INTER-AMERICAN TROPICAL TUNA COMMISSION

SCIENTIFIC ADVISORY COMMITTEE

14TH MEETING

La Jolla, California (USA)
15-19 May 2023

DOCUMENT SAC-14-05

EXPLORATORY ANALYSIS FOR THE STOCK ASSESSMENT OF BIGEYE TUNA IN THE EASTERN PACIFIC OCEAN

Haikun Xu, Mark N. Maunder, Carolina Minte-Vera, and Cleridy Lennert-Cody

CONTENT

EXECUTIVE SUMMARY

1. This document outlines six major modifications proposed by the staff to improve the stock assessment models for bigeye tuna in the eastern Pacific Ocean. These changes fall into three categories: fishery definitions, survey fleet characteristics, and fishery fleet characteristics.
2. The new "base" reference assessment model for bigeye tuna in the eastern Pacific Ocean is considered superior to the "base" reference assessment model from the last benchmark assessment based on a variety of model diagnostics:
   a. Improved fit to length composition data
   b. Improved fit to the index of abundance
   c. Included an additional source (Korean) of data in the calculation of longline length compositions
   d. Estimated more realistic initial conditions
   e. Reduced conflict between the information from the index of abundance and from length compositions about the scale of population abundance
   f. Reduced magnitude of the regime shift in recruitment (from 2.4 to 1.5)
   g. Showed very similar scales and trends of absolute abundance estimated by the new base model and its age-structured production model
3. The proposed improvements in this exploratory analysis show potential in significantly reducing or even resolving the bimodal pattern in model-combined joint distributions of management quantities. Specifically, these improvements result in more optimistic estimates of terminal year depletion for the pessimistic group of assessment models, and more pessimistic estimates of terminal year depletion for the optimistic group of assessment models.
4. It is essential to note that this exploratory analysis is considered preliminary and should not be used as the basis for providing any management advice.
5. The staff has identified several desirable research projects that can be conducted before the 2024 benchmark assessment to further improve the stock assessment of bigeye tuna in the eastern Pacific Ocean.

1. INTRODUCTION

Bigeye tuna (*Thunnus obesus*) is a highly-migratory species inhabiting tropical and temperate waters of the Pacific, Atlantic, and Indian Oceans (Collette *et al.* 2001). They are fished by various methods in the eastern Pacific Ocean (EPO). Bigeye tuna is the main target species of longline fisheries in the EPO since the 1970s, owing to its high commercial value in the global sashimi market (Matsumoto 2008). Prior to 1993, the distant water longline fishery was the primary method of harvesting bigeye tuna in the EPO, with an average annual catch of 88,000 metric tons during 1985-1992 (IATTC 2021). In contrast to longline fisheries that catch primarily large and mature bigeye, purse-seine fisheries catch mostly small and immature bigeye (Okamoto and Bayliff 2003, Xu *et al.* 2020). The three main types of purse-seine fisheries in the EPO include sets made on free-swimming tuna schools, on tunas associated with dolphin herds (primarily large yellowfin tuna), and on tunas associated with floating objects (Lennert-Cody and Hall 2000, Maunder and Harley 2006). Of the three types, bigeye tuna in the EPO is most vulnerable to the purse-seine sets on floating objects (hereinafter referred to as the OBJ fishery), which prior to 1993 were a coastal fishery based mostly on natural objects such as tree trunks and kelp paddies (Lennert-Cody and Hall 2000). The purse-seine fishery on floating objects during this period caught about five thousand metric tons of bigeye annually ,which is much lower than the level of longline catches (IATTC 2021).

With the rapid development of Fish Aggregation Devices (FADs) since 1993, the OBJ fishery gradually replaced longline fishery as the dominant fishery type for bigeye tuna in the EPO (IATTC 2021, Xu *et al.* 2020). FADs are man-made floating objects placed in the water to attract tunas. They are commonly

equipped with an echo-sounder to measure fish abundance and a GPS to report geographic location (Hall and Roman 2013). The OBJ fishery, which catches small bigeye tuna, has expanded substantially and rapidly in the tropical EPO from the coastal waters of the American continent to beyond the west management boundary (150°W) of the IATTC (Lennert-Cody and Hall 2000). This expansion of the OBJ fishery has a strong negative impact on the longline fisheries that catch large bigeye tuna of the same stock (Matsumoto 2008, Okamoto and Bayliff 2003, Sun *et al.* 2019). Specifically, the longline catch of bigeye tuna in the EPO has declined significantly from 88% in 1993 to a historical low level of 23% in 2020 (IATTC 2021).

The last benchmark assessment for bigeye tuna in the EPO was conducted in 2020 (SAC-11-06). This benchmark assessment represents a new approach to providing management advice and serves as the basis for a risk analysis (SAC-11-08). The new risk analysis methodology uses several reference models that represent various plausible states of nature (assumptions) about the biology of the fish, the productivity of the stocks, or the operation of the fisheries, effectively incorporating uncertainty into the management advice. In total, 48 reference models were developed for this benchmark assessment within a hierarchical framework to address three major uncertainties from the previous assessment. These uncertainties include the apparent regime shift in recruitment, the misfit to the length-composition data for the longline fishery that is assumed to have an asymptotic selectivity, and the steepness of the stock-recruitment relationship. With this risk analysis approach, the staff can explicitly evaluate the probability of breaching the reference points defined in the IATTC's harvest control rule for tropical tunas (C-16-02).

The last benchmark assessment for bigeye tuna in the EPO highlights a concern regarding the bimodal pattern observed in management quantities. The reference models used for bigeye tuna can be divided into two groups based on management quantities related to maximum sustainable yield (MSY): pessimistic and optimistic. The large difference between the MSY-related management quantities of these two groups results in model-combined joint distributions of management quantities, such as F/F_{MSY}, to have two distinct modes that are separate from each other. This bimodality presents two possible states of nature and can be interpreted such that the fishing mortality for bigeye should be greatly increased or greatly decreased from the recent level to achieve the target reference point. Moreover, the risk analysis indicated that neither of the two scenarios is significantly more likely than the other, making it challenging to provide effective management advice.

This report presents the outcomes of an exploratory analysis conducted to improve the reference models for the next benchmark stock assessment for bigeye tuna in the EPO. Since the last benchmark stock assessment, six improvements have been made to the reference models, which can be grouped into three aspects: fisheries definitions, survey fleet characteristics (including index of abundance and the corresponding length compositions), and fisheries characteristics (including selectivity assumptions and length compositions). In Section 2, we describe each improvement and its rationales in detail. In Section 3, we present the impact of each improvement on model estimates separately in a stepwise manner. In Section 4, we conduct a series of model diagnostics to comprehensively evaluate the performance of the new "base" model for bigeye tuna in the EPO. In Section 5, we present the effects of the six improvements on model-ensemble results, aiming to assess whether and to what extent the six improvements presented in this exploratory analysis can reduce the bimodality in management quantities. In Section 6, we present a list of desirable research projects that can lead to further model improvements, which the IATTC staff plans to undertake in preparation for the upcoming 2024 benchmark assessment.

2. IMPROVEMENTS MADE IN THIS EXPLORATORY ANALYSIS

This section provides a detailed description of the six improvements we have made to the reference models for bigeye tuna in the EPO, which can be classified into three categories: fishery definitions, survey fleet characteristics, and fishery fleet characteristics. First, we update the fishery definitions for both

longline and purse-seine fisheries by using *FishFreqTree*[1], an open-source R package recently developed at the IATTC to automate the regression tree analysis for making fishery definitions. Second, we improve the methodology applied to the standardization of the abundance index as well as the associated length composition for the survey fleet. Third, we make three modifications to the longline fishery fleets in the assessment model including 1) incorporating Korean longline data to compute joint (Japan + Korea) length compositions for longline fishery fleets; 2) adding another time block, which began in 2011, to longline fishery selectivity curves; and 3) standardizing the length compositions for longline fishery fleets by developing a length-specific spatiotemporal model.

2.1. Fishery definitions

Same as the assessment models considered in the last benchmark assessment, the exploratory assessment models are not spatially-structured and use the "areas-as-fleets" approach, which models geographic areas as separate fleets with different selectivity curves in a single-stock assessment model. Although this approach implicitly assumes that the stock is homogenously distributed throughout its range and any differences in composition data arise due to different contact selectivity (Hurtado-Ferro *et al.* 2014), it recognizes that fishing in different areas usually leads to different ages/sizes of fish being removed from the population due to spatial variation in age/size structure. As such, fisheries need to be defined spatially with the goal of achieving homogeneous distribution of fish across each area. This approach ensures that the length composition of each fishery is not influenced by the location of fishing activities (Punt 2019).

We use a regression tree approach for analyzing length frequency data to provide gear and set type-specific fishery definitions. The regression tree algorithm (Lennert-Cody *et al.* 2013, Lennert-Cody *et al.* 2010) uses recursive partitioning to search for hierarchical binary decision rules that divide the data into more homogeneous subgroups. The binary decision rules are selected to provide the greatest decrease in the heterogeneity of length composition data, which is measured based on the Kullback–Leibler divergence. The regression tree algorithm has been recently included in an R package *FishFreqTree*, where fisheries length-frequency data, separated by gear (longline/purse-seine) and purse-seine set type (floating object/unassociated/dolphin), are grouped by latitude, longitude, quarter, and cyclical-quarter.

There are two main differences between the regression tree analysis conducted for the last benchmark assessment and this exploratory analysis. The previous analysis is based on both catch-per-unit-effort (CPUE) and length frequency with the objective of finding compromised spatial boundaries across gear and set type. In contrast, this analysis is based solely on length frequency and is conducted for each gear and set type to provide uncompromised gear and set type-specific fishery definitions. The habitat preference of bigeye tuna is size-specific, so fish caught by different gear types (longline catches predominately adult bigeye and purse-seine catches predominantly juvenile bigeye) are likely to have distinct spatial patterns of age/size composition. As such, independent fishery definitions are more appropriate for this assessment model that utilizes the "areas-as-fleets" approach.

The second difference is the source of longline composition data. The regression tree analysis conducted for the last benchmark assessment is based partially on the longline length composition data that Japan submitted to the IATTC's public domain. This data is coarse and pre-aggregated by 5° latitude, 10° longitude, and 1 quarter (WSBET-02-02). In this exploratory analysis, the new longline length composition data submitted to the IATTC by Japan through a Memorandum of Understanding is used. This data has a much finer spatial and temporal resolution (1° latitude, 1° longitude, and month) and includes additional useful information, such as the bin size associated with each length measurement.

[1] https://github.com/HaikunXu/FishFreqTree

2.1.1. Longline fisheries

Longline fisheries are defined in this exploratory analysis using Japanese longline length composition data, which covers the period between 1986 and 2020. Before being analyzed by the regression tree algorithm, the data is filtered to include only commercial vessels' data collected at a spatial resolution of 1° x 1° and a bin size of 1 or 2 cm. Poorly sampled grids with less than four years of data between 1986 and 2020 are excluded from the dataset. The remaining data is aggregated by 5° x 5° spatially and one quarter temporally into fifteen length bins (<70 cm, 70-80 cm, 80-90 cm, ..., >200 cm).

We specify the regression tree algorithm to define four splits or five longline fisheries for the EPO except Hawaii, where a separate longline fishery is defined as in the previous benchmark assessment model. The regression tree is hierarchical and may exhibit a certain degree of instability. Instead of selecting only the best candidate for each split, we consider the top four and two competing candidates for the first and second splits, respectively, and rank the eight (4 x 2) 4-split combinations according to the proportion of variance in the length-frequency data explained.

Among the eight 4-split combinations, the best one selected for the longline fishery in the EPO explains 10.72% of the variance in the length-frequency data (Table 1). The first split (15°S) divides the EPO into tropical and temperate regions (Figure 1 and Table 1). The second split (105°W) divides the tropical EPO into the eastern and western portions (Figure 1 and Table 1). The third split (5°S) further separates the western tropical EPO into the northern and southern portions (Figure 1 and Table 1). The last split (90°W) divides the temperate EPO into the coastal and offshore portions (Figure 1 and Table 1). In general, the fish caught by longline are larger in tropical regions than in temperate regions and are larger in offshore regions than in inshore regions (Figure 1). The two offshore tropical regions have the largest mean size of bigeye for the longline, while the inshore region of Peru has the smallest mean size of bigeye for the longline (Figure 1).

Given that longline catches are reported to the IATTC in numbers by some fleets and in weight by others, two longline fleets, one in number and one in weight, are defined for each of the six longline fisheries. Same as the last benchmark assessment model, the exploratory assessment model includes twelve longline fishery fleets.

2.1.2. Purse-seine fisheries on floating-objects

The definition of OBJ fisheries is based on length composition data collected by port samplers from the OBJ sets made by Class-6 vessels (Suter 2010). Port samplers collect data only from wells with catch from the same set type, sampling area, and year-month. We remove the data before 2000 because the sampling protocol used by the IATTC port-sampling program changed in that year and the OBJ fishery was not fully expanded across the EPO during the 1990s. The raw data has a 5° x 5° spatial resolution and a 1 cm bin size from 1 cm to 201 cm. Poorly sampled grids with less than 4 years of data available since 2000 are removed from the dataset. The remaining data is then aggregated into fifteen 10 cm length bins (<30 cm, 30-40 cm, 40-50 cm, ..., >160 cm).

Same as in the last benchmark assessment model, the assessment model in this exploratory analysis includes five OBJ fishery fleets and therefore we specify the regression tree algorithm to include four splits. We also consider the top four and two competing candidates for the first and second splits, respectively, and rank the eight (4 x 2) split combinations according to the proportion of variance in the length-frequency data explained.

Among the eight 4-split combinations, the best one selected for the OBJ fishery in the EPO explains 13.93% of the variance in the length-frequency data (Table 2). The first three splits (110°W, 100°W, and 125°W) are all meridional (Table 2), which is consistent with the fact that the contrast in the mean size of bigeye

for the OBJ is most obvious in the East-West direction (Figure 2). Notably, the first split (110°W), which is the most important split, is identical to that selected in the last benchmark assessment. The mean size of bigeye tuna for the OBJ increases continuously from the management boundary (150°W) to the coastline, which is almost the opposite of the spatial pattern of mean size for the longline. The fourth split (15°S) divides the inshore EPO by latitude into tropical and temperate inshore regions (Figure 2 and Table 2). The temperate inshore region off Peru has the largest mean size of bigeye for the OBJ, while the westmost region adjacent to the management boundary has the smallest mean size (Figure 2).

We also conduct a sensitivity analysis on the OBJ fishery definition by fitting the same regression tree model to standardized length frequency data. Standardized length frequency is defined as the aggregated length frequency divided by the average length frequency for the year-quarter, which removes the influence of factors other than selectivity and availability on length frequency distribution. For example, the highly variable recruitment of bigeye tuna in the EPO occasionally creates strong cohorts moving through the population and may distort the length frequencies observed in the OBJ fishery. The regression tree provides an identical fishery definition for the OBJ fishery based on the standardized data, further validating the credibility of the selected fishery definition (Table 2).

2.1.3. Purse-seine fisheries on free schools

The definition of unassociated purse-seine (NOA) fisheries, as for the OBJ fisheries, is based on length composition data collected by port samplers, but from NOA sets, made by Class-6 vessels (Suter 2010). We remove the data before 2000 because the sampling protocol used by the IATTC port-sampling program changed in that year. The raw data is aggregated into fifteen 10 cm length bins (<30 cm, 30-40 cm, 40-50 cm, ..., >160 cm).

The length frequency data for NOA sets are sparse both spatially and temporally (Figure 3), and NOA sets contribute to only a small percentage of bigeye catch in the EPO. Therefore, we include only two NOA fishery fleets in the exploratory assessment model. Due to the lack of length composition data and a negligible percentage of total bigeye catch, we pool both pole-and-line and DEL sets into the NOA sets in the exploratory assessment model for convenience (but not in the tree analysis). The same simplification is made in the last benchmark assessment.

The best split selected for the NOA fishery in the EPO (i.e., 130°W) explains 9.14% of the variance in the length-frequency data (Table 3). The mean size of bigeye tuna for the NOA is generally larger in the onshore region than in the offshore region, although the spatial pattern of mean size is very noisy (Figure 3).

2.1.4. Summary

In total, twenty fisheries are defined for the exploratory assessment model for bigeye tuna in the EPO (Table 4). These fisheries comprise twelve longline fisheries, six OBJ fisheries (including one discard OBJ fishery), and two NOA fisheries. One unsolved issue identified in the last benchmark assessment is that some purse-seine fisheries, such as NOADEL-1, NOADEL-4, and OBJ-4 (Figure 5), have multiple modes in the aggregated length frequency. This is an indication that these fisheries are not well defined, which is not surprising given that the fishery definitions made in the last benchmark assessment are compromises among different gears and set types. In contrast, the fishery definitions made in this exploratory analysis are specific to gear and set type. As a result, no purse-seine fisheries in the exploratory assessment have a notable bimodal pattern in the aggregated length frequency (Figure 6), indicating that independent fishery definitions are more suitable for this assessment model. The third (south tropical offshore) longline fishery observes the highest proportion of large bigeye among the six longline fishery fleets (Figure 6) and is therefore assumed to have an asymptotic selectivity.

2.2. Survey fleet characteristics: definition

Since the last benchmark assessment, survey fleets are disconnected from the fisheries structure, total catch, and catch composition. In the EPO, there were no fishery-independent surveys of tuna abundance and size composition, with the term "survey" in this context referring to a fleet that has data (e.g., abundance index and size composition) but takes no catch (Methot and Wetzel 2013). For the "areas-as-fleets" approach on which the assessment is based, the abundance index and the associated composition data should reflect the conditions of the entire bigeye population in the EPO (Maunder *et al.* 2020a). Therefore, the abundance index for a survey fleet should be computed using an area-weighting approach for the entire spatial domain rather than for an area defined for the fishery. The composition data associated with the survey abundance index should be spatially weighted by catch rate and aggregated across the entire spatial domain as well.

In the assessment of bigeye tuna in the EPO, the survey fleet is based on fishery-dependent CPUE and length composition data collected by Japanese commercial longline vessels that persistently target bigeye tuna. Among all distant-water longline vessels operated in the EPO, Japanese longline vessels have the highest spatial coverage within the EPO and the longest history of high-quality logbook data, providing the information needed for the standardization of a reliable abundance index with a large contrast across time. In the exploratory assessment model, we revise the definition of longline survey fleet as well as the methodology used in the standardization of the abundance index and associated length compositions.

In the last benchmark assessment model, two longline survey fleets are defined based on the time of operation: 'early' (1979-1992) and 'late' (1995-2019). Catchability and selectivity are estimated separately for the two survey fleets and the coefficient of variation (CV) of the late index of abundance is fixed while that of the early index is estimated. The main reason for splitting the longline abundance index into two time periods is that gear configurations of Japanese longline vessels changed abruptly in 1993 and 1994. Specifically, both hooks-between-floats and mainline material, two key indicators of hooks' depth distribution in the water column, changed rapidly in 1993-1994. As the depth distribution of bigeye tuna in the EPO is influenced by body size (Schaefer and Fuller 2010), these notable changes in gear configurations may lead to a temporal change in catchability and selectivity for the survey fleet.

The current good practices for CPUE modeling advise against splitting the abundance index by time into separate non-overlapping time blocks unless there is clear evidence against constant survey catchability and selectivity (Hoyle *et al. in prep*). Splitting the abundance index by time wastes a large amount of information in the CPUE data, particularly the continuous trend of population abundance over a long period. Hoyle *et al.* argue that if the assessment model is misspecified, splitting the abundance index can introduce bias as the model may not be able to reliably scale the split abundance indices. Thus, analysts should at least consider whether the estimated change in catchability at the split makes sense. Regarding this point, we revisit the survey definition in the exploratory assessment by checking the estimated change in catchability and selectivity at the split.

Indeed, the last benchmark assessment model estimates similar catchability and selectivity for the early and late survey fleets. The estimated catchability for the early period (1.58 ± 0.39) is slightly higher than that for the late period (1.34 ± 0.13). The selectivity curves estimated for the two time periods are also closely aligned (Figure 7). This result is contrary to expectations, as the catchability of the main target species (Japanese longline fishery in the EPO persistently targets bigeye tuna) tends to increase over time due to continuous improvements in fishing technology and knowledge. This counterintuitive result suggests that the assessment model is likely mis-specified and is unable to reliably scale the two abundance indices. Consequently, an analysis is conducted to evaluate the sensitivity of model results to the decision of whether to split the abundance index by time.

2.3. Survey fleet characteristics: standardization methodology

Indices of relative abundance are a crucial input to stock assessment models as they directly inform the changes in population abundance over time (Francis 2011). Ideally, indices of abundance should be calculated using fishery-independent survey data, collected using the same fishing gear and operation across time to assure constant catchability and selectivity, and have a random or fixed sampling design in space. However, for most tuna species worldwide, including bigeye tuna in the EPO, survey data are not available. Therefore, indices of abundance are derived solely from fishery-dependent CPUE data. These data need to be standardized so that the abundance index is approximately proportional to population abundance (Maunder and Punt 2004). To achieve this, the standardization model needs to remove the part of the variation in the CPUE data that is not driven by changes in population abundance. Furthermore, the standardization model should impute fish abundance for unfished locations and use an area-weighting approach to compute the abundance index for the population for the entire spatial domain of the fishery (Thorson *et al.* 2015).

We use a delta-generalized linear mixed spatiotemporal model VAST (Thorson and Barnett 2017) to standardize the Japanese longline CPUE data for bigeye tuna in the EPO. VAST is an open-source R package[2] that has recently gained increasing popularity in standardizing fishery-dependent CPUE data for tunas (Ducharme-Barth *et al.* 2022, Maunder *et al.* 2020b, Satoh *et al.* 2021, Xu *et al.* 2019). Fishery-dependent CPUE data, including those for tunas, are not randomly distributed in space. They tend to concentrate in areas with high fish abundance or easy-to-fish locations (a phenomena referred to as preferential sampling), and do not cover the entire fishing ground within a quarter or even a year. VAST can estimate spatial and temporal correlations and use that information to impute fish abundance for unfished locations using neighboring CPUE data.

To standardize longline CPUE for the survey fleet, VAST separately models encounter probability and positive catch rate to account for zero-inflated catch rate observations. Specifically, we specify VAST to use the logit and log link functions for the linear predictors of encounter probability and positive catch rate, respectively. Both linear predictors include an intercept (year-quarter) term, a time-invariant spatial term, a time-varying spatiotemporal term, a catchability covariate (using HBF as a 3-knot spline) term, and a vessel effects term. Of these five terms, the intercept term and the catchability covariate term are estimated as fixed effects and the other three terms are estimated as random effects. This VAST model treats the four quarters equally (no seasonal component), consistent with the "quarters-as-years" approach used in the stock assessment model.

In addition to CPUE data, there is a need to standardize the composition data associated with the abundance index (Maunder *et al.* 2020a). In general, only a small portion of CPUE data has the corresponding composition data, indicating that composition data is distributed more sparsely in space than CPUE data. Since the composition data for the survey fleet should represent the condition of the entire population, a length-specific spatiotemporal model is needed to impute length frequency for unsampled locations and compute area-weighted CPUE-raised length frequencies for the survey fleet.

To standardize longline length-specific CPUE for the survey fleet, VAST models encounter probability and positive catch rate separately to account for zero-inflated length-specific catch rate observations. Specifically, we specify VAST to use the *logit* and *log* link functions for the linear predictors of encounter probability and positive catch rate, respectively. Both linear predictors include an intercept (year-quarter) term, a time-invariant spatial term, and a time-varying spatiotemporal term. Of these three terms, the intercept term is estimated as fixed effects and the other two terms are estimated as random effects. Neither catchability covariate (HBF) term nor vessel effects term is included in this model because they

[2] https://github.com/James-Thorson-NOAA/VAST

are not available in the Japanese longline length composition data. This VAST model treats the four quarters equally (no seasonal component) to be consistent with the "quarters-as-years" approach used in the stock assessment model.

The CPUE data for the standardization are collected from Japanese commercial longline vessels and aggregated at a resolution of 1° cell x month x vessel x HBF from an operational-level database. Japanese commercial longline vessels also provide associated composition data, which includes both length and weight compositions and is reported at various spatial resolutions and bin sizes. For the survey fleet, we select only the length composition data with a 1° x 1° spatial resolution and a 1 or 2 cm bin size.

2.3.1. The index of abundance

The first change implemented in the standardization methodology for the survey fleet is about the spatial domain on which the standardizations of the abundance index and the associated length frequencies are based. In the last benchmark assessment, the spatial domain is restricted to the "core" longline fishing ground, which includes only the 1° x 1° cells with at least 80 quarters of CPUE data between 1979 and 2019 (Figure 8). This was done to address the concern that the marked westward contraction of the Japanese longline fishing ground in the past decade may result in biased index for those years. By fitting the CPUE standardization model to data collected only from the core fishing ground, the potential impact of biased spatial imputation of fish abundance for unfished locations on the accuracy of the standardized abundance index was reduced.

Findings in recent studies (Xu *et al. in prep*) suggest that restricting the CPUE standardization to the core fishing ground, where depletion rate is relatively slow, likely leads to a hyper-stable abundance index for bigeye tuna in the EPO. In the past two decades, an obvious local depletion of the bigeye tuna population has been observed in the eastern EPO. Catch rates of bigeye in both longline (Xu *et al. in prep*) and OBJ (FAD-05 INF-D) fisheries have decreased pronouncedly faster in the tropical fishing ground east of 110°W than west of 110°W. During the same period, Japanese longline vessels gradually retreated from the eastern fishing ground, which is relatively data poor and excluded from the core fishing ground (Figure 8). Thus, the abundance index standardized for bigeye in the less-depleted core fishing ground does not reflect the population trend at the EPO-wide level. It likely underestimates the rate at which the bigeye population in the EPO decreased over time. In this exploratory assessment, we broaden the definition of the core fishing ground to include the 1° x 1° cells with at least 20 quarters of CPUE data between 1979 and 2019, allowing the eastern EPO to be included in the spatial domain for the CPUE standardization (Figure 8). As expected, the abundance index estimated based on the new spatial domain decreases faster between 1979 and 2019 (Figure 9).

The second change we make in the standardization methodology for the survey fleet is the assumption on which the imputation of fish abundance for unfished locations is based. Given that the spatial domain extends beyond the previous core fishing ground to encompass locations with relatively sparse CPUE data, the abundance index for this exploratory assessment is subject to greater influence by imputed fish densities for unfished locations. As such, it is crucial to address potential biases associated with the imputation process, particularly in this case where fishery-dependent CPUE data is preferentially sampled. Most CPUE standardization models, including the one we use in this exploratory assessment, cannot explicitly account for preferential sampling in the imputation process. Ignoring preferential sampling in fishery-dependent CPUE data results in positive bias in imputed fish density for unfished locations. As the extent of unfished locations expands over time due to the depletion-induced contraction of the Japanese longline fishery, the positively-biased imputation plays an increasingly more important role in the area-weighted abundance index, leading to a hyper-stable abundance index.

The spatiotemporal term, which describes how the spatial pattern of fish density changes over time, needs to be interpolated for each location and time. In the CPUE standardization model developed for the last benchmark assessment, the spatiotemporal term is assumed to be temporally independent but spatially correlated according to the Matérn function. Thus, the spatiotemporal terms for the unfished eastern EPO are interpolated solely based on data collected from the fished western EPO during the same year-quarter. This approach ignores the concurrence of local depletion and preferential sampling, leading to positively biased imputations of bigeye density in the eastern EPO. To achieve more realistic imputations of bigeye density for the eastern EPO, we modify the assumption for the spatiotemporal terms to be correlated in both space and time. Specifically, the spatiotemporal terms are now assumed to follow a random-walk process in time to capture the directional change in the spatial distribution of bigeye abundance over time (the pronounced local depletion pattern). Under this assumption, the spatiotemporal terms for the unfished eastern EPO are interpolated based on data collected not only from the fished western EPO in the same year-quarter but also from the eastern EPO in adjacent fished years. The spatiotemporal dynamics of bigeye density predicted by the improved CPUE standardization model demonstrates the evolution of bigeye depletion in the EPO (Figure 10). A recent simulation study conducted by the staff (Xu *et al. in prep*) shows that this assumption leads to a less-biased abundance index for bigeye tuna in the EPO than the previous assumption. As expected, the abundance index estimated based on this assumption indicates a more pessimistic population trend than the abundance index estimated based on the previous assumption (Figure 9).

2.3.2. Composition data associated with the index of abundance

In this exploratory analysis, the standardization methodology is also improved for the composition data associated with the abundance index. The length composition data for the survey fleet needs to be spatially weighted by CPUE. To ensure consistency with the CPUE standardization model, the spatial domain of the length-specific spatiotemporal model for composition standardization is also expanded (Figure 8). Another important change has been made to the computation of length composition data for the survey fleet. Previously, the length-specific model was fit to length-specific CPUE data obtained by matching Japanese longline CPUE data with Japanese longline length frequency data. Length frequency data are much sparser than CPUE data (Figure 11). In fact, less than 30% of CPUE grids contain length frequency data in any year-quarter, with a historical average of less than 15% (Figure 12). This data-matching procedure results in a substantial amount of CPUE data being wasted. To overcome this limitation, the new length-specific spatiotemporal model is fit to length frequency (i.e., proportion) data only. The predicted spatiotemporal field of length frequency from this model is subsequently matched with the predicted spatiotemporal field of fish density from the CPUE standardization model. This new approach includes all available CPUE data, instead of relying on average less than 15% of CPUE data, to provide a more accurate and precise spatiotemporal distribution of bigeye density for the purpose of weighing length frequencies for the survey fleet.

2.4. Fishery fleet characteristics: the source of longline composition data

In the last benchmark assessment, the computation of length composition data for longline fishery fleets relies solely on length composition data from Japanese commercial longline vessels. However, concerns have been raised about the representativeness of the Japanese longline length composition data collected in recent years. The contribution of Japanese longline catch to the total longline catch has continuously decreased over time from nearly 100% before 1985 to less than 25% since 2017 (Figure 13). Furthermore, both the spatial coverage (Figure 14) and the sample size (Figure 15) of the longline length composition data from Japan has decreased notably in the 2010s. As the composition data for fishery fleets should be weighted spatially by catch amount, it is reasonable to expand the source of composition data for longline fishery fleets to other CPCs.

In this exploratory assessment, we include Korean longline length composition data to provide joint length frequencies for longline fishery fleets. There are several reasons for choosing Korea as the additional source of composition data for longline fishery fleets. Firstly, Korea has recently replaced Japan as the most important longliner for bigeye tuna in the EPO (Figure 13). Secondly, Korea routinely submits high-resolution (1° x 1°) length composition data to the IATTC database. Lastly, a large portion of the grids where Korean length composition data are available are not covered by Japanese length composition data (Figure 14). Expanding the data source to multiple fleets allows for a more complete spatial coverage of the longline fishing ground in the EPO. The sample size of the Korean length composition data before 2011 is negligible, so only the data from 2011 onward are included in the computation of joint length frequencies for longline fisheries fleets. The longline fisheries corresponding to the second and third longline areas (see Figure 4) are the two most important longline fisheries with respect to catch amount. Since 2011, Korean length composition data collected are primarily for these two longline fisheries, contributing to the majority of sample size in the second longline area and around half of the sample size in the third longline area (Figure 15).

An algorithm[3] is developed to compute the joint length composition data for longline fishery fleets. First, we separately filter the raw length composition data from Japan and Korea by removing the 5° x 5° x quarter strata with a sample size of less than sixteen. We then weight country-specific length frequencies according to the proportion of catch in the same strata to obtain strata-specific joint length frequencies. Finally, these length frequencies are spatially weighted by total (sum across all countries) bigeye catch in numbers for each longline area, deriving catch-weighted joint length frequencies for each longline fishery fleet. The joint length frequencies are noticeably different from the Japanese length frequencies for the two most important longline fisheries. This difference could result from the two fleets having different contact selectivity or operating in different locations. Overall, the joint length frequencies suggest a higher proportion of the largest bigeye (Figure 16).

2.5. Fishery fleet characteristics: time blocks for longline selectivity

The selectivity curves for some, if not all, fishery fleets are subject to change over time. The regression tree algorithm on which fishery definitions are based is designed to find the greatest decrease in the heterogeneity of length composition data by dividing the data into more homogeneous subgroups. However, fishery length compositions are spatially weighted by catch amount. If the spatial distributions of fish abundance and fishing intensity vary over time within a fishery area, the fishery selectivity for that area can potentially be time-varying.

In the last benchmark assessment, all longline selectivity curves are estimated separately for two time periods: 1979-1993 and 1994-2019. The main reason for splitting all longline selectivity curves into two time blocks is the sudden change in longline fishing gear between 1993 and 1994, including an increased HBF and an upgrade of mainline material. In this exploratory assessment, we add another time block separated by 2011 to all longline selectivity curves due to the following three reasons: 1) longline length composition data are from one source (Japan) before 2011 and two sources (Japan and Korea) since 2011; 2) the composition of longline catch by CPC changes notably since 2011, in particular, the contribution of Japanese longline catch dropped rapidly since 2011; and 3) the spatial distribution of longline catch shifted sharply since 2011 (Figure 17). In fact, the aggregated fisheries length frequencies since 2011 are pronouncedly different from those before 2011, especially for the two most important longline fisheries 2 and 3 (Figure 18). This evidence further supports splitting the selectivity of longline fisheries temporally by 2011.

[3] https://github.com/HaikunXu/IATTCassessment/blob/master/R/ll_fisheries_lf_joint.R

2.6. Fishery fleet characteristics: the methodology of computing longline length frequencies

The methodology for computing length composition data for longline fishery fleets has been improved. In the last benchmark assessment, length composition data for longline fishery fleets are computed by spatially raising raw length compositions to catch amount. This methodology has a significant limitation, as a large proportion of longline catches do not contribute to the computation of length frequencies for fishery fleets. This is due to the sparse distribution of longline length composition data in space (Figure 14). This data only covers a small percentage of the grids with positive catches, with less than 15% of the grids with positive Japanese catch having associated length compositions, on average, as shown in Figure 12. As a result, length frequencies computed by raising raw length compositions spatially to catch may not adequately represent fishery removal.

To overcome this issue, we develop length-specific spatiotemporal models to impute length frequency for the catches without corresponding length compositions. This new approach allows the computation of length frequencies for longline fishery fleets based on all, rather than a small percentage, of longline catches. The joint longline length frequencies are based on data collected by Japan and Korea, so two length-specific spatiotemporal models are developed and fit separately to Japanese and Korean length composition data.

The standardized length frequencies for the two countries are raised to flag-specific catch and then aggregated for each longline area, deriving standardized length frequencies for longline fishery fleets. The two length-specific spatiotemporal models have the same specifications as those for survey length compositions, and the standardized length frequencies for longline fishery fleets also have a bin size of 10 cm. The Francis weights of standardized fisheries length frequencies are around twice of those of unstandardized fishery length frequencies. Considering that unstandardized longline length frequencies are already very influential to the estimated population scale, as suggested by the R_0 likelihood profile, 0.5 × Francis weights are applied to standardized fishery length frequencies.

3. IMPACT OF THE IMPROVEMENTS ON THE NEW "BASE" MODEL

We choose one reference model (Env-Fix; see Table 2 in SAC-11-06 for model definition) from the last benchmark assessment as the platform to illustrate the impacts of those proposed improvements on model results. In this reference model we assume that there is a regime shift in recruitment after 1993, only one longline fishery fleet during 1994-2019 has an asymptotic selectivity, and both growth and natural mortality are known. To evaluate the impact of each improvement separately, the six improvements are made progressively in a stepwise manner. A total of seven models, including the chosen SAC11 reference model (Env-Fix) and six stepwise models, are compared in this section (Table 5). The impact of the six improvements in combination on other reference models' results will be shown in the next section.

The trajectory of estimated depletion level is used to evaluate the impact of the six improvements on assessment results (Figure 19). Re-defining fishery fleets alone has a negligible impact on estimated depletion (M1 vs. M0). Removing the time block in the survey abundance index and associated length composition data also has a minor effect on the estimated depletion level (M2 vs. M1). This result is consistent with our expectation based on the similarities in both catchability and selectivity between the early and late periods. The use of an improved survey abundance index and associated length composition data results in a less depleted early period and a more depleted late period (M3 vs. M2). This change in the estimated depletion trend is mainly due to the new abundance index suggesting a faster decrease in population abundance over time. Adding Korea as the additional source of composition data for longline fisheries has a minor influence on the estimated depletion level (M4 vs. M3). Adding another time block in 2011 to longline fishery selectivity curves leads to pronouncedly more optimistic depletion estimates

(M5 vs. M4) across time. In comparison, the use of standardized length frequencies for longline fisheries causes a notable increase in the estimated depletion level (M6 vs. M5). In summary, assessment results are most sensitive to some changes made with respect to longline selectivity and length frequencies.

4. MODEL DIAGNOSTICS FOR THE NEW "BASE" MODEL

M6 incorporates all six improvements made to the model and is the final product of the stepwise improvement process. Consequently, it is chosen from M1-M6 as the new "base" model for this exploratory analysis. To evaluate the performance of the new "base" model, several model diagnostics are conducted, including R_0 likelihood profile, retrospective analysis, and age-structured production model.

4.1. R_0 likelihood profile

Unfished recruitment (R_0), defined as the equilibrium recruitment in the absence of fishing, is a key parameter in the stock-recruitment relationship that influences the scale of absolute abundance. The profile of model likelihood (i.e., the total negative log-likelihood and its components) against R_0 is referred to as the R_0 likelihood profile (Wang *et al.* 2009). The profile of model likelihood is derived from running the new "base" model several times with a range of fixed R_0 around the maximum likelihood estimate. The R_0 likelihood profile is a diagnostic tool widely used to compare the influence of composition data and indices of relative abundance on the estimate of absolute abundance.

The R_0 likelihood profile for the new "base" model suggests the existence of data conflict between fishery and survey fleets (Figure 20). Specifically, the abundance index and the associated composition data for the survey fleet indicate a higher absolute abundance than the maximum likelihood estimate (MLE). They, however, do not include information regarding how high the absolute abundance should be. In contrast, fishery composition data support a slightly lower absolute abundance than the MLE. The recruitment penalty, which is another important component of the total likelihood, strongly penalizes an estimate of absolute abundance lower than the MLE.

4.2. Retrospective analysis

Retrospective analysis is a diagnostic tool used to evaluate the consistency of a stock assessment model from year to year (Mohn 1999). Inconsistency in model estimates from year to year as new data is added indicates the stock assessment model may be mis-specified. It is usually carried out by progressively eliminating last year's data from the assessment model without changing the model structure and assumptions. Thus, the retrospective analysis can show the effect of including new data on estimated quantities.

The retrospective analysis for the new "base" model was conducted by removing the last four quarters' data four times. The estimates of spawning biomass show a notable retrospective pattern (Figure 21), suggesting that the new "base" model is likely mis-specified. We suspect that the main reason for this pattern is the discrepancy between the assumed and empirical selectivity for the third longline fishery since 2011. This selectivity is assumed to be asymptotic but is suggested by the empirical selectivity to be dome-shaped (see F3 in Figure 22). The scale of estimated spawning biomass is heavily influenced by the composition data in association with an assumed asymptotic selectivity. Since the empirical selectivity suggests the shape of the curve is very dome-shaped, the assessment model interprets this discrepancy as the stock being very depleted. Namely, the composition data for that fishery since 2011 tends to lower the scale of estimated spawning biomass. The number of years with an assumed asymptotic selectivity is relatively small, so the estimates of spawning biomass are noticeably reduced by the inclusion of new data.

4.3. Age-structured production model

Age-structured production model (ASPM) is a diagnostic tool used to evaluate whether the stock assessment model can provide a reliable production function for the population of interest (Maunder and Piner 2014). The ASPM is built by fixing all selectivity parameters at the values estimated by the reference model and removing all length composition likelihood components from the total model likelihood. The results, particularly spawning biomass, from the ASPM without recruitment deviates are then compared with those from the reference assessment model. If the ASPM without recruitment deviates is not able to mimic the trends in indices of abundance, it could be because the stock is recruitment-driven, the reference model is not correctly specified, or indices of abundance are not proportional to population abundance (Carvalho *et al.* 2017, Maunder and Piner 2014).

The trajectories of spawning biomass estimated by the new "base" model and the corresponding ASPM without recruitment deviates show similar scales and temporal trends (Figure 23). It suggests that the new "base" model can simulate the production dynamics of the population reasonably well. The index of abundance predicted by the ASPM follows closely with the observed index (Figure 24), demonstrating that the ASPM can accurately simulate the decreasing trend in population abundance in response to increasing fishing mortality. In summary, the good performance according to the ASPM diagnostics indicates that the model can reliably estimate the scale of population abundance. The higher biomass estimated by the ASPM is consistent with the R_0 likelihood component profile for which the index of abundance supports a higher R_0.

4.4. Summary

The R_0 likelihood profile for the new "base" model indicates the presence of data conflict between the index of abundance and fishery compositions. This observation is consistent with the previous base model. The new model has a notable retrospective pattern throughout the spawning biomass trajectory, which is not observed in the previous base model. The large retrospective bias in spawning biomass is an indication of model mis-specification, which we suspect is mainly contributed by the inconsistency between the assumed (asymptotic) and expected (dome-shaped) selectivity for the third longline fishery since 2011. The negative retrospective bias might be interpreted as indicating that the new "base" model is likely over-estimating spawning biomass. However, it is likely a consequence of conflicting information about absolute abundance in different datasets. As more years of length composition data in the final selectivity time block, which has an asymptotic longline selectivity, are added to the model, the estimates of absolute abundance decline and get closer to that supported by the length composition data. The ASPM analysis suggests that the new "base" model has an accurate production function for bigeye tuna in the EPO and can reliably estimate the scale of population abundance. Conversely, the previous base model lacks a realistic production function for bigeye tuna in the EPO (see Figure A7 in SAC-08-06). In summary, the new base model is more dependable than the previous model in estimating population abundance and trend.

5. EFFECTS OF THE IMPROVEMENTS ON ASSESSMENT RESULTS

5.1. Model-specific effects

We apply all six improvements proposed in the previous section to the twelve reference models (steepness=1) included in the last benchmark assessment for bigeye tuna in the EPO. These models have a hierarchical structure based on various assumptions, of which the most important one is whether the recruitment shift is real or due to model mis-specification. A recruitment shift statistic (ratio of median recruitment after 1993 to median recruitment before 1993) was used in the last benchmark assessment to test the recruitment shift hypothesis. We first check the recruitment shift statistic for the new reference models to which the statistic is applicable (Table 6). Of the four reference models in which the recruitment

shift is assumed to be real, three estimate a recruitment shift of less than 10%, which contradicts the assumption of a recruitment shift. Therefore, we remove the three models from the list of reference models considered in this exploratory analysis (Table 7).

We then compare the nine reference models selected in this exploratory analysis with those selected in the last benchmark assessment (steepness=1) to evaluate the impact of the six improvements on each model's estimates (Figure 25). To quantitatively evaluate the impact, we use the spawning biomass ratio, which is the ratio of spawning biomass to unfished spawning biomass. In general, the six improvements lead to a higher terminal spawning biomass ratio for the five pessimistic reference models (Env-Fix, Srt-Fix, Srt-Gro, Srt-Mrt, and Srt-Sel), and lower terminal spawning biomass ratio for the four optimistic reference models (Gro, Mov, Mrt, and Sel). In other words, the new changes reduce the discrepancy between the optimistic and pessimistic reference models (Figure 26).

5.2. Model-combined effects

One of the key uncertainties identified in the last benchmark assessment for bigeye tuna in the EPO is the bimodal pattern in management quantities (SAC-11-08). Reference models can be divided into pessimistic and optimistic groups based on the estimates of MSY-related quantities. The large contrast in MSY-related management quantities causes the model-combined joint likelihood profiles for management quantities to demonstrate two distinct modes. This bimodality in management quantities suggests that there are two possible states of nature, implying that the current fishing mortality for bigeye tuna in the EPO would need to be substantially increased or decreased to achieve the target reference points. Moreover, the risk analysis for bigeye tuna indicates that neither of the two states of nature is significantly more likely, making the provision of management advice even more challenging.

To investigate whether the bimodal pattern has been reduced in this exploratory analysis, we compute the model-combined joint distribution of terminal year depletion (i.e., spawning biomass ratio) based on the nine selected reference models for this exploratory analysis. However, it should be noted that this joint distribution is not suitable for providing management advice since it is not derived from a formal risk analysis that weights model estimates based on hypotheses and model diagnostics. For the joint distribution, the reference models are equally weighted under each overarching recruitment hypothesis. The primary objective of presenting the joint distribution is to demonstrate the level of divergence in model estimates, which can inform us about whether the bimodal pattern has been reduced due to the improvements made to the assessment models.

The bimodality in the joint distribution of terminal year depletion has been significantly reduced in this explorative analysis (Figure 27). This result is not surprising given that the improvements made in this exploratory analysis have reduced the discrepancy between the optimistic and pessimistic reference models (Figure 26). The joint distribution for the last benchmark assessment shows two distinct and distant modes, one at between 0.1-0.15 and another at around 0.3. In comparison, the joint distribution for this exploratory analysis does not have two distinct and distant modes. Although there are two noticeable peaks at around 0.16 and 0.26, the overall joint distribution is unimodal-like.

6. FUTURE DIRECTIONS

The IATTC staff is committed to further improving the assessment model for bigeye tuna in the EPO before the next benchmark assessment, which is scheduled to take place in 2024. Several desirable research projects have been identified for the next benchmark assessment, including:

1. Increasing the number of longline fishery fleets. Rationale: This exploratory analysis suggests that assessment results are highly sensitive to longline composition data, and the spatial distribution of

longline catch has changed pronouncedly over time, with fishing effort becoming concentrated in a few hotspots in late years.

2. Exploring alternative ways of dealing with longline composition data from Japan and Korea. Rationale: length frequencies from these two longline fleets show slightly different patterns. It could be due to different contact selectivity, different spatiotemporal distributions of fishing effort, or the combination of the two. If the two longline fleets have similar contact selectivity, they can be combined into a joint longline fishery. Otherwise, they need to be defined as separate longline fisheries.
3. Improving the longline index of abundance by using finer CPUE data. Rationale: according to the QQ-plot, the spatiotemporal model used in the assessment has shown inadequate fit to aggregated longline CPUE data for bigeye tuna in the EPO. The IATTC staff will explore whether the model fit can be improved by using set-by-set CPUE data.
4. Improving the specification of natural mortality. Rationale: a new parametric natural mortality curve is available in Stock Synthesis, the platform on which the stock assessment of bigeye tuna in the EPO is based. This new curve can account for both decreased natural mortality as juvenile fish grow larger and increased natural mortality as female fish reach maturity (Maunder *et al*. 2023).
5. Considering other changes recommended by the results of the CAPAM workshop on tuna stock assessment good practices.

REFERENCES

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

Collette, B.B., Reeb, C., and Block, B.A. 2001. Systematics of the tunas and mackerels (Scombridae). Fish physiology **19**: 1-33.

Ducharme-Barth, N.D., Grüss, A., Vincent, M.T., Kiyofuji, H., Aoki, Y., Pilling, G., Hampton, J., and Thorson, J.T. 2022. Impacts of fisheries-dependent spatial sampling patterns on catch-per-unit-effort standardization: A simulation study and fishery application. Fisheries Research **246**: 106169.

Francis, R.I.C.C. 2011. Data weighting in statistical fisheries stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences **68**(6): 1124-1138.

Hall, M., and Roman, M. 2013. Bycatch and non-tuna catch in the tropical tuna purse seine fisheries of the world. FAO fisheries and aquaculture technical paper(568): I.

Hurtado-Ferro, F., Punt, A.E., and Hill, K.T. 2014. Use of multiple selectivity patterns as a proxy for spatial structure. Fisheries Research **158**: 102-115.

IATTC. 2021. The tuna fisheries in the eastern Pacific Ocean. Inter-Amer.Trop. Tuna Comm., 12th Scient. Adv. Com. Meeting: SAC-12-03.

Lennert-Cody, C., and Hall, M. 2000. The development of the purse seine fishery on drifting Fish Aggregating Devices in the eastern Pacific Ocean: 1992-1998.

Lennert-Cody, C.E., Maunder, M.N., Aires-da-Silva, A., and Minami, M. 2013. Defining population spatial units: Simultaneous analysis of frequency distributions and time series. Fisheries Research **139**: 85-92.

Lennert-Cody, C.E., Minami, M., Tomlinson, P.K., and Maunder, M.N. 2010. Exploratory analysis of spatial–temporal patterns in length–frequency data: An example of distributional regression trees. Fisheries Research **102**(3): 323-326.

Matsumoto, T. 2008. A review of the Japanese longline fishery for tunas and billfishes in the eastern Pacific Ocean, 1998-2003. Bull IATTC **24**: 1-187.

Maunder, M.N., Hamel, O.S., Lee, H.-H., Piner, K.R., Cope, J.M., Punt, A.E., Ianelli, J.N., Castillo-Jordán, C., Kapur, M.S., and Methot, R.D. 2023. A review of estimation methods for natural mortality and their performance in the context of fishery stock assessment. Fisheries Research **257**: 106489.

Maunder, M.N., and Harley, S.J. 2006. Evaluating tuna management in the eastern Pacific Ocean. Bulletin of Marine Science **78**(3): 593-606.

Maunder, M.N., and Piner, K.R. 2014. Contemporary fisheries stock assessment: many issues still remain. ICES Journal of Marine Science **72**(1): 7-18.

Maunder, M.N., and Punt, A.E. 2004. Standardizing catch and effort data: a review of recent approaches. Fisheries research **70**(2-3): 141-159.

Maunder, M.N., Thorson, J.T., Xu, H., Oliveros-Ramos, R., Hoyle, S.D., Tremblay-Boyer, L., Lee, H.H., Kai, M., Chang, S.-K., and Kitakado, T. 2020a. The need for spatio-temporal modeling to determine catch-per-unit effort based indices of abundance and associated composition data for inclusion in stock assessment models. Fisheries Research **229**: 105594.

Maunder, M.N., Xu, H., Lennert-Cody, C., Valero, J.L., Aires-da-Silva, A., and Minte-Vera, C.V. 2020b. Implementing reference point-based fishery harvest control rules within a probabilistic framework that

considers multiple hypotheses. Inter-Amer.Trop. Tuna Comm., 11th Scient. Adv. Com. Meeting: SAC-11 INF-F.

Methot, R.D., and Wetzel, C.R. 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research **142**: 86-99.

Mohn, R. 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES Journal of Marine Science: Journal du Conseil **56**(4): 473-488.

Okamoto, H., and Bayliff, W.H. 2003. A review of the Japanese longline fishery for tunas and billfishes in the eastern Pacific Ocean, 1993-1997. Inter-american tropical tuna commission bulletin **22**(4): 221-431.

Punt, A.E. 2019. Spatial stock assessment methods: a viewpoint on current issues and assumptions. Fisheries Research **213**: 132-143.

Satoh, K., Xu, H., Minte-Vera, C.V., Maunder, M.N., and Kitakado, T. 2021. Size-specific spatiotemporal dynamics of bigeye tuna (Thunnus obesus) caught by the longline fishery in the eastern Pacific Ocean. Fisheries Research **243**: 106065.

Schaefer, K.M., and Fuller, D.W. 2010. Vertical movements, behavior, and habitat of bigeye tuna (Thunnus obesus) in the equatorial eastern Pacific Ocean, ascertained from archival tag data. Marine Biology **157**: 2625-2642.

Sun, C.-H., Maunder, M.N., Pan, M., Aires-da-Silva, A., Bayliff, W.H., and Compeán, G.A. 2019. Increasing the economic value of the eastern Pacific Ocean tropical tuna fishery: Tradeoffs between longline and purse-seine fishing. Deep Sea Research Part II: Topical Studies in Oceanography **169**: 104621.

Suter, J.M. 2010. An evaluation of the area stratification used for sampling tunas in the eastern Pacific Ocean and implications for estimating total annual catches.

Thorson, J.T., and Barnett, L.A.K. 2017. Comparing estimates of abundance trends and distribution shifts using single- and multispecies models of fishes and biogenic habitat. ICES Journal of Marine Science **74**(5): 1311-1321.

Thorson, J.T., Shelton, A.O., Ward, E.J., and Skaug, H.J. 2015. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. ICES Journal of Marine Science **72**(5): 1297-1310.

Wang, S.-P., Maunder, M.N., Aires-da-Silva, A., and Bayliff, W.H. 2009. Evaluating fishery impacts: application to bigeye tuna (Thunnus obesus) in the eastern Pacific Ocean. Fisheries Research **99**(2): 106-111.

Xu, H., Lennert-Cody, C.E., Maunder, M.N., and Minte-Vera, C.V. 2019. Spatiotemporal dynamics of the dolphin-associated purse-seine fishery for yellowfin tuna (Thunnus albacares) in the eastern Pacific Ocean. Fisheries research **213**: 121-131.

Xu, H., Maunder, M.N., Minte-Vera, C.V., Valero, J.L., Lennert-Cody, C., and Aires-da-Silva, A. 2020. Bigeye tuna in the eastern Pacific Ocean, 2019: benchmark assessment. Inter-Amer.Trop. Tuna Comm., 11th Scient. Adv. Com. Meeting: SAC-11-06.

TABLE 1. The best four-split combination selected by the regression tree algorithm for the longline fishery for bigeye tuna in the eastern Pacific Ocean. The last column shows the percentage of variance in the length-frequency data being explained.
TABLA 1. La mejor combinación de cuatro divisiones seleccionada por el algoritmo de árbol de regresión para la pesquería palangrera de atún patudo en el Océano Pacífico oriental. La última columna muestra el porcentaje de varianza explicada en los datos de frecuencia de talla.

Split	Key	Value	Variance explained
Split1	Latitude	15°S	5.86%
Split2	Longitude	105°W	7.83%
Split3	Latitude	5°S	9.79%
Split4	Longitude	90°W	10.72%

TABLE 2. The best four-split combination selected by the regression tree algorithm for the OBJ fishery for bigeye tuna in the eastern Pacific Ocean. The last column shows the percentage of variance in the length-frequency data being explained.
TABLA 2. La mejor combinación de cuatro divisiones seleccionada por el algoritmo de árbol de regresión para la pesquería OBJ de atún patudo en el Océano Pacífico oriental. La última columna muestra el porcentaje de varianza explicada en los datos de frecuencia de talla.

Split	Key	Value	Variance explained
Split1	Longitude	110°W	10.53%
Split2	Longitude	100°W	11.92%
Split3	Longitude	125°W	13.21%
Split4	Latitude	15°S	13.93%

TABLE 3. The best four-split combination selected by the regression tree algorithm for the NOA fishery for bigeye tuna in the eastern Pacific Ocean. The last column shows the percentage of variance in the length-frequency data being explained.
TABLA 3. La mejor combinación de cuatro divisiones seleccionada por el algoritmo de árbol de regresión para la pesquería NOA de atún patudo en el Océano Pacífico oriental. La última columna muestra el porcentaje de varianza explicada en los datos de frecuencia de talla.

Split	Key	Value	Variance explained
Split1	Longitude	130°W	9.14%

TABLE 4. A summary of the fishery fleets defined for the exploratory assessment of bigeye tuna in the EPO. PS = purse-seine; LL = longline; OBJ = sets on floating objects; NOA = sets on unassociated fish; DEL = sets on dolphins. Fleet-specific definition of column "Area" can be found in Figure 4.
TABLA 4. Un resumen de las flotas pesqueras definidas para la evaluación exploratoria del atún patudo en el OPO. PS = cerco; LL = palangre; OBJ = lances sobre objetos flotantes; NOA = lances no asociados; DEL = lances sobre delfines. Se presenta una definición de la columna "Área" para cada flota en la Figura 4.

Fleet Number	Gear	Set type	Area	Catch data	Unit
1	LL	-	1	Retained catch only	1,000s
2			2		
3			3		
4			4		
5			5		
6			6		
7	LL	-	1	Retained catch only	tons
8			2		
9			3		
10			4		
11			5		
12			6		
13	PS	OBJ	1	Retained catch + discards (inefficiencies)	tons
14			2		
15			3		
16			4		
17			5		
18	PS	OBJ	1-5	Discards (size-sorting)	tons
19	PS	NOA+DEL	1	Retained catch + discards (all)	tons
20			2		

TABLE 5. Model list for the stepwise analysis of model improvements on assessment results. The stepwise manner means each improvement is made additional to all previous improvements.
TABLA 5. Lista de modelos para el análisis escalonado del impacto de las mejoras de los modelos sobre los resultados de la evaluación. El método escalonado consiste en realizar cada mejora de manera adicional a todas las mejoras anteriores.

Model	Component	Improvement
M0		SAC11 "base" model (Env-Fix)
M1	Fishery definition	Improved fishery definitions
M2	Survey fleet	Do not split the abundance index and associated length composition data into two time periods
M3		Improved index of abundance and the associated length compositions
M4	Fishery fleets	Joint (Japan + Korea) length compositions for longline fishery fleets
M5		Add another time block in 2011 to the selectivity curves of longline fishery fleets
M6		Standardize the joint length compositions for longline fishery fleets

TABLE 6. Recruitment shift statistic for the eight reference models to which the statistic can be applied. SAC11 and SAC14 represent the models in the last benchmark assessment and this exploratory analysis, respectively.
TABLA 6. Estadística de cambio de reclutamiento para los ochos modelos de referencia a los que se puede aplicar esta estadística. SAC11 y SAC14 representan los modelos de la última evaluación de referencia y de este análisis exploratorio, respectivamente.

	Regime shift is real				Regime shift is not real			
Source	**Env-Fix**	**Env-Gro**	**Env-Mrt**	**Env-Sel**	**Gro**	**Mov**	**Mrt**	**Sel**
SAC11	2.4	1.5	1.8	1.6	1.2	1.3	1.4	1.3
SAC14	1.5	1.0	1.1	1.1	1.0	1.2	1.1	1.1

TABLE 7. The list of reference models being considered in this exploratory analysis. Regime means whether a regime parameter for recruitment is estimated in the model, and the selectivity is for Fishery 3 in 2011-2019.
TABLA 7. Lista de modelos de referencia considerados en este análisis exploratorio. El rubro "régimen" indica si se estima o no un parámetro de régimen para el reclutamiento en el modelo. La selectividad es la de la pesquería 3 en 2011-2019.

Model	**Years**	**Regime**	**Growth**	**Natural mortality**	**Selectivity**	**Note**
Env-Fix	1979-2019	Y	std (L1)	Fix	Asymptotic	
Gro		N	all	Fix	Asymptotic	Fit to age-at-length data
Mov		N	std (L1)	Est (quarter 13)	Asymptotic	
Mrt		N	std (L1)	Est (quarter 26)	Asymptotic	
Sel		N	std (L1)	Fix	Dome-shape	
Srt-Fix	2000-2019	N	std (L1)	Fix	Asymptotic	
Srt-Gro		N	all	Fix	Asymptotic	Fit to age-at-length data
Srt-Mrt		N	std (L1)	Est (quarter 26)	Asymptotic	
Srt-Sel		N	std (L1)	Fix	Dome-shape	

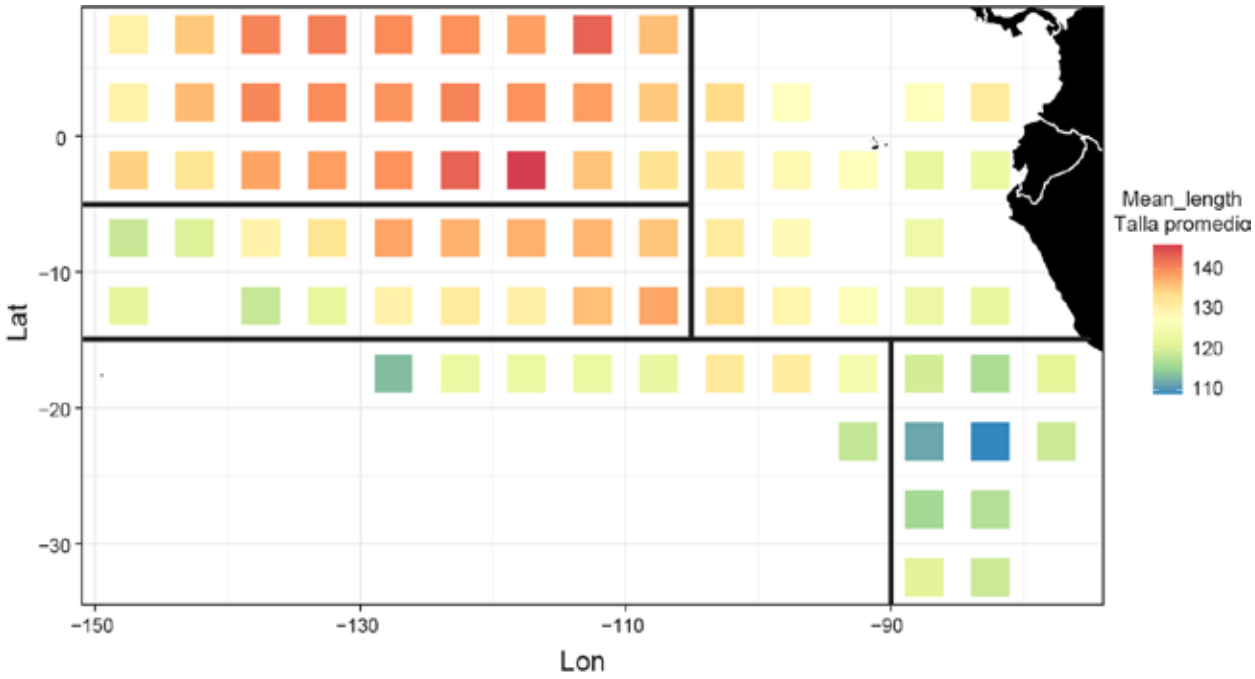

FIGURE 1. Map of average length (cm) of bigeye tuna caught by the longline fishery. The four solid lines are the best four-split combination selected by the regression tree algorithm.
FIGURA 1. Mapa de la talla promedio (cm) del atún patudo capturado por la pesquería palangrera. Las cuatro líneas sólidas corresponden a la mejor combinación de cuatro divisiones seleccionada por el algoritmo de árbol de regresión.

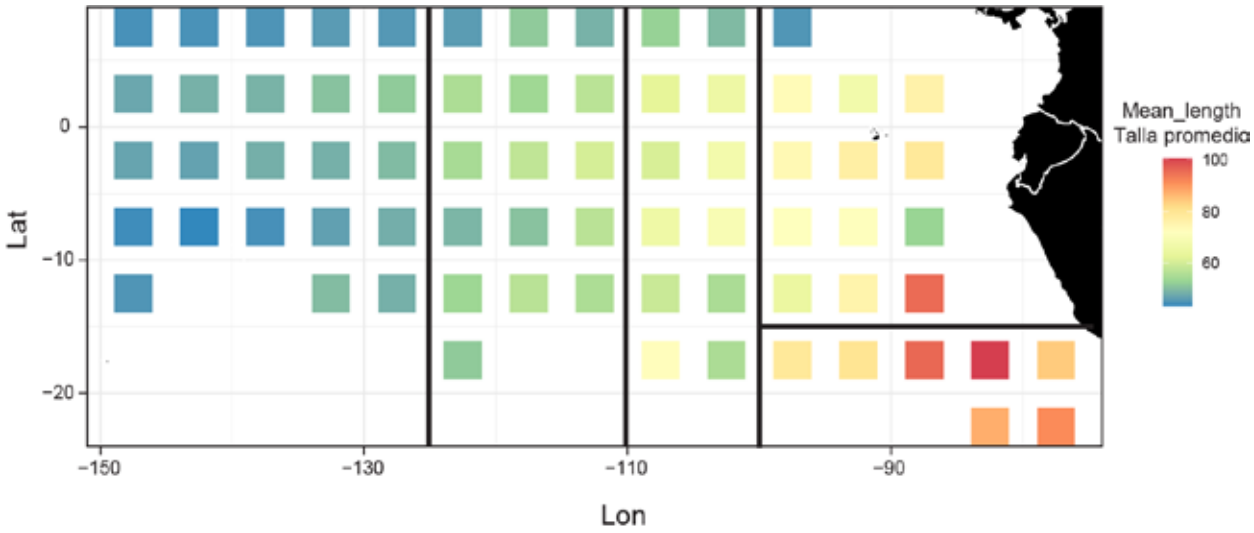

FIGURE 2. Map of average length (cm) of bigeye tuna caught by the floating-object fishery. The four solid lines are the best four-split combination selected by the regression tree algorithm.
FIGURA 2. Mapa de la talla promedio (cm) del atún patudo capturado por la pesquería sobre objetos flotantes. Las cuatro líneas sólidas corresponden a la mejor combinación de cuatro divisiones seleccionada por el algoritmo de árbol de regresión.

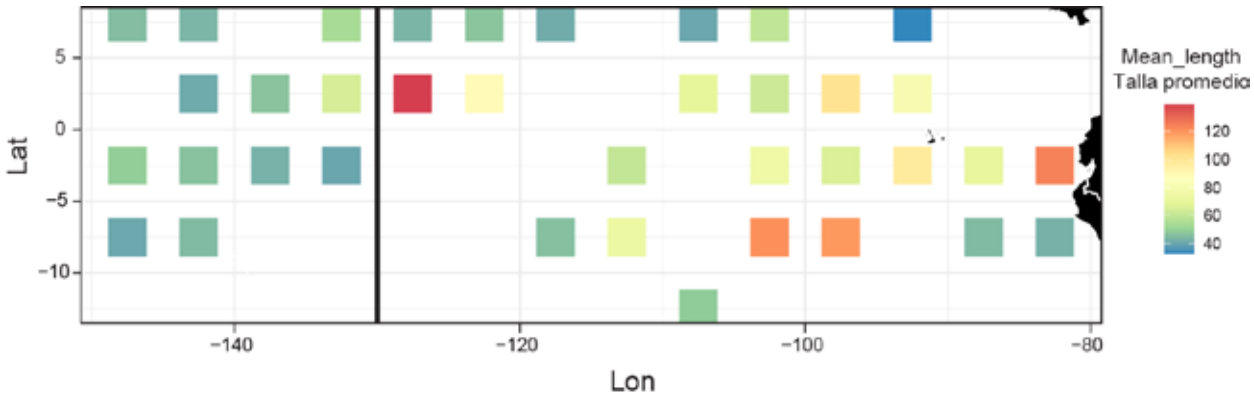

FIGURE 3. Map of average length (cm) of bigeye tuna caught by the unassociated fishery. The solid line is the best split selected by the regression tree algorithm.
FIGURA 3. Mapa de la talla promedio (cm) del atún patudo capturado por la pesquería no asociada. La línea sólida corresponde a la mejor división seleccionada por el algoritmo de árbol de regresión.

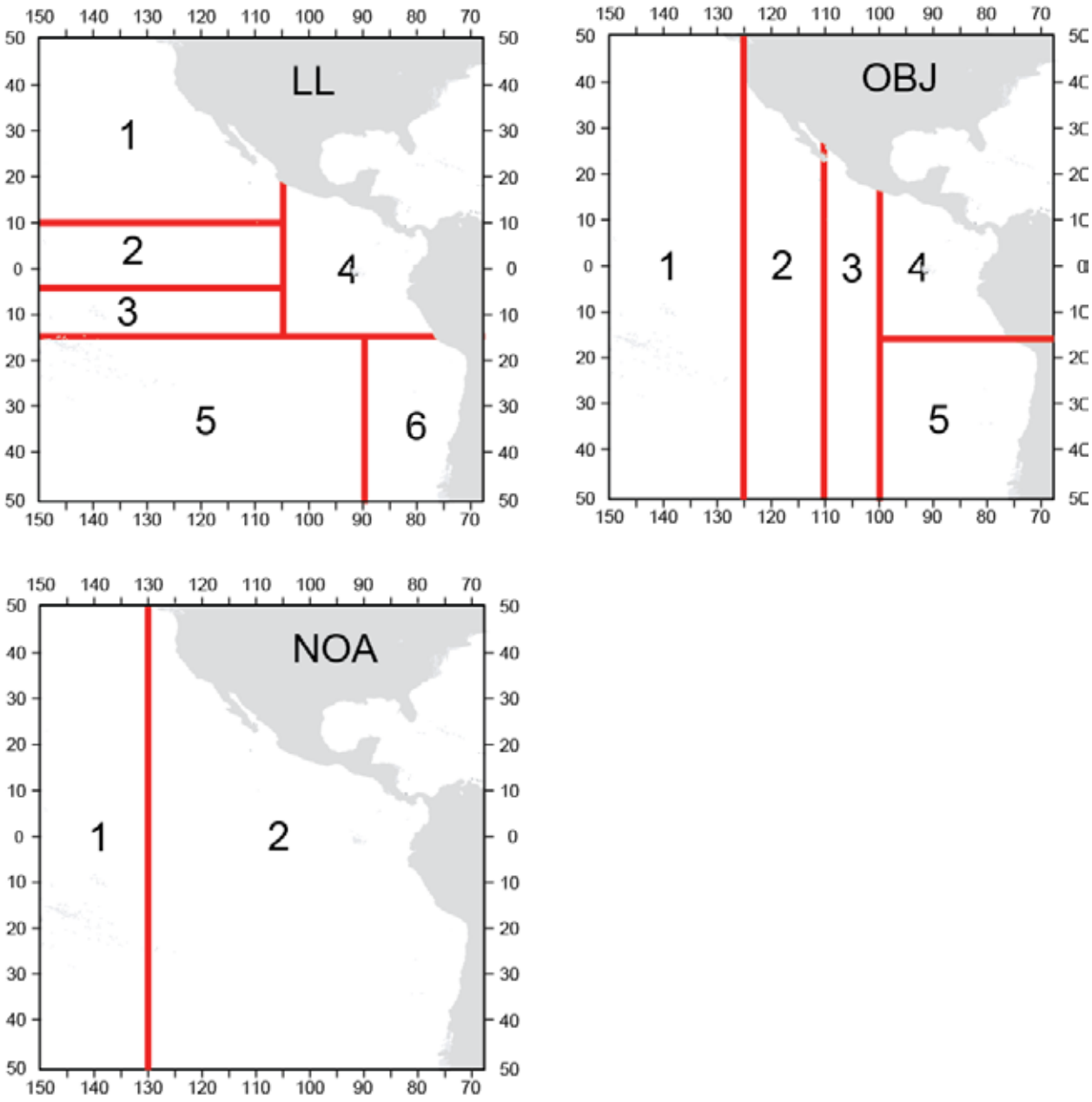

FIGURE 4. Summary of area definitions for the longline (LL), floating-object (OBJ), and unassociated (NOA) fishery fleets in the exploratory assessment model.

FIGURA 4. Resumen de las definiciones de áreas para las flotas de las pesquerías palangrera (LL), sobre objetos flotantes (OBJ) y no asociada (NOA) en el modelo de la evaluación exploratoria.

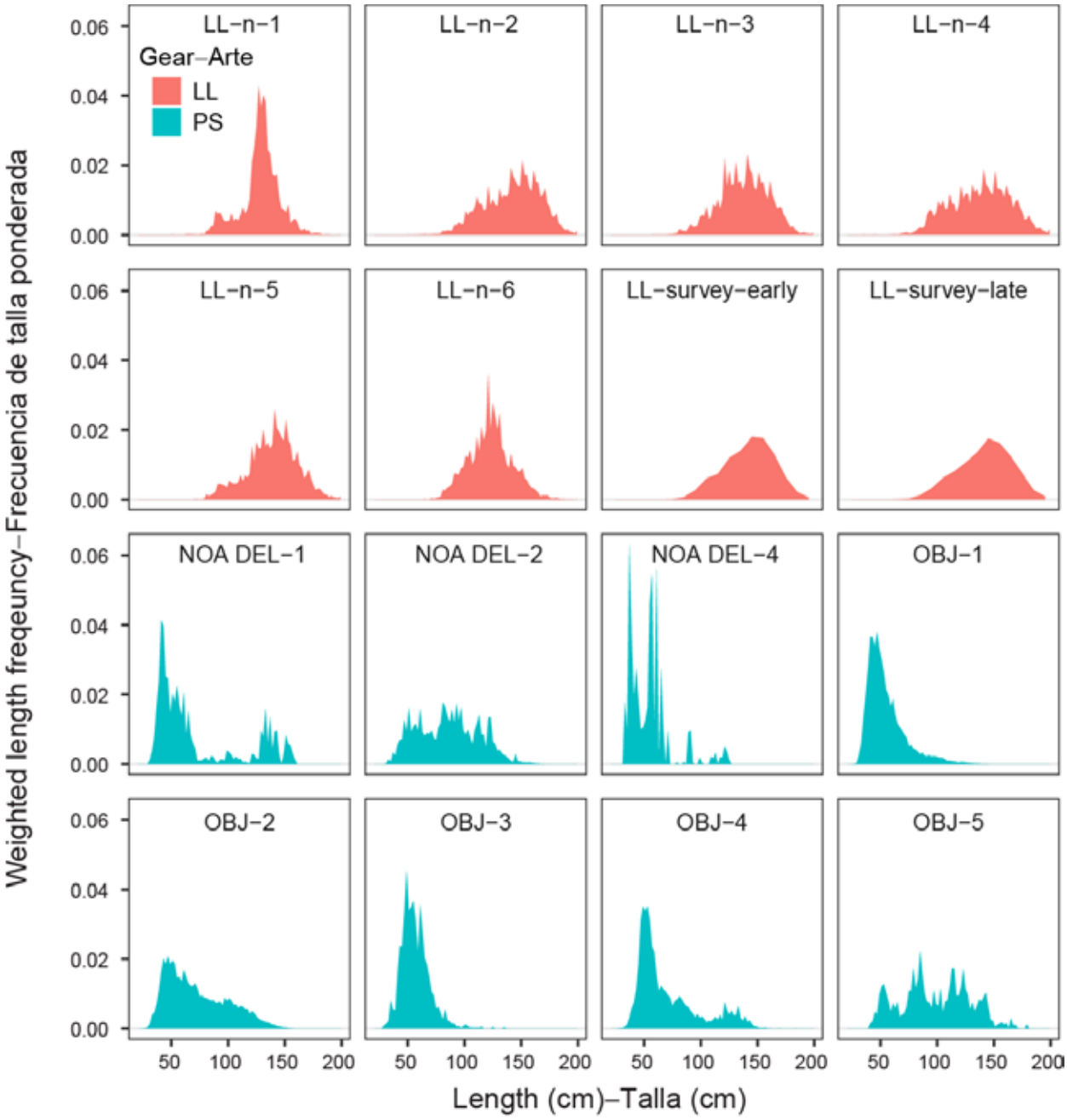

FIGURE 5. Sample-size weighted length frequency of bigeye tuna observed by each fishery and survey fleet in the last benchmark assessment model.

FIGURA 5. Frecuencia de talla ponderada por tamaño de muestra de atún patudo observada por cada pesquería y flota de estudio en el modelo de la última evaluación de referencia.

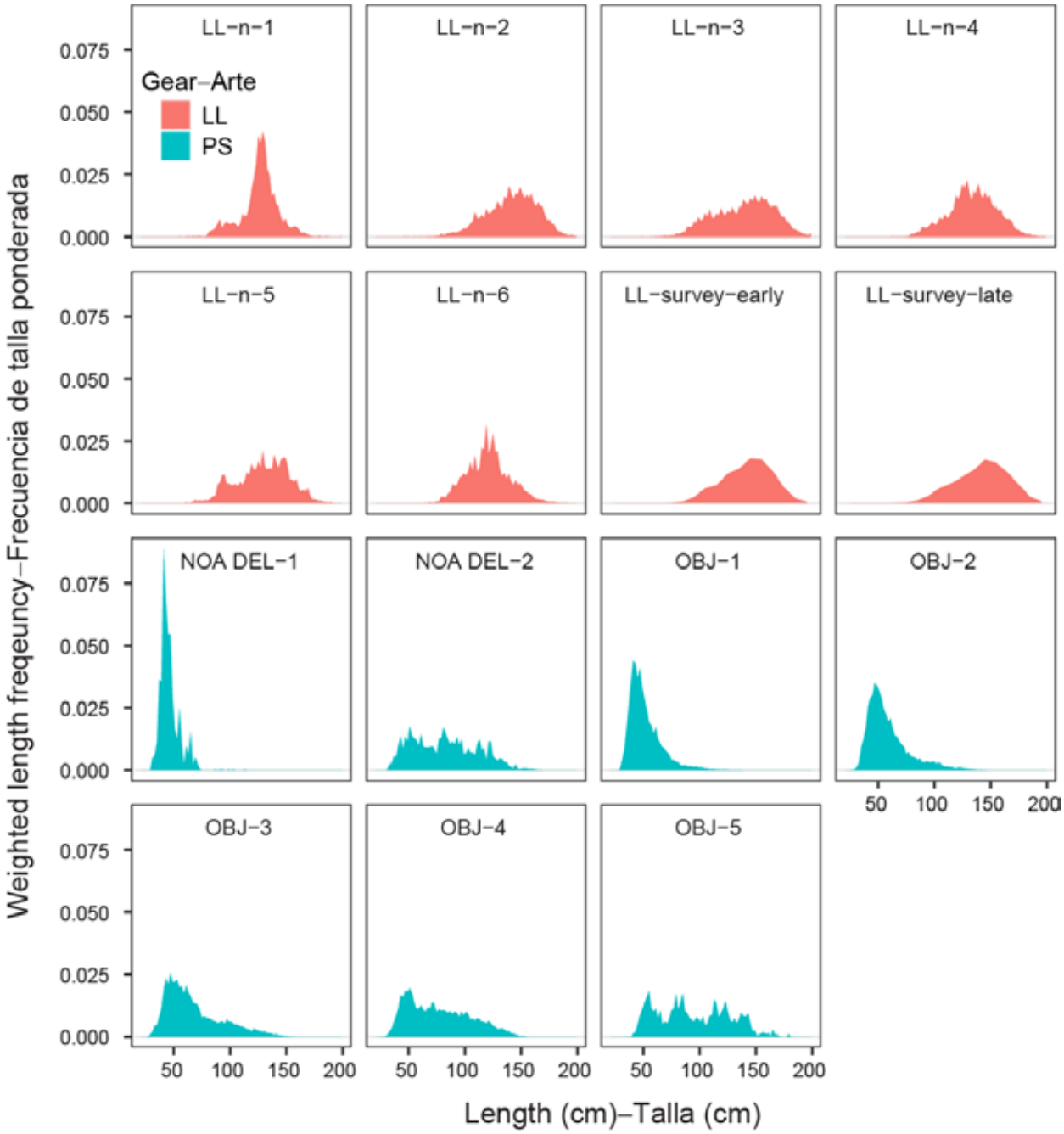

FIGURE 6. Sample-size weighted length frequency of bigeye tuna observed by each fishery and survey fleet in the exploratory assessment model.

FIGURA 6. Frecuencia de talla ponderada por tamaño de muestra de atún patudo observada por cada pesquería y flota de estudio en el modelo de la evaluación exploratoria.

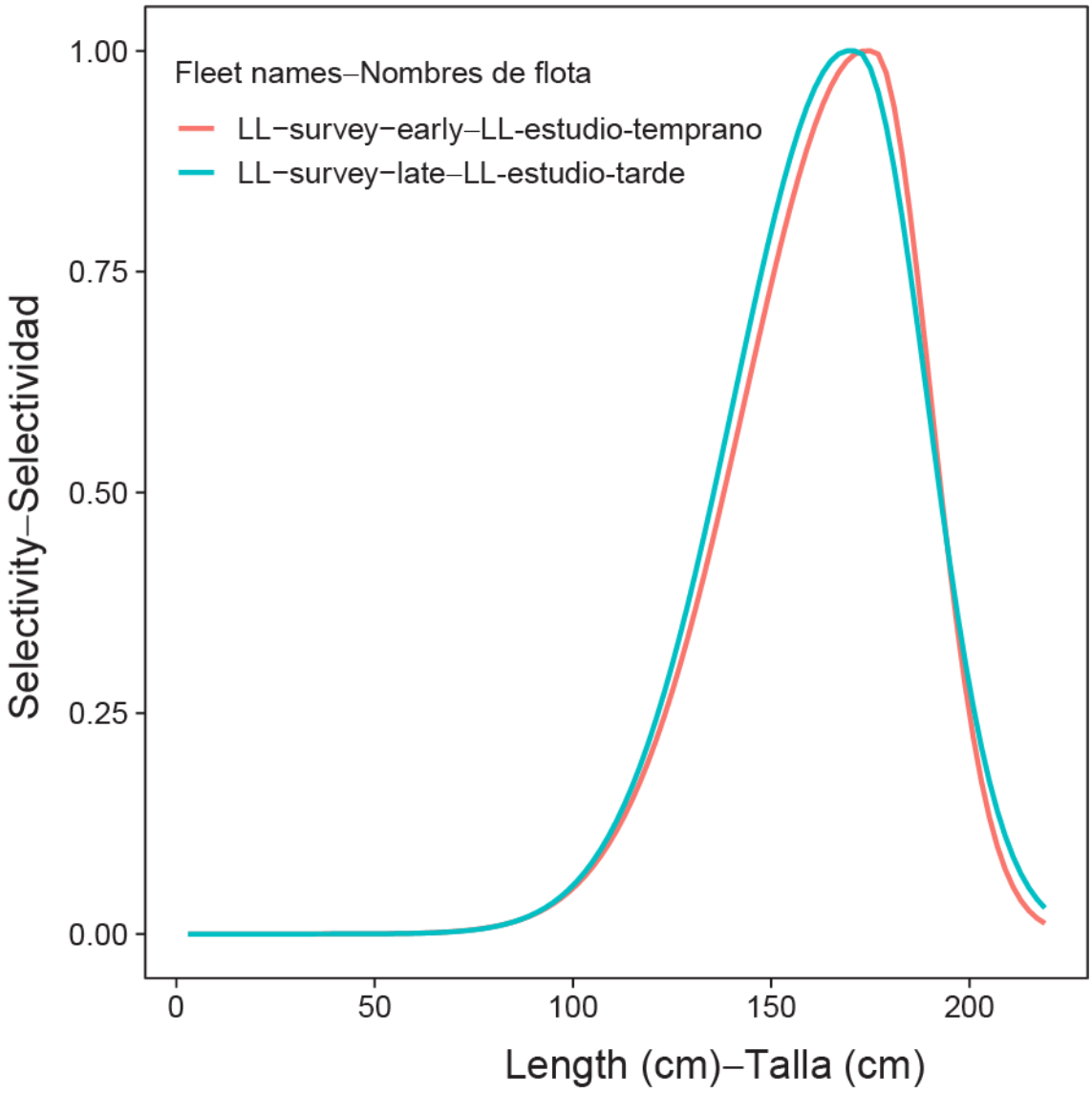

FIGURE 7. Estimated selectivity curves for the early and late longline survey fleets in the last benchmark assessment model.
FIGURA 7. Curvas de selectividad estimadas para las flotas de estudio de palangre temprana y tardía en el modelo de la última evaluación de referencia.

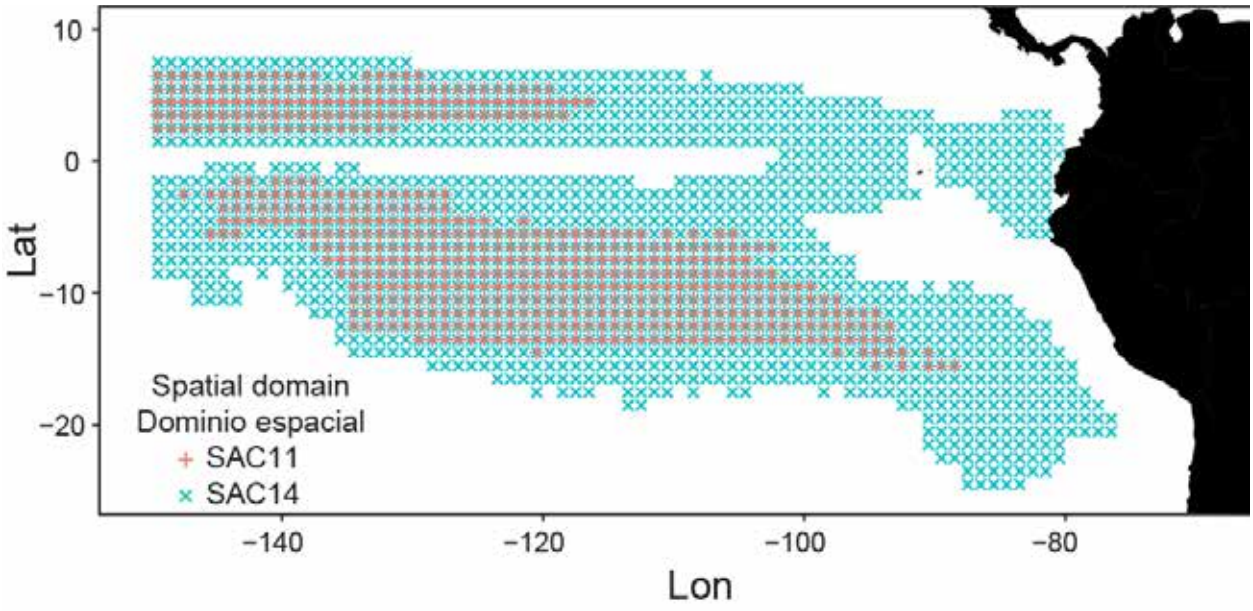

FIGURE 8. Comparison of the spatial domain on which the CPUE standardization for the last benchmark assessment (SAC11) and this exploratory assessment (SAC14) is based.
FIGURA 8. Comparación del dominio espacial sobre el cual se basa la estandarización de la CPUE para la última evaluación de referencia (SAC11) y esta evaluación exploratoria (SAC14).

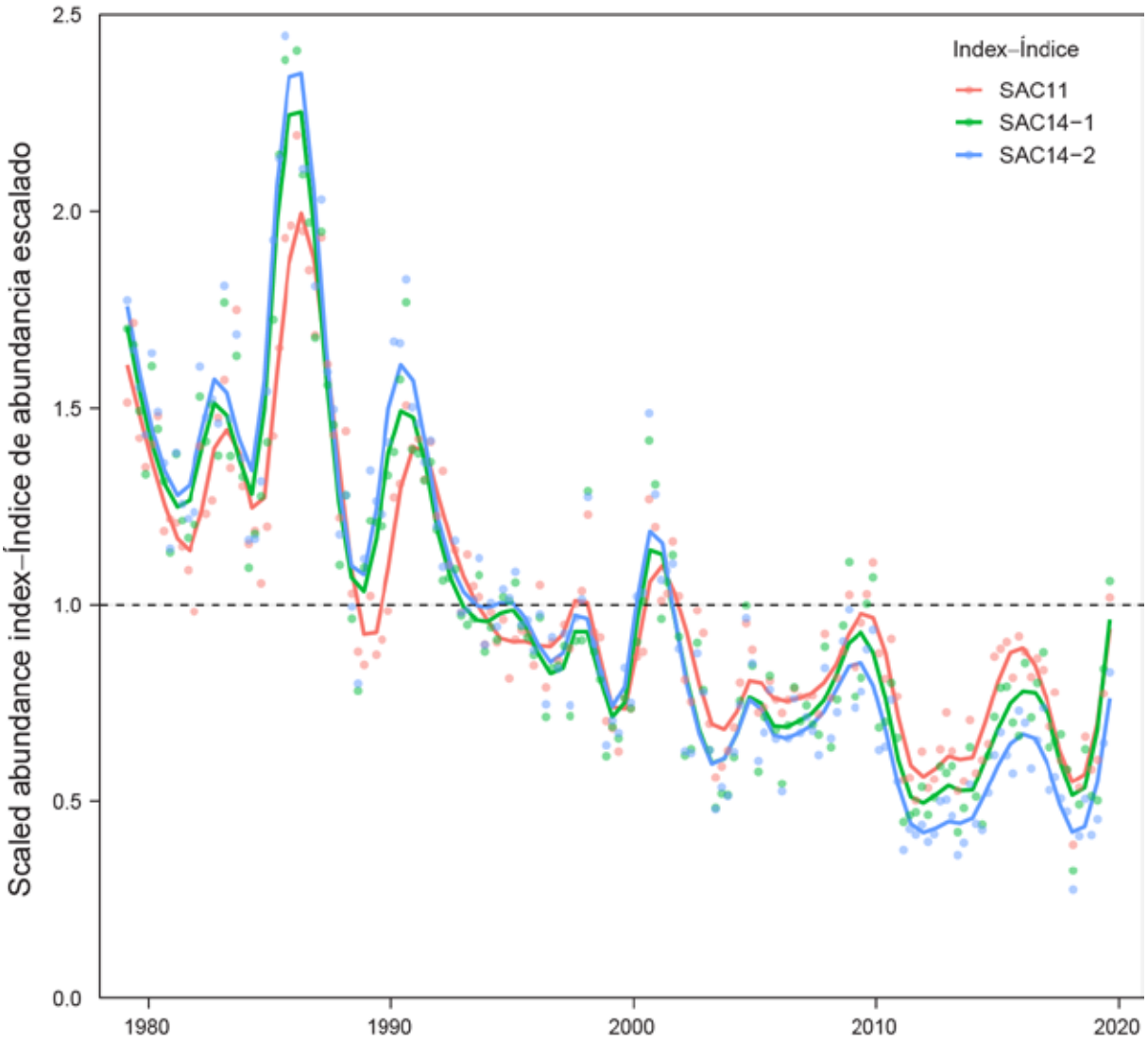

FIGURE 9. Comparison of the longline abundance indices standardized for bigeye tuna in the EPO. SAC11 represents the abundance index estimated for the last benchmark assessment; SAC14-1 represents the abundance index estimated for this exploratory assessment based on the assumption that spatiotemporal terms are independent in time but correlated in space; and SAC14-2 represents the abundance index estimated for this exploratory assessment based on the assumption that spatiotemporal terms are correlated in space and follow a random-walk process in time. The color dots and lines are the quarterly estimates and smoothed values, respectively. All three indices are scaled to have a mean of 1 for easy comparison.

FIGURA 9. Comparación de los índices de abundancia de palangre estandarizados para el atún patudo en el OPO. SAC11 representa el índice de abundancia estimado para la última evaluación de referencia. SAC14-1 representa el índice de abundancia estimado para esta evaluación exploratoria con base en el supuesto de que los términos espaciotemporales son independientes en el tiempo pero están correlacionados en el espacio. SAC14-2 representa el índice de abundancia estimado para esta evaluación exploratoria con base en el supuesto de que los términos espaciotemporales se correlacionan en el espacio y siguen un proceso de paseo aleatorio en el tiempo. Los puntos y líneas de color corresponden a las estimaciones trimestrales y a los valores suavizados, respectivamente. Se ajusta la escala de los tres índices para que tengan un promedio de 1 para facilitar su comparación.

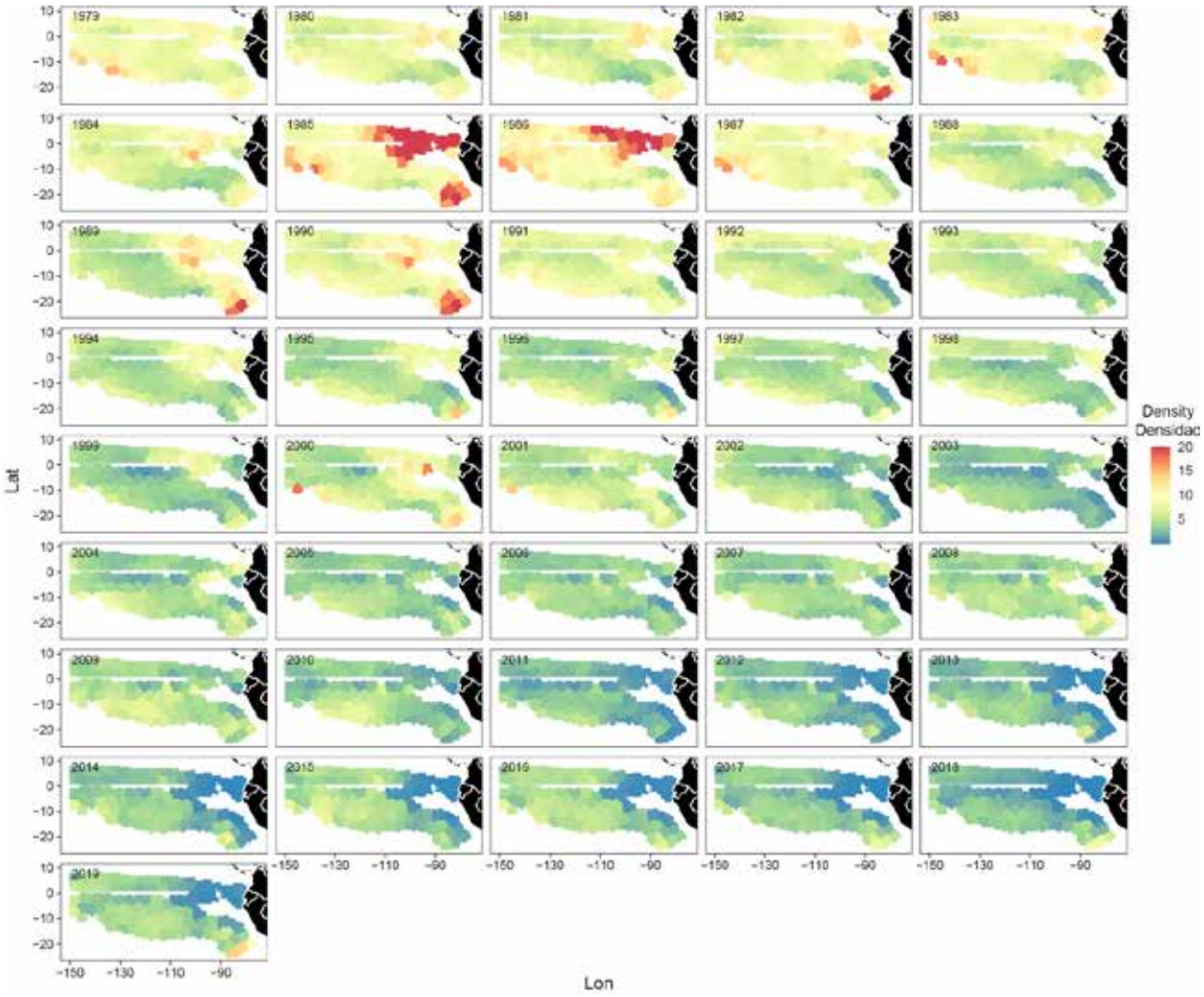

FIGURE 10. Maps of predicted bigeye density by year from the CPUE standardization model developed based on Japanese longline data.

FIGURA 10. Mapas de la densidad predicha de atún patudo por año a partir del modelo de estandarización de CPUE, desarrollado a partir de datos de palangre de Japón.

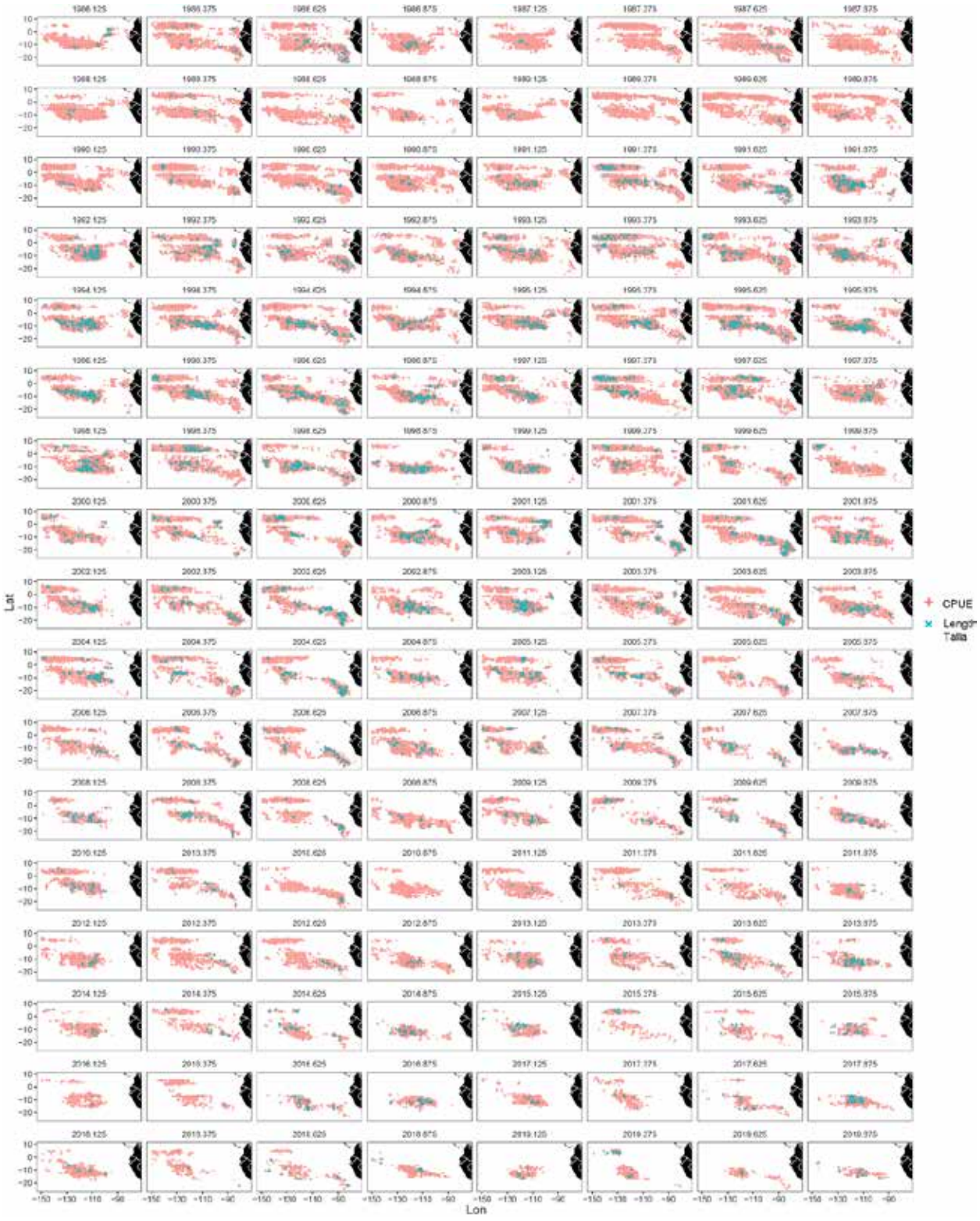

FIGURE 11. Comparison of the spatial distribution of Japanese longline CPUE data and Japanese length composition data by year-quarter.

FIGURA 11. Comparación de la distribución espacial de los datos de CPUE de palangre de Japón y los datos de composición por talla de Japón por año-trimestre.

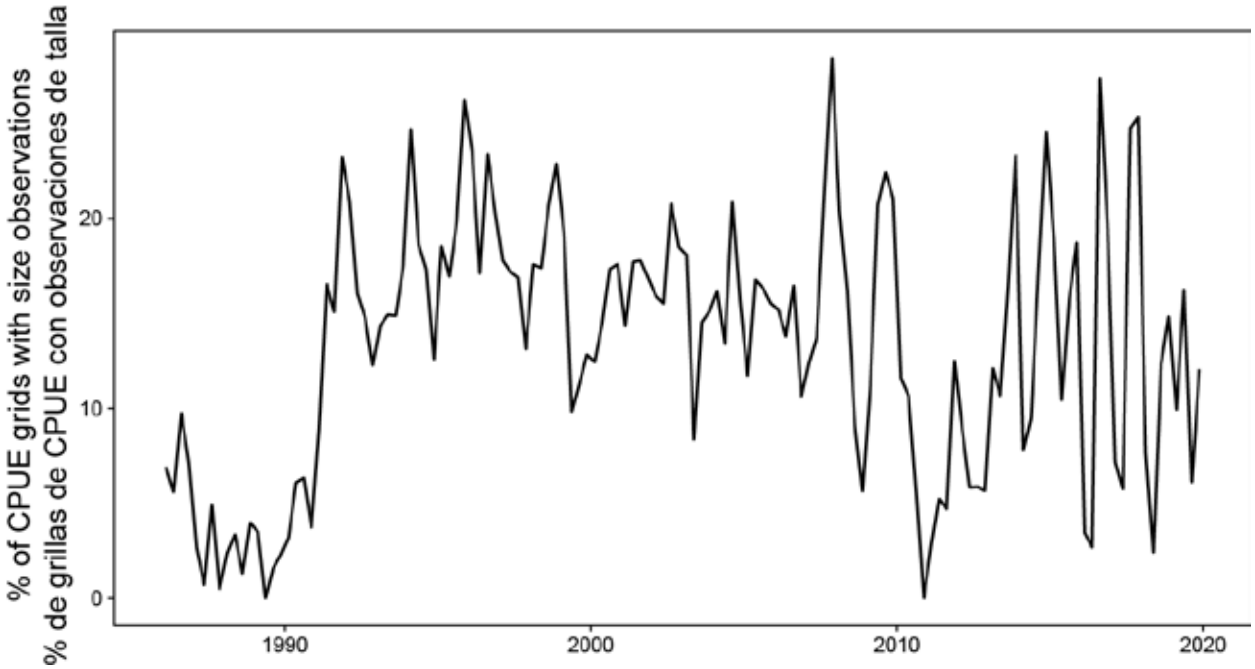

FIGURE 12. The percentage of Japanese longline CPUE grids containing Japanese longline length composition data in the same year-quarter.
FIGURA 12. Porcentaje de grillas de CPUE de palangre de Japón que contienen datos de composición por talla de palangre de Japón en el mismo año-trimestre.

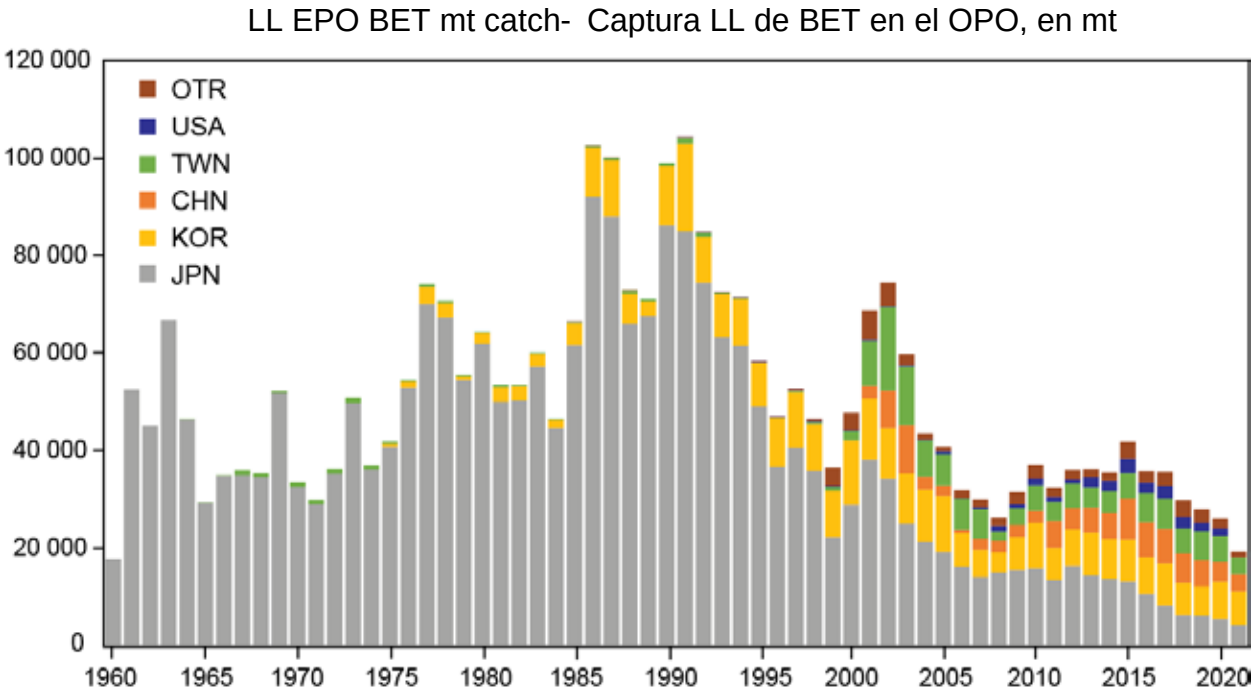

FIGURE 13. Annual longline catch of bigeye tuna in the eastern Pacific Ocean by CPC.
FIGURA 13. Captura palangrera anual de atún patudo en el Océano Pacífico oriental por CPC.

FIGURE 14. Comparison of the spatial coverage of Japanese and Korean longline length composition data by year-quarter between 2011 and 2019.
FIGURA 14. Comparación de la cobertura espacial de los datos de composición por talla de palangre de Japón y de Corea por año-trimestre entre 2011 y 2019.

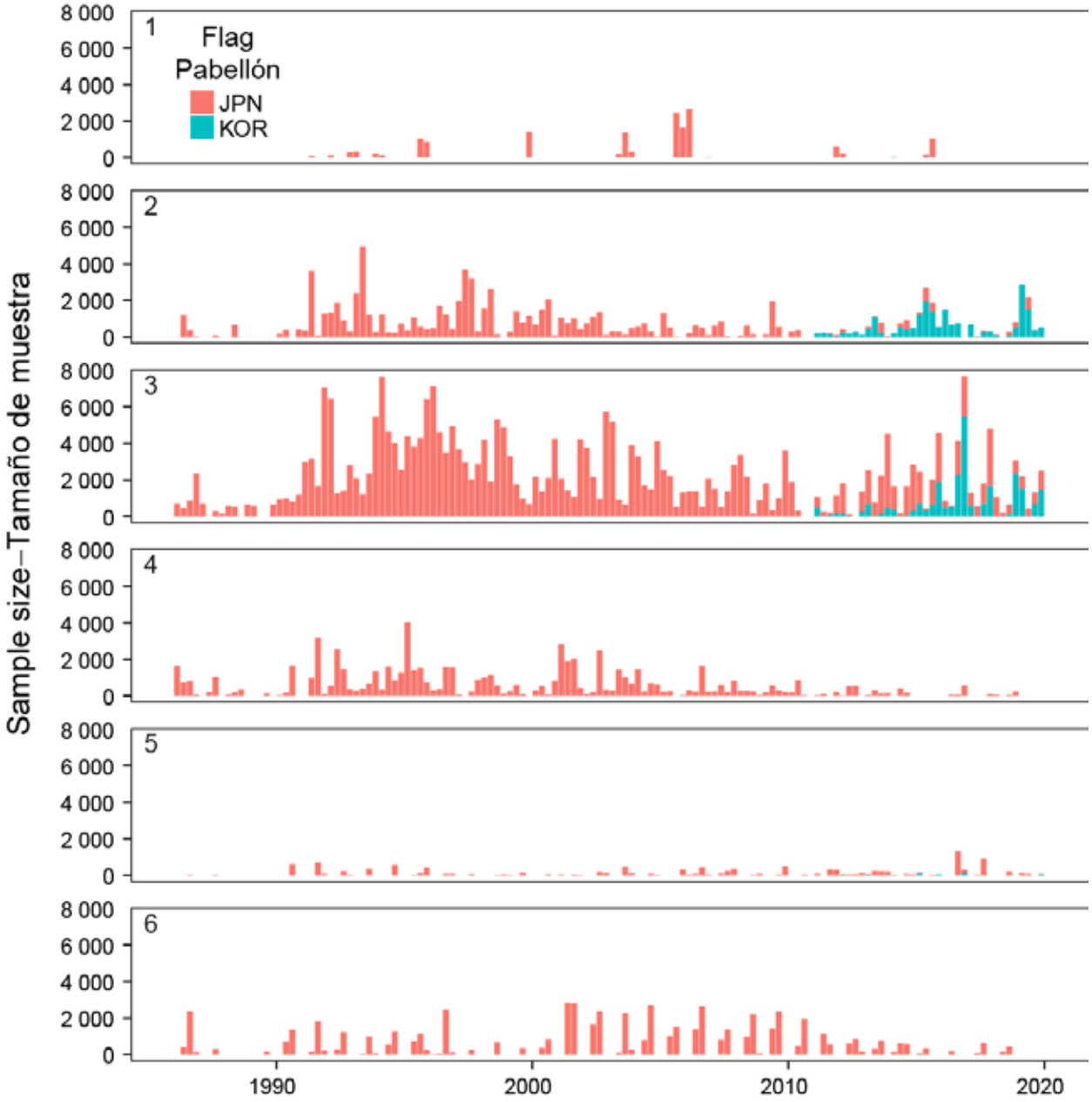

FIGURE 15. Quarterly sample size of the joint longline length composition data for fisheries fleets.
FIGURA 15. Tamaño de muestra trimestral de los datos de composición por talla de palangre conjuntos para flotas pesqueras.

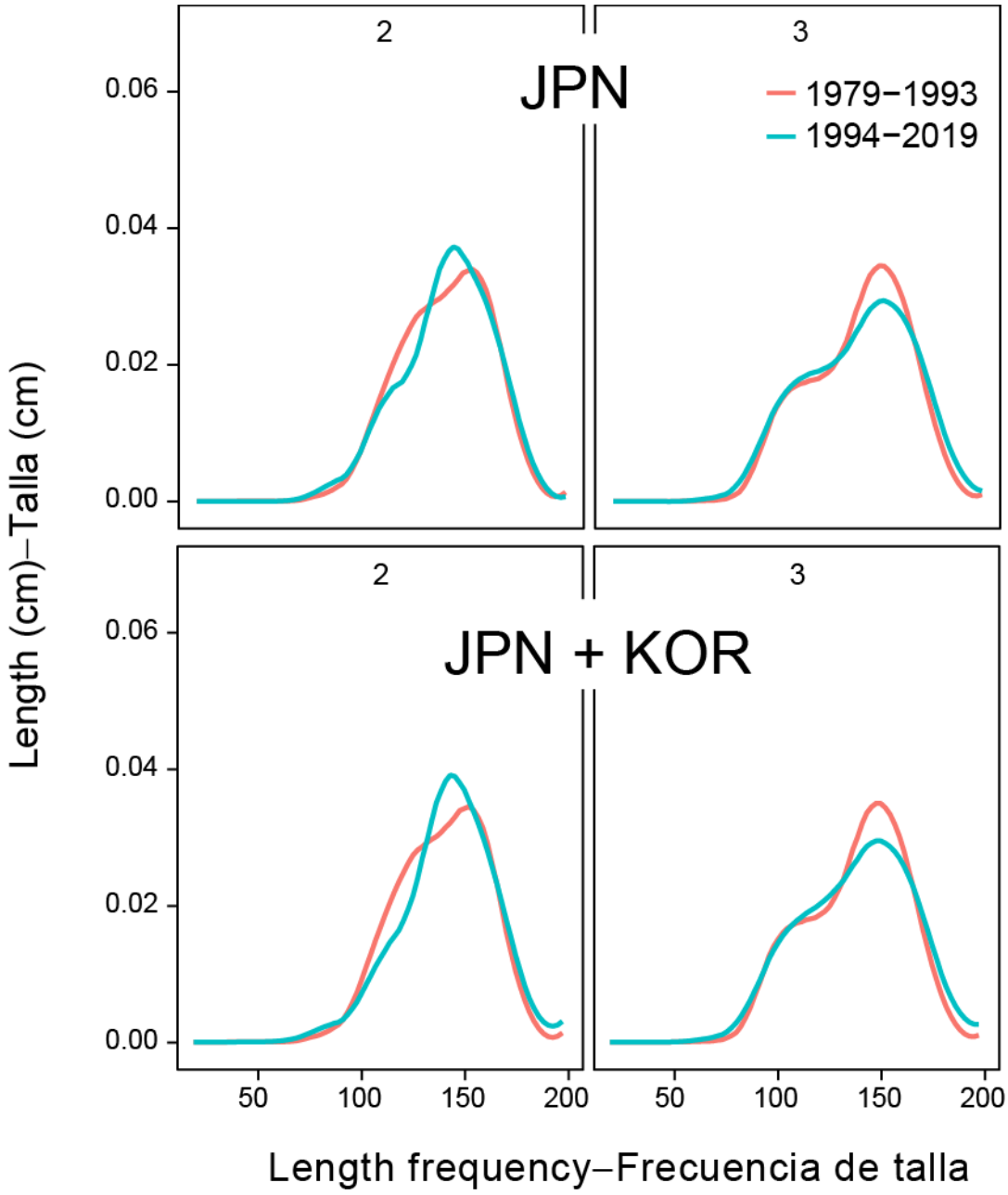

FIGURE 16. Comparison of the joint (Japanese and Korean) and Japanese length frequencies of bigeye tuna observed by the second and third longline fishery fleets. The red and blue profiles represent sample size weighted values for the early and late longline periods, respectively.

FIGURA 16. Comparación de las frecuencias de talla conjuntas (de Japón y de Corea) y de Japón del atún patudo, observadas por la segunda y tercera flotas pesqueras de palangre. Los perfiles rojo y azul representan valores ponderados por tamaño de muestra para los periodos de palangre temprano y tardío, respectivamente.

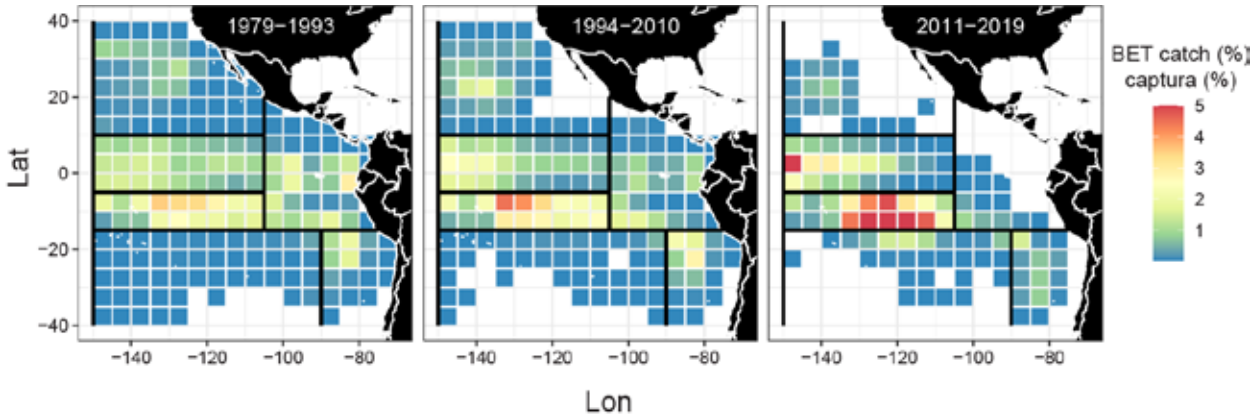

FIGURE 17. The spatial distribution (%) of bigeye catch in numbers of fish by Japanese and Korean longline vessels in three different time periods. The black lines are the area boundaries defined for longline fishery fleets.

FIGURA 17. Distribución espacial (%) de la captura de patudo, en número de peces, por buques palangreros japoneses y coreanos en tres periodos de tiempo diferentes. Las líneas negras corresponden a los límites de las áreas definidas para las flotas pesqueras de palangre.

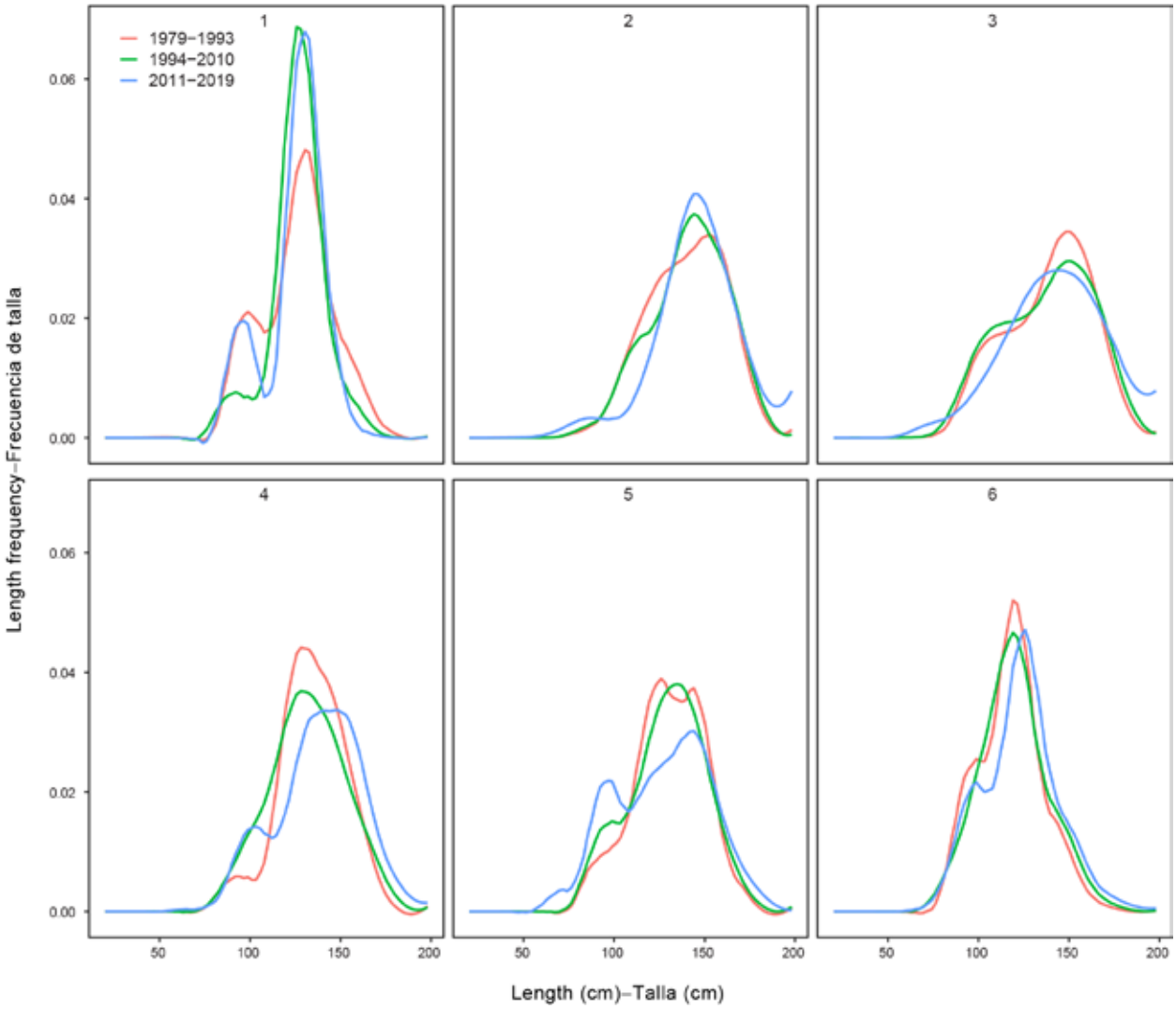

FIGURE 18. Comparison of joint length frequencies of bigeye tuna, by longline fishery, in the three time blocks proposed for the selectivity of longline fisheries.
FIGURA 18. Comparación de las frecuencias de talla conjuntas del atún patudo, por pesquería palangrera, en los tres bloques de tiempo propuestos para la selectividad de las pesquerías palangreras.

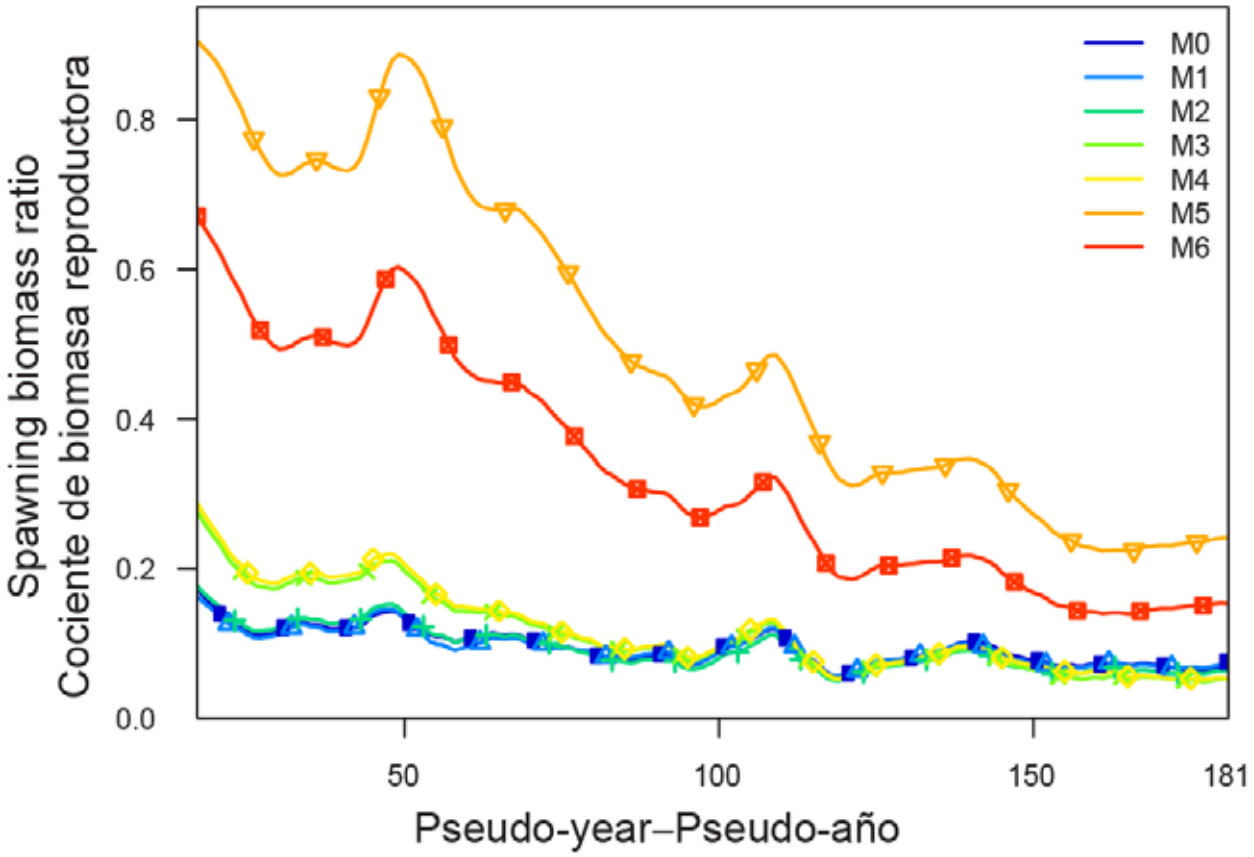

FIGURE 19. Comparison of estimated spawning biomass ratio from the seven stepwise assessment models for bigeye tuna in the eastern Pacific Ocean (see Table 4 for model definition).
FIGURA 19. Comparación del cociente de biomasa reproductora estimado de los siete modelos de evaluación escalonados para el atún patudo en el Océano Pacífico oriental (ver la Tabla 4 para conocer las definiciones de los modelos).

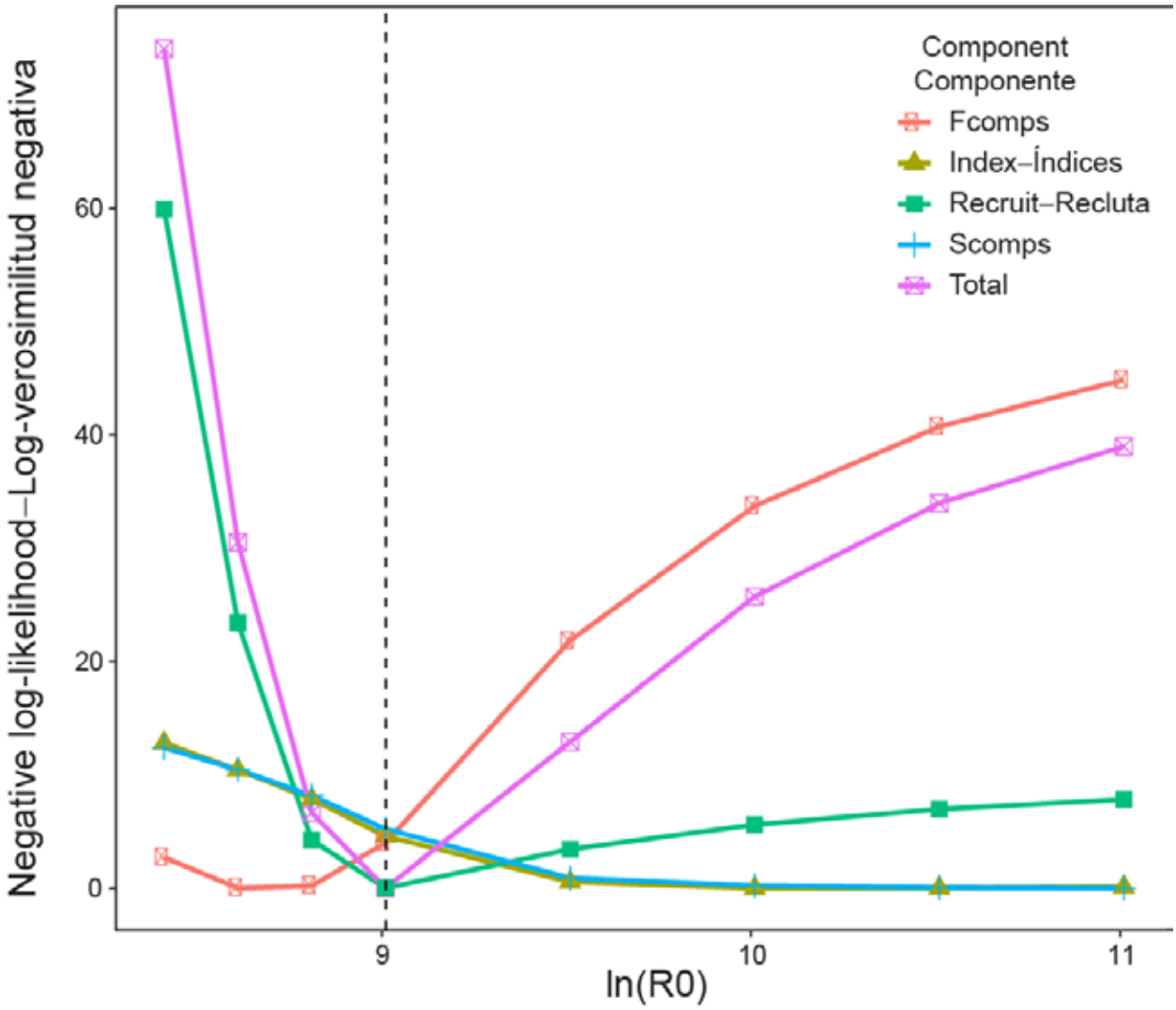

FIGURE 20. The R_0 likelihood profile for the new "base" model for bigeye tuna in the eastern Pacific Ocean. The components of the total negative log-likelihood include fishery compositions (Fcomps), survey compositions (Scomps), survey index of abundance (Index), and recruitment penalty (Recruit).
FIGURA 20. Perfil de verosimilitud de R_0 para el nuevo modelo "base" para el patudo en el Océano Pacífico oriental. Los componentes de la verosimilitud logarítmica negativa total comprenden las composiciones pesqueras ("Fcomps"), las composiciones de estudio ("Scomps"), el índice de abundancia de estudio ("Índice") y la penalización de reclutamiento ("Recluta").

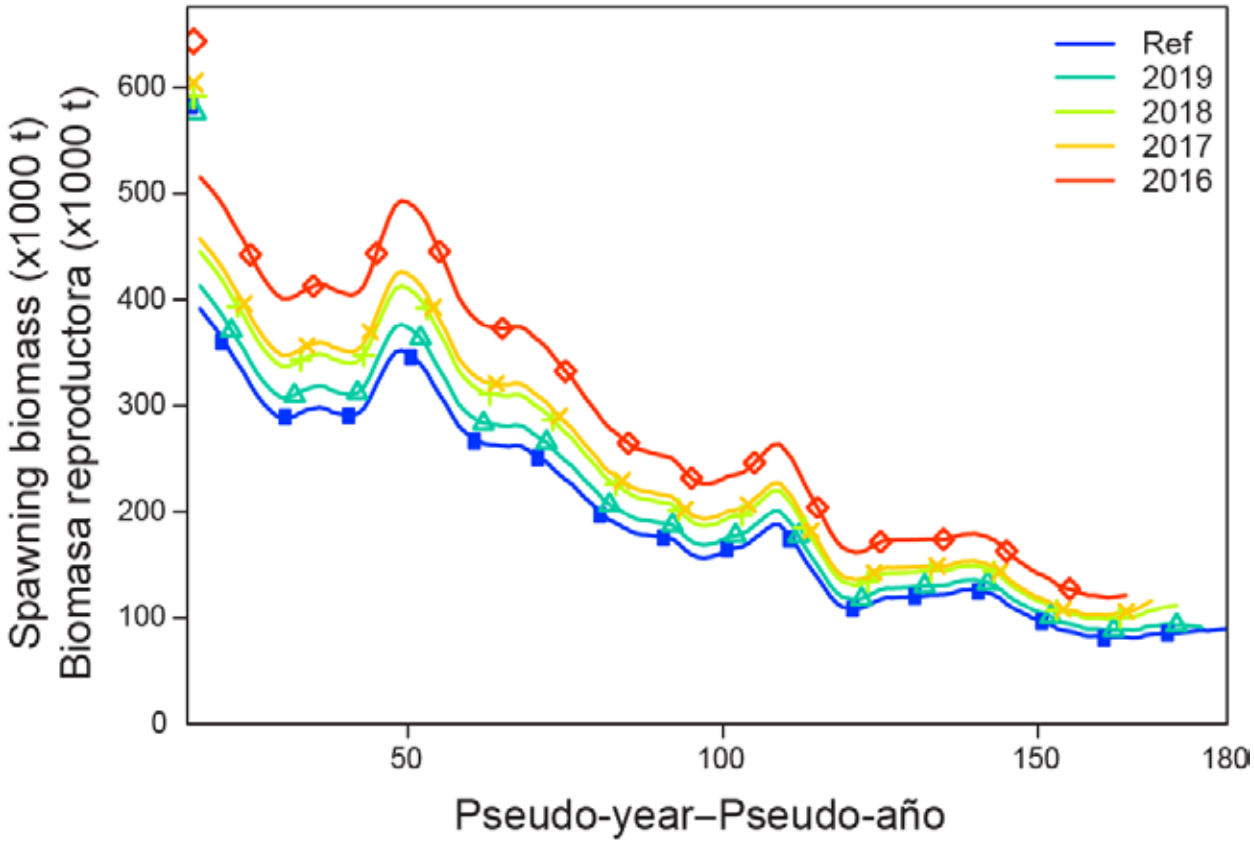

FIGURE 21. The retrospective pattern of spawning biomass for the new "base" model for bigeye tuna in the eastern Pacific Ocean.
FIGURA 21. Patrón retrospectivo de la biomasa reproductora para el nuevo modelo "base" para el atún patudo en el Océano Pacífico oriental.

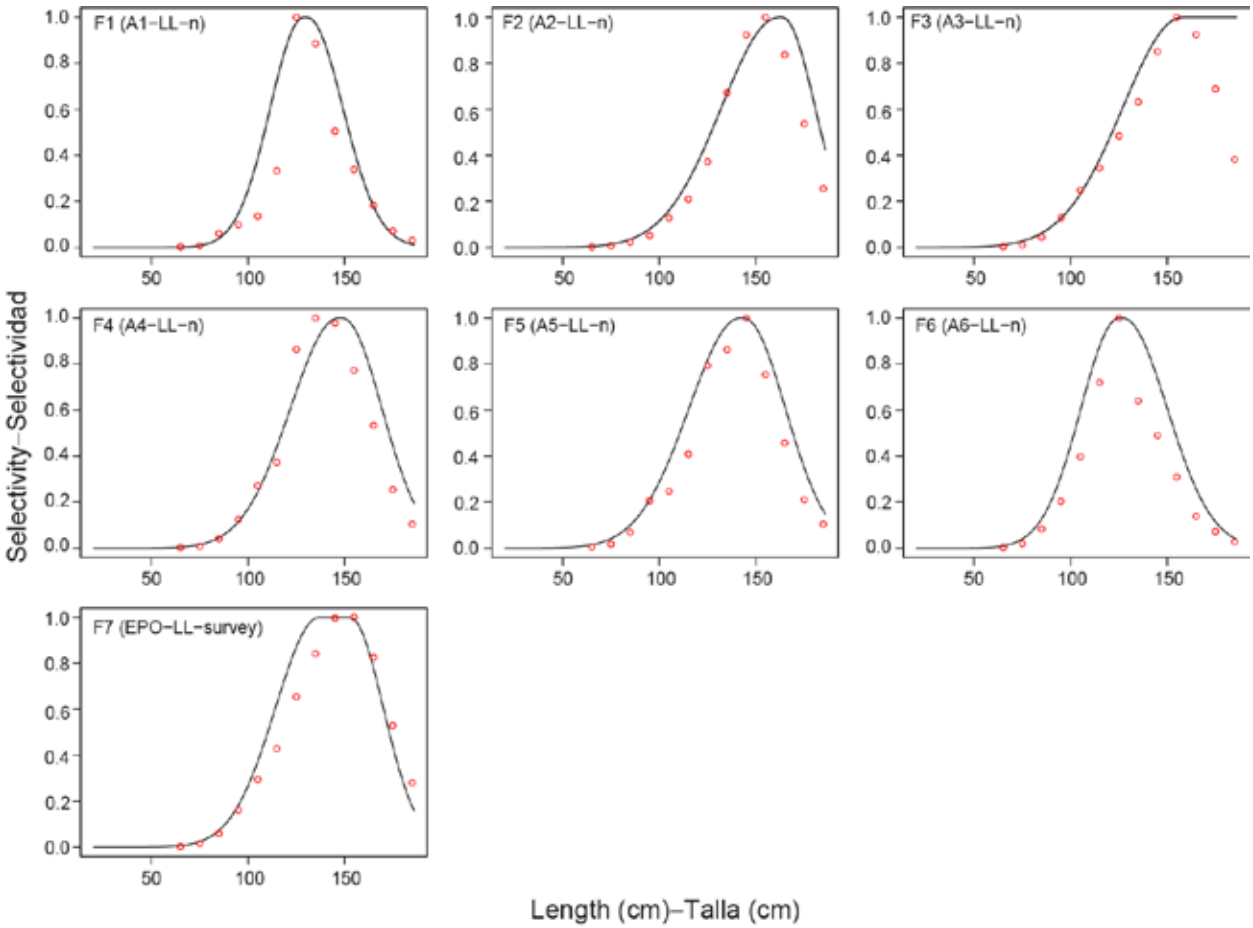

FIGURE 22. Comparison of estimated (black lines) and empirical (red dots) selectivity for every longline fishery between 2011 and 2019.
FIGURA 22. Comparación de la selectividad estimada (líneas negras) y empírica (puntos rojos) para cada pesquería palangrera entre 2011 y 2019.

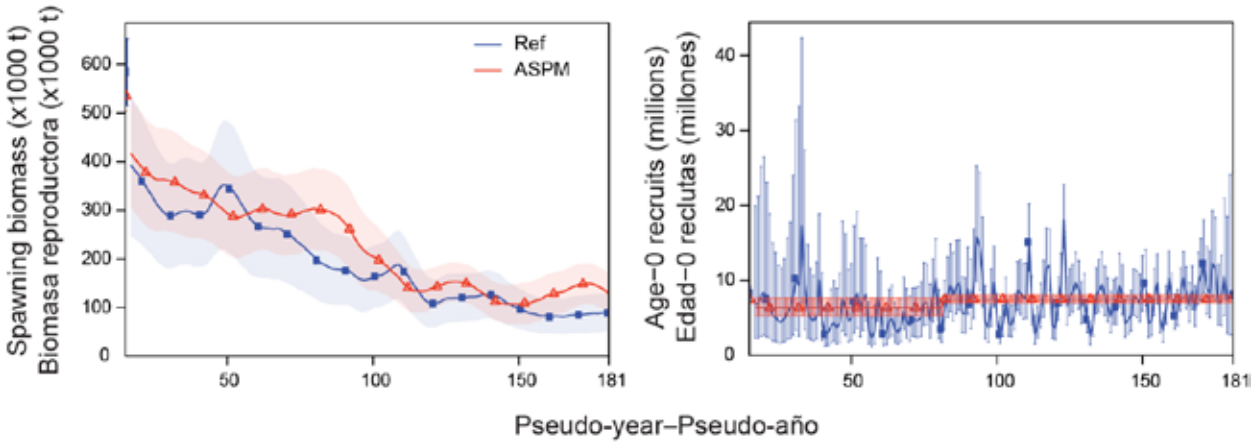

FIGURE 23. Age-structured production model (ASPM) diagnostics for the new "base" model for bigeye tuna in the eastern Pacific Ocean.
FIGURA 23. Diagnósticos del modelo de producción estructurado por edad (ASPM) para el nuevo modelo "base" para el atún patudo en el Océano Pacífico oriental.

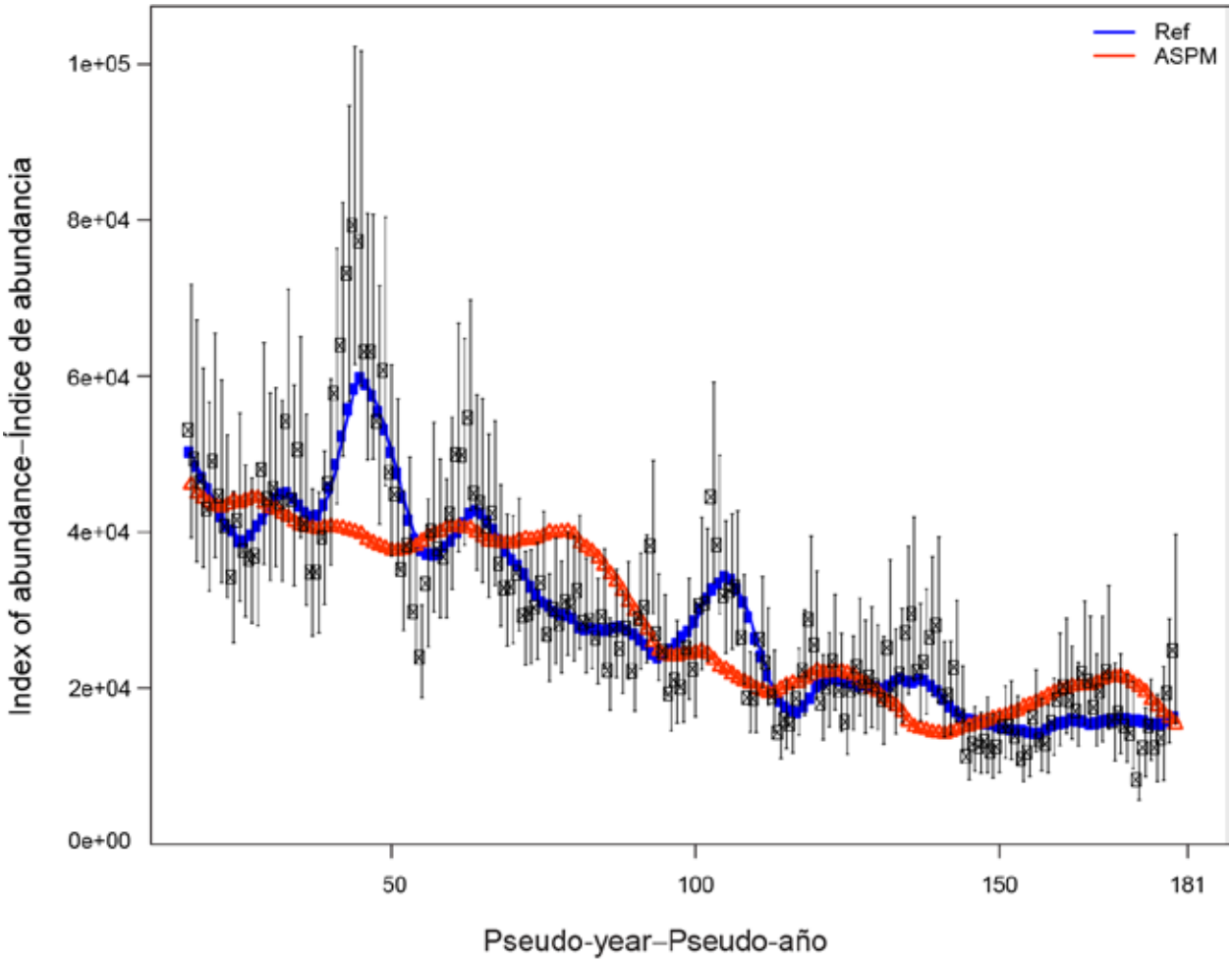

FIGURE 24. Comparison of the fit to the survey abundance index between the "base" model and the age-structured production model (ASPM). Black crosses and error bars are the observed abundance indices and the associated 95% confidence interval.

FIGURA 24. Comparación del ajuste al índice de abundancia de estudio entre el modelo "base" y el modelo de producción estructurado por edad (ASPM). Las cruces negras y las barras de error son los índices de abundancia observados y el intervalo de confianza de 95% asociado.

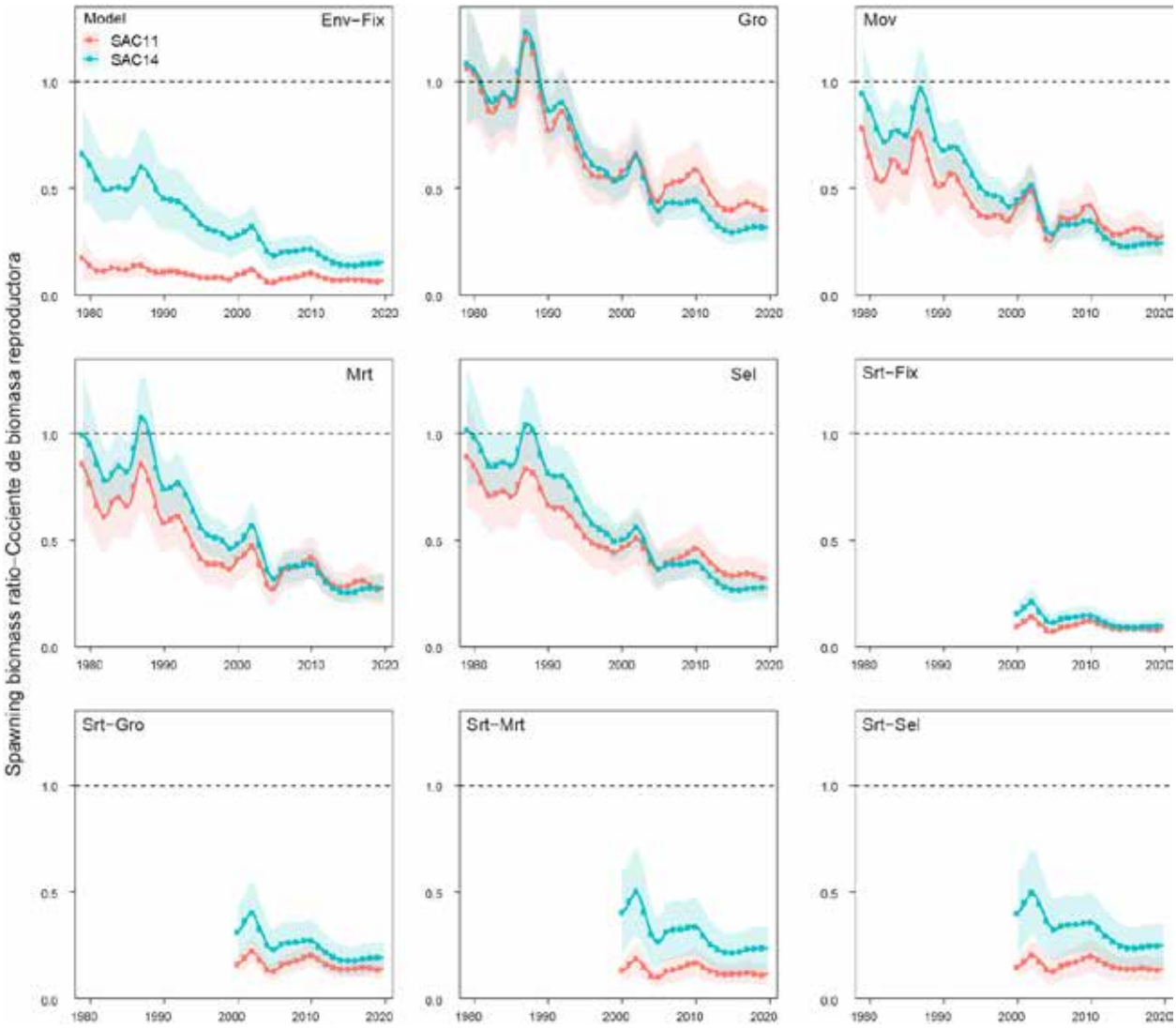

FIGURE 25. Comparison of spawning biomass ratio estimated by the nine reference models (steepness=1) from the last benchmark assessment (SAC11) and this exploratory analysis (SAC14). The shaded areas represent the 95% confidence intervals.

FIGURA 25. Comparación del cociente de biomasa reproductora estimado por los nueve modelos de referencia (inclinación=1) de la última evaluación de referencia (SAC11) y en este análisis exploratorio (SAC14). Las áreas sombreadas representan los intervalos de confianza de 95%.

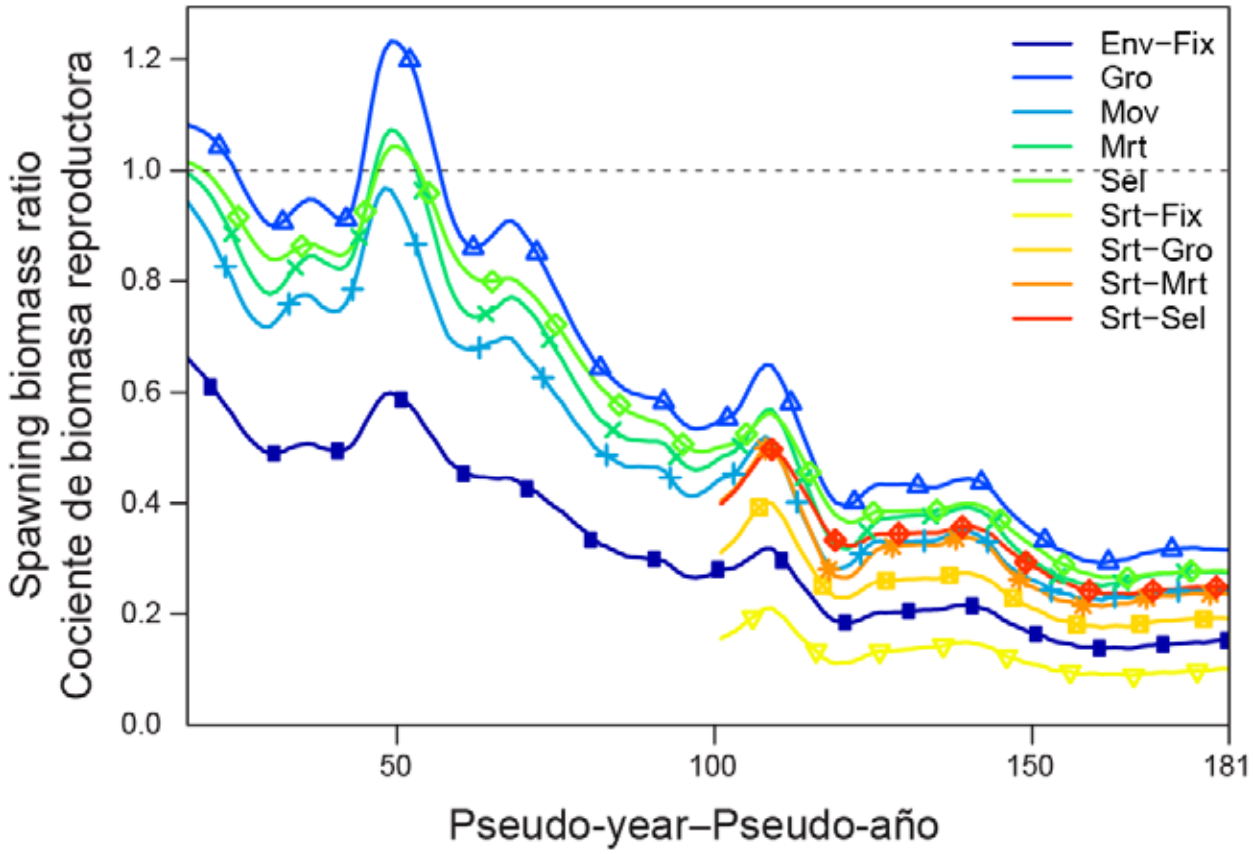

FIGURE 26. Comparison of spawning biomass ratio estimated by the nine reference models (steepness=1) considered in this exploratory analysis.
FIGURA 26. Comparación del cociente de biomasa reproductora estimado por los nueve modelos de referencia (inclinación=1) considerados en este análisis exploratorio.

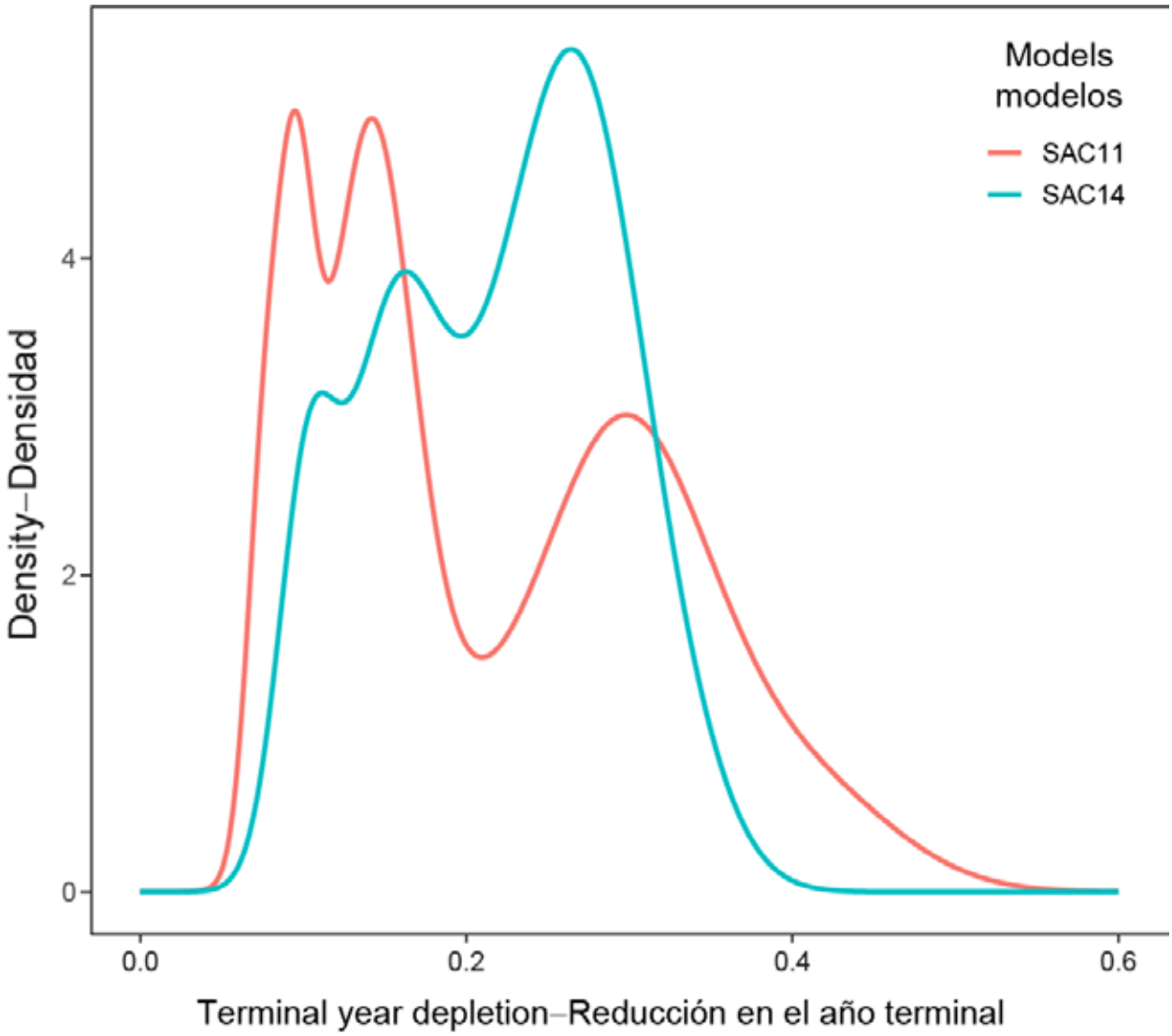

FIGURE 27. Comparison of model-combined joint distribution of terminal year depletion (spawning biomass ratio) between the last benchmark assessment (SAC11) and this exploratory analysis (SAC14). The reference models used to compute the joint distribution are equally weighted under each overarching recruitment hypothesis.

FIGURA 27. Comparación de la distribución conjunta de la reducción en el año terminal (cociente de biomasa reproductora) para los modelos en conjunto entre la última evaluación de referencia (SAC11) y este análisis exploratorio (SAC14). Los modelos de referencia utilizados para calcular la distribución conjunta se ponderan por igual bajo cada hipótesis de reclutamiento dominante.