Independent Peer Review of the Gulf of Mexico Gray Snapper, Stock Assessment

Laurence T. Kell*

May 2018

SEDAR: Southeast Data, Assessment, and Review
SEDAR 51: Gulf of Mexico Gray Snapper
Center for Independent Experts (CIE) Peer Review Report
Review workshop held March 20-22, 2018 in Tampa, Florida

*Email: laurie@seaplusplus.co.uk

Executive Summary

The assessment of gray snapper in the Gulf of Mexico assessment was conducted using Stock Synthesis, an integrated assessment model. $SPR_{30\%}$[1] was selected as the maximum sustainable yield (MSY) proxy used to determine whether the stock was over fishedand the minimum stock size threshold (MSST) is then defined as 50% of $SSB_{SPR_{30}}$. The maximum fishing mortality threshold (MFMT) to define overfishing was based on the corresponding fishing mortality value, i.e. $F_{30\%_{SPR}}$. On this basis the stock is currently undergoing overfishing ($F_{current}/F_{SPR_{30}} = 1.20$) and has been undergoing overfishing for most years since 1976, which has reduced stock biomass to 70.3% of $SSB_{SPR_{30}}$. Since the MSST is set at 50% of $SPR_{30\%}$, however, the stock remains above the MSST and its status is therefore determined as not overfished. There are no explicit target reference points or Harvest Control Rules in place.

Gray snapper in the Gulf of Mexico supports important commercial and recreational fisheries, which are managed differently in the Gulf compared to the Atlantic and in state versus federal waters. Management in the form of size limits and bag sizes have important impacts on the quality of the data. Problems were seen, however, with fits to the discards and to the Age length Key. Predicted discards were very different from actual observations, which may indicate a problem either with the data or model misspecification. The Data Workshop had noted various issues related to stock ID, such as low dispersion rates and regional variations in population parameters, which may explain the conflicts seen when an Age Length Key was used in the assessment. The uncertainies in the assessment are not unique to the Gulf of Mexico gray snapper and it is recognised that the Assessment Team had done an excellent job given the workload required to run the integrated assessment used to assess gray snapper.

Although integrated assessments are able to include a variety of datasets in order to try and simultaneously provide information on all estimated parameters conflicts between datasets are not uncommon and may indicate model misspecification resulting in biased estimates of parameters and quantities derived from them. Conducting a stock assessment is an iterative process that allows hypotheses to be tested, the impact of uncertainty on management advice to be identified and where a reduction of uncertainty can reduce risk, where risk is an uncertainty that impact the ability to achieve management objectives. Although the ALK was not used in the base case assessment due to problems with the fits this may indicate problems with the data or the model. It is important to determine which, especially since an ALK will be very informative about year-class strength, growth, mortality rate and productivity. It is important therefore to examine a number of goodness of fit diagnostics when interpreting results and to develop appropriate methods to validate integrated stock assessments.

If the problems seen are with the data collection then appropriate sampling schemes need to be designed, while if the problem is with the model then methods for model validation need to be developed While improved data is always welcome collection of data is expensive. A cost benefit exercise could be conducted using simulation based on the current assessment model, i.e. to show the risk of the stock being below the MSST or exploitation being above the MFMT and how this could be reduced through better discards data or an ALK. Or in other words is it better to invest in sampling programs for collecting information on discards and release morality rates or to develop ALKs over a broader range of depths and fisheries.

[1]calculated as the point on the spawner per recruit curve where SPR is 30% of the value when F=0

Contents

List of Figures

1. Background

This document contains an independent report of review activities and findings for the 51_{st} Southeast Data, Assessment and Review (SEDAR 51) Review Workshop, held between the 20^{th} and 22^{nd} March 2018 in Tampa, Florida where the assessment of grey snapper, including the findings of the data and assessment workshops were reviewed. This document contains a summary of my inputs to the discussions as well as some personal opinions about the assessment. It does not necessarily repeat the conclusions set forward in the consensus report, with which I agree.

Prior to the meeting a Statement of Work (section 6.2) was provided to panel members (section 6.3), this included the Terms of Reference (TOR) for the review panel (RP). Assessment documents and background material (section 6.1) were also provided via a website[2] and by email prior to the meeting. In addition I asked for the assessment inputs and outputs so that it was possible to review the assessment itself and to be able to run additional analyses as required. Some of these analyses are provided as appendices in section 7.4.

During the meeting there was a general consensus among the RP regarding most of the main discussion points and findings which are outlined in the Review Workshop Report. Since I have had wide experienced of a wide range of stock assessment methods and I have a particular interest in risk based management I focused on model diagnostics, the treatment of uncertainty and how it is propagated into advice.

Although the Data Workshop Report, Working Papers and background material were provided to RP in accordance with the Statement of Work, the Assessment Workshop Report was provided late. The stock assessment inputs and outputs were provided by the AT on request, however, and this helped me in reviewing the assessment especially the sensitivity analyses and evaluation of uncertainty. It had not been possible to perform projections before or during the meeting, these were only finished after the meeting. These problems are not a reflection on the AT more a consequence of the workload and resources required to complete a full assessment using integrated models.

2. Individual Reviewer Activities

Prior to the meeting I reviewed the background documents provided to the workshop. I then participated in the Review Workshop in Tampa, which benefited from the participation of a number of stakeholders who were able to provide both background and personal experience.

The analytical team (AT) from the Assessment Workshop (AW) presented the assessment results. The structure was fairly informal with discussion during each presentation, and a lot of interaction which I actively participated in with the other members of the RP. I was able to make a contribution by helping to run model diagnostics in following these discussions. This is an area which I will expanded upon in the next sections.

After the Review Workshop, I assisted in finalising the Review Workshop Report and in preparing my independent report, in which I summarise my activities during the RP meeting, including providing a detailed summary of findings, conclusions, and recommendations against each term of reference.

3. Summary of Findings, Conclusions and Recommendations

The folowing sections closely follow the Terms of Reference and in them I expand upon the panel consensus report, particularly in relation to goodness of fit diagnostics, model validation, and the treatment of uncertainty.

[2] `http://sedarweb.org/sedar-51`

3.1 Data

The method chosen for the assessment Stock Synthesis 3 (Methot, 2005, SS3) is able to integrate various sources of information, and variety of data were available, including commercial and recreational landings, length compositions, indices of abudance, discards and estimates of discard mortality from the main fisheries. Data on reproduction (Fitzhugh et al., 2017) and stock ID (S51 Stock ID Working Group, 2017) were also available

Indices covered a variety of areas and habitats and included both fisheries independent and dependent series (Pollack et al., 2017; Gardner et al., 2017; Campbell et al., 2017; Thompson et al., 2017b; Flaherty-Walia and Tyler-Jedlund, 2017; Muller, 2017; Thompson et al., 2017a). Fisheries dependent catch per unit effort (CPUE) had been standardised appropriately (Smith, 2017).

While accurate age and growth information is crucial to stock assessment (Thornton et al., 2017) because of various problems the age length keys (ALKs) although available were not used in the assessment.

Due to management measures (i.e size and bag limits) discards are a driving force in this fishery (Flaherty-Walia et al., 2017) making it important to use assessment models that incorporate length composition information (Chih, 2017; Hanisko and Pollack, 2017). A major problem in the catch and landings data and hence in the assessment itself is the incomplete record of discards.

In a stock assessment it is commonly assumed there is no immigration or emigration and that a stock is homogeneous. The latter assumption may be violated, however, in this case as tagging studies have shown relatively low dispersion (Germeroth et al., 2016) and there is evidence of spatial variation in the population structure of gray snapper and regional differences in mean size and mortality-at-age are seen (Allman and Goetz, 2017).

A fishery is normally defined as a relatively homogeneous fishing unit, where selectivity and catchability characteristics do not vary greatly over time, although some allowance can be made for time-series variation in catchability. The fisheries used in the assessment are relatively localised, which may help in ensuring fishing units ae the homogeneous but low dispersion and spatial differences in population structure mean it will be important to look at conflicts in the data due to stock definition issues when conducting the assessment.

With respect to the Terms of Reference; are

1. data decisions made by the DW and AW sound and robust?
2. data uncertainties acknowledged, reported, and within normal or expected levels?
3. data applied properly within the assessment model?
4. input data series reliable and sufficient to support the assessment approach and findings?

The main problems with the data were due to management measures (i.e. size and bag limits that results in discarding), regional differences in population dynamics and the definition of stock boundaries. These issues had been discussed in the DW and taken into consideration by the AW when setting up the stock assessment.

A Commercial Workgroup had considered the data and concurred that the majority of landings data was adequate for assessment analyses, however, they did note that data appeared to be most accurate and reliable in recent years due to the implementation of state trip ticket programs. Historic landings prior to 1962 were found to be the least reliable as there was missing data for some years and states. The Commercial Workgroup therefore developed CVs, with to smaller CVs over time, based on expert opinion that reflect improvements in data collection methodologies.

The DW considered that discards were insignificant for all fleets and that length samples were adequate for the assessment as there were a relatively high number of samples for most years and strata. Although estimates of discards and post discard mortality are available the DW did note that discard mortality of gray snapper has not been extensively studied.

The main uncertainies and problems with the data are not unique to the Gulf of Mexico gray snapper and can be seen in other stock assessments. conducting a stock assessment is an iterative process that allows hypotheses to be tested, the impact of uncertainty on management advice to be identified and show where a reduction in uncertainty can reduce risk, where risk is an uncertainty that impact the ability to achieve management objectives.

In conclusion the strengths and weaknesses of the various datasets are set out sufficiently to allow the assessment to be conducted and for the stock assessment analysts to be aware of them when examaning goodness of fit diagnostics and interpreting results.

3.2 Methods

SS3 is able to integrate several sources of information, it can be applied even where data are limited and is extensively used in the US and globally. It is a statistical method that uses the available data in a raw a form as appropriate in a single analysis. This allows for consistency in assumptions and permits the uncertainty associated with both data sources to be propagated to nal model outputs (Maunder and Punt, 2013).

When running an integrated Assessment it is important to correctly specify:

1. the observation model process, e.g. the form of selectivity and discarding;
2. systems dynamics, e.g. the values for steepness of the stock-recruitment relationship and natural mortality;
3. appropriate data weighting; and
4. model structure.

Currently, however, there are few standard diagnostic tools available for integrated stock assessment models that can provide an analyst with all the information needed to determine if there is substantial model misspecification.

Potential diagnostic tests to identify model misspecification include, i) residuals analysis and run tests; ii) retrospective analysis; iii) likelihood component profiles; and iv) Age structured production models. Carvalho et al. (2017) in a simulation exercise showed that residual analyses were easily the best detector of misspecification of the observation model while the ASPM test was the only good diagnostic for detecting misspecification of system dynamics model. While retrospective analysis and the likelihood component profile infrequently detected misspecified models. Applying multiple carefully selected diagnostics can increase the power to detect misspecification without substantially increasing the probability of falsely concluding there is misspecification when the model is correctly specified.

With respect to the Terms of Reference; are

1. methods scientifically sound and robust?
2. assessment models configured properly and consistent with standard practices?
3. the methods appropriate for the available data?

SS3 is widely used, both in the US and internationally, and the analysts were well versed in its application. Small tweaks to the input data or assumptions, however, can often result in substantial differences to advice and it is important to use appropriate diagnostics methods to detect potential data conflicts and model misspecification, since both may bias the results and underestimate the true uncertainty. Residual plots, retrospective analyses and likelihood profiles were presented in both the assessment report and at the RP.

I was able to run the assessment base case before the review meeting and to reproduce the results, I was also able to explore a range of diagnostics and to conduct a variety of sensitivity analyses. This was because the assessment model had been configured correctly and documented by the analysts.

In the case of gray snapper it is important to include length composition due to size and bag limits that mean discarding can be substantial in certain fisheries. Including length data in the assessment allows exploration of conflicts in the data due to spatial variations in the stock and temporal changes due to management, climate and changes in exploitation across the fisheries to be performed.

Although models like SS3 are able to use many sources of available data in an integrated framework it also means that stock assessment models have grown increasingly complex and their use and development is limited to a few experts Hilborn (2003). This can mean that too often discussions in a stock assessment focus on detailed options in a control file and not on how well the model fits the data. Methods like SS3 are also more computationally intensive than traditional methods because there are more data and parameters to consider. This means that development time is increased, especially as a model has usually to be run many times to check for implementation errors and to carry out sensitivity tests to evaluate the performance of an assessment and to determine how robust it is to its main assumptions.

Simpler models, could have been used to explore important issues with and to increase dialogue with data providers and other stakeholders. SS3 is basically an aged based model, length based models could also have been explored, particularly given problems with the ALK that may indicate model misspecifucation. However, given the limited time to conduct the integrated assessment this may be difficult to achieve.

Although many different datasets can be fitted to an integrated stock assessment model in some such stock assessments the index of abundance provides almost no information on population scale. Consequently, the estimates of the model outputs rely almost completely on the size- and age-composition data and model structure (Maunder and Piner, 2014). By configuring SS3 as an Age Structured Production Model (ASPM) a diagnostic tool could have been developed been developed to be used for comparison with a model estimating all of the model parameters and fitting to all the data as in the base case. In the SS3 assessment as run it is inferred that a production function is apparent in the data and that the catch data can explain the indices, i.e. that the indices are reasonable proxies of stock trend. However, there was little information on steepness and so the dynamics could have been driven by recruitment. Therefore if the ASPM cannot mimic the index, then either the stock is recruitment-driven, catch levels have not been high enough to have a detectable impact on the population, the model is incorrect, or the index of relative abundance is uncertain or not proportional to abundance. This diagnostic has only begun to be implemented, and its utility remains unknown, so it is not reasonable to have expected it to have been performed in this assessment. It does, however, illustrate the importance of running simpler methods.

3.3 Results

Although SS3 is able to include a variety of datasets in order to try and simultaneously provide information on all estimated parameters conflicts between datasets are common and may indicate model misspecification resulting in biased estimates of parameters and quantities derived from them. It is important therefore to examine a number of goodness of fit diagnostics when interpreting results, see Cass-Calay et al. (2014) for a check list i.e.

1. Does the model run?
2. Does the hessian converge?
3. Are there any parameters on bounds?
4. Plot model output. Anything obviously wrong? e.g. productivity way too low, selectivity patterns
5. Examine parameter estimates, plot parameter distributions with starting values and bounds
6. Look at trace plots of parameter estimates relative to phase of estimation? Do parameters change
7. Inspect mean and standard deviation of estimated parameters. Is CV less than 1?
8. Are any of the parameters highly correlated?
9. Plot model fits to data and diagnostics. Is model fitting data reasonably?
10. Check for model stability using Jitter analysis, does model converge
11. Profile leading model parameters such as stock-recruitment parameters (steepness/R0) or natural
12. Evaluate model sensitivity to key model assumptions, data weightings, and alternative data inputs.
13. Evaluate model sensitivity to the most recent years of data using a retrospective analysis. Did the
14. Run bootstrap amd plot distribution of parameter estimates
15. Evaluate model convergence using MCMC

Many of these diagnostics had been run. Diagnostics for integrated models is an active research area, however, and the efficacy of diagnostic tests depended on whether misspecification is in the observation or systems dynamics model (Maunder and Piner, 2017).

The accuracy and precision of any model depends on the quality of the model and the information in the data. The ALK was not used in the assessments presented to the RP due to problems with fittimg and the fits to discard were also problematic, possibly due to stock ID issues and incomplete record of discards and subsequent mortality after release. In the latter case this was despite the opinion of the DW that the discard data were adequate, SS3 consistently underestimated the discard fractions in most fleets. There was also an inherent conflict between discard rates and size composition data and to fit the reported discards requires fish above the size limit to be discarded, which was considered unlikely by stakeholders.

Issues related to stock ID (e.g. low dispersion rates and regional variations in population parameters) and the lack of size composition for discards may mean that there is model misspecification, which may explain the conflicts when the ALK was used in the assessment. Therefore when considering the results it is important to validate models run and to have used a range of goodness of for diagnostics.

The scenarios ran showed limited sensitivity of the estimates of biomass and fishing mortality to alternative data weighting, and so it was argued that the assessment was robust despite problems in fitting the discard data and when the ALK was used. A robust model, however, is one that performs correctly in the presence of invalid inputs, the fact that the assessment results did not change much in the sensitivity runs is not necessarily a test of robustness just consistency. Especially since stock biomass and fishing mortality can not actually be observed so it is not possible to validate models by comparing estimates of SSB and F, i.e. the unknowns that the model is trying to estimate but which can never actually be observed.

With respect to

1. Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?

2. Is the stock overfished? What information helps you reach this conclusion?

3. Is the stock undergoing overfishing? What information helps you reach this conclusion?

4. Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?

5. Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?

Despite the points raised above trends in abundance and exploitation were consistent across the sensitivity analyses and alternative scenarios presented and discussed during the RW. Although absolute abundance varied slightly amongst the scenarios, conclusions on stock status across scenarios were similar to those from the base model.

In the assessment the steepness of the stock recruitment relationship was estimated to be close to 1, i.e. there was no compensation at small stock size. Benchmarks and reference point were therefore based on a per recruit analysis. In the initial period for which there are no indices of abundance or length composition data no variation was estimated in recruitment, and in the 1980s extremely large and small recruitments were seen. Neither are likey to be realistic. Therefore in the estimation of reference points for the evaluation of productivity and for performing projections average recruitment from the period 1990-2015 was assumed.

$SPR_{30\%}$ was selected as the maximum sustainable yield (MSY) proxy used to determine whether the stock was over fished. The minimum stock size threshold (MSST) is then defined as 50% of $SSB_{SPR_{30}}$. Based on this definitions, the stock is currently not over fished($SSB_{2015}/SSB_{SPR_{30}} = 0.703$).

To define overfishing the maximum fishing mortality threshold (MFMT) is used, based on $F_{30\%_{SPR}}$. Using this reference point the stock is currently undergoing overfishing ($F_{current}/F_{SPR_{30}} = 1.20$) and has been undergoing overfishing for most years since 1976. This history of overfishing (fishing above $F_{30\%_{SPR}}$) has reduced stock biomass to 70.3% of $SSB_{SPR_{30}}$. Since the MSST is set at 50% of $SPR_{30\%}$ the stock remains above the MSST, however, and its status therefore determined as not overfished.

Per recruit reference points are also determined by the selection patterns, which will vary across fisheries, and may mean that there is implicit competition between them. It would be informative to see a summary of the selection patterns across fleets and how these impact on the theoretical optimal yield summarised.

The apparent contradiction that the stock has been subject to overfishing but is not over fishedcan be explained by the fact that MSST and MFST are limit reference points and do not need to be consistent with each other. This is because a main objective of reference points is to prevent overfishing, i.e. growth, recruitment, economic and target overfishing. Growth and recruitment overfishing are generally associated with limit reference points, while economic overfishing may be expressed in terms of either targets or limits. The difference between targets and limits is that indicators may fluctuate around targets but in general limits should not be crossed. Target overfishing occurs when a target is overshot, although variations around a target is not necessarily considered serious unless a consistent bias becomes apparent. In contrast even a single violation of a limit reference point may indicate the need for immediate action.

No explicit management targets are defined for this stocks, e.g. to achieve MSY, which if had been the case would potentially require a Harvest Control Rule (HCR) to determine management action, when the stock was overfished

A variety of emprircal reference points could also be used, a benefit of which is that they are more intuitive to understand in discussions between stakeholders. Example include length-composition (Kokkalis et al., 2015; Prince et al., 2015), age-composition (Thorson and Cope, 2015), fishery catch and fishing effort data (Roa-Ureta et al., 2015), abundance indices (Needle, 2015) or simple length-based reference points (Cope and Punt, 2009). These may be easier to understand by stakeholders but before being able to make management recommendations, a link between a trigger reference point and stock status has to be identified. For example using a HCR to link removals to the current state of the resource (Restrepo and Powers, 1999).

I would also have liked to have seen results from Monte Carlo Markov Chain (MCMC) and/or bootstap simulations in order to evaluate bias in the results.

3.4 Projections

A term of reference for the Assessment Workshop was to project future stock conditions (biomass, abundance, and exploitation; including probability density functions) and develop rebuilding schedules if warranted.

With respect to the Terms of Reference; are

1. methods consistent with accepted practices and available data?
2. methods appropriate for the assessment model and outputs?
3. results informative and robust, and useful to support inferences of probable future conditions?
4. key uncertainties acknowledged, discussed, and reflected in the projection results?

The provision of fisheries management advice requires first fitting a model to data to assess current stock status, then predicting the response of the stock to management, and then checking that predictions are consistent with reality. The accuracy and precision of the prediction depends on the quality the model, the information in the data and the prediction horizon, i.e. how far ahead we wish to predict.

The use of an integrated model such as SS3 allows for consistency in assumptions and permits the uncertainty associated with both data sources to be propagated to final model outputs, such as catch limits under harvest control rules.

3.5 Uncertainties

The main way that uncertainty was explored was through running a range of scenarios, by applying CVs to the time series of catches to reflect changes in the quality of data collection, and apply different weights to the various datasets. The main sources of uncertainty evaluated were i) discard length composition and mortality by exploring a range of discard mortality values in the sensitivity runs, and applying a large CV to discard rate; ii) Natural mortality; iii) in the growth model various CVs were used on growth curve parameters; and iv) index CV scalings. Data conflicts were explored with runs containing different weights for fishery independent and dependent indices.

A problem with computer intensive and complex models such as SS3, however, is that developing and then running an assessment can be a long process. This means that only a limited number of scenarios can be run and little time remains for validation of the assessment and considering goodness of fit diagnostics.

With respect to the Terms of Reference; are

- 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
- Ensure that the implications of uncertainty in technical conclusions are clearly stated

Important forms of uncertainty not explored were spatial variations in life history parameters and issues related to the age-length key. For example the issues in the age length key are possibly of fundamental importance, i.e. due to variations in regional growth resulting in model misspecification and needs further investigation.

Inclusion of parameter uncertainty can be carried through from the assessment into the forward projections using a variety of methods e.g. using the covariance matrix the bootstrap or MCMC. In the projections the delta method and the covariance matrix were used due to time limitations Magnusson et al. (2012), however, showed that this was the least reliable method. The bootstrap and MCMC are also important tools for detecting bias and should be run before the next assessment.

Presumably the data in the ALK were used to estimate growth, maturity and natural mortality via the Lorenzen natural mortality curve which is a function of mass-at-age (Lorenzen and Enberg, 2002). If there is a prblem in these data simply removing the ALK from the assessment may mean that while the assessment is consistent across scenarios it may not be robust.

Uncertainty about difficult to estimate parameters such as the form of the stock recruitment relationship, natural mortality and the form of discarding could have been included by runnuing many scenarios and then including them in the projections. This would have allowed the benefits of reducing uncertainty to be quantified. Using a single base case means that any improvement in data or knowlegde is likely to undermine faith in the stock assessment process, i.e. if results change as new data are collected.

3.6 Recommendations

Consider the research recommendations provided by the Data and Assessment workshops and make any additional recommendations or prioritizations warranted.

- Clearly denote research and monitoring that could improve the reliability of, and information provided by, future assessments
- Provide recommendations on possible ways to improve the SEDAR process

I am in broad agreement with the conclusions in the review panel report. I do believe, however, it would help to have an objective way to compare and validate models. Where validation examines if a model family should be modified or extended, and is complementary to model selection and hypothesis testing. Model selection searches for the most suitable model within a family, whilst hypothesis testing examines if the model structure can be reduced. The inability of the model to predict observations, i.e. of discards, is a cause for concern since if data are regarded as being representative of the dynamics of the stock then they can be used as a validation measure (Hjorth, 1993). The the best performing scenarios, i.e. choice of models and data, can be identified by comparing predictions with observations. This is something that needs more exploration in the future, to identify the limitations of the current assessment and what information is required to reduce uncertainty in order to avoid the risk of failing to meet management objectives. Such an approach would also allow a cost benefit analysis to be conducted.

A stock assessment is never finished, however, and is an iterative process, where hypotheses are tested, models validated and ways in which risk can be reduced by better information identified. I would therefore like to see the assessment framework used to look at ways of reducing risk, i.e. which are the key uncertainies to reduce and how this ca be done in a cost effective way before the next assessment.

3.7 Stock Assessment

The issue is whether the stock assessment constitutes the best scientific information available based on relevance, inclusiveness, objectivity transparency, timeliness, verification, validation, and peer review of fishery management information have been addressed above.

In summary the assessment was the best that could be done, given the constraints under which it was conducted. The data workshop provided a good summary the data used in the assessment and potential problems, while the Assessment Team had documented all the decisions made, were responsive to requests by the Review Panel and the assessment could be rerun by others

3.8 Improvements

The assessment model, SS3, is an integrated statistical method that is able to combine raw datasets into a single analysis, allowing for the use of limited or incomplete datasets, consistency in assumptions and permits uncertainty to be propagated to nal model outputs such stock status and potential management measures. Despite these advantages integrated models are not a panacea for poor quality data or model structure uncertainty in stock assessments (see, Maunder and Punt, 2013), as are several disadvantages related to model misspecication, the complexity of the resulting models, and the associated, often considerable, computational requirements.

Data analysis is an integral part of producing good, reliable stock assessments and there is value in analysing each component dataset in colaboration with stakeholders before conducting an integrated model of all data simultaneously. Especially as in this case where poor fits to the discards **Figure 1** and the ALK could indicate either model misspecification or problems with the way the data were collected. It is important to identify as the solutions are quite different, i.e. allocate resources to redesign new data sampling schemes or develop new modelling approaches.

While improved data is always welcome collection of data is expensive. A cost benefit exercise could be conducted using simulation based on the current assessment model, i.e. to show the risk of the stock being below the MSST or exploitation being above the MFMT and how this could be reduced through better discards or age data. Or in other words is it better to invest in sampling programs for collecting information on discards and release morality rates or to develop ALKs over a broader range of depths and fisheries.

Stock assessment is an iterative process, where the assessment model can be used to test hypotheses about the resource dynamics and identify where uncertainty can be reduced by better analysis or Monitoring, Control and Surveillance (MSC). An important aspect is that management can also affect data quality, as in this case where size and bag limits result in discards, which may mean it is harder to estimate stock processes

3.81 Validation

To develop a robust stock assessment requires a rigorous process of model validation to examine if a model family should be modified or extended, and is complementary to model selection and hypothesis testing. Model selection searches for the most suitable model within a family, whilst hypothesis testing examines if the model structure can be reduced. The inability of the model to predict observations, i.e. of discards, is a cause for concern since if data are regarded as being representative of the dynamics of the stock then they can be used as a validation measure (Hjorth, 1993). Especially as it is not possible to validate a model by comparing unobservable quantities such as stock biomass and fishing mortality.

3.82 Potential Analyses

Here I present some potential analyses using the base case that may help in developing future assessment scenarios. These were generated in R using Rmarkdown and can be readily reproduced for different stock assessments and sensitivity runs. In section 7.4 some future exploratory analyses are present, the code is available on request.

I would recommend the use of cluster analysis to look at the correlation and potential conflicts between datasets to help come up with hypotheses that could be tested or form the basis of different data weighting schemes. **Figure 2** show a cluster analysis of the CPUE series.

Given the difficulties with the conflicts in assessment datasets and the problem with correctly specifing the assessment model it is important that in the future the use of a range of diagnostics are explored, so that the impacts of model assumptions can be better understood

I would also suggest the use of runs tests (**Figure 3**) to compare residuals to fits across datasets. If the process of interest shows only random variation, the data points will be randomly distributed around the median. Non-random variation may present itself in several ways. If the process centre is shifting due to improvement or degradation unusually long runs of consecutive data points may be seen on the same side of the median or the graph crosses the median unusually few times. The red dashed line indicates when randomness is violated and the red points where the variation of fits to individual observations is greater than expected.

Diagnostics for integrated models is an active research area, however, and the efficacy of diagnostic tests depended on whether misspecification is in the observation or systems dynamics model (Maunder and Piner, 2017). Currently there are few standard diagnostic tools available for integrated stock assessment models that can provide an analyst with all the information needed to determine if there is substantial model misspecification.

Potential diagnostic tests to identify model misspecification include

- Residuals analysis, SDNR and run tests;
- Retrospective analysis
- Likelihood component profiles
- Age-structured production model (ASPM)
- Catch-curve analysis (CCA)
- Prediction residuals
- Hindcasting and
- One step ahead projections

As proposed by Cass-Calay et al. (2014) it is important to also plot model output to detect if anything is obviously wrong. In the absence of a stock recruitment relationship the patterns of incoming recruitments and subsequent year-class strengths are important for the estimation of references points and hence management advice. A particular concern is the patterns of recruitments in the 1980s which ranged from 0 to the biggest ever seen **Figures 4**.

There is no variability in the recruitments in the early period since the indices of abundance and length compositions start in 1980. To make comparisons easier the stock estimates are replotted from 1980 in **Figure 5**.

The Beverton and Holt stock recruitment relationship as estimated by SS in shown in **Figure 6**. Steepness is near 1 and so the dynamics are driven by incoming recruitments. This can be seen in **Figure 8** where the estimated numbers-at-age have been scaled within an age so that cohorts can be followed (red are values below the mean-at-age and black values greater than the mean). It appears that in the 1980s extreme recruitment events produced very strong year-classes followed by extremely weak year classes. It is difficult, however, to think of biological process that would produce such recruitment events.

The age-length-key from SS3 **Figure 7** shows that after about age 4 it is difficult to assign a given length to an age, i.e. there is little year-class signal in the older catches (i.e larger fish). **Figure 8**, however, shows very clear year-classes, if this is not coming from information in the larger lengths (i.e. older age classes) due to the absence of an ALK then it must be coming from a index of recruitment, either from an index or length compositions of mainly smaller size classes. To explore this in **Figure 9** the recruitments (black) and the mean size of the length compositions shows the mean size from the Recreational Private Fishing Fleet length compositions (estimates and shown as the red line and observed values as the red points).

There appears to be a strong correlation with recruitment (at age 0) and the mean size in the length compositions. This is at first surprising since it would be expected that a strong recruitment would mean that catches would be dominated by small fish and so mean size would decline. This initially anomalous seeming finding is explained by recruitment being at age 0 and length compositions are dominated from ages 1 and older. So there is a lag between recruitment and mean size.

The observedlength compositions (obs) and the expected values (Exp) as fitted by SS are shown in **Figure 10**. In the expected values, i.e. the fits by SS, after a strong year-class (i.e. recruitment) enters the fishery the mean size increases in susequent years, e.g. from 1988 through to 1990. It appears therefore that the estimates of recruitment and hence cohort strength is therefore coming from the length compostion data. **Figure 11** shows the cumulative distributions and that the fits in 1988 and 1989 were particularly poor.

The DW recognised there were periods of low sampling effort from all data sources prior to 2007 and so the assessment may be modelling noise rather than actual signal, i.e. is over fitting. This could be evaluated by estimating the prediction residuals using cross validation.

Cross validation evaluates the predictive error of a model by testing it on a set of data not used in fitting. There is often insufficient data, however, in stock assessment datasets to allow some of it to be kept back for testing. A more sophisticated way to create test datasets is, like the jackknife, to leave out one (or more) observation at a time. Cross validation then allows prediction residuals to be calculated, i.e. the difference between fitted and predicted values where the later is calculated from the out-of-sample predictions. A comparison of the variance of the model and prediction residuals can help identify over fitting.

4. Report

A Peer Review report was produced with the other panel membders summarizing the Panels evaluation of the stock assessment and addressing each Term of Reference.

5. Acknowledgments

I would like the acknowledge the friendly and constructive manner that the Review was conducted under. I would particularly like to mention the Analytic Representations Jeff Isely and Shannon Cass-Calay and for their forbearance and Julie Neer for her patience in dealing with our many requests.

References

R. Allman and L. Goetz. Regional variation in the population structure of gray snapper, lutjanus griseus, along the west florida shelf. *SEDAR Grey Snapper Reference Documents*, 2017. URL http://sedarweb.org/docs/wsupp/S51_RD_16_Allman%20and%20Goetz%202009.pdf.

M. Campbell, K. Rademacher, M. Hendon, P. Felts, B. Noble, R. Caillouet, J. Salisbury, and J. Moser. Seamap reef fish video survey: Relative indices of abundance of grey snapper. *SEDAR Grey Snapper Data Workshop*, 2017. URL http://sedarweb.org/docs/wpapers/S51_DW_07_SEAMAP%20reeffish%20gray%20snapper%20index%20of%20abundance.pdf.

F. Carvalho, A. E. Punt, Y.-J. Chang, M. N. Maunder, and K. R. Piner. Can diagnostic tests help identify model misspecification in integrated stock assessments? *Fisheries Research*, 192:28–40, 2017.

S. L. Cass-Calay, J. C. Tetzlaff, N. J. Cummings, and J. J. Isely. Model diagnostics for stock synthesis 3: examples from the 2012 assessment of cobia in the us gulf of mexico. *Collect. Vol. Sci. Pap. ICCAT*, 70(5):2069–2081, 2014.

C.-P. Chih. Length frequency distributions for gray snapper length and age samples collected from the gulf of mexico. *SEDAR Grey Snapper Data Workshop*, 2017. URL http://sedarweb.org/docs/wpapers/S51_DW_04_Gray%20snapper%20LFD%20report.pdf.

J. M. Cope and A. E. Punt. Length-based reference points for data-limited situations: applications and restrictions. *Marine and Coastal Fisheries: Dynamics, Management, and Ecosystem Science*, 1(1): 169–186, 2009.

G. Fitzhugh, Beech, H. V.C. Lyon, and P. Colson. Reproductive parameters for the gulf of mexico gray snapper, lutjanus griseus, 1991-2015. *SEDAR Grey Snapper Data Workshop*, 2017. URL http://sedarweb.org/docs/wpapers/S51_DW_06_Reproductive%20Parameters%20for%20Gulf%20of%20Mexico%20Gray%20Snapper.pdf.

S. T. Flaherty-Walia, Kerry E. and A. Tyler-Jedlund. Gray snapper abundance indices from inshore surveys of northeastern gulf of mexico estuaries. *SEDAR Grey Snapper Data Workshop*, 2017. URL http://sedarweb.org/docs/wpapers/S51_DW_11_FWRI_inshoreindex_Lgris_2017_v2.pdf.

W. B. Flaherty-Walia, Kerry E., A. Tyler-Jedlund, and . J. P. Davis. Short-term discard mortality estimates for gray snapper in a west-central florida estuary and adjacent nearshore gulf of mexico waters. *SEDAR Grey Snapper Reference Documents*, 2017. URL http://sedarweb.org/docs/wsupp/S51_RD_01_flaherty-waliaetal2016_graysnappermort.pdf.

C. Gardner, D. DeVries, K. Overly, and A. Pollack. Gray snapper lutjanus griseus findings from the nmfs panama city laboratory camera fishery-independent survey 2005-2015. *SEDAR Grey Snapper Data Workshop*, 2017. URL http://sedarweb.org/docs/wpapers/S51_DW_05_PC_reef_fish_survey_findings.pdf.

R. Germeroth, Flaherty-Walia, Kerry, B. Sauls, and T. Switzer. Brief summary of fwri-fdm tag-recapture program and brief summary of fwri-fim tag-recapture data. *SEDAR Grey Snapper Data Workshop*, 2016. URL http://sedarweb.org/docs/wpapers/S51_DW_01_Tag-Recap%20Program%20Summary%20Final.pdf.

D. Hanisko and A. Pollack. Summary of length and weight data for gray snapper (lutjanus griseus) collected during nmfs and seamap fishery-independent surveys in the gulf of mexico. *SEDAR Grey Snapper Data Workshop*, 2017. URL http://sedarweb.org/docs/wpapers/S51_DW_02_Gray_Snapper_Length_and_Weight_Observations_from_Fisheries-Independent_Data.pdf.

R. Hilborn. The state of the art in stock assessment: Where we are and where we are going. *Scientia Marina*, 67(S1):15–20, 2003.

J. U. Hjorth. *Computer intensive statistical methods: Validation, model selection, and bootstrap*. CRC Press, 1993.

A. Kokkalis, U. H. Thygesen, A. Nielsen, and K. H. Andersen. Limits to the reliability of size-based fishing status estimation for data-poor stocks. *Fisheries Research*, 171:4–11, 2015.

K. Lorenzen and K. Enberg. Density-dependent growth as a key mechanism in the regulation of fish populations: evidence from among-population comparisons. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1486):49–54, 2002.

A. Magnusson, A. Punt, and R. Hilborn. Measuring uncertainty in fisheries stock assessment: the delta method, bootstrap, and MCMC. *Fish and Fisheries*, pages 325–342, 2012.

M. N. Maunder and K. R. Piner. Contemporary fisheries stock assessment: many issues still remain. *ICES Journal of Marine Science*, 72(1):7–18, 2014.

M. N. Maunder and K. R. Piner. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. *Fisheries Research*, 192: 16–27, 2017.

M. N. Maunder and A. E. Punt. A review of integrated analysis in fisheries stock assessment. *Fisheries Research*, 142:61–74, 2013.

R. D. Methot. Technical description of the stock synthesis II assessment program version 1.17-March 2005. *Unpublished draft report provided on the CD-ROM of background materials for the STAR. NOAA Fisheries, Seattle, Washington, USA*, 2005.

R. Muller. Standardized reef fish visual census index for gray snapper, lutjanus griseus,for the florida reef track from biscayne bay through the florida keys for 1997-2016. *SEDAR Grey Snapper Data Workshop*, 2017. URL http://sedarweb.org/docs/wpapers/S51_DW_14_RVC%20index%20Gray%20Snapper%20S51.pdf.

C. L. Needle. Using self-testing to validate the surbar survey-based assessment model. *Fisheries Research*, 171:78–86, 2015.

A. Pollack, H. David S., and G. Ingram Jr. Gray snapper abundance indices from seamap groundfish surveys in the northern gulf of mexico. *SEDAR Grey Snapper Data Workshop*, 2017. URL http://sedarweb.org/docs/wpapers/S51_DW_03_Gray%20Snapper%20SEAMAP%20Groundfish_v2.pdf.

J. Prince, S. Victor, V. Kloulchad, and A. Hordyk. Length based spr assessment of eleven indo-pacific coral reef fish populations in palau. *Fisheries Research*, 171:42–58, 2015.

V. Restrepo and J. Powers. Precautionary control rules in us fisheries management: specification and performance. *ICES Journal of Marine Science: Journal du Conseil*, 56(6):846–852, 1999.

R. H. Roa-Ureta, C. Molinet, N. Barahona, and P. Araya. Hierarchical statistical framework to combine generalized depletion models and biomass dynamic models in the stock assessment of the chilean sea urchin (loxechinus albus) fishery. *Fisheries Research*, 171:59–67, 2015.

S51 Stock ID Working Group. Stock id working paper. *SEDAR Grey Snapper Data Workshop*, 2017. URL http://sedarweb.org/docs/wpapers/S51_DW_09_Stock_ID_Working%20paper_FINAL.pdf.

M. Smith. Standardized catch-per-unit effort index for gulf of mexico gray snapper lutjanus griseus commercial handline fishery (1993 2015). *SEDAR Grey Snapper Data Workshop*, 2017. URL http://sedarweb.org/docs/wpapers/S51_DW_12_Commercial_Handline_Index_Final.pdf.

K. Thompson, T. Switzer, M. C. Christman, S. Keenan, C. Gardner, M. Campbell, and P. Adam. Indices of abundance for gray snapper (lutjanus griseus) using combined data from three independent video surveys. *SEDAR Grey Snapper Data Workshop*, 2017a. URL http://sedarweb.org/docs/wpapers/S51_DW_15_Combined%20video%20survey%20index%20.pdf.

K. Thompson, T. Switzer, and S. Keenan. Indices of abundance for gray snapper (lutjanus griseus) from the florida fish and wildlife research institute (fwri) video survey on the west florida shelf. *SEDAR Grey Snapper Data Workshop*, 2017b. URL http://sedarweb.org/docs/wpapers/S51_DW_10_FWRI_cameraindex_Lgris_2017_update.pdf.

L. Thornton, L. Lombardi, and R. Allman. Description of age data and estimated growth of gray snapper from the northern gulf of mexico: 1982-1983 and 1990-2015. *SEDAR Grey Snapper Data Workshop*, 2017. URL http://sedarweb.org/docs/wpapers/S51_DW_08_GS_S51_agelengthgrowth_DWreport.pdf.

J. T. Thorson and J. M. Cope. Catch curve stock-reduction analysis: An alternative solution to the catch equations. *Fisheries Research*, 171:33–41, 2015.

6. Figures

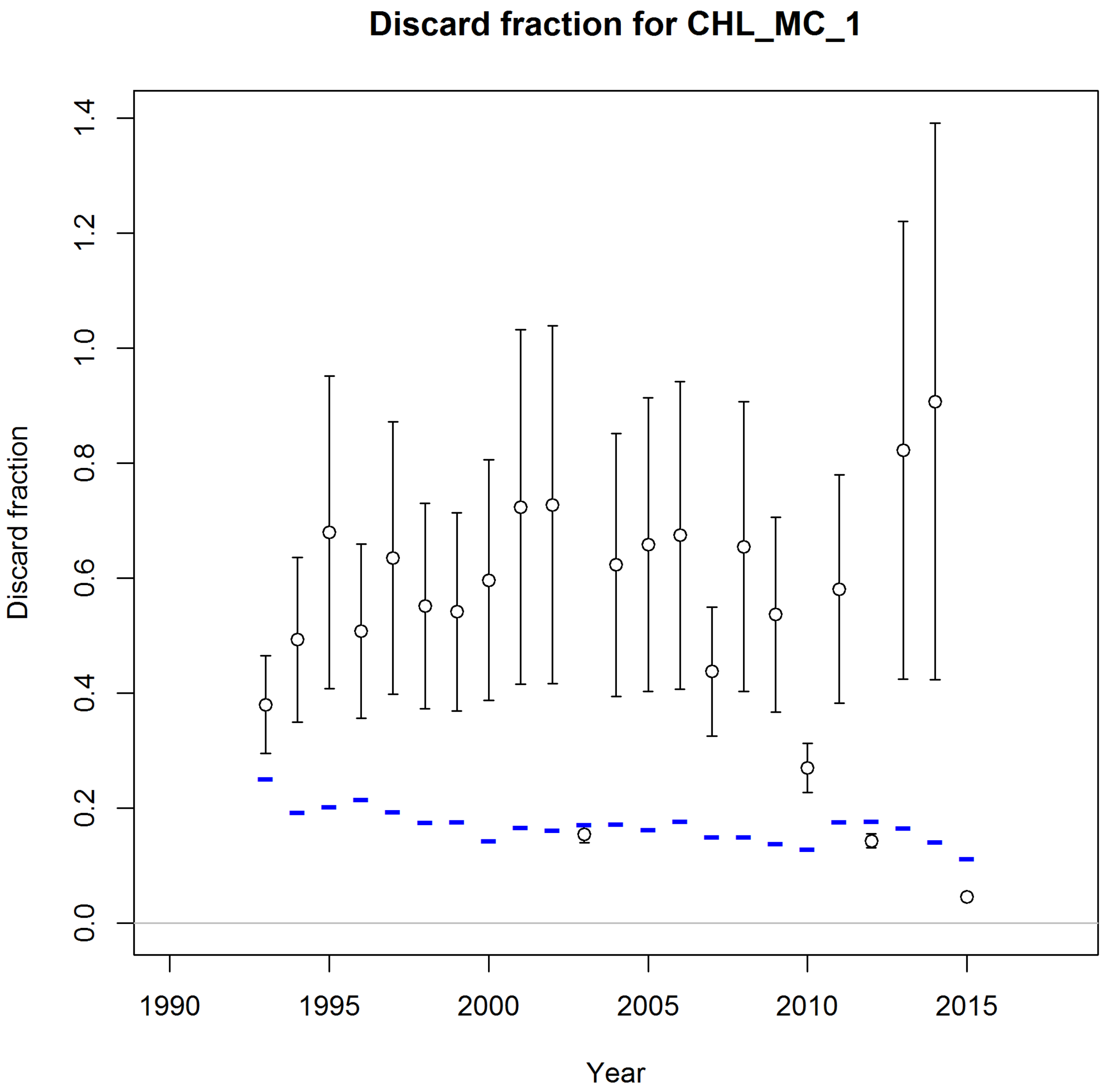


Figure 1: Model fit to discards f.

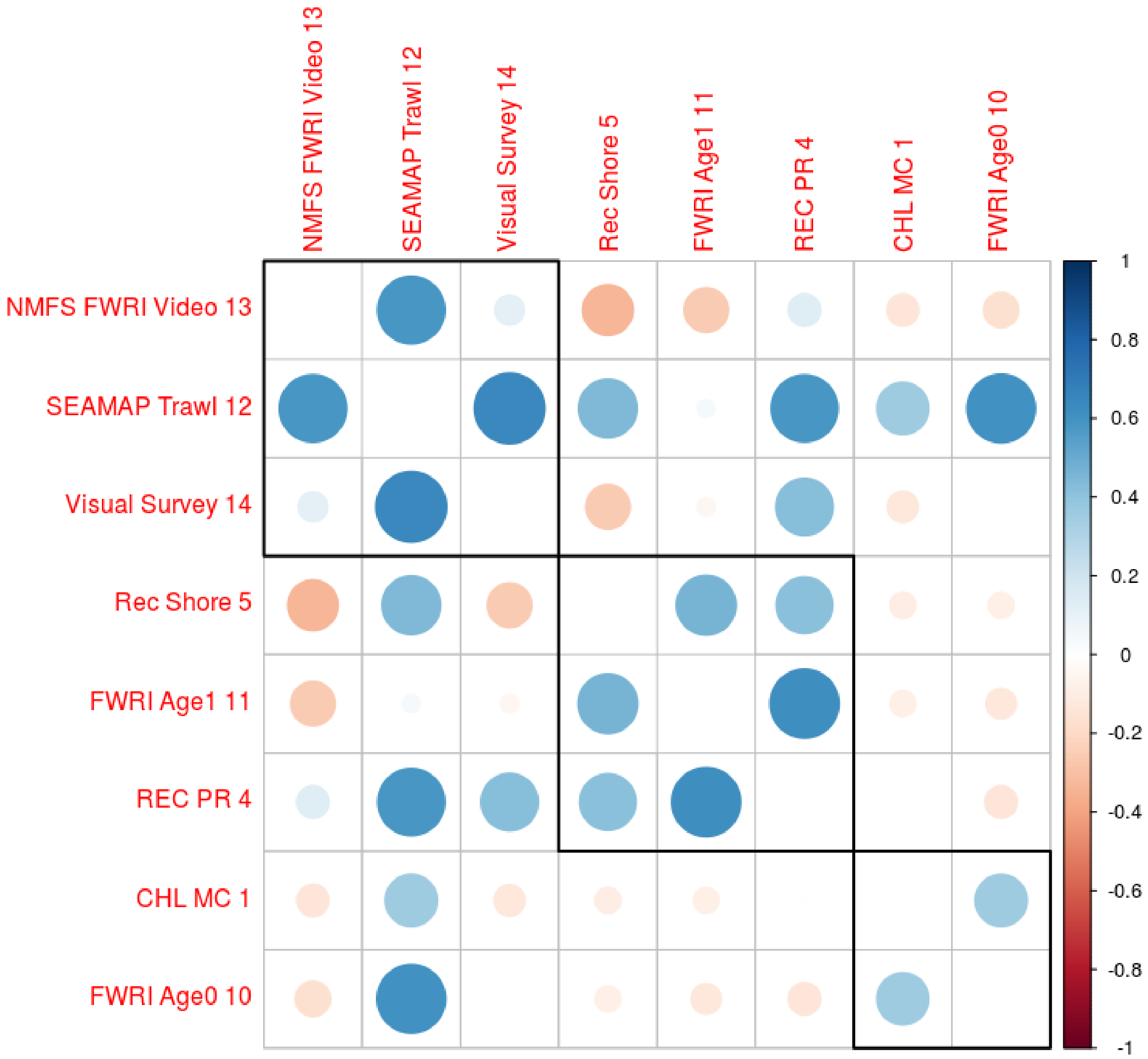


Figure 2: Plot of the correlation matrix for the Southern CPUE indices, blue indicate a positive correlation and red negative. the order of the indices and the rectanglur boxes are chosen based on a hierarchical cluster analysis using a set of dissimilarities for the indices being clustered.

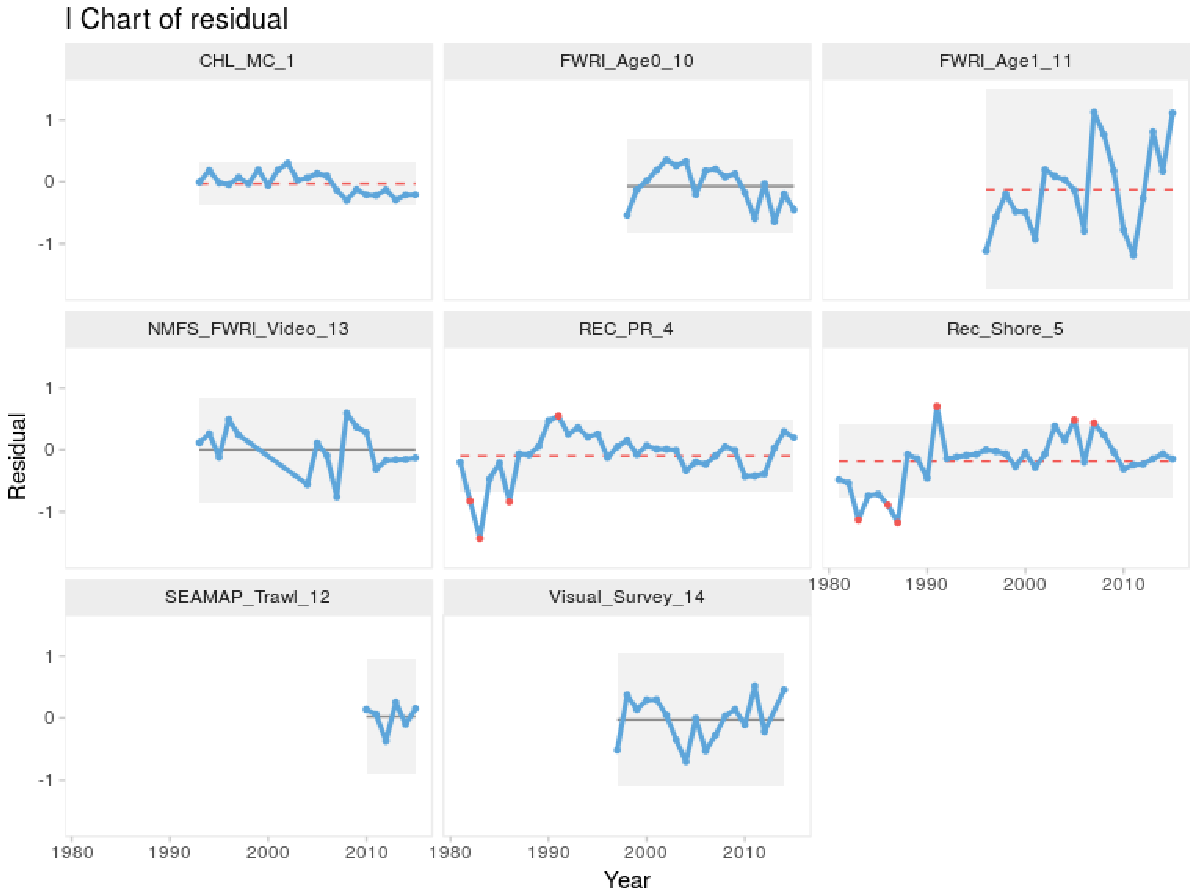


Figure 3: Runs chart showing the residuals by year; red points indicate points that violate the 3 sigma rule, and the red dashed line indicates unusually long runs or unusually few crossings.

Figure 4: Time series of recruitment, SSB, catch and fishing mortality, as estimated by the base case.

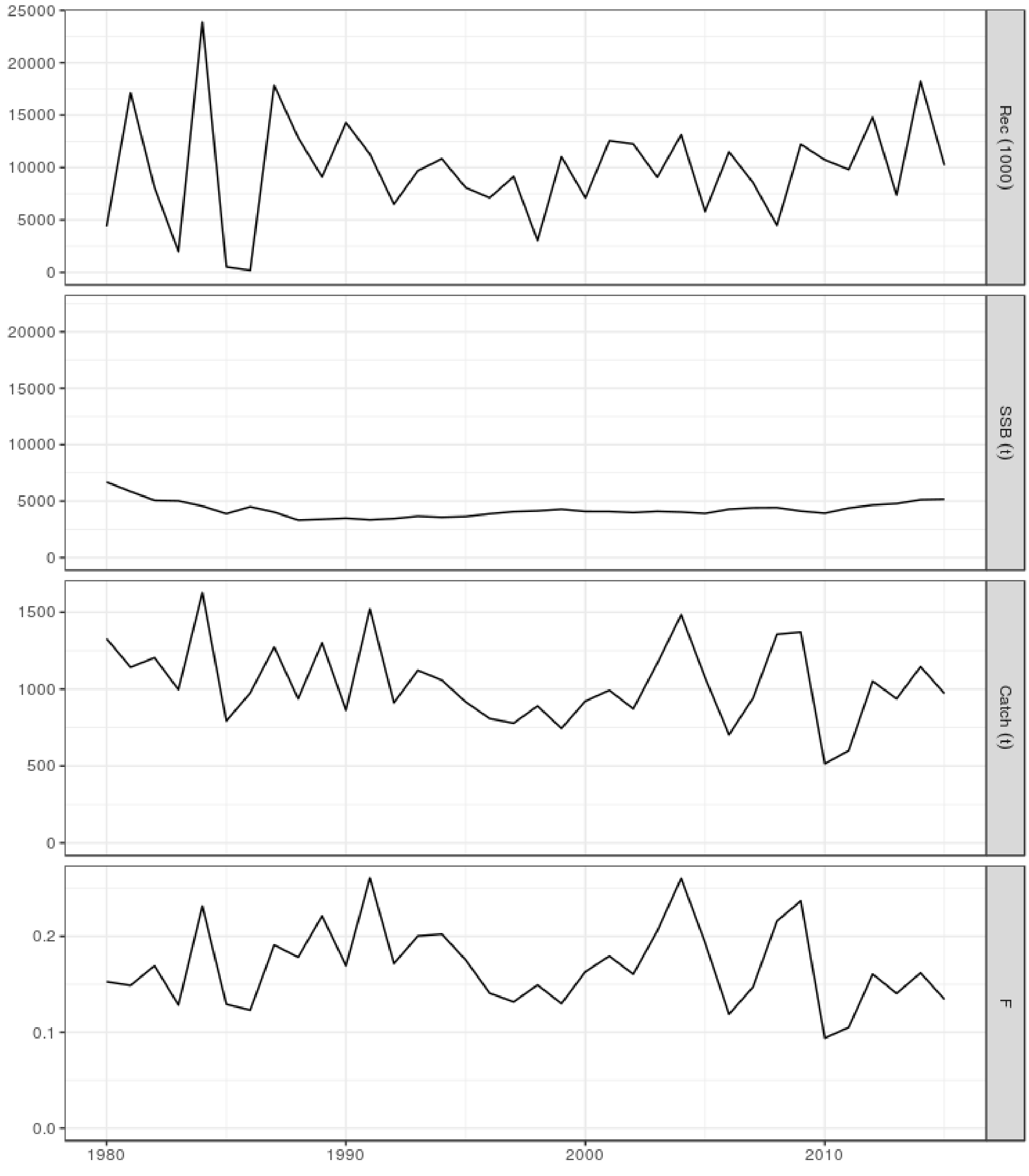


Figure 5: Time series plotted from 1980 of recruitment, SSB, catch and fishing mortality, as estimated by the base case.

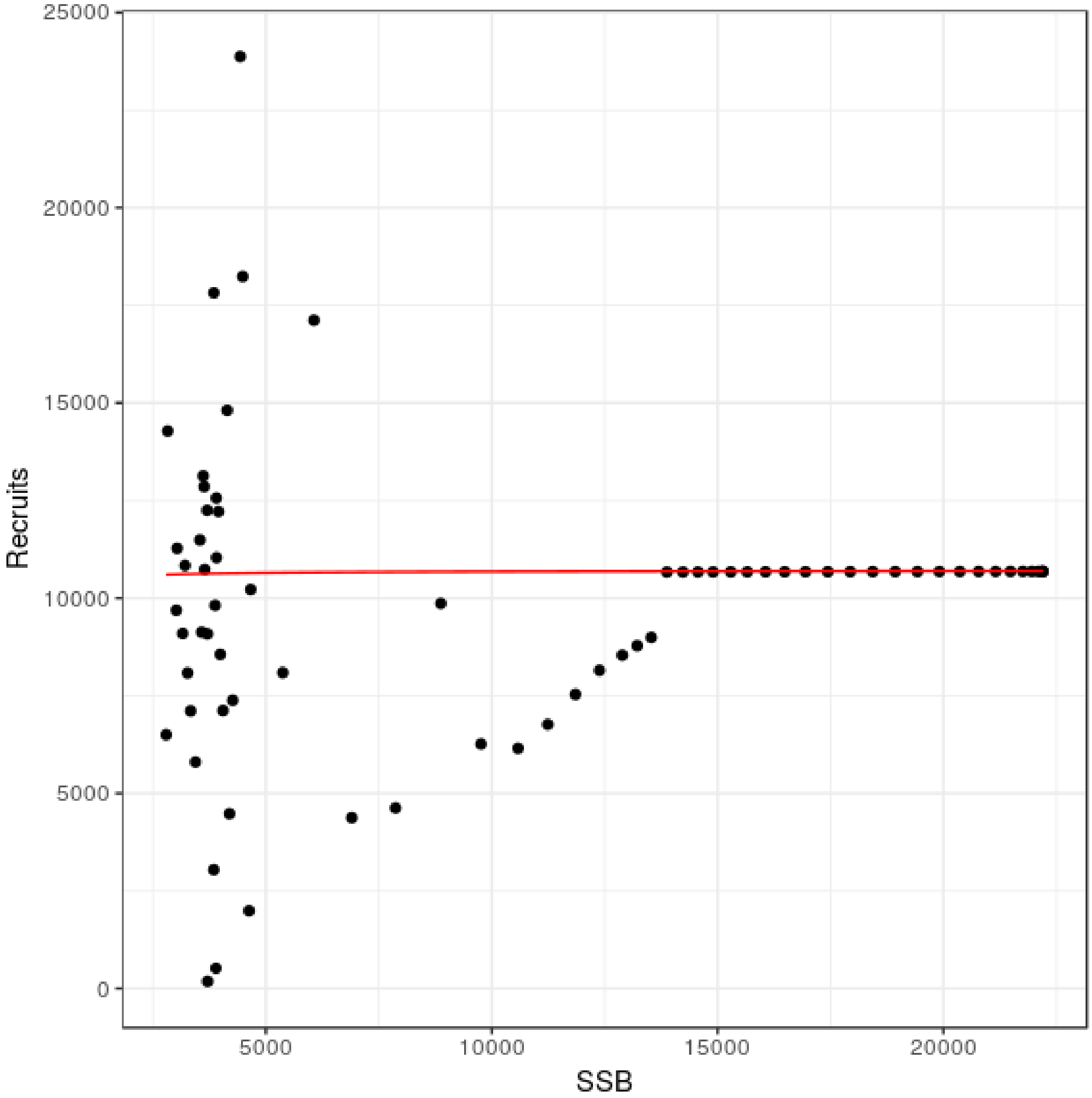


Figure 6: SSB, recruits and the Beverton and Holt relationship estimated by SS..

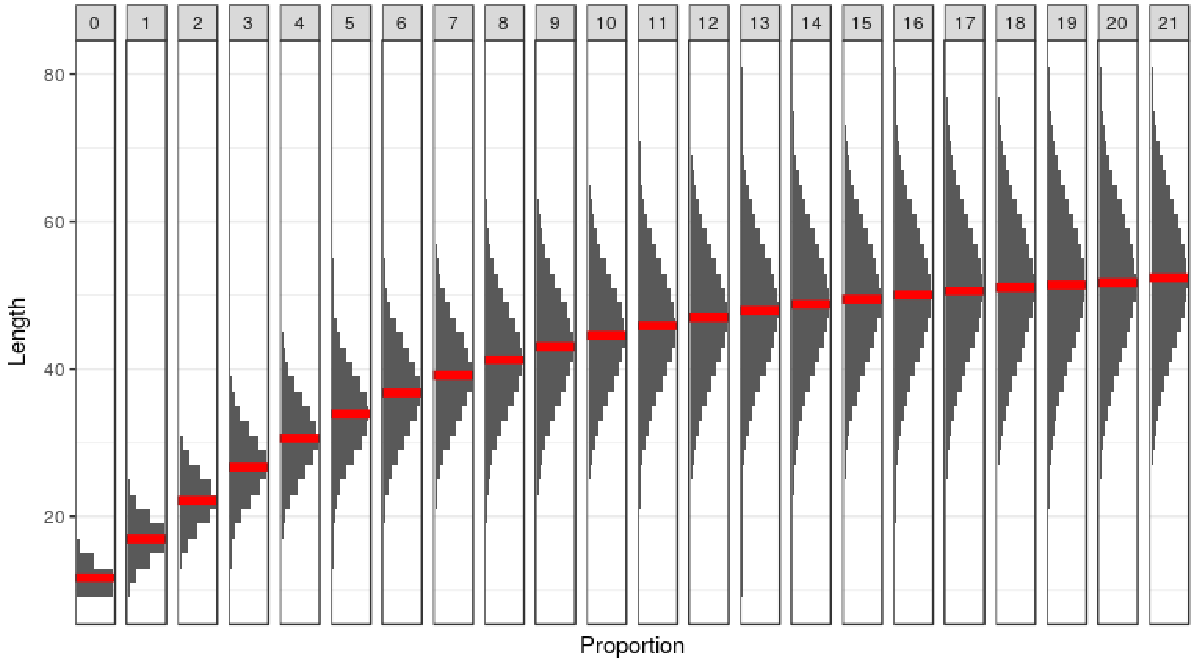


Figure 7: Length distributions by age from the age length key generated by SS3

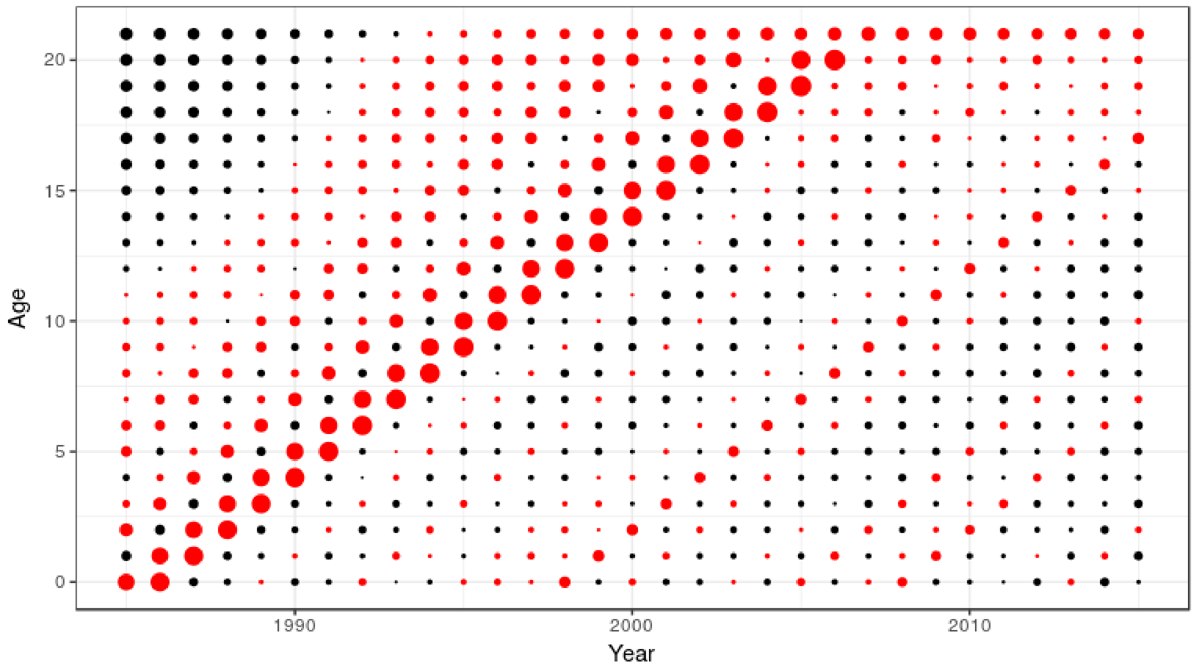


Figure 8: Log numbers-at-age scaled (by mean across age) to show year-class trends (red negative and black positive).

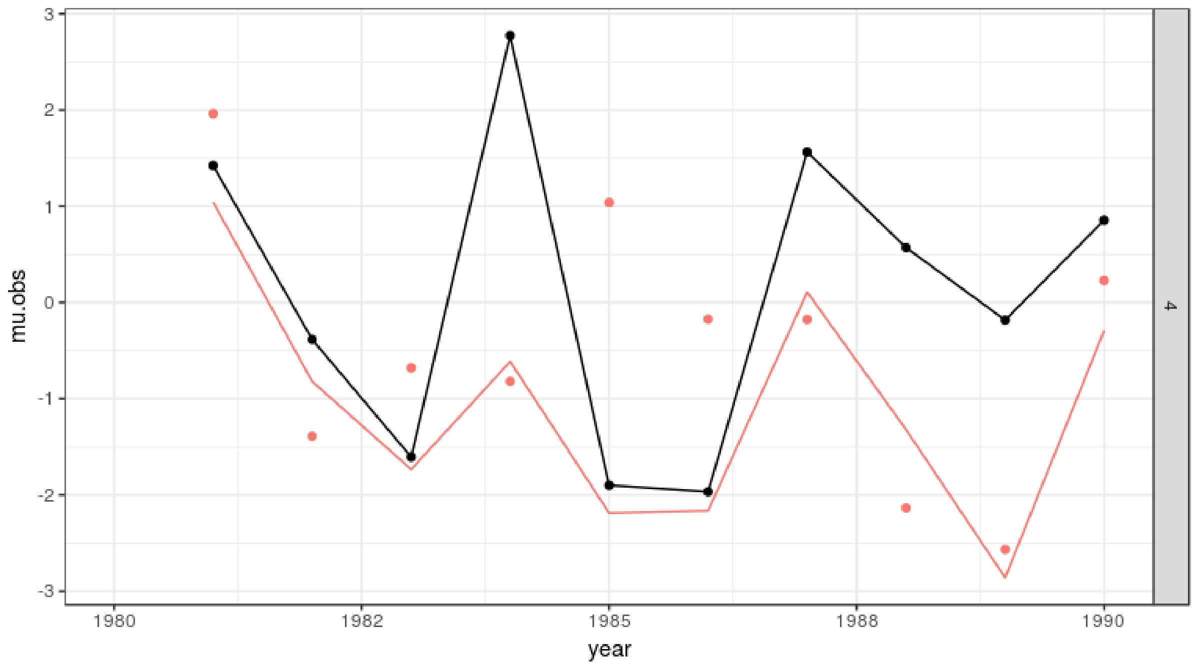


Figure 9: Plots of the recruitment deviates (black line) and the fitted mean length estimates (red line) and their observed values (red points).

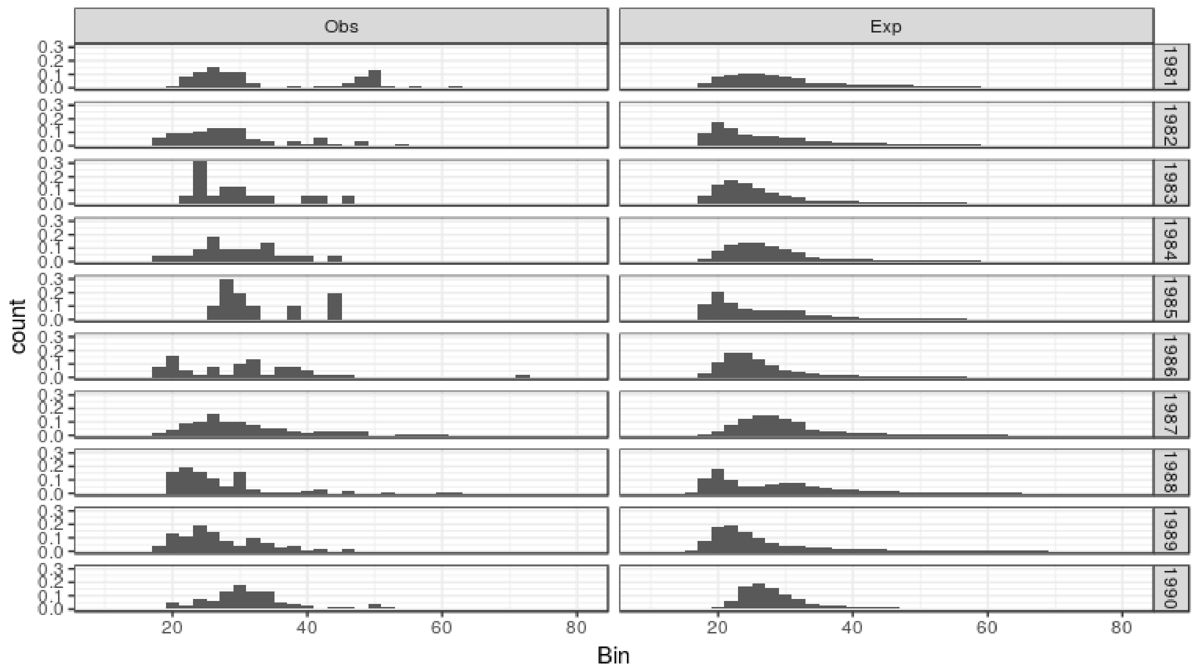


Figure 10: Plot of observed and fitted length distributions.

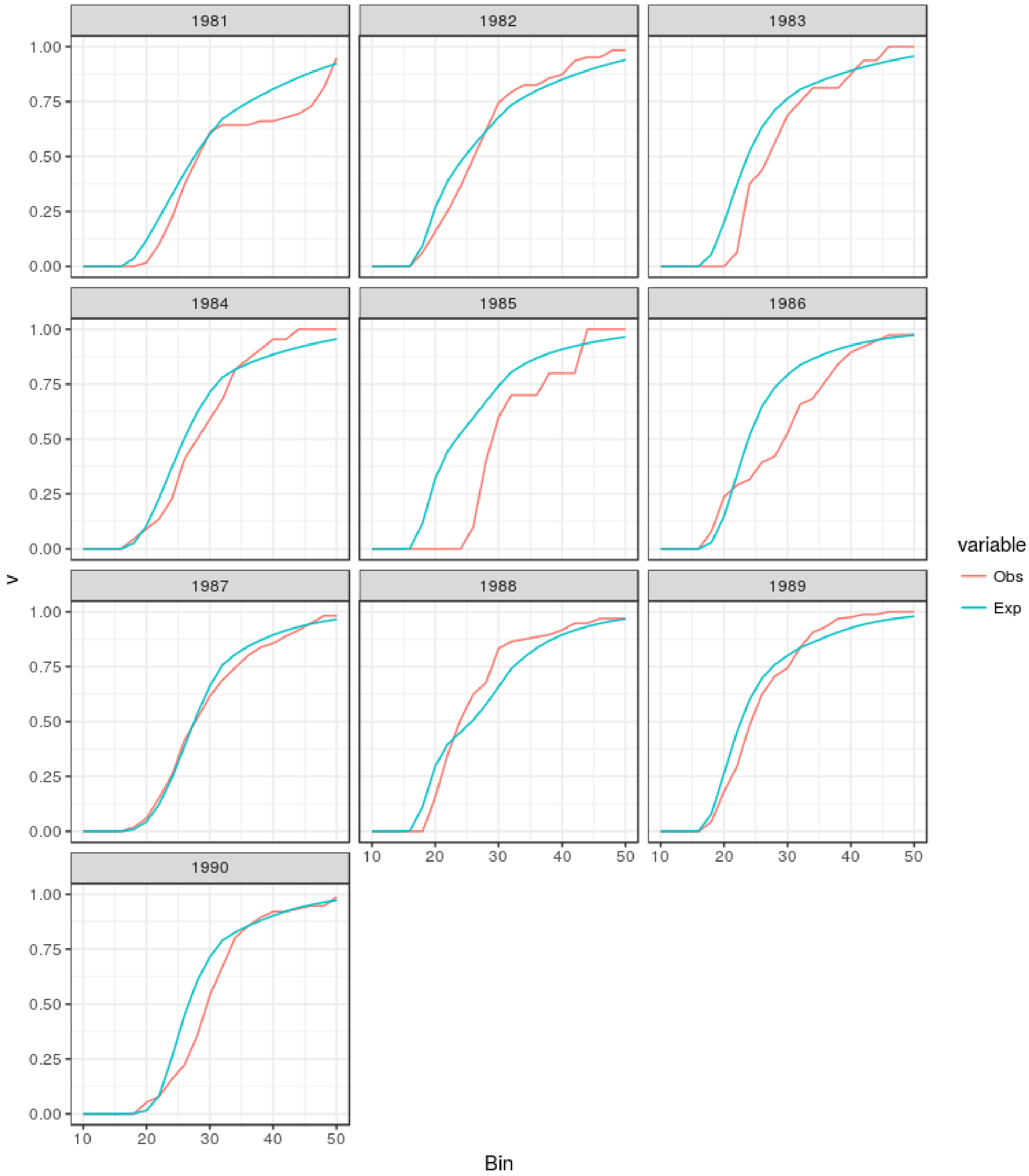


Figure 11: Plot of the cummulative distribution of the observed and fitted length distributions.

7. Appendices

7.1 Bibliography of materials provided for review

SEDAR 51
Gulf of Mexico Gray Snapper
Document List

Document #	Title	Authors	Date Submitted
Documents Prepared for the Data Workshop			
SEDAR51-DW-01	Brief Summary of FWRI-FDM Tag-Recapture Program and Brief Summary of FWRI-FIM Tag-Recapture Data	Rachel Germeroth, Kerry Flaherty-Walia, Beverly Sauls and Ted Switzer	4 Nov 2016
SEDAR51-DW-02	Summary of length and weight data for gray snapper (*Lutjanus griseus*) collected during NMFS and SEAMAP fishery-independent surveys in the Gulf of Mexico	David S. Hanisko and Adam Pollack	20 March 2017
SEDAR51-DW-03	Gray Snapper Abundance Indices from SEAMAP Groundfish Surveys in the Northern Gulf of Mexico	Adam G. Pollack, David S. Hanisko and G. Walter Ingram, Jr	7 April 2017 Updated: 12 June 2017
SEDAR51-DW-04	Length frequency distributions for gray snapper length and age samples collected from the Gulf of Mexico	Ching-Ping Chih	9 April 2017
SEDAR51-DW-05	Gray snapper *Lutjanus griseus* Findings from the NMFS Panama City Laboratory Camera Fishery-Independent Survey 2005-2015	C.L. Gardner, D.A. DeVries, K.E. Overly, and A.G. Pollack	7 April 2017
SEDAR51-DW-06	Reproductive parameters for the Gulf of Mexico gray snapper, *Lutjanus griseus,* 1991-2015	G.R. Fitzhugh, V.C. Beech, H.M. Lyon, and P. Colson	13 April 2017
SEDAR51-DW-07	SEAMAP Reef Fish Video Survey: Relative Indices of Abundance of Grey Snapper	Matthew D. Campbell, Kevin R. Rademacher, Michael Hendon, Paul Felts, Brandi Noble, Ryan Caillouet, Joseph Salisbury, and John Moser	10 April 2017
SEDAR51-DW-08	Description of age data and estimated growth of Gray Snapper from the northern Gulf of Mexico: 1982-1983 and 1990-2015	L.A. Thornton, L.A. Lombardi, and R.J. Allman	14 April 2017
SEDAR51-DW-09	SEDAR 51 Stock ID Working Paper	S51 Stock ID Working Group	February 2017
SEDAR51-DW-10	Indices of abundance for Gray Snapper (*Lutjanus griseus*) from the Florida Fish and Wildlife Research Institute (FWRI) video survey on the West Florida Shelf	Kevin A. Thompson, Theodore S. Switzer, and Sean F. Keenan	21 April 2017
SEDAR51-DW-11	Gray Snapper Abundance Indices from Inshore Surveys of Northeastern Gulf of	Kerry E. Flaherty-Walia, Theodore S.	24 April 2017 Updated: 27

	Mexico estuaries	Switzer, and Amanda J. Tyler-Jedlund	April 2017
SEDAR51-DW-12	Standardized Catch-Per-Unit Effort Index for Gulf of Mexico Gray Snapper *Lutjanus griseus* Commercial Handline Fishery (1993 – 2015)	Matthew W. Smith	26 April 2017
SEDAR51-DW-13	Commercial Landings of Gray or Mangrove Snapper (*Lutjanus griseus*) from the Gulf of Mexico	Refik Orhun and Beth Wrege	
SEDAR51-DW-14	Standardized Reef Fish Visual Census index for Gray Snapper, *Lutjanus griseus*, for the Florida reef track from Biscayne Bay through the Florida Keys for 1997-2016	Robert G. Muller	6 June 2017
SEDAR51-DW-15	Indices of abundance for Gray Snapper (*Lutjanus griseus*) using combined data from three independent video surveys	Kevin A. Thompson, Theodore S. Switzer, Mary C. Christman, Sean F. Keenan, Christopher Gardner, Matt Campbell, Adam Pollack	15 June 2017
Documents Prepared for the Assessment Process			
SEDAR51-AW-01			
SEDAR51-AW-02			
SEDAR51-AW-03			
SEDAR51-AW-04			
SEDAR51-AW-05			
Documents Prepared for the Review Workshop			
SEDAR51-RW-01			
SEDAR51-RW-02			
Final Stock Assessment Reports			
SEDAR51-SAR1	Gulf of Mexico Gray Snapper	SEDAR 51 Panels	
Reference Documents			
SEDAR51-RD01	Short-Term Discard Mortality Estimates for Gray Snapper in a West-Central Florida Estuary and Adjacent Nearshore Gulf of Mexico Waters	Kerry E. Flaherty-Walia, Brent L. Winner, Amanda J. Tyler-Jedlund & John P. Davis	
SEDAR51-RD02	Regional Correspondence in Habitat Occupancy by Gray Snapper (*Lutjanus griseus*) in Estuaries of the Southeastern	Kerry E. Flaherty & Theodore S. Switzer & Brent L. Winner & Sean F. Keenan	

SEDAR 51
Gulf of Mexico Gray Snapper
Document List

	United States	
SEDAR51-RD03	Improved Ability to Characterize Recruitment of Gray Snapper in Three Florida Estuaries along the Gulf of Mexico through Targeted Sampling of Polyhaline Seagrass Beds	Kerry E. Flaherty-Walia, Theodore S. Switzer, Brent L. Winner, Amanda J. Tyler-Jedlund & Sean F. Keenan
SEDAR51-RD04	Conservation Genetics of Gray Snapper (*Lutjanus griseus*) in U.S. Waters of the Northern Gulf of Mexico and Western Atlantic Ocean	John R. Gold, Eric Saillant, N. Danielle Ebelt, and Siya Lem
SEDAR51-RD05	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
SEDAR51-RD06	Age, growth, and mortality of gray snapper, *Lutjanus griseus*, from the east coast of Florida	Michael L. Burton
SEDAR51-RD07	Ingress of transformation stage gray snapper, *Lutjanus griseus (*Pisces: Lutjanidae) through Beaufort Inlet, North Carolina	Mimi W. Tzeng, Jonathan A. Hare, and David G. Lindquist
SEDAR51-RD08	Biological response to changes in climate patterns: population increases of gray snapper (*Lutjanus griseus*) in Texas bays and estuaries	James M. Tolan and Mark Fisher
SEDAR51-RD09	Returns from the 1965 Schlitz tagging program including a cumulative analysis of previous results	Dale S. Beaumariage
SEDAR51-RD10	Recruitment dynamics and otolith chemical signatures of juvenile gray snapper, *Lutjanus griseus,* among West Florida estuarine and coastal marine ecosystems	Cecelia Lounder
SEDAR51-RD11	Reproductive biology of gray snapper (*Lutjanus griseus*), with notes on spawning for other Western Atlantic snappers (Lutjanidae)	M.L. Domeier, C. Koenig, and F. Colman
SEDAR51-RD12	Climate-related, decadal-scale assemblage changes of seagrass-associated fishes in the northern Gulf of Mexico	F. Joel Fodrie, Kenneth L. Heck, Jr., Sean P. Powers, William M. Graham, and Kelly L. Robinson
SEDAR51-RD13	Response of coastal fishes to the Gulf of Mexico oil disaster	F. Joel Fodrie and Kenneth L. Heck Jr.
SEDAR51-RD14	Variation in the isotopic signatures of juvenile gray snapper (*Lutjanus griseus*) from five southern Florida regions	Trika Gerard and Barbara Muhling
SEDAR51-RD15	Temporal and spatial dynamics of spawning, settlement, and growth of gray	Robert J. Allman and Churchill B. Grimes

	snapper (*Lutjanus griseus*) from the West Florida shelf as determined from otolith microstructures	
SEDAR51-RD16	Regional variation in the population structure of gray snapper, *Lutjanus griseus,* along the West Florida shelf	R.J. Allman and L.A. Goetz
SEDAR51-RD17	Evaluating juvenile thermal tolerance as a constraint on adult range of gray snapper (Lutjanus griseus): A combined laboratory, field and modeling approach	Mark J. Wuenschel, Jonathan A. Hare, Matthew E. Kimball, and Kenneth W. Able
SEDAR51-RD18	Growth variation, settlement, and spawning of gray snapper across a latitudinal gradient	Kelly Denit and Su Sponaugle
SEDAR51-RD19	Age, growth, mortality, and radiometric age validation of gray snapper (Lutjanus griseus) from Louisiana	Andrew J. Fischer, M. Scott Baker, Jr., Charles A. Wilson, and David L. Nieland
SEDAR51-RD20	Southeast Florida reef fish abundance and biology: Five year performance report	Luiz R. Barbieri and James A. Colvocoresses
SEDAR51-RD21	Larval ecology of a suite of snappers (family: Lutjanidae) in the Straits of Florida, western Atlantic Ocean	E. K. D'Alessandro, S. Sponaugle, and J. E. Serafy
SEDAR51-RD-22	Multidecadal otolith growth histories for red and gray snapper (Lutjanus spp.) in the northern Gulf of Mexico, USA	Bryan A. Black, Robert J. Allman, Isaac D. Schroeder, and Michael J. Schirripa
SEDAR51-RD-23	Investigations on the Gray Snapper, Lutjanus griseus	Walter A. Starck II and Robert E. Schroede
SEDAR51-RD-24	Age-size Structure of Gray Snapper from the Southeastern United States: A Comparison of Two Methods of Back-calculating Size at Age from Otolith Data	A.G. Johnson, L.A. Collins, and C.P. Keim

7.2 Statement of Work

Statement of Work

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

SEDAR 51 Gulf of Mexico Grey Snapper 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

SEDAR 51 will be a compilation of data, an assessment of the stock, and CIE assessment review conducted for Gulf of Mexico Gray Snapper. 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 stock assessed through SEDAR 51 is within the jurisdiction of the Gulf of Mexico Fisheries Management Council and the states of Florida, Mississippi, Alabama, Louisiana, and Texas.

Requirements

NMFS requires three (3) CIE reviewers to conduct an impartial and independent peer review in accordance with the SoW, OMB guidelines, and the TORs below. The CIE reviewers shall have expertise 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.

Tasks for Reviewers

1) Review the following background materials and reports prior to the review meeting.
    a. Working papers, reference documents, and the Data Workshop and Assessment Process Reports will be available on the SEDAR website: http://sedarweb.org/sedar-51

2) Attend and participate in the panel review meeting. The meeting will consist of presentations by NOAA and other scientists, stock assessment authors and others to facilitate the review, to answer any questions from the reviewers, and to provide any additional information required by the reviewers.

3) After the review meeting, reviewers shall conduct an independent peer review report in accordance with the requirements specified in this SoW, OMB guidelines, and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus.

4) Each reviewer should assist the Chair of the meeting with contributions to the summary report.

5) Deliver their reports to the Government according to the specified milestones dates.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registration-system.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance

The place of performance shall be at the contractor's facilities, and at Tampa, FL.

Period of Performance
The period of performance shall be from the time of award through March 31, 2018. The CIE chair and each reviewer's duties shall not exceed 14 days to complete all required tasks.

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

Within two weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
February 13-15, 2018	Panel review meeting
Approximately 2 weeks later	Contractor receives draft reports
Within 2 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
All travel expenses shall be reimbursable in accordance with Federal Travel Regulations (http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not to exceed $10,000.

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

Project Contact:
Julie A. Neer, Ph.D.
SEDAR Coordinator
Science and Statistics Program
South Atlantic Fishery Management Council
4055 Faber Place Drive, Suite 201
North Charleston, SC 29405
(843) 571-4366 *Julie.neer@safmc.net*

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 Statement of Work
   Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Terms of Reference for the Peer Review

SEDAR 51 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, and consider the following:
    a) Are data decisions made by the DW and AW sound and robust?
    b) Are data uncertainties acknowledged, reported, and within normal or expected levels?
    c) Are data applied properly within the assessment model?
    d) Are input data series reliable and sufficient to support the assessment approach and findings?
2. Evaluate and discuss the strengths and weaknesses of the methods used to assess the stock, taking into account the available data, and considering the following:
    a) Are methods scientifically sound and robust?
    b) Are assessment models configured properly and consistent with standard practices?
    c) Are the methods appropriate for the available data?
3. Evaluate the assessment findings and consider the following:
    a) Are abundance, exploitation, and biomass estimates reliable, consistent with input data and population biological characteristics, and useful to support status inferences?
    b) Is the stock overfished? What information helps you reach this conclusion?
    c) Is the stock undergoing overfishing? What information helps you reach this conclusion?
    d) Is there an informative stock recruitment relationship? Is the stock recruitment curve reliable and useful for evaluation of productivity and future stock conditions?
    e) Are the quantitative estimates of the status determination criteria for this stock reliable? If not, are there other indicators that may be used to inform managers about stock trends and conditions?
4. Evaluate the stock projections, including discussing strengths and weaknesses, and consider the following:
    a) Are the methods consistent with accepted practices and available data?
    b) Are the methods appropriate for the assessment model and outputs?
    c) Are the results informative and robust, and useful to support inferences of probable future conditions?

7.3 Panel membership.

LIST OF PARTICIPANTS

Workshop Panel

Dr. Kai Lorenzen, Chair..Chair, GMFMC SSC (Univ. of FL)
Dr. Luiz Barbieri..GMFMC SSC (FWRI)
Dr. Yong Chen...CIE Reviewer (Univ. of Maine)
Bob Gill..GMFMC SSC
Dr. Laurence Kell...CIE Reviewer (Henstead, UK)
Dr. Peter Stephenson..CIE Reviewer (Perth, AUS)

Analytic Representation

Dr. Jeff Isely (Lead)..SEFSC, Miami
Dr. Shannon Cass-Calay..SEFSC, Miami

Appointed Observers

Ed Walker...Industry Representative

Observers

Beth Wrege...SEFSC, Miami

Staff

Dr. Julie Neer..SEDAR
Steven Atran...GMFMC
Doug Gregory...GMFMC
Karen Hoak..GMFMC
Ryan Rindone..GMFMC
Charlotte Schiaffo...GMFMC

7.4 Additional Analyses.

7.41 Indices of Abundance

Appendix 1: CPUE

SEDAR 51 Gulf of Mexico Gray Snapper Review

Laurence Kell

14 April, 2018

This document looks at the catch per unit effort (CPUE) data in order to indentify conflicts in the data and to help hypotheses that may help setting up stock assessment scenarios to run.

The CPUE time series are plotted in **figure 1**, to help compare trends by stock a lowess smoother is fitted to year using a general additive model (GAM).

To look at potential conflicts in the data, i.e. deviations from the overall trends, the residuals from the fits are compared in **figure 2**. Conflicts between indices can be idenfified by looking for patterns in the residuals.

If indices represent the same stock components then it is reasonable to expect them to be correlated, if indices are not correlated or negatively correlated, i.e. they show conflicting trends, this may result in poor fits to the data and bias in the estimates. Therefore the correlation between the indices is evaluated in **Figure 3**, the lower triangle show the pairwise scatter plots between the indicies with a regression line, the upper triangle the correlation coefficients and the diagonal the range of observations. A single influential point may cause a strong spurious correlation there it is important to look at trends as well as the correlation coefficients, since as a strong correlation could be found by chance if a few key points coincide.

The correlations can be used to select groups that represent a common hypotheses about the evolution of the stock, therefore **Figure 4** shows the results from a hierarchical cluster analysis using a set of dissimilarities.

Correlations between series may be lagged due to indices being dominated by particular age classes so the cross-correlations(lagged by -10 to 10 years) are plotted in **Figure 5**. The diagonals show the autocorrelations as an index is lagged against itself.

Figure 6 shows the selection patterns of the various indices used to calculate the indices, to help in interpreting any cross correlations

Figures

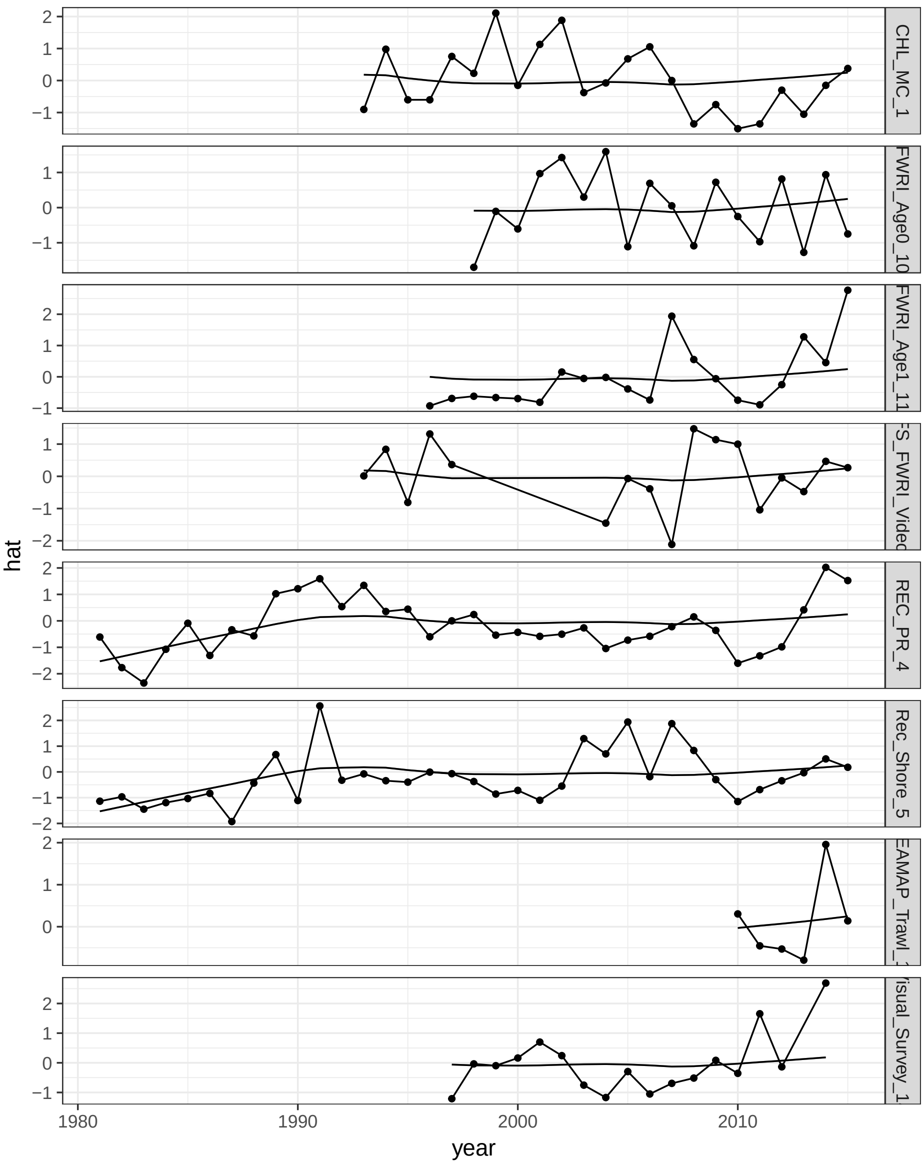


Figure 1. Time series of CPUE indices, continuous black line is a lowess smother showing the average trend by area (i.e. fitted to year for each area with series as a factor)

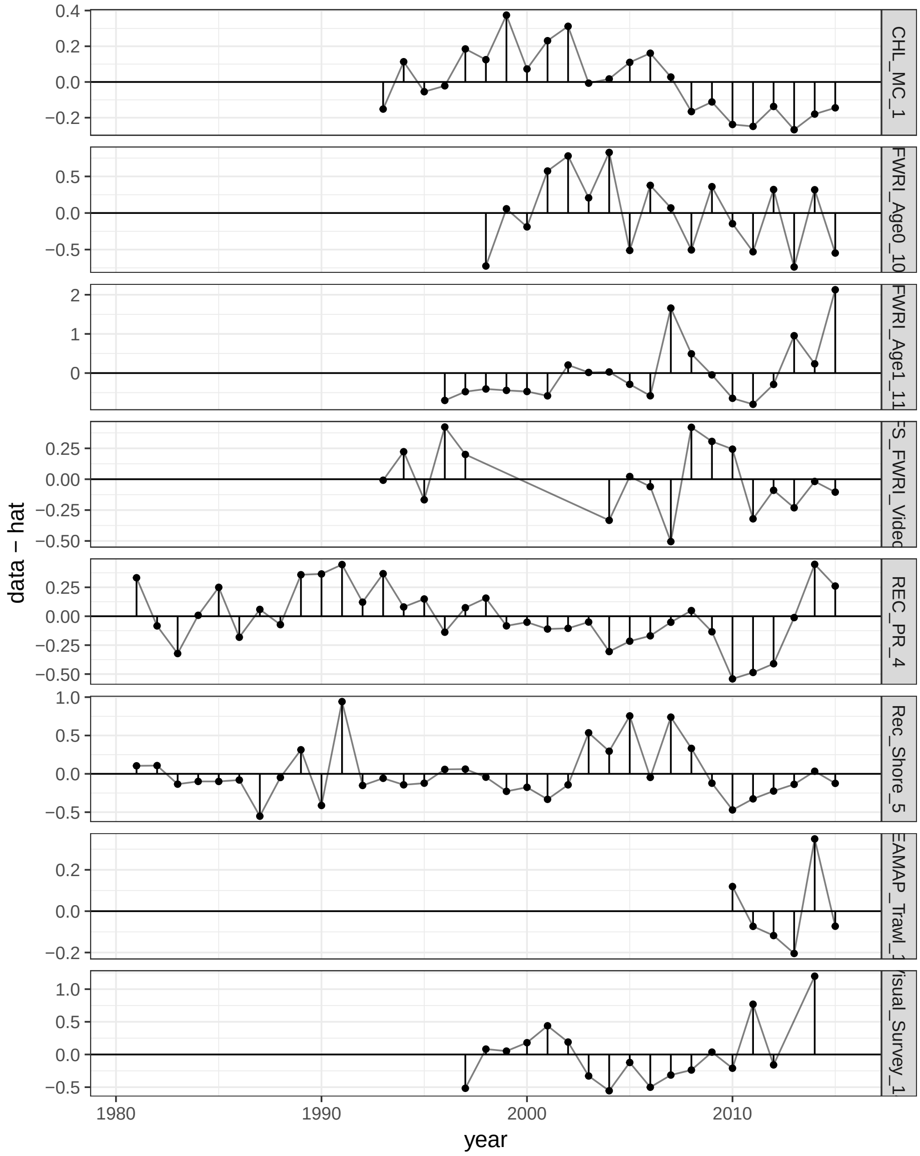


Figure 2. Time series of residuals from the lowess fit.

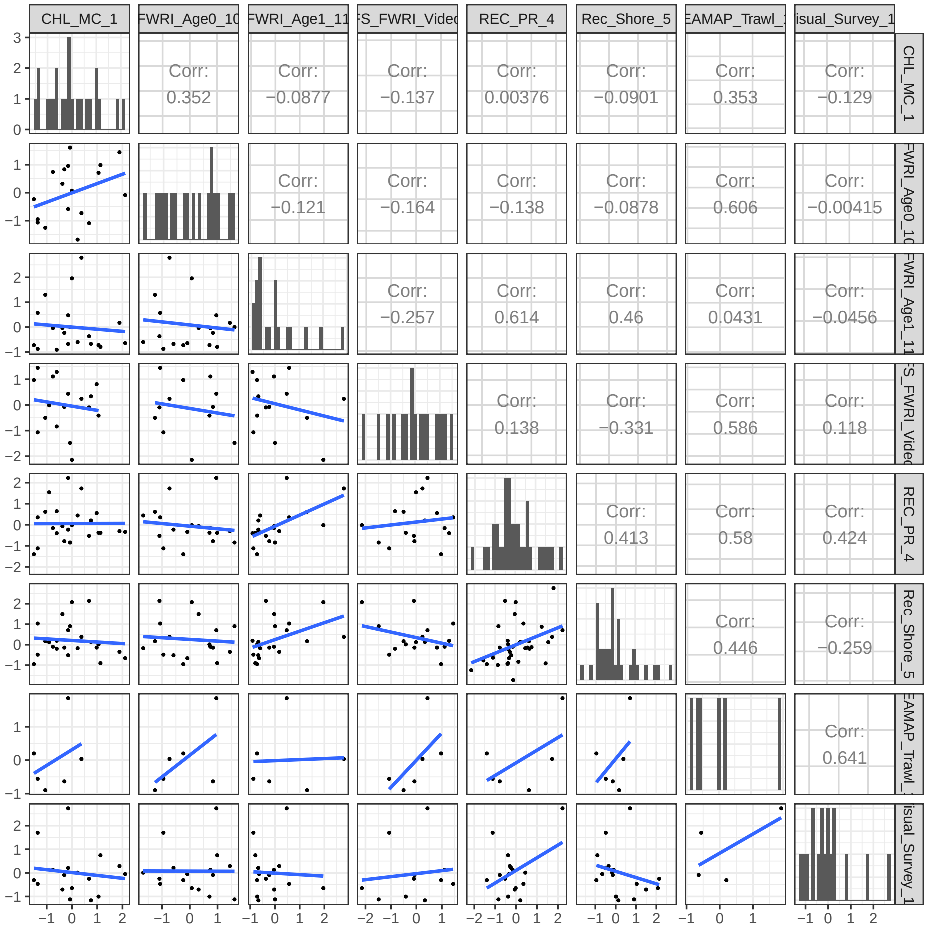
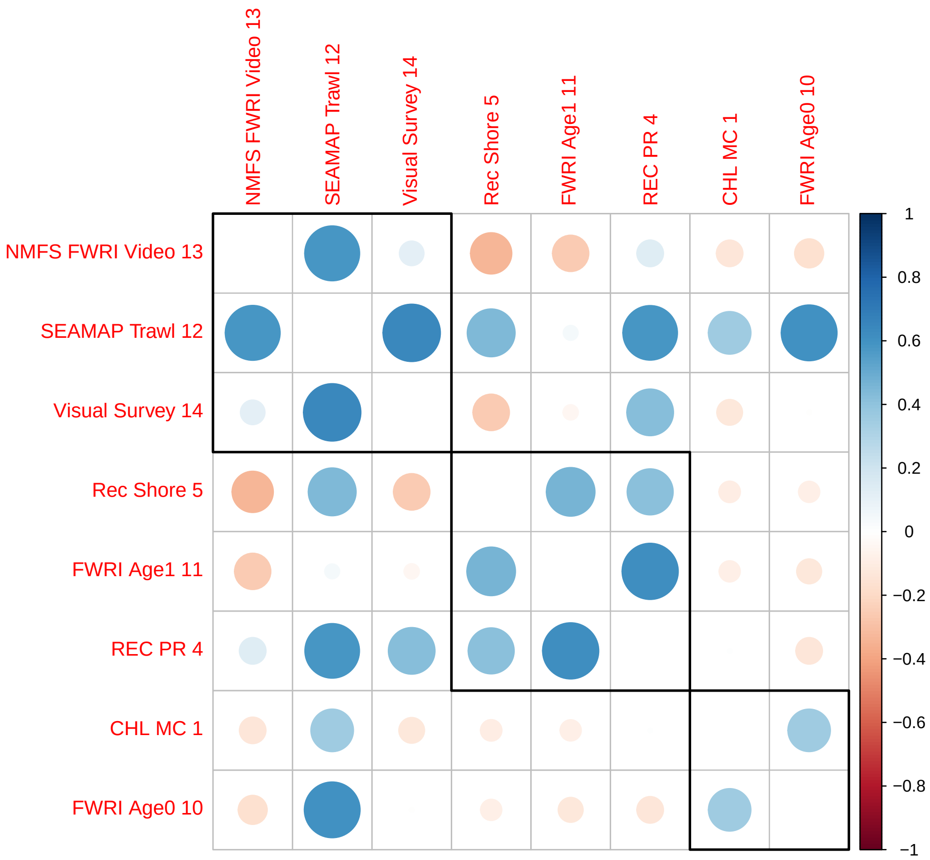


Figure 3. Pairwise scatter plots to look at correlations between Indices, North.



NULL

Figure 4. Plot of the correlation matrix for the Southern CPUE indices, blue indicate a positive correlation and red negative. the order of the indices and the rectanglur boxes are chosen based on a hierarchical cluster analysis using a set of dissimilarities for the indices being clustered.

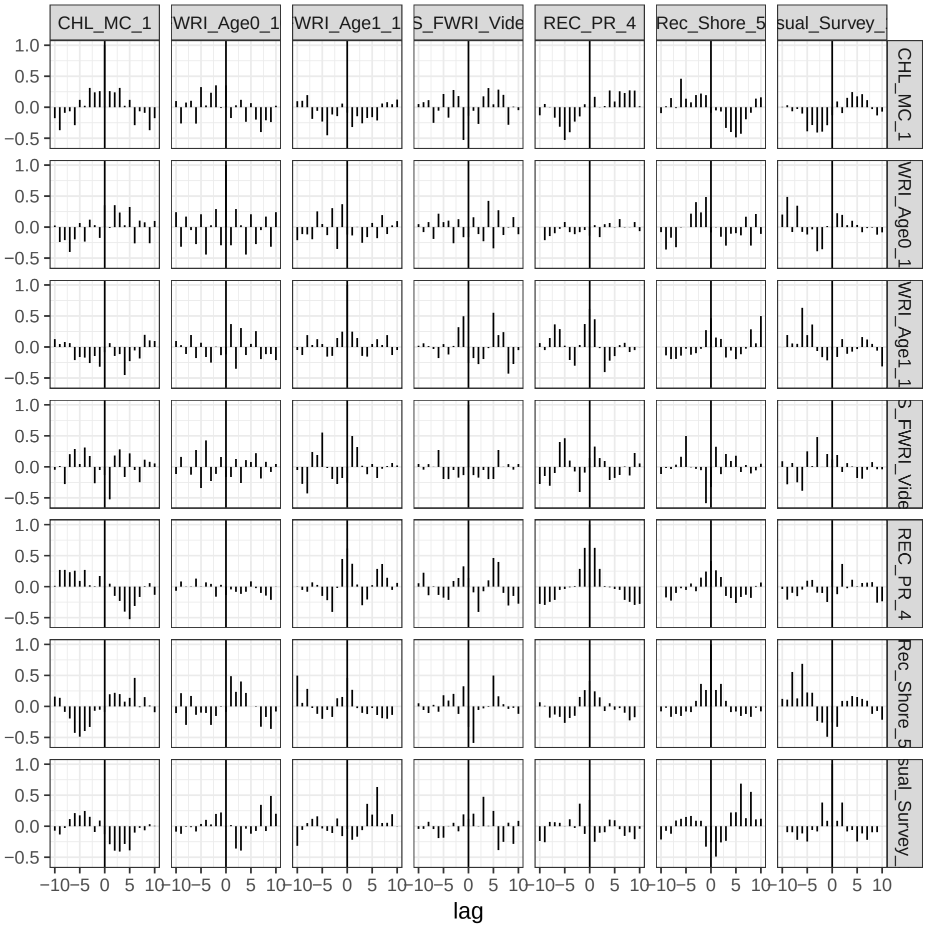


Figure 5 Cross correlations between Northern indices, to identify potential lags due to year-class effects.

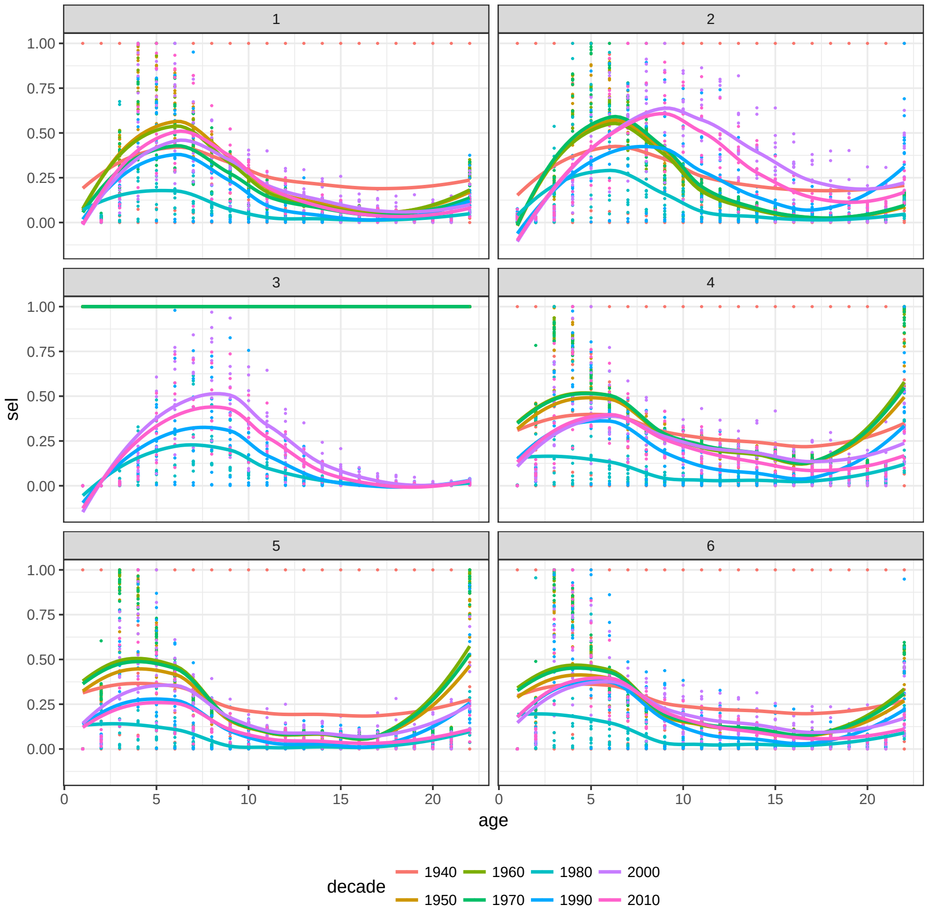


Figure 6 Catch curve analyses.

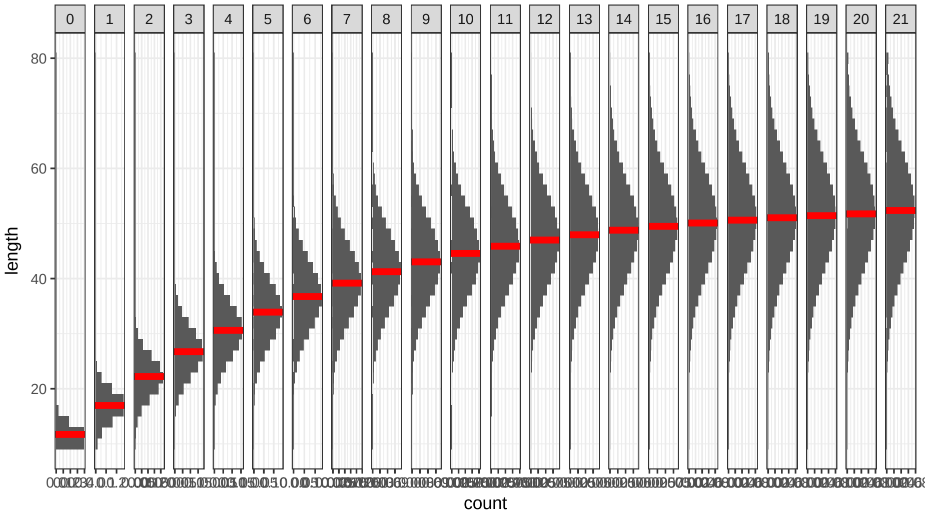


Figure 7 Growth from age length key generated by SS3.

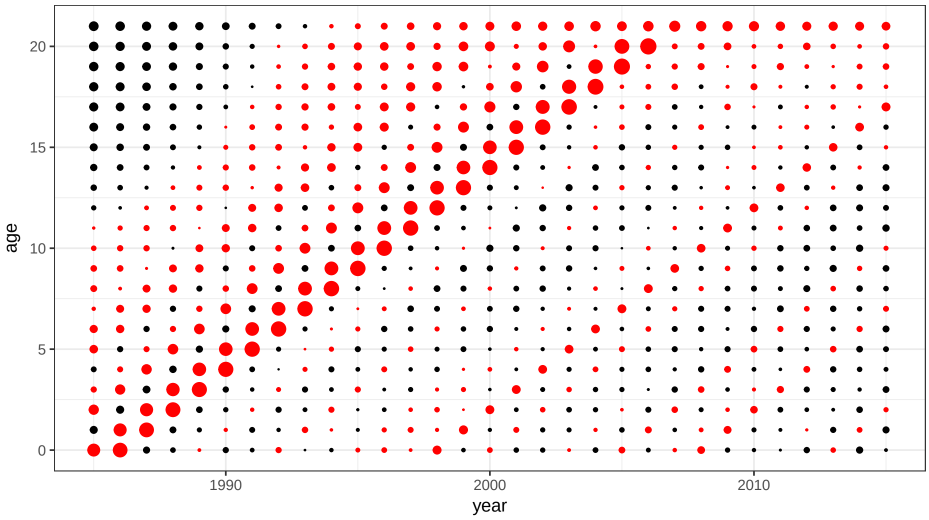


Figure 8 Numbers-at-age scaled to show year-class trends.

Software

All analysis was conducted using R and FLR and the `diags` package which provides a set of common methods for reading these data into R, plotting and summarising them.(http://www.flr-project.org/)

- R version 3.4.1 (2017-06-30)

- FLCore: 2.6.6.9006
- diags:
- ggplotFL:
- diags:
- r4ss:
- **Compiled**: Sat Apr 14 08:34:36 2018

Author information

Laurence KELL. laurie@kell.es

References

7.42 Residuals from Fits to Surveys

Appendix 2: Residuals from Fits to Surveys

SEDAR 51 Gulf of Mexico Gray Snapper Review

Laurence Kell

14 April, 2018

Analysis of residuals is perhaps the most common way to determine a model's goodness-of-fit (Cox and Snell 1968). Residuals are examined for patterns to evaluate whether the model assumptions have been met. Many statistics exist to evaluate the residuals for desirable properties. A non-random pattern of residuals may indicate that some heteroscedasticity is present, or there is some leftover serial correlation (serial correlation in sampling/observation error or model misspecification). Several well-known nonparametric tests for randomness in a time-series include: the runs test, the sign test, the runs up and down test, the Mann-Kendall test, and Bartel's rank test (Gibbons and Chakraborti 1992).

Departures from the assumption that the index is proportional to the stock can also be seen by plotting the residuals by time (**Figure 1**), there do appear to be patterns in the residuals that may suggest conflicts between the data series. **Figure 2** plots the observed CPUE against the fitted (the blue line is a linear regression fitted to the points and the black line is y=x) as the index is assumed to be proportional to abundance the points should fall either side of the $y = x$ line.

Autocorrelated residuals within indices may be due to year-class effects and allow the importance of factors not included in the standardisation of the CPUE to be identified. Autocorrelation may mean that the estimated parameters are biased, autocorrelation can be checked by plotting the residuals against each other with a lag of 1 (**Figure 3**), if there is no autocorrelation then the fit to the points (blue line) will be horizontal. The error distribution was checked by plotting the observed and the predicted quantiles for a given distribution e.g. for the normal distributuion (**Figure 4**)

The variance of the distribution can be checked by plotting the residuals against the fitted values (**Figure 5**)

Runs plot

If the process of interest shows only random variation, the data points will be randomly distributed around the median. Random meaning that we cannot know if the next data point will fall above or below the median, but that the probability of each event is 50%, and that the data points are independent. Independence means that the position of one data point does not influence the position of the next data point, that is, data are not auto-correlated.

If the process shifts, these conditions are no longer true and patterns of non-random variation may be detected by statistical tests. Non-random variation may present itself in several ways. If the process centre is shifting due to improvement or degradation we may observe unusually long runs of consecutive data points on the same side of the median or that the graph crosses the median unusually few times. The length of the longest run and the number of crossings in a random process are predictable within limits and depend on the total number of data points in the run chart Anhoej (n.d.).

A shift signal is present if any run of consecutive data points on the same side of the median is longer than the prediction limit, round(log2(n) + 3). Data points that fall on the median do not count, they do neither break nor contribute to the run Schilling (2012).

A crossings signal is present if the number of times the graph crosses the median is smaller than the prediction limit, qbinom(0.05, n - 1, 0.5) Chen (2010). n is the number of useful data points, that is, data points that do not fall on the median. The shift and the crossings signals are based on a false positive signal rate around 5% and have proven useful in practice.

Standard Deviation of the Normalised Residuals

A simple diagnostic statistic that allows a check that the residuals for each data set are about the right size is the standard deviation of the normalised residuals, or SDNR, which should have a value near 1.

To calculate the SDNRs first calculate the "normalised" residuals, which which are residuals that have been scaled (and transformed, if necessary) so that the expected distribution is normal (at least approximately) with standard deviation 1, given the error assumptions of the model.

One way is to calculate, for each abundance index, the standard deviation of the normalized (or standardized) residuals divided by the sampling (or assumed) standard deviation (SDNR) (Francis 2011). The SDNR is a measure of the fit to the data that is independent of the number of data points. A relatively good model fit will be characterized by smaller residuals (i.e. close to zero) and a SDNR close to 1. It is also necessary to conduct a visual examination between observed and predicted values to be sure that the fit is good even when SDNR values are not much greater than 1.

Survey

For a CPUE or survey time series the normalised residual for year y is given by

$r = log(U_y/\hat{U}_y)/\sigma$

where U_y and $\hat{U}_y$ are the observed and expected values of CPUE, c_y and

$\sigma = \sqrt{log(1 + c_y^2)}$

where c_y is the assumed CV for U_y.

Compostion data

For composition data, let P_{iy} be the observed proportion of fish in the i^{th} length (or age) bin in year y, with length (or age) x_i. Then the normalised residual for year y is

$r_i = (\mu_y - \hat{\mu}_y)/se(\hat{\mu}_y)$ where μ_y is the observed mean length (or age), calculated as

$\mu_y = \Sigma_i x_i P_{i,y}$ and $\hat{\mu}_y$ is the corresponding expected mean

$se(\mu_y) = \sqrt{[\Sigma_i (x_i - \mu_y)^2 P_{iy}]/N_y}$

and N_y is the assumed sample size.

If the SDNR are markedly greater than 1, this indicates that data weights are much too high. The next step is to calculate correction factors for each data set, which are estimates of how we much we might need to change the initial data weights to obtain SDNRs near 1. The initial weights can be divided by the CPUE CVs.

Figures

Residual PLots

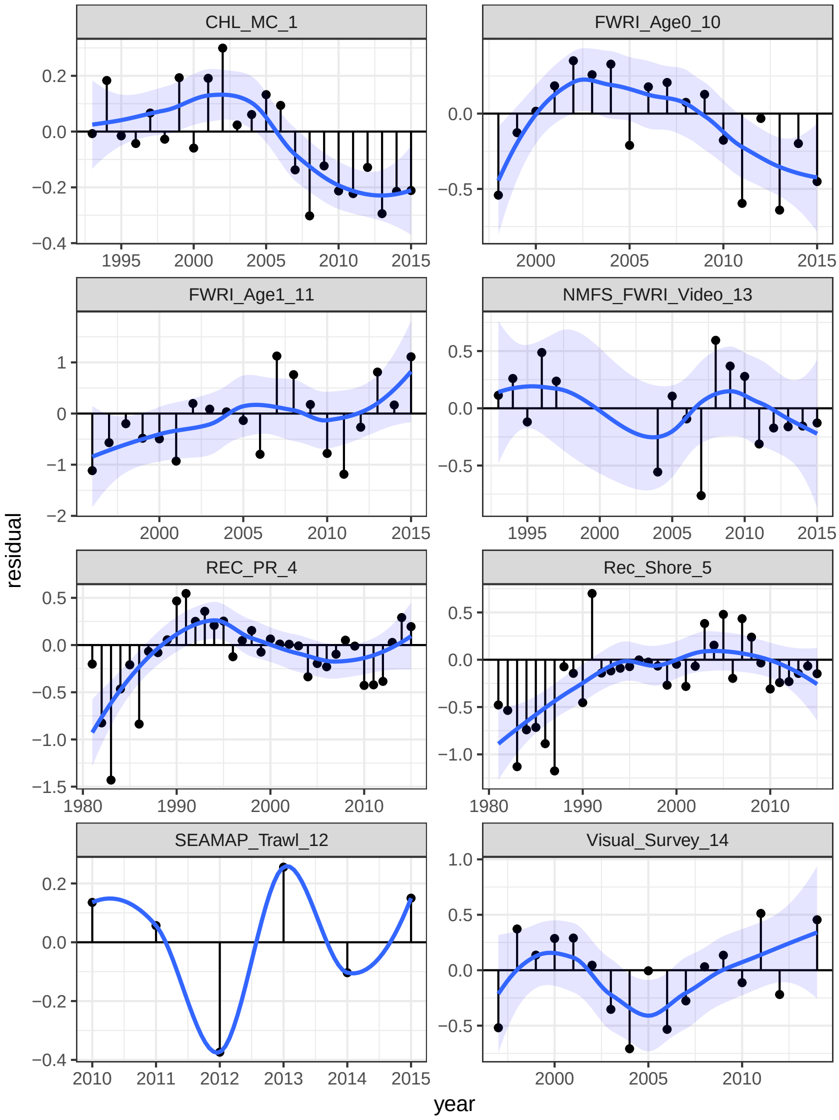


Figure 1 Residuals by year, with lowess smoother and SEs.

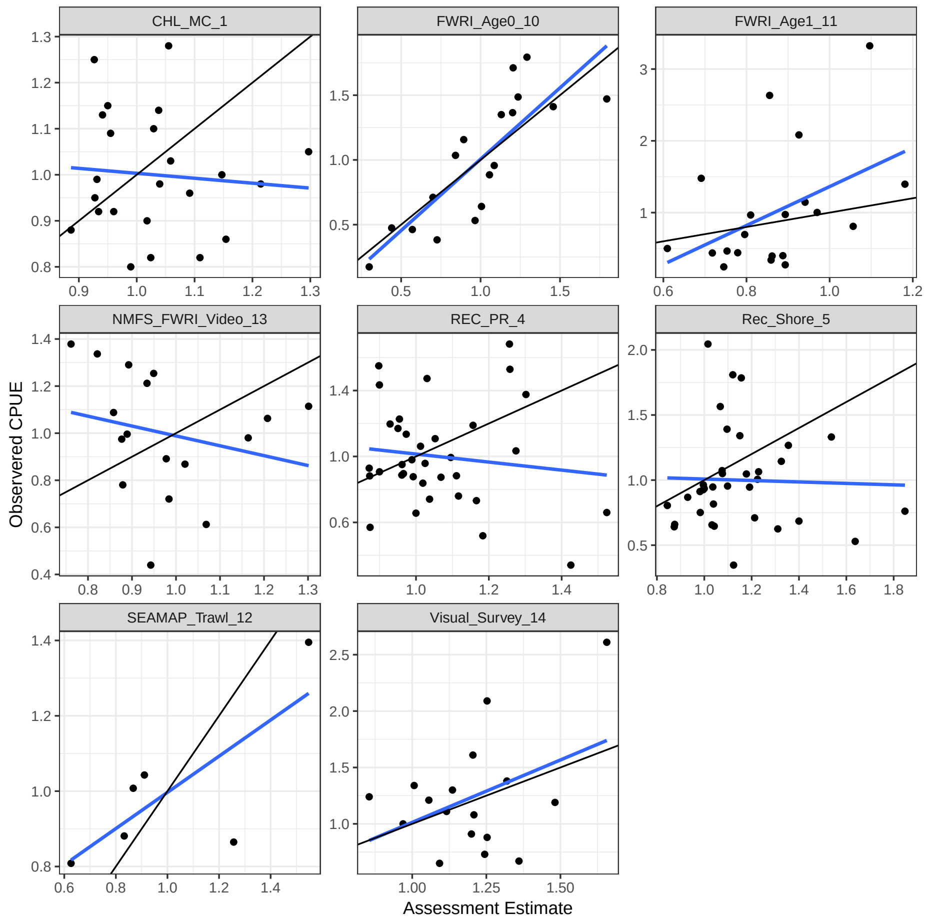


Figure 2 Observed CPUE verses fitted, blue line is a linear regression fitted to the points and the black line is y=x.

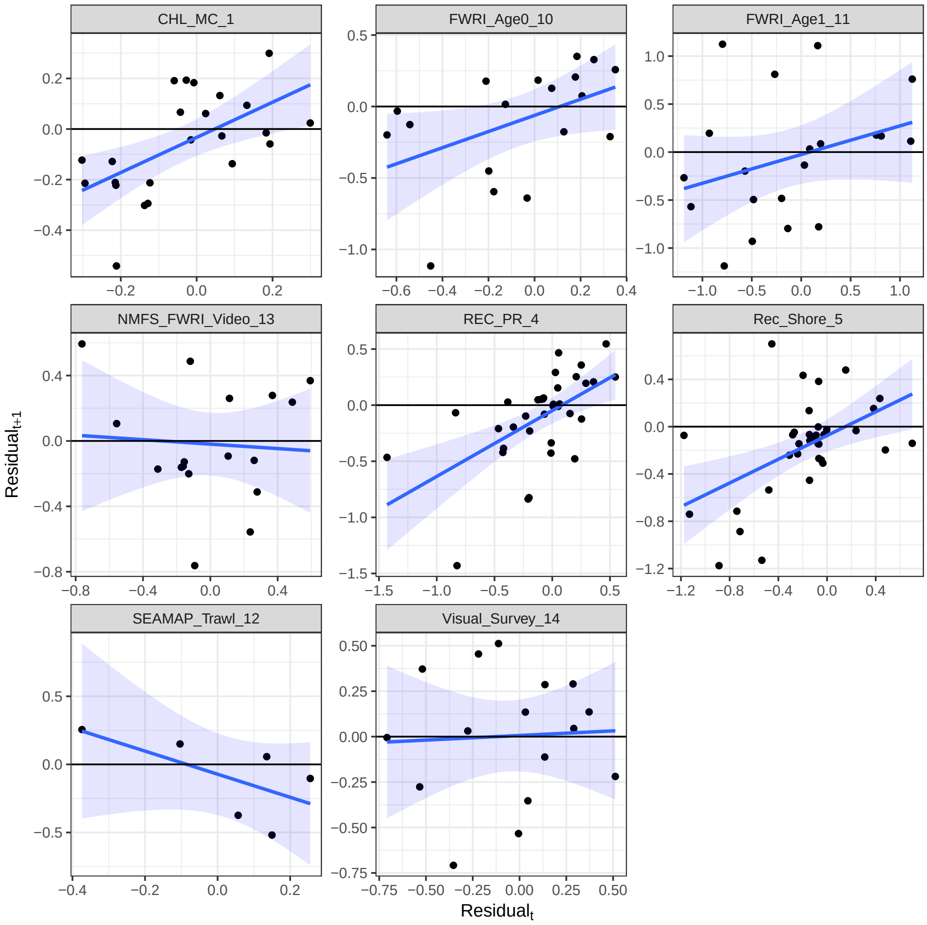


Figure 3 Plot of autocorrelation, i.e. $residual_{t+1}$ verses $residual_t$.

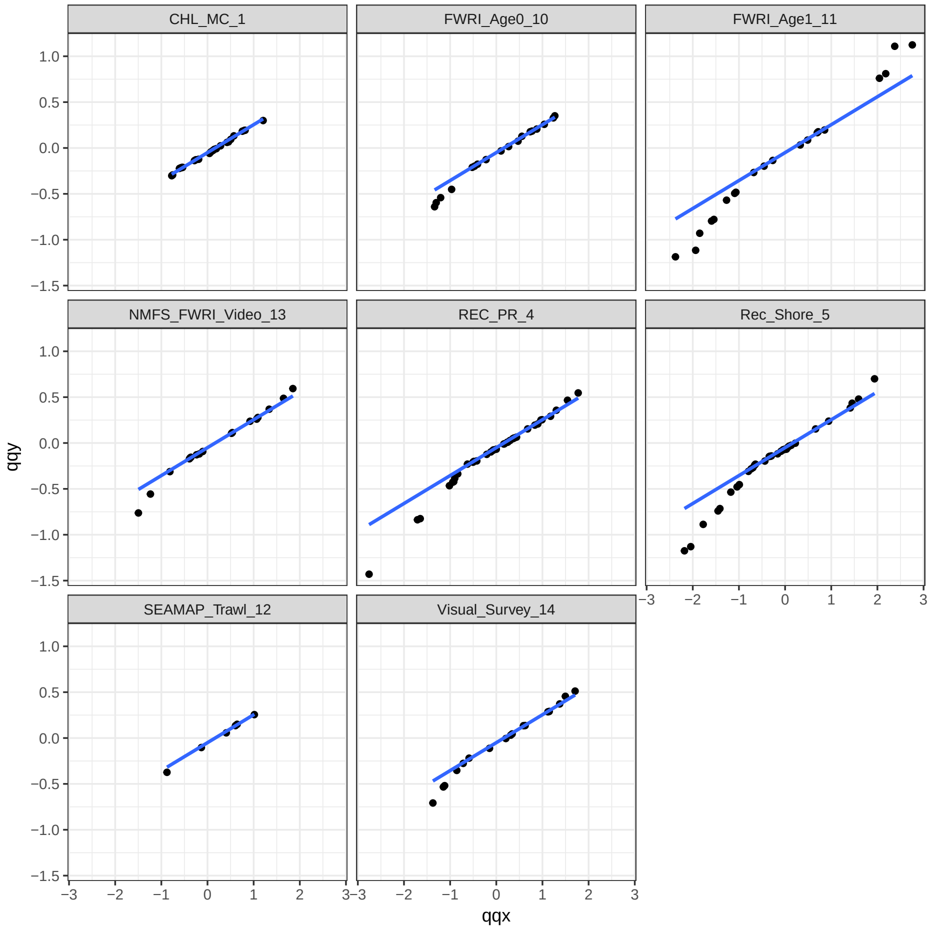


Figure 4 Quantile-quantile plot to compare residual distribution with the normal distribution.

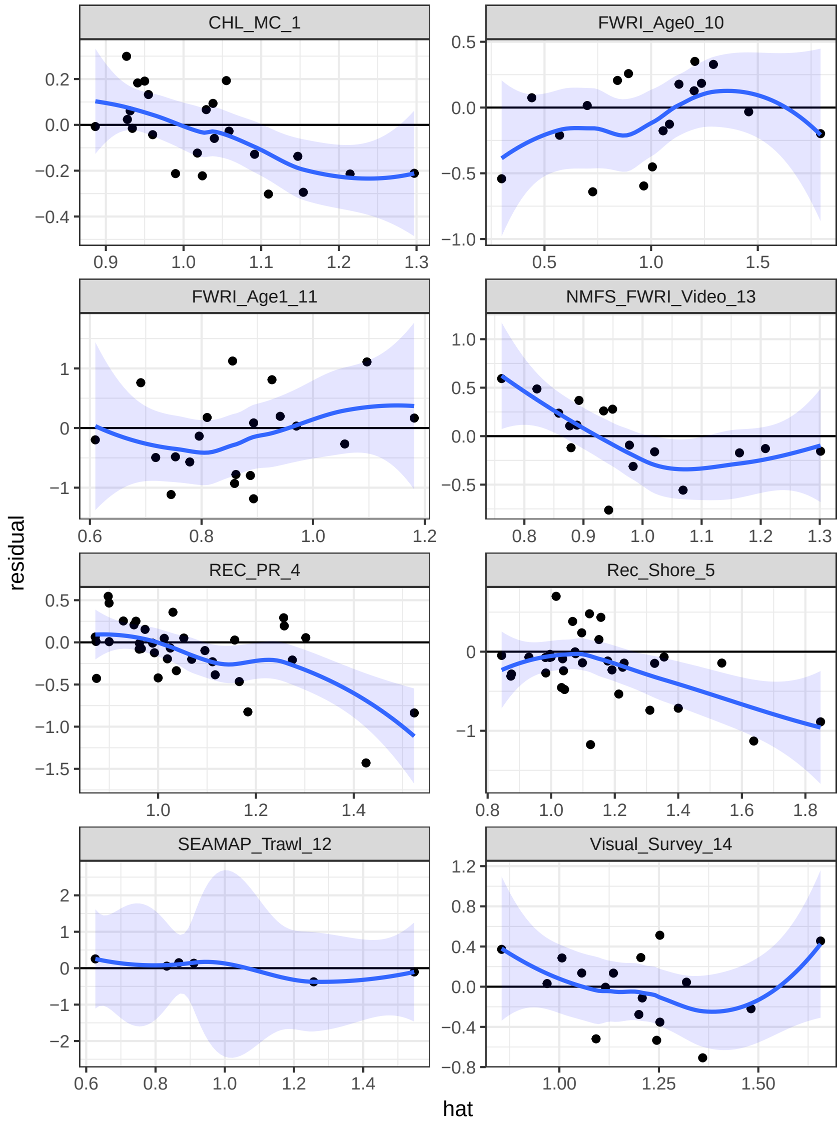


Figure 5 Plot of residuals against fitted value, to check variance relationship.

Runs plots

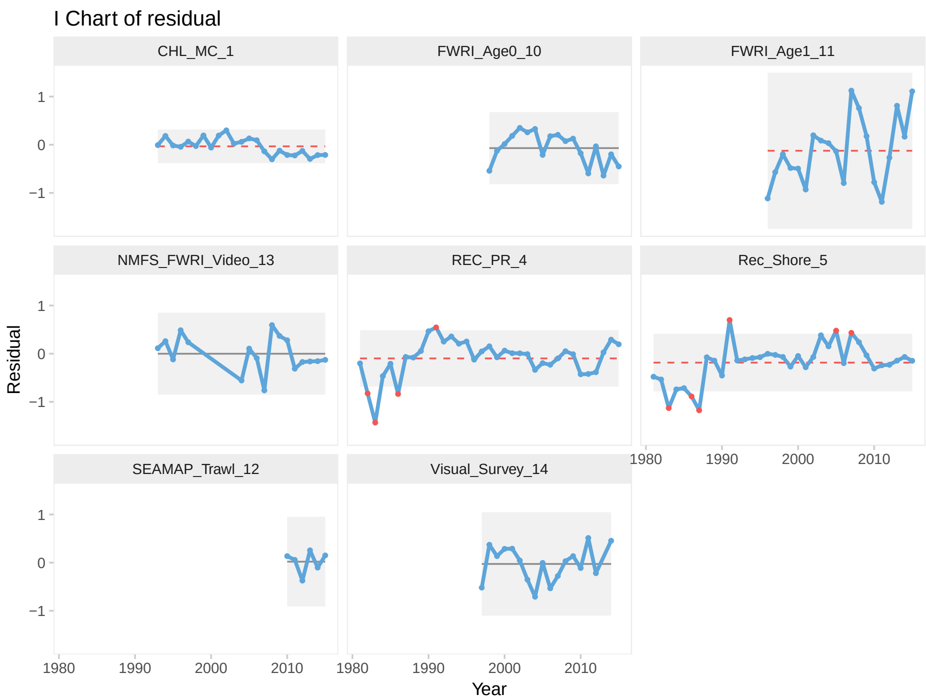


Figure 6 Runs chart showing the residuals by year; red points indicate points that violate the 3 sigma rule, and the red dashed line indicates unusually long runs or unusually few crossings.

```
               Name    V1
1          CHL_MC_1 0.953
2      FWRI_Age0_10 1.613
3      FWRI_Age1_11 6.535
4 NMFS_FWRI_Video_13 1.813
5          REC_PR_4 1.973
6        Rec_Shore_5 2.779
7    SEAMAP_Trawl_12 1.065
8   Visual_Survey_14 2.503
```

Software

All analysis was conducted using R and FLR and the `diags` package which provides a set of common methods for reading these data into R, plotting and summarising them.(http://www.flr-project.org/)

- R version 3.4.1 (2017-06-30)
- FLCore: 2.6.6.9006
- diags:
- ggplotFL:

- diags:
- r4ss:

Author information

Laurence KELL. laurie@kell.es

References

Anhoej, Jacob. n.d. "Run Charts with R." *Analysis* 21 (30): 8–10.

Chen, Zhenmin. 2010. "A Note on the Runs Test." *Model Assisted Statistics and Applications* 5 (2). IOS Press: 73–77.

Cox, David R, and E Joyce Snell. 1968. "A General Definition of Residuals." *Journal of the Royal Statistical Society. Series B (Methodological).* JSTOR, 248–75.

Francis, RIC Chris. 2011. "Data Weighting in Statistical Fisheries Stock Assessment Models." *Canadian Journal of Fisheries and Aquatic Sciences* 68 (6). NRC Research Press: 1124–38.

Gibbons, Jean D, and S Chakraborti. 1992. "Response to Zimmerman (1992)." *The Journal of Experimental Education* 60 (4). Taylor & Francis: 365–66.

Schilling, Mark F. 2012. "The Surprising Predictability of Long Runs." *Mathematics Magazine* 85 (2). Mathematical Association of America: 141–49.

Appendix 3: ALK and Cohort Strengths

SEDAR 51 Gulf of Mexico Gray Snapper Review

Laurence Kell

14 April, 2018

This document looks at the age length key (ALK) and estimates of cohort strength as estimated by the base case.

Figure 1 shows the ALK in the form of a growth curve, it can be seen that after about age 8 there is considerable overlap with lengths, The red bars give the mean size-at-age.

The log stock numbers-at-age are plotted in **figure 2**, these show stong cohort effects in the 1980s cohorts, The fact that the ALK shows that there is considerable overlap with older ages implies that the year-class information must be contained in the recruitment signal.

Figures

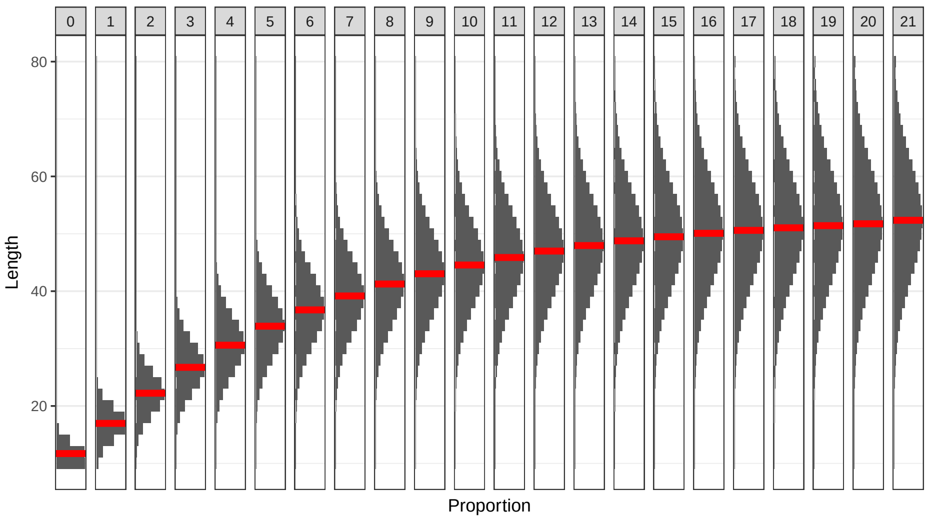


Figure 1 Length distributions by age from the age length key generated by SS3.

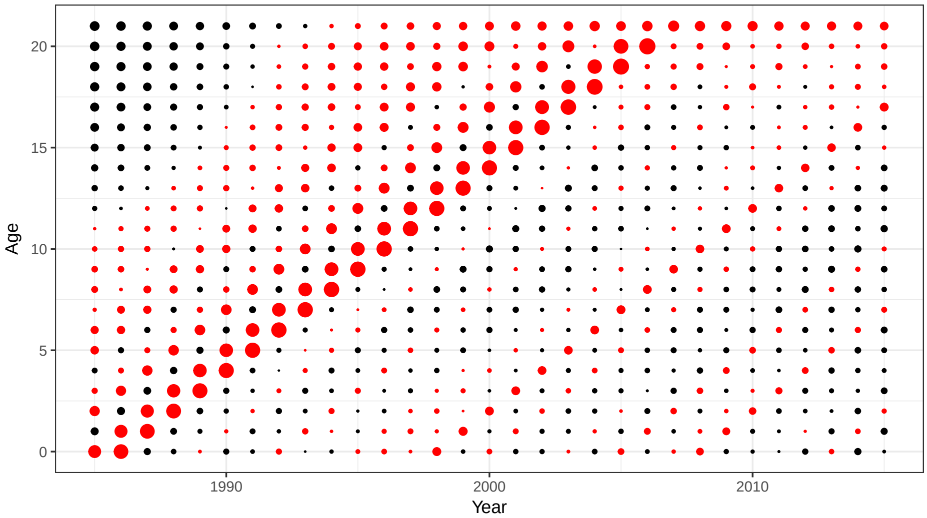


Figure 2 Log numbers-at-age scaled (by mean across age) to show year-class trends (red negative and black positive).

Software

All analysis was conducted using R and FLR and the `diags` package which provides a set of common methods for reading these data into R, plotting and summarising them.(http://www.flr-project.org/)

- R version 3.4.1 (2017-06-30)
- FLCore: 2.6.6.9006
- diags:
- ggplotFL:
- diags:
- r4ss:
- **Compiled**: Sat Apr 14 08:36:19 2018

Author information

Laurence KELL. laurie@kell.es

References

Appendix 4: Fits to length compostion data

SEDAR 51 Gulf of Mexico Gray Snapper Review

Laurence Kell

14 April, 2018

The mean lengths from the fits to the length compositions are plotting (line) and comparing to the observations with bars indicating the 95% CI.

Figures

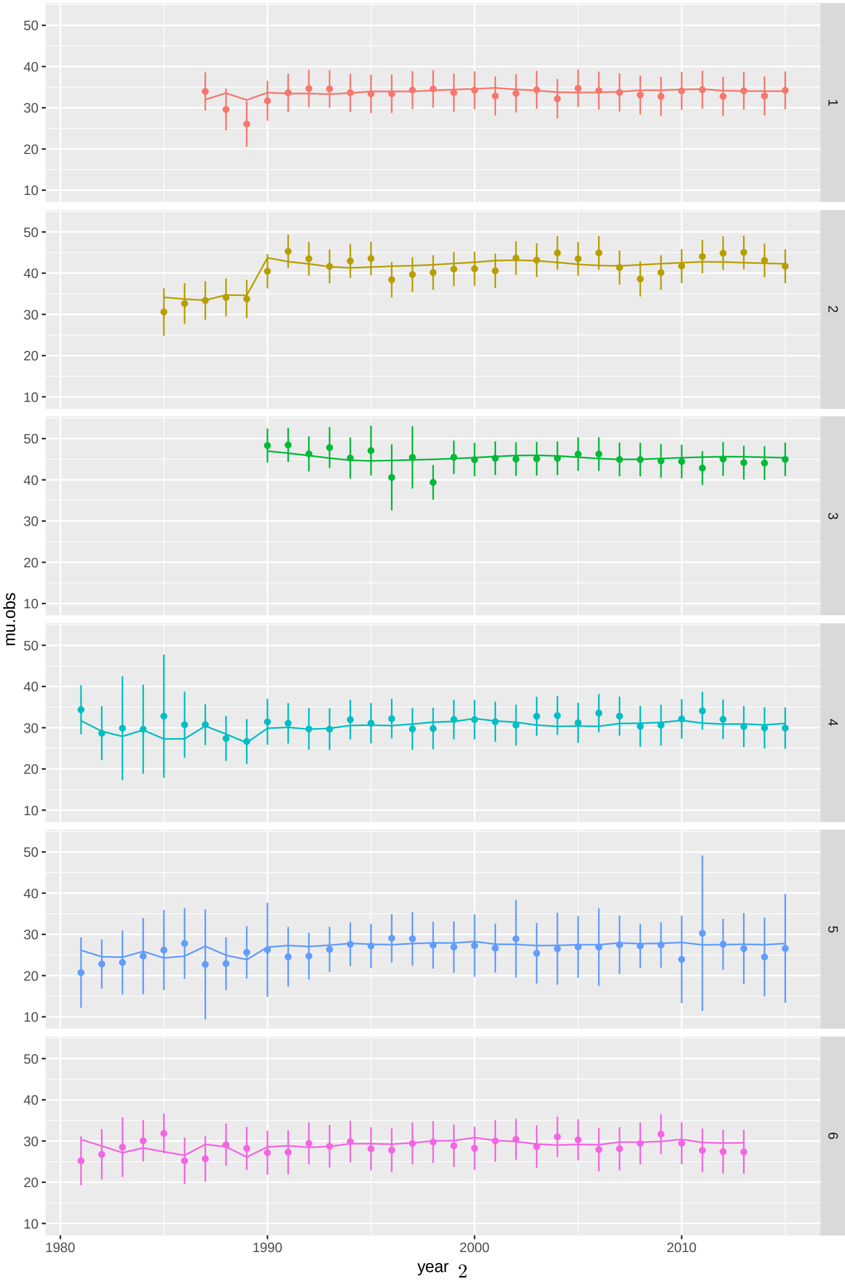


Figure 1 Mean lengths by indices, lines are estimates and points with bars are the observations with the 95% CIs.

Software

All analysis was conducted using R and FLR and the `diags` package which provides a set of common methods for reading these data into R, plotting and summarising them.(http://www.flr-project.org/)

- R version 3.4.1 (2017-06-30)
- FLCore: 2.6.6.9006
- diags:
- ggplotFL:
- diags:
- r4ss:

Author information

Laurence KELL. laurie@kell.es

References

7.45 Stock and recruitment

Appendix 5: Stock and recruitment

SEDAR 51 Gulf of Mexico Gray Snapper Review

Laurence Kell

14 April, 2018

The stock recruitment relationship is explored in this document, **Figure 1** shows the time series of the estimates of recruitment and SSB, and **Figure 2** the Beverton and Holt stock recruitment relationship estimated by SS3.

Figure 3 shows the fits to the stock recruitment relationship to the estimates of SSB and recruitment with goodness of fit diagnostics. The difference with the fit in **Figure 2** is since the SS3 estimates are derived from a model where the deviates are also estimated.

Even if a SRR exists it may be difficult to estimate it from stock assessment data sets a cross correlation analysis examines whether it is as likely that SSB is driven by recruitment as the existence of a stock recruitment relationship. **Figure 4** shows the cross correlations between recruitment and SSB. If there was a strong SRR then high recruitments follow high SSBs i.e. there would be a correlation with no lag, while if SSB was driven by strong recruitments, i.e. strong year-classes, then there would be a negative lag. Although negative lags are strong the sign varies and so the analysis is inconclusive. Possibly as a result of the early period when not variation was see due to lack of length compositions and CPUE in the early period.

Figure 5 shows the tock recruitment relationship fitted to estimates of SSB and recruitment from 1980. The discontinuity is due to a negative estimate of a, i.e. there is no infomation about a reduction of recruitment at small stock size.

Finally in **Figure 6** evidence of regimes was evaluated in the time series of recruits, boxes are regimes estimated by STARS algorithm (Rodionov 2004). Although there appears to be a recent increase in recruitment and higher variance in the 1980s bo regime was estimated

Figures

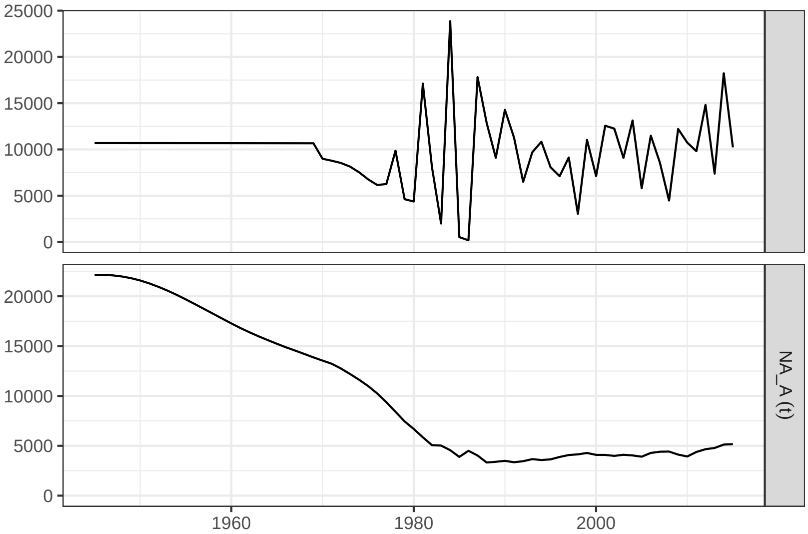


Figure 1 Time series of estimates of recruitment and SSB.

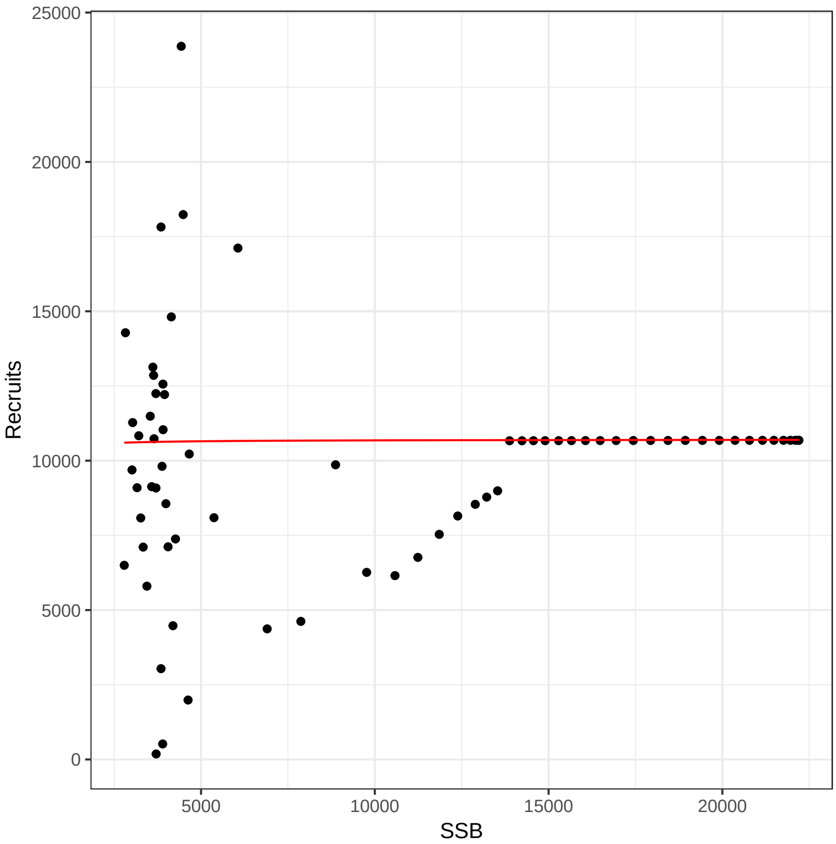


Figure 2 Stock recruitment relationship estimated by SS3.

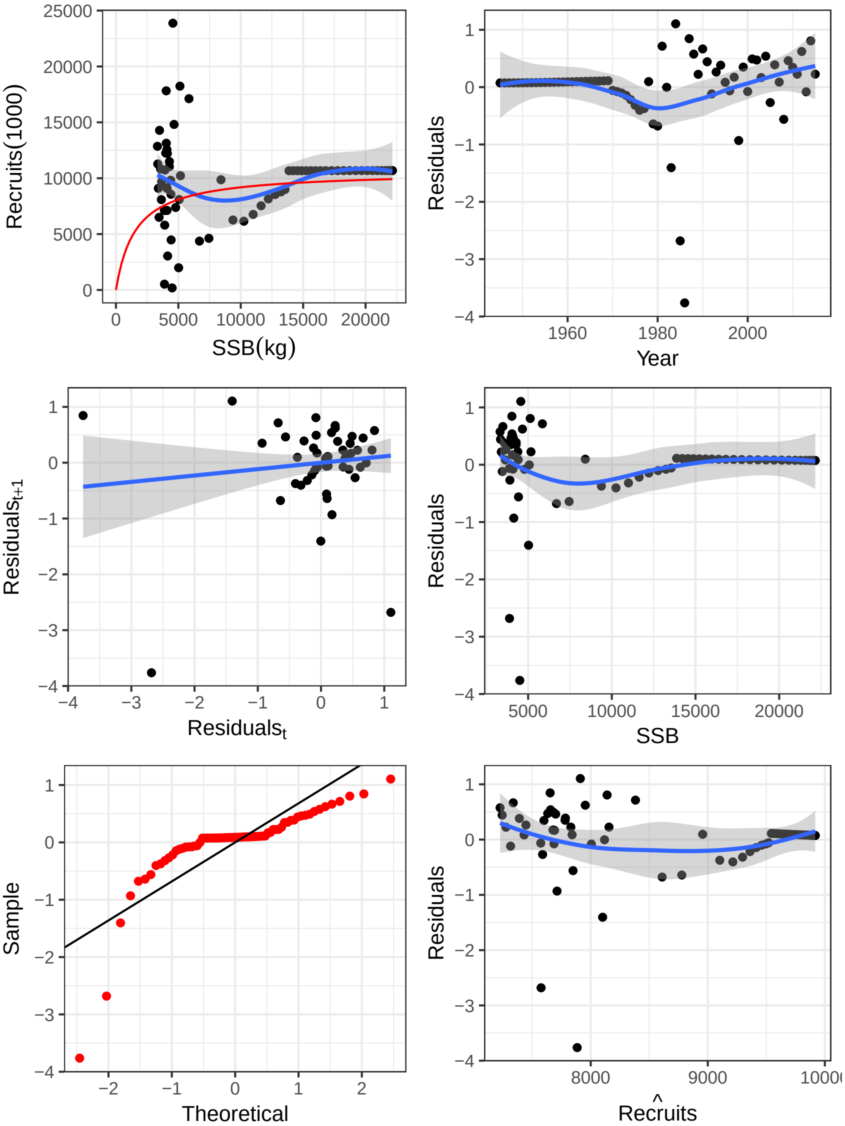


Figure 3 Stock recruitment relationship fitted to estimates of SSB and recruitment.

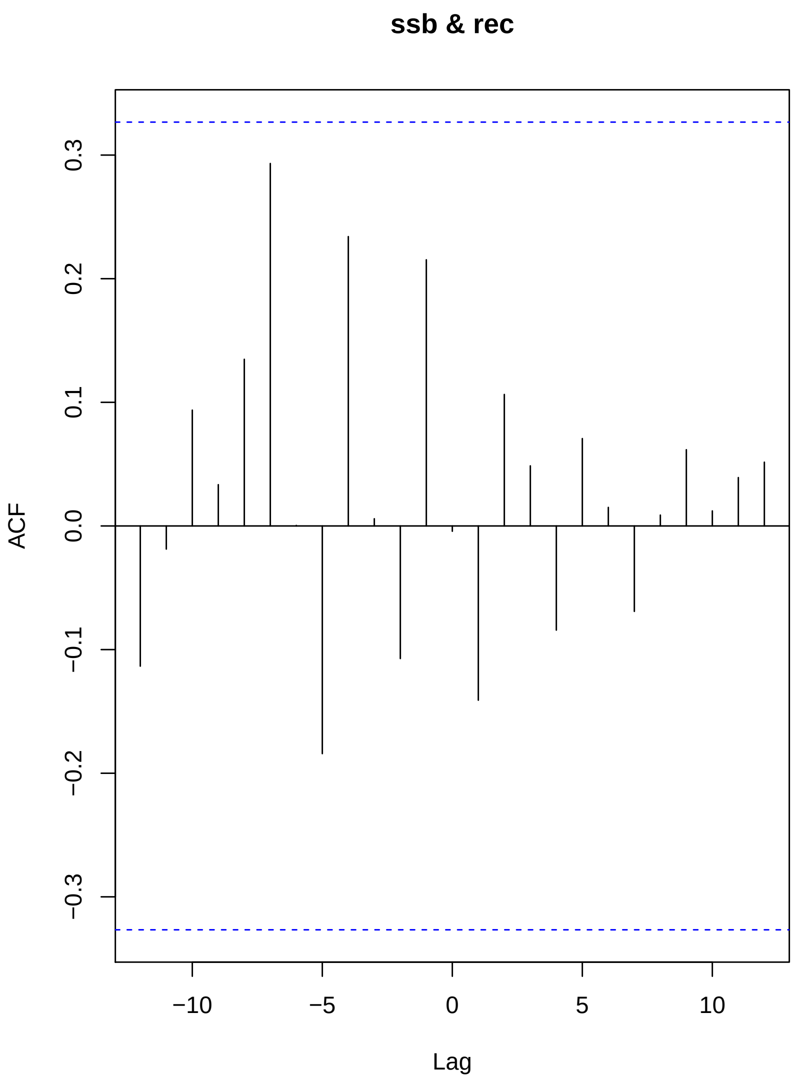


Figure 4 Cross correlations between recruitment and SSB.

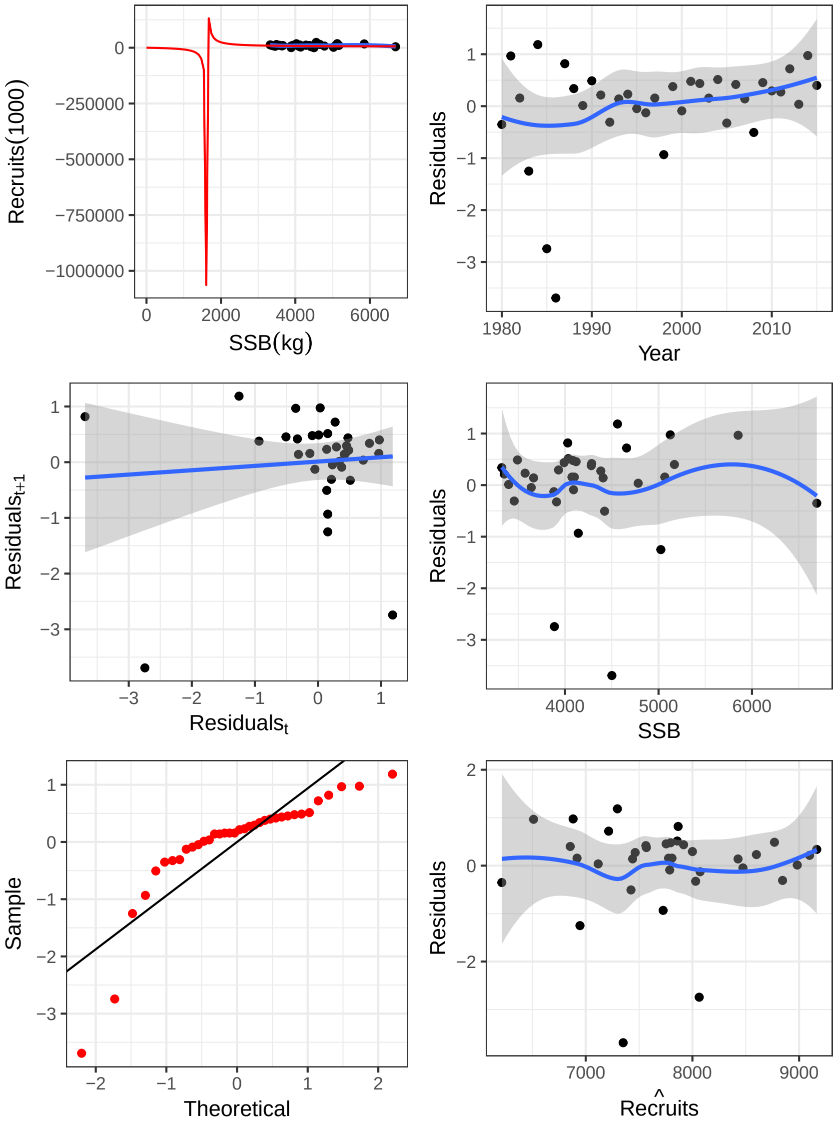


Figure 5
Stock recruiment relationship fitted to estimates of SSB and recruitment from 1980.

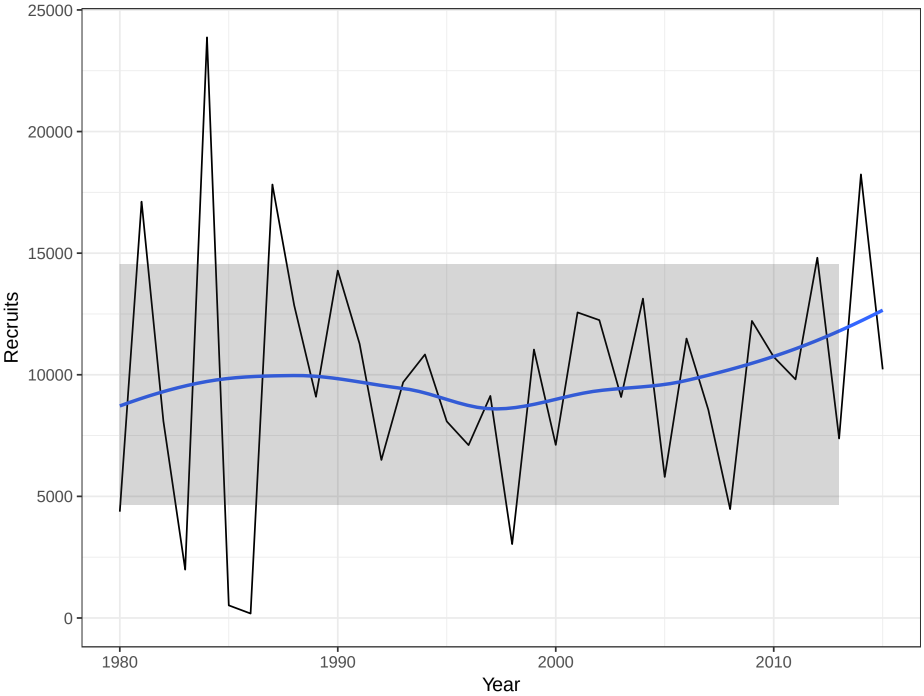


Figure 6 Time series of recruits, boxes are regimes estimated by STARS algorithm.

Software

All analysis was conducted using R and FLR and the `diags` package which provides a set of common methods for reading these data into R, plotting and summarising them.(http://www.flr-project.org/)

- R version 3.4.1 (2017-06-30)
- FLCore: 2.6.6.9006
- diags:
- ggplotFL:
- diags:
- r4ss:

Author information

Laurence KELL. laurie@kell.es

References

Rodionov, Sergei N. 2004. "A Sequential Algorithm for Testing Climate Regime Shifts." *Geophysical Research Letters* 31 (9). Wiley Online Library.