Canadian Science Publishing

Canadian Journal of Fisheries and Aquatic Sciences

Comparing performance of catch-based and length-based stock assessment methods in data-limited fisheries

Journal:	*Canadian Journal of Fisheries and Aquatic Sciences*
Manuscript ID	cjfas-2019-0276.R1
Manuscript Type:	Article
Date Submitted by the Author:	26-Nov-2019
Complete List of Authors:	Pons, Maite; University of Washington, Aquatic and Fishery Sciences Cope, Jason; National Marine Fisheries Service - NOAA, Northwest Fisheries Science Center Kell, Laurie; Comision Internacional para la Conservacion del Atun Atlantico
Keyword:	data-limited assessment methods, depletion, life-history, harvest rates
Is the invited manuscript for consideration in a Special Issue? :	Not applicable (regular submission)

Comparing performance of catch-based and length-based stock assessment methods in data-limited fisheries

Maite Pons [1*], Jason M. Cope [1,2], and Laurence T. Kell [3]

[1] School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, Washington 98195-5020, USA. Email: mpons@uw.edu

[2] National Oceanic and Atmospheric Administration Fisheries, Northwest Fisheries Science Center, Fishery Resource Analysis and Monitoring Division, 2725 Montlake Boulevard East, Seattle, Washington 98112-2097, USA. Email: jason.cope@noaa.gov

[3] Centre for Environmental Policy, Imperial College London, London SW7 1NE, UK. Email: laurie@kell.es

[*] Corresponding author: School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, Washington 98195-5020, USA. Email: mpons@uw.edu. Telephone number: +1 206 883 5102.

Abstract

The quantity of data from many small-scale fisheries is insufficient to allow for the application of conventional assessment methods. Even though in many countries they are moving to close-loop simulations to assess the performance of different management procedures in data limited situations, managers in most developing countries are still demanding information on stock status. In this study we use the common metric of harvest rate to evaluate and compare the performance of the following catch-only and length-only assessment models: Catch-Maximum Sustainable Yield (Catch-MSY), Depletion Based Stock Reduction Analysis (DBSRA), Simple Stock Synthesis (SSS), an extension of Catch-MSY (CMSY), Length Based Spawning Potential Ratio (LBSPR), Length-Based Integrated Mixed Effects (LIME), and Length-Based Bayesian (LBB). In general, results were more biased for slightly depleted than for highly depleted stocks, and for long-lived than for short-lived species. Length-based models, such as LIME, performed as well as catch-based methods in many scenarios and, among the catch-base models the one with the best performance was SSS.

Keywords: data-limited assessment methods, depletion, life-history, harvest rates

INTRODUCTION

Major commercial fish species usually have substantial sets of data that can be integrated by complex stock assessment models (e.g., Methot and Wetzel 2013); these data may include time series of total removals, catch-at-length or -age, relative or absolute abundance indices, fishing effort, and information on life-history parameters. Such datasets are typically unavailable for most of the small-scale fisheries and by-catch species around the world. Fisheries and stocks lacking comprehensive datasets are commonly known as "data-poor" or "data-limited" fisheries (Costello et al. 2012; Dowling et al. 2015). Recently, many data-limited approaches have been developed to meet an increasing demand for science-based fisheries management of unassessed fisheries where data and resources are limited (Wetzel and Punt 2011; Costello et al. 2012; Dowling et al. 2015, 2016; Chrysafi and Kuparinen 2016; Rosenberg et al. 2017).

Assessing stocks using only catch and life-history data started many years ago with the development of Stock Reduction Analysis (SRA; Kimura and Tagart 1982; Kimura et al. 1984). Since then, SRA has been extended to estimate productivity and reconstruct historical abundance trends by making assumptions about final biomass relative to unfished or initial biomass (i.e., stock depletion; Thorson and Cope 2015). SRA has been further extended to incorporate stochastic variability in population dynamics (Stochastic-SRA; Walters et al. 2006), a flexible shape for the production function (Depletion-Based SRA; Dick and MacCall 2011), prior information regarding resilience and population abundance at the start of the catch time series (Catch-Maximum Sustainable Yield, Catch-MSY; Martell and Froese 2013), bayesian approches (CMSY, Froese et al. 2017), and age-structured population dynamics (Simple Stock Synthesis, SSS; Cope 2013). Despite these differences, this family of catch-only models share a common dependence on prior assumptions about final stock depletion. Simulation testing has previously indicated that these methods perform well only when assumptions regarding final relative abundance are met (Wetzel and Punt 2015). Unsurprisingly, because final stock depletion is a prior assumption, the methods perform differently under different stock depletion levels (i.e., highly depleted or slightly depleted stocks, Walters et al. 2006) or under different harvest history or catch trends.

For many small-scale fisheries, obtaining reliable time series on historical total catch is difficult, whereas sampling lengths from the catch is easier. Mean-length mortality estimators (Beverton and Holt 1957) assume that fishing mortality directly influences the mean length of the catch under equilibrium conditions. This basic method has been extended by length-based spawning potential ratio (LBSPR, Hordyk et al. 2015a), length-based Integrated Mixed Effects (LIME, Rudd and Thorson 2017) and Length-Based Bayesian approach (LBB, Froese et al. 2018) models, among others. These allow for the estimation of instantaneous fishing mortality (F) and spawning potential ratio (SPR) when basic biological parameters are known. In contrast to LBSPR and LBB, LIME does not assume equilibrium conditions. The mixed-effects aspect of LIME extends length-based methods by estimating changes in recruitment and separating them from fishing mortality over time (Rudd and Thorson 2018).

It is good practice to simulation test the performance of assessment methods before applying them in practice (Cope 2008). This can be done using a variety of approaches, though it is most often accomplished using an Operating Model (OM) to generating pseudo-data with error to fit an assessment model (Punt et al. 2016). Simulation can either be an open loop or a closed loop with feedback. Carruthers (2016), using closed loop simulations, found that data-limited methods using observations of stock depletion offer the best overall performance across life history types, data quality and autocorrelation in recruitment strength. However, these management procedures are based on setting catch limits and were designed for use in data-limited fisheries for which annual catch data are available, sometimes together with a relative abundance index (Carruthers et al. 2014). In many data poor fisheries, measuring total removals is difficult, as is enforcing catch limits. Recently, Hordyk et al. (2015b) tested some harvest strategies using a simulation approach to assess the utility of LBSPR as a tool for management in data-limited fisheries using an effort-based harvest control rule. They found that the LBSPR assessment model with an iterative effort-based harvest control rule can be used to rebuild an overfished stock to sustainable levels or fish down a stock to the target SPR. So far, these two approaches, output-based limits versus input/effort-based limits, have not been directly tested.

In addition, there are some studies comparing the performance in determining stock status of catch-based (Wetzel and Punt, 2015; Rosenberg et al. 2017) and length-based assessment models (Chong et al. 2019). However, a comparison of the performance of both length-based and catch-based methods to estimate stock status is needed. Unfortunately, finding a common metric between catch-based and length-based stock status metrics is difficult; the former measures overfishing via fishing rates and the latter biomass-based catch limits.

Even though many countries are moving to close-loop simulations or Management Strategy Evaluation (MSE) to assess the performance of different management procedures (i.e., the combination of analytical method and control rule) for data-limited fisheries (Harford and Carruthers, 2017), it remains a need to understand how the individual analytical methods perform. Therefore, this study used a simulation approach to better estimate relative bias and precision of a range of data-limited methods so researchers can choose which one could be more appropriate to use for the fishery they need to assess, with the results providing general information needed to construct an appropriate control rule for a given method's performance.

We used OMs (here considered just as a simulated population model) to represent the main sources of uncertainty and to generate data for use in data-limited stock assessment methods and to evaluate how well the methods perform when the data and knowledge requirements are met. A common metric is then used for comparison across model types, namely exploitation or harvest rates, to evaluate the performance of both catch-based and length-based models in a simulation context. We evaluate performance considering three fish populations with contrasting life-history strategies, under three different harvest trends, and three different levels of final stock depletion.

METHODS

Simulation studies commonly make different operating model assumptions from those of the methods being tested to allow the evaluation of robustness; in some cases, however, the same population model is used for both simulation and estimation, i.e. self-testing (Deroba et al., 2015). Using the same model structure for simulation and estimation can result in more optimistic results that might not be true under many scenarios (Francis, 2012) since it is not possible to explore the impact of model assumptions when the model used for simulation and estimation is the same. If a method performs poorly, however, when the assumptions in the OM are the same as the assessment model, it is unlikely to perform well in practice. To evaluate robustness to model structure our approach evaluated multiple data-limited assessment methods with a range of assumptions about population and fishery dynamics using an OM decoupled from the tested methods.

Twenty-seven different OMs were created using a factorial design comprising 3 harvest rates, 3 life-history types, and 3 depletion scenarios. The different harvest rates scenarios, each considered a 20-year time series of fishing, correspond to fishing mortality histories commonly seen in many fisheries. In harvest rate scenario 1, fishing mortality increases until it reaches a maximum and starts declining afterwards. This is commonly seen once management measures are implemented to reduce fishing pressure. Harvest rate scenario 2 assumes that fishing mortality increases and remains constant after reaching a maximum. This harvest rate profile could result from the implementation of catch or effort management limits. Harvest rate scenario 3 has constantly increasing fishing mortality, which would occur for fisheries that are still developing (Figure 1a).

Three life history types of varying longevity and somatic growth rate were simulated, namely (i) a short-lived fast-growing species, Pacific chub mackerel (*Scomber japonicus*), (ii) a medium-lived medium-growing fish, albacore tuna (*Thunnus alalunga*), and (iii) a longer-lived slow-growing species, canary rockfish (*Sebastes pinniger*) (Table 1). Finally, the following three final depletion levels (D) were considered: (i) heavily fished (D = 0.2; *i.e.* stock biomass in the last year is 20% of virgin biomass), (ii) moderately fished (D = 0.4) and (iii) lightly fished (D = 0.6).

Operating model specifications

The OM was developed to simulate resource dynamics under the different fishing scenarios, life histories and final depletion levels. The OMs consist of an age-structured population with numbers at age over time modelled as follows:

$$N_{a,t} = \begin{cases} R_t, & a = 0 \text{ and } t = 0 \\ N_{a-1,t}\, e^{(-M - F_t S_{a-1})}, & 0 < a < A \text{ and } t = 1 \\ \dfrac{N_{a-1,t}\, e^{(-M - F_t S_{a-1})}}{1 - e^{(-M - F_t S_{a-1})}}, & a = A \text{ and } t = 1 \\ N_{a-1,t-1}\, e^{(-M - F_t S_{a-1})}, & 0 < a < A \text{ and } t > 1 \\ (N_{a-1,t-1} + N_{a,t-1})\, e^{(-M - F_{t-1} S_{a-1})}, & a = A \text{ and } t > 1 \end{cases}$$

R_t is the number of age-0 animals at the start of year t, $N_{a,t}$ is the number of fish of age a at the start of the year t, S_a is the selectivity at age, F_t is the instantaneous fishing mortality rate during year t, M is the instantaneous rate of natural mortality, and A is the age of the plus group. Fishing mortality deviations were included as $F_t \sim lognormal\ (F_{t-1}, \sigma_F^2)$. A Beverton–Holt spawner–recruitment function (Beverton and Holt 1957) and annual normally distributed recruitment deviations $N(0,\sigma_R)$ were assumed (Table 1).

The biomass in each year t was calculated as $B_t = \sum_{a=1}^{A} N_{a,t} w_a$ where $w_a = \alpha L_a^{\beta}$ (parameters in Table 1). In addition, the predicted total catch by year (C_t) was calculated as $C_t = \sum_{a=0}^{A} C_{a,t}$ with:

$$C_{a,t} = \frac{F_t S_a}{M + F_t S_a} N_{a,t}\left(1 - e^{(-M - F_t S_a)}\right)$$

For each OM (N=27), 100 time series of fishing mortality were simulated, and the harvest rate per year (U_t) as C_t/B_t calculated (Figures S1 to S3). Each simulated population began at the unfished biomass level and all fishing trend scenarios terminate at the specified depletion level (Appendix Figures A1 to A27).

To condition the OM, published life-history parameter values (Table 1) reported in formal stock assessments were used (Crone and Hill 2015 for the short-lived Pacific chub mackerel; Anon 2014 for the medium-lived albacore tuna; and Thorson and Wetzel 2015 for long-lived canary rockfish). Each population was assumed to be targeted in a single area, by one fleet with a selectivity pattern (Table 1) that was logistic and constant through time.

Length bins were defined as they were in their respective assessments; every 2 cm (Crone and Hill 2015; Anon 2014; Thorson and Wetzel 2015). To obtain the catch length frequency, the probability (p) of being in a length bin (j) given age (a) was calculated as:

$$p_{j,a} = \begin{cases} \emptyset\left(\dfrac{j - L_a}{L_a CV_L}\right), & j = 1 \\ \emptyset\left(\dfrac{j - L_a}{L_a CV_L}\right) - \emptyset\left(\dfrac{j - 1 - L_a}{L_a CV_L}\right), & 1 < j < J \\ 1 - \emptyset\left(\dfrac{j - 1 - L_a}{L_a CV_L}\right), & j = J \end{cases}$$

With the predicted probability of harvest by length bin being:

$$\pi_j = p_{j,a} \frac{\sum_{a=0}^{A} N_{a,t} S_a}{N_t}$$

One thousand fish per year were drawn using a multinomial distribution with a π_j probability (Rudd and Thorson 2018).

Comparing methods outputs

One of the challenges when comparing catch-based and length-based methods is that they produce different model outputs. Catch-only models calculate total and/or spawning stock biomass and MSY, whereas length-based models estimate exploitation and transient SPR, which can be used to infer stock status. These are fundamentally different measures of the population status. As a result, our performance metric is defined as the error relative (RE) to the OM, where RE = $(U_{Method} - U_{OM}) / U_{OM}$. This allows for a measure of uncertainty, in both bias and precision, for all methods under each scenario, and is used as a standardized metric of model performance. Bias in this study is how far, on average, the performance measure from each estimation model is from the OM. Imprecision is related to the variability (variance) around the central tendency.

We used U as a common measure for comparisons between each data limited method and the OM. For catch-only approaches it is defined as the ratio catch/biomass; while for the length-based models, the estimated F was transformed to an exploitation rate via $U = 1 - exp$

(- *F*). In addition, we present the average RE across the last five years of the time series, not along the entire time series of data, because we are interested in the estimation of the current exploitation rates. Also, multiple studies have shown that catch-based methods might be appropriate to predict sustainable catch at the end of the time series, but not to reconstruct a biomass time series (Carruthers et al. 2012; Wetzel and Punt 2015).

Estimation models

All simulations, and data-limited model calculations, were conducted using the open-source statistical software R (R Core Team 2018). Each catch-based and length-based method evaluated here are summarized below.

Catch-based data-limited methods

Catch-MSY (Martell and Froese 2013) is a SRA approach that assume a Schaefer biomass dynamic model. Inputs are a time series of removals, priors for the population rate of increase at low population size (r), carrying capacity (K), and a range of stock depletion in the final year (Table 1). Values of r and K are randomly sampled using a Monte Carlo approach to detect ‘viable’ r-K pairs. A parameter pair is considered ‘viable’ if the corresponding biomass trajectories calculated from a production model are compatible with the observed catches, so that the population abundance never falls below 0, and is compatible with the prior assumption of relative biomass (i.e., stock depletion; Martell and Froese 2013). r-K pairs are drawn from uniform prior distributions and the Bernoulli distribution is used as the likelihood function for accepting each r-K pair. Catch-MSY uses catch and productivity to estimate MSY. Here we use the modified version of Catch-MSY (Rosenberg et al. 2017) to extract biomass trends from all viable r-K pairs using the R package *datalimited* version 0.1.0 (Anderson et al. 2016). The biomass trajectory is calculated as the median of all viable biomass trajectories generated under the Monte Carlo process. We decided to include Catch-

MSY in this performance comparison analysis since it is a method that has become very popular in developing countries because of its easy implementation through R libraries.

CMSY (Froese et al. 2017) extends Catch-MSY by using a Monte-Carlo filter (instead of the SIR algorithm) that fixes systematic biases in the Catch-MSY method. It also, explicitly incorporates process error and estimates target reference points (MSY, F_{MSY}, B_{MSY}) as well as relative stock size (B/B_{MSY}) and exploitation (F/F_{MSY}) from catch data and priors for *r* and depletion at the beginning and the end of the time series. CMSY has an inbuilt piecewise "hockey-stick" to prevent over-estimating of rebuilding potential at very low abundance $B<0.25B_0$. The CMSY package implements a Bayesian state-space implementation of the Schaefer surplus production model (Winker 2019). CMSY was included in this paper because it has been described as an unbiased version of Catch-MSY and it has been already implemented in different parts of Europe (Froese et al. 2018) and explored by Regional Fisheries Management Organizations such as the International Commission for the Conservation of Atlantic Tunas, ICCAT (Winker et al. 2017, ICCAT 2017). We included a prior for depletion in this method (see Table 1).

DBSRA (Dick and MacCall 2011) modifies the SRA approach by using Monte Carlo draws from four parameter distributions (M, F_{MSY}/M, B_{MSY}/B_0, and depletion) and age at maturity (A_{mat}) to separate the total biomass into immature and mature biomass (fishery selectivity is also assumed to have an identical pattern to the age-at-maturity ogive). It uses a delay-difference production model with a time lag for recruitment and mortality as:

$$B_{t+1} = B_t + P(B_{t-Amat}) - C_t$$

where B_t is the biomass at the start of year t, $P(B_{t-Amat})$ is the latent annual production based on a function of adult biomass in year $t\text{-}Amat$ and C_t is the catch in year t. Biomass in the first year (B_0) is assumed to be equal to K. The package *fishmethods* version 1.10-3 was used to perform this analysis (Nelson 2017). For DBSRA we used the *Amat* and M as fixed inputs and three priors: final stock depletion, F_{MSY}/M, and B_{MSY}/B_0 (distributions in Table 1). Each of these is assigned a distribution from which the Monte Carlo draws are taken. We chose DBSRA to be included in our analysis because it is currently used for providing fisheries management advice on the US West Coast.

SSS is based on the Stock Synthesis age-structured stock assessment model (Methot and Wetzel 2013). SSS fix all parameters in the Stock Synthesis model except for initial recruitment ($\ln R_0$). It also sets up an artificial index of abundance that represents the relative stock biomass. The first value of the index is always 1, and the value in the final year represents the percent of the population left in that year (final depletion). The values of steepness (h) and the final year of the abundance survey are all randomly drawn from a specified distribution using a Monte Carlo approach (Cope 2013) and $\ln R_0$ is then estimated. Benefits of this approach are that it retains the same modelling framework as the data-rich stock assessments, but still allows for flexibility in a variety of parameter and model specifications, if desired. The input priors used for SSS were relative stock status and steepness and selectivity was matched to the OM (Cope 2019). We chose SSS to be included in our analysis because it is gaining use as a flexible platform to incorporate more data as it is collected, as well as providing an age-based alternative to the other catch-only models.

Length-based data-limited methods

SPR is the proportion of the unfished reproductive potential per recruit under a given level of fishing pressure (Goodyear 1993). In **LBSPR,** SPR in an exploited population is calculated as a function of the ratio of fishing mortality to natural mortality (F/M), selectivity, and the two life-history ratios M/k and L_m/L_∞; k is the von Bertalanffy growth coefficient, L_m is the size of maturity and L_∞ is asymptotic size (Hordyk et al. 2015a). The inputs of LBSPR are M/k, L_∞, the variability of length-at-age (CVL_∞), which is normally assumed to be around

10%, and the length at maturity specified in terms of L_{50} and L_{95} (the size at which 50% and 95% of a population matures). Given the assumed values for *M/k* and L_∞ and that length composition data come from an exploited stock, the LBSPR model uses maximum likelihood methods to estimate the selectivity ogive, which is assumed to be of a logistic form defined by the selectivity-at-length parameters S_{50} and S_{95} (the size at which 50% and 95% of a population is retained by the fishing gear), and *F/M*. The selectivity ogive and relative fishing mortality are then used to calculate SPR (Hordyk et al. 2015a, 2015b). Estimates of SPR are primarily determined by the length of fish relative to L_{50} and L_∞, but it also depends on life history parameters such as fecundity-at-age/length and selectivity. LBSPR is an equilibrium based method with the following assumptions: (i) asymptotic selectivity, (ii) growth is adequately described by the von Bertalanffy equation, (iii) a single growth curve can be used to describe both sexes which have equal catchability, (iv) length at-age is normally distributed, (v) rates of natural mortality are constant across adult age classes, (vi) recruitment is constant over time, and (vii) growth rates remain constant across the cohorts within a stock (Hordyk et al. 2015a). Analyses were conducted using the LBSPR package, version 0.1.3 in R (Hordyk 2018). We used the Rauch-Tung-Striebel smoother function available in the LBSPR package to smooth out the multi-year estimates of *F*. We decided to include LBSPR in the analysis as its easy implementation through R libraries and online apps have made it a popular, widespread choice.

LIME uses length composition of the catch and biological information to estimate *F* and SPR. LIME has the same data requirements as LBSPR, but does not assume equilibrium conditions. The mixed effects aspect of LIME extends length-based methods by estimating changes in recruitment and fishing mortality separately over time (Rudd and Thorson 2018). LIME uses automatic differentiation and Laplace approximations as implemented in Template Model Builder (TMB; Kristensen et al. 2016) to calculate the marginal likelihood for the mixed-effects. All other assumptions are the same as LBSPR but LIME estimates one selectivity curve for the entire time series of length data while LBSPR estimates one selectivity curve for each year since each time step estimation in LBSPR is independent (Hordyk et al. 2015a). The inputs to LIME (Rudd 2019) are: M, k, L_{∞}, t_0, CVL_{∞}, L_{50}, L_{95}, h, and the parameters of the length-weight relationship a and b (Table 1). We decided to include LIME in our analysis because is the only length-based method that allows for recruitment deviations and it has not been widely tested.

LBB is a simple and fast method for estimating relative stock size that uses a Bayesian Monte Carlo Markov Chain (MCMC) approach (Froese et al. 2018). In contrast to other length-based methods, LBB uses pre-specified priors on parameters, and thus, technically does not require further inputs in addition to length frequency data, if the user is willing to accept the life history defaults. However, it provides the user the option to specify priors for the inputs L_{∞}, length at first capture (Lc), and relative natural mortality (M/k). We specified the true M/k value and we let the model calculate Lc (length where 50% of the individuals are retained by the gear) and L_{∞}, which is approximated by the maximum observed length L_{max}. In addition, F/M is estimated as means over the age range represented in the length-frequency sample. We decided to include LBB because it has not been widely tested and compared with other length-based assessments methods. As well as CMSY, LBB is gaining consideration as a plausible method in some international commissions such as ICCAT (Anon, 2019).

During our simulation testing, we assumed that length models had the correct values for the von Bertalanffy length-at-age relationship, L_{∞}, k and t_0, length-weight parameters α and β, M, and the parameters L_{50} and L_{95} from a logistic maturity-at-length curve. We did not evaluated misspecifications in life history parameters inputs. The objective is only to compare the performance of different data-limited methods under different scenarios, not to evaluate each of them under different parameters misspecifications, which has already been done in the original publications of each method.

RESULTS

Large variability is seen in model performance across harvest scenarios, life-history types, and depletion levels (Figure 1b-d, Figure 2). Figures 1 and 2 can be used to search for the best model, i.e. to identify the least biased and least imprecise method based on the life history of the species, harvest trends, and knowledge of final depletion. The imprecision of each method is shown in the variability around those estimations in Figure 1 and the specific values in Figure 2. A robust method would show low bias and high precision for all life history and harvest scenarios; for example, among all catch-based methods, SSS appears to be the most robust method (Figure 2).

In general, catch-based methods tended to be more biased at relatively higher stock sizes (D=0.6) and for long-lived species (Figures 1 and 2). Catch-MSY tended to perform poorly across a range of scenarios. Among the length-based models, all were less biased and more precise for the medium-lived species (Figure 2) without a clear different performance among harvest scenarios. The most biased length-based method was LBSPR although more precise than LIME and LBB (Figure 2). Overall, LIME and SSS were the most robust methods.

Catch-only methods

It was to be expected that the various catch-based models considered in this study would perform differently among them because they have different model structure and assumptions. SSS performed best in most cases estimating unbiased exploitation rates across different scenarios of harvest trends, final stock depletion and life histories (Table 2). However, it tended to underestimate harvest rates by 30% for medium and short-lived species at relatively high stock sizes (D=0.6). For long-lived species the estimations were slightly overestimated for moderately fished stock sizes (D=0.4), but SSS was always the model that was least biased. DBSRA was the most precise but it underestimated harvest rates in general, and in contrast to the other catch-based methods, it was less biased for stocks at relatively high stock sizes (D=0.6, Figure 2). Catch-MSY was the most biased of the catch-based models tested, overestimating harvest rates in particular for stocks lightly fished (D=0.6). It was less biased to itself for medium-lived species and produced non-biased estimates of MSY for highly depleted (D=0.2) medium-lived or short-lived species at moderate stock sizes (D=0.4) (Table A1). CMSY was less biased in the relatively medium to high stock sizes (D=0.4 and D=0.6) and for medium-lived species (Table 2). In general, catch-based models were less biased and more precise when stocks were at relatively low stock sizes (i.e. using a prior centered around 0.2, Figure 1). In general, estimation models in harvest scenario 1 (ramp shaped) produced the most variable estimations in harvest rates (Figure 1, Table 2). The catch-based methods performed the best for the medium-lived species.

Length-based methods

In many cases, length-based models gave a less biased estimation of *U* than catch-based models (Figure 1). LIME was the least biased length-based method (Table 3). However, in general, LIME did not converge around 10% of the time. The three length-based methods used here (LIME, LBSPR and LBB) produced more variable estimations in harvest scenario 1, where the fishing intensity decreased at the end of the time series (Figure 1) but no clear pattern was observed in terms of harvest scenarios in bias (Figure 2). LBSPR in general underestimated harvest rates. Compared to itself, LBB performed better in scenarios where the stocks have relatively low to medium stock sizes (D=0.2 and D=0.4) and for medium-lived species (Figure 2). Both, LBB and LIME were highly variable for long-lived species and harvest scenario 1 (Figure 1). Multiple modes for harvest scenario 1 may suggest poor convergence. Overall, length-based models were less biased for the medium-lived species.

Short-lived species

For the short-lived life-history strategy, SSS was the least biased and most precise among the catch-based methods. The second most precise method was CMSY, but this method was positively biased for lightly fished or relatively high stock sizes (D=0.6). Among the three length-based models, LIME had better overall performance than LBSPR and LBB. However, LBSPR was less biased in harvest scenario 1. LBB was the one model that presented the most variability in harvest-rate estimations for the short-lived species.

Medium-lived species

Both catch-based and length-based models gained more precision as the stocks were more depleted. Among the catch-based methods, CMSY was the less biased followed by SSS, although SSS was more precise. DBSRA was highly negatively biased particularly for relatively low stock sizes (D=0.2). Catch-MSY was highly positively biased but less so for highly depleted stocks (Figure 1c). Between the length-based models, LIME showed very good performance with regards to bias for harvest scenarios 2 and 3 under the three depletion levels, although it was slightly biased for harvest scenario 1. LBSPR and LBB showed similar performance but LBSPR was more precise (Figure 1c).

Long-lived species

Both catch-based and length-based methods were less precise (more variability in RE) as the long-lived stocks had relatively higher stock sizes (D=0.6), as they did for the other life-history strategies. Among the catch-based methods, SSS was the most precise and least biased method except in harvest scenario 1 and D=0.2 where CMSY was the least biased. SSS underestimated harvest rates in scenarios 2 and 3, where the catch history is constant or increases at the end of the time series. Among assessment methods, a less biased and more precise estimation was observed for medium depleted stocks (D=0.4) than for other depletion levels. LBSPR and DBSRA was negatively biased in all cases (Figure 1d). LIME was highly imprecise in harvest scenario 1, but the least biased among the other length-based methods (Figure 1c, Figure 2).

Discussion

It is to be expected that the various methods would perform differently. Rosenberg et al. (2017) used four catch-based data-limited models and found that models frequently disagreed about population status estimations, with no model showing overall good performance, i.e. high precision and low bias across all case studies. When scenarios represent specific resource dynamics or particular stocks or fisheries, it may be difficult to draw any overall conclusions. Therefore, we chose scenarios that represented different fishing intensity trends, depletion levels, and life histories. We found that model performance is highly dependent on all these factors. More imprecision occurred where fishing pressure decreases at the end of the time series (harvest scenario 1), in comparison with harvest scenarios where *F* was either stable or increasing (harvest scenarios 2 and 3) that performed similarly in terms of bias and imprecision. In addition, most bias was found when relative stock abundance is high and/or for the slow-grow long-lived life history.

In particular, catch-based models performed better (i.e. were less biased and more precise) for stocks that were medium to highly depleted than for lightly depleted stocks. Walters et al. (2006) suggested that for SRA, stocks that have experienced extensive historical depletion gain precision due to a high rate of rejected parameter draws. The SSS approach, which tracks age-structure population dynamics, performed better than the models that are based on lumped biomass production functions such as Catch-MSY, CMSY, and DBSRA, even when priors for depletion were centered on the true values for all methods. Although SSS seems to be the least biased catch-based model, unlike other catch-based models, more detailed life-history information (e.g., age and growth estimates) are required by SSS to define age structure and remove catch according to age-/size-based selectivity patterns (Cope, 2013). So, when this information is highly uncertain, sensitivity analysis to parameter misspecifications is strongly recommended.

Catch-MSY performed poorly in all scenarios, overestimating harvest rates even when given a prior for depletion close to the true value. A key point of Catch-MSY is the ability to define a reasonable prior range for the parameters of the Schaefer model, in particular K. In our case, we have arbitrarily chosen 100 times the maximum catch as the upper bound for K based on Martell and Froese (2013). Other priors for K could be explored to see if this improves the outcome, but it remains a difficult parameter to specify under most conditions.

CMSY on the other hand performed particularly well with respect to bias and precision for medium-lived species, even better than SSS for medium to low depleted stocks (Figure 2). Also, CMSY was more accurate than the original Catch-MSY method (Martell and Froese, 2013). The difference is that Catch-MSY was designed to select the most probable r-K pair as the geometric mean of this distribution, but CMSY searches not in the center of the distribution but rather near the right tip of viable pairs. According to Froese et al. (2017) since r is defined as the maximum net productivity, the right tip of the distribution of r-K pairs is where these parameters should be found. So, between Catch-MSY and CMSY, CMSY is preferred, though its overall poor performance is noted.

Hordyk et al. (2015a) explained how LBSPR relies on detecting the signal of fishing mortality in the right-hand side of the length composition. Consequently, fishing is not likely to have a visible impact on the length composition until fishing mortality is very high and stocks are highly depleted. This is why LBSPR was less biased for more depleted populations and in fishing scenario 1.

Our study found that LIME was highly imprecise for long-lived species. Rudd and Thorson (2017) also showed that LIME is more imprecise for long lived species. The model is trying to track cohorts through the length data to estimate recruitment deviations and this is likely difficult for long-lived species when time series of length data are short or much of the population is found near the asymptotic size (Rudd and Thorson 2017).

Hordyk et al. (2019) suggested that LBB has not been sufficiently simulated tested and it can produce biased estimates of fishing mortality. We found that LBB was the most biased and imprecise length-based method, although for the less depleted stocks it generally performed better than LBSPR. One of criticisms of LBB by Hordyk et al. (2019) is that it *a priori* assumes that $M/k = 1.5$, however here, we specified the true M/k value. So, the main bias was associated with the estimations of L_∞ due to the approximation using maximum observed length L_{max}. There are many reasons why the maximum length may be a biased measure of L_∞. In addition to L_∞, L_c was always overestimated by the LBB assessment model (see Table A2), likely due to the bias in the L_∞ estimates.

In general, all catch-based and length-based methods seem to perform worse for long-lived life-history types, where there is likely to be less contrast in the dynamics over time (few years relative to their life span), than for medium and short-lived. The length of the time series for the long-lived canary rockfish is therefore probably too short in comparison to the age they reach the maximum length (64 years), to capture the true dynamics of the population and the response to different harvest rates.

The present study did not look at parameter misspecification but correctly specified (i.e., unbiased values) the life-history parameters. As we mentioned before, parameters misspecification testing was performed in the original publications for each method. With accurate prior information, length-based models showed better performance in many cases than some catch-based models, as the latter were more sensitive to the catch history scenarios and depletion levels.

We note that bias is not symmetrically interpretable, as it could, depending on management objectives, be better or worse to be biased high or low. In data-limited management, maximum sustainable yield is often hard to pinpoint, thus a more precautionary pretty good yield (Hilborn 2010) is likely a better target. When measuring exploitation rates, this would translate to being biased high as a preferable way of being wrong rather than underestimating the exploitation rate and possible increasing the fishing rate above the true target. The consideration of asymmetric management significance when measuring bias is an important consideration when using the results presented here to develop control rules.

Bias and precision are both important factors to consider when assessing fish stocks. Bias reflects how close an estimate is to a known value; precision reflects reproducibility of the estimate. For example, if an assessment is to be re-conducted every year to monitor the impact of a management measure, a precise but biased method would be able to detect a trend better than an unbiased but imprecise method. Like a scientific instrument, this trade-off require calibration to correct for the bias, and such calibration can be explored using closed-loop simulations such as management strategy evaluation, where the choice of parameters and reference points in a management procedure are tuned (i.e. calibrated) to meet the desired management objectives as represented by the operating model. Thus, a biased method (e.g., DBSRA) may be preferable to one that is less biased, but more imprecise (e.g., LIME). Alternatively, imprecision can be addressed through the choice of the percentile (e.g., median being the 50% percentile value) for the derived model output used by management (e.g., catch or SPR); assuming that the true value is contained within the parameter distribution. For example, instead of taking the median value, one could instead use the derived model output associated with the 40th percentile to incorporate risk tolerance as reflected in the calculated imprecision. Such an approach (Ralston et al. 2011) is used in fisheries management systems to directly incorporate scientific uncertainty (both bias and imprecision), and can also be explored and tuned using MSE.

Recommendations

To provide estimates of stock status for unassessed fisheries where data are limited, but reconstructing time series of catch is possible, SSS is recommended. The performance of SSS hinges on the correct specification of the input parameters such as stock depletion, productivity, maturity, and growth parameters. Knowledge about some of these parameters, especially stock depletion, is likely to be poor for data-limited stocks, resulting in misspecification of these parameters. Meta-analyses may offer some starting values for life history parameters (e.g., Myers 2001, Thorson *et al.* 2012, and Zhou *et al.* 2012, 2017), however other inputs remain difficult to specify (Chrysafi and Cope *in press*).

For fisheries where the time series of catch are unavailable or catches are not consistently monitored and managed, using length-composition data can provide good approximations of the status of the stock, in particular for medium-lived species. It has been shown here that in some cases, length-based models such as LIME can provide the same or less biased estimates of exploitation status than catch-based models. However, growth parameters are even more important for length-based than for any catch-based method, so it is important to have good estimates of those parameters before using any length-based assessment.

Making recommendations on which models should be applied to estimate exploitation intensity in different fisheries is challenging because model choice is dependent on data availability, trends in fishing intensity, and the biology of the species. If possible, simulation studies testing different data-limited methods with OMs based on the focus species and the dynamics of the fishery can greatly inform which method is most appropriate. Likewise, decision support tools such as FishPath (Dowling et al. 2015) can also help one weigh the input requirements and assumptions to identify the most appropriate methods given data and life history. Based on the OMs used in this study, we conclude that when only catch data is available, SSS should be considered. When only length data is available, LIME may be less biased that LBSPR and LBB if recruitment variability is an important consideration. However, Pons et al. (2019) found that neither LBSPR or LIME are good in all situations, and thus, both should be considered and compared for inconsistent results. LIME sometimes undergoes convergence issues and has difficulties separating changes in recruitment from changes in fishing mortality (Pons 2018; Pons et al. 2019).

For long-lived species it is necessary to have longer time series of data to draw more conclusions. However, Pons (2018) recommended SSS and LBSPR when long time series (i.e. 80 years) of data are available for a species that lives more than 60 years to evaluate changes in fishing intensity.

If both catch and length data are available, models that integrate both data types should be considered. LIME, although primarily length based, allows for the inclusion of catch data as well as an index of abundance if one is available. Moreover, integrated assessment models (that use catch as well as length information) like Stock Synthesis could also be considered (Methot and Wetzel 2013). Length information can therefore be added to the SSS data file, with the possibility of freeing up the stock status assumption input, and running the model more like a traditional statistical-catch-at-age model (Cope 2013; Thorson and Cope 2015).

For the scenarios analyzed here, including the specific life-histories considered, we do not recommend Catch-MSY for estimating exploitation rates, even with a good estimate of stock depletion. This method will however, as it was originally created to do, produce unbiased estimates of MSY, in particular for short and medium-lived and highly depleted species (Table A1).

Future directions

Dowling et al. (2019) in a review of data limited methods, noted the dangers in the indiscriminate use of generic methods (i.e., methods that are either imported as solutions or considered just because no others are understood, but are not rigorously considered for appropriateness given data and assumptions) and recommended obtaining better data, using care in acknowledging and interpreting uncertainties, developing harvest strategies (including control rules) that are robust to these higher levels of uncertainty and tailoring them to the species and fisheries specific data and context. Management actions to regulate fishing can be based on changes in harvest rates, mainly when catch limits are not an option. The steps, for each fishery, would include i) developing reference points, ii) identifying monitoring and assessments options, and iii) a qualitative and possibly quantitative (e.g., closed-loop simulation) evaluation of how to adjust harvest strategies and paired control rules to meet management objectives (Dowling et al. 2008). Control rules linked to methods explored in this work can be tailored and tested based on the bias and imprecision found in this study.

In addition, these data-limited methods could be tested using a fully-specified MSE with stakeholder input to specify the management objectives in order to determine management procedures to help ensure robust and sustainable fisheries management. This evaluation includes the benefits of adaptive improvement of the harvest strategy and management of the fishery.

This study provides steps to the above closed-loop simulation by way of conditioning OMs and generating pseudo data for use by the management procedure. Having a common metric (exploitation rate) to compare methods that are often decoupled in performance testing allows for a comparison of the components of uncertainty, bias and imprecision. Understanding the comparative degree each of these methods express component uncertainty under control simulation testing benefits the next steps of closed-loop performance testing if management procedures. The importance of considering assessment methods as part of a management procedure is that a method that provides biased, yet precise results could, if calibrated correctly, provide useful advice, whereas unbiased estimates with high imprecision would need specific consideration on how to manage the risk inherent in imprecision.

Acknowledgments

MP was funded by the Walton Family Foundation (#2018-51) and International Seafood Sustainability Foundation (ISSF). LK's involvement was funded through the MyDas project under the Marine Biodiversity Scheme which is financed by the Irish government and the European Maritime & Fisheries Fund (EMFF) as part of the EMFF Operational Programme for 2014-2020. We want to acknowledge Trevor Branch, Victor Restrepo, Dave Fluharty, Jim Ianelli and Ray Hilborn for their significant comments in the early stages of this study. We also want to acknowledge Merrill Rudd, Chantel Wetzel and Ian Taylor for their help with SS, r4ss, and other packages and models used in this study and Nicole Baker and Kelli Johnson for proofreading the final version of the manuscript.

REFERENCES

Anderson, S.C., Afflerbach, Jamie Cooper, A.B., Dickey-Collas, M., Jensen, O.P., Kleisner, K.M., Longo, C., Osio, G.C., Ovando, D., Minte-Vera, C., Minto, C., Mosqueira, I., Rosenberg, A.A., Selig, E., Thorson, J.T., and Walsh, J.C. 2016. datalimited: Stock Assessment Methods for Data-limited Fisheries. R package version 0.1.0. Available from https://github.com/datalimited/datalimited.

Anderson, S.C., Cooper, A.B., Jensen, O.P., Minto, C., Thorson, J.T., Walsh, J.C., Afflerbach, J., Dickey-Collas, M., Kleisner, K.M., Longo, C., Osio, G.C., Ovando, D., Mosqueira, I., Rosenberg, A.A., and Selig, E.R. 2017. Improving estimates of population status and trend with superensemble models. Fish Fish. **18**(4): 732–741. doi:10.1111/faf.12200.

Anon. 2014. Report of the 2013 ICCAT North and South Atlantic Albacore stock assessment meeting. Collect. Vol. Sci. Pap. ICCAT. **70**(3): 830–995.

Anon. 2017. Report of the 2017 ICCAT Albacore species group intersessional meeting (including assessment of Mediterranean Albacore). Collect. Vol. Sci. Pap. ICCAT. **76**(5): 1–43.Anon. 2019. Report of the 2019 ICCAT working group on stock assessment methods meeting (WGSAM). Collect. Vol. Sci. Pap. ICCAT. **70**(3): 830–995.

Beverton, R.J.H., and Holt, S.J. 1957. On the Dynamics of Exploited Fish Populations. Fishery Investigations (Great Britain, Ministry of Agriculture, Fisheries, and Food), London.

Carruthers, T.R., Walters, C.J., and McAllister, M.K. 2012. Evaluating methods that classify fisheries stock status using only fisheries catch data. Fish. Res. **119**: 66–79. doi:10.1016/j.fishres.2011.12.011.

Carruthers, T.R., Punt, A.E., Walters, C.J., MacCall, A., McAllister, M.K., Dick, E.J., and Cope, J. 2014. Evaluating methods for setting catch limits in data-limited fisheries. Fish. Res. **153**: 48–68. Elsevier B.V. doi:10.1016/j.fishres.2013.12.014.

Carruthers, T. R., Kell, L. T., Butterworth, D. D. S., Maunder, M. N., Geromont, H. F., Walters, C., McAllister, M. K., *et al.* 2016. Performance review of simple management procedures. ICES J. Mar. Sci. **73**(2): 464–482. doi: 10.1093/icesjms/fsv212.

Chong, L., Mildenberger, T.K., Rudd, M.B., Taylor, M.H., Cope, J.M., Branch, T.A., Wolff, M., and Sta¨bler, M. 2019. Performance evaluation of data-limited, length-based stock assessment methods. ICES J. Mar. Sci. doi:10.1093/icesjms/fsz212.

Chrysafi, A., and Kuparinen, A. 2016. Assessing abundance of populations with limited data: Lessons learned from data-poor fisheries stock assessment. Environ. Rev. **1**: 1–44. doi:10.1017/CBO9781107415324.004.

Chrysafi, A., and J.M. Cope. in review. Testing methods of determining relative stock abundance priors when setting catch recommendations using data-limited approaches. Fish. Res. under review.

Cope, J.M. 2008. Issues and Advances in Data-Limited Stock Assessment: Experimentation through Simulation. Diss. - Dr. Philos. degree: 1–235.

Cope, J.M. 2013. Implementing a statistical catch-at-age model (Stock Synthesis) as a tool for deriving overfishing limits in data-limited situations. Fish. Res. **142**: 3–14. Elsevier B.V. doi:10.1016/j.fishres.2012.03.006.

Cope, J. M. 2019. Simple Stock Synthesis code and examples. Available at https://github.com/shcaba/SSS.

Costello, C., Ovando, D., Hilborn, R., Gaines, S.D., Deschenes, O., and Lester, S.E. 2012. Status and solutions for the world's unassessed fisheries. Science **338**(6106): 517–20. doi:10.1126/science.1223389.

Crone, P.R., and Hill, K.T. 2015. Pacific mackerel (*Scomber japonicus*) stock assessment for USA management in the 2015-16 fishing year. Pacific Fish. Manag. Counc. 7700 NE Ambassad. Place, Suite 101, Portland, Oregon 97220, USA (May): 131.

Deroba, J.J., Butterworth, D.S., Methot Jr, R.D., De Oliveira, J.A.A., Fernandez, C., Nielsen, A., Cadrin, S.X., Dickey-Collas, M., Legault, C.M., Ianelli, J. and Valero, J.L., 2014. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES J. Mar. Sci. **72**(1): 19–30. doi: 10.1093/icesjms/fst237.

Dick, E.J., and MacCall, A.D. 2011. Depletion-Based Stock Reduction Analysis: A catch-based method for determining sustainable yields for data-poor fish stocks. Fish. Res. **110**(2): 331–341. Elsevier B.V. doi:10.1016/j.fishres.2011.05.007.

Dowling, N.A., Smith, D.C., Knuckey I., Smith, A.D.M., Domaschenz P., Patterson, H.M. and Whitelaw, W. 2008. Developing harvest strategies for low-value and data-poor fisheries: Case studies from three Australian fisheries. Fish. Res. **94**: 380–390.

Dowling, N.A., Dichmont, C.M., Haddon, M., Smith, D.C., Smith, A.D.M., and Sainsbury, K. 2015. Empirical harvest strategies for data-poor fisheries: A review of the literature. Fish. Res. **171**: 141–153. Elsevier B.V. doi:10.1016/j.fishres.2014.11.005.

Dowling, N., Wilson, J., Rudd, M., Babcock, E., Caillaux, M., Cope, J., Dougherty, D., Fujita, R., Gedamke, T., Gleason, M., Guttierrez, M., Hordyk, A., Maina, G., Mous, P., Ovando, D., Parma, A., Prince, J., Revenga, C., Rude, J., Szuwalski, C., Valencia, S. and Victor, S., 2016. FishPath: A Decision Support System for Assessing and Managing Data- and Capacity- Limited Fisheries, in: Quinn II, T., Armstrong, J., Baker, M., Heifetz, J., Witherell, D. (Eds.), Assessing and Managing Data-Limited Fish Stocks. Alaska Sea Grant, University of Alaska Fairbansk. https://doi.org/10.4027/amdlfs.2016.03

Dowling, N. A., Smith, A. D. M., Smith, D. C., Parma, A. M., Dichmont, C. M., Sainsbury K., Wilson, J. R., Dougherty, D. T. and Cope, J. M. 2019. Generic solutions for data-limited fishery assessments are not so simple. Fish Fish. **20**(1): 174–188. doi:

10.1111/faf.12329.

Francis, C. 2012. Reply to "The reliability of estimates of natural mortality from stock assessment models." Fish. Res. **119–120**: 133–134. Elsevier B.V. doi:10.1016/j.fishres.2012.02.015.

Free, C.M, Jensen, O.P., Wiedenmann, J., and Deroba, J.J. 2017. The refined ORCS approach: A catch-based method for estimating stock status and catch limits for data-poor fish stocks. Fish. Res. **193**: 60–70. doi: 10.1016/j.fishres.2017.03.017

Froese, R., Demirel, N., Coro, G., Kleisner, K.M., and Winker, H. 2017. Estimating fisheries reference points from catch and resilience. Fish Fish. **18**(3): 506–526. doi:10.1111/faf.12190.

Froese, R., Winker, H., Coro, G., Demirel, N., et al. 2018. A new approach for estimating stock status from length frequency data, ICES J. Mar. Sci. **75**(6). 2004–2015. doi: 10.1093/icesjms/fsy078.

Goodyear, C.P. 1993. Spawning stock biomass per recruit in fisheries management: foundation and current use. In Risk evaluation and biological reference points for fisheries management. Edited by S.J. Smith, J.J. Hunt, and D. Rivard. Can. Spec. Publ. Fish. Aquat. Sci. No. 120.

Harford, W.J., and Carruthers, T.R. 2017. Interim and long-term performance of static and adaptivemanagement procedures. Fish. Res. **190**: 84–94.

Hilborn, R. 2010. Pretty Good Yield and Exploited Fishes. Marine Policy **34**(1): 193–96. https://doi.org/10.1016/j.marpol.2009.04.013.

Hordyk, A., Ono, K., Valencia, S., Loneragan, N., and Prince, J. 2015a. A novel length-based empirical estimation method of spawning potential ratio (SPR), and tests of its performance, for small-scale, data-poor fisheries. **72**(1): 217–231.

Hordyk, A.R., Loneragan, N.R., and Prince, J.D. 2015b. An evaluation of an iterative harvest strategy for data-poor fisheries using the length-based spawning potential ratio assessment methodology. Fish. Res. **171**: 20–32. Elsevier B.V. doi:10.1016/j.fishres.2014.12.018.

Hordyk A. R. 2018. LBSPR: Length-Based Spawning Potential Ratio. R package version 0.1.3. https://github.com/AdrianHordyk/LBSPR.

Hordyk, A. R., Prince, J. d., Carruthers, T. R. and Walters, C. J. 2019. Comment on "A new approach for estimating stock status from length frequency data" by Froese et al. (2018). ICES J. Mar. Sci. **76**(2): 457–460. doi: 10.1093/icesjms/fsy168.

Kimura, D.K., and Tagart, J.V. 1982. Stock reduction analysis, another solution to the catch equations. Can. J. Fish. Aquat. Sci. **39**: 1467–1472. doi: 10.1139/f82-198.

Kimura, D.K., Balsiger, J.W., and Ito, D.H. 1984. Generalized stock reduction analysis. Can. J. Fish. Aquat. Sci. **41**: 1325–1333. doi: 10.1139/f84-162.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., and Bell B.M. 2016. TMB: Automatic Differentiation and Laplace Approximation. J. Stat. Software. **70**(5), 1-21. doi:10.18637/jss.v070.i05.

Martell, S., and Froese, R. 2013. A simple method for estimating MSY from catch and resilience. Fish Fish. **14**(4): 504–514. doi:10.1111/j.1467-2979.2012.00485.x.

Methot, R.D., and Wetzel, C.R. 2013. Stock synthesis: A biological and statistical framework for fish stock assessment and fishery management. Fish. Res. **142**: 86–99. doi:10.1016/j.fishres.2012.10.012.

Myers, R.A., 2001. Stock and recruitment: generalizations about maximum reproductive rate, density dependence, and variability using meta-analytic approaches. ICES J. Mar. Sci. **58**(5): 937–951 doi:10.1006/jmsc.2001.1109.

Nelson, G.A. 2017. fishmethods: Fishery Science Methods and Models in R. R package version 1.10-3. : https://CRAN.R-project.org/package=fishmethods.

Pons, M. 2018. Stock Status and Management in Tuna Fisheries: From Data-rich to Data-poor. University of Washington, ProQuest Dissertations Publishing, 2018: 10846859. 186 pp. https://search.proquest.com/docview/2126689812? accountid¼14784 (last accessed 15 November 2018).

Pons, M., Kell, L., Rudd, M. B., Cope, J. M., and Lucena Fredou, F. 2019. Performance of length-based data-limited methods in a multifleet context: application to small tunas, mackerels, and bonitos in the Atlantic Ocean. ICES J. Mar. Sci. **76**(4): 960–973, doi:10.1093/icesjms/fsz004.

Punt, A. 2008. Refocusing stock assessment in support of policy evaluation. Fish. Glob. Welf. Environ. 5th World Fish. Congr. (January 2008): 139–152. Available from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Refocusing+Stock+Assessment+in+Support+of+Policy+Evaluation#0%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Refocusing+stock+assessment+in+support+of+policy+evaluation#0.

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., Haddon, M., 2016. Management strategy evaluation: best practices. Fish Fish. **17**, 303–334. doi: 10.1111/faf.12104.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org.

Ralston, S., Punt, A.E., Hamel, O.S., DeVore, J.D., Conser, R.J. 2011. A meta-analytic approach to quantifying scientific uncertainty in stock assessments. Fish. Bull. **109**(2): 217–232.

Rosenberg, A.A., Kleisner, K.M., Afflerbach, J., Anderson, S.C., Dickey-Collas, M., Cooper, A.B., Fogarty, M.J., Fulton, E.A., Gutiérrez, N.L., Hyde, K.J.W., Jardim, E., Jensen, O.P., Kristiansen, T., Longo, C., Minte-Vera, C. V., Minto, C., Mosqueira, I., Osio, G.C., Ovando, D., Selig, E.R., Thorson, J.T., Walsh, J.C., and Ye, Y. 2017. Applying a New Ensemble Approach to Estimating Stock Status of Marine Fisheries Around the World. Conserv. Lett. **00**(April): 1–9. doi:10.1111/conl.12363.

Rudd, M.B., and Thorson, J.T. 2018. Accounting for variable recruitment and fishing mortality in length-based stock assessments for data-limited fisheries. Can. J. Fish. Aquat. Sci. **75**(7): 1019–1035. doi:10.1139/cjfas-2017-0143.

Rudd, M. B. 2019. Length-based integrated mixed effects model code. Available at: https://github.com/merrillrudd/LIME.

Thorson, J. T., Cope, J. M., Branch, T. A., Jensen, O. P., and Walters, C. J. 2012. Spawning biomass reference points for exploited marine fishes, incorporating taxonomic and body size information. Can. J. Fish. Aquat. Sci. **69**(9): 1556–1568. doi: 10.1139/f2012-077.

Thorson, J.T., and Wetzel, C. 2015. The status of canary rockfish (Sebastes pinniger) in the California Current in 2015. Pacific Fish. Manag. Counc. (June): 22.

Thorson, J.T., and Cope, J.M. 2015. Catch curve stock-reduction analysis: An alternative solution to the catch equations. Fish. Res. **171**: 33–41. Elsevier B.V. doi:10.1016/j.fishres.2014.03.024.

Walters, C.J., Martell, S.J.D., and Korman, J. 2006. A stochastic approach to stock reduction analysis. Can. J. Fish. Aquat. Sci. **63**(1): 212–223. doi:10.1139/f05-213.

Wetzel, C.R., and Punt, A.E. 2011. Model performance for the determination of appropriate harvest levels in the case of data-poor stocks. Fish. Res. **110**(2): 342–355. doi:10.1016/j.fishres.2011.04.024.

Wetzel, C.R., and Punt, A.E. 2015. Evaluating the performance of data-moderate and catch-only assessment methods for U.S. west coast groundfish. Fish. Res. **171**: 170–187. Elsevier B.V. doi:10.1016/j.fishres.2015.06.005.

Winker, H., Carvalho, F., Sharma, R., Parker, D., and Kerwath, S. Initial results for North and South Atlantic Shortfin Mako (*Isurus oxyrinchus*) stock assessment using the Bayesian Surplus Production Model JABBA and the Cacth-Resilience method CMSY. Collect. Vol. Sci. Pap. ICCAT. **74**(4): 1836–1866.

Winker, Henning. 2019. The CMSY method for data-limited stock assessment. Package available at https://github.com/Henning-Winker/cmsy.

Zhou, S., Yin, S., Thorson, J.T., Smith, A.D.M., Fuller, M., and Walters, C.J. 2012. Linking fishing mortality reference points to life history traits: an empirical study. Can. J. Fish. Aquat. Sci. **69**(8): 1292–1301. doi:10.1139/f2012-060.

Zhou, S., Punt, A.E., Ye, Y., Ellis, N., Dichmont, C.M., Haddon, M., Smith, D.C. and Smith, A.D. 2017. Estimating stock depletion level from patterns of catch history. Fish and Fisheries. **18**(4): 742–751. doi: 10.1111/faf.12201.

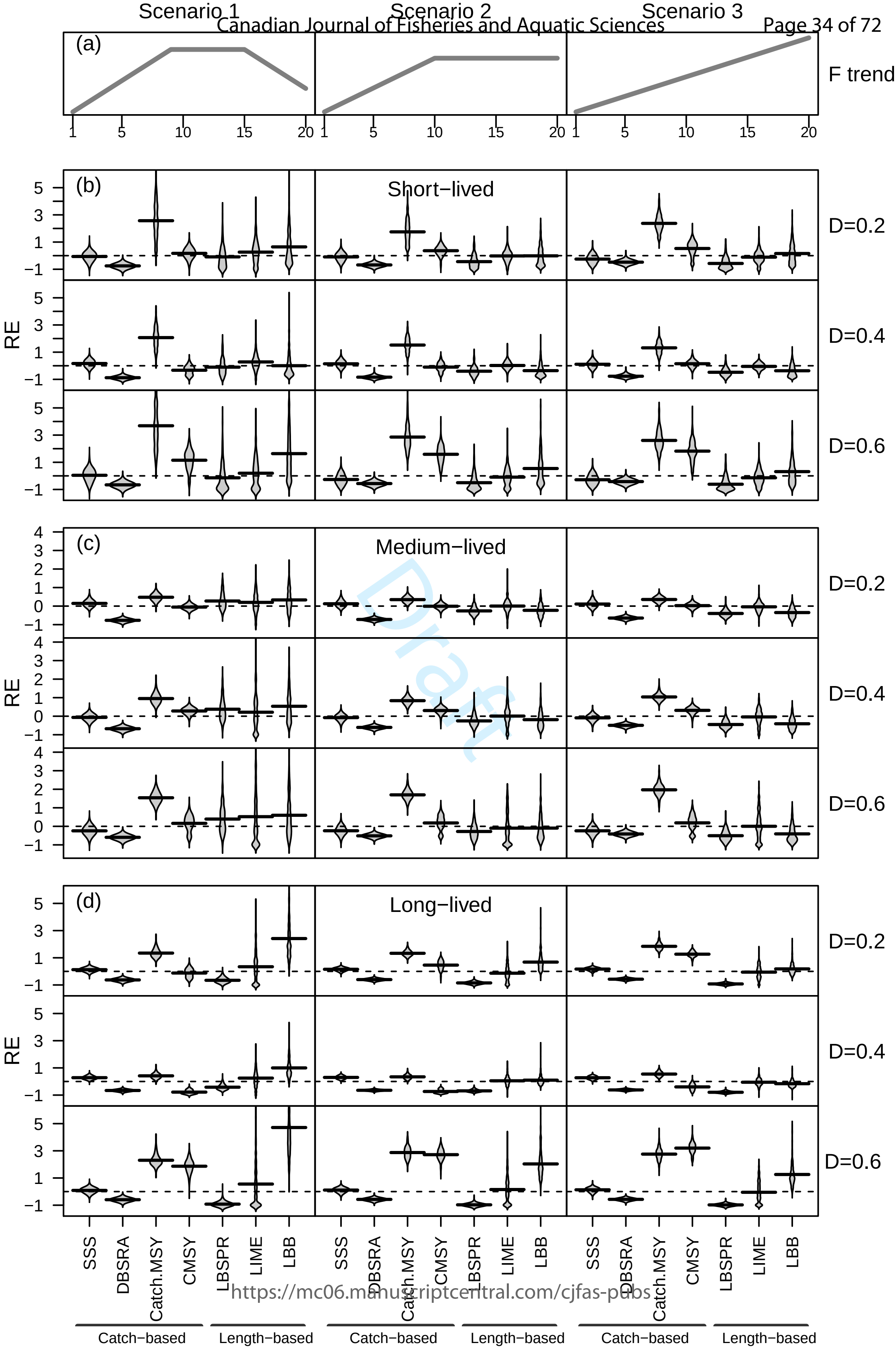
Scenario 1
Scenario 2
Scenario 3
(a)
F trend
(b)
Short−lived
(c)
Medium−lived
(d)
Long−lived
RE
D=0.2
D=0.4
D=0.6
SSS
DBSRA
Catch.MSY
CMSY
LBSPR
LIME
LBB
Catch−based
Length−based
Draft

Table 1. Life-history information and priors for the three species used in the study. Notation is as follows: *Lognormal* (μ, σ^2); Uniform U (a, b). Priors for K were Uniform between the maximum catch in the time series and 100 times the maximum catch. For the Catch-MSY method, the depletion priors were Uniform centered on the true value with a minimum of *true - 0.1* and a maximum of *true + 0.1*.

Operating model inputs	**Symbol**	**Short-lived**	**Medium-lived**	**Long-lived**
Maximum age	Age_{max}	10	15	64
Age at 50% maturity (years)	A_{mat}	3	5	16
Length where 50% of the fish are mature (FL cm)	L_{50}	29	90	55
Length where 95% of the fish are mature (FL cm)	L_{95}	34	100	57
Length-weight scaling parameter	α	2.73×10^{-6}	1.34×10^{-5}	1.80×10^{-5}
Length-weight allometric parameter	β	3.444	3.107	3.094
Von Bertalanffy Brody growth coefficient (1/years)	k	0.40	0.21	0.14
Von Bertalanffy asymptotic length (cm)	L_∞	38.2	122.2	60.0
Theoretical age at length=0	t_0	-0.6	-1.3	-1.9
Coefficient of variation of length at age for all ages	CVL	0.1	0.1	0.1
Natural mortality (1/years)	M	0.60	0.30	0.05
Steepness	h	0.5	0.9	0.8
Selectivity at 50% (cm)	S_{50}	25	60	45
Selectivity at 95% (cm)	S_{95}	30	75	50
Recruitment deviations	σ_R	0.3	0.4	0.5
Fishing mortality deviations	σ_F	0.2	0.2	0.2
Estimation models prior distributions				
Depletion (used for all catch-based models)	XB_0	*Lognormal* (true, 0.1)	*Lognormal* (true, 0.1)	*Lognormal* (true, 0.1)
Carrying capacity (used for Catch-MSY, CMSY and SSCOM)	K	U (max(catch), max(catch)x100)	U (max(catch), max(catch)x100)	U (max(catch), max(catch)x100)
Population rate of increase (used for Catch-MSY, CMSY and SSCOM)	r	U (0.8, 1.2)	U (0.2, 0.6)	U (0.05, 0.4)
Steepness (used for SSS)	h	*Normal* (0.5,0.1)	*Normal* (0.9,0.1)	*Normal* (0.8,0.1)
Vulnerability (used for DBSRA)	F_{MSY}/M	U (0, 2)	U (0, 2)	U (0, 2)
Compensation (used for DBSRA)	B_{MSY}/B_0	U (0, 1)	U (0, 1)	U (0, 1)

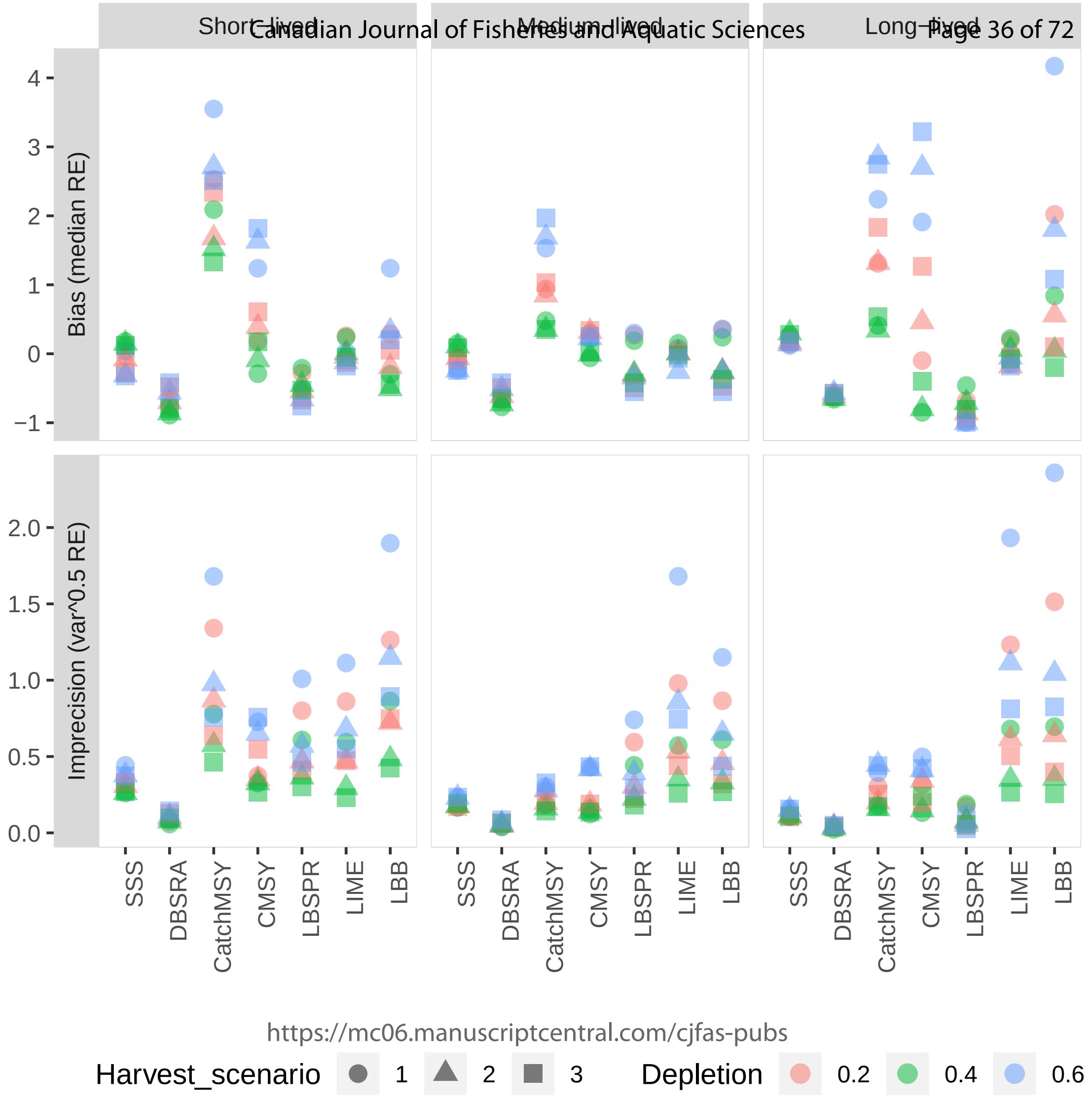
Short-lived
Medium-lived
Long-lived
Bias (median RE)
Imprecision (var^0.5 RE)
4
3
2
1
0
−1
2.0
1.5
1.0
0.5
0.0
SSS
DBSRA
CatchMSY
CMSY
LBSPR
LIME
LBB
Harvest_scenario
1
2
3
Depletion
0.2
0.4
0.6

Table 2 Bias in performance for the catch-based models. The values in between brackets are the

Scenarios	Life-history	Harvest trend	Final depletion	SSS (RE = 0.17)	DBSRA (RE = 0.63)
1	Short-lived (RE = 1.15)	Scenario 1 (RE = 1.53)	0.2	**-0.06**	-0.75
2			0.4	**0.17**	-0.88
3			0.6	**0.04**	-0.66
4		Scenario 2 (RE = 0.97)	0.2	**-0.09**	-0.68
5			0.4	0.14	-0.84
6			0.6	**-0.27**	-0.57
7		Scenario 3 (RE = 0.95)	0.2	**-0.25**	-0.48
8			0.4	**0.11**	-0.77
9			0.6	**-0.29**	-0.42
10	Medium-lived (RE = 0.50)	Scenario 1 (RE = 0.57)	0.2	**-0.06**	-0.67
11			0.4	0.15	-0.77
12			0.6	-0.24	-0.59
13		Scenario 2 (RE = 0.46)	0.2	**-0.08**	-0.60
14			0.4	0.13	-0.72
15			0.6	-0.24	-0.51
16		Scenario 3 (RE = 0.47)	0.2	**-0.09**	-0.49
17			0.4	0.11	-0.65
18			0.6	-0.24	-0.41
19	Long-lived (RE = 1.33)	Scenario 1 (RE = 1.10)	0.2	**0.13**	-0.63
20			0.4	0.28	-0.66
21			0.6	**0.09**	-0.60
22		Scenario 2 (RE = 1.44)	0.2	**0.15**	-0.60
23			0.4	0.30	-0.65
24			0.6	**0.11**	-0.57
25		Scenario 3 (RE = 1.43)	0.2	**0.17**	-0.58
26			0.4	0.28	-0.62
27			0.6	**0.13**	-0.57

: mean absolute relative error per factor: harvest rates scenarios, species and final de

Catch-MSY (RE = 1.62)	CMSY (RE = 0.71)	Absolute mean RE
2.57	0.17	1.36
2.07	-0.33	1.05
3.69	1.15	2.19
1.75	0.36	0.79
1.52	**-0.10**	0.60
2.85	1.59	1.52
2.37	0.53	0.95
1.32	0.15	0.52
2.61	1.82	1.37
0.95	0.28	0.47
0.48	**-0.05**	0.38
1.54	**0.16**	0.86
0.84	0.31	0.41
0.35	**-0.01**	0.36
1.70	**0.18**	0.62
1.04	0.32	0.45
0.36	**0.03**	0.35
1.97	**0.19**	0.60
1.34	-0.12	0.72
0.41	-0.78	0.48
2.31	1.87	2.11
1.33	0.46	0.85
0.34	-0.74	0.46
2.87	2.72	3.03
1.84	1.27	1.60
0.55	-0.40	0.38
2.76	3.20	2.31

Table 2 Bias in performance for the length-based methods. The values in between brackets are t

Scenarios	Life-history	Harvest trend	Final depletion	LBSPR (RE = 0.51)	LIME (RE = 0.15)
1	Short-lived (RE = 0.80)	Scenario 1 (RE = 0.39)	0.2	-0.42	**0.25**
2			0.4	-0.69	**0.05**
3			0.6	-0.80	**-0.06**
4		Scenario 2 (RE = 0.69)	0.2	-0.66	**0.34**
5			0.4	-0.85	**-0.13**
6			0.6	-0.92	**-0.06**
7		Scenario 3 (RE = 1.29)	0.2	-0.92	**0.55**
8			0.4	-0.97	**0.15**
9			0.6	-0.98	**-0.05**
10	Medium-lived (RE = 0.31)	Scenario 1 (RE = 0.23)	0.2	-0.10	0.28
11			0.4	-0.40	**0.02**
12			0.6	-0.48	**-0.05**
13		Scenario 2 (RE = 0.26)	0.2	**-0.09**	0.26
14			0.4	-0.44	-0.02
15			0.6	-0.57	**-0.11**
16		Scenario 3 (RE = 0.46)	0.2	**-0.14**	0.20
17			0.4	-0.50	**-0.09**
18			0.6	-0.62	**-0.14**
19	Long-lived (RE = 0.28)	Scenario 1 (RE = 0.23)	0.2	0.27	**0.20**
20			0.4	-0.26	**0.01**
21			0.6	-0.40	**-0.04**
22		Scenario 2 (RE = 0.27)	0.2	0.38	**0.22**
23			0.4	-0.25	**0.01**
24			0.6	-0.45	**-0.04**
25		Scenario 3 (RE = 0.32)	0.2	0.39	0.52
26			0.4	-0.28	**-0.01**
27			0.6	-0.51	**0.01**

he mean absolute relativ

LBB (RE = 0.73)	Absolute mean RE
1.00	0.56
0.10	0.28
-0.16	0.34
2.41	1.14
0.68	0.56
0.18	0.39
4.71	2.06
2.04	1.05
1.26	0.76
0.01	0.13
-0.36	0.26
-0.37	0.30
0.64	0.33
-0.01	0.16
0.16	0.28
1.64	0.66
0.54	0.38
0.31	0.36
0.33	0.27
-0.23	0.17
-0.35	0.26
0.54	0.38
-0.19	0.15
-0.41	0.30
0.60	0.50
-0.10	0.13
-0.41	0.31

Appendix

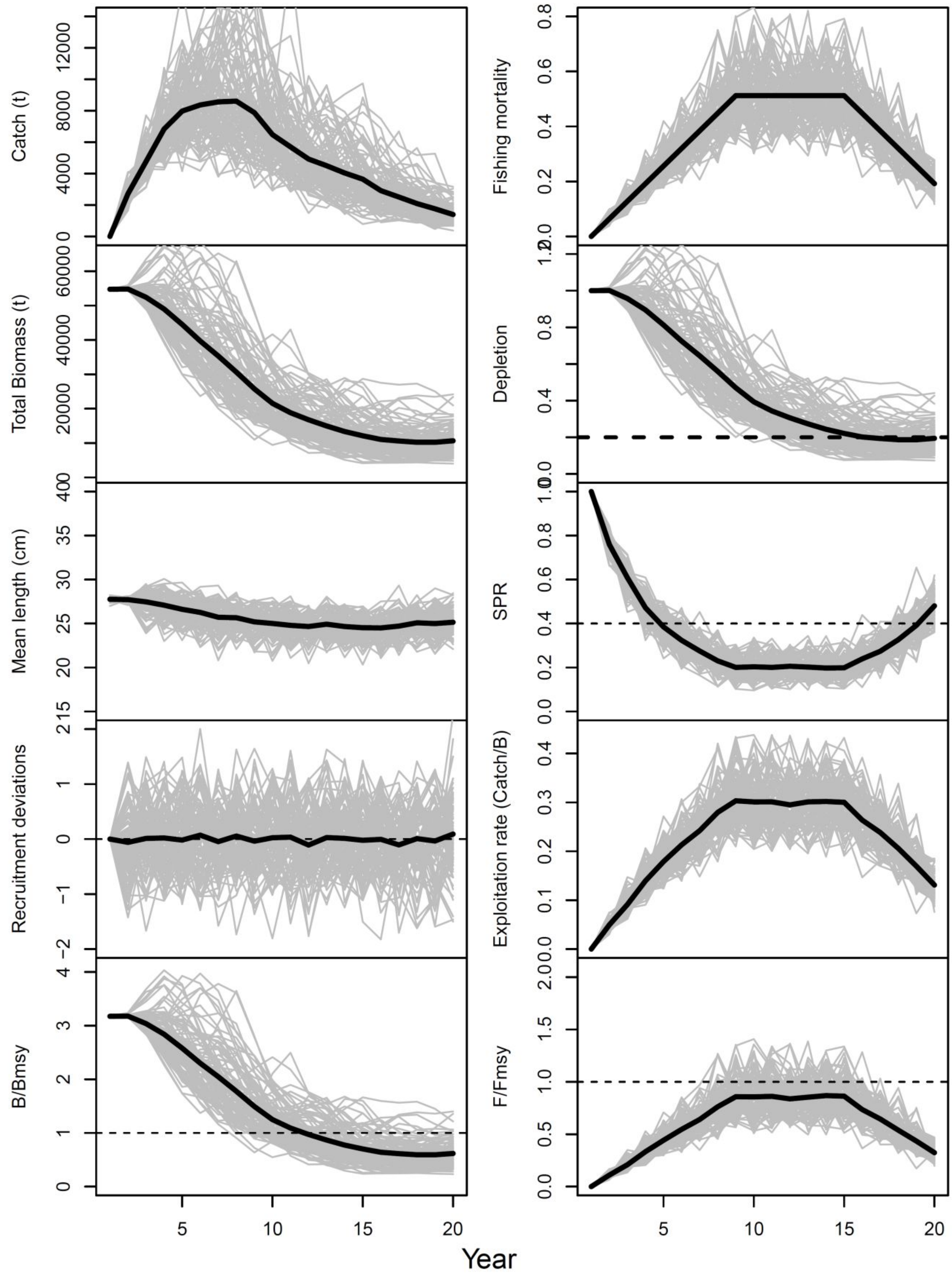

Figure A1. Time series for each simulated fast-grow chub mackerel population for harvest scenario 1 and depletion = 0.2. The black solid lines represent the mean value for all runs.

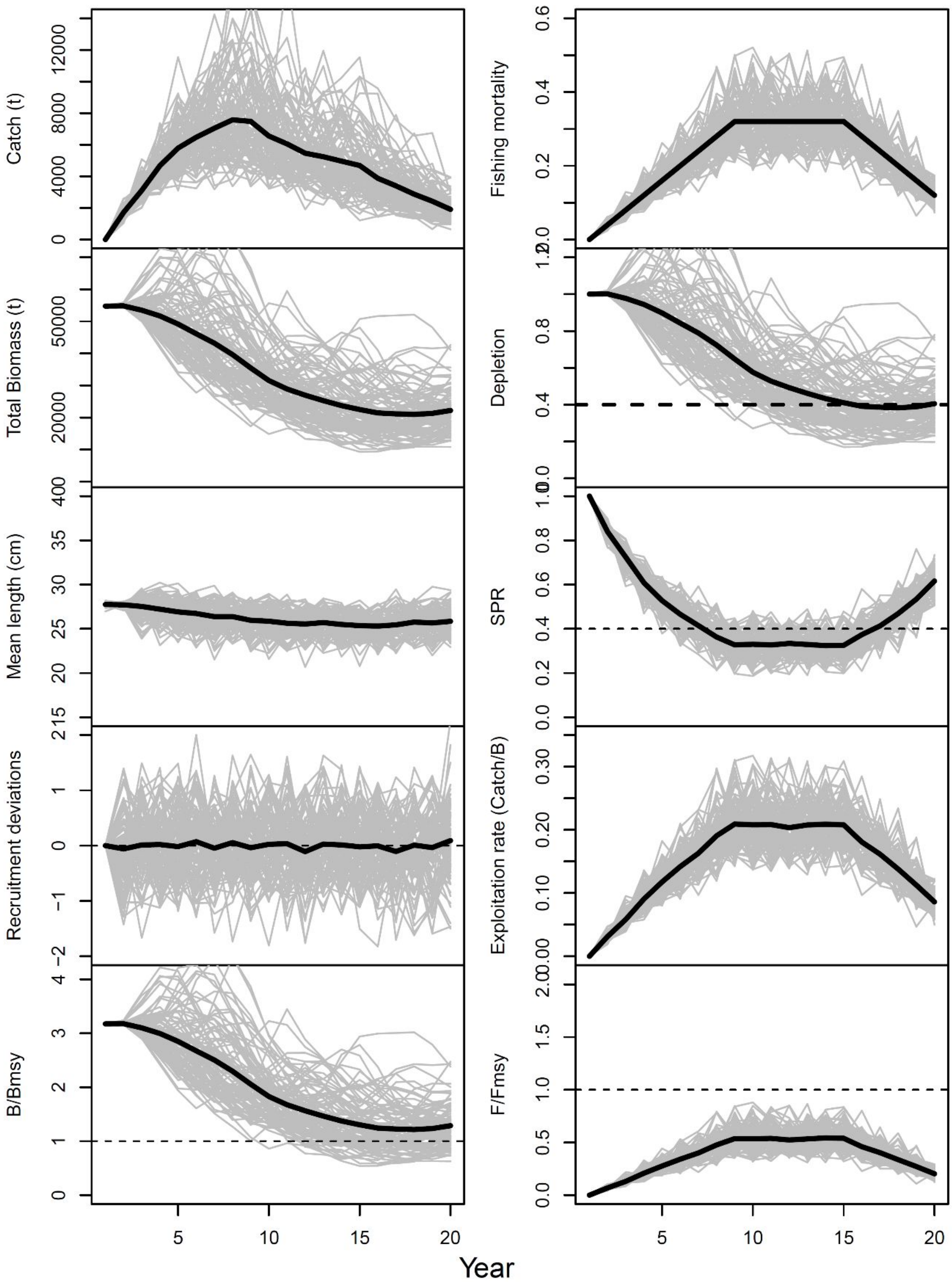

Figure A2. Time series for each simulated fast-grow chub mackerel population for harvest scenario 1 and depletion = 0.4. The black solid lines represent the mean value for all runs.

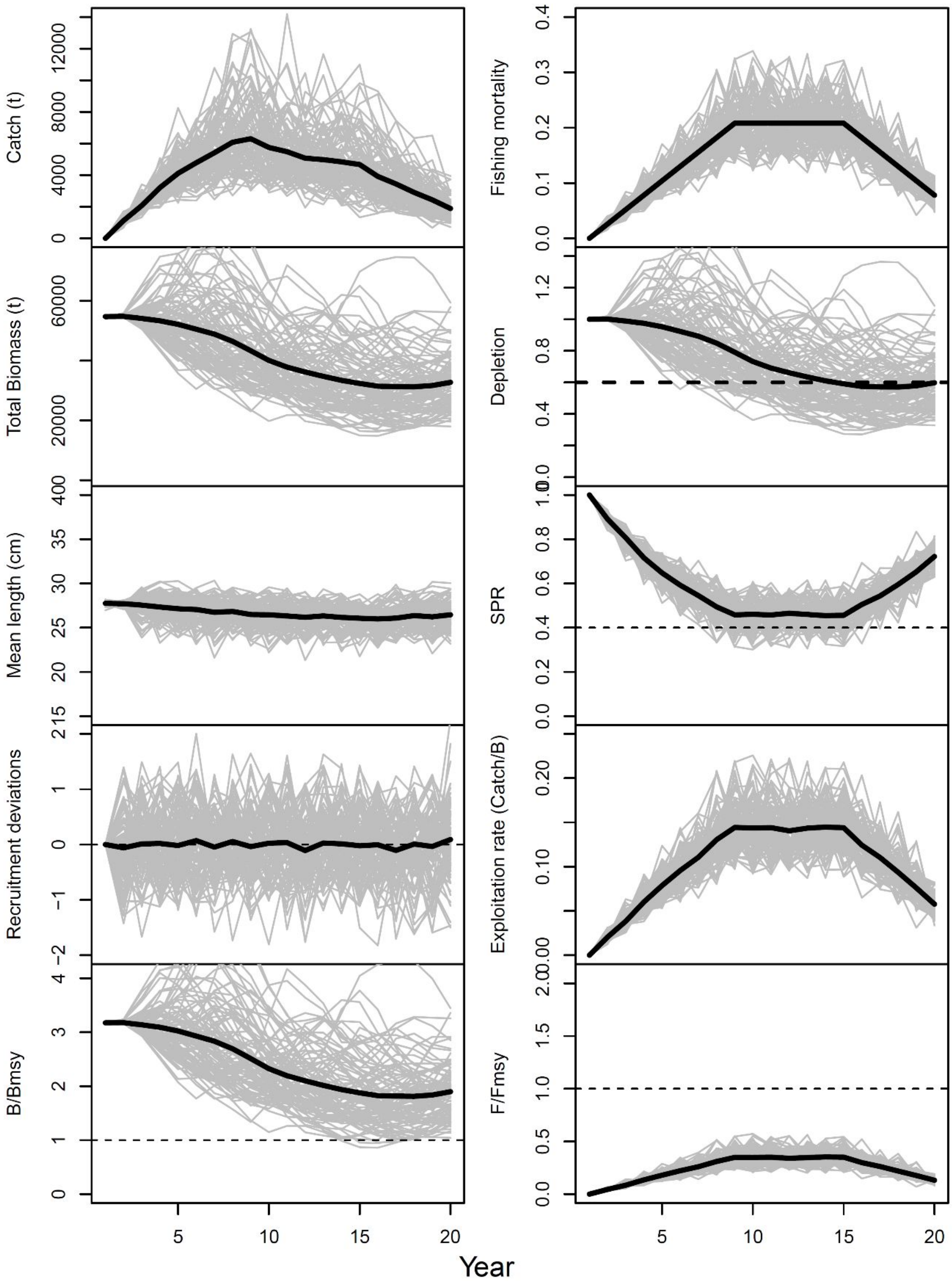

Figure A3. Time series for each simulated fast-grow chub mackerel population for harvest scenario 1 and depletion = 0.6. The black solid lines represent the mean value for all runs.

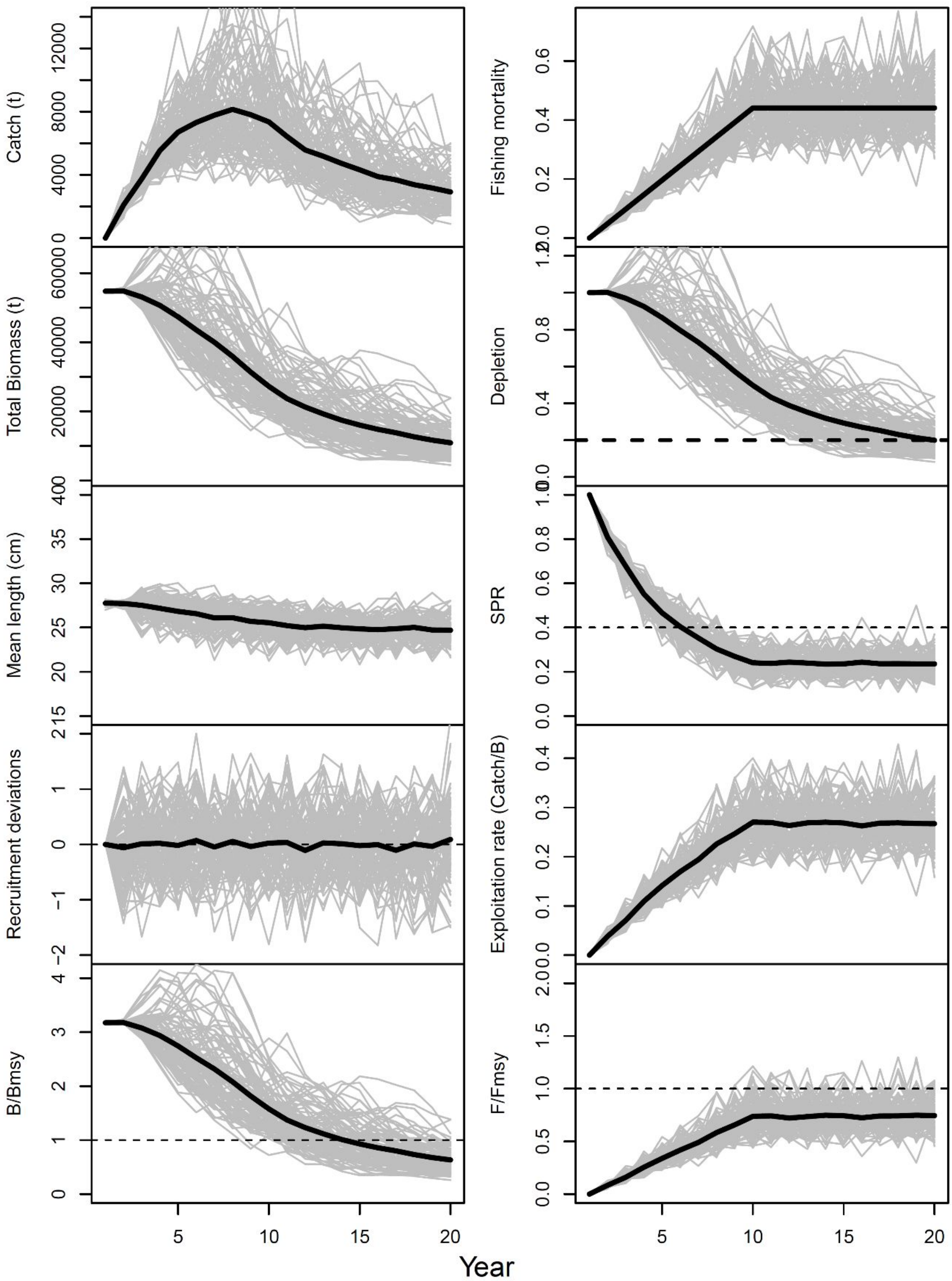

Figure A4. Time series for each simulated fast-grow chub mackerel population for harvest scenario 2 and depletion = 0.2. The black solid lines represent the mean value for all runs.

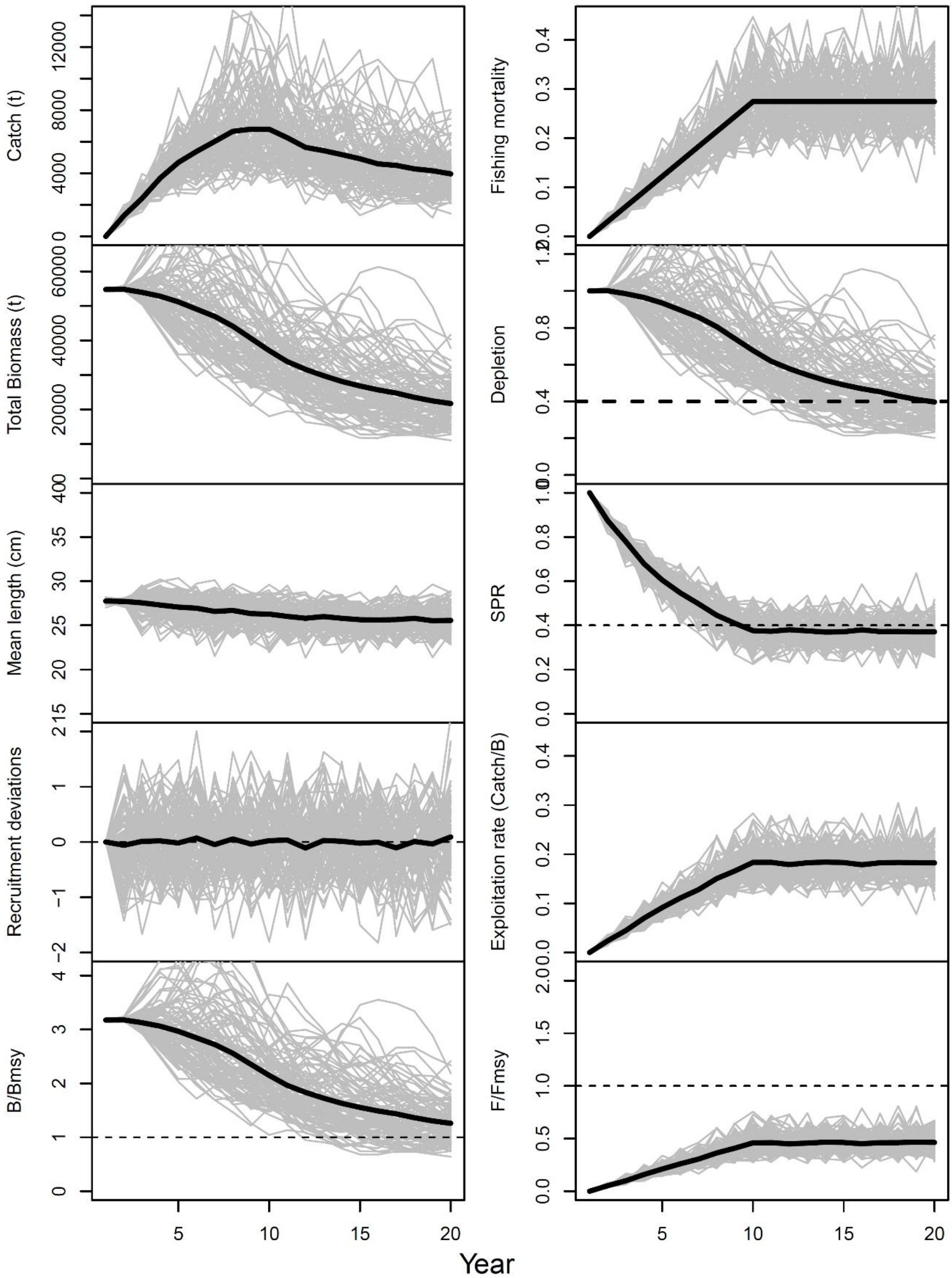

Figure A5. Time series for each simulated fast-grow chub mackerel population for harvest scenario 2 and depletion = 0.4. The black solid lines represent the mean value for all runs.

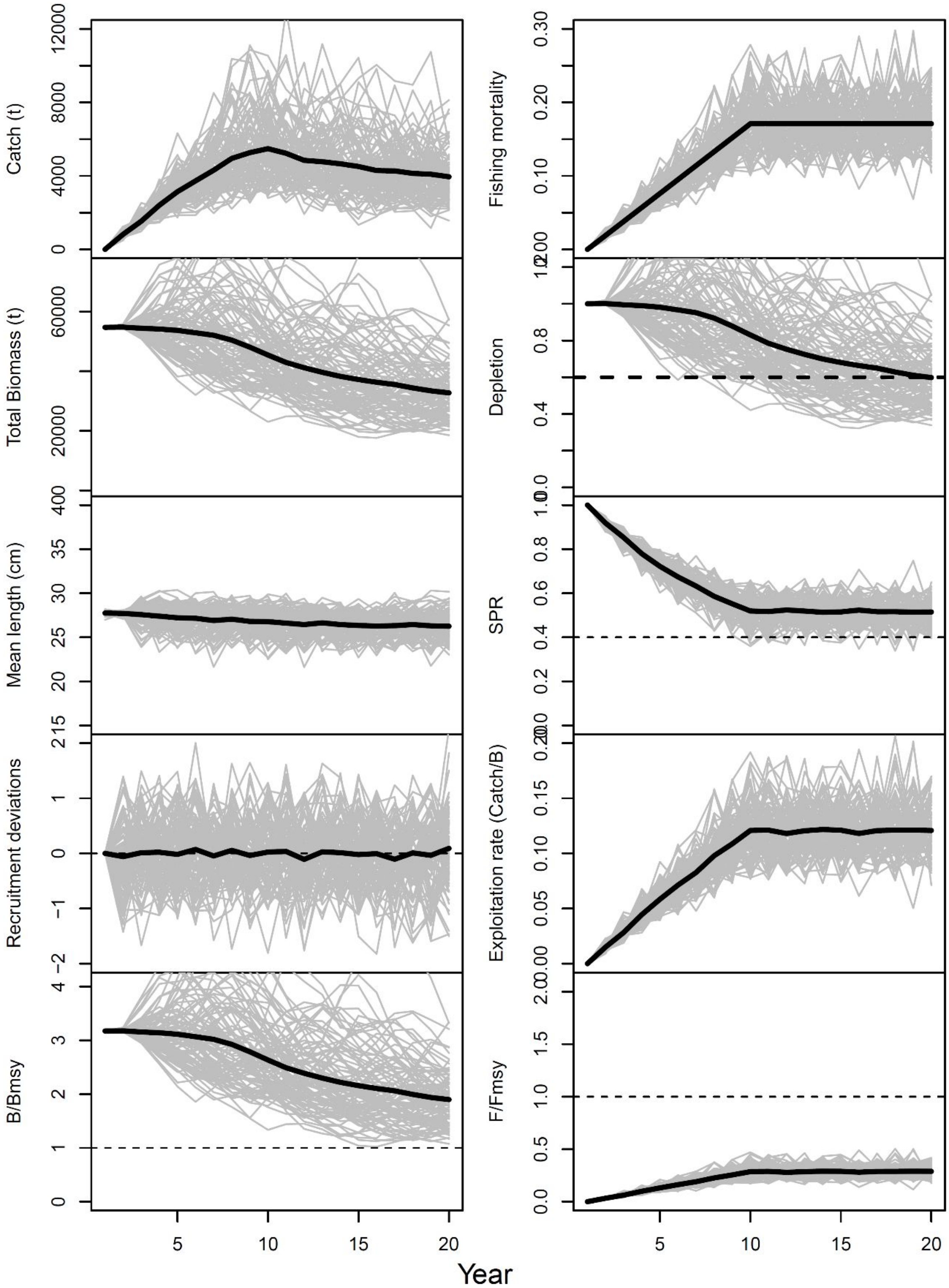

Figure A6. Time series for each simulated fast-grow chub mackerel population for harvest scenario 2 and depletion = 0.6. The black solid lines represent the mean value for all runs.

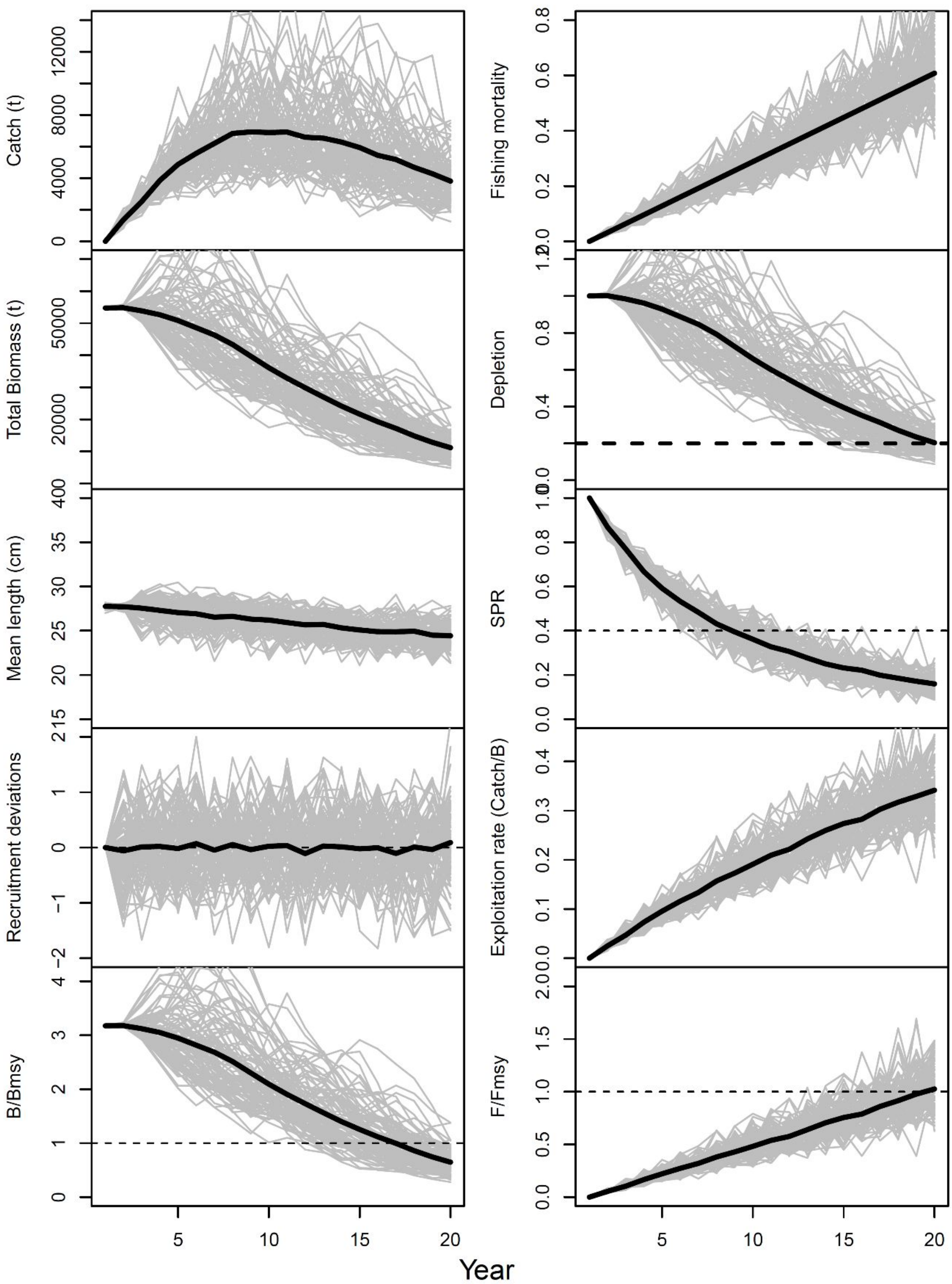

Figure A7. Time series for each simulated fast-grow chub mackerel population for harvest scenario 3 and depletion = 0.2. The black solid lines represent the mean value for all runs.

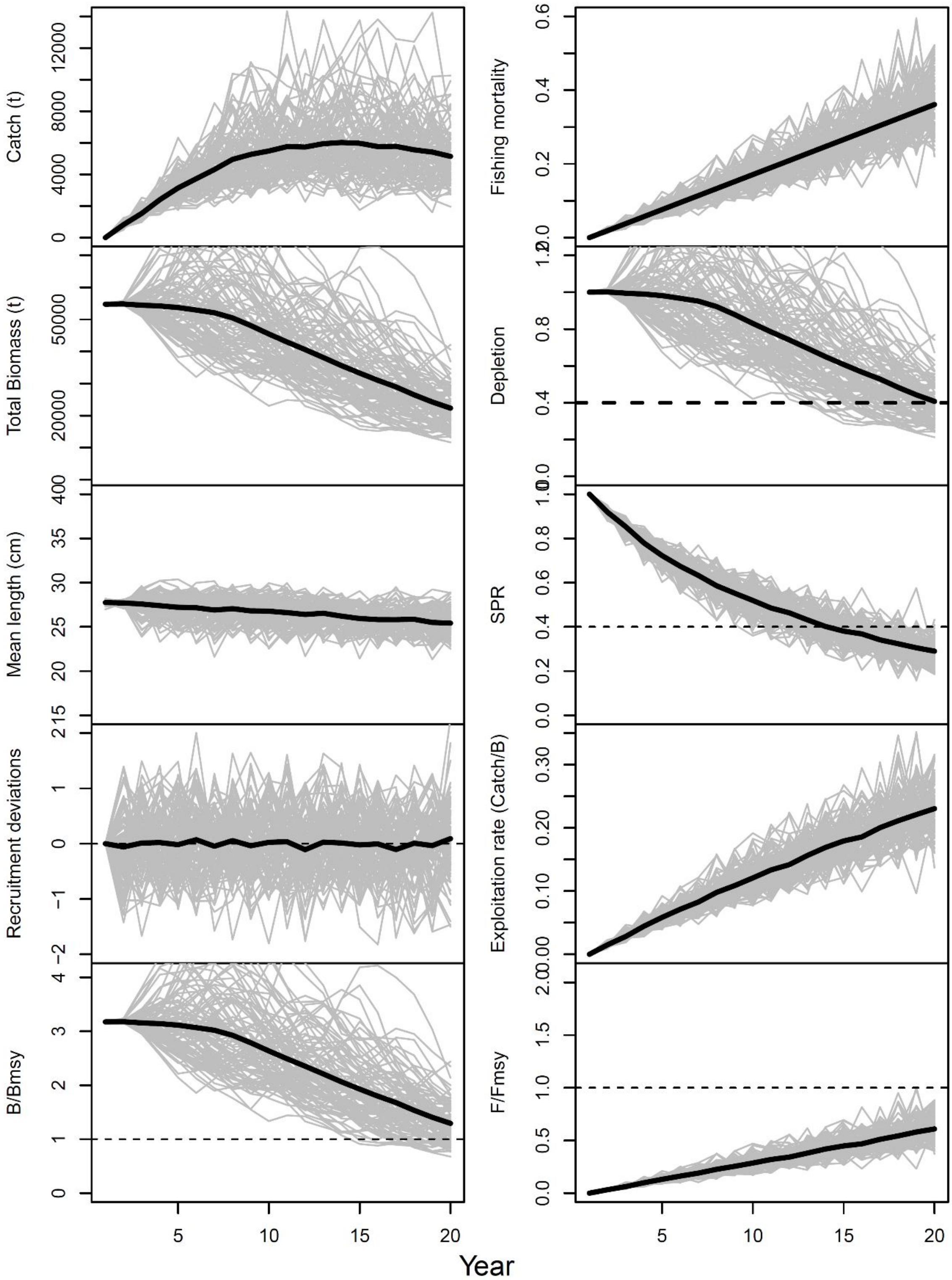

Figure A8. Time series for each simulated fast-grow chub mackerel population for harvest scenario 3 and depletion = 0.4. The black solid lines represent the mean value for all runs.

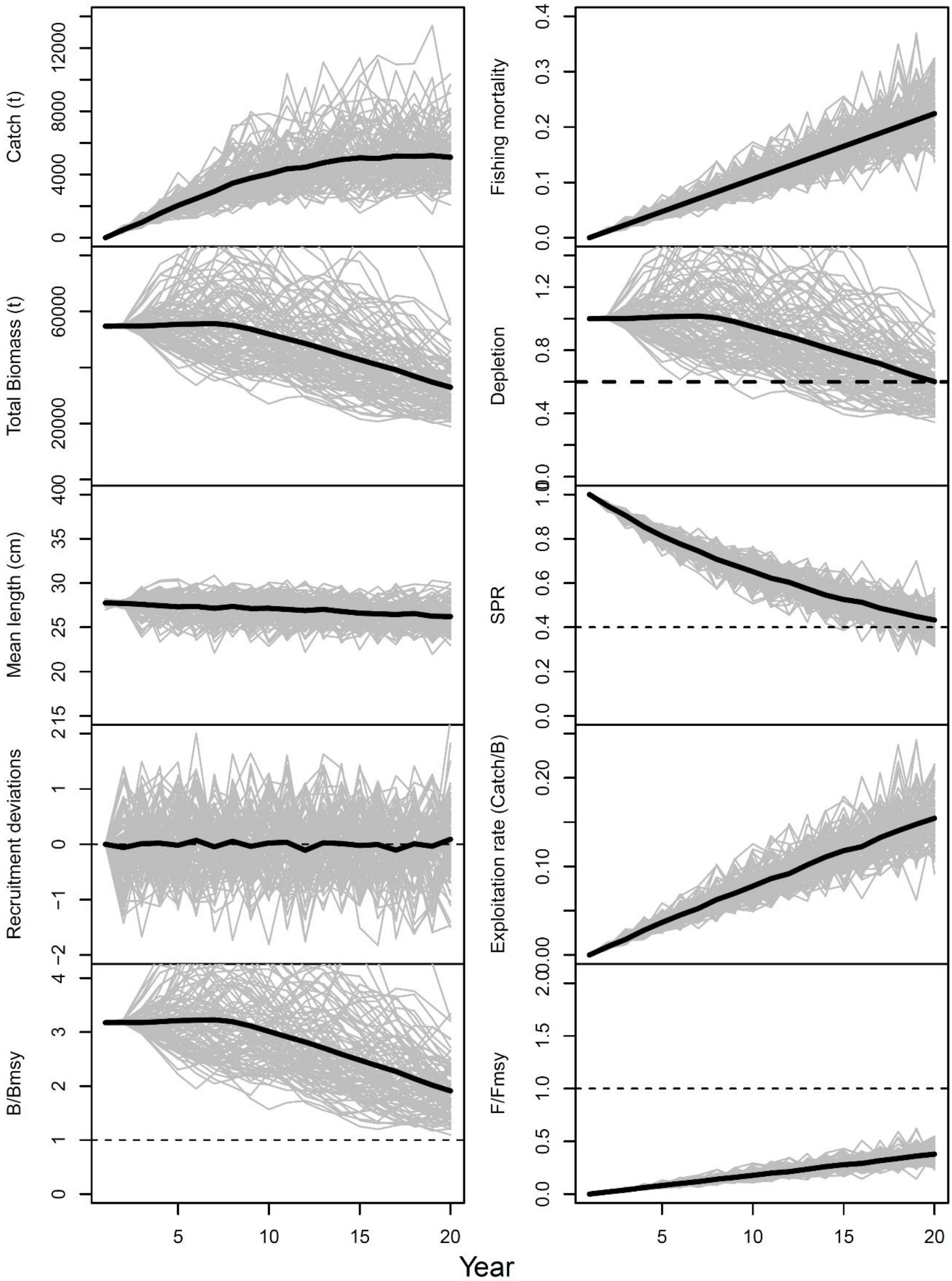

Figure A9. Time series for each simulated fast-grow chub mackerel population for harvest scenario 3 and depletion = 0.6. The black solid lines represent the mean value for all runs.

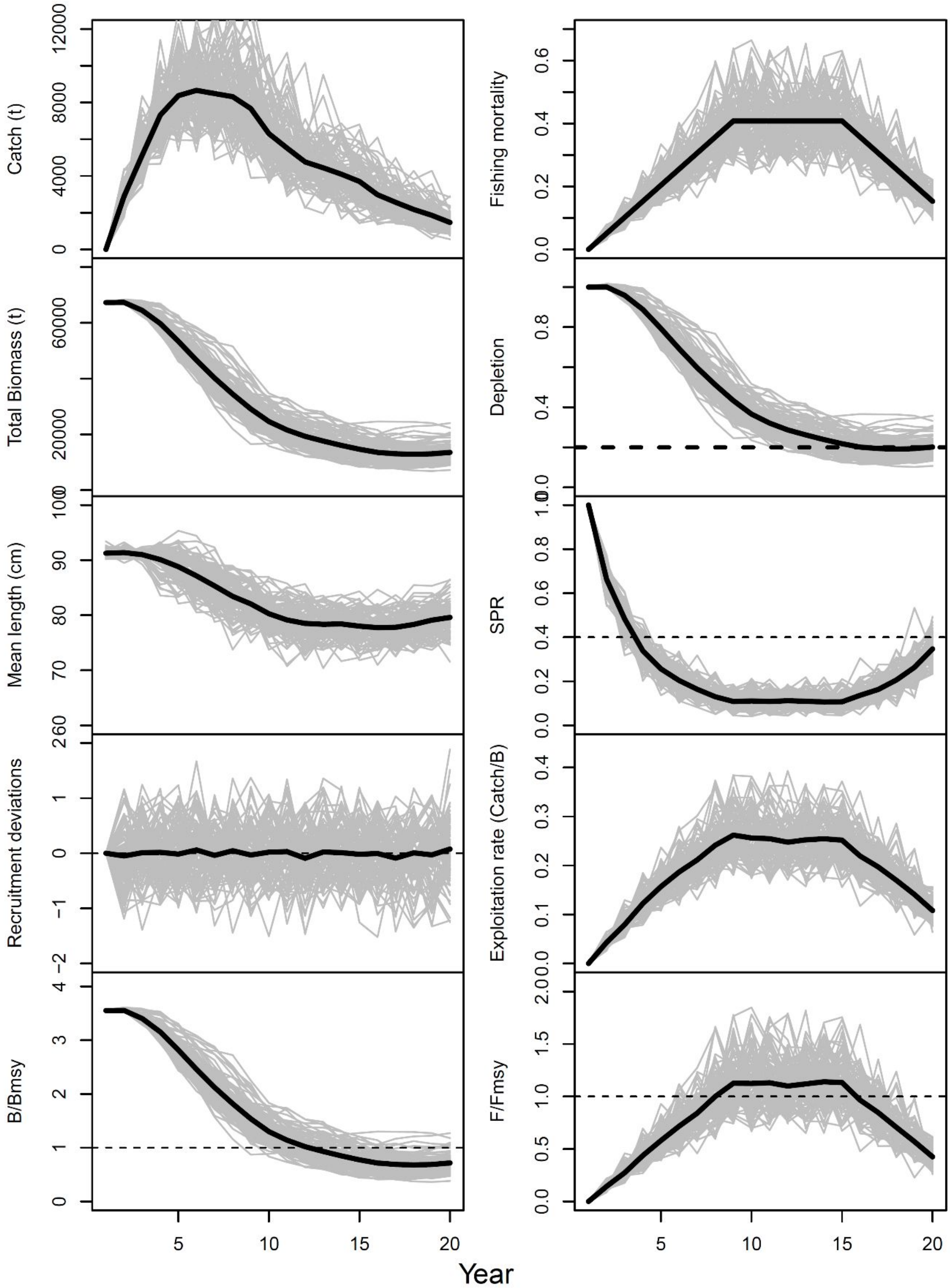

Figure A10. Time series for each simulated medium-grow albacore tuna population for harvest scenario 1 and depletion = 0.2. The black solid lines represent the mean value for all runs.

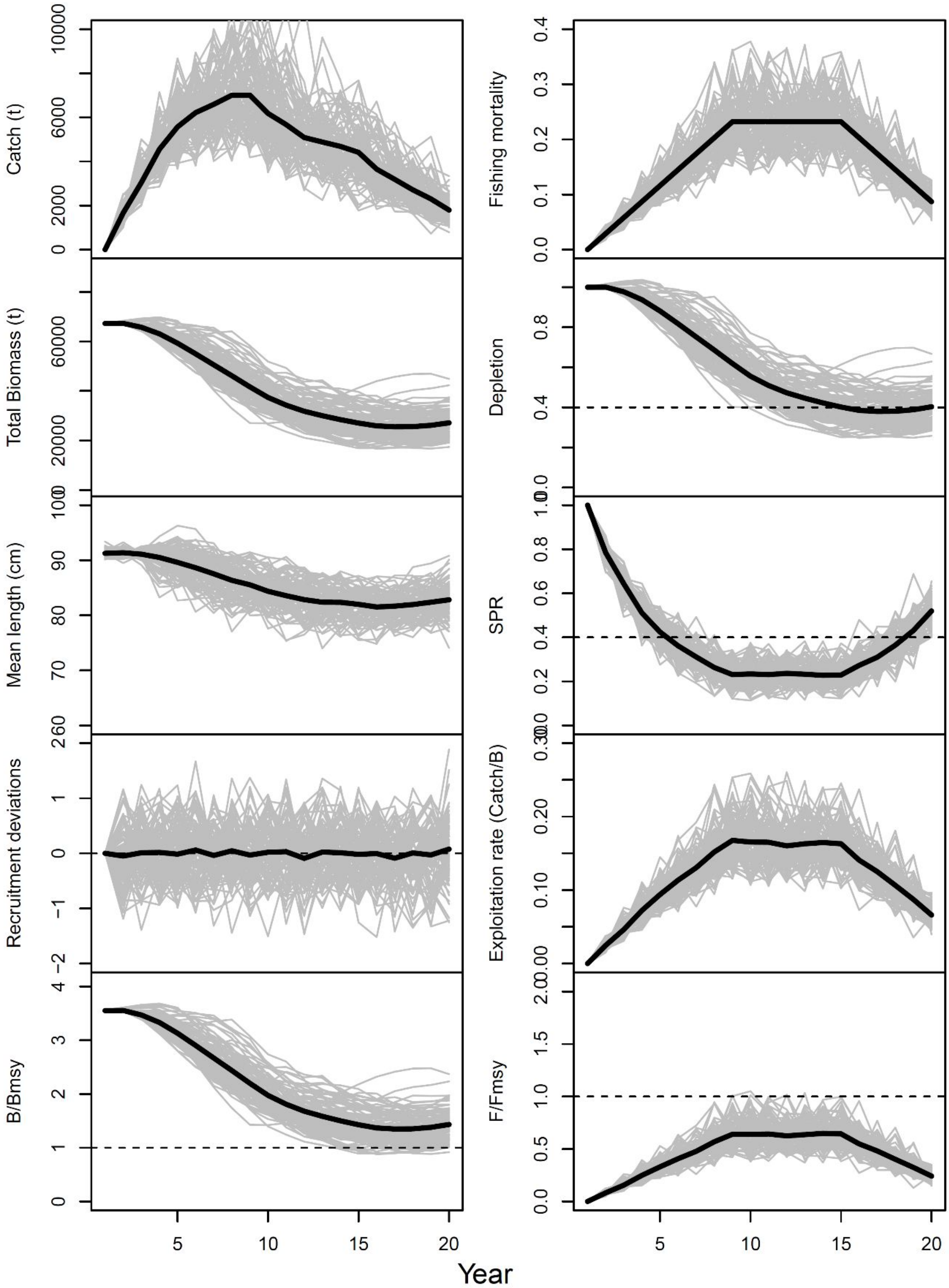

Figure A11. Time series for each simulated medium-grow albacore tuna population for harvest scenario 1 and depletion = 0.4. The black solid lines represent the mean value for all runs.

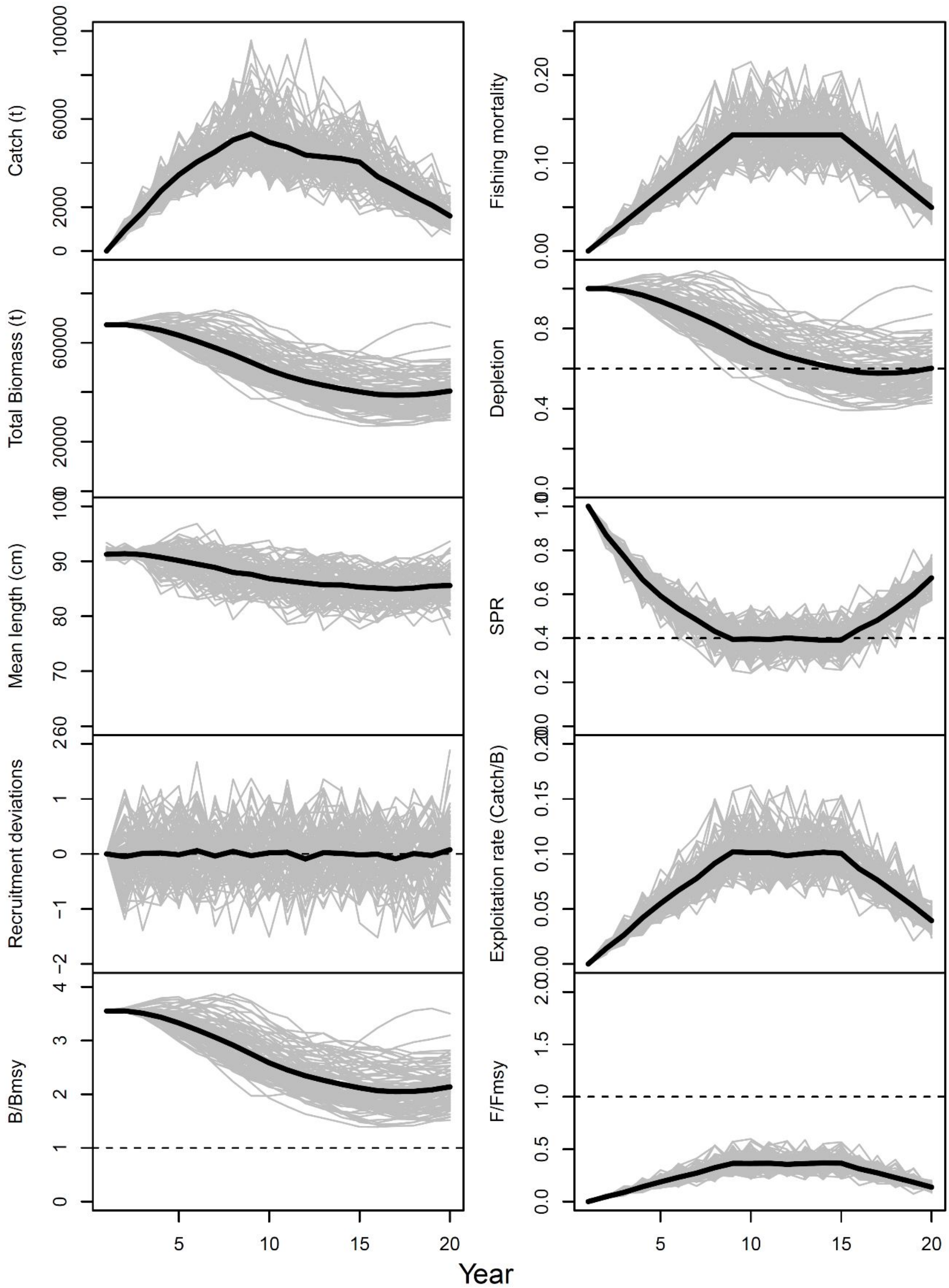

Figure A12. Time series for each simulated medium-grow albacore tuna population for harvest scenario 1 and depletion = 0.6. The black solid lines represent the mean value for all runs.

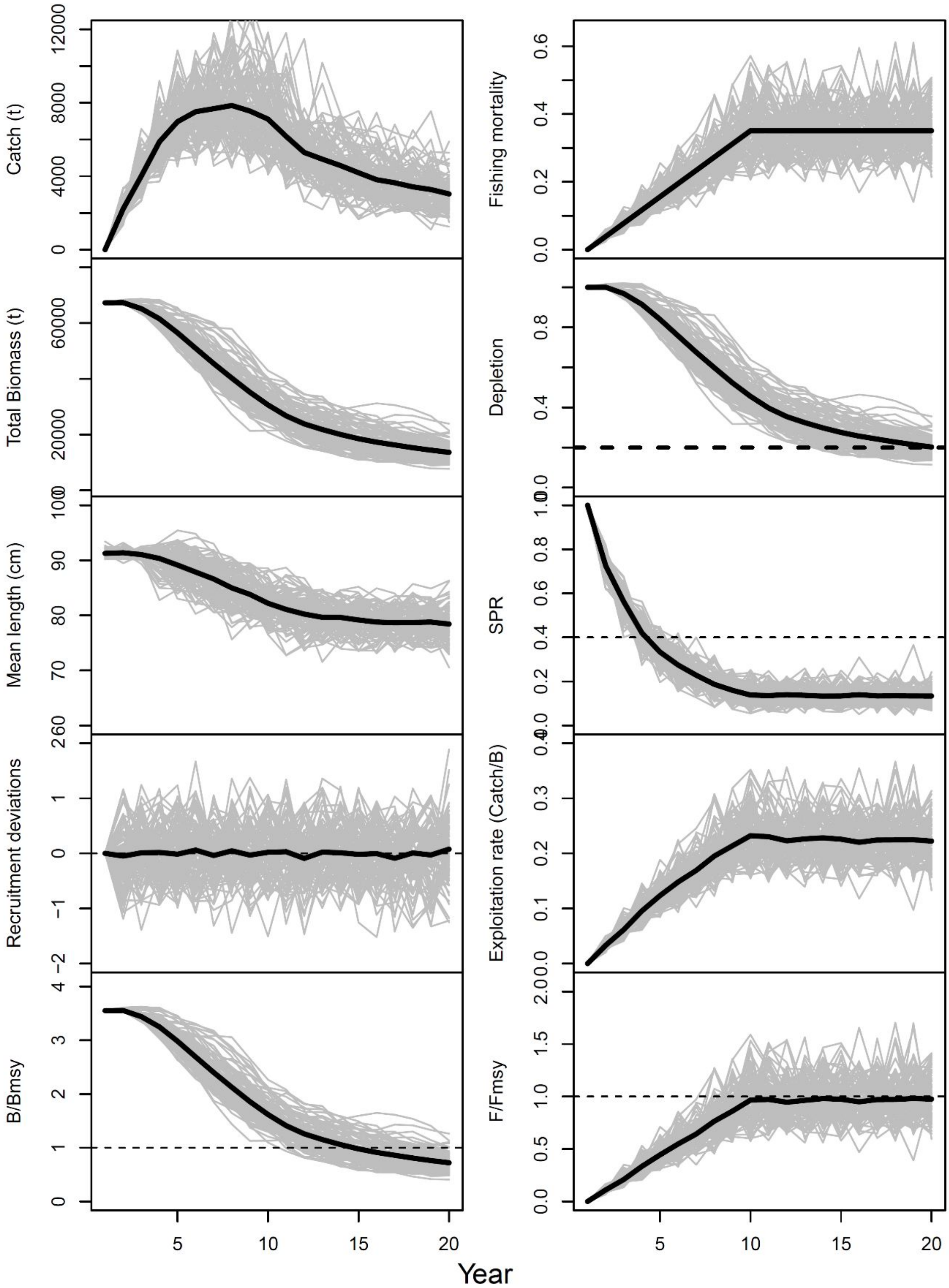

Figure A13. Time series for each simulated medium-grow albacore tuna population for harvest scenario 2 and depletion = 0.2. The black solid lines represent the mean value for all runs.

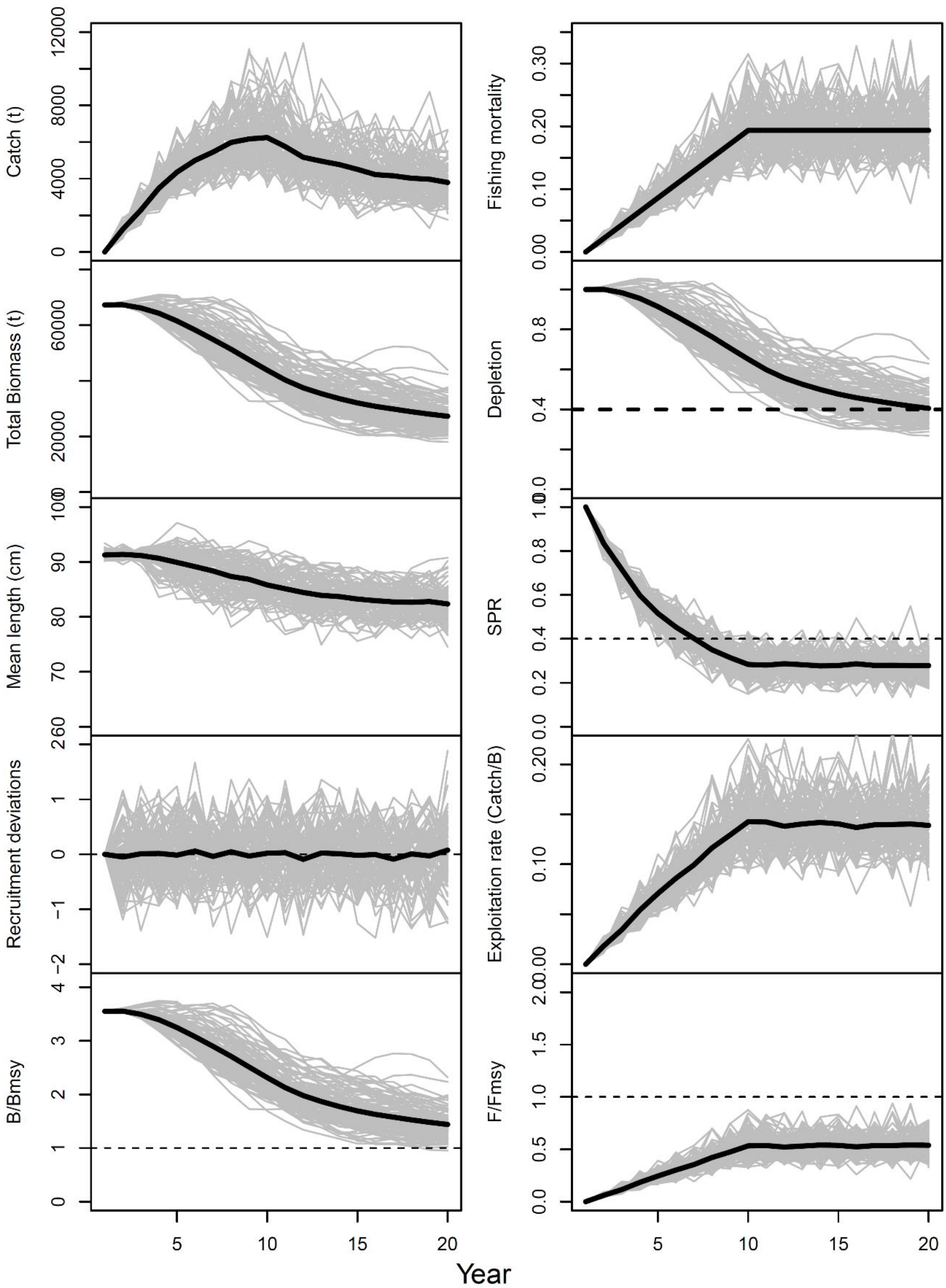

Figure A14. Time series for each simulated medium-grow albacore tuna population for harvest scenario 2 and depletion = 0.4. The black solid lines represent the mean value for all runs.

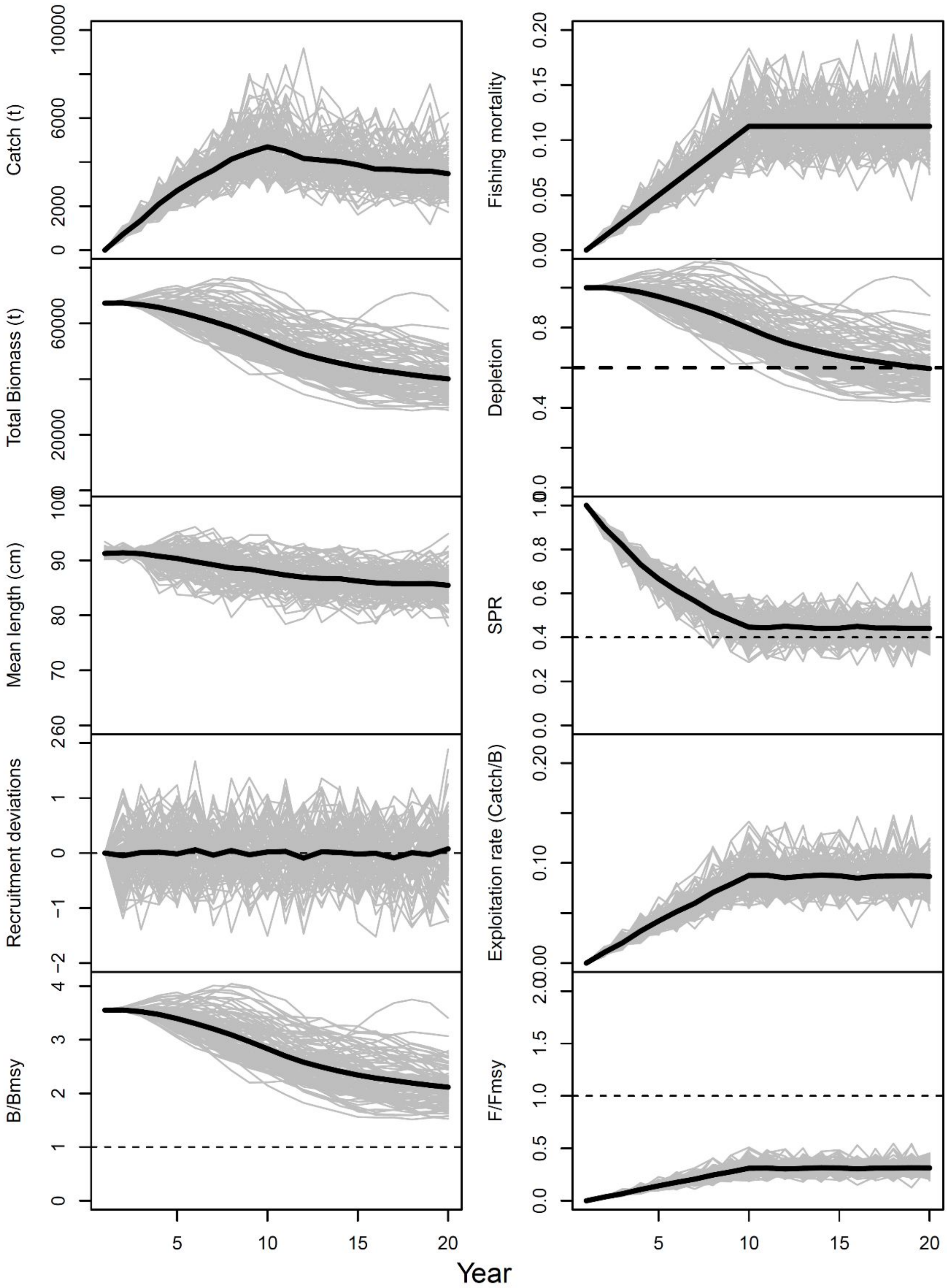

Figure A15. Time series for each simulated medium-grow albacore tuna population for harvest scenario 2 and depletion = 0.6. The black solid lines represent the mean value for all runs.

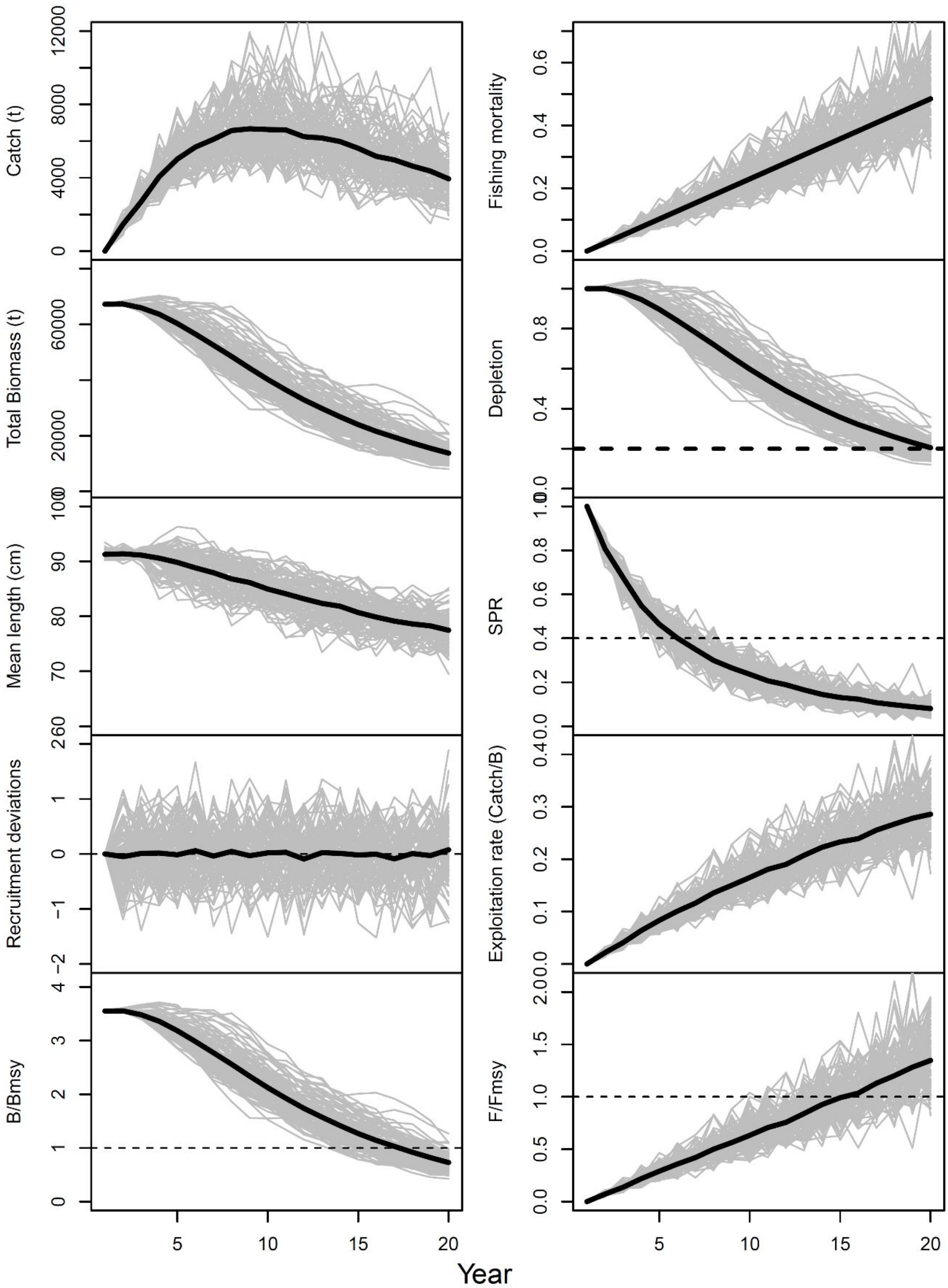

Figure A16. Time series for each simulated medium-grow albacore tuna population for harvest scenario 3 and depletion = 0.2. The black solid lines represent the mean value for all runs.

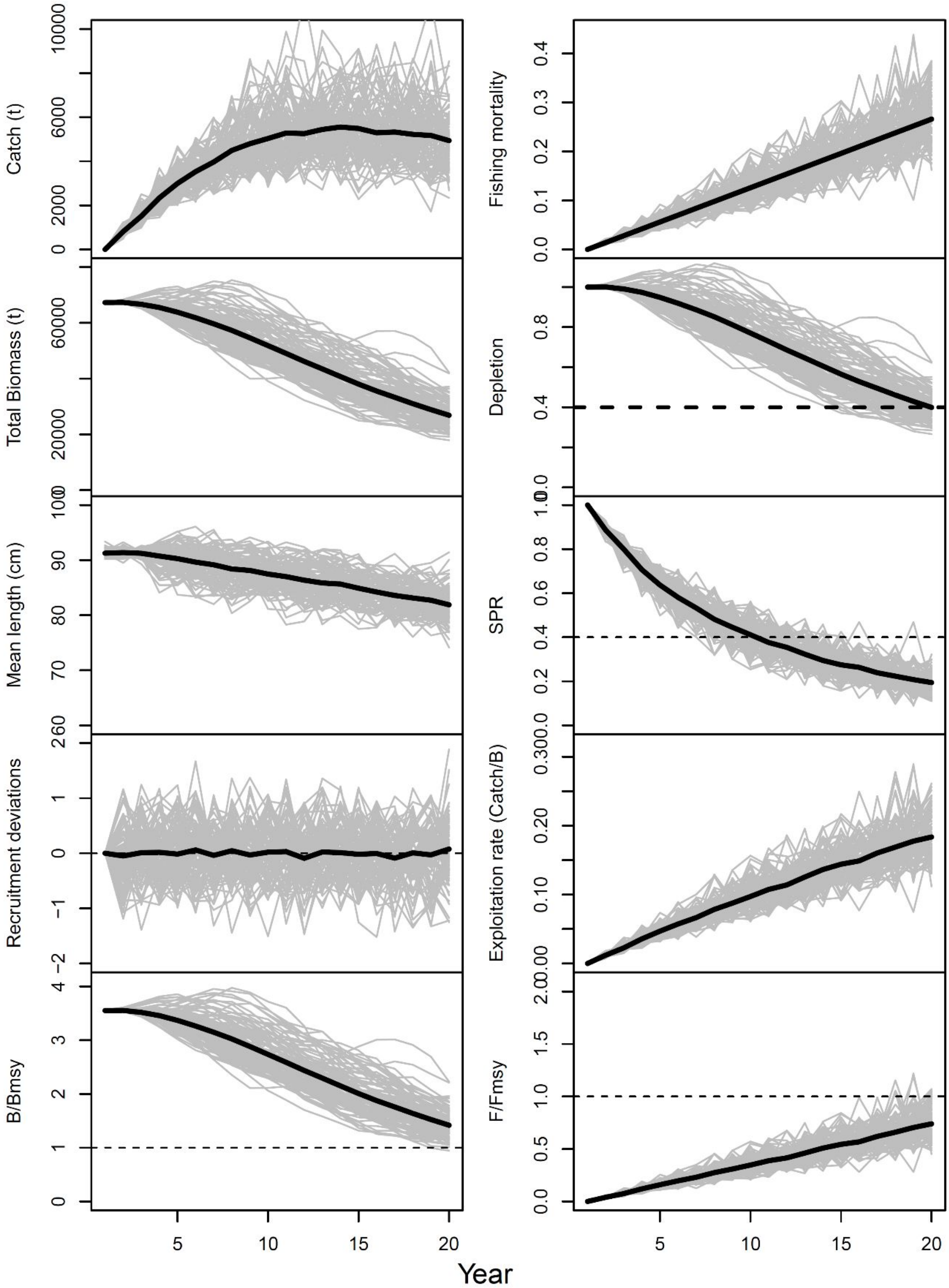

Figure A17. Time series for each simulated medium-grow albacore tuna population for harvest scenario 3 and depletion = 0.4. The black solid lines represent the mean value for all runs.

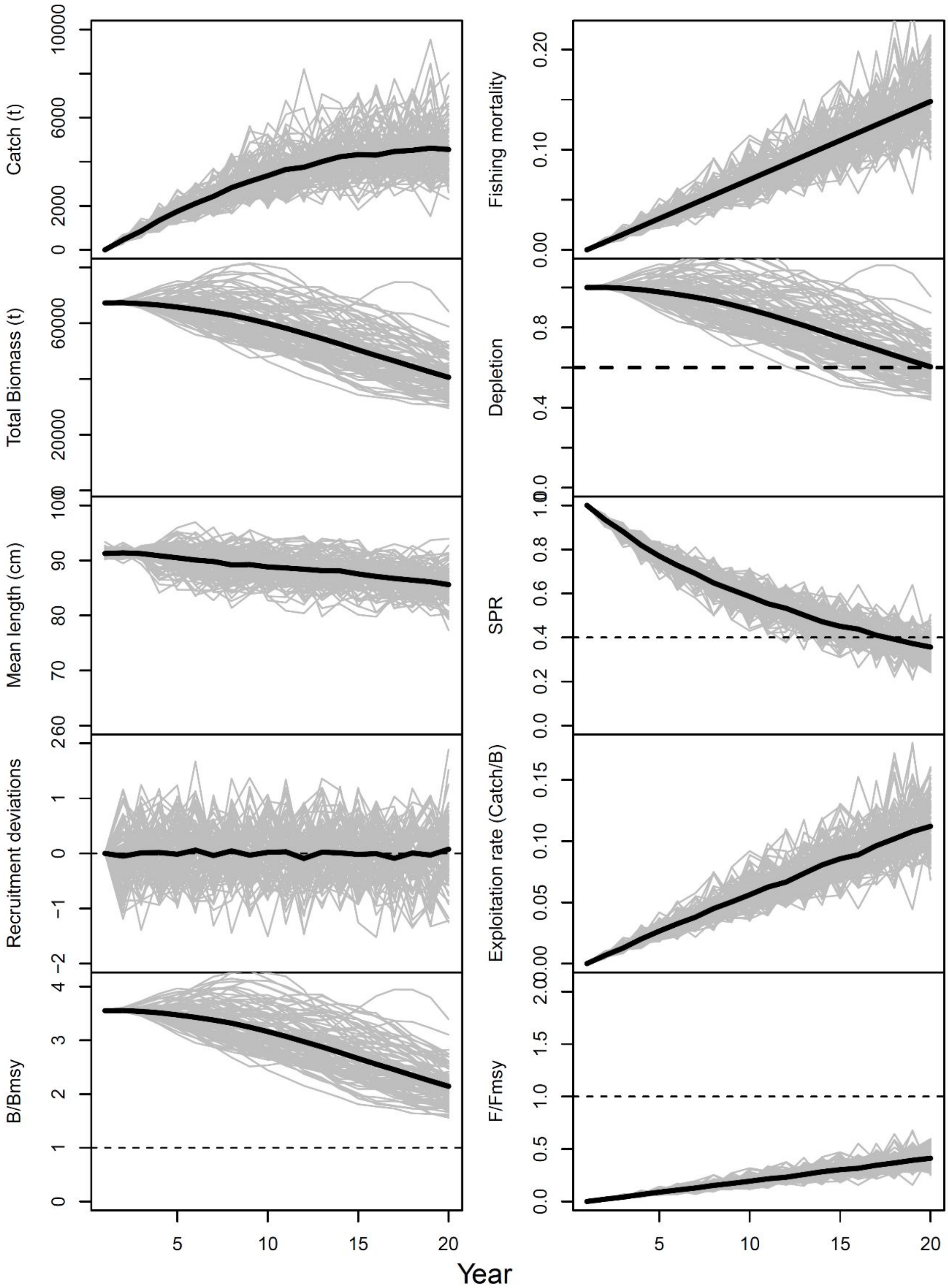

Figure A18. Time series for each simulated medium-grow albacore tuna population for harvest scenario 3 and depletion = 0.6. The black solid lines represent the mean value for all runs.

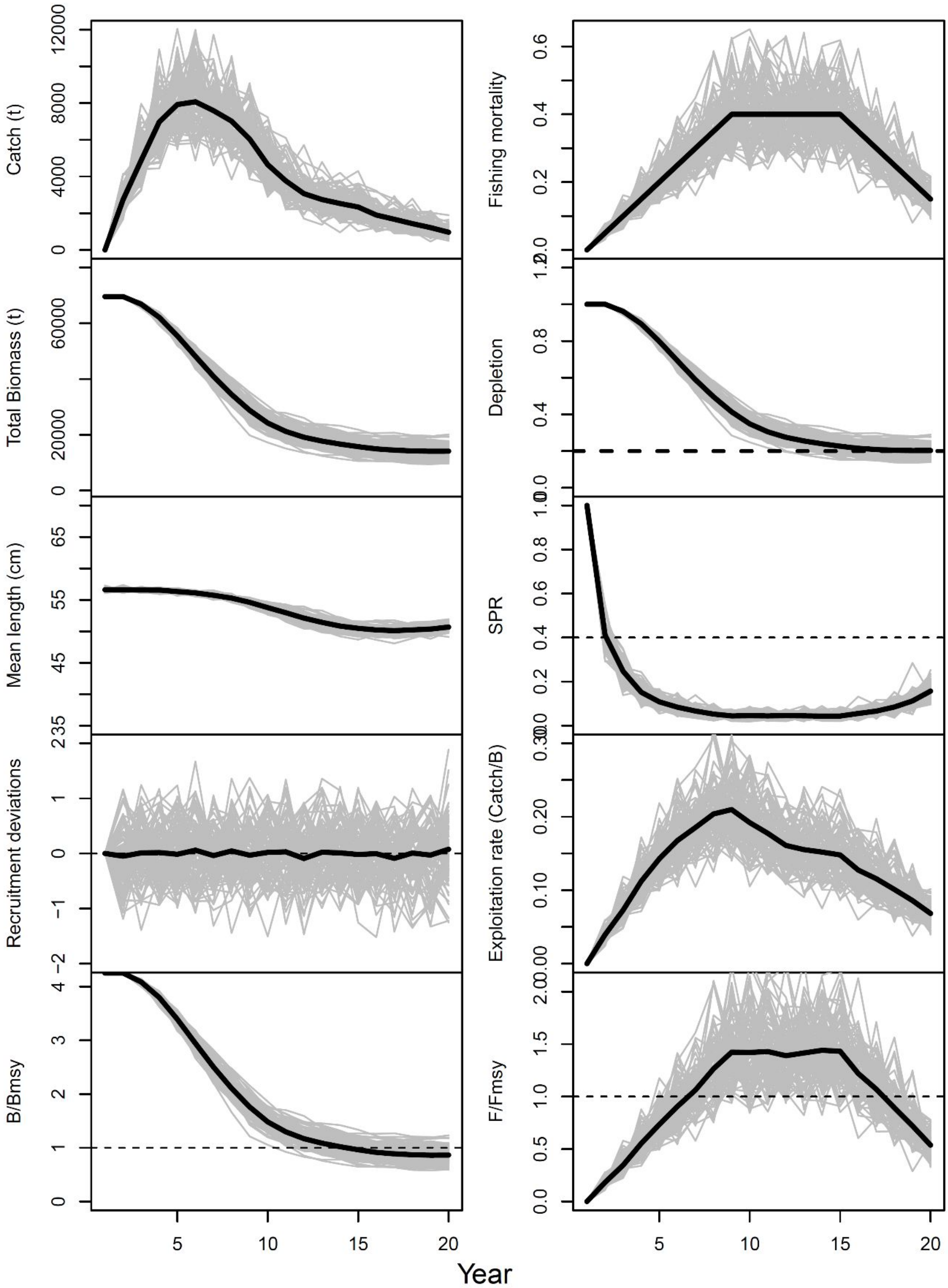

Figure A19. Time series for each simulated slow-grow canary rockfish population for harvest scenario 1 and depletion = 0.2. The black solid lines represent the mean value for all runs.

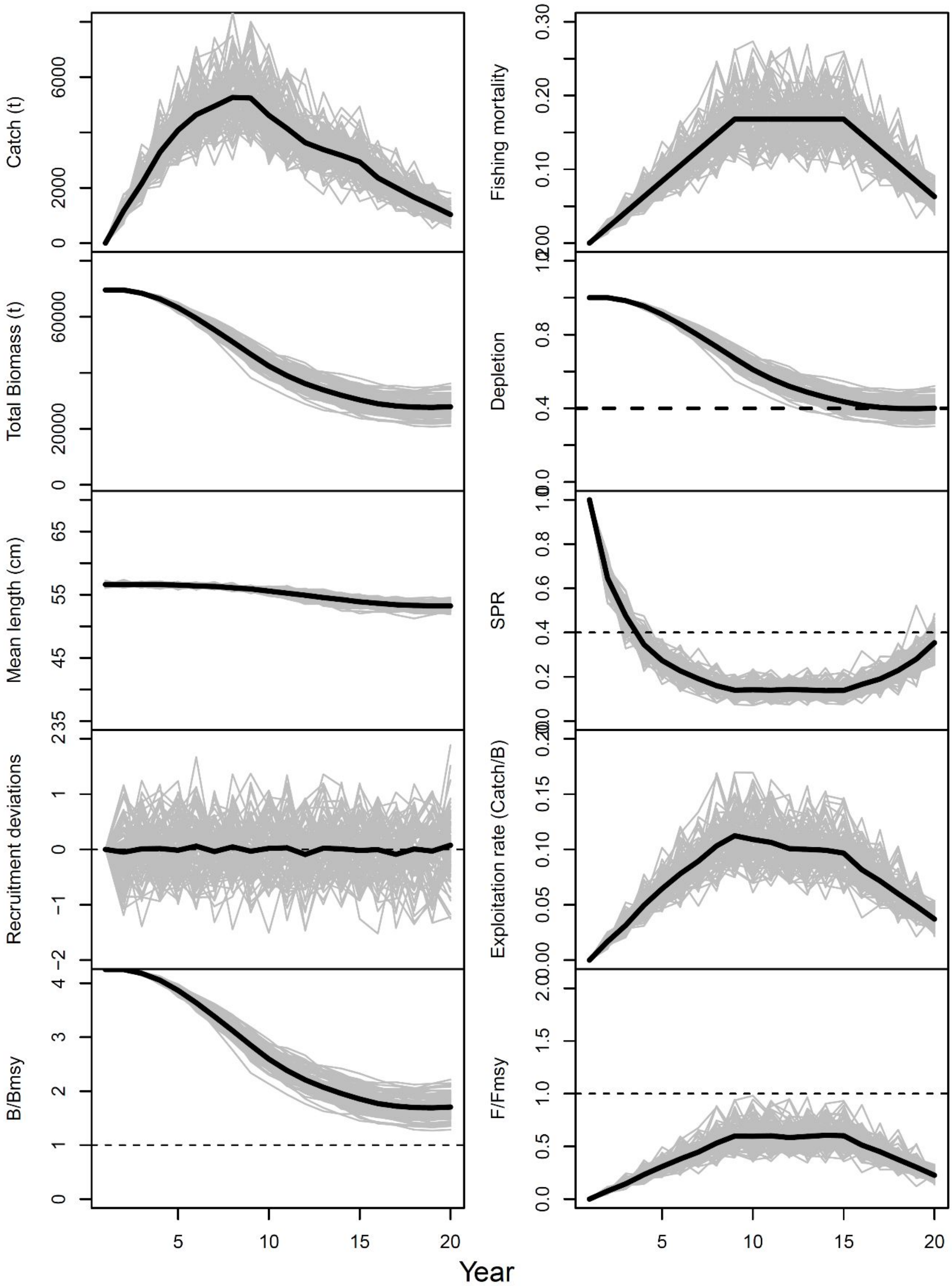

Figure A20. Time series for each simulated slow-grow canary rockfish population for harvest scenario 1 and depletion = 0.4. The black solid lines represent the mean value for all runs.

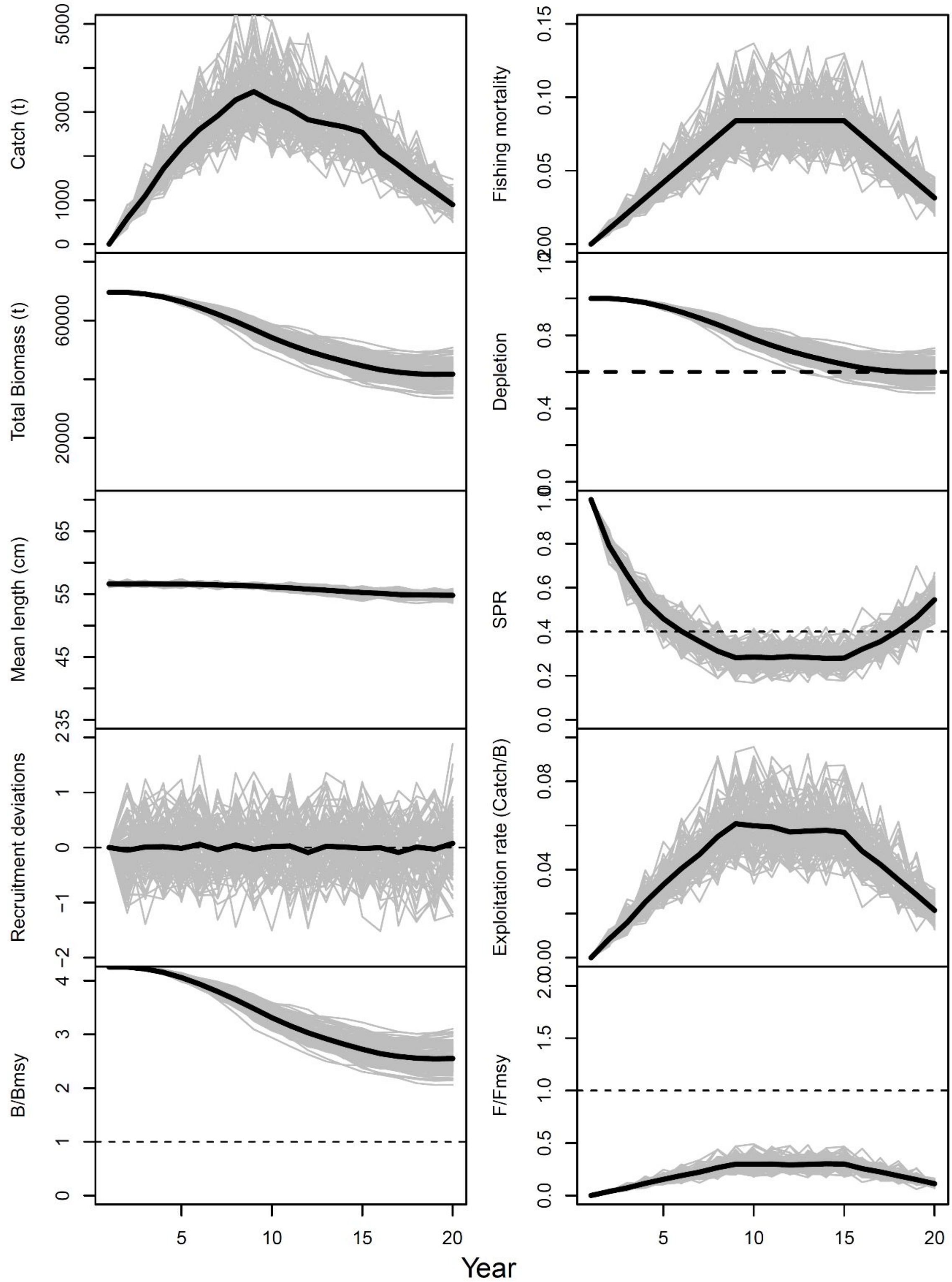

Figure A21. Time series for each simulated slow-grow canary rockfish population for harvest scenario 1 and depletion = 0.6. The black solid lines represent the mean value for all runs.

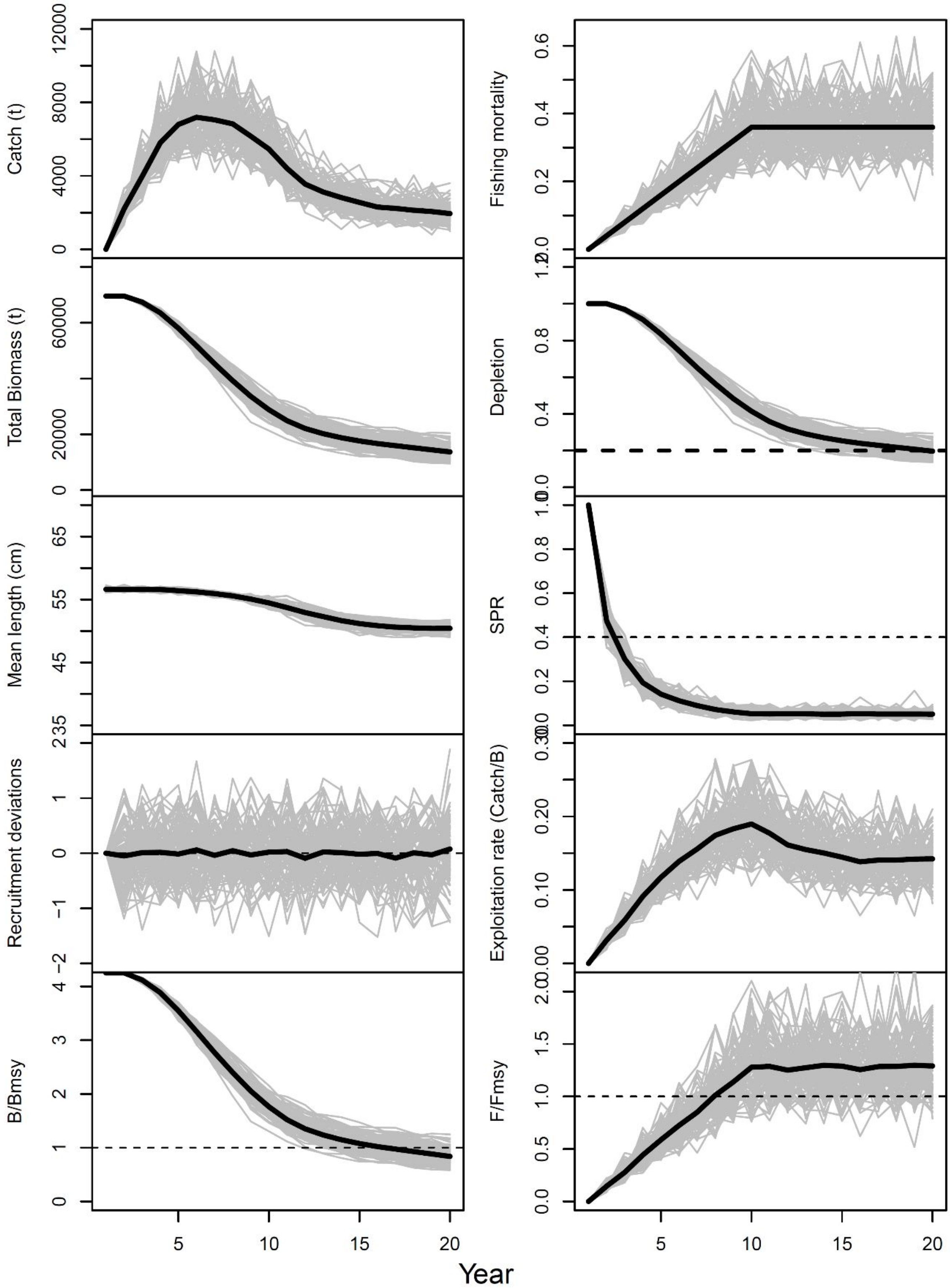

Figure A22. Time series for each simulated slow-grow canary rockfish population for harvest scenario 2 and depletion = 0.2. The black solid lines represent the mean value for all runs.

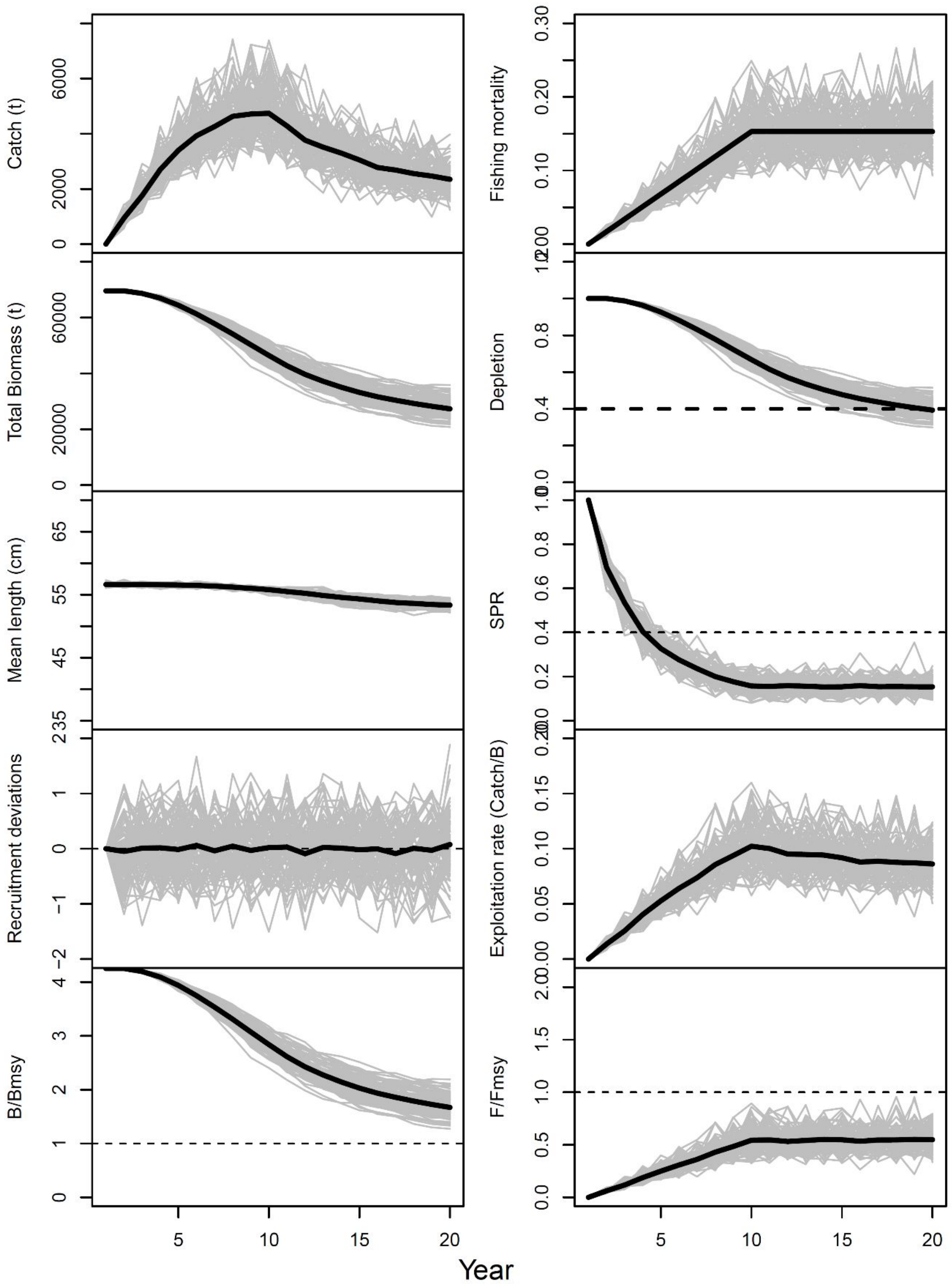

Figure A23. Time series for each simulated slow-grow canary rockfish population for harvest scenario 2 and depletion = 0.4. The black solid lines represent the mean value for all runs.

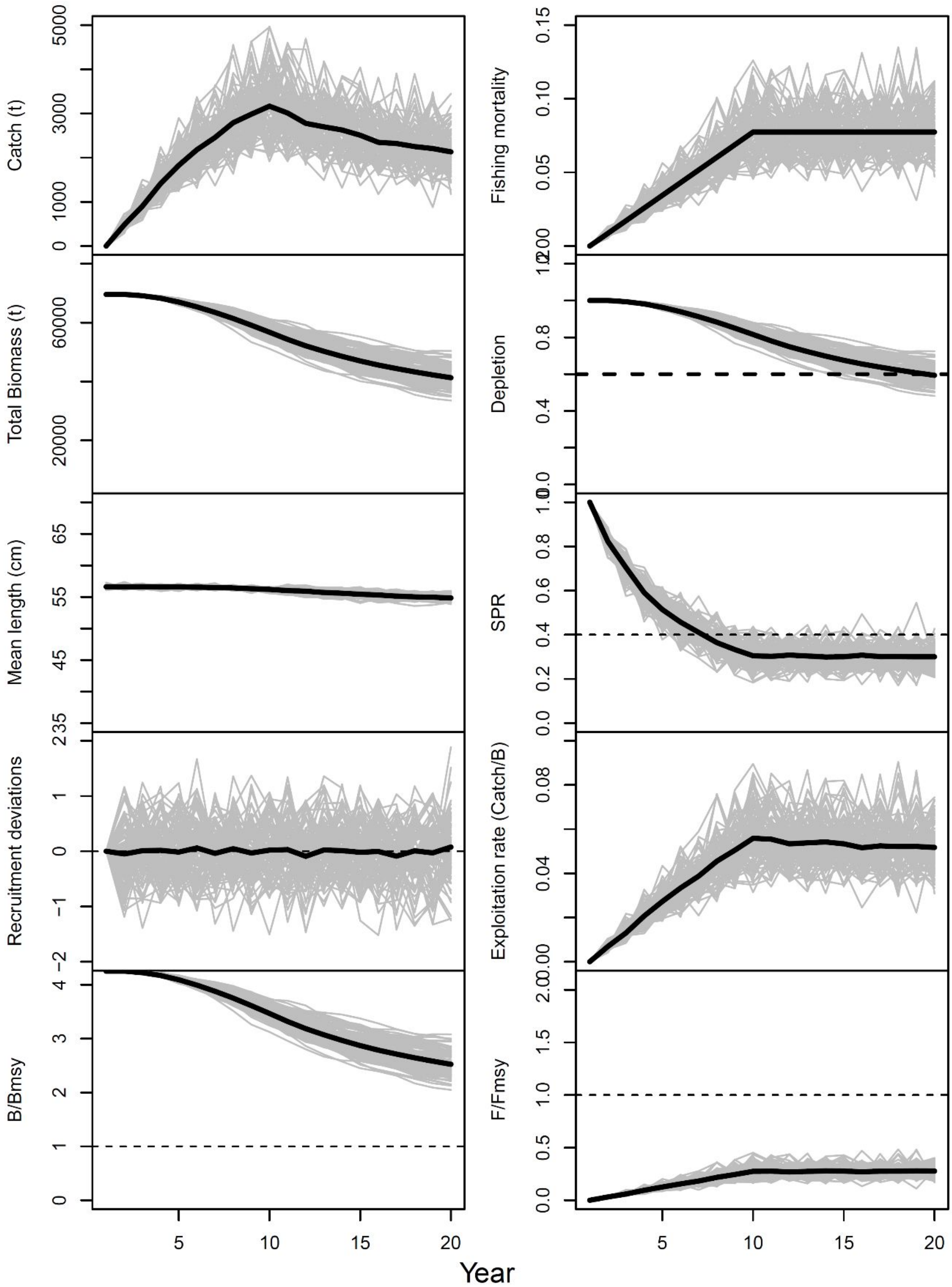

Figure A24. Time series for each simulated slow-grow canary rockfish population for harvest scenario 2 and depletion = 0.6. The black solid lines represent the mean value for all runs.

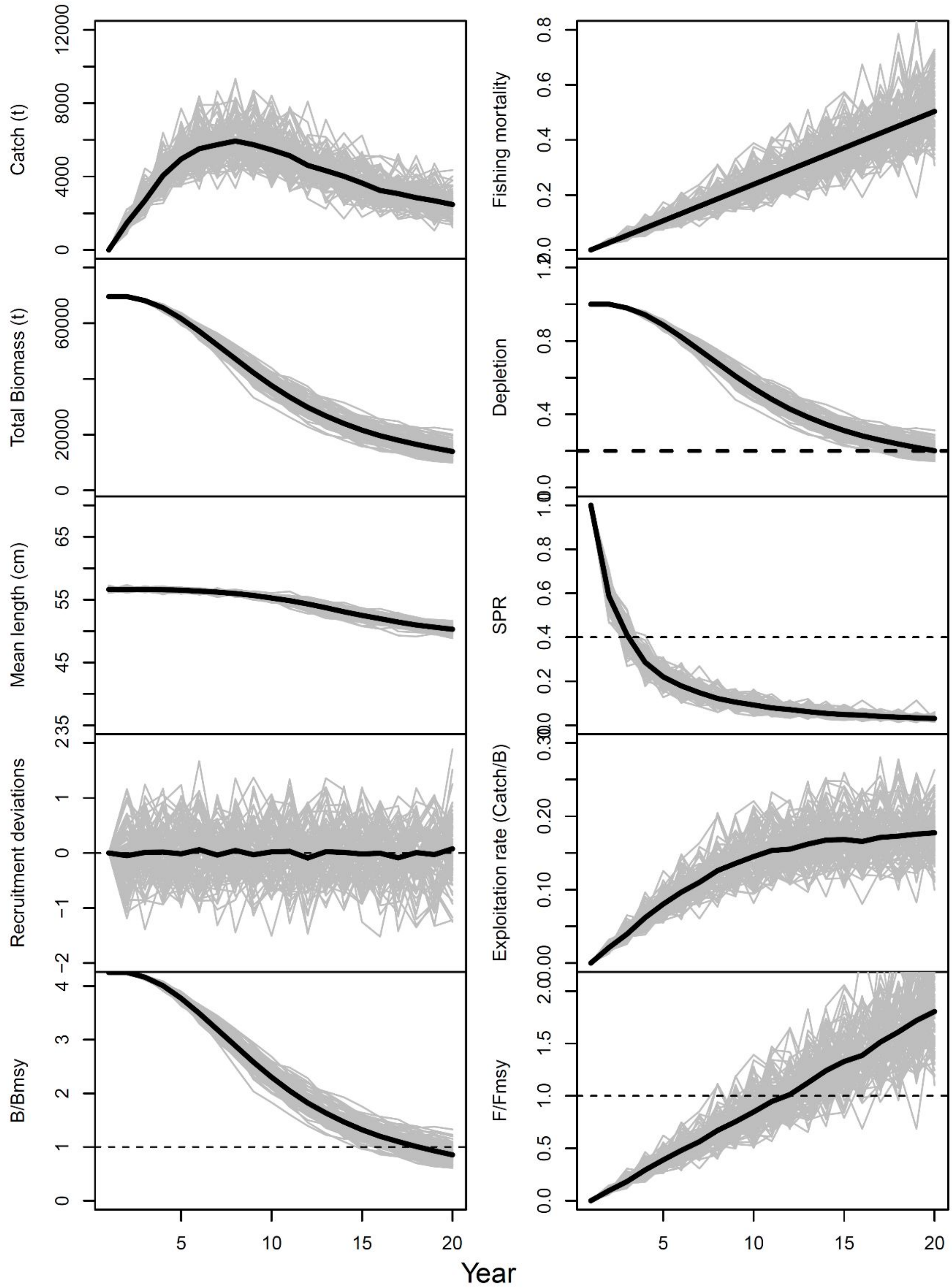

Figure A25. Time series for each simulated slow-grow canary rockfish population for harvest scenario 1 and depletion = 0.2. The black solid lines represent the mean value for all runs.

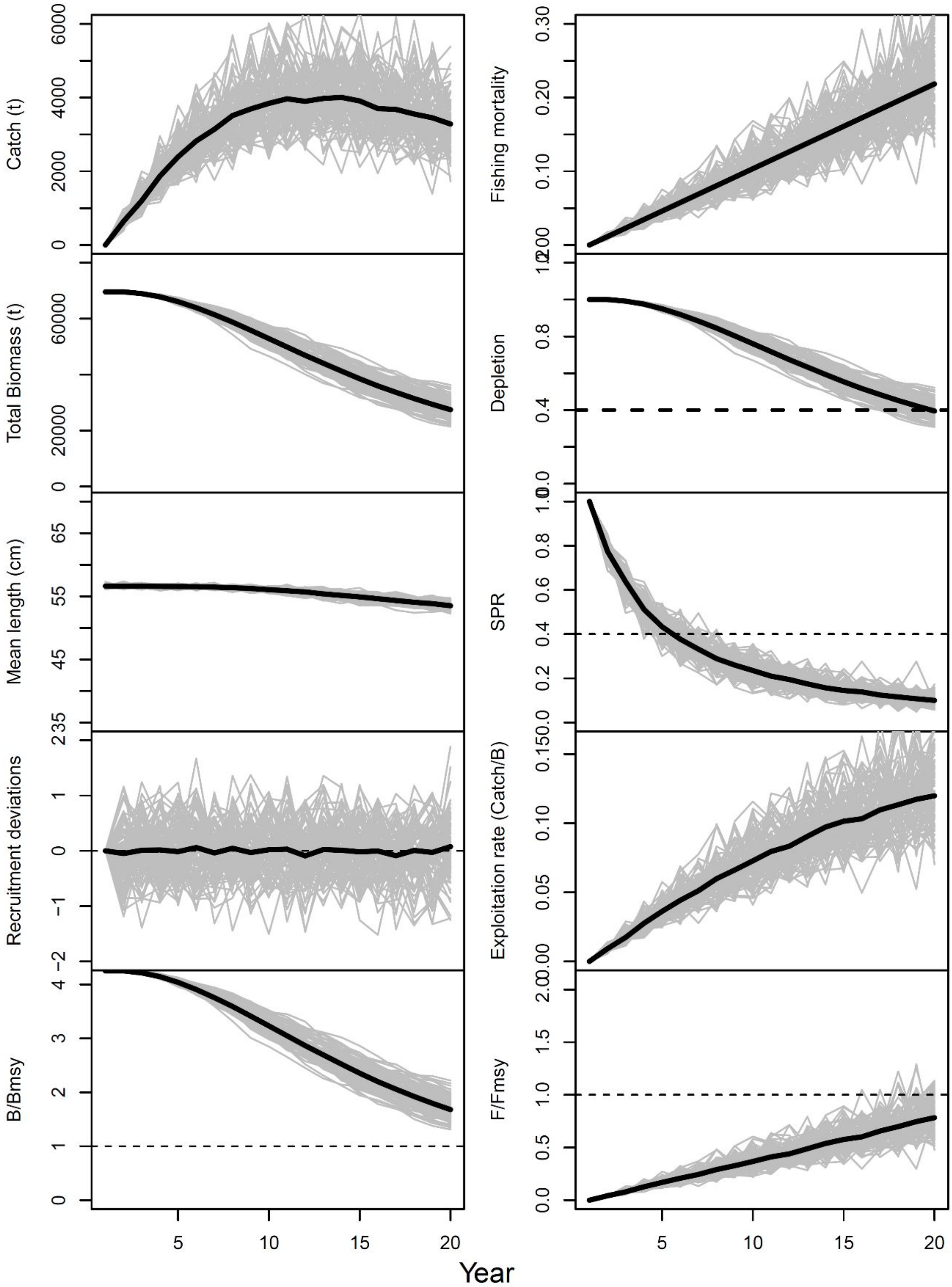

Figure A26. Time series for each simulated slow-grow canary rockfish population for harvest scenario 3 and depletion = 0.4. The black solid lines represent the mean value for all runs.

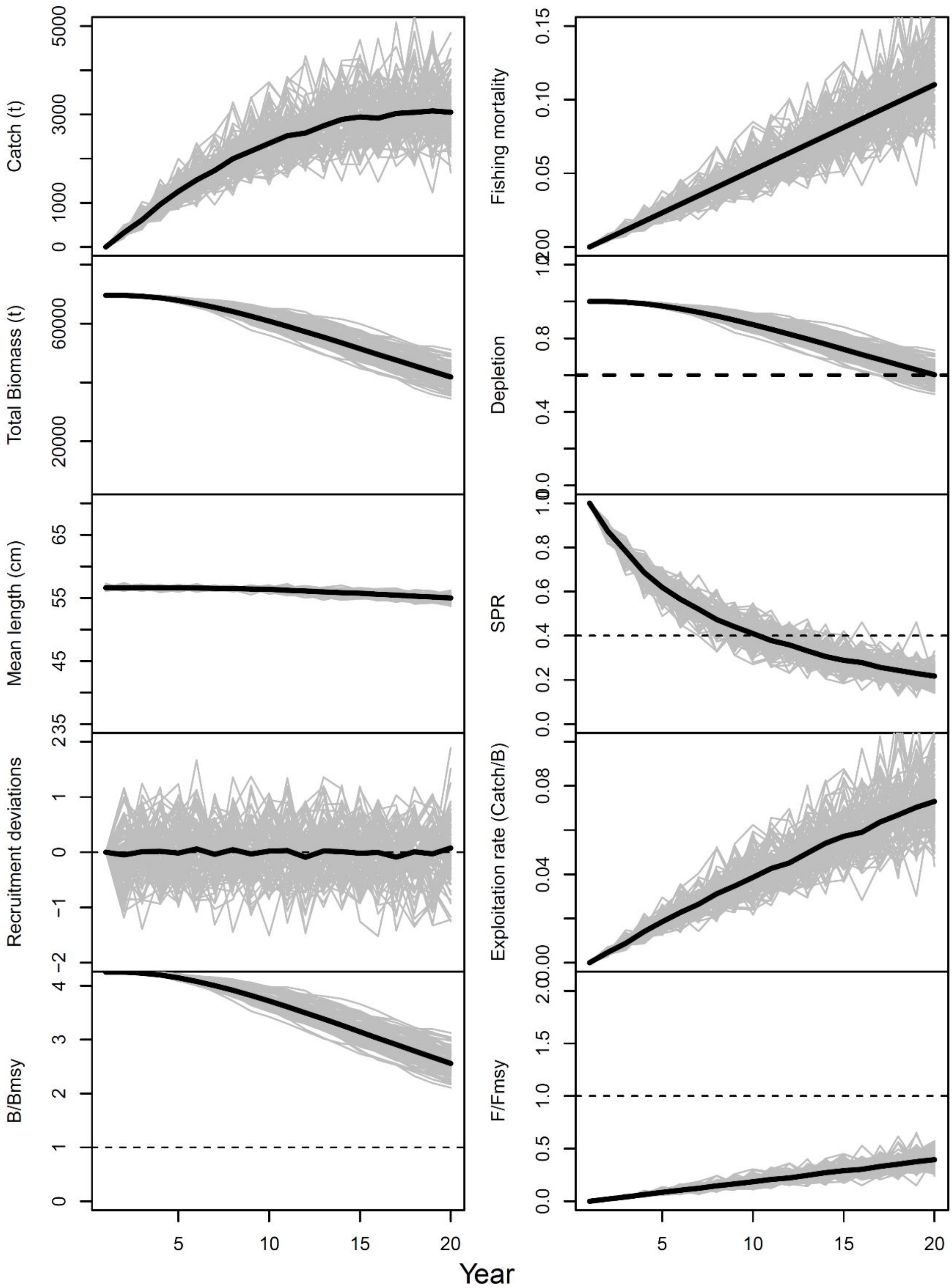

Figure A27. Time series for each simulated slow-grow canary rockfish population for harvest scenario 3 and depletion = 0.6. The black solid lines represent the mean value for all runs.

Table A1. Mean and standard deviation of relative error (RE) between the maximum sustainable yield (MSY) estimated by the operating model (OM) and the MSY estimated by the CMSY method. Values are proportions. Positive values mean that the MSY was overestimated and negative values that it was underestimated.

Species	Harvest Scenario	Mean ± Standard deviation		
		Depletion level		
		0.2	0.4	0.6
Mackerel	Scenario 1	0.37 ± 0.08	0.18 ± 0.09	-0.13 ± 0.15
	Scenario 2	0.28 ± 0.08	0.03 ± 0.09	-0.22 ± 0.13
	Scenario 3	0.33 ± 0.11	0.05 ± 0.11	-0.26 ± 0.14
Albacore	Scenario 1	-0.07 ± 0.02	-0.18 ± 0.03	-0.46 ± 0.02
	Scenario 2	-0.25 ± 0.02	-0.42 ± 0.02	-0.42 ± 0.02
	Scenario 3	-0.12 ± 0.02	-0.24 ± 0.02	-0.43 ± 0.02
Rockfish	Scenario 1	1.61 ± 0.27	1.13 ± 0.31	0.80 ± 0.32
	Scenario 2	1.67 ± 0.26	1.27 ± 0.33	0.86 ± 0.34
	Scenario 3	1.77 ± 0.33	1.29 ± 0.45	0.94 ± 0.45

Table A2. True OM and LBB estimated values for L_∞ and S_{50} (~Lc). LCL is the lower confidence limit and UCL the upper confidence limit for the estimated values.

Scenarios	Life-history	Harvest trend	Final depletion	True L_∞	True S_{50}	Estimated L_∞	Estimated LCL L_∞	Estimated UCL L_∞	Estimated Lc	Estimated LCL Lc	Estimated UCL Lc
1	Short-lived	Scenario 1	0.2	38.2	25	44.5	44.0	45.1	16.9	16.5	17.4
2			0.4	38.2	29	43.6	43.3	44.0	15.6	15.5	15.8
3			0.6	38.2	29	45.1	44.6	45.6	20.1	19.5	20.7
4		Scenario 2	0.2	38.2	29	44.7	44.2	45.2	18.0	17.4	18.6
5			0.4	38.2	29	43.7	43.4	44.1	15.7	15.5	15.8
6			0.6	38.2	29	45.3	45.0	45.8	20.3	19.7	20.9
7		Scenario 3	0.2	38.2	29	45.5	45.2	46.0	20.3	19.8	21.0
8			0.4	38.2	29	44.5	44.0	45.1	15.9	15.7	16.1
9			0.6	38.2	29	45.7	45.4	46.2	20.7	20.1	21.3
10	Medium-lived	Scenario 1	0.2	122.2	60	142.5	141.4	143.6	62.4	61.7	63.2
11			0.4	122.2	60	141.4	140.1	142.5	60.4	59.7	61.0
12			0.6	122.2	60	143.2	142.7	143.9	64.1	63.2	64.9
13		Scenario 2	0.2	122.2	60	143.0	141.8	143.9	62.8	62.0	63.6
14			0.4	122.2	60	141.9	141.1	143.3	60.7	60.0	61.4
15			0.6	122.2	60	143.3	143.0	144.0	64.6	63.8	65.4
16		Scenario 3	0.2	122.2	60	143.3	143.0	144.0	63.3	62.5	64.1
17			0.4	122.2	60	142.5	141.4	143.7	61.1	60.4	61.8
18			0.6	122.2	60	143.5	143.1	144.3	65.4	64.5	66.3
19	Long-lived	Scenario 1	0.2	60.0	60	76.7	76.0	77.6	50.4	49.9	51.0
20			0.4	60.0	45	75.3	74.5	76.3	47.0	46.5	47.4
21			0.6	60.0	45	76.9	76.2	77.8	51.2	50.7	51.8
22		Scenario 2	0.2	60.0	45	76.5	75.8	77.4	50.5	50.0	51.1
23			0.4	60.0	45	75.4	74.6	76.4	47.8	47.2	48.3
24			0.6	60.0	45	77.0	76.3	78.0	51.4	50.8	52.0
25		Scenario 3	0.2	60.0	45	76.8	76.1	77.7	51.0	50.4	51.6
26			0.4	60.0	45	76.0	75.3	76.9	49.5	49.0	50.1
27			0.6	60.0	45	77.1	76.3	78.0	51.7	51.1	52.2