Collect. Vol. Sci. Pap. ICCAT, 82(2), SCRS/2025/128: 1-29 (2025)

STOCK ASSESSMENT OF THE SOUTH ATLANTIC SHORTFIN MAKO SHARK, USING BAYESIAN SURPLUS PRODUCTION MODELS (JABBA) AND LARGE GRID MODEL ENSEMBLES

Rui Coelho[1,2,*]

SUMMARY

Bayesian Surplus Production Models were fitted to South Atlantic shortfin mako using JABBA (Just Another Bayesian Biomass Assessment). Four base models were constructed with combinations of base vs. low productivity and reported vs. estimated catches. Models were checked for goodness of fit and validated, and a sensitivity analysis was conducted. A large model grid (500 models) was run by randomly selecting priors from distributions built from the plausible and agreed limits for their values, and alternatively using each of the two catch scenarios. Stock status from these four main models ranged depending on the productivity and the catch type used. Stock status for the large grid ensemble was weighted in two alternative ways (equal-weighting and DIC-weighting) and resulted in a stock status of overfished and subject to overfishing ($B<B_{MSY}$ and $F>F_{MSY}$).

RÉSUMÉ

*Des modèles bayésiens de production excédentaire ont été ajustés au requin-taupe bleu de l'Atlantique Sud avec JABBA (*Just Another Bayesian Biomass Assessment*). Quatre modèles de base ont été construits avec des combinaisons de productivité de base par rapport à faible productivité et de captures déclarées par rapport aux captures estimées. L'adéquation des modèles a été vérifiée et validée, et une analyse de sensibilité a été réalisée. Une grande grille de modèles (500 modèles) a été exécutée en sélectionnant de manière aléatoire des distributions a priori à partir de distributions construites à partir des limites plausibles et convenues pour leurs valeurs, et en utilisant alternativement chacun des deux scénarios de capture. L'état des stocks selon ces quatre modèles principaux varie en fonction de la productivité et du type de capture utilisés. L'état des stocks pour l'ensemble de la grande grille a été pondéré de deux manières différentes (pondération égale et pondération DIC) et a abouti à un état des stocks surexploités et victimes de surpêche ($B<B_{PME}$ et $F>F_{PME}$).*

RESUMEN

Se ajustaron modelos bayesianos de producción excedente al marrajo dientuso del Atlántico sur utilizando JABBA (Just Another Bayesian Biomass Assessment). Se construyeron cuatro modelos base con combinaciones de productividad base frente a baja y de capturas declaradas frente a estimadas. Se comprobó la bondad del ajuste y la validación de los modelos, y se realizó un análisis de sensibilidad. Se ejecutaron un conjunto de modelos de matriz grande (500 modelos) seleccionando aleatoriamente los valores de distribución previa a partir de distribuciones elaboradas basándose en los límites plausibles y acordados para sus valores, y utilizando alternativamente cada uno de los dos escenarios de capturas. El estado del stock, partiendo de estos cuatro modelos principales, variaba en función de la productividad y del tipo de captura utilizado. El estado del stock para el conjunto de modelos de matriz grande se ponderó de dos formas alternativas (ponderación igual y ponderación DIC) y dio como resultado un estado del stock de sobrepesca y sujeto a sobrepesca ($B<B_{MSY}$ & $F>F_{MSY}$).

KEYWORDS

Bayesian statistics, ensemble model grids, shortfin mako, stock assessment, surplus production models

[1] Instituto Português do Mar e da Atmosfera (IPMA, I.P.), Av. 5 de Outubro s/n, 8700-305 Olhão, Portugal.
[2] Centro de Ciências do Mar (CCMAR), Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal.
* Corresponding author: Rui Coelho (rpcoelho@ipma.pt).

1. Introduction

The shortfin mako is a widely distributed pelagic shark captured mostly as bycatch in oceanic pelagic fisheries worldwide, including in the South Atlantic Ocean, under the jurisdiction area of ICCAT.

Previous assessments have been carried out by ICCAT for the South Atlantic shortfin mako. The last assessment was carried out in 2017, with the stock status given by production models (BSP2JAGS) and data-limited catch only methods (CMSY). Using those models, the combined probability of the stock being overfished was 32.5% and that of experiencing overfishing was 41.9% (ICCAT, 2017). The SCRS considered in general those assessment results to be highly uncertain, owing to the conflict between catch and CPUE data, with the CPUE showing a similar trend to that of catches which was problematic especially for production models (ICCAT, 2017).

ICCAT has also conducted Ecological Risk Assessments for sharks, with the shortfin mako shark receiving one of the highest vulnerability ranking, mainly due to its low productivity and high susceptibility to pelagic longline gear (Cortes *et al*., 2015).

The purpose of this paper is to present the JABBA stock assessment configuration, inputs and results for the ICCAT South Atlantic shortfin mako shark, to support the management advice for this pelagic shark species to ICCAT.

2. Material and methods

In March 2025, an ICCAT Sharks Species Group (Sharks-SG) data-preparatory meeting was held, where the group agreed on the technical specifications that should be used for the stock assessment, and that were followed in this work and presented here. The group also noted that a grid approach over several variables would be appropriate, for exploring different life history and productivity options, CPUE scenarios and catch histories.

In terms of life history, in addition to the data preparation meeting there was some additional intersessional work, and some final work during the stock assessment meetings. The final results with the biology and life history is provided in Cortés (2025).

2.1 SMA nominal catches and catch reconstructions

Two options were discussed by the Sharks-SG and considered for the assessment. One is based on the nominal SMA catches that are reported by the CPCs to ICCAT (Option A). The other (Option B) is based on a catch reconstruction, following one of the catch options presented in Mejuto *et al*. (2021), namely C3_6, which was a mixed scenario, assuming ratio C1/C2 (from previous assessment, with C2 based on Coelho and Rosa, 2017) of underreported catches between 1950-1985, and full reporting 1986-2015. This series is described in full detail in Mejuto *et al*. (2021).

The two series have some differences, but those are mostly in the initial period of the catches, until approximately the mid-1990s (**Figure 1**). This is consistent with the fact that it was since that period that the shark catches recording and reporting became more consistent for many CPCs, so there is expectation that in the earlier period the differences between reported and estimated are larger. In terms of absolute scale, the values between the two series are relatively similar in the more recent period, with the reported series achieving its highest value in 2014 with 3274 tons, and the estimated series achieving its maximum value in 2011 with 3893 tons (**Figure 1**).

It is also noted that after 2015 the only series available was the reported catches (option A) as the estimation method ends in 2015. As such, when using option B (estimated time-series), for the period 2015-2023 the reported catches were used.

2.2 Life history and demographic analysis

The decisions on the life history parameters and population dynamics values to use were based on the work of Cortés (2025). Those were initially discussed and presented at the data-preparatory meeting, followed by intersessional work and developments, and with the final values were presented at the beginning of the assessment meeting. The final work with the final values is presented in Cortés (2025). The values that are most relevant for the JABBA production models and used in these assessment models are summarized here:

- r prior (intrinsic population growth rate): values of 0.049 and 0.114.
- Bmsy/K prior (used for the shape parameter of the production function): values of 0.578 and 0.637.

2.3 *Standardized CPUEs series*

The CPUEs series that were available were those either available and/or presented at the Sharks SG meeting or revised after that meeting. Those included series from the following CPCs fleets (**Figure 2**):

- Japan LL – time block 1 (1994-2011)
- Japan LL – time block 2 (2012-2023, with a gap in 2020-2021)
- Chinese Taipei LL (2007-2022)
- Brazil/Uruguay combined LL index (1978-2022)
- South Africa (2000-2023)
- Spain (1990-2023)

The CVs for each series in the models were those produced in the CPUE standardization analysis. In the initial analysis, for cases where the CPUE CV values were lower than 0.2, then a minimum CV of 0.2 was defined. This allowed some flexibility in fitting the models to the CPUEs.

The first phase of the model building process was to carry out a sensitivity analysis for determining the influence of each CPUE in the overall stock trends, and look for any major conflicts between the CPUEs, and between CPUEs and the catch history. This was done by running models using only one CPUE at a time as provided in the CPUE papers. This process helped the initial stage of decision making on which CPUES should be suggested for used in the base case models, and also on needs to introduce time-blocks in the various CPUEs, to take into account different catchability in different time periods that might not be fully taken into account during the CPUE standardization process.

2.4 *Stock assessment*

2.4.1 Assessment platform

The assessment models were implemented in JABBA, a Bayesian state-space surplus production model framework (Winker *et al*., 2018). JABBA is implemented in R and available from: github.com/jabbamodel/JABBA.

JABBA is a flexible Bayesian stock assessment modeling framework with various options, that include: 1) automatic fitting of multiple CPUE time series and associated standard errors, 2) estimating or fixing the process variance, 3) optional estimation of additional observation variance for individual or grouped CPUE time series 4) specifying the production function, i.e., Fox, Schaefer or Pella-Tomlinson, this last one by setting the inflection point from Bmsy/K and converting it into the shape parameter m, 5) setting priors for various parameters, including r and K, that can range from more to less informative depending on the confidence in the previously available information, 6) model diagnostics and goodness-of-fit features with associated tests and plots (e.g. residuals run tests, hindcast and retrospective analysis) and, 7) projections for constant catches (TACs) in the future to achieve management objectives over certain timeframes.

JABBA is implemented in R (R Core Team, 2025) and uses the JAGS software (Plummer, 2003) to estimate model parameters in a Bayesian framework, by means of Markov Chains Monte Carlo (MCMC) simulation. JAGS is executed from R using the library “r2jags” (Su and Yajima, 2012).

All analysis in this paper was conducted using R v.4.4.3. (R Core Team, 2025). Some additional libraries were used for manipulating and plotting data, including libraries “reshape” (Wickham, 2007), “doBy” (Højsgaard and Halekoh, 2023), “tidyr” (Wickham *et al*., 2023), “tidyverse” (Wickham *et al*., 2019), “ggplot2” (Wickham, 2016), “dplyr” (Wickham *et al*., 2023), “gridExtra” (Auguie, 2017) and “cowplot” (Wilke, 2024).

2.4.2 Stock assessment model specifications

The model specifications were based on an ensemble grid of 4 main models, given the current uncertainty that is associated with the shortfin mako sharks, and following the decisions of the Sharks Species Group. Those 4 models accounted for the following general combinations:

- Base productivity with estimated catches (considered as a general base case for this work)

- Base productivity with reported catches
- Low productivity with estimated catches
- Low productivity with reported catches

The *r* (intrinsic population growth rate) values used as priors came from by Cortés (2025). Specifically, the values used for the base model scenario were r = 0.114, while for the low productivity scenario were r = 0.049. In both scenarios, those values were used as the center of lognormal priors with a CV of 0.2.

The production function used was a Pella-Tomlinson model, with the shape parameter (*m*) estimated based on the life history and population dynamics as described and calculated by Cortés (2025). In JABBA this was inputted in the models as informative priors of the inflection point at Bmsy/K, with values of 0.578 (base case) and 0.637 (low productivity) and with CVs set at 0.3, rather than using fixed values. This allowed some further variability to be included, given the uncertainties associated with this parameter.

As described, two time-series of catches were available, namely one with the data as reported to ICCAT and the other with estimated catches. Those time series have different values in terms of magnitude, so there are some differences in the estimations of absolute values such as B0 and MSY. Nonetheless, the Sharks Species Group agreed to provide based on both catch histories, and as such both were run in the final model ensemble. In all cases, the catches were inputted in JABBA with an associated CV of 0.2, therefore allowing some deviation from the observer catches to reflect the likelihood that the catches may not be accurately estimated, recorded and reported to ICCAT.

In the model specifications, the *K* prior (carrying capacity) was kept as vaguely informative, given the lack of prior knowledge on these values and to allow for more emphasis to be put in the *r* parameter), which is derived from biological data. Specifically, the *K* prior used the default settings of JABBA, namely the use of a lognormal prior with a large CV (100%) and a central value corresponding to 8 times the maximum total catch. This is consistent with other types of models, such as the approach used in catch-MSY (Martell and Froese, 2013).

For all models the same initial depletion (B_{1971}/K) was considered, using a prior with beta distribution with a mean of 0.9 and CV of 10%. Sensitivities were run by setting this value to higher initial depletion levels.

Catchability parameters were formulated as uninformative priors and the CPUEs were scaled externally by their respective means before inputting into the models.

The process error was defined by an uninformative inverse-gamma distribution with the shape and scaling parameters set at 4 and 0.01 (JABBA default). Sensitivities were carried out by setting both those parameters at 0.001 (Gelman, 2006; Mourato *et al*., 2023), and by fixing the sigma of the process error to CVs of 5% and 10%.

In addition to the CPUE variance associated with the data, the base case grid models configuration allowed the internal estimation of additional observation variance for each CPUE, allowing therefore for a larger divergence between the observed and model predicted CPUEs. A sensitivity analysis was carried out by disabling this process.

For the parameter estimation in the Bayesian models, various distinct specifications were used in each phase. For the development phase and exploration of the models, and to prioritize computing efficiency and speed, we used 3 MCMC chains with 20,000 iterations each, a burn-in period of 4,000 and a thinning rate of 4, resulting in a total of 12,000 posterior samples. For the final 4 base case grid models we used 3 MCMC chains with 40,000 iterations each, a burn-in period of 5,000 and a thinning rate of 5, resulting in a total of 21,000 posterior samples. This configuration allowed for a thorough exploration of the posterior distributions of the main base models, ensured convergence of the model key parameters, and provided effective sample sizes well above recommended thresholds (ESS > 1,000 for all parameters). Finally, for the large ensemble grid of models using random input values from pre-determined distributions, we used 3 MCMC chains, with 20,000 iterations each, a burn-in period of 5,000 and a thinning rate of 5, resulting in a total of 9,000 posterior samples. This configuration was chosen to balance computational efficiency with posterior stability in high-performance computing environments.

2.4.3 Model diagnostics

Basic diagnostics of model convergence included MCMC trace-plots and other statistics (Heidelberger and Welch, 1992; Geweke, 1992; Gelman and Rubin, 1992) implemented in the "CODA" package (Plummer *et al*., 2006).

To evaluate the CPUE fits, the model predicted CPUE indices were compared to the observed CPUE. Additionally, residual plots were used to examine the residuals of observed versus predicted CPUE indices for all fleets and boxplots with the median and quantiles of all residuals for each year (the area of each box indicates the strength of the discrepancy between CPUE series, with larger box indicating higher degree of conflicting information), and a loess smoother through all residuals to aid detection of the presence of systematic residual patterns.

Additionally, the root-mean-squared-error (RMSE) was used as a goodness-of-fit statistic, and runs tests were conducted to quantitatively evaluate the randomness of residuals (Carvalho *et al*., 2017). The runs test diagnostic was applied to residuals of the CPUE fit on log-scale considering the 2-sided p-value of the Wald-Wolfowitz runs test and is visualized in JABBA to illustrate which time series passed or failed the test, as well as highlighting individual data points that fall outside the three-sigma limits (Anhøj and Olesen, 2014).

To check for systematic bias in the stock status estimates, a retrospective analysis was carried out for all the base case Grid models. This analysis was carried out by sequentially removing one year of data at a time, over a total period of 5 years, and then refitting the model without those years. The parameters of interest (i.e., biomass, fishing mortality, B/Bmsy, F/Fmsy, B/K and MSY) were then compared to the original models fitted using the full time series. The presence of possible retrospective bias between the models was analyzed visually with plots, and statistically with the Mohn's rho (ρ) statistic (Mohn, 1999), using the formulation defined by Hurtado-Ferro *et al*. (2014). In this analysis, the more the values diverge from zero the stronger there is the presence of a retrospective bias. In general, values that fall between - 0.15 and 0.2 are widely deemed as having an acceptable retrospective bias (Huerto *et al*., 2014).

The analysis included several sensitivity model runs, namely based on the following scenarios: 1) a catch only model without using information from the CPUE time series; 2) leave-one-out CPUE analysis where each CPUE was dropped at a time starting either with the full model using all available CPUEs; 3) using one CPUE at a time, 4) sensitivity analysis to the sigma of the process error (fixed at 5% and 10%) and inclusion of additional CPUE variance and; 5) a sensitivity analysis using the estimated catch time series; 6) sensitivity analysis with different initial depletion levels. For the sensitivity analysis, the main base model (base productivity and estimated catches) was used.

Finally, and following a request from the Sharks-DG at the assessment meeting, we prepared and provided correlations pairs plots for the final models.

2.5 Model ensemble

2.5.1 Model ensemble specifications

After building the base grid model with the 4 main models as specified above, we then constructed and carried out a larger grid ensemble approach, drawing random values from pre-defined distributions between the limits or based on the model underlying assumptions that had been used in the main base models. This allowed increasing the dispersion of values between the limits that were considered acceptable for the main assessment models.
The parameters and respective statistical distributions that were used in building this large ensemble grid were:

- *r* prior (intrinsic population growth rate): Values from a random uniform distribution varying between 0.049 and 0.114, the values used in the main base models;
- Bmsy/K prior (Bmsy at K, used to set the shape of the production function): Values from a random uniform distribution varying between 0.578 and 0.637, the values used in the main base models;
- Catch CV prior (coefficient of variation of the catch series): Values from a random uniform distribution varying between 0.15 and 0.25, around the values used in the main base models;
- PSI prior (initial depletion): Values from a random uniform distribution varying between 0.90 and 0.92, around the values used in the main base models;
- Catch series (estimated vs reported): Each set of values taken from the distributions above was run in models using either the estimated or the reported catch series.

The large model ensemble with those characteristics was run 500 times (250 times using each of the two alternative catch series). Each individual model was validated and checked for convergence for key biological parameters (K, *r*) using the Geweke diagnostics and the Heidelberger and Welch tests. Models with failure in either diagnostic were excluded from the final ensemble.

Running such large grid model ensembles requires high computational power, and to achieve that we used the High Performance Computing platform services of Inductiva (https://inductiva.ai/), setting high speed parallel computing and running multiple models in parallel. In this specific case, we run the entire 500 model grid in cloud machines of the type "c2-standard-8", each with 8 vCPUs and 32Gb RAM. Running the entire grid with those specifications took approximately 24 mins, resulting in an output of 12.88 GB and had an estimated cost of 2.29 $US. This can be compared to running this task in a single PC (e.g., a modern laptop with an Intel Core Ultra 7 Processor (4.8 GHz) and 64Gb RAM) that would take approximately 2 mins per model, meaning a total time of approximately 16 hours to run this entire model grid ensemble.

2.5.2 Model weighting and averaging

Following the convergence diagnostics, the retained ensemble models were summarized using two alternative approaches:

- Equal model-weighing: Each individual model was given equal weight, assuming therefore that all models are equally plausible without favoring any particular type of fit.
- DIC-weighted ensemble: Models were weighted inversely by their Deviance Information Criterion (DIC) values. The Delta-DIC was computed relative to the best-fitting model and the derived model weights as exp(-0.5*ΔDIC), normalized to sum to one (Burnham & Anderson, 2002). This approach gives higher weights to better fitted models, while still incorporating ensemble uncertainty for the less fit models

2.6 Preliminary projections

Preliminary deterministic projections were carried out individually for the 4 base models. Those projections assumed a delay of 3 years in implementation after the terminal year, i.e., in this case with the models terminating in 2023 and the fixed TACs implemented from 2026 onwards. The catches for the intermediate years (2024-2025), were assumed as the average catches of the previous 3 years (mean for years 2021-2023).

3. Results and discussion

3.1 Initial model development and configuration

Some initial model runs were carried out using all CPUEs as provided, and using the CPUEs one-at-a-time, but those were not used as it was deemed extremely unreasonable given the conflicts between some CPUEs and, more importantly, the fact that the catch history and the CPUEs had similar increasing trends over parts of the time series, which is not biologically plausible. Those CPUEs conflicts are highlighted in **Figure 3**, where depending on the CPUE used the stock trajectories would range from presently being close to a virgin biomass state to being collapsed, in some cases with extremely large CIs.

As such, the procedure that followed was to explore in more detail the inclusion of the various CPUEs, and the need to split them into different time periods, allowing therefore the estimation of different catchability (q) parameters for each time period. This split into time-blocks could be justified by differences in the fishing gear, technologies or practices over time that were not fully taken into account in the CPUEs standardization, or also due to changes in data collection and reporting procedures over time.

The Japanese CPUE series was already split in 2011 by the CPUE analysists, so that series was used as provided. In addition to that, there was the need to also split the series of South Africa, Spain and Brazil-Uruguay at around the same time period, as there were peaks and inflections in the CPUEs that cannot be easily explained or fitted in the production model. Specifically, the South Africa series was split in 2012 when there seemed to be an inflection point, the Spanish series in 2011, and Brazil-Uruguay in 2015. Additionally, the initial 4 years of the Brazil-Uruguay series show a very strong increase in standardized CPUEs which are not biologically plausible, and for that reason and for that series a 3rd (earlier) time period was set for the years previous to 1982.

Finally, and after various trials, it doesn't seem reasonable to add the Chinese Taipei series in the base case model, given this series is relatively short and shows very strong and continuous declines in the entire period, which results in very high values of fishing mortality but with extremely large and unreasonable CIs. As such, the Chinese Taipei CPUE series was not used on the base case model and was only used in the sensitivity analysis.

In summary, the various CPUE series used for the base case models, with their respective time blocks were:

- Japan LL – time block 1 (1994-2011)
- Japan LL – time block 2 (2012-2023, with a gap in 2020-2021)
- Brazil/Uruguay combined LL index – time block 1 (1978-1981)
- Brazil/Uruguay combined LL index – time block 2 (1982-2014)
- Brazil/Uruguay combined LL index – time block 3 (2015-2022)
- South Africa – time block 1 (2000-2011)
- South Africa – time block 2 (2012-2023)
- Spain – time block 1 (1990-2010)
- Spain – time block 2 (2011-2023)

In terms of the CVs of the CPUEs, there was a great disparity between the CPUE series, which can result more from the CPUE data structure and methods used for the CPUE standardization, rather than the actual variability associated with the series. As such, a first approach was established to set a minimum CV of 0.2 to all CPUEs, keeping the cases when higher values were estimated by the CPUE standardization analysis. In addition to that, for the series that had to be split in early and later time periods, a minimum CV of 0.4 was set to the earlier period, again keeping any higher values if estimated and provided in the CPUE analysis directly. This was done to take into account and reflect the possibility that there is expectation of higher uncertainty in the initial years of the series, and more certainty and data quality in the more recent years.

After this procedure, a base case model with estimated catches was created that seemed to provide reasonable results. From that, the 4 main base case models (base/low productivity; estimated/reported catches) were developed.

3.2 Model goodness of fit and validation

3.2.1 Model convergence

The MCMC convergence tests by Heidelberger and Welch (1992) and Geweke (1992) all passed with regards to the MCMC estimation of the parameters for all models. An adequate convergence of the MCMC chains was also corroborated visually by checking the trace plots, which showed good mixing and random deviations around the parameters space, without any detectable bias or patterns that could result from autocorrelations in the estimations (**Figure 4**).

3.2.2 Fit to CPUE indices

The fits of the base case grid models to each of the standardized CPUE indices used are shown in **Figure 5**. The goodness-of-fit of those residuals were similar between all base case grid models used, with the RMSE statistic ranging between 41.3% and 41.9% (**Figure 6**).

The runs test for the CPUE residuals from each of the main 4 grid models are provided in **Figure 7**. Some CPUEs passed the tests, while others failed, showing that in some cases there are patterns of non-randomness in the residuals. All 4 base case grid models showed similar results in terms of the runs tests. Specifically, the Japanese and Spanish CPUEs failed in the early time-period but passed in the later period, the South Africa CPUE passed on both time periods, while the Brazil-Uruguay index passed in the early period and failed in the recent periods. In some cases, some outliers were identified in the residuals, defined as points outside a 3-fold limit around the overall residuals means (Anhøj and Olesen, 2014).

3.2.3 Process error deviations

The deviations from the process error show similar patterns for all 4 main grid models. There were some patterns with lower deviates in the very early time period until around 1985, followed by several years of positive process error deviates until around 2010, and then a more stable period for the more recent years, with the deviates oscillating and centered around zero. Nonetheless, the 95% credibility intervals always included the zero value during the entire time series period, suggesting that there is no major evidence of structural model misspecifications (**Figure 8**).

3.2.4 Retrospective analysis

The results of the retrospective analysis applied to all the 4 main base grid models are shown in **Figure 9**, and the corresponding summaries of the estimations of the Mohn's rho are summarized in **Table 1**. Almost all models have values within the acceptable range of -0.15 to 0.20 for almost all the parameters, as defined by Hurtado-Ferro *et al*. (2014) and Carvalho *et al*. (2017). This analysis confirms that there are some retrospective patterns in the models with regards to the main parameters.

3.2.5 Hindcast cross validation

The hindcast cross-validation procedure was conducted for the more recent time periods of the indices. The results show that the predictions, when 1-year at a time for the last 5-years are removed, mostly fall within the limits of the 95% CIs, but not for all cases (**Figure 10**). The mean absolute scaled error (MASE) estimates were under the reference level (MASE < 1) for the cases of Japan, Spain and South Africa, indicating that the average forecasts for those indices have good predictive skills (Carvalho *et al*., 2021), with the observed values both above and under the predictions. By the contrary, for the Brazil-Uruguay index the MASE value was well above 1, indicating poor predictive skills for that index, with the observed values alway above the predictions.

3.2.6 Correlations of the posteriors

As requested at the assessment meeting, correlations pairs-plots were prepared and added to the analysis, represented in **Figure 11**. The main patterns detected were a strong negative correlation between *K* and *r*, and some positive correlation between *m* and *K*, and between *m* and *r*, which are all expected from life history population dynamics.

3.3 Sensitivity analysis

With regards to the sensitivity analysis, and given that the 4 main base grid models developed all had similar behavior, goodness of fit and validation performance, only one of the models was used as a comparison to the options defined as sensitivities. Specifically, all sensitivity analysis was compared to the model using a base productivity and estimated catches.

3.3.1 Catch only model

The results of the sensitivity analysis conducted for a model with catch only information is shown in **Figure 12**. The catch only model is mostly informed by the biological prior information, the priors set for initial and final depletion, and the history and trends from the times series of the catches. For this sensitivity, catch-only models using both the estimated and the reported catches were done.

There are some differences when the CPUE data is entirely excluded, but the overall results have some similarities in terms of general trends, with the biomass showing an overall and continuous decrease, while the fishing mortality shows an increase until around 2020, and then a decrease in the more recent years. This final status with the catch-only models are more pessimistic, resulting in a stock that is more depleted and has worse stock status.

3.3.2 Using one CPUE at a time

Another sensitivity analysis was conducted by using one CPUE at a time, where each model was run using only one of the available CPUEs at a time. This was configured in relation to the base case model, with the CPUEs used with their different time-blocks as used for the base case model development. The results are represented in **Figure 13**.

The results show overall and relatively similar trends in the stock status over time, even though depending on the CPUE series that is used, the depletion and reductions in relative biomass are more or less stronger.

This indicates that all CPUE series are providing relatively similar signals to the stock trends and status, and that each one individually is not too strong to introduce signals that are very different from the others.

3.3.3 Leave out one CPUE at a time

A sensitivity analysis was conducted with leave-one-out CPUE scenarios, where each model was run excluding one CPUE series at each time, starting with the base case model. The results are represented in **Figure 14**.

Overall, there are no major differences when excluding one CPUE at a time from the base case model, indicating that there is not any CPUE series that has a particular major influence in the final trends and results. This means that if any of those series is removed from the base case model, the model still maintains its overall trends and status and is not overly being influenced by each of the individual CPUE series.

3.3.4 CPUEs by fleet type (Shark/SWO target vs bycatch)

Another sensitivity analysis was conducted by creating models using only indices for the fleets that are more assumed to be targeting SWO/Sharks (Brazil-Uruguay, South Africa and Spain) compared to fleets where sharks are mostly an occasional bycatch (Japan), and those results are represented in **Figure 15**.

Again, and overall, there were no major differences in this sensitivity, with the overall stock trends and stock status very similar in both scenarios.

3.3.5 Using ICCAT reported catches

Another sensitivity analysis was conducted for using the ICCAT nominal catches instead of the reconstructed catches, with the results represented in **Figure 16**.

In this case, because the catch series are not the same, there is a direct effect in terms of total stock biomass and overall MSY estimates, even though they are not majorly different in this case. In terms of trajectories, both are relatively similar in both models, using either the estimated or the reported catches.

3.3.6 Process error and CPUE variance

An additional sensitivity analysis carried out was with regards to the process error and additional estimation of CPUEs variance. In the base case model the process error is estimated within the models with uninformative igamma priors set as the default values, and the option to allow for additional CPUEs variance internally in JABBA is also allowed. Sensitivities were run for options on fixing the sigma of the process error to CVs of 5% and 10%, and another to turn off the additional inclusion of CPUE variance. The results of this analysis are presented in **Figure 17**.

Overall, those sensitivities show no major differences in the overall trends compared to the base case model. By fixing the process error with a lower CV the trajectories become smoother, but the end results are still similar. The main caveat with this option is related to the model validation procedures, as using such configuration results in worse model fits, in general with poorer fits to the CPUEs and worse performance in terms of retrospective analysis. As such, it is in general preferable to allow the process error to be estimated internally by the models, as that will optimize the posterior of the process error based on the rest of the data that is providing information to the models.

With regards to the additional CPUE variance, when that option is disabled, the trends and results also do not change much compared to the base case model. This is likely because the CPUEs individually do not have too much weight in the overall model trends, so the additional CPUE variance is also not having a large impact.

3.3.7 Initial depletion levels

Following a request from the Sharks-SG, an additional sensitivity analysis was carried out using different initial depletion levels. The results are presented in **Figure 18**, and show that with the different depletions, the starting point at 1971 is very different but then the trajectories tend to converge around 1995 and are consistent from that point forward. The final year stock status remains similar on all tested scenarios.

3.4 Stock status for final models

The parameters estimated in the final models (base case and low productivity) are represented in **Table 2**.

The trajectories of the final main 4 base models of both biomass and fishing mortality in relation to MSY reference points are indicated in **Figure 19**, and the Kobe phase plots for those scenarios are represented in **Figure 20**. The overall trajectories of all scenarios are all relatively similar in terms of biomass, and show more differences in terms of fishing mortality. In terms of the final year stock status there are also important differences, as the final year falls in the green when using base productivity with reported catches, in the yellow when using base productivity with estimated catches, and in the red when using the low productivity both with reported and estimated catches.

3.5 Model ensemble

The large model grid ensemble was run with 500 models, with 250 using each of the 2 optional catch series (reported or estimated). The range and distribution of the input prior values that were randomly generated for the large grid ensemble is shown in **Figure 21**.

Of the 500 initial modes, 47 failed to converge, while the remaining 453 models converged. The models that converged were used for the remainder of this analysis, descriptive statistics and determining stock status. The trajectories of all the models that converged and are used for analysis are described in **Figure 22**.

The summaries of the main quantities from the large grid models are summarized in **Table 3**. The quantities are summarized in 2 alternative ways: 1) using equal weights across all models that converged; 2) using DIC-weights across models that converged, giving more weight to models with better DIC values.

The summarized Kobe status plot with the 2 model weighing options is given in **Figure 23**. The stock status is relatively similar when using either model weighting options, in both cases with the stock status determined to be in the red quadrant, overfished and subject to overfishing.

3.6 Preliminary projections

The preliminary deterministic projections of the 4 base models are represented in **Figure 24**. It is possible to verify that in those 4 main models that there are some differences depending on whether base or low productivity are considered and depending on whether reported or estimated catches are considered.

4. Conclusions

Bayesian Surplus Production Model (JABBA) were applied for determining the stock status, in support of the provision of management advice for the South Atlantic (ICCAT) shortfin mako shark.

Four main base models were developed, using combinations of base/low productivity scenarios, and reported/estimated catches, as agreed at the Shark-SG data preparatory meeting (ICCAT, 2025), and additional intersessional work. Stock status from these 4 main models ranged from overfished and subject to overfishing (B<Bmsy & F>Fmsy), overfished but not subject to overfishing (B<Bmsy & F<Fmsy), and not overfished and not subject to overfishing (B>Bmsy & F<Fmsy).

In addition to these four base models, a large model grid ensemble (500 models) was run, selecting randomly prior values from distributions built around the plausible limits for their values, and using alternatively each of the 2 catch scenarios. The stock status from this large grid ensemble was weighed in 2 alternative ways, namely using equal model-weighting and DIC model-weighting. In both cases, the stock status were similar and resulted in the stock being overfished and subject to overfishing (B<Bmsy & F>Fmsy), with slight worse stock status when using the DIC-weighting criteria.

5. Acknowledgments

We thank INDUCTIVA (https://inductiva.ai/) for all their support and for providing access to their High-Performance Computing platform, which was instrumental in conducting the large grid model ensemble simulations and analysis presented in this work. Rui Coelho (CCMAR) receives partial support from FCT through projects UIDB/04326/2020, UIDP/04326/2020 and LA/P/0101/2020.

References

Anhøj, J., Olesen, A.V. 2014. Run charts revisited: A simulation study of run chart rules for detection of non-random variation in health care processes. *PLoS One*, 9: 1–13. https://doi.org/10.1371/journal.pone.0113825.

Auguie, B. 2017. gridExtra: Miscellaneous Functions for "Grid" Graphics. R package version 2.3, https://CRAN.R-project.org/package=gridExtra.

Burnham, K.P., Anderson, D.R. 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach (2nd ed.). Springer

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

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., Methot, R.D. 2021. A cookbook for using model diagnostics in integrated stock assessments. *Fisheries Research*, 240: 105959.

Coelho, R., Rosa, D. 2017. An alternative hypothesis for the reconstruction of time series of catches for North and South Atlantic stocks of shortfin mako shark. Collect. Vol. Sci. Pap. ICCAT, 74(4): 1746-1758.

Cortés, E., Domingo, A., Miller, P., Forselledo, R., Mas, F., Arocha, F., Campana, S., Coelho, R., Da Silva, C., Holtzhausen, H., Keene, K., Lucena, F., Ramirez, K., Santos, M.N., Semba-Murakami, Y., Yokawa, K. 2015. Expanded ecological risk assessment of pelagic sharks caught in Atlantic pelagic longline fisheries. Collect. Vol. Sci. Pap. ICCAT, 71(6): 2637-2688.

Gelman, A. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3): 515-533.

Gelman, A., Rubin, D.B. 1992. Inference from iterative simulation using multiple sequences. Statistical Science, 7, 457–472. https://doi.org/10.1214/ss/1177011136.

Geweke, J. 1992. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Berger, J.O., Bernardo, J.M., Dawid, A.P., Smith,A.F.M., (Eds.), Bayesian Statistics 4: Proceedings of the 4th Valencia International Meeting. Clarendon Press, Oxford, pp. 169–193.

Heidelberger, P., Welch, P.D., 1992. Simulation run length control in the presence of an initial transient. Operations Research, 31. DOI: 1109–1144. https://doi.org/10.1287/opre.31.6.1109.

Højsgaard S., Halekoh., U. 2023. doBy: Groupwise Statistics, LSmeans, Linear Estimates, Utilities. R package version 4.6.20, https://CRAN.R-project.org/package=doBy.

Hurtado-Ferro, F., Szuwalski, C.S., Valero, J.L., Anderson, S.C., Cunningham, C.J., Johnson, K.F., Licandeo, R., McGilliard, C.R., Monnahan, C.C., Muradian, M.L., Ono, K., Vert-Pre, K.A., Whitten, A.R., Punt, A.E. 2014. Looking in the rear-view mirror: Bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science, 99–110. https://doi.org/10.1093/icesjms/fsu198.

ICCAT. 2017. Report of the 2017 ICCAT Shortfin Mako Assessment Meeting. Collect. Vol. Sci. Pap. ICCAT, 74(4): 1465-1561.

ICCAT. 2025. Report of the 2025 ICCAT Shortfin Mako Data Preparatory Meeting. Hybrid/Malaga, Spain, 10-14 March 2025. Collect. Vol. Sci. Pap. ICCAT, 82(2), SCRS/2025/002: 1-91

Kell, L.T., Sharma, R, Winker, H. 2022. Artefact and artifice: evaluation of the skill of Catch-Only methods for classifying stock status. Frontiers in Marine Science, 9: 762203. doi: 10.3389/fmars.2022.762203

Mejuto, J., Fernández-Costa, J., Ramos-Cartelle, A., Carroceda, A. 2021. Plausibility and uncertainty of basic data and parameter selection on stock assessments: a review of some input data used in the 2017 assessment of the shortfin mako (*Isurus oxyrinchus*) of the northern Atlantic stock. Collect. Vol. Sci. Pap. ICCAT, Vol. 78(5): 119-170.

Mohn, R. 1999. The retrospective problem in sequential population analysis: an investigation using cod fishery and simulated data. ICES Journal of Marine Science, 56: 473–488. https://doi.org/10.1080/02755947.2014.882452.

Mourato, B., Sant'Ana, R., Kikuchi, E., Cardoso, L.G., Ngom, F., Arocha, F., Kimoto, A., Ortiz, M. 2023. Assessment of the eastern Atlantic sailfish stock using JABBA model. Collective Volumes of Scientific Papers ICCAT, 80(8): 315-336.

Plummer, M. 2003. JAGS: A Program for Analysis of Bayesian Graphical Models using Gibbs Sampling, 3rd International Workshop on Distributed Statistical Computing (DSC 2003); Vienna, Austria.

Plummer, M., Nicky, B., Cowles, K., Vines, K. 2006. CODA: convergence diagnosis and output analysis for MCMC. *R News*, 6 (1): 7–11.

R Core Team. 2025. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Su, Yajima. 2012. R2jags-a package for running jags from R. https://cran.r-project.org/web/packages/R2jags/index.html.

Wickham, H. 2007. Reshaping data with the reshape package. Journal of Statistical Software, 21(12), 2007.

Wickham, H. 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L.D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T.L., Miller, E., Bache, S.M., Müller, K., Ooms, J., Robinson, D., Seidel, D.P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., Yutani, H. 2019. Welcome to the tidyverse. Journal of Open Source Software, 4 (43): 1686. doi:10.21105/joss.01686.

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D. 2023. dplyr: A Grammar of Data Manipulation. R package version 1.1.4, https://CRAN.R-project.org/package=dplyr.

Wickham, H., Vaughan, D., Girlich, M. 2023. tidyr: Tidy Messy Data. R package version 1.3.0, https://CRAN.R-project.org/package=tidyr.

Winker, H., Carvalho, F., Kapur, M. 2018. JABBA: Just Another Bayesian Biomass Assessment. *Fisheries Research* 204, 275–288. https://doi.org/http://doi.org/10.1016/j.fishres.2018.03.01.

Wilke, C. 2024. cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.3, https://CRAN.R-project.org/package=cowplot.

Table 1. Summary of the Mohn's rho statistic computed from the retrospective analysis pattern evaluated for the 4 main base grid models (Mod.07: base with estimated catches; Mod.08: low productivity with estimated catches; Mod.09: base with reported catches; Mod.10: low productivity with reported catches). Values that have an acceptable retrospective bias (Huerto *et al*., 2014) are highlighted in green.

Model	B	F	Bmsy	Fmsy	B/K	MSY
Mod.07	-0.133	0.162	-0.029	0.100	0.001	-0.038
Mod.08	0.006	0.044	-0.101	0.125	-0.012	0.029
Mod.09	0.038	0.054	-0.039	0.155	-0.002	-0.067
Mod.10	-0.123	0.169	-0.040	0.219	-0.004	-0.138

Table 2. Estimates (mean, lower and upper confidence intervals) of the point estimates for the various parameters estimated in the 4 main base grid models, developed for the 2025 ICCAT SMA South Atlantic stock assessment.

	Base productivity			Low productivity			
Parameters	**mu**	**lci**	**uci**	**mu**	**lci**	**uci**	
K	54415	33696	81976	62335	41971	90478	**Estimated catches**
r	0.166	0.093	0.298	0.069	0.034	0.139	
psi	0.895	0.522	0.996	0.916	0.646	0.997	
sigma.proc	0.063	0.039	0.117	0.081	0.044	0.139	
m	2.352	1.356	4.094	3.339	1.800	6.199	
Hmsy	0.070	0.032	0.154	0.020	0.007	0.059	
SBmsy	28842	15938	48430	36945	23282	56685	
MSY	2030	1223	3026	773	282	1781	
bmsyk	0.531	0.425	0.634	0.597	0.480	0.704	
P1971	0.874	0.508	1.068	0.895	0.629	1.108	
P2023	0.489	0.295	0.735	0.505	0.311	0.803	
B_Bmsy.cur	0.922	0.552	1.428	0.852	0.533	1.354	
H_Hmsy.cur	0.772	0.320	1.789	2.153	0.751	7.092	
	Base productivity			**Low productivity**			
Parameters	**mu**	**lci**	**uci**	**mu**	**lci**	**uci**	
K	43494	29702	69484	79482	51573	146200	**Reported catches**
r	0.183	0.095	0.334	0.059	0.031	0.110	
psi	0.903	0.522	0.996	0.885	0.537	0.995	
sigma.proc	0.071	0.042	0.128	0.067	0.040	0.122	
m	2.349	1.348	4.187	3.695	2.024	6.747	
Hmsy	0.078	0.032	0.173	0.016	0.006	0.039	
SBmsy	22991	14354	40451	48021	29665	93618	
MSY	1850	926	3002	806	320	1763	
bmsyk	0.531	0.424	0.638	0.616	0.502	0.717	
P1971	0.881	0.504	1.085	0.859	0.524	1.062	
P2023	0.537	0.348	0.763	0.536	0.318	0.823	
B_Bmsy.cur	1.015	0.639	1.502	0.879	0.509	1.373	
H_Hmsy.cur	0.772	0.313	2.048	2.046	0.679	6.513	

Table 3. Summaries of the estimates (mean, lower and upper confidence intervals) of the main parameters across the large grid of ensemble models. The values reported represent the mean, SD and limits of the 95% CIs across the estimations (posterior values) of all ensemble models. The summaries are calculated by 2 options: 1) giving equal weights across all models; 2) using DIC-weights across models.

	Equal-Weighted			**DIC-Weighted**		
Parameter	**Mean**	**LCI**	**UCI**	**Mean**	**LCI**	**UCI**
K	56411	37751	83782	59655	40033	88887
r	0.123	0.065	0.226	0.112	0.059	0.204
psi	0.909	0.530	0.998	0.907	0.531	0.998
sigma.proc	0.071	0.041	0.129	0.071	0.042	0.128
m	2.766	1.563	4.983	2.785	1.569	5.045
Hmsy	0.046	0.018	0.105	0.041	0.016	0.094
SBmsy	31558	19320	50702	33456	20570	53920
MSY	1427	655	2434	1377	626	2345
bmsyk	0.561	0.451	0.667	0.562	0.452	0.669
P1971	0.882	0.515	1.086	0.880	0.516	1.085
P2023	0.509	0.316	0.765	0.496	0.304	0.757
B_Bmsy.cur	0.911	0.562	1.399	0.886	0.540	1.375
H_Hmsy.cur	1.206	0.463	3.438	1.301	0.498	3.746

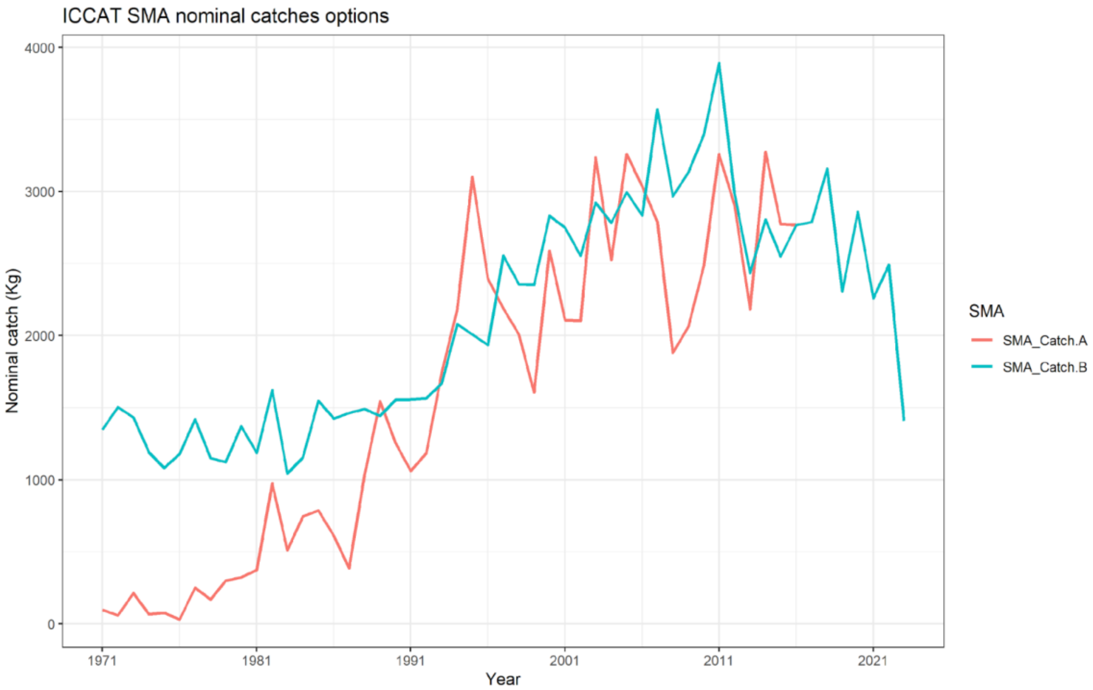

Figure 1. Time series of the ICCAT SMA catch data used for the assessment. Two options are available: Option A - nominal catches that are reported to ICCAT; Option B - catch reconstruction using ratio methods.

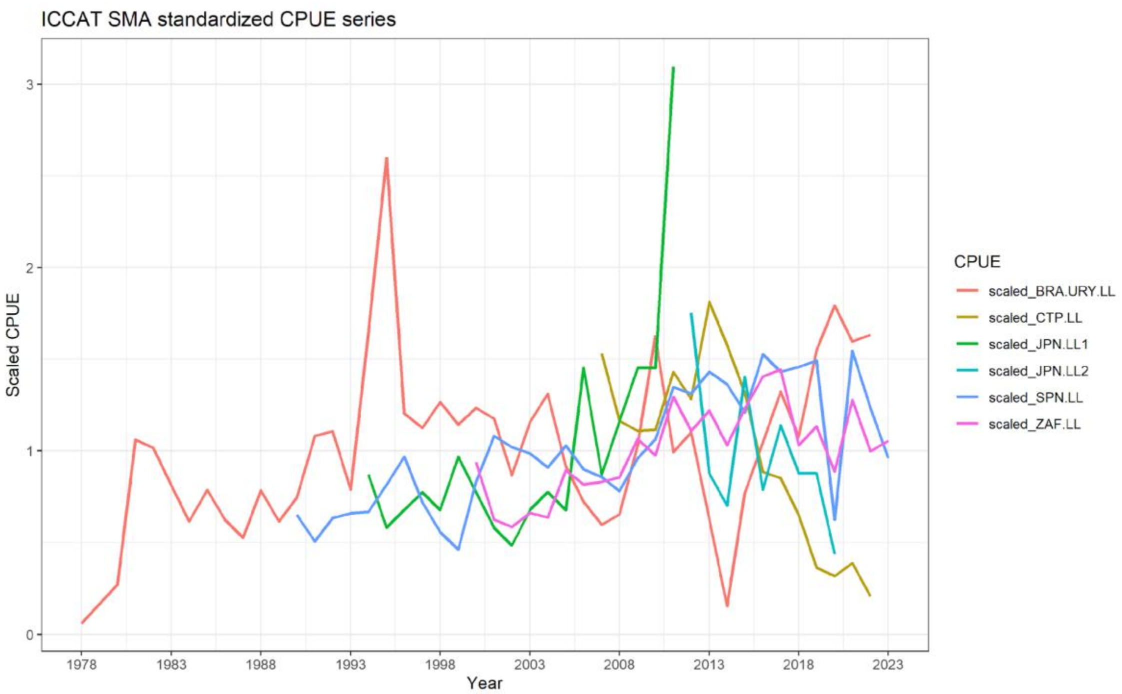

Figure 2. Standardized CPUE series available for the ICCAT South SMA stock assessment. For better visualization and comparison, each series is scaled by its respective average.

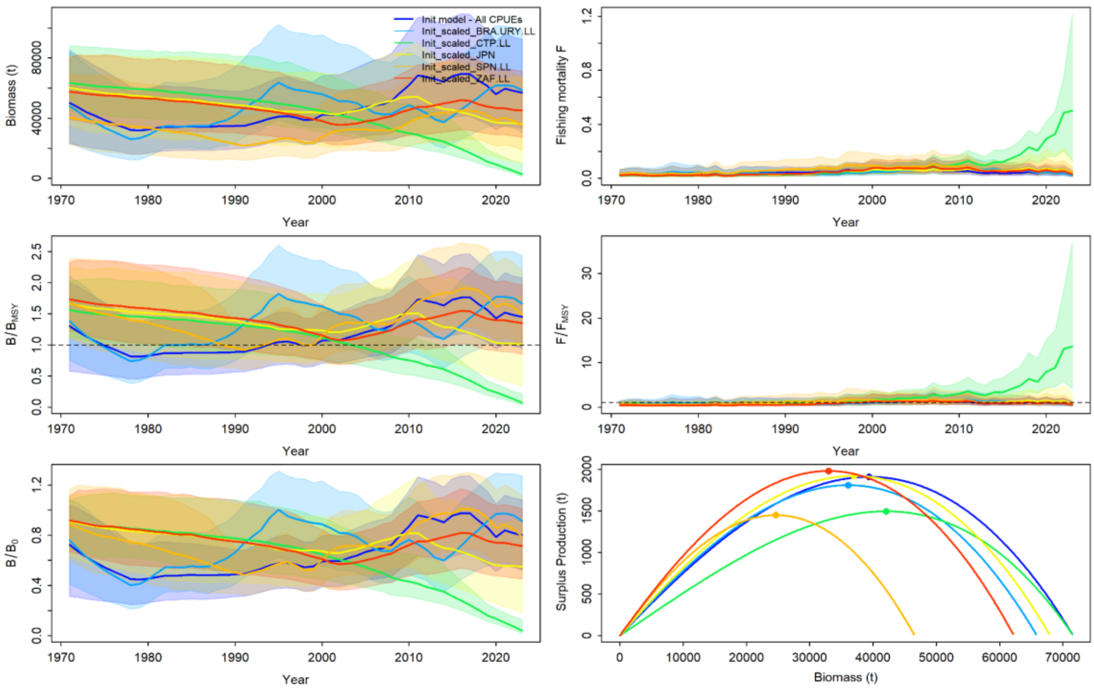

Figure 3. Initial model runs using all available CPUEs and using one CPUE at-a-time, for the south Atlantic stock of shortfin mako shark.

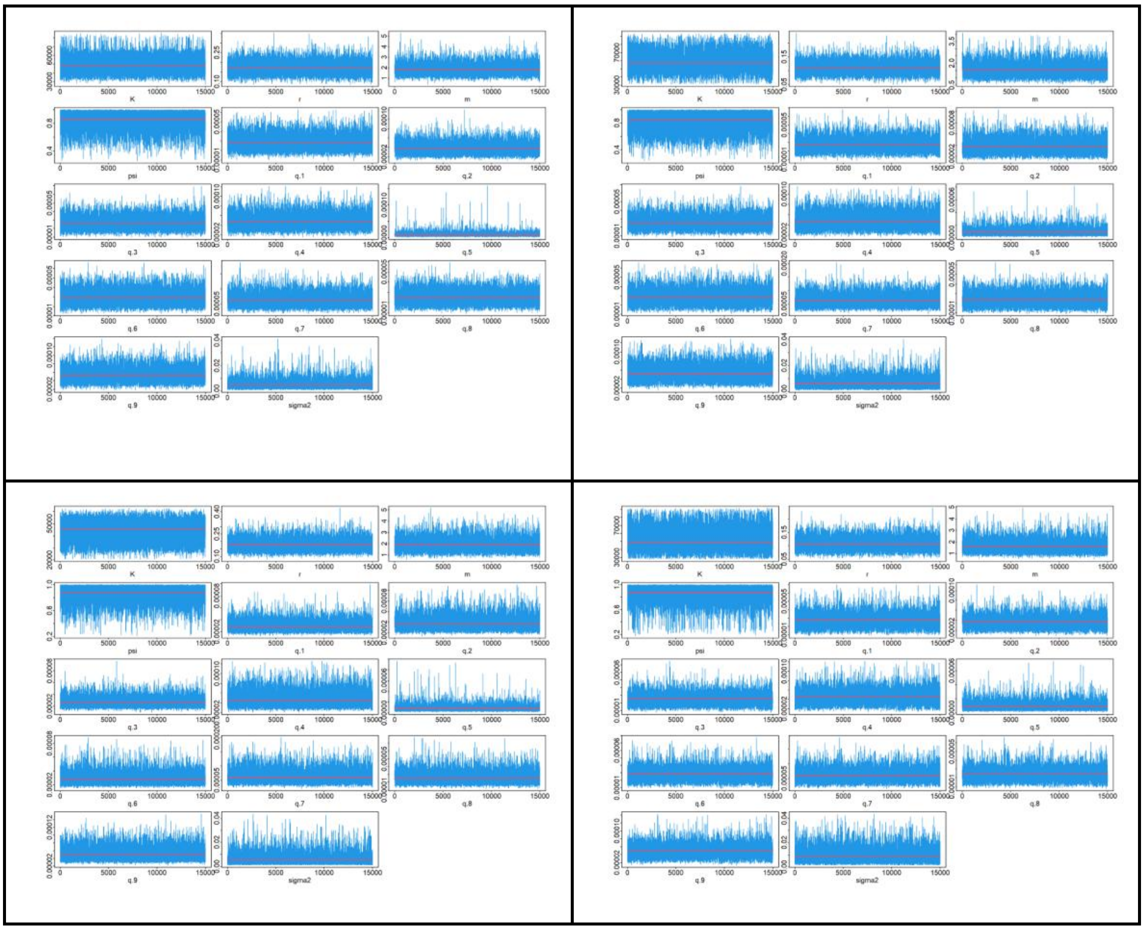

Figure 4. Trace plots of the following base grid models: base with estimated catches (Mod.07 - top left); low productivity estimated catches (Mod.08 - top right); base with reported catches (Mod.09 - bottom left); low productivity reported catches (Mod.10 - bottom right).

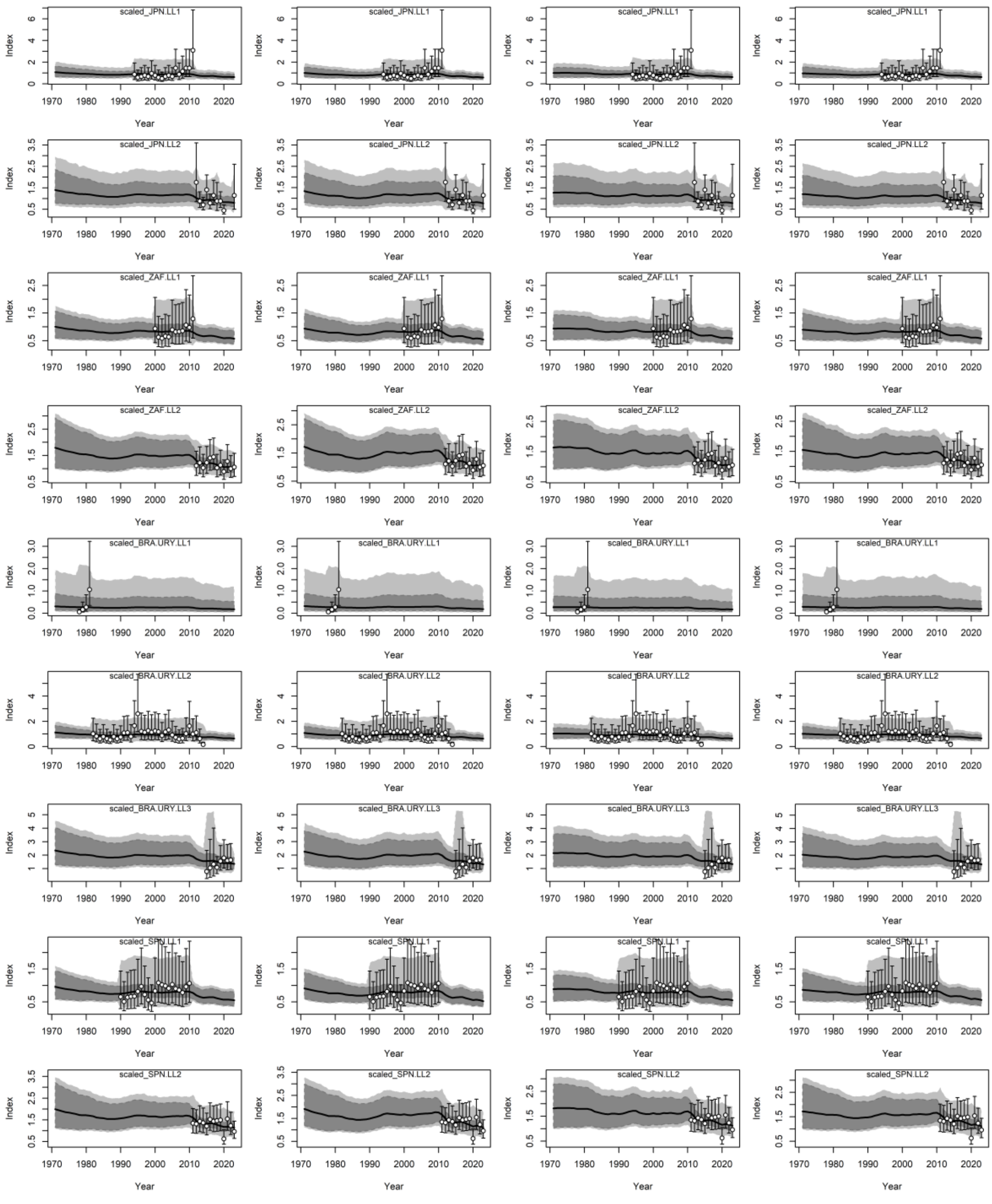

Figure 5. Time series of observed (circles) and predicted (solid line) CPUEs for the ICCAT South Atlantic SMA stock assessment models, namely the 4 main grid models defined. The dark shaded areas represent the 95% credibility intervals of the expected mean CPUE, and the light shaded areas represent the 95% posterior predictive distribution intervals. The error bars are the 95% confidence intervals (CIs) from the CPUE observations.

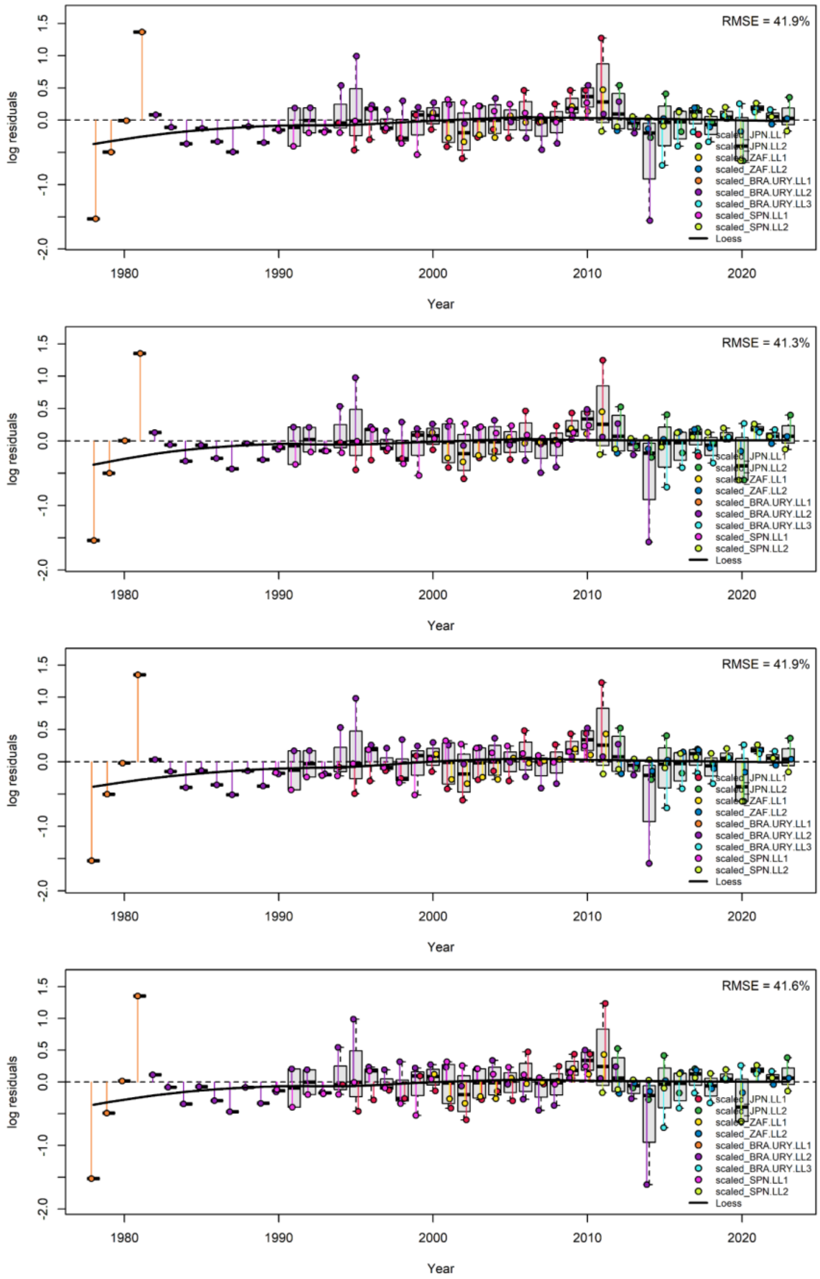

Figure 6. Residuals diagnostic plots for the main 4 base grid models, for the ICCAT South Atlantic SMA. Each individual CPUE index and its respective residuals are represented by a different color. The solid black lines represent loess smoothers through all residuals combined.

Figure 7. Runs tests for the CPUE index for all the base case and low productivity models, used for the ICCAT South Atlantic SMA stock assessment.

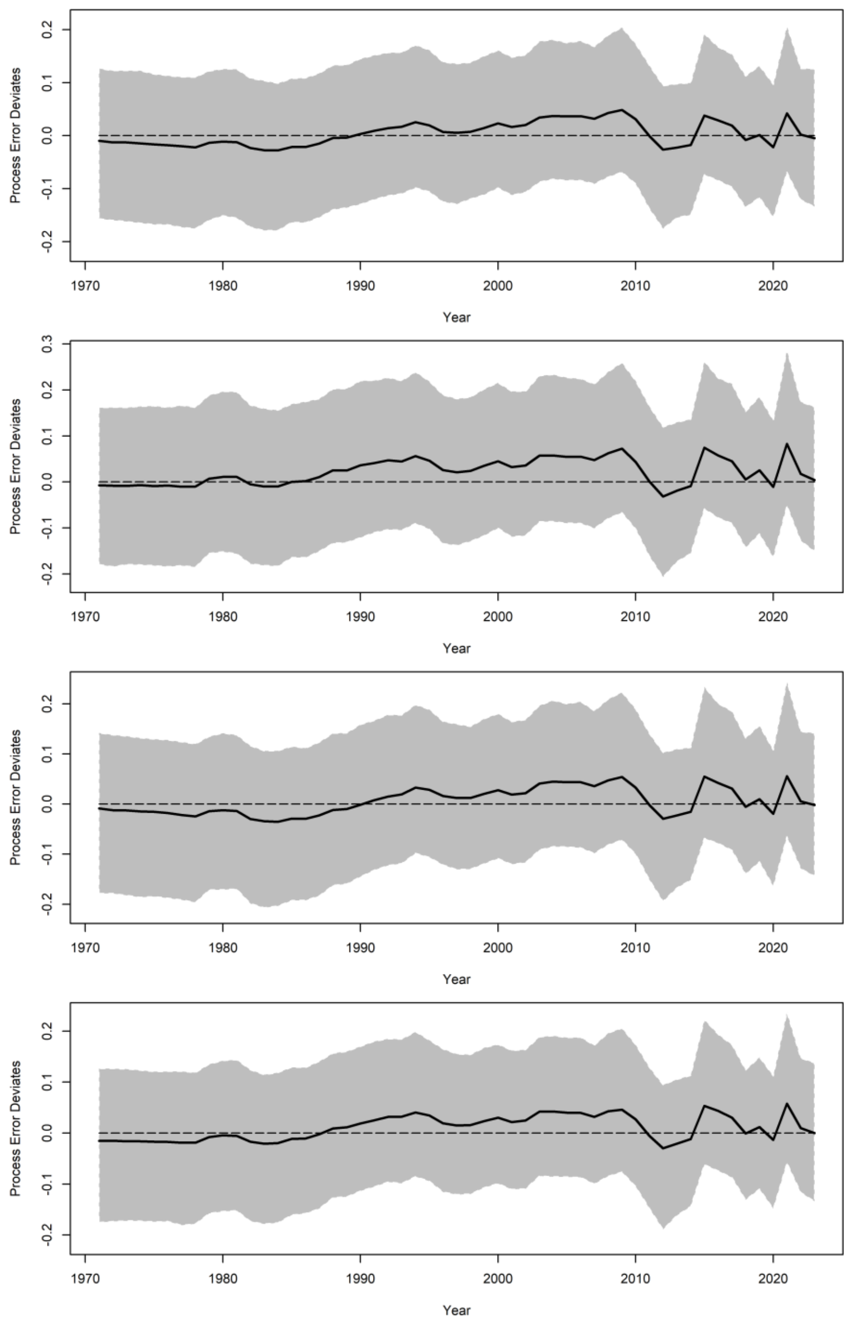

Figure 8. Process error deviates for the main 4 grid models used for the ICCAT SMA South Atlantic stock assessment. The solid line represents the median, and the shaded gray area the 95% credibility intervals.

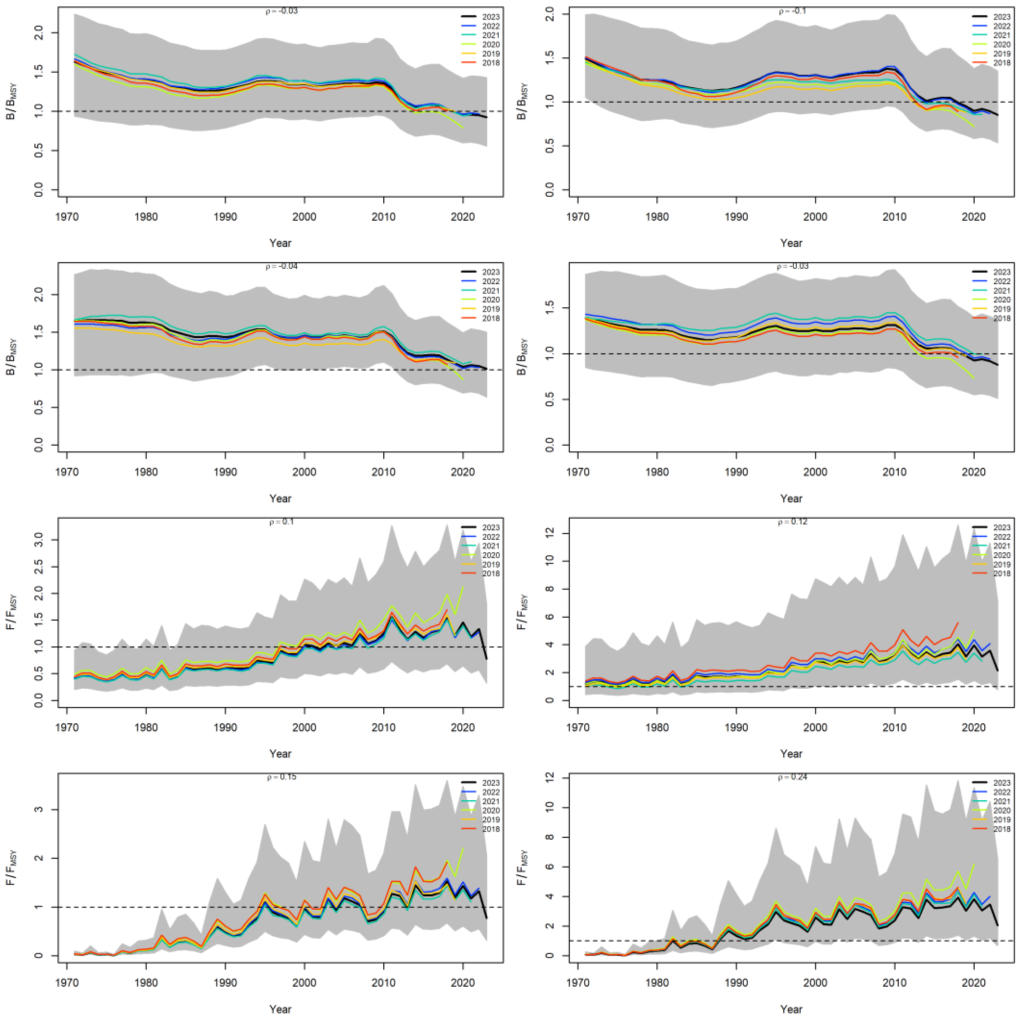

Figure 9. Retrospective analysis conducted for the 4 main base grid models developed for the 2025 ICCAT SMA South Atlantic stock assessment, by removing 1-year at a time sequentially (n=5) and predicting the trends in biomass and fishing mortality relative to MSY (i.e, B/Bmsy and F/Fmsy).

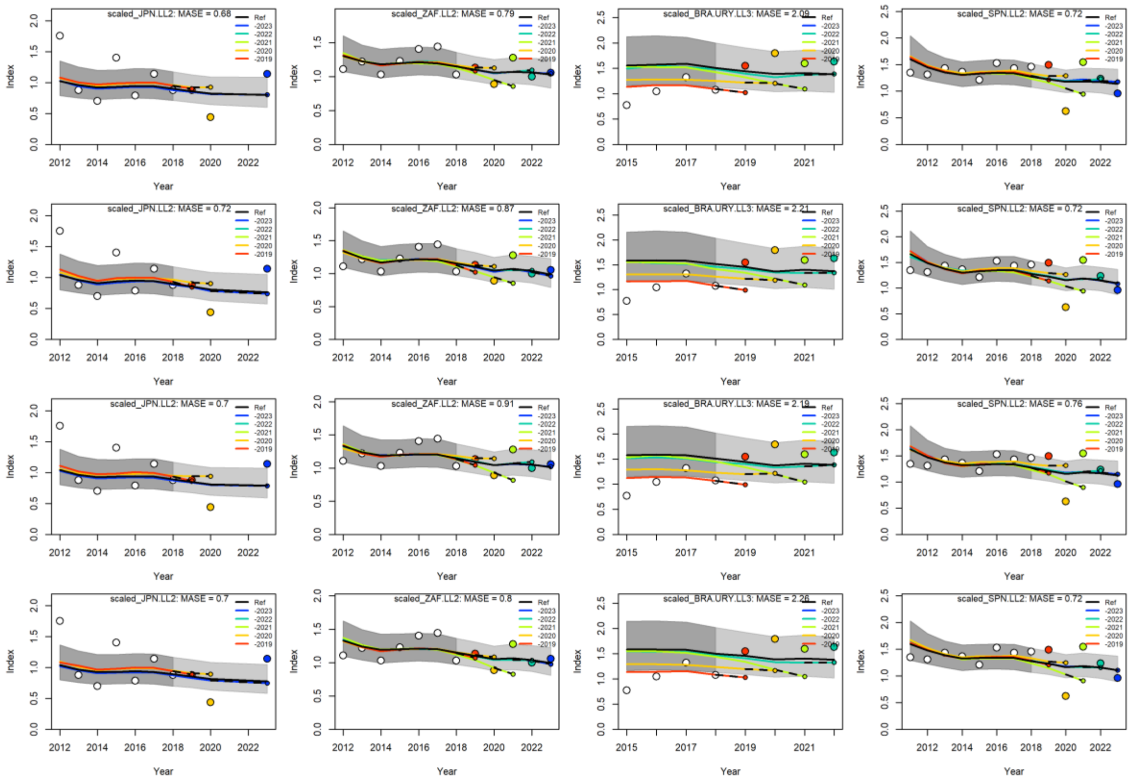

Figure 10. Hindcasting cross-validation results for the index available in the last years of the model, run for the 4 main base grid models. The plots show 1-year-ahead forecasts of CPUE values, when the last years are removed one at a time, relative to the observed CPUE using all data. The CPUE observations, used for cross-validation are highlighted as the color-coded solid circles with associated light-grey shaded 95% confidence interval.

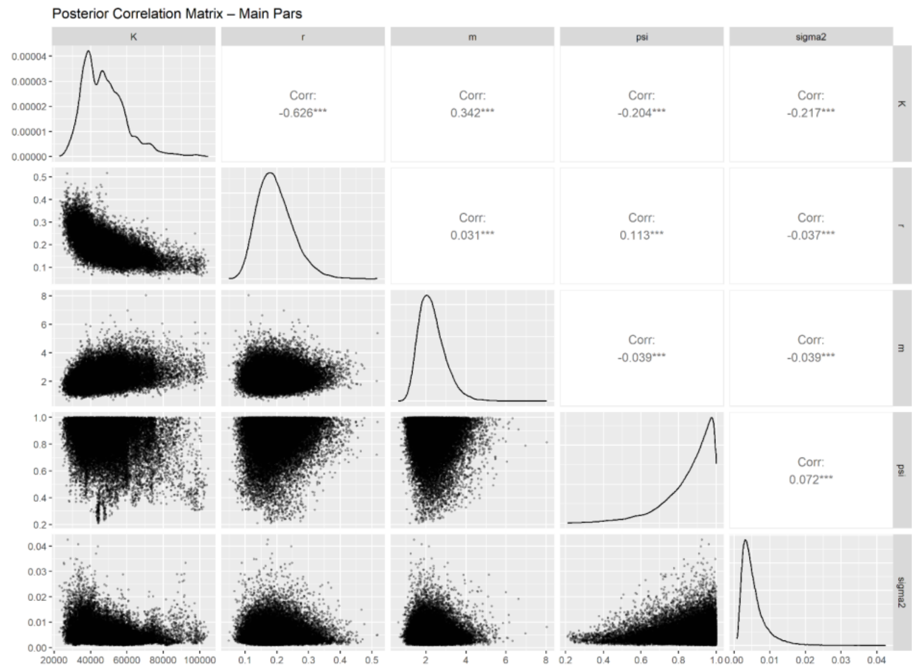

Figure 11. Correlation pairs plots for the posterior estimates of the base case model.

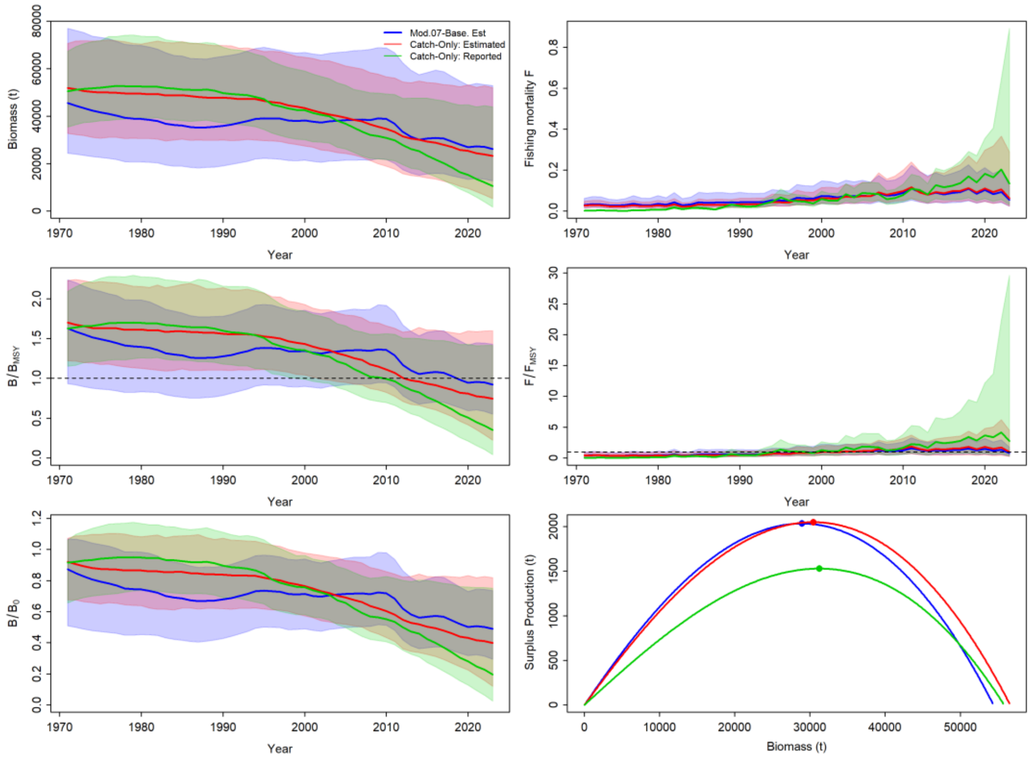

Figure 12. Sensitivity analysis relative to a catch-only model (using both estimated and reported catches) and without any CPUE information. The analysis was carried out in relation to the base case model (Mod 07).

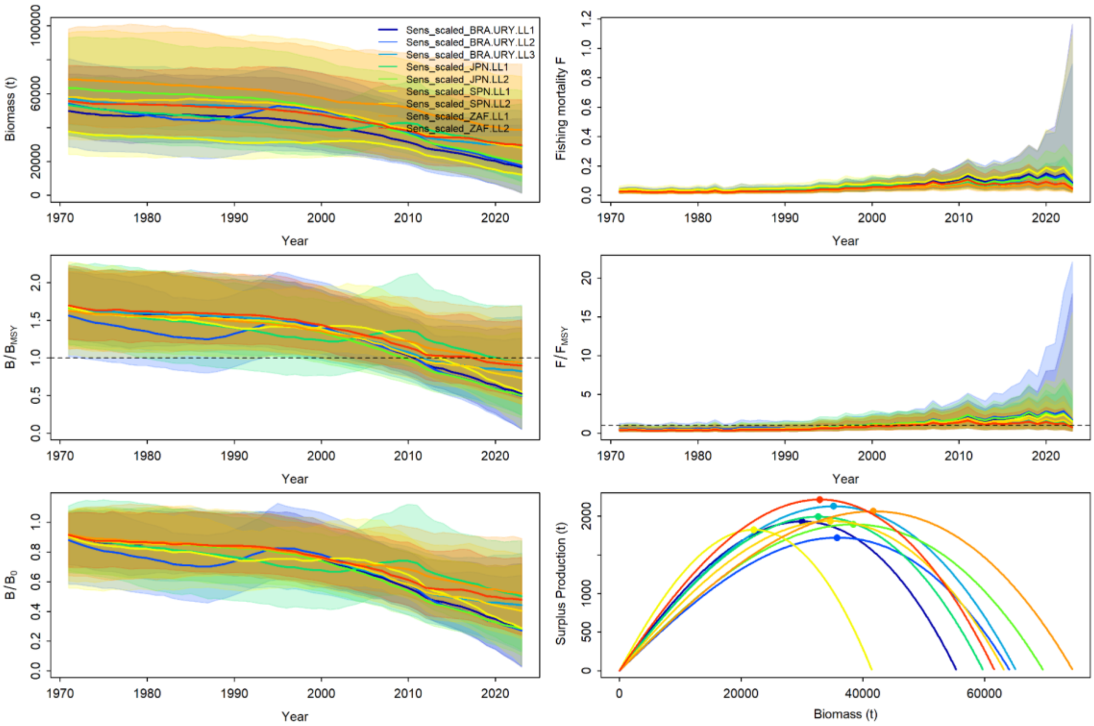

Figure 13. Sensitivity analysis to using one CPUE at a time, for all available CPUE series after being split in the various time periods proposed.

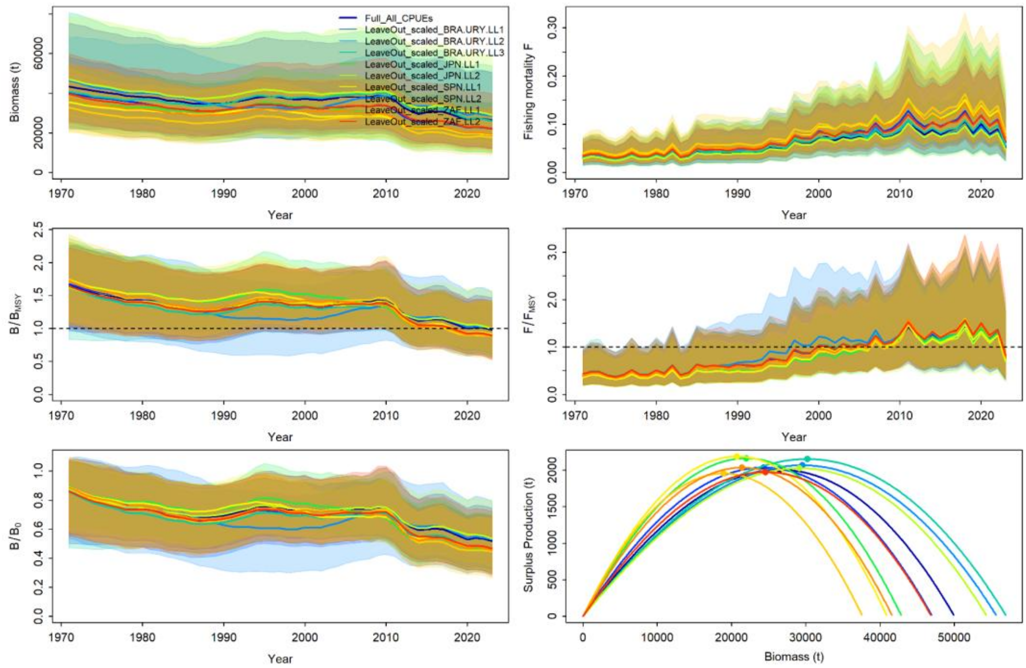

Figure 14. Sensitivity analysis for leaving out one CPUE at a time, for all available CPUEs with the split time periods, as used in the base case model for the ICCAT SMA South Atlantic assessment.

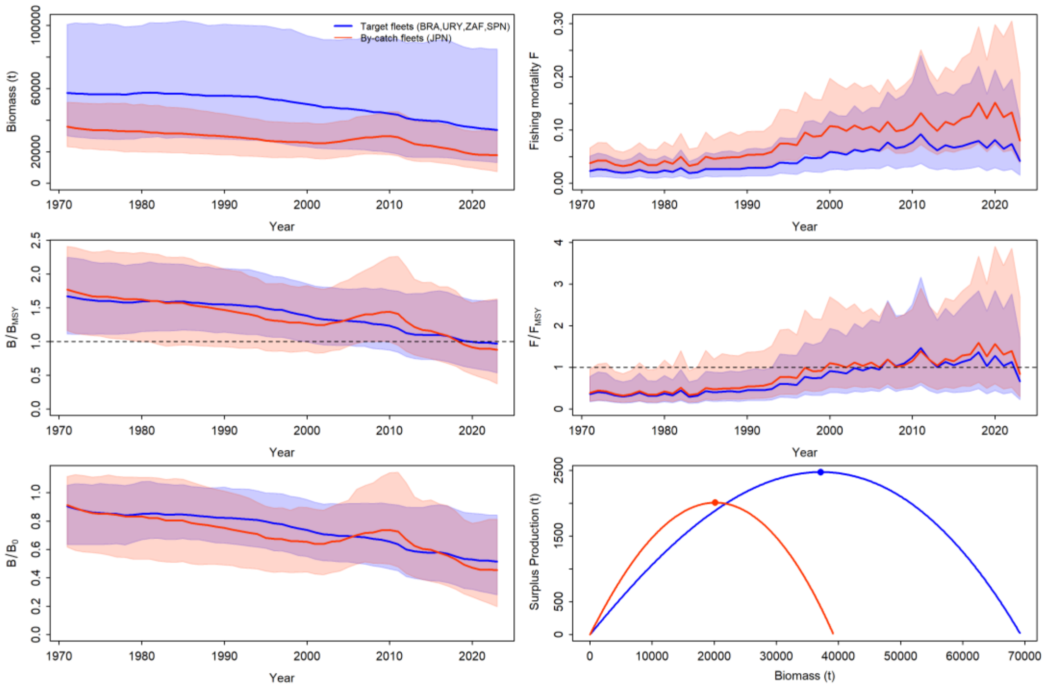

Figure 15. Sensitivity analysis using CPUEs from Sharks/SWO targeting fleets (BRA-URY, ZAF, SPN) versus fleets where those species are mainly bycatch (JPN).

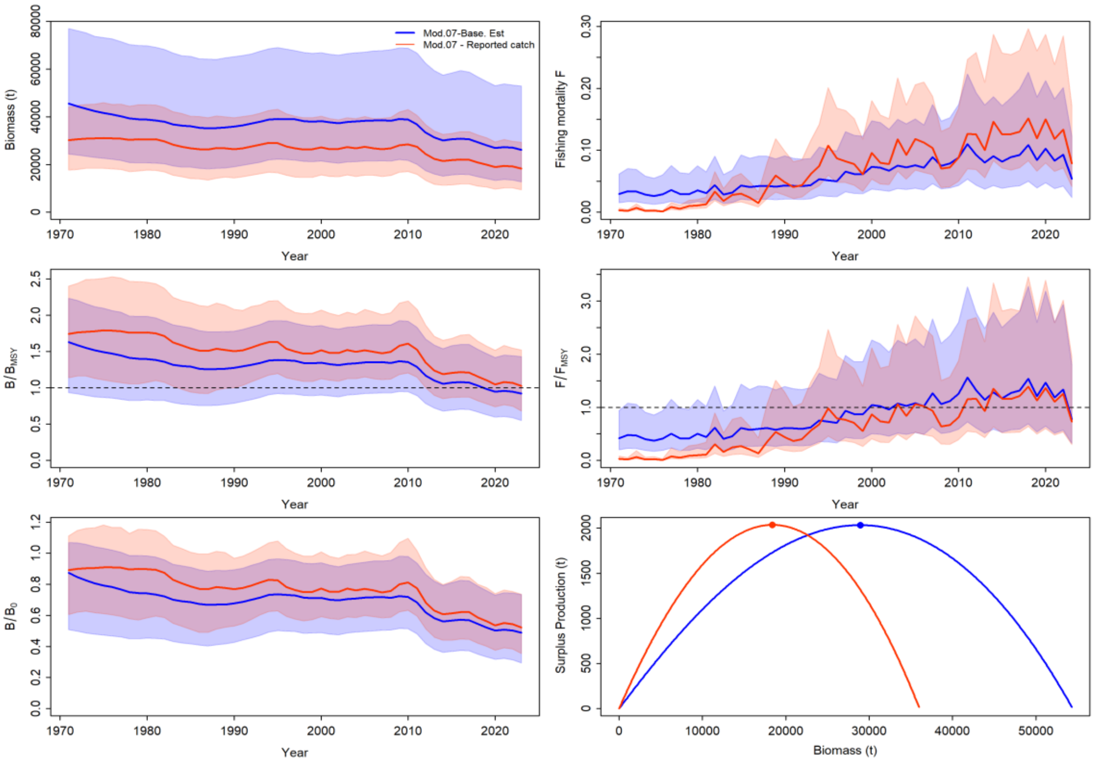

Figure 16. Sensitivity analysis using the ICCAT reported catches instead of the reconstructed catches, for the base case model for the ICCAT SMA South Atlantic stock assessment.

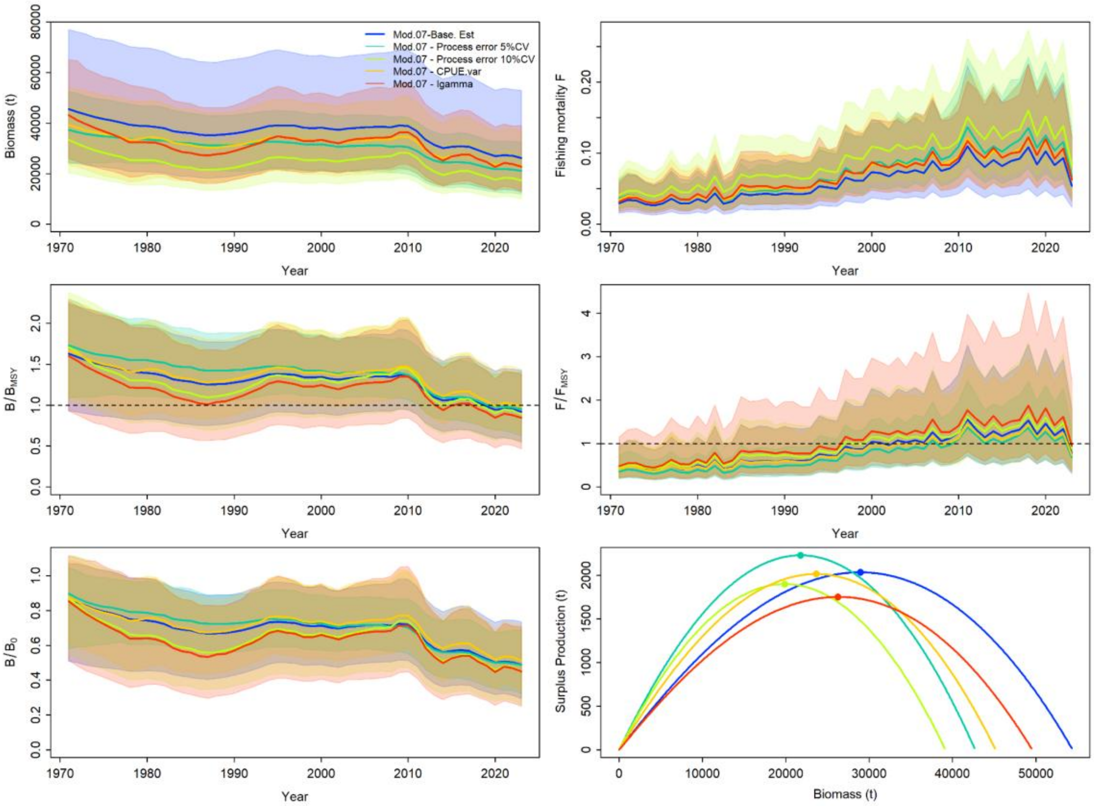

Figure 17. Sensitivity analysis relative to various options for the sigma of the process error ("igamma" = using an igamma vaguely informative prior; "process error 5%" = fixing at 5%; "process error 10%" = fixing at 10%) and turning off the estimation of additional CPUE variance in the models (CPUE.var), performed for the 2025 ICCAT SMA South Atlantic stock assessment. The analysis was carried out in relation to the base case model (Mod.07).

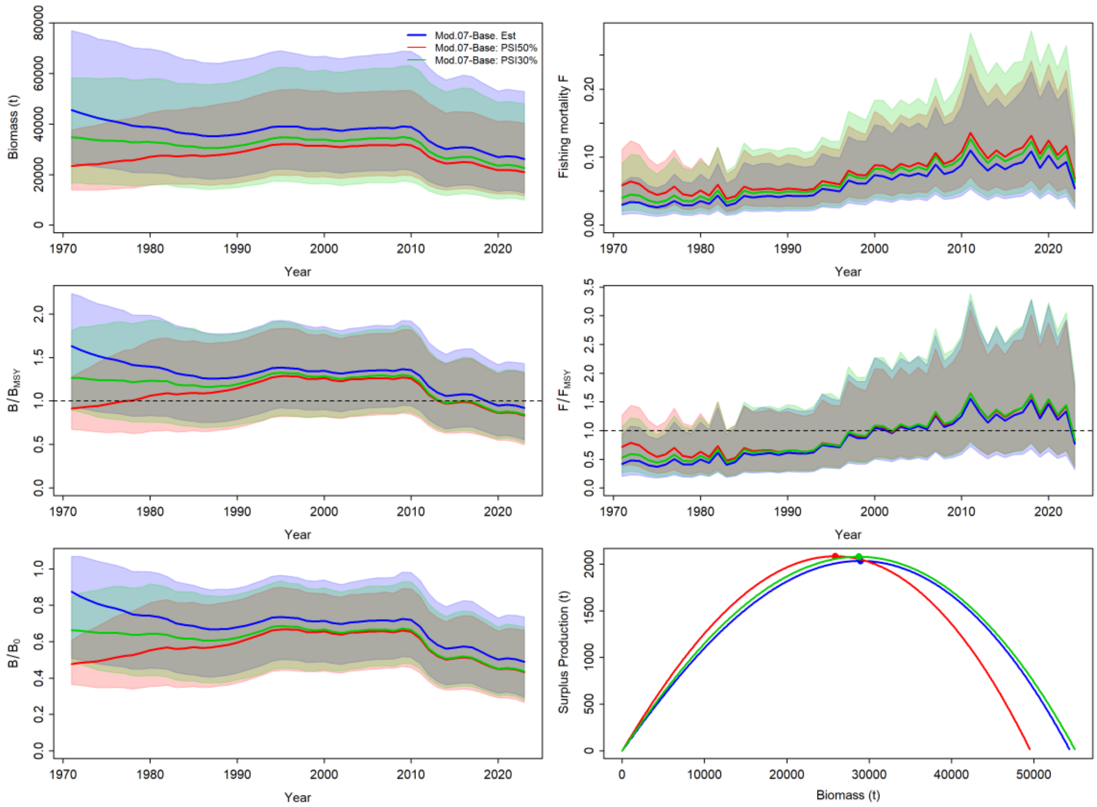

Figure 18. Sensitivity analysis relative to various options in terms of initial depletion levels, namely the base model (prior with 10% depletion) and additional options uwing initial depletion at 30% and 50%. The analysis was carried out in relation to the base case model (Mod.07).

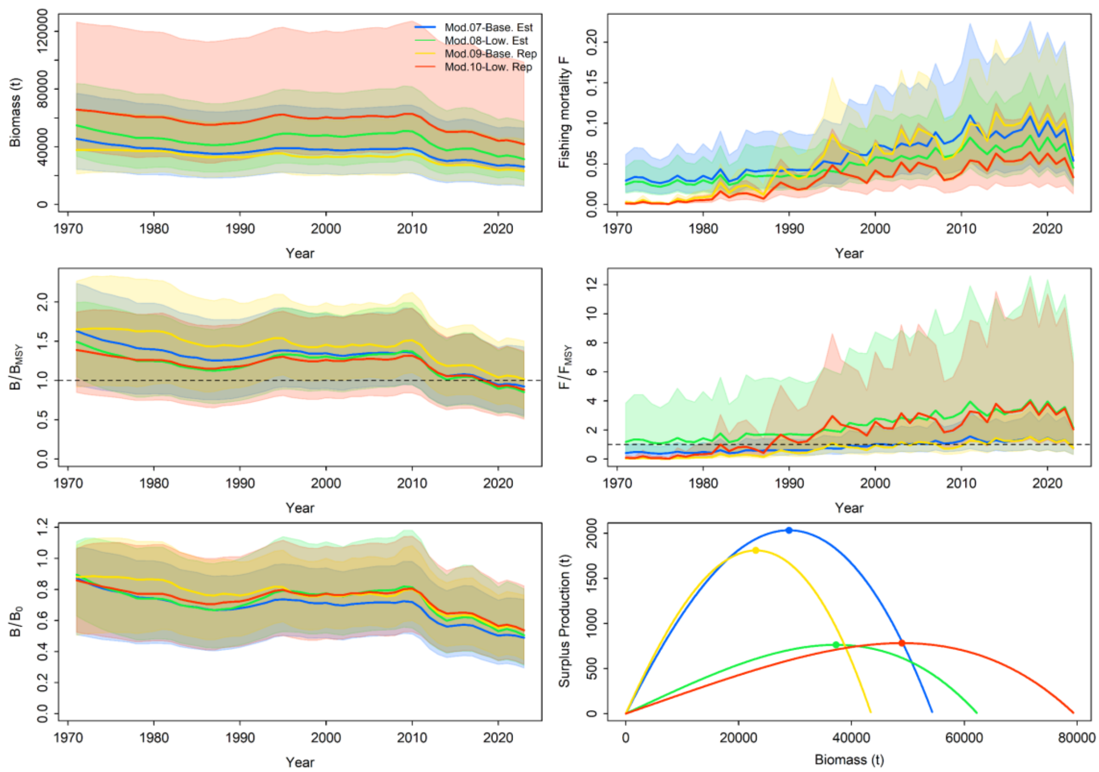

Figure 19. Comparative trends and trajectories of the 4 main base grid models, run for the 2025 ICCAT SMA South Atlantic stock assessment.

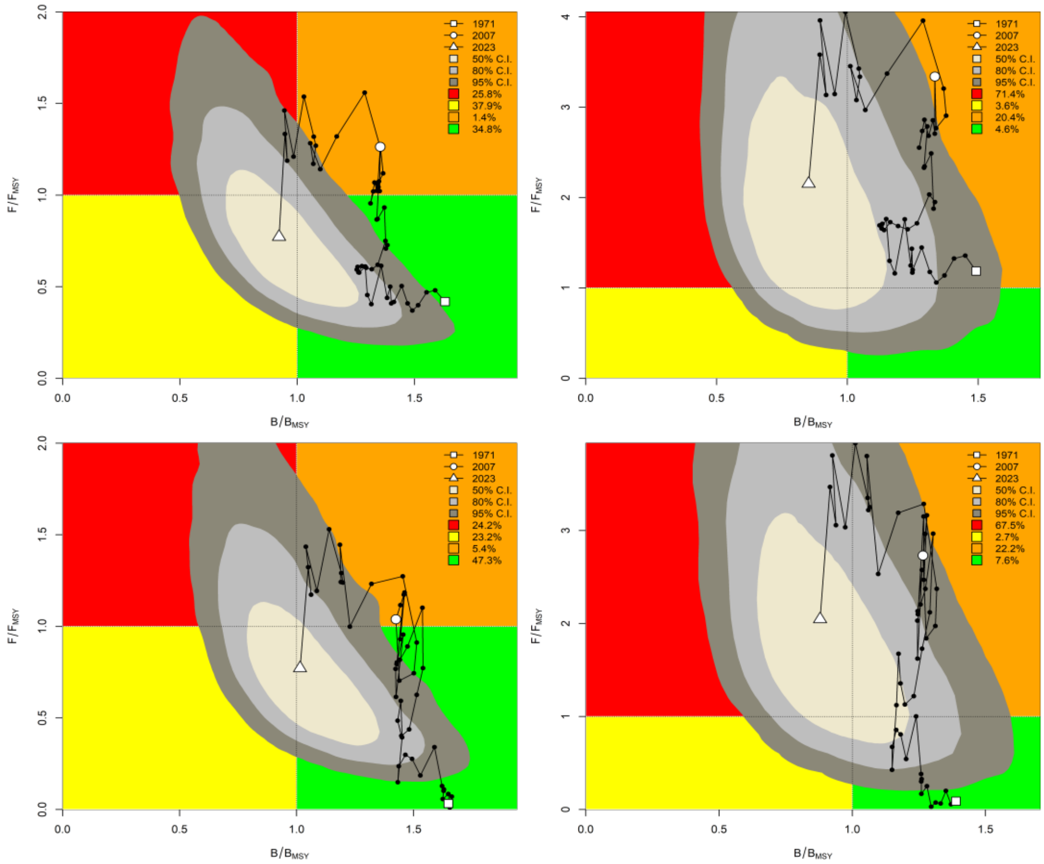

Figure 20. Kobe phase plot with the estimated trajectories (1971-2023) of B/Bmsy and F/Fmsy for the 4 main base grid models considered for the 2025 ICCAT SMA South Atlantic stock assessment. The different gray shaded areas denote the 50%, 80%, and 95% credibility intervals for the terminal year of the assessment data (2023). The probability of the terminal year stock status falling within each quadrant of the Kobe phase plot is indicated in the figure legend, for each of the grid models.

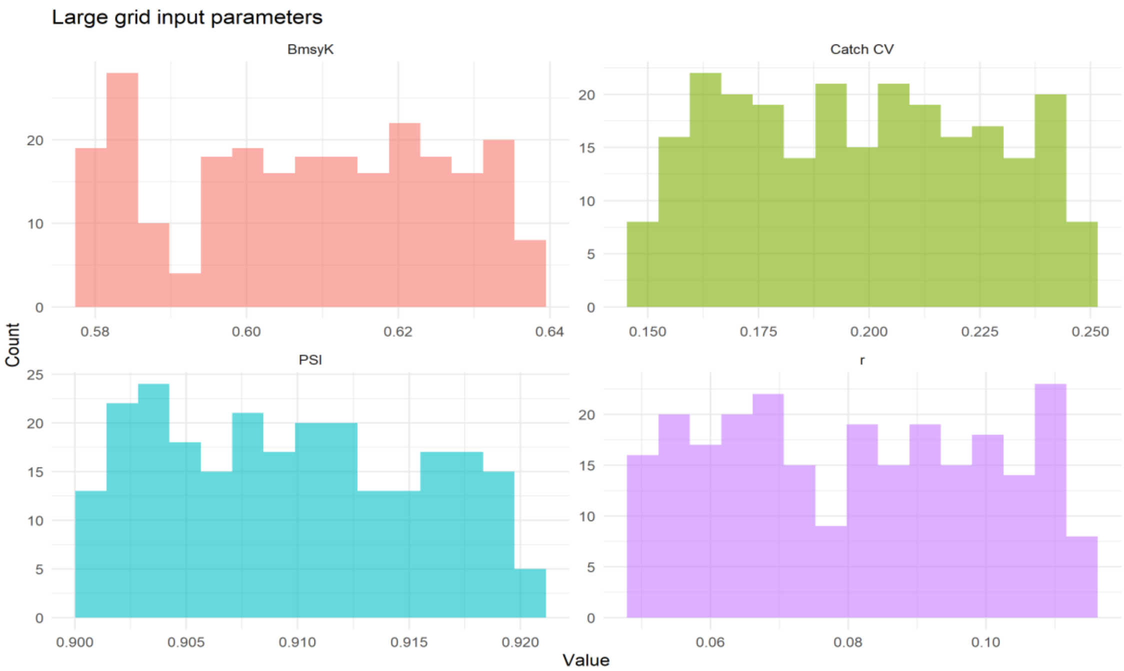

Figure 21. Distribution of the input prior values that was randomly generated for the large model grid ensemble.

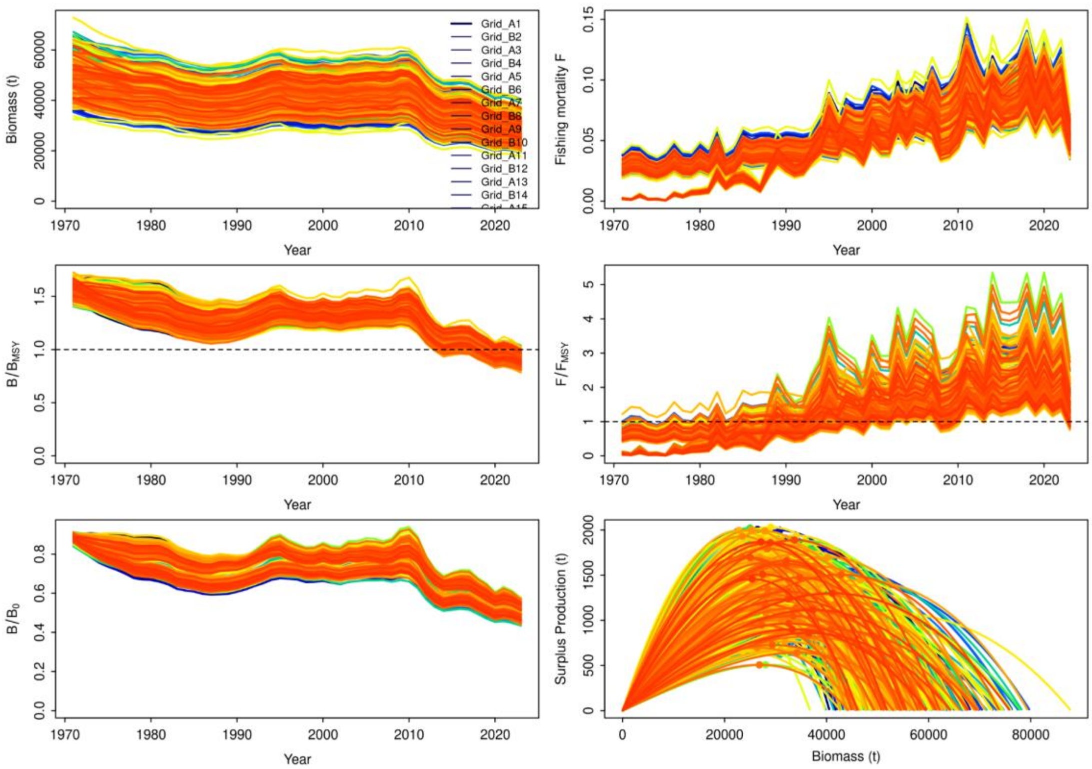

Figure 22. Results of the trajectories and main results from the large ensemble grid models (453 models that have converged).

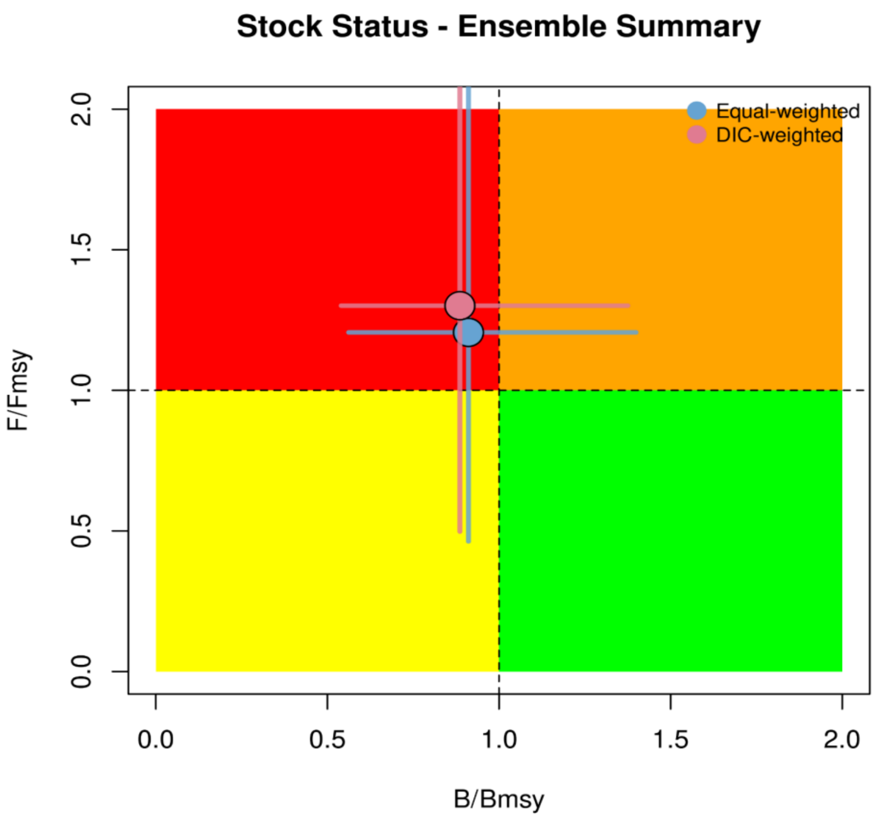

Figure 23. Kobe plot with the point estimates and CIs of the summarized large grid models, using two options for model weighing: 1) equal weights across all models; 2) using DIC-weights across models. The summaries are done with the 453 large grid models that have converged and are used for this final analysis.

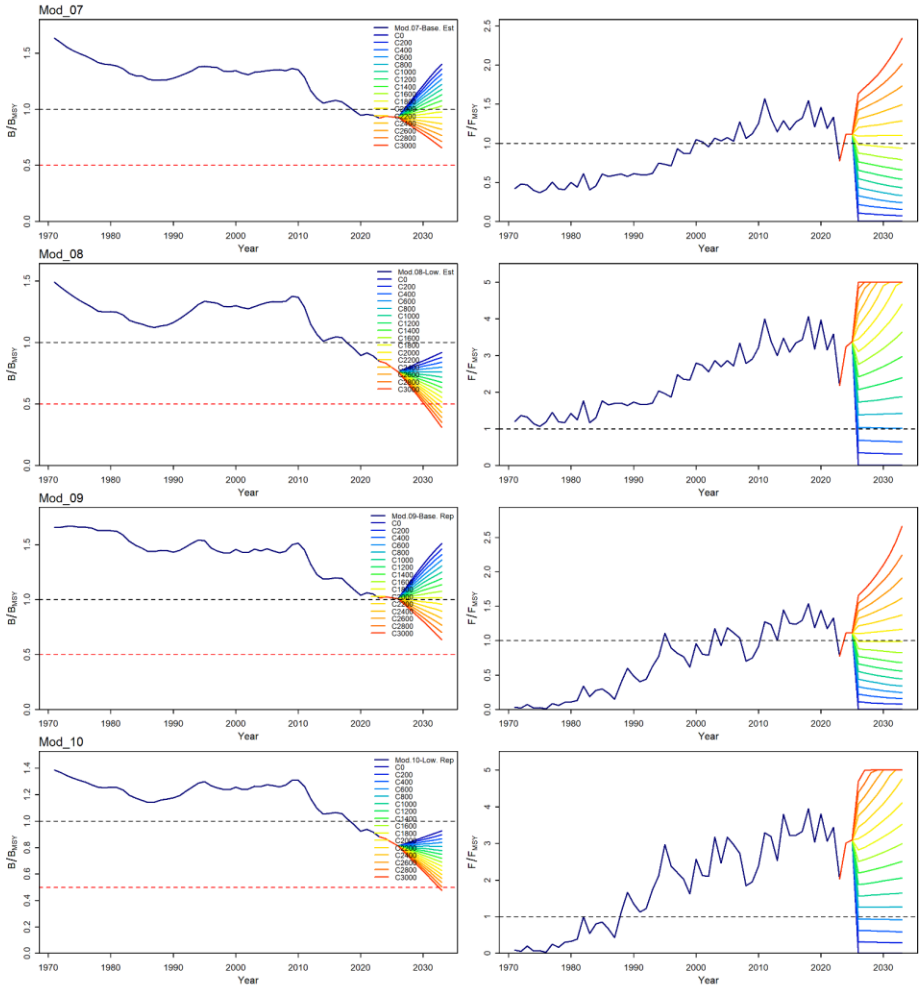

Figure 24. Preliminary individual deterministic projections for each of the 4 main base case models considered. The 4 scenarios are: Mod_07: base productivity with estimated catches; Mod_08: low productivity with estimated catches; Mod_09: base productivity with reported catches; Mod_10: low productivity with reported catches.