NCDMF Stock Assessment Program
External Peer Review

Assessment Information

Assessment Species:	Southern flounder (*Paralichthys lethostigma*)
Stock Assessment Report:	Stock Assessment of Southern Flounder, *Paralichthys lethostigma*, in North Carolina Waters—2014
Date Sent:	November 10, 2014

Dear Reviewer-

Thank you for agreeing to review the 2014 stock assessment of the North Carolina southern flounder stock. The purpose of the external peer review process is to ensure that the assessment and results presented are scientifically sound and that decision makers are provided adequate advice. Peer reviewers are asked to address the terms of reference in the terms of reference report that follows. Please be as specific as possible in recording your comments and suggestions for revision and improvement. Any additional suggestions to improve the stock assessment are appreciated. Reviewers are also welcome to make comments directly in the assessment report using the Track Changes feature in Microsoft Word.

Please return this form, the terms of reference report, and any additional comments to laura.lee@ncdenr.gov. We would like to have your review by December 5, 2014. A copy of the final report will be provided after it has been presented to the North Carolina Marine Fisheries Commission.

Thank you,

Laura M. Lee
Senior Stock Assessment Scientist
North Carolina Division of Marine Fisheries

TERMS OF REFERENCE REPORT FOR EXTERNAL PEER REVIEW

Reviewer Information

Reviewer Name:	Dr. Erik H. Williams
Business Mailing Address:	NOAA/NMFS, 101 Pivers Island Road, Beaufort, NC 28516
Business E-Mail:	Erik.Williams@noaa.gov
Business Phone:	252-728-8603

1) Evaluate the thoroughness of data evaluation and presentation including:
   a) Justification for inclusion or elimination of available data sources

      The report seems to focus on data that were included, with not much being considered, but ultimately rejected. It is not clear how the age data are being used in the assessment model. It seems this whole document relies on a person with some experience with Stock Synthesis. The document says the age data are entered into the model, yet the growth parameters are fixed and the selectivity curves are estimated as functions of length, not age. This leaves me wondering if the age data is being used for anything in the assessment model. Are there annual age composition fits? None are shown in the document. This is a major weakness of this document and potentially the stock assessment.

      Generally the justification for including data seems appropriate. Aside from the concerns mentioned above about the age data, there are some concerns about the index data being used (see below).

   b) Consideration of survey and data strengths and weaknesses (e.g., temporal and spatial scale, gear selectivities, sample size)

      Stock Definition:

      The documentation for the stock definition is clear. The limited tagging data are presented and used in an appropriate manner. The issue remains that this is clearly an open population being modeled in this stock assessment.

      Movement and Migration:

      Section 1.2.2 in the report should include more detail. The 15% emigration rate is mentioned later in the report, but not in this section, where it clearly belongs. This is a critical issue for this stock assessment and I would expect the write-up to reflect that with much more detail and discussion of available data, alternate hypotheses, anecdotal information, etc.

      Age, Size, Growth:

      The age and length data for this species seems fairly extensive based on the sample size tables presented in the document, yet it is completely unclear how all this data is being used or not used in the assessment. There is extensive discussion about stratifying the length data into six month periods, but only one overall growth curve is presented for

each sex. Were six month growth curves fit? The model uses two seasons, but the same growth curve for each season? It appears that the growth curves were fixed in the model.

The sampling design, or lack of sampling design for collecting length and age samples could use some more detail. Are the age data useless because of the biased sampling design? I am not clear on this. Are the lengths biased in some way? How exactly were the fishery specific length composition data put together? Are there fishery specific age composition data available? I realize later in the report that there is a statement about the age sampling being non-random. But, were there any attempts to correct the data for use in the assessment?

Maturity:

The maturity data seems fairly well documented and one of the few sections of the report that was actually clear and concise.

Mortality:

The choice of natural mortality values for an assessment is usually one of the most important choices made. In this report a single paragraph is devoted to this topic. That is woefully inadequate. The choice of M estimators needs to be discussed and justified. It looks as if the analyst just picked a few formulae off the shelf and then averaged them all, without regard to their utility for this species. For instance, the Ralston estimator is almost never used and should be discarded. The Jensen estimator is just one estimator and should not appear twice in the averaging. Basically these M estimators should be cast into categories of estimators, such as max age based (Hoenig), life-history invariant (Jensen, Charnov), and size based (Lorenzen). The strengths and weaknesses of each should be discussed, with particular note to the species at hand. Which methods may have more utility for Southern Flounder? Why?

Habitat:

It seems clear from the document that the knowledge about movement and habitat is limiting, as it can be for many fish species. However, it would have been nice to see a little more data from GA and SC to see if anything from those regions can be gleaned for this assessment. Perhaps that was done and it is not clear to me. There is mention of tagging studies from GA and SC, but not much detail about what is known in those areas about Southern Flounder.

Descriptions of Fisheries:

This is very clear and understandable. It seems that getting better discard estimates should be a high priority for research.

Fishery-Independent Data:

It would have been nice to see a summary table comparing the Program 120, 135, 195, and 915 data side by side. Perhaps indicating months of data collection, amount of area covered, number of stations, etc. In any event, the documentation seems clear enough to understand the data, but what is unclear is why the choices were made for inclusion or exclusion of the data, and why the data was subset in the ways that it was for Southern Flounder. For example, Program 120 seems inferior to Program 195, which seems to cover more time and space, yet two indices, carrying equal weight in the model are used

from these programs. Why? And, May and June months were used for Program 120 data, yet the Program 195 data was limited to September. They are both measuring age-0 Southern Flounder. Why the discrepancy in months used?

The Program 135 and 915 data are both measuring adult Southern Flounder, but I see no length composition data. This leaves me wondering how this data is being treated in the model. Maybe if I was able to read Stock Synthesis input files I might be able to find it, but if that was an expectation for this review, then you called upon the wrong reviewer. More discussion of the indices is below.

c) Calculation and standardization of indices and other statistics

There are several issues with the indices that were produced for this assessment that warrant further consideration or explanation. The document is limiting in its description of why index data were subset and ultimately selected. For instance, there is no clear linkage between the months chosen for modeling index data and the months during which one would expect to see Southern Flounder. There is discussion of taking the months with the highest catch rates, but how high and what percentage of the catch is excluded in the months that were dropped? This point obviously does not apply to all the index data, because some of the fishery-independent sampling was already limited by design.

The potential biases and uncertainties sections of the index reports are incomplete and border on the useless. Saying that a survey is not designed for the species in question and not explaining how that might affect the relationship between survey abundance and true abundance is a dereliction of duty. I would guess that 95% of surveys used in fisheries are never designed for the species index that is being calculated. What is more important is to discuss the spatial and temporal overlap of the species and survey. This is not done very effectively in this assessment.

The justification for choice of indices to be put into the model is not clear. These indices should be compared in two ways, (1) relative to each other and (2) based on the properties of how well they capture abundance. Having more than one index that measures the same thing should not be put into the model. The model does not have a good means to sort out the differences, especially when no age data is being used. I do not understand the reason for including both program 120 and 195 data since they both are measuring age-0 Southern Flounder. Either combine them into a single index or select the better one. Because the length and age composition data is not presented for any of the indices, I am not sure whether the same goes for program 135 and 915 index data. If they have the same selectivity (or even roughly the same), then combine them or select the best.

The documentation and diagnostics for the indices is too little for me to offer any advice about how reflective they are of abundance. The Mann-Kendall analysis in section 2.3 is pointless. The pairwise correlation analysis is good, but those results showing very little correlation are the very results you are looking for to question the choice of indices for inclusion in the model, yet none of that is done in this report. Lack of correlation among your indices (assuming they are lagged appropriately) should be a big red flag and warrants further investigation, discussion, and justification than what is in this report.

2) Evaluate the adequacy, appropriateness, and application of data used in the assessment.

A stock assessment relies on the accuracy of each input data source. In order to have a useful model you must at a bare minimum have an understanding of the total removals from the stock. In the case of Southern Flounder, there is a good set of landings estimates, but the discards are highly uncertain, or even ignored in the case of shrimp trawl bycatch. Nonetheless, that is a common problem for many stock assessments. The best way to address this is through sensitivity runs and uncertainty estimates.

The next critical piece of information for a size/age-structured model is size and age structured data. In this case length composition data is being used and it remains unclear to this reviewer what is being done with the age data. In any event this data needs to be plotted and visually examined to determine if (1) year class strengths/weaknesses can be observed moving through the data, (2) if shifts in the smaller ages/sizes correspond to changes in minimum size regulations, and (3) if there are notable shifts in the maximum size/age of fish over time. None of this appears to have been done in this assessment. It should have been noted upon examination of the length composition data that there is no sign of year classes and that there seems to be little response in the data to changes in the minimum size limit. These are concerns when going forward with a size/age-structured model like Stock Synthesis. It may have suggested that a surplus-production model was more appropriate, but I caveat that statement with the issue of not being able to adequately evaluate the age data. The age data may have been more valuable in this case. It is often the case that age data contains far more important population dynamic information in it compared to length data, especially when the variation in size at age is as great as it seems to be for Southern Flounder.

Of course a very critical piece of information for any assessment, especially one that seeks to provide long-term, equilibrium benchmarks is abundance data, usually in the form of relative CPUE indices. In this case there appears to be both juvenile (age-0) and adult (age range not documented) indices. Unfortunately there are two of each and they appear to conflict. This is a problem for any stock assessment and usually needs to be resolved by the analyst by either combining the data or eliminating an index. Forcing the model to choose between the indices just adds noise and potential bias to the model and results.

3) Evaluate the adequacy, appropriateness, and application of method(s) used to assess the stocks.

The stock synthesis assessment model is a powerful model that must be applied with care and full understanding of the underlying processes. For the set of data, which includes landings, discards, size composition, age composition (???), and indices, this model is appropriate. Unfortunately the way this report is written, determining more precisely how various settings were structured for this model is near impossible without being able to read stock synthesis input files, which I cannot read. Some details that I am left wondering about include: how are the discard selectivity curves (retention curves) being estimated? What selectivity function is being used for all the curves? What selectivity is being used for the adult indices? How is the age data being fit in the model? What are the likelihood components being estimated? What are the values of those likelihood components? What are the sample size inputs for the multinomial components? What are the CV values being used in the lognormal

likelihood components? There is a lot left unknown to me because of poor documentation and the inability to interact with the analyst during a desk review.

If I could read stock synthesis input files I might be able to find some of this information, but I cannot do that.

The choice of dome-shaped selectivity curves is a tricky one. In all assessment models like this one there is a confounding between the degree of doming in the curve (the descending limb of the selectivity curve) and the fishing mortality rate. I would have liked to see sensitivity runs with alternate hypotheses about selectivity shape and form.

The likelihood components are very poorly documented in this report, more specifically the error levels assumed for each data set. What are the multinomial n's and the index and landings CV's? The model is set with all likelihood multipliers at 1.0. This assumes that all error levels are specified correctly, which is almost never the case. Recent trends in the literature (see Francis 2011) suggest that likelihoods should be re-weighted such that residuals conform to assumptions of N(0,1). At a minimum some other likelihood weighting scheme should have been explored. The typical pattern seen in these models is that the multinomial composition components are over-weighted, which by the looks of the index fits, seems to be the case in this model. But, again the poor documentation makes this insight difficult.

Uncertainty is characterized through sensitivity runs. This is inadequate, especially if the range of sensitivity values is not directly tied to the amount of uncertainty in the input data. What is the uncertainty in the F level that corresponds to $F_{35\%}$? This is probably the most critical output, yet there is no specified level of uncertainty. Sensitivity analyses are really meant to determine behavior of the model to different assumptions, not characterize uncertainty. The author(s) should look into delta-method, bootstrap, Monte Carlo, and MCMC approaches to characterizing uncertainty.

4) Evaluate the methods used to estimate stock status determination criteria. Evaluate the adequacy and appropriateness of recommended stock status determination criteria.

   Because of the open nature of this stock assessment, equilibrium based stock status criteria are not appropriate. The SPR rates are useful, but the uncertainty about selectivity, natural mortality, and even maturity should be factored into this calculation.

5) Does the stock assessment provide a valid basis for management for at least the next five years given the available data and current knowledge of the species stock dynamics and fisheries?

   (•) Yes

   ( ) No

   Comment on response.

   Although this stock assessment has its issues, as almost any stock assessment does, it is important to realize that like most stock assessments it represents a compilation of all available data for the fishery and species. Because it is a compilation of all information, it is ultimately useful for management. The question and concern is what type of management

can be justified with this assessment? As a scientist and reviewer of this assessment, I do not want to offer management advice, as it clearly is not my place. Instead what I will try to do is highlight some of the important results and data that reflect on the population dynamics/sustainability of the stock, while keeping an eye on the associated level of uncertainty with each. The best way to think of this is as indicators of stock condition, with some being clearer than others.

First, it should be clear that the first major shortcoming of this stock assessment is the modeling of an open population, which means there is uncertainty in the feedback between spawning stock, recruitment, and ultimately fishing mortality rates. The assessment assumes there is no stock-recruit relationship, which means that it assumes recruitment is essentially constant (with some environmental noise around it). Of course that is never the case because ultimately a stock must have some level of spawning to produce some recruits, they don't just appear out of thin air. So, we know that at some low levels of spawning stock, recruitment must become impaired. The question is how low? Because of the open population we have of Southern Flounder in NC, another layer of uncertainty is added to the whole system. Certainly this type of situation is not conducive to estimating long term sustainability benchmarks, and managers should take note of that.

In the absence of long term sustainable benchmarks there are some things that managers can continue to keep an eye on for sustainability issues. For instance what are recent trends in the abundance indices, what are the recent trends in recruitment and fishing mortality, what are the recent trends in age/length composition? Downward trends in abundance indices, downward trends in recruitment, increases in fishing mortality, and truncation of the older/bigger fish from the age/length composition should trigger concern from managers. Other changes in the data should be noted as well, like shifts in maturity or size-at-age.

What I am suggesting here is that this assessment is not going to help with any traditional benchmarks because the open population renders them useless. But, the recent trends in the data and assessment model output are useful. In evaluating these trends it is important to try and understand the root cause if possible. For instance, if recruitment is declining, we must ask why? Is there increased fishing from the other areas? Is there increased fishing in NC? Is there a change in the environment? Find the most credible explanation and this will help in deciding what can be controlled and what is uncontrollable. But recognize that controllable or uncontrollable signals may still require action on the part of managers.

What do I see in these results? Fishing mortality is showing an overall long term decline, with a recent upturn in the last few years. This is largely a reflection of the landings trend, which means any direct actions on landings levels should result in changes in fishing mortality. SPR and F are directly inversely related. Recruitment is showing a general long term decline, with a recent upturn. This could be concerning, especially since the most recent estimates of recruitment are the most uncertain. The total biomass trend is essentially flat which suggests things are fairly sustainable, but recognize that the flat trend in biomass came during a period of decreasing landings. If the decreased landings were having really positive results, we should have seen biomass trend upward. Of course recruitment has been trending downward, hence it balances out into a seemingly flat biomass time series.

The length frequency (composition) data is amazingly consistent over time, showing very few shifts or modes due to year classes. This is likely a result of the apparent large variation

in size-at-age in the growth curve data. Unfortunately this also points out a severe limitation in the utility of the length data. It does not reflect any year classes moving through the population and does not show any shifts in structure due to changes in fishing mortality or even due to changes in regulations. I would expect to see a little more shift in the small fish being kept, reflecting the increase in minimum size limits over time. It almost appears that the minimum size limits are simply being ignored.

However, the most unexpected and concerning trend in the data is the apparent shift in the age data toward younger fish. This is during a time of increasing size limits (which seem to be ignored) and stable length compositions. The catch since 2008 seems to be predominated by age-1 fish, with an increasing fraction of age-0 fish. This is of concern from a long term sustainability perspective. The reason for this needs to be explored and management measures should attempt to reverse this trend.

6) Evaluate appropriateness of research recommendations. Suggest additional recommendations warranted, clearly denoting research and monitoring needs that may appreciably improve the reliability of future assessments.

   The research recommendations seem appropriate.

7) Are you aware of any reference material not cited in this report that should be included?

   No references that should have been cited, but plenty of material that should have been considered, applied, and possible cited. Without looking up the exact references, I would direct the author(s) to work on selectivity estimation (Sampson, Legault, and a recent Fish Res issue), adjusting age comps for biased sampling (Chih), likelihood re-weighting (Francis), and index selection (Hilborn, Walters).

8) Would you be willing to act as an external peer reviewer for a future NCDMF stock assessment?

   (•) Yes

   ( ) No

9) Do you have any additional comments?

   Because the review of this stock assessment is being conducted as a desk review, which does not allow for interaction between the reviewers and stock assessment analyst, it is incumbent that the documentation be very thorough and inclusive. I find that not to be the case for this assessment. It is a major shortcoming of this assessment and severely impinges my ability to adequately review it. Half the document is attachments of Stock Synthesis (SS) files. By themselves they serve no utility for non-SS users. Therefore I am left to rely on the Tables and Figures alone, which are very incomplete. For future reviews, consider a face-to-face review, or have a much more thorough set of documents. I appreciate the need to keep a stock assessment report that is readable and understandable, but there are many supplemental things that are needed for a good technical review.

   For instance, almost all input data should appear in Table form, with some appearing in Figure form as well. Consider actually including a spreadsheet of the assessment input data and even the files and software for running the model. Or provide some digital copies of the

input and output files, not just copy and pasted versions in the report. Diagnostics and more details are needed for index GLM analyses and data subsetting.

It is amazing that this assessment report has a single equation in it for maturity and nothing else. I recognize that the SS documentation is referenced, but at the same time that does not provide enough information to know how preliminary data may have been processed, nor does it allow the reviewer the details necessary to know exactly which equations are being used for selectivity, how F's are being estimated, and other important details.