Independent peer review report of the SEDAR 68 Atlantic and Gulf of Mexico Scamp assessment review

Prepared for the Center for Independent Experts

Professor Massimiliano Cardinale

November 2021

Executive Summary

- This document is the individual CIE Reviewer report of the SEDAR 68 Atlantic and Gulf of Mexico Scamp assessment conducted during August-September 2021 and provided at the request of the Center for Independent Experts (CIE) (see Appendix 2 aka Attachment A).
- This report solely represents the view of the independent reviewer (Prof. Massimiliano Cardinale). The text in this report is mainly based on the original assessment reports and background documents provided to the reviewer in advance of the meeting. Additional comments based on discussion during the SEDAR 68 web meeting and presentation of alternative model configurations are included in the meeting report.
- The Assessment team tackled all the assigned terms of reference (TORs).
- The reviewer considers that the Assessment team has done a satisfactory job in carrying out the assessment, analysing all available data, modelling part of the uncertainty and providing some sensitivity analyses of both the data and the models. However, the reviewer does not completely agree with all of the findings reported in the SEDAR 68 Atlantic and Gulf of Mexico Scamp assessment report. Taking into account all available information, the reviewer considers that a single "best model" cannot be singled out to be used for advice. Instead, given the uncertainties in the data used as input, and in key biological parameters and processes, a model ensemble should be developed in the future (see also Recommendations section).
- The reviewer also considers that the diagnostics tools used by the Assessment team to evaluate the robustness of the model are incomplete and should be augmented following recent publications (see details in the full Report below).
- Findings that are reported in the SEDAR 68 Atlantic and Gulf of Mexico Scamp assessment report are not necessarily fully repeated in this individual report. This report focuses on clarification of elements contained in the SEDAR 68 Atlantic and Gulf of Mexico Scamp assessment (including the Data Workshop Report and the backgrounds documents) and some additional views of the individual reviewer about how available data could have been better explored to derive more robust estimates of the exploitation rate and stock status of Atlantic and Gulf of Mexico Scamp stocks.
- Further recommendations aimed at improving the assessment of Atlantic and Gulf of Mexico Scamp as presented in the SEDAR 68 were made and included in the full report below. These are mainly based on additional re-analysis and modelling of the original data set made by the reviewer.

Introduction

SEDAR 68 Atlantic and Gulf of Mexico Scamp assessment report, associated background documents containing detailed information on the data used in the assessment and input files of the assessment models were provided to the independent reviewer (Prof. Massimiliano Cardinale) well in advance of the deadline. The reports and documentations were reviewed at the request of the Center for Independent Experts (CIE).

Description of review activities

This review was undertaken by Prof. Massimiliano Cardinale during August-September 2021 at the request of the Center for Independent Experts (CIE) (see Appendix 2 aka Attachment A).

Relevant documents (see bibliography, Appendix 1 aka S68 Doc List) and background information were made available four weeks prior to the deadline through email and via a link to the SEDAR 68 website (http://sedarweb.org/sedar-68-scamp-assessment-process). The assessment report was made available four weeks' prior the deadline via a link to the SEDAR 68 website. The documentation was reviewed prior to the deadline and the deadline was met. The background information and assessment report of Atlantic and Gulf of Mexico Scamp was presented through several documents (see Appendix 1 aka S68 Doc List). Background information relevant to this review is presented in a series of appendices, including: CIE Statement of Work (Appendix 2 aka Attachment A); a bibliography (Appendix 1 aka S68 Doc List) and Terms of Reference (TORs, Annex 2 under Appendix 2 aka Attachment A). Comments included here are provided following the TORs and are those of the independent reviewer only.

Summary of findings

Main recommendations

- The report should include the estimated key biological parameters for each stock of the two species separately. The biology of the pseudo-species in the models should resemble Scamp Grouper, as it represents by far the largest part of the catches, or should be weighed accordingly to their proportion in the catches.

- Regression model standardization procedures that account for an unbalanced sampling between depth and space (and many other covariates) should be preferred for deriving relative abundance indices to be used in stock assessment models.

- The model diagnostic toolbox should be greatly expanded to include as a minimum runs test of the residuals, retrospective and forecast Mohn´s rho, hindcasting and MCMC.

- An ensemble of different plausible model configurations selected using hypothesis testing and weighed by a comprehensive diagnostic against performance criteria agreed beforehand should be used to provide stocks status and management advice for both stocks. As best practice, and as a minimum, the ensemble should integrate the three main sources of uncertainty, process uncertainty, parameter uncertainty and observation error in the data.

- For the Gulf of Mexico Scamp stock steepness cannot be estimated reliably due to the lack of strong contrast in the spawning stock time series. Thus, steepness is either fixed at the species prior estimated by FishLife or, at best, is used as a dimension (e.g, low, med, high) of the ensemble together with others key parameters.

- For the Gulf of Mexico Scamp stock, reference point should be based on SPR or on a fraction of B_0 as MSY derived reference points in conjunction with the high steepness estimated by the model will most likely imply large risk of stock collapse.

Terms of Reference (TORs) (In italics is a condensed answer of the reviewer to each specific TOR; detailed elaborations of each identified issue can be found in the full Report below)

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

- Are data decisions made by the Data Workshop (DW) and Assessment Workshop (AW) 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?

I consider the data used within the chosen assessment models as generally appropriate and data uncertainty properly acknowledged. The models used to conduct the data preparation for the assessments are suitable for the available data as well as the data series are adequate to support the assessment models used. The choice of the analytical tools used to derive the data is well justified in the background documents presented for both stocks. Also, uncertainty associated with the different data sources is estimated and well presented. However, I found the presentation of the available data somewhat incomplete. Concerning the biology of the two species, which are assessed as a pseudo-species for two separate stocks, Gulf of Mexico and Atlantic, it would be beneficial to add, as a minimum, a figure with the estimate of the proportion by year of the two species in the catches. Ideally, it would be informative to include in the reports also the estimated key biological parameters for each stock of the two species separately. All biological parameters are combined in the assessment but it is hard to evaluate to which of the two species the biological parameters used in the assessment are mostly similar. The only information I found about the proportion between the two species is that Yellowmouth Grouper represents at most less than 3% of the catches (e.g., page 8 of the Gulf of Mexico report). In this case, I consider that the biology of the pseudo-species in the models should resemble Scamp Grouper or should be weighed accordingly to their proportion in the catches.

For the Atlantic stock, catches of the commercial fleets were pooled in a single pseudo-fleet and the same was done for the recreational fleets. This is in theory fine if the pooled fleets have similar selectivity and/or if the proportion between the fleets is more or less constant between years. However, it would be useful to add more information to substantiate this choice, especially the catch size composition of the different fleets that compose the commercial and recreational pseudo-fleet and their landings (or proportion) over time.

For the Atlantic stock, I found a reference on page 3 of the assessment report on size regulations to come into place for the first time in 1992 but it would be useful to add some more information on what kind of regulations and to which fleets they were applied over time (see for example Figure 10 in the Gulf of Mexico assessment report).

For both stocks**,** several processes that lead to the estimation of the indices of relative abundance have been modelled using a Delta model (i.e. combining two separate generalized linear models, GLMs for encounter probability and positive catch rates). The proportion of the 0s has the potential to affect the performance of an ad-hoc method such as the Delta model. As a rule of thumb, a Tweedie distribution should be used when the ration is larger than 1/3. In the absence of the information about the proportion of 0s by year in the catch data set used, it is not possible to evaluate how the use of the Delta model might have affected the estimation of the indices of abundance.

The choice of a survey-design ratio-of means estimator instead of a regression model estimator seems to be guided mostly by the unbalance in effort between stratum-area combinations (i.e. depth and space). However, regression model standardization procedures that account for an unbalanced sampling between depth and space (and many other covariates) are widely used for deriving indices of relative abundance to be used in stock assessment models and should be preferred.

Historical catch data, even if not used in the assessment and albeit incomplete, should be presented and included in the report for both stocks for completeness and to verify the assumption for the initial conditions of the models.

2. **Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account 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?

The models (i.e. BAM and Stock Synthesis) used to assess the two stocks are appropriate, robust and in general properly configured, and in line with standard practices. However, some of the choices concerning key parameters or processes are somewhat arbitrary, based on inconsistent analysis and most importantly not substantiated by any model diagnostic. In particular, the choice of the CV associated with the relative abundance indices, the form of the selectivity process, the assumption of the initial conditions and others. As matter of fact, all model configurations are plausible as long as they achieve similar performances when tested against the data and therefore choosing one among all equally plausible configurations will run the risk of "cherry picking" and will affect the final results. This can be avoided by using model diagnostics, which should be greatly expanded to include, as a minimum, runs test of the residuals, retrospective and forecast Mohn´s rho, hindcasting and MCMC.

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

- Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods.
- Comment on the likely relationship of this variability with possible ecosystem or climate factors and possible mechanisms for encompassing this into management reference points.

The major concerns I have for both stocks concern the use of the sensitivity analysis and ensemble and thus how uncertainty is treated in the modelling context and used to provide advice. For the Atlantic stock, while parameter uncertainty and observation error are integrated, a key part of the uncertainty (i.e. structural uncertainty) is not included in the ensemble but only presented as sensitivity analysis and thus has no effect on the stock status and on the management advice. For the Gulf of Mexico stock, different model configurations are presented only as sensitivity analysis and no attempt has been made to integrate parameter uncertainty and observation error. I consider that an ensemble of different plausible configurations selected and weighed by a comprehensive diagnostic against performance criteria agreed beforehand should be developed to provide stocks status and management advice for both stocks in the future. As best practice, and as a minimum, the ensemble should integrate the three main sources of uncertainty, process uncertainty, parameter uncertainty and observation error in the data (sensu Punt et al. (2016)).

The ensemble should also be used for deriving catch forecast scenarios, in which plausible assumptions on the productivity of the stock (e.g., recruitment, growth, mortality, etc) can be integrated to mimic variability of the ecosystem and possible effects of climate factors.

No particular ecosystem considerations or inclusion of ecosystem indicators into the models were made for the two stocks. Concerning the possible effect of climate factors on management reference points, as the assessments encompass also the revision of the reference points, which are estimated within the assessment models, and it is done within a relative short time frame (in average 5 years), there is no compelling need to account for climate changes (which typically occur over a longer time frame than the assessment update). This is especially valid also in the context of the use of SPR to derive management reference points since SPR can be calculated using most recent biological parameters and taking into account most recent conditions of stock productivity.

4. Provide, or comment on, recommendations to improve the assessment.

- Consider the research recommendations provided by the Data and Assessment workshops 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. These recommendations should be described in sufficient detail for application, and should be practical for short-term implementation (e.g., achievable within ~6 months). Longer-term recommendations should instead be listed as research recommendations above.

The model diagnostic toolbox should be greatly expanded to include as a minimum runs test of the residuals, retrospective and forecast Mohn´s rho, hindcasting and MCMC.

For the Gulf of Mexico Scamp stock steepness cannot be estimated reliably due to the lack of strong contrast in the spawning stock time series. For the Atlantic stock, steepness is estimated but retrospective analysis shows that it is greatly dependent on the latest estimates of recruitment. However, even if we are unable to estimate it in the model, the SR curve has a functional form and therefore ignoring steepness (e.g., using a statistical catch at age approach) might have consequences on the fit and the predictive capability of the model. I recognize that ignoring the existence of a functional form of the SR curve used in conjunction with average recruitment in the projections and SPR as reference points has limited impact on the short term forecast advice, however, it has a large effect when modelling long term dynamic as for example when conducting an MSE (i.e. Management Strategy Evaluation). Thus, I consider that as steepness (and therefore a SR functional form) exist in fishes, it should be estimated or fixed based on best available knowledge. In this particular case, steepness should be either fixed at the species prior estimated by FishLife or, at best, is used as a dimension (e.g. low, med, high) of the ensemble together with others key parameters for both stocks.

There is an evident conflict between observed length compositions and estimated growth parameters, especially the length compositions that included fish up to 129 cm while the assumed L_{inf} is around 70. A full revision of the length composition data and the associated growth parameters is needed to resolve this conflict.

For the Gulf of Mexico Scamp stock, the reference point should be based on SPR or on a fraction of B_0 as MSY derived reference points in conjunction with the high steepness estimated by the model, which will most likely imply large risk of stock collapse.

For the Atlantic stock, the base case model estimated that selectivity of small fish has increased after the entrance into force of the size regulation (i.e. A_{50} is smaller after 1992). This result is counterintuitive and might be linked to existing conflict between the different data sources or other model or data misspecification. As selectivity is affecting MSY reference points, this issue would need a full revision of the size and age data and the model configuration used. In principle, selectivity should be modelled as a length process but this option does not exist in the BAM model. Thus, an option might be to use Stock Synthesis as a supporting model in the future for testing the effect of model configuration that are not achievable with present BAM architecture.

An ensemble of different plausible model configurations selected using hypothesis testing and weighed by a comprehensive diagnostic against performance criteria agreed beforehand should be used to provide stock status and management advice for both stocks. As best practice, and as a minimum, the ensemble should integrate the three main sources of uncertainty, process uncertainty, parameter uncertainty and observation error in the data.

5. **Provide recommendations on possible ways to improve the Research Track Assessment process.**

None

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

See above.

Detailed report of of the SEDAR 68 Atlantic and Gulf of Mexico Scamp assessment review

General comments

SEDAR 68 present two separate assessments for a species complex, where Scamp (*Mycteroperca phenax*) and yellowmouth Grouper (*Mycteroperca interstitialis*) are assessed as a pseudo-species for two separate stocks, Gulf of Mexico and Atlantic. I can understand the logic of doing so since distinguishing the two species is deemed to be difficult, however, it would be beneficial to add, as a minimum, a figure with the estimate of the proportion by year of the two species in the catches. Ideally, it would be informative to include in the reports also the estimated key biological parameters for each stock of the two species separately. All biological parameters are combined in the assessment but it is hard to evaluate to which of the two species the biological parameters used in the assessment are mostly resembling. In theory, although the biology appears to be similar between the species, you might expect the reference points to differ and thus a species aggregated assessment might cause depletion of the species more sensitive to exploitation. The only information I found about the proportion between the two species is that Yellowmouth Grouper represents at most less than 3% of the catches (e.g. page 8 of the Gulf of Mexico report). In this case, I consider that the biology of the pseudo-species in the models should resemble Scamp Grouper or should be weighed accordingly to the proportion of the two species in the catches. However, using life trait history parameters within the R package *SPMpriors*, a value of F_{MSY} can be derived for the species assuming a length at first capture around 25 cm (eye-balled from Figure 13 of the Gulf of Mexico stock report). In this case, M is estimated to be 0.21 and 0.30 and F_{MSY} is 0.09 and 0.12 for *M. interstitialis* and *M. phenax,* respectively. Given the close similarity in life history trait parameters (also the derived steepness is very similar for the two species, see section on Gulf of Mexico stock below), merging the two species in a pseudo-species and assuming an equally weighed biology is most likely to be fine although some of the key parameter (e.g. t*max* and thus M) are different between the two species but t*max* used in the Gulf of Mexico stock model (i.e. 34 years) is more similar to Yellowmouth Grouper than to Scamp.

The major concerns I have for both stocks is about the use of the sensitivity analysis and ensemble and thus how uncertainty is treated in the modelling and advice context. For the Atlantic stock, an ensemble is used to estimate stocks status and provide management advice, which is my preferred option given the substantial uncertainty in the dataset and for several of the key parameters. However, for this stock, while parameter uncertainty and observation error are integrated, a key part of the uncertainty (i.e. structural uncertainty) is not included in the ensemble but only presented as sensitivity analysis and thus has no effect on the stock status and on the management advice. Moreover, for this stock the diagnostic toolbox used is limited and even for the only diagnostic used (i.e. retrospective analysis), the bias has not been quantified.

For the Gulf of Mexico stock, the diagnostic toolbox is more extensive and includes several of the key tests but some are missing, in particular runs test of the residuals and, most importantly, hindcasting of relative abundance indices and size and age compositions, forecast Mohn´s rho of SSB and F and MCMC. Those have been indicated by recent papers (Carvalho et al., 2021; Kell et al., 2021) as key diagnostics to evaluate how well an assessment model is fitting the data, is stable in retrospective analyses and most importantly, how good is the model in predicting the future. Generating predictions to be used in management advice is the key objective of any assessment model and thus a model that is unable to predict has limited use in fisheries management. Also for this stock, different model configurations are presented only as sensitivity analysis. Thus, I suggest that an ensemble of different plausible configurations selected and weighed by a comprehensive diagnostic against performance criteria agreed beforehand should be developed to provide stock status and management advice for both stocks. As best practice, and as a minimum, the ensemble should integrate the three main sources of uncertainty, process uncertainty, parameter uncertainty and observation error in the data (*sensu* Punt et al. (2016). The ensemble should also be used for deriving catch forecast scenarios, in which plausible assumptions on the productivity of the stock (e.g. recruitment, growth, mortality, and others) can be integrated to mimic variability of the ecosystem and possible effects of climate factors.

I also noted that informative priors for key quantities such as steepness were used but that the derived quantities were presented in the form of maximum likelihood estimates (MLEs) and associated estimates of asymptotic normal standard estimates from the Hessian. From a strict statistical point of view, the use of informative priors requires Bayesian MCMC estimation to correctly integrate out the expected values from posterior distribution, given the informative prior knowledge and the data. In particular, steepness is directly linked to F_{MSY} so that MSY reference points and their estimates can be affected by the choice of the steepness value and distribution. The distribution of the steepness prior is a symmetric beta with a large penalty near the bounds (SD=1). The estimated distribution of steepness from the model is very different because MLE uses a truncated normal approximation with an upper at h = 1. This is why each time you use a prior for key parameters such as steepness then you should as a minimum run an MCMC to corroborate that the posterior mode and maximum likelihood estimate do not differ considerably. Where these differ considerably, model behavior should be investigated further before strict quantitative interpretation is made of either the point estimates or the uncertainty in those estimates (Stewart et al., 2013). Ideally, if informative priors are used, you should run your model with MCMC and present the results as derived from the MCMC for estimated quantities.

Atlantic Scamp and Yellowmouth Grouper stock

SEDAR 68 Atlantic Scamp Grouper, SECTION III: assessment report

Catches of the commercial fleets were pooled in a single pseudo-fleet and the same was done for the recreational fleets. This is in theory fine if the pooled fleets have similar selectivity and/or if the proportion between the fleets is more or less constant between years. However, it would be useful to add more information to substantiate this choice, especially the size composition of the different fleets that compose the commercial and recreational pseudo-fleet and their landings (or proportion) over time.

Length compositions were removed in years for which age compositions were available. This is in principle fine but running the model with length compositions as a ghost fleet (*sensu* Stock synthesis) might provide important information on the consistency between growth (i.e. fixed) and yearly age compositions (i.e. yearly estimated), which is a good indication of possible changes in growth over time.

The model assumes that discards are null before 1992. I found a reference on page 3 on size regulations to come into place for the first time in 1992 but it would be useful to add some more information on what kind of regulations and to which fleets they were applied (see for example

Figure 10 in the Gulf of Mexico assessment report). Discards might occur not only for reasons linked to (if fully enforced) management regulations and therefore an **alternative configuration would be to model retention and discard selection ogives separately and thus allow the model to predict discards also before 1992**.

As already explained for **SEDAR 68-AW-04**, processes that lead to the estimation of the indices of abundance have been modelled using a Delta model (i.e. combining two separate generalized linear models, GLMs for encounter probability and positive catch rates). The proportion of the 0s has the potential to affect the performance of an ad-hoc method such as the Delta model. As a rule of thumb, a Tweedie distribution should be used when the ratio is larger than 1/3 (Shono, 2008). In the absence of the information about the proportion of 0s by year in the catch data set used, it is not possible to evaluate how the use of the Delta model might have affected the estimation of the indices of abundance.

The commercial handline and the headboat indices of abundance were truncated in 2009 (i.e. years after 2010 were excluded from the model). However, I wonder if any attempt to standardize the recent years of those two indices for management regulations in the GLM has been made. The only CPUE index present in the model in recent years is the SERFS/MARMAP CPUE index. However, the model fit of this index between 2005 and 2015 is rather poor, which may be evidence of a conflict between catches, and most likely its size compositions, with the SERFS/MARMAP index. Thus, it would be desirable to attempt to extend the commercial handline and the headboat indices of abundance to help resolving that conflict. Also, a likelihood profile of different components for the key derived quantities should be conducted to help understanding what might be the cause the observed conflict.

The report states at page 14 that SERFS/MARMAP index is the primary source of information of the population trend (at least this is my interpretation of the text). However, if this is the case, an effort should have been made to somehow force the model to fit the SERFS/MARMAP index, which is not the case as showed by Figure 10. On the other hand, the poor fit of the relative abundance index might be also a symptom of model misspecification instead of real conflict between the different data sources.

The indices have a CV=0.2, so they are considered to be rather precise. Generally, estimated CV of relative abundance indices rarely achieve that level of precision. **Thus, an alternative model configuration would be to allow for additional variance of the relative abundance indices** to understand the effect of this assumed high precision on the assessment model in terms of diagnostics and most importantly on the results.

In Figure 5, there is an apparent inconsistency between the label on the y-axis and the caption. From the text at page 16, recreational catches should have been expressed in 1000s fish while the label of Figure 5 is referring to 1000 lbs.

Initial abundance at age is assumed from the equilibrium age structure, so implicitly F in 1969 and before is assumed to be 0. However, Figure 24 does not present estimated landings from the model for the period 1969-1975 so that it is difficult to evaluate the difference between the estimated and the inputted landings and therefore if the equilibrium assumption is justified. This might theoretically have an impact on the estimated biomass reference points and therefore I suggest adding the estimated landings between 1965 to 1975 (or even before if available) for completeness.

I got confused about the underlying assumptions of growth and I might be wrong with my comment here. My understanding is that growth is estimated externally and it is time invariant, which is fine. But then a different growth curve is used for the population and for the fishery, which confuses me. The true growth is a characteristic of the population so I would expect that it does not depend on the data source used to estimate it (i.e. fish growth is independent of the fishing gear). I think some extra lines should be added to the report to explain what are the benefits (and drawbacks) of having two separated growth curves in the model. Also, parameters related to the growth curve including the CV of growth curve are generally estimated when conditional age at length (CAAL) are included in the model. This does not seem to be the case here, so I wonder how you can confidently estimate the CVs for the two growth curve.

It would be worthwhile to include recruitment and recruitment deviation estimates in the report as well as to specify when (i.e., in which year) the model starts to estimate recruitment deviations.

BAM is inherently an age structured model so that selectivity can be translated into an age derived process. However, from the report it was not clear to me if selectivity was modelled as a length-based (which should be generally the case as selectivity is almost invariably a length dependent more than age dependent process) or as an age derived process. Also, it is unclear if the resulting logistic selectivity stems from the model or the parametric form of the selectivity was factually superimposed by the analyst. In this context, I found the justification for the use of a logistic selectivity (i.e. the presence of old fish in the data) a bit arbitrary. Theoretical work has shown that selectivity in models like BAM (i.e., gear selectivity plus fish availability) are invariably dome shaped (e.g., Sampson and Scott 2011) but the extent of the dome might vary, which justify the presence of old fish in the data but not the assumption that they are fully selected. Thus, it is important to clarify if logistic selectivity as shown in Figure 15,17 and 19 is freely estimated or superimposed. For example, if logistic selectivity has been superimposed, **dome shaped selectivity might be an alternative configuration** in a model ensemble.

The base case model included time blocks for selectivity of the commercial and recreational fleet. My understanding of the change in the regulation that occurred in 1992 (although I have to admit that limited information is available in SEDAR 68 to fully grasp the details of the change in regulation) was *de facto* aimed to reduce the exploitation on small fish. However, figure 15 and 17 show that selectivity of small fish has increased instead (i.e. A_{50} is smaller after 1992). This result is counterintuitive and might be even spurious or linked to existing conflict between the different data sources. Unfortunately, the report does not include any sensitivity on this particular aspect and indeed it would have been useful to compare the base case to a model without time blocks on selectivity or using random walk assuming yearly changes in selectivity. Comparison though should have ideally been done using diagnostics such as those listed in the **Uncertainty and measures of precision** section below.

The weight assigned to the relative abundance indices is estimated using an iterative procedure *sensu* Francis 2011. However, the COM index is up weighed substantially (1.4), which is counterintuitive as generally effective sample size is smaller than observed sample size.

Uncertainty and measures of precision

Sensitivity analysis

I consider that the alternative models examined as part of the sensitivity analysis do cover most of the parameter uncertainty and observation error and some of the structural uncertainty. However, they remain solely as part of the sensitivity and as such they do not affect the stock status and the management advice. Characterizing uncertainty of management related quantities is a crucial part of modern fisheries science and a pre-requisite for any form of risk assessment (Magnusson et al., 2013; Maunder et al., 2020). Uncertainty can be quantified in terms of process and observation errors that are estimated by a statistical stock assessment model (i.e. estimation uncertainty) and structural uncertainty that describes variations in outcomes due to alternative assumptions about the model structure. Typical examples of uncertainty in the structural formulation of stock-assessment models pertain to functions describing the stock recruitment relationship, somatic growth and selectivity pattern (e.g. logistic or dome-shaped). In addition, it is often not possible to estimate highly influential population parameter, such as natural mortality (M), and fixing those to a range of alternative values will evidently contribute to structural uncertainty (Punt et al., 2021). This has been partially tackled in what the analysts here define as an “ensemble modelling approach” where bootstrap of observed data and resampling of two of the key parameters (i.e. M and discard mortality) were integrated. However, the ensemble used here deals only with parameter uncertainty and observation error (and only key ones) but surely not with structural uncertainty, which is partially dealt with instead under the “Sensitivity analysis” section. Ideally, the three main source of uncertainty should be integrated in the ensemble so that probabilistic statement of the stock status (i.e. Figure 36 of the Atlantic stock report) but also of consequences in terms of stock trajectories under different management choices can be properly derived. This is particularly important for F/F_{MSY}, which is substantially affected by M, ageing error and partially steepness (see Figure 39, 41 and 43).

Retrospective analysis

The only diagnostic presented in the report is the retrospective analysis (Figure 44). However, estimated retrospective bias is not quantified (e.g. Mohn´s rho; Hurtado-Ferro, 2015) so that it is difficult to determine its magnitude (which seems to be substantial judging from Figure 44). This is important because the magnitude and direction of the retrospective bias is one of the key diagnostics used to identify model specification in stock assessments (Carvalho et al., 2021). A ’rule of thumb’, proposed by Hurtado-Ferro et al. (2015), suggests values of Mohn´s rho that fall outside (-0.15 to 0.20) for SSB for longer-lived species, or outside (-0.22 to 0.30) for shorter-lived species indicates an undesirable retrospective pattern. In addition, the direction of the retrospective bias has implications for characterizing risk associated with management advice. A positive Mohn´s rho for SSB is of particular concern because it implies a systemic overestimation of biomass, which would lead to over-optimistic quota advice if not taken into consideration (Hurtado-Ferro et al., 2015).

Other key diagnostics are the quantitative analysis of the residuals (i.e. runs test and root mean square error of the residuals), which can be used for CPUEs, age and length compositions, jittering (i.e. sensitivity to initial values of the parameters and evaluate whether the model has converged to a global solution rather than a local minimum), Age-Structured Production Model (ASPM), hindcast Mohn´s rho and hindcasting (Carvalho et al., 2021; Kell et al., 2021). For example, judging from the residuals showed in Figure 8-10, I suspect that several of the relative abundance indices would not pass a runs test.

Hindcasting and hindcast Mohn´s rho is particularly important because the provision of fisheries management advice requires the assessment of stock status relative to reference points, the prediction of the response of a stock to management, and checking that predictions are consistent with reality. A major uncertainty in stock assessment models is the difference between model estimates and reality. To evaluate uncertainty often a number of scenarios are considered corresponding to alternative model structures and dataset choices (Hilborn, 2016). It is difficult, however, to empirically validate model prediction, as fish stocks can rarely be observed and counted. Kell et al. (2016, 2021) showed how hindcasting can be used to evaluate model prediction skill of the CPUE but it has been recently extended also to size and age compositions (Carvalho et al., 2021). When conducting hindcasting, a model is fitted to the first part of a time series and then projected over the period omitted in the original fit. Prediction skill can then be evaluated by comparing the predictions from the projection with the observations using for example the MASE indicator (Kell et al., 2021). Therefore, hindcasting is used to estimate prediction skill, a measure of the accuracy of a predicted value unknown by the model relative to its observed value, and to explore model misspecification and data conflicts. Prediction skill can also be used to identify alternative hypotheses, weight models in an ensemble and agree on reference sets of operating models when conducting Management Strategy Evaluation.

This part of the analysis is unsatisfactory and I suggest that as a minimum, the analysis would add the key diagnostics presented by Carvalho et al., 2021 and Kell et al., 2021. Importantly, any of the diagnostics should not be used in isolation (Carvalho et al., 2021 and Kell et al., 2021) but they should be used as a set of tools to assess whether it is plausible that a system identical to the model generated the data and thus how equally plausible are the different model configurations and ultimately help the analyst to select those to be used for providing management advice.

Benchmarks and reference points

It would be important to add the plot of the production function so that it would be possible to evaluate the stock status also in terms of B_0 but most significantly the shape of the production curve. This is particularly important because when the production function is highly skewed to the left (e.g. Fox type of curve), the resultant B_{MSY} might be close to levels of biomass that are usually considered as a limit, typically 20%B_0, and in those cases biomass target reference points would be more appropriate (Punt et al., 2013).

Gulf of Mexico Scamp and Yellowmouth Grouper stock

Working documents

Estimation of commercial abundance index for Gulf of Mexico Scamp & Yellowmouth Grouper using reef fish observer data (SEDAR68-AW-04)

A standardized catch-per-unit-effort (CPUE) index of the Gulf of Mexico Scamp and Yellowmouth Grouper is estimated and used in the assessment model. Although a survey-design ratio-of means estimator is used instead of a regression model estimator, it would be important to report the proportion of 0s by year in the catch data set used. The proportion of the 0s has the potential to affect the performance of an ad-hoc method such as the Delta model (i.e. combining two separate generalized linear models, GLMs for encounter probability and positive catch rates). Delta models were used to guide the specification of various aspects of the estimation process. As a rule of thumb, a Tweedie distribution should be used when the ration is larger than 1/3 (Shono, 2008). In recent years, however, a "Poisson-link" model has been suggested as a better alternative to Delta models for CPUE standardization (Thorson, 2018). The "Poisson-link" model is a derivation of the compound Poisson-gamma (CPG) distribution (Foster and Bravington 2013), which is a special case of the Tweedie distribution. In the absence of the information about the proportion of 0s by year in the catch data set used, it is not possible to evaluate how the use of the Delta model might have affected the estimation process.

The choice of a survey-design ratio-of means estimator instead of a regression model estimator seems to be guided mostly by the unbalance in effort between stratum-area combinations (i.e. depth and space). However, regression model standardization procedures that account for an unbalanced sampling between depth and space (and many other covariates) are widely used for deriving CPUE to be used in stock assessment models (e.g., VAST; https://github.com/James-Thorson-NOAA/VAST). As a minimum, a comparison between the two methods would be ultimately useful as the use of regression model standardization procedures would avoid having to make several semi-subjective choices during the process implicit in a survey-design ratio-of means estimator (e.g. the depth-space stratification). The use of regression modeling would facilitate the presentation and interpretation of the covariate effect on the CPUE. Moreover, it will also allow for a comprehensive standardization of the length compositions (e.g., exploring and eventually including additional covariates in the standardization process; see Berg and Kristensen (2013)), which currently are only weighted by the depth and space strata.

Finally, it would be useful to add to the report the complete formula and results for any GLMs used during the estimation process and most importantly the comparison between the survey-design ratio-of means estimator and the unstandardized CPUE index to allow for the evaluation of the effect of the standardization procedure on the time trend of the relative abundance of Gulf of Mexico Scamp and Yellowmouth Grouper stocks.

Gulf of Mexico Scamp (*Mycteroperca phenax*) and yellowmouth Grouper (*Mycteroperca interstitialis*) commercial and recreational length and age compositions (SEDAR 68-AW-01)

As explained for SEDAR 68-AW-04, the use of regression models would allow for a comprehensive (e.g. exploring and eventually including additional covariates in the standardization process) standardization of the length compositions (see Berg and Kristensen (2013)), which currently are only weighted by spatial strata for commercial and recreational length and age compositions.

In Table 1, it is not clear if samples refer to the number of fish or the number of sampling events. This is a bit confusing as in Table 4 it is clearly specified that samples refer to number of aged individuals. For clarity, it would be better to use a consistent terminology. Also, for clarity, it would be better to use either acronyms or full name of the fleets in the captions of the figures.

Commercial discard mortality estimates based on observer data (SEDAR 68-AW-03)

While immediate discard mortality is estimated from observer data and therefore a value for each gear and depth combination can be derived, delayed mortality is based on literature values and it has a rather wide range. However, it is not clear how the range value of delayed mortality has been used when estimating the total discard mortality and thus the number of removed fishes to be used in the assessment model. I realized reading the assessment report that uncertainty values in the discard mortality estimates were used in the ensemble modelling (page 22 of **SEDAR 68**) but it is unclear which uncertainty was used (i.e. immediate discard mortality or delayed discard mortality) or, if a combination of the two estimates was used, how those were combined. Some additional explanation here would be beneficial.

Modelling of recreational landings in Gulf of Mexico (SEDAR 68-RW-01)

I appreciate the investigation of the effect of how recreational landings are incorporated in the model. But as for the rest of the sensitivity analysis, I think it should be part of an ensemble model (see comments below).

SEDAR 68 Gulf of Mexico Scamp Grouper Research Track Assessment

Technical notes

It looks like you are using a rather old version of Stock Synthesis, I recommend using the latest version, which has numerous fixes and improvements compared to older versions.

I noted you use Fmethod=2. Usually, the hybrid (Fmethod=3) is recommended unless you have a bycatch fleet (in that case the hybrid method does not actually work) or very high F. I don't mind to use Fmethod=2 but I wonder what is the reason you do so. I have done some trials and the use of Fmethod=3 achieve the same convergence (0.003; any of your models achieve good convergence level anyhow).

Why discards are only females and retained only males? I understand that the species in a protogynous hermaphrodites but is really such a clear cut between the two sexes? It would be good to add any observations (i.e. observed sex at length) to substantiate this assumption.

2.2.2 Age and growth

The model contains substantial amount of conditional age-at-length (i.e. CAAL) data but the CV at age in the growth curve is fixed at 0.13. First, it would be useful to show the profiling of those parameters (which has been used to justify the choice of a CV=0.13). Second, given the considerable amount of CAALs, it would be relevant to attempt to estimate the CVs instead of fixing them also because CAALs are particularly suitable for that task (Methot et al., 2020) as also stated in page 12 of the report.

2.2.3 Natural mortality

To my knowledge, you don't need first age at vulnerability and peak spawning to scale M according to Lorenzen 1996 and Lorenzen 2000 (which is based on the allometric relationship between natural mortality and body weight). To avoid confusion, it would be best to add the formula in the report to show how first age at vulnerability and peak spawning have been used for scaling M.

Natural mortality assumed for the youngest age is rather small when compare to the values obtained for the species using Lorenzen (Jason Cope website; http://barefootecologist.com.au/shiny_m). Is that intentional and could you add an explanation for that? If you want to use Lorenzen to input M, I suggest to use the option within Stock Synthesis (i.e. it just requires the M of a reference age, which can be derived using Jason Cope website or by any other method) to avoid any confusion.

2.2.7. Discard mortality

See above comments on SEDAR68-AW-03.

2.3.1. Commercial landings

In general, I am reluctant to exclude historical landings from an assessment model as they might have an impact on biomass reference points and provide key information on the history of the fisheries and of the stock. However, although I can understand that historical landings are excluded based on expert opinion, I would suggest that they are included in the report and ideally an **alternative model configuration including historical catches could be also presented for completeness**.

For most of the fleets, very few individuals are sampled for fish larger than 80 cm. I suggest to use *CompressBins* options in the data file to create pseudo +groups of large fish. This usually improves the estimation of selectivity, especially the descending parameter.

2.3.7. Commercial Age compositions

Using length compositions in combination with CAALs is surely the best way to model fisheries data as it allows one to propagate uncertainty through the model and the derived quantities and it also allows for time varying growth in the model (if desired by the analyst). However, it is important to note that CAALs do not inform only growth, but they also contain information on cohort strength and thus key population dynamics and stock trends. This is why using CAAL with fixed growth parameters (with the exception of L_{Amin} here) is somewhat illogic because it complicates the model without achieving the advantages of the combined use of CAALs and length compositions. In this case, inputting directly derived number at age instead of CAALs and length compositions would be more appropriate if there is no intention by the analyst to estimate growth.

An alternative model configuration where growth parameters are estimated within the model is used in the sensitivity analysis, but as I have written for the Atlantic Scamp and Yellowmouth grouper stock, those alternative model configurations do not affect the management advice although they show different results compared to the base case. This reiterates the need of using a model ensemble (see also general comments for both stocks, detailed comments for Atlantic stock and section 3.4 below) where different model configurations defined by the analyst through hypothesis testing are integrated, weighted by model performances and all contributing to the estimation of stock status and management advice.

On the technical side, some of the length class proportions do not sum to 1 (i.e. several sum less than 1 and few more than 1). As Synthesis always translate size and age compositions into proportions, it is more convenient to input real numbers than transforming them into proportions just to avoid those kinds of minor mistakes.

There are few very large fish in the size compositions data (up to 129 cm) but L_{inf} is fixed at 70 cm. It is fine to have large fish than L_{inf} in the model as the CV at age will compensate for that, but those individuals are really outside the bound of L_{inf} (L_{inf} = 0.8L_{max} as a rule of thumb; Morais and Bellwood (2018). I suggest adding a sentence or two on this aspect of the data in the report. Indeed, the model estimate of L_{inf} is much more in line with the observed data than with the prior and the fixed value.

2.3.9. and 2.3.10

See comments on standardization procedure on SEDAR 68-AW-04 above.

The relative abundance indices have a CV=0.2, which I consider to be rather precise. Generally, true CV of relative abundance indices rarely achieve that level of precision. Thus, an **alternative model configuration would be to allow for additional variance of the relative abundance indices** to understand the effect of a CV=0.2 of the relative abundance indices on the assessment model in terms of diagnostics and most importantly on the results. See comment under 3.2 section below.

3.1.7. and 3.1.8.

Results of a catch curve analysis was used to justify the use of logistic selectivity for the commercial fleets. Theoretical work has shown that selectivity in models like Stock Synthesis (i.e., gear selectivity plus fish availability) are invariably dome shaped (e.g., Sampson and Scott 2011) but the extent of the dome might vary. Also, catch curves *sensu* Quinn and Deriso (1999) require the assumption that fishing mortality is constant over time (Thorson and Cope, 2017), which not fulfilled here (Figure 47). I was also looking for a reference concerning the relationship between slope of the catch curve and shape of selectivity for older ages (i.e. rule of thumb of slope larger than 1 indicates dome shaped selectivity) but I could not find any in the literature. The slope of the descending limb of the catch curve is very similar (and much less than 1 for all fleets) when comparing commercial and recreational fleets but notwithstanding selectivity was assumed to have a fundamental different shape between two kind of fleets. Also, two aspects are important to note here concerning the double normal selectivity as parametrized in Stock Synthesis. First, it can be used with only 4 parameters and not necessarily 6 (in reality even 3 and still allowing for both shape to be estimated but 4 is the preferred set up in Stock Synthesis). Second, in general, the top-logit parameter is the hardest to estimate while, when the estimation of the descending parameter is problematic, a pseudo plus group of the largest size groups can be created to facilitate its estimation. And third, as also stated in the report, the double normal allows for domed or logistic selectivity depending on the data and therefore its use avoids having to make an a priori choice of the selectivity shape and still keeping the necessary estimation flexibility. Bearing in mind all the above, **I consider that an alternative model configuration that can be tested should use double normal selectivity for more fleets than used now**. Especially, the longline fleet (ComLL) and the RFOP_Index.

Figures 15B, 15D, 16B and 16D shows that selectivity of small fish has increased somewhat after 1990, decreased again after 1999 and increased albeit just slightly after 2003 for commercial fleets, which is a bit counterintuitive when looking at Figure 10 (although I can partially follow the logic described in section 4.4). On the other hand, selectivity of small fish has increased over time for recreational fisheries. Given the estimated parameters in Table 10, I wonder how much the blocks are tracking noise as opposed to a real significant change in selectivity for the different time periods and how much the blocks are depending on the fact that most of the parameters are indeed fixed. Moreover, the first time block contains only 4 years of data (1986-1989) and the estimation of the historical selectivity parameters might be difficult for some fleets. **I suggest that, at least a model without blocks is run and the likelihood presented to support the choice of the use of the blocks.** Also, if blocks are to be used, to test a model in which the first block starts in 1986 (thus increasing the number of observation used to estimate the first block) is compared to the base case model to verify that there is a significant improvement in the likelihood of the fit of the size compositions when using the block from 1990.

Also, changes in selectivity, even when occurring because of management regulations, are rarely a stepwise process and thus it might worth trying using random walk instead of blocks.

The inflection point in the table at page 19 should be easier to understand if expressed always as fish length. I don’t understand what an inflection equal to 0 means here. The text is also confusing, from the Table 10 and Figures 15B, 15D, 16B and 16D the inflection point is time varying using time blocks but the table at page 19 states that it is “Fixed at Maximum”. Further clarifications would be required here.

3.2

Here a reweighting iterative procedure of the relative abundance indices variance was used. All iterative procedures require a pinch of subjectivity. Instead, I suggest to use the estimated “extra_se” option in the Q_setup as it is done automatically within the model.

3.4

The model diagnostic is broad and encompasses some of the key tools for individuate possible model misspecification and poor model performances. However, as for the Atlantic stock, it is still incomplete and there is lack of a quantitative estimation of some the key diagnostic indices. In particular, several key diagnostics have recently been used to evaluate how well an assessment model is fitting the data, is stable in retrospective and most importantly, how good is the model in predicting the future. Given the large uncertainty in several of the key parameters and data sources, a single base model is not warranted here (and is very seldom warranted) and thus an ensemble of plausible model configurations should be used (see also general comments for both stocks and detailed comments for Atlantic stock). As matter of fact, all model configurations are plausible as long as they achieve similar performances when tested against the data and therefore choosing one among all equally plausible configurations will run the risk of "cherry picking" and will affect the final results.

4.3.

All discard ogives parameters are fixed and not estimated but size compositions of the discards are included in the model. On what basis are those parameters fixed and why are they not instead estimated within the model? My understanding from the Synthesis manual is that at least the first two parameters are estimated and the 3th and the 4th are fixed (generally the *Asymptotic retention* is set to 999 if large fish are all retained and the *Male offset To inflection* to 0). However, as you have both discards and retained size compositions, in theory all parameters could be estimated (but you don't necessarily need to do so).

4.5

Recruitment deviations in early years are all negative, which might be true but also might be an artefact of too low initial catches (i.e. initial catches were the average of the first 5 years of the time series). This why it would be important to show the historical landings even if they were not used in the model. Also, you have size compositions already in 1986, i.e. the first year of the model. In this case, it is correct to start the main recruitment deviation in 1986 (although given the argument used for the ramp, you should possibly start 2-3 years later as your size composition do not contain 0 fishes) but it is also appropriate to use early recruitment deviations. This is because your size compositions contain information of the cohorts born much earlier than 1986, at least 10 but up to 20 years back in time (e.g. from 1966 onwards). For example, setting 10 to 15 years of early recruitment deviation might shed light on the series of negative residuals in the first recruitment deviation of the base case model.

4.6

Figure 24 should be shown for the population and not for the data. Thus, I recommend using age groups up to 34 (the true +group in this assessment) in the figure.

4.7.1

Landings of the Charter Private has a rather large discrepancy between observed and estimated values. The issue is that they are not simply scaled but also have rather different pattern over the years, with a large observed set of landings not estimated by the model (i.e. 2006) and vice versa. In these circumstances, **I consider that an alternative model configuration with a low CV** (0.3 used in the sensitivity analysis is still too high) of the landings of Charter private would should be tested to assess the effect of reported landings on the model results.

4.7.2.

Modelling discards separately from retained fish allows the model to estimate discards also in years where discards data are absent. While discard for Charter Private approach null close to the beginning of the time series of discard observations, so the assumption of low discards for the historical period is justified, the same is not true for the other fleets. It seems unlikely that discards of the commercial fleets will go to 0 in 1989 and before from rather high values in 1990. Also, large discrepancy in discards estimation are observed for the recreational fleets. **Thus, an alternative model configuration is to allow the model to estimate discards for the commercial fleets also for the historical period.**

The clear mismatch between length and age compositions might be due to rigid assumptions of growth in the model or changes in growth over time. Discrepancy between estimated and observed age compositions are observed for the recreational catches and they are mostly evident for younger fish (i.e. age 1 and 2). This might point out either that L_{Amin} is mis-specified, that growth changes over time or that there are spatial differences in growth as commercial and recreational fish *de facto* in different areas. Charter boat is by far the most important fleet in recent years so that a mismatch between observed and estimated age compositions for this fleet (Figure 38) can have important consequences on the short term forecast.

4.7.5

I appreciated the use of a prior for the Theta in the Dirichlet-Multinomial setting, it is rarely done. However, the upper bound should be set at 5. An upper bound of 5 may help identify cases that otherwise would have convergence issues as indicated at page 216 of the Stock Synthesis manual (Methot et al., 2020). As five of the DM parameters are at the 5 bound, this will reduce the estimated model parameters by 5 simply fixing the DM Theta of those parameters.

4.8.4

Retrospective analysis is included as one of the key diagnostics but estimated retrospective bias is not quantified (e.g. Mohn´s rho; Hurtado-Ferro, 2015) so that it is difficult to discern the extent of the retrospective bias (which seems not to be anyhow too substantial judging from Figure 41).

Steepness and sigmaR

A special section is dedicated to two of the key parameters of the model steepness (h) and sigmaR (σR), especially if MSY reference points would be used.

In general, uncertainty regarding steepness is the major source of variation in the final size of the resource and whether it is below the overfished threshold, although the extent of recruitment variability also impacts these quantities (Punt et al., 2008). Steepness prior from Shertzer and Conn 2012 (i.e. 0.84) is the mode of a meta-analysis of 75 stocks from very different areas and genus, with one single *Mycteroperca* species (i.e. *M. microlepsis*). A much closer match for the prior would be to use the R package SPMpriors, which uses FishLife (https://github.com/James-Thorson-NOAA/FishLife) for the assessed species, ideally tuned for their stock specific life history parameters. As those were not available in the report, I have derived *h* and *σR* using SPMpriors for the two species separately. For *M. interstitialis*, *h* is 0.79 and *σR* is 0.42, while for M. phenax, *h* is 0.77 and *σR* is 0.41. Those values are rather consistent between species but different from the prior derived by Shertzer and Conn (2012).

In general, steepness cannot be estimated reliably in assessment models unless there is strong contrast in the spawning stock time series (i.e. resulting from a level of spawning biomass decline that should probably be avoided) (Kolody et al., 2019). Simulations structured to be similar to the assessment models have demonstrated that steepness estimates are often imprecise and biased, often converging to the upper bound (i.e. close to 0.99 as it is the case here when freely estimated), even when the true h is considerably lower (e.g., Magnusson and Hilborn, 2007; Lee et al., 2012). It is easy to demonstrate that due to the flat yield curve when the true h=1 (which is anyhow biologically rarely the case), under-specifying h (i.e. assuming a lower h when the true h is indeed larger, e.g. assuming h=0.75 when the true h is close to 1) results in less lost catch than over-specifying h. In addition, fishing at F_{MSY} based on h=0.75 maintains the biomass at a much higher level. Theoretically, MCMC could be used to integrate over random effects as done in ISSF 2011 which might resolve the issue often encountered with the use of Beverton and Holt curves in integrated models that steepness tends towards the upper bound of 1.0. However, I would not recommend it here. Instead, given the fact the data here are un-informative for steepness, I suggest that steepness is either fixed at the species prior estimated by FishLife or, at best, is used as a dimension (e.g. low, med, high) of the ensemble together with others key parameters as discussed above for the Atlantic stock.

SigmaR is the stochastic recruitment process error and the estimation of this parameter within integrated models is generally recognised to be problematic (Kolody et al., 2019) so that *σR* is generally fixed at a values that is large enough to prevent the stock and recruits from constraining individual recruitment estimates (e.g. analogous to traditional VPA) (Kolody et al., 2019). The ISSF (2011) preliminary meta-analysis (based on time series of spawning biomass and recruitment point estimates) mostly indicated $0.2 \leq \sigma R \leq 0.5$, which suggests that σR is often being inflated in stock assessment models (intentionally or not). The corresponding estimates for B/B_{MSY} and MSY indicate a moderate sensitivity to *σR* (Kolody et al., 2019), which is somehow reassuring. Here the estimated *σR* is around 0.36 (Table 10) which is in line with both the meta-analysis and the stock specific estimated from *SPMpriors*. Anyhow, my two preferred options are either to fix *σR* at the species prior after profiling, assuming that the different values do not affect significantly the stock status, or, at best, to use *σR* as a dimension (e.g, low, med, high) of the ensemble. Finally, *h* and *σR* are often correlated and thus a combined likelihood profile that considers plausible combination of these two parameters should be done.

Reference points

When setting steepness close to 1, the ratio between B_{MSY} and B_0 is generally close to 0.20. In this cases, SPR or biomass targets should be used instead, which are independent on both steepness and selectivity. For example, biomass targets in the range of 35–40% of carrying capacity minimizes the potential loss in yield compared with what would arise if B_{MSY} was known exactly (Punt et al., 2013). Here I fully agree with the report on the use of the results from Harford et al., 2019, which are specific for groupers and thus an SPR ratio between 40 and 50% is a suitable candidate for both stocks. I have myself compared different harvest control rules system based on different ways of deriving limit and target reference points and the analysis confirmed that within an F_{MSY} context, using the steepness estimated by the model will doom the stock to collapse while SPR40 and SB40 (i.e. setting $0.2B_0$ as limit and $0.4B_0$ as target) has the same performance in terms of long term catches and risk to fall below the limit reference point.

Inference on stock status

As explained in the under **General comments**, each time you use a prior for key parameters as steepness then you should as a minimum run an MCMC to corroborate that the posterior mode and maximum likelihood estimate do not differ considerably. Moreover, MCMC is a powerful diagnostic for detecting model misspecifications in the broader sense and to regularize the model, i.e., to check that all parameters are identifiable (Monnahan et al., 2019). I ran an MCMC using the NUTS algorithm, which confirmed that all parameters are identifiable, the two estimates are rather similar and stock status is unaffected (99% using MLE against 100% when using MCMC of probability of being in the green quadrant of the kobe plot), which reassured at least myself that the model is fairly robust.

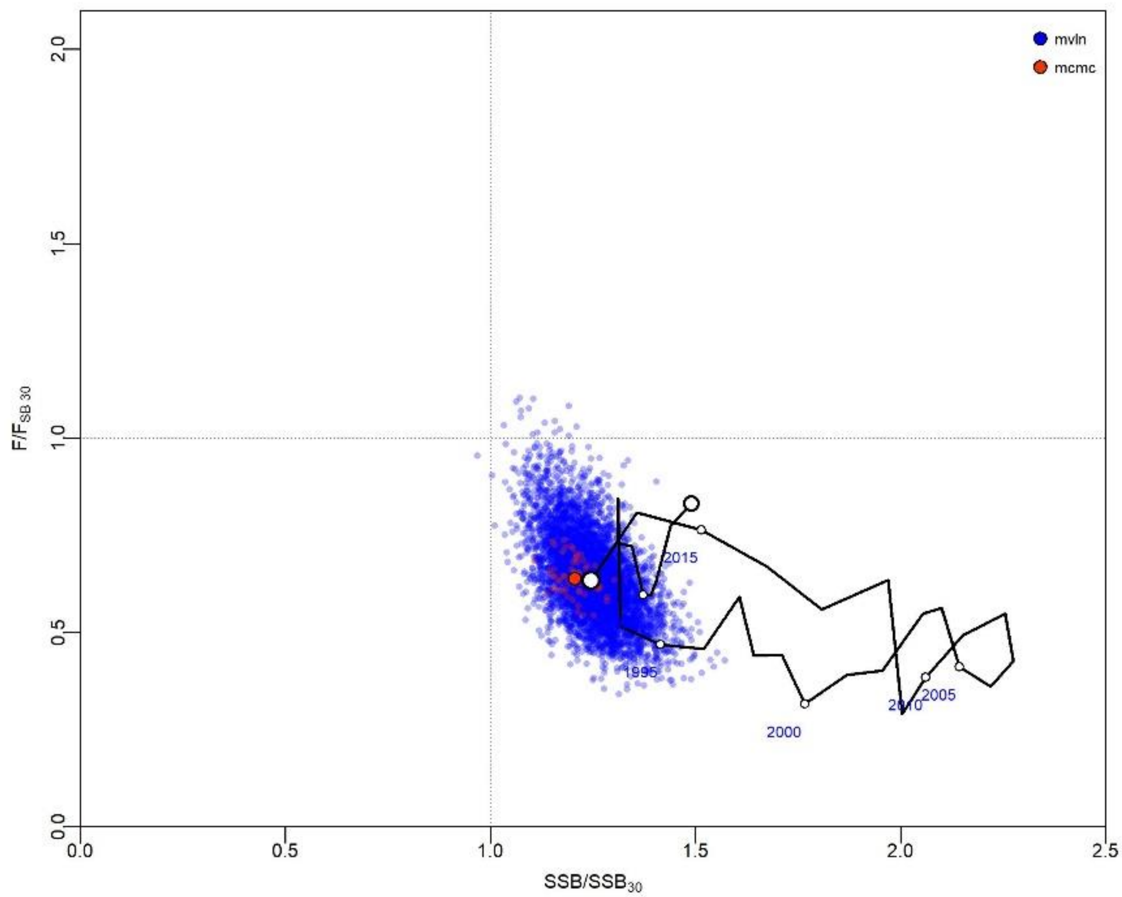

Figure 1

Suggested alternative model configurations to be included in an ensemble

Atlantic scamp grouper

1. Model retention and discard selection ogives separately and thus allow the model to predict discards also before 1992.
2. Allow for additional variance of the relative abundance indices.
3. Allow for dome shaped selectivity for all fleets.

Gulf of Mexico scamp grouper

1. Include historical catches in the model for completeness.

2. Assume a low CV for the Charter Private landings.
3. Allow for additional variance of the relative abundance indices.
4. Use double normal selectivity for more fleets than used now, especially, the longline fleet (ComLL) and the RFOP_Index.
5. Run a model without blocks and include the likelihood of the size composition to support the choice of the use of the blocks.
6. Increase the years in the first block to avoid having too little data to estimate the historical selectivity.
7. Allow the model to estimate discards for the commercial fleets in the historical period.

My own analysis

I have run some of the diagnostics (see Carvalho et al., 2021) for the base case model configuration, in particular retrospective, runs test, forecast Mohn´rho and hindcasting. The retrospective Mohn´s rho of the SSB is over the -0.15 limit (i.e. 0.29), while F retrospective bias is low (0.07). It is to be noted that the run-3 does not converge. The results of the runs test and hindcasting are of mixed nature, which is normal when models deal with multiple indices and size/age compositions. In general, the model has good prediction skill, both for the size compositions and for the relative abundance indices, with the exception for the age compositions of the Charter Privat and for the relative abundance index of RFOP. The combined MASE for age and relative abundance index fail to pass the test.

I run also “my favorite” base case model (Reference_new in the table below), which integrates the following changes:

1. Creates +groups for size larges than 84cm as there are no fish observed in those size classes.
2. Use Lorenzen option for M within Synthesis with reference age sets to 10 years and M=0.18 (estimated from http://barefootecologist.com.au/shiny_m).
3. Estimate growth parameters with priors and prior SD as specified in the SEDAR 68 report. When estimating both, k is at the bound (0.05) and much smaller than the prior while L_{inf} is much larger than the prior but the priors are uninformative anyhow and this aspect might need some attention. It might be needed to fix either k or L_{inf} and estimate only one of them.
4. Fix steepness and sigma R (i.e. do not estimate) to priors as specified in the SEDAR 68 report.
5. Use FMethod=3.
6. Use extra SD option of relative abundance indices.

7. Estimate Retain_L_infl and Retain_L_width for all fleets.
8. Reduce bound of Dirichelet to 5 and fix parameters (5 parameters) that hit the bound.

A further field of exploration would be:

1. Simplify the blocks, setting to 10 all Retain_L_asymptote_logit and do no estimate (Reference_blocks).
2. Remove all blocks except block 1 for the ComLL commercial fleet (Reference_noblocks).

However, this last option will likely generate some of the selectivity parameters to be hitting bounds and/or to have a large standard deviation. While this is an issue that requires attention and some extra work, it will also tell you which parameters are really estimable given the data and which are indeed not. In table 1 and 2, show an example with a hypothetical ensemble of 5 alternative models.

Table 1: Summary table of alternative models.

Label	Reference	Reference_new	Reference_blocks	Reference_noblocks	Reference_DN
TOTAL_like	16651	15266	15247	15440	15248
Survey_like	-47.3	-43.6	-56.0	-58.9	-43.8
Length_comp_like	6831.6	5890.9	5890.1	6060.2	5875.6
Age_comp_like	9774.5	9288.5	9294.7	9314.6	9286.6
Parm_priors_like	33.5	31.2	31.6	31.7	31.3
Recr_Virgin_millions	1.69	2.41	2.54	2.44	2.44
SR_LN(R0)	7.43	7.79	7.84	7.80	7.80
SR_BH_steep	0.948756	0.84	0.84	0.84	0.84
NatM_Lorenzen_Fem_GP_1	NA	0.18	0.18	0.18	0.18
NatM_Lorenzen_Mal_GP_1	NA	0.18	0.18	0.18	0.18
NatM_p_1_Fem_GP_1	NA	0.18	0.18	0.18	0.18
NatM_p_1_Mal_GP_1	NA	0.18	0.18	0.18	0.18
SSB_Virgin_thousand_mt	3.911	4.209	4.445	4.35	4.226

SSB_2018_thousand_mt	1.431	1.195	1.444	1.431	1.200
F_2017	0.11	0.13	0.10	0.10	0.13
SSB_MSY_thousand_mt	0.734	0.836	0.882	0.876	0.837
Bratio_2018	0.366	0.284	0.325	0.329	0.284
SPRratio_2018	0.497	0.518	0.468	0.460	0.519

Table 2: Summary table of the diagnostics of alternative models (number of parameters is not directly comparable when switching from FMethod =2 to FMethod=3 in Stock Synthesis; F as parameters and F from hybrid are comparable notions, just estimated in a different way.

Run	Reference	Reference_new	Reference_blocks	Reference_noblocks	Reference_DN
Convergence	0.0839	0.0023	0.0005	0.0019	0.0156
Total_LL	16651	15266	15247	15248	15440
N_Params	220	99	101	107	81
Runs_test	0	0	0	0	0
Runs_test_1	1	0	1	0	1
Runs_test_2	1	0	0	0	0
Runs_test_3	1	1	1	1	1
Runs_test_4	0	0	1	0	0
Runs_test_5	1	1	1	1	1
Runs_test_6	0	0	0	0	0
Runs_test_7	1	1	1	1	1
Runs_test_8	1	1	1	1	1
Runs_test_9	0	0	0	0	1
Runs_test_10	0	0	0	0	1
Runs_test_11	1	1	1	1	1
RMSE_Perc	0	0	0	0	0
RMSE_Perc_1	1	1	1	1	1
RMSE_Perc_2	1	1	1	1	1
MASE1	1	1	0	1	1
MASE2	0	0	1	0	0
MASE3	0	0	0	0	0
MASE_Combines	1	1	1	1	1
Retro_Rho_SSB	0	1	0	1	0

Forecast_Rho_SSB	0	1	0	1	0
Retro_Rho_F	1	1	0	1	1
Forecast_Rho_F	1	1	0	1	1
MASE41	1	1	1	1	1
MASE42	1	1	1	1	1
MASE43	1	1	1	1	1
MASE44	1	1	1	1	1
MASE45	1	1	1	1	1
MASE46	1	1	1	1	1
MASE47	1	1	1	1	1
MASE48	0	0	0	0	0
MASE49	1	1	1	1	1
MASE_combined	0	0	1	0	0
Average weighed by diags	**0.458**	**0.597**	**0.458**	**0.597**	**0.472**

Literature cited

Berg, C., Kristensen, K., 2013. Spatial age-length key modelling using continuation ratio logits, 2012. Fisheries Research 129, 119–126. https://doi.org/10.1016/j.fishres.2012.06.016

Carvalho, F., Winker, H., Courtney, D., Kapur, M., Kell, L., Cardinale, M., Schirripa, M. et al. 2021. A cookbook for using model diagnostics in integrated stock assessments. Fisheries Research, 240: 105959.

Foster, S.D., and Bravington, M.V. 2013. A Poisson–Gamma model for analysis of ecological non-negative continuous data. Environ. Ecol. Stat. 20(4): 533–552. doi:10.1007/s10651-012-0233-0.

Francis, R. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences 68:1124-1138.

Harford WJ, SR Sagarese and M Karnauskas. 2019. Coping with information gaps in stock productivity for rebuilding and achieving maximum sustainable yield for grouper–snapper fisheries. Fish and Fisheries 20(2):303-321.

Hilborn, R. (2016). Correlation and causation in fisheries and watershed management. Fisheries, 41(1), 18-25.

Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F.,

ISSF, 2011. Report of the 2011 ISSF Stock Assessment Workshop Rome, Italy, March14-17, 2011. ISSF Technical Report 2011-02. International Seafood Sustainability Foundation, Washington, D.C., USA.

Kell, L. Kimoto, A., Kitakado, T., 2016. Evaluation of the prediction skill of stock assessment using hindcasting. Fisheries Research, 183: 119-127

Kell, L.T., Sharma, R., Kitakado, T., Winker, H., Mosqueira, I., Cardinale, M., and Fu D., 2021. Validation of stock assessment methods: is it me or my model talking? ICES Journal of Marine Science (2021), 00(0), 1–12. https://doi.org/10.1093/icesjms/fsab104.

Kolody, D. S., Eveson, J. P., Preece, A. L., Davies, C. R., & Hillary, R. M. (2019). Recruitment in tuna RFMO stock assessment and management: A review of current approaches and challenges. Fisheries Research, 217, 217-234.

Lee, H.-H., Maunder, M.N., Piner, K.R., Methot, R.D., 2012. Can steepness of the stock–recruitment relationship be estimated in fishery stock assessment models? Fish. Res. 125–126, 254–261. https://doi.org/10.1016/j.fishres.2012.03.001.

Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-Pre, K.A., Whitten, A.R., Punt, A.E., 2015. Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. Ices J. Mar. Sci. 72, 99–110. https://doi.org/10.1093/icesjms/fsu198.

Lorenzen K. 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. Journal of Fish Biology 49:627−647.

Lorenzen K. 2000. Allometry of natural mortality as a basis for assessing optimal release size in fish-stocking programmes. Canadian Journal of Fisheries and Aquatic Sciences 57(12):2374-2381.

Magnusson, A., Hilborn, R., 2007. What makes fisheries data informative? Fish Fish. 8, 337–358. https://doi.org/10.1111/j.1467-2979.2007.00258.x.

Magnusson, A., Punt, A. E., and Hilborn, R. 2013. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC. Fish and Fisheries, 14, Blackwell Publishing Ltd. http://doi.wiley.com/10.1111/j.1467-2979.2012.00473.x.

Maunder, M. N., Xu, H., Lennert-Cody, C. E., Valero, J. L., Aires-da-Silva, A., and Minte-Vera, C. 2020. Implementing reference point-based fishery harvest control rules within a probabilistic framework that considers multiple hypotheses. Scientific Advisory Commitee, Inter-American Tropical Tuna Commission, SAC-11-INF-F. San Diego. 1–54 pp.

Methot RD, CR Wetzel, IG Taylor and K Doering. 2020. Stock Synthesis User Manual Version 3.30.16. NOAA Fisheries, Seattle Washington. 225 pp.

Monnahan, C.C., Branch, T. A., Thorson, J. T., Stewart, I. J., Szuwalski, C. S., 2019. Overcoming long Bayesian run times in integrated fisheries stock assessments. ICES Journal of Marine Science, fsz059, https://doi.org/10.1093/icesjms/fsz059.

Morais, R.A. and Bellwood, D.R., 2018. Global drivers of reef fish growth. https://doi.org/10.1111/faf.12297.

Punt, A.E., Dorn, M.W., Haltuch, M.A., 2008. Evaluation of threshold management strategies for groundfish off the U.S. West Coast. https://doi.org/10.1016/j.fishres.2007.12.008.

Punt, A.E., Smith, A.D.M., Smith, D.C., Tuck, G.N. and Klaer, N.L. 2013. Selecting relative abundance proxies for B_{MSY} and B_{MEY}. ICES Journal of Marine Science; doi:10.1093/icesjms/fst162.

Punt, A. E., Butterworth, D. S., de Moor, C. L., De Olivera, J. A. A., & Haddon, M. (2016). Management strategy evaluation: Best Practices. Fish & Fisheries, 17, 303–334. https://doi.org/10.1111/faf.12104.

Punt, A. E., Castillo-Jordán, C., Hamel, O. S., Cope, J. M., Maunder, M. N., and Ianelli, J. N. 2021. Consequences of error in natural mortality and its estimation in stock assessment models. Fisheries Research, 233: 105759. http://www.sciencedirect.com/science/article/pii/S0165783620302769.

Quinn T and R Deriso. 1999. Quantitative fish dynamics. Oxford University Press, New York.

Sampson, D.B. and Scott, R.B. 2011. A spatial model for fishery age-selection at the population level. Canadian Journal of Fisheries and Aquatic Sciences, 68 (2011), pp. 1077-1086.

Shono, H. 2008. Application of the Tweedie distribution to zero-catch data in CPUE analysis. doi:10.1016/j.fishres.2008.03.006.

Stewart, I.J., Hicks, A.C., Taylor, I.G., Thorson, J.T., Wetzel, C., Kupschus, S., 2013. Comparison of stock assessment uncertainty estimates using maximum likelihood and Bayesian methods implemented with the same model framework. Fisheries Research 142 (2013) 37– 46.

Thorson, J. 2018. Three problems with the conventional delta-model for biomass sampling data, and a computationally efficient alternative. https://doi.org/10.1139/cjfas-2017-0266.

Thorson, J.T., Cope, J.M., 2017. Catch curve stock-reduction analysis: An alternative solution to the catch equations. https://doi.org/10.1016/j.fishres.2014.03.024.

Appendix 1: Background material provided for review

Document #	Title	Authors	Date Submitted
Documents Prepared for the Stock ID Process			
SEDAR68-SID-01	Brief Summary of FWRI-FDM Tag Recapture Program	Rachel Germeroth	8 April 2019 Updated: 3 September 2019
SEDAR68-SID-02	Larval dispersal of scamp (*Mycteroperca phenax*) in the waters off the southeastern United States: Connectivity within and between the Gulf of Mexico and Atlantic Ocean	J. R. Brothers, M. Karnauskas, C.B. Paris, and K.W. Shertzer	28 September 2019
SEDAR68-SID-03	Preliminary Genetic Stock Assessment of Scamp (Mycteroperca phenax) in Florida Waters	Elizabeth Wallace	26 July 2019 Updated: 20 September 2019
SEDAR68-SID-04	Population Genetic Analyses of Scamp	Darden, T. and M. Walker	26 July 2019 Updated: 22 August 2019
SEDAR68-SID-05	Gulf of Mexico and Atlantic Scamp Stock ID Process Final Report	Stock ID Panel	31 March 2020
Documents Prepared for the Data Workshop			
SEDAR68-DW-01	Standardized video counts of Southeast U.S. Atlantic scamp and yellowmouth grouper (*Mycteroperca phenax* and *Mycteroperca interstitialis*) from the Southeast Reef Fish Survey	Rob Cheshire and Nathan Bacheler	7 February 2020
SEDAR68-DW-02	Standardized catch rates of scamp and yellowmouth grouper (*Mycteroperca phenax* and *Myteroperca interstitialis*) in the southeast U.S. from headboat logbook data	Sustainable Fisheries Branch	4 March 2020
SEDAR68-DW-03	Standardized catch rates of scamp and yellowmouth grouper (*Mycteroperca phenax* and *Myteroperca interstitialis*) in the southeast U.S. from commercial logbook data	Sustainable Fisheries Branch	2 March 2020 Updated: 9 March 2020; 13 April 2020

SEDAR68-DW-04	Scamp/Yellowmouth Grouper Fishery-Independent Indices of Abundance in US South Atlantic Waters Based on a Chevron Video Trap Survey and a Short Bottom Longline Survey	Walter J. Bubley, Dawn Glasgow, and Tracey I. Smart	20 February 2020

SEDAR68-DW-05	Reproductive Parameters for South Atlantic Scamp and Yellowmouth Grouper in Support of the SEDAR 68 Research Track Assessment	David M. Wyanski, Dawn M. Glasgow, Keilin R. GamboaSalazar, and Wally J. Bubley	4 March 2020 Updated: 31 October 2020
SEDAR68-DW-06	Fisheries-independent data for Scamp (*Mycteroperca phenax*) from reef-fish visual surveys in the Florida Keys and Dry Tortugas, 1999-2018	Jessica Keller, Jennifer Herbig, and Alejandro Acosta	19 February 2020
SEDAR68-DW-07	Indices of abundance for Scamp (*Mycteroperca phenax*) using combined data from three independent video surveys	Kevin A. Thompson, Theodore S. Switzer, Mary C. Christman, Sean F. Keenan, Christopher Gardner, Katherine E. Overly, Matt Campbell	19 February 2020 Updated: 21 October 2020
SEDAR68-DW-08	Recreational Survey data for Scamp and Yellowmouth Grouper in the South Atlantic	Vivian M. Matter and Matthew A. Nuttall	2 March 2020 Updated: 11 March 2020 Updated: 25 August 2020 Updated: 27 October 2020
SEDAR68-DW-09	Recreational Survey data for Scamp and Yellowmouth Grouper in the Gulf of Mexico	Vivian M. Matter and Matthew A. Nuttall	2 March 2020 Updated: 11 March 2020 Updated: 25 August 2020 Updated: 27 October 2020
SEDAR68-DW-10	SEFSC computation of variance estimates for custom data aggregations from the Marine Recreational Information Program	Kyle Dettloff, Vivian M. Matter, and Matthew Nuttall	11 March 2020

SEDAR68-DW-11	Estimates of Historic Recreational Landings of Scamp and Yellowmouth Grouper in the South Atlantic Using the FHWAR Census Method	Ken Brennan	25 February 2020 Updated: 29 May 2020
SEDAR68-DW-12	Estimates of Historic Recreational Landings of Scamp and Yellowmouth Grouper in the Gulf of Mexico Using the FHWAR Census Method	Ken Brennan	25 February 2020 Updated: 29 May 2020
SEDAR68-DW-13	Marine Recreational Information Program Metadata for the Atlantic, Gulf of Mexico, and Caribbean regions	Vivian M. Matter and Matthew A. Nuttall	2 March 2020

SEDAR68-DW-14	SEAMAP Reef Fish Video Survey: Relative Indices of Abundance of Scamp	Matthew D. Campbell, Kevin R. Rademacher, Paul Felts, Brandi Noble, Joseph Salisbury, and John Moser	20 February 2020
SEDAR68-DW-15	Scamp (*Mycteroperca phenax*) age comparisons between aging labs in the Gulf of Mexico and South Atlantic	Andrew D. Ostrowski, Jennifer C. Potts, and Eric Fitzpatrick	31 March 2020
SEDAR68-DW-16	Commercial Discard Length Composition for South Atlantic Scamp and Yellowmouth Grouper	Sarina F. Atkinson	5 March 2020 Updated: 27 August 2020
SEDAR68-DW-17	Commercial Discard Length Composition for Gulf of Mexico Scamp and Yellowmouth Grouper	Sarina F. Atkinson	5 March 2020 Updated: 27 August 2020
SEDAR68-DW-18	Standardized Catch Rate Indices for Scamp (*Mycteroperca phenax*) and Yellowmouth Grouper (*Mycteroperca interstitialis*) during 1986-2017 by the U.S. Gulf of Mexico Headboat Recreational Fishery	Gulf and Caribbean Branch	2 March 2020 Updated: 9 June 2020 Updated: 10 December 2020
SEDAR68-DW-19	Scamp grouper reproduction on the West Florida Shelf	Susan LowerreBarbieri, Hayden Menendez, Ted Switzer, and Claudia Friess	4 March 2020 Updated: 2 April 2020
SEDAR68-DW-20	Summary of preliminary age, length, and reproduction data for U.S. Gulf of Mexico scamp, *Mycteroperca phenax*, submitted for SEDAR68	Veronica Beech, Laura Thornton, Beverly Barnett	3 March 2020

SEDAR68-DW-21	Summary of preliminary age and length data for U.S. Gulf of Mexico yellowmouth grouper, *Mycteroperca interstialis*, submitted for SEDAR68	Laura Thornton, Veronica Beech, Beverly Barnett	3 March 2020
SEDAR68-DW-22	Preliminary Non-Technical Fishery Profile and Limited Data Summary for Scamp, Mycteroperca phenax with Focus on the West Florida Shelf: Application of Electronic Monitoring on Commercial Snapper Grouper Bottom Longline Vessels	Carole L. Neidig, Daniel Roberts, Max Lee, Ryan Schloesser	12 March 2020
SEDAR68-DW-23	Scamp Length Frequency Distributions from At-Sea Headboat Surveys in the South Atlantic, 2005 to 2017	Dominique Lazarre, Chris Wilson, Kelly Fitzpatrick	1 April 2020
SEDAR68-DW-24	A Summary of Observer Data from the Size Distribution and Release Condition of Scamp Discards from Recreational Fishery Surveys in the Eastern Gulf of Mexico	Dominique Lazarre	1 April 2020
SEDAR68-DW-25	Summary of the SAFMC Scamp Release Citizen Science Pilot Project for SEDAR 68	Julia Byrd	16 April 2020 Updated: 26 August 2020
SEDAR68-DW-26	Voluntary reports of Scamp caught by private recreational anglers in MyFishCount for SEDAR 68	Chip Collier	7 April 2020
SEDAR68-DW-27	Assigning fates in telemetry studies using hidden Markov models: an application to deepwater groupers released with descender devices	Brendan J. Runde, Theo Michelot, Nathan M. Bacheler, Kyle W. Shertzer, and Jeffrey A. Buckel	27 February 2020
SEDAR68-DW-28	Scamp grouper reproduction in the Gulf of Mexico	Susan Lowerre-Barbieri, Veronica Beech, and Claudia Friess	22 May 2020 Updated: 2 September 2020
SEDAR68-DW-29	Standardized Catch Rate Indices for Scamp (*Mycteroperca phenax*) and Yellowmouth Grouper (*Mycteroperca interstitialis*) during 1993-2017 by the U.S. Gulf of Mexico Vertical Line and Longline Fisheries	Gulf and Caribbean Branch, SFD	11 September 2020
SEDAR68-DW-30	CPUE Expansion Estimation for Commercial Discards of Gulf of Mexico Scamp & Yellowmouth Grouper	Steven G. Smith, Kevin J. McCarthy, Stephanie Martinez	23 September 2020

SEDAR68-DW-31	SEFSC Computation of Uncertainty for Southeast Regional Headboat Survey and Total Recreational Landings Estimates, with Applications to SEDAR 68 Scamp and Yellowmouth Grouper	Matthew A Nuttall, Kyle Dettloff, Kelly E Fitzpatrick, Kenneth Brennan, and Vivian M Matter	27 October 2020
SEDAR68-DW-32	Discards of scamp (*Rhomboplites aurorubens*) for the headboat fishery in the US South Atlantic	Fisheries Ecosystems Branch, National Marine Fisheries Service, Southeast Fisheries Science Center, Beaufort, NC	30 October 2020
SEDAR68-DW-33	Discards of scamp (*Mycteroperca phenax*) for the headboat fishery in the US Gulf of Mexico	Fisheries Ecosystems Branch, National Marine Fisheries Service, Southeast Fisheries Science Center, Beaufort, NC	30 October 2020

SEDAR68-DW-34	South Atlantic U.S. scamp (*Mycteroperca phenax*) age and length composition from the recreational fisheries	Fisheries Ecosystems Branch, National Marine Fisheries Service, Southeast Fisheries Science Center	10 December 2020
SEDAR68-DW-35	Commercial age and length composition weighting for Southeast U.S. scamp and yellowmouth grouper (*Mycteroperca phenax and Mycteroperca interstitialis)*	Sustainable Fisheries Branch, National Marine Fisheries Service, Southeast Fisheries Science Center	12 November 2020
	Documents Prepared for the Assessment Process		
SEDAR68-AP-01	Gulf of Mexico Scamp (*Mycteroperca phenax*) and Yellowmouth Grouper (*Mycteroperca interstitialis*) Commercial and Recreational Length and Age Compositions	Molly H. Stevens	27 January 2021

SEDAR68-AP-02	A description of system dynamics of scamp populations in the Gulf of Mexico and South Atlantic to support ecosystem considerations in the assessment and management process	Matt McPherson and Mandy Karnauskas	29 January 2021
SEDAR68-AP-03	SEDAR 68 Commercial Discard Mortality Estimates Based on Observer Data	Jeff Pulver	9 March 2021
SEDAR68-AP-04	Estimation of a Commercial Abundance Index for Gulf of Mexico Scamp & Yellowmouth Grouper Using Reef Fish Observer Data	Steven G. Smith, Skyler Sagarese, Stephanie MartinezRivera, Kevin J. McCarthy	29 March 2021
	Documents Prepared for the Review Workshop		
SEDAR68-RW-01	Modeling of recreational landings in Gulf stock assessments	Gulf Branch – Sustainable Fisheries Division	10 August 2021
	Final Stock Assessment Reports		
SEDAR68-SAR1	Gulf of Mexico Scamp	SEDAR 68 Panels	
SEDAR68-SAR2	Atlantic Scamp	SEDAR 68 Panels	
	Reference Documents		
SEDAR68-RD01	A retrospective (1979-1996) multispecies assessment of coral reef fish stocks in the Florida Keys	Ault et al. 1997	

SEDAR68-RD02	Spawning Locations for Atlantic Reef Fishes off the Southeastern U.S.	Sedberry et al. 2006
SEDAR68-RD03	Site Fidelity and Movement of Reef Fishes Tagged at Unreported Artificial Reef Sites off NW Florida	Addis et al. 2007
SEDAR68-RD04	Implications of reef fish movement from unreported artificial reef sites in the northern Gulf of Mexico	Addis et al. 2013
SEDAR68-RD05	Comparison of scamp grouper (*Mycteroperca phenax*), growth off of the West Florida shelf and the coast of Louisiana	Bates 2008

SEDAR68-RD06	Aspects Of The Life History Of The Yellowmouth Grouper, *Mycteroperca interstitialis*, In The Eastern Gulf Of Mexico	Bullock and Murphy, 1994
SEDAR68-RD07	Memoirs of the Hourglass Cruises: Seabasses (Pisces: Serranidae)	Bullock and Smith, 1991
SEDAR68-RD08	Groupers on the Edge: Shelf Spawning Habitat in and Around Marine Reserves of the Northeastern Gulf of Mexico	Coleman et al. 2014
SEDAR68-RD09	Decadal fluctuations in life history parameters of scamp (*Mycteroperca phenax*) collected by commercial handline vessels from the west coast of Florida	Lombardi-Carlson et al.
SEDAR68-RD10	A Description of Age, Growth, and Reproductive Life History Traits of Scamps from the Northern Gulf of Mexico	Lombardi-Carlson et al. 2012
SEDAR68-RD11	Incorporating Mortality from Catch and Release into Yield-per-Recruit Analyses of Minimum-Size Limits	Waters and Huntsman 1986
SEDAR68-RD12	Population genetic analysis of red grouper, *Epinephelus morio*, and scamp, *Mycteroperca phenax*, from the southeastern U.S. Atlantic and Gulf of Mexico	Zatcoff et al. 2004
SEDAR68-RD13	Population Assessment of the Scamp, *Mycteroperca phenax*, from the Southeastern United States	Mancooch et al. 1998
SEDAR68-RD14	A Preliminary Assessment of the Populations of Seven Species of Grouper (Serranidae, Epinephelinae) in the Western Atlantic Ocean from Cape Hatteras, North Carolina to the Dry Tortugas, Florida	Huntsman et al.

SEDAR68-RD15	Color Variation And Associated Behavior In The Epinepheline Groupers, *Mycteroperca microlepis* (Goode And Bean) And *M. Phenax* Jordan And Swain	Gilmore and Jones 1992
SEDAR68-RD16	Age, Growth, and Reproduction of Scamp, *Mycteroperca phenax*, in the Southwestern North Atlantic, 1979 – 1997	Harris et al. 2002

SEDAR68-RD17	Age, Growth, Mortality, Food and Reproduction of the Scamp, *Mycteroperca phenax*, Collected off North Carolina and South Carolina	Matheson et al. 1986
SEDAR68-RD18	Tagging Studies and Diver Observations of Fish Populations on Live-Bottom Reefs of the U.S. Southeastern Coast	Parker 1990
SEDAR68-RD19	Age and growth of the yellowedge grouper, *Epinephelus flavolimbatus*, and the yellowmouth grouper, *Mycteroperca interstitialis*, off Trinidad and Tobago	Manickchand-Heileman and Phillip 2000
SEDAR68-RD20	Multi-decadal decline in reef fish abundance and species richness in the southeast USA assessed by standardized trap catches	Bachelor and Smart 2016
SEDAR68-RD21	Aspects Of The Life History Of The Yellowmouth Grouper, *Mycteroperca interstitialis*, In The Eastern Gulf Of Mexico	Bullock and Murphy 1994
SEDAR68-RD22	Age, Growth, and Mortality of Yellowmouth Grouper from the Southeastern United States	Burton et al. 2014
SEDAR68-RD23	South Carolina Marine Game Fish Tagging Program 1978 -2009	Robert K. Wiggers
SEDAR68-RD24	Decadal-scale decline of scamp (Mycteroperca phenax) abundance along the southeast United States Atlantic coast	Nathan M. Bacheler and Joseph C. Ballenger
SEDAR68-RD25	Timing and locations of reef fish spawning off the southeastern United States	Nicholas A. Farmer, William D. Heyman, Mandy Karnauskas, Shinichi Kobara, Tracey I. Smart, Joseph C. Ballenger, Marcel J. M. Reichert, David M. Wyanski, Michelle S. Tishler, Kenyon C. Lindeman, Susan K. Lowerre-Barbieri, Theodore S. Switzer, Justin J. Solomon, Kyle McCain, Mark Marhefka, George R. Sedberry
SEDAR68-RD26	Developmental patterns within a multispecies reef fishery: management applications for essential fish habitats and protected areas	Kenyon C. Lindeman, Roger Pugliese, Gregg T. Waugh, and Jerald S. Ault
SEDAR68-RD27	Ingress of postlarval gag, Mycteroperca microlepis (Pisces: Serranidae)	Paula Keener, G. David Johnson, Bruce

		W Stender, Edward B. Brothers and Howard R. Beatty
SEDAR68-RD28	Survival estimates for demersal reef fishes released by anglers	Mark R. Collins
SEDAR68-RD29	Commercial catch composition with discard and immediate release mortality proportions off the southeastern coast of the United States	Jessica A. Stephen, Patrick J. Harris
SEDAR68-RD30	Discard composition and release fate in the snapper and grouper commercial hook-and-line fishery in North Carolina, USA	P.J. Rudershausen, J.A. Buckel, and E.H. Williams
SEDAR68-RD31	Sink or swim? Factors affecting immediate discard mortality for the Gulf of Mexico commercial reef fish fishery	J.R. Pulver
SEDAR68-RD32	SEDAR 33-DW-19: A meta-data analysis of discard mortality estimates for gag grouper and greater amberjack	Linda Lombardi, Matthew D. Campbell, Beverly Sauls, and Kevin J. McCarthy
SEDAR68-RD33	Potential survival of released groupers caught deeper than 40 m based on shipboard and in-situ observations, and tag-recapture data	Raymond R. Wilson, Jr. and Karen M. Burns
SEDAR68-RD34	Scamp Fishery Performance Report	SAFMC Snapper Grouper Advisory Panel
SEDAR68-RD35	Hierarchical analysis of multiple noisy abundance indices	Paul B. Conn
SEDAR68-RD36	SAFMC SSC MRIP Workshop Report	SAFMC SSC
SEDAR68-RD37	Catch Characterization and Discards within the Snapper Grouper Vertical Hook-and-Line Fishery	Gulf and South Atlantic Fisheries Foundation
SEDAR68-RD38	A Continuation of Catch Characterization and Discards within the Snapper Grouper Vertical Hook-and-Line Fishery	Gulf and South Atlantic Fisheries Foundation
SEDAR68-RD39	Continuation of Catch Characterization and Discards within the Snapper Grouper Vertical Hook-and-Line Fishery	Gulf and South Atlantic Fisheries Foundation
SEDAR68-RD40	Descender Devices are Promising Tools for Increasing Survival in Deepwater Groupers	Brendan J. Runde and Jeffrey A. Buckel
SEDAR68-RD41	Something’s Fishy with Scamp Response Summary	GMFMC

SEDAR68-RD42	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 III1
SEDAR68-RD43	Length selectivity of commercial fish traps assessed from in situ comparisons with stereo-video: Is there evidence of sampling bias?	Tim J. Langlois, Stephen J. Newman, Mike Cappo, Euan S. Harvey, Ben M. Rome, Craig L. Skepper, Corey B. Wakefield
SEDAR68-RD44	Changes in Reef Fish Community Structure Following the Deepwater Horizon Oil Spill	Justin P. Lewis, Joseph H. Tarnecki, Steven B. Garner, David D. Chagaris &William F. Patterson III

Appendix 2: Performance Work Statement

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

SEDAR 68 Atlantic and Gulf of Mexico Scamp Assessment Review

Background

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

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

Scope

The **SouthEast Data, Assessment, and Review (SEDAR)** is the cooperative process by which stock assessment projects are conducted in NMFS' Southeast Region. SEDAR was initiated to improve planning and coordination of stock assessment activities and to improve the quality and reliability of assessments.

SEDAR 68 will be a CIE assessment review conducted for Atlantic and Gulf of Mexico Scamp Grouper. There are two separate models to be reviewed: one for the US Atlantic, and one for the Gulf of Mexico. 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 sensitivity runs of the assessment models provided by the assessment panel. The review panel is ultimately responsible for ensuring that the best possible assessment is provided through the SEDAR process. The specified format and contents of the individual peer review reports are found in **Annex 1**. The Terms of Reference (TORs) of the peer review are listed in **Annex 2**. The tentative agenda of the panel review meeting is attached in **Annex 3** and the technical specifications required for this review are listed in **Annex 4.**

Requirements
NMFS requires three (3) reviewers to conduct an impartial and independent peer review in accordance with the Performance Work Statement (PWS), OMB guidelines, and the TORs 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 fisheries stock assessment.

Tasks for Reviewers

1) Two weeks before the peer review, the Project Contacts will send (by electronic mail or make available at an FTP site) to the CIE reviewers the necessary background information and reports for the peer review. In the case where the documents need to be mailed, the 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 to the PWS scheduled deadlines specified herein. The CIE reviewers shall read all documents in preparation for the peer review.
2) Additionally, during the week of August 16, 2021 prior to the peer review, the CIE reviewers will participate in a test to confirm that they have the necessary technical (hardware, software, etc.) capabilities to participate in the virtual panel in advance of the review meeting. This review's Project Contacts will provide the information for the arrangements for this test.
3) Attend and participate in a virtual 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.
4) After the review meeting, reviewers shall conduct an independent peer review report in accordance with the requirements specified in this PWS, OMB guidelines, and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus.
5) Each reviewer should assist the Chair of the meeting with contributions to the summary report.
6) Deliver their reports to the Government according to the specified milestones dates.

Place of Performance
The place of performance shall be online via GoToWebinar.

Period of Performance
The period of performance shall be from the time of award through October 2021. 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.

Schedule	Milestones and Deliverables
Within two weeks of award	Contractor selects and confirms reviewers
2 weeks prior to the panel review	Contractor provides the pre-review documents to the reviewers
August 30-31 and September 1-3 2021	Panel will attend and participate in review webinars lasting approximately 7 hours one Days One and Two, and four and a half hours each of the remaining days. Webinars will be held between the hours of 8 am -8 pm ET
Approximately 3 weeks later	Contractor receives draft reports
Within 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 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
Since this is a virtual panel review travel is neither required nor authorized for this contract.

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

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

Julie A Neer - SEDAR Coordinator
Science and Statistics Program
South Atlantic Fishery Management Council
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 the best scientific information available.

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 TORs.

   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 that they believe might require further clarification.

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

   e. The 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 68 Atlantic and Gulf of Mexico Scamp Assessment
Review Workshop Terms of Reference

Review Workshop Terms of Reference

1. Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions. Consider the following:
   - Are data decisions made by the DW and AW 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 and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account 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?

3. Consider how uncertainties in the assessment, and their potential consequences, are addressed.
   - Comment on the degree to which methods used to evaluate uncertainty reflect and capture the significant sources of uncertainty in the population, data sources, and assessment methods.
   - Comment on the likely relationship of this variability with possible ecosystem or climate factors and possible mechanisms for encompassing this into management reference points.

4. Provide, or comment on, recommendations to improve the assessment
   - Consider the research recommendations provided by the Data and Assessment workshops 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. These recommendations should be described in sufficient detail for application, and should be practical for short-term implementation (e.g., achievable within ~6 months). Longer-term recommendations should instead be listed as research recommendations above.

5. Provide recommendations on possible ways to improve the Research Track Assessment process.

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

Annex 3: Tentative Agenda - SEDAR 68 Atlantic and Gulf of Mexico Scamp Assessment Review
Via webinar
August 30 - September 3, 2021

Each of the first two days will likely consist of a 7-hour long webinar held between the times of 8 am and 8 pm ET
The remaining days will likely consist of 4.5 hour long webinars
The start and end times of each webinar are dependent on CIE and analyst availability

August 30- Introductions and Opening Remarks		Coordinator
	- Agenda Review, TOR, Task Assignments	
	Assessment Presentations	Lead Analysts
August 31 – Assessment Presentation continued		Lead Analysts

August 30 - 31 Goals: Initial presentations completed, sensitivities and modifications identified.

September 1 -	Panel Discussion	Chair
	- Review additional analyses, sensitivities	
	- Consensus recommendations and comments	Chair

September 1 Goals: Final sensitivities identified, preferred models selected, projection approaches approved, Summary report drafts begun

September 2 - Panel Discussion		Chair
	- Final sensitivities reviewed.	
	- Projections reviewed.	
September 3 Panel Discussion or Work Session		Chair
	- Review Consensus Reports	

September 2 and 3 Goals: Complete assessment work and discussions. Final results available. Draft Summary Report reviewed.

Annex 4: SEDAR 68 Atlantic and Gulf of Mexico Scamp Review Workshop minimum technical requirements

1. Computer
2. Microphone and speakers (headset recommended)
3. GoToWebinar desktop app (JavaScript enabled) available for download here: https://support.goto.com/webinar/help/download-now-g2w010002
4. Internet: 1 Mbps or better (wired preferred)
5. Web browser:
   a. Google Chrome v57 or later
   b. Mozilla Firefox v52 or later
   c. Internet Explorer v10 or later
   d. Microsoft Edge v12 or later
   e. Apple Safari v10 or later
6. Operating system
   a. Windows 7 - Windows 10
   b. Mac OS X 10.9 (Mavericks) - macOS 10.15 (Catalina)
7. 2GB of RAM (minimum), 4GB or more of RAM (recommended)
8. Smart phone for use as audio backup and internet hotspot (recommended)