SCIENTIFIC COMMITTEE
SEVENTEENTH REGULAR SESSION

Electronic Meeting
11-19 August 2021

Reference model of bigeye tuna using SEAPODYM with catch, length and conventional tagging data

WCPFC-SC17-2021/EB-IP-08

Senina, I.[1], Briand, G.[1], Lehodey, P.[1], Nicol, S.[2], Hampton, J.[2] and Williams, P[2]

[1] Marine Ecosystems Modelling, Sustainable Fisheries Management Division, CLS. 11 rue Hermes, 31520 Ramonville, France
[2] Oceanic Fisheries Programme, SPC, BPD5, 98848 Noumea, New Caledonia

Reference model of bigeye tuna using SEAPODYM with catch, length and conventional tagging data.

Inna Senina[1], Guillaume Briand[1], Patrick Lehodey[1], Simon Nicol[2], John Hampton[2], Peter Williams[2]

[1]Marine Ecosystems Modelling, Sustainable Fisheries Management Division, CLS. 11 rue Hermes, 31520 Ramonville, France

[2]Oceanic Fisheries Programme, SPC, BPD5, 98848 Noumea, New Caledonia

Contents

1 Executive Summary

1.1 Scope of work

This paper describes the updated reference model for Pacific bigeye tuna spatiotemporal dynamics as a result of integration of catch, length and tagging data into the SEAPODYM model to inform model parameters.

1.2 Key Outcomes

SC17 is advised of the following:

1. A next generation (NG) reference model has been prepared for Pacific bigeye tuna describing its spatiotemporal dynamics as a result of integration of catch, length and tagging data into the SEAPODYM model to estimate model parameters.

2. The inclusion of the tagging data estimates habitats and movement rates that are quite different from those estimated from fisheries data alone.

3. The NG reference model simulates reasonably well the observed distributions of tag recapture, which the previous bigeye reference model failed to do.

4. It (in comparison to previous reference models) predicts the spatial structure of bigeye that is more consistent with the distribution of catch data (Fig. 1). It provides better fits to major longline fisheries targeting bigeye and better or similar fit to the purse-seine fleet data.

1.3 Report details

The bigeye stock dynamics is described by a next generation SEAPODYM optimisation with parameter estimation based on catch, length and tagging data. The report provides a detailed technical summary of the work completed together with revised parameter estimates and relevant model diagnostics.

The development of the NG reference model was split in four major phases including i) global sensitivity analysis, ii) a parameter estimation from conventional tagging data only iii) update of historical fisheries data with structure adapted for SEAPODYM, and iv) optimisation study with full likelihood.

A global sensitivity analysis was performed for each type of data that is used in the parameter estimation. The results showed that catch and effort data mostly inform the model on reproduction and mortality processes, while length frequency data control the recruitment. Model parameters controlling movement rates are also observable from fisheries data, however integration of tagging data in the likelihood enhances this observability.

A comprehensive optimization study with the full likelihood was undertaken to minimize the biases due to fixed (model and forcing) parameters and to achieve the MLE solution characterized by biologically meaningful parameters, model validity for the independent datasets, prediction of the fish stock with the spatiotemporal structure that sustains fishing pressure and describes best the catch and length frequency data variability and observed distributions of tag recaptures.

a) CL model **b)** CLT model

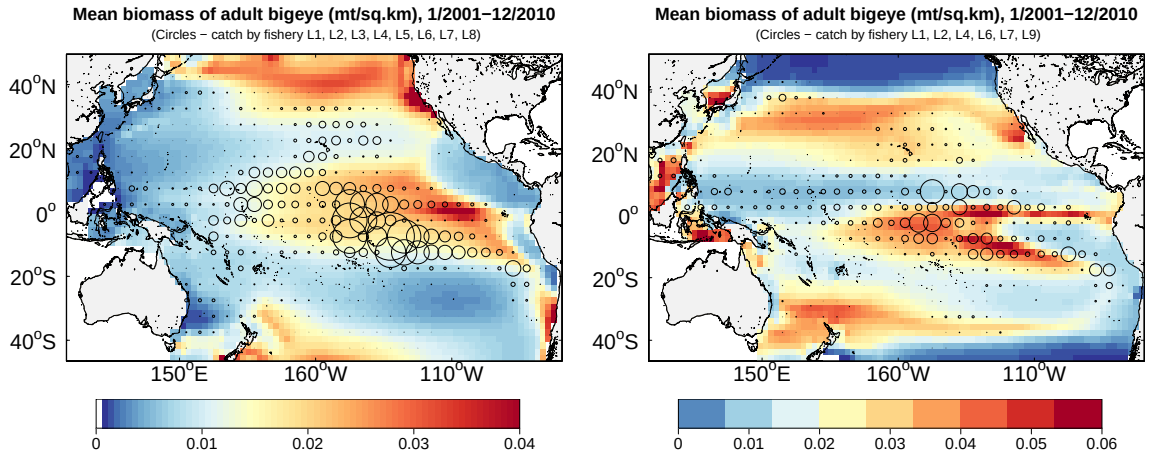


Figure 1: Biomass density of adult bigeye (in mt/km^2) predicted by CL and CLT models. The circles indicate positions of 5-degree cells with aggregated catch of selected fisheries. Note that fisheries are selected based on the mean length in catches estimated in each model. It is considered that fishery is targeting more adult tuna if the mean length in catches is higher than the those corresponding to the age at 50% maturity.

1.4 Remaining Actions

The following sources of uncertainty in the obtained model results need to be verified in further work:

1. The estimation of stock-recruitment and mortality parameters through extending the time period in optimisation runs;
2. The impact of the chosen primary production model on the MLE solution;
3. The role of climatological oxygen in the habitat and movement estimations;
4. The estimation of parameters describing the early life stages dynamics;
5. The degradation method of environmental forcing data.

1.5 Acknowledgments

The model forcing data was obtained from CMEMS reanalysis (add link). Special thanks to Olivier Titaud from CLS for sharing the integrated 2D fields of physical variables and satellite-derived primary production. We are thankful also to Marc Ghergariu from SPC for providing consolidated tagging dataset including information from various tagging campaigns and data reliability flag. The Inter American Tropical Tuna Commission has provided access to non-public domain data for the purposes of progressing the work programme of the WCPFC-SC. This work is supported by Pacific Community under a contract agreement between CLS and SPC.

2 Introduction

SEAPODYM is a numerical model developed for investigating population dynamics of tunas under the influence of environment and fishing. The underlying continuous equations of SEAPODYM are classical advection-diffusion-reaction equations with ageing term, describing the population dynamics in time, age and two-dimensional space. The quantitative modelling of tuna population dynamics with SEAPODYM has been continuously improving, including development of reference models that integrate fisheries catch and length data for Pacific Ocean populations of skipjack, yellowfin, bigeye and South Pacific albacore.

Besides complete geo-referenced fisheries datasets for a given species, the 'next-generation' of SEAPODYM reference models include integration of tagging data in the likelihood estimation approach, implementation of robust statistical methods for global sensitivity analysis and cost function profiling, enhanced algorithm of the fisheries data use within likelihood function and complete validation on independent data sets. This leads to better estimates of both stock size and stock spatial structure (Senina et al. 2020). Next generation reference model have been completed for the Pacific Ocean populations of skipjack and South Pacific albacore.

This Information Paper describes the developed "next-generation" reference model for Pacific bigeye tuna (*Thunnus obesus*). The previous reference model constrained by fisheries data only was presented in Senina et al. (2018, 2020). Here we report the main results from updated reference model and demonstrate how integrating tagging dataset influences the model parameters and allows improving model consistency with the data. In particular, we present the results of comprehensive sensitivity analysis based on an ensemble of $\approx 0.65 \cdot 10^6$ simulations preformed with 1) the reference model configuration evaluating separately catch and length-frequency likelihoods and 2) the model of tag dynamics with conventional tagging data in the likelihood 3) LF likelihoods and fisheries selectivity parameters. Then, we present the results of the optimization study performed with the full-likelihood, CLT, model. Finally we evaluate the fit and validate the CLT model using catch, length and tagging data. To do the latter, we use independent tagging data, and validate three optimal solutions of: 1) the previous CL reference model, 2) the model of tag movement integrating 6 years of tag recaptures and 3) the CLT model integrating 7 years of tag recaptures. The results of this study are discussed, highlighting both the improvement and deficiencies and suggesting the next steps aimed at improving the new reference model.

3 Data

3.1 Update of bigeye tuna fisheries data

The new structure of SEAPODYM fisheries for bigeye tuna is provided in Table 1. A fishery in SEAPODYM is defined by the homogeneous fishing catchability and selectivity, unique fishing gear, having the same target species and fishing strategy. To achieve this, the long-line fleet data were processed individually and split into bigeye, yellowfin and albacore target "fleets" depending on the proportions of each species in the total catch. Then the fleet data were aggregated into fisheries by similar statistical properties

of the CPUE derived from effort-and-catch (EC) dataset and of the length frequency (LF) distribution in the same area (L1, L2, L3, L4, L7). Long-line fisheries with heterogeneous CPUE but similar LF properties were aggregated to so-called 'dummy' fisheries, accounted in the model without fishing effort and catchability using catch removal method (L6, L8 - L11). For purse-seine fleets, the statistical properties of catch and length frequency were computed by the school type: unassociated, log, drifting or anchored FADs, animal association. Then the fleets with the best data coverage in both, EC and LF, datasets were selected to be used for fishing mortality and catch prediction based on fishing effort (S12, S14, S15, S17). The other fleets with same school type and similar selectivity were aggregated to 'dummy' fisheries (S13, S16, S18). All pole-and-line catches, which can be considered rare and accidental for bigeye, were combined into fishery P19. The mixed-gear domestic fisheries of Indonesia, Philippines and Vietnam were combined into O20 fishery.

All longline catch and effort fishing data are available at a resolution of 5° x 5° x month while for surface gears (purse seine and pole-and-line) the resolution is 1° x 1° x month, excepted for Philippine and Indonesia fisheries. Size frequency data are at a resolution of 5° x 5°, 5° x 10° and 10° x 20°. A screening for the outliers based on the CPUE variance in 5° was performed with the aim to detect and correct incidental and erroneous fishing effort. Finally, we checked that the geo-referenced catch dataset for the Pacific fisheries matches closely the total landings declared by countries (Fig. 2).

As seen on Fig. 2 and Fig. 3) providing the overview of spatial distribution of catches in the two previous decades, there was a persistent decline in long-line catches since the beginning of 2000, moreover, the surface gears extended towards the central tropical Pacific Ocean, resulting in higher catches in this area and the decline in the EPO.

3.2 Conventional tagging data

Conventional tagging data are integrated into the optimization method in SEAPODYM essentially to improve the estimates of habitat and movement parameters that are critical to control the overall population dynamics. The approach considers only fish that have been recaptured, as these are the only data that contain information about potential movement (for details, see 5.2). Tags are aggregated into groups by their common time (month, quarter) of recapture. The groups of tags were defined using compiled tagging datasets provided by SPC and IATTC from different tuna tagging programs between 1967 and 2019, which contains 18735 records on released and recaptured bigeye. Note that 99% of all records in this dataset occur between 2000 and 2017, with only 23 released bigeye before 2000 and 39 after 2017. Besides, 38% of 2000-2017 tag recaptures occurred before 2008 and 57% were recaptured between 2008 and 2013. In 2012 number of releases dropped and from 2014 the number of recaptured bigeye started to decline rapidly. For more details on tagging programs see [SPC-OFP]. The tagging data temporal coverage and the distribution in terms of mean length and time at liberty are illustrated in Fig. 4. Two characteristic periods with massive tagging of bigeye can be selected within 2000-2013 time range, clearly distinguished by the length of released tunas, the positions of release and the distributions of recaptures. During the first one, from early 2000 to mid-2007, the mean length of bigeye at release was 77 cm, they were mostly tagged at three release position around equator at 95°W longitude and a few dozens of tunas were tagged

and recaptured in the warm pool area. During the second period, from mid-2007 to the end of 2013, much smaller bigeye, 56 cm on average, were tagged extending the area of release towards central Pacific Ocean. The maps and the data statistics are depicted in Fig. 4, the red and blue lines corresponding to the first and the second tagging periods respectively.

Unfortunately, neither sub-set of conventional data represents the whole population, but essentially the juvenile fish that are associated to surface schools and caught with surface gears. First, juvenile and immature bigeye tunas were tagged and released primarily in six lon-lat positions along the equator, three in the western and central equatorial region and three in the EPO. Multiple minor releases were done all over the warm pool area. Second, 89% of all tagged and recaptured tunas are recaptured before one year of liberty at sea, 10% between one and two years and only 1% more than two years. In consequence, with size at 50% maturity being 115 cm, the majority (93%) of recaptures are still immature tunas. As a consequence, observing only part of the population may present difficulties for estimating model parameters, in particular those responsible for dynamics of an unobserved fraction of a population.

For the purposes of reducing computational costs and integrating maximum information to inform dynamics processes through all model dimensions, the second sub-set of recaptures, from July 2008 to June 2014, constituting about 50% of the dataset was used in the present work in order to estimate model parameters, while the first one, with recaptures from May 2000 to June 2007, was left for validation.

3.3 Environmental forcing

SEAPODYM uses ocean physical (temperature and horizontal currents) and biogeochemical (primary production, euphotic depth and dissolved oxygen concentration) variables. Physical variables are provided by ocean general circulation models (OGCM), either from hindcast simulations or reanalyses. They both provide the same outputs but in the first case the ocean model is forced by atmospheric variables only. In reanalyses, the simulation assimilates observations of oceanic variables (e.g. Argo profilers, satellite sea surface temperature and altimetry) to produce more realistic circulation patterns, especially at mesoscale resolution. Primary production and euphotic depth can be simulated by a biogeochemical model coupled to the physical model or estimated from satellite ocean colour, solar radiation and sea surface temperature data [Behrenfeld and Falkowski, 1997]. The euphotic depth is used for averaging the physical data and the dissolved oxygen concentration data over three vertical layers representing i) the epipelagic layer, between the surface and 1.5 the euphotic depth; ii) the upper mesopelagic layer, between 1.5 and 4.5 the euphotic depth; and iii) the lower mesopelagic layer, between 4.5 and max[1000m; 7.5 euphotic depth]. Dissolved oxygen concentration is provided by most biogeochemical (BGCH) models. However, in the case when the BGCH model was not coupled with the physical model, providing forcing variables to SEAPODYM, a monthly climatology based on all available observations can be also used, although in that case, it cannot represent the interannual variability (e.g. ENSO). SEAPODYM tuna habitats rely of the biomass distributions of micronekton functional groups, which need to be computed with the SEAPODYM-LMTL (Low and Mid Trophic Levels) model. There are 6 micronekton groups inhabiting the epipelagic, upper and lower mesopelagic layers.

In this study we use ocean environmental forcing datasets to cover the period 1998-2019, with the start of the period determined by the beginning of satellite observation of ocean color. The physical variables are the ocean hindcast simulated with the Mercator-Ocean global eddy-permitting NEMO OGCM in the ORCA025 configuration under project GLORYS. This simulation was forced by the atmospheric reanalysis ERA-INTERIM and did not have data assimilation. To distinguish it from GLORYS reanalysis with data assimilation hereafter this dataset is called GLORYS-free. These physical variables are associated with the primary production derived from satellite data and the WOA climatology for dissolved oxygen concentration (see Table 2).

All forcing variables were interpolated to a regular 2° grid and degraded to monthly time step. See Table 2 for details on the original NEMO-PISCES-INTERIM outputs and Figures A1 - A3 showing the mean state of key environmental variables of each forcing data set.

4 Model

4.1 Model structure

The model PDE equations and the functional links between species habitats and environmental drivers are described in [Senina *et al.*, 2020a]. As for skipjack, the seasonal spawning migrations were set off, hence assuming that bigeye is an opportunistic spawner, does not undertake the distant spawning migrations and spawns whenever the feeding habitat conditions become also favorable for spawning. Three life stages are considered: larvae (0-1 month of age), small juveniles (1-3 months) and adults (older than 3 months of age, including both immature and mature part of the population).

Unlike in earlier SEAPODYM applications, both parameters of thermal function, optimal temperature and preferred thermal range were set to be estimated independently for spawning and feeding habitat. The previous assumption of the link between the thermal preferences for spawning (and hence larvae survival) and those for feeding of the oldest individuals, driven by the individual's length [Lehodey et al., 2008] could not be verified in skipjack and albacore applications [Senina et al., 2020a,b].

The fisheries mortality was computed using both methods: using fishing effort according to Gordon-Schaefer model, or using catch removal method (Senina et al., 2020a), applied to the fisheries according to their data quality and target species properties (Table 1).

4.2 Numerical configuration

The model PDE equations [Senina *et al.*, 2020a] are numerically solved on a 2° regular grid on the spatial domain covering the Pacific Ocean domain $\Omega = \{x \in (88°E, 69°W), y \in (55°S, 65°N)\}$ and monthly time step. The age is discretized between 0 and $a_{max} = 14$ (years) into monthly cohorts resulting for bigeye in 85 cohorts, so that the first seven years are split into 84 monthly age classes and the oldest individuals are aggregated into a single A+ cohort of length 7 years.

4.3 Static model parameters

The estimates of length-at-age, weight-at-age and maturity-at-age relationships were taken from the 2014 MULTIFAN-CL assessment report [Farley et al., 2014] and interpolated to the model age structure (Fig. 5). The more recent estimates of the bigeye growth [**?**], implying faster growing, but substantially smaller adult bigeye, were implemented within the CL reference model and found to provide unsatisfactory fit to the EPO fisheries data. In the current model application the continuous maturity-at-age function was provided to compute the spawning biomass. According to this function, the 50% of stock is mature at the age 36 months (mean age of the age class 36.5), which corresponds to the mean length of 115.7 cm and mean weight 34.4 kg.

Since accessibility to the vertical layer and therefore the abundance of accessible prey depend on the species size, it is practical to link the habitat computation to the growth rate in order to avoid unnecessary computations when the growth slows down. In the current model configuration, the habitats were updated only if the mean length of the age class has been changed more than 4 cm compared to the previous age class. The result of this technique was the reduction from 85 two-dimensional habitat fields, equal to the number of monthly age classes, to only 33 habitat fields (see Fig. 5).

4.4 Initial conditions

During the first set of optimisation runs the estimated state of the INTERIM model was used as the initial conditions. Furthermore, the predictions of the first five years of simulation were not included into the likelihood to reduce the effect of the initial conditions of the MLE solution.

5 Methods

The aim of the current study was to develop a quantitative modelling application for bigeye tuna using all available geo-referenced fisheries data and historical conventional tagging data. Since this task requires extensive computations given the use of highly dimensional model of spatial dynamics and involves solving complex non-linear optimization problem with dozens of unknown parameters, the preliminary exploratory study was done in order to compute parameters sensitivity given different types of observational data and to explore the capacity of the tag movement model alone to predict observed distributions of tag recaptures. The results of this exploratory study can be seen in previous SC report [Senina *et al.*, 2020c]. Further, the following steps were undertaken to achieve the optimal parametrisation of the full population dynamics model:

- Processing and compilation of the updated fisheries dataset, which included the new definitions of fisheries based on the target species and statistical properties of the CPUE fleet data. The Hampel rule [Pearson, 2011] was used to detect the effort outliers, which were either removed (corresponding catch records were moved to so-called "dummy" fisheries) or corrected to the value giving the average CPUE in the 5-degree cell or its vicinity.

- Global sensitivity analysis was extended and performed for the fisheries selectivity parameters to leave only parameters, which can be estimated from the available data and hence to achieve more rapid convergence of optimisation method;
- Maximum likelihood estimation of parameters minimizing the log-negative likelihood with four terms: catch, length-frequency, tag recaptures and the average stock constraint over the WCPO region;
- Model validation based on statistical metrics computed for each type of data.

5.1 The global sensitivity analysis

A global sensitivity analysis (GSA) implemented in SEAPODYM (Senina et al., 2020a,c) is based on variance methods [Saltelli et al., 2008, Pianosi et al., 2016]. It allows evaluating model sensitivity to its parameters through the i) first-order, or main effect, indices S_i^F measuring the direct and independent from other parameters contribution of each parameter to the output variance and ii) the total-order, or total effect, indices S_i^T measuring the overall contribution from a parameter including its interactions with other parameters. The main and total effect indices are useful to rank and to exclude the non-influential parameters respectively. The parameter θ_i is not influential if and only if the index $S_i^T = 0$. Having $S_i^T > S_i^F$ indicates existing correlation with other parameters.

According to their definitions, the indices S_i^F can be computed in so called All-At-a-Time (AAT) SA experiments, i.e. randomly sampling all parameters at every model run. The evaluation of S_i^T requires One-At-a-Time (OAT) SA experiments, where only one parameter is randomly varied in a series (here 25) of model runs while others are fixed. In order to evaluate model sensitivity to its parameters given the information contained in each type of data, it is practical to set-up SA simulation study with three model configurations integrating either catch, length frequency or tag recaptures. Since the complementary SA in this study focused on selectivity parameters, only the likelihood terms dependent on fisheries data were used. Although selectivity parameters mostly define the length distributions of catch, they influence catch predictions and mortality as well, therefore the output function was set as $L^- = L_C^- + L_L^- + \beta$ (see next section for description of likelihood terms).

First, a large number $m = 2500$ of AAT simulations was run in parallel, each with a different set of $n = 58$ parameters sampled randomly. Second, the set of parameter combinations of length $l << m$ providing the lowest cost function values found in the AAT simulations was selected to conduct OAT simulations, $l = 50$. Selecting the lowest function values allows the profiles that are too far from the optimum to be excluded from the OAT simulations. Overall, OAT ensembles use l parameter sets of length n, which each parameter sampled $k = 25$ times, hence resulting in $n \times l \times k$ simulations and same number of output function values. In addition, the following algorithm has been found very efficient to subsequently improve the likelihood function towards its optimum within OAT simulation study. It consists in fixing each parameter at its value providing the best among profiled output values after each of n series of $k = 25$ runs. Thus, at every iteration, the OAT is moving in a three-dimensional space of the likelihood projection on a pair of parameters conditioned by the best values of previously sampled parameter(s)

and the starting point. In the current SA study, overall 145000 simulations were run in both AAT and OAT ensembles.

5.2 Integration of conventional tagging data

Tagging data can primarily inform the model parameters controlling movement. The measurement model for tag recaptures describes the observed density of cohorts including tagged individuals only. Thus, re-defining the state variable of the advection-diffusion-reaction model used in SEAPODYM, which is the density of fish population $N(a,t,x,y)$ at age a, in time t and located at position $\mathbf{x} = (x,y) \in \Omega \in \mathbf{R}^2$, so that it represents the density of tagged cohort k, denoted $R_k(a,t,\mathbf{x})$, this model can describe movements of tagged sub-population only, $R_k \in N$. However, there are some caveats to consider before integrating individual movement data into a Eulerian model, which is not designed to predict the displacement of individual fish. First, after release, only a small number of tagged fish is recaptured, and the reporting rate is unknown. Second, the fish are recaptured by the fishing gear, the exact characteristics of which (fishing technique, effort) are poorly observed. To remove these sources of uncertainty, in SEAPODYM we use only released and recaptured fish. This allows simplifying the model as we do not need to deal with mortality term. Third, one needs to have enough individual data to make an assumption on validity of a Eulerian model, which is suitable to describe movement of a large number of individuals. It is therefore important to select the time period with massive release-recapture data, providing hundreds of tag recaptures for each model time step. For example, in the dataset selected in the present study, each cohort has 134 tagged fish on average. Last but not least, to transform the individual data to fish density and to account for the uncertainty of recapture positions, a bivariate Gaussian kernel for two independent variables (longitudinal and latitudinal coordinates) is applied to the observed recapture records. Also, the tags are aggregated into larger spatiotemporal strata and the tag movement model is solved on coarse spatiotemporal resolutions.

The cohort k in the tag movement model is not defined by the same origin (classical definition), but by the same ending, with k denoting the common time of recapture. Namely, a cohort of tags includes all tagged fish released at $\forall t_{\text{rel}} \in [t_0, k-1]$, and recaptured at time k, t_0 being the time of release of the first tag in the cohort. Then, modelling the dynamics of recaptured fish density $R_k(x,y)$ through times $[t_0,k]$ can be viewed as the inverse design of the problem: instead of initialising the PDE with the observed density of tagged cohorts of fish (with cohort being defined as ensemble of fish tagged and released at the same time and position) and then recapturing fish one by one at observed positions and times, we initialise the PDE with individual tags and find the solution of the PDE equation that yields the density of a given tag distribution at the moment of recapture. An important advantage of this approach is that it accounts precisely for the time at liberty of all tags within Eulerian model, which is otherwise impossible as the modelled quantity cannot be traced.

Note that the age dynamics, the movement rates, the habitat indices and the control parameters are the same as in the main SEAPODYM model. Hence, solving the tag recapture measurement model and population model simultaneously in the MLE framework, allows estimating the model parameters controlling movement. Obviously, this approach requires the assumption, that the movement of tagged fish obeys the same physical prin-

ciples and is led by the same environmental drivers that control the movement of all fish in the population.

5.3 Maximum likelihood estimation

The MLE approach used in SEAPODYM has been detailed for the skipjack tuna application in (Senina et al. 2020). The log-negative likelihood function, denoted as $\bar{L} = -ln(L)$ to be minimized consists of four terms $\bar{L} = \bar{L_C} + \bar{L_Q} + \bar{L_R} + \beta$, where the definitions of likelihoods describing the contribution of catch data $L_C\left(\theta|C^{obs}\right)$ was chosen to follow Poisson distribution for all fisheries with catch prediction method depending on the fishing effort and the normal distribution for all fisheries with catch removal method. The length-frequency of catch likelihoods $L_Q\left(\theta|Q^{obs}\right)$ were set to follow robustified normal likelihood for all fisheries (Hampton and Fournier, 2001). The normal likelihood function was used for the tag recapture density $L_R\left(\theta|R^{obs}\right)$. The stock constraint is a single contribution $\beta = \overline{B_{reg}} - B^*$, so the average over the entire time series biomass $\bar{B_{reg}}$ was computed over the WCPO region $(110E, 150W, 40S, 50N)$, and the fixed values B^* varied in the optimisation runs within 1500-1900Mt, with the lower value being the average total WCPO biomass estimated by the Multifan-CL model [McKechnie et al., 2017].

5.4 Optimisation runs

The optimization runs were configured over the period 2002-2018, with the first five years of model predictions and data not augmented to the likelihood to reduce the impact of the population structure imposed by the initial conditions. Therefore, the model parameter estimation was essentially driven by twelve full years with fisheries data over 2007-2018 and six years of tag recapture data over 2008-2014, including sixty monthly cohorts. Leaving four years out in the beginning of the time series (the forcing data availability period is 1998-2019) was dictated by the need to have enough environmental forcing to regenerate the initial conditions and to avoid starting with the strong ENSO event during 1998-2000. As for the last year, the geo-referenced fisheries data represent 86% of total landings in 2018 and is likely less complete in 2019 (Fig. 2). Nevertheless, optimization runs with the full CLT likelihoods and 17-years long simulations demand 62Gb of RAM and take 22-26 minutes per function evaluation (FE) depending on machine's CPU.

Besides the need to search for global the minimum solution, which requires so-called "jitter" runs, it is important to explore different sources of uncertainty in the estimated solution associated to fixed parameters and their boundaries, the model structure, the choice of the likelihoods, forcings and observational data. Since, it is practically impossible to make the exhaustive optimisation study in such complex, computationally demanding and highly-dimensional problem, the achievable objective is to find such an MLE solution, which satisfies pre-defined model validity criteria and provides reasonable fits to the independent data (next section). A total of 88 CL and CLT optimisations were performed with different model configurations in order to explore:

- The impact of the forcing: CL likelihood were optimised with previously estimated model parameters (INTERIM CL reference solution) and fixed or released fisheries parameters; 22-78 parameters; 18 optimisation runs with CL likelihood only.

- The validity of the habitat and movement parameters estimated from the tagging data only (Senina et al., 2020) to describe fisheries data; 12 CL and 15 CLT runs.
- The impact of the fixed selectivity parameters set to values found in the OAT profiling on the MLE solution; 65-76 control parameters; 2 CL and 1 CLT run.
- The sensitivity to the stock-recruitment parameters; 63 control parameters; 15 CLT runs.
- The use of SST instead of integrated epipelagic temperature in the spawning habitat computation; 63 control parameters; 3 CLT runs.
- The use of seasonal spawning migration mechanism in the movement habitat; 63 control parameters; 5 CLT runs.
- The sensitivity to the boundaries for the optimal temperature at age 0; 61-63q control parameter; 2 CLT runs.
- The use of catch removal method to predict purse-seine catch; 57 control parameters; 6 CLT runs.
- Sensitivity to the stock constraint; 57 control parameters; 4 CLT runs.

  Additional 5 CLT runs were launched with the reset initial conditions and the calibrated linear trends for the catchability parameters to assure zero trend in the residual $C^{obs} - C^{pred}$ time series.

5.5 Model validation

To conclude whether the obtained solution of the optimisation problem provides the best parametrisation given the model, the forcing and the data, each optimisation run has to be analysed and validated. First, the quality of the fit to the data being used in the minimization, is evaluated. This step is done with help of statistical metrics, which are selected depending on the type of the data: i) the R-squared goodness of fit, measuring how much the model is a better predictor to the data than the mean of the data; ii) the squared Pearson correlation coefficient, measuring the proportion of the variation in data described (explained) by the model iii) the root-mean-squared-error (RMSE) and the normalized root-mean-squared-error (NMSE); iv) residual variance and temporal bias; v) relative error; vi) model to data variance ratio.

Second, the parameter estimates are examined and confronted with the existing knowledge on the modelled species. Some important biological characteristics, such as thermal preferences, spawning sites and seasonality, and the species life span are reported in scientific literature.

Finally, the model is validated using independent sets of fisheries and tagging data, which were not used in the likelihood.

6 Results

The selected MLE reference solution of the CLT model was obtained with the configuration in which the catch and LF of catch from five longline fisheries targeting bigeye (L1 - L4 and L7) were predicted based on Gordon-Shaefer approach, i.e. using fishing effort. The catch of all other longline fisheries as well as of all purse-seine fisheries was predicted based on the catch removal method. The initial condition for January 2002 was generated with the optimal parameters obtained in the series of optimizations with the above set-up. The results of GSA (see Appendix, Fig. A4- A8) allowed significant reduction of selectivity parameters to estimate without any impact on the MLE solution. Among 58 selectivity function parameters controlling the prediction of catch and LF data of twenty fisheries, only 34 were left to control the fit to corresponding data.

6.1 Optimal parameters

This section describes the estimates of main model parameters and their role in the key dynamic processes: reproduction, survival, movement. The results of estimation of fisheries parameters are briefly described as well.

6.1.1 Spawning habitat and reproduction

Despite of the use of the most comprehensive and complete fisheries and tagging datasets, in the absence of observational data to constrain the dynamics of the early life stages, the estimates of the spawning parameters remain highly uncertain. The optimal temperature for spawning was stuck to the lower boundary fixed at $26°C$, while the standard deviation of the Gaussian function was stuck to the upper boundary, $3.75°C$, meaning that function minimizer basically tends to extend the spawning to larger geographic zones. Such behaviour of the function minimizer obviously indicates the lack of signal in fisheries and tagging data for the spawning success, so the spatial distributions of larvae are simply driven by the back-tracing of the exploited ages observed by the data. Obviously, the observed part of the population may constrain the spatial distribution of early life stages only partially, leaving the model free to predict unobserved stock emerging from excessively extended spawning sites.

The boundaries for the thermal preferences for spawning were set to $(26, 3.75)$ to remain in the range of surface temperatures where the bigeye larvae are observed, hence providing plausible estimation of larvae distribution. Indeed, taking into account that the resulting number of larvae depends not only on temperature, but also on density of prey (primary production converted to the wet weight of plankton), predator (surface micronekton density during the day and the twilight) and the density of reproducers (spawning biomass), the overall reproduction parameters provide the maximal number of larvae associated with SST $25 - 26°C$ with 85% of the frequency distribution confined between $21°C$ and $30°C$ (Fig. 7). To compare with observations, Reglero et al. (2014) reported that bigeye larvae worldwide are found in waters with SST between $21.7°C$ and $30.2°C$. The rest of the spawning parameters, α_P, α_F and β_F are well estimated and the estimated functional links are shown in Fig. 6 (see *prey* and *predator* functions). Note that spawning habitat is modulating the spawning outcome, which depends not only on larvae survival, but also the density of adults and hence on the suitability of the spawning

site for adults. Very high density of micronekton is obviously positive for concentrating adults, which prey on the micronekton, however, it is negative for larval survival as the micronekton prey on the larvae. So, the form of the predator function provides the optimal range of micronekton density that is favorable for both, tuna larvae and the spawners.

The seasonal dynamics of spatial distributions of (recruited at age 30 days) larvae in the sub-tropical regions are clearly driven by the temperature (Fig. 8). The apparent discontinuity in the larvae density along the equator, showing near zero larvae density is likely a model artefact, especially in the central part of the Pacific. First, bigeye larvae were observed in the equatorial Pacific between 180E and 130W [Nishikawa et al., 1985]. Second, the divergence between the north and south equatorial currents is known to create an upwelling phenomenon and an increased productivity in the surface layer (Fig. A2), which should a-priori create favorable conditions for larvae feeding. However, the density of surface micronekton being larvae predators is estimated to be optimal for the spawning success at values $0.4-2g/m^2$ (Fig. 6). This density predicted by the SEAPODYM-LMTL model, is characterised by very low values in the equatorial band west of Galapagos Islands (Fig. A3 in the Appendix). This result can only be explained by the strong sub-surface currents, so that the integrated epipelagic currents are pushing the micronekton off the equator.

The reproduction rate in the Beverton-Holt function is well estimated, while the stock-recruitment relationship cannot be estimated given a very short time period in the optimization. In the last phase of optimization, it converged to the optimal value an optimal value close to the one previously estimated with the CL reference model (Table **??**, parameter b). Its estimation remains subject to further work once the longer forcing datasets will be available.

6.1.2 Species demography and population structure

All four parameters of mortality-at-age function were estimated within their boundaries, the coefficient of variation with age was fixed to provide 10% variability of local mortality rate depending on the habitat index. The latter mean that mortality decreases by 0-10% in favorable habitat index, $H \geq 0.5$, and increases by 0-10% if habitat $H \leq 0.5$, with value $H = 0.5$ not affecting local mortality rates. The mortality estimation gives the bigeye mortality rates ranging from 0.11 mo^{-1} at age 0.5 months down to 0.02 mo^{-1} at age 7 years. These rates are equivalent to only 25% survival through the first month of life and 79% survival each year starting at the age of seven years. The non-linear shape of the mortality curve implies that 95% reduction of the recruited cohort occurs by the age of four years in fished population and 5.5 years if not fished (Fig. A9, lower right panel). The decline in abundance after the age of seven years is very slow, resulting in a significant biomass of the last A+ age class, which includes all individuals between 7 and 14 years of age (Fig. A9, lower left panel). In terms of biomass, the modelled population is composed of 20% of juveniles (small juveniles and immature adults in SEAPODYM definitions) and 80% of mature adults.

6.1.3 Feeding habitats and movement

Feeding habitat is the key variable determining the spatial structure of the stock as it drives the fish movement. The estimation of feeding habitat parameters gives the estima-

tion of thermal preferences and feeding preferences for bigeye tuna through ages 3 months to 10 years (the mean age of the A+ class). The thermal preferences curve is shown on Fig. 6 (panel *Adult habitat*). The preferred ambient temperatures of the young fish range from 18°C and 21°C, the later being the lowest estimated temperature for larval habitat. The mature fish are predicted to prefer temperatures $15 - 18$°C. Taking into account the estimated temperature tolerances, bigeye occupy waters with temperatures ranging between 13°C and 28.7°C (Table 3), which is consistent with findings based on acoustic telemetry and archival tagging data analysis (e.g. Brill, 1994; Schaefer and Fuller, 2002; Graham and Dickson, 2004). Note that the feeding habitat thermal parameter estimates are not very far from those estimated for the movement model of tagged fish.

The oxygen requirement parameter is usually well estimated from the fisheries data alone, given that thermal habitat parameter estimates are not biased. The critical oxygen value is well informed from the observed CPUE reduction in the habitat, which is otherwise accessible in terms of thermal preferences. The 50% accessibility requirement for dissolved oxygen of 1.49 ml/l is stipulated by the oxygen distribution in the mesopelagic layer (Fig. A1). Note that this estimation is obtained with the climatological fields of oxygen (Table 2).

In terms of feeding preferences, the estimation of relative contribution of micronekton shows that bigeye prefers feeding on resident epipelagic and upper and lower mesopelagic groups migrating to epipelagic layer at night. However, given much higher densities of micronekton groups in the mesopelagic layers predicted by the SEAPODYM-LMTL model (Fig. A3), the feeding habitat distribution of bigeye is driven by mesopelagic forage groups, including upper mesopelagic resident migrant groups as well as highly migrant lower mesopelagic group. The wide range of preferred temperatures and the contributions from resident and migrant forage are consistent with a feeding behaviour targetting the epipelagic layer at night and mesopelagic layer during the day. The lower mesopelagic resident forage is inaccessible to bigeye due to the low oxygen levels in almost the entire EPO and in the tropical part of the WCPO fishing grounds, which is why the parameter eF_{33} was fixed to 0. The very large standard deviation in Gaussian thermal function, 5.5°C together with the estimates for the coefficients eF_{dn} provide too extended favorable habitats for young tuna, which varies essentially between 0.8 and 1 from 45°S to 40°N. At the same time the feeding habitats of large bigeye vary between 0 and 1 essentially driving the spatial distribution of adult bigeye shown in Fig. 11 (see panel without fishing). Therefore for the old age classes, the spatial heterogeneity of the feeding habitat and the biomass density is due to the estimated narrowing of thermal preferences based on the standard deviation decreasing from 5.5°C to 2.43°C with age.

The velocities of directed and random movements modelled as advection and diffusion of population density depend on the habitat estimation at age. As shown on Fig. 3 (panel *Movement rates*), the mean directed movement (including both passive and active transport) velocities range between 1 nmi/d[1] and 3 nmi/d through species age. Besides, the juvenile fish are predicted to move faster, meaning less residency than adult fish, whose horizontal movements are negatively affected by deep-diving vertical behaviour. The dispersive movements are estimated to increase with age from $< 1\text{nmi}^2/\text{d}$ at 3 months to $460\text{nmi}^2/\text{d}$ at the age of 7 years and 600 nmi^2/d being an average rate of the A+ group.

[1] nmi – nautical mile.

Note that these estimates are very close to those estimated by UKFsst model from archival tagging data provided the median values for horizontal speeds of bigeye 2.45 nmi/d and rate of dispersal 496.7 nmi^2/d [Schaefer et al., 2014].

6.2 Validation

The fit to the catch and length data provided by the MLE solution of the CLT reference model can be seen in Appendix, section A.4 *Fit to the catch and LF data*. The time series of aggregated catch by fisheries whose catch is predicted with Gordon-Shaefer method is highly correlated to observations, the residual mean is close to zero and the residual variance is small. Also, the three statistical metrics are evaluated for each fishery based on geo-referenced data and shown on Taylor diagram (Fig. 9). Note, the metrics on this diagram were computed over the full time series with the data, i.e. 1998-2019, while only 2007-2018 data was used in the likelihoods. The extension of the time series by 10 years improves the metrics: the MLE solution obtains (0.6,0.62,0.63) compared to (0.64,0.71,0.6) for all catch data, confirming the model validity for the independent catch data. The validation of fit for the length frequency data showed the persistent bias in the model predicting more large tuna in the catches. Further work is needed to verify the robustness of the mortality rates estimation.

Also we compared the skills of the model with three optimal parametrizations in describing the movement of tagged tunas: 1) reference CL model with only fisheries data in the likelihood, 2) tag dynamics model with its optimal parameters estimated on the sub-set 2008-2013 and 3) the reference CLT model and with fisheries data over 2007-2018 and tagging data sub-set 2008-2014. Then, the validation of optimal solutions was performed on the independent dataset, i.e. releases and recaptures data from 2000 to 2007, which were not used in parameter estimation in either of three models. So the MLE parameters of each model were used to run tag simulations for the 2000-2007 period, then the fit was compared between models. Note that in the optimization runs, the resolution used to compute the likelihood term for the tagging data was 6° in longitude, 6° in latitude and a 3-month time step, while the validation of the optimization results was undertaken on the model's spatio-temporal resolution, i.e. 2° and 1 month.

Neither sub-set of data can be described by the CL MLE solution, i.e. obtained with fisheries data only (see Fig. 10 and A12). The distributions of tags is driven by diffusion, hence the spatial structure has a significant south-eastern bias, showing the displacements of tagged tunas around three release positions with noticeable drift by equatorial countercurrent and the tag density distortion by the Peru current pushing the density patch from south-east. Also, spatial distributions for both time periods are characterised by excessive density extension in the latitudinal direction.

The optimal solution achieved with the 2008-2013 tagging data sub-set presents significant improvement in simulating the observed dynamics of tags released and recaptured before 2008 compared to the reference CL model. Although still presenting some eastern bias, more that two thirds of tags released at 95°W moved towards central Pacific. The latitudinal extension is well reproduced with more bigeye moving to the southern hemisphere both from eastern and western positions of release. The tag simulation with the CLT model parameters is certainly improved compared to the CL model, however, it lacks the westward movement and shows over-dispersion of density in latitudinal direction. The

evaluation of the fit obtained by the MLE method demonstrated the observed movement of bigeye within the equatorial band 10S-10N, although some latitudinal over-dispersion is also apparent.

6.3 Model predictions

The model state vector gives the number of fish at age, time and position over the entire life cycle of the species, the time period when environmental data are available and over model domain respectively. The young bigeye distributions in the EPO are mostly driven by the recruitment and ocean circulation. This result is attributed to mostly homogeneous and highly favorable feeding habitat in this area. The moderate movement in the latitudes $10 - 20°$N of the central Pacific, results in the presence of immature bigeye in the the north Pacific gyre. The spatial structure is also different from those at recruitment in the warmpool, where young bigeye are estimated to be highly exploited by the purse-seine fisheries, which is clearly visible comparing the maps of this life stage with and without fishing (Fig. 11).

The mature adult distributions are consistent with the spatial distribution of long-line data. Given the similarity of estimations for key thermal and movement parameters provided by TAGS and CLT models (Tableřefparameters), adult spatial structure seems to be well constrained by the tag release-recapture data. The model clearly shows the existence of three sub-stocks in bigeye Pacific population, with the largest adult sub-stock in the tropical Pacific and two smaller sub-stocks in the sub-tropical zones of southern and northern Pacific, with the latter extending towards tropics in the central Pacific and south-east of Hawaii (see Fig 11). Note that high densities predicted in the Indonesian EEZ and in the South China Sea are highly uncertain due to closed boundaries and coarse resolutions used in the simulation.

6.3.1 Fisheries impact

Fishery impact estimated as $FI = 1 - B_F(t)/B_{F=0}(t)$ was computed from the total biomass $B(t)$ predicted by SEAPODYM with actual fishing mortality F and without fishing F=0. Note that the simulations with and without fishing were run starting from the same initial conditions representing the fished stock, so the fishing impact is evaluated only for the years 1998-2019. Since the stock was already reduced prior to 1998, our estimations are likely lower than expected if the virgin stock was modelled starting the simulation from pre-industrial times. The monthly percentage fishery impact on bigeye depicted on Fig. 12, shows the evolution of fishery impact on immature and mature adults, with an average value of 49% in early 2018. The maximal percentage fishery impact on young bigeye is estimated to be 30% Pacific-wide, however, local fishery impacts are estimated to be very high, up to 90%, in the warm pool area. Spatial maps of fishery impact computed from SEAPODYM outputs for the adult life stage indicate local depletion of adult biomass in 2019 from 35% in the sub-tropical fishing grounds to 65-75% in the tropical areas, with lower values in the EPO.

6.3.2 Pacific-wide stock estimation

SEAPODYM estimates bigeye stock biomass between 1.25-2.5Mt in the WCPO and nearly the same biomass, 1.2-2.2Mt in the EPO (Fig. 13). Compared to the stock assessment with the Multifan-CL model for WCPO [McKechnie et al., 2017], SEAPODYM shows higher biomass in the beginning of time series and more rapidly declining stock of bigeye in WCPO, with similar biomass estimations only in 2011. Although the rate of decline depends mostly on the stock structure (in space and age dimensions) and the impact of fishing, it should be noted that in the absence of forcing prior to 1998, the initial condition for model state vector in SEAPODYM, i.e. the biomass in 1998 remains uncertain. The increasing trend estimated by Multifan-CL starting 2012 is not captured by SEAPODYM. This increasing trend in Multifan-CL is estimated to occur in region 4 (170°E-150°W, 10°S-20°N), region 5 (140°E-170°E,40°S-10°S) and region 7 (110°E-140°E,10°S-20°N), while SEAPODYM indicates only a slowing of the decline in regions 3, 5 and 7 (Fig. 14).

7 Conclusion

In the present study we used geo-referenced fisheries data and conventional tagging data to inform parameters of spatiotemporal model SEAPODYM with the aim to build quantitative model for Pacific bigeye population. The global sensitivity analysis performed for all types of data and all model parameters was very helpful in detecting, fixing unobservable control parameters, configuring the optimisation runs and facilitating the process of searching for global minimum. Thanks to conventional tagging data integrated into the MLE method in SEAPODYM, we obtained improved estimates of model dynamics parameters, better representation of population spatial structure and plausible fit to the independent datasets.

Note, that as in all SEAPODYM reference models, the MLE solution is tightly linked to the environmental forcing and the structure of data being used. It might be interesting to test the GLORYS reanalysis based on data assimilation coupled with the biochemical model providing primary production, euphotic depth and dissolved oxygen. Also, the use of the high-resolution Japanese long-line CPUE data in the MLE approach might better constrain the habitat parameters throughout the selected ages. The biases in the estimation of feeding habitats of young tunas still need to be explored.

Other work and developments are planned to verify the robustness of the obtained solution and to further improve the current bigeye model. They include incorporation of linear trends parameters for catchabilities into parameter estimation; a better estimation of the initial state vector using extended forcing time series and backtracking technique relying on MLE estimates; the use of surface currents instead of mean epipelagic currents to model the passive drift of larvae; estimation of the uncertainty of model predictions. Additional analyses, such as evaluation of connectivity between identified bigeye substocks will be helpful in understanding the mechanism of their interactions.

References

[Behrenfeld and Falkowski, 1997] Behrenfeld, Michael J; Falkowski, Paul G. 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography*, 42(1), 1-20, https://doi.org/10.4319/lo.1997.42.1.0001.

[Brill, 1994] Brill, R. 1994. A review of temperature and oxygen tolerance studies of tunas pertinent to fisheries oceanography, movement models and stock assessments. *Fisheries Oceanography* 3:3, 204-216.

[Graham and Dickson, 2004] Graham, J.B., Dickson, K.A. 2004. Tuna comparative physiology. *Journal of Experimental Biology*. doi:10.1242/jeb.01267.

[Harley et al., 2014] Harley, S., Davies, N., Hampton, J., McKechnie, S. Stock assessment of bigeye tuna in the Western and Central Pacific Ocean. WCPFC-SC10-2014/SA-WP-01, Scientific Committee tenth regular session, Majuro, Republic of the Marshall Islands, 6-14 August 2014.

[Farley et al., 2014] Farley, J., Eveson, P. Krusic-Golub, K., Sanchez, C., Roupsard, F., McKechnie, S., Nicol, S., Leroy. B., Smith, N, Chang, S-K. Project 35: Age, growth and maturity of bigeye tuna in the western and central Pacific Ocean. WCPFC-SC13-2017/SA-WP-01, Scientific Committee thirteenth regular session, Rarotonga, Cook Islands, 9 – 17 August 2017.

[Lehodey et al., 2008] Lehodey, P. and Senina, I. and Murtugudde, R. 2008. A spatial ecosystem and populations dynamics model (SEAPODYM) – Modeling of tuna and tuna-like populations. *Progress in Oceanography*, 78:304–318.

[McKechnie et al., 2017] McKechnie, S., Pilling, G., Hampton, J. Stock assessment of bigeye tuna in the western and central Pacific Ocean. WCPFC-SC13-2017/SA-WP-05, Scientific Committee thirteenth regular session, Rarotonga, Cook Islands, 9 – 17 August 2017.

[Morel and Berthon, 1989] Morel, A, J-F Berthon. 1989. Surface pigments, algal biomass profiles, and potential production of the euphotic layer: Relationships reinvestigated in view of remote-sensing applications. *Limnol. Oceanogr.*, Volume 34: 1545-1562.

[Nishikawa et al., 1985] Nishikawa, Y., Honma, M., Ueyenagi, S., Kikawa, S., 1985. Average Distribution of Larvae of Oceanic Species of Scombrid Fishes, 1951–81. Contribution of the Far Seas Fisheries Research Laboratory, 236. Fishery Agency of Japan, pp. 1–99.

[Pearson, 2011] Pearson, R. K.: Exploring Data in Engineering, the Sciences, and Medicine, Oxford University Press, New York, 2011.

[Pianosi et al., 2016] Pianosi, F., Beven, K., Freer, J., Hall, J., Rougier, J., Stephenson, D., and Wagener, T. 2016. Sensitivity analysis of environmental models: A systematic review with practical workflow. Environ. Modell. Softw. 79: 214–232. http://dx.doiorg/10.1016/j.envsoft.2016.02.008.

[Reglero et al., 2014] Reglero, P., Tittensor, D.P., Alvarez-Berastegui, D., Aparicio-Gonzalez, A., Worm, B. 2014. Worldwide distributions of tuna larvae: revisiting hypotheses on environmental requirements for spawning habitats. *Marine Ecology Progress Series*, Vol. 501: 207–224. doi: 10.3354/meps10666.

[Saltelli et al., 2008] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al. 2008. Global sensitivity analysis. The Primer. John Wiley & Sons.

[Schaefer et al., 2014] Schaefer, K., Fuller, D., Hampton, J., Caillot, S., Leroy, B., Itano, D. 2014. Movements, dispersion, and mixing of bigeye tuna (Thunnus obesus) tagged and released in the equatorial Central Pacific Ocean, with conventional and archival tags. *Fisheries Research*, 161: 336-355.

[Senina *et al.*, 2020a] Senina, I., Lehodey, P., Sibert, J., Hampton, J., 2020. Integrating tagging and fisheries data into a spatial population dynamics model to improve its predictive skills. Can. J. Fish. Aquat. Sci. doi:10.1139/cjfas-2018-0470.

[Senina *et al.*, 2020b] Senina, I., Lehodey, P., Hampton, J., Sibert, J. 2020. Quantitative modelling of the spatial dynamics of South Pacific and Atlantic albacore tuna populations. Deep Sea Research II, doi.org/10.1016/j.dsr2.2019.104667.

[Senina *et al.*, 2020c] Senina, I., Lehodey, P., Nicol, S., Scutt Phillips, J., Hampton, J., 2020. SEAPODYM: revisiting bigeye reference model with conventional tagging data. WCPFC-SC16-2020.

[SPC Yearbook 2019] Tuna Fisheries Yearbook 2018. 2019. Western and Central Pacific Fisheries Commission, Pohnpei, Federated States of Micronesia. 155 pp.

[SPC-OFP] Project 42: Pacific tuna tagging project report and work-plan for 2020-2023. WCPFC-SC16-2020/RP-PTTP-01.

List of Tables

8 Tables

Table 1: Bigeye Fishing Dataset (DS2020). Definition of SEAPODYM fisheries in Pacific Ocean. The column 'Use' indicates the catch prediction method: FE) with effort or CR) catch removal.

ID	Description	Nation	Resolution	Use
L1	LL traditional BET target	Japan, Korea	5°, month	FE
L2	LL targeting BET	China and Taiwan	5°, month	FE
L3	LL targeting YFT	Japan, Korea	5°, month	FE
L4	LL targeting YFT	China and Taiwan	5°, month	FE
L5	Distant-water LL, albacore target	Asian fleets	5°, month	CR
L6	LL targeting BET and YFT	Vietnam, Philippines, Indonesia	5°, month	CR
L7	Longline BET target	USA, Australia, New Zealand	5°, month	FE
L8	Mixed-target LL	USA, Australia, New Zealand	5°, month	CR
L7	Longline bigeye target	PICs	5°, month	FE
L10	Mixed-target longline	PICs	5°, month	CR
L11	Albacore target LL	PICs	5°, month	CR
S12	PS free schools, WCPO	Korea, Taiwan, PNG	1°, month	FE
S13	PS free schools, Pacific	Others	1°, month	CR
S14	PS logs, WCPO	ALL	1°, month	FE
S15	PS FADs, WCPO	Korea, Taiwan, PNG, China	1°, month	FE
S16	PS FADs, Pacific	Others	1°, month	CR
S17	PS FADs, CPO and EPO	EC, SV flags	1°, month	FE
S18	PS marine mammals, Pacific	ALL	1°, month	CR
P19	Pole-end-line, Pacific	ALL	5°, month	CR
O20	Domestic fisheries, multiple gears	Philippines, Indonesia	5°, month	CR

Table 2: Forcing variables used in current SEAPODYM application. Note that table shows original resolutions, all variables were then interpolated onto SEAPODYM spatial and temporal resolutions 2°, 30 days.

Variable	**Description**	**Resolution**	**Time period**
	GLORYS-free		
T, u, v	Global ocean reanalyses without data assimilation, NEMO OGCM forced by atmospheric ECMWF reanalyses	ORCA025	1/1998 -12/2019
	Observations		
P, Z	EPPLEY-VGPM primary production and euphotic depth computed by Morel's model* from satellite-derived Chl-a	1/4°, 7 days	1/1998 -12/2019
O_2	WOA monthly climatology	1/4°, 30 days	clim. year
	SEAPODYM-LMTL		
F	Simulated six micronekton groups**	1/4°, 7 days	1/1998-12/2019

*Morel and Berthon, 1989; **QUID 2019.

Table 3: Parameter estimates for the following model configuraions: CL-I - population model with INTERIM forcing and likelihood with Catch and Length data only, TAGS-I - tag density dynamics model with INTERIM environmental forcing and 2008-2010 tagging dataset in the likelihood, TAGS-G - tag density dynamics model with GLORYS-free environmental forcing and 2008-2013 tagging dataset in the likelihood. Parameters marked by asterisks were fixed in optimization run. Parameter with [or] were estimated at their lower or upper boundary respectively.

θ	**Description**	**CL-I**	**TAGS**	**CLT**
	Reproduction			
σ_0	standard deviation in temperature Gaussian function at age 0, $^\circ C$	3]		3.75]
$T_0^\star$	optimal surface temperature for larvae, $^\circ C$	28.9		[26.0
α_P	prey encounter rate in Holling (type III) function, day^{-1}	0.073		0.024
α_F	Log-normal mean parameter predator-dependent function, g/m^2	[0.05		0.5
β_F	Log-normal shape parameter in predator-dependent function, g/m^2	1.054		0.77
R	reproduction rate in Beverton-Holt function, mo^{-1}	0.0038		0.0065
b	slope parameter in Beverton-Holt function, nb/km^2	16.05*		15.73
	Mortality			
$\bar{m}_p$	predation mortality rate age age 0, mo^{-1}	0.05*		0.11
β_p	slope coefficient in predation mortality	0.434		0.024
$\bar{m}_s$	senescence mortality rate at age 0, $\text{mo}^{-1-\beta_s}$	0.02		1e-10
β_s	slope coefficient in senescence mortality	[0		3.85
ϵ	variability of mortality rate with habitat index $M_H \in (\frac{M}{(1+\epsilon)}, M(1+\epsilon))$	0.5]		0.1*
	Habitats			
σ_1	standard deviation in temperature Gaussian function at first young age, °C	5.5]	4]	5.5]
T_1	optimal temperature (if Gaussian function), or temperature range for the first young cohort, $^\circ C$	[25	24.5	[23.2
σ_K	standard deviation in temperature Gaussian function at age K, °C	6.5]	[1	2.43
T_K	optimal temperature (if Gaussian function), or temperature range for the oldest adult cohort, $^\circ C$	[10	15]	15.48
b_T	allometric power coefficient for thermal preferences at age	[1	2.51	[0.7
$\hat{O}$	threshold value of dissolved oxygen, ml/l	0.758	[1.0	1.49
eF_{11}	contribution of epipelagic forage to the habitat	[0	[0.5	1]
eF_{22}	contribution of upper mesopelagic forage to the habitat	[0.1	3]	0.49
eF_{21}	contribution of migrant upper mesopelagic forage	0.05*	0.34	1]
eF_{33}	contribution of lower mesopelagic forage to the habitat	0.005	0*	0*
eF_{32}	contribution of migrant lower mesopelagic forage	0.05*	2]	0.37
eF_{31}	contribution of highly migrant lower mesopelagic forage	0.079	[0	1]
	Movement			
V	velocity at maximal habitat gradient and $A = 1$, BL/s	1.7895	0.13	0.15
A	slope coefficient in allometric function for tuna velocity	0.75*	0.73*	0.75*
σ	multiplier for the maximal diffusion rate	1.4091	1e-4	1.05
c	coefficient of diffusion variability with habitat index	0.5*	0.93*	0.9*

List of Figures

9 Figures

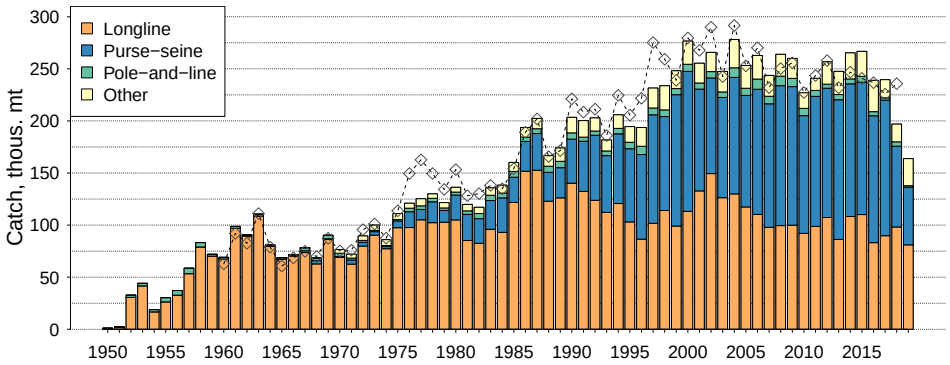


Figure 2: Total annual bigeye catch aggregated from geo-referenced catch (Pacific-wide) used in SEAPODYM analyses. Dashed line corresponds to total landings of bigeye (SPC Yearbook, 2019).

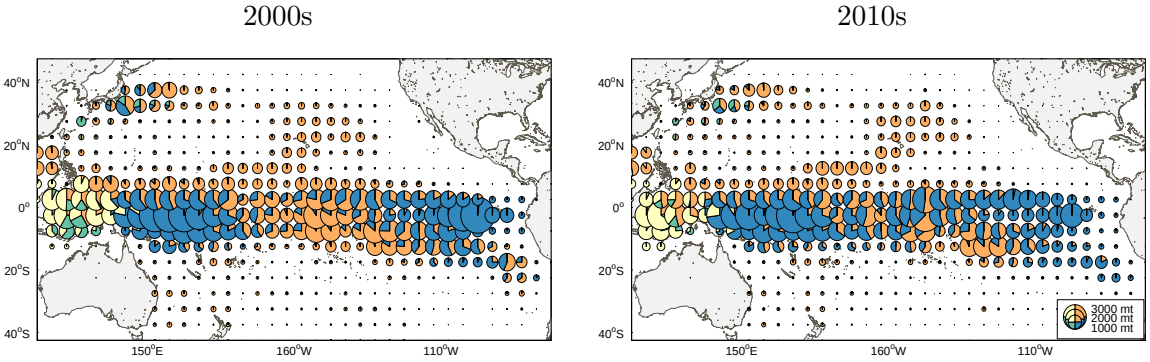


Figure 3: Spatial distributions of catches by decade and by gear: longline (orange), purse-seine (blue), pole-and-line (green), and others (yellow).

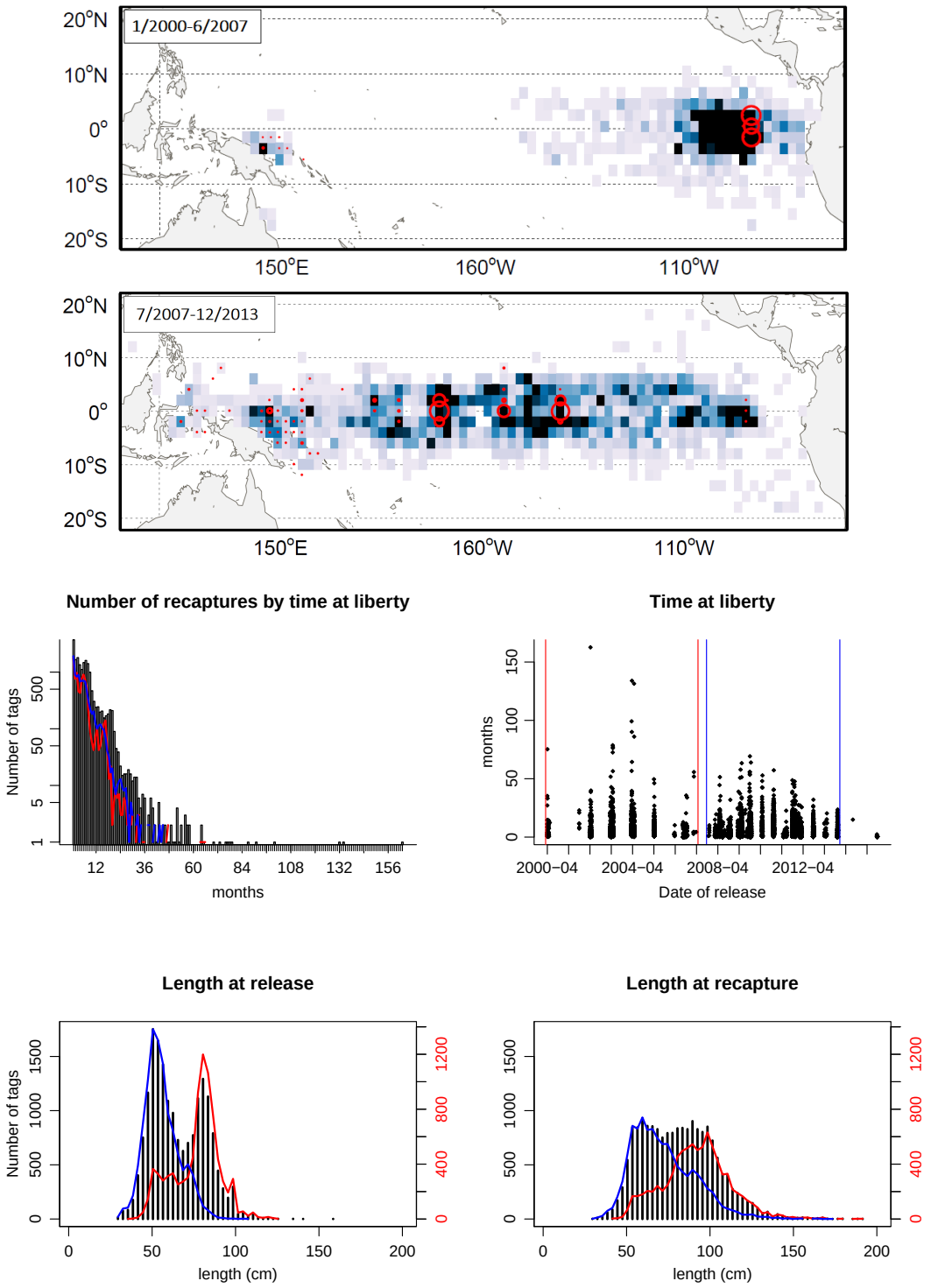


Figure 4: Top panel maps: number of bigeye tuna tagged and recaptured during conventional tagging campaigns separated into two periods (using linear color bar from white to blue indicating 0 to 40 and more tag returns respectively). The recapture data from the later, 2007-2013, period were used to inform SEAPODYM model parameters, while the recaptures from the earlier, 2000-2007, period were used in the model validation. Center panel: time at liberty histogram and the time at liberty of the tags depending on their date of release. Bottom panel: bigeye size distributions at release and recapture (black bars). The color-coded distributions and vertical lines indicate the data used in optimization (blue) and validation (red).

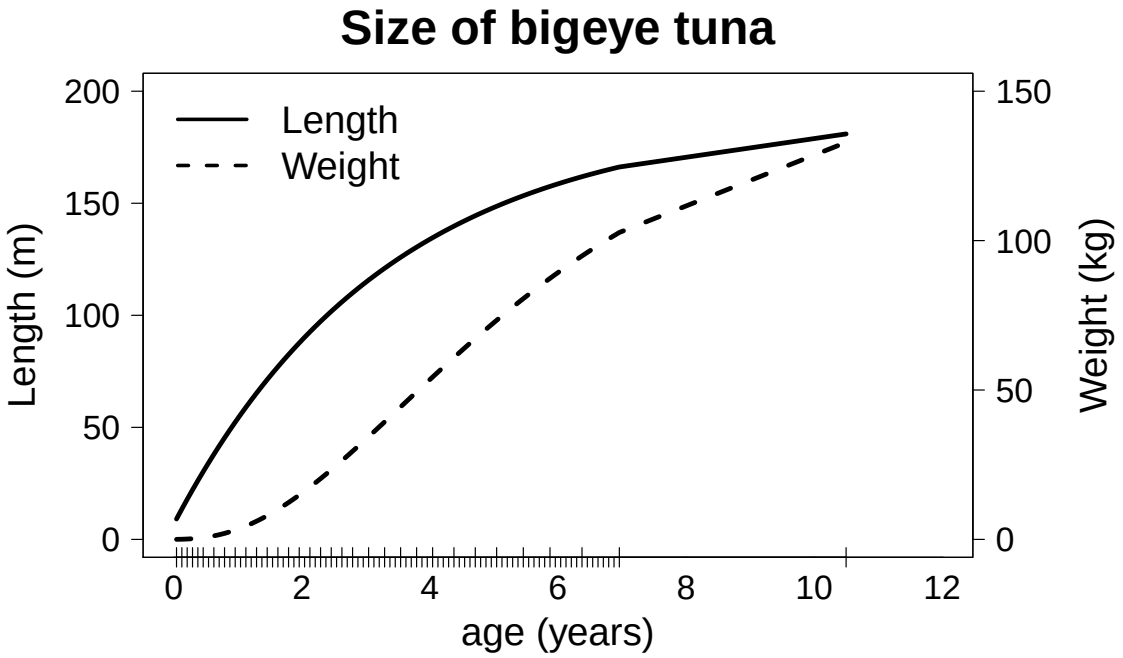


Figure 5: Static model parameters, mean length and weight interpolated from the Multifan-CL estimated functions (Harley et al., 2014) at the mid-point of each age class indicated with the outer tick marks of the x-axis. The inner ticks of the x-axis show the ages at which the habitat indices were evaluated in the current reference model.

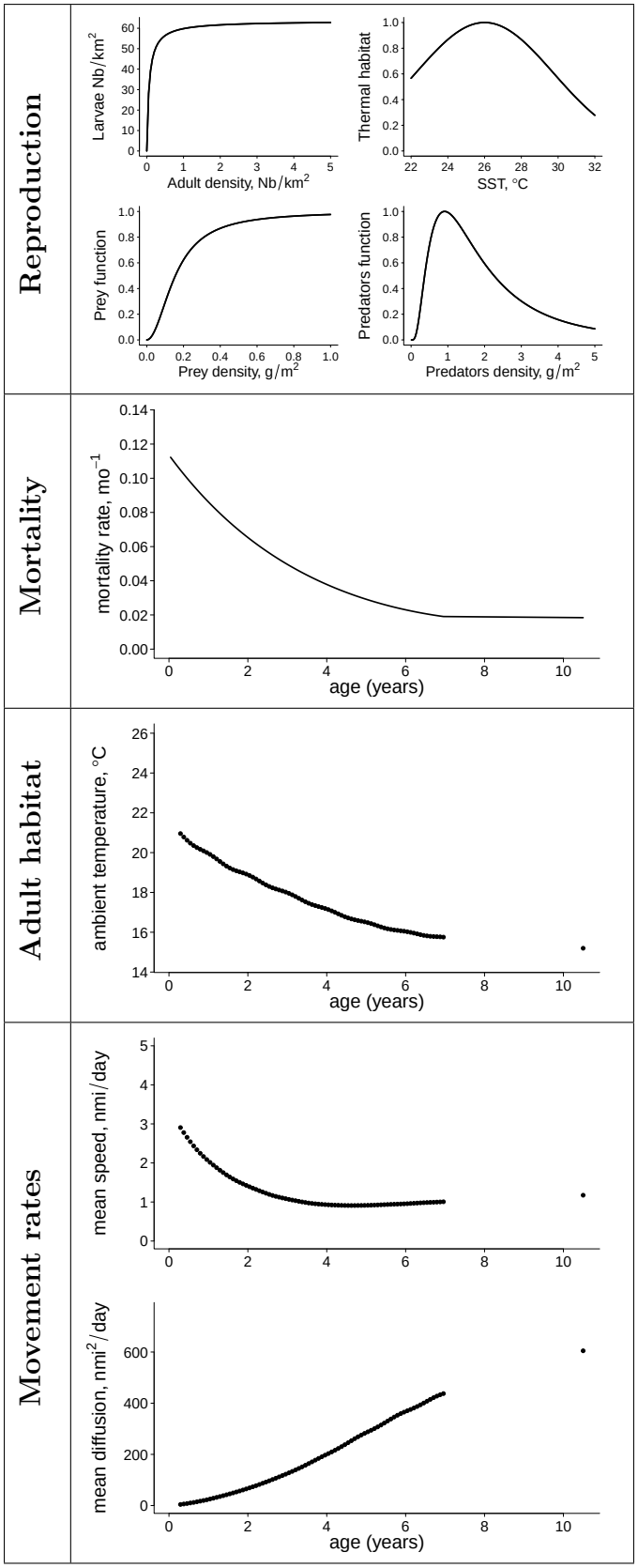


Figure 6: Estimated functional relationships in main dynamical processes (reproduction, natural mortality and movement) of reference MLE model constrained by fisheries data only. Habitat temperature and movement rates are computed as weighted spatio-temporal average with weights being the population density at age.

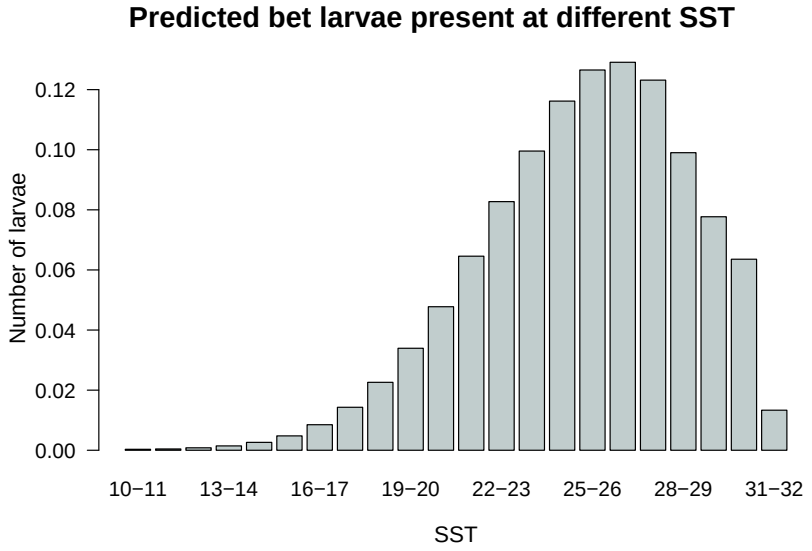


Figure 7: Predicted mean number of bigeye larvae in association with sea surface temperature.

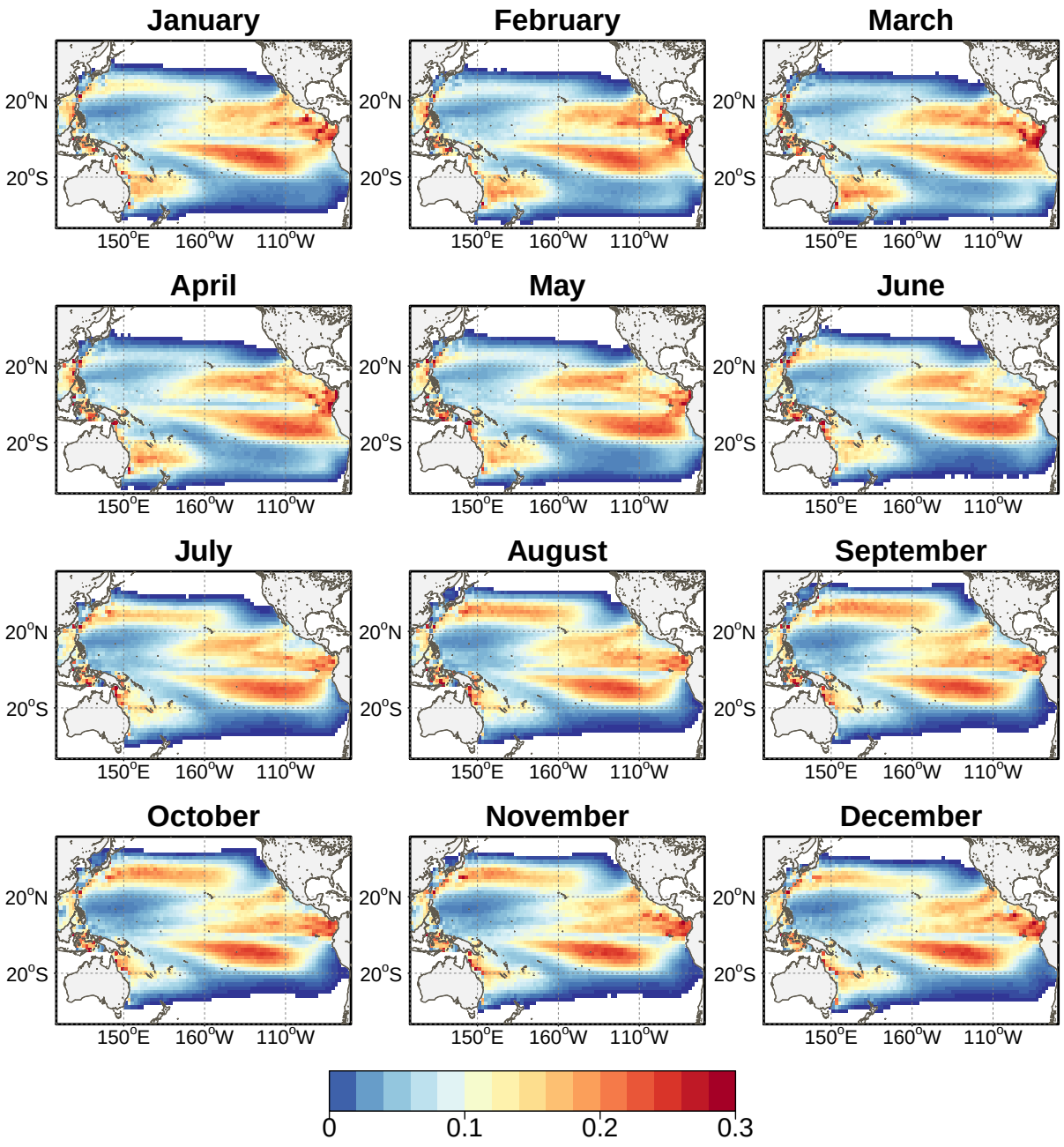


Figure 8: Mean monthly distributions of density of bigeye larval recruits (2001-2010 average).

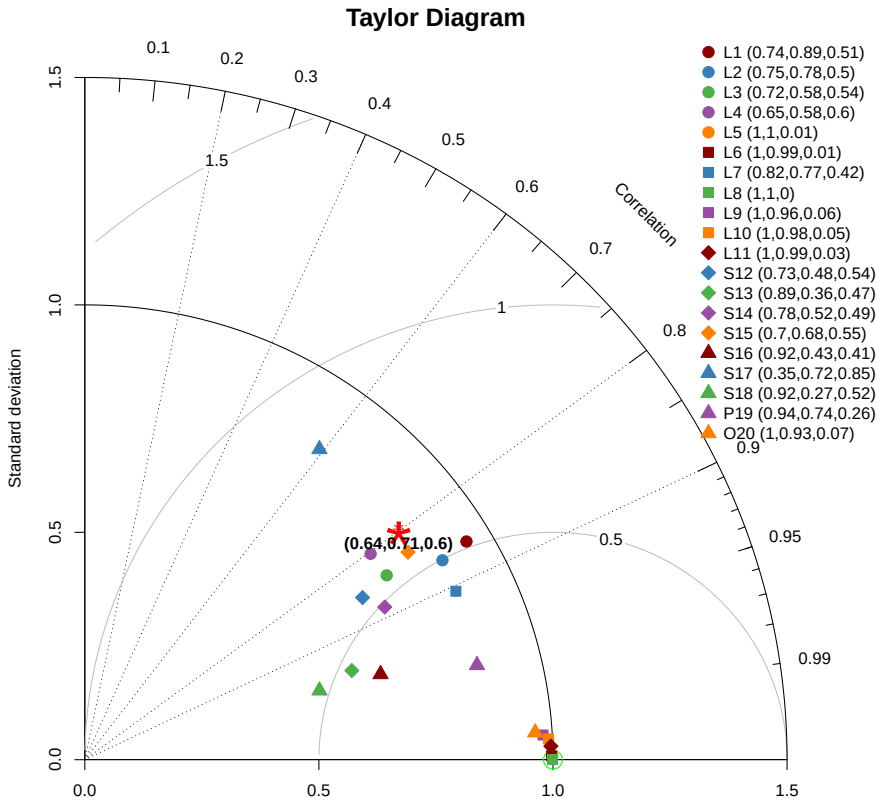


Figure 9: Taylor diagram, providing three aggregated metrics of model fit to the data: correlation (angular coordinates) between predictions and observations, standard deviation ratio (distance from (0,0) point depicts the ratio between model and data standard deviation) and normalized mean squared error (concentric circles with the green bullet being the center). Each point on the graph shows three metrics of the fit to the catch data by each fisheries (Table 1). The metrics are evaluated using the fisheries data over 1998-2019.

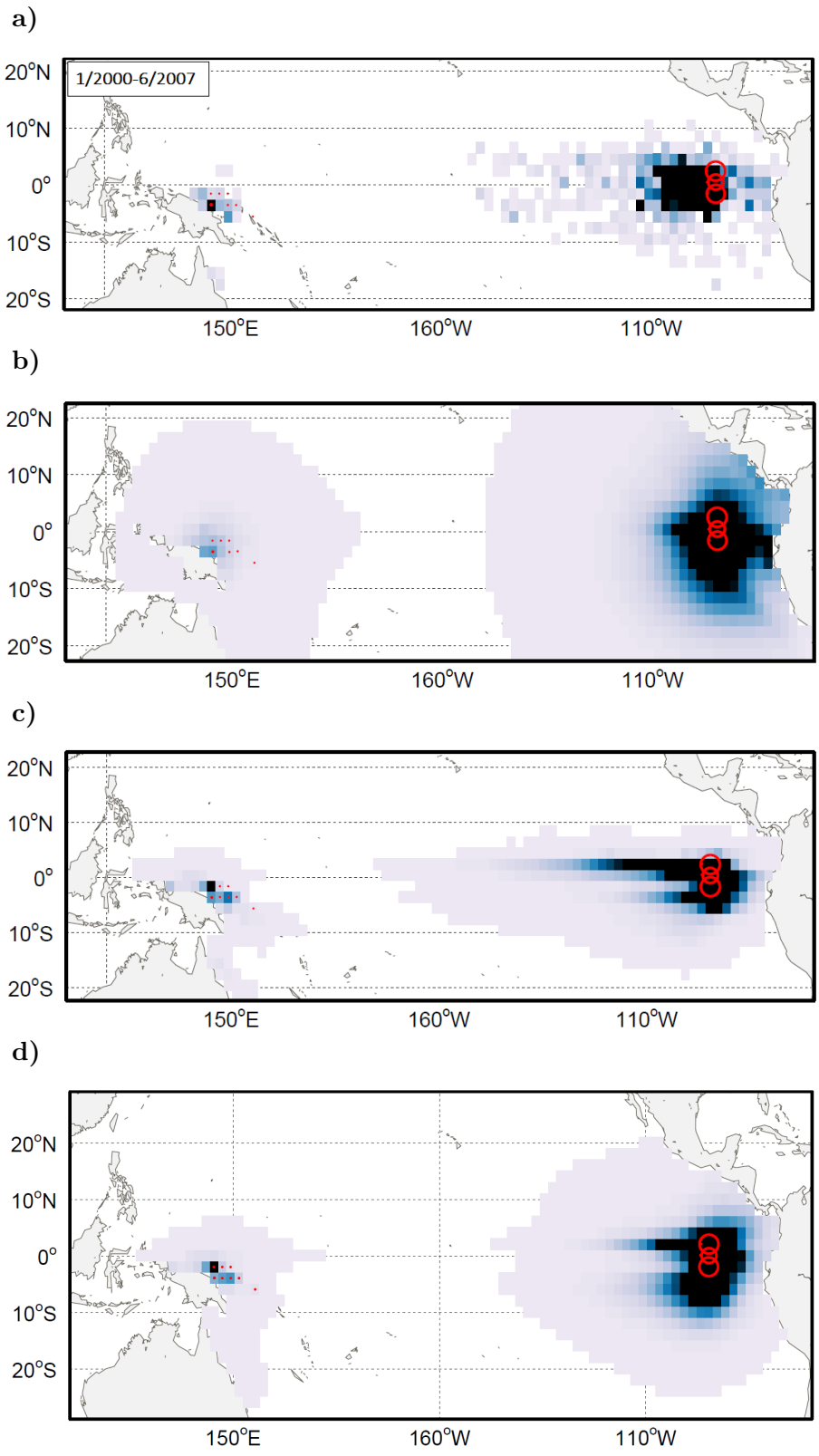


Figure 10: a) Number of bigeye tuna recaptured between January 2000 and June 2007 (linear color bar from white to blue indicating 0 to 40 and more tag returns respectively). b) Distribution of tag recaptures predicted with MLE parameters of current reference model, estimated with fisheries data only. c) Distribution of tag recaptures predicted with MLE solution obtained with the 2008-2013 sub-set of tagging data and GLORYS-free forcing. d) Distribution of tag recaptures predicted with MLE parameters estimated including the 2008-2014 sub-set of tagging data by the CLT model.

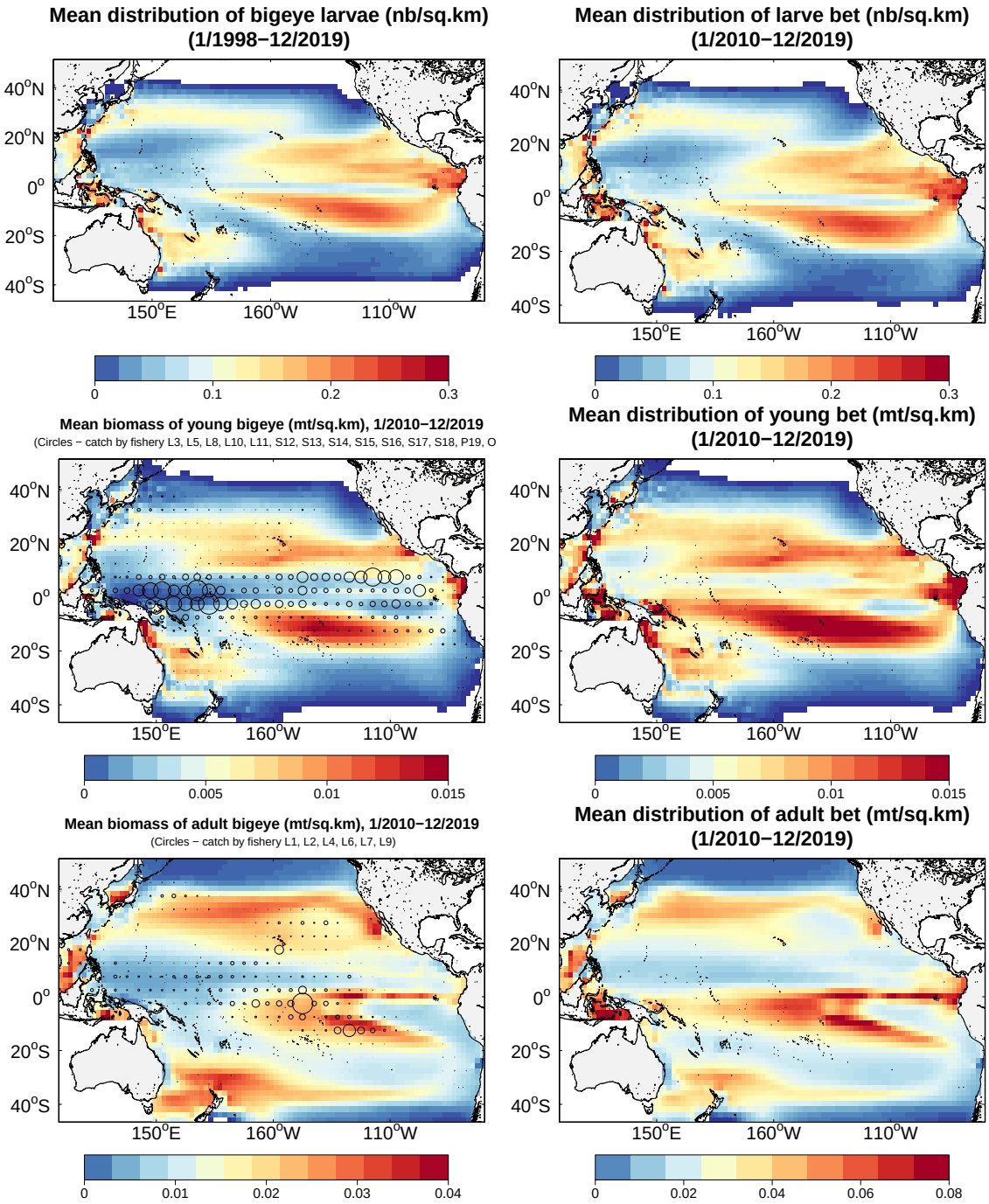


Figure 11: From top to bottom: average density of larval (Nb/km^2), young (mt/km^2) and adult (mt/km^2) bigeye tuna predicted with (left) and without fishing (right). Note different range of values of adult density shown for exploited and virgin stock.

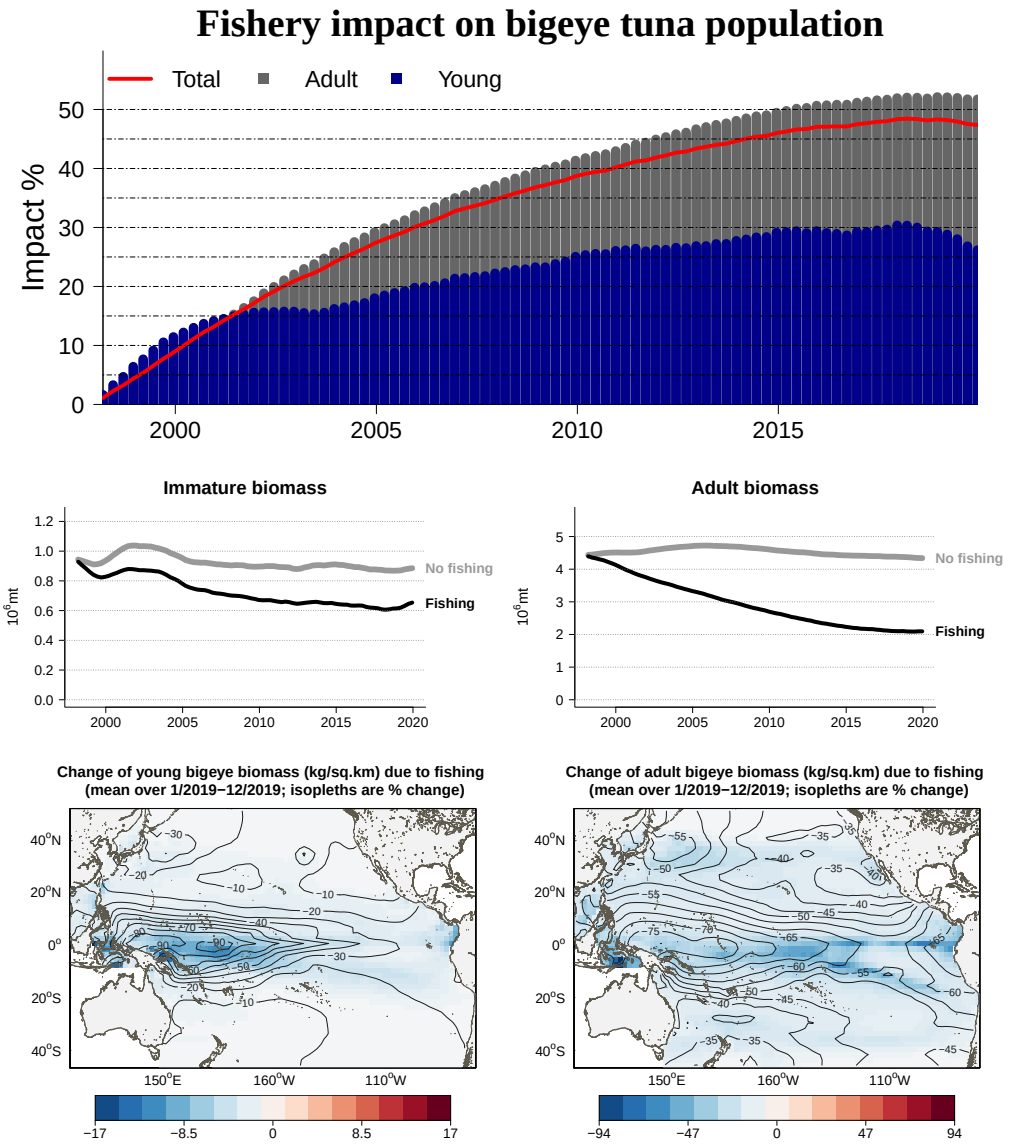


Figure 12: Spatial fishing impact on young and adult population stages of bigeye. Contour lines show the index $\frac{B_{F0}-B_{ref}}{B_{F0}}$ and colour shows the average biomass reduction due to fishing.

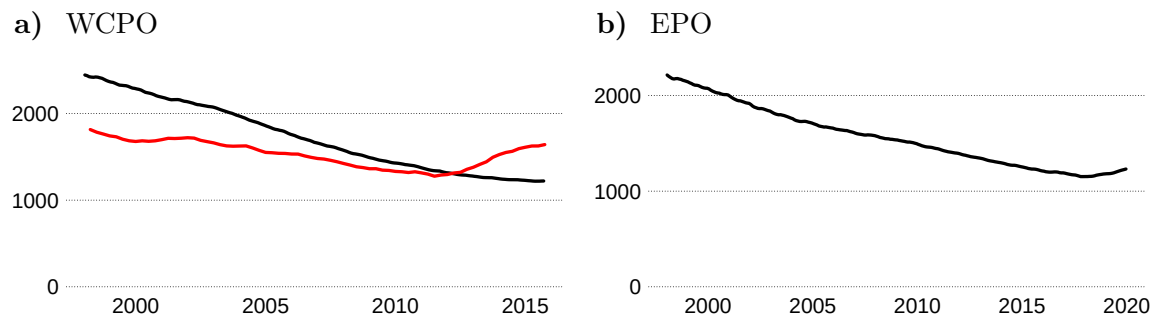


Figure 13: Biomass of bigeye (in thousand metric tons) including immature and mature individuals predicted by SEAPODYM (black) and estimated in WCPO by Multifan-CL (red).

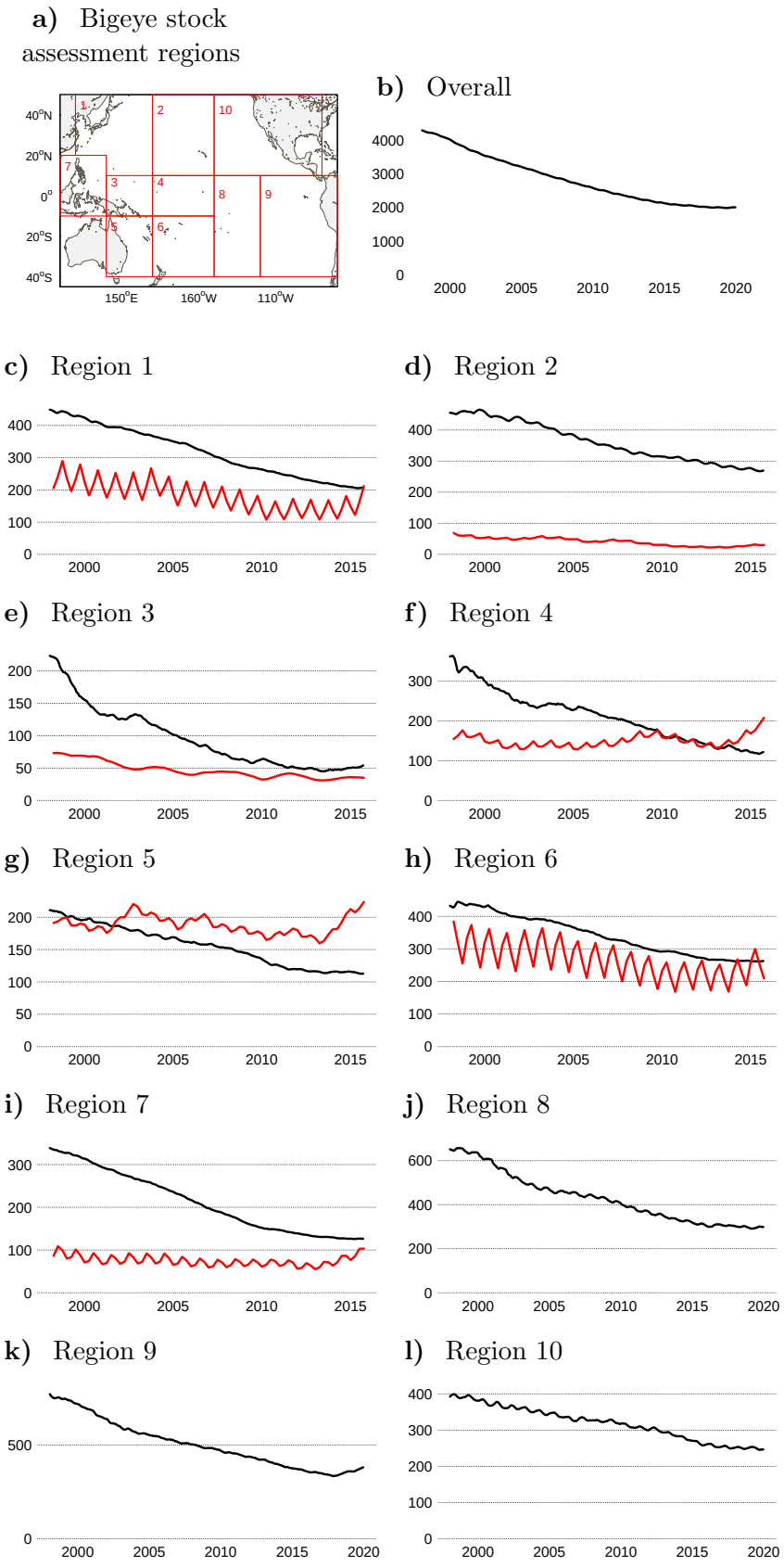


Figure 14: Comparison between SEAPODYM (black) and Multifan-CL (red) stock assessment model predictions for the Western and Central Pacific stock of mature adult bigeye (in thousand metric tons).

A Appendices

A.1 Model forcing

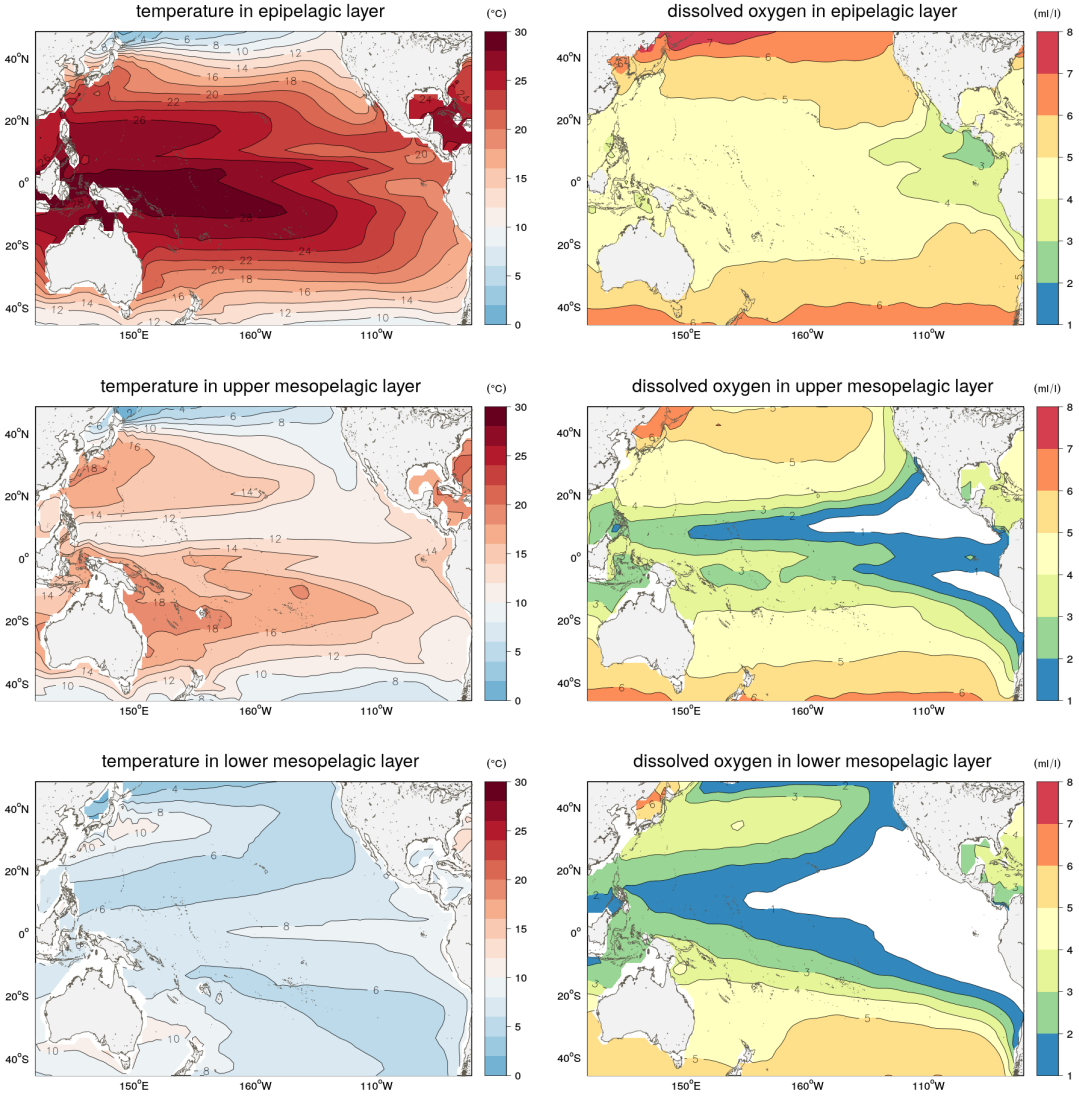


Figure A1: left) Water temperature from GLORYS-free forcing of SEAPODYM, and right) dissolved oxygen climatology (source: World Ocean Atlac), integrated over three pelagic layers, epipelagic, upper and lower mesopelagic layers. These two variables control species accessibility to prey organisms inhabiting respective pelagic layers. Maps show the mean variable over the decade 2001-2010.

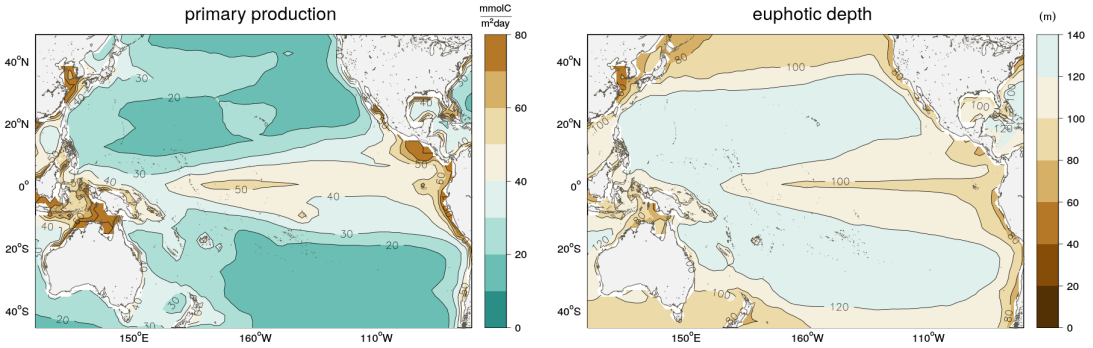


Figure A2: Primary production and euphotic depth from GLORYS-free forcing of SEAPODYM. Maps show the mean variable over decade 2001-2010.

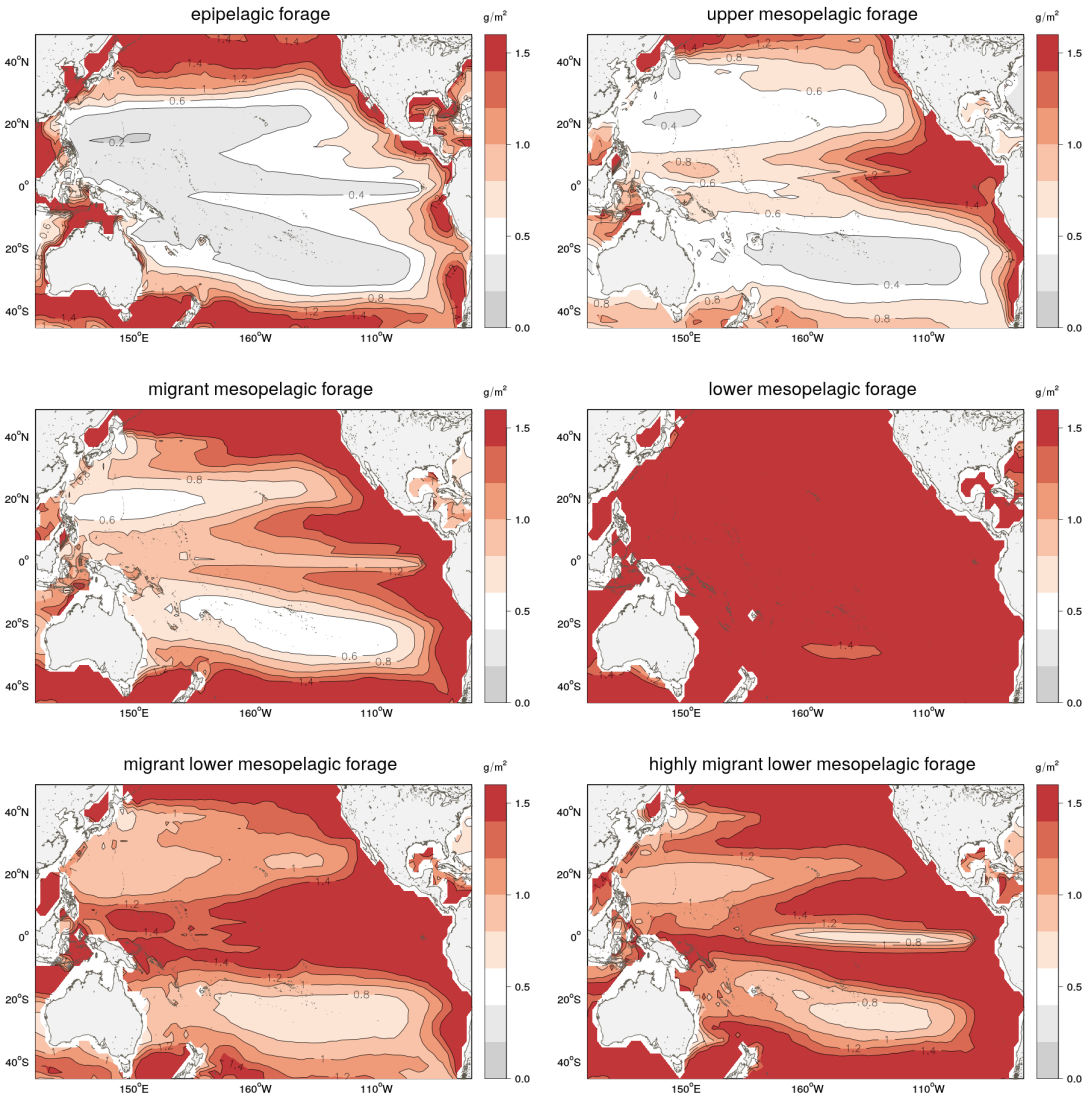
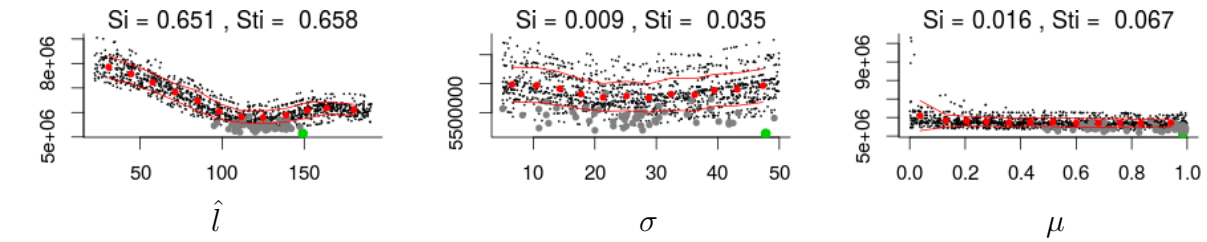


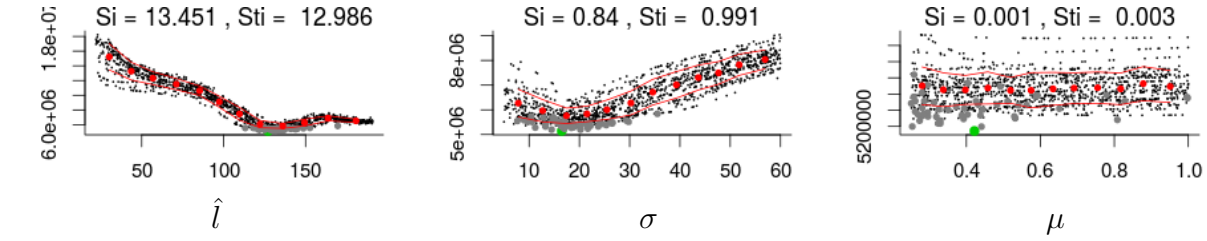
Figure A3: Six functional groups of micronekton, either resident in indicated pelagic layer or migrating to above (migrant) or to sub-surface (highly migrant) pelagic layers at night, simulated by SEAPODYM-LMTL model with GLORYS-free forcing.

A.2 Sensitivity to selectivity parameters

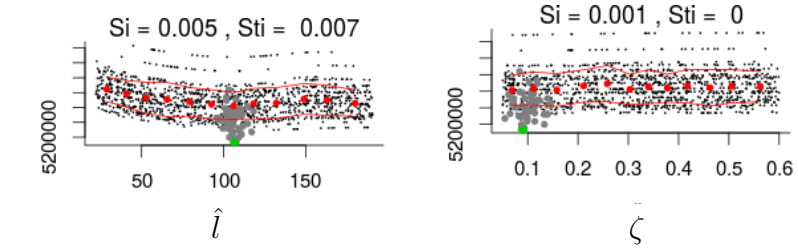
L_1, selectivity function of type III



L_2, selectivity function of type III



L_3, selectivity function of type II



L_4, selectivity function of type III

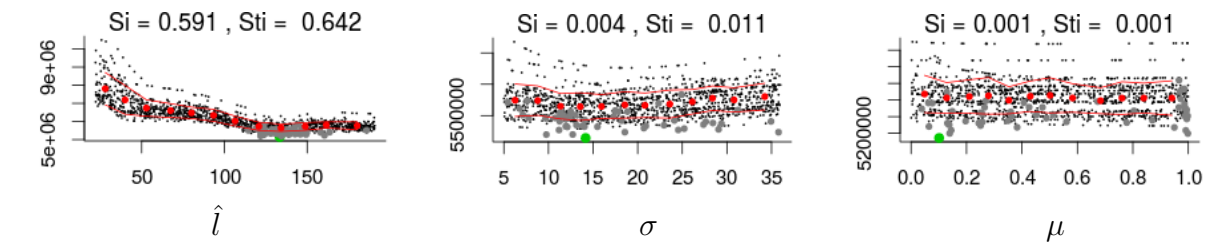
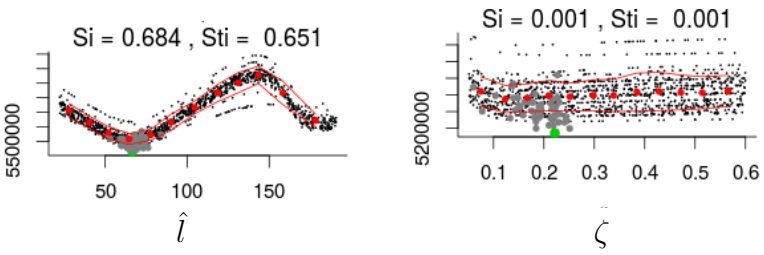
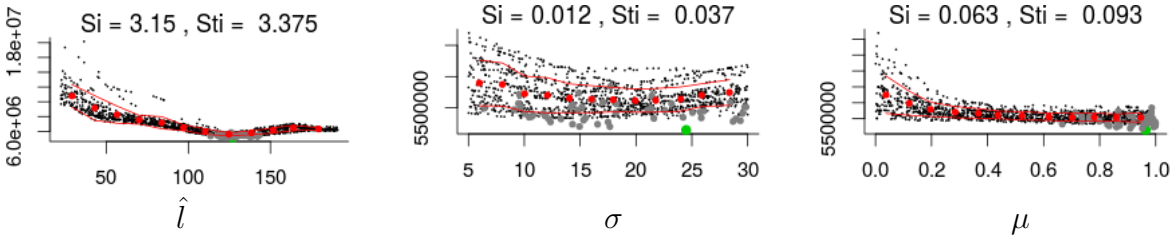


Figure A4: OAT profiles of selectivity function parameters of SEAPODYM fisheries. Green dots correspond to the lowest value of the likelihood function found in the OAT SA simulations.

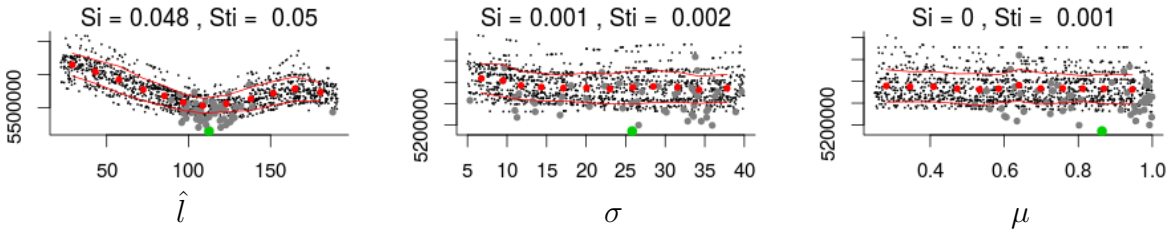
L_5, selectivity function of type II



L_6, selectivity function of type III



L_7, selectivity function of type III



L_8, selectivity function of type III

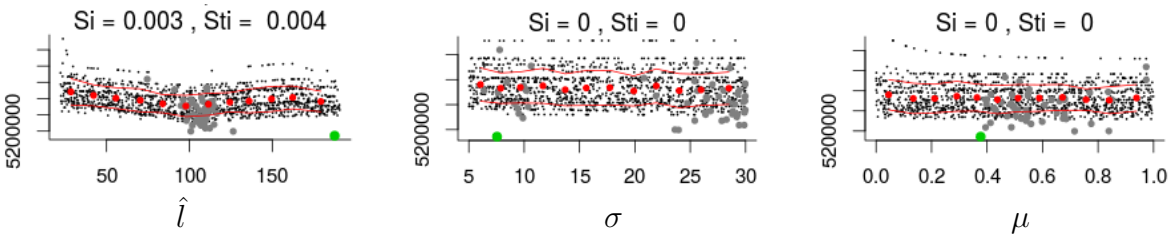
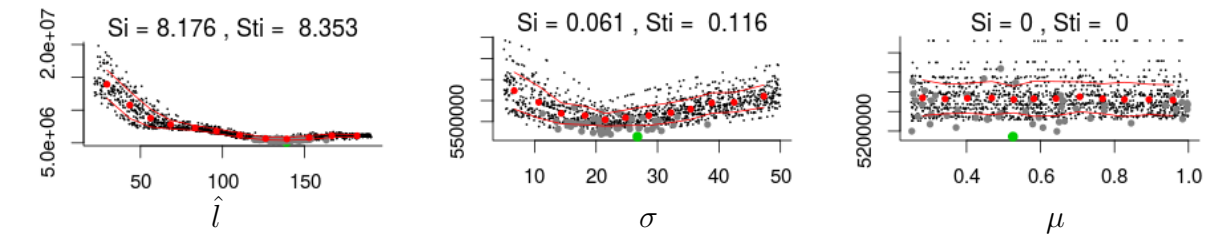
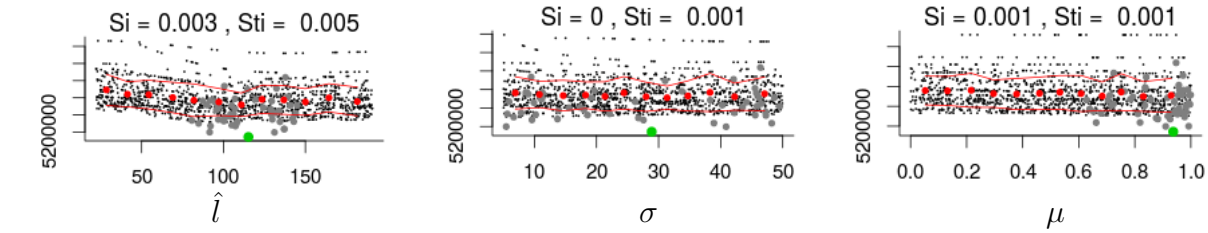


Figure A5: OAT profiles of selectivity function parameters of SEAPODYM fisheries. Green dots correspond to the lowest value of the likelihood function found in the OAT SA simulations.

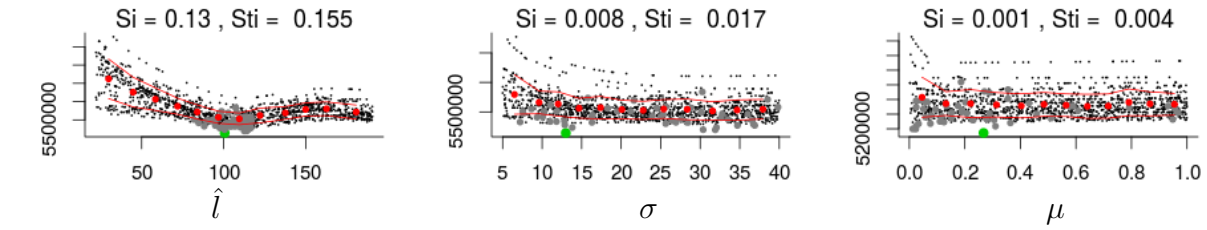
L_9, selectivity function of type III



L_{10}, selectivity function of type III



L_{11}, selectivity function of type III



S_{12}, selectivity function of type III

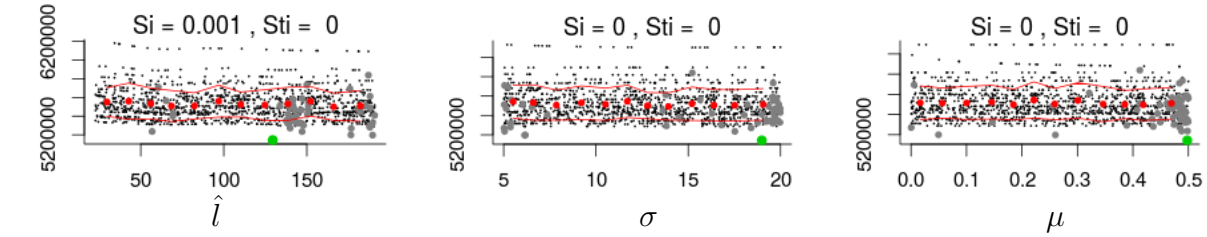
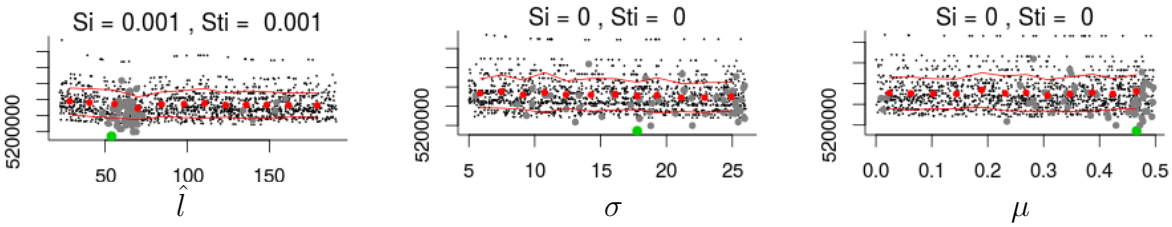
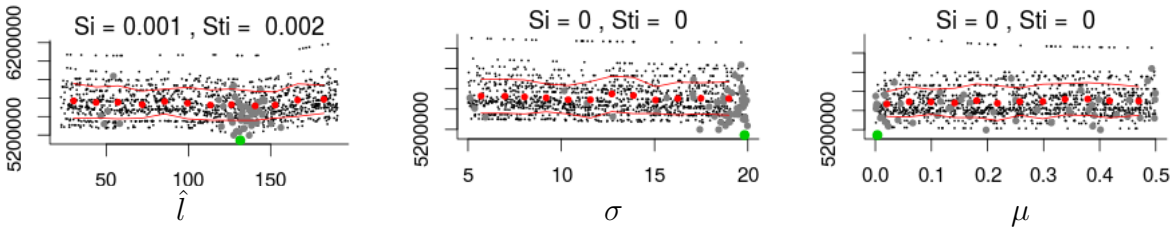


Figure A6: OAT profiles of selectivity function parameters of SEAPODYM fisheries. Green dots correspond to the lowest value of the likelihood function found in the OAT SA simulations.

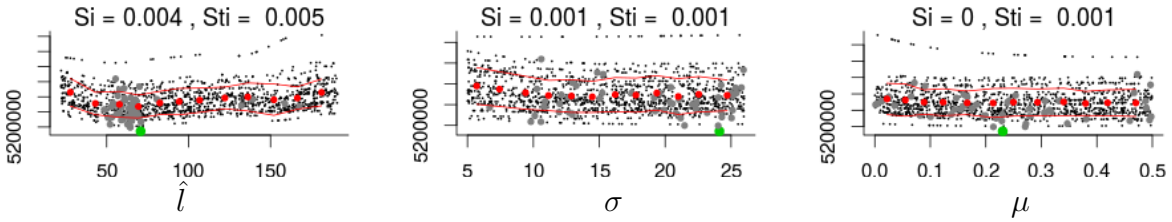
S_{13}, selectivity function of type III



S_{14}, selectivity function of type III



S_{15}, selectivity function of type III



S_{16}, selectivity function of type III

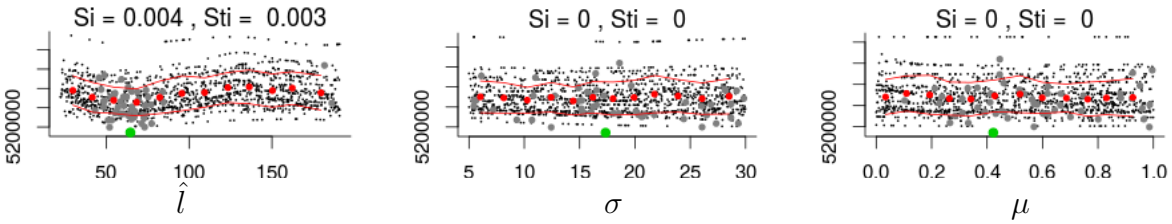
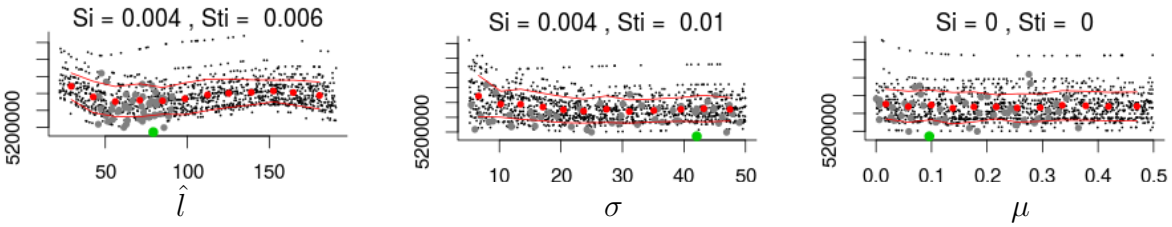
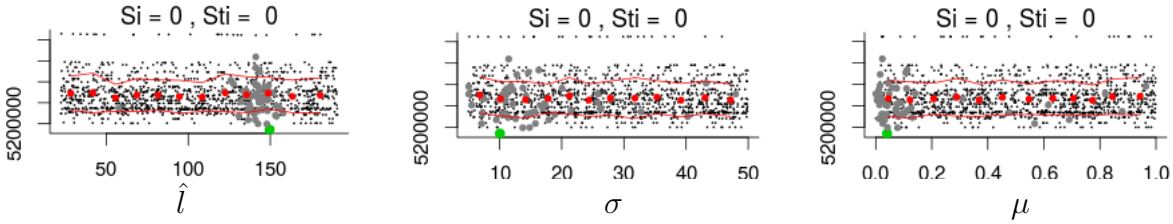


Figure A7: OAT profiles of selectivity function parameters of SEAPODYM fisheries. Green dots correspond to the lowest value of the likelihood function found in the OAT SA simulations.

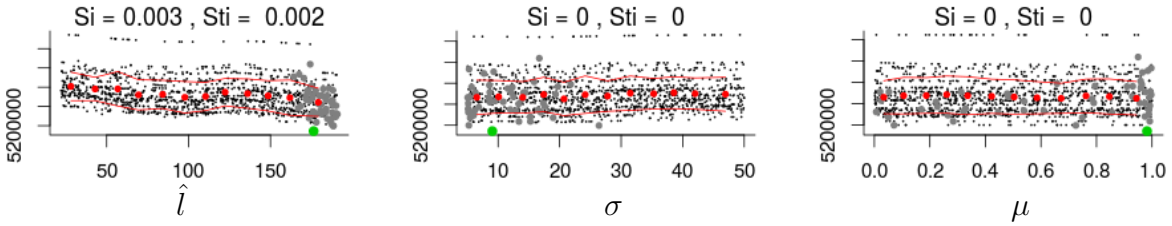
S_{17}, selectivity function of type III



S_{18}, selectivity function of type III



P_{19}, selectivity function of type III



O_{20}, selectivity function of type III

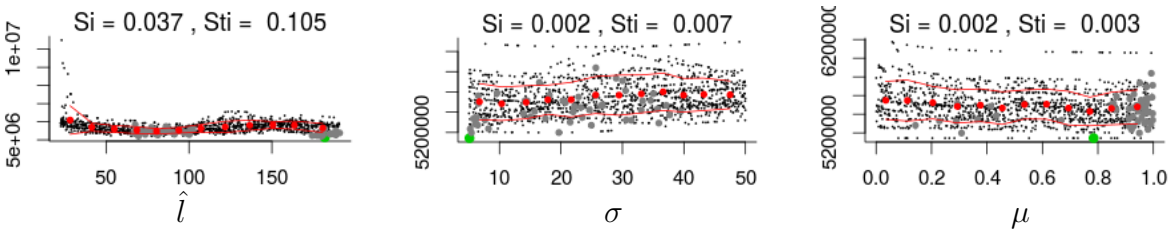


Figure A8: OAT profiles of selectivity function parameters of SEAPODYM fisheries. Green dots correspond to the lowest value of the likelihood function found in the OAT SA simulations.

A.3 Estimated population age structure

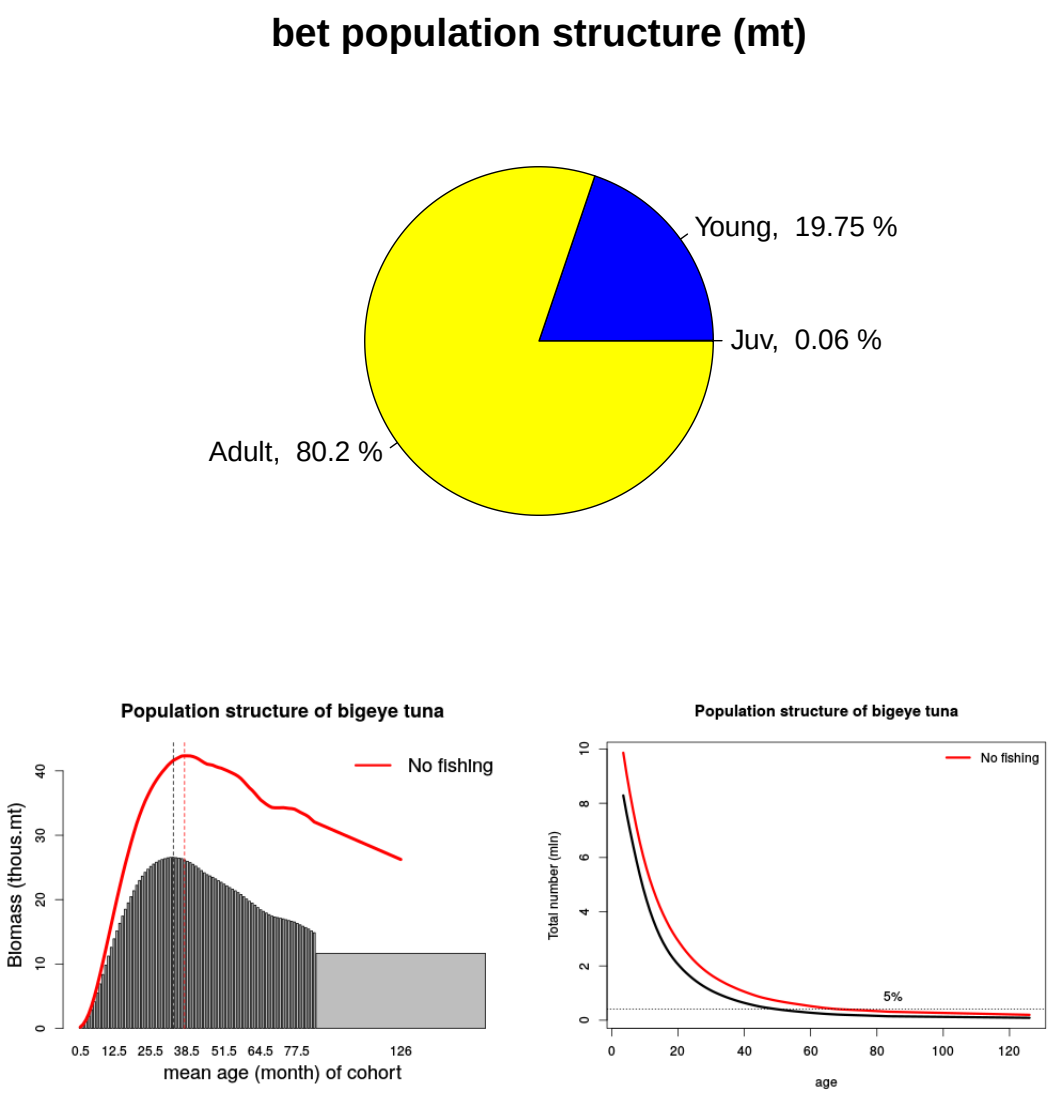


Figure A9: Population structure estimated in the CLT model. The life stage proportions of biomass are derived from the entire 1998-2019 simulation, the age structure of the population with and without fishing in biomass units (lower left panel) and in number of individuals (lower right panel).

A.4 Fit to the fisheries data

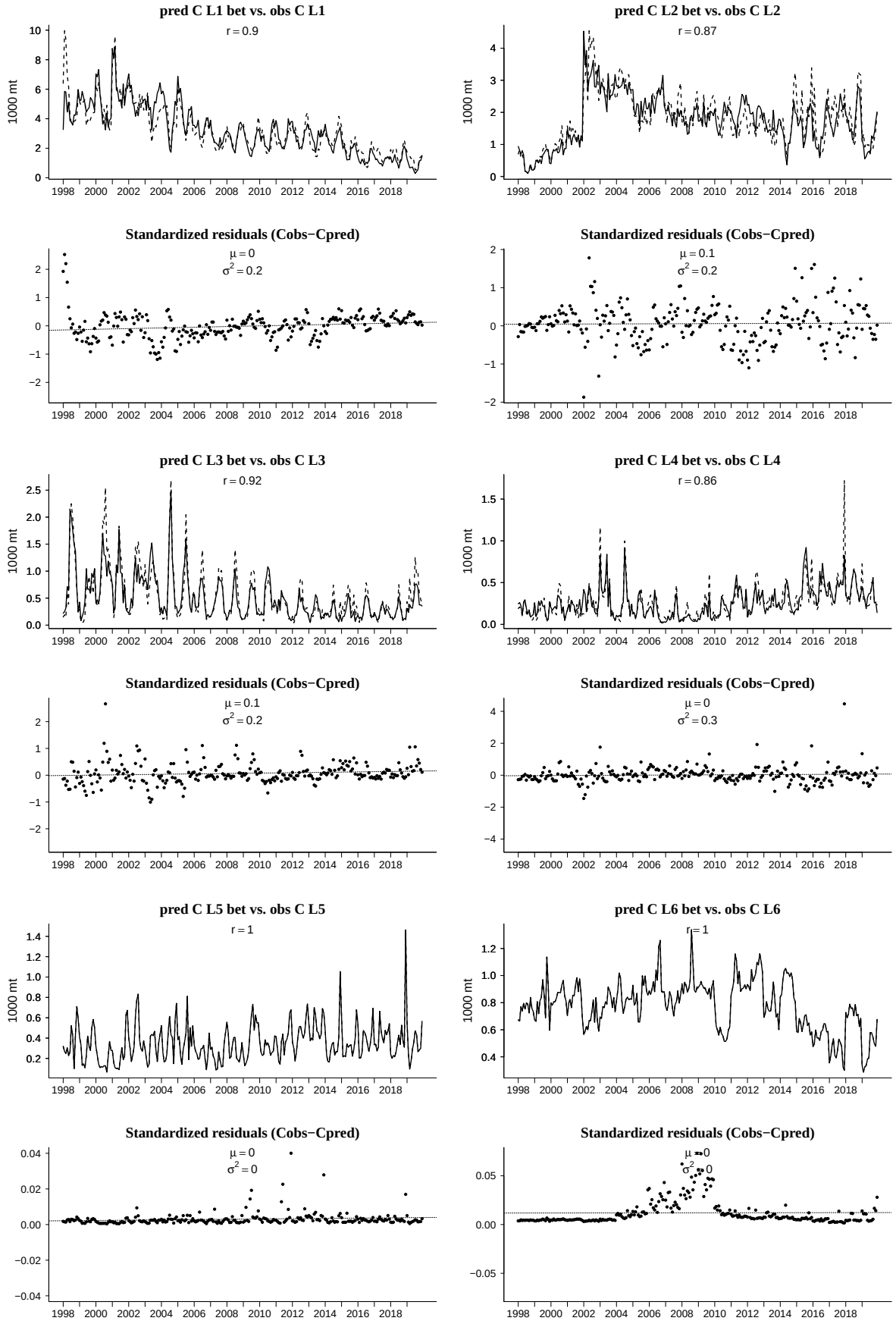


Figure A10: Monthly time series of observed (dashed) vs. predicted (solid) catch by fishery and standardized residuals. Three statistical scores shown on the plots are Pearson correlation coefficient (r) between predicted and observed catches, mean (μ) and variance (σ^2) of standardized residuals).

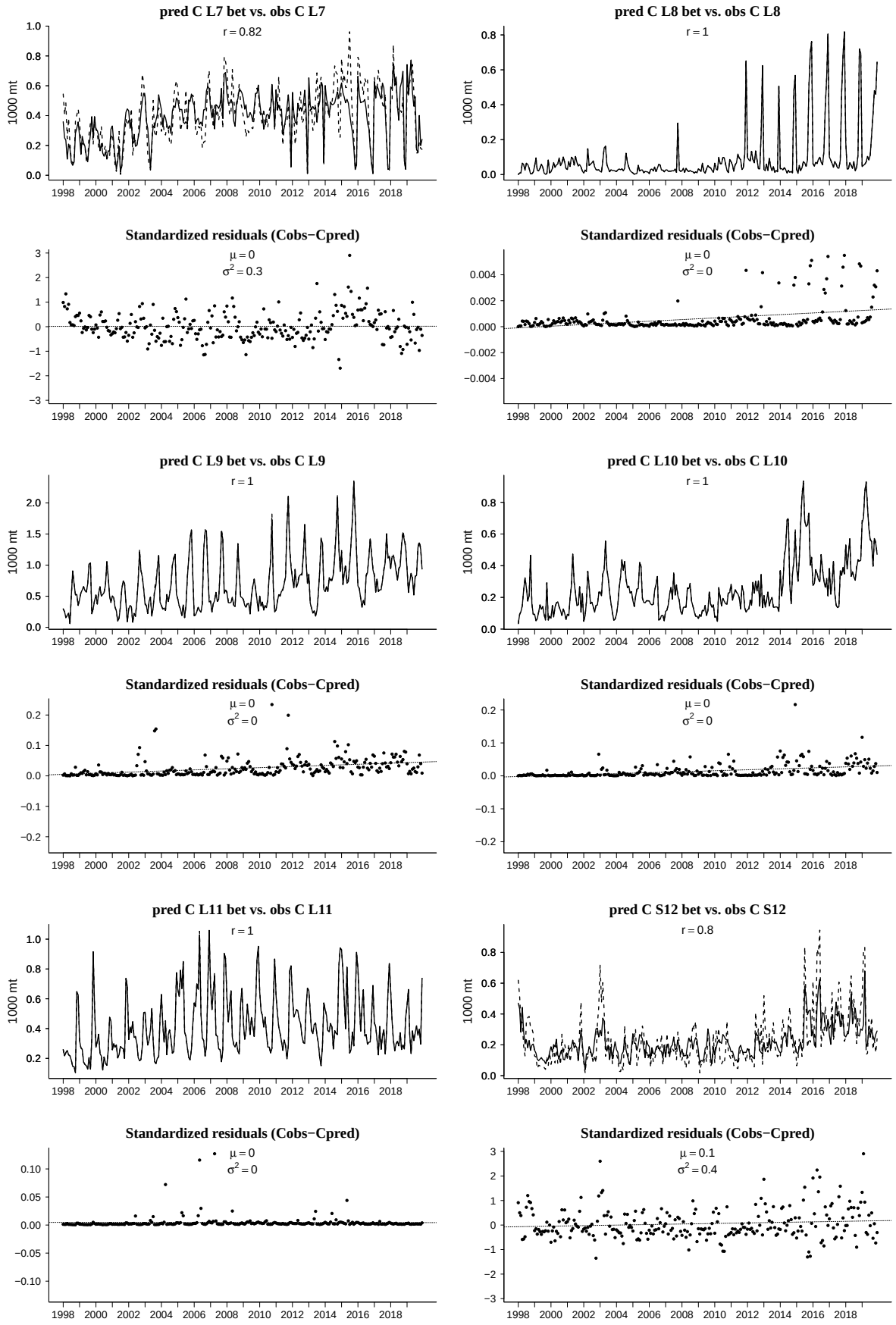


Figure A10: Monthly time series of observed and predicted catch by fishery (Continued)

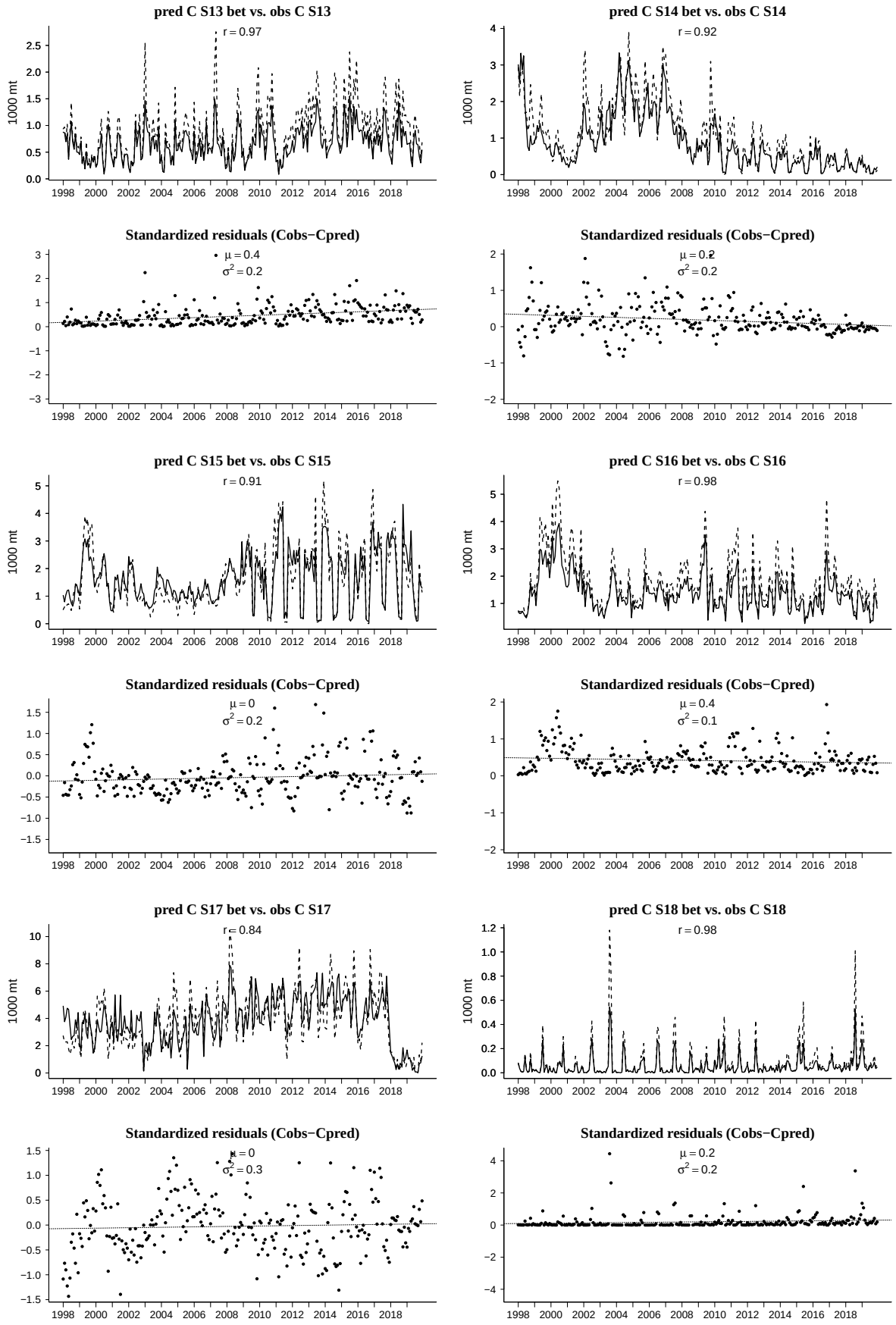


Figure A10: Monthly time series of observed and predicted catch by fishery (Continued)

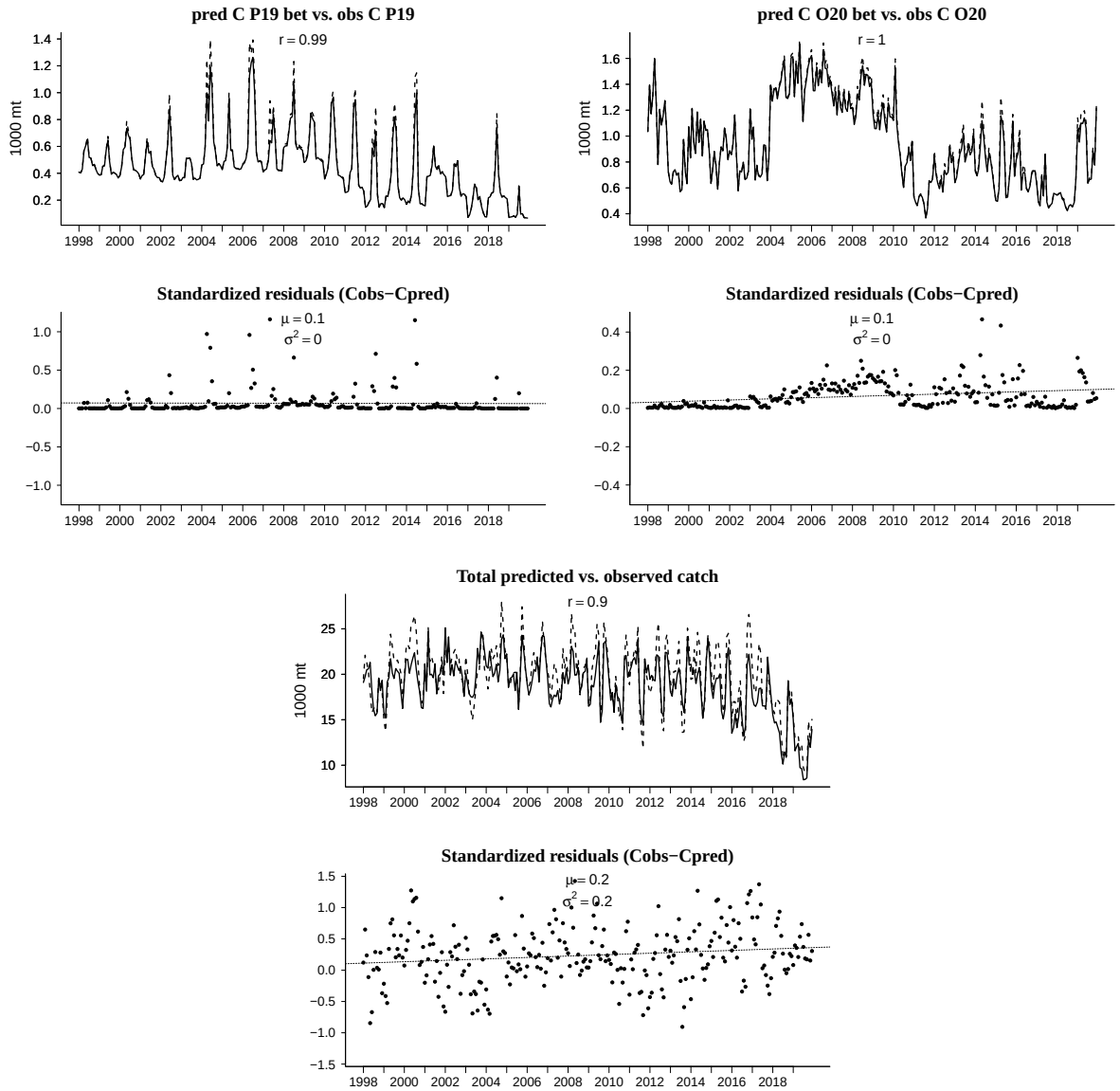


Figure A10: Monthly time series of observed and predicted catch by fishery (Continued)

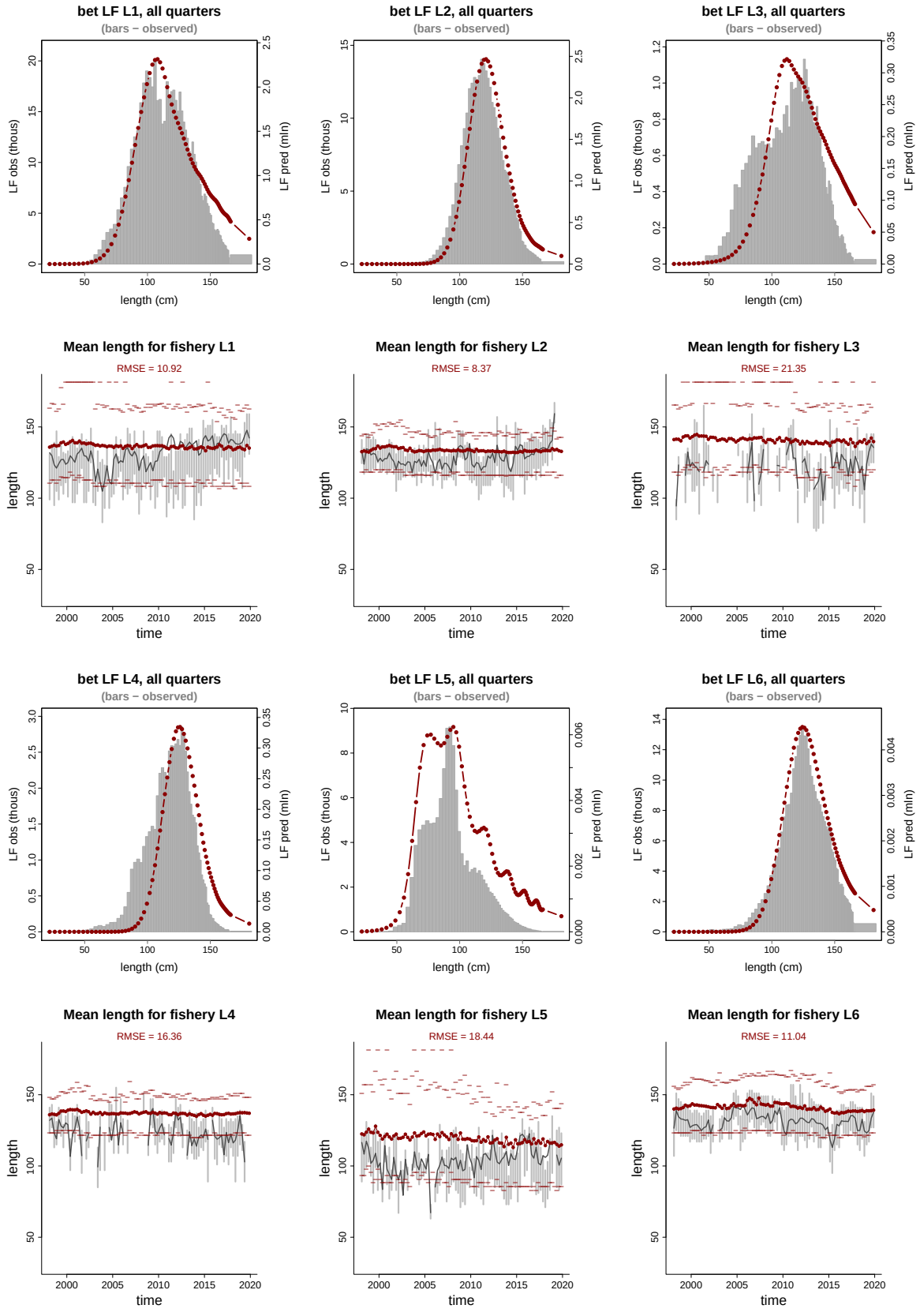


Figure A11: Observed (grey) and predicted (red) length frequencies distribution and mean length in catches.

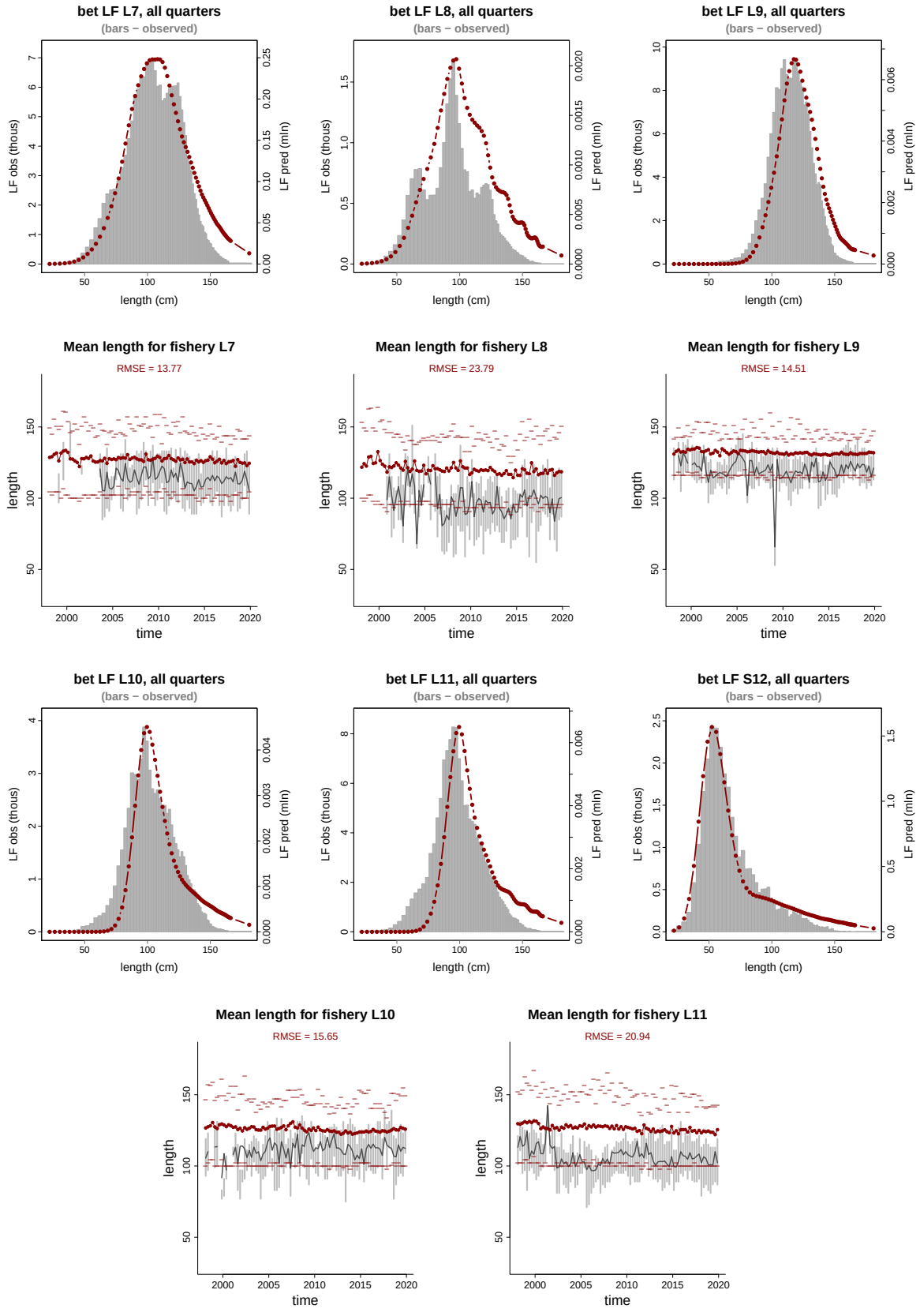


Figure A11: Fit for the length frequencies data. Continued.

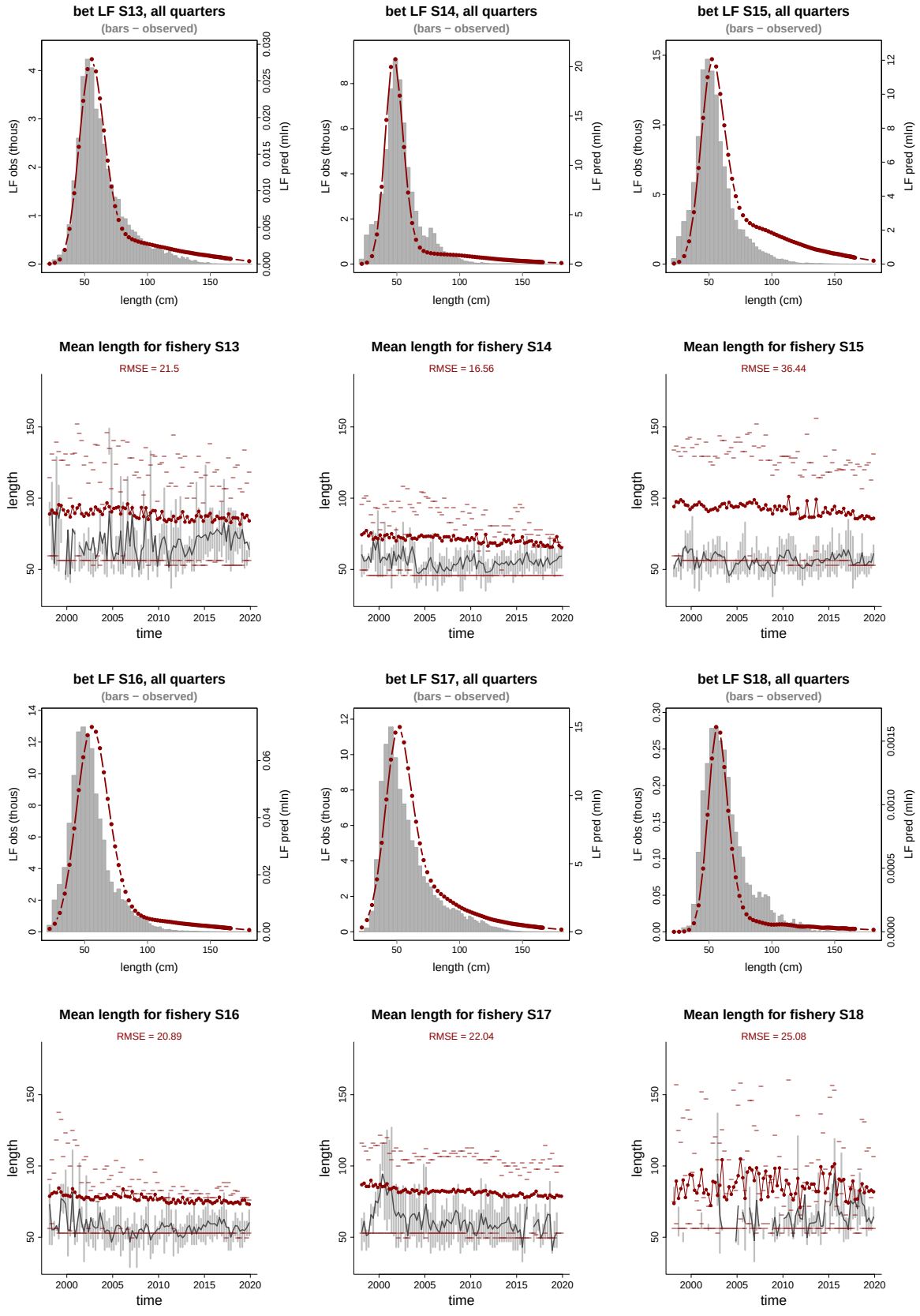


Figure A11: Fit for the length frequencies data. Continued.

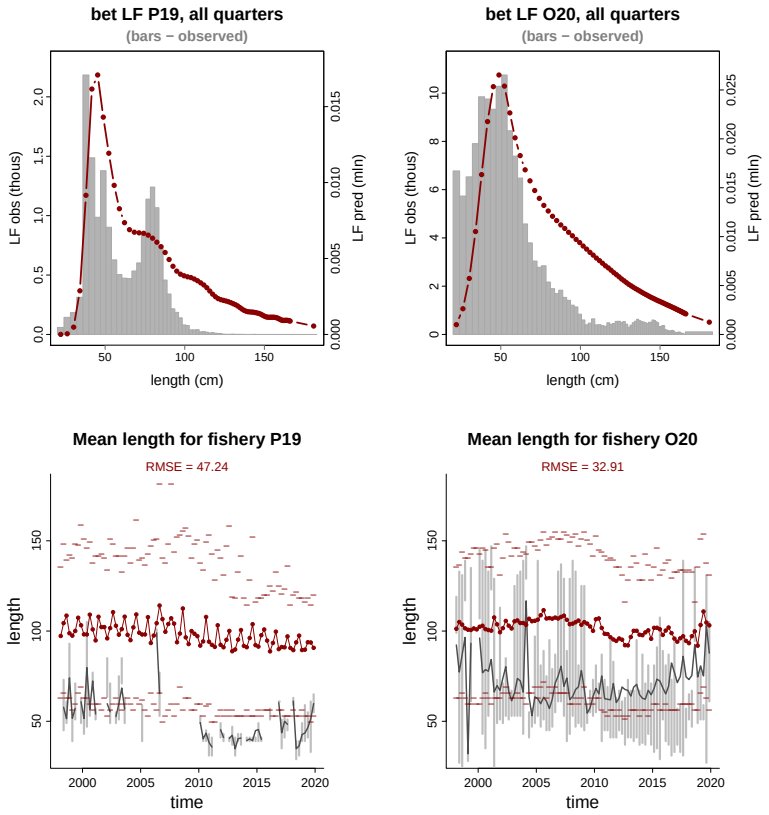


Figure A11: Fit for the length frequencies data. Continued.

A.5 Fit to the tagging data

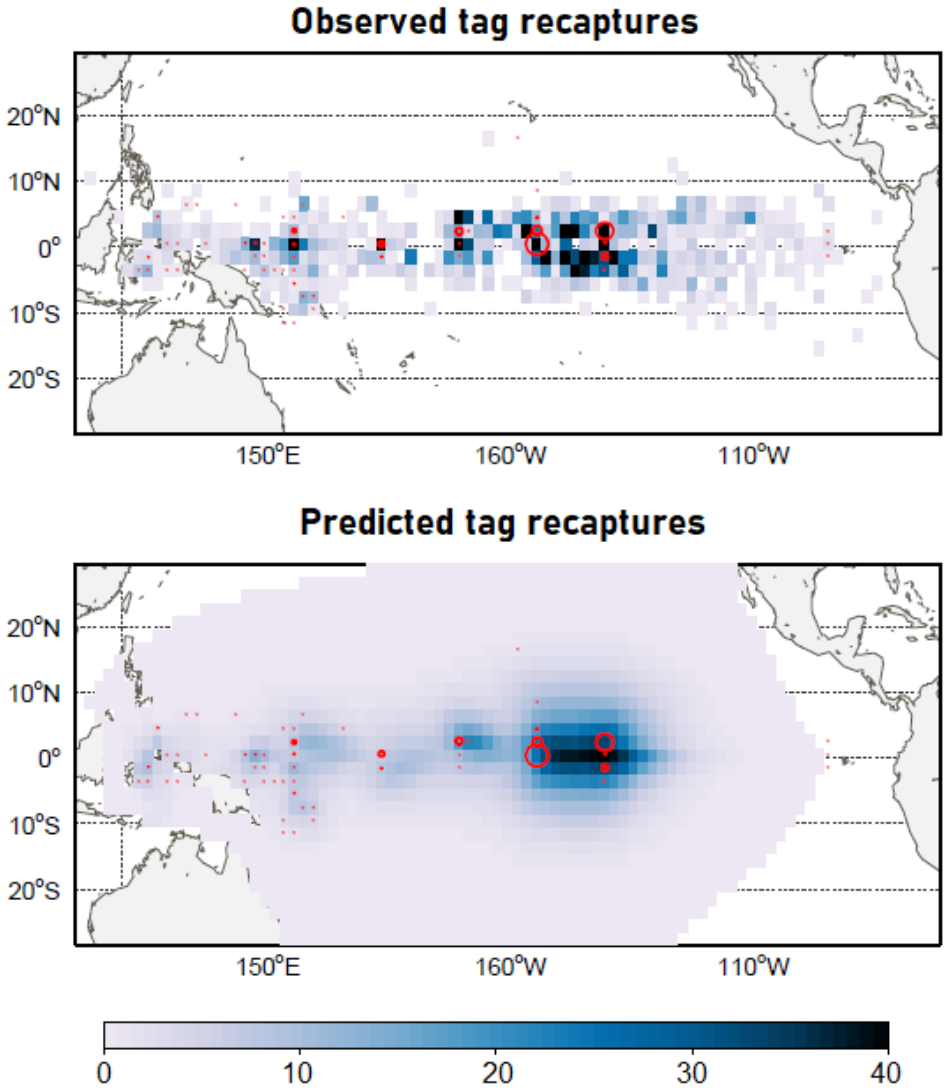


Figure A12: Number of bigeye tuna recaptured between July 2007 to December 2010 (top). Distribution of tag recaptures predicted for the same time period by the CL reference model, i.e. with MLE parameters, estimated from fisheries data only (bottom).

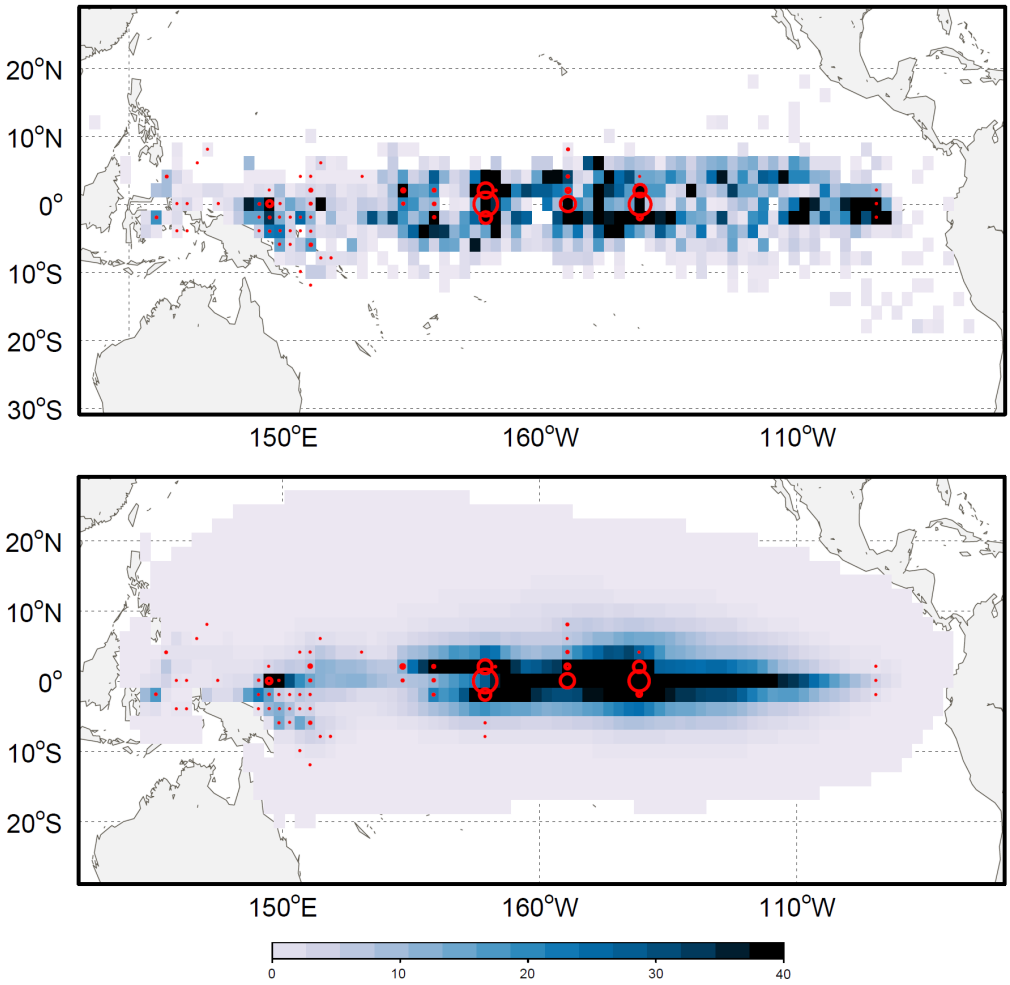


Figure A13: Number of bigeye tuna recaptured between July 2007 to December 2014 integrated into current reference model (top). Distribution of tag recaptures predicted for the same time period by the CLT model with MLE parameters, estimated from fisheries and tagging data (bottom).

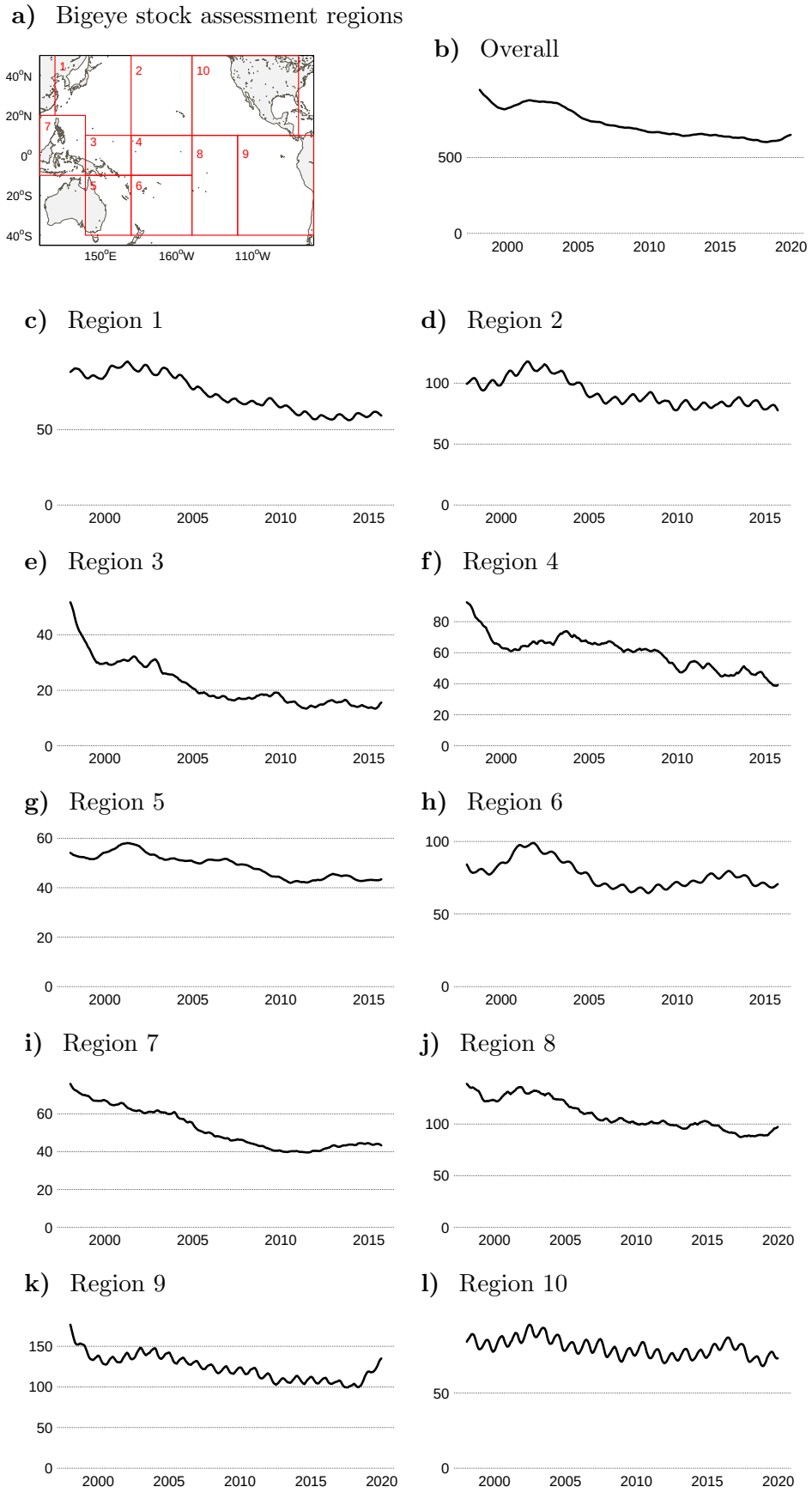


Figure A14: SEAPODYM model predictions for immature bigeye (in thousand metric tons) over the stock assessment regions.