REPORT OF THE 2021 MEETING OF
THE WORKING GROUP ON STOCK ASSESSMENT METHODS (WGSAM)
(Online, 5-10 May 2021)

The results, conclusions and recommendations contained in this Report only reflect the view of the Working Group on Stock Assessment Methods. Therefore, these should be considered preliminary until the SCRS adopts them at its annual Plenary meeting and the Commission revise them at its Annual meeting. Accordingly, ICCAT reserves the right to comment, object and endorse this Report, until it is finally adopted by the Commission.

1. Opening, adoption of Agenda and meeting arrangements and assignment of rapporteurs

The meeting of the Working Group on Stock Assessment Methods (WGSAM: "the Group") was held online from 5 to 10 May 2021. Dr. Michael Schirripa (USA), the Working Group ("the Group") rapporteur and meeting Chair, opened the meeting, and the Executive Secretary, Mr. Camille Manel, welcomed the participants to the meeting. The Chair proceeded to review the Agenda which was adopted after some changes (**Appendix 1**).

The List of Participants is included in **Appendix 2**. The List of Documents presented at the meeting is attached as **Appendix 3**. The abstracts of all SCRS documents and presentations provided at the meeting are included in **Appendix 4**. The following served as rapporteurs:

Sections	*Rapporteur*
Items 1, 7	A. Kimoto
Item 2	K. Gillespie, D. Rosa
Item 3	D. Courtney, H. Winker
Item 4	R. Scott, E. Babcock
Items 5-6	M. Schirripa, A. Kimoto

2. Harvest Control Rules, Limit Reference points and Management Strategy Evaluation

Presentation SCRS/P/2021/028 provided an overview of recent progress and ongoing work related to the North Atlantic swordfish (N-SWO) Management Strategy Evaluation (MSE). The initial Operating Model (OM) grid had seven axes of uncertainty (natural mortality, recruitment deviation, steepness, CPUE CV, effective sample size of the length comps, catchability increase, and environmental modulation of catchability among fleets) with 288 OMs. In 2020 and early 2021, the grid was revised in three respects: i) the OM grid was migrated from SS3.24 to SS3.30, ii) the CPUE CV and length composition effective sample size axes were combined into a single axis reducing the grid to 216 OMs, and iii) the OMs were modified to estimate discarding among fleets and account for discard mortality (as requested by the Commission). This inclusion of discard mortality estimation is the most consequential of the changes as it allows for Candidate Management Procedures (CMPs) to investigate effectiveness of minimum size limits as components of harvest control rules. CMP development is underway and will continue through 2021. In early 2021, the MSE code began external peer-review. N-SWO MSE sub-groups are developing candidate exceptional circumstances (ECs) and performance metrics which will be considered by the SWO Species Group, the SCRS, and PA4 later in 2021.

The use of the swordfish joint CPUE for MPs was discussed. It was noted that a strength in the use of this CPUE would be that if a singular fleet is not able to provide data the joint CPUE can still be used, probably avoiding an EC that could lead to a revision of the applied MP. It was further noted that a sensitivity analysis with a "leave-one-out" approach (dropping a fleet one at a time) could be done to test the robustness of the results when a fleet's data is not available.

It was asked if the conflict between CPUEs in the East/West areas is being considered for the CMP development, as only one CPUE would be used in the MP, especially if it is data-based. It was noted that the conflict between CPUEs has been included in the OM by having an uncertainty axis related to environmental effects, where in half the OMs there is an Atlantic Multi-decadal Oscillation (AMO) effect that modulates catchability between conflicting indices. It was further noted that most effort so far on N-SWO MSE has been on the conditioning of the OM, so far there was not much effort in the development of CMPs, therefore it is still open for both model or empirical based CMP development.

It was noted that with an empirical based MP it may not be possible to track stock status (stock assessments would need to run periodically to monitor stock status). It was noted that model-based MPs can create computational challenges, these could be overcome to a certain point with the use of a production model that can run in a few minutes and is still robust in providing MSY estimates.

It was further noted that an update of the MSE will be presented to the Panel 4 meeting in July. The Group agreed that when presenting the different MSE efforts at the Commission level these should be presented in a uniform manner to avoid having mixed messages between stocks.

Presentation SCRS/P/2021/025 provided an update on the developments of the Atlantic tropical tuna MSE in relation to bigeye (BET), yellowfin (YFT) and eastern skipjack (E-SKJ) stocks. The first phase of the project was carried out in 2018 and a summary of the first steps was provided with details on the model that is going to be used and the working documents that were produced. In 2021, the project has re-started with a review of the main uncertainties of tropical tuna fisheries with the aim of defining the axes of uncertainty of the MSE, which will be used to condition OMs.

The Group thanked the authors for their presentation and noted the progress this MSE has made despite limited resources. The Group asked if this effort represents an MSE for each stock or a combined MSE that accounts for all three stocks (BET, YFT, E-SKJ) in the OM grid and management procedures. The authors clarified that separate OMs will be built for each species, but unlike other ICCAT MSE processes, this MSE will incorporate the three species in a single MSE as there are many similarities in the fleets, gears, and distributions among these three stocks and that the simulation framework would need to account for significant overlap among the fisheries. They further noted that western SKJ is being evaluated in a separate MSE. It was asked if ECs have been discussed within this MSE process and the authors noted that they have not been considered yet as this MSE is in relatively early stages.

SCRS/P/2021/023 provided a summary on the North Atlantic ALB (N-ALB) MSE over the last few years. Definition of the list of performance statistics was agreed with the Panel 2 in 2016 (Anon., 2016), prior to the adoption of the harvest control rule (HCR) in 2017. However, mostly 4 performance metrics are used, expressed in the form of radar plots, to illustrate the performance across the reference case, which includes uncertainty around biology (natural mortality: M and steepness), data (size, CPUE, and tagging) and modelling options (time range and catchability trends). The 132 OMs are equally weighted. The code was peer reviewed in 2018 and comments were addressed in 2019. Several HCR variants (as specified in Rec. 17-04 (Anon., 2017a)) have been tested (as components of a CMP), the SCRS advice incorporates the elements to adopt a full MP, and feedback to Panel 2 has been provided on proposed indicators to detect ECs. The primary document for the ALB MSE is the ALB MSE consolidated report (Merino *et al.*, 2020).

The Group thanked the ALB MSE team for pioneering several MSE analyses and concepts within the SCRS and at the Commission level. The Group asked how the ALB MSE framework was considering MP implementation errors. It was noted that both the current carry-over scheme as well as a banking and borrowing scenario where TAC was implemented with an error of 20% had been tested with positive results, thus could be useful to define some EC indicators. This led to discussion on whether short-term deviations in implementation may lead to an EC finding. The Group noted that many scenarios, not limited to implementation error, may lead to indicators registering values outside the “full range of values” and that it was not practical to plan for all eventualities that could lead to ECs. The Group agreed on the need for flexibility in the draft EC protocol and noted the importance of inclusion of both science and management responsibilities within the protocol.

SCRS/P/2021/027 detailed recent progress for the Atlantic bluefin tuna (BFT) MSE and the work plan for 2021 and 2022. The presentation addressed recent developments including interim adoption of the reference grid, plausibility weighting of the factor levels within the grid, development tuning and reconditioning. The BFT MSE has six CMP development teams working mainly empirical (or index based) procedures with model based CMPs. Development tuning is the process of having all CMPs tuned to achieve one common performance metric (Br_{30} or B_{MSY} in year 30) so as to then be able to evaluate performance across other metrics. It is a process to both refine individual CMPs and to evaluate and vet their performance. It is a critical step in the evolutionary process of CMP development and in the process of narrowing down the number of CMPs to a few top-performing ones to present to the Commission and Panel 2 in November. The author noted the decisions such as how to develop OMs, empirical versus model based CMPs and whether the MP should adjust the previous TAC directly or re-estimate scale may be context and stock-dependent but should be explicit considerations. For stocks where estimating scale has proven challenging (e.g., eastern BFT) adjusting previous TAC provides a grounding in an absolute scale and may be preferable to allowing a model-based MP to re-estimate scale at each application of the MP.

The Group commended the BFT Species Group for confronting significant stock complexity in their analysis. The Group asked how advice would be balanced between both MSE and assessments. It was noted that even after implementation of an MP, assessments would continue to occur on a 3-5 year cycle to act as a check on the MSE. It was also noted that parallel advice would likely be provided to the Commission in 2022 in the form of MSE and an eastern BFT assessment. The Group asked how ECs were being dealt with in the BFT MSE process. The author noted that ECs have not been a focus and will likely not be included as in initial CMP adoption; rather BFT MSE will likely follow the lead of other ICCAT MSE groups and their EC frameworks.

An interactive MSE visualization tool intended to aid in consultation and decision making (harveststrategies.org; Slick Decision Analysis) was presented to the Group. The decision analysis tool presents outcomes of potential policy options across various states of nature. This tool allows for the simultaneous presentation of various performance metrics and can account for uncertainty in the states of nature in an interactive format, allowing users to filter results interactively in order to explore robustness and performance. While the tool can be applied to any decision analysis context it was specifically designed to investigate the performance of fisheries management procedures tested by MSE.

The Group noted that the tool appeared to be very powerful for demonstrating trade-offs between management procedures. The volume of information contained within the tool, while very useful for fisheries scientists intimately familiar with the assessment and simulation process, may present significant hurdles for more lay audiences. It was suggested that science-manager focus groups identify key metrics for MP performance and trade-offs, as was done previously for the development of the Kobe matrix. It was noted that an important characteristic of MSE is the additional information available and that this information should be considered by decision makers. It was suggested that two visualization versions could be developed: an executive summary version that includes key metrics, and a full version for those who would like additional detail. The Group concluded the discussion by suggesting that national scientists discuss the tool with their respective resource managers to solicit feedback on key decision-making metrics, plot types and other features within the visualization tool.

The Group discussed broad themes among the ICCAT MSEs. There was a strong emphasis on the importance of clear communication between scientists, managers, and other stakeholders throughout the process, for example via a standard set of terminologies and tradeoff visualizations. It was noted that clear and consistent communication was an important recommendation in other RFMO MSE reviews (e.g., ICES). The Group discussed the continued need for strong SCRS and Secretariat support for communication (e.g., workshops at the Commission meetings, MSE training courses for scientists and managers, etc.). The Group noted that a new website (www.harveststrategies.org) is compiling communications tools on MSE, including Slick. It was suggested that ICCAT work plans include very clear and explicit goals for communication steps within each step of the MSE process.

There was brief discussion on other methods that may have shorter timelines to implementation, while still have HCRs that consider risk relative to limit reference points (e.g., Maunder *et al.,* 2020). Such methods, in many cases, are careful to focus on uncertainty that has meaningful impacts on management advice (i.e., expending less effort on areas of the grid that are not considered plausible or have significant biological impacts). The Chair suggested that these alternative methods may warrant further examination given the demanding workloads currently facing the SCRS.

3. Recommended standard diagnostics for stock assessment models

Two presentations were provided by the developers of an important new compilation of diagnostic tools for integrated age-structured stock assessment models within an R package (ss3diags; https://github.com/jabbamodel/ss3diags). The ss3diags R package is documented within a recent publication (Carvalho *et al.,* 2021), which also provides worked examples of each diagnostic method implemented for two recent age-structured stock assessment models completed in Stock Synthesis.

Presentation SCRS/P/2021/022 identified practical guidelines for the review and interpretation of the diagnostics described in Carvalho *et al.*, 2021. The guidelines include four main components related to the "plausibility" of a model: Model convergence; Goodness of fit; Model consistency; and Prediction skill. A flow chart from the publication identified "detours" to consider during model development when particular diagnostics do not show satisfactory results. Plots available from the ss3diags R package aid diagnostic interpretation: Model convergence–jitter–; Goodness of fit–individual residuals and JABBA joint residuals–; Model consistency–R0 profile, age structured production model (ASPM), and retrospective analysis bias–; Prediction skill–hindcasting cross-validation (HCxval)–. It was also noted that plausibility of a model must also be evaluated independently of diagnostics, on a case-by-case basis, for example relative to stock specific exploitation history and population biology.

The Group discussed implementation and interpretation of the jitter diagnostic, but noted that there is no single convergence criterion for the diagnostic, and that multiple diagnostics should be evaluated together. The Group discussed implementation and interpretation of profiling diagnostics, which could help determine which parameters should be within an uncertainty grid (i.e., parameters with a flat profile, and presumably uninformed by data).

The Group discussed that diagnostic results should not necessarily be interpreted individually on a pass/fail basis, because of the possibility of false positive results. Instead, diagnostic results should be interpreted as an aid to identify potential model diagnostic issues for continued model development. The Group also discussed that diagnostic should evaluated as early as possible within the model development process. It was noted that one goal of the publication was to serve as an aid to improve workflow within an inclusive model development process, for example, within a stock assessment team. The Group noted that there are other diagnostics available for integrated analyses, which were not included in the publication. Examples include using the fitted integrated model likelihood to generate pseudo data given model estimates (Self-test), and MCMC diagnostics under ongoing development.

Presentation SCRS/P/2021/020 demonstrated practical applications to ensemble weighting and projections using model validation and prediction skill based on the objective criteria for evaluating model plausibility proposed by Carvalho *et al.*, 2021. Three emerging issues were identified: 1) Propagating estimation error and structural uncertainty into the ensemble; 2) Selecting models for inclusion in the ensemble composition (hypothesis testing); and 3) Assigning relative weights (plausibility) to the model outcomes for ensemble weighting. The presentation focused on some practical aspects of ensemble weighting with Stock Synthesis using new R package 'ss3diags' (github.com/JABBA-model/ss3diags). Results from the delta-multivariate-lognormal (delta-MVLN) Monte-Carlo approach to combine model outcomes and future projections under any weighting scheme, while also accounting for within model uncertainty, were provided for a range of examples. An example of selecting models for inclusion in the ensemble composition (hypothesis testing) was provided. Plausibility of alternative recruitment hypotheses was evaluated with automated diagnostic tools. Examples of assigning relative weights (plausibility) to the model outcomes for ensemble weighting were provided, and upcoming publications were identified (Kell *et al.*, 2021). The presented results highlight the importance of model validation using prediction skill as an objective way of comparing plausibility of alternative model scenarios and across models.

The Group acknowledged the effort for the development of the 'ss3diags' R package (https://github.com/jabbamodel/ss3diags), which offers a range of important model diagnostic routines for Stock Synthesis. The Group noted that ss3diags also provides an automated and further refined implementation of the delta-MVLN Monte-Carlo approach to rapidly estimate uncertainty of key quantities for Stock Synthesis models (Walter *et al.* 2019; Walter and Winker 2020 and reviewed in Anon. 2019).

The Group also noted that key diagnostics, such as residuals run tests, retrospective analysis, and hindcast cross-validation are also available in the Bayesian state-space surplus production model 'JABBA' (Winker *et al.*, 2018) and in the new R package 'a4adiags' (github.com/flr/a4adiags) for the statistical catch-at-age (sca) model FLa4a.

The Group discussed potential limitations of using likelihood or Akaike's Information Criteria (AIC) for stock assessment model selection. AIC, for example, is only a relative measure of the appropriateness of models, however, for model validation, additional diagnostic tests are required. It was noted that information criteria may be challenged by complex models, for example if many parameters are fixed or if the model includes process errors. Use hindcasting to estimate prediction skill, a measure of the accuracy of a predicted value unknown by the model relative to its observed value, can aid detecting if the model is not overfitted by determining that the model is able to make future predictions (Kell *et al.*, 2021). The Group noted that hindcasting can be, in principle, applied to any stock assessment model that provides an option for a forecast under alternative catch scenarios.

The Group discussed that plausibility of alternative somatic growth and historical catch scenarios could be evaluated based on the objective model diagnostic criteria proposed in Carvalho *et al.*, 2021. In particular, hindcast cross-validation may be an aid to validate the model scenarios and compare their plausibility. The Group noted that, to improve contrast among models, the hindcast horizon (currently one step ahead) may need to be extended for long lived species (for example, at least three years) to account for strong inertia in the population forecasts as a result of the long generation time.

The Group discussed that different diagnostic criteria are often utilized for MSE compared to those discussed during this meeting, which were tailored to a stock assessment. The Group recognized that there is a place for diagnostic criteria in both individual models and MSE, but the Group only discussed in detail diagnostics as applied for stock assessment models. Consequently, the suggestions on diagnostics from the Group are relevant primarily for stock assessment models but may be explored for applications to MSE model grid development, which should probably be discussed in more detail separately in a future WGSAM meeting.

Document SCRS/2021/050 provided a review of plausibility and uncertainty of mainly three key data inputs used in the 2017 Shortfin Mako Stock Assessment meeting (Anon., 2017c): catches, abundance, and biology. The main conclusions regarding catches were that the historical Task 1 catches used in the assessment base case scenario (C1) greatly underestimated the magnitude of catches in several of the initial decades when taking into consideration the history of the respective fisheries, the fleets' capacity, and the fishing effort exerted by each fleet; whereas the alternative catch scenario considered in the assessment (C2) probably considerably overestimated the catch of some fleets and fleets combined at least for the most recent period of that series. For indices of abundance, qualitative and/or quantitative limitations in some of the CPUE series would likely affect their representativeness as indicators of abundance. In terms of biology, the growth model and female age at first reproduction used in the assessment were compared to those from other studies and those obtained from preliminary estimations based on tagging-recapture data, noting that the estimates used in the assessment showed slower growth dynamics than those from all these other studies.

A discussion on the presentation and paper ensued. A comment was made about whether this work, which the Shark Species Group was unaware of, would also be presented at the Shark Species Group meeting in September and why the work was not presented during or after the 2017 Shortfin Mako Data Preparatory meeting (Anon., 2017b). The first author responded that it is not possible to conduct this kind of work prior to the stock assessment when the SCRS and other scientific contributions are not available, and also because it required review of the stock assessment itself and papers presented during previous decades of the history of ICCAT that are usually forgotten, but provide useful information. Unfortunately, it is not possible to anticipate the approach that will be used in an assessment with new input data, time periods considered, and models implemented; or what documents and contents will be presented. Therefore, a simple

contribution like this one can only be carried out *a posteriori*. On the other hand, the shortfin mako assessment process was initiated in 2017 and concluded in 2019, and the agendas of the different Working Groups were cancelled or modified in 2019 and 2020 due to COVID-19, which also affected the work plans and capacities of the already very complex workload of national scientists. Therefore, in addition to domestic limitations probably affecting many scientists, this was the earliest opportunity to present this contribution on the plausibility and uncertainty of the input data and results achieved in the assessments between 2017-2019, hoping that they will be of use at this point considering the restricted agenda of the SSG September 2021 that was already established and communicated.

The former Chair of the Shark Species Group clarified that issues with the Task 1 catches were acknowledged in the stock assessment, which is why an exercise was conducted to reconstruct the catches by some CPCs historically and also ratios of SMA to other species were used to construct an alternative catch series (C2) noting that most CPCs need to improve their catch data reporting to the Secretariat, which was something out of the control of the Shark Species Group. Until CPCs provide better catch records, the Shark Species Group must rely on its own estimations. For indices of abundance, it was acknowledged that more in-depth vetting of individual CPUE series is always possible, but that the series used in the assessment had been subject to the battery of criteria developed by the WGSAM to determine inclusion in the assessment. It was further pointed out that cross-validation and a hierarchical cluster analysis had been run for the CPUE series included and that the general trend of these series was remarkably consistent, something generally unusual in shark assessments. For biology, specifically the growth model used, it was pointed out that it came from a stock-wide study in the Atlantic conducted as part of the activities of the SRDCP (Shark Research and Data Collection Program), which was the most comprehensive information on growth for the North Atlantic stock available at the time. It was recognized that relatively little time was available to review all the data sources used in the assessment, but that the data used were the best data available to the Shark Species Group at the time of the assessment. It was also noted that these and other alternative hypotheses about the main data inputs could be considered during the next stock assessment, for example as alternative sensitivity scenarios or plausible states of nature (e.g., low/high catch runs, low/high productivity runs).

The author further noted that for bycatch species, like shortfin mako, understanding the history of catches and other essential aspects of the stocks being assessed is a process that takes many years, and is in a much more advanced stage for targeted tuna stocks compared to shark stocks, whose history of assessment is much shorter. It is thus important to fully evaluate all data sources and modeling approaches before providing advice that may affect the status of the stocks and the fishing operations of the fleets involved. In that respect, another comment noted that, although older period catch estimates are available for some Contracting Parties, reported catch records for sharks only started being formally reported in the 1990s, with mandatory reporting only starting in the 2000s.

A comment was made as to why the SRDCP had not conducted any work, such as the one presented, that could have been used in the shortfin mako assessment. It was clarified that the SRDCP did indeed conduct work, such as the growth study and size distribution revisions, whose results were used in the assessment, as well as other activities that helped inform the assessment (e.g., population genetics). It was also noted that specific work on catch reconstruction, which can also be considered part of the SRDCP activities, had not been conducted because it involves extensive collaborative work from not only active members contributing to the SRDCP Group, but also from other CPC scientists and support from the Secretariat, and that such a project requires considerable time and could perhaps be contracted. Nevertheless, some work in this regard, which resulted in the C2 alternative catch scenario used in the 2017 shortfin mako stock assessment (Anon., 2017c), has been undertaken by national scientists as well as additional work for other shark species in the past.

The question was asked of whether the model diagnostics described in another presentation (SCRS/P/2021/022) could be used to examine the plausibility of catch scenarios for SMA or other species in general. It was noted that it is becoming increasingly important to estimate total catch (not just landings) and that information is only available for a few fleets. The catch determines the scale of abundance.

It was also noted that the 2017 shortfin mako stock assessment (Anon., 2017c) represented the first application of an integrated assessment model for SMA and also that the catch series extended further back in time compared to previous assessments. The foundation now exists for going forward into the future. Very specific hypotheses can be formulated with Stock Synthesis and projections will be particularly important for considering rebuilding plans.

4. CPUE standardization/incorporation of oceanographic and environmental changes into the assessment process

The Group acknowledged that CPUE standardization has always been an important topic for the WGSAM as indices are an important input to assessment and can always be improved (e.g., potentially by including environmental data in the standardization).

4.1 Results of study on Investigation into decadal changes in swordfish habitat distribution

Document SCRS/2021/075 presented preliminary analyses of spatiotemporal patterns in swordfish habitat distributions using a swordfish species distribution model (SDM). Habitat distribution of SWO varied spatiotemporally throughout the SDM model grid based on species affinity curves for environmental and biological variables. Model results can be used as input data for a longline simulator (LLSIM) to generate simulated CPUE data.

Because swordfish can tolerate a broad temperature range, temperature was found to have little predictive value for swordfish distribution. Habitat suitability was mostly structured by oxygen, zooplankton density, and sea surface height. When summed across all depth layers, SDM results showed areas of high habitat density in the North and South Atlantic (referred to as 'core habitat') and a lower density area near the equator that roughly coincides with current ICCAT management boundaries. Predicted quarterly habitat was in agreement with observer data and known migratory patterns. Decadal patterns in core habitat showed an expansion in suitable habitat area over time but a decrease in the density of particularly favorable habitat, suggesting increasing habitat homogeneity. This habitat expansion in the North and South Atlantic could result in two larger areas of swordfish habitat, although both with reduced swordfish density. This shift may be correlated with changes in the Atlantic Multidecadal Oscillation (AMO) and may have impacts on MSY metrics, and further examination was recommended.

SCRS/2021/075 concluded that the swordfish SDM is sufficiently parameterized to describe swordfish distribution over space and time and recommended that monthly habitat coefficients associated with each SDM grid square be a covariate included in CPUE standardization.

The Group acknowledged the utility of this work and that it could be applied to other species, such as bluefin tuna. In this case species-specific affinity curves would need to be created. The Group discussed the need for a new affinity curve related to the density of a direct prey species (as prey distribution may determine the distribution of predators), but this will require distribution data for prey species such as forage fish. There was concern among the Group that the model may not be relatable to the behavioral patterns of fleets as vessels may not be looking for areas of high swordfish concentration because of low catch quotas.

The Group raised the issue of differences in habitat preferences between swordfish age classes, but the creation of a juvenile swordfish stanza in the SDM is not currently feasible due to low pop-up satellite archival tag (PSAT) sample size of small swordfish. It was suggested that a future BFT SDM would be a good opportunity to test out a multi-stanza model as the PSAT sample size and size range is much larger for that species. Additionally, it was put forth that the LLSIM model can incorporate multiple species to examine swordfish management objectives related to bycatch reduction of non-targeted species such as blue marlin (e.g., area closures, fishing at night vs day hours). At the present time multiple SDM outputs cannot be generalized into a single SDM model. However, the LLSIM can include many different species and/or life stages, each with a unique SDM input, which may be a future area of interest.

There was a debate regarding the incorporation of environmental data in CPUE standardization. It was argued that analysts should not use environmental covariates in standardization, unless they believe the environment is influencing catchability. If there are large scale influences of the environment on the productivity of the stock, including the variable in standardization may lead to a hyperstable index that doesn't correctly follow change in abundance. However, it was noted that the SDM reports relative swordfish abundance, in that relative abundance across all model grid cubes sum to one. The SDM only demonstrates the distribution of the stock, the quantification of which doesn't occur until incorporation into the LLSIM. Because the habitat variable from the SDM sums to one, incorporating this into the CPUE standardization may be less problematic than using, for example, sea surface temperature (SST), which could change throughout the entire region over time and be confounded with productivity. This is an ongoing topic of discussion that should be addressed by the WGSAM.

The Group noted that SST is not likely to be a good predictor of swordfish habitat suitability as swordfish spend much time in deep water and are caught at depth. As SST is broadly used as an environmental predictor of habitat, an analysis of the spatiotemporal conditions in which it may be a good predictor would be of interest. The Group suggested that this would be a useful future research idea for a surface-oriented species such as blue marlin (BUM) using the LLSIM as a tool.

There was discussion regarding the appropriate resolution of spatial models, and it was suggested that operating models (OM), such as the SDM and LLSIM, should be the finest resolution possible to capture true environmental variation, whereas estimation models should be on a coarser scale. For example, the length of a longline is often greater than the width of an SDM cell, so data must be analyzed on a realistic scale. It was posed that the SDM model grid can be made coarser by averaging across existing cells rather than reconfiguring the existing model structure. The Group suggested that the WGSAM should provide guidance on the spatial resolution of models as well as a cross check on the relationship between the spatial distribution of fish and fishing effort for directed fisheries such as swordfish.

Regarding the affinity curves used in development of the swordfish SDM, it was noted that temperature and depth are currently in the model as independent curves, whereas the two are not truly independent. Probabilities are currently calculated by multiplying the probability of occurrence in each depth layer by the probability of occurrence at a given temperature, which may result in the underestimation of occurrence probabilities at deeper depths and overestimation near the surface. There is an ongoing tagging work for swordfish to fill spatial gaps, and the new PSAT technology will have more capability for detecting depth and temperature utilization, thereby providing better movement predictions.

4.2 Results of study of the addition of a second fishery to a longline simulator

Document SCRS/2021/049 examined swordfish habitat patterns from the SDM model presented in SCRS/2021/075 in 4 dimensions (latitude, longitude, time, and depth). The study defined the term habitat 'envelope', which is the volume that includes the densest concentration of a given fraction of the population. The volume of a given envelope will vary seasonally and over time. Similarly, to the habitat expansion seen in SCRS/2021/049, this work demonstrates an increase in the size of Envelope99 (habitat envelopes containing 99% of the population: E99) over time (there is intra-annual variation but an overall upward interannual trend). Simulated distributions in the North and South direction were quantified by the mean latitude of the habitat envelope in ICCAT regions SWO-N and SWO-S, and the mean fraction of the population in each envelop in each region.

Results from this work show a decrease in latitude SWO-N and an increase in latitude in SWO-S, suggesting an overall southward shift in E50 (habitat envelopes containing 50% of the population) and E99 throughout time that has increased in scale since the mid-1970s. The fraction of habitat in each envelope (E50 and E99) in SWO-N also decreased over time beginning in the mid-1970s, likely accompanied by an increase of the population fraction in SWO-S. This trend does not appear to be cyclical over the available time series. Each environmental factor considered in the model appeared to influence distribution in some way, although temperature seemed to be most responsible for the southward trend in the South Atlantic. The document concludes that the SWO SDM replicated conventional wisdom on swordfish movement and therefore provides a reasonable basis for LLSIM.

It was reiterated that the SDM model calculated the distribution of relative abundance and therefore does not look at the change in population abundance over time. The Group again discussed the difference in habitat utilization between adult and juvenile swordfish, as well as the change in SWO age structure. It was suggested that age structure could introduce bias into CPUE standardization, but that the SDM does not yet have age structure capabilities to test for this. The LLSIM could accommodate multiple life stages if data is available to parameterize separate SDMs for each.

Document SCRS/2021/048 presented the simulation of longline CPUE estimates using the LLSIM. The current LLSIM assembly includes three simulated fleets ('Japan-like', 'Brazil-like', 'US-like'), three gear types (shallow, mid-depth, and deep), and two species (SWO and BUM). Fleets were patterned after Brazilian and Japanese data in the ICCAT Task 2 database in regard to the number of hooks fished in each 5-degree latitude/longitude (lat/lon) grid cell and the species composition of catch. Each set was assigned to either shallow, mid-type, or deep gear type, with fishing depths adopted from gears described in a previous study on US fisheries. The number of sets per gear was estimated from total hooks using the number of hooks along the gear mainline.

Sets from each 5-degree lat/lon ICCAT grid cell were randomly assigned to SDM 1-degree lat/long grid squares, and resulted in a monthly number of sets per fleet per gear within each SDM cell. The LLSIM also included a gear file which incorporated miscellaneous identifiers, the probability of hooks in each depth layer, and gear specific catchability. When combined with output from the SDM model and stock assessment population trends, this produced a time series of simulated longline catch data. It was stressed that users should input multiple trends if using this model for simulation testing studies.

The Group discussed the feasibility of simulating sets that spanned 1-degree squares, and determined that this would require running another program. Habitat coefficients calculated for the LLSIM data are calculated based on knowledge of the depth distribution of hooks over time. The habitat coefficients could be calculated at different scales and could include adjacent 1-degree squares. The Group also discussed including the idea of areas with more or less productivity in a later version of the simulator, which could be related to population trends over time.

The Group returned to the ongoing topic of the desirability of including environmental variables in standardization. The SDM can be used to make datasets with variability in productivity and/or habitat preference to test whether including the environment (e.g., SST) in standardization is appropriate, and whether models can detect catchability versus productivity influences of environmental variables. Since many indices (e.g., BFT) are now using SST, this is an important topic to address. This could be accomplished in the simulator by using the same environmental variable to drive both the abundance trend and the distribution. Environmental variables (e.g., temperature) could also be used as a driver of catchability. The environmental variables operate on different spatiotemporal scales for productivity (large scale) versus catchability (small scale). The hypothesis that looking at multiple scales allows for discriminating between productivity and catchability drivers was proposed. In an estimation model this would require the ability to track range increases or decreases versus a change in population linked to environmental variables (e.g., temperature).

The Group highlighted inconsistencies in the term 'catchability'. In the context of stock assessment, catchability is the constant of proportionality between a standardized index and global abundance. At the level of individual sets, catchability is the probability of catching a fish given local density, and is related to the gear. Availability of fish to gear is useful terminology, and is an interesting research question. The Group suggested that the term 'catchability' should be clearly defined at the beginning of each study to ensure consistency.

4.3 Using LLSIM to evaluate CPUE methods and estimation of bycatch

Document SCRS/2021/078 discussed testing a generalized bycatch estimation software in development using data for blue marlin generated by the longline simulator. Multiple GLMs were fit to observer data and the best fitting model for each distribution was selected using information criteria (e.g., BIC). Ten-fold cross validation was employed to determine which of the top-performing models best predicted CPUE. The best model was then used to predict bycatch in each unsampled trip/set from logbooks and summed to get an annual bycatch estimate. Prediction interval variances were also summed across trips. This method was tested using the blue marlin data generated by LLSIM, with a simulated observer program applied to 5% of trips. In cross-validation, the delta gamma slightly outperformed the other candidate models while the delta lognormal performed the worst. Bycatch estimates produced by the delta gamma model matched observer bycatch data closely and had very small variances.

It was noted that, unlike for index estimation, for total bycatch estimation, it does not matter whether an environmental variable influences catchability, distribution, or total abundance. Because the model is used to predict the catches in unsampled trips, rather than to extract a standardized index, any variable that improves the prediction of individual CPUE values improves the estimate of total bycatch.

Abundance indices for the North and South Atlantic were then estimated using the same delta-gamma GLM and showed a steep decline in BUM abundance at the start of the time series followed by a slower decline in recent decades, consistent with the true trend input into LLSIM. Suggested future directions to improve index estimation from the bycatch estimation tool included the addition of random effects, GAMs, and spatial models to cope with interactions between spatial patterns and the temporal trend.

The LLSIM is a very useful operating model for testing bycatch estimation and CPUE standardization methods because it includes a realistic amount of complexity in spatiotemporal patterns and variability in gear usage. This allows the design of simulated data sampling programs, such as observer programs, with realistic limitations and potential biases, to use in testing the effectiveness of the estimation software.

The Group commended the utility of this work and discussed the challenges of estimating historical bycatch, particularly for fleets with insufficient logbooks and limited observer data. Changes in gear and fishing areas over time makes it difficult to assume constant values across years retrospectively. Such data issues could be simulated in LLSIM, so that the potential biases could be quantified. The Group expressed interest in the development of scenarios that stretch the ability of the model to fill in past values and give spatial estimates of variance, and indicated that results from such scenarios could be useful to assessments. This could generate reasonable hypotheses and provide an objective way to suggest bounds for historical bycatch estimates, the magnitude of which tend to vary among existing models.

The question was raised of whether this tool could be used to determine best CPUE standardizations. It was explained that the best model for index CPUE standardization would need to be more complex and include interactions between variables such as area and year, as well as a random effect for confounding time and space, which is not the current functionality. It was also pointed out that observer data covers only a small fraction of a given fleet and is not randomly spatially distributed, which limits the complexity of candidate models, particularly if there are significant interactions between categorical variables.

The Group discussed the importance of having a clean and updated database and analysis of fishing effort. Increased specificity in effort estimates would improve total bycatch estimates, as any variable included in a bycatch estimate model must also be present in effort data.

The Group acknowledged that this work provides a useful tool for bycatch estimation with direct applications to ICCAT research interests. Several current Workplans include better estimates of bycatch of those billfish and shark species, and it was recommended that this tool be used for such in a future study. It was noted that this tool is ready to be applied to bycatch estimation of any species for which the fishery has observer data and logbook info. It was noted that CPUE data could be taken from either observer or logbook data, although logbooks may not be as useful as observer data as they tend to only record target species. Application to other species will require access to this data, knowledge of fishery, and collaborators familiar with the fleet and potential confounding factors.

The Group was reminded that CPCs were recommended to report their methodology for estimating bycatch, with the intent for SCRS to evaluate methodologies and suggest best practices. It was suggested that this estimation tool could be used to design a standard method for how bycatch estimates are to be conducted.

It was proposed that the Group recommend the use of this bycatch estimation tool along with LLSIM to address ICCAT objectives for estimating billfish and shark bycatch in the final recommendations and Workplan. The Group was reminded that an estimation of effort will need to be provided for any final recommendation with associated costs. Potential funding of this work with remaining 2021 funds was discussed, and a workplan for 2022 funds will be created.

5. Other matters

Current health conditions related to the Covid pandemic has required that meeting times be reduced for all Working and Species Groups. However, the reduction of WGSAM to only 4 days (4.5 hours per day) was found to be limiting. Not all SCRS paper submissions could be accommodated and not all topics could be given full consideration by the Group (e.g., Work Plan and Terms of reference discussions). Should online SCRS meetings be required in 2022, careful consideration should be given to allow adequate time.

The Group noted that the prioritization of stock assessment methods development needs remains dominated by emerging challenges from case studies based Northern Hemisphere stocks. The Group also noted that this can be partially explained in that prioritization of MSE work identified a number of exclusively Northern stocks (e.g., N-ALB, N-SWO, ATL BFT), which continue to require large portion of the limited time and capacity of the Group. The Group agreed to more actively seek to encourage updates of ongoing research based on case studies also for Southern Hemisphere stocks, and in particular, encourages

the submission of papers the cover topic, such simulation testing of stock assessment methods, CPUE standardization and the development of Harvest Control Rules (HCRs), including, but not limited to, full MSE frameworks.

The Chair presented the tentative terms of reference (TORs) for an investigation into bycatch estimation methodologies to be developed in the second half of 2021 and throughout 2022, that the Group adopted with some changes during the meeting (**Appendix 5**). The Group also discussed the work plan and its budget estimate for 2022, and they are adopted during the meeting (**Appendix 6**).

6. Recommendations

Recommendation with financial implications

1. The Group recommended that tools similar to those presented during the meeting (i.e., SDM/LLSIM and the bycatch estimation tool) be further explored as a means to address the SCRS general needs to estimate bycatch of species such as, but not limited to, billfish and shark. The Group further recommends that this work be carried out using the WGSAM 2021 funds. Therefore, the Group recommends an expert to be contracted to further develop and evaluate these tools.

2. The Group recognizes the difficulty in researching the large amount of information contained within the ICCAT Col. Vol. Sci. Pap. The Group further notes that each SCRS Report must contain keywords associated with the work. The Group therefore recommends that the Collective Volume of Scientific Papers be made searchable based on such fields as author, year, keyword or other helpful index. For that purpose, funds should be made available for hiring a dedicated staff or, as an alternative, issue a short-term contract to ensure indexing all SCRS papers published.

The table below contains the overall funding requests made by the WGSAM for 2022:

Working Group on Stock Assessment Methods	**2022**
Other fisheries related studies (including data recovery, experts, etc.)	
Tool for estimate bycatch of species	**35,000€**
Published SCRS documents to be OCRed to make the contents searchable and indexed	**10,000€**
TOTAL	**45,000€**

General recommendations

1. Two-way communication between managers, scientists and stakeholders is a key part of the MSE process, particularly when a request to develop and test management procedure is being drafted. The Group recognized that this two-way communication between the SCRS and the Commission needs to increase as all SCRS MSEs continue to progress. The Group recommended several ways to increase this two-way communication: (1) ensure that terminology used in MSE communications adheres the tRFMO MSE glossary of term (Anon., 2018); (2) reinstating regular meetings of the Standing Working Group to Enhance Dialogue Between Fisheries Scientists and Managers (SWGSM); (3) creating a greater connection between the ICCAT Secretariat representative and the tRFMO MSE Working Group; and (4) supporting the existing outreach efforts of the ICCAT Secretariat; (5) utilize existing communication and visualization tools such as the Shiny App "SLICK'. Furthermore, the Group recommends that a second, "Executive Summary" version of the interactive MSE visualization tool intended to aid in consultation and decision making (harveststrategies.org; Slick Decision Analysis) be developed that includes only key metrics and graphics essential to the understanding of the MSE results, geared towards a more lay audience.

2. The Group recommends the SCRS routinely apply objective criteria for model plausibility for all ICCAT stock assessments that are intended for management (e.g., TAC) advice. These criteria shall be based on best practice in using model diagnostics for evaluating (1) model convergence, (2) fits to the data, (3) model consistency (e.g., retrospective patterns) and (4) prediction skill, as well as biological plausibility criteria. The Group recommends the model diagnostics applied are similar, but not limited to those

described in Carvalho *et al.,* (2021). The Group noted that key diagnostics, such as residuals run tests, retrospective analysis, and hindcast cross-validation are available in the R package 'ss3diags', within the JABBA modeling framework, as well as 'a4adiags' for the statistical catch-at-age (sca) model FLa4a and that these packages be included in the ICCAT website stock assessment software catalogue to facilitate this process.

3. The Group recommends that SCRS meetings in preparation for stock assessment evaluations routinely include a presentation and discussion of model and the diagnostics of the previous assessment being used to provide management advice. The presentations should identify model uncertainties, biases and/or misspecifications, which should be considered when specifying the uncertainty grid to be submitted at the subsequent stock assessment meeting.

4. The Group recommended that the products created via the 2020/21 funding (Species Distribution Model (SDM) and Longline Simulator (LLSIM), along with the simulated CPUE data sets be included in the ICCAT software catalogue for use in future studies.

5. The Group recommended that the scheduling of the WGSAM meeting concurrent with the Sub-committee on Ecosystems meeting be discontinued. The concurrent timing was an impediment to the progress of both Groups. CPC scientists were forced to attend either one or the other of the meetings despite the relevance of their contributions to each meeting. While the Group did see some utility of having perhaps one overlapping day, the advantages of this option would be dependent on the overlap of the two Groups specified agenda items.

6. The Group noted the importance of having the historical sex information on the conventional tagging database. Such data are usually reported for sharks, but currently it is only available in the ICCAT database for the most recent years. Therefore, the Group recommends that the Secretariat makes a revision of the available historical sex information for inclusion in the conventional tagging database and make it available in the cases where such information was reported.

7. Adoption of the report

The Report of the meeting of the Working Group on Stock Assessment Methods (WGSAM) was adopted. Dr Michael Schirripa thanked the participants and the rapporteurs for their work and collaboration to finalize the report on time. The meeting was adjourned.

References

Anon. 2016. Report for biennial period, 2016-17 Part I (2016) – Vol 2 SCRS.

Anon. 2017a. Rec 17-04 Report for biennial period, 2008-09 Part I (2008) – Vol 2 SCRS.

Anon. 2017b. Report of the 2017 ICCAT Shortfin Mako Data Preparatory meeting (Madrid, Spain 28-31 March 2017). ICCAT Collect. Vol. Sci. Pap. 74(4): 1373-1464.

Anon. 2017c. Report of the 2017 ICCAT Shortfin Mako Stock Assessment meeting (Madrid, Spain 12-16 June 2017). ICCAT Collect. Vol. Sci. Pap. 74(4): 1465-1561.

Anon. 2018. Glossary of terms for harvest strategies, management procedures and management strategy evaluation. https://www.tuna-org.org/Documents/MSEGlossary_tRFMO_MSEWG2018.pdf. The 2018 Joint tuna RFMO Management Strategy Evaluation Working Group Meeting in Seattle, USA, 13-15 June 2018.

Anon. 2019. Report of the 2019 ICCAT Yellowfin Tuna Stock Assessment meeting (Grand-Bassam, Côte d'Ivoire, 8-16 July 2019). ICCAT Collect. Vol. Sci. Pap. 76(6): 344-515.

Carvalho, F., Winker H., Courtney D., Kapur M., Kell L., Cardinale M., Schirripa M., Kitakado T., Yemane D., Piner K. R., Maunder M. N., Taylor I. Wetzel C. R., Doering K., Johnson K. F., and Methot R. D.2021. A cookbook for using model diagnostics in integrated stock assessments. Fisheries Research Volume https://doi.org/10.1016/j.fishres.2021.105959

Kell, L.T., Sharma, R., Kitakado, T., Mosqueira, I., Winker, H., Cardinale, M., Fu, D., 2021. *In press.* Validation of stock assessment models using prediction skill: Is it me or my model talking? ICES J. Mar. Sci.

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

Merino G., Kell L.T., Arrizabalaga H. and Santiago S. 2020. Updated consolidated report for North Atlantic albacore management strategy evaluation. ICCAT Collect. Vol. Sci. Pap. 77(7): 428-461

Walter J., Yokoi H., Satoh K., Matsumoto T., Winker H., Ijurco A.U., and Schirripa, M. 2019. Atlantic bigeye tuna Stock Synthesis projections and Kobe 2 matrices. ICCAT Collect. Vol. Sci. Pap. 75(7), 2283–2300.

Walter J., and Winker H., 2020. Projections to create Kobe 2 Strategy Matrices using the multivariate log-normal approximation for Atlantic yellowfin tuna. ICCAT Collect. Vol. Sci. Pap. 76(6), 725–739.

Winker H., Kerwath S., Merino G., and Ortiz M. 2018. Bayesian State-Space Surplus production model JABBA assessment of Atlantic bigeye tuna (*Thunnus obesus*) stock. ICCAT Collect. Vol. Sci. Pap. 75(7): 2129-2168.

Appendix 1

Agenda

1. Opening, adoption of agenda and meeting arrangements
2. Harvest Control Rules, Limit Reference points and Management Strategy Evaluation
3. Recommended standard diagnostics for stock assessment models
4. CPUE standardization/incorporation of oceanographic and environmental changes into the assessment process
    - 4.1 Results of study on Investigation into decadal changes in swordfish habitat distribution
    - 4.2 Results of study of the addition of a second fishery to a longline simulator
    - 4.3 Using LLSIM to evaluate CPUE methods and estimation of bycatch
5. Other matters
6. Recommendations
7. Adoption of the report and closure

Appendix 2

List of Participants

CONTRACTING PARTIES

ALGERIA
Bennoui, Azzeddine
Chercheur au Centre de Recherche et de Développement pour la Pêche et de l'Aquaculture (CNRDPA), 11 Bouevard Colonel Amirouche, 42415 Bou Ismail, Tipaza
Tel: +213 24 32 64 10, Fax: +213 24 32 64 11, E-Mail: bennoui_azeddine@yahoo.fr

Ferhani, Khadra
Centre National de Recherche et de Développement de la Pêche et de l'Aquaculture (CNRDPA), 11 Boulevard Colonel Amirouche, BP 67, 42415 Tipaza Bou Ismail
Tel: +213 550 735 537, Fax: +213 24 32 64 10, E-Mail: ferhani_khadra@yahoo.fr

Kouadri-Krim, Assia
Sous-Directrice infrastructures, industries et services liés à la pêche, Ministère de la Pêche et des Productions Halieutiques, Direction du développement de la pêche, Route des Quatre Canons, 1600
Tel: +213 558 642 692, Fax: +213 214 33197, E-Mail: assiakrim63@gmail.com

BRAZIL
Leite Mourato, Bruno
Profesor Adjunto, Laboratório de Ciências da Pesca - LabPesca Instituto do Mar - IMar, Universidade Federal de São Paulo - UNIFESP, Rua Carvalho de Mendoça, 144, Encruzilhada, 11070-100 Santos, SP
Tel: +55 1196 765 2711, Fax: +55 11 3714 6273, E-Mail: bruno.mourato@unifesp.br; bruno.pesca@gmail.com; mourato.br@gmail.com

Sant'Ana, Rodrigo
Researcher, Laboratório de Estudos Marinhos Aplicados - LEMA Ecola do Mar, Ciência e Tecnologia - EMCT, Universidade do Vale do Itajaí - UNIVALI, Rua Uruquai, 458 - Bloco E2, Sala 108 - Centro, Itajaí, CEP 88302-901 Santa Catarina Itajaí
Tel: +55 (47) 99627 1868, E-Mail: rsantana@univali.br

CANADA
Gillespie, Kyle
Fisheries and Oceans Canada, St. Andrews Biological Station, Population Ecology Division, 125 Marine Science Drive, St. Andrews, New Brunswick, E5B 0E4
Tel: +1 506 529 5725, Fax: +1 506 529 5862, E-Mail: kyle.gillespie@dfo-mpo.gc.ca

EUROPEAN UNION
Arrizabalaga, Haritz
Principal Investigator, AZTI Marine Research Basque Research and Technology Alliance (BRTA), Herrera Kaia Portualde z/g, 20110 Pasaia, Gipuzkoa, España
Tel: +34 94 657 40 00; +34 667 174 477, Fax: +34 94 300 48 01, E-Mail: harri@azti.es

Biagi, Franco
Senior Expert Marine & Fishery Sciences, Directorate General for Maritime Affairs and Fisheries (DG-Mare) - European Commission, Unit C3: Scientific Advice and data collection, Rue Joseph II, 99, 1049 Brussels, Belgium
Tel: +322 299 4104, E-Mail: franco.biagi@ec.europa.eu

Garibaldi, Fulvio
University of Genoa - Dept. of Earth, Environmente and Life Sciences, Dipartimento si Scienze della Terra, dell'Ambiente e della Vita (DISTAV), Corso Europa, 26, 16132 Genova, Italy
Tel: +39 335 666 0784; +39 010 353 8576, Fax: +39 010 357 888, E-Mail: largepel@unige.it; garibaldi.f@libero.it

Gordoa, Ana
Senior scientist, Centro de Estudios Avanzados de Blanes (CEAB - CSIC), Acc. Cala St. Francesc, 14, 17300 Blanes, Girona, España
Tel: +34 972 336101; +34 666 094 459, E-Mail: gordoa@ceab.csic.es

Mejuto García, Jaime
Ministerio de Economía y Competitividad, Instituto Español de Oceanografía, C.O de A Coruña, Paseo Marítimo Alcalde Francisco Vázquez, 10 - P.O. Box 130, 15001 A Coruña, España
Tel: +34 981 218 151, Fax: +34 981 229 077, E-Mail: jaime.mejuto@ieo.es

Merino, Gorka
AZTI - Tecnalia /Itsas Ikerketa Saila, Herrera Kaia Portualdea z/g, 20100 Pasaia - Gipuzkoa, España
Tel: +34 94 657 4000; +34 664 793 401, Fax: +34 94 300 4801, E-Mail: gmerino@azti.es

Ortiz de Zárate Vidal, Victoria
Investigadora, Ministerio de Ciencia, Innovación y Universidades, Instituto Español de Oceanografía, C.O. de Santander, Promontorio de San Martín s/n, 39004 Santander, Cantabria, España
Tel: +34 942 291 716, Fax: +34 942 27 50 72, E-Mail: victoria.zarate@ieo.es

Pappalardo, Luigi
Scientific Coordinator, OCEANIS SRL, Vie Maritime 59, 84043 Salerno Agropoli, Italy
Tel: +39 081 777 5116; +39 345 689 2473, E-Mail: gistec86@hotmail.com; oceanissrl@gmail.com

Pierucci, Andrea
University of Cagliari, 09126 Cagliari, Italy
Tel: +39 328 651 6025, E-Mail: andrea.pierucci@hotmail.it

Pinto, Cecilia
16132 Genova, Italy
Tel: +39 340 496 6905, E-Mail: cecilia.pinto@edu.unige.it

Rosa, Daniela
PhD Student, Portuguese Institute for the Ocean and Atmosphere, I.P. (IPMA), Av. 5 de Outubro s/n, 8700-305 Olhao, Portugal
Tel: +351 289 700 504, E-Mail: daniela.rosa@ipma.pt

Rueda, Lucía
Ministerio de Ciencia, Innovación y Universidades, Instituto Español de Oceanografía Málaga, Puerto pesquero s/n, 29640 Fuengirola Málaga, España
Tel: +34 952 197 124, E-Mail: lucia.rueda@ieo.es

Thasitis, Ioannis
Department of Fisheries and Marine Research, 101 Vithleem Street, 2033 Nicosia, Cyprus
Tel: +35722807840, Fax: +35722 775 955, E-Mail: ithasitis@dfmr.moa.gov.cy

Winker, Henning
Joint Research Centre - European Commission, Ispra, Italy, TP 051, Via Enrico Fermi 2749, 21027 Ispra, VA, Italy
Tel: +27 725 126 558, E-Mail: henning.winker@ec.europa.eu

HONDURAS
Cardona Valle, Fidelia Nathaly
Colonia Lomo Linda Norte, Avenida FAO, edificio SENASA, 11101 Tegucigalpa Francisco Morazán
Tel: +504 877 88713, E-Mail: investigacion.dgpa@gmail.com

JAPAN
Kitakado, Toshihide
Professor, Faculty of Marine Science, Tokyo University of Marine Science and Technology, Department of Marine Biosciences, 4-5-7 Konan, Minato, Tokyo 108-8477
Tel: +81 3 5463 0568, Fax: +81 3 5463 0568, E-Mail: kitakado@kaiyodai.ac.jp; toshihide.kitakado@gmail.com

MOROCCO
Serghini, Mansour
Institut national de recherche halieutique, Route Sidi Abderrahmane Club équestre Ould Jmel, 20000 Casablanca
Tel: 0660 455 363, E-Mail: serghini@inrh.ma; serghini2002@yahoo.com; serghinimansour@gmail.com

Yassir, Anass
Ingénieur Statisticien, Institut National de Recherche Halieutique, Route Sidi Abderrahmane Club équestre Ould Jmel, 20000 Casablanca
Tel: +212 618 392 738, E-Mail: yassiranas.insea@gmail.com

NORWAY
Nottestad, Leif
Principal Scientist, Institute of Marine Research Research Group on Pelagic Fish, P.O. Box 1870 Nordnesgaten, 33, 5817 Bergen, Hordaland county
Tel: +47 5 99 22 70 25, Fax: +47 55 23 86 87, E-Mail: leif.nottestad@hi.no

SENEGAL
Ba, Kamarel
Docteur en Sciences halieutiques et modélisation, Ministère de l'Agriculture et de l'Equipment Rural, Institut Senegalais de Recherches Agricoles (ISRA), Centre de Recherches Oceanographiques de Dakar Thiaroye (CRODT), Pôle de Recherches de Hann, Route du Front de Terre, 2241 Dakar
Tel: +221 77 650 52 32, Fax: +221 338 328 262, E-Mail: kamarel2@hotmail.com

Sow, Fambaye Ngom
Chercheur Biologiste des Pêches, Centre de Recherches Océanographiques de Dakar Thiaroye, CRODT/ISRA, LNERV - Route du Front de Terre - BP 2241, Dakar
Tel: +221 3 0108 1104; +221 77 502 67 79, Fax: +221 33 832 8262, E-Mail: famngom@yahoo.com; ngomfambaye2015@gmail.com

TUNISIA
Zarrad, Rafik
Chercheur, Institut National des Sciences et Technologies de la Mer (INSTM), BP 138 Ezzahra, Mahdia 5199
Tel: +216 73 688 604; +216 972 92111, Fax: +216 73 688 602, E-Mail: rafik.zarrad@gmail.com

UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND
De Oliveira, José
The Centre for Environment, Fisheries and Aquaculture Science, CEFAS, Pakefield Road, Lowestoft - Suffolk, IP19 8JX
Tel: +44 150 252 7727, E-Mail: jose.deoliveira@cefas.co.uk

Walker, Nicola
Centre for Environment, Fisheries and Aquaculture Science (CEFAS), Lowesfolk Suffolk NR33 0HT
Tel: +44 1502 524450, E-Mail: nicola.walker@cefas.co.uk

UNITED STATES
Babcock, Elizabeth
Professor, Rosenstiel School of Marine and Atmospheric Science, University of Miami, Department of Marine Biology and Ecology, 4600 Rickenbacker Causeway, Miami Florida 33149
Tel: +1 305 421 4852, Fax: +1 305 421 4600, E-Mail: ebabcock@rsmas.miami.edu

Brown, Craig A.
Chief, Highly Migratory Species Branch, Sustainable Fisheries Division, NOAA Fisheries Southeast Fisheries Science Center, 75 Virginia Beach Drive, Miami, Florida 33149
Tel: +1 305 586 6589, Fax: +1 305 361 4562, E-Mail: craig.brown@noaa.gov

Carvalho, Felipe
NOAA Regional Center, 1845 Wasp Boulevard, Honolulu Hawaii 96818
Tel: +1808 397 9210, E-Mail: felipe.carvalho@noaa.gov

Cortés, Enric
Research Fishery Biologist, NOAA-Fisheries, Southeast Fisheries Science Center, Panama City Laboratory, 3500 Delwood Beach Road, Panama City, Florida
Tel: +1 850 234 6541; +1 850 814 4216, Fax: +1 850 235 3559, E-Mail: enric.cortes@noaa.gov

Courtney, Dean
Research Fishery Biologist, NOAA/NMFS/SEFSC Panama City Laboratory, 3500 Delwood Beach Road, Panama City Beach Florida 32408
Tel: +1 850 234 6541, E-Mail: dean.courtney@noaa.gov

Cummings, Nancie
NOAA, NMFS, Southeast Fisheries Science Center, Sustainable Fisheries Division, 75 Virginia Beach Drive, Miami, Florida 33143
Tel: +1 305 361 4234, Fax: +1 305 361 4299, E-Mail: nancie.cummings@noaa.gov

Schirripa, Michael
Research Fisheries Biologist, NOAA Fisheries, Southeast Fisheries Science Center, 75 Virginia Beach Drive, Miami, Florida 33149
Tel: +1 305 445 3130; +1 786 400 0649, Fax: +1 305 361 4562, E-Mail: michael.schirripa@noaa.gov

Scott, Rebecca
University of South Florida College of Marine Science, 830 1st St South, Florida 33701
Tel: +1 484 707 2461, E-Mail: rscott13@mail.usf.edu

Walter, John
Research Fishery Biologist, NOAA Fisheries, Southeast Fisheries Center, Sustainable Fisheries Division, 75 Virginia Beach Drive, Miami, Florida 33149
Tel: +305 365 4114; +1 804 815 0881, Fax: +1 305 361 4562, E-Mail: john.f.walter@noaa.gov

Zhang, Xinsheng
NOAA/NMFS/SEFSC, 3500 Delwood Beach Rd., Florida 32408
Tel: +1 850 234 6541 ext. 264, Fax: +1 850 235 3559, E-Mail: Xinsheng.Zhang@noaa.gov; Xinsheng.Zhang0115@gmail.com

OBSERVERS FROM NON-GOVERNMENTAL ORGANIZATIONS

THE OCEAN FOUNDATION
Miller, Shana
The Ocean Foundation, 1320 19th St., NW, 5th Floor, Washington, DC 20036, United States
Tel: +1 631 671 1530, E-Mail: smiller@oceanfdn.org

SCRS CHAIRMAN
Melvin, Gary
SCRS Chairman, St. Andrews Biological Station - Fisheries and Oceans Canada, Department of Fisheries and Oceans, 285 Water Street, St. Andrews, New Brunswick E5B 1B8, Canada
Tel: +1 506 652 95783, E-Mail: gary.d.melvin@gmail.com; gary.melvin@dfo-mpo.gc.ca

SCRS VICE-CHAIRMAN
Coelho, Rui
Researcher, SCRS Vice-Chairman, Portuguese Institute for the Ocean and Atmosphere, I.P. (IPMA), Avenida 5 de Outubro, s/n, 8700-305 Olhão, Portugal
Tel: +351 289 700 504, E-Mail: rpcoelho@ipma.pt

INVITED EXPERT
Goodyear, Philip
1214 North Lakeshore Drive, Niceville Florida 32578, United States
Tel: +1 850 897 2666, Fax: +1 850 897 2666, E-Mail: phil_goodyear@msn.com

*****

ICCAT Secretariat
C/ Corazón de María 8 – 6th floor, 28002 Madrid – Spain
Tel: +34 91 416 56 00; Fax: +34 91 415 26 12; E-mail: info@iccat.int

Manel, Camille Jean Pierre
Neves dos Santos, Miguel
Ortiz, Mauricio
Kimoto, Ai
Peña, Esther

Appendix 3

List of Papers and Presentations

Number	Title	Authors
SCRS/2021/048	Development of new model fisheries for simulating longline catch data with LLSIM	Goodyear C.P.
SCRS/2021/049	Investigations into spatiotemporal patterns in swordfish habitat distributions	Goodyear C.P.
SCRS/2021/050	Plausibility and uncertainty of basic data and parameter selection on stock assessments: a review of some input data used in the 2017 assessment of shortfin mako (*Isurus oxyrinchus*) of the Northern Atlantic stock	Mejuto J., Fernández-Costa J., Ramos-Cartelle A., and Carroceda A.
SCRS/2021/075	A preliminary analysis of spatiotemporal patterns in swordfish habitat distributions	Schirripa M.J., Forrestal F., Goodyear C.P., Abascal F., Bubley W., Coelho R., and Hanke A.
SCRS/2021/078	Testing a bycatch estimation tool using simulated blue marlin longline data	Babcock E.A., and Goodyear C.P.
SCRS/P/2021/020	Ensemble weighting and projections using model validation and prediction skill with ss3diags	Winker H., Carvalho F., Cardinale M., and Kell L.
SCRS/P/2021/022	A Cookbook for Using Model Diagnostics in Integrated Stock Assessments	Carvalho F., Winker H., Courtney D., Kapur M., Kell L., Cardinale M., Schirripa M., Kitakado T., Yemane D., Piner K.R, Maunder M.N., Taylor I., Wetzel C.R., Doering K., Johnson K.F., and Methot R.D.
SCRS/P/2021/023	Summary on North Atlantic ALB MSE	Arrizabalaga H., and Merino G.
SCRS/P/2021/025	Atlantic tropical tuna MSE	Merino G., Die D., Urtizberea A., and Laborda A.
SCRS/P/2021/027	Atlantic bluefin tuna MSE progress	Walter J.
SCRS/P/2021/028	North Atlantic Swordfish MSE update	Gillespie K.

Appendix 4

SCRS Document and Presentations Abstracts as provided by the authors

SCRS/2021/048 Two idealized fishing fleets were developed to augment LLSIM simulations of complex longline data. They were patterned after Brazilian and Japanese data in the ICCAT Task ii database. These data provided the numbers of hooks fished in 5° latitude-longitude squares for 1958- 2018. Each fleet had shallow, intermediate, or deeply-fishing gears. These simulated gears were assigned to individual sets based on previously-reported fishing patterns or the species composition of fleet catches in the ICCAT database. Fishing depths for each gear were adopted from an earlier study of the US fishery. The number of sets in each square were estimated from total hooks using the number of hooks per set for the gear. The sets for each fishery in each 5° square were randomly assigned to habitable 1° LLSIM cells (no land). The resulting file provided the monthly number of sets for each fleet by gear within each 1 square from 1958 to 2018. When used in combination with Swordfish and Blue Marlin species-distribution-model inputs, these fisheries will allow investigators to simulate catches with three-fishery, two species runs of the LLSIM software.

SCRS/2021/049 The time-varying three-dimensional distribution of the Swordfish pelagic habitat was previously studied using a species distribution model (SDM) that integrated multiple habitat features using habitat suitability methods. The SDM predicted the 1958-2019 monthly relative abundances during hours of daylight and darkness in each of 46 depth layers from the surface down to about 2 km depth. Results suggested a long-term poleward shift in habitat relative densities. However, model-predicted distributions exhibited complex spaciotemporal patterns that confounded analysis. The current study developed methods to define the physical boundaries of pelagic habitat envelopes that contained specified fractions of the population. These habitat envelopes allowed important features such as habitat volume and density to be tracked in time and space along with the average locations of individuals in the population. These methods were implemented with two computer programs, “surfaces” and “envelopes”, that accompany this report. Predicted Swordfish distributions exhibited large seasonal fluctuations with a small but persistent annual shift southward that began in the mid-1970s. Analyses found that, among the environmental variables included, temperature was the dominant factor contributing to the complex spatial effects.

SCRS/2021/050 Three key-elements for the case of shortfin mako used in 2017 assessment of North Atlantic stock are reviewed. The catch scenarios implemented indicate that historical T1 considered in base case scenario (C1) have greatly underestimated the level of catches during several of the initial decades, taking into consideration the history of the fisheries, fleet’s capacity and fishing effort by fleet. A hypothetical catch scenario also used (C2) probably overestimated in an important amount the catch levels of some fleets and fleets combined for the most recent period of that series. The review of CPUE series suggests that there may be qualitative and/or quantitative limitations in some of them which would likely affect some series being considered as indicators of abundance. Some key biological parameters considered in the assessment are also reviewed and discussed, such as the growth model implemented and the age of first reproduction of the females, within a context compared to other studies, preliminary estimations from tagging-recapture data and those parameters applied in other stock of the same species, and in other species from the same family.

SCRS/2021/075 A species distribution model (SDM) for swordfish that was in the development stage has been finalized. The model used detailed biological and oceanographic data to define the spatial distribution of Swordfish. The SDM adequately predicted Swordfish habitat (and thus fish) distributions such that it was found suitable for investigations into the spatiotemporal distribution of habitat. Results of this preliminary investigation supports the current hypothesized stock boundaries between the north and south Atlantic stocks used for management. Both the north and south Atlantic may be experiencing an expansion of habitat. This could result in decreased density of swordfish into a larger area and/or change MSY production metrics. A more detailed examination of this possibility is recommended.

SCRS/2021/078 The species distribution model and longline simulate developed by Goodyear et al. (2017) was used to generate simulated longline sets for three fleets, catching both swordfish and blue marlin. These simulated data were used to test the effectiveness of a bycatch estimation tool in development. The tool allows for semi-automated model selection using information criteria to select the best set of predictor variables, and using cross validation to choose between Tweedie, negative binomial, delta-lognormal, and delta-gamma models. The simulated data allowed for a nuanced evaluation of the model decisions, such as whether to use trips or sets as a sample unit. As new functions are added to the bycatch estimation tool, the simulated data will continue to be key to adequate testing.

SCRS/P/2021/020 presented practical applications to ensemble weighting and projections using model validation and prediction skill. Assessing stock status is typically associated with substantial structural uncertainty, which may not be captured adequately by a single 'best' model. To address this, ensemble modelling is increasingly advocated as a potentially more risk-adverse approach by representing uncertainty by a range alternative stock assessment models that can vary in terms model structure, parameterization and data use. Although arguably attractive, ensemble modelling faces three challenges pertaining to: (1) how to best propagate both structural uncertainty and estimation error, (2) on which basis should models be included in the 'ensemble composition' and (3) how to assign relative weights to the models 'ensemble weighting' . Here we focus on some practical aspects of ensemble weighting with Stock Synthesis using new R package 'ss3diags' (github.com/JABBA-model/ss3diags). We introduce new delta-multivariate-lognormal (delta-MVLN) Monte-Carlo function as a resampling approach to combine model outcomes and future projections under any weighting scheme, while also accounting for within model uncertainty. For model selection and weighting, we propose the following four properties as objective criteria for evaluating the plausibility of a model: (1) model convergence, (2) fit to the data, (3) model consistency, and (4) prediction skill. We suggest that the here presented diagnostic tools for evaluating these criteria are widely applicable across modelling platforms beyond ensemble weighting applications with Stock Synthesis.

SCRS/P/2021/022 Integrated analysis has increasingly been the preferred approach for conducting stock assessments and providing the basis for management advice for fish and invertebrate stocks around the world. Many decisions are required when developing integrated stock assessments. For example, the analyst needs to decide whether the model fits the data, if the optimization was successful, if estimates are consistent retrospectively, and if the model is suitable to predict future stock responses to fishing. This study provides practical guidelines for implementing selected diagnostic tools that can assist analysts in identifying problems with model specifications and alternatives that can be explored to minimize or eliminate such problems. Emphasis is placed on reviewing the implementation and interpretation of contemporary model diagnostic tools. We first describe each diagnostic approach and its utility. We then proceed by providing a "cookbook recipe" on how to implement each of the diagnostics, together with an interpretation of the results, using two worked examples of integrated stock assessments with Stock Synthesis. Further, we provide a conceptual flow chart that lays out a generic process of model development and selection using the presented model diagnostics. Based on this, we propose the following four properties as objective criteria for evaluating the plausibility of a model: (1) model convergence, (2) fit to the data, (3) model consistency, and (4) prediction skill. It would greatly benefit the stock assessment community if the next generation of stock assessment models could include the diagnostic tests presented in this study as a set of open source tools.

SCRS/P/2021/023 provided a summary on the North Atlantic ALB MSE over last few years. Definition of the list of performance statistics was agreed with Panel 2 in 2016, prior to the adoption of the HCR in 2017. However, mostly 4 performance metrics are used, in the form of radar plots, to illustrate the performance across the reference case, which includes uncertainty around biology (M, steepness), data (size, cpue, tagging) and modelling options (time range and catchability trends). The 132 Oms are equally weighted. The code was peer reviewed in 2018 and comments were addressed in 2019. Several HCR variants (as specified in Rec 17-04) have been tested, the SCRS advice incorporates the elements to adopt a full MP, and feedback to Panel 2 has been provided on proposed indicators to detect exceptional circumstances. The primary document for the ALB MSE is the ALB MSE consolidated report (SCRS/2020/153Rev).

SCRS/P/2021/025 provided an update on the developments of the Atlantic tropical tuna MSE in relation to bigeye, yellowfin and Eastern skipjack stocks. In 2021, the project has re-started with a review of the main uncertainties of tropical tuna fisheries with the aim of defining the axes of uncertainty of the MSE, which will be used to condition Operating Models.

SCRS/P/2021/027 detailed recent progress for the Atlantic Bluefin tuna MSE and the work plan for 2021 and 2022. The presentation addressed recent developments including interim adoption of the reference grid, plausibility weighting of the factor levels within the grid, development tuning and reconditioning. The BFT MSE has six candidate management procedure (CMP) development teams working mainly empirical (or index based) procedures with model based CMPs. It is a process to both refine individual CMPs and to evaluate and vet their performance. It is a critical step in the evolutionary process of CMP development and in the process of narrowing down the number of CMPs to a few top-performing ones to present to the Commission and Panel 2 in November.

SCRS/P/2021/028 provided an overview of recent progress and ongoing work related to the North Atlantic swordfish MSE. The initial OM grid had seven axes of uncertainty (natural mortality, recruitment deviation, steepness, CPUE CV, effective sample size of the length comps, catchability increase, and environmental modulation of catchability among fleets). Some modifications were made in 2020 and 2021. CMP development is underway and will continue through 2021. In early 2021, the MSE code began external peer-review. N-SWO MSE sub-groups are developing candidate exceptional circumstances and performance metrics which will be considered by the SWO species group, the SCRS, and PA4 later in 2021.

Appendix 5

Terms of Reference for Investigation into Bycatch Estimation Methodologies

1. Background and Objectives

The complete and total accounting of the bycatch of non-targeted species within the ICCAT arena is becoming an increasingly important source of concern for the ICCAT Commission. The bycatch of species such as billfish (Blue Marlin, White Marlin and Round-scale Spearfish) and several species of Sharks are not always well documented, yet account for a non-negligible percentage of the total fish kill associated with the fishing operations. Not properly accounting for this bycatch, either kept or discarded, in the assessment process can lead to errors and/or biases in the management advice provided to the Commission. Recognizing the importance of an accurate accounting of bycatch, the Commission established Rec. 19-05 which states in the following paragraphs:

1. CPCs shall provide their estimates of total live and dead discards of blue marlin, white marlin/roundscale spearfish, based on fishing logbooks, landing declarations, or equivalent document for the sport/recreational fisheries, as well as scientific observer reports, as part of their Task I and II data submission to support the stock assessment process.

2. No later than 2020, CPCs shall present to the SCRS the statistical methodology used to estimate dead and live discards. CPCs with artisanal and small-scale fisheries shall also provide information about their data collection programs.

3. The SCRS shall review these methodologies and if it determines that a methodology is not scientifically sound, the SCRS shall provide relevant feedback to the CPCs in question to improve the methodologies.

4. The SCRS shall evaluate the completeness of Task I and II data submissions, including estimates of total dead and live discards, and determine the feasibility of estimating fishing mortalities by industrial fisheries (including longline and purse seine), artisanal fisheries and recreational fisheries. If after conducting such evaluation, the SCRS determines that significant gaps in data reporting exist, the SCRS should explore approaches to estimate the level of unreported catches to include in future stock assessments in order to enhance the basis on which to provide management advice to the Commission.

At the 2021 meeting of the Working Group on Stock Assessment methods (WGSAM) the Group was presented with the products of the work funded under the 2020 contract entitled "Investigation in decadal changes in Swordfish habitat distribution and the furthering of a longline simulator". The suit of products included a Species Distribution Model (SDM) for Atlantic Swordfish and the addition of two more fishing fleets to the longline simulator. The Group was also presented with a demonstration of how this suite of models, combined with a bycatch estimation package, can be used to test the efficacy of various bycatch estimation methods. The preliminary results of the bycatch estimation package was a proof of concept as it was able to successfully estimate the simulated bycatch values from the SDM/LLSIM simulated data.

The Group recommended that the bycatch estimation package represented a valuable way forward to the estimation of billfish bycatch. The Group recommended that tools similar to those presented during the meeting (i.e. SDM/LLSIM and the bycatch estimation tool) be further explored as a means to address the SCRS general needs to estimate bycatch of species such as, but not limited to, billfish and shark. The Group further recommends that this work be carried out using the WGSAM 2021 funds. Therefore, it is recommended an expert to be contracted to further develop and evaluate these tools.

2. Contractor Tasks

Development and testing of a bycatch estimation package that is suitable for generalized use by ICCAT CPCs for the purposes of bycatch estimation.

A statistical vetting of the above-mentioned package, via simulation testing and using LLSIM platform, to quantify the efficacy of the methods. This testing is to be conducted to reflect current, realistic scenarios of CPC data collection programs (i.e. percent observer coverage, spatiotemporal coverage, etc.).

The contractor will evaluate at least two CPCs bycatch estimation methodology via simulated data to determine if that method is scientifically sound. Bycatch estimates using the CPCs method will be compared to the bycatch estimates obtained using the contractor's bycatch estimation package in an effort to elucidate the most accurate approaches.

Appendix 6

Work Plan for 2022

1. Evaluation of the products provided by the bycatch estimation methodology contract

2. Development of advice and/or guidelines on bycatch estimation

3. Report on a review of the practices for constructing the stock assessment uncertainty grid in terms of, but not limited to, grid size, parameter selection and range, hypothesis and model plausibility weighting.