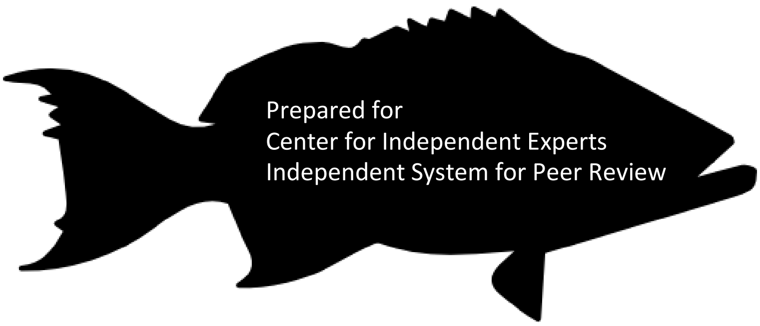
Report of the SEDAR 68 Atlantic and Gulf of Mexico Scamp Assessment Review

Robin Cook



Contents

Executive summary 2
Background 4
Description of the individual reviewers' roles in the review activities 4
Findings under the Review Workshop Terms of Reference 4
Gulf scamp 4
Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions. 4
Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account the available data. 5
Consider how uncertainties in the assessment, and their potential consequences, are addressed. .7
Provide, or comment on, recommendations to improve the assessment 7
Atlantic scamp 8
Evaluate the data used in the assessment, including discussion of the strengths and weaknesses of data sources and decisions. 8
Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account the available data. 9
Consider how uncertainties in the assessment, and their potential consequences, are addressed. 11
Provide, or comment on, recommendations to improve the assessment 11
Provide recommendations on possible ways to improve the Research Track Assessment process. 12
Conclusions and recommendations 12
NMFS review process...13
References 13
Appendix 1. Materials provided for the review...14
Appendix 2. Performance Work Statement (PWS)...23
Appendix 3. Panel Membership 31
Appendix 4. An assessment of Gulf and Atlantic scamp using a Schaefer surplus production model .32

Executive summary

1) A review of draft research track assessments for Gulf of Mexico and Atlantic scamp was held during a virtual meeting from the 30th August-3rd September 2021 as part of the SEDAR 68 process. Three CIE reviewers participated in the review panel.
2) For both assessments the panel supported the decisions made by the data workshop (DW) and Assessment workshop (AW). Available data comprised landings, estimates of discards, one fishery independent survey, a number of fishery dependent surveys, and age and length compositions. These data are sufficient to support the assessment methods for each stock.
3) In both assessments the landings data are assumed to be known with very high precision and this is likely to be influential in the assessment results. It is noteworthy that in the Gulf assessment the variance of the expected values for all catch components and the commercial vertical line survey was higher than the raw data and this is somewhat counter-intuitive result needs to be investigated.

Gulf assessment

4) The Gulf assessment adopted Stock Synthesis as the main assessment tool. In the draft base model, a number of data sets had poor residual patterns, notably for the commercial vertical line cpue index and the RFOP index. An overall improved fit was achieved by including age compositions rather than conditional age at length data, estimating more of the selectivity parameters, reducing the number of size bins and allowing the model to estimate additional variance for the surveys. Nevertheless, the improvements were modest and much lack of fit was still evident.
5) Trends in summary quantities such as F and SSB appear insensitive to a range of model configurations and assumptions. However, jitter analysis suggests the minimum of the objective function is not well defined and retrospective analysis shows major revisions to F and SSB as data are removed, primarily to the scale of the SSB. Hence the assessment results cannot be considered robust to the addition of new data.
6) It was not possible to estimate steepness in the Beverton-Holt stock recruitment function and this was fixed at a weighted mean of the value estimated from meta-analyses and the Atlantic scamp assessment. Estimates of MSY reference points will be conditioned on this assumption.

Atlantic assessment

7) The Atlantic assessment used the Beaufort Assessment Model (BAM) which is appropriate for the data available as it can integrate a range of different data types. The draft assessment had modelled landings and discards as separate fleets. This is probably not the most realistic way to model the observations. A revised base model combined landings and discards into total catch by fleet to avoid this problem.
8) Compared to the Gulf assessment, the model better fitted the data, especially the abundance indices. Jitter analysis indicated a well-defined minimum for the objective function. Retrospective analysis also indicated consistency in both F and SSB.
9) An area of concern was the change in selectivity pattern after 1992. Here more recent selectivity as estimated by the model retained more younger fish than the earlier period which appears to be in conflict with the increase in permitted landing size. It was not possible to resolve this issue at the meeting, but it appeared that the adoption of two selectivity time blocks to reflect the change in size regulations was necessary to obtain a satisfactory model fit.

10) It was possible to estimate steepness for the stock-recruitment function but this value was sensitive to the specification of the number of selectivity time blocks.

11) The assessment included an ensemble modelling approach to quantify uncertainty in quantities of interest. The method included bootstrapping the input data and using Monte Carlo methods to explore uncertainty in biological constants such as discard mortality and natural mortality. This is a useful additional tool to quantify uncertainty around the base model and should give more realistic estimates of uncertainty than asymptotic CVs.

Conclusions

12) Both assessments represent high quality science and the best available. Improvements were made during the review but further work is still in progress. Of the two assessments, the Atlantic appears more robust.

Recommendations

13) Only one fishery independent survey is available for each stock. More work is needed to refine this index which needs to be continued to support future assessments. The fishery dependent surveys should be investigated for evidence of technological creep and corrected if necessary as the headboat indices play an important role in the assessments.

14) Selectivity parameters are crucial in these assessments but there was no experimental evidence to inform the choice of selectivity function. It would be desirable to undertake field experiments to support the form of the selection function in the assessments.

15) As these were research track assessments, in future it would be desirable to explore the data more thoroughly using a range of models before identifying a final model and how best to configure it.

Background

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 is a CIE assessment review conducted for Atlantic and Gulf of Mexico Scamp Grouper. Two separate models were reviewed: one for the US Atlantic, and one for the Gulf of Mexico. The panel (Appendix 3) held a virtual meeting from the 30th August to the 3rd September where draft assessments were reviewed. Additional assessment runs were requested by the panel to investigate alternative model configurations and hypotheses concerning model structure and data inputs. This led to revised base models for both assessments

Description of the individual reviewers' roles in the review activities

Prior to the meeting the draft Assessment Workshop (AW) reports for the two assessments were received and reviewed. A list of additional documents was also received and reviewed as appropriate (Appendix 1). Approximately one week before the full Review Workshop (RW) a preliminary virtual meeting was held to test the GoToWebinar arrangements and have preliminary discussions with the assessment analysts. The reviewer gave some initial thoughts on the Atlantic assessment to the analysists at this point. At the full RW meeting the reviewer participated fully and made a number of requests for additional runs. For the Gulf this included runs without the RFOP survey and changes to the data weighting. For the Atlantic I requested an additional run to combine the landings and discard data in order to avoid modelling these as separate fleets.

During the meeting the reviewer provided preliminary text for the Summary Report on both assessments relating to ToR#3. Following the meeting the reviewer assisted with the preparation of the Summary Report. The statement of work is given in Appendix 2.

Findings under the Review Workshop Terms of Reference

Gulf scamp

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

The data available consisted of reported landings from the commercial fleets, estimates of catch derived from surveys for the recreational fleets, length and age compositions covering some years and both fishery independent and fishery dependent indices. Estimates of discards and their mortality were also available. Biological information on growth and transition from female to male were obtained external to the assessment. Natural mortality estimates were derived from standard methods (Lorenzen and Then) based on life history characteristics and meta-analyses.

For the catch data and the survey indices, CVs were estimated externally from the assessment model and used as in initial estimate of uncertainty. These were used in the assessment model and affect the weighting of data in the objective function. Such estimates are based either on the sampling design implemented for the acquisition of data or modelling and do not necessarily capture all sources of uncertainty, such as changes in the spatial distribution of fish or the influence of time varying environmental effects. They are therefore generally minimum estimates. To some degree this is recognised in the assessment model used which can estimate additional variability. For the commercial landings data, a very low CV was assumed which, in effect, means that these data are treated as constants for the purposes of assessment. There is a justifiable belief that these data are the most precise but the extremely low CV assumed is very low even for the most sophisticated landings recording system and a figure closer to 10% or even 20% seems more likely.

The abundance indices included fishery dependent series. These were derived from model based analyses that seek to standardize CPUE by accounting for fleet behaviour, such as changes in fleet distribution and target species. This is a necessary and desirable procedure that, *inter alia*, can provide estimates of precision. However, the indices are not corrected of technological creep and there is a danger that recent CPUE over-estimates abundance relative to early estimates. Some analysis of this issue would be desirable especially as the headboat index appears to be influential in the assessment.

Given the data available a number of methods could be applied that include data limited methods making use of survey and catch data only, simple surplus production models that can account for density dependence in biological traits, length only or age only methods, and fully integrated approaches that seek to include all sources of data in a single framework. The analyst chose the latter option in the form of Stock Synthesis (SS). This approach was designed for data of this type and hence the data are sufficient for the chosen method. It does, however, require a number of quite strong assumptions about growth, maturation, natural mortality and, in this case the transition from female to male. These were assumed to be time invariant but are likely to vary and there is naturally uncertainty about the extent to which the simplifying assumption of constancy is adequate.

Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account the available data.

The principal assessment model used was the current version of Stock Synthesis which is based around an age structured population dynamics model. It is well established and has been in use for many years. It has been published in peer reviewed literature and can be regarded as scientifically sound in the hands of a competent practitioner. It may be viewed as a modelling framework within which a wide variety of models can be configured. Hence the validity of the model implemented and its robustness will depend upon the way it is configured and the data involved. For the Gulf assessment the model was configured following standard practice with high precision given to the landings data, Dirichlet-Multinomial weighting given to the composition data, reweighting of the abundance indices, and assumed Beverton-Holt stock-recruitment function with fixed steepness. This is a scientifically sound approach. The robustness of the model can in part be judged from the diagnostics of model fit and performance discussed below. Robustness needs also to be evaluated in the context in which it is to be used. Overall, I felt that if the model was to be used for advice then considerable caution is necessary due to the uncertainty in the scale of the biomass. At present the assessment is for research purposes and some of the issues arising may be resolved by the time of an operational assessment.

The principal strength of the assessment approach is the ability to incorporate a disparate range of data into a single modelling framework and hence extract as much information as possible. The specific software tool also provides a sophisticated range of diagnostics that help understand model performance and quantify much of the uncertainty. In this particular assessment there is information on abundance from four surveys, information on age structure from sampled length and age compositions and a long series of landings data that are assumed to be precise. Thus while some of the abundance and composition data are missing in some years, there is a reasonably comprehensive range of data in the assessment. Results from a range of sensitivity runs and additional model runs undertaken during the review to test alternative configurations all produced very similar trends in summary population quantities such as spawning stock biomass and fishing mortality. Thus the gross trends appear robust to a range of alternative SS models. At least some of this robustness is attributable to the assumption that landing data are known with high precision.

Model diagnostics illustrate a range of potential weaknesses in the assessment. Perhaps two of these provide reasons for caution in interpretation of the results. Firstly, the results of jitter analysis indicate that a relatively small number of runs converged on the same minimum of the base model log-likelihood. While these runs all tended to show a similar trend in SSB and F there are changes of scale that are quite large. Furthermore, it is unclear whether some of the latent parameters such as virgin biomass, selectivity and retention are unique as a number of these were poorly estimated in the jitter runs. Unless these parameters show consistency between jitters, there is a danger that reference point calculations will be compromised. Secondly, the results of retrospective runs show marked changes of scale as successive years of data are removed and this is consistent with the low proportion of convergence to the base run minimum for virgin biomass in the jitter analysis. Although the gross trends remain similar this is indicative of a problem that when new data are added to the assessment, there is likely to be a revision of scale, and this may be important for reference point calculations and assessment of stock status. Overall, it is clear there is insufficient information in the data to estimate the scale of the biomass well.

An important element of SS assessments is the choice of stock recruitment function which is ultimately critical in the calculation of reference points. In this assessment attempts were made to estimate steepness, a crucial parameter in the Beverton-Holt stock recruitment function. It is clear there is no information in the data to estimate this parameter and the RW agreed that it should be fixed as a constant based on meta-analyses and the value estimated in the Atlantic assessment where this parameter was estimable. From the perspective of estimating stock trends steepness is not important in this assessment but it does have a bearing on the use of the model for forecasting and the evaluation of stock status.

An analysis performed by the reviewer presented in Appendix 4 using a surplus production model (SPM) re-enforces some of the properties of the assessment. This analysis uses only survey and landings data. It suggests that the RFOP survey has increasing catchability relative to the video survey which would explain the difficulty fitting these data in the SS model. The analysis also highlights the indeterminacy in unexploited biomass. The SPM generally fits the survey data well and while estimating a similar biomass trajectory, the estimated trend in F is rather different with higher values in recent years compared with the SS assessment. However, the CVs of the estimated quantities are very large suggesting substantial uncertainty in the assessment.

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

There are perhaps three classes of consideration in relation to uncertainty. These are the conditioning assumptions, the overall goodness of fit, and estimates of precision of the model parameters and quantities of interest (SSB, F etc). Some of the uncertainty in the conditioning assumptions were addressed with sensitivity runs that considered alternative values of M, steepness, male contribution to SSB and the estimation of growth parameters. Given that there are multiple plausible alternative assumptions it is extremely difficult to explore fully the range of uncertainty. With the collection of sensitivity runs undertaken it appears that the estimated trends are insensitive to alternative hypotheses. Additional runs undertaken at the request of the review panel that included estimating the selectivity parameters and omitting the RFOP survey did improve model fit but with relatively small changes to the main stock trends.

Although a number of changes made to the AW base model resulted in a new base model that better fitted the data, there was still a lack of fit notably to the commercial vertical line survey and the age compositions. This will affect the quality of the estimated CVs for the quantities of interest since these are based on the shape of the likelihood at its minimum where it is assumed the underlying statistical assumptions are satisfied. This will not be the case where there are poor fits to the data. The jitter runs suggest the likelihood at the minimum is poorly defined in which case the approximation used to estimate parameter variance may not be adequate and a MCMC approach would be a more robust method.

The assumed CVs for the landings are very low. Despite this, the fitted catches for all categories of catch showed higher variability than the input data. In a trial run when the CVs on the landings was increased to 0.3, this problem became even worse. Clearly if the CV of the landings data is under-estimated the CVs of the estimated quantities of interest will be too low. In my opinion this is a likely occurrence and the apparent inability to fit the catch data without imposing a very low CV is a cause for concern.

Sensitivities to the data were explored in relation to the survey data and the age data. Better model fits were obtained with the age data entered as age compositions rather than conditional age-at-length. The RFOP survey shows a slightly increasing biomass trend in recent years in contrast to other surveys covering the same time period. Runs without the RFOP survey did not show any major change. After much discussion the panel agreed it should be included as the survey was down-weighted in the likelihood and had little overall influence.

Provide, or comment on, recommendations to improve the assessment

Recommendations from the DW and AW were briefly reviewed at the meeting and supported. I missed any strong recommendation to develop or strengthen survey abundance indices. Such indices are essential for a well-conditioned assessment. The current Gulf scamp assessment only makes effective use of the video survey and the headboat survey. The former is not fit particularly well yet is the only fishery independent survey. Further work to refine this index and a commitment to its future operation are important. The headboat index is also important as it is the only other index that is reasonably well fit by the model. Although it is standardized it is not explicitly corrected for technological creep, an issue that deserves investigation to avoid bias over time.

Selectivity parameters are influential parameters in the model. In this assessment *a priori* assumptions were made about the shape of the selection curve which, while reasonable, need to be supported by experimental evidence. It would be desirable to conduct field experiments to identify the appropriate selectivity for the fleets concerned.

Since this is a research track assessment it would be useful to explore a number of simpler assessment approaches to explore the data in order to understand them better before building a complex and highly parameterized SS model. For example, a full age structured model can be used to analyse the survey data alone (e.g. Cotter et al 2007) while almost any surplus production model can be used to investigate biomass indices, with or without catch data (Pedersen and Berg, 2017, Cook et al, 2021, Appendix 4). These models can be applied to explore the consistency of information in the catch and index data and help in the specification of more complex models. It would also be useful to investigate the length frequency data with length based methods such as LIME (Rudd and Thorson, 2017) to gain insights into the quality of the data.

Atlantic scamp

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

The data available consisted of reported landings from the commercial fleets, estimates of catch derived from surveys for the recreational fleets, length and age compositions covering some years and both fishery independent and fishery dependent indices. Estimates of discards and their mortality were also available. Biological information on growth and transition from female to male were obtained external to the assessment. Natural mortality estimates were derived from standard methods (Lorenzen and Then) based on life history characteristics and meta-analyses.

For the catch data and the survey indices, CVs were estimated externally from the assessment model and used as in initial estimate of uncertainty. These were used in the assessment model and affect the weighting of data in the objective function. Such estimates are based either on the sampling design implemented for the acquisition of data or modelling and do not necessarily capture all sources of uncertainty, such as changes in the spatial distribution of fish or the influence of time varying environmental effects. They are therefore generally minimum estimates. To some degree this is recognised in the assessment model used which can estimate additional variability. For the commercial landings data, a very low CV was assumed which, in effect, means that these data are treated as constants for the purposes of assessment. There is a justifiable belief that these data are the most precise but the extremely low CV assumed is very low even for the most sophisticated landings recording system and a figure closer to 10% or even 20% seems more likely.

The abundance indices included fishery dependent series. These were derived from model based analyses that seek to standardize CPUE by accounting for fleet behaviour, such as changes in fleet distribution and target species. This is a necessary and desirable procedure that, *inter alia*, can provide estimates of precision. However, the indices are not corrected of technological creep and there is a danger that recent CPUE over-estimates abundance relative to early estimates. Some analysis of this issue would be desirable especially as the recreational headboat index appears to be influential in the assessment.

Given the data available a number of methods could be applied that include data limited methods making use of survey and catch data only, simple surplus production models that can account for density dependence in biological traits, length only or age only methods, and fully integrated approaches that seek to include all sources of data in a single framework. The analyst chose the latter option in the form of the Beaufort Assessment Model (BAM). This approach was designed for data of this type and hence the data are sufficient for the chosen method. It does, however, require a number of quite strong assumptions about growth, maturation, natural mortality and, in this case the transition from female to male at older ages. These were assumed to be time invariant but are likely to vary and there is naturally uncertainty about the extent to which the simplifying assumption of constancy is adequate.

Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account the available data.

The principal assessment model used was the current version of BAM which is based around an age structured population dynamics model and shares many of the characteristics of Stock Synthesis. It is well established and has been in use for many years in this region. It has been published in peer reviewed literature and can be regarded as scientifically sound in the hands of a competent practitioner. It may be viewed as a modelling framework within which a wide variety of models can be configured. Hence the validity of the model implemented and its robustness will depend upon the way it is configured and the data involved. For the Atlantic assessment the model was configured following standard practice with high precision given to the landings data, Dirichlet-Multinomial weighting given to the composition data, reweighting of the abundance indices, and assumed Beverton-Holt stock-recruitment function with estimated steepness. This is a scientifically sound approach. The robustness of the model can in part be judged from the diagnostics of model fit and performance discussed below. Robustness needs also to be evaluated in the context in which it is to be used. Overall, I felt that the model was robust given the satisfactory convergence in the jitter runs and the retrospective analysis. There is some uncertainty in steepness estimates which appeared to be sensitive to the choice of selectivity time blocks and this may be important in reference point calculations. At present the assessment is for research purposes and this issue may be resolved by the time of an operational assessment.

The principal strength of the assessment approach is the ability to incorporate a disparate range of data into a single modelling framework and hence extract as much information as possible. The specific software tool also provides a range of diagnostics that help understand model performance and quantify much of the uncertainty. In this particular assessment there is information on abundance from three surveys, information on age structure from sampled length and age compositions and a long series of landings data that are assumed to be precise. Thus, while some of the abundance and composition data are missing in some years, there is a reasonably comprehensive range of data in the assessment. Results from a series of sensitivity runs and additional model runs undertaken during the review to test alternative configurations all produced similar trends in summary population quantities such as spawning stock biomass and fishing mortality. The gross trends, therefore, appear robust to a range of alternative models. At least some of this robustness is attributable to the assumption that landing data are known with high precision. In contrast to the Gulf assessment, the jitter analysis for the BAM model showed consistency in reaching the same minimum log-likelihood, thus providing more confidence in the estimated quantities.

In addition to the conventional approach of using sensitivity analysis to investigate uncertainty, this assessment included an ensemble modelling method that considered errors in the data as well as uncertainties in some of the conditioning assumptions (e.g. M, discard mortality etc). This is an important step forward in deriving more realistic posterior distributions of the quantities of interest.

The results of retrospective runs showed a high degree of consistency in SSB and F but with a notable spike in F in 2014. Thus the assessment appears reliable in its estimate of SSB but with some modest uncertainty on the terminal estimate of F.

An important element of the assessment is the choice of stock recruitment function which is ultimately critical in the calculation of reference points. In this assessment steepness, a crucial parameter in the Beverton-Holt stock recruitment function, was estimable. It is clear that there is some information in the data to estimate this parameter since recent recruitment and SSB values appear to lie close to the descending limb of the stock-recruitment relationship. While this is encouraging it is noteworthy that the estimates of steepness were sensitive to other model assumptions such as the number of time blocks used to estimate selectivity parameters. Recent recruitment will also be subject to higher uncertainty which will inevitably affect the estimate of steepness. From the perspective of estimating stock trends steepness is less important in this assessment but it does have a bearing on the use of the model for forecasting and the evaluation of stock status.

One of the main issues explored at the RW related to the choice of time blocks. These were originally chosen as two blocks coinciding with a change in fishery regulations where the minimum landing size was increased from 1992 onwards. This might be expected to result in a selectivity curve shifted to the right in recent years. However, the model estimated the reverse despite observed landing length frequencies where the mode has shifted right over time. BAM estimates selectivity by age rather than length and this may explain the result if size at age has changed. Exploratory runs that considered from one to six time blocks did not resolve the apparent inconsistency but it was clear that the two time block model fit the data best and was retained in the new base model. Further investigation of this apparent anomaly is still required, however.

In the AW base model discards and landings were modelled as two separate fleets with independent selectivity functions. This means that the estimated selectivities are a combination of fleet selectivity and a retention function. By estimating overall selectivity independently for each catch category (i.e., landings and discards) there may be inconsistency with the actual fleet selectivity and retention function between the two categories for a given fleet. As BAM does not currently offer an option to model discards as arising from a single fleet with a retention function, the panel requested a run with landings and discards combined for each fleet. This run gave similar results to the original AW base run but without the need to estimate additional selectivity parameters and the configuration was retained for a revised based model.

An analysis performed by the reviewer presented in Appendix 4 using a surplus production model (SPM) supports some of the properties of the assessment, and in particular, its greater consistency compared to the Gulf assessment. This analysis uses only survey and landings data. It suggests that fishery dependent surveys have increasing catchability relative to the video survey. Here unexploited biomass is well determined and the retrospective pattern (Appendix 4, Figure 4) is well behaved as with the BAM assessment. The SPM generally fits the survey data well and while estimating a similar biomass trajectory, the estimated trend in F is rather different with much higher values in recent years compared with the BAM assessment, perhaps as a result of the fishing power increment adjustment for the commercial CPUE indices.

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

There are perhaps three classes of consideration in relation to uncertainty. These are the conditioning assumptions, the overall goodness of fit, and estimates of precision of the model parameters and quantities of interest (SSB, F etc). Some of the uncertainty in the conditioning assumptions were addressed with sensitivity runs that considered alternative values of M, steepness, male contribution to SSB and aging error. Given that there are multiple plausible alternative assumptions it is extremely difficult to explore fully the range of uncertainty. With the collection of sensitivity runs undertaken it appears that the estimated trends are relatively insensitive to alternative hypotheses. Additional runs undertaken at the request of the review panel included changing the selectivity blocks, using domed selectivity, and combining landings and discards. Changing selectivity blocks had one of the largest effects on the estimate of steepness as a result of much lower values of recruitment in recent years. These steepness values were around 0.27 compared to ca 0.57 for the base model and appear unrealistically low. The two block configuration was therefore retained.

Although a number of changes made to the AW base model resulted in a new base model that better fitted the data, there was still some lack of fit notably to the composition data. This will affect the quality of the estimated CVs for the quantities of interest since these are based on the shape of the likelihood at its minimum where it is assumed the underlying statistical assumptions are satisfied.

Sensitivities to the data and some biological constants (e.g. M, discard mortality, etc) were explored using an ensemble modelling approach. Here the underlying model structure was retained and disturbances around the base model and data simulated. This gives an insight into the uncertainty in the quantities of interest that includes uncertainty well beyond asymptotic CVs. This is a major step forward but because the population dynamics are kept unchanged uncertainty around these dynamics is not fully explored. Some of the uncertainty in the ensemble analysis, for example in M and discard mortality, could be included in a full Bayesian model with these quantities treated as parameters with priors. Posterior distributions estimated from MCMC sampling would provide some insight as to whether the data contained any information on these quantities as well as quantifying uncertainty.

Provide, or comment on, recommendations to improve the assessment

Recommendations from the DW and AW were briefly reviewed at the meeting and supported. I missed any strong recommendation to develop or strengthen survey abundance indices. Such indices are essential for a well-conditioned assessment. The current Atlantic scamp assessment makes effective use of three surveys. The video survey is the only fishery independent survey and further work to refine this index, and a commitment to its future operation are important. The headboat index is also important as it is the longest time series that is well fit by the model. Although it is standardized it is not explicitly corrected for technological creep, an issue that deserves investigation to avoid bias over time.

Selectivity parameters are influential parameters in the model. In this assessment *a priori* assumptions were made about the shape of the selection curve which, while reasonable, need to be supported by experimental evidence. It would be desirable to conduct field experiments to identify the appropriate selectivity for the fleets concerned.

Currently BAM does not support an option to model discards with a retention function and appears to require this catch category to be modelled as a separate fleet. This does not reflect the way the observations arise and the model needs to be enhanced to allow discards to be modelled with a separate retention function for the fleet concerned.

Since this is a research track assessment, it would be useful to explore several simpler assessment approaches to explore the data in order to understand them better before building a complex and highly parameterized SS model. For example, a full age structured model can be used to analyse the survey data alone (e.g. Cotter et al 2007) while almost any surplus production model can be used to investigate biomass indices, with or without catch data (Pedersen and Berg, 2017, Cook et al, 2021, Appendix 4). These models can be applied to explore the consistency of information in the catch an index data and help in the specification of more complex models. It would also be useful to investigate the length frequency data with length-based methods such as LIME (Rudd and Thorson, 2017) to gain insights into the quality of the data.

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

The focus of the review was very much on the fit of the assessment model to the data. This is an important component of evaluating a model. However, uncertainty needs to be considered also in the context of the ability of the model to estimate stock status relative to reference points. In the Gulf assessment, for example, there were moderate changes in biomass scale and F even though the qualitative trends remained similar under different model configurations. It would be useful to know whether these differences also affected the perception of stock status. I therefore recommend that research track assessments include estimation of reference points so that the robustness of stock status evaluation can be made.

I had the impression that the decision to use SS or BAM as the assessment framework had been made before any analysis of the data. These assessment frameworks are, of course, appropriate and state of the art. Nevertheless, as this is a research track, more exploration of the data before deciding on the modelling approach is desirable to encourage innovation and better inform the final analytical approach. Some ways of exploring the data are outlined in the previous section.

Conclusions and recommendations

Both assessments are thorough and high-quality analyses and represent the best science available. The Atlantic assessment is probably the more reliable on the basis of the model diagnostics, both in terms of fit to the data and retrospective pattern. The Gulf assessment is more challenging as two of the surveys are not fit well and there is uncertainty about the estimate of unexploited biomass which affects its scale. This is reflected in a rather poor retrospective pattern. Improvements were made to both assessment models during the review, but it is acknowledged that more improvements are possible before an operational assessment is established.

Recommendations are made in the Findings section as required by the Review Workshop terms of Reference.

NMFS review process

The SEDAR 68 was an effective way to review research track assessments of scamp. The review workshop was well organized with material available in advance of the meeting which facilitated productive discussions at the meeting. The virtual meeting facilities functioned well and the work was conducted in a positive and fruitful manner. The analytical team provided excellent cooperation. The review panel comprised experts with a wide range of relevant expertise that covered the main disciplines involved in the analysis. No major disagreements emerged during the meeting.

References

Cotter, John, Rob Fryer, Benoit Mesnil, Coby Needle, Dankert Skagen, Maria-Teresa Spedicato, Verena Trenkel. (2007). A review of Fishery-Independent assessment models, and initial evaluation based on simulated data. ICES CM 2007/O:04

Cook, Robin and Acheampong, Emmanuel and Aggrey-Fynn, Joseph and Heath, Mike (2021) A fleet based surplus production model that accounts for increases in fishing power with application to two West African pelagic stocks. Fisheries Research, 243. 106048. ISSN 01657836

Pedersen, M. W., and Berg, C. W. 2017. A stochastic surplus production model in continuous time. Fish and Fisheries 18: 226–43. https://doi.org/10.1111/faf.12174.

Rudd, MB and Thorson, JT. 2017. Accounting for variable recruitment and fishing mortality in length-based stock assessments for data-limited fisheries. Canadian Journal of Fisheries and Aquatic Sciences https://doi.org/10.1139/cjfas-2017-0143.

Appendix 1. Materials provided for the review

Document #	Title	Authors	Date Submitted
Documents Prepared for the Stock ID Process			
SEDAR68-SID-01	Brief Summary of FWRI-FDM TagRecapture 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	Dominique Lazarre	1 April 2020

	Scamp Discards from Recreational Fishery Surveys in the Eastern Gulf of Mexico		
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 (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 four and a half hours each day held between the hours of 8 am -8 pm CT
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)

Appendix 3. Panel Membership

Doug Gregory (Chair)
Robin Cook, CIE
John Neilson, CIE
Massimiliano Cardinale, CIE
Luiz Barbieri, GMFMC SSC
Anne Lange, SAFMC SSC
George Sedberry, SAFMC SSC

Appendix 4. An assessment of Gulf and Atlantic scamp using a Schaefer surplus production model

Robin Cook
University of Strathclyde, Glasgow, UK

Introduction

This appendix summarises an exploratory analysis of some of the data used in the Gulf and Atlantic scamp assessments using a simple Schaefer model. These model runs are not intended as alternative assessments but hopefully to shed some light on the qualities of the data and characteristics of the assessment.

Model summary

The Schaefer model equations are set out in Table 1. The biomass projection equation is parameterised in terms of MSY, m, and carrying capacity, K, and includes a lognormal random process error. The catch, Y, is a function of biomass, B, where the fishing mortality, f, follows a random walk. Data included are the abundance surveys (denoted by u) and the landings. The fishery independent survey is assumed to be proportional to biomass with a constant catchability, q. The fishery dependent surveys are modelled in a similar fashion except that q is allowed to increase over time with a mean annual power increment, ∂. The data are assumed to be observed with lognormal errors. The model was fit using the Bayesian statistical R package rstan with 3 chains over 30000 iterations and a thinning rate of 100. The parameters and their priors are listed in Table 2.

Results

Gulf

Model fits to the data are shown in Figure 1 and generally fit both the surveys and the landings data well. Fishing mortality fluctuates with little trend until 2010 when it increases sharply. Biomass shows a long term decline, reducing by around 50% over the time period. Table 3 shows the estimated power increment for the surveys. Only the RFOP survey shows a significant increase of about 5% per year and explains the apparently increasing trend seen for this survey. Figure 2 shows a retrospective plot. The successive peels show large changes in scale as data are removed although the trends are similar. The 95% CI is very large and most runs fall within this range but it illustrates the problem that the data contain very little information on the scale of the biomass and that the estimates are subject to large uncertainty.

Atlantic

Model fits to the data are shown in Figure 3 and generally fit both the surveys and the landings data very well. Fishing mortality fluctuates with little trend until 2000 when it increases sharply and levels off around 2010. Biomass is initially stable but shows a long term decline from 1990, reducing by around 80% over the time period. Table 4 shows the estimated power increment for the surveys. The fishery dependent surveys show a significant increase of about 4-6% per year. Figure 4 shows a retrospective plot. The successive peels show a high degree of consistency in the estimate of terminal F and biomass, falling well withing the 95%CI of the base run.

Discussion

Even with very few parameters the model can fit the survey and catch data well for both stocks. For the Gulf, this comes with the cost of an extra parameter to reconcile changes in catchability for the RFOP survey. The large estimated power increment provides some reason to exclude the survey from the main SS assessment. However, it is perhaps worth noting that the Schaefer model is able to fit all the other surveys well including the commercial vertical line survey. For the latter the SS assessment was unable to fit the data, yet the Schaefer model finds a common signal for all the surveys. This might suggest problems with the age/length comps for this survey or mis-specification of the selectivity function in SS.

In the Atlantic the surveys are fit closely but there is a large estimated power increment associated with the fishery dependent surveys. As the raw indices are not corrected for technological creep there is a danger the indices will overestimate relative biomass in recent years. This is consistent in the Schaefer model (with power correction) estimating an increase in fishing mortality in recent years whereas the BAM model (without power correction) suggests it has been fluctuating without trend.

The retrospective plot for the Gulf highlights the problem that there is little information on the scale of the biomass and this is also evident in the SS assessment. It suggests that a more robust way of expressing F and biomass over time would be to use ratio estimators such as B/Bmsy and F/Fmsy. In the Atlantic assessment there is a small improvement in the retrospective pattern for F compared to the BAM assessment. Here the large 2014 deviation in the BAM retrospective is absent. This may be the result of a somewhat closer fit to the survey than in the BAM assessment.

Both assessments estimate the stocks to be over-fished with over-fishing occurring.

Table 1. Model equations

Model equation	Description
$B_{t+1} = \left(\left(1+\frac{4\text{m}}{K}\right)B_t - \frac{4\text{m}B_t^2}{K} - f_t B_t\right)\exp\left(\epsilon_t\right)$	Schaefer surplus production model expressed in terms of MSY (m) and carrying capacity, K, with random effect ϵ_t
$f_t \sim lognormal\left(\log(f_{t-1}), \sigma_f\right)$	Fishing mortality (yield-biomass ratio) follows a random walk
$\epsilon_t \sim normal(0, \sigma_B)$	Process error for biomass
$Y_t = f_t B_t$	Catch is a function of biomass
$u_{k,t} = q_k B_t (1-\partial_k)^{t-1}$	Abundance index, u, is proportional to biomass with an annual mean power correction ∂. For the fishery independent survey ∂=0.
$\text{u}'_{k,t} \sim lognormal(u_{k,t}, \sigma_k)$	Observed index, u' is subject to lognormal errors
$Y'_t \sim lognormal(Y_t, \sigma_y)$	Observed catch, Y' is subject to lognormal errors

*Table 2. Model parameters and their description. Where applicable, priors used in the base models for anchovy and bonga shad are shown. For K, the limits a and b are defined as a=√(minimum observed catch), b=√(10*maximum catch).*

Parameter	Description	Prior
m	Maximum sustainable yield (MSY)	Uniform(0.001, 2*maximum catch)
K	Carrying capacity or virgin biomass	Uniform(a,b) on square root scale
q_k	Catchability coefficient for index k	Uniform(0.001,100)
B_1	Biomass in first year	Uniform(0,100)
∂_k	Mean annual fishing power increment for fleet k	Uniform(-0.05,0.1)
σ_f	Standard deviation of fishing mortality process error	Uniform(0,1)
σ_B	Standard deviation of biomass process error	Uniform(0,1)
σ_y	Standard deviation of catch observation errors	Uniform(0,1)
σ_k	Standard deviation of observation errors on abundance index, k	Uniform(0,10)
B_t	Biomass in year t	NA
Y_t	Catch (yield) in year t	NA
Y'_t	Observed catch in year t	NA
$u_{k,t}$	Abundance index in year t for index k	NA
$u'_{k,t}$	Observed index in year t for index k	NA

Table 3. Gulf scamp. Estimated power increment for the surveys and the log scale standard deviation for the measurement error.

Survey	Mean annual power increment	Measurement error (SD)
Comb Video	Fixed at 0	0.35
Headboat	0.00 (-0.01,0.01)	0.28
Comm VL	-0.01 (-0.01,0.00)	0.19
RFOP	0.05 (0.03,0.10)	0.25
Landings	NA	0.21

Table 4. Atlantic scamp. Estimated power increment for the surveys and the log scale standard deviation for the measurement error.

Survey	Mean annual power increment	Measurement error (SD)
Comb Video	Fixed at 0	0.37
Commercial	0.06 (0.04,0.08)	0.07
Recreational	0.04 (0.03,0.06)	0.19
Landings	NA	0.06

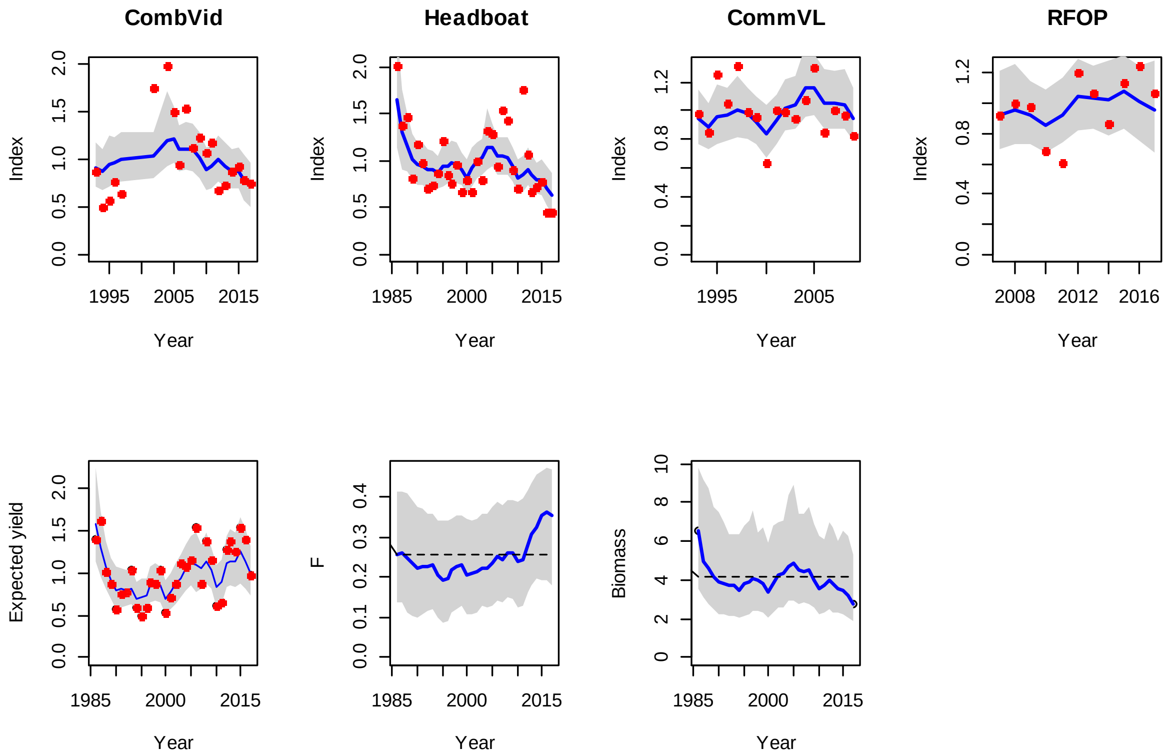


Figure 1. Model fit (blue line) to the survey and landings data (red dots) for the Gulf stock. Shaded area is the 95% CI. The two lower right panels show the estimated fishing mortality and biomass with F_{MSY} and B_{MSY} shown as dashed lines.

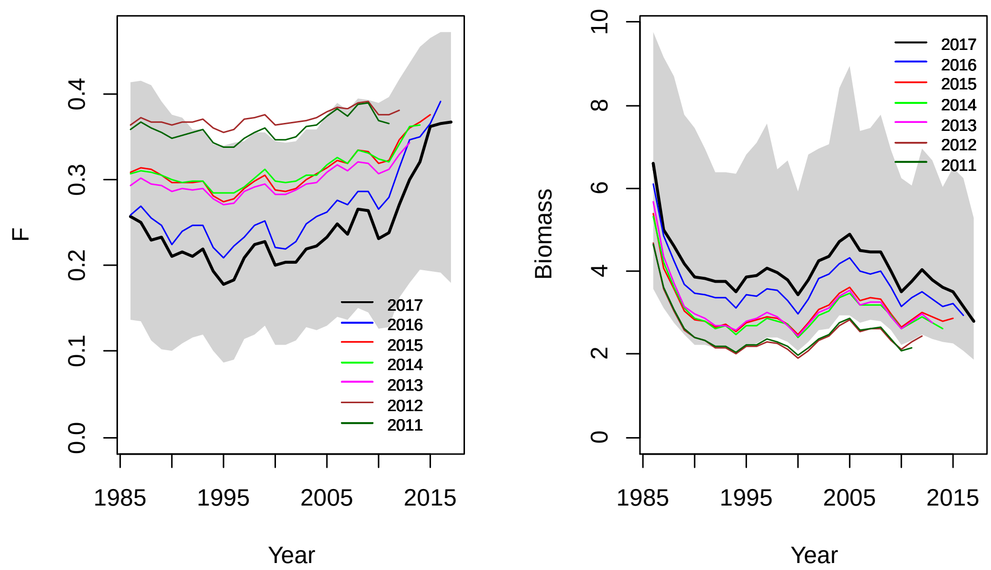


Figure 2. Gulf scamp. Retrospective plots for fishing mortality and biomass. Shaded area is the 95% CI for the full data set.

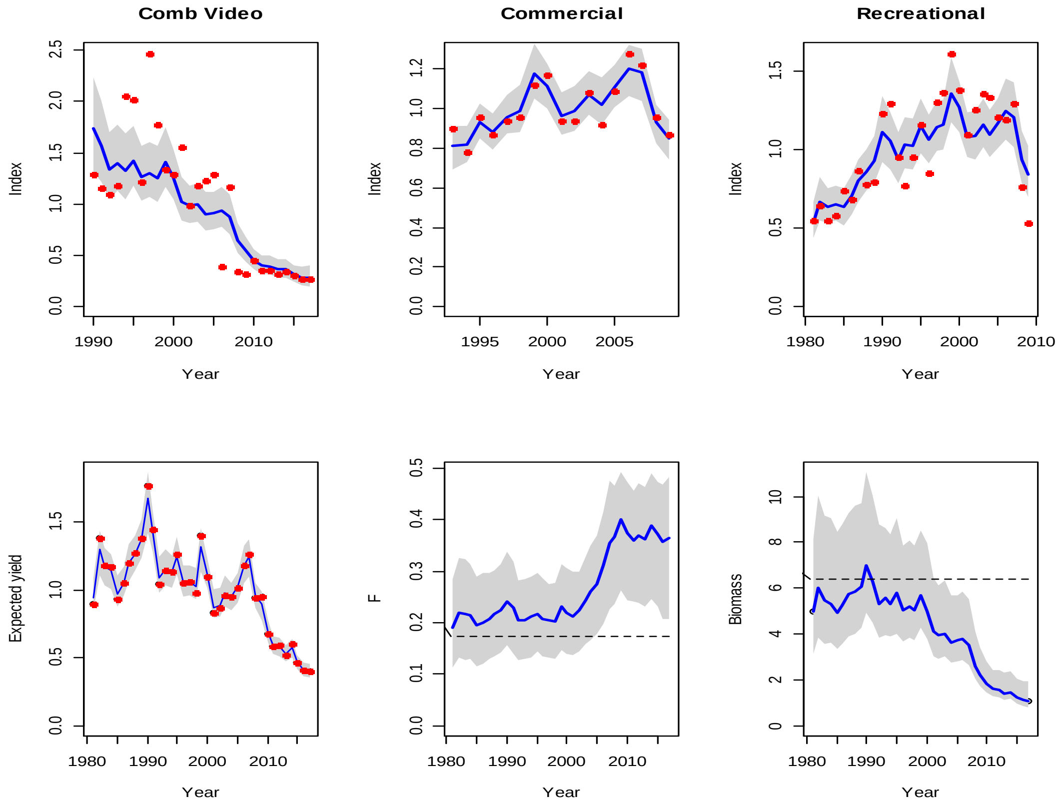


Figure 3. Model fit (blue line) to the survey and landings data (red dots) for the Atlantic stock. Shaded area is the 95% CI. The two lower right panels show the estimated fishing mortality and biomass with F_{MSY} and B_{MSY} shown as dashed lines.

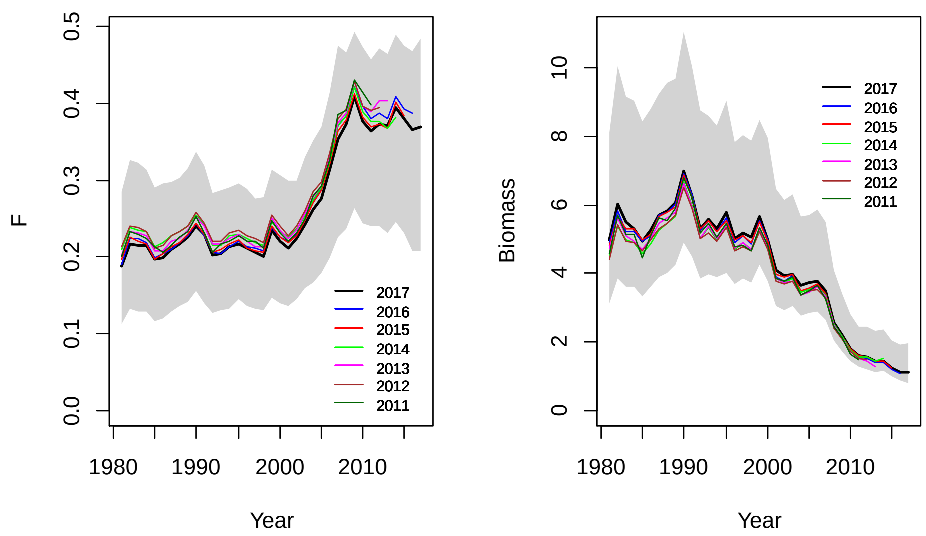


Figure 4. Atlantic scamp. Retrospective plots for fishing mortality and biomass. Shaded area is the 95% CI for the full data set.