SCRS/2021/132 Collect. Vol. Sci. Pap. ICCAT, 78(2): 499-522 (2021)

JUST ANOTHER ATLANTIC BIGEYE TUNA STOCK ASSESSMENT: PRELIMINARY RESULTS USING A BAYESIAN STATE-SPACE SURPLUS PRODUCTION MODEL (JABBA)

Sant'Ana, R.[1], Mourato, B.[2], Kimoto, A.[3], Ortiz, M.[3] and Winker, H.[4]

SUMMARY

Bayesian State-Space Surplus Production Models were fitted to Atlantic bigeye tuna catch and CPUE data using the 'JABBA' R package. The ten scenarios were based on the previous assessment and on the uncertainty grid proposed during the 2021 BET Data Preparatory Meeting, which in summary corresponded to a continuity run based on a Fox production function and a Pella-Tomlinson production function from an Age-Structured Equilibrium Model (ASEM). All scenarios showed similar trend for the trajectories of B/B_{MSY} and F/F_{MSY} over time, with a stepwise decreasing trend marked by a slower decrease among two sharply decrease patterns. Kobe stock status plots show a typical anti-clockwise pattern with the median quantities estimated for the last data year in the green quadrant. However, the continuity run based on the Fox model was notably more pessimistic than all alternative ASEM scenarios and associated with a relatively higher cumulative probability of 41.3% (red and yellow quadrants) that current biomass levels fall below B_{MSY}.

RÉSUMÉ

Des modèles bayésiens de production excédentaire d'état-espace ont été ajustés aux données de capture et de CPUE du thon obèse de l'Atlantique à l'aide du paquet R "JABBA". Les dix scénarios étaient basés sur l'évaluation précédente et sur la grille d'incertitude proposée lors de la réunion de préparation des données sur le thon obèse de 2021, qui correspondaient en résumé à un scénario de continuité basé sur une fonction de production de Fox et une fonction de production de Pella-Tomlinson d'un modèle d'équilibre structuré par âge (ASEM). Tous les scénarios ont montré une tendance similaire pour les trajectoires de B/B_{PME} et de F/F_{PME} au fil du temps, avec une tendance à la baisse progressive marquée par une diminution plus lente parmi deux schémas de forte diminution. Les diagrammes de l'état du stock de Kobe montrent un schéma typique dans le sens inverse des aiguilles d'une montre, avec la médiane des quantités estimées pour la dernière année de données dans le quadrant vert. Toutefois, le scénario de continuité basé sur le modèle de Fox était nettement plus pessimiste que tous les autres scénarios ASEM et associé à une probabilité cumulée relativement plus élevée de 41,3% (quadrants rouge et jaune) que les niveaux de biomasse actuels tombent en dessous de la B_{PME}.

RESUMEN

Se ajustaron modelos bayesianos de producción excedente estado-espacio a los datos de captura y CPUE del patudo del Atlántico utilizando el paquete R "JABBA". Los diez escenarios se basaron en la evaluación anterior y en la matriz de incertidumbre propuesta durante la reunión de preparación de datos de patudo de 2021, que en resumen correspondían a un ensayo de continuidad basado en una función de producción de Fox y una función de producción de Pella-Tomlinson de un modelo en equilibrio estructurado por edad (ASEM). Todos los escenarios mostraron una tendencia similar para las trayectorias de B/B_{RMS} y F/F_{RMS} a lo largo del tiempo, con una tendencia decreciente gradual marcada por una disminución más lenta entre dos patrones de disminución pronunciada. Los diagramas del estado del stock de Kobe muestran un patrón típico en sentido contrario a las agujas del reloj, con la mediana de las cantidades estimadas para el último año de datos en el cuadrante verde. Sin embargo, el ensayo de

[1] Universidade do Vale do Itajaí, Escola do Mar, Ciência e Tecnologia, Laboratório de Estudos Marinhos Aplicados. Rua Uruguai, 458, Itajaí, SC, Brazil.
[2] Universidade Federal de São Paulo, Instituto do Mar, Rua Carvalho de Mendonça, 144, Encruzilhada - Santos/SP, 11070-100.
[3] ICCAT Secretariat. Calle Corazón de Maria 8, Madrid Spain 28002.
[4] Joint Research Centre (JRC), European Commission, Ispra, Italy.

continuidad basado en el modelo de Fox fue notablemente más pesimista que todos los escenarios alternativos de ASEM y se asoció con una probabilidad acumulada relativamente más alta del 41,3% (cuadrantes rojo y amarillo) de que los niveles actuales de biomasa se sitúen por debajo de B_{RMS}.

KEYWORDS

Bigeye, stock status, CPUE fits, hindcast, life history priors

1. Introduction

The bigeye tuna (*Thunnus obesus*) is one of the largest tuna species. It is widely distributed in the tropical and subtropical waters of the Atlantic, Indian and Pacific Oceans (ICCAT, 2007). The species has habitat preferences for an epipelagic and mesopelagic realm generally inhabiting open waters with optimum temperature range varying between 17 ºC and 22 ºC (ICCAT, 2007). As a function of its wide distribution, bigeye tuna has been intensively exploited by various fisheries around the world (ICCAT, 2007). For management purposes, the International Commission for the Conservation of Atlantic Tunas (ICCAT) considers a single stock for the entire Atlantic based on molecular markers analysis, fish movements and migration patterns (ICCAT, 2007). The Atlantic bigeye tuna stock is mainly exploited by two general fleets, the longliners (45.27% in 2019) and the purse seiners (37.26% in 2019). Together, these two fleets are responsible for more than 83% of the total catches in the past 14 years.

The last Atlantic bigeye tuna stock assessment was carried out in 2018 (ICCAT, 2018) and included outputs from three distinct models, (a) Biomass Dynamic Model using *mpb* approach (Merino et al., 2018; Kell, 2016); (b) Bayesian State-Space Surplus Production Model using JABBA framework (Winker et al., 2018a; Winker et al., 2018b), and; (c) Integrated Age-Structured Model using Stock Synthesis 3 (Walter et al., 2018; Method and Wetzel, 2013). Those models showed consistent results in terms of stock status (ICCAT, 2018). All models indicated that the Atlantic bigeye tuna stock was overfished and experiencing overfishing. Reference points among those models were also similar, with MSY varying between 76,232 and 80,359 metric tons, F_{2017}/F_{MSY} at 1.21 and 1.63 and B_{2017}/B_{MSY} ranging from 0.59 to 0.82 (ICCAT, 2018).

Here, we present the 2021 preliminary stock assessment results for Atlantic bigeye tuna stock based on the Bayesian State-Space Surplus Production Model framework, JABBA (Just Another Bayesian Biomass Assessment; https://github.com/jabbamodel/JABBA; Winker et al., 2018). The JABBA model is a fully documented, open-source R package (https://github.com/JABBAmodel) that has been formally included in the ICCAT stock catalogue (https://github.com/ICCAT/software/wiki/2.8-JABBA) and has been widely applied in a number of recent ICCAT stock assessments, including: South Atlantic blue shark (ICCAT, 2016b), Mediterranean albacore (ICCAT, 2017c), South Atlantic swordfish (ICCAT, 2017a; Winker et al., 2018), Atlantic shortfin mako shark stocks (south and north) (ICCAT, 2017d; Winker et al., 2017, 2019a), Atlantic blue marlin (Mourato et al., 2019), Atlantic bigeye tuna (Winker et al., 2019b), Atlantic white marlin (Mourato et al., 2020), Atlantic yellowfin tuna (Sant'Ana et al., 2020), Mediterranean swordfish (Winker et al. 2020; ICCAT, 2017b) and South Atlantic albacore (Winker et al., 2020b).

This preliminary assessment of the Atlantic bigeye tuna stock is guided by the 2020 SCRS work plan (ICCAT, 2020). A grid scenario was built based on the discussions and recommendations that raised during the 2021 Bigeye Data Preparatory Meeting and 2021 Bigeye Stock Assessment Meeting. In this way, extensive model diagnostics, retrospective pattern analysis and model prediction skillness were provided to evaluate the fitted models. In addition, this document explores the sensitivity of the base case scenarios to the inclusion of alternative and additional standardized CPUE indices that have been made available for this assessment.

2. Material and Methods

2.1. JABBA inputs

This stock assessment is implemented using the Bayesian state-space surplus production model framework called JABBA (Winker et al., 2018), which is now available as '*R package*' that can be installed from github.com/jabbamodel/JABBA. JABBA's inbuilt options 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, and (4) specifying a Fox, Schaefer or

Pella-Tomlinson production function by setting the inflection point B_{MSY}/K and converting this ratio into a shape parameter *m*, (5) extensive diagnostic procedures and associated plots (e.g. residual run tests) and (6) a routine to conduct retrospective analysis. A full JABBA model description, including formulation and state-space implementation, prior specification options and diagnostic tools is available in Winker et al. (2018).

2.2. Fishery data

The ICCAT Secretariat provided fishery catch data for Atlantic bigeye tuna from 1950 to 2019 (Figure 1). Relative abundance indices were made available, principally, in the form of joint standardized CPUE time series. These indices cover various periods and represent the main longline fleets operating in the Atlantic Ocean (*e.g.* Chinese-Taipei fleet, Japanese fleet and Korean fleet). A summary of the available indices is described below:

- Early Joint Index (1959 – 1978) used in the 2018 assessment;
- Joint Index – provided for the 2021 assessment (1979 – 2019);
- Joint Index – used in the 2018 assessment (1979 – 2017).

The CV's for all indices were scaled to an 0.2 average. In addition, the analysis included a comparison between the old joint index used in the last BET assessment and the new one. The general idea with this comparative analysis was to evaluate the possible effects of the differences in methodologies adopted to standardize both indices.

2.3. Model specifications

The model specifications were based on two structures: (*i*) the first aimed to provide a continuity run with the same priors and model structure defined and used in the 2018 Atlantic bigeye tuna stock assessment (ICCAT, 2018), and; (*ii*) the second was based in the discussions and recommendations agreed during the 2021 BET Data Preparatory Meeting and includes a uncertainty grid for maximum age, natural mortality and steepness (**Table 1**). Consistent with the 2018 Bigeye assessment, , the Fox surplus production function was assumed for the continuity scenario.. The priors of *K* and *r* were kept uninformative to convey minimal prior information on the parameters estimates. For *K* and *r*, a lognormal distribution was implemented using JABBA "range" option. For *K*, lower and upper values ranged from 500,000 t to 5,000,000 t, which resulted in an approximated mean value of 1,581,138 t and a CV of 172%. For *r*, the range was set from 0.05 to 0.5, which resulted in an approximated mean of $r = 0.12$ and an associated CV of 166% (**Table 1**). For the scenarios based on uncertainty grid proposed during the 2021 BET Data Preparatory Meeting, we developed an alternative *r* prior distribution with an associated shape parameter of a Pella-Tomlinson production function from an Age-Structured Equilibrium Model (ASEM) approach with Monte-Carlo simulations (Winker et al., 2019b). The stock parameters used as inputs for the ASEM models included the following configuration: (a) Maximum age equal to 17, 20 and 25 years with the corresponding natural mortality values, and; (b) steepness values equal to 0.7, 0.8 and 0.9. This approach resulted in more informative priors to *r* following a lognormal distribution (**Table 1**; **Figure 2**) and the shape parameter *m* directly derived from the ASEM output of EB_{MSY}/EB_0 (**Table 1**; see details in Winker et al., 2019). **Table 1** provide a summary of the five scenarios initially tested.

For all scenarios, the same initial depletion prior ($\varphi = B_{1950}/K$) was defined by a beta distribution with mean = 0.93 and CV of 5%. All catchability parameters were formulated as uninformative uniform priors. Even as, the process error of $\log(B_y)$ in year *y* for all scenarios were defined by an inverse-gamma distribution with shape parameter equal to 9.606 and rate parameter equal to 0.03 as used by Winker et al. (2018b) in the 2018 Atlantic bigeye stock assessment.

JABBA is implemented in R (R Development Core Team, https://www.r-project.org/) with JAGS interface (Plummer, 2003) to estimate the Bayesian posterior distributions of all quantities of interest by means of a Markov Chains Monte Carlo (MCMC) simulation. The JAGS model is executed from R using the wrapper function jags() from the library r2jags (Su and Yajima, 2012), which depends on rjags R package. In this study, three MCMC chains were used. Each model was run for 30,000 iterations, sampled with a burn-in period of 5,000 for each chain and thinning rate of five iterations. Basic diagnostics of model convergence included visualization of the MCMC chains using MCMC trace-plots as well as Heidelberger and Welch (1992), Geweke (1992), and Gelman and Rubin (1992) diagnostics as implemented in the coda package (Plummer et al., 2006).

2.4. Model diagnostics and sensitivity runs

To evaluate CPUE fits, the model predicted CPUE indices were compared to the observed CPUE. JABBA-residual plots were used to examine (1) colour-coded lognormal residuals of observed versus predicted CPUE indices for all fleet together with (2) boxplots indicating the median and quantiles of all residuals available for any given year; the area of each box indicates the strength of the discrepancy between CPUE series (larger box means higher degree of conflicting information), and (3) a loess smoother through all residuals aids to detect the presence systematic residual patterns. In addition, it depicts the root-mean-squared-error (RMSE) as a goodness-of-fit statistic. We conducted a runs test 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 using the function runs.test in the R package tseries, considering the 2-sided *p*-value of the Wald-Wolfowitz runs test. The runs test results can be visualized within JABBA using a specifically designed plot function that illustrates which time series passed or failed the runs test and highlights individual data points that fall outside the three-sigma limits (e.g. Anhøj and Olesen, 2014).

To check for systematic bias in the stock status estimates, we also performed a retrospective analysis for the two production function scenarios (Base Case: Fox 2021 and ASEM M=17 | h=0.7: Pella m), by sequentially removing one year of data at a time over a period of eight years (n = 8), refitting the model after each data removal and comparing quantities of interest (*i.e.* biomass, fishing mortality, B/B_{MSY}, F/F_{MSY}, B/B_0 and MSY) to the reference model that is fitted to full data time series. To compare retrospective bias between the models, we computed Mohn's (1999) rho (ρ) statistic, specifically the commonly used formulation defined by Hurtado-Ferro et al. (2014).

Although the above model diagnostics are important to evaluate the goodness of fit to the data and the consistency of benchmarking retrospectively, providing scientific advice should also involve checking that the model has prediction skill of future states under alternative management scenarios. To do this, the model-free hindcasting cross-validation (HCXval) technique by Kell et al. (2016) was applied, where observations are compared to their predicted future values. The HCXval algorithm has in common with retrospective analysis that requires the same two routine procedures of sequential removal the observations and re-fitting the model to the so truncated data series, but HCXval involves the additional steps of projecting ahead over the missing years and then cross-validating these forecasts against observations to assess the model's prediction skill. A robust statistic for evaluating prediction skill is the Mean Absolute Scaled Error (MASE) proposed by Hyndman and Koehler (2006), which scales the mean absolute error of prediction residuals to a naïve baseline prediction, where a 'prediction' is said to have 'skill' if it improves the model forecast when compared to the naïve baseline. A widely used baseline forecast for time series is the 'persistence algorithm' that takes the value at the previous time step to predict the expected outcome at the next time step as a naïve in-sample prediction, e.g., tomorrow's weather will be the same as today's. The MASE score scales the mean absolute error of the prediction residuals to the mean absolute error of a naïve in-sample prediction. A MASE score higher than one can then be interpreted such that the average model forecasts are no better than a random walk. Conversely, a MASE score of 0.5 indicates that the model forecasts twice as accurately as a naïve baseline prediction; thus, the model has prediction skill.

Finally, to examine the sensitivity of the updated stock assessment results to the inclusion of the alternative Joint index used in the last assessment, a direct comparison of the trajectories management quantities estimated by the addiction of the old Joint index aside from the indices used in the continuity model structured for this assessment was conducted (Table 2). This sensitivity run was derived from the discussions that raised during the 2021 Bigeye Stock Assessment Meeting as a form to evaluate the impact of the changes adopted in the assumptions of the methodology used to build the standardization of the joint CPUE index in 2021. It is important to refer here, that these changes were made necessary as an adequation to the cooperative work between CPC's during the pandemic of COVID-19.

3. Results and Discussion

The MCMC convergence tests by Heidelberger and Welch (1992), Geweke (1992), and Gelman and Rubin (1992) were passed by all estimable key parameters for all models. Adequate convergence of the MCMC chains was also corroborated by visual inspection of trace plots (results available on request), which showed good mixing in general (*i.e.*, moving around the parameter space).

The model fits to each of the two standardized joint CPUE longline indices (Early Joint Index 1959 – 1978 and Joint Index 1979 – 2019) are shown in **Figure 3** for each of the ten scenarios. All scenarios appeared to fit reasonably well to the both joint indices (*i.e.* early and late joint indices), with few large deviations for some particular years. The results of the log-residuals runs tests for each CPUE and each scenario are shown in **Figure 4**. Green panels indicate CPUE indices that passed the runs test with no evidence of a non-random residual pattern ($p > 0.05$) and red panels indicating a failed runs test. In addition, the inner shaded area shows 3-sigma limits around the overall mean as proposed by Anhøj and Olesen (2014) and the red circles identify each specific year where the residuals are larger than the threshold limit. In the scenarios S01, S02, S04, S05, S07 and S09 was observed a failed behavior in the runs test diagnostic procedure, in general for the early joint CPUE only. The other four scenarios have passed completely in this diagnostic (**Figure 4**). The goodness-of-fit were comparable among all scenarios, in general, the RMSE statistics were consistent ranging from 11.9% to 12.1% (**Figure 5**). The annual process error deviation estimated for all scenarios shown a similar stochastic pattern with a constant average centered around the zero and 95% credibility intervals always covering the zero value (**Figure 6**), which suggest no evidence of structural model misspecifications.

The medians of the marginal posteriors for *K* ranged between 1,259,079 t (S04: ASEM M = 17 | h = 0.9: Pella m) and 1,542,554 t (S01: Continuity run: Fox 2021) (**Table 3**). The relatively small posterior to prior median (PPMR) and (PPVR) variance ratios observed to *K* parameter can be an evidence that the posteriors estimates of *K* were largely informed by the data (**Figure 7**). The estimated median of the marginal posteriors for *r* were very similar between the distinct scenarios, ranged from 0.129 to 0.166 (**Table 3**; **Figure 7**). The initial depletion ($\varphi = B_{1950}/K$) marginal posteriors for each scenario were also similar and largely informed by the priors distributions.

The range of *MSY* median estimates were narrow between all ten scenarios, reaching the lower value in the S01 scenario (79,325 metric tons) and the higher value in the S10 scenario (83,946 metric tons) (**Table 3**). Furthermore, the marginal posterior medians for B_{MSY} varied between 371,518 (S10: ASEM M = 25 | h = 0.9: Pella m) and 567,757 (S01: Continuity run: Fox 2021) metric tons, and estimates of F_{MSY} were also similar among all scenarios with median values varying from 0.140 (S01: Continuity run: Fox 2021) to 0.227 (ASEM M = 25 | h = 0.9: Pella m) (**Table 3**).

In general, all scenarios showed similar trend for the trajectories of B/B_{MSY} and F/F_{MSY} over time (**Figure 8**). The trajectory of B/B_{MSY} shown a stepwise decrease trend marked by a slower decrease among two periods stronger decreases in all scenarios. The first sharp decrease moment can be observed between the years 1965 and 1975, the second one between the years 1990 and 2000, and the soft decrease among both (1975 – 1990). The F/F_{MSY} trajectory show a gradually increasing trend from the beginning of the time series until late 1980s, followed by a sharply increase, crossing the reference level ($F/F_{MSY} = 1$) at the middle 1990s to all scenarios, except for the scenarios at maximum age 25 years (**Figure 8**). For all scenarios, except for the S01, the overfishing period was short and the fishing mortality declined below and/or close to the reference level ($F/F_{MSY} = 1$) after the middle of 2000's (**Figure 8**). In general, Fox model scenario presented a more conservative stock status when compared to the ASEM scenarios (**Figure 8**). This distinction could be explained by the different model assumptions with respect to the productivity of the stock and the associated life history input parameters to the ASEM model scenarios.

The results of an eight year retrospective analysis applied to scenarios S01 and S06 are depicted in **Figures 9** and **10**, respectively. In general, both scenarios show a negligible retrospective pattern. However, for the S01 scenario, a slightly retrospective pattern was notable for F/F_{MSY} for the 2011 and 2012 retrospective models, but within of 95% Credibility Intervals (CRIs) of the reference model (**Figure 9**). The estimated Mohn's rho for all stock quantities fell within the acceptable range of -0.15 and 0.20 (Hurtado-Ferro et al., 2014; Carvalho et al., 2017) and these results confirm the absence of an undesirable retrospective pattern (**Table 4**). The hindcasting cross-validation results for joint CPUE index 1979 – 2019 show predictions within limits of the 95% CRI's suggesting a good prediction skills for both scenarios or model-types (**Figure 11**). However, the mean absolute scaled error (MASE) estimated for both scenarios (S01 = 1.207; S06 = 1.117) were slightly above of the reference level (MASE > 1), which indicates that the average model forecasts are not better than a naïve baseline prediction – like a random walk process (Carvalho et al., 2021). Nonetheless, for the index with a flat trend with low variation at the end of the time series is expected that the MASE estimation will be close to reference level one.

The sensitivity analysis of model fits and log-residuals runs tests for the S01: Continuity run: Fox 2021 scenario and in the face of the inclusion of the joint index used in the last assessment (**Table 2**) are shown in **Figure 12**. The two scenarios show a distinct difference in the stock status trajectories as a result of the adding the joint index used in 2018 stock assessment (**Figure 12**). In general, and considering only the central tendencies, the addition of the 2018 joint index show more pessimistic stock trajectories for all management quantities estimated (*e.g.* Biomass, B/B_{MSY}, F/F_{MSY}) (**Figure 12**).

The surplus production phase plots show a similar behavior for the updated scenarios to 2019 with the last year observed in a green area (**Figure 13**). The Kobe biplots for all scenarios were shown in the **Figure 14**. The continuity run (S01) shows a typical anti-clockwise pattern with the median quantity estimated for the last year observed in a green quadrant. However, the continuity run based on the Fox model was notably more pessimistic than all alternative ASEM scenarios with a relatively higher cumulative probability of 41.3% (red and yellow quandrants) that current biomass levels fall below B_{MSY}. On the other hand, the ASEM scenarios show more optimistic status with the cumulative probabilities of green and orange quadrants always above of 80% (**Figure 14**). In the case of F/F_{MSY}, all updated scenarios to 2019 shown a cumulative probabilities of red and orange regions always below of 37% (**Figure 14**). This pattern suggests that the fishing mortality are below the limit of the F_{MSY}.

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

Babcock, E. 2012. Application of a Bayesian surplus production model to preliminary data for south Atlantic and Mediterranean albacore. Collect. Vol. Sci. Pap. ICCAT, 68(2): 519-528.

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? Fish. Res. 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., Rubin, D.B., 1992. Inference from Iterative Simulation Using Multiple Sequences. Stat. Sci. 7, 457–472. https://doi.org/10.2307/2246093

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 Fourth 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. Oper. Res. 31, 1109–1144. https://doi.org/10.1287/opre.31.6.1109

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, in: ICES Journal of Marine Science. pp. 99–110. https://doi.org/10.1093/icesjms/fsu198

Hyndman and Koehler, 2006. Another look at measures of forecast accuracy. International Journal of Forecasting 22(4), 679-688

ICCAT, 2007. ICCAT Manual – Species: Chapter 2.2.2 – Bigeye tuna. Available on-line at https://iccat.int/Documents/SCRS/Manual/CH2/2_1_2_BET_ENG.pdf.

ICCAT, 2016a. Report of the 2016 ICCAT north and south Atlantic albacore stock assessment meeting. Collect. Vol. Sci. Pap. ICCAT, 100p.

ICCAT, 2016b. Report of the 2015 ICCAT blue shark stock assessment session. Collect. Vol. Sci. Pap. ICCAT 72, 866–1019.

ICCAT, 2017a. Report of the 2017 ICCAT Atlantic swordfish stock assessment session. Collect. Vol. Sci. Pap. ICCAT 74, 841–967.

ICCAT, 2017b. Report of the 2016 Mediterranean swordfish stock assessment meeting. Col. Vol. Sci. Pap. ICCAT 73, 1005–1096.

ICCAT, 2017c. Report of the 2017 ICCAT albacore species group intersessional meeting (including assessment of Mediterranean albacore). Collect. Vol. Sci. Pap. ICCAT 74, 45.

ICCAT, 2017d. Report of the 2017 ICCAT shortfin mako assessment meeting. Collect. Vol. Sci. Pap. ICCAT 74, 1465–1561.

ICCAT, 2018. Report of the 2018 ICCAT bigeye tuna stock assessment meeting. Pasaia, Spain. 92 p.

Kell, L. T., Kimoto, A. and Kitakado., T. 2016. Evaluation of the prediction skill of stock assessment using hindcasting. Fisheries Research, 183:119–127

Mourato, B.L., Winker, H., Carvalho, F., Ortiz, M., 2019. Stock Assessment of blue marlin (Makaira nigricans) using a Bayesian State-Space Surplus Production Model JABBA. Collect. Vol. Sci. Pap. ICCAT 75, 1003–1025.

Mourato, B.L., Winker, H., Carvalho, F., Kimoto, A., Ortiz, M., 2020. Developing of Bayesian State-Space Surplus Production JABBA for Assessing Atlantic white marlin (Kajikia albida) stock. Col. Vol. Sci. Pap. ICCAT 76, 235–254.

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 Best, Cowles, K., Vines, K., 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News 6, 7–11.

Sant'Ana, R., Mourato, B., Kimoto, A., Walter, J., Winker, H., 2020. Atlantic Yellowfin tuna stock assessment: An Implementation of a Bayesian State-Space Surplus Production Model using JABBA. Col. Vol. Sci. Pap. ICCAT 76, 699–724.

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

Winker, H., Carvalho, F., Sharma, R., Parker, D., Kerwath, S., 2017. Initial results for North and South Atlantic shortfin mako (Isurus oxyrinchus) stock assessments using the Bayesian surplus production model JABBA and the catch-resiliance method CMSY 74, 1836–1866.

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

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

Winker, H., Carvalho, F., Kerwath, S., 2019a. Age-structured biomass dynamics of north Atlantic shortfin mako with implications for the interpretation of surplus production models. ICCAT-SCRS 098, 1–19.

Winker, H., Mourato, B., Chang, Y. 2019b. Unifying parametrizations between age-structured and surplus production models: an application to Atlantic white marlin (*Kajika albida*) with simulation testing. Col. Vol. Sci. Pap. ICCAT SCRC/2019/103.

Winker, H., Kerwath, S.E., Merino, G., Ortiz, M., 2019b. Bayesian State-Space Surplus Production Model JABBA of Atlantic bigeye tuna (*Thunnus obesus*) stock. Col. Vol. Sci. Pap. ICCAT 75, 2129–2168.

Winker, H., Kimoto, A., Mourato, B., Tserpes, G., and Ortiz, M. 2020b. Development of Bayesian state-space surplus production model JABBA for assessing the Mediterranean swordfish (*Xiphias gladius*) stock. SCRS/2020/082.

Table 1. Summary of the uncertainty grid scenarios for Atlantic bigeye tuna.

Scenario	Model	r	B_{MSY}/K (m)
S01	Continuity run: Fox 2021	Range (0.05, 0.5)	
S02	ASEM M=17 | h=0.7: Pella m	Lognormal (0.230, 0.349)	0.330
S03	ASEM M=17 | h=0.8: Pella m	Lognormal (0.251, 0.389)	0.330
S04	ASEM M=17 | h=0.9: Pella m	Lognormal (0.272, 0.413)	0.340
S05	ASEM M=20 | h=0.7: Pella m	Lognormal (0.186, 0.288)	0.320
S06	ASEM M=20 | h=0.8: Pella m	Lognormal (0.198, 0.331)	0.310
S07	ASEM M=20 | h=0.8: Pella m	Lognormal (0.211, 0.369)	0.300
S08	ASEM M=25 | h=0.7: Pella m	Lognormal (0.149, 0.239)	0.310
S09	ASEM M=25 | h=0.8: Pella m	Lognormal (0.155, 0.260)	0.290
S10	ASEM M=25 | h=0.9: Pella m	Lognormal (0.161, 0.293)	0.270

Table 2. Summary of sensitivity analysis runs for Atlantic bigeye tuna (*Thunnus obesus*).

Scenario	Model	Type	Indices
S01	Fox	Continuity run	• Joint LL index (1959-1978); • Joint LL index (1979-2019).
S01 + Joint Index 2018	Fox	Continuity run	• Joint LL index (1959-1978); • Joint LL index (1979-2019); • Joint LL index – Last assessment.

Table 3. Summary of posterior quantiles presented in the form of marginal posterior medians and associated the 95% credibility intervals of parameters for the Bayesian state-space surplus production models for Atlantic bigeye tuna.

S01: Continuity run: Fox 2021				S02: ASEM M=17 | h=0.7: Pella m			
Estimates	**Median**	**LCI (2.50%)**	**UCI (97.50%)**	**Estimates**	**Median**	**LCI (2.50%)**	**UCI (97.50%)**
K	1,542,554	957,894	2,902,892	K	1,299,585	914,940	1,967,796
r	0.140	0.066	0.238	r	0.157	0.100	0.232
ψ(psi)	0.963	0.818	0.999	ψ(psi)	0.963	0.821	0.999
σ_{proc}	0.053	0.041	0.072	σ_{proc}	0.054	0.041	0.073
F_{MSY}	0.140	0.066	0.237	F_{MSY}	0.193	0.123	0.286
B_{MSY}	567,757	352,565	1,068,448	B_{MSY}	428,976	302,010	649,544
MSY	79,325	63,209	95,205	MSY	82,687	71,483	97,663
B_{1950}/K	0.368	0.368	0.368	B_{1950}/K	0.330	0.330	0.330
B_{2018}/K	0.953	0.794	1.076	B_{2019}/K	0.953	0.796	1.081
B_{2018}/B_{MSY}	0.383	0.277	0.524	B_{2019}/B_{MSY}	0.386	0.286	0.518
F_{2018}/F_{MSY}	1.039	0.754	1.424	F_{2019}/F_{MSY}	1.168	0.866	1.570
S03: ASEM M=17 | h=0.8: Pella m				**S04: ASEM M=17 | h=0.9: Pella m**			
Estimates	**Median**	**LCI (2.50%)**	**UCI (97.50%)**	**Estimates**	**Median**	**LCI (2.50%)**	**UCI (97.50%)**
K	1,295,431	915,770	1,976,656	K	1,259,079	880,690	1,931,649
r	0.157	0.100	0.230	r	0.166	0.105	0.246
ψ(psi)	0.963	0.825	0.999	ψ(psi)	0.963	0.821	0.999
σ_{proc}	0.054	0.041	0.074	σ_{proc}	0.054	0.042	0.073
F_{MSY}	0.194	0.124	0.283	F_{MSY}	0.194	0.123	0.287
B_{MSY}	427,605	302,284	652,469	B_{MSY}	428,207	299,519	656,945
MSY	82,652	71,518	99,916	MSY	82,828	71,417	98,781
B_{1950}/K	0.330	0.330	0.330	B_{1950}/K	0.340	0.340	0.340
B_{2019}/K	0.953	0.802	1.082	B_{2019}/K	0.952	0.793	1.082
B_{2019}/B_{MSY}	0.385	0.291	0.538	B_{2019}/B_{MSY}	0.391	0.286	0.537
F_{2019}/F_{MSY}	1.166	0.881	1.629	F_{2019}/F_{MSY}	1.150	0.842	1.580
S05: ASEM M=20 | h=0.7: Pella m				**S06: ASEM M=20 | h=0.8: Pella m**			
Estimates	**Median**	**LCI (2.50%)**	**UCI (97.50%)**	**Estimates**	**Median**	**LCI (2.50%)**	**UCI (97.50%)**
K	1,284,136	939,889	1,874,208	K	1,313,070	940,614	1,999,410
r	0.153	0.101	0.216	r	0.148	0.095	0.214
ψ(psi)	0.964	0.822	0.999	ψ(psi)	0.963	0.824	0.999
σ_{proc}	0.054	0.041	0.073	σ_{proc}	0.054	0.042	0.073
F_{MSY}	0.200	0.132	0.282	F_{MSY}	0.204	0.131	0.296
B_{MSY}	411,019	300,835	599,886	B_{MSY}	407,157	291,666	619,977
MSY	82,104	71,621	96,483	MSY	82,852	71,853	100,966
B_{1950}/K	0.320	0.320	0.320	B_{1950}/K	0.310	0.310	0.310
B_{2019}/K	0.952	0.799	1.080	B_{2019}/K	0.953	0.796	1.080
B_{2019}/B_{MSY}	0.370	0.273	0.502	B_{2019}/B_{MSY}	0.378	0.279	0.523
F_{2019}/F_{MSY}	1.155	0.853	1.568	F_{2019}/F_{MSY}	1.220	0.898	1.687

Continues on the next page…

Table 3. Continued from the previous page.

S07: ASEM M=20 | h=0.9: Pella m				S08: ASEM M=25 | h=0.7: Pella m			
Estimates	**Median**	**LCI (2.50%)**	**UCI (97.50%)**	**Estimates**	**Median**	**LCI (2.50%)**	**UCI (97.50%)**
K	1,313,276	929,370	2,010,445	K	1,420,759	1,027,115	2,033,436
r	0.144	0.090	0.211	r	0.135	0.093	0.190
$\psi_{(psi)}$	0.962	0.823	0.999	$\psi_{(psi)}$	0.963	0.820	0.999
σ_{proc}	0.054	0.042	0.073	σ_{proc}	0.053	0.041	0.072
F_{MSY}	0.211	0.132	0.310	F_{MSY}	0.187	0.129	0.263
B_{MSY}	393,828	278,701	602,896	B_{MSY}	440,549	318,488	630,528
MSY	83,096	71,644	98,521	MSY	82,056	71,204	99,655
B_{1950}/K	0.300	0.300	0.300	B_{1950}/K	0.310	0.310	0.310
B_{2019}/K	0.951	0.797	1.079	B_{2019}/K	0.953	0.795	1.081
B_{2019}/B_{MSY}	0.376	0.279	0.506	B_{2019}/B_{MSY}	0.378	0.285	0.520
F_{2019}/F_{MSY}	1.254	0.929	1.687	F_{2019}/F_{MSY}	1.220	0.919	1.675
S09: ASEM M=25 | h=0.8: Pella m				**S10: ASEM M=25 | h=0.9: Pella m**			
Estimates	**Median**	**LCI (2.50%)**	**UCI (97.50%)**	**Estimates**	**Median**	**LCI (2.50%)**	**UCI (97.50%)**
K	1,353,754	1,002,481	1,982,140	K	1,375,997	989,492	2,046,832
r	0.135	0.090	0.191	r	0.129	0.084	0.185
$\psi_{(psi)}$	0.962	0.816	0.999	$\psi_{(psi)}$	0.964	0.820	0.999
σ_{proc}	0.054	0.041	0.072	σ_{proc}	0.054	0.041	0.072
F_{MSY}	0.211	0.141	0.297	F_{MSY}	0.227	0.148	0.326
B_{MSY}	392,580	290,713	574,808	B_{MSY}	371,518	267,162	552,642
MSY	82,761	71,837	98,111	MSY	83,946	72,435	101,152
B_{1950}/K	0.290	0.290	0.290	B_{1950}/K	0.270	0.270	0.270
B_{2019}/K	0.951	0.793	1.079	B_{2019}/K	0.954	0.798	1.080
B_{2019}/B_{MSY}	0.370	0.278	0.496	B_{2019}/B_{MSY}	0.373	0.275	0.508
F_{2019}/F_{MSY}	1.275	0.957	1.712	F_{2019}/F_{MSY}	1.381	1.020	1.882

Table 4. Summary Mohn's rho statistic computed for a retrospective evaluation period of eight years for two of the five scenarios fitted to the Atlantic bigeye tuna stock assessment 2021. The scenarios used in retrospective analysis were the S01: Continuity run: Fox 2021 and S06: ASEM M = 20 | h = 0.8: Pella m models. The more the values diverge from the zero, the stronger is the retrospective bias. Values falling between -0.15 and 0.2 are widely deemed as acceptable retrospective bias (Huerto et al., 2014).

Scenario	***Stock Quantity***					
	B	***F***	***B/B_{MSY}***	***F/F_{MSY}***	***B/K***	***MSY***
S01	-0.0363	0.0380	-0.0236	0.0414	-0.0236	-0.0123
S06	0.001	-0.0002	0.0064	-0.0093	0.0064	0.0067

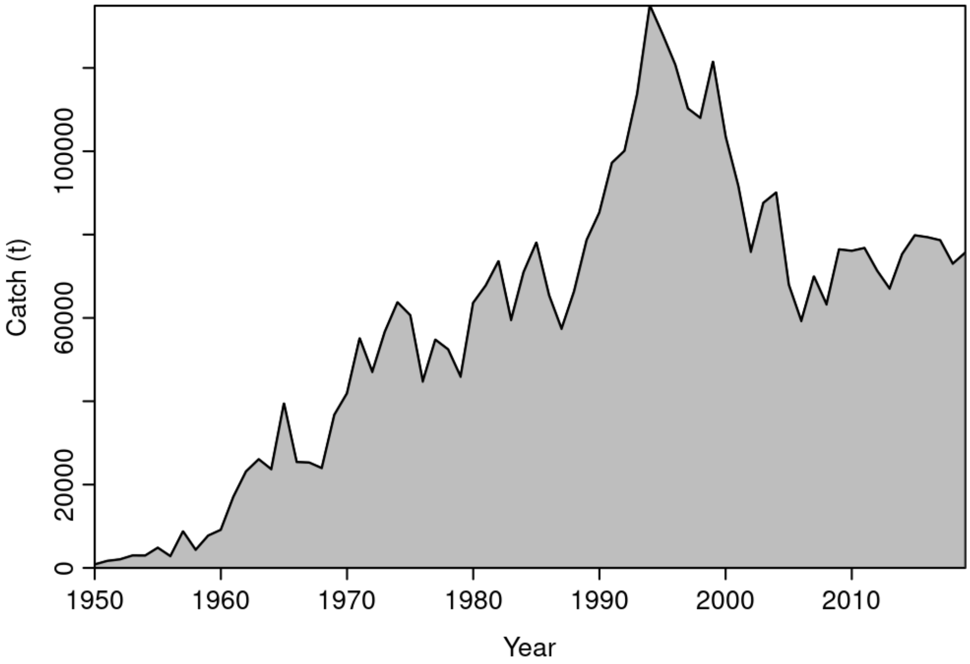

Figure 1. Catch time series in metric tons (t) between 1950 and 2019 for Atlantic bigeye tuna.

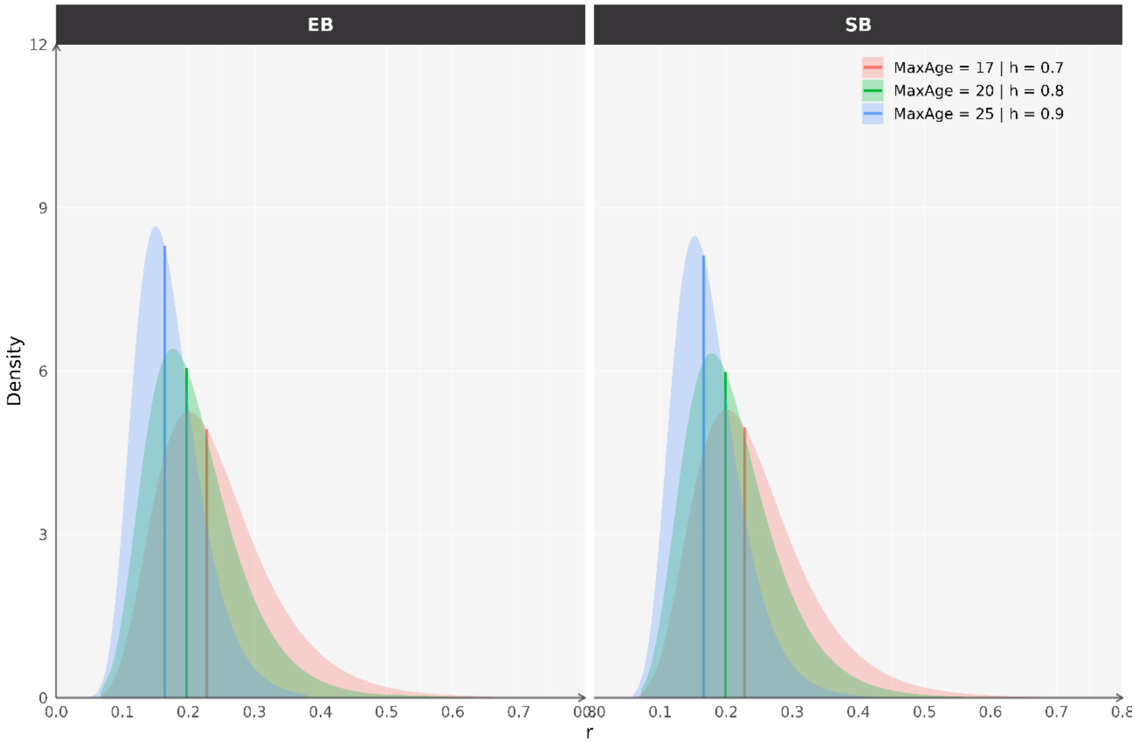

Figure 2. Comparison between *r* priors distributions derived from Age-Structured Equilibrium Models (ASEM) approach.

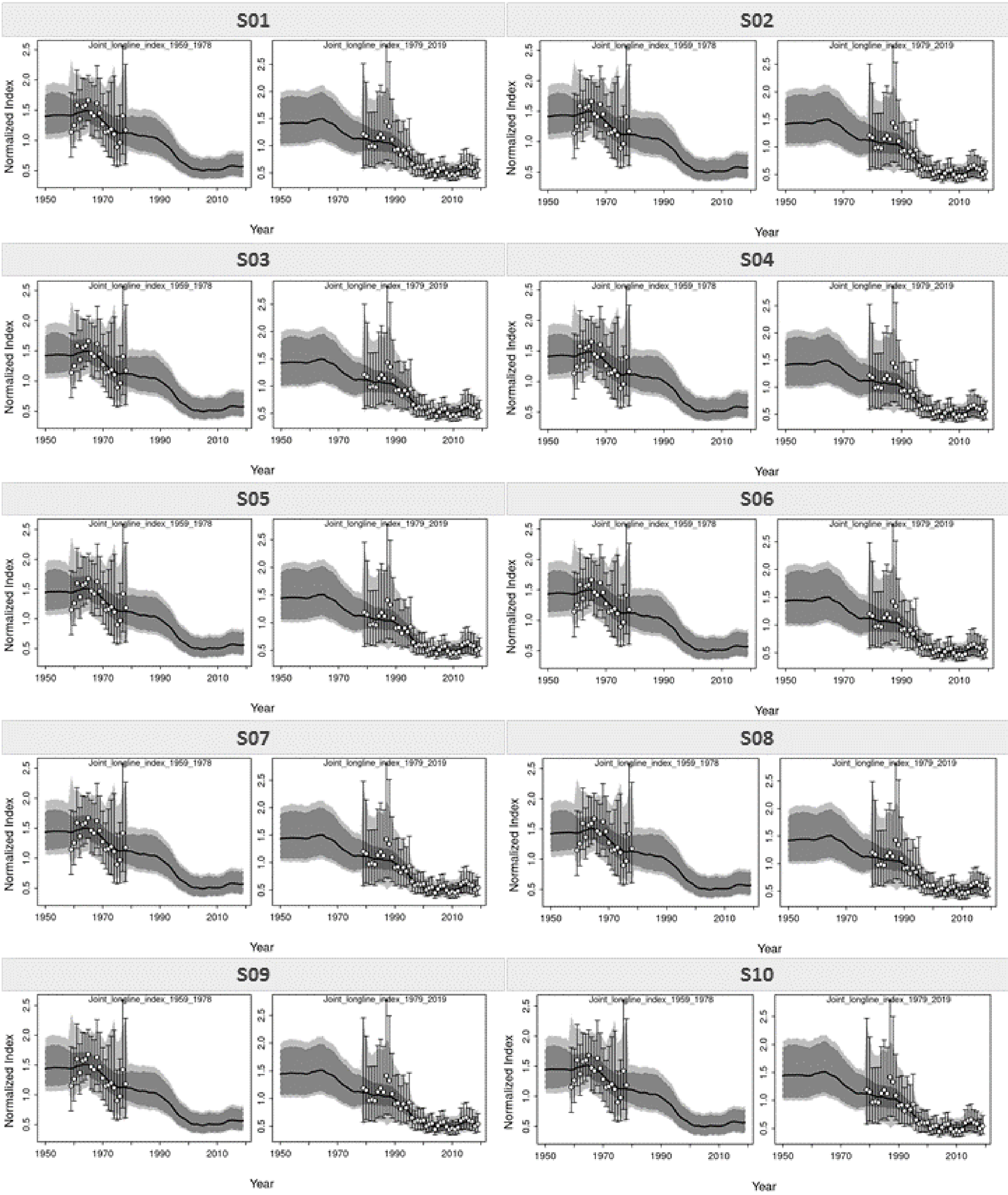

Figure 3. Time series of observed (circle) with error 95% Cis (error bars) and predicted (solid line) CPUE of Atlantic bigeye tuna for the Bayesian state-space surplus production model JABBA for each scenario fitted. Dark shaded grey areas show 95% credibility intervals of the expected mean CPUE and light shaded grey areas denote the 95% posterior predictive distribution intervals.

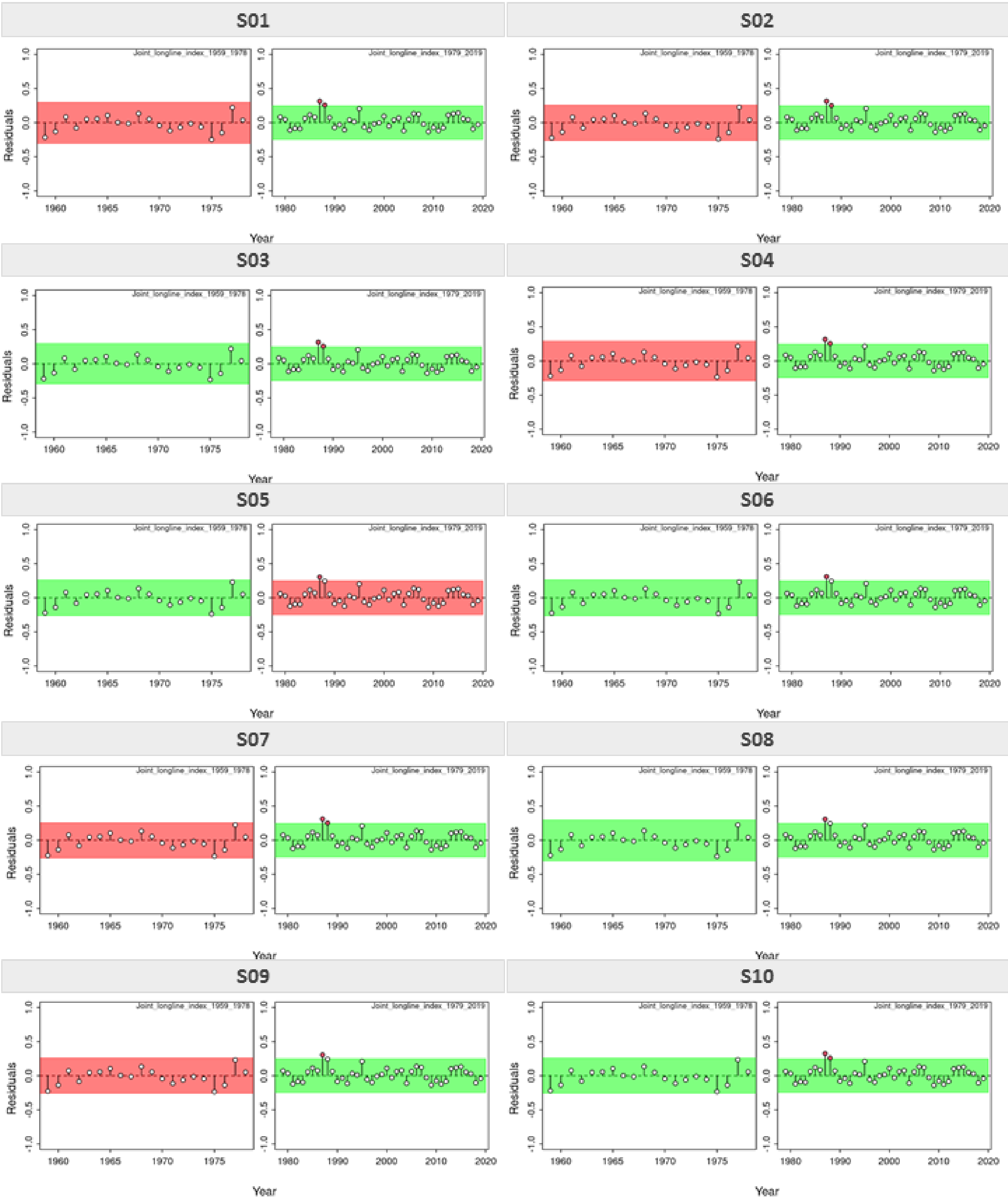

Figure 4. Runs tests to quantitatively evaluate the randomness of the time series of CPUE residuals for each scenario fitted for the Atlantic bigeye tuna. Green panels indicate no evidence of lack of randomness of time-series residuals (p>0.05) while red panels indicate the opposite. The inner shaded area shows three standard errors from the overall mean and red circles identify a specific year with residuals greater than this threshold value (3x sigma rule).

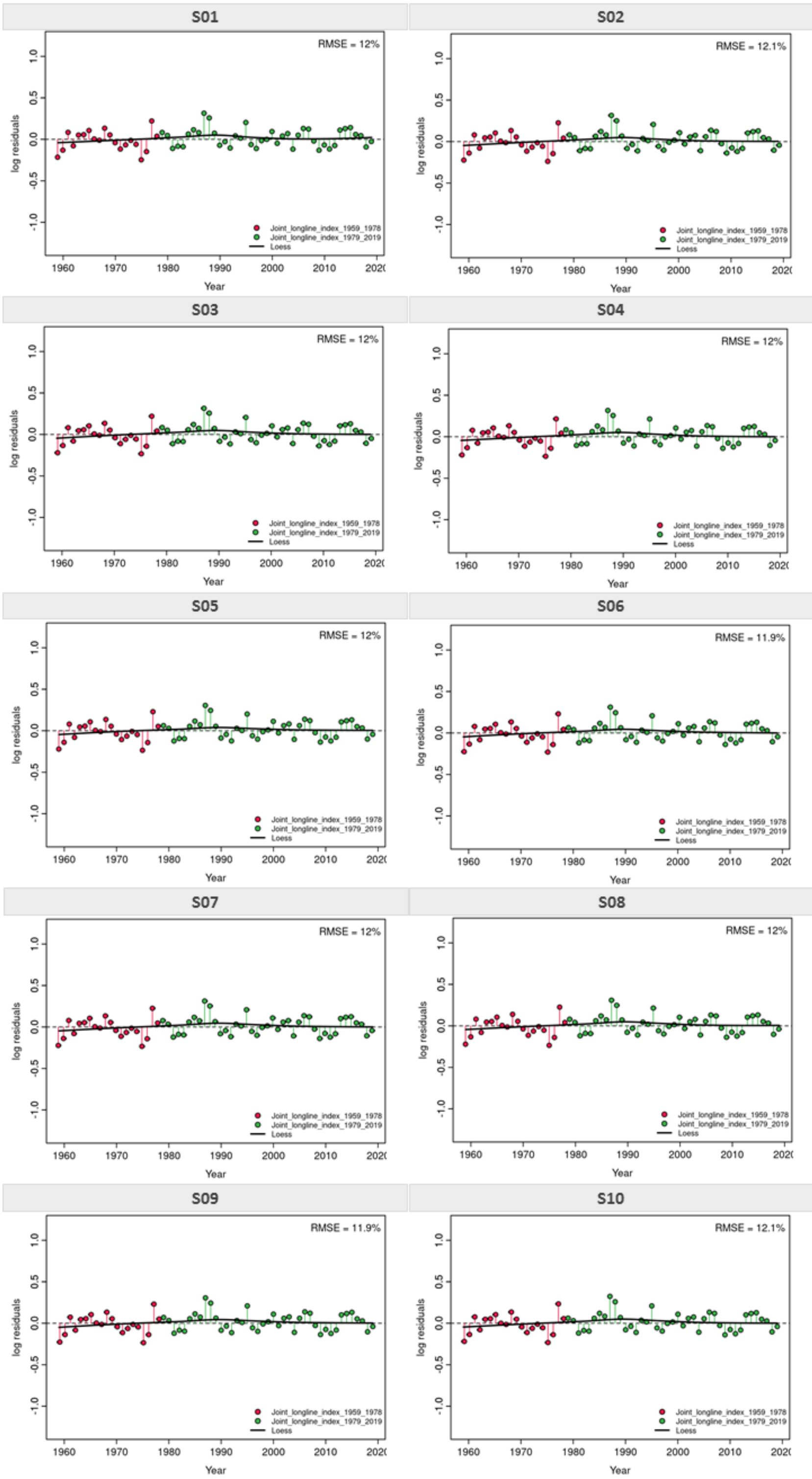

Figure 5. JABBA residual diagnostic plots for alternative sets of CPUE indices examined for each scenario fitted for the Atlantic bigeye tuna. Boxplots indicate the median and quantiles of all residuals available for any given year, and solid black lines indicate a loess smoother through all residuals.

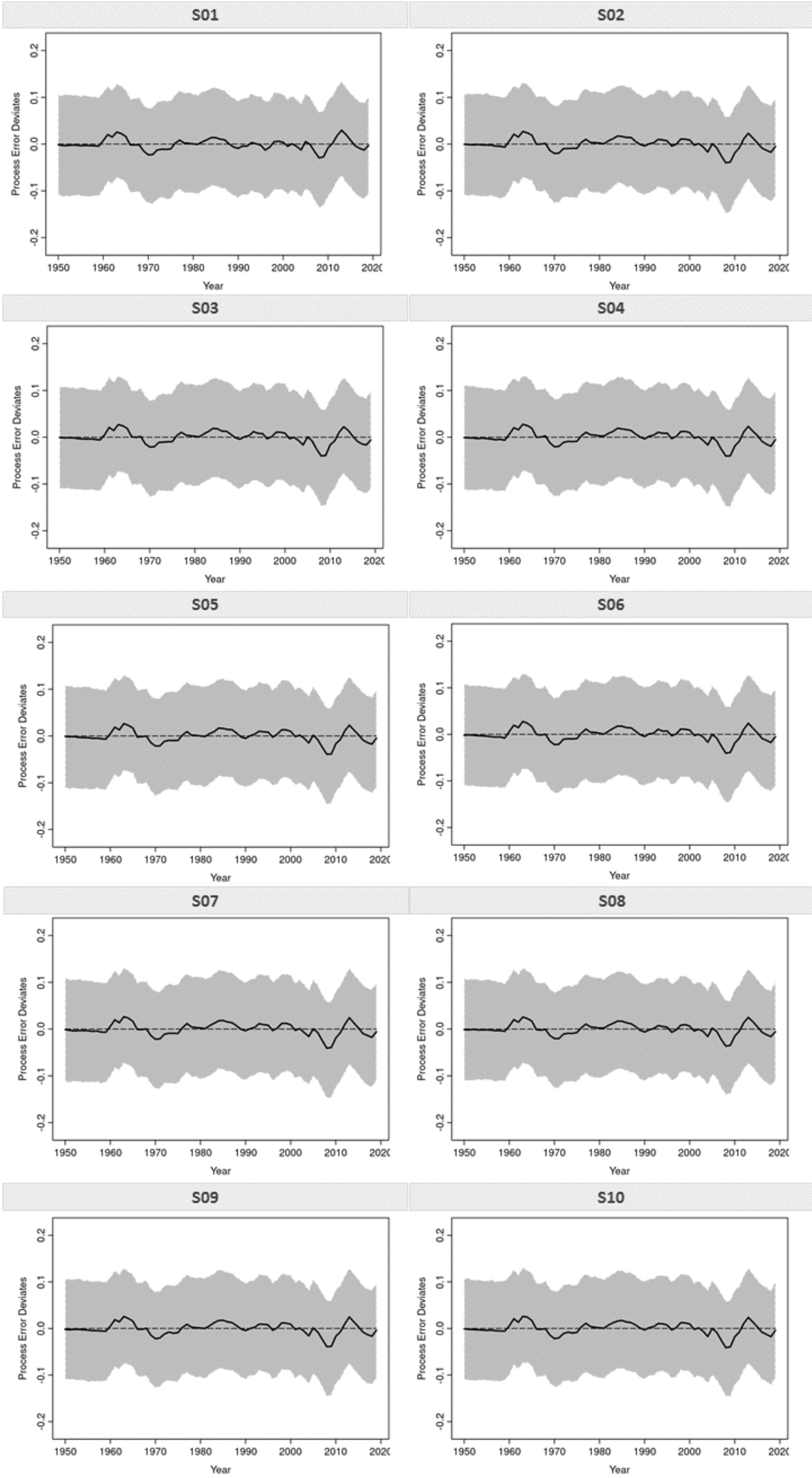

Figure 6. JABBA residual diagnostic plots for alternative sets of CPUE indices examined for each scenario fitted for the Atlantic bigeye tuna. Process error deviates (median: solid line) with shaded grey area indicating 95% credibility intervals.

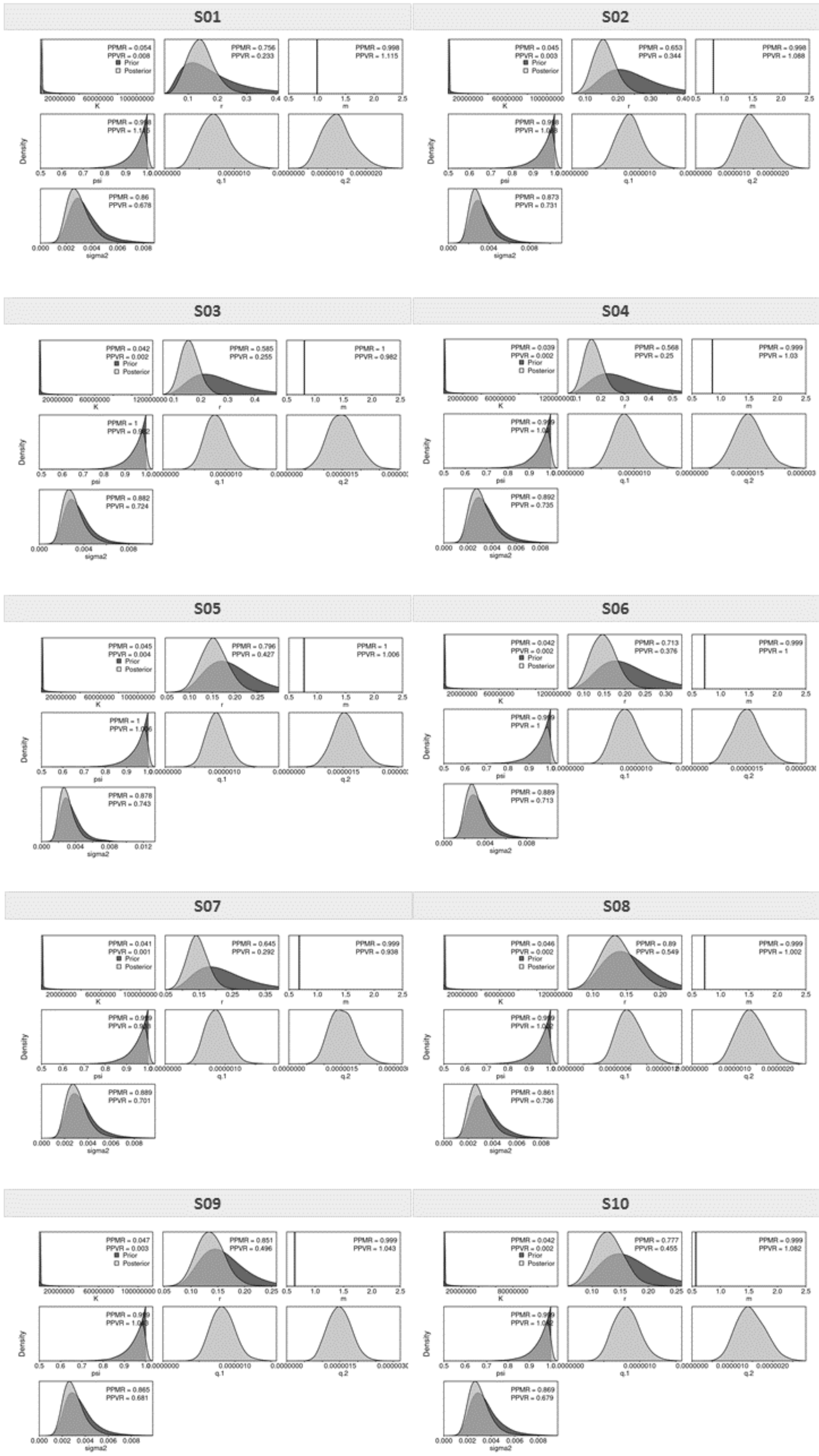

Figure 7. Prior and posterior distributions of various model and management parameters for the Bayesian state-space surplus production fitted for the Atlantic bigeye tuna. PPRM: Posterior to Prior Ratio of Medians; PPRV: Posterior to Prior Ratio of Variances.

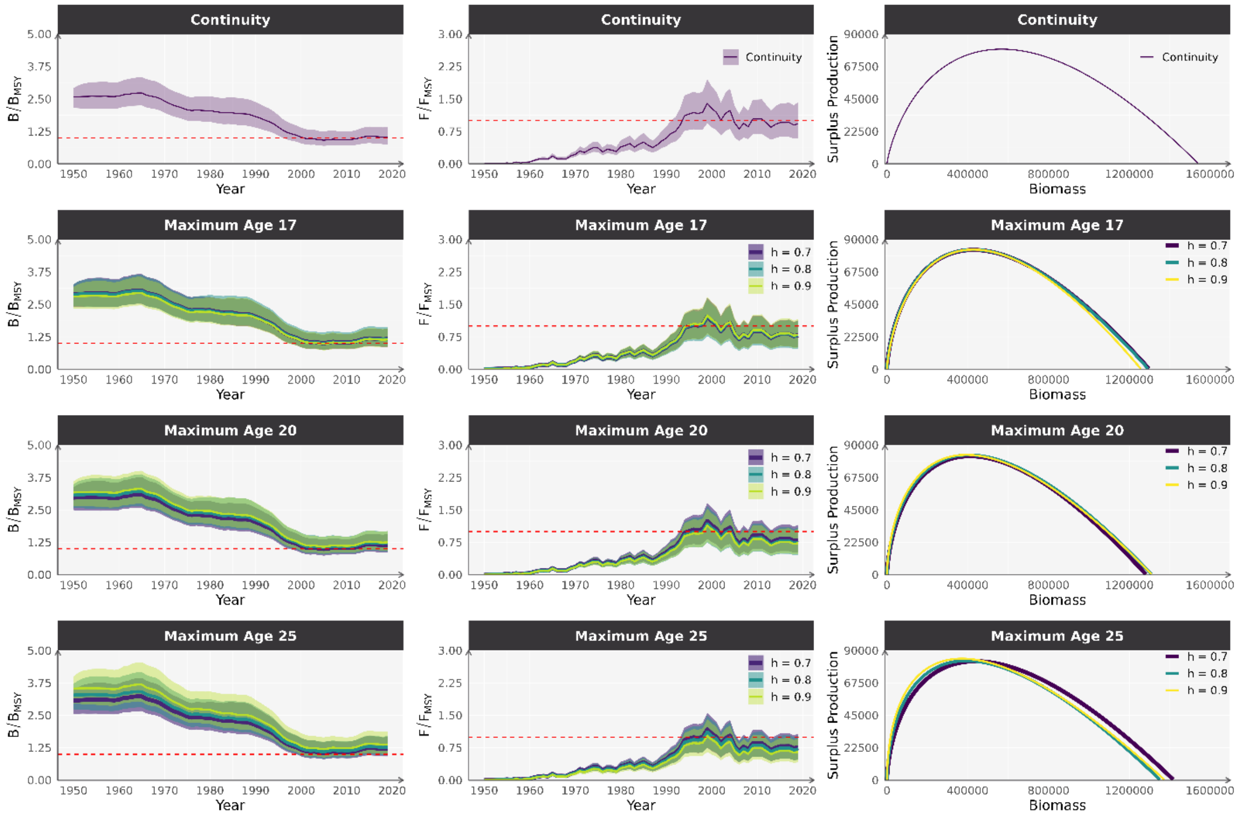

Figure 8. Trends in biomass relative to B_{MSY} (B/B_{MSY}), fishing mortality relative to F_{MSY} (F/F_{MSY}) and surplus production curve (bottom panels) for each scenario from the Bayesian state-space surplus production JABBA model fits to Atlantic bigeye tuna.

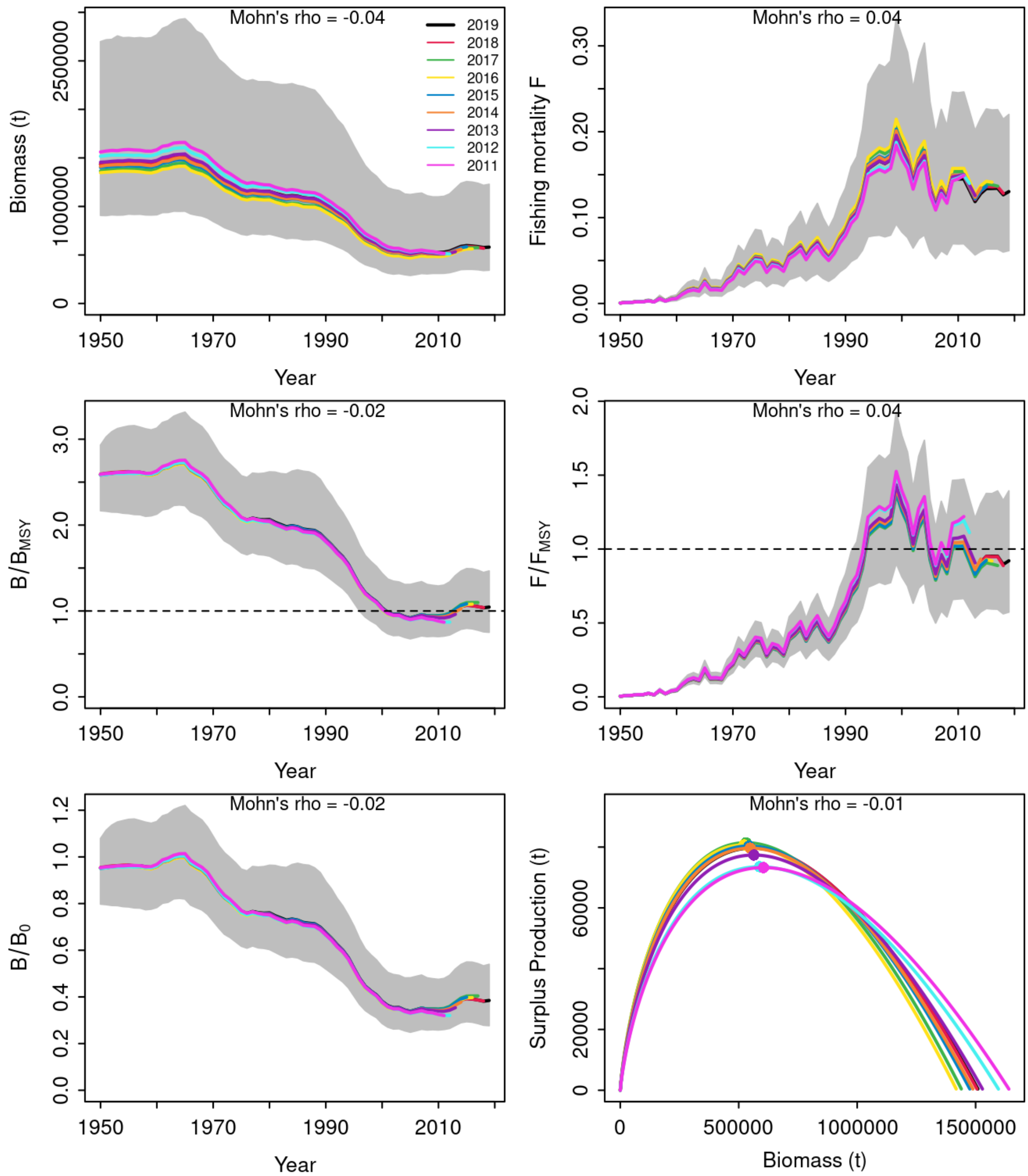

Figure 9. Retrospective analysis conducted for scenario S01: Continuity run: Fox 2021 for Atlantic bigeye tuna, by removing one year at a time sequentially (n=8) and predicting the trends in biomass and fishing mortality (upper panels), biomass relative to B_{MSY} (B/B_{MSY}) and fishing mortality relative to F_{MSY} (F/F_{MSY}) (middle panels) and biomass relative to K (B/K) and surplus production curve (bottom panels) from the Bayesian state-space surplus production model fits.

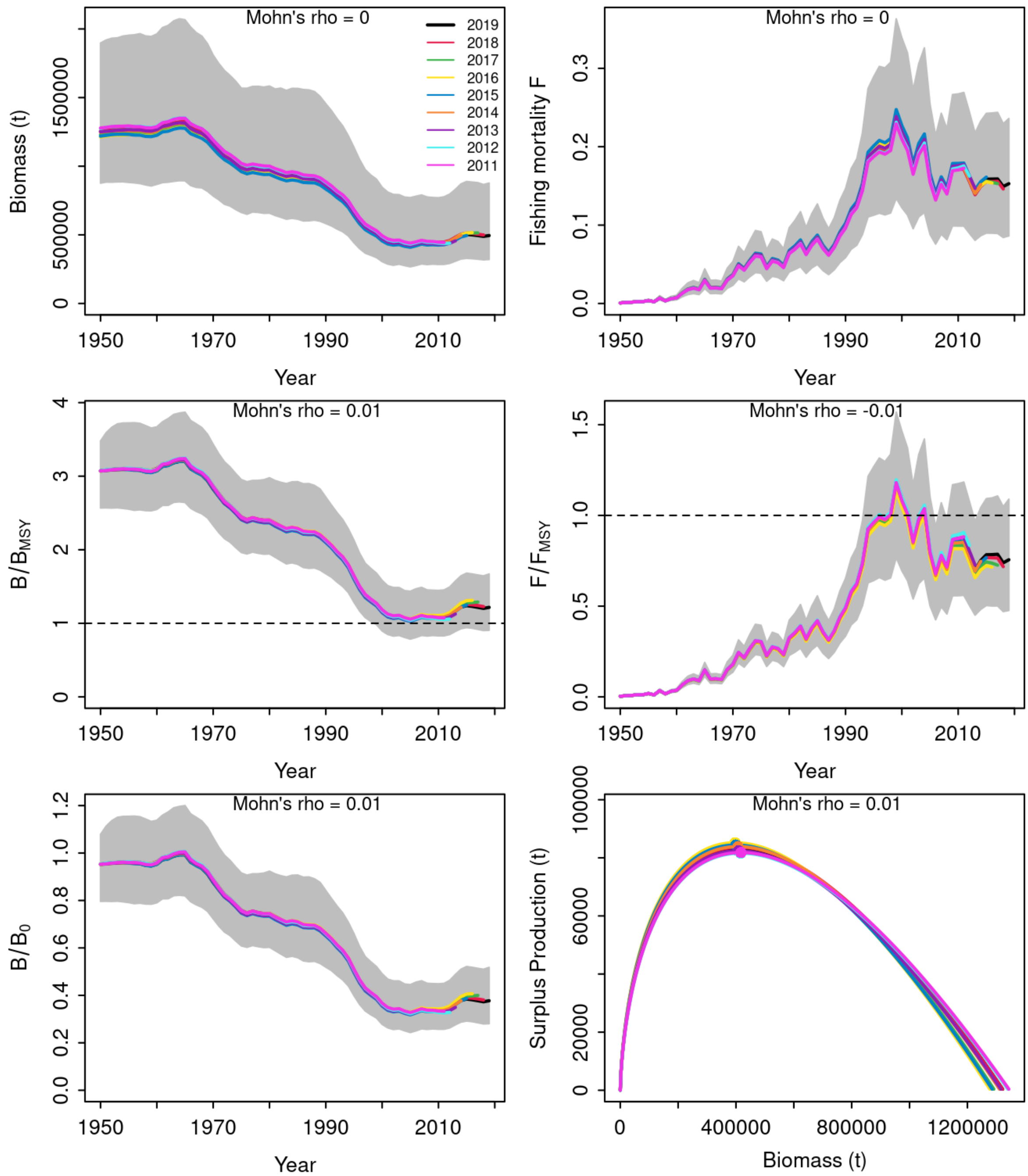

Figure 10. Retrospective analysis conducted for scenario ASEM M = 20 | h = 0.8: Pella m for Atlantic bigeye tuna, by removing one year at a time sequentially (n=8) and predicting the trends in biomass and fishing mortality (upper panels), biomass relative to B_{MSY} (B/B_{MSY}) and fishing mortality relative to F_{MSY} (F/F_{MSY}) (middle panels) and biomass relative to K (B/K) and surplus production curve (bottom panels) from the Bayesian state-space surplus production model fits.

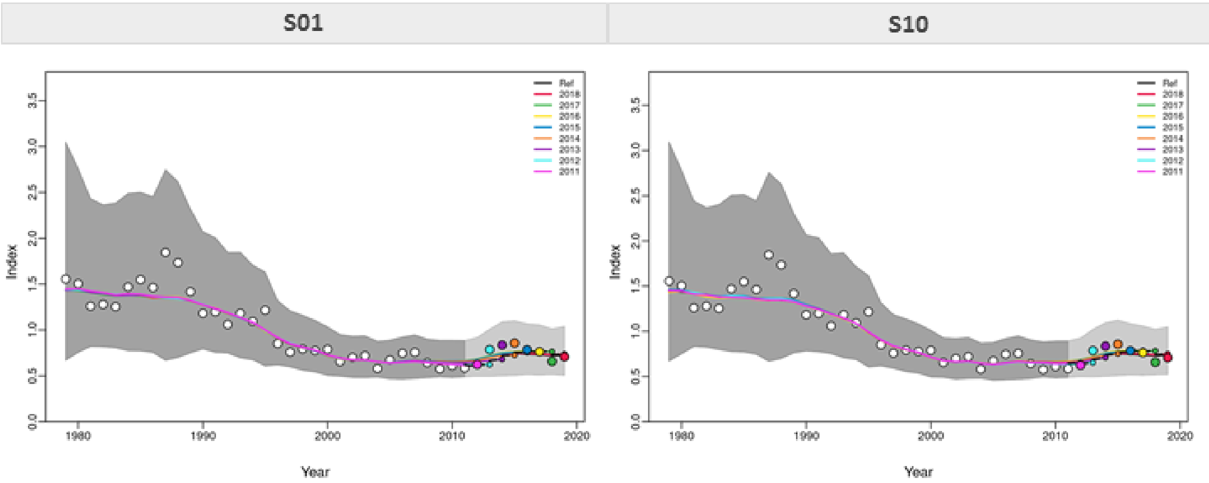

Figure 11. Hindcasting cross-validation results (HCxval) for the two scenarios S01: Continuity run: Fox 2021 and ASEM M = 20 | h = 0.8: Pella m for Atlantic bigeye tuna, showing one-year-ahead forecasts of CPUE values (2011-2019), performed with eight hindcast model runs relative to the expected CPUE. The CPUE observations, used for cross-validation, are highlighted as color-coded solid circles with associated light-grey shaded 95% confidence interval. The model reference year refers to the end points of each one-year-ahead forecast and the corresponding observation (i.e. year of peel + 1).

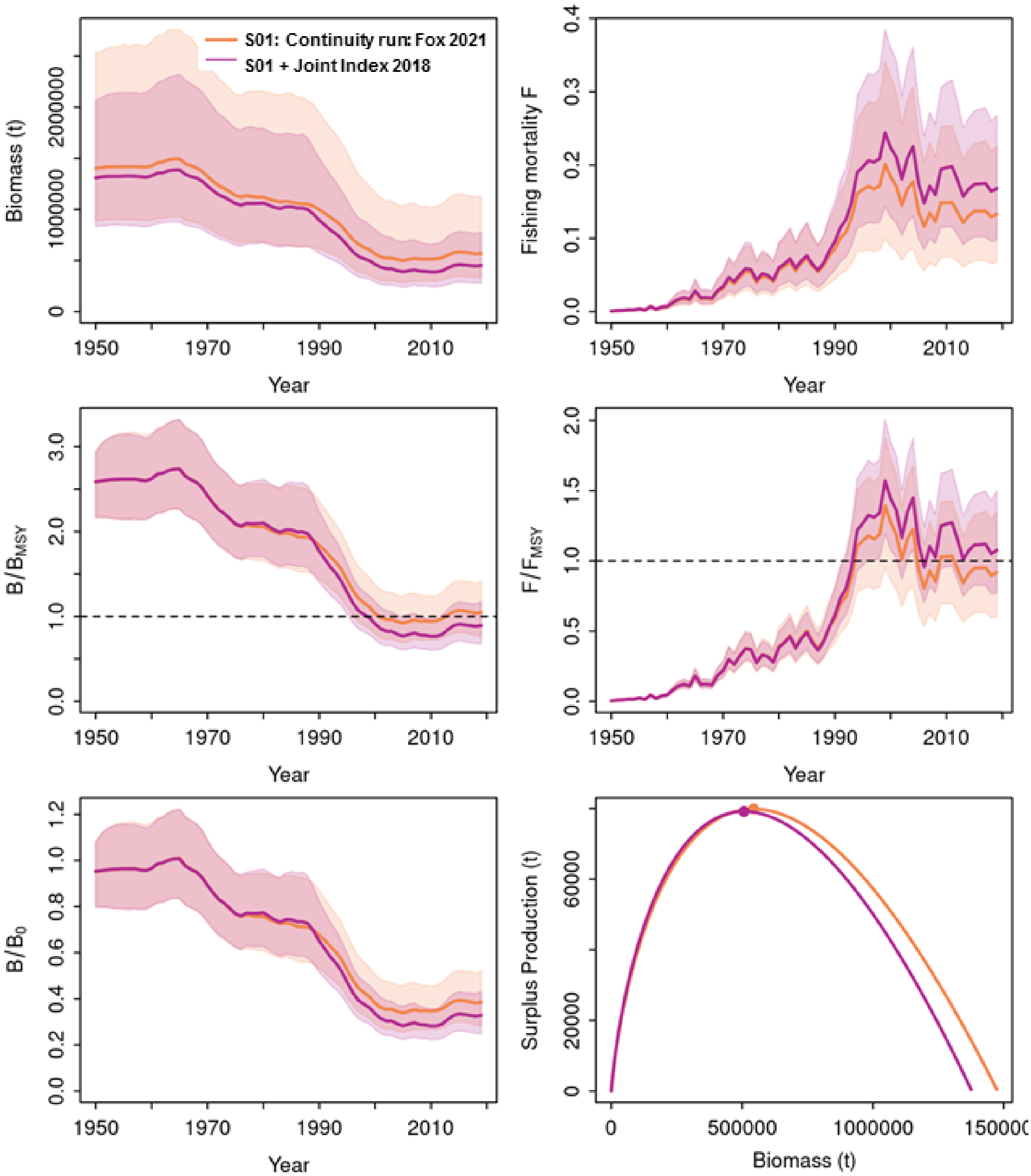

Figure 12. Sensitivity analysis performed for scenarios S01: Continuity run: Fox 2021 and S01: Continuity run: Fox 2021 including the joint index used in the last BET stock assessment showing the trends in biomass and fishing mortality (upper panels), biomass relative to B_{MSY} (B/B_{MSY}) and fishing mortality relative to F_{MSY} (F/F_{MSY}) (middle panels) and biomass relative to K (B/K) and surplus production curve (bottom panels) for the Atlantic bigeye tuna.

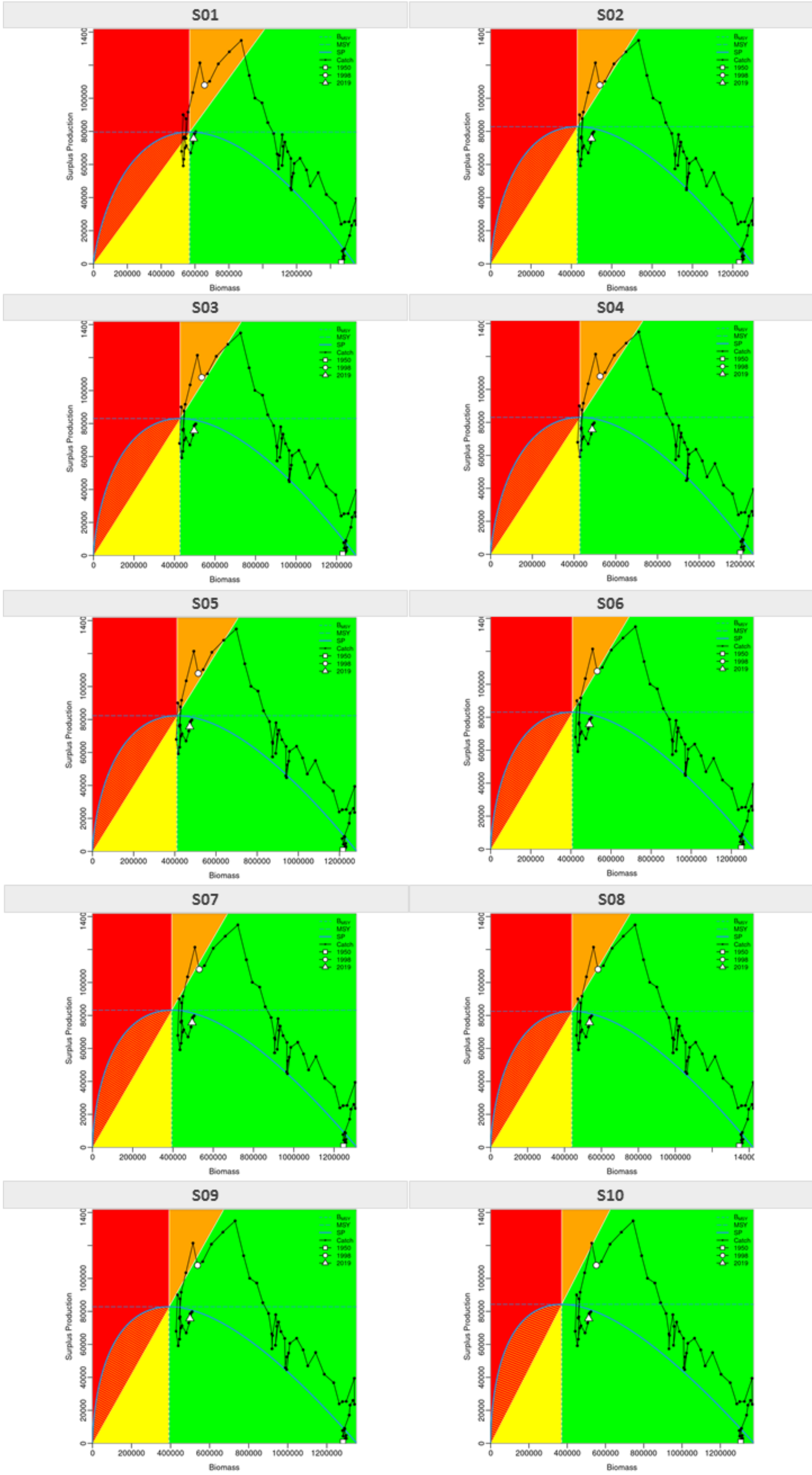

Figure 13. JABBA surplus production phase plot for the Bayesian state-space surplus production for each scenario showing trajectories of the catches in relation to B_{MSY} and MSY for the Atlantic bigeye tuna.

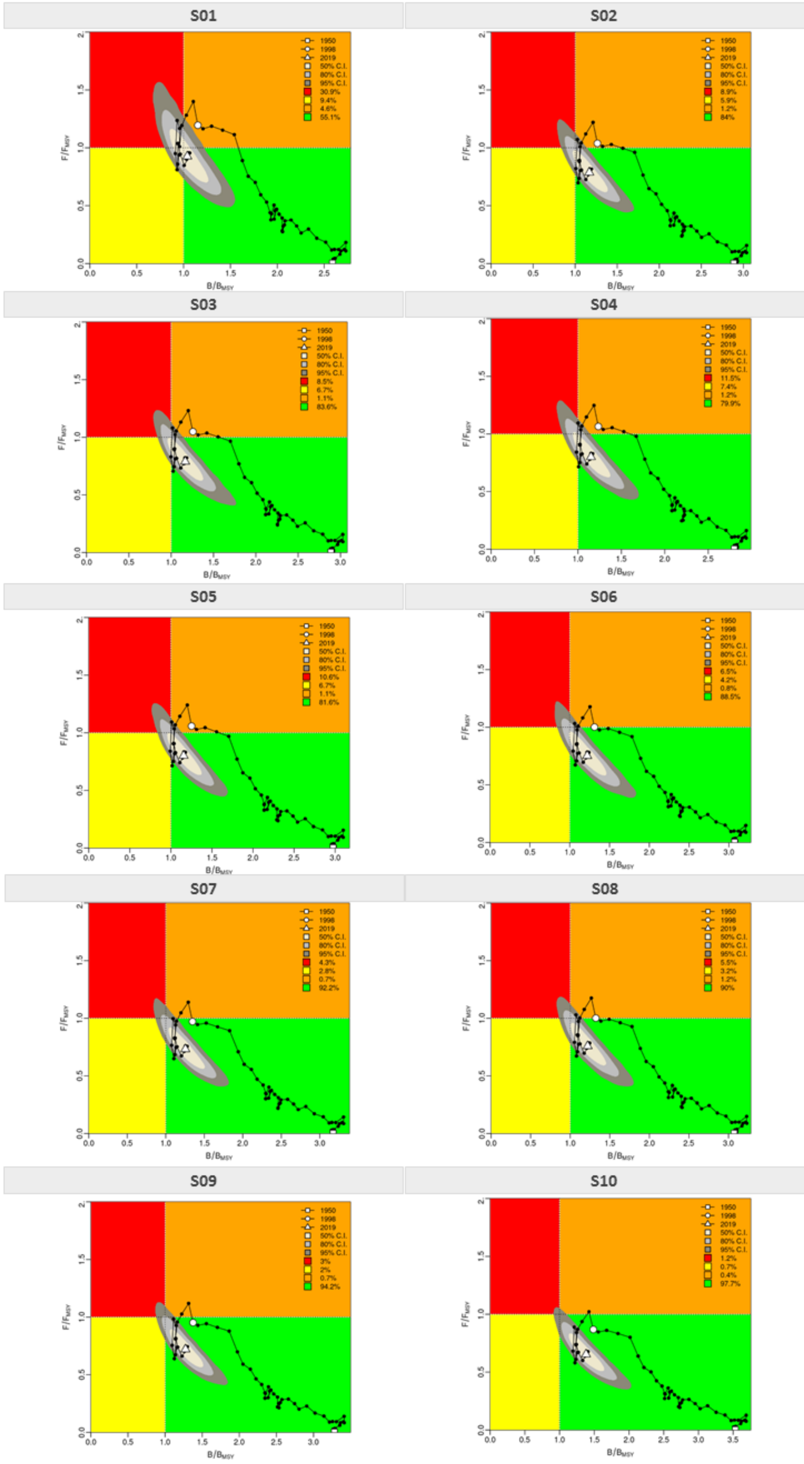

Figure 14. Kobe phase plot showing estimated trajectories (1950-2019) of B/B_{MSY} and F/F_{MSY} for the Bayesian state-space surplus production model for the Atlantic bigeye tuna. Different grey shaded areas denote the 50%, 80%, and 95% credibility interval for the terminal assessment year. The probability of terminal year points falling within each quadrant is indicated in the figure legend.