INDIAN OCEAN BLUE SHARK STOCK ASSESSMENT USING BAYESIAN SURPLUS PRODUCTION MODELS (JABBA): MODEL DEVELOPMENT, VALIDATION, SENSITIVITY ANALYSIS AND LARGE GRID MODEL ENSEMBLES

Rui Coelho[1,2,*]

SUMMARY

This paper provides the stock assessment for Indian Ocean blue shark using JABBA, a Bayesian surplus production models. Four base models were constructed with combinations of two productivities and two options for catch series. Models were checked for goodness of fit and validated, and sensitivity analysis was conducted. A large model grid (600 models) was run, by randomly selecting priors from distributions built from the plausible for their values, and using alternatively each of the two-catch series. Stock status for the main base models ranged from not overfished and not undergoing overfishing (B>Bmsy & F<Fmsy), to not overfished but currently undergoing overfishing (B>Bmsy & F>Fmsy). The stock status for the large grid ensemble was weighed in 2 alternative ways (equal-weighting and DIC-weighting), and resulted in a stock status not overfished and not subject to overfishing (B>Bmsy & F<Fmsy).

KEYWORDS: Bayesian statistics, blue shark, ensemble model grids, model weighing, 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)

Contents

1. Introduction 4
2. Material and methods 4
  2.1. IOTC BSH nominal catches and catch reconstructions 4
  2.2. Life history and demographic analysis 5
  2.3. Standardized CPUEs series 5
  2.4. Stock assessment 7
    2.4.1. Assessment platform 7
    2.4.2. Stock assessment model specifications 7
    2.4.3. Model diagnostics 9
  2.5. Model ensemble 10
    2.5.1 Model ensemble specifications 10
    2.5.2 Model weighting and averaging 10
  2.6. Projections 11
3. Results 11
  3.1. Initial model development and configuration 11
  3.2. Model goodness of fit and validation 12
    3.2.1. Model convergence 12
    3.2.2. Fit to CPUE indices 13
    3.2.3. Process error deviations. 16
    3.2.4. Retrospective analysis 17
    3.2.5. Hindcast cross validation 19
  3.3. Sensitivity analysis 20
    3.3.1. Catch only models 20
    3.3.2. Using one CPUE at a time 21
    3.3.3. Leave out one CPUE at a time 22
    3.3.5. Comparison of catch options series 23
    3.3.6. Process error and CPUE variance 24
  3.4. Stock status of base models 25
  3.5. Model ensemble 28
  3.6. Preliminary projections 33
4. Conclusions 35

5. Acknowledgments 36
6. References 36

1. Introduction

The blue shark is the most widely distributed pelagic shark species. It is captured both as target species and as bycatch is many oceanic pelagic fisheries worldwide, including in the Indian Ocean under the jurisdiction of IOTC.

Previous assessments have been carried out by IOTC for the blue shark, with the last one carried out in 2021. At the time, the stock status was based on an integrated age-structured model (SS3), using data up to 2019, with the base case model using GAM-based catch history estimates and CPUE series from South Africa, EU-Portugal, EU-France (Reunion), EU-Spain, Taiwan and Japan. The major sources of uncertainty identified at the time were the catches and CPUE indices of abundance. At the time and given the results of that stock assessment, the stock status was determined to be not overfished and not subject to overfishing (IOTC, 2024).

In 2025 the IOTC is conducting a new stock assessment for the Indian Ocean blue shark. While it is most likely that the base case advice will again be based on SS3 models, it is useful to have additional stock assessment models run, that can be used to either confirm the stock status, or help identify any issues. In that sense, the use of production models that are simpler and track the exploitable biomass of the stock, can be useful.

The purpose of this paper is to present the configuration, inputs and results of production model (JABBA) for the IOTC blue shark, to support the management advice for this pelagic shark species to IOTC.

2. Material and methods

In March 2025, an IOTC Working Party on Ecosystems and Bycatch (WPEB) 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 (IOTC, 2025). 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.

2.1. IOTC BSH nominal catches and catch reconstructions

In the last stock assessment for IOTC blue shark, GAM based estimated catches were used. During the WPEB data-preparatory meeting some discussions on the catches to be used were carried out, but there was no definitive agreement on what series was to be used. In the intersessional period, the secretariat and contractor updated the nominal and estimated catch series, that are represented in **Figure 1**. Those options contain 3 datasets (D1, D2 and D3), each with Nominal catches, Ratio based estimates and GAM based estimates.

Some preliminary observations from the contractor mentioned that the D2 dataset provided strange and unexplainable results, with very high values in the earlier period. In addition, in the last assessment, the Ratio based estimates were determined to be worse than the GAM based estimates. As such, for this preliminary analysis, we used the GAM based catch estimates based on both the D1 and D3 datasets.

Those 2 series are considerably different, especially in the more recent years after around 2000, with the D1 estimates continuing to show higher catches until the recent period, while the D3 catches decrease considerably after that period.

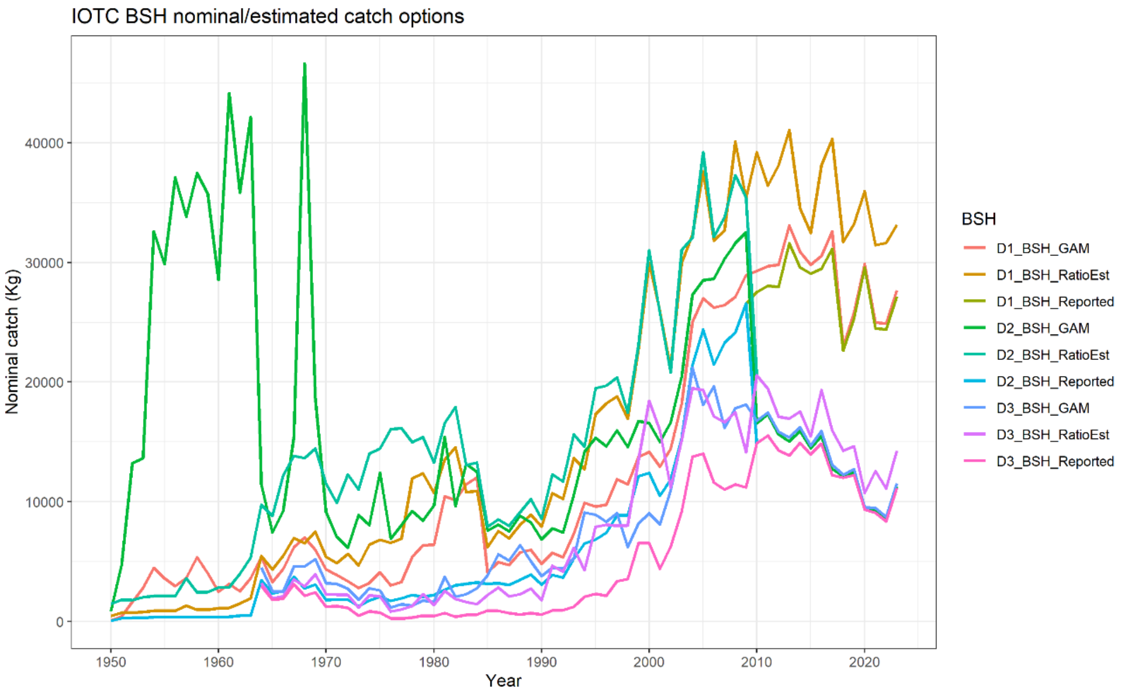

Figure 1: Time series of the IOTC BSH nominal and estimated catch data options considered.

2.2. Life history and demographic analysis

The life history and demographic parameters for the blue shark Indian Ocean have been previously studied, using demographic methods, specifically Leslie matrices analysis. We followed the guidelines from the data preparatory meeting, and based the initial model on the values from Rosa and Coelho (2016) that estimated *r* values between 0.32 and 0.35. Additionally, we also used a 2nd scenario using the demographic models from Geng et al. (2021) with slightly lower estimates of *r*, specifically varying between 0.26 and 0.32.

2.3. Standardized CPUEs series

The CPUEs series that were initially considered were those available and discussed at the data preparatory meeting (IOTC, 2025). Those included series from the following CPCs fleets (**Figure 2**):

- South Africa LL (2000-2023)
- France La Reunion LL (2007-2023)
- Portugal LL (2000-2023)
- Spain LL (2001-2023)
- Japan LL – time block 1 (1994-2008)
- Japan LL – time block 2 (2009-2023)
- Taiwan LL (2005-2023)
- USSR historical (1966-1989)

The CVs for each series in the models were the ones produced in the CPUE standardization analysis, except in cases where the values were lower than 0.2, in which cases 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 bae case models.

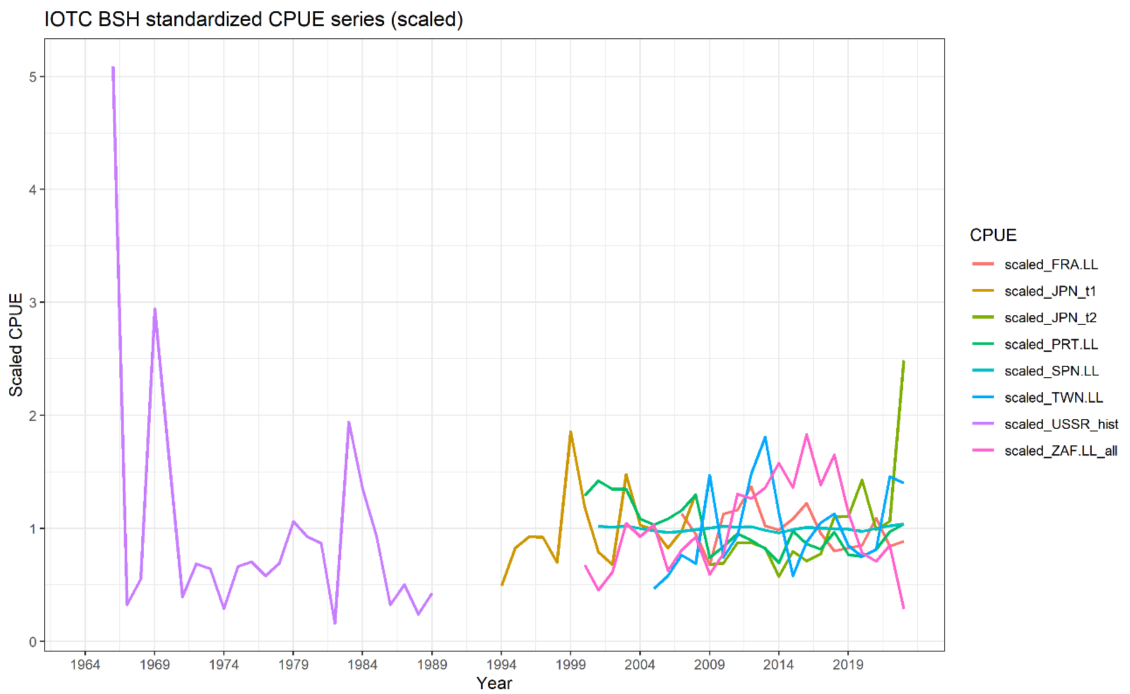

Figure 2: Standardized CPUE series available for the IOTC BSH stock assessment. For better visualization and comparison, each series is scaled by their respective means.

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 2 main models, mostly given the current uncertainty that are associated with the blue shark demographic parameters, and specifically in the calculation of *r*.

As such, 2 base models were constructed with different assumptions on *r*, specifically with one using a *r* value centered at 0.33 (following Rosa and Coelho, 2016) and the other with *r* centered at 0.29 (following Geng et al., 2021). In both scenarios, those values were used as the center of lognormal priors with a CV of 0.1.

The production function used was a Schaefer model, with the shape parameter (*m*) assumed to be 2, which corresponds to a value of Bmsy/K of 0.5, which seems to be a reasonable assumption for shark species.

As described, various time-series of catches were available, but 2 main ones were considered in this assessment, namely the GAM-estimations from Dataset 1 (D1) and Dataset 3 (D3). Those time series have different values in terms of magnitude, so there are some scale differences in the estimations of absolute values such as B0 and MSY. In all cases, the catches were inputted in JABBA with an associated CV of 0.2, therefore allowing some deviation from the observer and reported catches to reflect the likelihood that the catches may not be accurately estimated, recorded and reported to IOTC.

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 0.85 and CV of 10%. This was done, given that the models were initiated in 1964 when some of the estimated catch series started, and with the oldest historical CPUEs available from 1966.

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, we used 3 MCMC chains with 30,000 iterations each, a burn-in period of 10,000 and a thinning rate of 5, resulting in a total of 12,000 posterior samples for each model. 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 the minimum recommended thresholds (ESS > 1,000 for all parameters). This specification was used consistently in the model development phase, as well as in the sensitivity analysis and ensemble grid of models.

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 various estimated catch time series. For the sensitivity analysis, the main base model (base productivity and GAM estimated D1 catches) was used.

2.5. Model ensemble

2.5.1 Model ensemble specifications

After building the base models 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 us to increase 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 0.29 and 0.33, 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;
- PSI prior (initial depletion): Values from a random uniform distribution varying between 0.80 and 0.90.
- Catch series (D1 and D3 GAM estimated catch series): Each set of values taken from the distributions above was run in models using both of those 2 optional catch series.

The large model ensemble with those characteristics was run 600 times (300 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. To achieve this, we used the High-Performance computing platform services of Inductiva (https://inductiva.ai/), enabling high-speed parallel computing of multiple models simultaneously. In this case, the entire 600-model grid was run on 50 cloud machines of type "c2-standard-8", each with 8 vCPUs and 32 GB RAM. The full run took approximately 58 minutes to complete, produced 23.3 GB of outputs, and had an estimated cost of 5.10 USD. For comparison, running this task on a single modern laptop (Intel Core Ultra 7, 4.8 GHz, 64 GB RAM) would take about 3 mins per model, or approximately 30 hours for all 600 models in the ensemble grid. Thus, the cloud computing setup that was used provided an approximate 30× speedup over local sequential model execution.

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 $\exp(-0.5 \times \Delta DIC)$, 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. 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 of the intermediate years (2024-2025), were assumed as the average catches of the previous 3 years (mean 2021-2023).

3. Results

3.1. Initial model development and configuration

The initial model development phase started with a configuration similar to what was used in the last IOTC BSH stock assessment. At the time the JABBA model was used only as an auxiliary model and not used for advice, but some runs were carried out to support the main SS3 model.

In terms of catch data, we started with the series represented as D1-GAM Est, that uses a GAM modeling estimation and follows what was used in the last (2021) stock assessment.

In terms of CPUE series, we started with a model using all available CPUEs from the data preparatory meeting, that was followed by an analysis using CPUEs one-at-a-time. This was done to investigate possible inconsistencies in individual CPUE series, that could be problematic if used in the main models. Most CPUE series have somewhat similar behaviors in terms of stock trends, except the series from South Africa that produces some unreasonable and very different behavior compared to the other CPUEs around the period in the 2010s (**Figure 3**).

It is also noted that two of the series produced very high confidence intervals in the fishing mortality (F) estimation for recent years, namely USSR historical and Japan time block 1. Given that those 2 series are historical series that only contain data from the past and not any recent data, such high confidence intervals in estimations for recent years are expected, given that those series do not contain any information for the recent period.

Given this initial exploration, the final models were based on the use of all CPUE series, except for the South African series, given the conflicts with the other series trends.

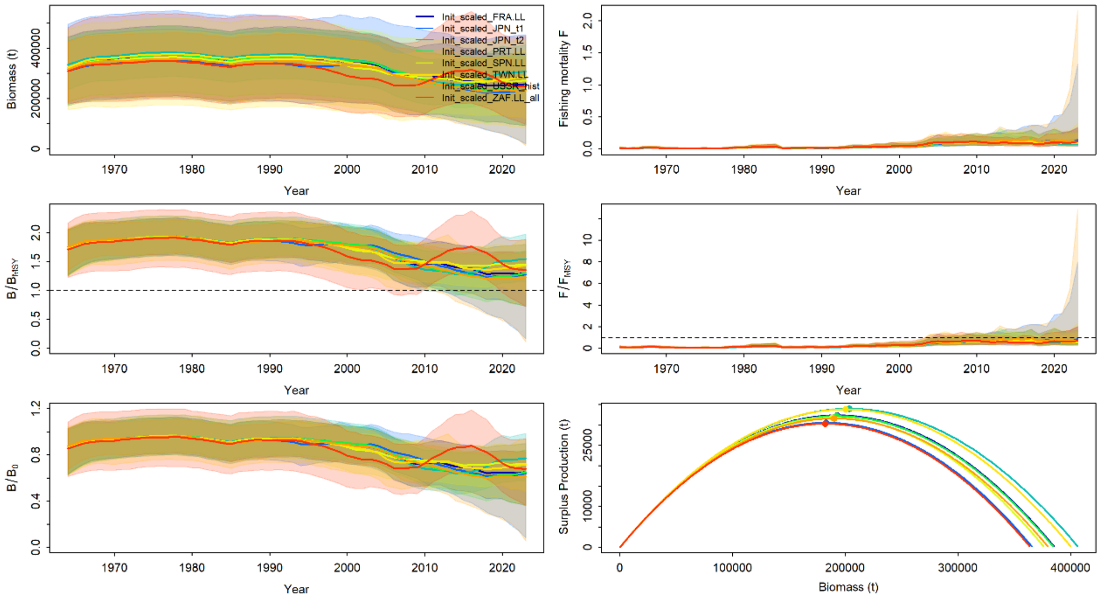

Figure 3. Initial exploratory model runs using one CPUE at-a-time, for the Indian Ocean blue shark.

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**).

Figure 4: Trace plots of the base models used for the IOTC blue shark stock assessment.

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 41.8% and 43.0% (**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. On all the models, the series from Japan (t1) and Taiwan passed the tests, while France, Spain, Japan (t2) and USSR historical failed. In the case of the Portuguese series, it passed the model using D3 catches and higher r, and failed the other tests. 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).

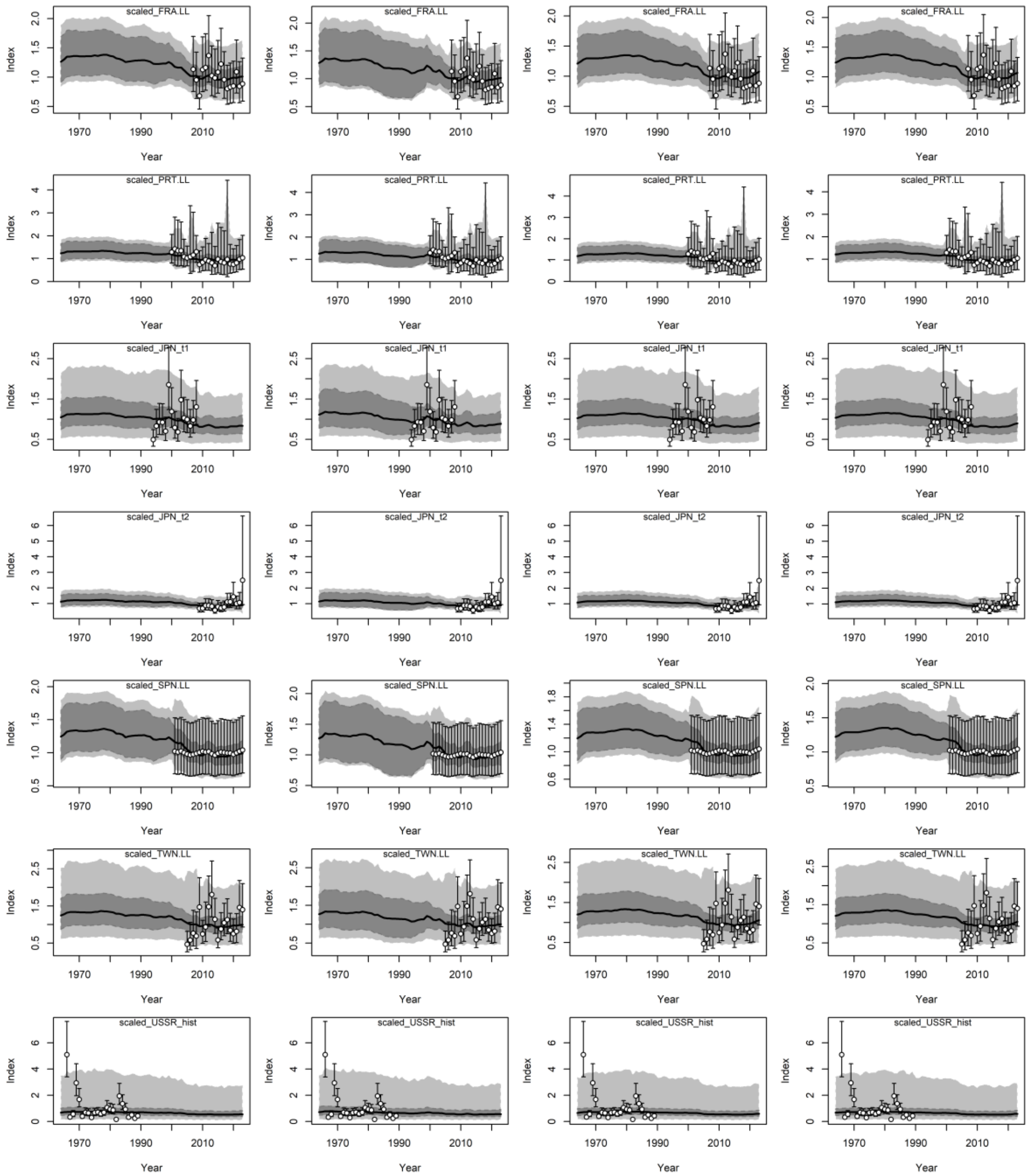

Figure 5: Time series of observed (circles) and predicted (solid line) CPUEs for the IOTC blue shark stock assessment models for the main base models. The dark shaded areas represent the 95% credibility intervals of the expected mean CPUE, and the light shaded areas represent the 95% posterior predictive distribution intervals. The error bars are the 95% confidence intervals (CIs) from the CPUE observations

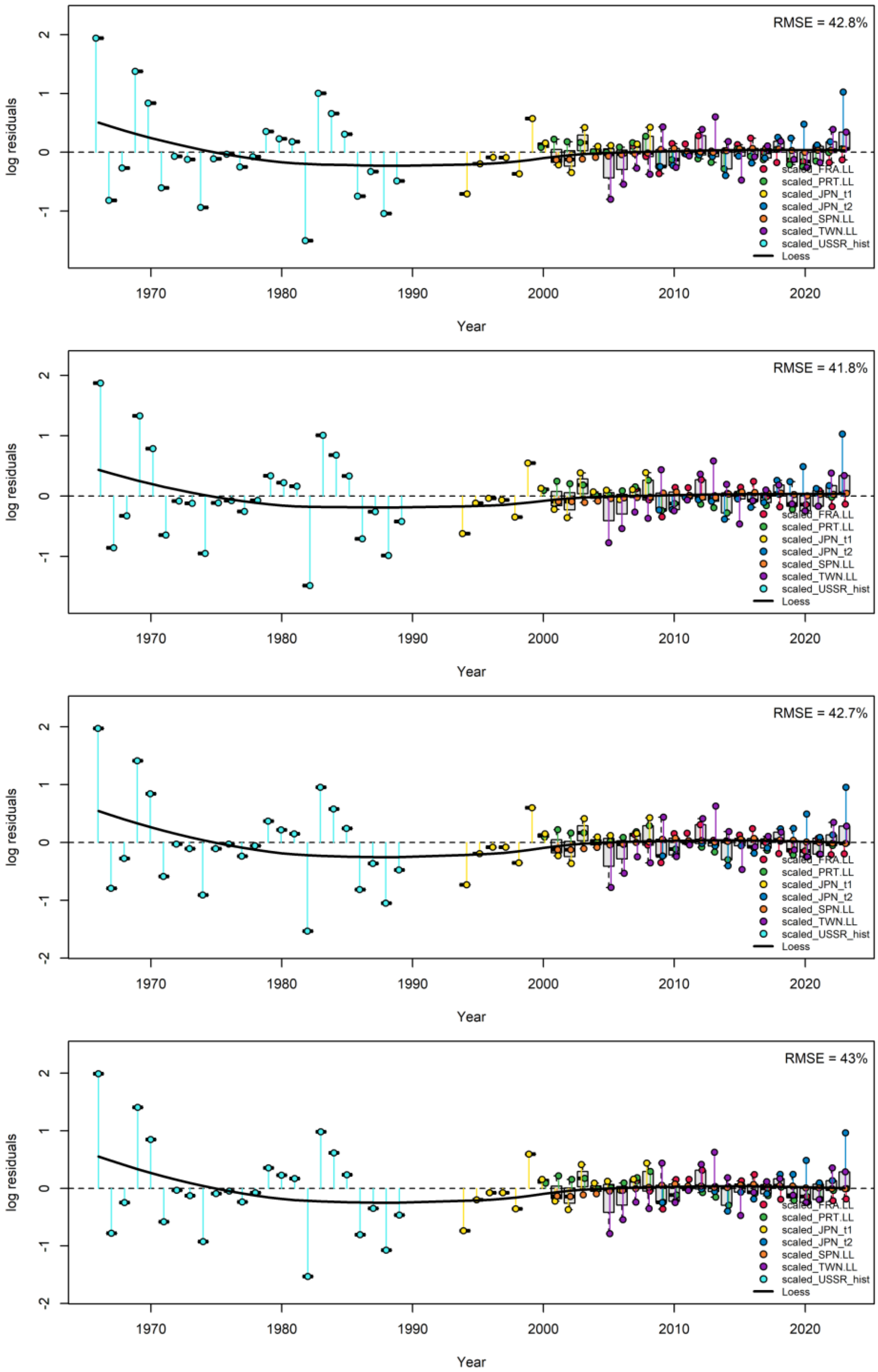

Figure 6: Residuals diagnostic plots for the base models used for the IOTC blue shark stock assessment. Each individual CPUE index and its respective residuals are represented by a different color. The solid black lines represent loess smoothers through all residuals combined.

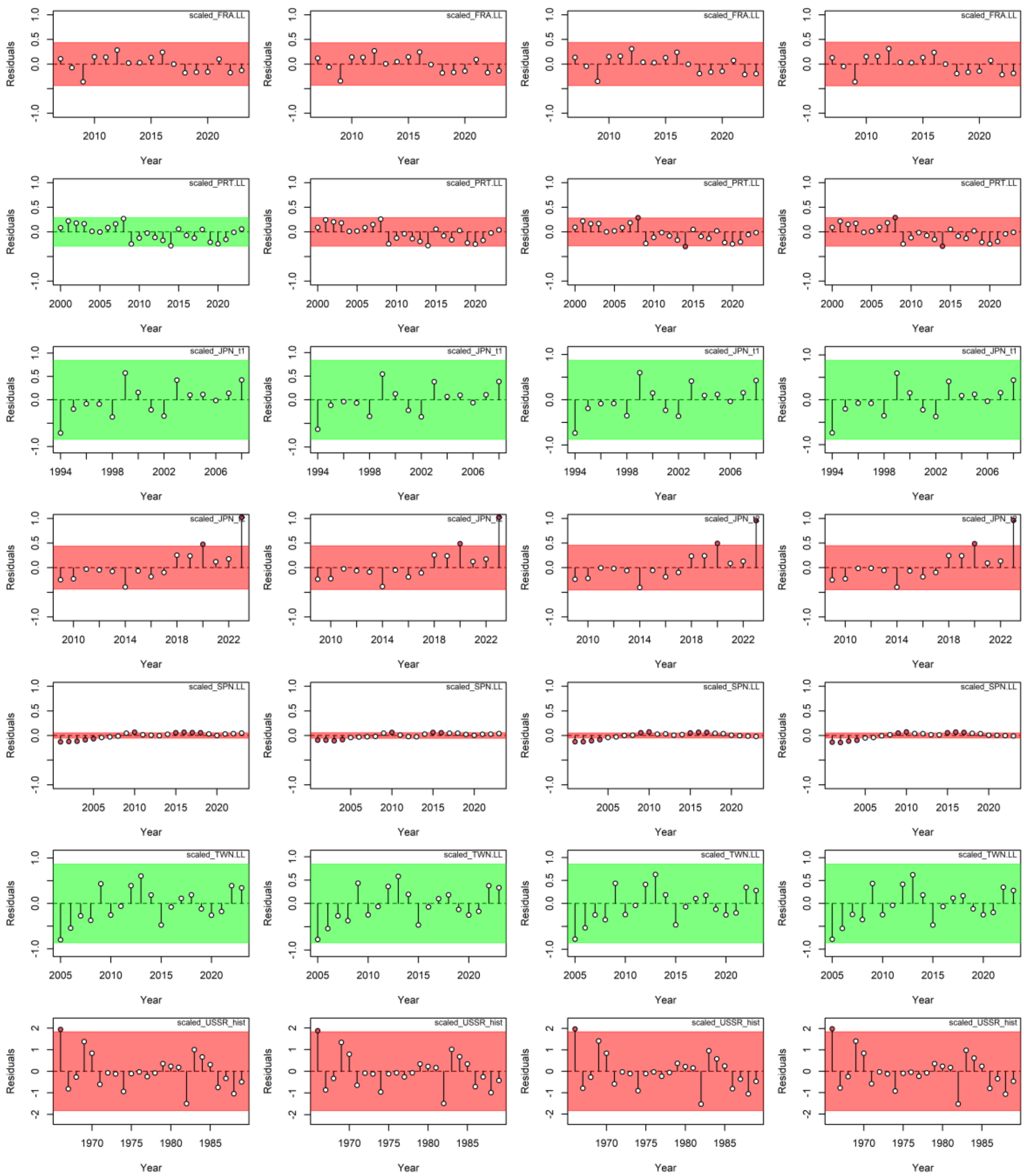

Figure 7: Runs tests for the CPUE index for the base models used in the IOTC blue shark stock assessment.

3.2.3. Process error deviations.

The deviations from the process error show similar patterns for both main models. There were some patterns with lower deviates in the middle time-period, followed by higher deviates in the more recent years, since around 2005 (**Figure 8**). 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.

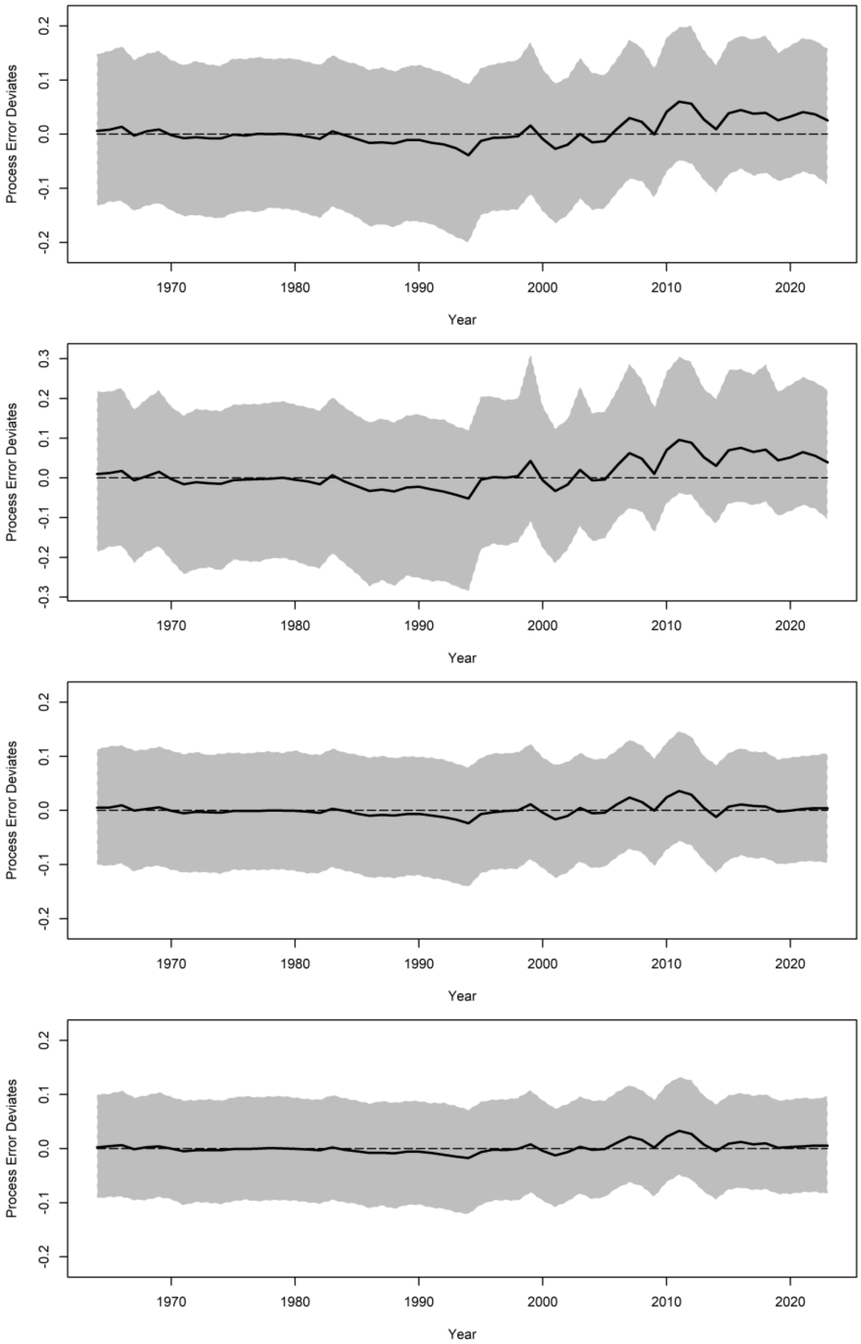

Figure 8: Process error deviates from the main base models used for the IOTC blue shark stock assessment. The solid line represents the median, and the shaded gray area the 95% credibility intervals.

3.2.4. Retrospective analysis

The results of the retrospective analysis applied to the 2 main base models are shown in **Figure 9**, and the corresponding summaries of the estimations of the Mohn's rho are summarized in **Table 1**. All models fell within the acceptable range of -0.15 to 0.20 for 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.

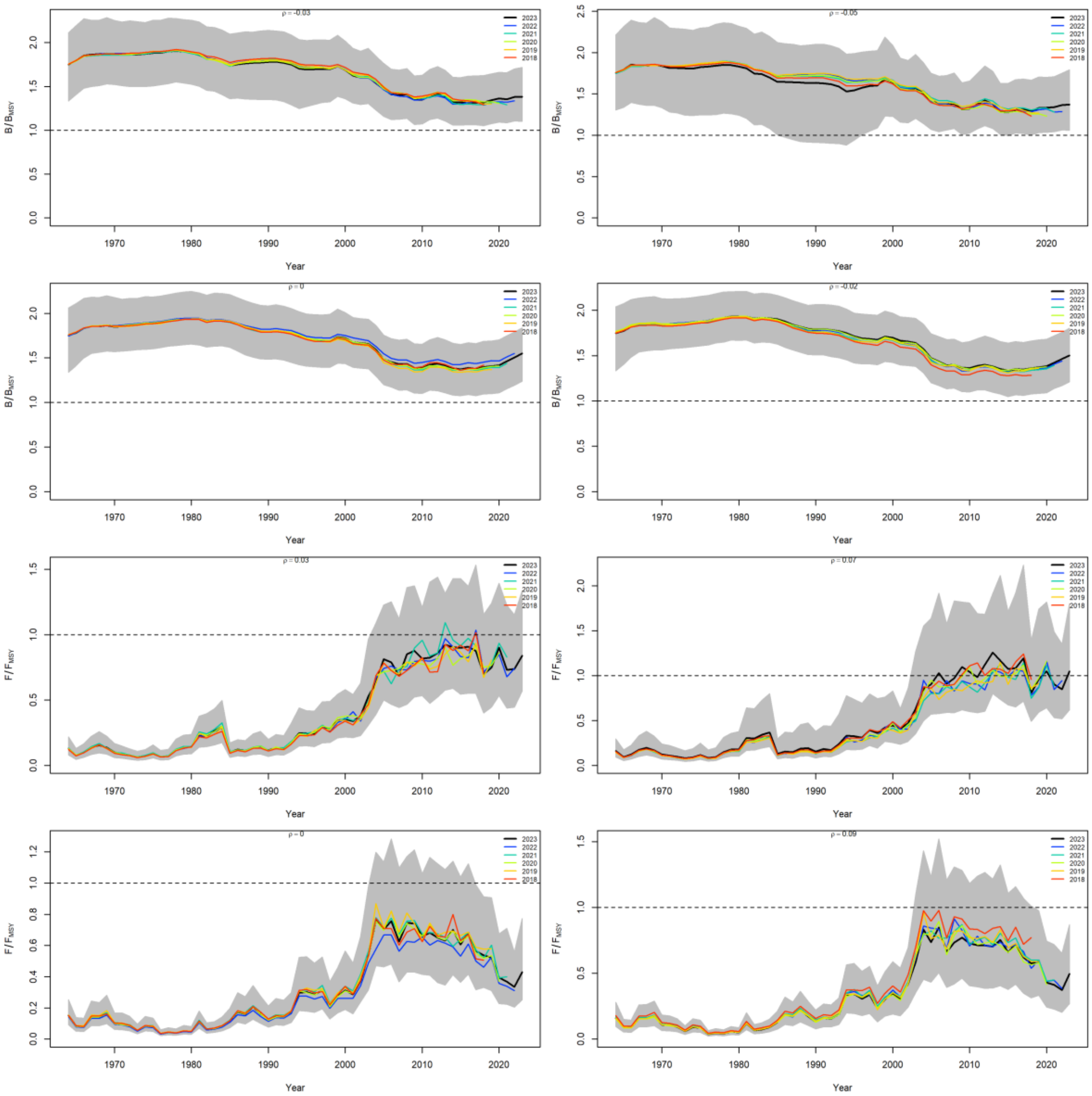

Figure 9: Retrospective analysis conducted for the main base models developed for the IOTC blue shark stock assessment, by removing 1-year at a time sequentially (n=5) and predicting the trends in biomass and fishing mortality relative to MSY (i.e, B/Bmsy and F/Fmsy).

Table 1: Summary of the Mohn's rho statistic computed from the retrospective analysis pattern evaluated for the main base models used for the IOTC blue shark assessment. All values fall between -0.15 and 0.20 and are considered to have an acceptable retrospective bias and are therefore highlighted in green in this table.

Model	B	F	Bmsy	Fmsy	B/K	MSY
Mod.06	-0.002	0.031	-0.034	0.035	-0.019	0.029
Mod.07	0.017	0.068	-0.054	0.068	-0.031	0.051
Mod.08	-0.015	0.005	0.000	0.002	-0.001	-0.015
Mod.09	-0.019	0.082	-0.021	0.087	-0.002	-0.006

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, in many cases fall outside the 95% CIs (**Figure 10**). The mean absolute scaled error (MASE) estimates were under or around the reference level (MASE <= 1) for the cases of France, Portugal, Taiwan. The MASE values were a little higher for the case of Japan (time block 2) and much higher for the case of Spain when using catches D1but better when using catches D3.

When values of MASE are < 1, this indicates that the average forecasts for those indices have good predictive skills (Carvalho et al., 2021), so in this case, the indices that had better prediction skills were from France, Portugal and Taiwan, for the recent period. The worse CPUE in terms of predicted skills were from Spain when using the catch series D1.

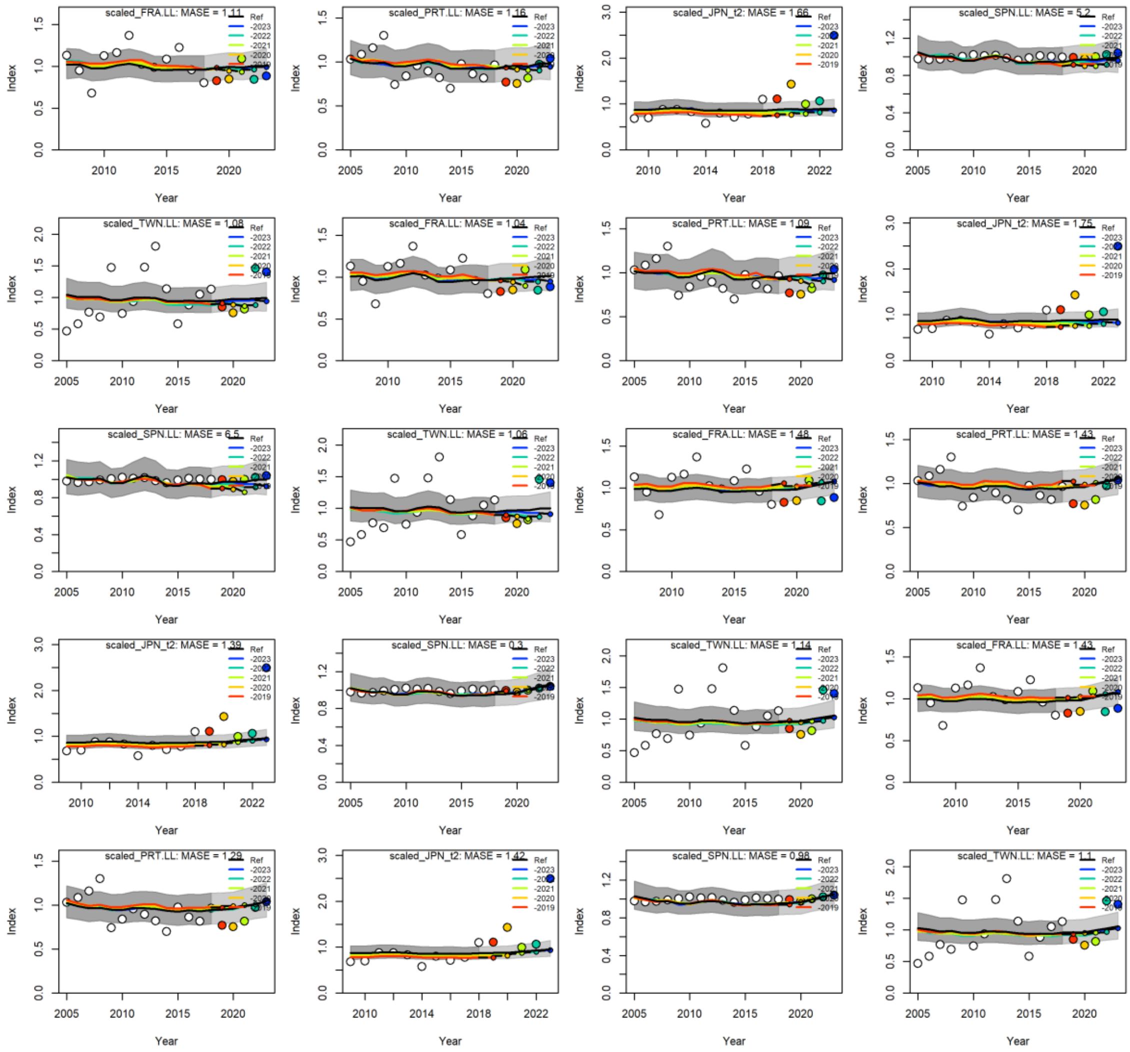

Figure 10. Hindcasting cross-validation results for the index available in the last years of the model, run for the main base models. The plots show 1-year-ahead forecasts of CPUE values, when the last years are removed one at a time, relative to the observed CPUE using all data. The CPUE observations,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

3.3. Sensitivity analysis

With regards to the sensitivity analysis, and given that the 4 main base grid models developed, 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 higher productivity scenario, and the estimated catches from the D1 dataset, which represents the main continuity from the 2021 assessment.

3.3.1. Catch only models

The results of the sensitivity analysis conducted for models with catch only information is shown in **Figure 11**. Those models are mostly informed by the biological prior information, the priors set for initial and final depletion (if used), and the history and trends from the times series of catches.

There are some differences when the CPUE data is entirely excluded. When the D1 catch dataset is used, the biomass trends show an overall and continuous decrease, while the fishing mortality shows an increase. When using D3 catch series the trends are similar until around 2010, but then increase from that point forward, which is reflective of that catch history information, that reduces substantially from that point forward. In both cases, the final status with the catch-only models is more pessimistic, resulting in a stock that is more depleted and has worse stock status.

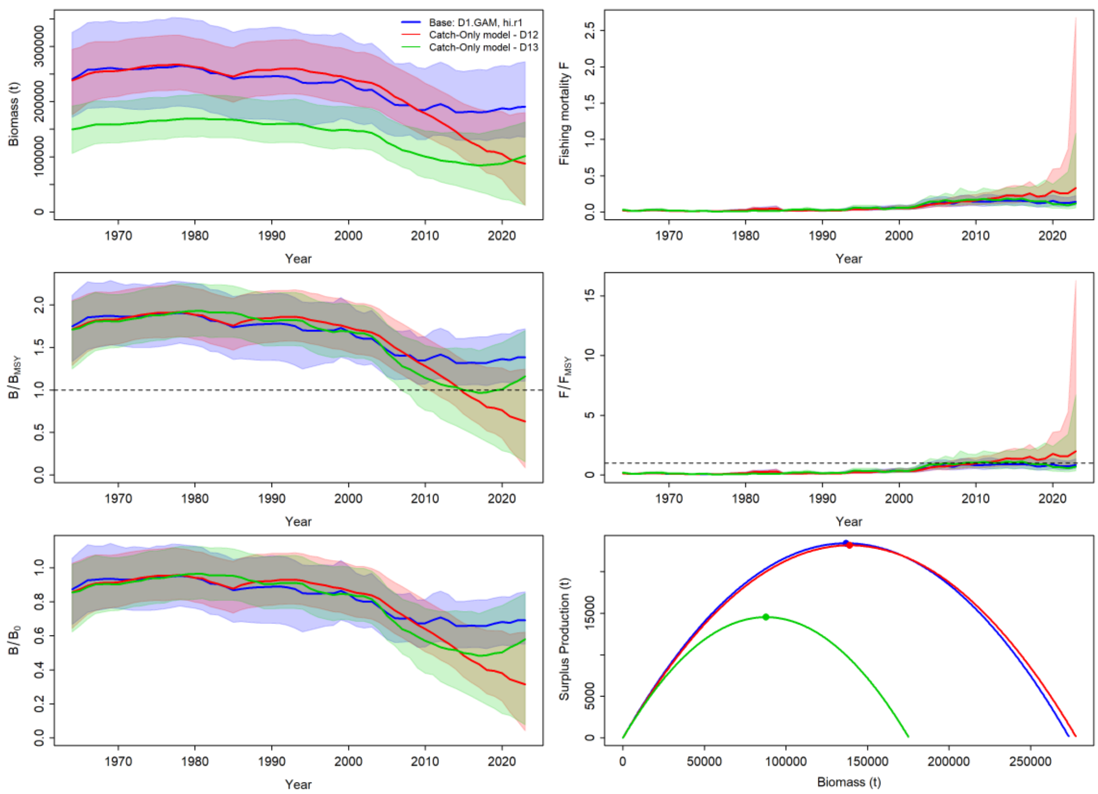

Figure 11: Sensitivity analysis relative to catch-only models (without any CPUE information) for the IOTC BSH stock assessment. The analysis was carried out in relation to the base case model (Mod 06), and contains 2 catch model options, using both D1 and D2 GAM estimated catch series.

3.3.2. Using one CPUE at a time

A sensitivity 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. The results are represented in **Figure 12**.

The results show overall and relatively similar trends in the stock status over time. This indicates that all CPUE series produce relatively similar signals to the stock, and that each one individually is not too strong to introduce signals that are very different from the others.

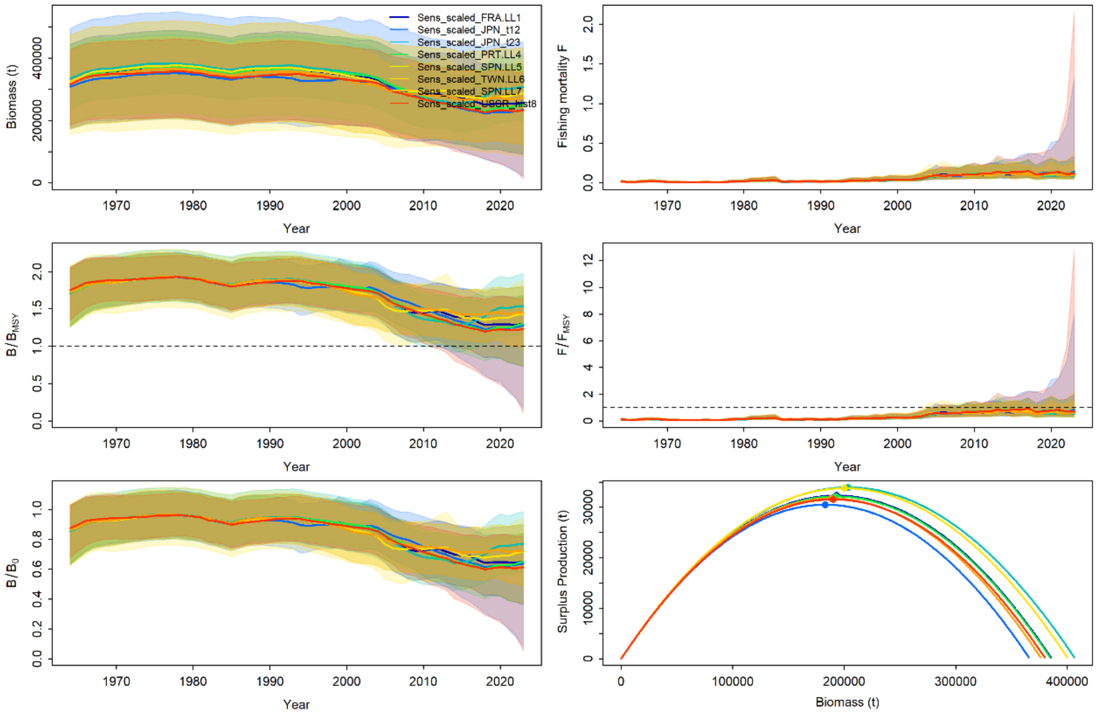

Figure 12: Sensitivity analysis to using one CPUE at a time, for all available CPUE series used in the base models for the IOTC BSH stock assessment.

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 13**.

Overall, there are relatively minor 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.

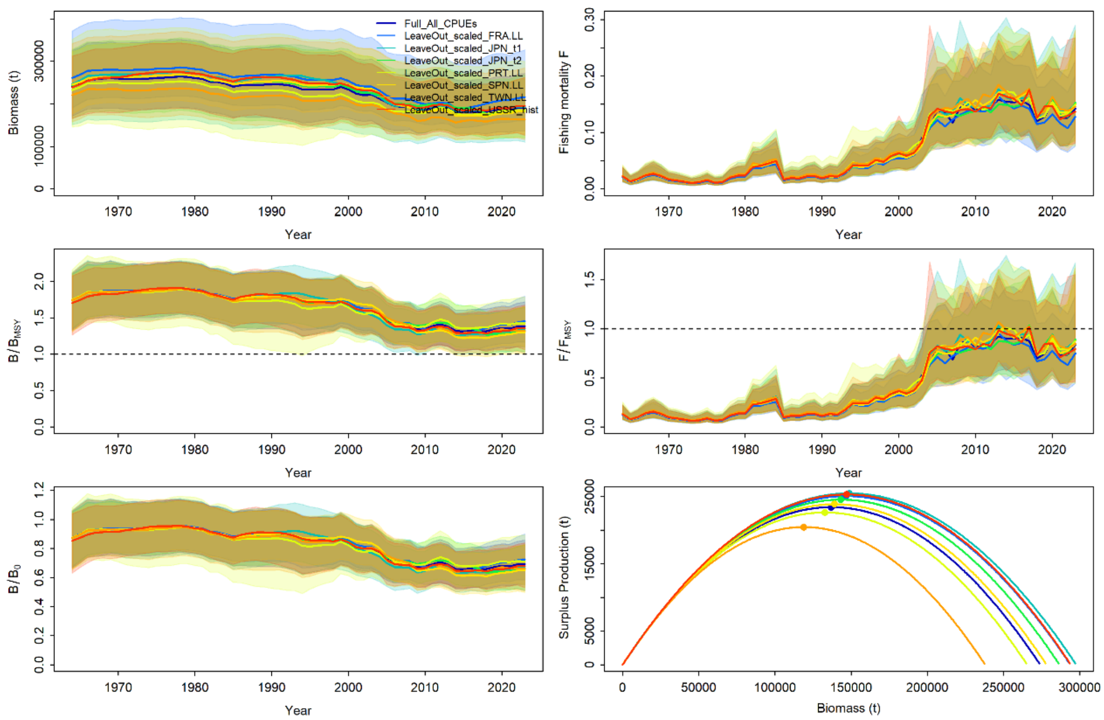

Figure 13: Sensitivity analysis for leaving out one CPUE at a time, for all available CPUEs used in the base case models for the IOTC BSH stock assessment.

3.3.5. Comparison of catch options series

Another sensitivity analysis was conducted for using various catch options for IOTC BSH series. In this case we have 2 options in terms of datasets (D1 and D3), each with the nominal catches, ratio estimations and GAM estimations (this last one used as the base cases).

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. The summaries of those estimates are described in **Table 2**. As expected, the models using the catch series D1 have overall higher MSY estimates given that the catches achieve higher absolute values, compared to catches from the D3 dataset. Nonetheless, in terms of relative trajectories, both are relatively similar on all models in terms of biomass, and the main differences are seen in terms of fishing mortality (**Figure 15**).

Table 2: Summary of the MSY estimations from the various catch series options, considered for the IOTC blue shark stock assessment.

Model	MSY estimation		
	Mean	Low CI	Upp CI
D1 - GAM est.	23583	19227	29651
D1 - Ratio est.	29605	24915	37848
D1 - Reported	23794	19506	29894
D3 - GAM est.	17411	12941	21463
D3 - Ratio est.	15347	13136	17897
D3 - Reported	15885	8801	21158

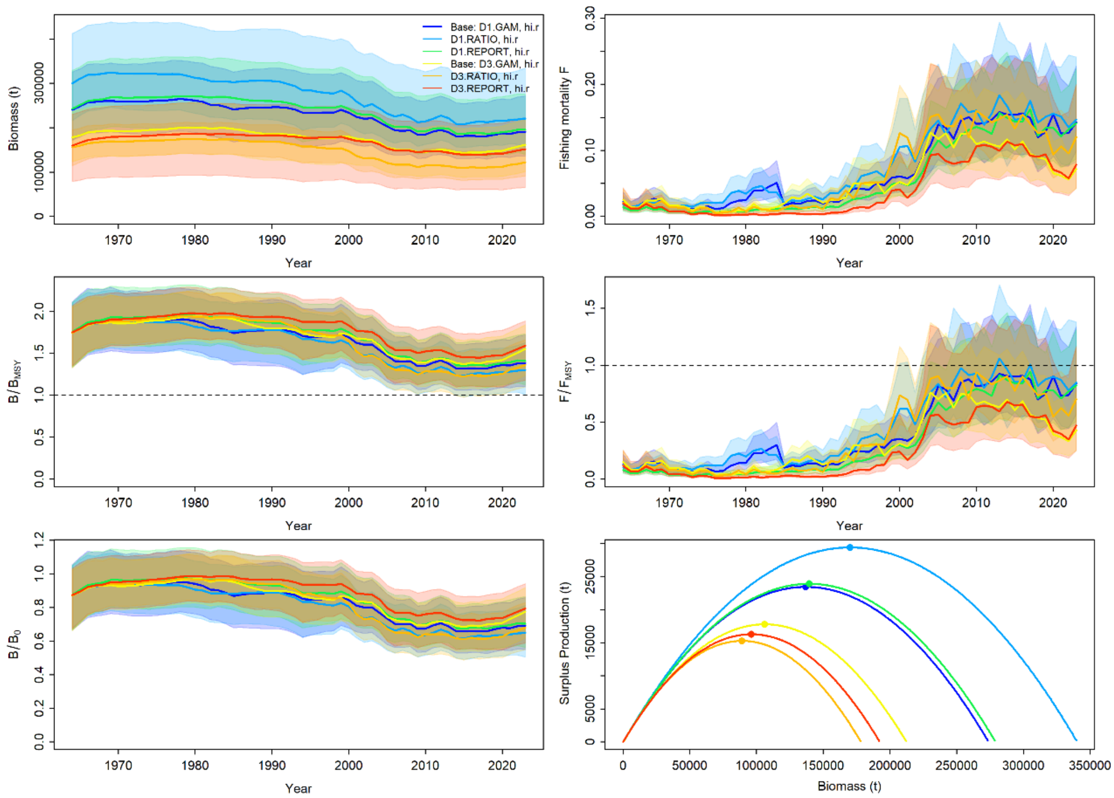

Figure 15: Sensitivity analysis using various catch series options for the IOTC BSH assessment.

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 models 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 16**.

Overall, the sensitivities in terms of the process error show no major differences in the overall trends compared to the base case model. In general, it is 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, by turning off the additional CPUE variance the results are substantially different, as the trajectories are forced to follow much more closely all the jumps that are present in the CPUE indices, providing quite unreasonable results. As such and in general, it is also preferable to allow this additional variance to be estimated within the models.

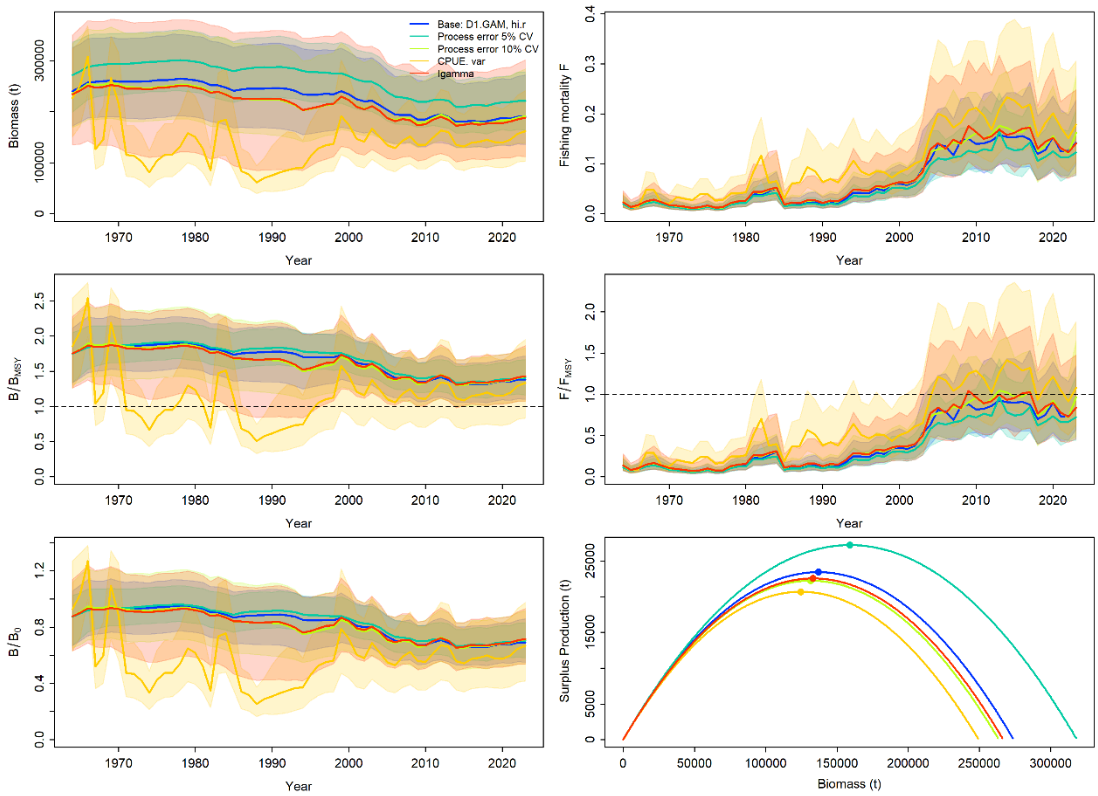

Figure 16: Sensitivity analysis relative to various options for the sigma of the process error ("igamma" = using an igamma vaguely informative prior; "process error 5%" = fixing at 5%; "process error 10%" = fixing at 10%) and turning off the estimation of additional CPUE variance in the models (CPUE.var), performed for IOTC blue shark stock assessment.

3.4. Stock status of base models

The parameters estimated in the final base models (2 options for productivity and 2 options for catches) 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 17**, and the Kobe Phase plots for those 2 scenarios are represented in **Figure 18**. The overall trajectories of all scenarios are very similar. On all cases the biomass values are above Bmsy, while the F is below Fmsy for 3 scenarios and slightly above in one model.

Table 2: Estimates (average, 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 IOTC blue shark stock assessment.

Parameters	Higher *r*			Lower *r*			
	mu	lci	uci	mu	lci	uci	
K	273953	231630	338392	274387	191968	319483	D1 GAM estimated catches
r	0.342	0.309	0.377	0.300	0.272	0.332	
psi	0.875	0.683	0.976	0.875	0.684	0.973	
sigma.proc	0.066	0.039	0.112	0.086	0.044	0.162	
m	2.000	2.000	2.000	2.000	2.000	2.000	
Hmsy	0.171	0.155	0.189	0.150	0.136	0.166	
SBmsy	136977	115815	169196	137193	95984	159742	
MSY	23583	19227	29651	20202	14105	24975	
bmsyk	0.500	0.500	0.500	0.500	0.500	0.500	
P1971	0.874	0.667	1.056	0.877	0.652	1.106	
P2023	0.692	0.551	0.859	0.686	0.530	0.896	
B_Bmsy.cur	1.384	1.102	1.717	1.372	1.059	1.791	
H_Hmsy.cur	0.840	0.570	1.339	1.046	0.617	1.801	
Parameters	Higher *r*			Lower *r*			
	mu	lci	uci	mu	lci	uci	
K	212459	152407	256027	209324	172421	272218	D3 GAM estimated catches
r	0.334	0.304	0.368	0.296	0.269	0.327	
psi	0.877	0.677	0.976	0.877	0.675	0.976	
sigma.proc	0.053	0.035	0.084	0.045	0.032	0.075	
m	2.000	2.000	2.000	2.000	2.000	2.000	
Hmsy	0.167	0.152	0.184	0.148	0.134	0.163	
SBmsy	106229	76203	128013	104662	86211	136109	
MSY	17411	12941	21463	15596	12727	20719	
bmsyk	0.500	0.500	0.500	0.500	0.500	0.500	
P1971	0.876	0.669	1.029	0.875	0.667	1.019	
P2023	0.775	0.619	0.918	0.751	0.606	0.902	
B_Bmsy.cur	1.549	1.238	1.836	1.503	1.211	1.804	
H_Hmsy.cur	0.430	0.251	0.774	0.495	0.269	0.869	

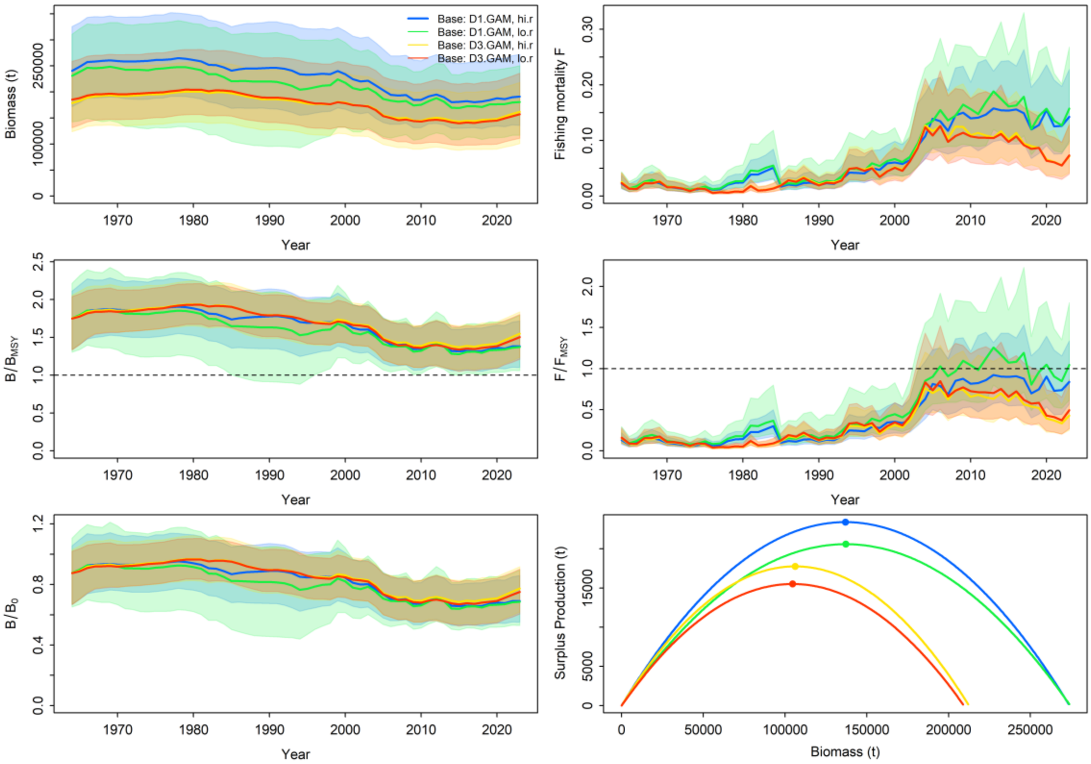

Figure 17: Comparative trends and trajectories of the 4 main base models, run for the 2025 IOTC blue shark stock assessment.

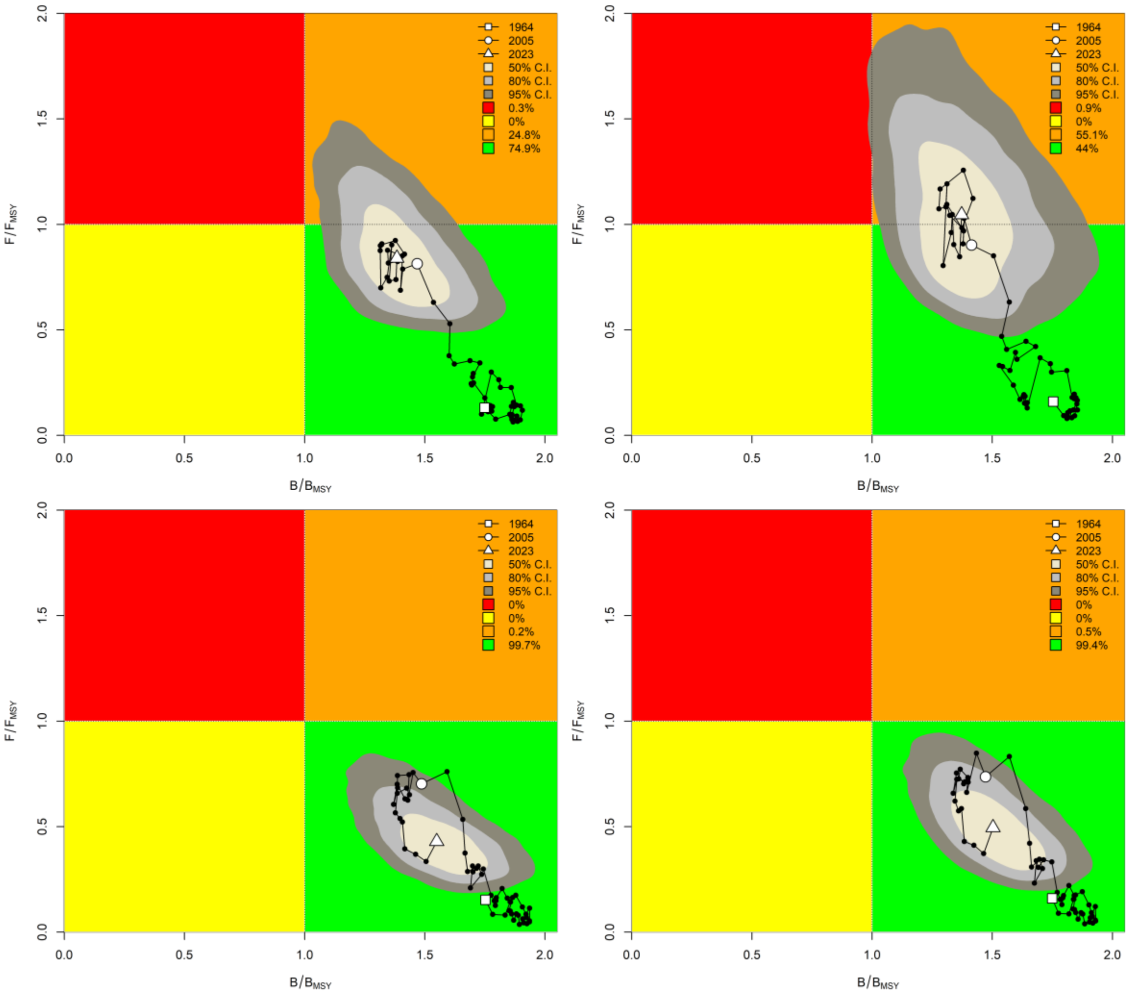

Figure 18: Kobe phase plot with the estimated trajectories (1964-2023) of B/Bmsy and F/Fmsy for the 4 main base grid models considered for the 2025 IOTC blue shark stock assessment. The different gray shaded areas denote the 50%, 80%, and 95% credibility intervals for the terminal year of the assessment 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. The plots represent the following scenarios: Top left: higher r with D1 catches; Top right: lower *r* with D1 catches; Bottom left: higher r with D3 catches; Bottom right: lower *r* with D3 catches.

3.5. Model ensemble

The large model grid ensemble was run with 600 models, comprising 300 models with each of the options catch series (GAM estimated D1 and D3). The range and distribution of the input prior values generated for this large grid ensemble of models is shown in **Figure 19**.

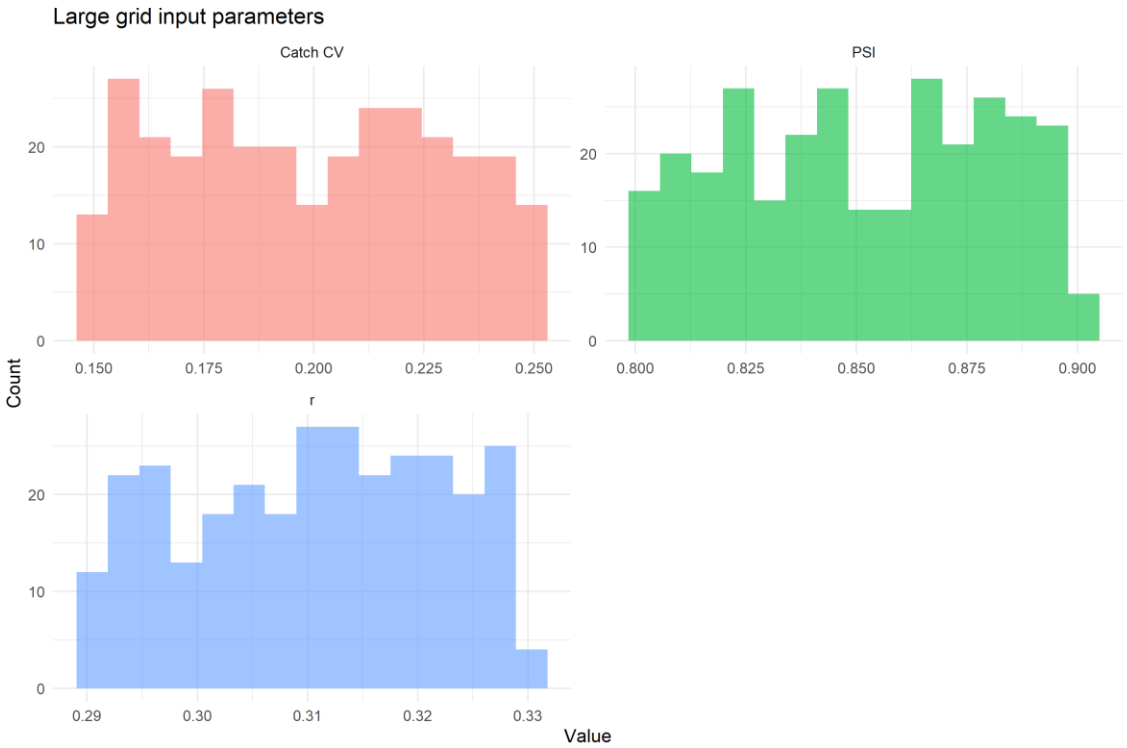

Figure 19: Distribution of the priors input values that were randomly generated for the large model grid ensemble, for the IOTC blue shark stock assessment.

Of the 600 ensemble models run, 32 failed to converge, namely 19 using the D1 catch data and 13 using the D3 catches, while the other 568 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 20**, while the Kobe plots with the terminal year status of all individual models relative to B/Bmsy and F/Fmsy (separated by the D1 and D3 catch ensembles)) are shown in **Figure 21**.

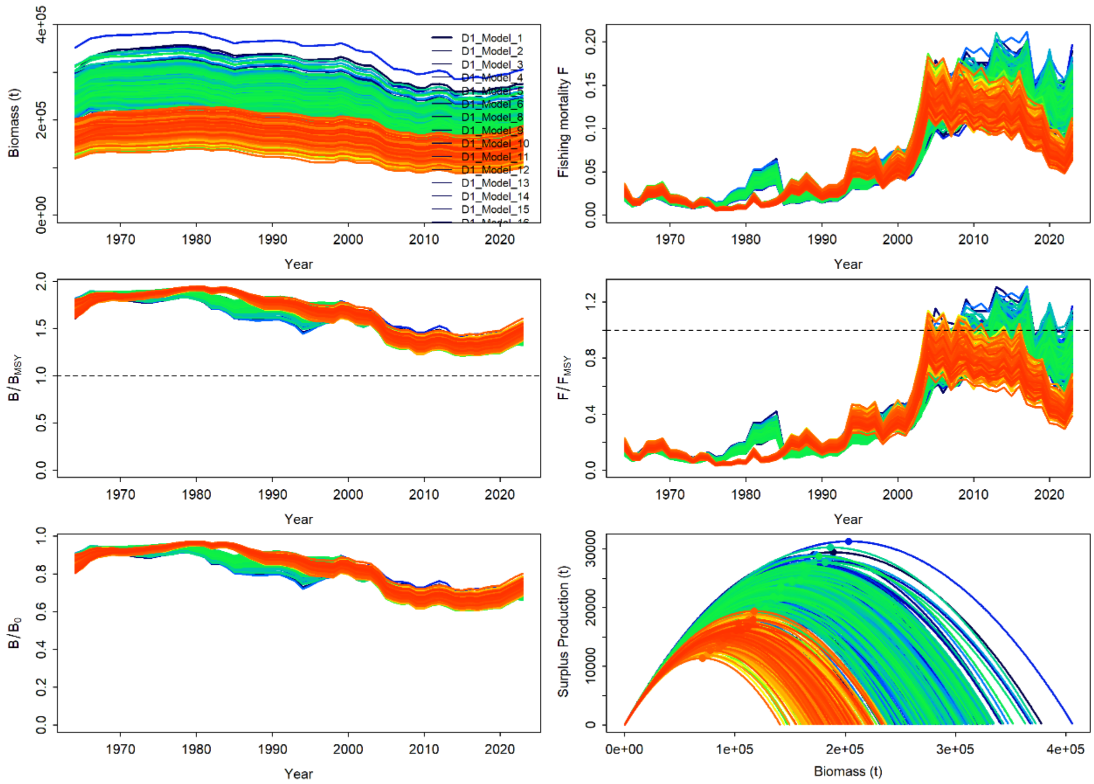

Figure 20. Results of the trajectories and main results from the large ensemble grid models, using both D1 and D3 GAM estimated catches. Only models that have converged are plotted (568 models).

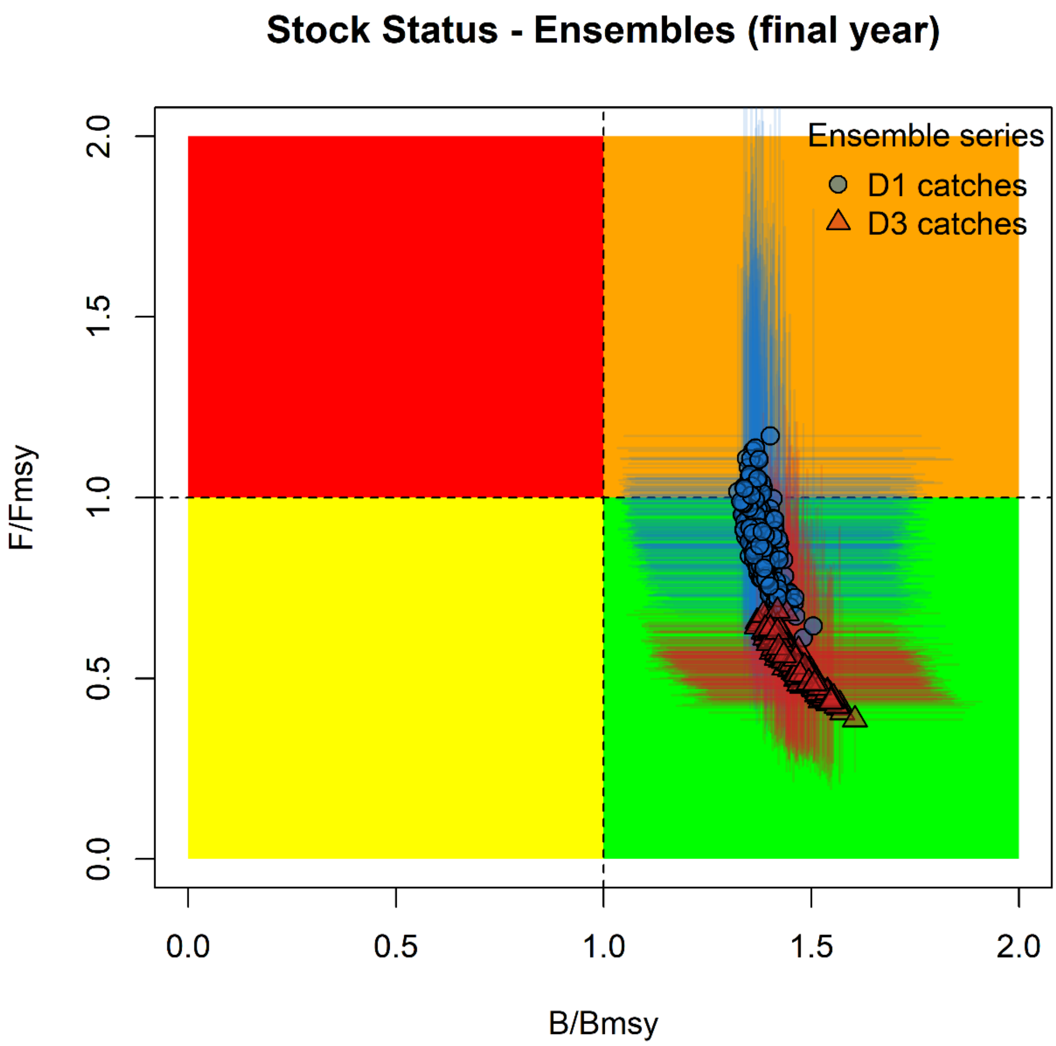

Figure 21. Kobe plot with the point estimates and CIs of the large ensemble grid models that have converged and are used for the analysis. Each ensemble model (based on D1 or D3 catches) is separated in the plot.

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. Additionally, the values are separated for the ensemble grids using D1 and D3 catches.

The summarized Kobe status plot with the 2 model weighing options, and for each pf the option on catches (D1 and D3) is given in **Figure 22**. The stock status is relatively similar when using either model weighting option, even though when using the DIC criteria for model weighing there is a slight tendency for a slightly better stock status.

Table 3: Summaries of the estimates (mean, lower and upper confidence intervals) of the main parameters across the large grid of ensemble models, using the two options catch series (D1 and D3). 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.

Catch series	Parameter	Equal weighted			DIC weighted		
		Mean	LCI	UCI	Mean	LCI	UCI
D1 GAM est	K	289200	226506	359953	309281	240387	388061
	r	0.322	0.291	0.355	0.322	0.291	0.355
	psi	0.851	0.831	0.871	0.853	0.832	0.872
	sigma.proc	0.068	0.04	0.128	0.062	0.038	0.117
	m	2	2	2	2	2	2
	Hmsy	0.161	0.146	0.178	0.161	0.146	0.178
	SBmsy	144600	113253	179977	154641	120193	194031
	MSY	23221	17721	29566	24798	18894	31770
	bmsyk	0.5	0.5	0.5	0.5	0.5	0.5
	P1964	0.857	0.737	1.006	0.858	0.747	0.993
	P2023	0.691	0.542	0.864	0.701	0.553	0.866
	B_Bmsy.cur	1.382	1.084	1.728	1.403	1.105	1.732
	H_Hmsy.cur	0.882	0.511	1.552	0.813	0.469	1.441
D3 GAM est	K	189885	153863	238686	199728	160364	251291
	r	0.318	0.289	0.351	0.319	0.289	0.352
	psi	0.852	0.832	0.872	0.854	0.833	0.874
	sigma.proc	0.051	0.034	0.088	0.05	0.033	0.083
	m	2	2	2	2	2	2
	Hmsy	0.159	0.144	0.176	0.159	0.145	0.176
	SBmsy	94942	76932	119343	99864	80182	125645
	MSY	15079	12088	19228	15864	12644	20302
	bmsyk	0.5	0.5	0.5	0.5	0.5	0.5
	P1964	0.856	0.766	0.961	0.858	0.771	0.958
	P2023	0.734	0.588	0.884	0.749	0.603	0.895
	B_Bmsy.cur	1.468	1.176	1.768	1.497	1.205	1.791
	H_Hmsy.cur	0.527	0.305	0.917	0.492	0.285	0.855

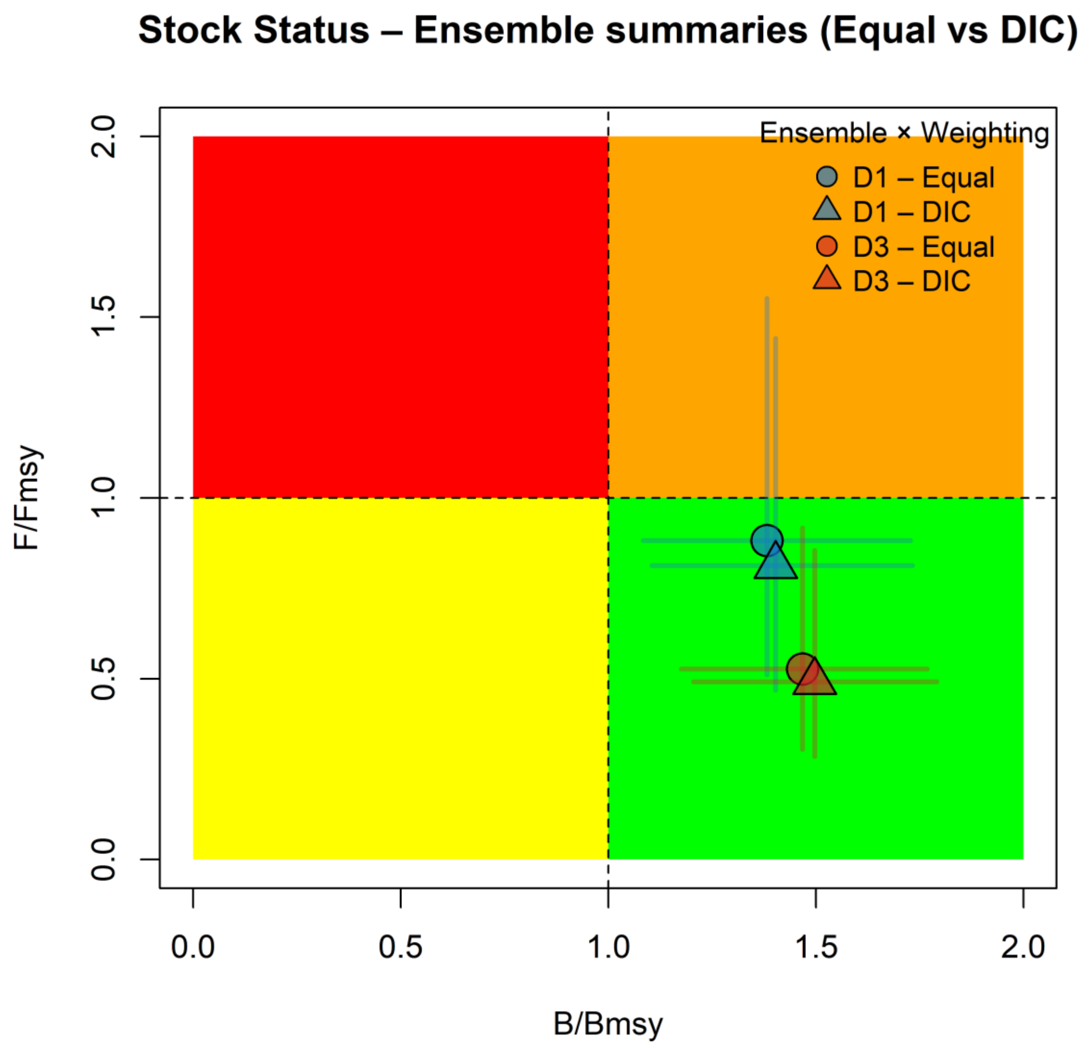

Figure 22. Kobe plot with the point estimates and CIs of the summarized large grid models, using two options for model weighing: 1) equal weights across all models; 2) using DIC-weights across models, and the 2 ensemble grids (based on D1 and D3 catches). The summaries are done with all the grid models that have converged.

3.6. Preliminary projections

The preliminary deterministic projections of the 4 main base models are represented in **Figure 23** and **Figure 24**. Given that the 4 main models use two very distinct catch series, the absolute values are different and therefore not entirely comparable in terms of the absolute scale for the projections. As such, for these preliminary projections there are 2 options, one using models with the D1 catch dataset and the other with the D3 catch dataset.

It is noted that under the D1 catch scenarios, the current catches (i.e., average of the last 3 years) corresponds to 25,848 ton/year, while in the D3 catch scenario the same current catch (i.e., average of the last 3 years) corresponds to 9,920 ton/year.

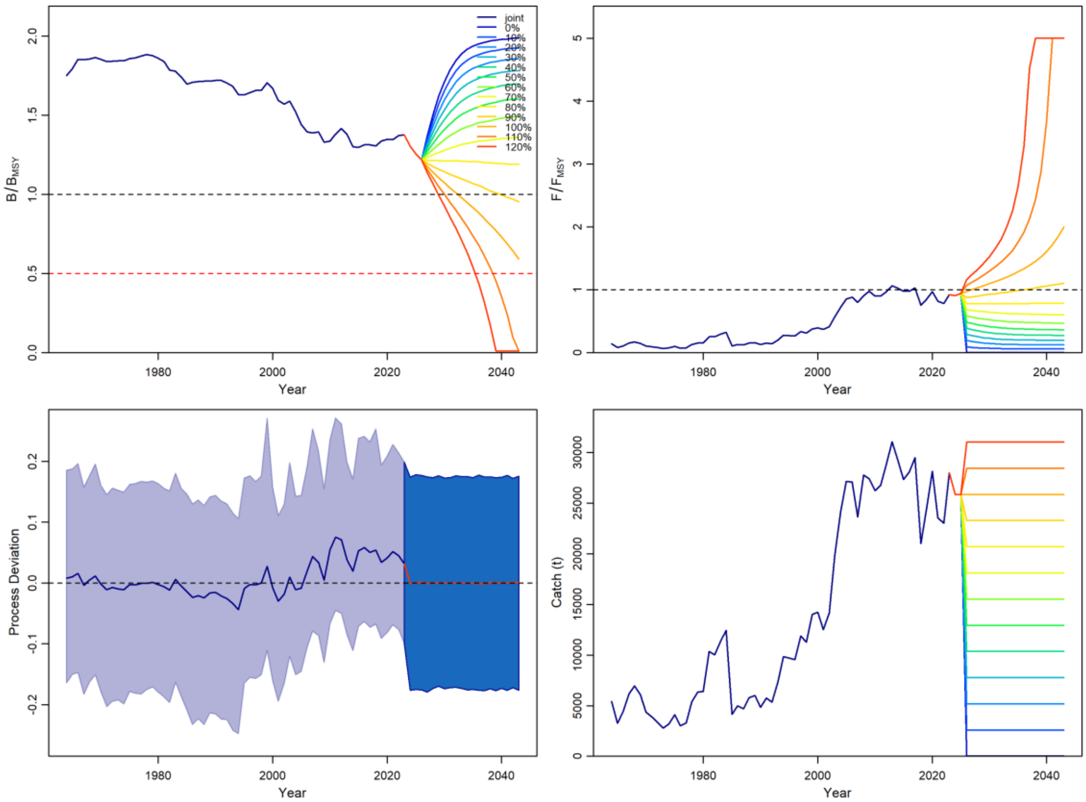

Figure 23. Preliminary deterministic projections for the 2 models using the D1 GAM estimated catch dataset (D1 current catches: average of the last 3 years of catches = 25,848 ton/year).

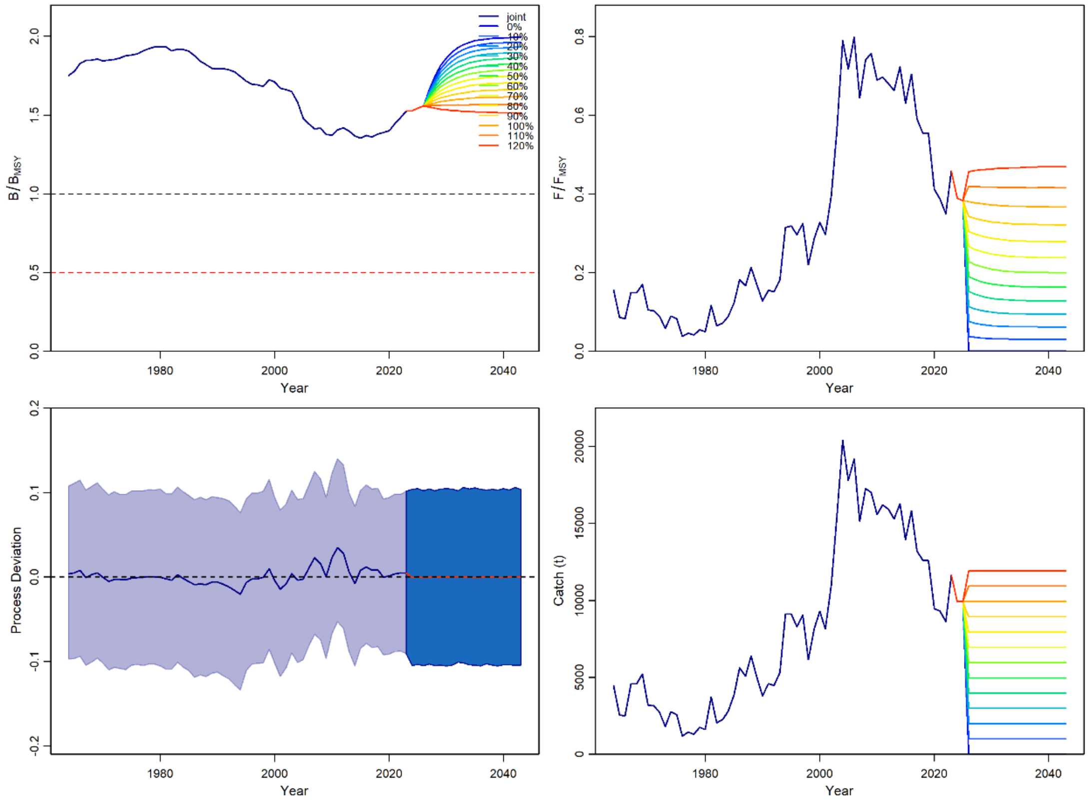

Figure 24. Preliminary deterministic projections for the 2 models using the D3 GAM estimated catch dataset (D3 current catches: average of the last 3 years of catches = 9,920 ton/year).

4. Conclusions

Bayesian Surplus Production Model (JABBA) were applied for determining the stock status, in support of the provision of management advice for the Indian Ocean (IOTC) blue shark.

Four main base models were developed, using two combinations of productivity and two options in terms of estimated catches. The stock status from these 4 main based models ranged not overfished and not undergoing overfishing (B>Bmsy & F<Fmsy), to not overfished but currently undergoing overfishing (B>Bmsy & F>Fmsy).

In addition to the base models, large model grid ensembles (600 models) were run, selecting randomly prior values from distributions built around plausible limits for their values, and using alternatively each of the 2 catch scenarios possibilities. The stock status from this large grid ensemble was weighed in 2 alternative ways, namely using equal model-weighting and DIC model-weighting. On all cases, the stock status resulted in the stock not being overfished and not subject to overfishing (B>Bmsy & F<Fmsy).

In summary, all the models and scenarios tested for the Indian Ocean blue shark show that the current biomass values are likely higher than Bmsy. However, in some scenarios, the current fishing mortality (F) is very close or even higher than Fmsy, suggesting that future catches should be regulated by means of a catch limit and carefully monitored.

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.

6. 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.

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.

Geng, Z., Wang, Y., Kindong, R., Zhu, J., Dai, X., 2021. Demographic and harvest analysis for blue shark (*Prionace glauca*) in the Indian Ocean. *Regional Studies in Marine Science*, 41: 101583. https://doi.org/10.1016/j.rsma.2020.101583.

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.

IOTC, 2024. Report of the 27th session of the IOTC Scientific Committee. South Africa/Hybrid, 2 – 6 December 2024. IOTC–2024–SC27–R[E]: 221 pp.

IOTC, 2025. Report of the 21st session of the IOTC Working Party on Ecosystems and Bycatch – Data Preparatory meeting. Online via Zoom, 28 – 30 April 2025. IOTC–2025–WPEB21(DP)–R[E]: 30pp.

Martell, S., Froese, R., 2013. A simple method for estimating MSY from catch and resilience. *Fish and Fisheries*, 14 (4): 504-514. https://doi.org/10.1111/j.1467-2979.2012.00485.x

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/.

Rosa, D., Coelho, R., 2016. Estimates of intrinsic rate of population change and steepness for blue shark (*Prionace glauca*) in the Indian Ocean. Technical document. Working Party on Ecosystems and Bycatch. IOTC–2016–WPEB12–18.

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.

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

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.