SCRS/2020/104 Collect. Vol. Sci. Pap. ICCAT, 77(7): 352-376 (2020)

PRELIMINARY STOCK ASSESSMENT OF SOUTH ATLANTIC ALBACORE TUNA (*THUNNUS ALALUNGA*) USING THE BAYESIAN STATE-SPACE SURPLUS PRODUCTION MODEL *JABBA*

Winker H.[1], Mourato B.[2], Parker D.[3], Sant'Ana R.[4], Kimoto A.[5], and Ortiz M.[5]

SUMMARY

Bayesian State-Space Surplus Production Models were fitted to South Atlantic albacore (Thunnus alalunga) catch and CPUE data using the 'JABBA' R package. In accordance with the 2019 SCRS work plan (update of the 2016 assessment), this document presents four preliminary scenarios that explore two production functions (Schaefer or Fox) as well as two CPUE data weighting scenarios (equal or model-internal weighting). Model diagnostics indicated reasonable fits to the data, no evidence of an undesirable retrospective pattern and a satisfying prediction skill to forecast into the future. Notable differences were the change in scale of absolute biomass between Schaefer and Fox models and a slightly more pessimistic stock depletion during in the late 1990s when applying equal CPUE weighting. The current stock status estimates were found to be insensitive model weighting and the production function, indicating a 98.4% - 100% probability that stock is not overfished or subject to overfishing. The consistency in current status estimates and model diagnostic results provides a degree of confidence in the updated assessment of the stock status of South Atlantic albacore for scientific advice.

RÉSUMÉ

*Les modèles de production excédentaire état-espace de type bayésien ont été ajustés aux données de capture et de CPUE du germon de l'Atlantique Sud (*Thunnus alalunga*) au moyen du progiciel JABBA R. Conformément au plan de travail de 2019 du SCRS (mise à jour de l'évaluation de 2016), ce document présente quatre scénarios préliminaires qui explorent deux fonctions de production (Schaefer ou Fox) ainsi que deux scénarios de pondération des données de CPUE (pondération égale ou interne au modèle). Les diagnostics du modèle ont indiqué des ajustements raisonnables aux données, aucune preuve d'un modèle rétrospectif indésirable et une capacité de prévision satisfaisante pour l'avenir. Les différences notables ont été le changement d'échelle de la biomasse absolue entre les modèles de Schaefer et de Fox et un épuisement du stock légèrement plus pessimiste à la fin des années 1990, lorsqu'une pondération égale des CPUE était appliquée. Les estimations actuelles de l'état du stock se sont avérées insensibles à la pondération du modèle et à la fonction de production, indiquant une probabilité de 98,4 % à 100 % que le stock ne soit pas surexploité ou victime de surpêche. La cohérence des estimations de l'état actuel et des résultats du diagnostic du modèle permet d'avoir un certain degré de confiance dans l'évaluation actualisée*

RESUMEN

Los modelos de producción excedente bayesianos de estado espacio se ajustaron a los datos de captura y CPUE del atún blanco (Thunnus alalunga) del Atlántico sur utilizando el paquete R de «JABBA». De conformidad con el plan de trabajo del SCRS para 2019 (actualización de la evaluación de 2016), este documento presenta cuatro escenarios preliminares que exploran dos funciones de producción (Schaefer o Fox), así como dos escenarios de ponderación de los

[1] Joint Research Centre (JRC), European Commission, Ispra, Italy.
[2] Instituto do Mar, Universidade Federal de São Paulo, Av. Doutor Carvalho de Mendonça, 144, 11070-100, Santos, Brazil.
[3] Department of Agriculture, Forestry and Fisheries (DAFF), Fisheries Branch, 8012 Foreshore, Cape Town, South Africa.
[4] Laboratório de Estudos Marinhos Aplicados, Escola do Mar, Ciência e Tecnologia, Universidade do Vale do Itajaí, Rua Uruguai, 458, 88302-901, Itajaí, Santa Catarina, Brazil.
[5] ICCAT Secretariat. Calle Corazón de Maria 8, Madrid Spain 28002.

datos de la CPUE (ponderación igual o interna del modelo). Los diagnósticos de los modelos indicaron ajustes razonables a los datos, ninguna evidencia de un patrón retrospectivo indeseable y una capacidad de predicción satisfactoria para pronosticar el futuro. Las diferencias notables consistieron en el cambio de escala de la biomasa absoluta entre los modelos de Schaefer y Fox y una estimación ligeramente más pesimista de la merma del stock a finales de los años noventa cuando se aplicó la ponderación igual de CPUE. Se determinó que las actuales estimaciones del estado del stock no eran sensibles a la ponderación del modelo y a la función de producción, lo que indicaba una probabilidad del 98,4 % al 100% de que el stock no esté sobrepescado ni siendo objeto de sobrepesca. La coherencia de las estimaciones del estado actual y de los resultados de los diagnósticos del modelo proporcionan cierto grado de confianza en la evaluación actualizada

KEYWORDS

South Atlantic albacore, stock status, JABBA, CPUE fits, hindcast, surplus production function

1. Introduction

The albacore tuna (*Thunnus alalunga*) is widely distributed in temperate and tropical waters of all oceans, including the Mediterranean Sea (Collette and Nauen, 1983). This species has habitat preferences for an epipelagic and mesopelagic realm and prefers cooler sea temperatures in comparison to other tropical tuna species. In the Atlantic Ocean, due to its ample distribution, albacore has been intensively exploited by a variety of fisheries. For management purposes, the International Commission for the Conservation of Atlantic Tunas (ICCAT) considers three stocks, a North Atlantic, a South Atlantic and a Mediterranean stock. The northern and southern Atlantic stocks are currently separated by the 5°N line of Latitude. The southern stock is mainly exploited by pelagic longline fleets and, to a lesser extent, by baitboat fleets. Over the past two decades, 97% of the total annual South Atlantic albacore landings were attributed to these two fleet types, with the Chinese Taipei longline fleet landing 52% of this total, followed by baitboat fleets from South Africa and Namibia, with 15% and 9% respectively and the Brazilian longline fleet with 5%. During this period, average landings have declined from approximately 25,000 t (in 2000-2010) to 18,000 t (in 2011-2018).

In 2016, the ICCAT carried out a stock assessment for South Atlantic albacore (ICCAT, 2016a), which included outputs from the non-equilibrium production model, ASPIC (Matsumoto, 2017; Prager, 2002), and from the Bayesian Production Model, BSP (McAllister, 2014), that were fitted to catch time series and standardized catch-per-unit-effort (CPUE) indices through 2014. Although both models consistently indicated that the southern albacore stock had been undergoing overfishing and had been in an overfished state for an extended period since the late 1990s, the 2014 combined model (ASPIC and BSP) stock status estimates indicated a 66% probability that the stock was in the green quadrant of the Kobe plot (ICCAT, 2016a). The estimates of B_{2014}/B_{MSY} was 1.10 (ranging between 0.51 and 1.80) while of F_{2014}/F_{MSY} was 0.54 (ranging between 0.31 and 0.87), however the high level of uncertainty in stock status model estimates were acknowledged (ICCAT, 2016a).

Here, we present the 2020 preliminary stock assessment results for South Atlantic albacore 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 (www.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) and Mediterranean swordfish (Winker *et al*. 2020; SCRS/2020/082).

This preliminary assessment of the South Atlantic albacore stock is guided by the 2019 SCRS work plan (ICCAT 2020) that recommended to, at a minimum, update the surplus production models up until 2018 following the procedures of the 2016 stock assessment. Extensive model diagnostics are provided to evaluate the model fits, retrospective patterns and prediction skill. In addition to the SCRS workplan's minimum recommended update of the 2016 stock assessment, this document explores the sensitivity of reference 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 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

Fishery catch data for South Atlantic albacore were available by the ICCAT Secretariat for the period 1956-2018 (**Figure 1**). Relative abundance indices were made available in the form of standardized CPUE time series for several fleets. The CPUE time series cover a variety time periods for several fleets, including longline (LL) and baitboat (BB) fisheries (**Figure 2**), as follows:

- Chinese-Taipei LL (CTP-LL) in 1967 - 2018,
- Japan LL in 1959-1969 (JPN-LL1), in 1969-1975 (JPN-LL2), and in 1976-2018 (JPN-LL3),
- Japan LL in core area in 1976-2018 (JPN-LLcore),
- Brazil LL (BRA-LL) in 2002-2018,
- Uruguay LL (URY-LL) in 1983-2011,
- South Africa BB (ZAF-BB) in 2003-2018

It should be noted that the 2016 stock assessment only included CTP-LL (1967-2014), JPN-LL3 (1976-2011), and URY-LL (1983-2011) in the final models, and sensitivity analyses were conducted with the "Japan early" CPUE series (JPN-LL1).

2.3. Model specifications

For the unfished equilibrium biomass K, we used default settings of the JABBA R package in the form of vaguely informative lognormal prior with a large CV of 100% and a central value that corresponds to eight times the maximum total catch, which is consistent with parameterization procedures followed when using other platforms such as Catch-MSY (Martell and Froese, 2013) or SPiCt (Pederson and Berg 2017). The initial depletion prior ($\varphi = B_{1956}/K$) was defined by a beta distribution with mean = 0.9 and CV of 10%. All catchability parameters were formulated as uninformative uniform priors. The process error of $\log(B_y)$ in year y was estimated "freely" by the model using an uninformative inverse-gamma distribution with both scaling parameters set at 0.001.

To provide continuity, initial JABBA runs included the same combination of standardized CPUE time series as applied in the previous assessments (ICCAT, 2013, 2016a), that is: CTP-LL, JPN-LL3 (but removing years after 2011 due to changes in species targeting) and URY-LL.

CPUE input data were characterized according to two alternative data weighting scenarios: 1) equal weighting (EqW), which correspond to a single observation variance estimate to all CPUE indices and; 2) Model-internal weighting (ModW), with index-specific variances being estimated internally by the model. In both instances, the observation variance is estimated assuming inverse-gamma priors (see details in Winker *et a*l., 2018). For the shape of the production function, we considered two model-types: the Schaefer model ($B_{MSY}/K = 0.5$) and the

Fox model (B_{MSY}/K = 0.37). As per previous assessments, all models assume a vaguely informative prior for $r \sim LN(\log(0.2),1)$, which broadly resembles the Bayesian Surplus Production (BSP) model prior developed by Babcock (2012). Accordingly, we formulated the following four reference model scenarios for South Atlantic albacore:

- S1: Schaefer EqW
- S2: Fox EqW
- S3: Schaefer ModW
- S4: Fox ModW

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. 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 (Heidelberger and Welch, 1992) and 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 (S3 and S4), by systematically removing one year of data at a time sequentially 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 assessment results to alternative indices, we also considered 6 sets of CPUE series and provided sensitivity analyses (**Table 1**) based on the Schaefer and Fox scenarios with ModW (i.e. S3 and S4). We choose the two ModW scenarios for the sensitivity runs to avoid many of the problems with equal weighting, such as ignoring the model goodness of fit for the CPUE data and potentially further down-weighting a well fitting index by adding an additional index leading to inconsistencies in cases where the number indices vary among scenarios.

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 fit to each of the three standardized CPUE LL indices are shown in **Figures 3** for each of the four reference scenarios (S1-S4**).** All scenarios appeared to fit the CTP-LL and JPN-LL3 (1975-2011) data reasonably well, with exceptions of large, occasional deviations in the JPN-LL3 index. In contrast, CPUE from the URY-LL fleet indicated a fairly poor fit, in particular to the CPUE observations over the period 2000-2005, which showed a sudden systematic decrease over this period that was in conflict with the other indices. The results of the log-residuals runs test for each CPUE fit by year and model are provided in **Figure 4,** whereby green panels indicate CPUE indices that passed the runs test, with no evidence for a non-random residual pattern (p>0.05) and red panels indicating a failed runs test. The inner shaded area shows 3-sigma limits (Anhøj and Olesen, 2014) around the overall mean and red circles identify a specific year with residuals greater than this threshold limit. In almost all models, CPUE time series from CTP-LL, JPN-LL3 and URY-LL failed the runs test diagnostic procedure, with the exception of scenario S4 fitted to JPN-LL3 (**Figure 4**). The reason for failing the runs tests is probably related to data-conflicts caused by the URY-LL indicating opposite trends compared to the CTP-LL and JPN-LL3 over period 2000-2001, but also before around 1990 (**Figures 4** and **5**). The goodness-of-fit were comparable among all scenarios, ranging from 34.4% (S2) to 36.9% (S3) (**Figure 5**). Annual process error deviation on log biomass (**Figure 6**) indicated similar stochastic patterns, associated with relatively small process error estimates (< 0.05), which suggest no evidence of structural model misspecification causing conflicts with the predicted population dynamics.

The medians of marginal posteriors for r ranged between 0.513 (S1) and 0.299 (S2) for the Schaefer models and 0.396 (S3) and 0.268 (S4) for the Fox models (**Table 2**). The range of posterior to prior median ratios (PPMR) for r (0.83 - 1.62), as well as the range of posterior to prior variance ratios (PPVR) of r (0.039 - 0.070) indicate that the estimate of r is largely informed by the data, because of a much higher precision relative to the prior (**Figure 7**). The estimated median of marginal posterior for K was slightly lower for the Schaefer models (S1 = 218,999; S2 = 249,585 metric tons) than that for the Fox models (S3 = 285,454; S4 = 285,231 metric tons) (**Table 2**). The relatively small PPVRs indicate that the posterior estimates of K were also largely informed by the data. The range of MSY median estimates was narrow among all four scenarios (27,219 (S2) - 28,016 (S3) metric tons). Likewise, the marginal posterior medians for B_{MSY} varied between 92,365 (S2) and 142,727 (S3) metric tons and estimates of F_{MSY} were also similar among scenarios with median values varying from 0.198 (S3) to 0.295 (S2) (**Table 2**).

In general, all models showed similar trends for the medians of B/B_{MSY} and F/F_{MSY} over time, with scenarios S2 and S4 producing slightly more optimistic stock status estimates (**Figure 8**). The trajectory of B/B_{MSY} decreased sharply in the late-1990s to an overfished status; a trend which continued until 2005. After 2005, the relative biomass rapidly recovered, reaching B_{MSY} around 2015 and has remained above B_{MSY}. The F/F_{MSY} trajectory gradually increased from the beginning of time series until late 1980s, followed by a relatively stable period at around the MSY level. In 2000, a substantial increase in fishing mortality was observed, however this overfishing period was short lived and fishing mortality declined until dipping below F_{MSY} in the late-2000s where it has remained ($F_{2018}/F_{\mathrm{MSY}}$ < 1). The rapid rebuilding in the biomass estimated in recent years can be attributed to the fact that fishing mortality rate has remained below F_{MSY} since late-2000s and recent catches have been well below MSY.

The results of an eight year retrospective analysis applied to scenarios S3 and S4 are depicted in **Figures 9** and **10,** respectively, and show a negligible retrospective pattern for both models. The estimated Mohn's rho for *B* and B/B_{MSY} (**Table 3**) fell within the acceptable range of -0.15 and 0.20 (Hurtado-Ferro *et al*. 2014; *Carvalho et al*. 2017) and confirm the absence of an undesirable retrospective pattern. Hindcasting cross-validation results for the CTP-LL suggest that the both model-types have good prediction skills as judged by the MASE scores of approximately 0.5 (**Figure 11**), which indicates that future projections are consistent with reality of model-based scientific advice.

The surplus production phase plots were similar for all four scenarios, corroborating that the stock was likely being overfished in the mid to late-1990s (**Figure 12**). However this period of overfishing was relatively short and catches returned to below the surplus production curve relatively quickly. Accordingly, the Kobe biplots for all scenarios show the typical anti-clockwise pattern with the stock status moving from underexploited through a period of unsustainable fishing to the overexploited phase (**Figure 13**). All scenarios indicate that the current stock status is in the green quadrant of the Kobe biplot ($B_{2018}>B_{MSY}$ and $F_{2018}<F_{MSY}$; **Table 2**) with almost 100% probability (range: 98.4% - 100%). As such, all four models conclusively estimate that stock not overfished nor subject to overfishing (**Figure 13**).

The sensitivity analysis of model fits and log-residuals runs tests for the alternative scenarios (**Table 1**) are shown in **Figure 14** for the two ModW scenarios (S3 and S4). The inclusion of the updated JPN-LL3 index through 2018 (excluding 2011-2012) resulted in reasonable fit that passed the runs test for both the Schaefer and Fox ModW scenarios. By contrast, the inclusion of JPN-LLcore and JPN-LL1 resulted in poor model fits to these indices. In particular, JPN-LLcore resulted in a notably worse fit compared to the updated JPN-LL3. The model fits to the CPUE indices from BRA-LL and ZAF-BB provided a somewhat conflicting fit to the model predictions in recent years, but these deviations, yet there was no violation of randomness in the residuals (**Figure 14**). The only CPUE which had not passed in the runs test diagnostic was JPN-LLcore (red panels) for both model-types (**Figure 14**). The sensitivity analysis confirmed that the inclusion of BRA-LL and ZAF-BB CPUE time series had little effect on the trajectories of B/B_{MSY}, B/B_0 and F/F_{MSY}, or the overall stock status estimate, all of which remained similar to the reference model scenarios. However, including these CPUE time series inflated the *K* estimate, which could be a result of conflicting CPUE trends between these fleets and CTP-LL in recent years. Yet, with respect to the stock status and MSY estimates, scenarios S4-BRA and S4-ZAF are in agreement with the reference model outputs. In contrast, the inclusion of the poorly fitting JPN-LL core CPUE was the most influential in that it resulted in notably more optimistic stock status trajectories as well as higher MSY estimates (**Figure 15**).

Our results suggest that all candidate models provide reasonably robust fits to the data as judged by the presented model diagnostics, with current stock status estimates being fairly insensitive to variations in model weighting and the shape of the production function. This consistency in current status estimates together with generally favourable model diagnostic results provide a degree of confidence in the updated assessment of the stock status of South Atlantic albacore stock for quota advice.

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.

Collette, B. B. and C. E. Nauen. 1983. FAO species catalogue, vol. 2, Scombrids of the world. FAO. Fisheries synopsis 125 (2): 137 p.

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, 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, 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, 2013. Report of the 2013 ICCAT north and south Atlantic albacore stock assessment meeting. Collect. Vol. Sci. Pap. ICCAT, 114p.

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

Martell, S., Froese, R., 2013. A simple method for estimating MSY from catch and resilience 504–514. https://doi.org/10.1111/j.1467-2979.2012.00485.x

Matsumoto, T. 2017. Stock assessment for South Atlantic Albacore using a non-equilibrium production model. Collect. Vol. Sci. Pap. ICCAT 73, 1454-1469.

McAllister, M.K., 2014. A generalized Bayesian surplus production stock assessment software (BSP2). Collect. Vol. Sci. Pap. -ICCAT 70, 1725–1757.

Mohn, R., 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES J. Mar. Sci. 56, 473–488. https://doi.org/10.1006/jmsc.1999.0481

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.

Pedersen, M.W., Berg, C.W., 2017. A stochastic surplus production model in continuous time. Fish Fish. 18, 226–243. https://doi.org/10.1111/faf.12174

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.

Prager, M.H., 2002. Comparison of logistic and generalized surplus-production models applied to swordfish, *Xiphias gladius*, in the north Atlantic Ocean. Fish. Res. 58, 41–57. https://doi.org/10.1016/S0165-7836(01)00358-7

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., 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., 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 sensitivity analysis runs for South Atlantic albacore

Scenario	Model	Type	Indices
S3.Ref	Schaefer	ModW	CTP LL, JPN LL3 (1976-2011), URY LL
S3.JPN2018	Schaefer	ModW	CTP LL, JPN LL3 (1976-2011 and 2014-2018), URY LL
S3.JPNcore	Schaefer	ModW	CTP LL, URY LL, JPN LLcore
S3.JPNearly	Schaefer	ModW	CTP LL, JPN LL3 (1976-2011), URY LL, JPN LL1
S3.BRA	Schaefer	ModW	CTP LL, JPN LL3 (1976-2011), URY LL, BRA LL
S3.ZAF	Schaefer	ModW	CTP LL, JPN LL3 (1976-2011), URY LL, ZAF BB
S4.Ref	Fox	ModW	CTP LL, JPN LL3 (1976-2011), URY LL
S4.JPN2018	Fox	ModW	CTP LL, JPN LL3 (1976-2011 and 2014-2018), URY LL
S4.JPNcore	Fox	ModW	CTP LL, URY LL, JPN LLcore
S4.JPNearly	Fox	ModW	CTP LL, JPN LL3 (1976-2011), URY LL, JPN LL1
S4.BRA	Fox	ModW	CTP LL, JPN LL3 (1976-2011), URY LL, BRA LL
S4.ZAF	Fox	ModW	CTP LL, JPN LL3 (1976-2011), URY LL, ZAF BB

Table 2. 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 South Atlantic albacore.

Estimates	*S1: Schaefer EqW*			*S2: Fox EqW*		
	Median	LCI (2.50%)	UCI (97.50%)	Median	LCI (2.50%)	UCI (97.50%)
K	218999	136710	407454	249585	161510	405964
r	0.513	0.257	0.845	0.299	0.175	0.477
ψ(psi)	0.925	0.656	0.997	0.927	0.675	0.997
σ_{proc}	0.049	0.020	0.108	0.050	0.021	0.111
F_{MSY}	0.256	0.129	0.422	0.295	0.173	0.471
B_{MSY}	109500	68355	203727	92365	59771	150237
MSY	27997	25018	30957	27219	24695	30282
B_{1956}/K	0.917	0.642	1.070	0.919	0.662	1.068
B_{2018}/K	0.734	0.519	0.899	0.605	0.450	0.777
B_{2018}/B_{MSY}	1.468	1.038	1.798	1.635	1.217	2.100
F_{2018}/F_{MSY}	0.416	0.322	0.634	0.384	0.280	0.548
Estimates	*S3: Schaefer ModW*			*S4: Fox ModW*		
	Median	LCI (2.50%)	UCI (97.50%)	Median	LCI (2.50%)	UCI (97.50%)
K	285454	165526	608784	285231	175117	493831
r	0.396	0.175	0.707	0.268	0.151	0.460
ψ(psi)	0.927	0.660	0.997	0.931	0.671	0.997
σ_{proc}	0.057	0.025	0.110	0.052	0.023	0.102
F_{MSY}	0.198	0.088	0.353	0.265	0.149	0.454
B_{MSY}	142727	82763	304392	105557	64807	182755
MSY	28016	24207	32769	27989	24923	32662
B_{1956}/K	0.918	0.648	1.082	0.921	0.665	1.073
B_{2018}/K	0.730	0.554	0.900	0.641	0.494	0.806
B_{2018}/B_{MSY}	1.460	1.107	1.800	1.732	1.335	2.177
F_{2018}/F_{MSY}	0.418	0.308	0.602	0.352	0.251	0.495

Table 3. Summary Mohn's rho statistic computed for a retrospective evaluation period of eight years for the two scenarios (S3) SchaeferModW and (S4) FoxModW for South Atlantic albacore. 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).

	Stock Quantity					
Scenario	B	F	B/B_{MSY}	F/F_{MSY}	B/K	MSY
S3: SchaeferModW	-0.02586	0.02878	0.01338	-0.01971	0.01338	0.00745
S4: FoxModW	-0.02851	0.03167	0.01995	-0.02608	0.01995	0.00655

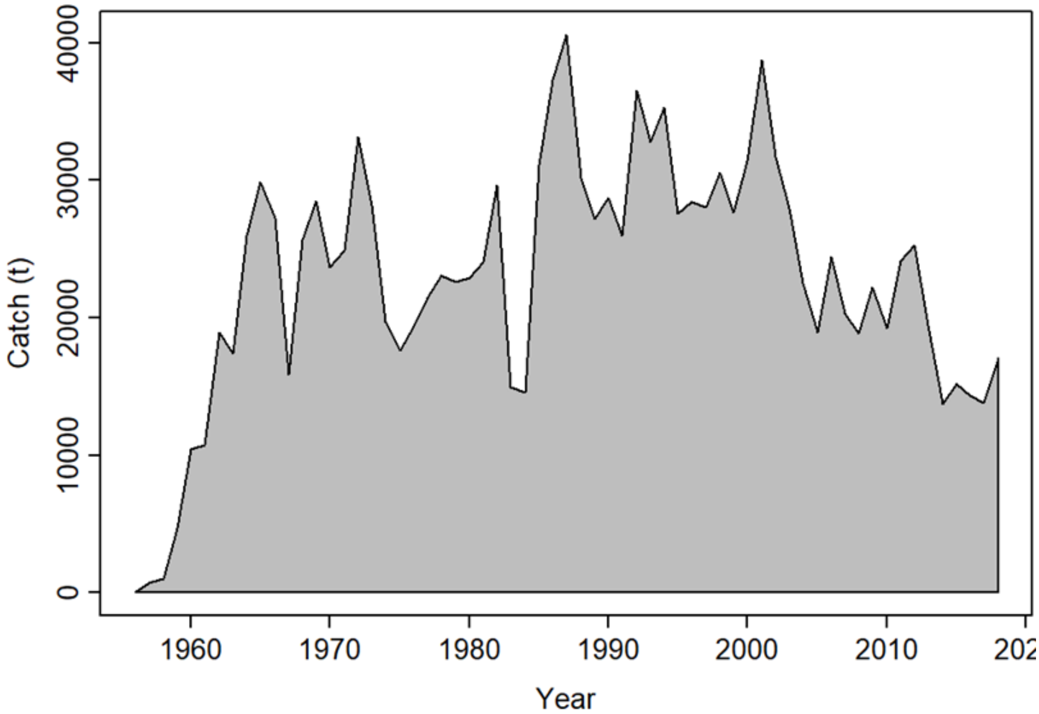


Figure 1. Catch time series in metric tons (t) between 1953 and 2018 for South Atlantic albacore.

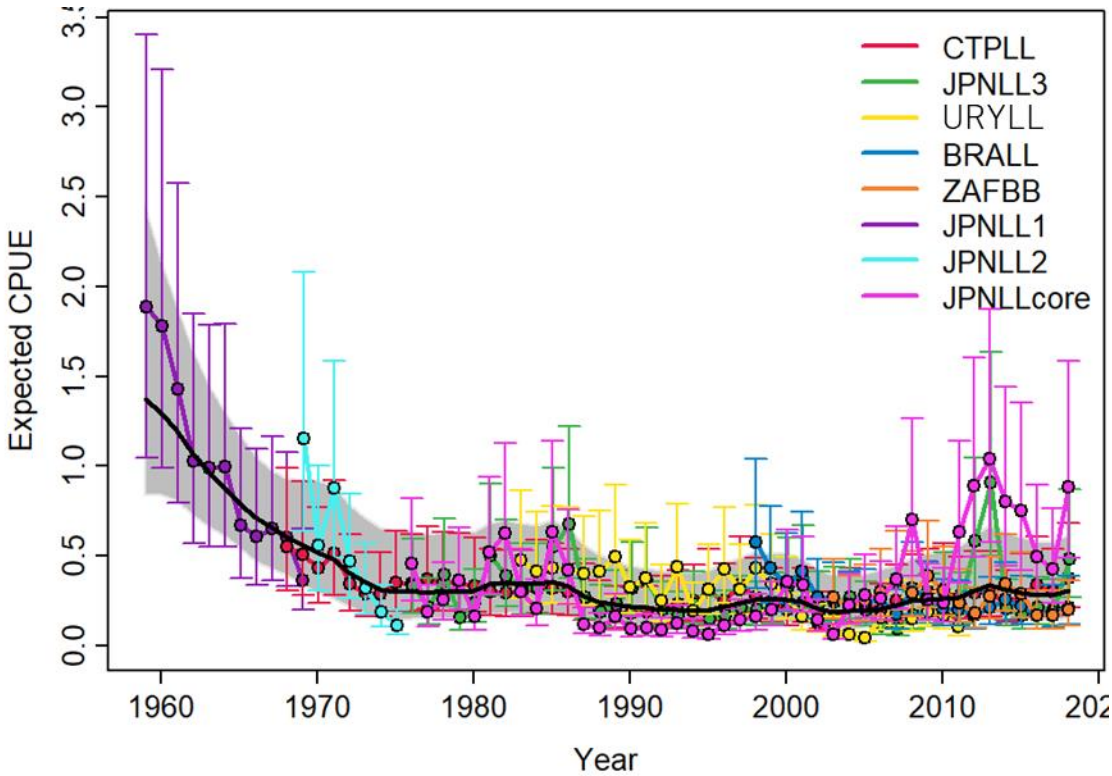


Figure 2. Time-series of four standardized CPUE series for South Atlantic albacore with and assumed Standard Errors of 0.2. Solid black line is the averaged CPUE with associated confidence intervals (shaded area).

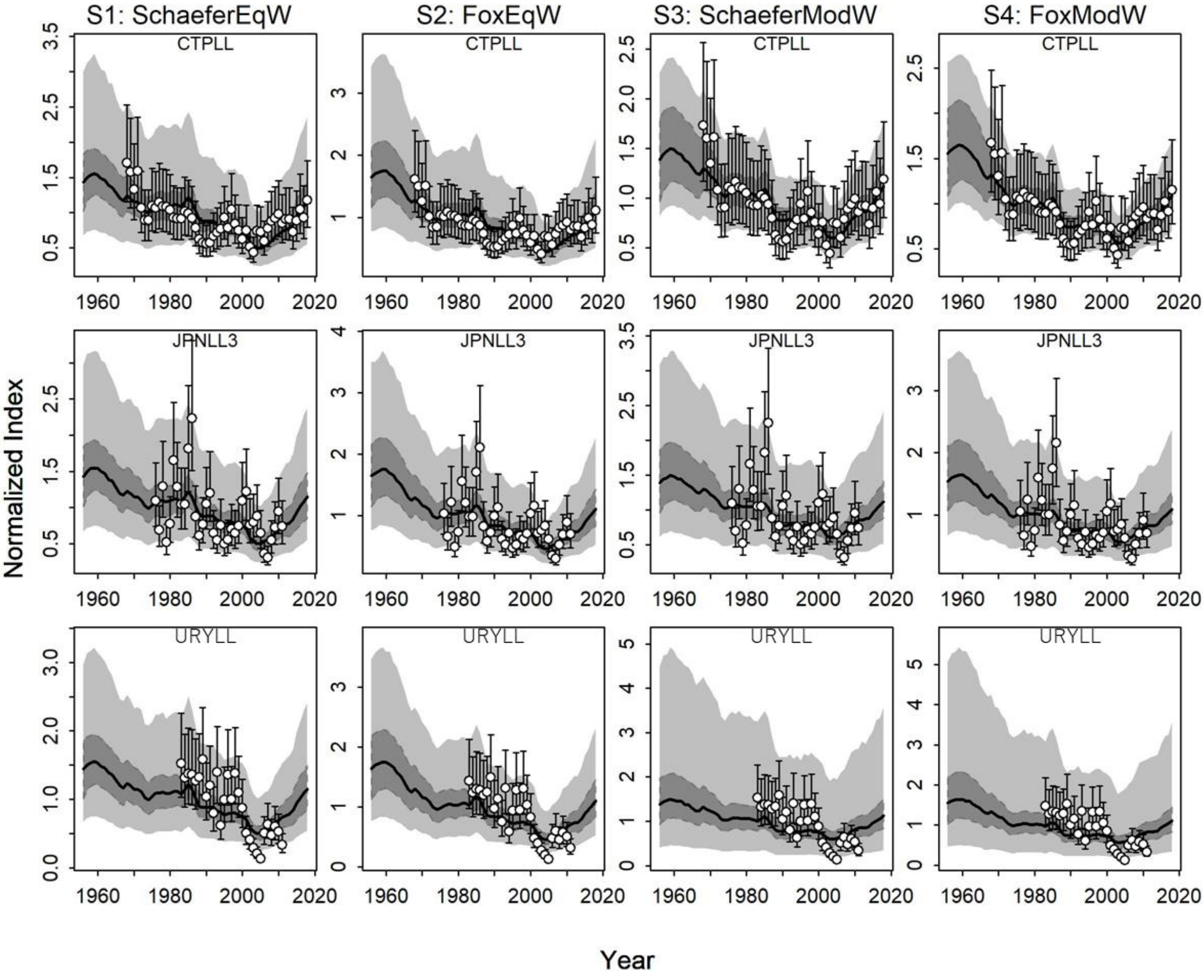


Figure 3. Time-series of observed (circle) with error 95% CIs (error bars) and predicted (solid line) CPUE of South Atlantic albacore for the Bayesian state-space surplus production model JABBA for each reference scenario (S1) SchaeferEqW; (S2) FoxEqW; (S3) SchaeferModW; (S4) FoxModW. Dark shaded grey areas show 95% credibility intervals of the expected mean CPUE and light shaded grey areas denote the 95% posterior predictive distribution intervals.

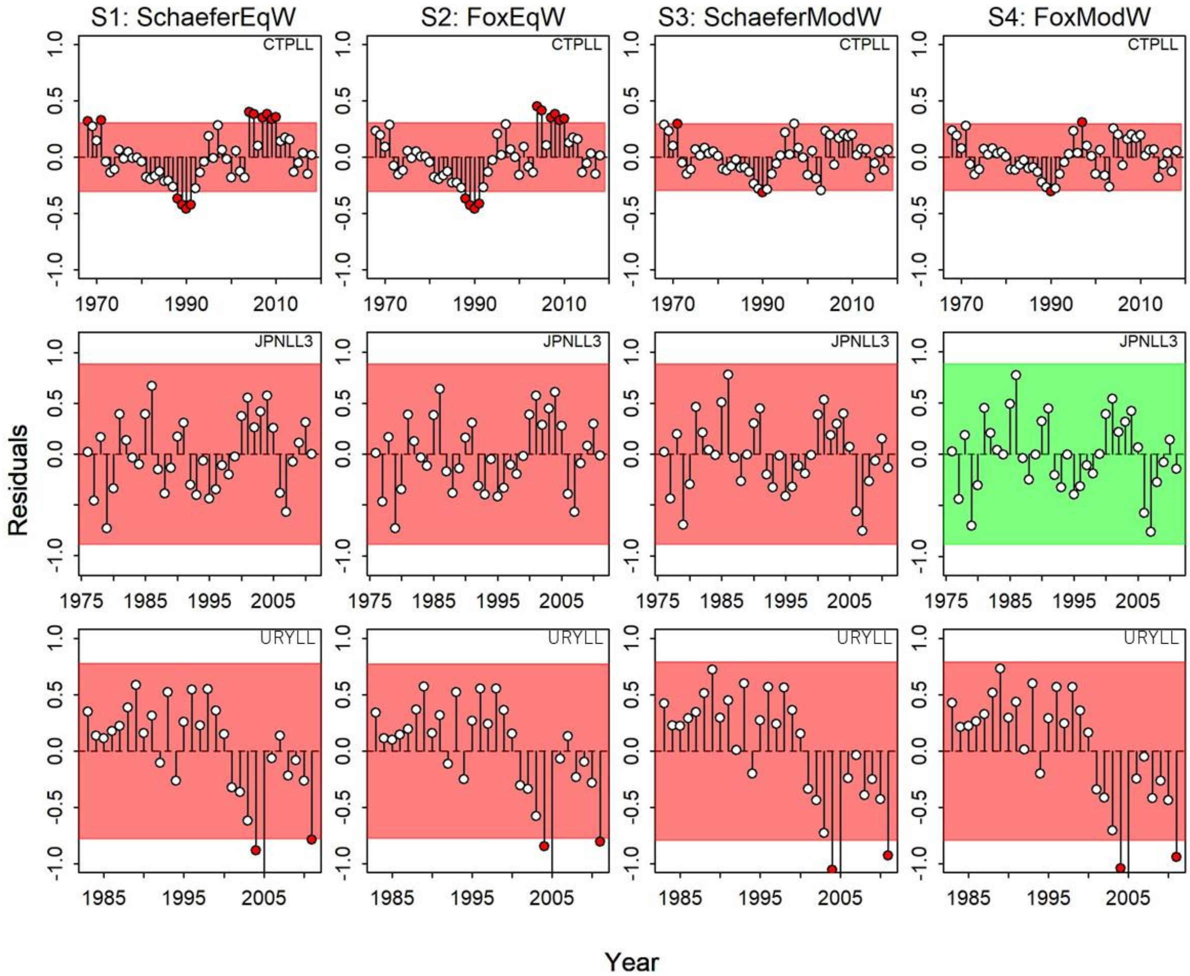


Figure 4. Runs tests to quantitatively evaluate the randomness of the time series of CPUE residuals by fleet for each reference scenario (S1) SchaeferEqW; (S2) FoxEqW; (S3) SchaeferModW; (S4) FoxModW. Green panels indicate no evidence of lack of randomness of time-series residuals (p>0.05) while red 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).

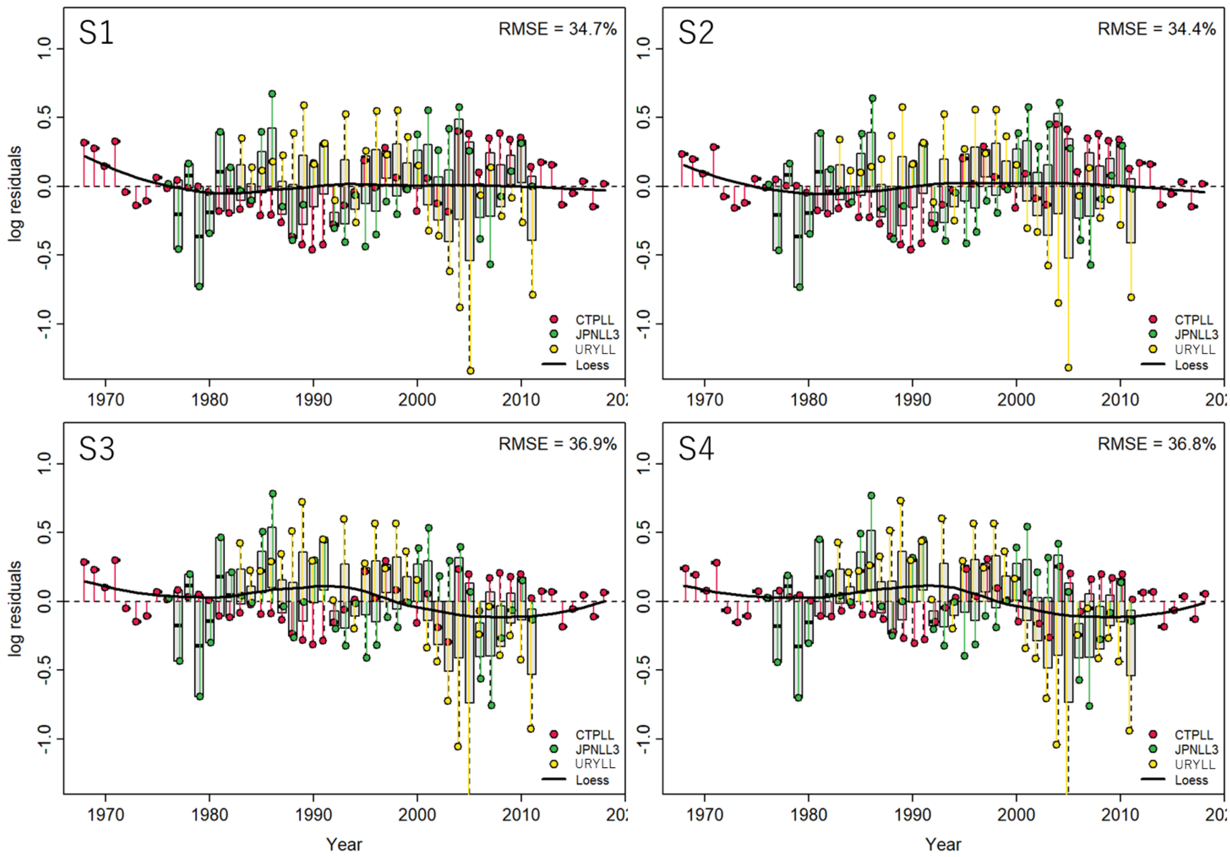


Figure 5. JABBA residual diagnostic plots for alternative sets of CPUE indices examined for each reference scenario (S1) SchaeferEqW; (S2) FoxEqW; (S3) SchaeferModW; (S4) FoxModW for the South Atlantic albacore. Boxplots indicate the median and quantiles of all residuals available for any given year, and solid black lines indicate a loess smoother through all residuals.

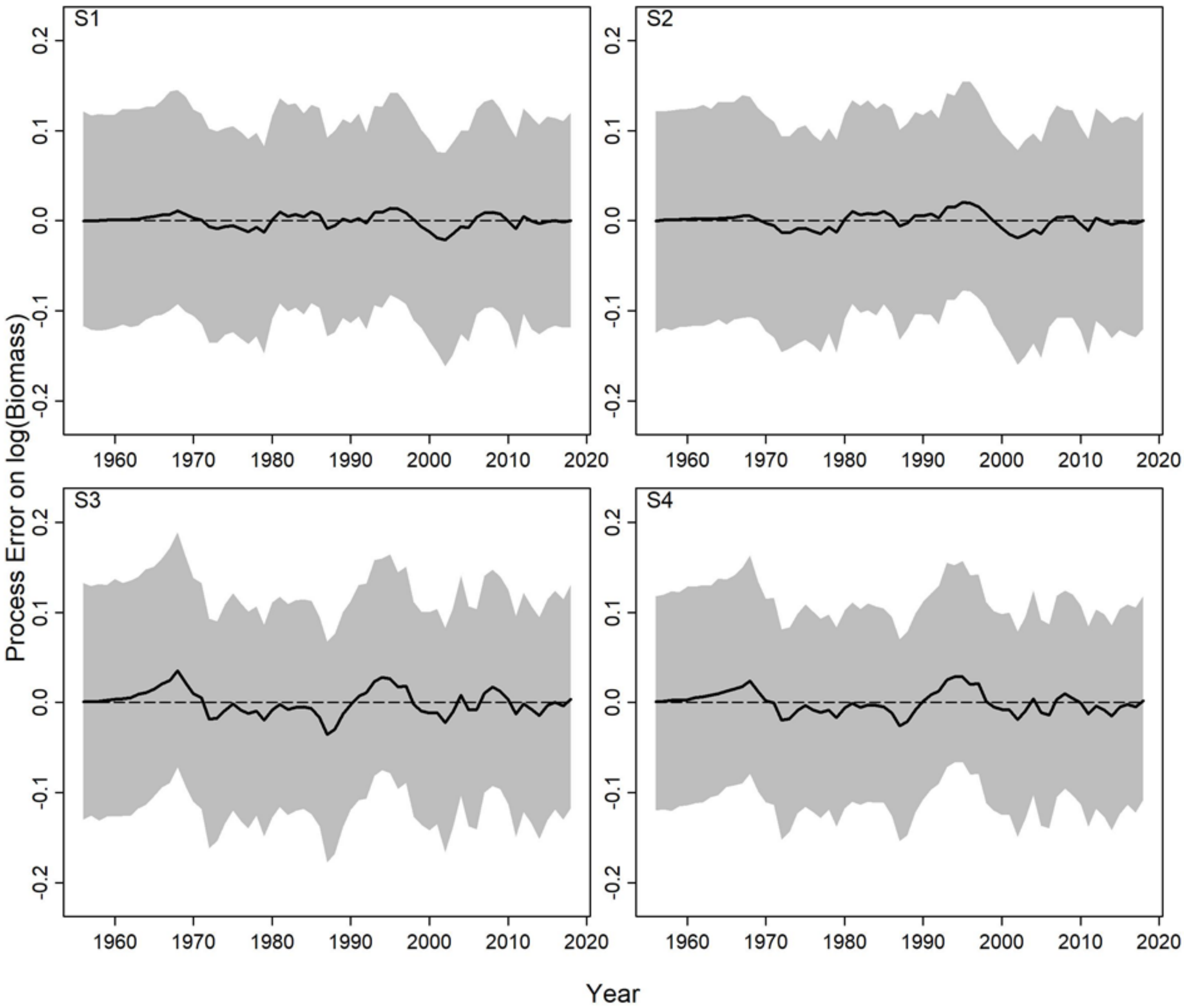
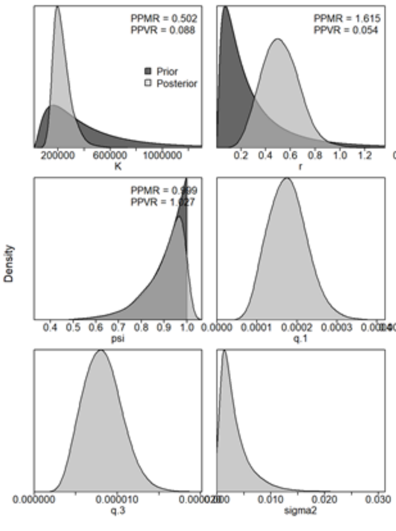
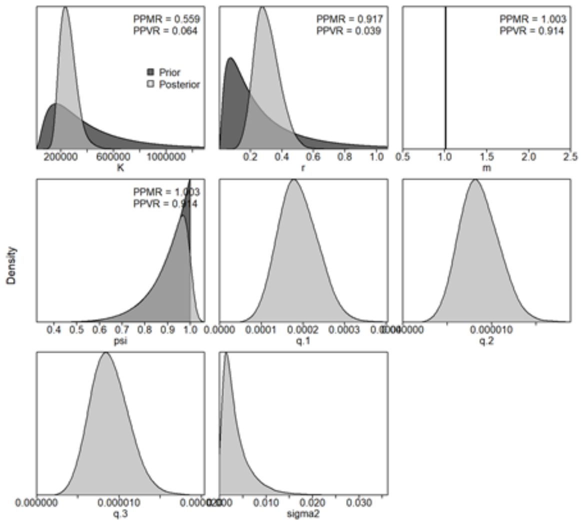


Figure 6. JABBA residual diagnostic plots for alternative sets of CPUE indices examined for each reference scenario (S1) SchaeferEqW; (S2) FoxEqW; (S3) SchaeferModW; (S4) FoxModW for the South Atlantic albacore. Process error deviates (median: solid line) with shaded grey area indicating 95% credibility intervals.

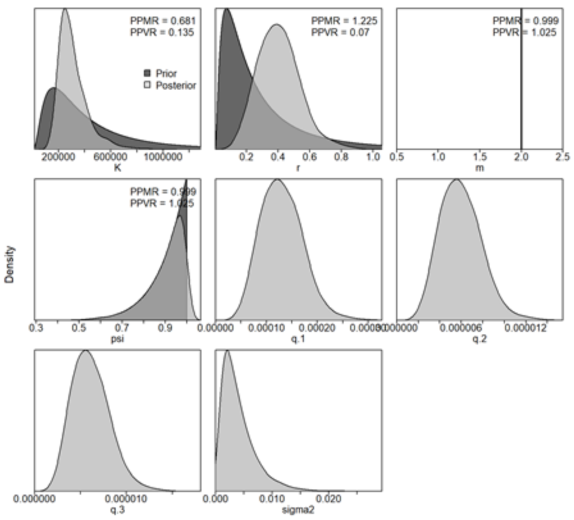
S1



S2



S3



S4

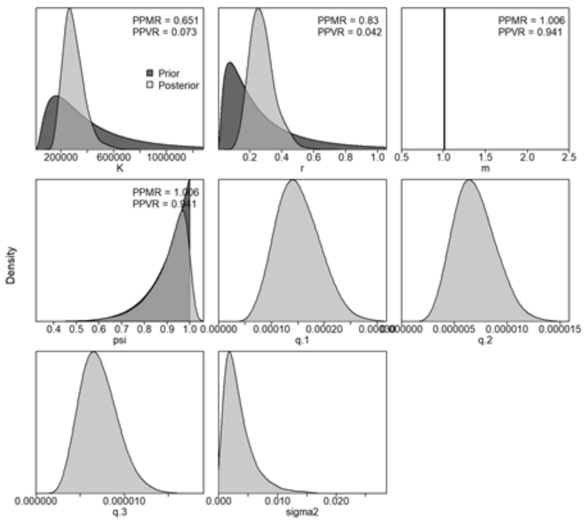


Figure 7. Prior and posterior distributions of various model and management parameters for the Bayesian state-space surplus production models (S1) SchaeferEqW; (S2) FoxEqW; (S3) SchaeferModW; (S4) FoxModW) for South Atlantic albacore. PPRM: Posterior to Prior Ratio of Medians; PPRV: Posterior to Prior Ratio of Variances.

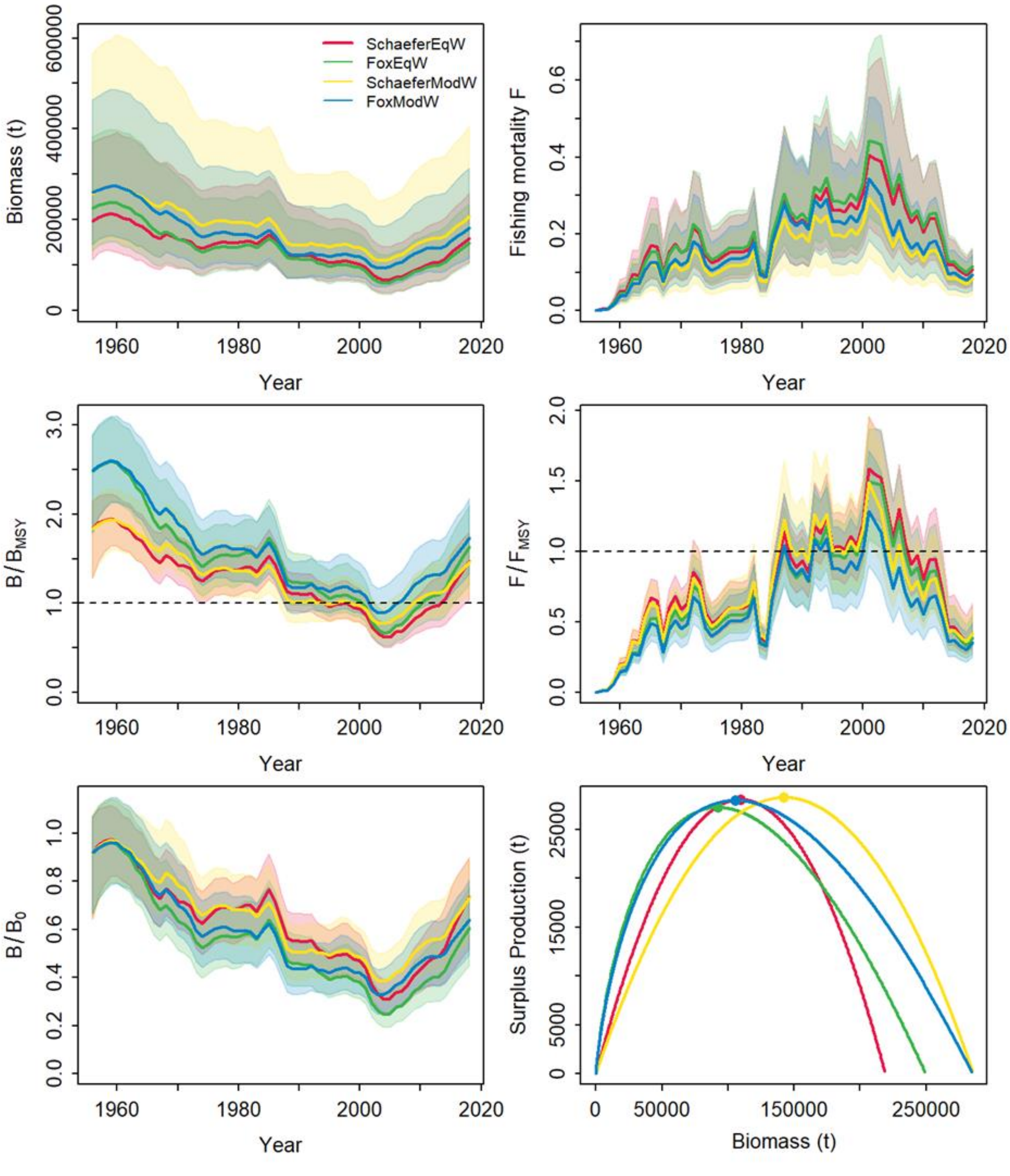


Figure 8. Trends in biomass and fishing mortality (upper panels), biomass relative to B_{MSY}(B/B_{MSY}) and fishing mortality relative to F_{MSY} (F/F_{MSY}) (middle panels) and biomass relative to K (B/K) and surplus production curve (bottom panels) for each reference scenario (S1: SchaeferEqW, S2: FoxEqW, S3: SchaeferModW, S4: FoxModW) from the Bayesian state-space surplus production JABBA model fits to South Atlantic albacore.

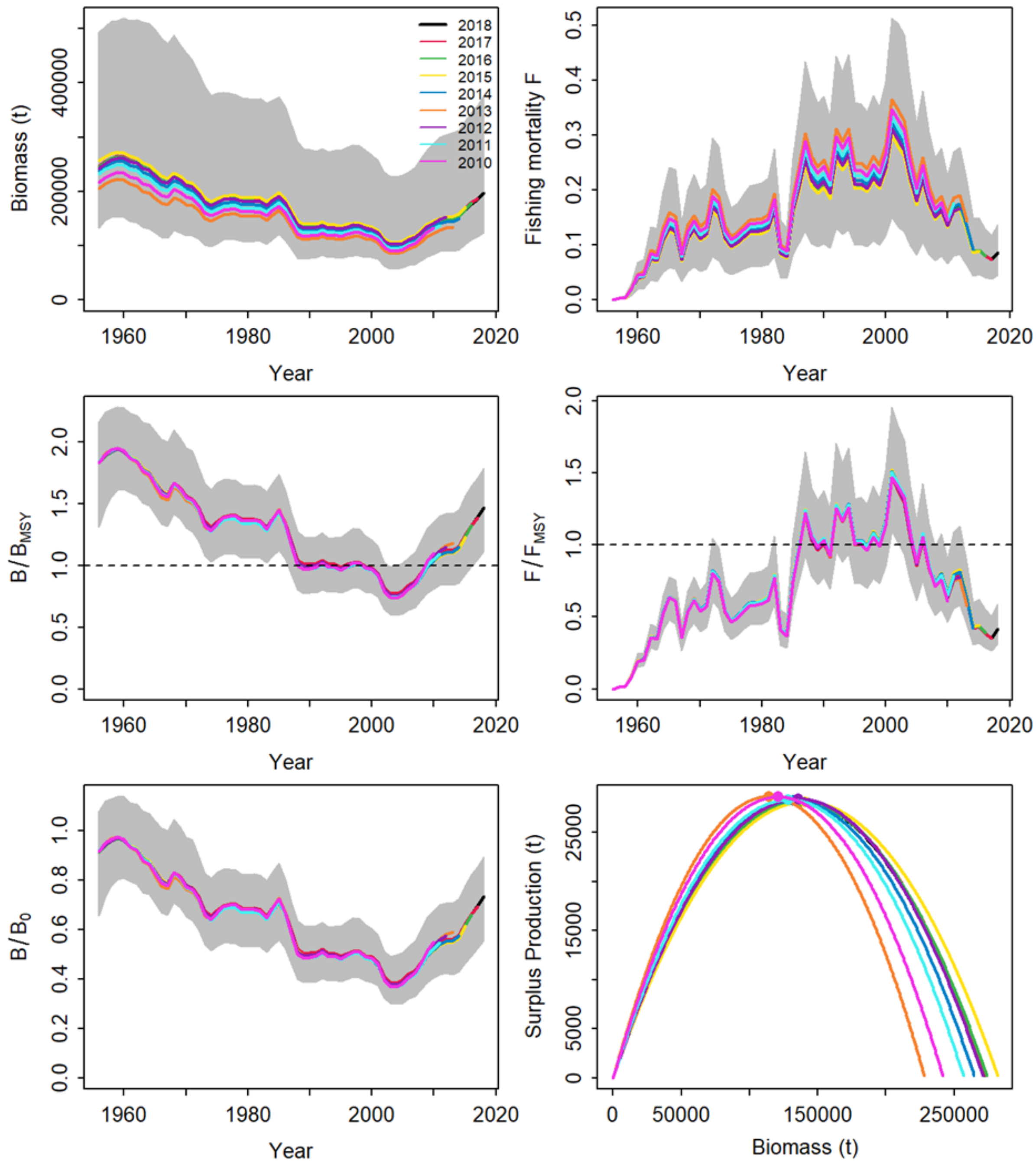


Figure 9. Retrospective analysis conducted for scenario (S3) SchaeferModW for South Atlantic albacore, by removing one year at a time sequentially (n=8) and predicting the trends in biomass and fishing mortality (upper panels), biomass relative to B_{MSY} (B/B_{MSY}) and fishing mortality relative to 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.

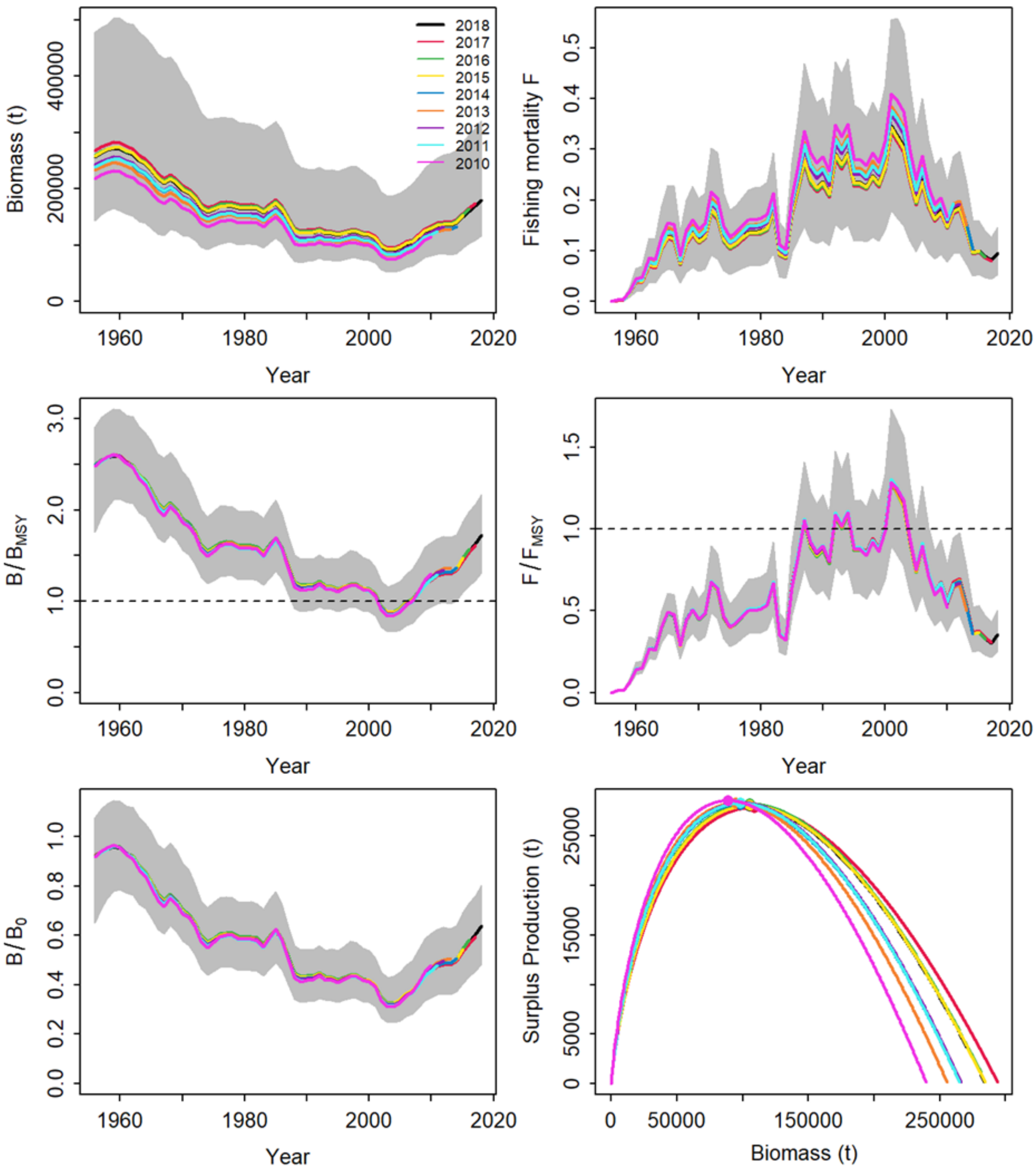


Figure 10. Retrospective analysis conducted for scenario (S4) FoxModW for South Atlantic albacore, by removing one year at a time sequentially (n=8) and predicting the trends in biomass and fishing mortality (upper panels), biomass relative to B_{MSY} (B/B_{MSY}) and fishing mortality relative to F_{MSY} (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.

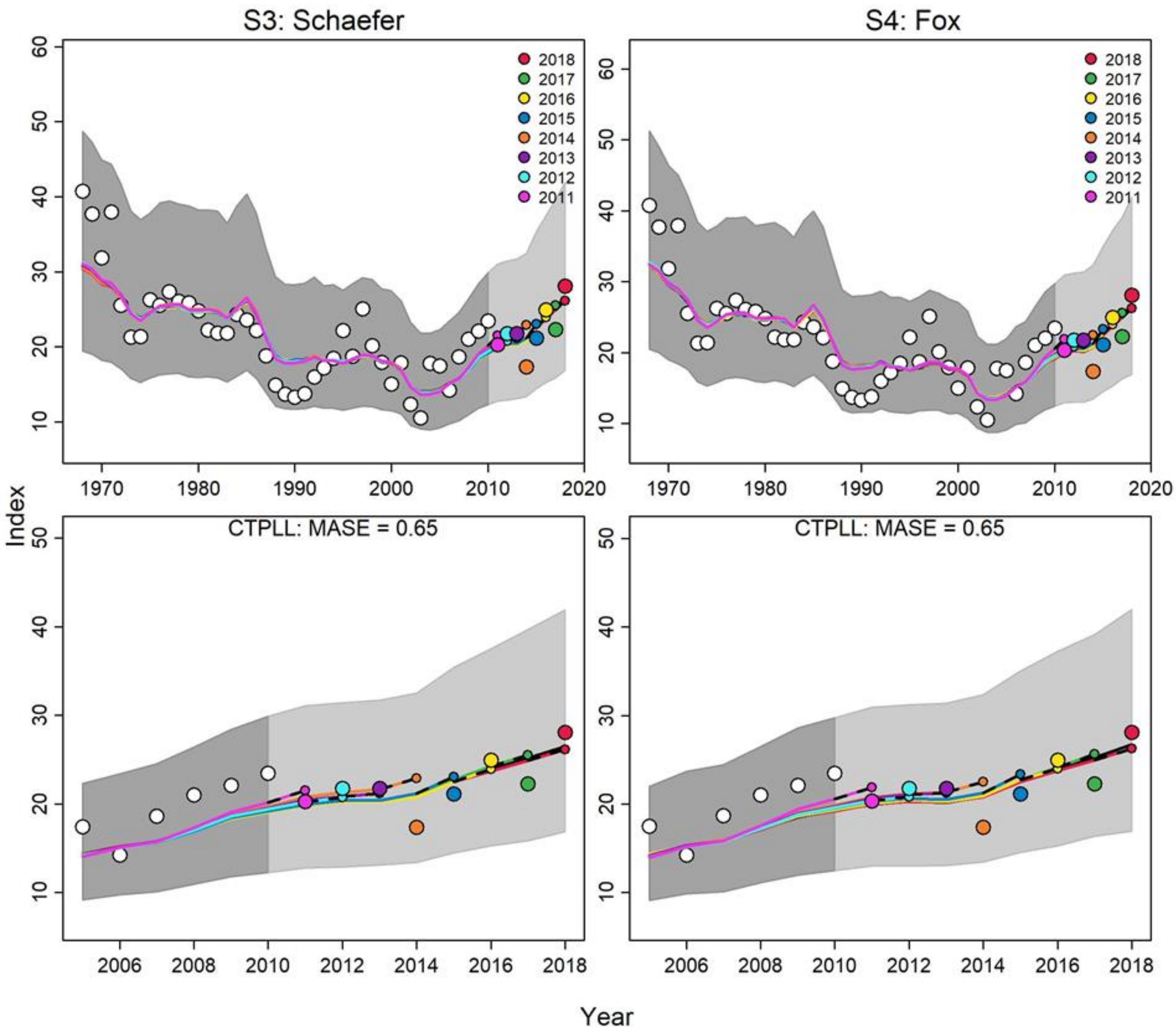


Figure 11. Hindcasting cross-validation results (HCxval) for the two scenarios (S3) SchaeferModW and (S4) FoxModW for South Atlantic albacore, showing one-year-ahead forecasts of CPUE values (2013-2018), performed with eight hindcast model runs relative to the expected CPUE. The CPUE observations, used 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).

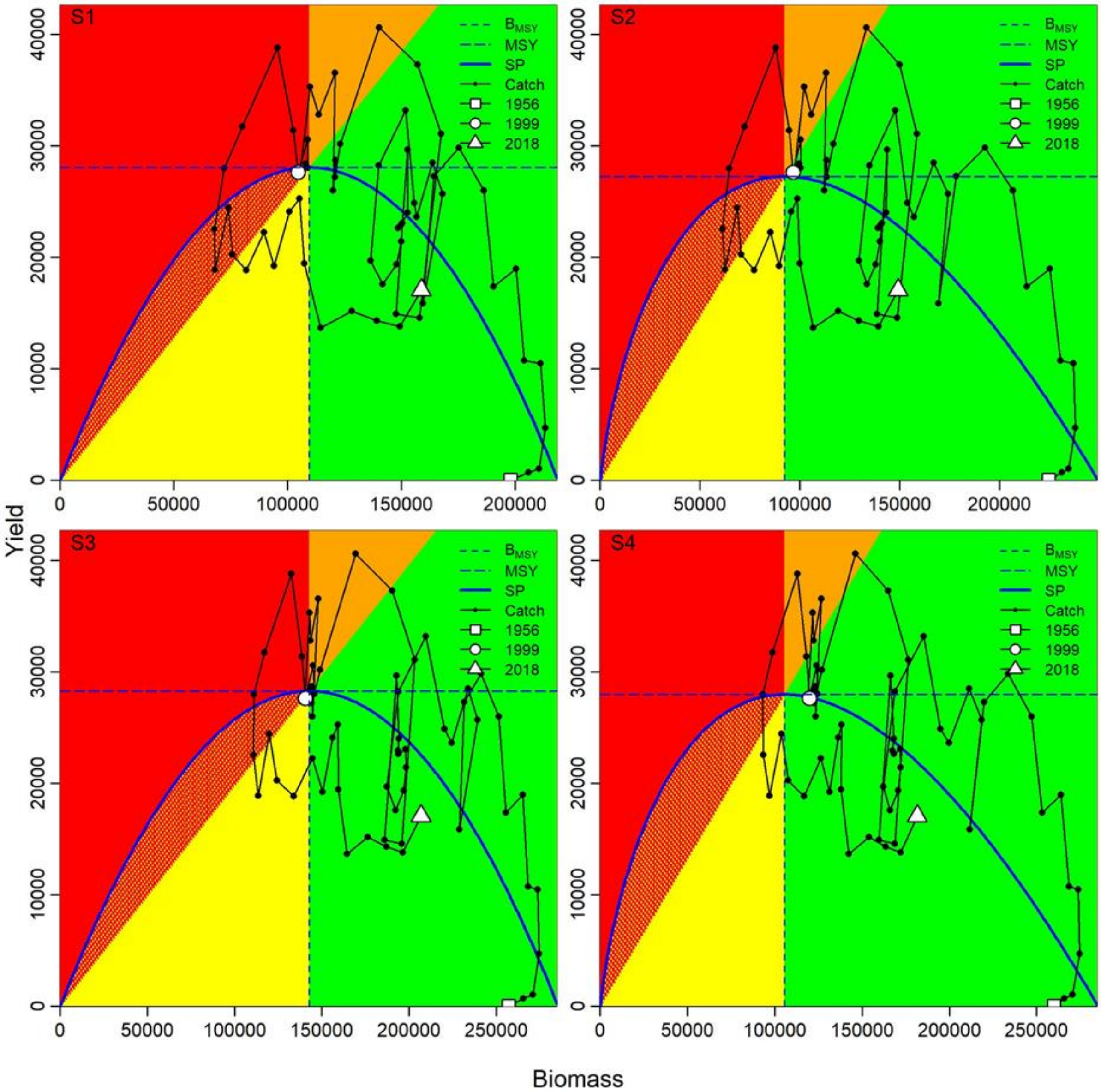


Figure 12. JABBA surplus production phase plot for the Bayesian state-space surplus production models (S1) SchaeferEqW; (S2) FoxEqW; (S3) SchaeferModW; (S4) FoxModW showing trajectories of the catches in relation to B_{MSY} and MSY for the South Atlantic albacore.

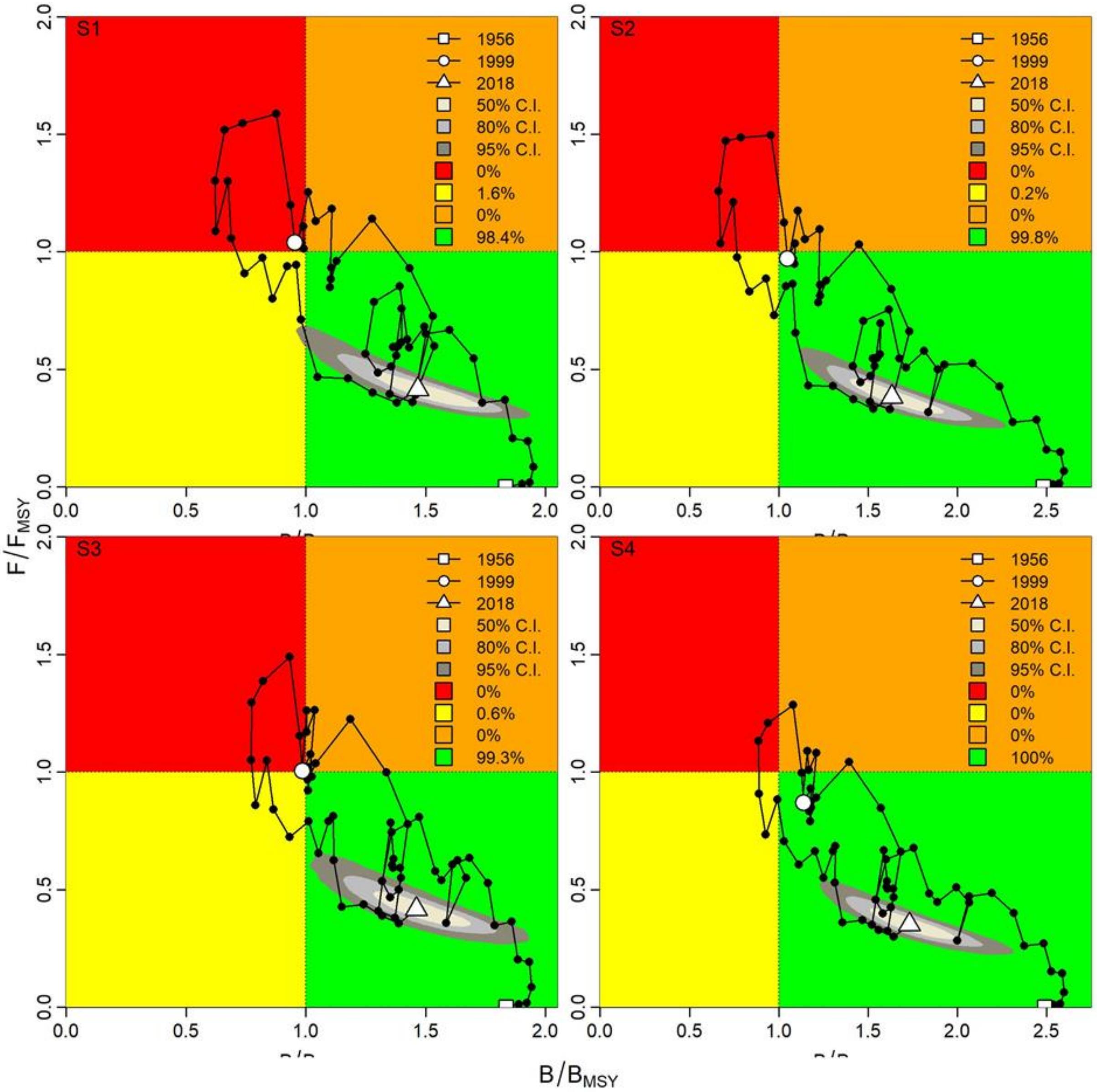


Figure 13. Kobe phase plot showing estimated trajectories (1956-2018) of B/B_{MSY} and F/F_{MSY} for the Bayesian state-space surplus production model for the South Atlantic albacore. Different grey shaded areas denote the 50%, 80%, and 95% credibility interval for the terminal assessment year. The probability of terminal year points falling within each quadrant is indicated in the figure legend.

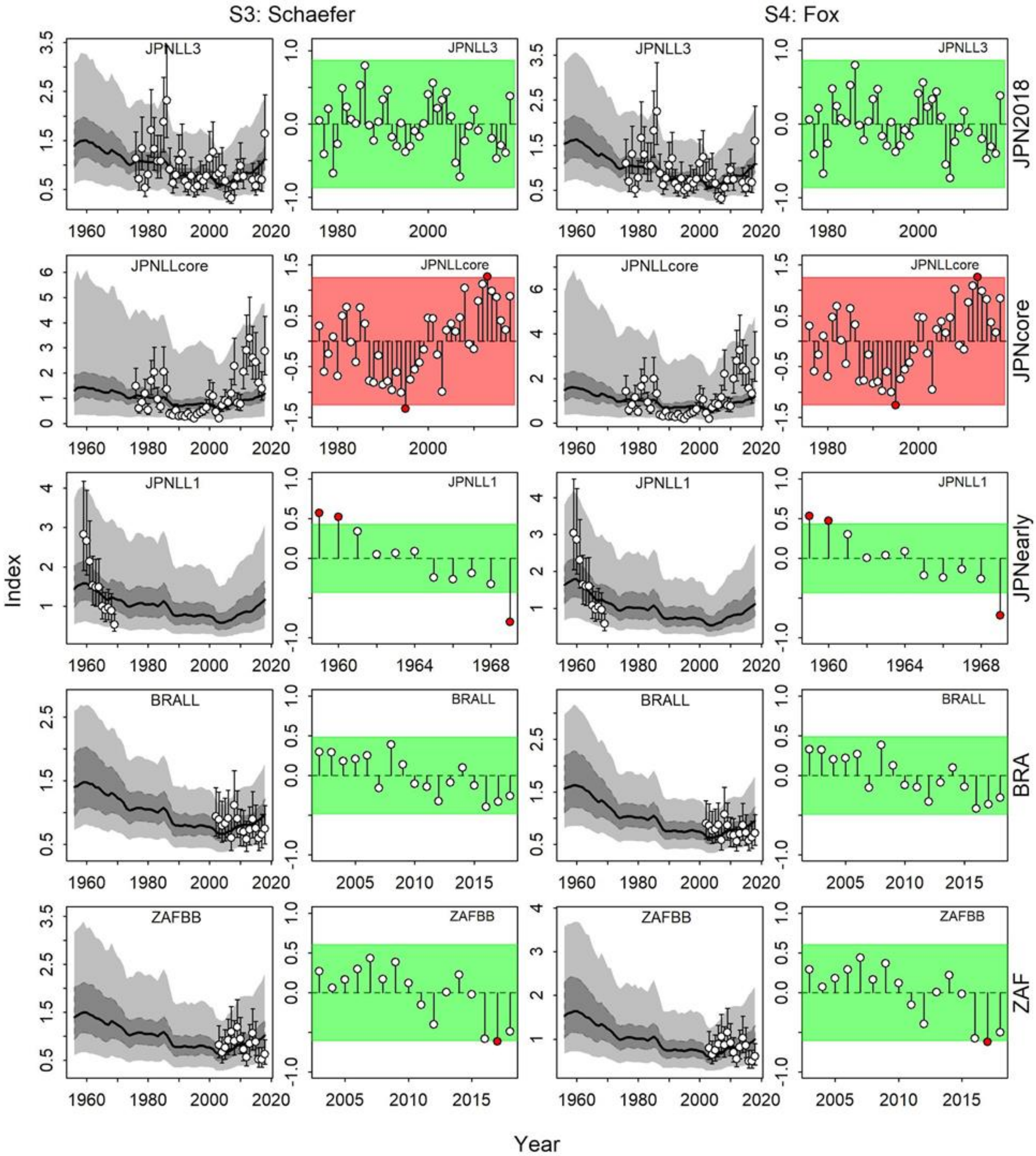


Figure 14 - Sensitivity analysis performed for scenarios (S3) SchaeferModW and (S4) FoxModW showing the time-series of observed (circle) with error 95% CIs (error bars) and predicted (solid line) CPUEs with the respective runs tests to quantitatively evaluate the randomness of the time series of CPUE residuals for South Atlantic albacore.

S3 S4

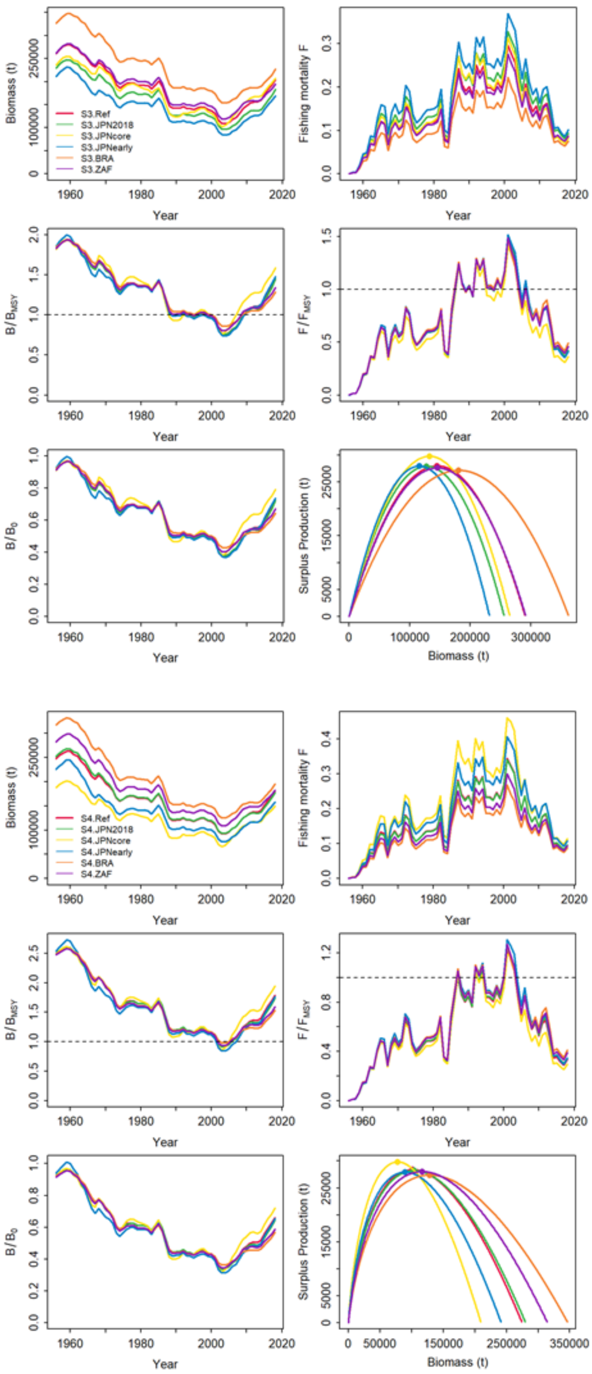


Figure 15. Sensitivity analysis performed for scenarios (S3; left panels) SchaeferModW and (S4; right panels) FoxModW showing the trends in biomass and fishing mortality (upper panels), biomass relative to B_{MSY} (B/B_{MSY}) and fishing mortality relative to F_{MSY} (F/F_{MSY}) (middle panels) and biomass relative to K (B/K) and surplus production curve (bottom panels) for the South Atlantic albacore.