SCRS/2017/221 Collect. Vol. Sci. Pap. ICCAT, 74(3): 968-974 (2017)

EXTERNAL REVIEW OF ICCAT ATLANTIC SWORDFISH STOCK ASSESSMENT

Richard D. Methot Jr.[1]

SUMMARY

This report provides the finding of the external review of the 2017 Atlantic Swordfish Stock Assessment. The assessment used a combination of biomass dynamic models without process error, biomass dynamics models with process error, and an age-structured model with process error. All were implemented with no obvious technical problems. Of these, the JABBA biomass dynamic model with process error and the age-structured SS model provide advanced capability to investigate factors affecting swordfish in the Atlantic and their continued use is encouraged. The various models have good agreement in the finding that the status of SWO-N has improved and now shows $B >= B_{MSY}$ *and* $F <= F_{MSY}$*. The models used for SWO-S also are in agreement that status has improved but B remains slightly less than* B_{MSY} *and* $F > F_{MS}Y$*. Continued research on methods to improve confidence in the CPUE standardization and spatial weighting is encouraged, in addition to research on alternative methods for abundance estimation. Future work could also investigate projections across a range of uncertainty in model configuration, as well as more stochastic projections.*

RÉSUMÉ

Le présent rapport fournit les conclusions de l'examen externe de l'évaluation du stock d'espadon de l'Atlantique de 2017. L'évaluation utilisait une combinaison de modèles de dynamique de la biomasse sans erreur de processus, de modèles de dynamique de la biomasse avec erreur de processus et d'un modèle structuré par âge avec erreur de processus. Ils ont tous été mis en œuvre sans problèmes techniques évidents. Parmi ceux-ci, le modèle dynamique de la biomasse JABBA avec erreur de processus et le modèle SS structuré par âge fournissent une capacité avancée d'étudier les facteurs affectant l'espadon dans l'Atlantique et il est encouragé de continuer à les utiliser. Les différents modèles concordent sur le fait que l'état de l'espadon du Nord s'est amélioré et présente maintenant $B \geq B_{PME}$ *et* $F \leq F_{PME}$*. Les modèles appliqués à l'espadon du Sud coïncident également sur le fait que l'état du stock s'est amélioré, mais B reste légèrement inférieur à* B_{PME} *et* $F > F_{PME}$*. La poursuite des recherches sur les méthodes visant à améliorer la confiance dans la standardisation de la CPUE et la pondération spatiale est encouragée, en plus de la recherche de méthodes alternatives pour estimer l'abondance. Les travaux futurs pourraient également étudier les projections d'une gamme d'incertitudes concernant la configuration du modèle, ainsi que des projections plus stochastiques.*

RESUMEN

Este informe presenta las conclusiones de la revisión externa de la evaluación de stock del pez espada del Atlántico de 2017. La evaluación utilizó una combinación de modelos de dinámica de biomasa sin error de proceso, modelos de dinámica de biomasa con error de proceso y un modelo estructurado por edad con error de proceso. Todos fueron implementados sin problemas técnicos obvios. De ellos, el modelo de dinámica de biomasa JABBA con error de proceso y el modelo estructurado por edad SS proporcionan capacidades avanzadas para investigar los factores que afectan al pez espada en el Atlántico y se insta a continuar su uso. Los diversos modelos concuerdan en el hallazgo de que el estado del SWO-N ha mejorado y ahora presenta $B >= B_{RMS}$ *y* $F <= F_{RMS}$*. Los modelos utilizados para el SWO-S también concuerdan en que el estado del stock ha mejorado, pero B continúa siendo ligeramente inferior a* B_{RMS} *y* $F > F_{RMS}$*. Se insta a continuar investigando sobre métodos para mejorar la confianza en la estandarización de la CPUE y en la ponderación espacial, además de investigar sobre métodos alternativos para la estimación de la abundancia. El trabajo futuro podría también centrarse en la investigación de las proyecciones a lo largo de un rango de incertidumbre en la configuración del modelo, así como en proyecciones más estocásticas.*

[1] Northwest Fisheries Science Center, 2725 Montlake Boulevard East, Seattle, WA 98112, USA

KEYWORDS

Swordfish, Xiphias gladius, Atlantic, Stock assessment, Review

1 . Review Terms of Reference

1. Evaluate the adequacy, appropriateness, and application of data used in the assessment.
2. Evaluate the adequacy, appropriateness, and application of methods used to assess the stock and if appropriate recommend alternative approaches to be accomplished in the future.
3. Evaluate the methods used to estimate population benchmarks and stock status (e.g., MSY, FMSY, BMSY or their proxies).
4. Evaluate the adequacy, appropriateness, and application of the methods used to evaluate future population status, given the commissions objectives.
5. Evaluate the adequacy, appropriateness, and application of methods used to characterize the uncertainty in estimated parameters. Comment on whether the implications of uncertainty in technical conclusions are clearly stated.
6. Comment on whether the stock assessment results are clearly and accurately presented in the detailed report of the Stock Assessment.
7. Comment on potential improvements on the stock assessment SCRS process (CPC participation, transparency, objectivity, documentation, uncertainty characterization, etc.). as applied to the reviewed assessments.
8. Comment on the adequacy of the workplan for the assessment and whether it was adequately addressed by the Data or Assessment Working groups.
9. Consider the research recommendations provided by the working group and suggest any additional recommendations or prioritizations warranted. Clearly denote research and monitoring needs that could improve the reliability of future assessments. Recommend an appropriate interval for the next assessment considering control rules or management strategy in effect.

2. General Findings

The assessment has used a combination of biomass dynamic models without process error, biomass dynamics models with process error, and an age-structured model with process error. All are implemented with no obvious technical problems. Of these, the JABBA[2] biomass dynamic model with process error and the age-structured SS model provide advanced capability to investigate factors affecting swordfish in the Atlantic and their continued use is encouraged. Work on the habitat model is also encouraged to better understand the spatial distribution of SWO-N, especially the effect of changing oceanographic conditions on SWO-N distribution.

The various models have good agreement in the finding that the status of SWO-N has improved and now shows $B >= B_{MSY}$ and $F <= F_{MSY}$. The models used for SWO-S also are in agreement that status has improved but B remains slightly less than B_{MSY} and $F > F_{MSY}$.

The assessment and resultant status findings are driven by observed trends in CPUE from several national fishing fleets operating within sub-regions of the wide distribution of SWO-N. These trends are assumed to be representative of spatially weighted stock abundance throughout its range. Continued research on methods to improve confidence in the CPUE standardization and spatial weighting is encouraged, in addition to research on alternative methods for abundance estimation.

Projection of catch levels that will maintain or increase stock abundance are provided for base case models. Future work could investigate projections across a range of uncertainty in model configuration, as well as more stochastic projections.

The proposed work on a MSE should start with good scoping sessions with the working group, managers and clients to develop a common understanding of goals for the MSE. For example, some goals could be with regard to investigation of the best approach to processing of spatial CPUE information. Other goals could be with regard to status goals and Harvest Control Rule (HCR) goals.

[2] all acronyms used in this review are in accord with their usage in the assessment report

TOR #1 – Data

Catch

- I trust that the continued ICCAT efforts to improve catch data have been successful in getting sufficiently complete accounting of each CPS' retained SWO catch. However, any shortcomings in the catch time series cannot be reviewed from the readily available information.
- It is somewhat troubling that no mention is made of dead discards in the catch section. If dead discards are a relatively small and constant fraction of retained catch, then the assessment advice is probably accurate with regard to sustainable levels of retained catch. However, if dead discards are a large fraction of total catch or if the dead discard / retained catch ratio has changed over time, then the assessment may be biased to some degree. The document should present the current knowledge regarding the degree of dead discards.

Biology

- The distribution model based on habitat variables is a promising approach as presented in SCRS/2017/133. That paper mentions north-south migrations and clearly the habitat model is designed to understand these changing distributions, but the information is buried in the modeling. I recommend that additional figures be included in the future to show average CATDIS maps by season. It would be useful to do this separately for high AMO vs. low AMO years. These graphical augmentations might provide further insight for improving the model.
- The work with the SS model has used the oceanographic information on AMO to modulate the catchability coefficient for some fleets CPUE. This work results in improved modeling of trends in the CPUE for some fleets, hence lessens biases that result when there is substantial pattern in the residuals of the fit to some fleets. I recommend continued efforts to refine this use of oceanographic information in the assessment.

Length Frequencies

- This statement in the SS section is unclear: "For a given size frequency observation, a minimum of 100 measured fish was required for use in the SS model." Does "observation" refer to data from a single fishing trip? Or does it refer to the overall size composition for the year from that fleet? This practice seems OK in the latter sense, but in the former sense it would bias the overall size frequency towards the more common smaller fish.
- I see that it is standard ICCAT practice to report the catch at size in weight equivalents. It is feasible in SS to use weight frequencies. This is done routinely for HMS assessments in the Pacific. However, most SS applications use length composition in numbers. I presume the length compositions used in the SS model for SWO-N are by numbers, not weight.

CPUE standardization

- Clearly there is a very long history of working papers supporting the current approach to CPUE standardization. I could not realistically attempt to review this aspect of the assessment. However, the accuracy of the assessment is highly dependent on the degree to which this work has turned raw logbook data into a spatially weighted index of SWO abundance.
- The degree to which major technological and fleet behavioral changes have occurred and been incorporated in the standardization is mentioned in some of the text, but is not completely or clearly presented. An appendix summarizing the time trends in major calibration factors would provide much needed transparency to this assessment.
    - ✓ For example, the following statement on page 3 is very difficult to evaluate as a reviewer: "Despite the overall flat trend, there was an increase in the CPUE values at the end of 1990, when the Spanish fleet changed the gear from multifilament to monofilament. Catch ratios were used as a proxy for targeting. The potential disadvantages (or advantages) of using this approach were discussed, but the group considered that the estimations presented are the best available information, and that it is worth to using them in the stock assessment."
- The rapid change in CPUE during first few years for several fleets is troubling. The Japanese CPUE in their first 3 years (1974-1976) are anomalously high. This seems an indicator that catchability can change over time. Perhaps the first few years concentrate on small hot spots and in latter years only more dispersed fish are fished. Perhaps the first few years are dominated by a few vessels with highly skilled crews and latter years have more average vessels participating. I do not know that either phenomenon has occurred, but I offer these thoughts only because they are possibilities that would cause changing fleet catchability relatively to the entire SWO stock. It is very difficult to test the assumption that catchability is in fact constant over time. While some gear and area effects are incorporated in the standardization, there is nothing to adequately track technological changes that might allow vessels to improve, or lose, ability to find good fishing locations.

- The figures state that CPUE was scaled based on the average for 1995-2015, but the spreadsheets in the CPUE folder do the scaling based on average across all years.
- There is no obvious explanation why the Japanese index is presented as three time stanzas in some tables and as a single time series in other tables. The effect of this on the standardization is unknown
- The recent 10-year trend by Portugal in the east seems much different that the recent trend of other fleets. This could be due to difference in spatial distribution and as affected by the AMO.
- Catch was lower prior to 1978 and there is a big gap in CPUE data then also. Has there been sensitivity runs of the BSP without including these early years? While long time series are good, their performance is contingent on very well standardized data over time. It is difficult to be confident in the combined index over this time during which it is dominated by the Canadian data. Over 50% of the catch from 1962-1970 was Canadian. They fish a small area and target large fish, so it is logical that their CPUE dropped sharply after 1963. No Canadian fishing from 1971-1978, then resumed at lower level of catch in 1979. The Canadian CPUE in 1979 when their fishing resumed was at a somewhat higher level. This Canadian CPUE increase from 1970 to 1979 dominates the calibration of the overall CPUE index. The Canadian index increase 44% from the era 1965-1970 to the era 1979-1981. Over the same period, the overall biomass index increases 39%. But the Canadian data are for a limited area and for larger fish, so there is some doubt as to whether it is a good index of the whole stock. Model runs that break the Canadian index between 1970 and 1979 should be considered in the set of sensitivities.

Combined CPUE index

- A single combined biomass index in the north is blurring together some disparate trends. Fleets that operate in different areas (i.e. Canada vs Spain areas; slide 8 in Ortiz powerpoint) and fleets that catch different sizes of fish compositions (Figure 13 in Ortiz et al., in press) should not be combined into a single overall index because they are measuring different components of the stock. The area index in the delta GLM is not going to correct for such a pattern. It is equally problematic to include a collection of disparate trends as separate indicators of the same overall stock trend. The SS approach to investigating oceanographic influence on area-specific CPUE seems better at avoiding potential bias due to oversimplification of a complex distribution pattern.
- For SWO-N, it seems better to create at least two indices, as is done for the SS model, each combining only fleets that have similar characteristics. In the biomass dynamic modeling using JABBA, it might be possible to introduce a time lag when using the index for larger fish. The growth curve shown in the SS model work indicates that the larger fish caught by CAN and JAP are 2 to 3 years older than the SWO caught by fleets operating in the eastern Atlantic.
- U.S. data since 1979 show a big drop in early years in Ortiz et al. (in press), but combined index only seems to include more stable U.S. data since 1992. There is no explanation for this difference.
- The combined index also seems to not use the 1962 Canadian data point which was lower and consistent with trend from 1964-1970. But the combined index does use the very high 1963 Canadian observation
- The state-space procedure used for SWO-S with JABBA could be considered as an alternative method to calculate a combined index for the SWO-N. Especially if done separately for the East and West and Southwest regions so only combining fleets with similar characteristics.

TOR #2 – Assessment Methods

- Three classes of models are used: biomass dynamics with no process error (ASPIC and biodyn), biomass dynamics with process error (BSP2 and JABBA), and age-structured with process error (SS). Some have results presented in the main document and some only in appendices. Is this difference meaningful? Does inclusion in the main document imply a greater degree of support by the Group? An overview of the models used would be a helpful addition to the model section and would improve the clarity of the document.
- Time series is getting longer and now has a down – up pattern. Consider doing base model runs without the early years for which CPUE calibration is less confident, or with a break in Canadian CPUE between 1970 and 1979. Kell (2017) does some model runs with a shorter time series for the biomass dynamic model. However, few detailed results are presented in Kell (2017), but it seems inconsistent to find that the early data points are highly influential in Figure 5 of that paper, but the model estimates shown in Figure 6 seem identical for the short and all model runs. Something seems amiss.
- The logistic Biodyn model produces a very high B1/K estimate.
- SS seems to treat CAN early separately from later CAN, but other models treat as continuous?

- It is good to see, on pg. 7, discussion of the steepness parameter used for the age-structured modeling with SS. Fixing steepness is akin to fixing r in ASPIC. It is good to see in the JABBA presentation recognition that a third parameter is needed for the biomass dynamics productivity function in order to align it with the skewness of the age-structured productivity shape. Alternatively, this can be done with SS by replacing the Beverton-Holt function with a 3 parameter Shepard function.
- For the SS model configuration - The Chinese Taipei fish are slightly smaller than Canada and Japan mean size. This suggests that a domed selectivity be used for Chinese Taipei also.
- For the JABBA model of SWO-S, scenario 3 and 4 seem to provide good use of the data. The results show the stock with biomass less than the target level and recent overfishing, catches have declined and recent CPUE trends look good.
- Important development in JABBA is explicit linkage of Beverton-Holt production function to the biomass dynamic shape parameter. However, I do not see where the assertion that JABBA accounts for selectivity comes from.

TOR #3 – Benchmarks and Status

- The work presented in Sharma and Arocha (in press) seeks to provide guidance regarding plausible spawner-recruitment steepness values. This is detailed work and appears to have been conducted very competently. It is useful that they have included recruitment autocorrelation in the analysis. I recommend that the degree of autocorrelation be measured from the time series of recruitment deviations in the SS analysis and used to refine the probable range of results for the risk analysis in Sharma and Arocha (2017). The risk analysis in Sharma and Arocha (in press) that leads to a recommendation of approximately 0.6 S_{MSY} as the biomass limit is sound. This work seems a valuable precursor to the proposed MSE.
- The status of SWO-N is estimated by be near or slightly above B_{MSY} and with F near or slightly below F_{MSY}. The Biodyn, BSP2, and SS models are highly consistent in this finding. This is not surprising because all three models use basically the same CPUE and catch time series and these time series now have reasonable contrast from which to estimate stock abundance and productivity. However, the validity of this finding is dependent on the validity of the models' common assumptions regarding the quality of the CPUE time series. I recommend that the robustness of the status finding be tested further in the future by considering possible density-dependence in catchability, and by considering possibility that CPUE data prior to 1975 are not equivalent to subsequent CPUE data. Model runs that start in 1975 seem valuable to consider and should be estimable because of the increasing trend in CPUE since 1995.
- In order to assure good understanding of the current status and recent increase in SWO-N stock biomass, I recommend that future assessment documents show observed and estimated trends over the past 10 years. This presentation could include each individual CPUE index over that time period, the combined CPUE index, and the estimated biomass index of each of the model's used. The results of the SS model with the AMO index are important to consider as it provides valuable insight regarding the expected trends in different areas of the N. Atlantic.
- For SWO-S, all CPUE trends appear quite flat since about year 2000. Hence, the current status of the stock is highly dependent on the data that indicates the degree of stock decline prior to 2000. Only the Japanese LL index provides this information. In the north also, the Japanese LL downtrend seems more extreme than more stable trends exhibited by most other fleets. This pattern suggests that either the Japanese fleet has experienced deteriorating fleet catchability over time, or other fleets have hyper-stable CPUE and increasing catchability at low stock abundance. This is worth exploration in future assessments. Using only the combined index masks such discrepancies.

TOR #4 – Projections

- The projections with a range of constant catch levels are adequate to indicate the expectation of stock decrease, stability, or increase. However, these simple projections do not adequately convey the risk of constant catch strategies, nor the risk associated with uncertainty in the current assessment. The simulation approach in Sharma and Arocha (in press) is an improvement on this simple approach, and the MSE that investigates improved HCR would be a further improvement.

TOR #5 – Uncertainty Characterization

- For Biodyn / ASPIC the jackknife approach does not seem robust given the noticeable autocorrelation in the residuals. Removing one data point does not fully remove the influence of that point because of autocorrelation with its neighbor.
- Cook's D for cross-validation show that the first, high CPUE point is highly influential. Like the jackknife, cross-validation seems susceptible to autocorrelation of residuals.
- When Biodyn was performance tested against MCMC data, did any of the residual patterns of the fit look as bad as the pattern in the fit to the actual SWO-N data?

TOR #6 – Clear Presentation

- Several comments regarding clarity are found in the sections above.
- Figures 5 and 7 show CPUE series "to be considered". The clarity would be improved if CPUE series used in each model's base case and the combined index be presented also. It would provide even more clarity if there also was a presentation on trends over the past ~10 years.
- It is difficult to appreciate and understand from the summary document the extraordinary amount of work that has gone into CPUE standardization. Unfortunately, this also makes it difficult to review any potential shortcomings in that standardization. A summary of the major gear changes and other factors that cause the processed CPUE trends to differ from the raw CPUE trends would be useful. It would be much more transparent if, for each nation and each area, the raw trend and processed trend be presented, as well as trends in major calibration factors relevant to each CPUE series. Also, the rationale for breaking some series into two or more segments. Some of this information is mentioned in the report, but it is not comprehensive.

TOR #7 – SCRS Process

- The SCRS process of assembling and presenting a set of working papers in addition to a concise summary document is pragmatic and appears successful in building consensus among the working group. This is laudable.
- Unfortunately, the conciseness of the summary document makes it difficult for an external reviewer, who has not attended the working group meeting, to see the big picture of how the working papers support the summary report. This difficult is compounded by the large number of blank folders on the ownCloud site and unclear dating of documents and their relationship to each other. An outline of the sequence of steps that the working group has undertaken would help external reviewers as well as subsequent readers of the assessment report.

TOR #8 – Workplan and Progress

- I have not explicitly reviewed the proposed workplan. Nevertheless, the working group is obviously not static and is making good progress. It is investigating important issues such as changing spatial distributions, it is bringing new models to the assessment, and it is working towards developing a MSE.

TOR #9 – Research Recommendations

- I support the group's conclusions on technical merits regarding which models to use in future. JABBA is good and flexible biomass dynamics formulation and SS provides an age-structured approach that is ready to use size data and oceanographic indicators. I encourage full documentation of JABBA and inclusion in the ICCAT catalogue of methods. There does not seem a need to update BSP2 given the capabilities of JABBA. While ASPIC and Biodyn currently give similar overall results regarding stock status, they lack process error so there is risk that their advice will not be able to track future trends influenced by changing oceanographic conditions.
- I encourage more explicit, clear presentation and comparison of CPUE trends by fleet and area and season. Outliers need to be identified and potentially down-weighted in combined indices and assessments.

- MSE needs to be able to incorporate AMO effect and spatial distribution and changing catchability in the operating model. From this, it seems feasible to test whether a simple combined CPUE could be an accurate indicator of stock trends.
- The MSE could either take a detailed and technical approach (e.g. spatial and oceanographic effects on the CPUE indices and subsequent effect on the assessment), or it could take a management oriented approach to investigate possible changes in the HCR. While both goals could be done at the same time, it might be better to tackle these as different projects in order to have high client engagement in the HCR project.
- Conduct sensitivity analysis with estimated total catch, including plausible degree of discard / retained catch ratio changing over time.

References

Goodyear C.P., Schirripa M., and Forrestal F. (in press). Creating a species distribution model for swordfish: evaluations of initial habitat variables. Document SCRS/2017/131. 16 p.

Kell L.T. 2017. Model validation using prediction residuals. Document SCRS/2017/127 (withdrawn)

Ortiz M., Mejuto J., Hanke A., Ijima H., Walter J., Coelho R., and Ikkiss A. (in press). Updated combined biomass index of abundance of North Atlantic Swordfish stock 1963-2016. Document SCRS/2017/137. 19 p.

Sharma R., and Arocha F. (in press). Resiliency for North Atlantic Swordfish using life history parameters. Document SCRS/2017/143. 16 p.