triggerfish.

The principal methodological issue concerns the treatment of natural mortality (*M*) and stock-recruitment steepness (*h*). The assessment's original configuration fixed both parameters, thereby fixing the two principal dimensions governing stock productivity. Additional analyses presented during the review demonstrated that fixing *h* at 0.99 was not as well supported as estimating it internally. When *M* was fixed at 0.36, the model estimated *h* at approximately 0.42, with similar estimates across several sensitivity analyses. The model estimating *h* also produced improved retrospective diagnostics, better fits to the data, and greater robustness to alternative recruitment assumptions. Consequently, I support, in consensus with the Review Panel, a revised base model that fixes *M* at 0.36 and estimates *h* internally.

Importantly, however, the estimated *h* should not be interpreted as a reliable biological parameter for long-term projections or equilibrium reference-point calculations. The observations informing the stock-recruitment relationship cover a relatively limited range of spawning stock biomass (SSB). Extrapolating the estimated Beverton-Holt relationship beyond that observed range produced biologically and operationally problematic results, including unrealistically low recruitment and an inability of the projected stock to maintain the intended SPR30% equilibrium. I therefore view estimation of *h* primarily as a means of improving the historical assessment and producing a more stable and plausible reconstruction of stock dynamics, rather than as justification for extrapolating the estimated stock-recruitment relationship into the future.

The assessment's treatment of uncertainty is another important area for improvement. Although numerous sensitivity analyses were conducted, several influential parameters were fixed, meaning their uncertainty does not propagate into derived quantities. In addition, the standard Hessian/delta-method approach used by Stock Synthesis provides only an asymptotic approximation to model uncertainty. The Dirichlet-multinomial treatment of composition data may also underestimate uncertainty, because it cannot accommodate the positive correlations apparent among neighboring composition bins, potentially resulting in excessive effective sample sizes and artificial certainty. I therefore recommend greater use of sensitivity analyses, prior distributions rather than fixed parameter values where appropriate, and consideration of bootstrap or MCMC approaches to better characterize total assessment uncertainty.

The principal implication of these findings is for the stock projections and management reference points. The current projection framework is technically appropriate and consistent with standard Stock Synthesis procedures, but the projections are overly dependent on the assumed recruitment relationship and do not adequately propagate the assessment's important sources of uncertainty.

In particular, projections based on extrapolating with either the former $h = 0.99$ relationship or the newly estimated h should not be used as the principal basis for catch advice.

I recommend instead that reference points and projections be based on empirical recent recruitment estimated in the assessment model. Specifically, F30%SPR should be estimated through per-recruit analysis, while SSBF30%SPR should be calculated using a justified central measure of recruitment from recent, well-estimated cohorts, without extrapolating the stock-recruitment relationship beyond its historically informed SSB range. Projections should then sample from recent empirical recruitment, retaining observed recruitment variability for stochastic projections. Alternative recruitment windows—particularly 5-, 10-, and 15-year periods—should be evaluated and their effects on projected stock status and management advice reported.

Several improvements to the underlying data streams would further strengthen future assessments. Highest priorities include improving age determination, including exploration of epigenetic ageing approaches; improving calibration of SEAMAP and commercial shrimp-trawl observations to reduce uncertainty in shrimp bycatch estimates; conducting tagging studies to improve estimates of recreational discard mortality and other sources of mortality; and investigating the potential relationship between sargassum habitat and Gray Triggerfish recruitment.

For the near term, I recommend incorporating the ageing-bias relationship into the assessment and associated data-preparation models, using bootstrap methods to derive more objective composition-data sample sizes, considering spatial and spatiotemporal standardization of abundance indices, expanding assessment diagnostics to include one-step-ahead residuals and standardized residual diagnostics, and using priors rather than fixed values for parameters such as natural mortality and recruitment variation where feasible.

In summary, SEDAR 100 provides a strong and substantially improved historical assessment of Gulf Gray Triggerfish, but the current projection framework should be revised before it is used to provide definitive catch advice. The key near-term priorities are to adopt the revised model with M fixed at 0.36 and h estimated, reformulate projections around empirical recent recruitment rather than extrapolated stock-recruitment relationships, recalculate management reference points accordingly, and more explicitly characterize uncertainty.

Background

The SEDAR 100 Gulf of America Gray Triggerfish Stock Assessment review workshop (RW) took place from August 10-14, 2026, in Tampa, Florida. The RW is part of the Southeast Data, Assessment, and Review (SEDAR) cooperative process for assessments conducted in NOAA Fisheries' Southeast Region. The Review Panel (RP) was chaired by Luiz Barbieri and included five other reviewers: Yong Chen (CIE), Anders Nielsen (CIE), Erik Lang, Jie Cao, and myself (Nicholas Fisch-CIE). The draft assessment report and all associated background documents (Appendix 1) were made available to the reviewers two weeks before the RW. At the RW, the assessment authors presented conclusions from the data workshop and assessment process, including details of the data inputs and the assessment methods. Additional model runs were produced at the RP request. The workshop was prepared and organized by SEDAR coordinator Julie Neer. This report is my independent review of the stock assessment for the CIE and reflects my own views, which are largely consistent with the RP conclusions in the summary report.

Individual Reviewer's Role

My role as an individual CIE reviewer in SEDAR 100 was to review the SEDAR 100 Data Workshop Report, the SEDAR 100 Assessment Process Report, and all SEDAR 100 Working papers and Reference documents (Appendix 1). The role also included conducting an independent peer review in accordance with the PWS (Appendix 2) and Terms of Reference (ToRs , below) via attending and actively participating in a professional and respectful manner as a member of the meeting review panel, with peer review tasks focused on the ToRs as specified below. My role as a CIE reviewer also extended to assisting the Chair of the meeting with contributions to the summary report.

In accordance with this role, I independently read the documents provided for review and background in preparation, traveled to the review workshop, participated actively in physical review meetings in Tampa, reviewed each term of reference and assisted the chair and other reviewers in drafting the chair summary report, and independently authored this review report.

Summary of Findings for TORs

1. ***Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions. Consider the following:***

   - ***The context of the data availability in the region. Are the procedures appropriate for our data limitations?***

     I find that the procedures utilized were largely appropriate for the data availability in the region. The assessment processed data such that there existed landings, discards, indices of abundance, length compositions, and conditional age-at-length compositions. The assessment also prepared fixed estimates of natural mortality (based on longevity), maturity at length, fecundity, discard mortality, and ageing error.

     The one additional data input I would have like to see would be age composition data. These could be used to evaluate the sensitivity to fitting conditional age-at-length data in the assessment. I recommend processing the data in the future such that there exist length compositions, age compositions, and conditional age-at-length compositions. This would allow for evaluating the impact of fitting conditional age-at-length data via substituting in age compositions in years that were previously fitting length and conditional age-at-length data (in years without age compositions, the model would still be fitting length compositions). One of the concerns with conditional age-at-length data is that it spreads the data quite thin (few observations per length bin, Stewart and Monnahan, 2025) and one way to evaluate this may be to substitute in full age compositions in those years with age data by removing length and conditional-age-at-length compositions.

   - ***Are data decisions made by the Data Workshop (DW) justified?***

     I found most of the key decisions made by the Data Workshop to be justified. Namely:

     The assessment chose a single area model and as such processed data Gulf-wide. The assessment cites lack of differences in growth and in length and age compositions between the east and west areas of the Gulf as a rationale. I support this decision, as it is a parsimonious choice given the lack of evidence for separate stocks (not to mention the additional movement assumptions that accompany multiple-area models).

     The decision to initiate the model in 1945 was made as exploitation was assumed to be negligible prior to this period. This makes intuitive sense given global events during that time (World War II), and the data certainly seem to suggest minimal removals around that timeframe. Starting the model later would require the difficult estimation of initial conditions, such as equilibrium fishing mortality or initial numbers at age, which I find are more tenuous assumptions than assuming an unfished state in 1945.

The decision to model the assessment using a single sex was made as growth models did not show any differences in mean length at age by sex. This finding was in contrast to previous studies (Chamberlin et al., 2025). However, I find it is justified in light of the observed data.

Fixed natural mortality in the assessment was assumed using the longevity-based methods of Hamel and Cope (2022), while a Lorenzen function was used to scale the natural mortality at age based on mean length at age. The Data Workshop recommended utilizing a max age of 15, which I understand was based on the fact that there were several age 15 specimens and only one older, which was deemed an outlier and potentially due to ageing error. This choice is consistent with recommendations from Hamel and Cope (2022). In contrast, the age data used in the assessment were combined for older fish into a 10+ age group. The assessment authors noted that observations above age 10 were sparse and that retaining these observations as separate age classes could add unnecessary complexity and potentially influence model estimates without providing substantial additional information. I support this decision as well.

The assessment assumed maturity as a function of length, under the rationale that age was highly uncertain. In addition, using a similar rationale, the assessment assumed fecundity at age was proportional to the weight at age of the fish. I believe each of these decisions to be justified in light of the uncertainty in ageing. However, in the future, growth, maturity, and fecundity models could be developed that make use of the ageing error matrix directly in their model structure.

The assessment assumed a 40% discard-mortality rate, obtained from a synthesis of the literature on tagging and acoustic telemetry. The assessment evaluated the sensitivity to this assumption by examining discard mortalities of 26% and 60% in sensitivity runs. I believe both the choice of the base discard rate and the evaluation of the sensitivity to the assumption to be justified and standard.

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

  The data uncertainties are mostly acknowledged, reported, and within normal or expected levels. One of the strengths of this assessment is the explicit acknowledgement of uncertainty in the input data series and their consideration in the actual model. In particular, I commend the fitting of recreational landings time series with error, which is not commonly done in integrated stock assessments conducted in Stock Synthesis.

  There are certainly some aspects of the data uncertainty that could be improved. The assessment utilizes the number of trips sampled or the number of fish sampled as the input sample size for the composition data sets. Although a simple heuristic, the input sample sizes could be more objectively developed using a bootstrapping approach such as that presented in Stewart and Hamel (2014). Using bootstrapping (resampling the trips, then the specimen within trips) allows for an input sample size estimate for compositions that can serve as a

good theoretical upper bound where the Dirichlet-Multinomial can further down-weight if deemed necessary. Another suitable alternative would be to employ a spatial and spatiotemporal standardization of the different composition data sets such as that presented in Thorson and Haltuch (2019), which also results in estimates of input sample sizes.

- ***Is the appropriate model applied properly to the available data?***

  I understand this to describe the models applied in processing the data series (not the assessment model). The data processing models were largely appropriate for the available data. The multiple abundance index time series are generally standardized using common methods for the region (delta-lognormal generalized linear models). However, spatial and spatiotemporal standardizations of abundance index time series have become more common across many other regions and have been shown to outperform design-based indices (Thorson et al., 2015). Where possible, i.e., where the data allow, future assessments should consider developing spatial and spatiotemporal standardizations of abundance index time series.

  In addition, for any data processing that included age ogives (e.g., mean length-at-age, maturity-at-age, etc.), the processes could be improved by incorporating the ageing error matrix (Section 2.6 in Data Workshop Report). The matrix was found to exhibit a small linear bias in ageing in addition to imprecision. These properties, upon consideration in the modeling, could improve growth and maturity estimation. (I do realize the assessment team ended up choosing maturity-at-length).

- ***Are input data series sufficient to support the assessment approach?***

  The available input data series are certainly sufficient to support an integrated age-structured model using Stock Synthesis. The assessment used standard data inputs required for an age-structured model, including landings, discards, relative abundance indices, length compositions, and conditional age-at-length compositions. In addition, the fixed inputs of mean length-at-age (not used), ageing error, and maturity-at-length were largely standard and sufficient for the integrated age-structured assessment.

  The shrimp-fishery bycatch of Triggerfish was highly uncertain, as back-calculation based on effort and linear interpolation were used to obtain estimates back to 1945 (among other limitations). These limitations do not prevent their inclusion in the integrated assessment model, but they reduce the reliability of the information used to estimate some key population and fishery processes. It is my understanding that the shrimp bycatch estimation procedure is currently under improvement. Any procedures to improve the understanding of historical removals of Gulf Gray Triggerfish would certainly improve the stock assessment.

Summary

Overall, I found the data used and the processing of them prior to use in the stock assessment to be largely sound and consistent with regional standards. A commendable amount of effort went in to re-ageing fish in the assessment according to a revised protocol which greatly decreased the bias in age reading (although did not totally eliminate it). All of the removals time series were re-analyzed since the last assessment to come up with the best point estimates and uncertainties. Lastly, all of the life history information was updated (in tandem with new ages), as were abundance indices and composition datasets.

I do find that in many aspects of processing the data prior to use in the stock assessment, procedures could be improved. I reiterate my main recommendations:

- I recommend the use of the ageing bias relationship in the assessment model and prior data-preparatory models. The assessment included the ageing imprecision, but not the bias into its formulation. I recommend including the bias as well in the future. In addition, this ageing error matrix could be used in growth, maturity, and fecundity models prior to use in the assessment.
- I recommend the use of bootstrapping to estimate more objective input sample sizes for composition datasets. Stewart and Hamel (2014) provide a very good blueprint to conduct such bootstraps.
- I recommend the consideration of spatial and spatiotemporal generalized linear mixed-model standardizations of abundance indices and composition datasets.
- I recommend processing the data in the future such that there exist length compositions, age compositions, and conditional age-at-length compositions. This would allow for evaluating the impact of fitting conditional age-at-length data via substituting in age compositions in years that were previously fitting length and conditional age-at-length data. (In years without age compositions, the model would still be fitting length compositions.) One of the concerns with conditional age-at-length data is that it spreads the data quite thin (few observations per length bin, Stewart and Monnahan, 2025) and one way to evaluate this may be to substitute in full age compositions in those years by removing the corresponding length and conditional-age-at-length compositions (just for years with age data).

2. ***Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions. Consider the following:***

    - ***The context of the data availability in the region. Are the procedures appropriate for our data limitations?***
    - ***Are data decisions made by the DW justified?***
    - ***Are data uncertainties acknowledged, reported, and within normal or expected levels?***
    - ***Is the appropriate model applied properly to the available data?***

- ***Are input data series sufficient to support the assessment approach?***

This TOR is equivalent to #1. As such, please refer to my comments on TOR 1.

3. ***Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, considering the available data. Consider the following:***
    - ***Are methods scientifically sound and robust?***

The methods used to assess the stock of Gulf of America Gray Triggerfish for SEDAR 100 are largely sound. The stock assessment used an assessment program commonly used in the United States, Stock Synthesis 3, to fit an age-structured assessment model to data on landings, discards, indices of abundance, length compositions, and conditional age-at-length compositions from several fleets. The authors evaluated a broad suite of model diagnostics including examining fits to the data, retrospective analyses, hindcasting, likelihood profiles, jitter analysis, and jackknife analyses in common with many assessments in the region and the United States. Results from some model diagnostics were certainly suboptimal, such as fits to the composition data and some of the fits to indices of abundance, however, in the context of the data quality and similar assessments in the region I find them acceptable.

I do have some concerns in the methodology for the assessment. Primarily, the stock assessment fixed both steepness (*h*) and natural mortality (*M*), and this fixes the two main axes determining the productivity of the stock. Shown at the review meeting were a suite of additional analyses regarding assessment performance, when *h* and *M* were estimated or fixed. When *h* was fixed at 0.99 and *M* was estimated, the model converged on a value for M of 0.076 without a prior and 0.128 using the prior from Hamel and Cope (2022), each of which appear biologically implausible given the observed longevity of the species. In light of this, I understand that fixing *M* at 0.36 (the Hamel and Cope (2022) estimate) is a defensible assumption. However, when *h* was estimated and *M* fixed at 0.36, the assessment converges on a value of $h = 0.422$ with a negative log likelihood solution lower by approximately 22 units. This topic was discussed at length in the review meeting, where we were shown model diagnostics and sensitivity to key assumptions of each assessment configuration. The model that estimated steepness exhibited improved retrospective diagnostics, fits to data, and largely similar hindcast predictions and jitter stability. Notably, the model that estimated *h* exhibited increased robustness to some key recruitment assumptions, such as applying the sum-to-zero constraint for recruitment deviations. In the SEDAR 100 AP model with *h* fixed at 0.99, the assessment was highly sensitive to relaxing the sum-to-zero constraint, estimating markedly different recruitment deviations (that summed to -14 in log space) in relation to when the constraint was applied. This suggests the recruitment deviations, when unconstrained, are accounting for some process error unaccounted for in the model rather than conforming to their specified normal distribution (with mean 0 in log space). In contrast, the assessment model that estimated *h* did not exhibit sensitivity to the sum-to-zero constraint.

In light of the improved diagnostics and robustness, I recommend the assessment adopt a base model where *h* is estimated internally (RW model). This model has the additional advantage of integrating the uncertainty in the *h* estimation into the model and its derived

outputs. I do stress however, that the paired SSB-recruitment estimates from the RW model occur only over a very small range of SSB values for the stock-recruitment curve in the time period when recruitment deviations are estimated, and in particular when SSB is low (Figure 1). This suggests the part of the stock-recruit curve at higher levels of SSB is not well defined. Therefore, I do not consider this *h* estimate useful for deriving management benchmarks or projections. Rather, I find the estimate of *h* to be a computational convenience leading to a more stable model that better fits the observed data, and consequentiy estimates a more likely historical time series.

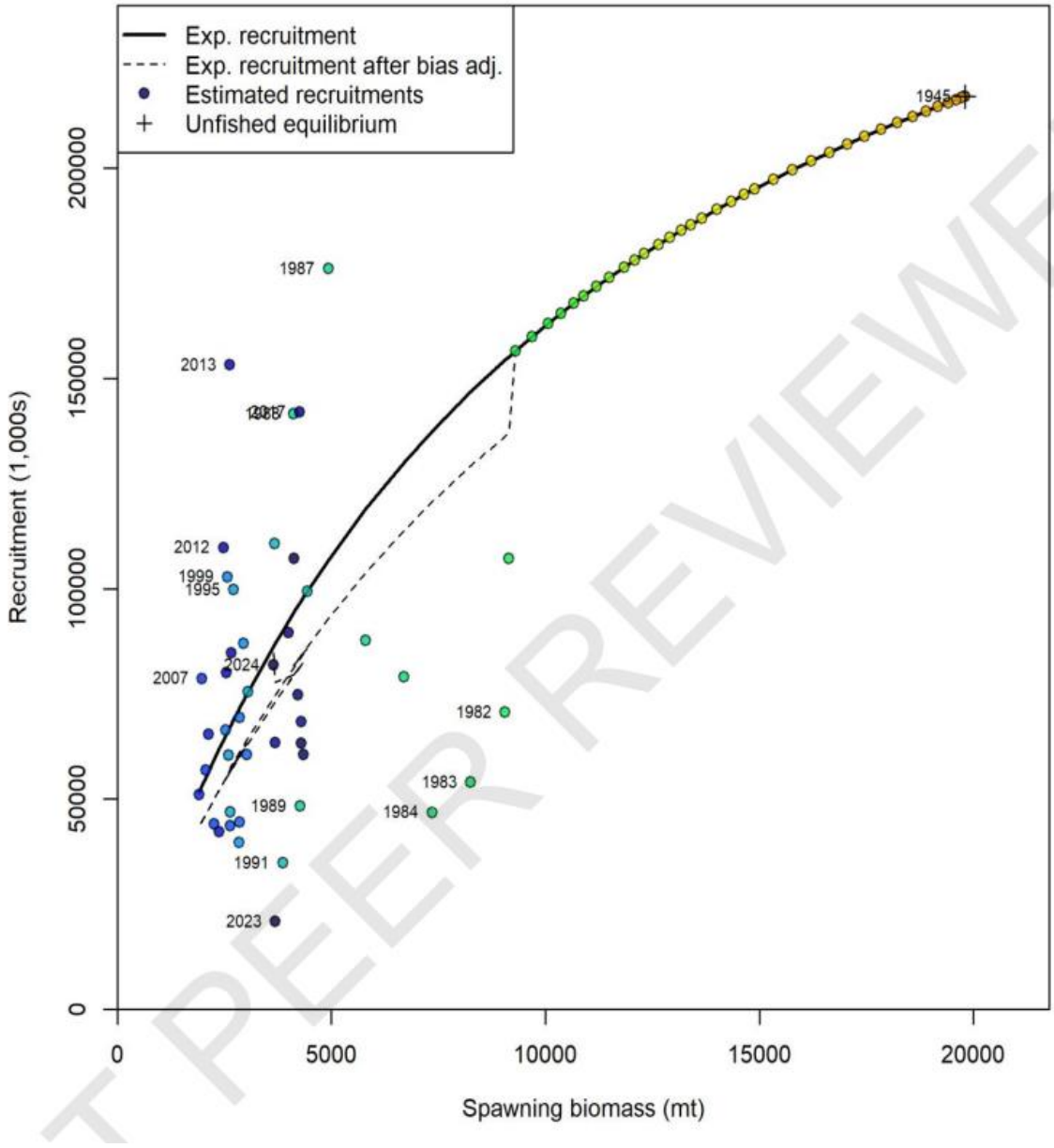


Figure 1. Reproduction of Figure 112 from the assessment document. Predicted stock-recruit relationship for the RW model, which estimated steepness without a prior. Plotted are predicted annual recruitments from Stock Synthesis (circles), expected recruitment from the stock-recruit relationship (black line), and bias adjusted recruitment from the stock-recruit relationship (dashed line).

T PEER REVIEWED

The provided diagnostics for the assessment model included Pearson residuals for the composition datasets. Unfortunately, these residuals are improper for composition datasets due to the fact that compositions must sum to 1 (thus each observation is correlated and violates Peason assumptions). For this reason, the composition fits for this assessment were very difficult to properly evaluate. I strongly recommend future assessments produce One step-ahead (OSA) residuals as diagnostics for evaluating fits to the composition datasets (Trijoulet et al., 2023). In addition, recent research (Stewart and Monnahan, 2025) has proposed some promising objective criteria to evaluate whether or not OSA residuals are showing good or poor fits to composition datasets. I strongly recommend the adoption of criteria from Stewart and Monnahan (2025) in future assessments when evaluating OSA residuals for composition datasets.

Root mean squared errors (RMSEs) were presented as a diagnostic for fits to the indices. These statistics are informative; however, they do not tell the analyst how well a model is fitting relative to expectations (i.e., what is set or estimated as the uncertainty in the data). I recommend in the future that the standard deviation of the normalized residuals (SDNRs) be presented as well as the RMSEs. The SDNRs should allow authors and reviewers to examine if a model if fitting as well as expected.

- ***Are priority modeling issues clearly stated and addressed?***

Several priority modeling issues were identified and addressed. The assessment authors did address the priority modeling issue of when to begin the assessment model. They examined a start date in 1945 in an unfished state, and two others in fished states: one in 1963 and another in 1981. Although the models had different starting estimates, their SSB time series converged to be very similar to the model that started in 1945 (and thus were largely insensitive to this assumption). In addition, when the models with different start dates attempted to estimate *h*, the resulting estimates of steepness were largely similar. Given the constraints of Stock Synthesis, I find the start date of 1945 in an unfished state to be defensible given the other two start dates required estimating highly uncertain equilibrium initial *F* parameters. An alternative approach in another assessment program that allows for mixed-effects would be to utilize random effects to aid the model in estimating initial numbers at age, recruitment estimates for early years (not fixed to SR curve), or fishing mortalities for early years where data interpolation was necessary. These functionalities may also be built into the Fisheries Integrated Modeling System (FIMS) in the future.

The assessment authors re-evaluated multiple assumptions that were made in the previous Gulf Gray Triggerfish assessment of SEDAR 43. The previous assessment assumed a fleets-as-areas (East/West) structure. The authors noted the data imbalance by region (composition data overwhelmingly from the East); however, they presented evidence that growth and selectivity did not differ between the regions, supporting the modeling choice in SEDAR 100 of a single, Gulf-wide model (and Gulf-wide data sources from each fleet). The authors also showed, in contrast to previous studies (Chamberlin et al., 2025), that there was no sexual dimorphism evident in growth of Gulf Gray Triggerfish. In light of this, a single sex-structure was used in the assessment. I support the decision and rationale for these choices. These are both data and assessment decisions; as such, they are listed in each section.

The assessment authors chose to model length-based selectivity, maturity, and fecundity in contrast to the previous assessment. This decision was largely on account of the sizable level of ageing error expected from the age data. The authors noted that although a new protocol was implemented for ageing triggerfish that greatly reduced bias, uncertainty in ageing remains a key challenge. I support modelling length-based processes where possible in the assessment and recommend that improvements in ageing remain a key research priority for Gray Triggerfish.

The assessment authors chose to fix the standard deviation of recruitment at 0.6, under the rationale that it could not be estimated and that 0.6 reflected a realistic standard deviation in recruitment estimates. I note that meta-analyses of recruitment variation (Thorson et al., 2014) typically show the parameter above 0.6 (notably Thorson et al. (2014) did not evaluate *Tetraodontiformes*). In addition, the likelihood profile suggests more support for the parameter at higher values. I recommend future assessments explore additional sensitivity evaluations to this parameter at higher levels. This is another area where mixed-effects models may improve estimation, as they are better able to estimate variance terms compared to penalized maximum likelihood estimation (Thorson, 2019).

- ***Are the methods appropriate for the available data?***

The methodology was largely appropriate for the available data. Again, moving from a penalized maximum likelihood framework such as Stock Synthesis to a state-space assessment framework such as the Woods Hole Assessment Model (WHAM), State-space Assessment Model (SAM), or the Fisheries Integrated Modeling System (FIMS), may help in multiple aspects of the assessment. These include, but are not restricted to, limiting concerns regarding the model starting year (e.g., estimating initial numbers at age, or initial recruitment deviations), incorporating environmental linkages, and estimating variance terms (e.g., standard deviation in recruitment).

One of the model diagnostics used was a jitter analysis, changing initial parameter estimates and re-running the model (100 times). Many jitter iterations came to local minima suggesting the model is sensitive to starting parameters (and not as robust as would be ideal). The RW model did however come to the solution with the lowest negative log likelihood. This type of robustness issue could be further explored in the future by conducting self-test simulations to ensure the model can reproduce parameter estimates in a correctly specified scenario (given the data available).

- ***Are assessment models configured properly and used in a manner consistent with standard practices?***

The assessment models are largely configured properly and used in a manner consistent with standard practices for the region and the United States. There are some aspects that could be improved, such as:

Incorporating ageing error bias directly into the Stock Synthesis assessment model. The data workshop recommended utilizing an ageing error matrix that includes both ageing

imprecision in terms of standard deviations at age and also a slight positive bias in ageing. The assessment model included in the report incorporated the imprecision into the Stock Synthesis assessment program, but not the bias. I recommend future assessments consider incorporating the bias in age reading as well into their definition of ageing error in the stock assessment.

Utilization of a prior distribution for natural mortality rather than a fixed point estimate. The stock assessment utilized the point estimate of *M* from the longevity-based methods of Hamel and Cope (2022). The reality is that Hamel and Cope (2022) provide a distribution of *M* given a maximum age estimate. This distribution is lognormal with mean log(5.4/tmax) and standard deviation of 0.31, both in log space. This advice would be more properly incorporated into the assessment as a prior or a penalty term (in maximum likelihood parlance), rather than a fixed value. This would have the advantage of still allowing the assessment model to estimate *M*, and to incorporate uncertainty in its estimation into the model and its derived outputs.

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

- ***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.***
- ***Comment on the relationship of this variability with ecosystem or climate factors and mechanisms for encompassing this into management reference points.***

The method used to evaluate uncertainty in the Gray Triggerfish stock assessment model is the default in Stock Synthesis and the standard across many integrated assessments in the region and United States. The assessment quantifies model uncertainty by inverting the Hessian matrix at the maximum likelihood estimates to get the variance-covariance matrix of the estimated parameters. Standard errors for derived quantities are then obtained using the delta method. This approach, although standard, suffers from some drawbacks:

- It is asymptotic, thus only valid when the number of observations informing each parameter is very high (a situation rarely met in stock assessments).
- Priors or penalty terms are included on many parameters (Dirichlet-multinomial parameters, recruitment deviations) meaning the reported uncertainty from the Hessian is not necessarily frequentist and conforming to the efficiency property under the theory of maximum likelihood estimation.

Some alternative methods could be included in the future in order to better capture uncertainty in the assessment model. These include Bayesian Markov Chain Monte Carlo (MCMC) methods and bootstrap simulations. MCMC techniques, via sampling the posterior distribution, can be used to quantify the uncertainty of said

distribution, and notably incorporate priors directly into the theory and calculation (in contrast to penalized maximum likelihood). Parametric bootstraps, where a large number of data sets are simulated from the model, and each is used to re-estimate all parameters, renders a distribution of the estimates to quantify uncertainty. The other alternative is a non-parametric bootstrap, where the original data are resampled with replacement, and the assessment refit *N* times (again producing distributions of parameters and derived values). Although each method is quite computationally intense, if time permits, I recommend their exploration in future assessments.

In addition to the drawbacks of penalized maximum likelihood and using the Hessian coupled with the delta method to get uncertainty, several key parameters in the assessment were fixed (CV length at age, recruitment variation, natural mortality scalar, etc.). Fixing parameters removes the uncertainty in those parameters from propagating into the uncertainty of derived values (not to mention uncertainty in the other estimated parameters). This understates the total model uncertainty. There are several ways to mitigate this. Where possible, such as for the fixed *M* scalar parameter or recruitment variation, utilizing a prior distribution to inform the parameter rather than fixing it can aid in its estimation. Another approach is to evaluate this uncertainty through sensitivity models, e.g., fitting assessment models with several fixed values of *M*, and presenting resulting advice to managers and acknowledging the uncertainty. Finally, one could produce an ensemble as the outcome of the assessment. To take just the natural mortality example, fit assessment models with 3 fixed estimates of *M*, and combine their results in final advice to management.

The Dirichlet-multinomial (DM) distribution for compositional data, although an improvement over previous iterative re-weighting procedures, does not describe the compositional observations in the most appropriate manner. The DM can only produce negative correlations between age bins, whereas the residuals appear to show that most compositions are positively correlated with neighboring classes. The DM not accounting for the positive correlations increases the weight given to the composition dataset (as it cannot down weight on account of the positive correlations). This also affects the uncertainty, as the model will have too high of an effective sample size and will be artificially certain. An alternative to the DM would be to fit the data with the logistic-normal which allows for a suite of correlation patterns in residuals (Fisch et al., 2023) although is not available in the Stock Synthesis program.

In addition to model uncertainty estimates, a reasonable set of sensitivity runs were presented (including scenarios on recreational removals, discard mortality, bycatch in the shrimp fishery, growth platoons, steepness, starting year, and leaving out indices). Most sensitivity runs showed similar overall trends, but some, such as using growth platoons, exhibited a considerable axis of uncertainty. Incorporating growth platoons resulted in the model estimating a decrease in SSB near the end of the time series. I think the total uncertainty could have been better captured by having some more sensitivity runs, such as with different values of some key fixed parameters like natural mortality or recruitment variation.

The projections and management reference points did not appear to incorporate uncertainty. A better approach than deterministic projections would be to incorporate uncertainty in some fashion in order to give managers probabilistic statements. This could be achieved in several ways. The first would be drawing from the joint precision matrix at the MLE (using normality and the Hessian). Secondly and thirdly, one could use MCMC or bootstrapping methods as discussed earlier.

The projections also did not directly include ecosystem or climate factors. There were several comments during the review meeting that there was an active or desired area of research to improve the understanding of how or if sargassum affects Gray Triggerfish recruitment, and to build such an index into the assessment model and projections if it were to improve predictions of recruitment. I support this research recommendation.

5. ***Provide, or comment on, recommendations to improve the assessment.***

- ***Consider the research recommendations provided by the Data Workshop in the context of overall improvement to the assessment and make any additional research recommendations warranted.***

  I think all of the recommendations made by the Data Workshop will lead to an improved data stream to eventually improve the stock assessment. Of those listed in the report, I would prioritize:

  - Explore epigenetic ageing as an approach to improve confidence in age-class assignments. This would help in many aspects of the assessment, notably uncertainty in age compositions, sex-specificity, growth, maturity, fecundity, among others.
  - Conduct field experiments to improve calibrations of SEAMAP and commercial shrimp trawls. There was much uncertainty regarding the shrimp bycatch estimates, and any calibration data would help immensely in this regard.
  - Implement tagging studies to estimate discard mortalities from recreational fisheries as well as additional sources of mortality, including shrimp trawl bycatch. Discarding is a major part of this fishery and population, and getting accurate discard mortality information for Gray Triggerfish would certainly aid the assessment.
  - Conduct research on the potential connection between sargassum habitat and year class strength of Gray Triggerfish. Any predictors for recruitment would help the assessment decrease uncertainty in recruitment. Sargassum appears to be the most promising predictor identified.

- ***If applicable, provide recommendations for improvement or for addressing any inadequacies identified in the data or assessment modeling. Describe these***

recommendations in sufficient detail for application and should be practical for short term implementation (e.g., achievable within ~6 months). List longer-term recommendations as research recommendations above.

- I recommend the use of the ageing bias relationship in the assessment model and prior data-preparatory models. The assessment included the ageing imprecision, but not the bias into its formulation. I recommend including the bias as well in the future. In addition, this ageing error matrix could be used in growth, maturity, and fecundity models prior to use in the assessment.
- I recommend the use of bootstrapping to estimate more objective input sample sizes for composition datasets. Stewart and Hamel (2014) provide a very good blueprint to conduct such bootstraps.
- I recommend the consideration of spatial and spatiotemporal generalized linear mixed-model standardizations of abundance indices and composition datasets.
- I recommend processing the data in the future such that there exist length compositions, age compositions, and conditional age-at-length compositions. This would allow for evaluating the impact of fitting conditional age-at-length data via substituting in age compositions in years that were previously fitting length and conditional age-at-length data. (In years without age compositions, the model would still be fitting length compositions). One of the concerns with conditional age-at-length data is that it spreads the data quite thin (few observations per length bin, Stewart and Monnahan, 2025) and one way to evaluate this may be to substitute in full age compositions in those years while removing the corresponding length and conditional-age-at-length compositions (just for years with age data).
- For the current assessment, I recommend estimating steepness and fixing natural mortality at the Hamel and Cope (2022) estimate, as discussed at length in the review meeting.
- I recommend future assessment diagnostics include OSA residuals for fits to the composition datasets (along with objective criteria), and SDNRs for fits to the indices.
- I recommend future assessments utilize priors rather than fixing parameters, for example with natural mortality and recruitment variation.
- I recommend more sensitivity run be explored to better characterize total uncertainty such as with different values of some key fixed parameters like natural mortality or recruitment variation.
- I recommend consideration of MCMC or bootstrap methods to better quantify uncertainty in the assessment and projections.
- I recommend the recalculation of reference points using a form of empirical recruitment. This can be achieved by estimating F30%SPR through per-recruit analysis and calculating SSBF30%SPR using the mean or other justified central recruitment measure from the most recent 10 well-estimated cohorts, without applying the estimated Beverton–Holt relationship outside the observed SSB range.

- I recommend rerunning the projections using the recent empirical recruitment values. Use a selected recruitment period mean (5-, 10-, 15-yr) for the deterministic projection and resample recruitment from those cohorts for the stochastic projection.
- Evaluate the sensitivity of projections to the recruitment-period. Compare recent 5-, 10-, and 15-year recruitment periods and report their effects on projection results.
- Conduct a management strategy evaluation (MSE) that evaluates alternative empirical-recruitment windows (5-, 10-, 15-yr) and the F30%SPR control rule. The MSE can examine whether using recent recruitment to scale SSBF30%SPR and generate forecasts remains robust under assessment and catch implementation error.

6. ***Evaluate the stock projections, including discussing strengths and weaknesses, and consider the following:***

As noted in the assessment section, the assessment authors presented a suite of additional analyses regarding estimating or fixing steepness and natural mortality. After reviewing the analyses, I, in consensus with the review panel, support a revised base model in which natural mortality is fixed at (M=0.36) and h is estimated rather than fixed at 0.99. The additional analyses consistently supported a substantially lower h, with the revised assessment model estimating h at approximately 0.42. This change improves the historical assessment as it allows the stock–recruitment relationship to be informed by the data instead of imposing an assumption of effectively constant recruitment in the early period.

However, I, again in consensus with the review panel, do not recommend using the estimated h directly in projections or equilibrium reference-point calculations. The assessment contains observations over a limited range of SSB (Figure 1), while projections built on the fitted Beverton–Holt curve would need to extrapolate well past that range. When this extrapolation was incorporated into the projections, it generated results that were difficult to justify both biologically and operationally — notably, there were unrealistically low recruitment forecasts resulting in an inability of the projected stock to hold at the target SPR30% equilibrium.

The additional analyses produced by the assessment authors clarified several important results which helped clarify the rationale for moving away from a fixed h = 0.99:

- Estimating h with a fixed M at 0.36 produced a value near 0.42.
- The likelihood profile of h had a minimum near 0.43, with a confidence interval range of approximately 0.39–0.47.
- The model beginning at a later date (in 1963) estimated h = 0.42, and further the model that started even later produced an estimate of h near 0.51.
- Removing recreational compositions and refitting the model produced an h estimate near 0.45.

- Joint estimation of *M* and *h* produced estimates of *M* near 0.366 and *h* = 0.431 (this model was very unstable).
- Estimating *h,* relative to fixing it at 0.99, reduced the negative log-likelihood by approximately 22 units.
- The model that estimated *h* improved the retrospective pattern in recruitment that was present in the original AP assessment model with a fixed *h*.
- The equilibrium calculation from the model that estimated *h* could not maintain the intended 30% SPR condition. Projected recruitment became too low as SSB moved outside the historically informed range.

Collectively, these results support estimating *h* in the historical model, but do not demonstrate that the resulting stock–recruitment curve can be projected reliably outside the observed SSB range (from the period where recruitment deviations were estimated in the assessment model).

Otherwise, the projection framework is technically appropriate and consistent internally with the Stock Synthesis model. The projections are informative for comparing management scenarios. However, as is common, they are conditional on several influential assumptions and should not be interpreted as precise forecasts. I support the use of the projection framework for management, provided that alternative recruitment scenarios (see ideas below), stochastic uncertainty, and major structural sensitivities are also presented.

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

The projection procedures are standard, commonly used Stock Synthesis procedures. Use of recent selectivity, retention, discard mortality, interim catches, sector allocation, and an SPR-based fishing-mortality target is transparent and operationally reasonable. However, the primary projections are deterministic and do not fully represent the uncertainty evident in the assessment. I recommend verifying that the 21:79 commercial–recreational allocation is compatible with current MRIP-FES catch accounting and evaluating alternatives to the assumed 2025–2027 catches, fixed shrimp-bycatch mortality, and constant future fishery patterns.

I find the projection mechanics acceptable; however, recruitment should not be generated from either the former fixed *h* stock-recruit relationship or an extrapolation of the newly estimated *h* in a stock-recruit curve. Projections should ideally use a form of drawing from recent recruitments empirically (I elaborate on this below). I find the methods acceptable as standard deterministic projection procedures but additional uncertainty analyses are needed for risk-based management advice.

- ***Are the methods appropriate for the assessment model and outputs?***

The Stock Synthesis forecast process is appropriate and maintains consistency with the population structure, biological parameters, fleets, and selectivity estimated by the assessment. The *F*30%SPR framework is a reasonable basis for projections under concerns about productivity estimation. The primary concern is the specification of recruitment with *h*=0.99. The *h*=0.99 value should therefore be considered as a projection convention rather than an empirically supported biological estimate. The framework is appropriate, but a single recruitment specification, and in particular one assuming a stock-recruit curve with *h*=0.99, is not sufficient. Long-term mean, recent mean and low-recruitment scenarios should be explored.

- ***Are the results informative and robust, and are they useful to support inferences of probable future conditions?***

  The additional analyses demonstrate that projected stock conditions and reference points are highly sensitive to how recruitment is specified.

  The estimated steepness produced different reference points and stock status when the fitted stock–recruitment curve was extrapolated. These results should not be interpreted as reliable evidence of future conditions because they depend on recruitment predictions outside the historically informed SSB range.

  Projections based on recent recruitment estimated from the stock assessment model would be more informative about probable near-term conditions, provided that:
  - The recruitment averaging period is justified;
  - Only sufficiently estimated cohorts are included;
  - Recruitment variability is retained; and
  - Projected SSB is monitored relative to the SSB range used to estimate the historical relationship.

  The most recent 10-year period is a reasonable starting choice, because it balances relevance to current productivity against sensitivity to a few anomalous cohorts. Nevertheless, the assessment authors should evaluate alternative periods, such as recent 5-, 10-, and 15-year periods, and select the window using objective criteria.

  I conclude that projections can be useful for management, but only after recruitment is reformulated empirically. Results based on extrapolating the estimated steepness should not be used as the principal basis for catch advice.

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

  The assessment acknowledges many important uncertainties, particularly those related to recruitment and alternative catch assumptions. However, these uncertainties are not adequately reflected in the numerical projection results.

The primary projections do not fully incorporate recruitment process variation, parameter covariance, natural mortality, growth, discard mortality, shrimp bycatch, video-index structure, model selection, or catch implementation error.

The additional analyses substantially improved understanding of recruitment uncertainty. They demonstrated that:

- Fixing *h* at 0.99 was inconsistent with the likelihood information;
- *h* can be estimated within the revised historical model;
- The estimate is reasonably consistent across several sensitivity runs;
- *h* and natural mortality can be confounded; and
- The fitted relationship is unreliable when extrapolated outside the observed SSB range**.**

These uncertainties must now be reflected directly in the projection design. Using a single deterministic recent mean would avoid stock–recruit extrapolation, but would still understate recruitment variability. The preferred forecasts should draw recruitment from an empirical distribution of recent, well-estimated cohorts.

The projection report should show:

- The years and cohorts included in the recruitment period;
- Arithmetic, geometric mean, and median recruitment used;
- Variability among years;
- Sensitivity to alternative recruitment periods; and
- Annual projection intervals.

Recommendations

1. Recalculate reference points using a form of empirical recruitment. Estimate F30%SPR through per-recruit analysis and calculate SSBF30%SPR using the mean or other justified central recruitment measure from the most recent 10 well-estimated cohorts, without applying the estimated Beverton–Holt relationship outside the observed SSB range.
2. Rerun the projections using the recent empirical recruitment. Use the selected recruitment period mean for the deterministic projection and resample recruitment from those cohorts for the stochastic projection.
3. Evaluate the sensitivity of projections to the recruitment-period. Compare recent 5-, 10-, and 15-year recruitment periods and report their effects on projection results.

Research recommendation

Evaluate alternative empirical-recruitment windows and the F30%SPR control rule through a management strategy evaluation (MSE). The MSE can examine whether using recent recruitment to scale SSBF30%SPR and generate forecasts remains robust under assessment and catch implementation error.

7. ***Prepare a Review Workshop Summary Report describing the Panel's evaluation of the stock assessment and addressing each Term of Reference.***

This report is my individual report of the SEDAR 100 Gulf of America Gray Triggerfish Stock Assessment. A summary report highlighting overall conclusions and recommendations has been prepared jointly by all members of the review panel.

References

Chamberlin, D.W., Siders, Z.A., Potts, J.C., Rogers, W.D., Taylor, M.A. and Patterson III, W.F., 2025. Bayesian estimation of von Bertalanffy growth parameters for gray triggerfish, Balistes capriscus, incorporating multiple readers and ageing structures. *Canadian Journal of Fisheries and Aquatic Sciences*, *82*, pp.1-12.

Fisch, N., Shertzer, K., Camp, E., Maunder, M. and Ahrens, R., 2023. Process and sampling variance within fisheries stock assessment models: estimability, likelihood choice, and the consequences of incorrect specification. *ICES Journal of Marine Science*, *80*(8), pp.2125-2149.

Hamel, O.S. and Cope, J.M., 2022. Development and considerations for application of a longevity-based prior for the natural mortality rate. *Fisheries Research*, *256*, p.106477.

Stewart, I.J. and Hamel, O.S., 2014. Bootstrapping of sample sizes for length-or age-composition data used in stock assessments. *Canadian journal of fisheries and aquatic sciences*, *71*(4), pp.581-588.

Stewart, I.J. and Monnahan, C.C., 2025. Diagnosing common sources of lack of fit to composition data in fisheries stock assessment models using one-step-ahead (OSA) residuals. *Canadian Journal of Fisheries and Aquatic Sciences*, *82*, pp.1-13.

Thorson, J.T. and Haltuch, M.A., 2019. Spatiotemporal analysis of compositional data: increased precision and improved workflow using model-based inputs to stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, *76*(3), pp.401-414.

Thorson, J.T., Jensen, O.P. and Zipkin, E.F., 2014. How variable is recruitment for exploited marine fishes? A hierarchical model for testing life history theory. *Canadian Journal of Fisheries and Aquatic Sciences*, *71*(7), pp.973-983.

Thorson, J.T., Shelton, A.O., Ward, E.J. and Skaug, H.J., 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES Journal of Marine Science*, *72*(5), pp.1297-1310.

Thorson, J.T., 2019. Perspective: Let's simplify stock assessment by replacing tuning algorithms with statistics. *Fisheries Research*, *217*, pp.133-139.

Trijoulet V., Albertsen C.M., Kristensen K., Legault C.M., Miller T.J., Nielsen A. 2023. Model validation for compositional data in stock assessment models: calculating residuals with correct properties. Fish. Res. **257**: 106487.

Appendix 1.

SEDAR 100 Document List

Document #	Title	Authors
Documents Prepared for the Data Workshop		
SEDAR100-DW-01	Headboat Data for Gray Triggerfish in the Gulf of America	Rob Cheshire
SEDAR100-DW-02	General Recreational Survey Data for Gray Triggerfish in the Gulf of America	Samantha M. Binion-Rock
SEDAR100-DW-03	Historical (1955-1980) Recreational Landings for Gulf of America Gray Triggerfish (Balistes capriscus) estimated using the FHWAR Census Method	Samantha M. Binion-Rock
SEDAR100-DW-04	Gulf of America Gray Triggerfish (Balistes capriscus) length and age compositions from the recreational fishery	Samantha M. Binion-Rock
SEDAR100-DW-05	Commercial Landings of Gulf of America Gray Triggerfish (Balistes capriscus) from 1949 - 2024	Micki Pawluk and Sarina Atkinson
SEDAR100-DW-06	CPUE Expansion Estimation for Commercial Discards of Gulf of America Gray Triggerfish	Kevin Thompson, Sarina Atkinson, Gary Decossas
SEDAR100-DW-07	Gulf of America Gray Triggerfish (Balistes capriscus) Length and Age Compositions for the Commercial Handline and Longline Fisheries	Micki Pawluk
SEDAR100-DW-08	Proxy Discard Estimates of Gray Triggerfish (Balistes capriscus) from the US Gulf of America Headboat Fishery	Nuttall, MA
SEDAR100-DW-09	Standardized Catch per Unit Effort for US Gulf of America Gray Triggerfish (Balistes capriscus) from the Southeast Region Headboat Survey	Nuttall, MA
SEDAR100-DW-10	Social Dimensions of Gray Triggerfish (Balistis capriscus) in the Gulf and South Atlantic	David Griffith
SEDAR100-DW-11	A Summary of Gulf Gray Triggerfish Discard Length Data Collected from At-Sea Observers in For-Hire Fishery Surveys in Florida 2005-2024	Ellie Corbett
SEDAR100-DW-12	A ratio-based method for calibrating estimates of total landings (numbers and pounds of fish), releases (numbers of fish), and total trips from MRIP-FCAL to SRFS for Gulf Gray Triggerfish (Balistes capriscus)	Maria McGirl and Chloe Ramsay
SEDAR100-DW-13	Length and age compositions of Gulf Gray Triggerfish, Balistes capriscus, collected in association with fishery-dependent projects	Maria McGirl, Jessica Carroll, and Bridget Cermak
SEDAR100-DW-14	A Review of Gray Triggerfish (Balistes capriscus) Age-length Data in the Gulf of America, 1999-2024	Ashley Pacicco, Steve Garner, Ryan Nichols
SEDAR100-DW-15	Standardized catch rates of Gray Triggerfish from the United States Gulf of America commercial handline fishery, 1993-2024	Kevin Thompson and Michaela Pawluk
SEDAR100-DW-16	Gray Triggerfish Abundance Indices from SEAMAP Groundfish Surveys in the Northern Gulf of America	Adam G. Pollack and David S. Hanisko

SEDAR100-DW-17	Gulf of America gray triggerfish (Balistes capriscus) post-release mortality based on long-term conventual tagging	A. Challen Hyman, Chloe Ramsay, and Thomas K. Frazer
SEDAR100-DW-18	Indices of abundance for Gulf Gray Triggerfish (Balistes capriscus) using data from multiple video surveys	Katherine E. Overly, Heather M. Christiansen, Sean F. Keenan, Theodore S. Switzer, Justin Lewis, Matthew D. Campbell
SEDAR100-DW-19	Gray Triggerfish (Balistes capriscus) indices of relative abundance from SEAMAP Fall Plankton Surveys, 1986 to 2023	David S. Hanisko, Glenn Zapfe, Adam G. Pollack and Denice M. Drass
SEDAR100-DW-20	Public comments received via the online comment form during the data process	
SEDAR100-DW-21	Fisherman Feedback: Gray Triggerfish Response Summary	Gulf Council Staff
SEDAR100-DW-22	Estimation of Gray Triggerfish Bycatch from Gulf of America Shrimp Trawls	Cheston T. Peterson, Carly Daiek, Kyle Dettloff, Steve G. Smith, and Sarina Atkinson
Documents Prepared for the Assessment Process		
SEDAR100-AP-01	Gray Triggerfish Post-Release Mortality (SEDAR 100)	S100 Ad Hoc Discard Mortality Working Group
Final Stock Assessment Reports		
SEDAR100-SAR1	Gulf Gray Triggerfish	SEFSC
Reference Documents		
SEDAR100-RD01	Application of three-dimensional acoustic telemetry to assess the effects of rapid recompression on reef fish discard mortality	Erin Collings Bohaboy, Tristan L. Guttridge, Neil Hammerschlag, Maurits P. M. Van Zinnicq Bergmann, and William F. Patterson III
SEDAR100-RD02	Cooperative Research Program Final Report: Estimating Gulf of Mexico Gray Triggerfish Release Mortality and Its Mitigation with Three-Dimensional Acoustic Telemetry	William F. Patterson III, Derek W. Chamberlin, and Miaya Taylor
SEDAR100-RD03	Bomb 14C validates Gray Triggerfish (Balistes capriscus) dorsal spine and otolith ageing protocols	Derek W. Chamberlin, Jennifer C. Potts, Walter D. Rogers, Zachary A. Siders, William F. Patterson III
SEDAR100-RD04	Bayesian estimation of von Bertalanffy growth parameters for gray triggerfish, Balistes capriscus, incorporating multiple readers and ageing structures	Derek W. Chamberlin, Zachary A. Siders, Jennifer C. Potts, Walter D. Rogers, Miaya A. Taylora, and William F. Patterson, III
SEDAR100-RD05	Low discard survival of gray triggerfish in the southeastern US hook-and-line fishery	Brendan J. Runde, Paul J. Rudershausen, Beverly Sauls, Chloe S. Mikles, Jeffrey A. Buckel
SEDAR100-RD06	Characterization of the U.S. Gulf of Mexico and South Atlantic Penaeid and Rock Shrimp Fisheries Based on Observer Data	Elizabeth Scott-Denton, Pat F. Cryer, Matt R. Duffy, Judith P. Gocke, Mike R. Harrelson, Donna L. Kinsella, James M. Nance, Jeff R. Pulver, Rebecca C. Smith, and Jo A. Williams

Appendix 2.

Performance Work Statement (PWS)
National Oceanic and Atmospheric Administration (NOAA)
NOAA Fisheries
Center for Independent Experts (CIE) Program
External Independent Peer Review

SEDAR 100 Gulf Gray Triggerfish Assessment
August 11 - 14, 2026

Background

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 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 **SouthEast Data, Assessment, and Review (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.

The SEDAR 100 review workshop will be a CIE assessment review of the stock assessment of Gulf gray triggerfish. The review workshop provides an independent peer review of SEDAR stock

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

assessments. The term review is applied broadly, as the review panel may request additional analyses, error corrections and additional runs of the assessment models provided by the assessment panel. The review panel is ultimately responsible for providing advice to ensure the assessment is appropriate for use by fishery managers.

Requirements
NOAA Fisheries requests three reviewers to conduct an impartial and independent peer review in accordance with this Performance Work Statement (PWS), OMB Guidelines, and the ToRs below. Modifications to this Performance Work Statement (PWS) and ToRs cannot be made during the peer review, and the Contracting Officer's Representative (COR) and the CIE contractor shall approve any modifications prior to the peer review. All ToRs must be addressed in each reviewer's report.

The reviewers shall have a working knowledge of stock assessment modeling, 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 fisheries stock assessment. Expertise with Stock Synthesis modeling platform, and the associated model diagnostics would be helpful. Familiarity with evaluating length-and-age composition data, selectivity patterns, and model diagnostics in data-moderate or data-limited contexts is preferred.

The chair, who is in addition to the panel of three 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.

The specified format and contents of the individual peer review report are found in **Annex 1**. The Terms of Reference (ToRs) of the peer and review are included in **Annex 2**. The tentative agenda of this virtual panel review and process review meetings is included in **Annex 3**.

Tasks for Reviewers
Each CIE reviewer shall complete the following tasks in accordance with the PWS and Schedule of Milestones and Deliverables herein.

Pre-review Background Documents: Two weeks before the peer review, the NOAA Fisheries Project Contact will email the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the NOAA Fisheries Project Contact will consult with the CIE on where to send documents. The 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.

Documents to be provided to the CIE reviewers prior to this review include:
S100 Data Workshop Report (https://sedarweb.org/documents/sedar-100-gulf-gray-triggerfish-data-workshop-report/)
S100 Assessment Process Report (to be provided two weeks prior to review)

S100 Working papers and Reference documents (https://sedarweb.org/assessments/sedar-100-gulf-triggerfish/)

Panel Review Meeting: The CIE reviewers shall conduct the independent peer review in accordance with the PWS and ToRs and shall not serve in any other role unless specified herein. The CIE reviewers shall attend and actively participate in a professional and respectful manner as members of the meeting review panel, and their peer review tasks shall be focused on the ToRs as specified herein. The NOAA Fisheries Project Contact is responsible for any meeting arrangements (e.g., conference room reservations, video conferencing logistics, etc.).

Chair Summary Report: The reviewers shall assist the Chair of the meeting with contributions to the summary report.

Contract Deliverables - Independent CIE Peer Review Report: The CIE reviewers shall complete an independent peer review report in accordance with this PWS. Reviewers are not required to reach a consensus. The CIE reviewer shall complete the independent peer review according to required format and content as described in **Annex 1**. The CIE reviewer shall complete the independent peer review addressing each ToR as described in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3**.

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 Project Contact for the purpose of their security clearance, and this information shall be submitted at least two weeks in advance. For additional information, please see the following link: https://www.commerce.gov/osy/programs/foreign-access-management. All CIE reviewers must have a REAL ID-compliant form of identification to access Federally owned or leased facilities The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance
The place of performance shall be Tampa, FL.

Period of Performance
The CIE reviewers' 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:

Schedule	Milestones and Deliverables
Within two weeks of award	Contractor secures participation of the CIE reviewers.
At least two weeks prior to the panel review meeting	Pre-review documents are provided to the reviewers.
August 11 - 14, 2026	The reviewers participate and conduct an independent peer review during the panel review meeting.
Approximately 2 weeks later	Contractor receives draft reports
Within 3 weeks of receiving draft reports	Contractor submits final reports to the Government

* The Peer Review Summary Report will not be submitted to, reviewed, or approved by the Contractor.

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 in **Annex 1**; (2) The reports shall address each ToR as specified **Annex 2**; 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 (Travel resources | GSA), and all contractor travel must be approved by the COR prior to the actual travel. Any travel conducted prior to the receipt of proper written authorization from the COR will be done at the Contractor's own risk and expense. International travel is authorized for this contract. Travel is not to exceed the agreed upon price in accordance with the Price Schedule.

Confidentiality and Data Privacy

This contract may require that services contractors have access to Privacy Information. Services contractors are responsible for maintaining the confidentiality of all subjects and materials and may be required to sign and adhere to a Non-disclosure Agreement (NDA).

NOAA Fisheries Project Contact:

Katie Siegfried, Branch Chief, Gulf Fisheries Branch
NOAA Fisheries Service, Southeast Fisheries Science Center
101 Pivers Island Rd, Beaufort, NC 28516
Kate.Siegfried@noaa.gov

Annex 1: Peer Review Report Requirements

1. The CIE independent report shall be prefaced with an Executive Summary providing a concise summary of the findings and recommendations as specified in the ToR.

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

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

   c. Reviewer should elaborate on any points raised in the Summary Report that they feel might require further clarification.

   d. The CIE independent report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the panel summary report. The CIE independent report shall be an independent peer review of each ToRs and shall not simply repeat the contents of the summary report.

3. The reviewer report shall include the following appendices:

   Appendix 1: Bibliography of materials provided for review
   Appendix 2: A copy of the CIE 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 100 Gulf Gray Triggerfish Assessment

CIE reviewers are contracted to complete their independent peer review based on the ToRs. Therefore, the CIE-NOAA Fisheries review and approval process is based on whether the CIE independent reports addressed each of the ToRs.

1. Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions. Consider the following:
    - The context of the data availability in the region. Are the procedures appropriate for our data limitations?
    - Are data decisions made by the Data Workshop (DW) justified?
    - Are data uncertainties acknowledged, reported, and within normal or expected levels?
    - Is the appropriate model applied properly to the available data?
    - Are input data series sufficient to support the assessment approach?
2. Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions. Consider the following:
    - The context of the data availability in the region. Are the procedures appropriate for our data limitations?
    - Are data decisions made by the DW justified?
    - Are data uncertainties acknowledged, reported, and within normal or expected levels?
    - Is the appropriate model applied properly to the available data?
    - Are input data series sufficient to support the assessment approach?
3. Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, considering the available data. Consider the following:
    - Are methods scientifically sound and robust?
    - Are priority modeling issues clearly stated and addressed?
    - Are the methods appropriate for the available data?
    - Are assessment models configured properly and used in a manner consistent with standard practices?
4. Consider how uncertainties are addressed in the assessment, and their potential consequences.
    - 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.
    - Comment on the relationship of this variability with ecosystem or climate factors and mechanisms for encompassing this into management reference points.
5. Provide, or comment on, recommendations to improve the assessment.
    - Consider the research recommendations provided by the Data Workshop in the context of overall improvement to the assessment and make any additional research recommendations warranted.
    - If applicable, provide recommendations for improvement or for addressing any inadequacies identified in the data or assessment modeling. Describe these

recommendations in sufficient detail for application and should be practical for short term implementation (e.g., achievable within ~6 months). List longer-term recommendations as research recommendations above.

6. Evaluate the stock projections, including discussing strengths and weaknesses, and consider the following:
    - Are the methods consistent with accepted practices and available data?
    - Are the methods appropriate for the assessment model and outputs?
    - Are the results informative and robust, and are they useful to support inferences of probable future conditions?
    - Are key uncertainties acknowledged, discussed, and reflected in the projection results?
7. Prepare a Review Workshop Summary Report describing the Panel's evaluation of the stock assessment and addressing each Term of Reference.

Annex 3: Tentative Agenda for SEDAR 100 Gulf Gray Triggerfish

August 11 – 14, 2026

Tampa, Florida

NOAA Fisheries Project Contact: Katie Siegfried

Tuesday:

8:30 a.m.	**Introductions and Opening Remarks**	**Coordinator**
	- Agenda Review, TOR, Task Assignments	
9:00 a.m. – 12:00 p.m.	**Assessment Presentations**	**Analytic Team**
	- *Assessment Data & Methods*	
	- Identify additional analyses, sensitivities, corrections	
12:00 p.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 4:30 p.m.	**Assessment Presentations (continued)**	**Analytic Team**
	- *Assessment Data & Methods*	
	- Identify additional analyses, sensitivities, corrections	
4:30 p.m. – 5:30 p.m.	**ToR Review and Daily wrap up**	**Chair**
5:30 p.m. – 6:00 p.m.	**Public comment**	**Chair**

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

Wednesday:

8:30 a.m. – 12: p.m.	**Panel Discussion**	**Chair**
	- *Assessment Data & Methods*	
	- Identify additional analyses, sensitivities, corrections	
12:00 p.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 5:00 p.m.	**Panel Discussion/Panel Work Session**	**Chair**
	- Continue deliberations	
	- Review additional analyses	
	- Recommendations and comments	
5:00 p.m. – 5:30 p.m.	**ToR Review and Daily wrap up**	**Chair**

5:30 p.m. – 6:00 p.m.	**Public comment**	**Chair**

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

Thursday

9:00 a.m. – 12: p.m.	**Panel Discussion**	**Chair**
	- *Assessment Data & Methods*	
	- Identify additional analyses, sensitivities, corrections	
12:00 p.m. – 1:30 p.m.	**Lunch Break**	
1:30 p.m. – 4:30 p.m.	**Panel Discussion/Panel Work Session**	**Chair**
	- Continue deliberations	
	- Review additional analyses	
	- Recommendations and comments	
4:30 p.m. – 5:00 p.m.	**ToR Review and Daily wrap up**	**Chair**
5:30 p.m. – 6:00 p.m.	**Public comment**	**Chair**

Thursday Goals: sensitivities and modifications identified, preferred models selected, projection analysis reviewed, Report draft continued

Friday

8:30 a.m. – 11:30 a.m.	**Panel Discussion or Work Session**	**Chair**
	- Final sensitivities reviewed.	
	- Projections reviewed.	**Chair**
	- *Review Reports*	
11:30 a.m. – 12:00 p.m.	**Public comment**	**Chair**
12:00 p.m.		**ADJOURN**

Friday Goals: Complete assessment work and discussions, final base configuration available. Draft Reports reviewed.

Appendix 3.

LIST OF PARTICIPANTS

Review Panel

Luiz Barbieri (Chair)	GFMC SSC
Jie Cao	GFMC SSC
Yong Chen	CIE Reviewer
Nicholas Fisch	CIE Reviewer
Erik Lang	GFMC Appointee
Anders Nielsen	CIE Reviewer

Analytic Team

Skyler Sagarese	SEFSC
Katie Siegfried	SEFSC

Staff

Julie A Neer	SEDAR
Alyssa Clift	GFMC Staff
Lisa Hollensead	GMFC Staff
Matt Freeman	GMFC Staff
John Froeschke	GMFC Staff
Jesscia Matos	GMFC Staff
Ryan Rindone	GFMC Staff
Carrie Simmons	GFMC Staff

Council Representative

Kevin Anson	Gulf Council

Workshop Observers via Webinar

Lisa Ailloud	SEFSC
Shanae Allen	USF/FWC-FWRI
Jordan Bajema	FWC-FWRI

Anessa Carter	FWC-FWRI
Rob Cheshire	SEFSC
Manuel Coffill-Rivera	
LaTreese Denson	SEFSC
Leonardo Eguia	GFMC Staff
Alena Figueroa	FWC-FWRI
Francesca Forrestal	SEFSC
Sarah Gardiner	GFMC Staff
Challen Hyman	USF
Tiffany Hopper	TXPWD
Emily Ott	SEDAR
Chloe Ramsay	FWC-FWRI
Seth Leimgruber	FWC-FWRI
Molly Stevens	SEFSC
James Swarthout	AL DCNR
Nathan Vaughan	Vaughan Analytics
Kym Walsh	LAWLF