SOUTH ATLANTIC BLUE SHARK STOCK: JUST ANOTHER BAYESIAN BIOMASS ASSESSMENT

R. Sant'Ana[1], B. Mourato.[2], L.G. Cardoso.[3], A. Kimoto.[4], and M. Ortiz[4]

SUMMARY

Bayesian State-Space Surplus Production Models were fitted to South Atlantic blue shark tuna catch and CPUE data using the 'JABBA' R package. The thirty-six distinct scenarios were based on a life history parameters, steepness and model weighting. All scenarios were based on a Pella-Tomlinson production function from an Age-Structured Equilibrium Model (ASEM). All scenarios showed similar trends for the trajectories of B/B_{MSY} and F/F_{MSY} over time. In general, B/B_{MSY} showed a decreasing pattern in the first half of the time series followed by a slight increase after 1998. The F/F_{MSY} showed a general pattern with a sharp increasing trend during 1990s, followed by stable trend. Kobe stock status plots had shown median quantities estimated for the last data year in the green quadrant. However, the scenarios based on a more conservative values of steepness (0.5) were more pessimistic than others.

RÉSUMÉ

Des modèles de production excédentaire état-espace de type bayésien ont été ajustés aux données de capture et de CPUE du requin peau bleue de l'Atlantique Sud au moyen du progiciel JABBA R. Les trente-six scénarios étaient basés sur les paramètres du cycle vital, la pente et la pondération du modèle. Tous les scénarios étaient basés sur une fonction de production Pella-Tomlinson d'un modèle d'équilibre structuré par âge (ASEM). Tous les scénarios affichaient des tendances similaires pour les trajectoires de B/B_{PME} et F/F_{PME} au fil du temps. En général, B/B_{PME} présentait une tendance à la baisse dans la première moitié de la série temporelle, suivie d'une légère augmentation après 1998. F/F_{PME} présentait un schéma général de forte tendance à la hausse au cours des années 1990, suivie d'une tendance stable. Les diagrammes de l'état du stock de Kobe montraient les quantités médianes estimées pour la dernière année de données dans le quadrant vert. Toutefois, les scénarios fondés sur des valeurs plus conservatrices de la pente (0,5) étaient plus pessimistes que les autres

RESUMEN

Se ajustaron modelos bayesianos de producción de excedentes estado-espacio a los datos de captura y CPUE de tiburón azul del Atlántico sur utilizando el paquete R "JABBA". Los treinta y seis escenarios distintos se basaron en los parámetros del ciclo vital, la inclinación y la ponderación del modelo. Todos los escenarios se basaron en una función de producción de Pella-Tomlinson a partir de un modelo en equilibrio estructurado por edad (ASEM). Todos los escenarios mostraron tendencias similares para las trayectorias de B/B_{RMS} y F/F_{RMS} a lo largo del tiempo. En general, la B/B_{RMS} mostró un patrón decreciente en la primera mitad de la serie temporal, seguido de un ligero aumento después de 1998. La F/F_{RMS} mostró un patrón general con una marcada tendencia al alza durante la década de 1990, seguida de una tendencia estable. Los gráficos de Kobe sobre el estado del stock mostraban las cantidades medias estimadas para el último año de datos en el cuadrante verde. Sin embargo, los escenarios basados en valores más conservadores de inclinación (0,5) fueron más pesimistas que los demás

KEYWORDS

Blue shark, stock status, CPUE fits, hindcast, life history priors

[1] Universidade do Vale do Itajaí, Escola Politécnica, 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] Universidade Federal de Rio Grande, Instituto de Oceanografia, Laboratório de Recursos Pesqueiros Demersais, Rio Grande, Brazil.

[4] ICCAT Secretariat. Calle Corazón de Maria 8, Madrid, Spain 28002.

1. Introduction

Blue shark is an oceanic-epipelagic species that can be found close to the coast in some areas and at certain times, particularly where the shelf is narrow or even in ports and marinas. It is found in deep waters of tropical, warm and temperate seas from the surface to a depth of at least 1,291.1 m, with greater abundancy in areas outside the platform (ICCAT, 2022). It is an oceanic and epipelagic species distributed in all oceans, in tropical, subtropical and temperate waters between 62ºN and 54ºS (ICCAT, 2022). As a function of its wide distribution, blue shark has been intensively exploited by various fisheries around the world. For management purposes, the International Commission for the Conservation of Atlantic Tunas (ICCAT) considers three distinct stocks in Atlantic Ocean, the North, South and Mediterranean stock (ICCAT, 2022). Aimed in the South Atlantic blue shark stock, nine longline fisheries fleets (EU-Spain, EUPortugal, Chinese Taipei, Namibia, Brazil, Japan, Uruguay, China (P.R.) and South Africa) made 99% of the total landings between 1990 and 2014 (ICCAT, 2022).

The last South Atlantic blue shark stock assessment was carried out in 2015 (ICCAT, 2015) and included outputs from distinct production models frameworks. The synthesis of the assessment indicated that the South Atlantic blue shark stock was not overfished (B_{2013}/B_{MSY}=1.50 to 1.96) and not experiencing overfishing (F_{2013}/F_{MSY}= 0.04 to 0.50) (ICCAT, 2015). On the other hand, estimates obtained with the state-space BSP were generally less optimistic, especially when process error was not included, predicting that the stock could be overfished (B_{2013}/B_{MSY}= 0.78 to 1.29) and that overfishing could be occurring (F_{2013}/F_{MSY} = 0.54 to 1.19) (ICCAT, 2015).

Here, we present the 2023 preliminary stock assessment results for South Atlantic blue shark 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 South Atlantic blue shark stock is guided by the SCRS work plan. Some insights for an uncertainty grid scenario was built based on the discussions and recommendations that raised during the 2023 Blue shark Data Preparatory 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 South Atlantic blue shark from 1971 to 2022 (**Figure 1**). Relative abundance indices were made available. These indices cover various periods and represent the main longline fleets operating in the South Atlantic Ocean (*e.g. Spain, Japan, Chinese-Taipei, Brazil and Uruguay longline fleets*). The Brazil and Uruguay index was made available in a form of joint index. A summary of the available indices is described below:

- Spain LL index – 1997 to 2021;
- Japan LL index – 1994 to 2021;
- Chinese-Taipei LL index – 2007 to 2021;
- Joint LL index (Brazil and Uruguay) – 1992 to 2021.

The CV's for each index was treated as a source of uncertainty to model weighting process. Thus, the scenarios tested were (a) using the original CV's from the standardizations; (b) re-estimate the CV's based on the Courtney *et al*. (2016), and; (c) fixed in 20% and gave the opportunity to model-based weighting internally.

2.3 Model specifications

The model specifications were based on three main sources of uncertainties. The first one was based on the life-history parameters presented during the Blue shark data preparatory meeting and included by the results from the preliminary SS3 runs; Second source was based on the impact of the steepness on the productivity parameters (r and B_{MSY}/K). For this source was used the confidence interval and central tendency proposed by Cortes *et al.* (2023) (h = 0.5, h = 0.8 and h = 0.9), and; The last source of uncertainty, as described before, it was tested the model weigthing process (**Table 1**).

The priors of *K* was kept similar to those used in the last assessment of the species. For this parameter, it was used vaguely informative lognormal prior with a large CV of 30% 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). For *r*, were developed priors 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 here were based on the proposals made by Cortes *et al*. (2023) and those proposed by Cardoso *et al.* (2023) as inputs for the ASEM models included the uncertainty grid configuration citet before and presented in **Table 1**. The stock parameters used as inputs for the ASEM models included the following configuration: (a) Maximum age equal to 16 and 22 years with the corresponding natural mortality values, and; (b) steepness values equal to 0.5, 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 B_{MSY}/B_0 (**Table 1**; see details in Winker *et al*., 2019). **Table 1** provides a summary of all scenarios initially tested.

For all scenarios, the same initial depletion prior ($\varphi = B_{1950}/K$) was defined by a beta distribution with mean = 0.9 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 0.001 and rate parameter equal to 0.001.

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 first scenario (ASEM_SS3_F_h_0.5) of each model weighting uncertainty source of the grid, 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.

Additionally, the analysis included sensitivity model runs based on forward stepwise inclusion of each index one-by-one in the model. Taking as prior the Spain LL index in the small model. The general idea with this comparative analysis was to evaluate the possible effects of the inclusion of each index over estimated biomass dynamic of this stock. Finally, an additional nine scenarios were implemented based on the exploration of time-blocks in relative abundance indices as a form to evaluate possible influences of recent targeting to this species and stock.

3 Results and Discussion

In the sections Tables and Figures are presented the results of the first runs that will be presented and updated during the Blue shark stock assessment meeting.

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).

Figure 3 displays the fitting of the model to each of the four standardized CPUE indices for the thirty-six scenarios. In all scenarios, the models capture the overall trends in the relative abundance indices, except for a few outlier or influential points noted in the Chinese-Taipei, Brazil-Uruguay, and Japan longline indices. By applying time-blocks, the model's performance improves, leading to a better capture of the input data's tendencies. But still remaining the same outliers issues observed before.

Figure 4 displays the log-residuals runs tests results for each CPUE and scenario. Green panels indicate CPUE indices that passed the runs test with no evidence of a non-random residual pattern ($p > 0.05$), while red panels indicate a failed runs test. The inner shaded area shows 3-sigma limits around the overall mean as suggested by Anhøj and Olesen (2014), and red circles identify each specific year where the residuals are larger than the threshold limit. The test revealed that the Japanese standardized CPUE had non-random residuals, without time-block applied to the relative abundance indices. For those scenarios with an internal model weighting, the Spain index also showed non-random residual patterns. Moreover, for scenarios based on time-block and internal model-weighting, only the Spain index exhibited non-random residual issues. The goodness-of-fit were comparable among all scenarios, in general, the RMSE statistics were consistent ranging from 23.1% to 26.6% (**Figure 5**). In general, 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. Despite not being statistically significant, there seemed to be a slight upward trend towards the end of the series. This trend was less noticeable in the scenarios that were adjusted with time-block and internal model-weighting.

Overall, the various scenarios produced similar trends for the trajectories of B/B_{MSY} and F/F_{MSY} over time (**Figure 7**; **Figure 8**). However, scenarios with higher values of steepness tended to show more positive outcomes than those with lower steepness values. The B/B_{MSY} trajectory showed a decrease between 1985 and 1995 in all scenarios tested, but after that period, the trends varied depending on the model-weighting decision implemented. Scenarios based on the original CV values or Francis correction proposal showed an increase between 1996 and 2011, followed by a small decline at the end of the series (**Figures 7A** and **7B**). Scenarios fitted with model-internal weighting only showed a stable trend between 1996 and 2005, followed by further increase, stabilization, and a smooth decline at the end of the time series (**Figure 7C**). The model-internal weighting with time-block structure showed a more stable trend after 1996. The F/F_{MSY} trajectory showed a sharp increase at the same time that the B/B_{MSY} trajectory decreased (**Figure 8**). For all scenarios evaluated here, the models do not evidenced periods of overfishing ($F/F_{MSY} > 1$) or even the stock are being overfished ($B/B_{MSY} < 1$) (**Figure 7**; **Figure 8**).

The results of an eight year retrospective analysis applied to the first scenario (ASEM_SS3_F_h_0.5) for each of the four model weighting process tested were depicted in **Figure 9**, respectively. In general, a negligible retrospective pattern were observed. 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 2**). The hindcasting cross-validation results for all updated indices show predictions within limits of the 95% CRI's suggesting a good prediction skills for the scenario tested within distinct model weighting structure (**Figure 10**). However, the mean absolute scaled error (MASE) estimated were above of the reference level (MASE > 1) for Japan index when fitted with original weighting and Francis correction (MASE > 2) and for Chinese-Taipei index when fitted with model-internal weighting with time-block approach (MASE > 3). And slightly above (1 < MASE < 2) for Japan index when fitted with model-internal weighting and Japan and Spain indices when they fitted using model-internal weighting with time-block.

The results of the sensitivity analysis based on forward stepwise indices in scenarios S01, S10 and S19 are shown in **Figure 11**. These results shown a similar trend when the stepwise process were implemented for S01 and S10. For S19 scenario, the stepwise process shown some changes in the middle term period. Although, the begging and end of the time-series do not shown high discrepancy based on the interactive inclusion of the relative abundance indices tested.

The Kobe biplots for all scenarios were shown in **Figure 12**. All scenarios show optimistic status with probabilities of the stock are being stable on green area (**Figure 12**). However, these results are preliminary and will need to be further explored during the blue shark stock assessment meeting.

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

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

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.

ICCAT. 2022. ICCAT Manual – Species: Chapter 2.2.1.1 – Blue shark. Available on-line at https://www.iccat.int/Documents/SCRS/Manual/CH2/2_2_1_1_BSH_ENG.pdf.

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 South Atlantic blue shark.

Scenario	Type	Model Weighting	r	BmsyK
S01	ASEM_SS3_F_h_0.5	Original CV	lognormal(0,226, 0,324)	0,35
S02	ASEM_SS3_F_h_0.8	Original CV	lognormal(0,278, 0,352)	0,23
S03	ASEM_SS3_F_h_0.9	Original CV	lognormal(0,282, 0,374)	0,18
S04	ASEM_SS3_M_h_0.5	Original CV	lognormal(0,293, 0,34)	0,34
S05	ASEM_SS3_M_h_0.8	Original CV	lognormal(0,366, 0,378)	0,22
S06	ASEM_SS3_M_h_0.9	Original CV	lognormal(0,377, 0,415)	0,16
S07	ASEM_CT_Both_h_0.5	Original CV	lognormal(0,122, 0,288)	0,37
S08	ASEM_CT_Both_h_0.8	Original CV	lognormal(0,148, 0,295)	0,26
S09	ASEM_CT_Both_h_0.9	Original CV	lognormal(0,149, 0,297)	0,22
S10	ASEM_SS3_F_h_0.5	Courtney *et al* (2016)	lognormal(0,226, 0,324)	0,35
S11	ASEM_SS3_F_h_0.8	Courtney *et al* (2016)	lognormal(0,278, 0,352)	0,23
S12	ASEM_SS3_F_h_0.9	Courtney *et al* (2016)	lognormal(0,282, 0,374)	0,18
S13	ASEM_SS3_M_h_0.5	Courtney *et al* (2016)	lognormal(0,293, 0,34)	0,34
S14	ASEM_SS3_M_h_0.8	Courtney *et al* (2016)	lognormal(0,366, 0,378)	0,22
S15	ASEM_SS3_M_h_0.9	Courtney *et al* (2016)	lognormal(0,377, 0,415)	0,16
S16	ASEM_CT_Both_h_0.5	Courtney *et al* (2016)	lognormal(0,122, 0,288)	0,37
S17	ASEM_CT_Both_h_0.8	Courtney *et al* (2016)	lognormal(0,148, 0,295)	0,26
S18	ASEM_CT_Both_h_0.9	Courtney *et al* (2016)	lognormal(0,149, 0,297)	0,22
S19	ASEM_SS3_F_h_0.5	Internal Model Weight	lognormal(0,226, 0,324)	0,35
S20	ASEM_SS3_F_h_0.8	Internal Model Weight	lognormal(0,278, 0,352)	0,23
S21	ASEM_SS3_F_h_0.9	Internal Model Weight	lognormal(0,282, 0,374)	0,18
S22	ASEM_SS3_M_h_0.5	Internal Model Weight	lognormal(0,293, 0,34)	0,34
S23	ASEM_SS3_M_h_0.8	Internal Model Weight	lognormal(0,366, 0,378)	0,22
S24	ASEM_SS3_M_h_0.9	Internal Model Weight	lognormal(0,377, 0,415)	0,16
S25	ASEM_CT_Both_h_0.5	Internal Model Weight	lognormal(0,122, 0,288)	0,37
S26	ASEM_CT_Both_h_0.8	Internal Model Weight	lognormal(0,148, 0,295)	0,26
S27	ASEM_CT_Both_h_0.9	Internal Model Weight	lognormal(0,149, 0,297)	0,22
S28	ASEM_SS3_F_h_0.5	Internal Model Weight – Time block	lognormal(0,226, 0,324)	0,35
S29	ASEM_SS3_F_h_0.8	Internal Model Weight – Time block	lognormal(0,278, 0,352)	0,23
S30	ASEM_SS3_F_h_0.9	Internal Model Weight – Time block	lognormal(0,282, 0,374)	0,18
S31	ASEM_SS3_M_h_0.5	Internal Model Weight – Time block	lognormal(0,293, 0,34)	0,34
S32	ASEM_SS3_M_h_0.8	Internal Model Weight – Time block	lognormal(0,366, 0,378)	0,22
S33	ASEM_SS3_M_h_0.9	Internal Model Weight – Time block	lognormal(0,377, 0,415)	0,16
S34	ASEM_CT_Both_h_0.5	Internal Model Weight – Time block	lognormal(0,122, 0,288)	0,37
S35	ASEM_CT_Both_h_0.8	Internal Model Weight – Time block	lognormal(0,148, 0,295)	0,26
S36	ASEM_CT_Both_h_0.9	Internal Model Weight – Time block	lognormal(0,149, 0,297)	0,22

Table 2. Summary Mohn's rho statistic computed for a retrospective evaluation period of eight years for the firs four scenarios of each model weighting fitted to the South Atlantic blue shark stock assessment 2023. 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}***	***Proc(B)***	***MSY***
S01	-0.771	0.094	-0.006	0.045	-0.004	-0.027
S10	-0.077	0.100	-0.002	0.028	-0.006	-0.012
S19	-0.045	0.050	0.0002	0.014	-0.009	-0.013
S20	-0.118	0.140	-0.081	0.152	-0.010	-0.048

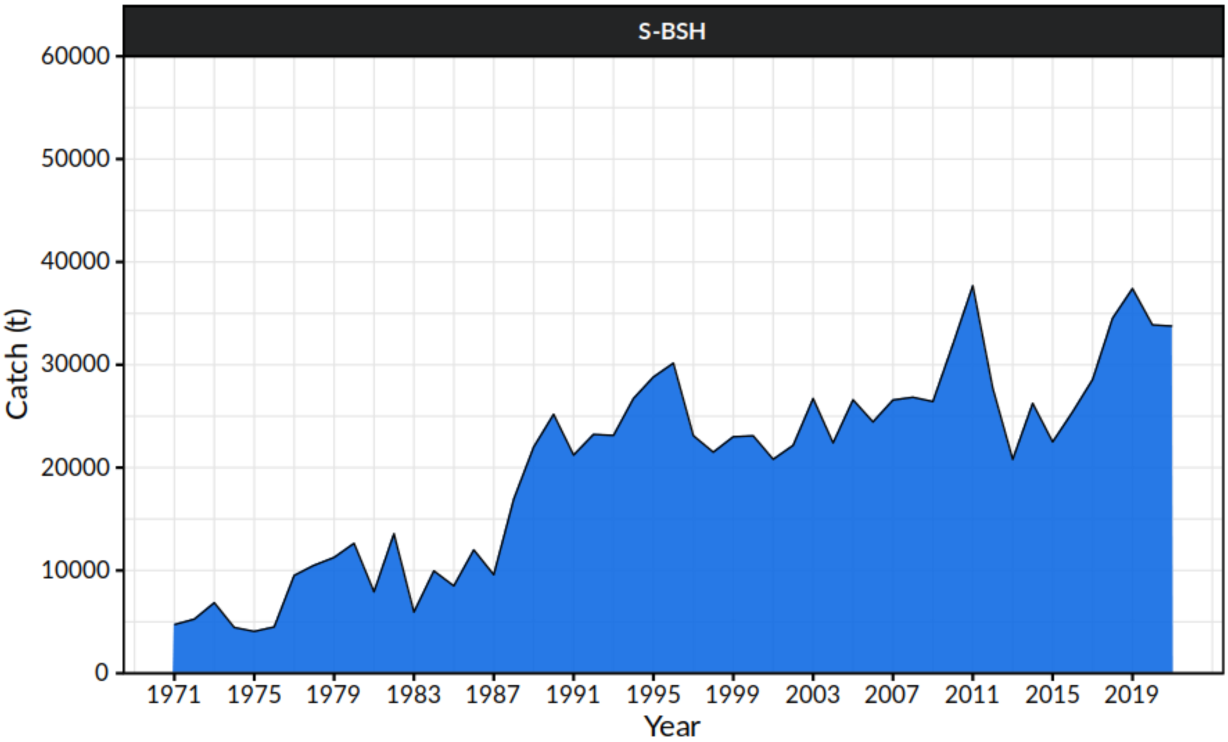


Figure 1. Catch time series in metric tons (t) between 1971 and 2022 for South Atlantic blue shark.

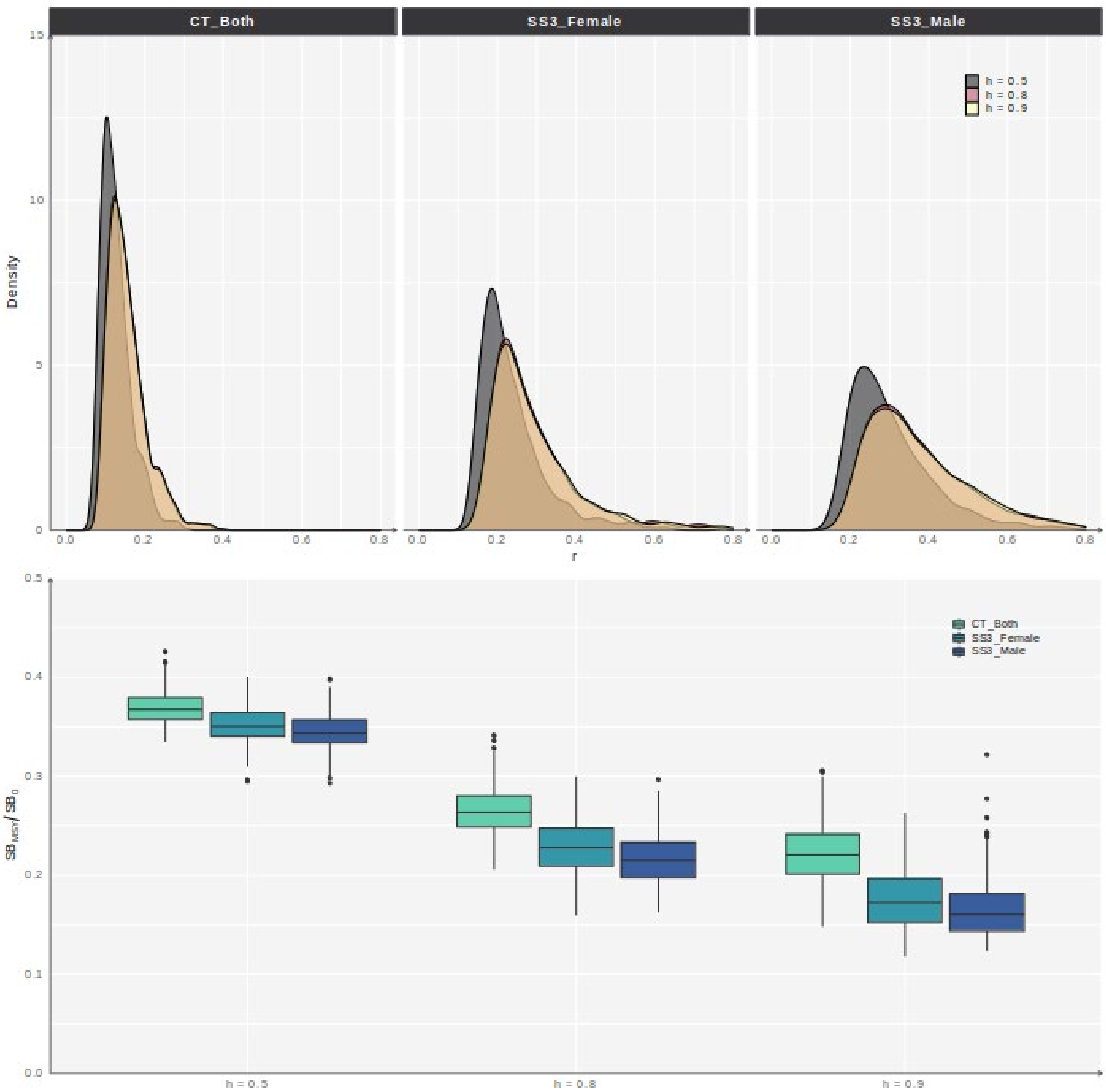
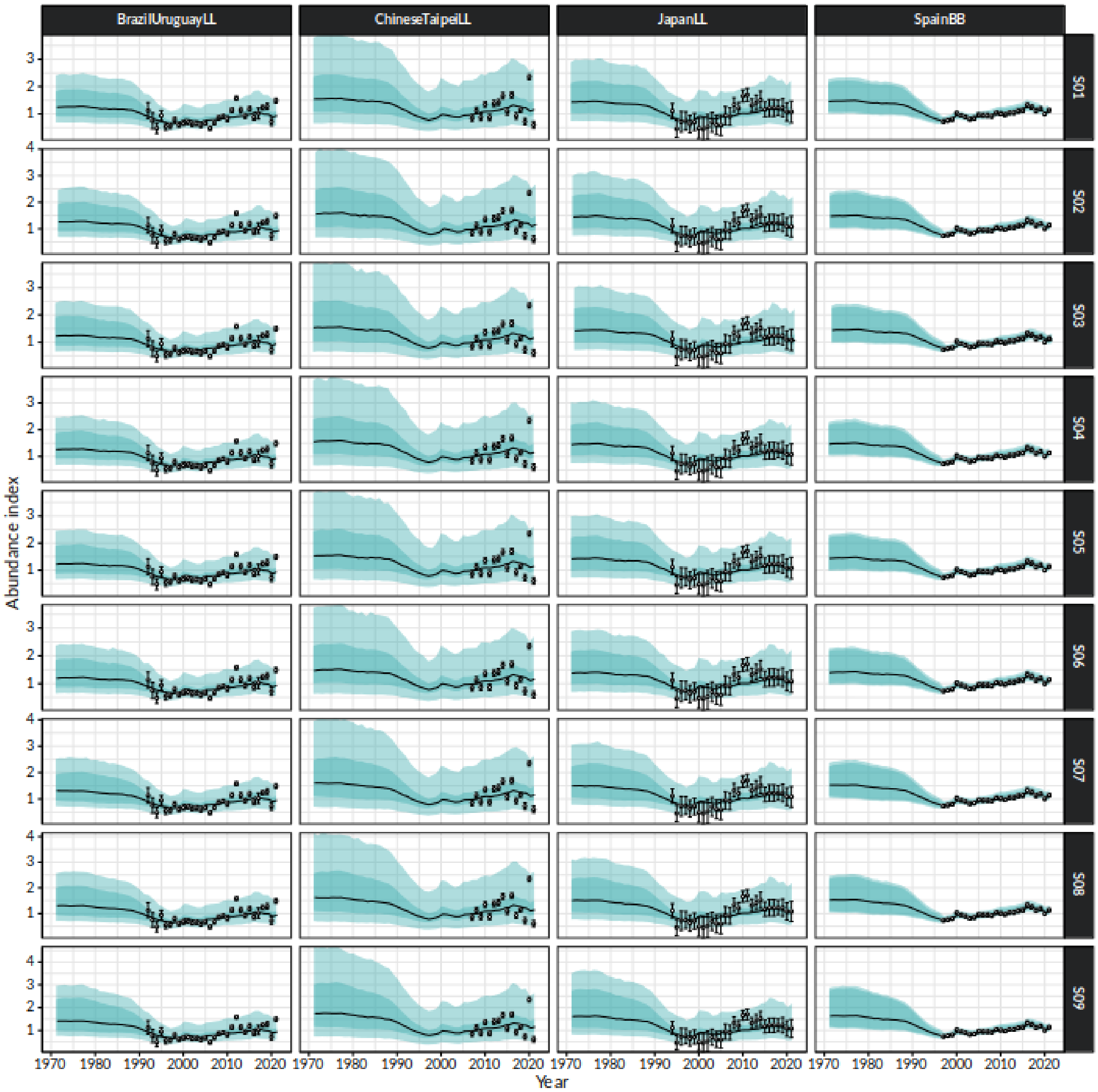
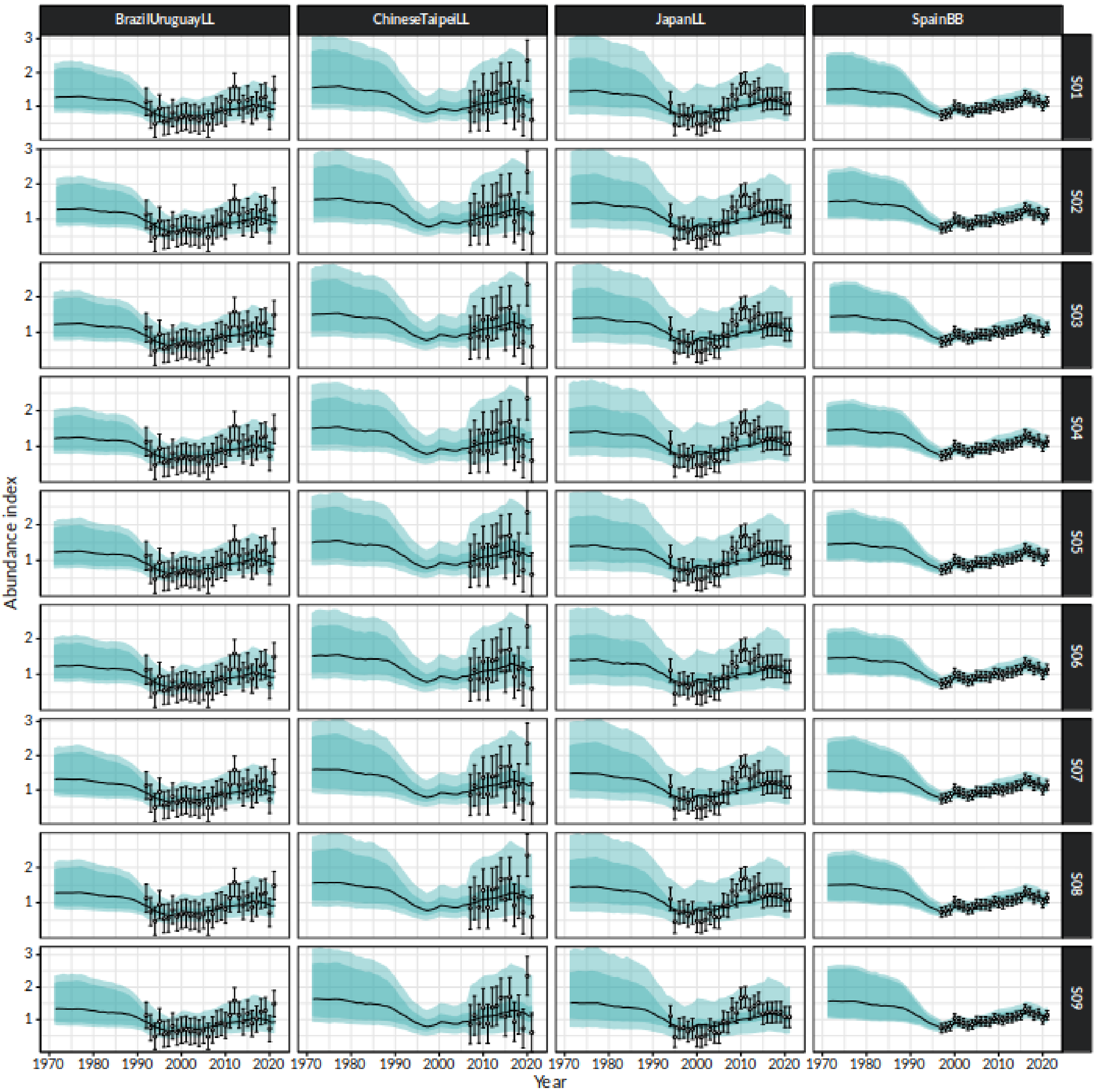


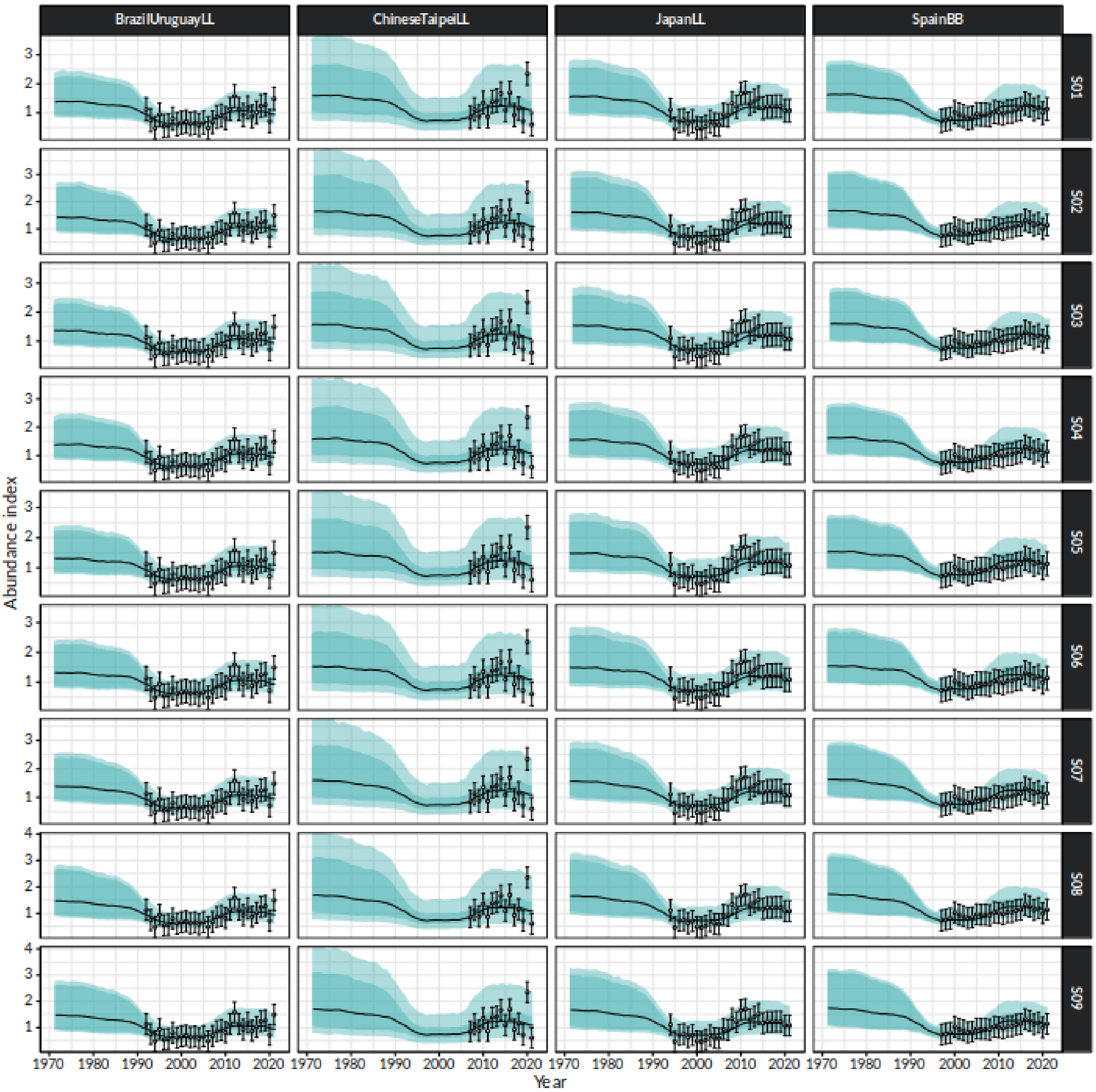
Figure 2. Comparison between *r* and B_{MSY}/K priors distributions derived from Age-Structured Equilibrium Models (ASEM) approach.



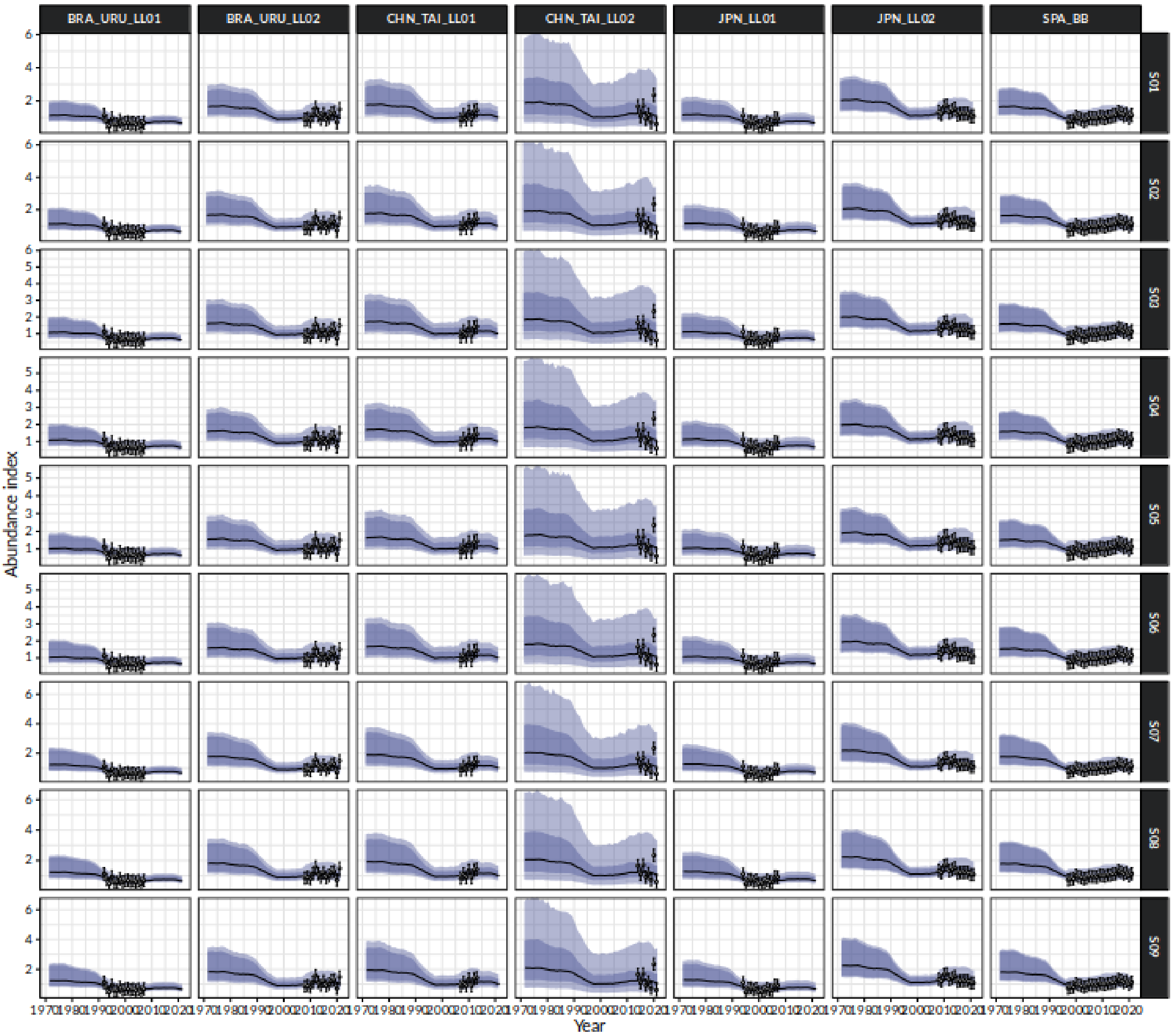
(a) Original weighting



(b) Courtney *et al* (2016)

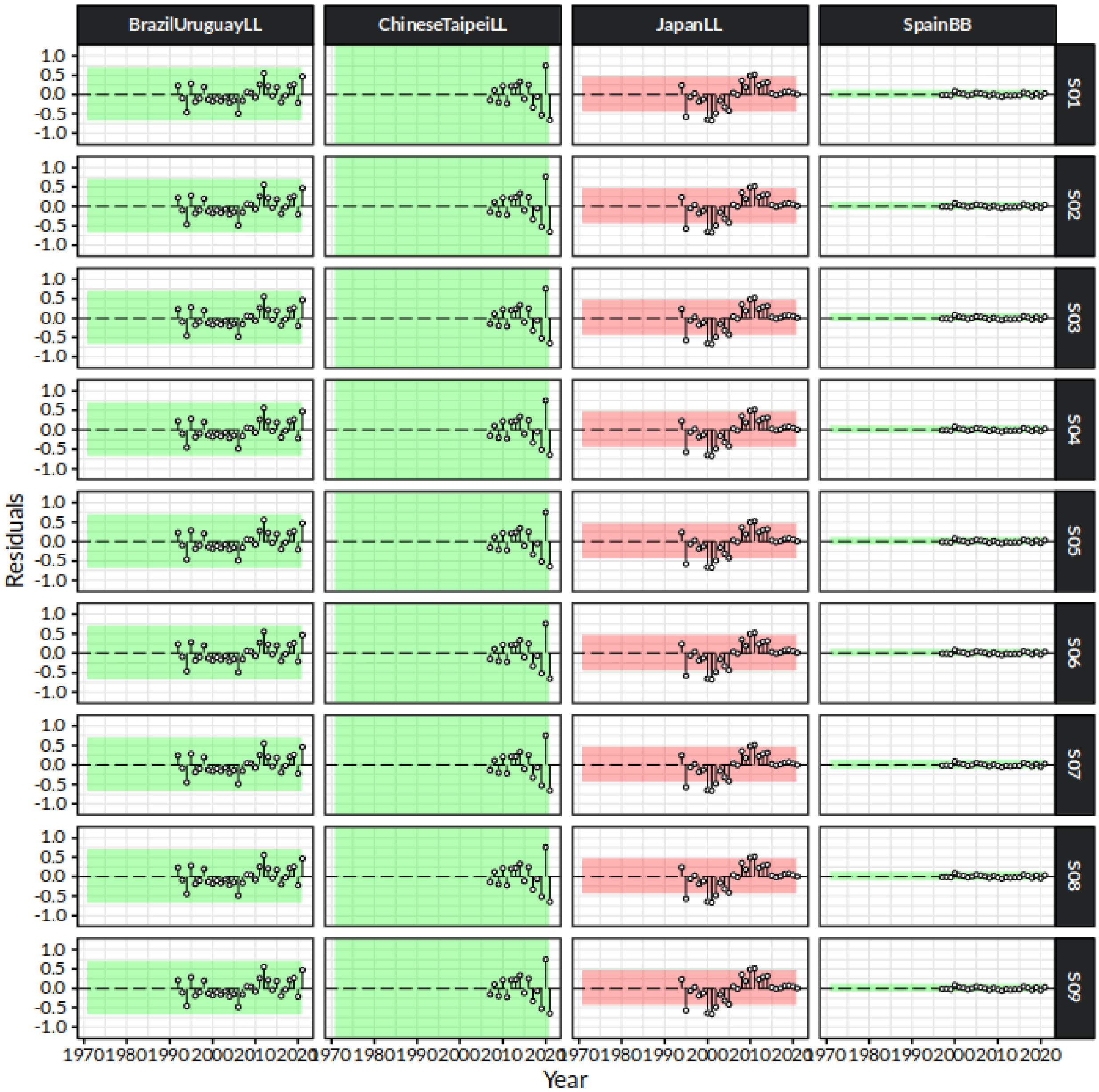


(c) Model-internal weighting

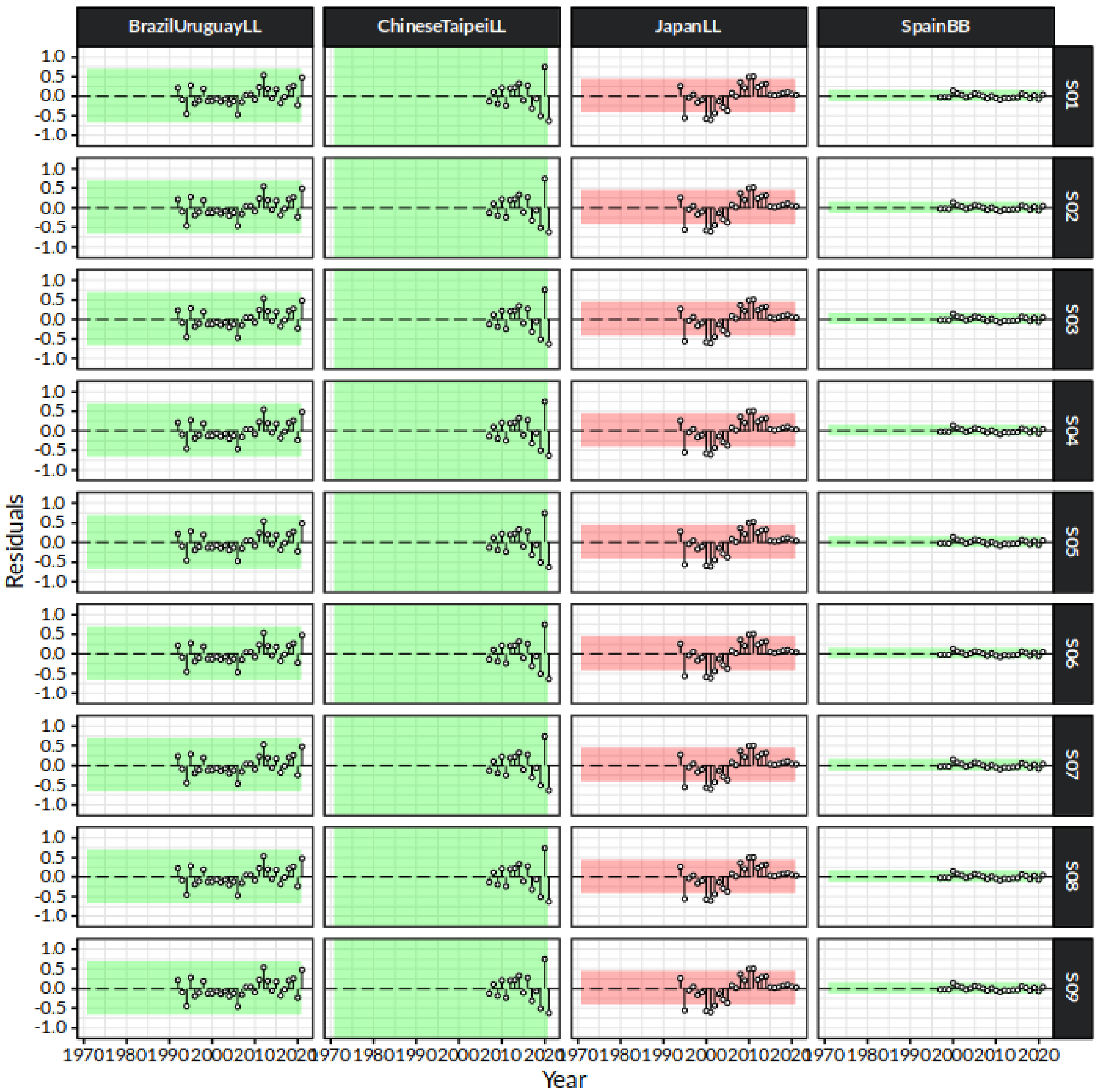


(d) Model-internal weighting – Time block

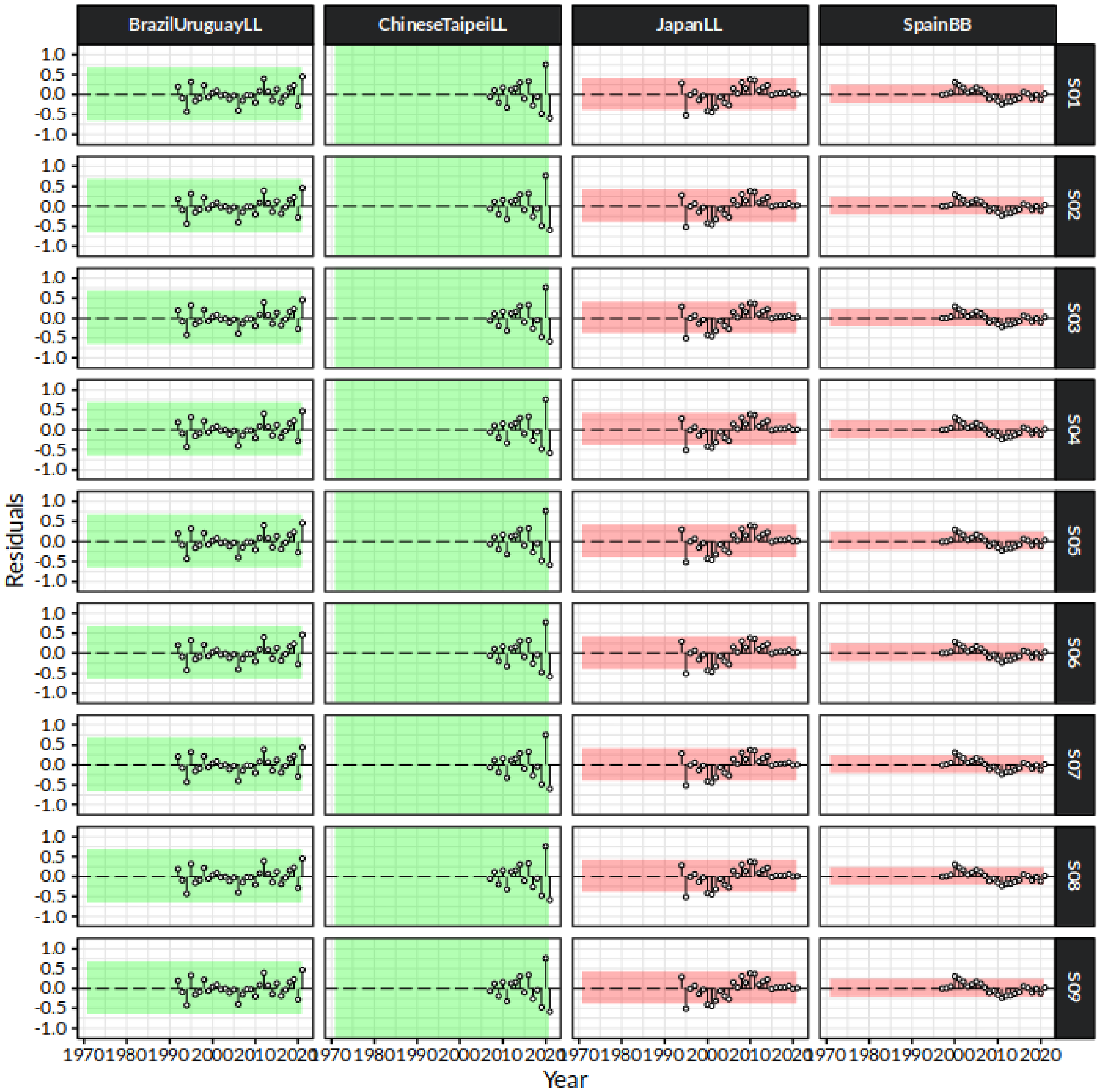
Figure 3. Time series of observed (circle) with error 95% Cis (error bars) and predicted (solid line) CPUE of South Atlantic blue shark for the Bayesian state-space surplus production model JABBA for each scenario fitted. Dark shaded grey areas show 95% credibility intervals of the expected mean CPUE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



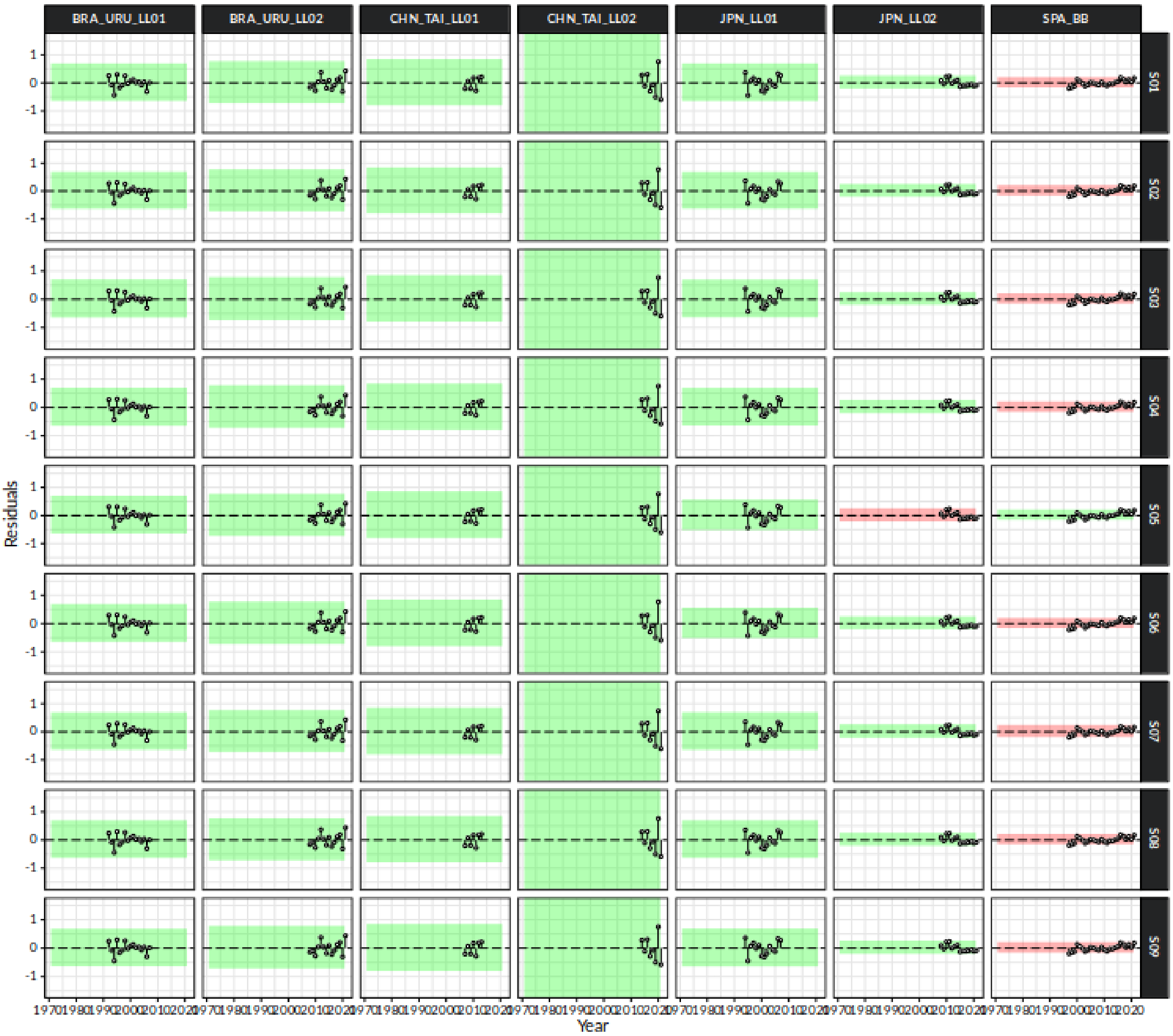
(a) Original weighting



(b) Courtney *et al* (2016)

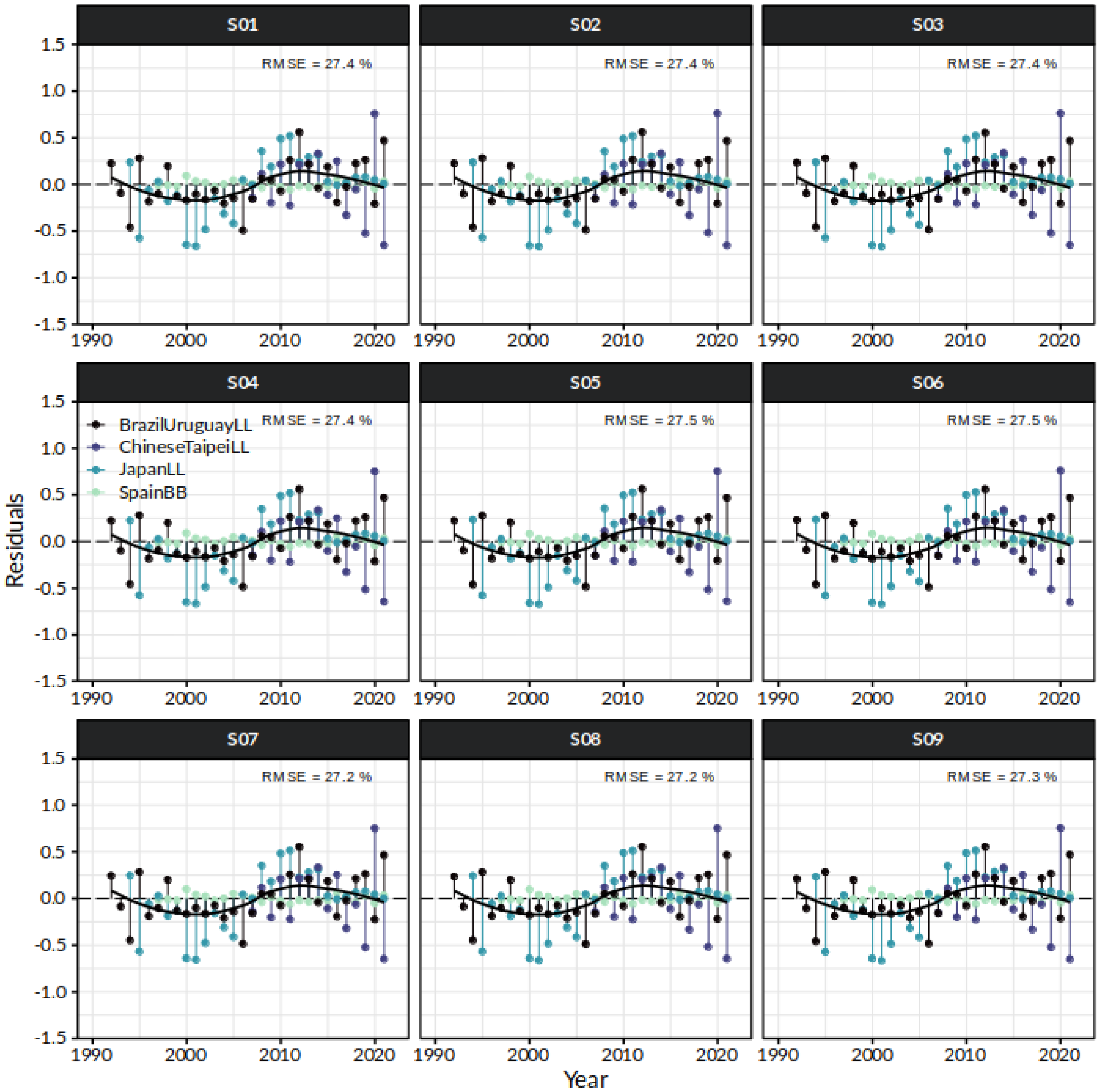


(c) Model-internal weighting

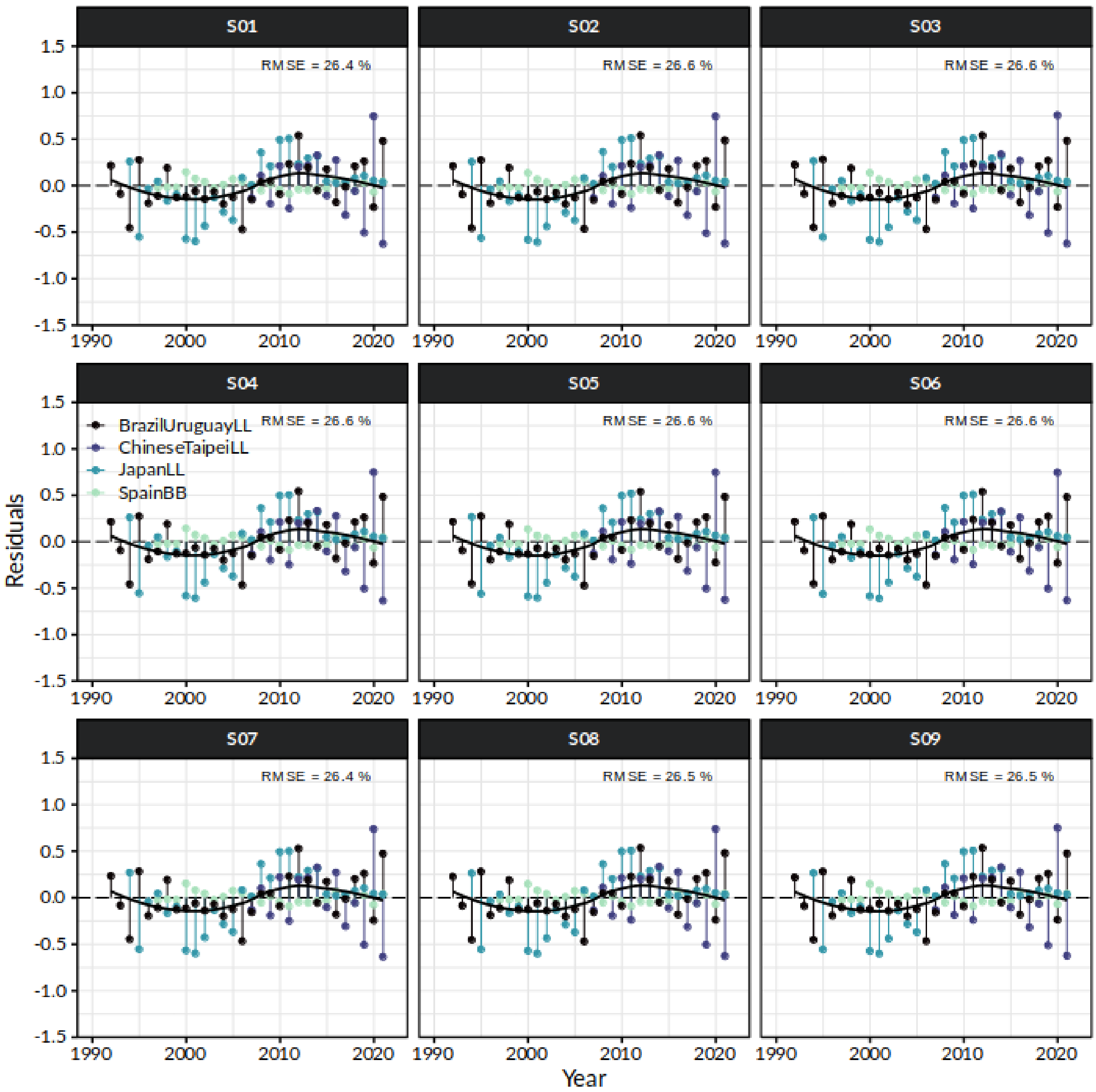


(d) Model-internal weighting – Time block

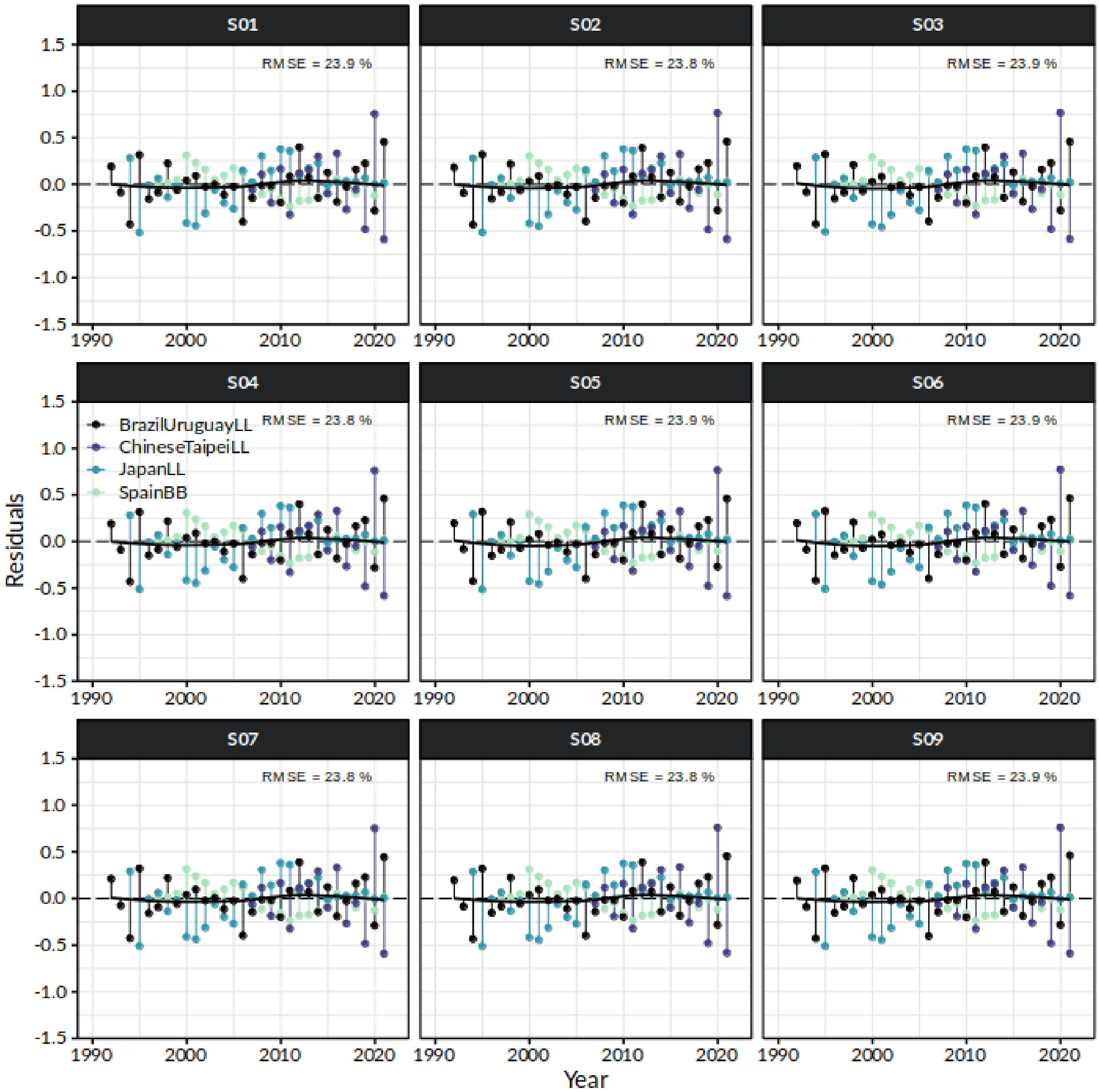
Figure 4. Runs tests to quantitatively evaluate the randomness of the time series of CPUE residuals for each scenario fitted for the South Atlantic blue shark. Green panels indicate no evidence of lack of randomness of time-series residuals (p>0.05) while red panels indicate the opposite. The inner shaded 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).



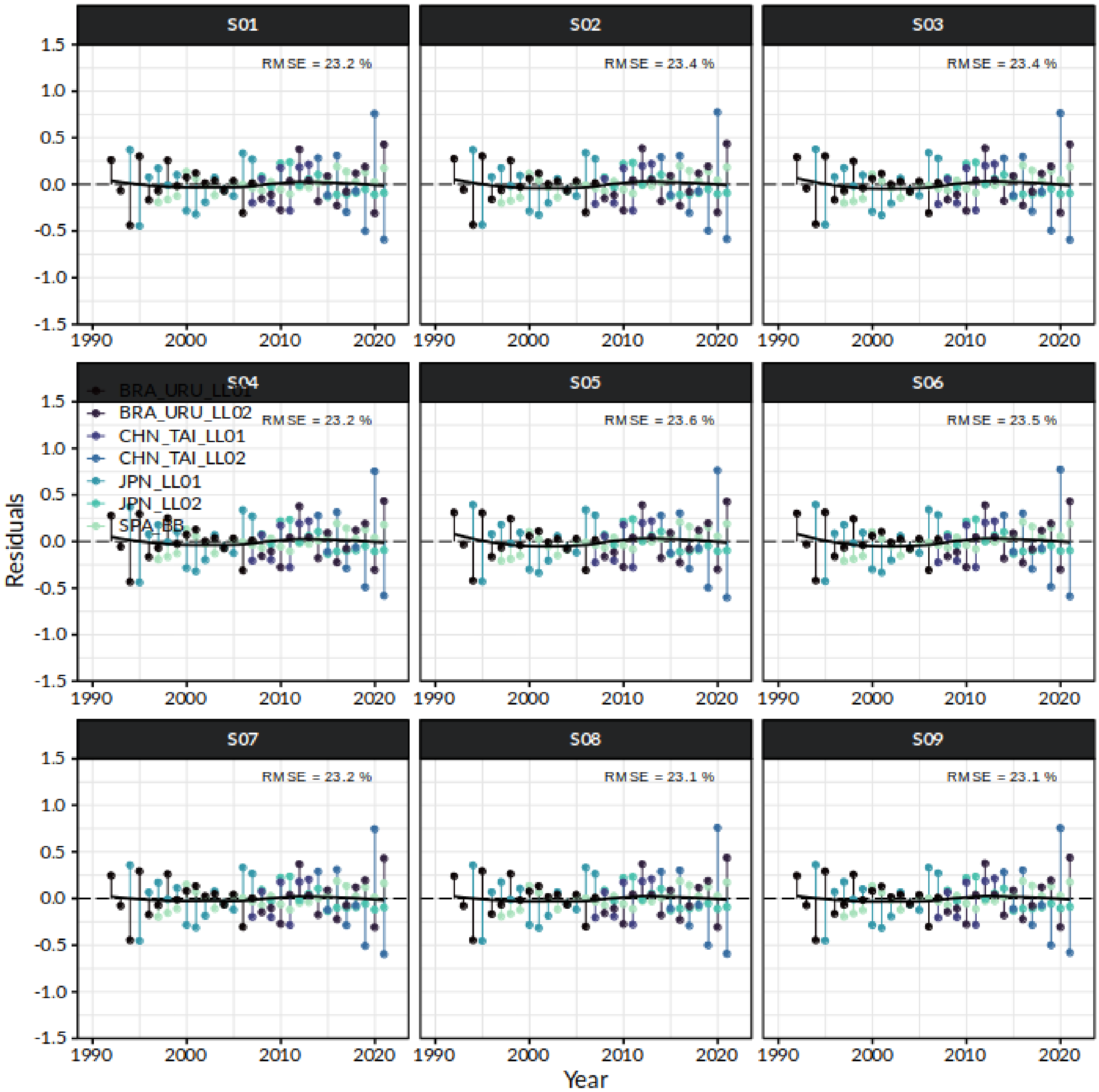
(a) Original weighting



(b) Courtney *et al* (2016)

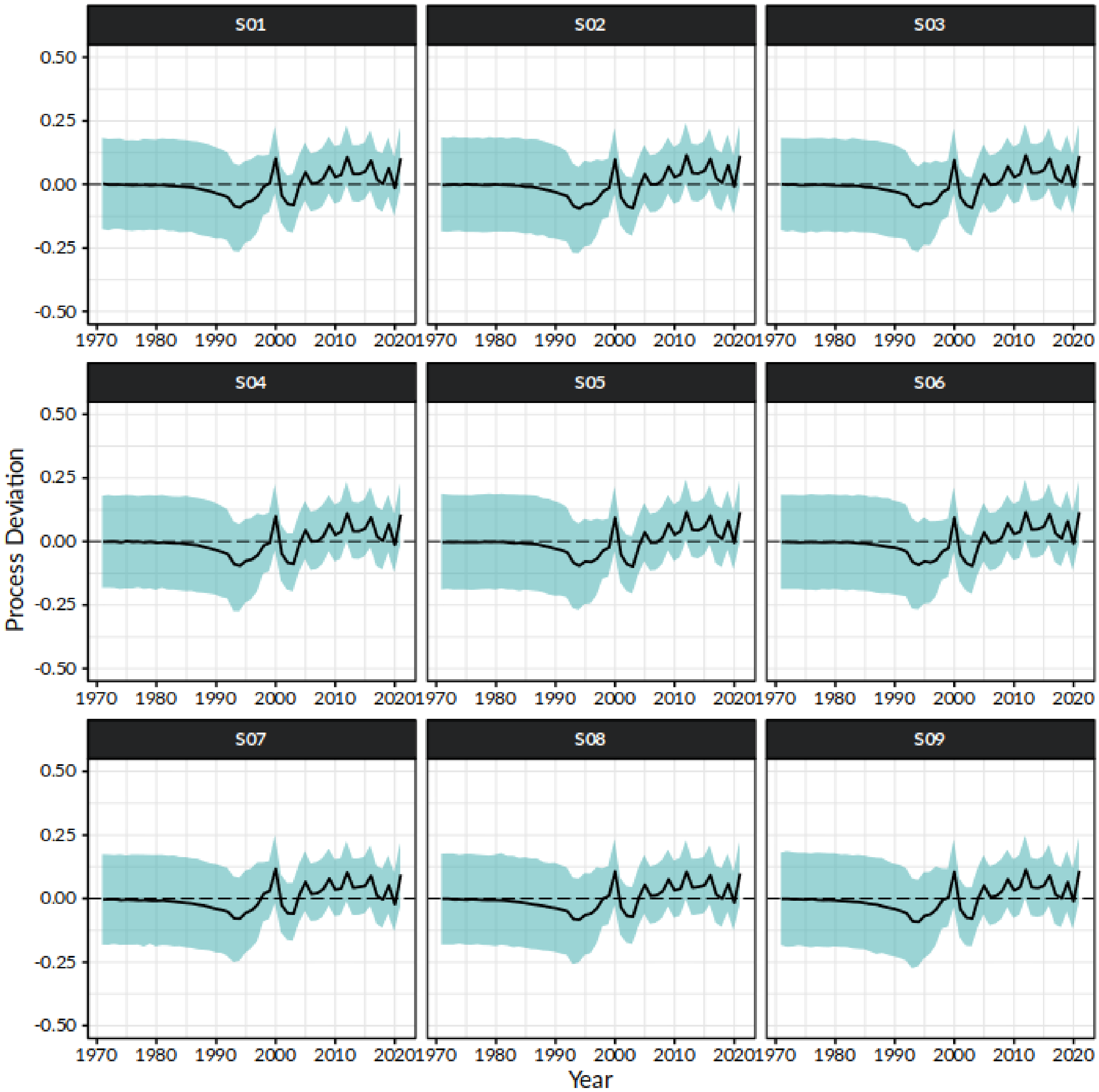


(c) Model-internal weighting

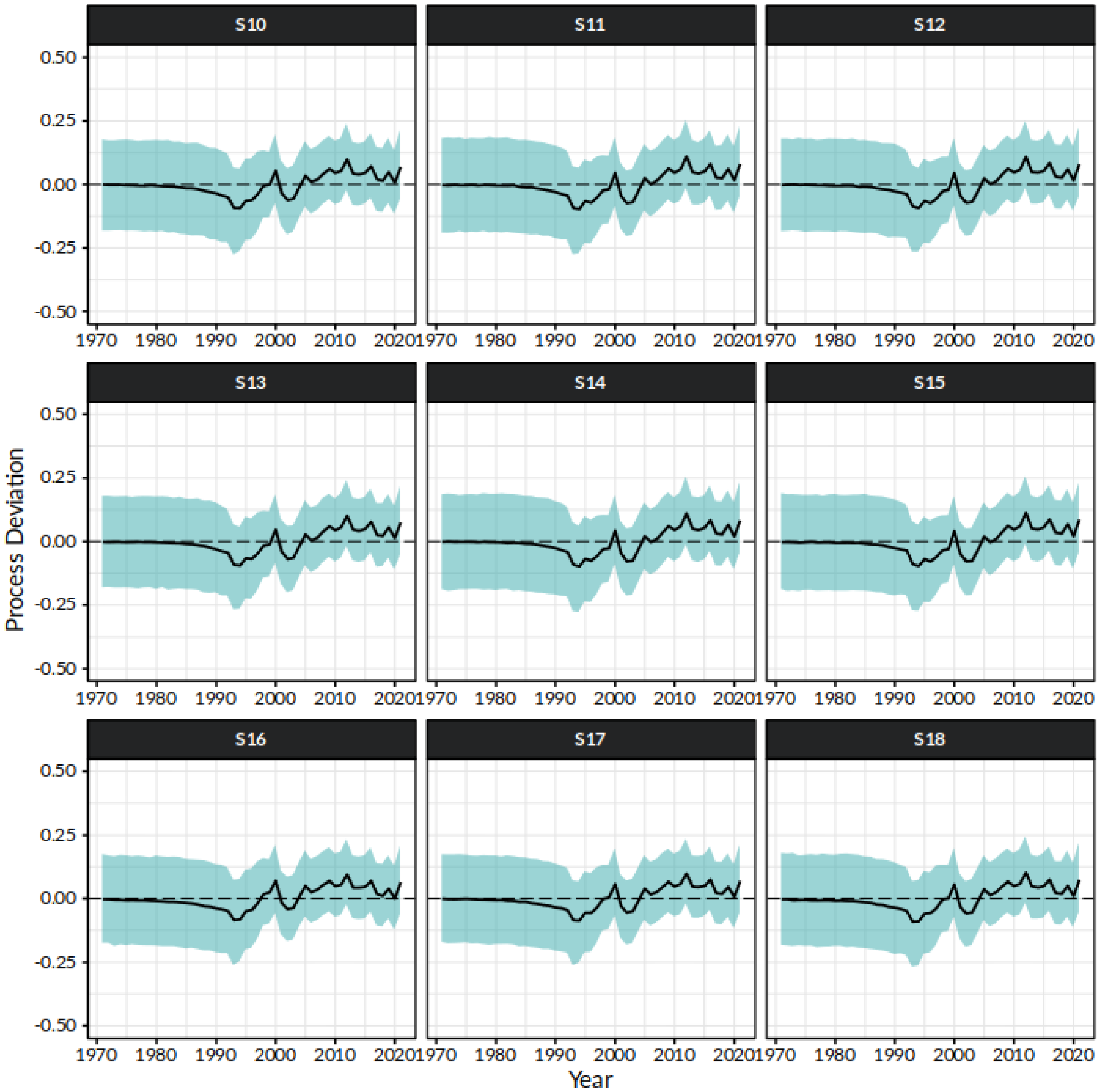


(d) Model-internal weighting – Time block

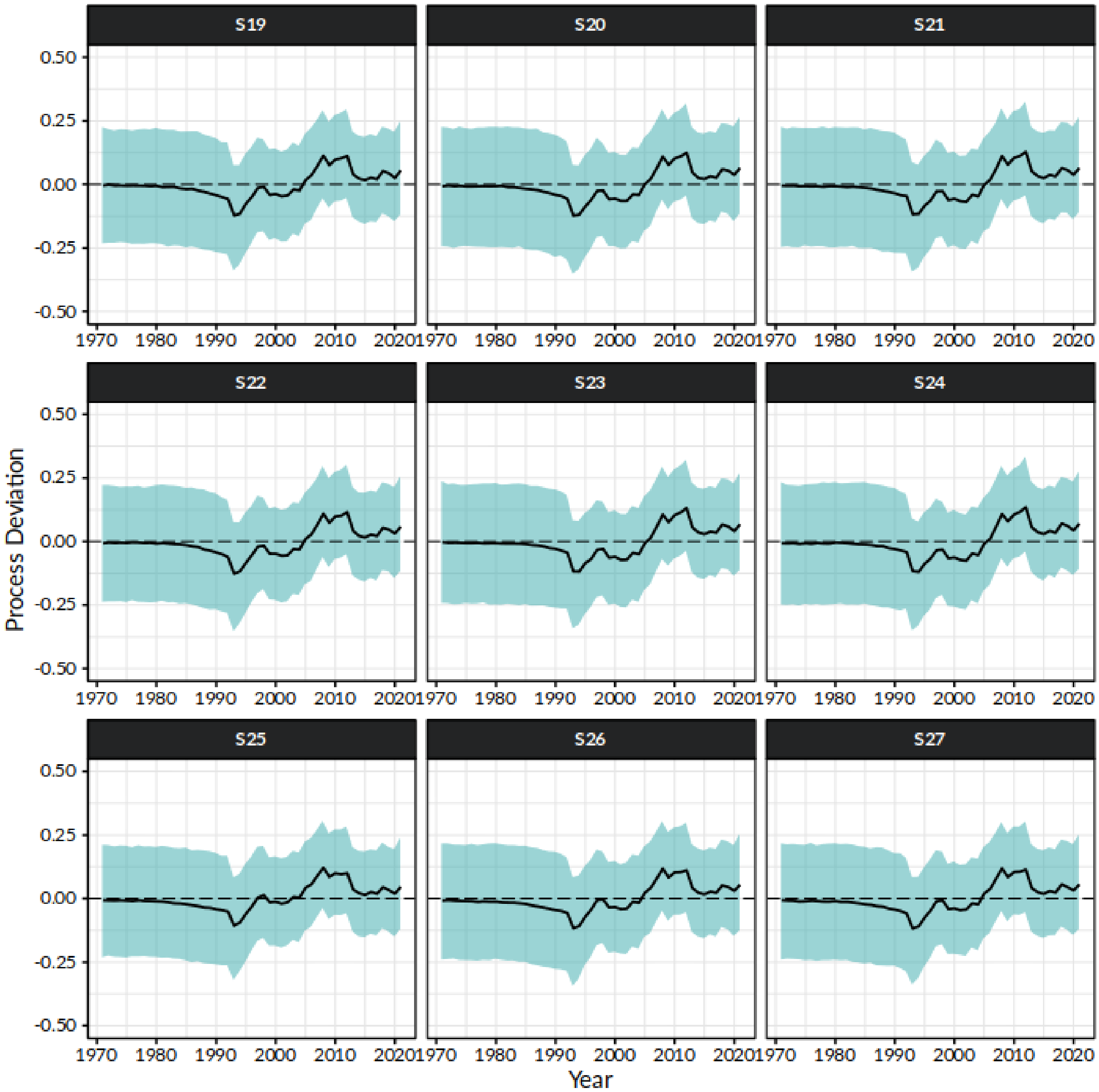
Figure 5. JABBA residual diagnostic plots for alternative sets of CPUE indices examined for each scenario fitted for the South Atlantic blue shark. Boxplots indicate the median and quantiles of all residuals available for any given year, and solid black lines indicate a loess smoother through all residuals.



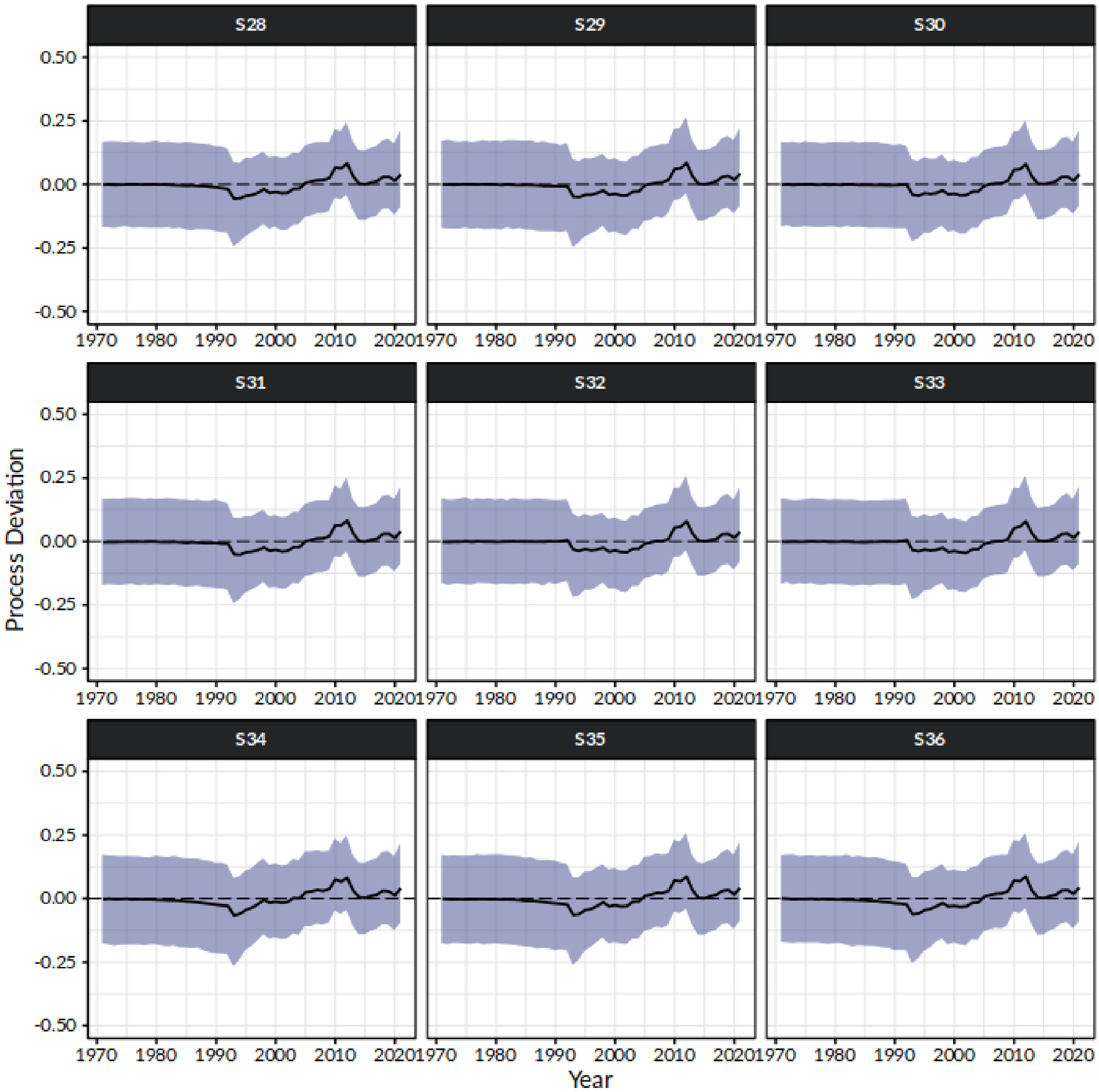
(a) Original weighting



(b) Courtney *et al* (2016)

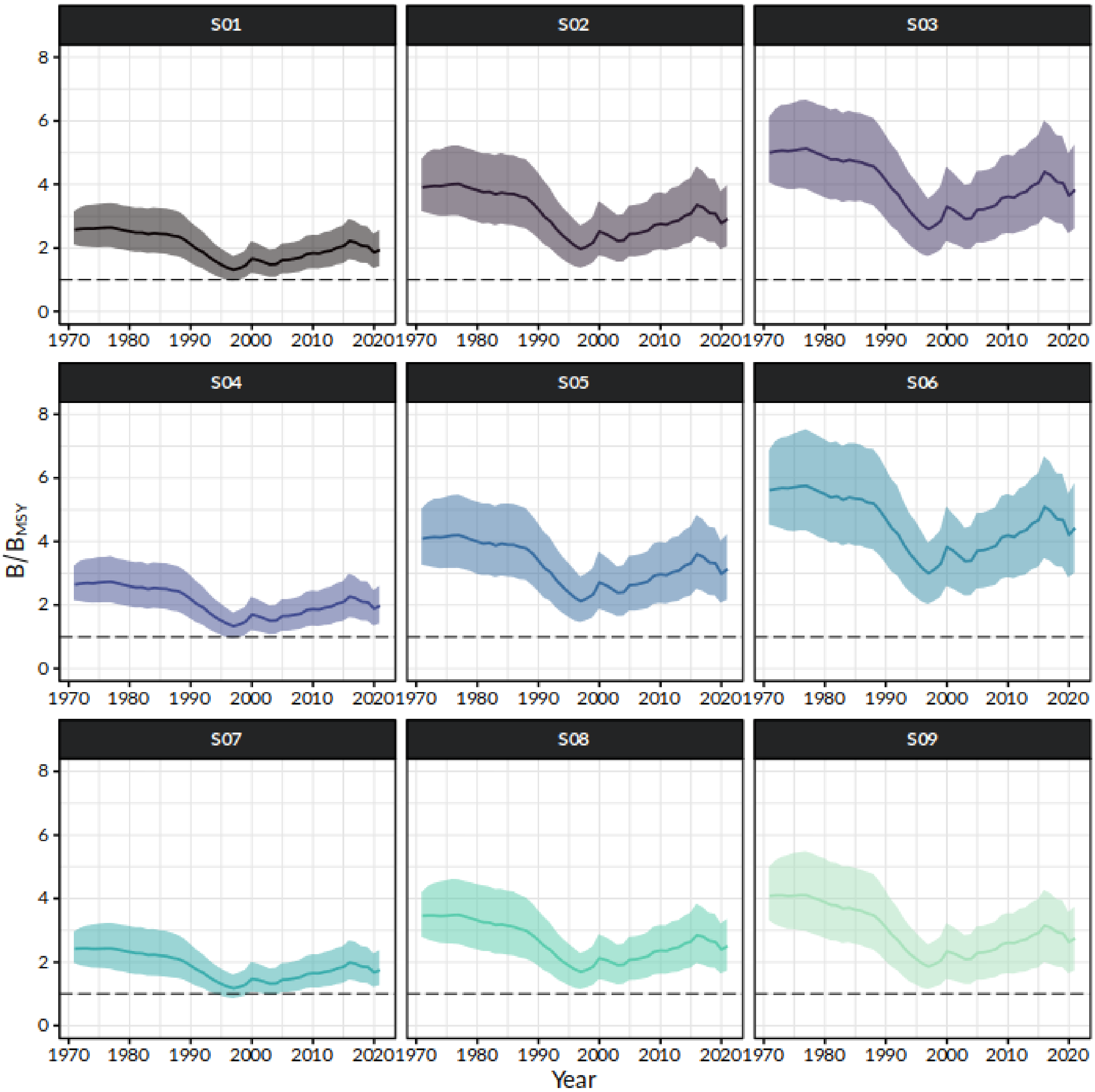


(c) Model-internal weighting

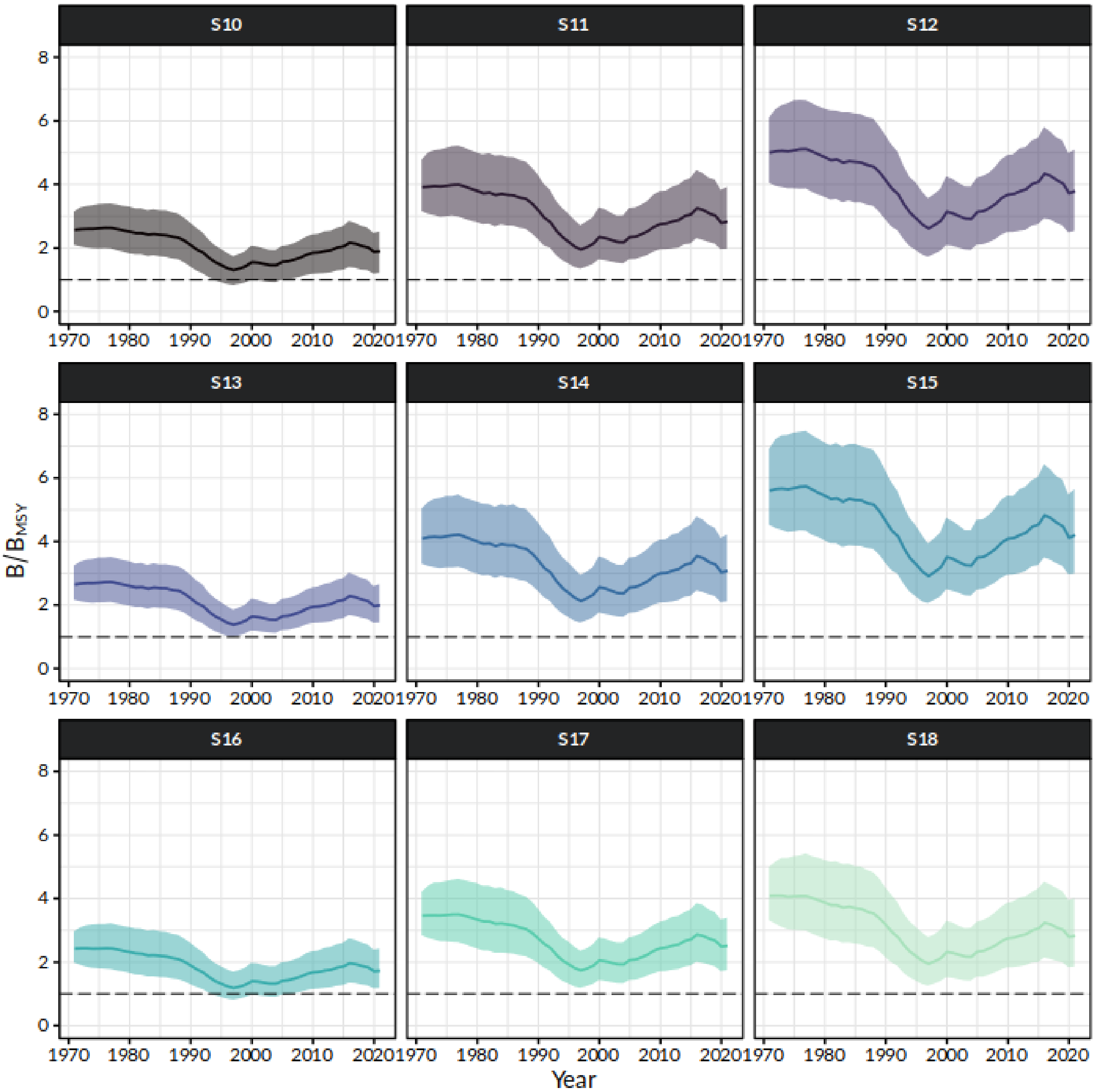


(d) Model-internal weighting – Time block

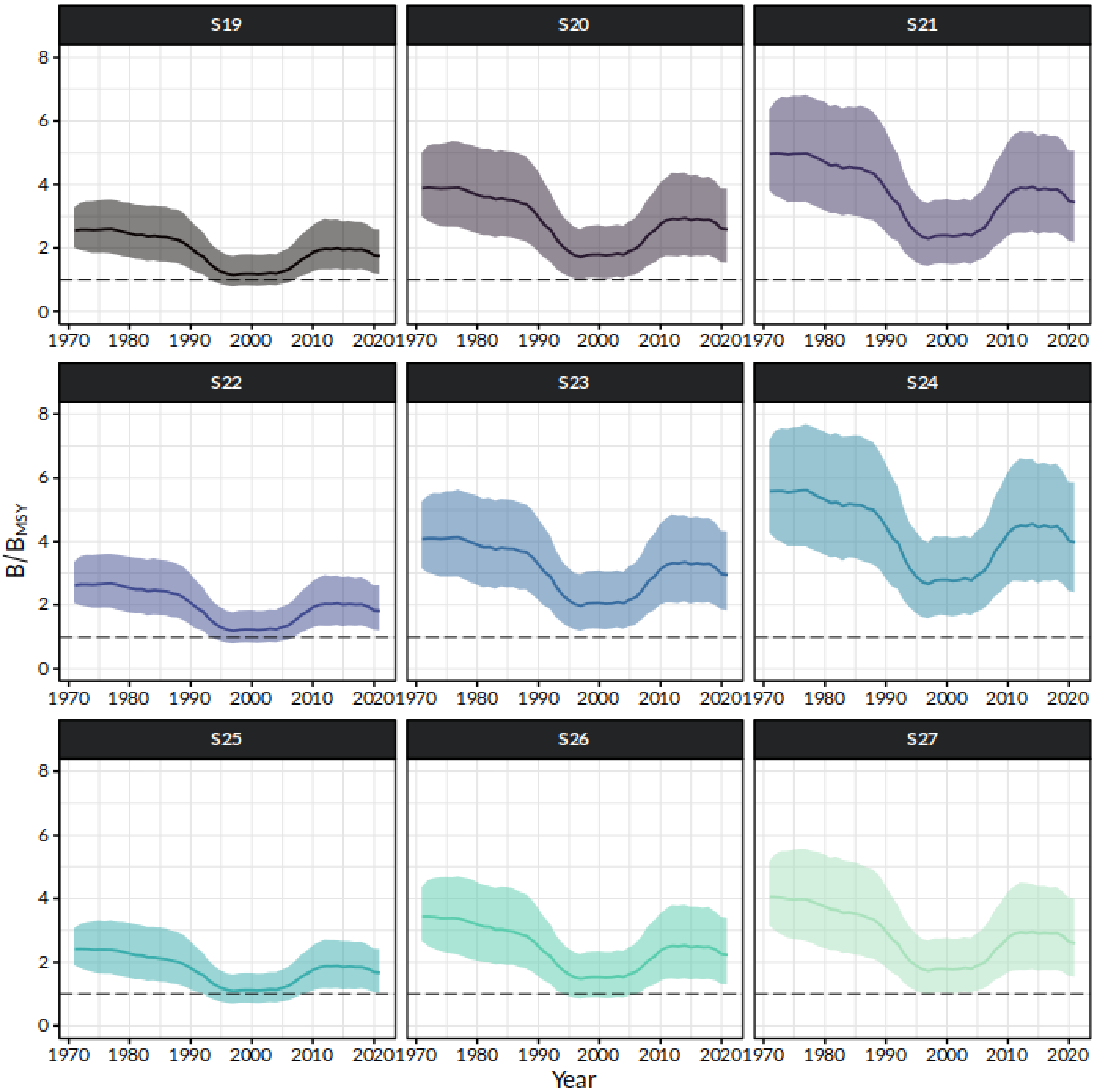
Figure 6. JABBA residual diagnostic plots for alternative sets of CPUE indices examined for each scenario fitted for the South Atlantic blue shark. Process error deviates (median: solid line) with shaded grey area indicating 95% credibility intervals.



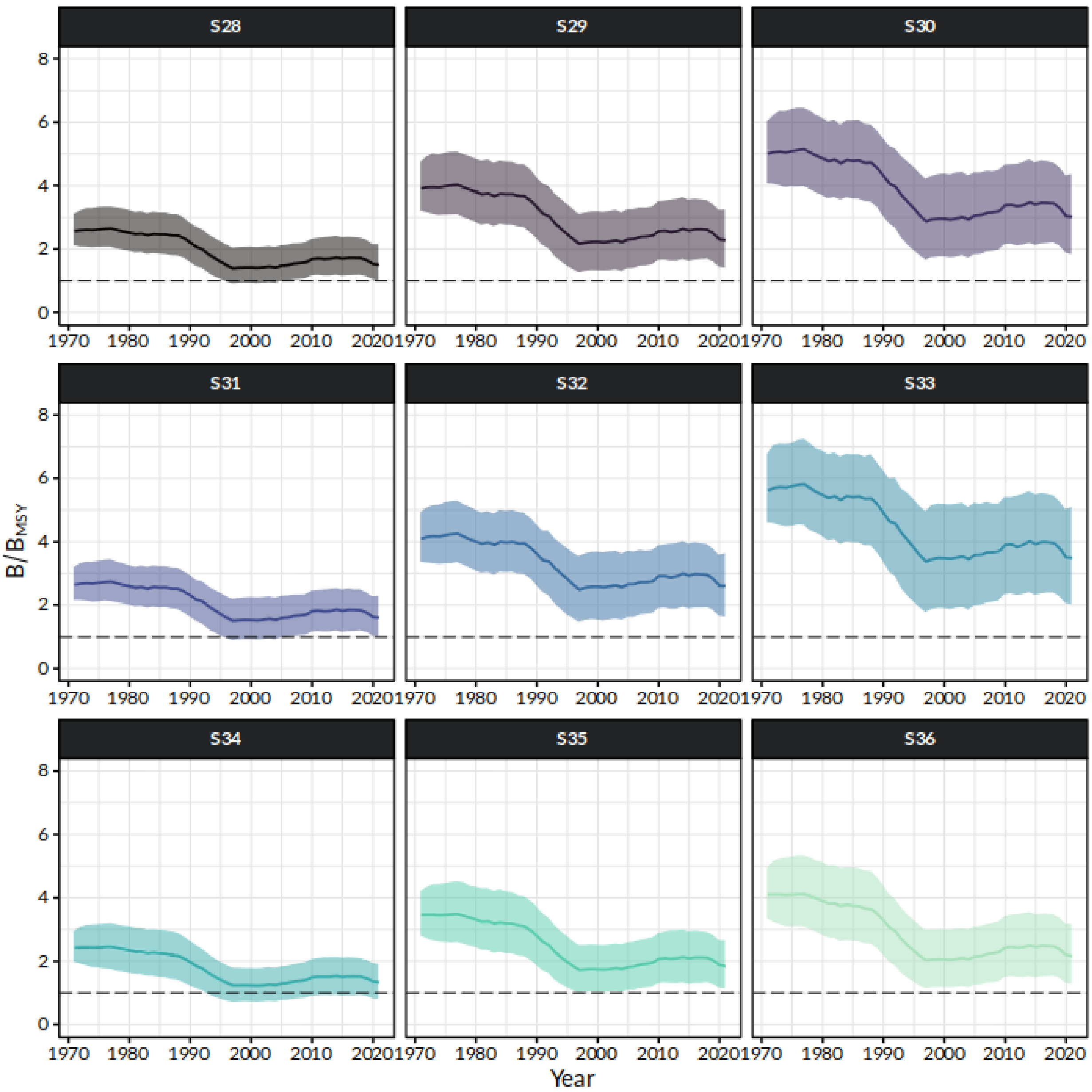
(a) Original weighting



(b) Courtney *et al* (2016)

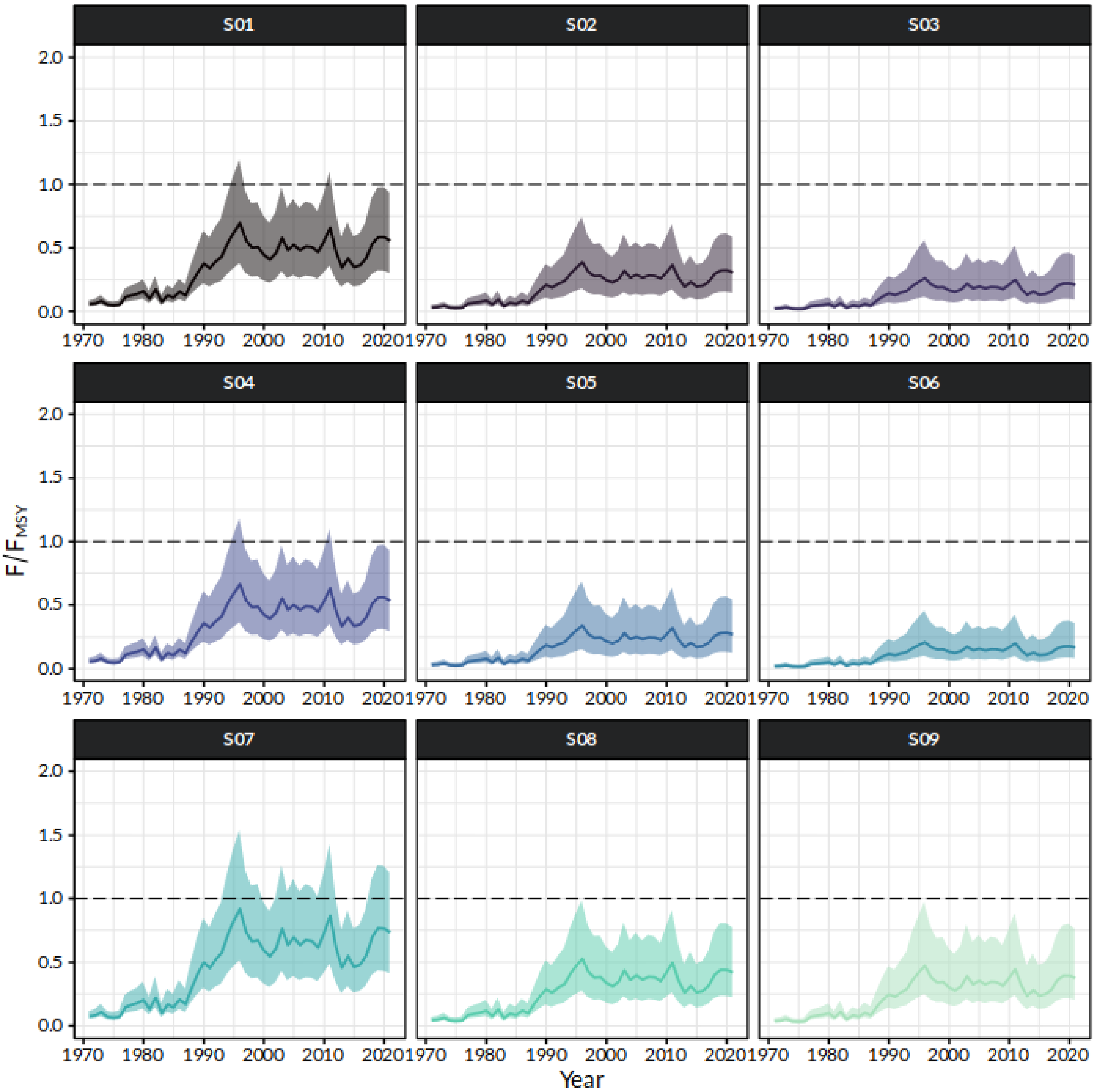


(c) Model-internal weighting

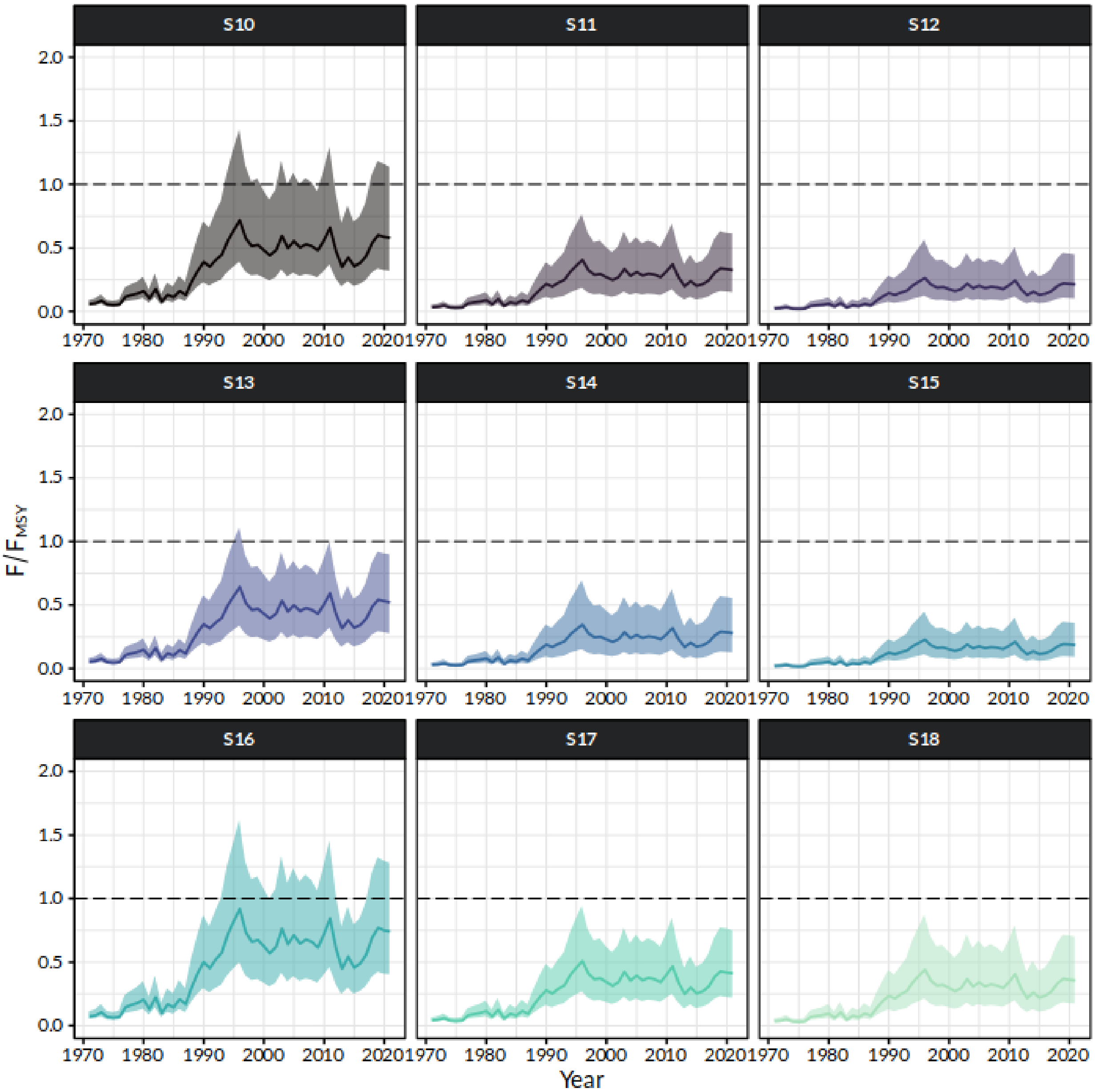


(d) Model-internal weighting - Time block

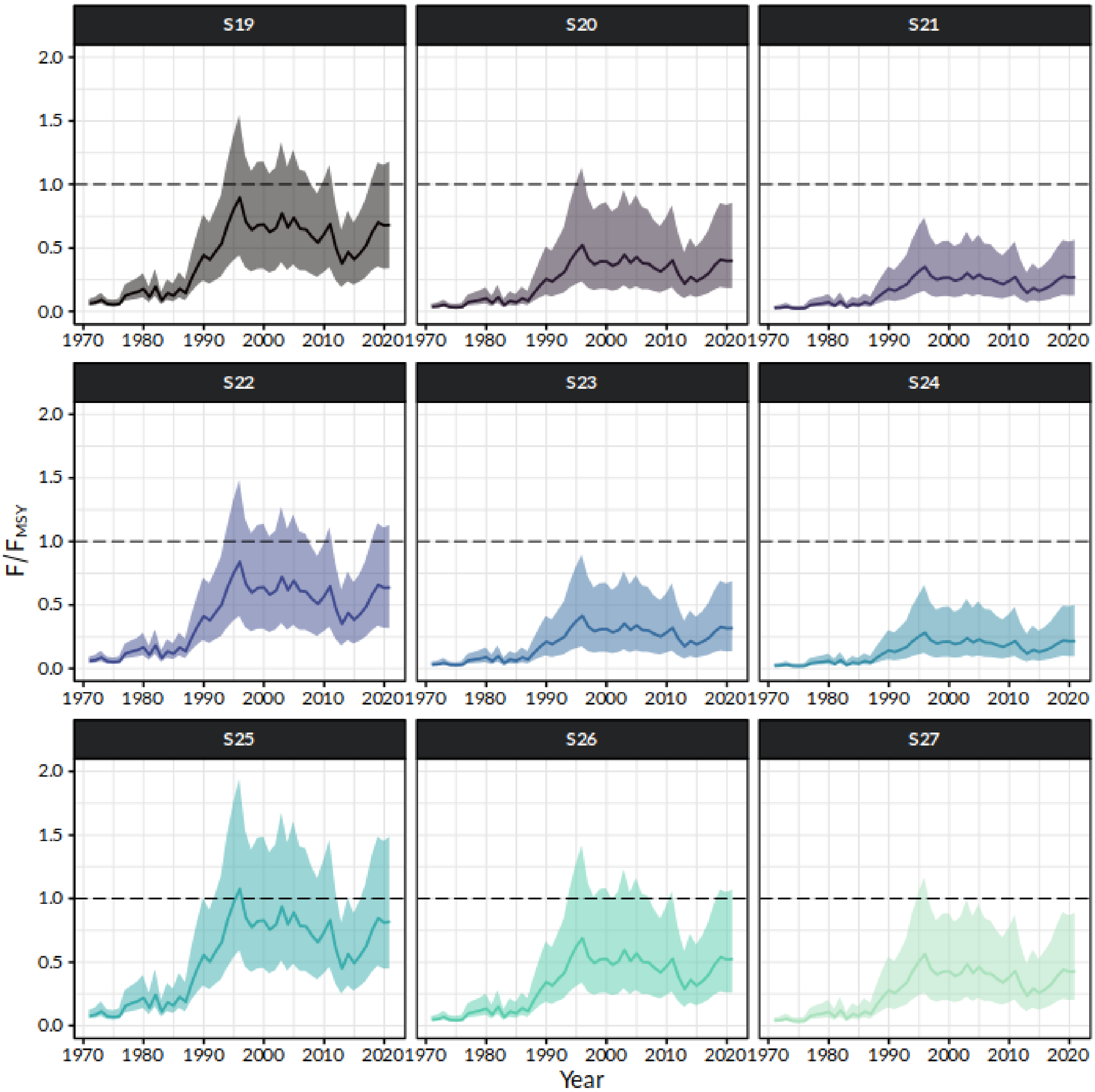
Figure 7. Trends in biomass relative to B_{MSY} (B/B_{MSY}) for each scenario from the Bayesian state-space surplus production JABBA model fits to South Atlantic blue shark.



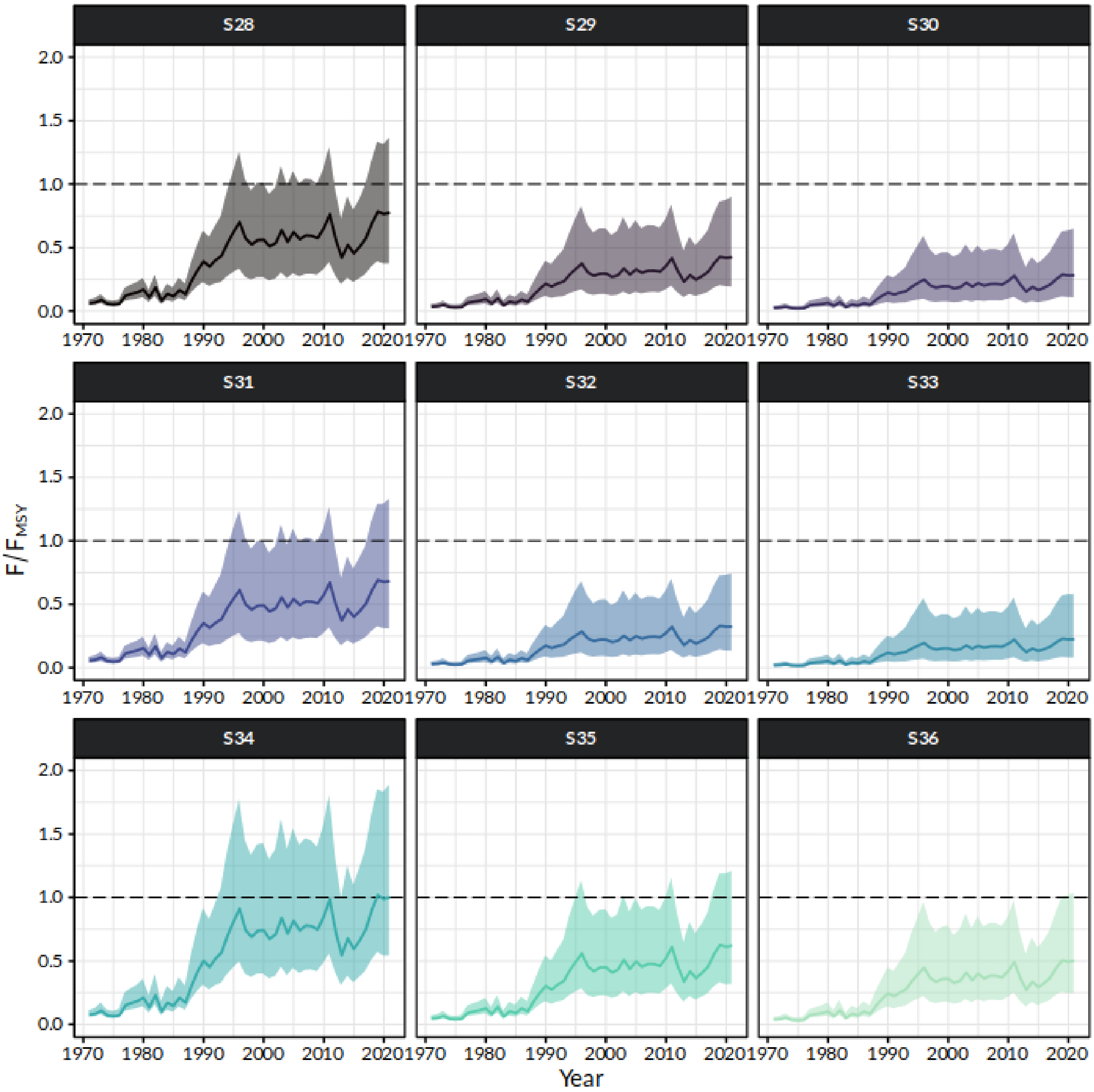
(a) Original weighting



(b) Courtney *et al* (2016)

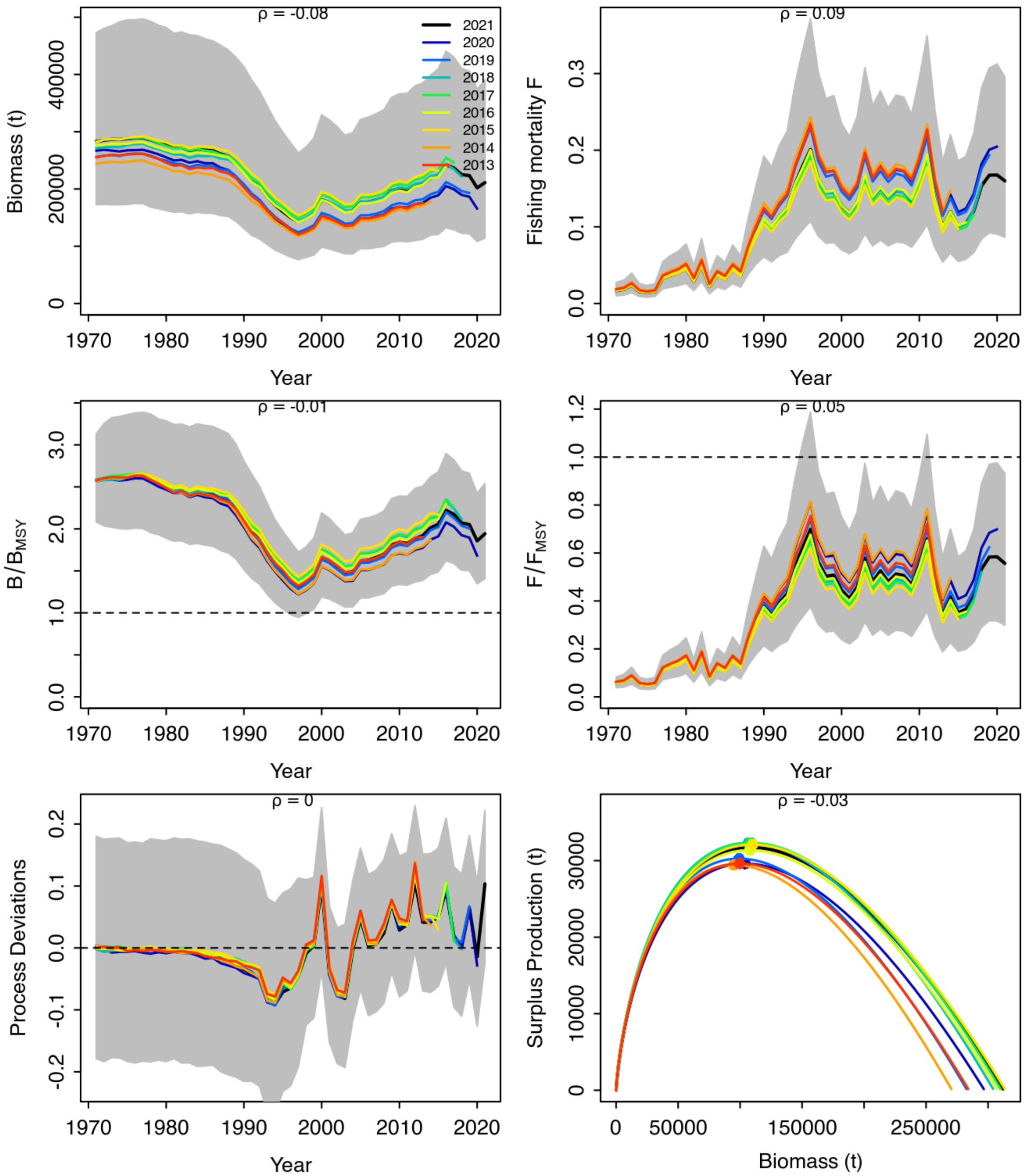


(c) Model-internal weighting

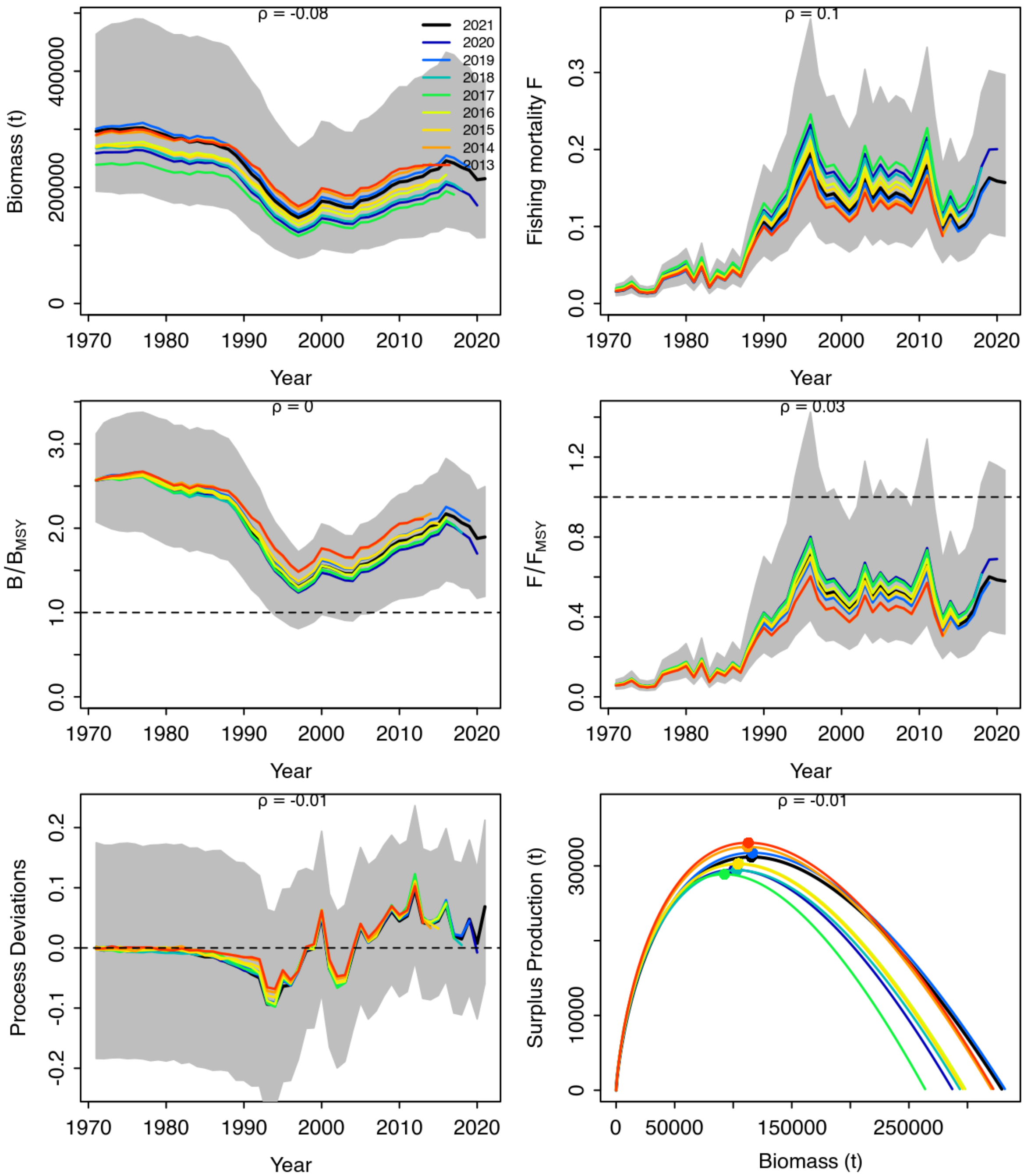


(d) Model-internal weighting - Time block

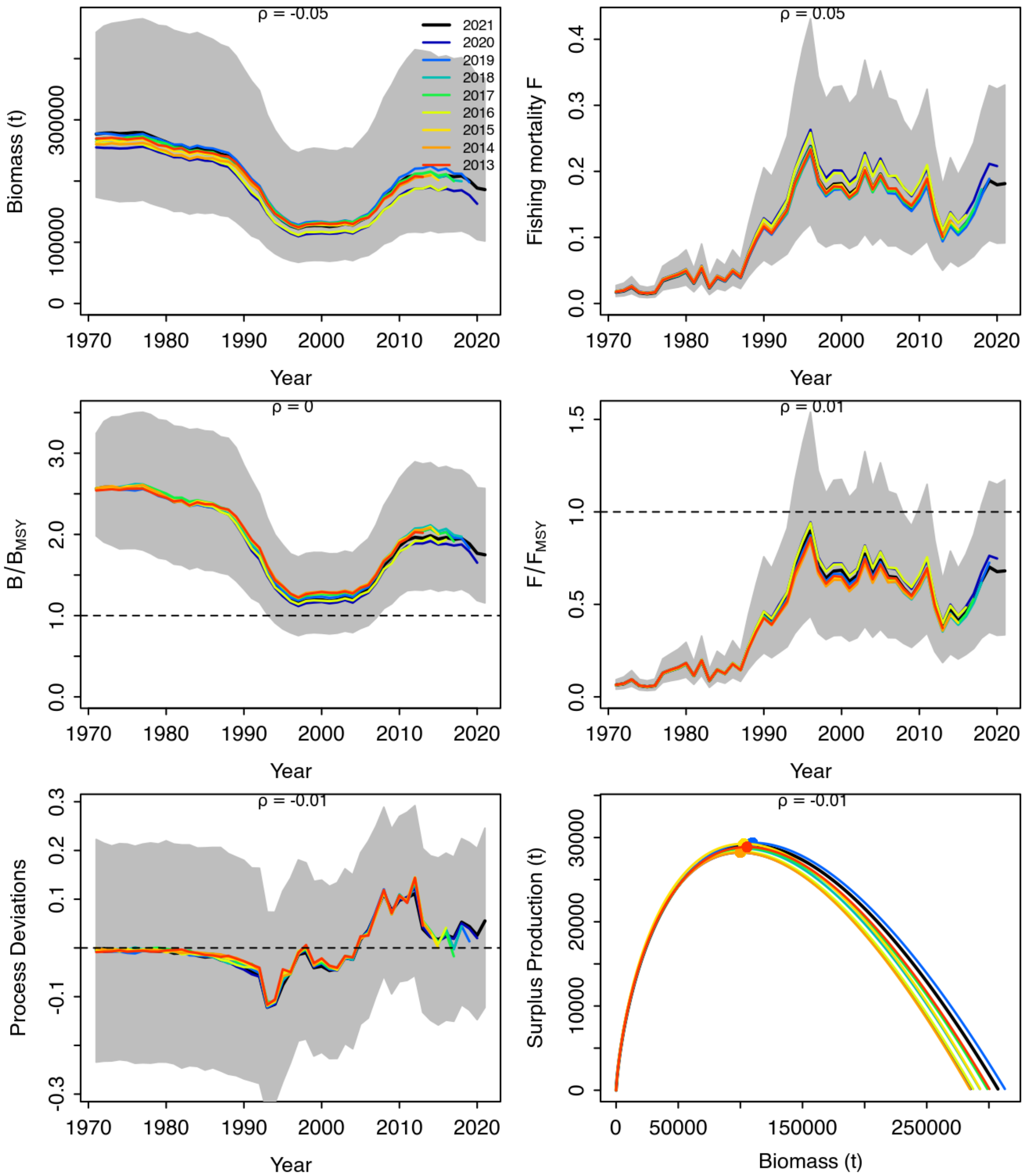
Figure 8. Trends in biomass relative to F_{MSY} (F/F_{MSY}) for each scenario from the Bayesian state-space surplus production JABBA model fits to South Atlantic blue shark.



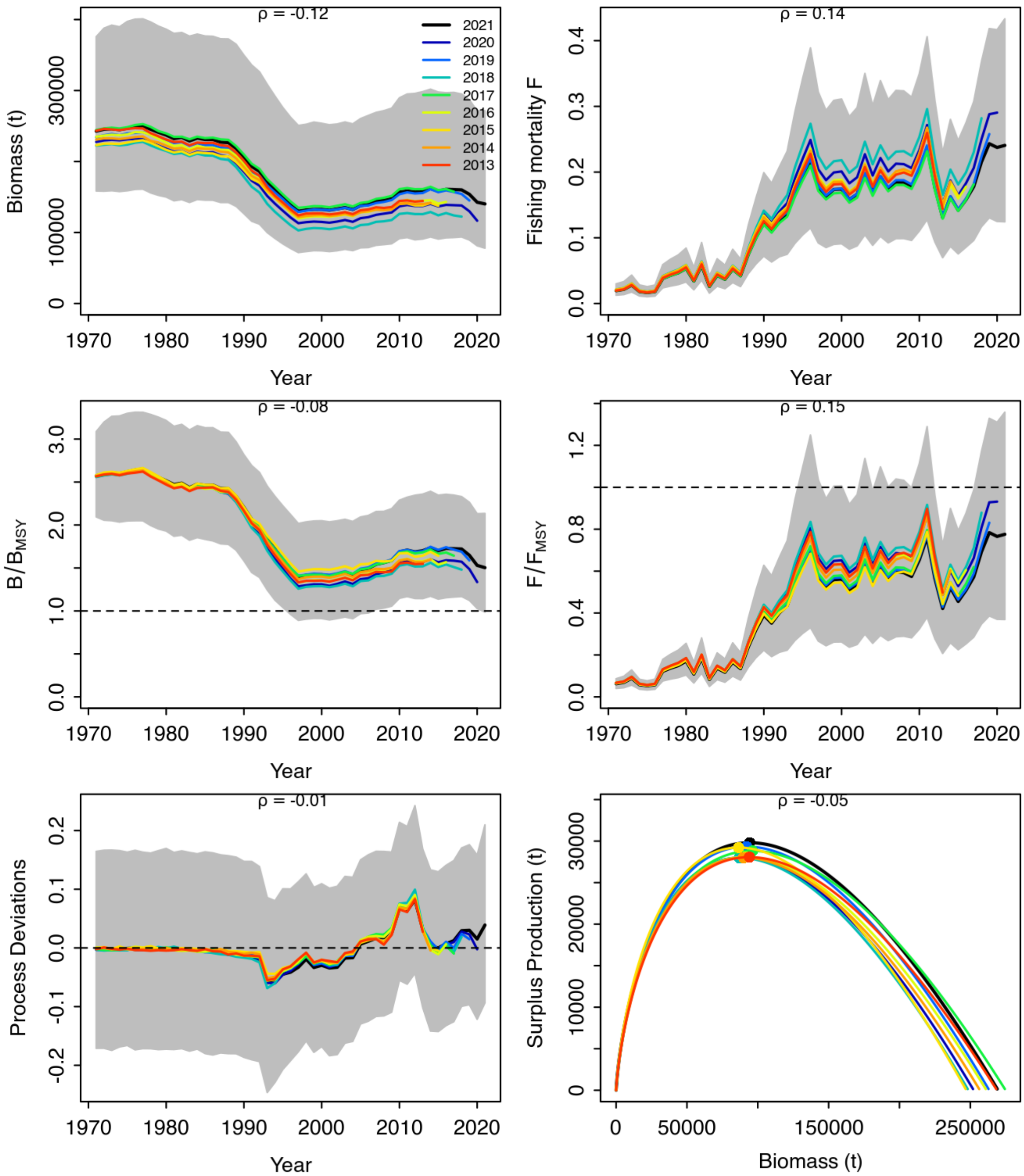
(a) Original weighting



(b) Courtney *et al* (2016)

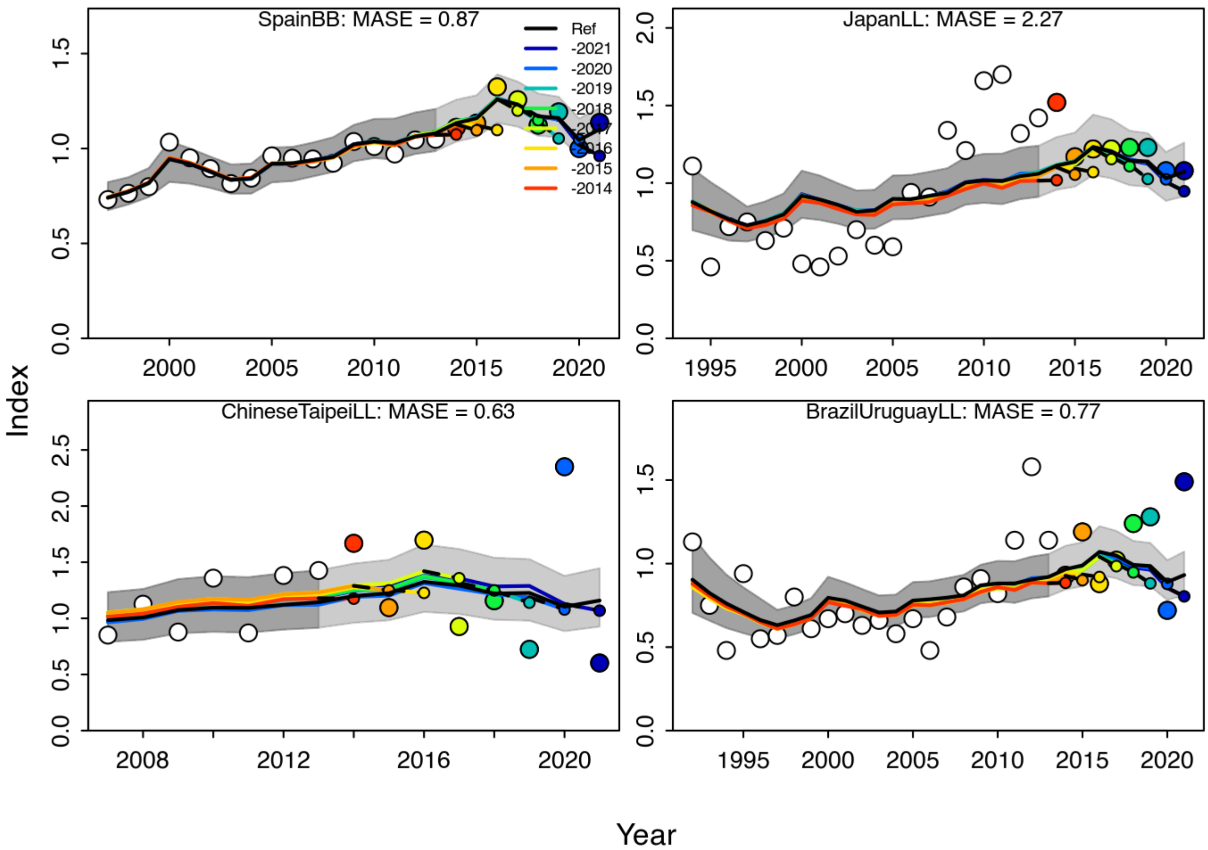


(c) Model-internal weighting

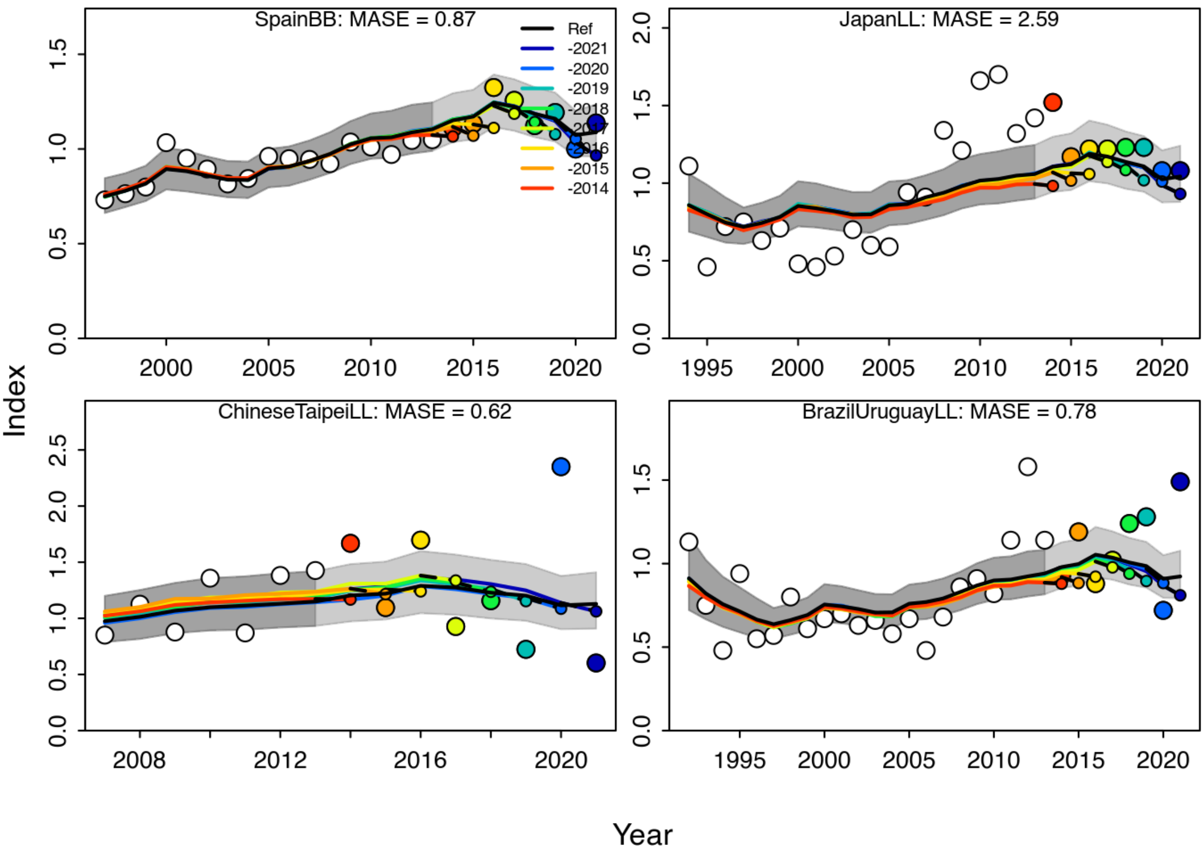


(d) Model-internal weighting - Time block

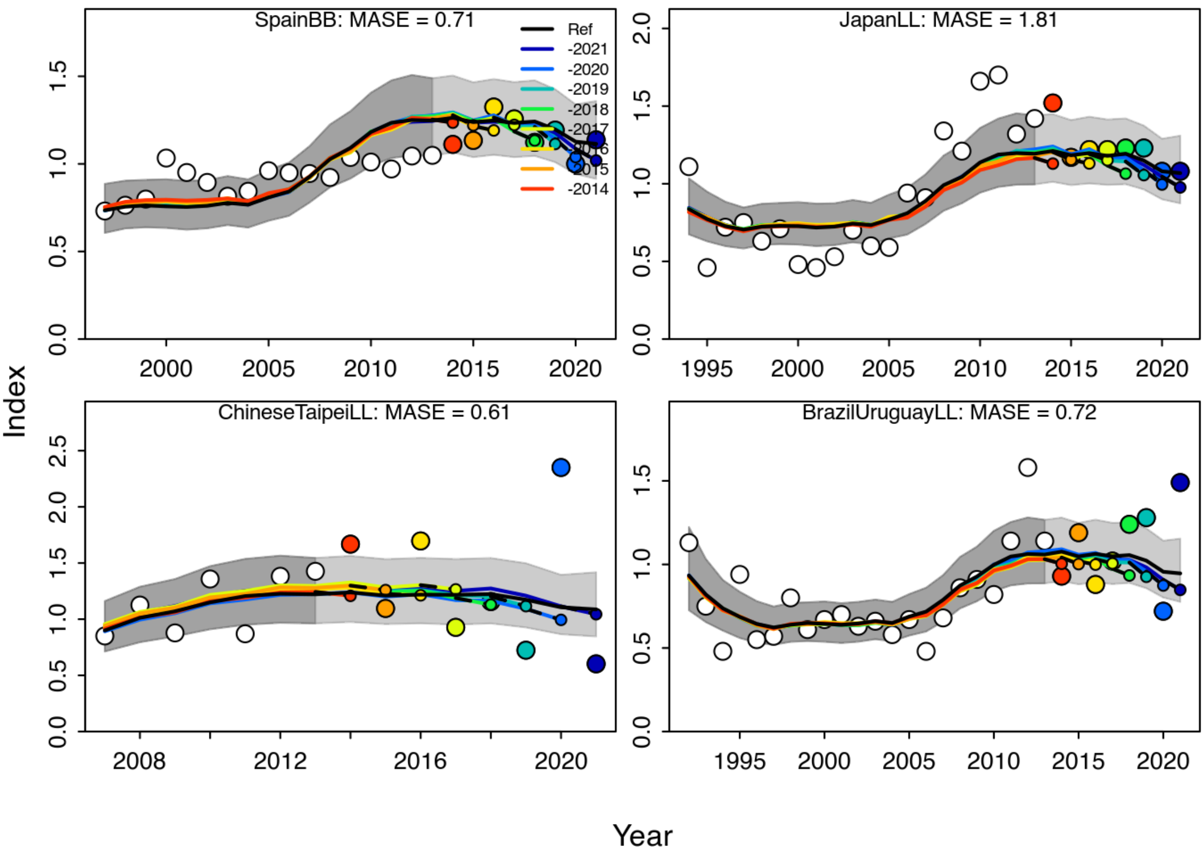
Figure 9. Retrospective analysis conducted for the first scenario of each model-weighting process tested for South Atlantic blue shark, by removing one year at a time sequentially (n=8) and predicting the trends in biomass and fishing mortality (upper panels), biomass relative to B_{MSY} (B/B_{MSY}) 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.



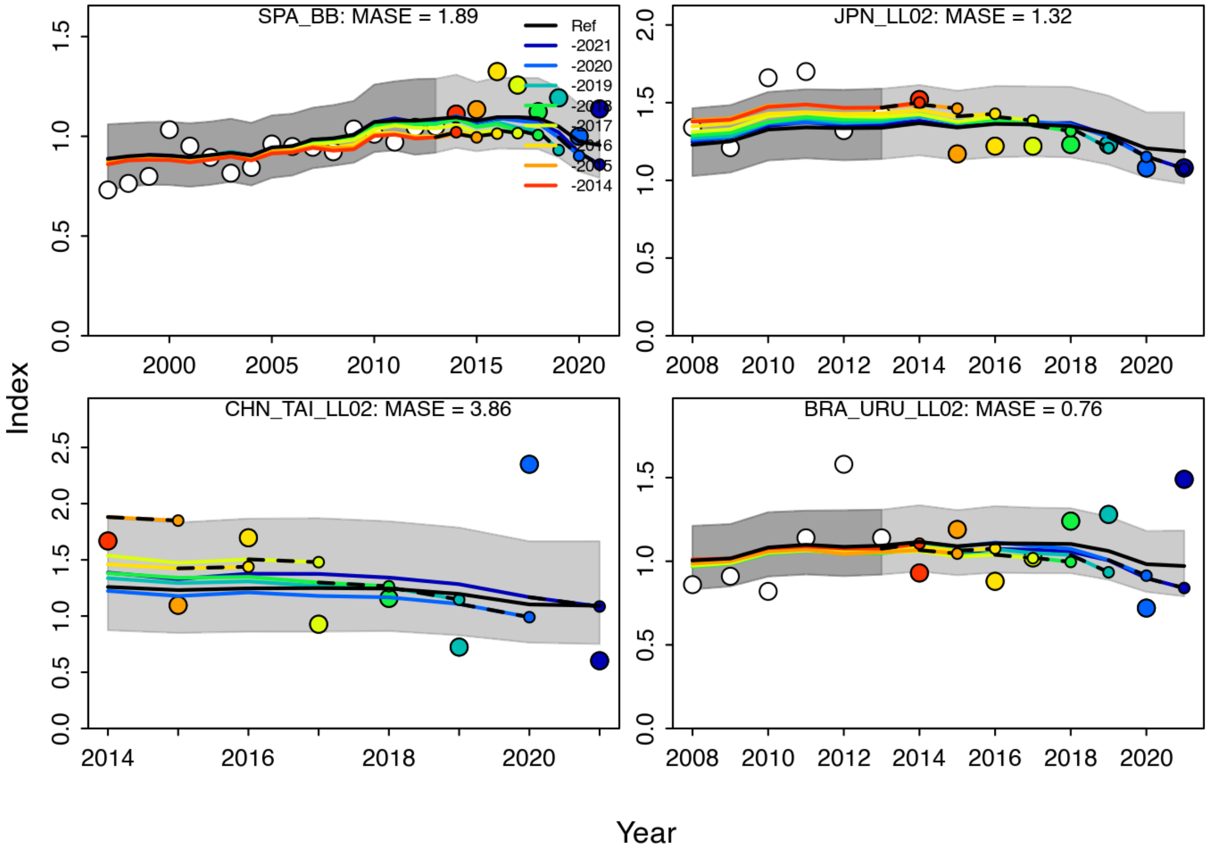
(a) Original weighting



(b) Courtney *et al* (2016)

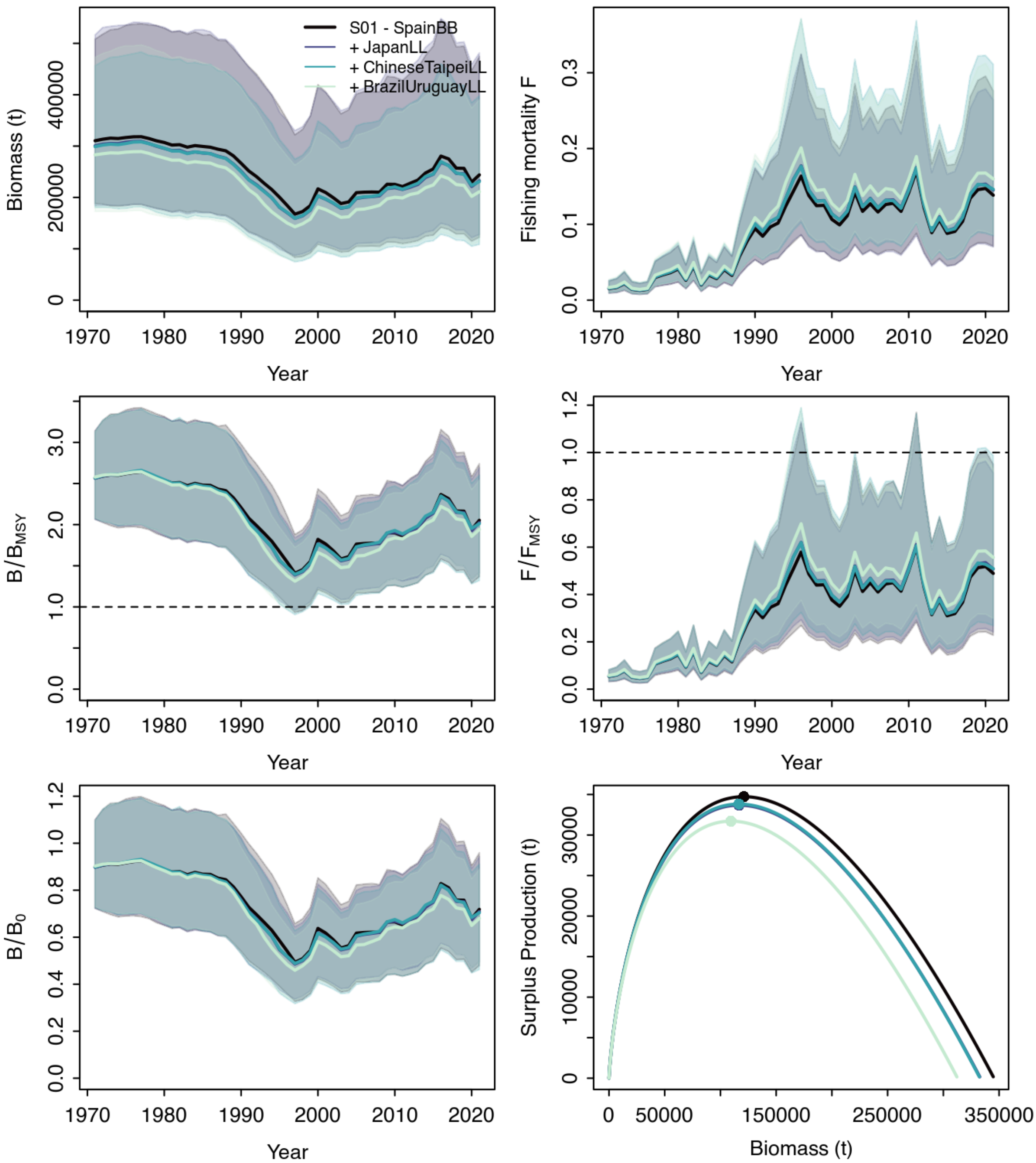


(c) Model-internal weighting

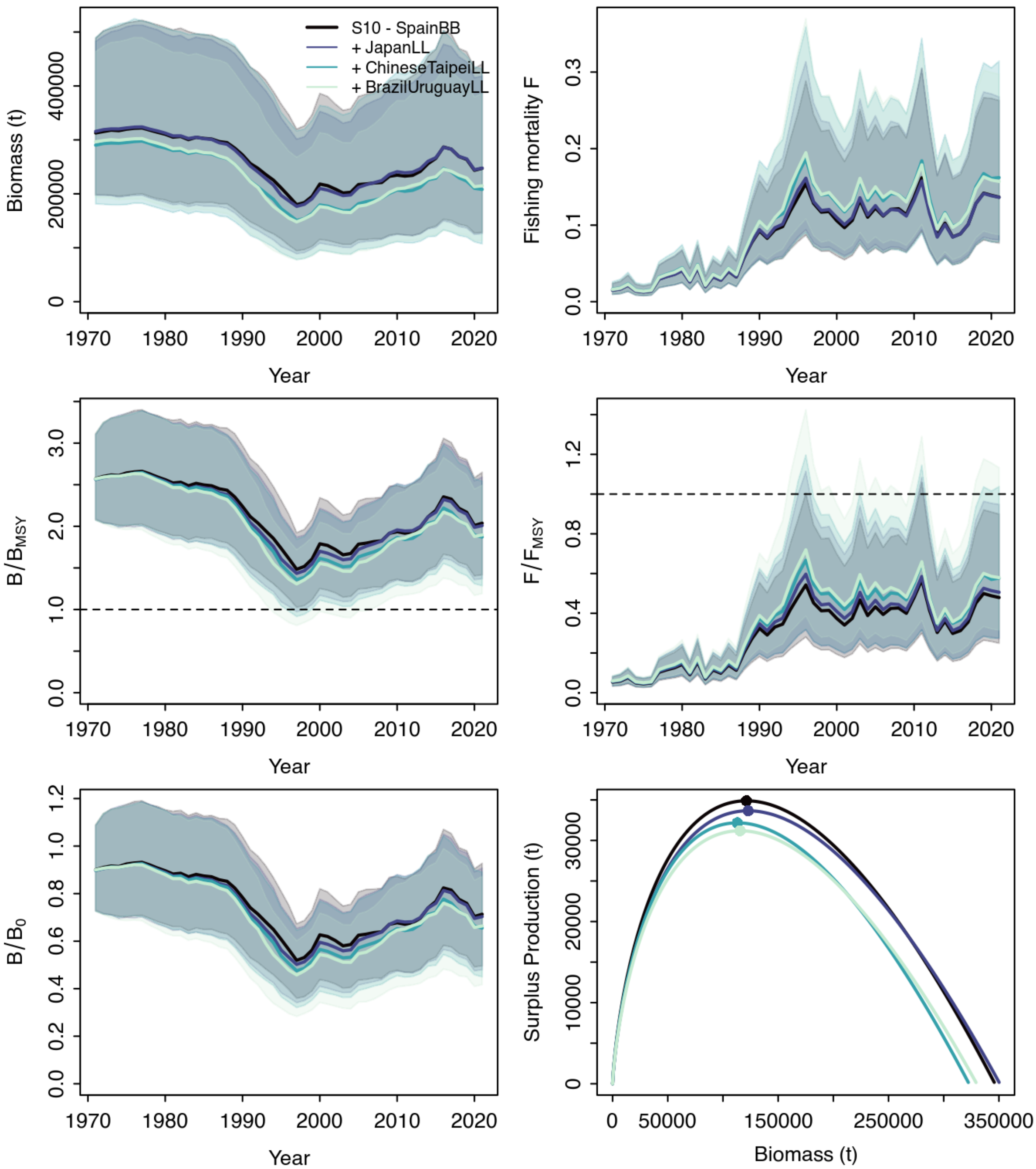


(d) Model-internal weighting - Time block

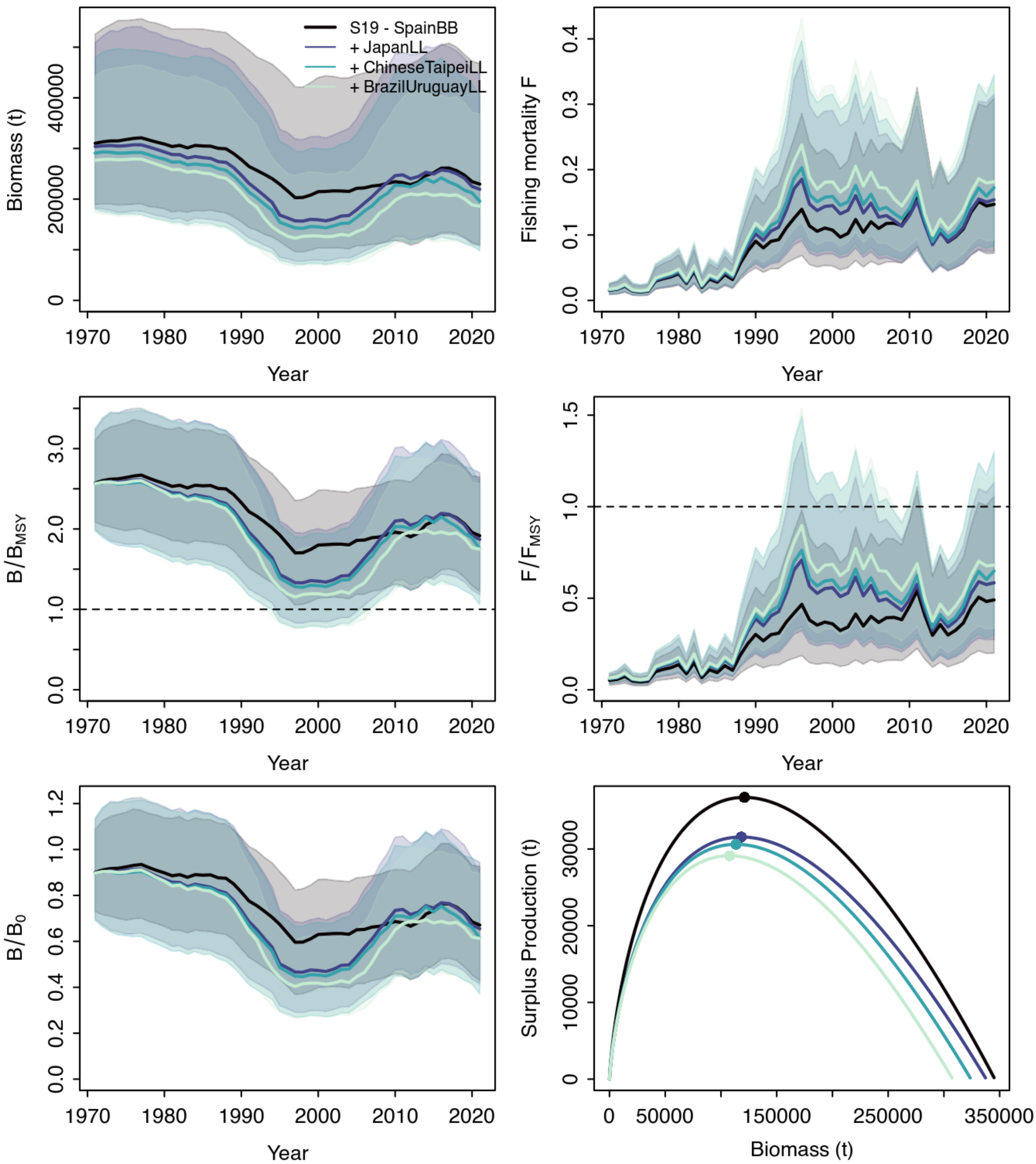
Figure 10. Hindcasting cross-validation results (HCxval) for the first scenario of each model-weighting process tested for South Atlantic blue shark, showing one-year-ahead forecasts of CPUE values, performed with eight hindcast model runs relative to the expected CPUE. The CPUE observations, used for 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).



(a) Original weighting

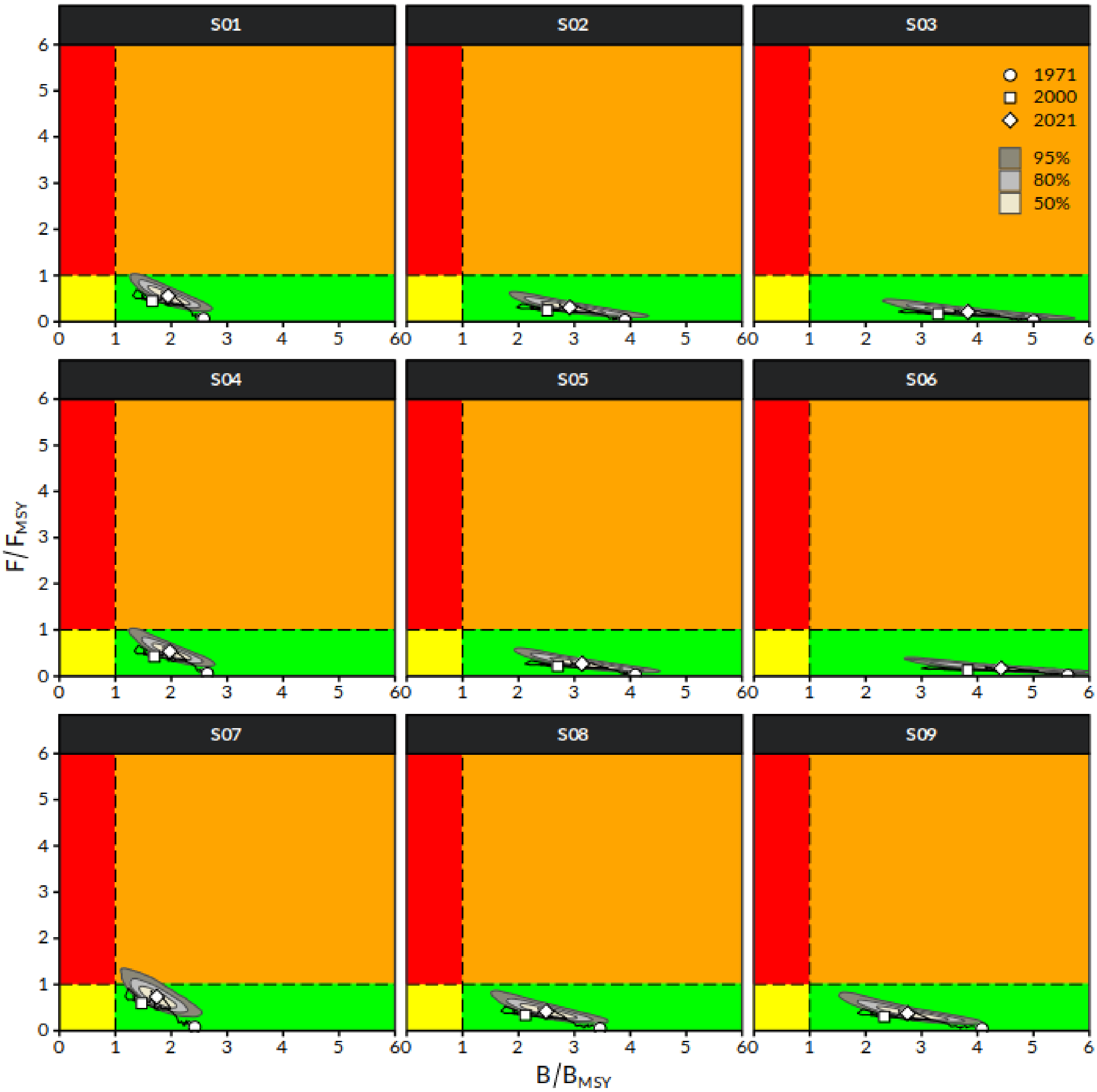


(b) Courtney *et al* (2016)

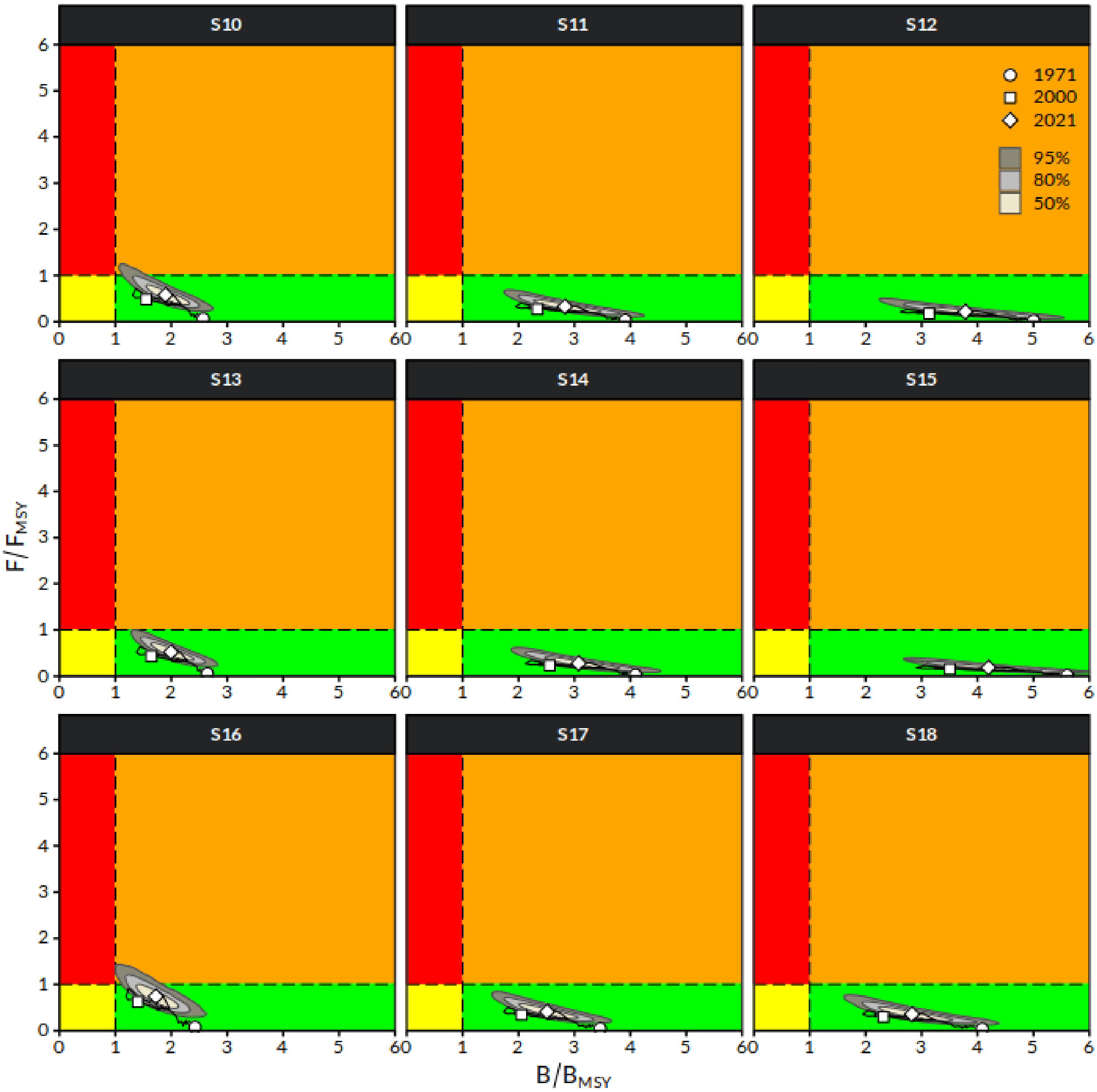


(c) Model-internal weighting

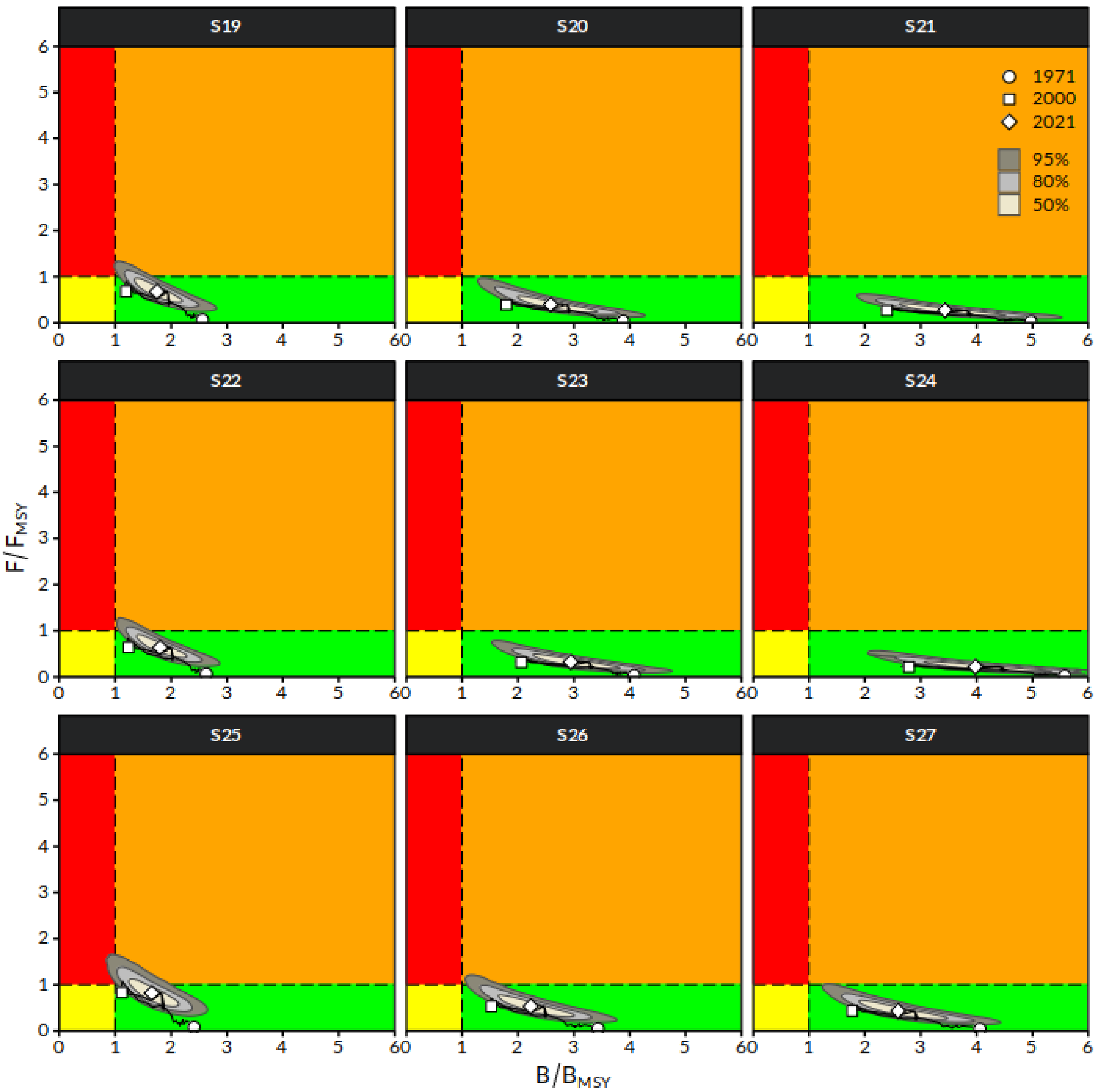
Figure 11. Sensitivity analysis performed for the first scenario of each model-weighting process tested for South Atlantic blue shark showing the trends in biomass and fishing mortality (upper panels), biomass relative to B_{MSY} (B/B_{MSY}) and fishing mortality relative to F_{MSY} (F/F_{MSY}) (middle panels) and biomass relative to K (B/K) and surplus production curve (bottom panels).



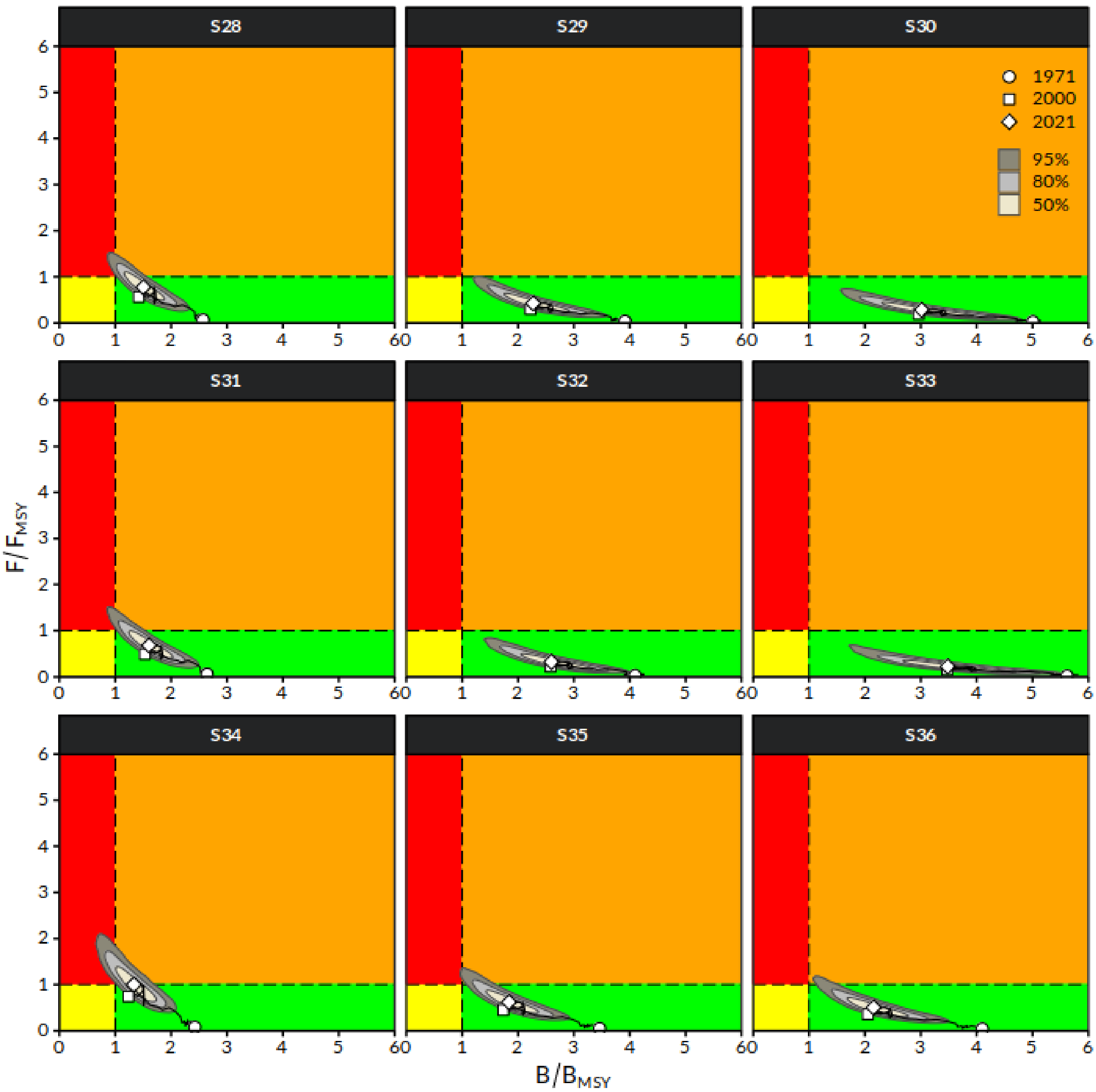
(a) Original weighting



(b) Courtney *et al* (2016)



(c) Model-internal weighting



(d) Model-internal weighting - Time block

Figure 12. Kobe phase plot showing estimated trajectories (1971-2022) of B/B_{MSY} and F/F_{MSY} for the Bayesian state-space surplus production model for the South Atlantic blue shark. 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.