Report on updates to the age-structured, integrated Catch-

MSY assessment approach

Merrill Rudd[1] and Steve Martell[2]

[1]School of Aquatic and Fishery Sciences Box 355020

University of Washington

Seattle, WA 98195-5020 U.S.A.

[2]Sea State, Inc., 9401 NE Tidal Ct. Bainbridge Island, WA 98110 U.S.A.

Report contents

I.   Summary of work conducted

The Catch-MSY approach was initially developed by Martell and Froese (2013) as a method for estimating MSY from catch data, species resilience, and assumptions about relative stock sizes in the first and last year of catch data available. The initial program followed a biomass-dynamic model, estimating the carrying capacity (K) and intrinsic rate of growth (r). More recently, improvements on the Catch-MSY approach have included management-oriented parameterization, moving from an underlying biomass-dynamic to age-structured model, and the ability to integrate more data types if available and estimate model parameters with a Bayesian Sampling Importance Resampling (SIR) algorithm. *Management oriented parameterization* transforms the scale parameter from K to MSY and the rate of change parameter from r to Fmsy (Forrest *et al.*, 2008). The management-oriented parameters more easily allow for development of priors, as MSY can be based on the previous catch, whereas it was more difficult to develop a prior for K. The *age-structured model* allows the Catch-MSY approach to reflect age-structured population dynamic processes such as growth and selectivity. The *integrated* nature of the Catch-MSY approach allows for an abundance index or biomass survey to further narrow down the possible parameter values based on fitting the model to the observed time series.

The Catch-MSY method with these new features underwent a CIE review in 2015. The three reviewers recommended a list of points for future work before the Catch-MSY method could be applied to inform management of U.S. stocks. This report summarizes the results of the new additions – including new features and testing.

New features of the Catch-MSY approach include (i) model fitting to mean length and length composition data, (ii) output of residuals between observed and predicted values included in the likelihood, (iii) option to estimate age at 50% selectivity, (iv) output of the spawning potential ratio (SPR) reference point, for comparison purposes with other data-limited assessments in the region, and (v) option to assume a dome-shaped selectivity curve.

This work has also tested the Catch-MSY approach more thoroughly. As requested by the CIE reviews, we (a) self-tested and (b) cross-tested the Catch-MSY method (Deroba *et al.*, 2014; Needle, 2014). Self-tests provide consistency checks within the assessment model, whereas cross-checks provide consistency check among assessment models. Self-testing was conducted by using a pre-specified operating model based on Namibian hake, including catch and an abundance index for Namibian hake. However, mean length and length composition data were not available for that dataset, so we generated time series for those indices from the operating model itself. We could then run the Catch-MSY method with that pseudo-data to test that the estimated MSY was near the MSY calculated when the mean length and/or length composition data were not included. Cross-testing was done using the operating model from the Length-based Integrated Mixed Effects (LIME) method, an age-converted-to-length-based stock assessment method that has the ability to fit to length composition, catch, and an abundance index assuming basic biological information is known to estimate annual fishing mortality, recruitment variation, and age at 50% selectivity (Rudd and Thorson, *in prep*; package available at: https://github.com/merrillrudd/LIME).

Rosenberg et al. (2014) conducted simulation testing on the biomass-dynamics-based Catch-MSY method (Froese *et al.*, 2016; Martell and Froese, 2013) in comparison to other catch-only assessment methods and found it to be the best performer of the approaches tested. Here, the age-structured Catch-MSY (in this report, referred to as Catch-MSY) is simulation tested with the other new features and compared to the biomass-dynamics approach (referred to as CMSY) (Froese *et al.*, 2016).

II.   New routines

Fitting to length data

The Catch-MSY method can now be fitted to mean length time series and length composition data (one year or multiple years). When a sample parameter combinations passes the non-statistical criteria, it can then calculate the age-length key and compositional predictions. The age-length routines are not calculated when the sample parameter combination does not pass the non-statistical criterion because in many cases there infinite or NA values in the population predictions that will cause an error in the age-length predictions. The age-length key represents the probability of being in a length bin given age. To get a predicted length composition in the catch each year, the age-length key is multiplied by the vulnerable abundance of fish at age in each year. The predicted mean length in the population each year is then calculated from the predicted proportion of catch in each length bin. This predicted mean length over time is added to the model output to be able to compare to the observed mean length in a dataset, and to

generate data from the model in the case of self-testing. For the purposes of self-testing, length composition data is generated from the Catch-MSY model using a multinomial distribution with sample size 1,000 and the predicted probability of being harvested in each length bin. A sample size of 1,000 has been shown to be the minimum sample size where the attributes of the length distribution can be clearly observed (Gerritsen and McGrath, 2007).

With the predicted values for mean length and length composition in the model, the method can then fit to observed data with assumptions about observation error. The mean length likelihood component is lognormally distributed:

(1) $$L(l_t) \sim Lognormal(\overline{l}_t, \sigma_l)$$

where l_t is the observed mean length time series, $\overline{l}_t$ is the predicted mean length time series, and σ_l is the mean length observation error. The mean length time series can be non-continuous. This routine requires assumed values for σ_l and variation around the growth curve.

The length composition likelihood component follows a multivariate logistic function:

(2) $$logL(LF_{xt}) = \sum_{t=1}^{T} \frac{1}{2}\log(X) - (X-1)\left(\frac{1}{2}\log(2\pi) + \log(\sigma_{LF})\right) - \frac{1}{2\sigma_{LF}^2}\sum_{x=1}^{X} \delta_{xt}^2$$

where X is the maximum length bin (converted from the maximum age which was treated as a plus group), σ_{LF} is the length frequency observation error, and δ_{xt} are the residuals between the observed and predicted length frequency at length over time, defined as:

(3) $\delta_{xt} = \log(LF_t) - \log(\overline{LF_t}) - \frac{1}{X}\sum_{x=1}^{X}\log(LF_t) - \log(\overline{LF_t})$

Additional outputs

Prior to this work, the output of the Catch-MSY method included, for each parameter combination, a) the "code" signifying whether the parameter combination could have led to an extant population harvested by the given catch time series, b) unfished biomass, c) steepness, d) negative log likelihood (when auxiliary data were available), e) prior, f) total biomass time series, g) depletion time series, h) spawning biomass time series, i) fishing mortality time series, and j) weight of the particular parameter combination for the SIR algorithm. We have now added k) predicted mean length, l) predicted length frequency, m) spawning potential ratio at Fmsy, n) spawning potential ratio given the population was fished at the estimated level of fishing mortality annually (i), and when available, o) residuals for the observed and predicted biomass survey, p) residuals for the observed and predicted abundance index, q) residuals for the observed and predicted length composition data used in the multivariate logistically-distribution (Equation 3), and r) residuals for the observed and predicted mean length.

The addition of the spawning potential ratio (SPR) output facilitates comparison of the Catch-MSY approach with other data-limited stock assessment methods that do not output MSY reference points. A benefit of the integrated Catch-MSY assessment is that it can take advantage of catch data that is available to narrow down a possible value for MSY to inform catch limit recommendations, and integrate other data types to narrow down the posterior distribution for MSY when they are available. However, in many areas where data-limited stock assessment methods are required, catch data is very

difficult to reliably obtain due to issues of capacity limitation and the decentralized nature

of many small-scale fisheries. In this context, length measurements may be more reliable,

and length composition and/or mean length data may hold information on the level of

fishing pressure with assumptions about biological characteristics and selectivity.

However, length data does not hold information on the scale of the population, which is

required to produce MSY-based reference points. Alternatively, the SPR reference point

compares the lifetime potential egg production per recruit in the fished to unfished state,

and does not require a measure of scale. A harvest strategy that aims for 40% SPR, or a

fishing mortality rate that is expected to result in 40% SPR (F_{40}), is considered risk averse

for many species (Clark, 2002). F_{40} can then serve as a proxy for Fmsy, and SPR below

40% can provide a reference for whether the population is overfished.

From this work, the Catch-MSY method now outputs the SPR evaluated at a)

each value of Fmsy drawn from the prior distribution and b) the theoretical SPR given the

population continued to be fished at the level of fishing mortality estimated annually, for

each sample of parameter values. Because the Catch-MSY method does not assume

equilibrium conditions (i.e. the fishing mortality can vary between years), SPR is

essentially the ratio of relative reproductive potential at fished and unfished states. The

most important value of SPR output from is the derived SPR in the terminal year. This is

a reasonable proxy for the fishing mortality.

Selectivity parameter estimation

With the possibility of length composition data integration, there may be enough

information to estimate the age at 50% selectivity into the estimated parameter list along

with natural mortality, Fmsy, and MSY. This would allow the analyst to better

characterize the uncertainty in the selectivity curve. The age-at-50% selectivity parameter

can be estimated based on information in the composition data conditional on

assumptions about fish growth and natural mortality. This option is added within the

functions 'catchMSYModel', 'sample.sid', and 'sir.sd' – specifying whether the fourth

column, draws from the prior on the age at 50% selectivity, should be added to the list of

samples. When the selectivity parameter is included in the prior distributions to be

sampled, the analyst must specify the type of prior distribution, its mean/minimum, and

its standard deviation/maximum, for the case of a lognormal or uniform prior,

respectively.

Dome-shaped selectivity option

When setting up the operating model (e.g. specifying the life history information,

putting data in the format required for Catch-MSY, etc.) there is now an option to specify

the selectivity function used. The argument *smodel* is set to "logistic" as a default.

However, the analyst can change the *smodel* to "dome" to specify a double-normal

selectivity function, in which case the model also uses the argument *dome_sd* in the

operating model to set the standard deviation for the normal distribution applied to the

older age groups. In both cases, the parameter *sel1* is the age at 50% selectivity as young

individuals become vulnerable to the gear. In the logistic model, the parameter *sel2* is the

age at 95% selectivity. When using the dome-shaped selectivity model, *sel2* becomes the

age at maximum selectivity, and the *dome_sd* is the standard deviation of this normal

distribution, specifying the rate at which older individuals become less vulnerable to the gear.

III.   Simulation testing

Self-testing

Self-testing is used to evaluate a model by using the model itself to generate simulated data, then analyze the pseudo-data using the same model. The idea is that a reasonable model should be able to estimate the parameters without bias if the "true" and "predicted" values are generated from the same model. We conducted a simple self-test of the new routines to fit to mean length and length composition data using a pre-specified operating model (OM) based on Namibian hake. The hake dataset includes catch and an abundance index, but no length information. We generated length composition and mean length data using the Catch-MSY model based on the hake OM, and then tested these new routines by running the Catch-MSY method with the pseudo-dataset to ensure the estimated MSY was near the MSY calculated when the mean length and/or length composition data were not included.

We used self-testing to a) explore how much the posterior distributions for MSY and age at 50% selectivity were updated with each data type from the Namibian hake dataset and b) test the new routines fitting to mean length and length composition data, generated from the same model. We ran the Catch-MSY method for five data scenarios: i) catch-only, ii) catch and abundance index, iii) catch and mean length data, iv) catch, mean length, and abundance index, and v) catch and length composition data. For each of

these five data scenarios, we ran the method twice: once fixing the age at 50% selectivity,

and another time adding a prior on the selectivity parameter and including it in the list of

parameters from which to draw samples.

In the case of the Namibian hake dataset (Figure 1), the Catch-MSY method with

catch data only updated the posterior distribution of MSY from the prior, but not by

much. The median for the prior distribution was 249 with lower 95% credible interval at

109 and upper 95% credible interval at 392 based on the same quantiles from the catch.

The posterior distribution using only catch data updated to a median of 258 (lower 95%

CI=137, upper 95% CI=391). When the model included catch and an abundance index,

however, the posterior distribution updated much more from the prior distribution. The

median of the posterior distribution including both catch and abundance index was 227

(lower 95% CI=102, upper 95% CI=280). The abundance index appears to provide

information to the model that narrows down the MSY to be lower than that expected

using catch data alone. Including mean length in the integrated model did not update the

posterior distribution on MSY any more than the catch data alone (median 260, intervals

same as the model using catch only). This is likely because the mean length is not

changing much over the time period (Figure 1). When the Catch-MSY method fit to both

mean length and the abundance index, however, the resulting posterior distribution was

even more narrow than the scenario with catch and abundance index, no mean length

data. Both the abundance index and mean length provide information on the rate at which

the population is changing over time, however the abundance index is more informative.

The mean length data likely reinforces the trajectory at which the abundance index helps

predict, thus better informing the parameter estimates for Fmsy and MSY (related to their correlation).

Self-testing of the routine to fit to length composition data revealed lack of fit between observed and predicted distributions, leading to zero of the candidate combinations of M, Fmsy, and MSY chosen as plausible values based on the inclusion of the length composition data. The values for M, Fmsy, and MSY govern the predicted population, and since the possible values of each of these parameters are drawn from a prior distribution, it is possible that the values will be far enough away from the truth that the shape of the length composition data could look different from the observed data. The likelihood of the data given the model is then very low, close to zero. In a maximum likelihood context or other Bayesian algorithm besides SIR (e.g. MCMC, Hamiltonian), the low likelihood would mean that the length composition data contributes little to the inference about the estimated parameters. In the context of MCMC or the Hamiltonian algorithms, the posterior distribution would not be updated much further from the posterior distribution that would arise if the length composition data was not included. In the SIR context, however, the likelihoods from the fits to the auxiliary data (e.g. index, mean length, or length composition) given the draws from the parameter distributions are used to sample from the possible parameter combinations not ruled out from catch data alone. When the likelihood is very low, the sampling algorithm does not choose any of the parameter combinations. This results in the same lack of update of the posterior distribution as would occur for other algorithms. However, if one parameter combination allows the predicted length composition to fit the observed quite well, the algorithm will

choose that parameter combination above all others, resulting in unrealistically high certainty in the posterior distribution for MSY.

The estimation of the age at 50% selectivity demonstrated the information held by each data type respectively. Catch data did not update the posterior distribution on the selectivity parameter from the prior distribution very much, as the median and credible intervals were effectively identical (Figure 3). Unlike the estimation of the MSY parameter, the abundance index was not very informative in estimating the selectivity curve. This makes sense, since the selectivity function is an age-structured process, and the abundance index does not hold any information on the age structure in the catch or population. The mean length, however, did contribute information to update the posterior distribution on the age at 50% selectivity. The uncertainty in the parameter decreased with the inclusion of this data type (with mean length: median=3.71, lower 95% CI=2.73, upper 95% CI=5.32; as opposed to the prior distribution with median =3.93, lower 95% CI=2.02, upper 95% CI=8.7). However, the addition of both the abundance index and mean length data led to more uncertainty in the age at 50% selectivity parameter than when only the mean length data was included. This could be due to data conflict, or uncertainty in other parameters for those runs. The length composition data should theoretically hold the most information on the age at 50% selectivity, but unfortunately there was not much update between the posterior and prior distributions due to the difficulty fitting to length composition data.

Self-testing indicated that the abundance index is very important to decrease uncertainty in the posterior distribution for MSY. Mean length information did not serve

to update the posterior distribution of MSY much more than was updated using catch data

alone. One caveat to this finding is that the mean length data may not have been very

informative based on lack of changes in the Namibian hake population. If the mean

length had more of a trend during the time series, it would be more informative of the

trend in fishing mortality and thus, how hard the population has been fished (Fmsy). Due

to correlations between Fmsy and MSY, the posterior distribution for MSY may have

been updated more from the information held in the mean length data if the mean length

data changed more over the time period.

Diagnostics

The residuals between the observed and predicted data are now output from the

Catch-MSY method (when an abundance index, biomass survey, mean length, or length

composition data are available). Analysts using the integrated Catch-MSY method should

look at the residuals to ensure they are normally distributed. If they are not, this is an

indicator that the distributions for the observed and predicted data do not match, and that

the Catch-MSY method is not an adequate model to predict the observed data. For

example, the residuals for the abundance index are relatively normally distributed, but the

residuals for the length composition data show patterns that indicate the fits are not good

(Figure 4). This indicated samples from the Fmsy and MSY prior distributions must be

close enough to the truth for the predicted length composition data to match the observed,

which is difficult given the Catch-MSY algorithm.

Another diagnostic would be to compare the prior and posterior distributions for

estimated parameters. If the posterior distribution is not updated from the prior

distribution, then this means the data types are not providing any information on the

parameter of interest (e.g. MSY, Fmsy, M) (Figure 2).

Cross-testing

We used an age-structured operating model from the LIME package

(https://github.com/merrillrudd/LIME) to generate data from a different model than the

Catch-MSY model. The main differences between the two age-structured models are that

the LIME operating model does not necessarily start at a lowly-fished state, whereas the

Catch-MSY assessment assumes there is no fishing prior to the first year. The LIME

method assumes a single-parameter logistic function for the selectivity curve, as opposed

to the Catch-MSY method's two-parameter logistic function (or dome-shaped function if

that option is turned on).

We tested the Catch-MSY method's performance in estimating MSY against

"true" simulated populations based on four life history types, four fishing mortality

patterns, three recruitment patterns. This rigorous simulation testing helps identify

strengths and weaknesses in the Catch-MSY method, where different data types improve

inference, and how the assessment method may deal with different possible scenarios in

the real world. We assumed three data availability scenarios, all assuming 20 years of

data were available: a) catch-only, b) catch and abundance index, and c) catch and mean

length data. Unfortunately due to time constraints and poor model fits to the length

composition data (discussed above in *Self testing*) we did not simulation test the Catch-

MSY method integrated with length composition data at this time. We ran 50 iterations of

each combination of data availability, life history, fishing mortality, and recruitment

scenarios to understand the variability in the ability of the model to estimate the true MSY.

The three life history types mimicked: (a) spotted rose snapper (*Lutjanus guttatus*), a medium-lived fish (asymptotic length L_∞=64.6 cm, von Bertalanffy k=0.21, natural mortality M=0.43, and length at 50% maturity l_{m50}=34 cm), (b) rabbitfish (*Siganus sutor*), a short-lived fish (L_∞=36.2 cm, k=0.87, M=1.49, and l_{m50}=20.2 cm), and (c) Namibian hake (*Merluccius capensis*), a long-lived fish (L_∞=100 cm, k=0.13, M=0.15, and l_{m50}=29.5 cm) (Figure 5).

The four fishing mortality scenarios included a) endogenous, or a pseudo-bioeconomic model where the effort dynamics follows the population dynamics, b) increasing, c) constant, and d) ramped patterns (Figure 6). All fishing mortality scenarios include a burn-in period where the populations are fished constantly at a rate of 0.05 to be similarly depleted across scenarios in the initial year. After the burn-in period, the constant fishing mortality scenario stayed at a rate of 0.25. The endogenous fishing mortality scenario used an effort dynamics model from Thorson et al. (2013) that approximates dynamics for a developing fishery, with an equilibrium fishing mortality rate of 0.25, initial fishing mortality of 0.05, and the F_{rate} value of 0.2. The increasing fishing mortality pattern increased from the initial value of 0.05 to the maximum rate of 0.7. The ramp-up pattern represents a situation where the fishing mortality increases from an initial value of 0.05, to a maximum rate of 0.7, and then a "managed" state at 1/3 of the maximum fishing mortality rate. We included process variability in the fishing mortality scenarios, lognormal error with standard deviation 0.3.

We also explored three different patterns of time-varying recruitment (Figure 6), including constant, where essentially the same number of recruits are added to the population each year regardless of the local stock size, pulsed recruitment, indicative of phase shifts over time (Vert-pre *et al.*, 2013), and Beverton-Holt asymptotic dynamics (Hordyk *et al.*, 2014), where the steepness parameter in the stock-recruit function is fixed at 0.7 (as opposed to 1.0 which would result in constant recruitment regardless of stock size). We included process variability in the recruitment scenarios, lognormal error with standard deviation 0.6. We assumed the level of recruitment variability was known in the Catch-MSY method; there is not currently the capacity to estimate this value, only to include some assumed value of recruitment variation.

For each scenario, we generated data from the LIME method and provided the Catch-MSY operating model with the life history information (assumed known to be true) and the generated input data (i.e. catch time series, abundance index, mean length, and length composition, depending on the data availability scenario). We used a lognormal prior distribution on natural mortality with the mean at the true value of natural mortality and standard deviation of 5% of the true value of natural mortality. This provides some uncertainty around the value of natural mortality, instead of just assuming the input value is the truth. We used a uniform distribution for the prior on Fmsy, between 20-150% of the natural mortality. We based the uniform prior distribution for MSY on the quantiles from the catch time series, assuming the true MSY would fall between 5-95% of the observed catch. We assumed all observation errors to be 0.2. This value should be adjusted based on the dataset, but the method does not currently have the capability to estimate these parameters.

For each scenario, we output the relative error between the true MSY and the median MSY from the posterior distribution ((observed – true)/true) and the coverage interval, or the proportion of iterations out of 50 where the true MSY was within the 95% credible intervals.

One of the main take-away points of the simulation study is that the Catch-MSY method works best for the long-lived life history type when an abundance index or mean length time series is integrated along with the catch time series. For the hake life history type, we see biases in estimates of MSY under the catch-only scenario. When recruitment follows a Beverton-Holt stock recruit function where the steepness parameter is fixed to 0.7 in the operating model, mean length does not improve accuracy in estimates of MSY when fishing mortality was either increasing or ramped pattern. This is likely due to the fact that the steepness parameter is misspecified in the Catch-MSY estimation model, and since mean length does not hold information on MSY, the misspecification was detrimental to the accuracy of the estimate of MSY.

We see better performance of the catch-only version of catch-MSY for the medium-lived life history type under the endogenous fishing mortality scenario. Overall, however, there are positive biases in estimates of MSY for the medium- and short-lived life history types (Figures 8 and 9). For the medium-lived life history type, the inclusion of an abundance index improves accuracy, but the inclusion of mean length data does not necessarily improve the accuracy of MSY estimates from the catch-only scenario, on average. The Catch-MSY method performs well for the short-lived life history type when fishing mortality is constant over time. For the other fishing mortality scenarios,

however, there are positive biases in the estimates of MSY, and the estimates of MSY are more variable between iterations in the other scenarios. In the endogenous, increasing, and ramped fishing mortality patterns, the inclusion of an abundance index actually decreased precision in estimates of MSY across iterations of generated data. One possible reason for this is because the selectivity curve is set up such that the age at 95% selectivity is specified as one year older than the age at 50% selectivity. The age at 95% selectivity in the short-lived species model is also the maximum age, meaning that no age group in the population is ever fully selected to the gear. This may cause high uncertainties in estimating the scale of the population, when there is some proportion of the population un-accounted for.

Another major take-away is that even for one of the best performing scenarios (the long-lived hake life history type, endogenous fishing mortality and constant recruitment, with an abundance index), there is high variability between iterations of generated data in the ability to accurately and precisely estimate MSY (Figure 10). While looking at the median estimates across 50 iterations shows that MSY can be estimated without bias on average, the true MSY was within the 95% credible intervals for only 74% of the iterations. We would expect this number to be 95% based on the definition of the 95% credible interval. This still occurs even though the median relative error between median estimates of MSY is relatively low (0.12). The same pattern is true when fishing mortality is constant over time, with 76% of the iterations having the true MSY within the 95% credible intervals, when the median relative error between the true and estimated MSY is only -0.03 (Figure 11). Even when the method is predicted to be unbiased on

average across iterations, there is variability in accuracy and precision for a single

iteration of generated data and run of the Catch-MSY assessment method.

Comparison to the biomass-dynamic version of Catch-MSY

The biomass-dynamics version of the Catch-MSY (CMSY) is still under

development, with a recent publication in advancements on the method (Froese *et al.*,

2016). This method estimates the Schaefer parameters r and K instead of management-

oriented parameters Fmsy and M. When an abundance index or biomass survey is

available, the CMSY method includes a Bayesian state-space algorithm (as does the

Catch-MSY method) but uses the Markov Chain Monte Carlo algorithm instead of the

SIR algorithm. The CMSY method cannot integrate mean length or length composition

data. We tested the biomass-dynamic CMSY method against the same simulated

populations and generated datasets as were used to test the Catch-MSY method.

When tested against our age-structured true population, the age-structured

integrated Catch-MSY method performed better than the biomass-dynamic CMSY

method both when only catch data was available and when the methods integrated an

abundance index (Figures 12 and 13). When only catch data was available (under

endogenous fishing mortality, constant recruitment, and recruitment standard deviation

0.6, for example), the median relative error in estimates of MSY across scenarios was

1.76 for the biomass-dynamics approach and 0.35 for the age-structured approach (Figure

12). For the age-structured approach, the true MSY was within the 95% credible intervals

for 94% of the iterations, while the coverage rate was lower for the biomass-dynamics

approach at only 44% (Figure 12). When an abundance index was included, the overall

message was the same (Figure 13). While the abundance index improved the accuracy of estimates of MSY for both methods, the age-structured Catch-MSY had lower median relative error (0.12 compared to 0.40 for CMSY) and higher coverage rates (74% compared to 36% for CMSY).

IV.   User manual

The R package for the age-structured, integrated Catch-MSY method can be found at https://github.com/smartell/CatchMSY. The documentation includes details on the objective, arguments, and output of each function. We also included examples of how to run the Catch-MSY method using the package, including how to set up the data in the operating model, how to view results, and suggestions on diagnostics.

V.   Future work

When the number of parameter estimates increases from two (MSY and Fmsy) to three or four (natural mortality and age at 50% selectivity), the sampling-importance-resampling (SIR) algorithm becomes less efficient. This algorithm has very poor performance when trying to estimate a larger number of parameters; the importance function becomes incredibly sensitive to minor changes in model parameters if the variance terms are mis-specified. The addition of length composition increases this sensitivity because it adds another dimension to the likelihood function. Future work should investigate additional or alternative algorithms for sampling the joint posterior distribution when more data are available. One alternative Bayesian algorithm is the Markov Chain Monte Carlo (MCMC, as is used by the CMSY Bayesian state-space

method) or a Hamiltonian algorithm available in STAN. The algorithm update is outside

the scope of this work, but the issues in fitting to length composition data identified in

this study could spur the addition of alternative gradient-based algorithms for cases when

the data are informative.

Future work should also include ways of coming up with logical catch advice

based on the results of several assessment methods. A benefit of the Catch-MSY method

is that is uses available catch data to narrow down possible values of MSY in order to set

catch limits, or at least hazard a guess as to what the current catch is relative to MSY.

Other methods that rely on mean-length or length-composition data to estimate fishing

mortality and spawning potential ratio (SPR) proxy reference point are less informative

about model scale (i.e., how large is the stock). As more types of assessment methods

become available, including integrating existing and new methods and types of data,

methods for making sensible recommendations on catch limits should also be developed;

by analogy, hurricane landfall predictions are now based on an ensemble of models.

Sometimes the output from a single assessment model just don't make sense and

in practice we ignore the model output and rely on more precautionary methods. But you

can't always tell from a single assessment if the results make sense. For example, in the

Catch-MSY method, the simulation study demonstrated that adding an abundance index

or mean length data does not always improve accuracy and precision in estimates of

MSY. This seems counter-intuitive, as more data should mean better precision. In the

mean-length mortality estimators that utilize mean length data from the same stocks, it is

possible that some values of mean length could lead the model to think that fishing

mortality is (implausibly) negative. Or incorrect structural assumptions (e.g., shape of the selectivity curve) could result in creating contradiction among composition data and abundance indices. Future work should look into whether inconsistencies between stock assessment methods are always attributed to the same reason, including data conflict or structural assumptions, to tease apart why the multiple assessment methods may be telling different stories about the same stock. The statistical challenge, in data-poor situations, will be assigning different weights to each of the candidate models.

References

Clark, W. G. 2002. F 35% revisited ten years later. North American Journal of Fisheries Management, 22: 251–257. http://www.tandfonline.com/doi/abs/10.1577/1548-8675(2002)022<0251:FRTYL>2.0.CO;2\npapers2://publication/uuid/42C84EAA-DE5C-48C3-977D-8AAF29A16583.

Deroba, J. J., Butterworth, D. S., Methot, R. D., De Oliveira, J. A. A., Fernandez, C., Nielsen, A., Cadrin, S. X., *et al.* 2014. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES Journal of Marine Science.

Forrest, R. E., Martell, S. J. D., Melnychuk, M. C., and Walters, C. J. 2008. An age-structured model with leading management parameters, incorporating age-specific selectivity and maturity. Canadian Journal of Fisheries and Aquatic Sciences, 65: 286–296.

Froese, R., Demirel, N., Coro, G., Kleisner, K. M., and Winker, H. 2016. Estimating

fisheries reference points from catch and resilience. Fish and Fisheries: 1–21.

Gerritsen, H. D., and McGrath, D. 2007. Precision estimates and suggested sample sizes

for length-frequency data. Fishery Bulletin, 105: 116–120.

Hordyk, A., Ono, K., Valencia, S., Loneragan, N., and Prince, J. 2014. A novel length-

based empirical estimation method of spawning potential ratio (SPR) and tests of its

performance, for small-scale, data-poor fisheries. ICES Journal of Marine Science.

Martell, S., and Froese, R. 2013. A simple method for estimating MSY from catch and

resilience. Fish and Fisheries, 14: 504–514. http://doi.wiley.com/10.1111/j.1467-

2979.2012.00485.x (Accessed 18 December 2013).

Needle, C. L. 2014. Using self-testing to validate the SURBAR survey-based assessment

model. Fisheries Research, 171: 78–86. Elsevier B.V.

http://dx.doi.org/10.1016/j.fishres.2015.03.001.

Rosenberg, A. A., Fogarty, M. J., Cooper, A. B., Dickey-Collas, M., Fulton, E. A.,

Gutierrez, N. L., Hyde, K. J. W., *et al.* 2014. Developing new approaches to global

stock status assessment and fishery production potential of the seas. 175 pp.

Thorson, J. T., Minto, C., Minte-vera, C. V, Kleisner, K. M., and Longo, C. 2013. A new

role for effort dynamics in the theory of harvested populations and data-poor stock

assessment, 1844: 1829–1844.

Vert-pre, K. A., Amoroso, R. O., Jensen, O. P., and Hilborn, R. 2013. Frequency and

intensity of productivity regime shifts in marine fish stocks. Proceedings of the

National Academy of Sciences of the United States of America, 110: 1779–84.

http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3562848&tool=pmcentr

ez&rendertype=abstract (Accessed 15 October 2014).

List of Figures

observed and predicted value were further from each other, blue circles represent

negative residuals, and red circles represent positive residuals.

Figure 5. Growth (blue), selectivity (black), and maturity (gray dotted) curves for three

life history types: (I) medium-lived spotted rose snapper (*Lutjanus guttatus*; asymptotic

length L_∞=64.6 cm, von Bertalanffy k=0.21, natural mortality M=0.43, and length at 50%

maturity l_{m50}=34 cm), (II) short-lived rabbitfish (*Siganus sutor*; L_∞=36.2 cm, k=0.87,

M=1.49, and l_{m50}=20.2 cm), and (III) long-lived Namibian hake (*Merluccius capensis*,

L_∞=100 cm, k=0.13, M=0.15, and l_{m50}=29.5 cm)

Figure 6. Simulated fishing mortality (red) and recruitment (blue) patterns over time, in

the deterministic (black line) and stochastic (color lines) scenarios. In the stochastic

scenarios, fishing mortality has lognormally distributed error with a standard deviation of

0.3, and recruitment has lognormally distributed error with a standard deviation of 0.6.

Figure 7. Relative error between the median MSY from the posterior distribution and the

true MSY from the simulated population under the long-lived hake life history type and

50 iterations of generated data. The boxplots show results for each data availability

scenario: catch, catch + abundance index, catch + mean length (ML), fishing mortality

pattern across the columns, and recruitment pattern across the rows, with stochasticity in

the fishing mortality (standard deviation = 0.3) and recruitment time series (standard

deviation = 0.6).

Figure 8. Relative error between the median MSY from the posterior distribution and the true MSY from the simulated population under the medium-lived snapper life history type and 50 iterations of generated data. The boxplots show results for each data availability scenario: catch, catch + abundance index, catch + mean length (ML), fishing mortality pattern across the columns, and recruitment pattern across the rows, with stochasticity in the fishing mortality (standard deviation = 0.3) and recruitment time series (standard deviation = 0.6).

Figure 9. Relative error between the median MSY from the posterior distribution and the true MSY from the simulated population under the short-lived rabbitfish life history type and 50 iterations of generated data. The boxplots show results for each data availability scenario: catch, catch + abundance index, catch + mean length (ML), fishing mortality pattern across the columns, and recruitment pattern across the rows, with stochasticity in the fishing mortality (standard deviation = 0.3) and recruitment time series (standard deviation = 0.6).

Figure 10. Relative error between the MSY posterior distribution and the true MSY including catch and an abundance index across 50 iterations of generated data for the long-lived hake life history type, under the endogenous fishing mortality pattern and constant recruitment. The median relative error across iterations is designated by RE in the upper-right corner. The gray boxplots indicate iterations where the true MSY was within the 95% credible intervals (proportion indicated by Cover in the upper-right

corner). Red boxplots indicate iterations where the true MSY was not within the 95%

credible intervals.

Figure 11. Relative error between the MSY posterior distribution and the true MSY

including catch and an abundance index across 50 iterations of generated data for the

long-lived hake life history type, under the constant fishing mortality pattern and constant

recruitment. The median relative error across iterations is designated by RE in the upper-

right corner. The gray boxplots indicate iterations where the true MSY was within the

95% credible intervals (proportion indicated by Cover in the upper-right corner). Red

boxplots indicate iterations where the true MSY was not within the 95% credible

intervals.

Figure 12. Comparison of the biomass-dynamics (top) and age-structured catch-MSY

(bottom) methods to estimate MSY when only catch data is available. The figures show

relative error between the estimates of MSY and the true MSY across 50 iterations of

generated data for the long-lived hake life history type, under endogenous fishing

mortality pattern, constant recruitment, and recruitment standard deviation 0.6. The

median relative error across iterations is designated by RE in the upper-right corner. The

gray boxplots (or points) indicate iterations where the true MSY was within the 95%

credible intervals (proportion indicated by Cover in the upper-right corner). Red boxplots

(or points) indicate iterations where the true MSY was not within the 95% credible
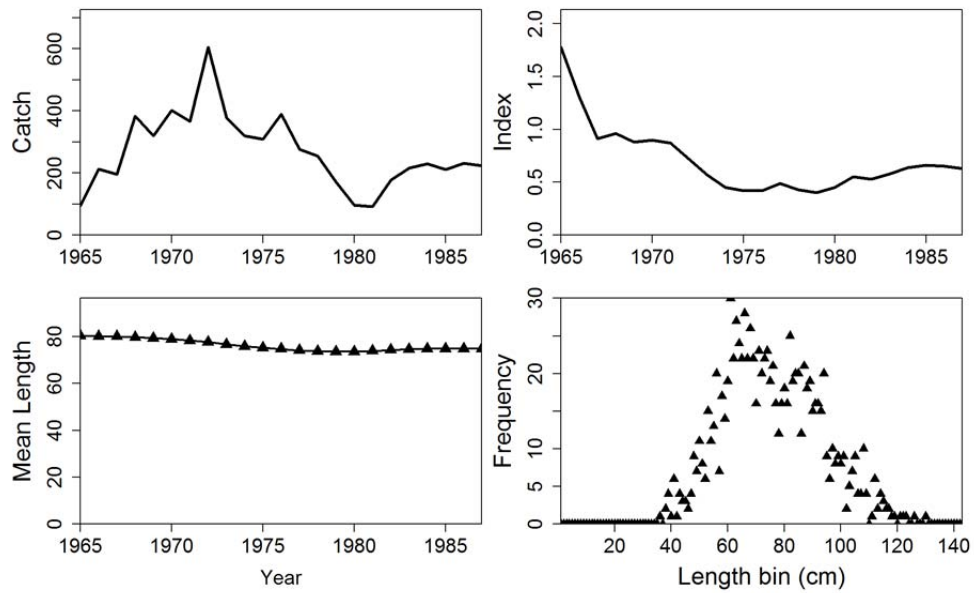
intervals.

Figure 13. Comparison of the biomass-dynamics (top) and age-structured catch-MSY (bottom) methods to estimate MSY when both catch and an abundance index were available. The figures show relative error between the estimates of MSY and the true MSY across 50 iterations of generated data for the long-lived hake life history type, under endogenous fishing mortality pattern and constant recruitment, with recruitment standard deviation of 0.6. The median relative error across iterations is designated by RE in the upper-right corner. The gray boxplots (or points) indicate iterations where the true MSY was within the 95% credible intervals (proportion indicated by Cover in the upper-right corner). Red boxplots (or points) indicate iterations where the true MSY was not within the 95% credible intervals.

Figures

Figure 1. Namibian hake data (catch and index; solid lines) and self-generated (triangles)

mean length time series and length composition data from the last year.
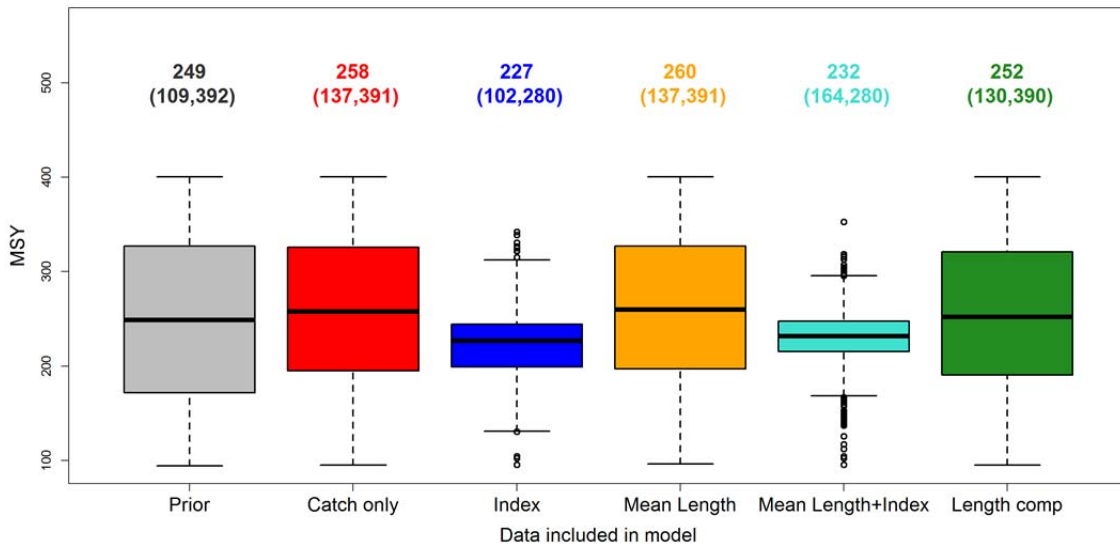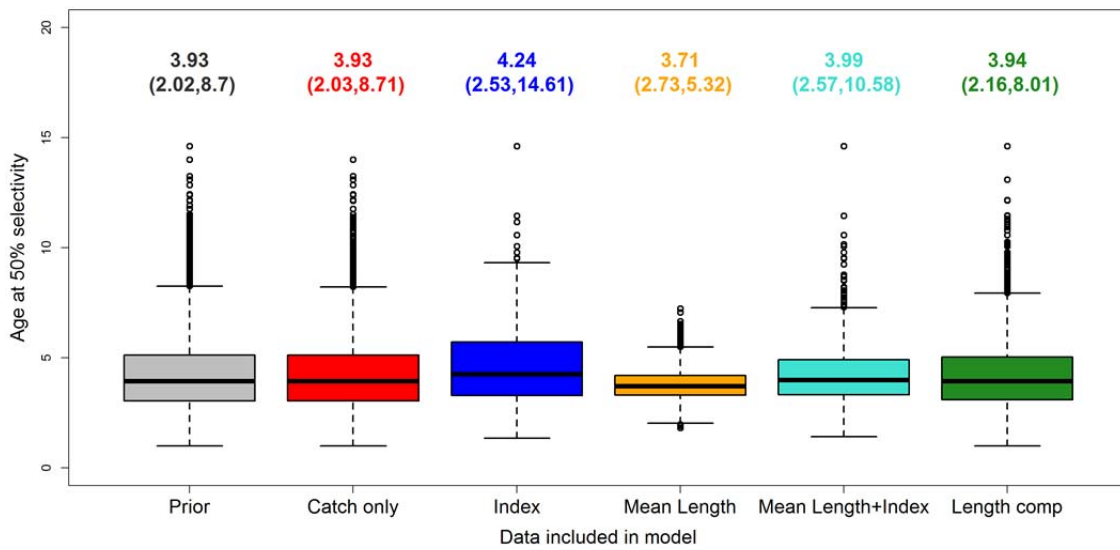
Figure 2. Boxplots of the prior on MSY (gray) and posterior distributions given the data provided to the Catch-MSY model: catch (red), catch + abundance index (blue), catch + mean length (orange), catch + mean length + abundance index (light blue), and catch + length composition (green). The median MSY from the prior (gray) and posterior distributions (colors) is above each boxplot, with the 95% lower and upper credible interval for each distribution in parentheses.

Figure 3. Boxplots of the prior on age at 50% selectivity (gray) and posterior distributions

given the data provided to the Catch-MSY model: catch (red), catch + abundance index

(blue), catch + mean length (orange), catch + mean length + abundance index (light blue),

and catch + length composition (green). The median age at 50% selectivity from the prior

(gray) and posterior distributions (colors) is above each boxplot, with the 95% lower and

upper credible interval for each distribution in parentheses.
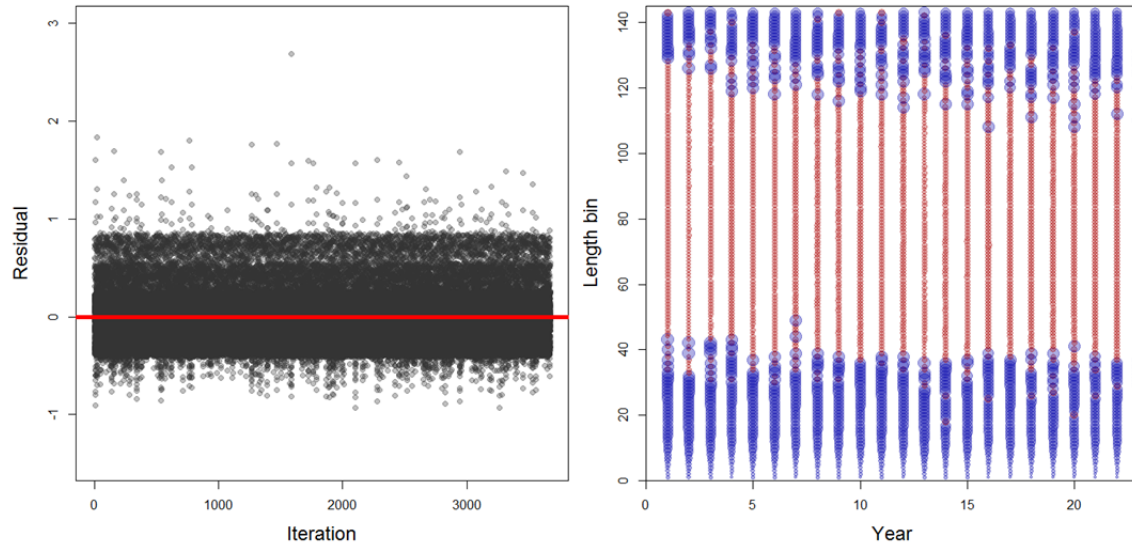
Figure 4. Residuals between the observed and predicted abundance index (left) and the

observed and predicted length composition (right). The residuals for the abundance index

represent the overall residuals across years for each iteration, and the residuals for the

length composition data show an example for one iteration, for each length bin across

years. A larger circle size for the length composition residuals represents where the

observed and predicted value were further from each other, blue circles represent

negative residuals, and red circles represent positive residuals.

Figure 5. Growth (blue), selectivity (black), and maturity (gray dotted) curves for three life history types: (I) medium-lived spotted rose snapper (*Lutjanus guttatus*; asymptotic length L_∞=64.6 cm, von Bertalanffy k=0.21, natural mortality M=0.43, and length at 50% maturity l_{m50}=34 cm), (II) short-lived rabbitfish (*Siganus sutor*; L_∞=36.2 cm, k=0.87, M=1.49, and l_{m50}=20.2 cm), and (III) long-lived Namibian hake (*Merluccius capensis*, L_∞=100 cm, k=0.13, M=0.15, and l_{m50}=29.5 cm)

Figure 6. Simulated fishing mortality (red) and recruitment (blue) patterns over time, in the deterministic (black line) and stochastic (color lines) scenarios. In the stochastic scenarios, fishing mortality has lognormally distributed error with a standard deviation of 0.3, and recruitment has lognormally distributed error with a standard deviation of 0.6.
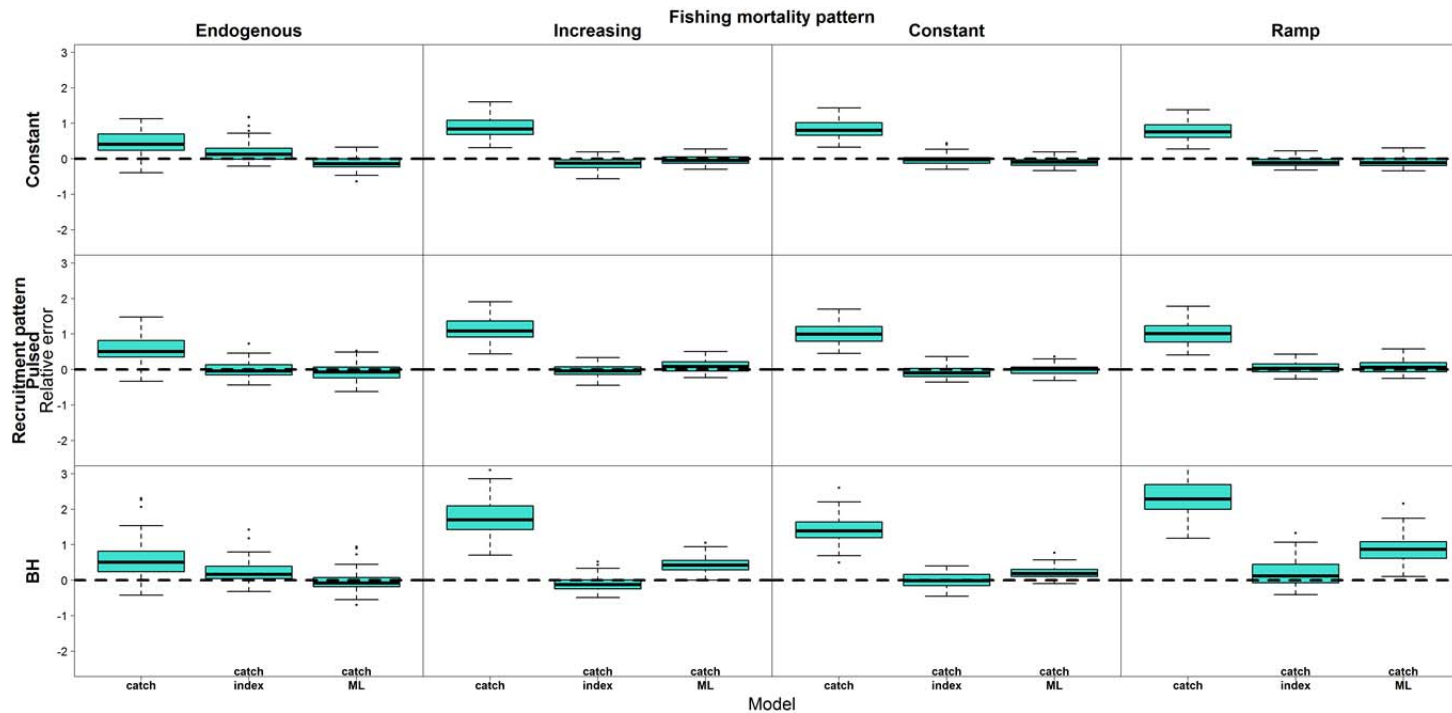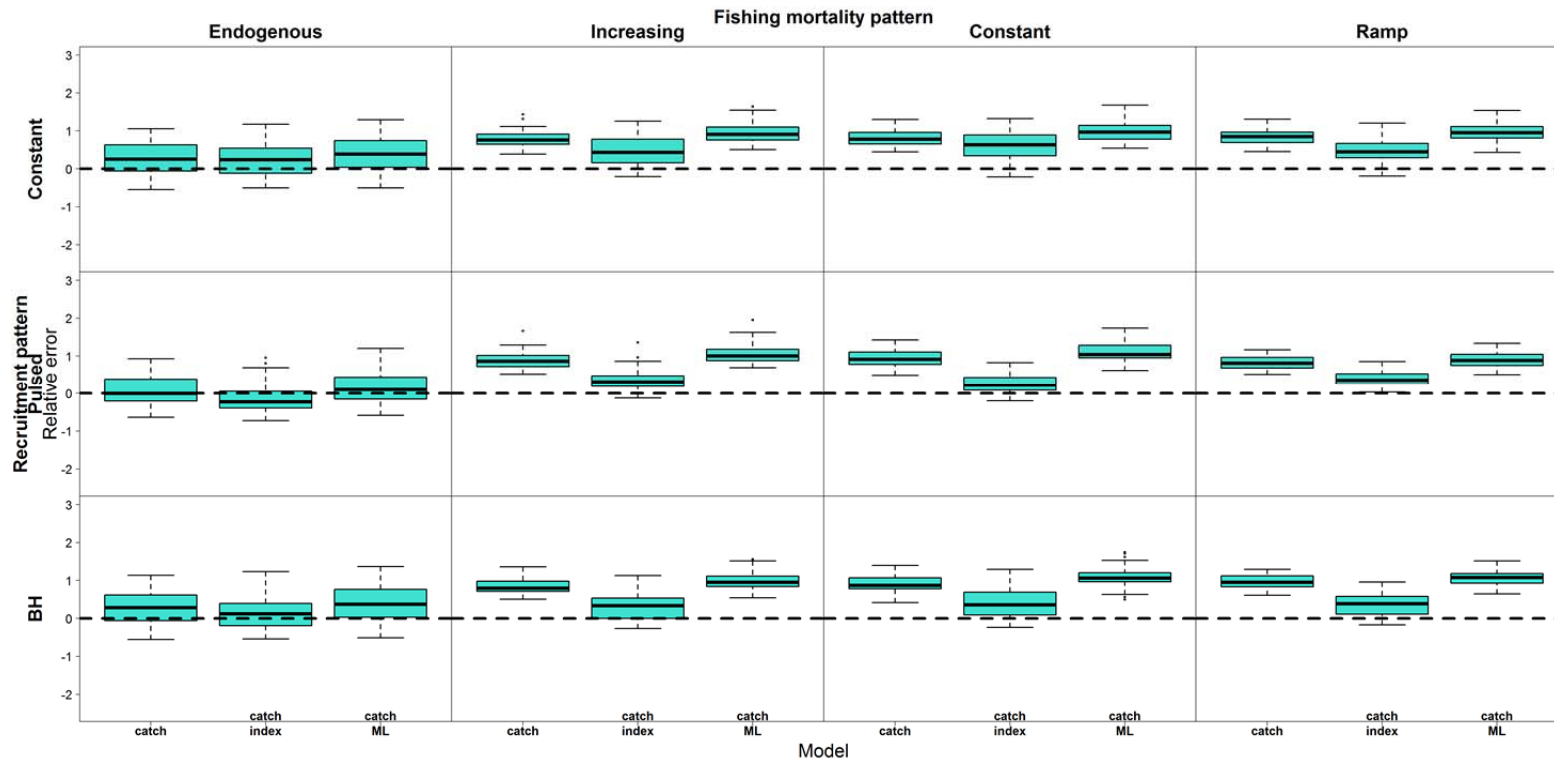
Figure 7. Relative error between the median MSY from the posterior distribution and the true MSY from the simulated population under the long-lived hake life history type and 50 iterations of generated data. The boxplots show results for each data availability scenario: catch, catch + abundance index, catch + mean length (ML), fishing mortality pattern across the columns, and recruitment pattern across the rows, with stochasticity in the fishing mortality (standard deviation = 0.3) and recruitment time series (standard deviation = 0.6).
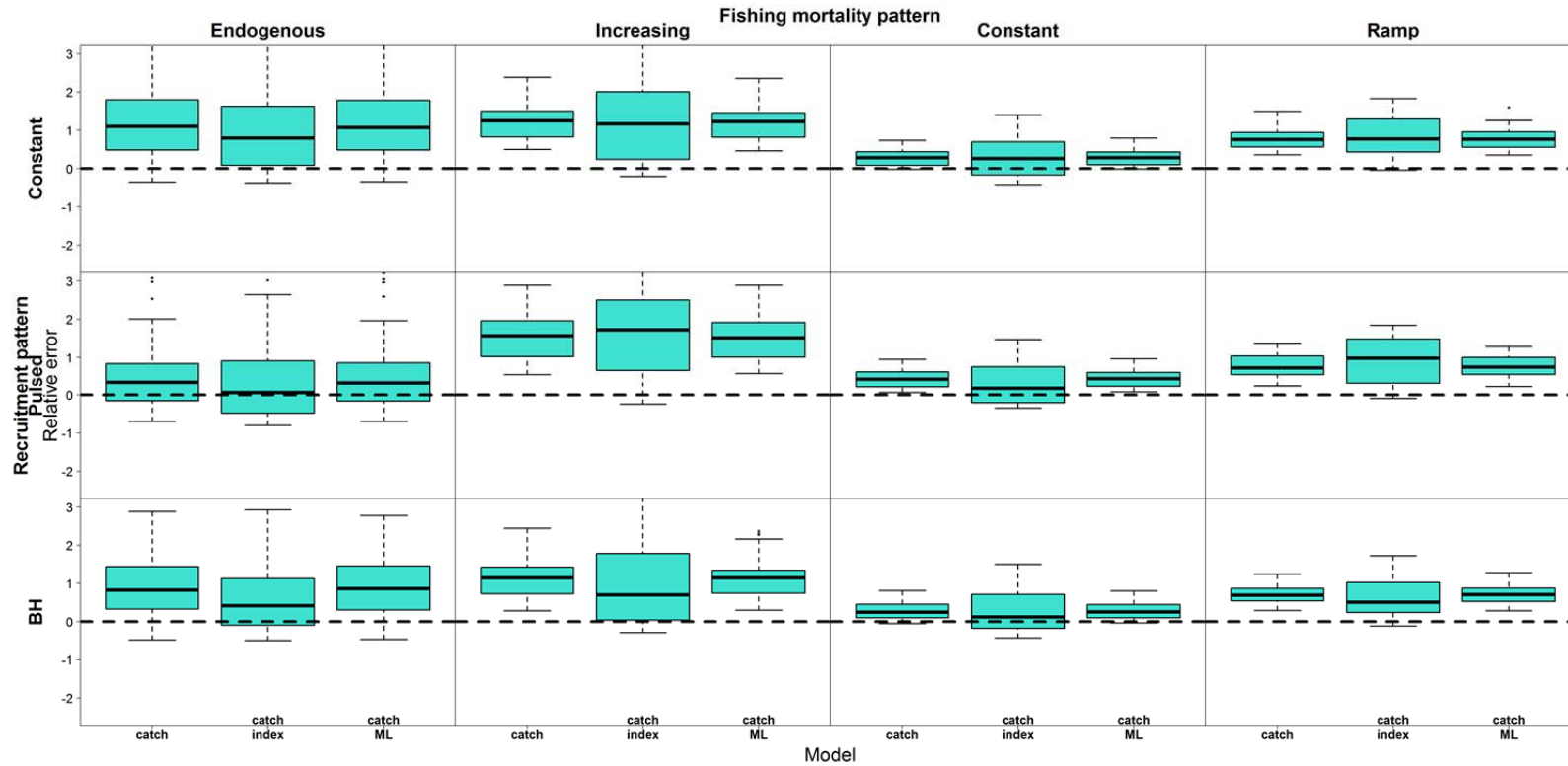
Figure 8. Relative error between the median MSY from the posterior distribution and the true MSY from the simulated population under the medium-lived snapper life history type and 50 iterations of generated data. The boxplots show results for each data availability scenario: catch, catch + abundance index, catch + mean length (ML), fishing mortality pattern across the columns, and recruitment pattern across the rows, with stochasticity in the fishing mortality (standard deviation = 0.3) and recruitment time series (standard deviation = 0.6).

Figure 9. Relative error between the median MSY from the posterior distribution and the true MSY from the simulated population under the short-lived rabbitfish life history type and 50 iterations of generated data. The boxplots show results for each data availability scenario: catch, catch + abundance index, catch + mean length (ML), fishing mortality pattern across the columns, and recruitment pattern across the rows, with stochasticity in the fishing mortality (standard deviation = 0.3) and recruitment time series (standard deviation = 0.6).
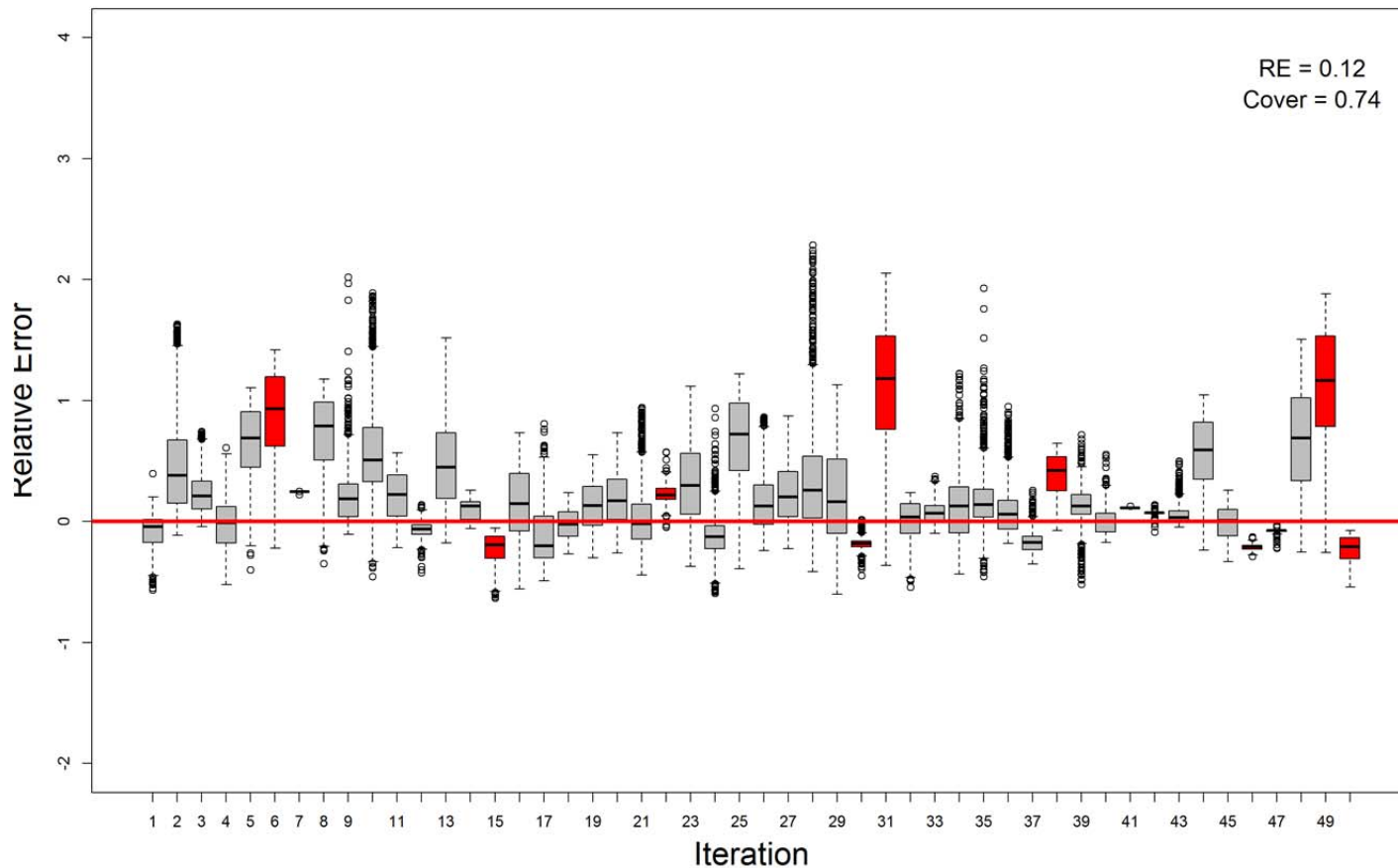
Figure 10. Relative error between the MSY posterior distribution and the true MSY including catch and an abundance index across 50 iterations of generated data for the long-lived hake life history type, under the endogenous fishing mortality pattern and constant recruitment. The median relative error across iterations is designated by RE in the upper-right corner. The gray boxplots indicate iterations where the true MSY was within the 95% credible intervals (proportion indicated by Cover in the upper-right corner). Red boxplots indicate iterations where the true MSY was not within the 95% credible intervals.
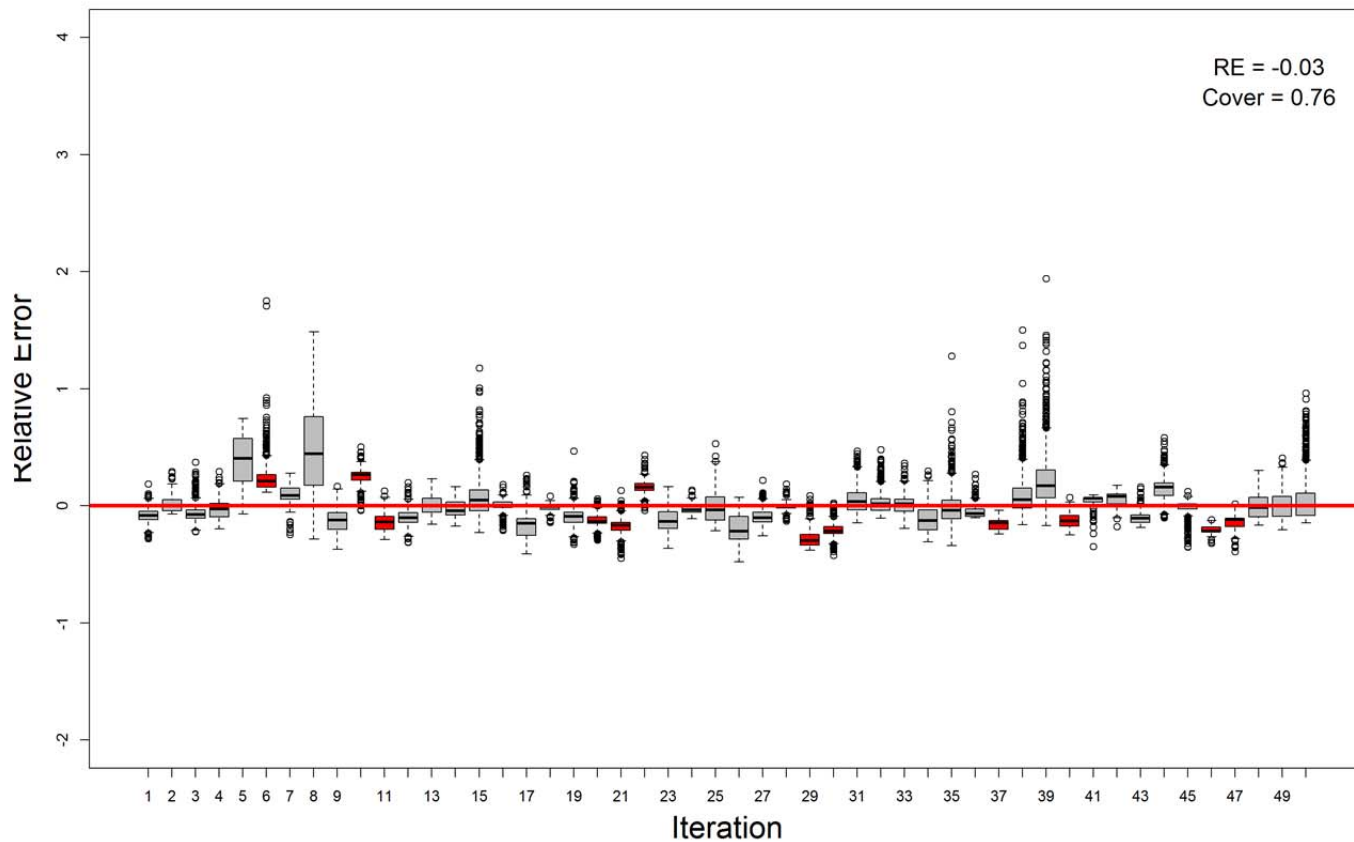
Figure 11. Relative error between the MSY posterior distribution and the true MSY including catch and an abundance index across 50 iterations of generated data for the long-lived hake life history type, under the constant fishing mortality pattern and constant recruitment. The median relative error across iterations is designated by RE in the upper-right corner. The gray boxplots indicate iterations where the true MSY was within the 95% credible intervals (proportion indicated by Cover in the upper-right corner). Red boxplots indicate iterations where the true MSY was not within the 95% credible intervals.
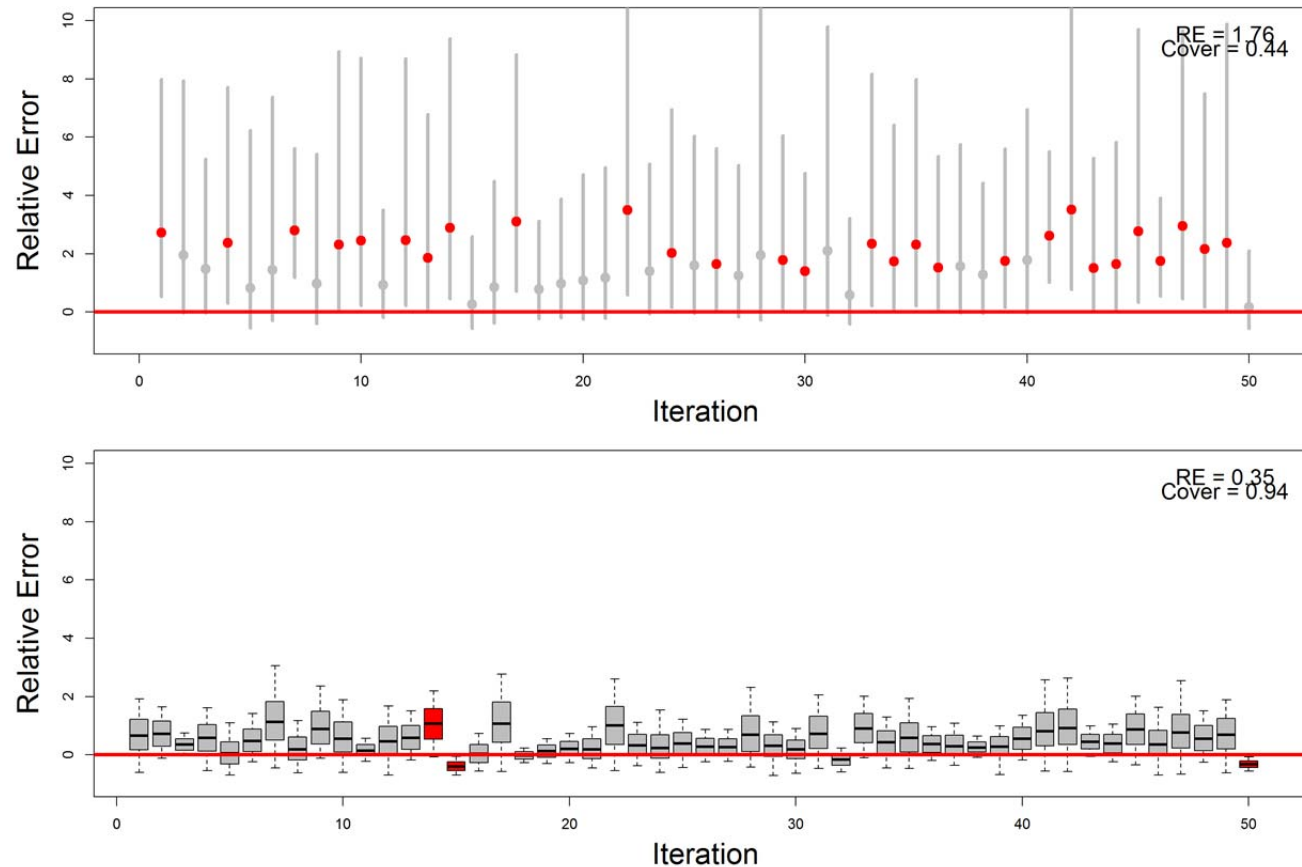
Figure 12. Comparison of the biomass-dynamics (top) and age-structured catch-MSY (bottom) methods to estimate MSY when only catch data is available. The figures show relative error between the estimates of MSY and the true MSY across 50 iterations of generated data for the long-lived hake life history type, under endogenous fishing mortality pattern, constant recruitment, and recruitment standard deviation 0.6. The median relative error across iterations is designated by RE in the upper-right corner. The gray boxplots (or points) indicate iterations where the true MSY was within the 95% credible intervals (proportion indicated by Cover in the upper-right corner). Red boxplots (or points) indicate iterations where the true MSY was not within the 95% credible intervals.
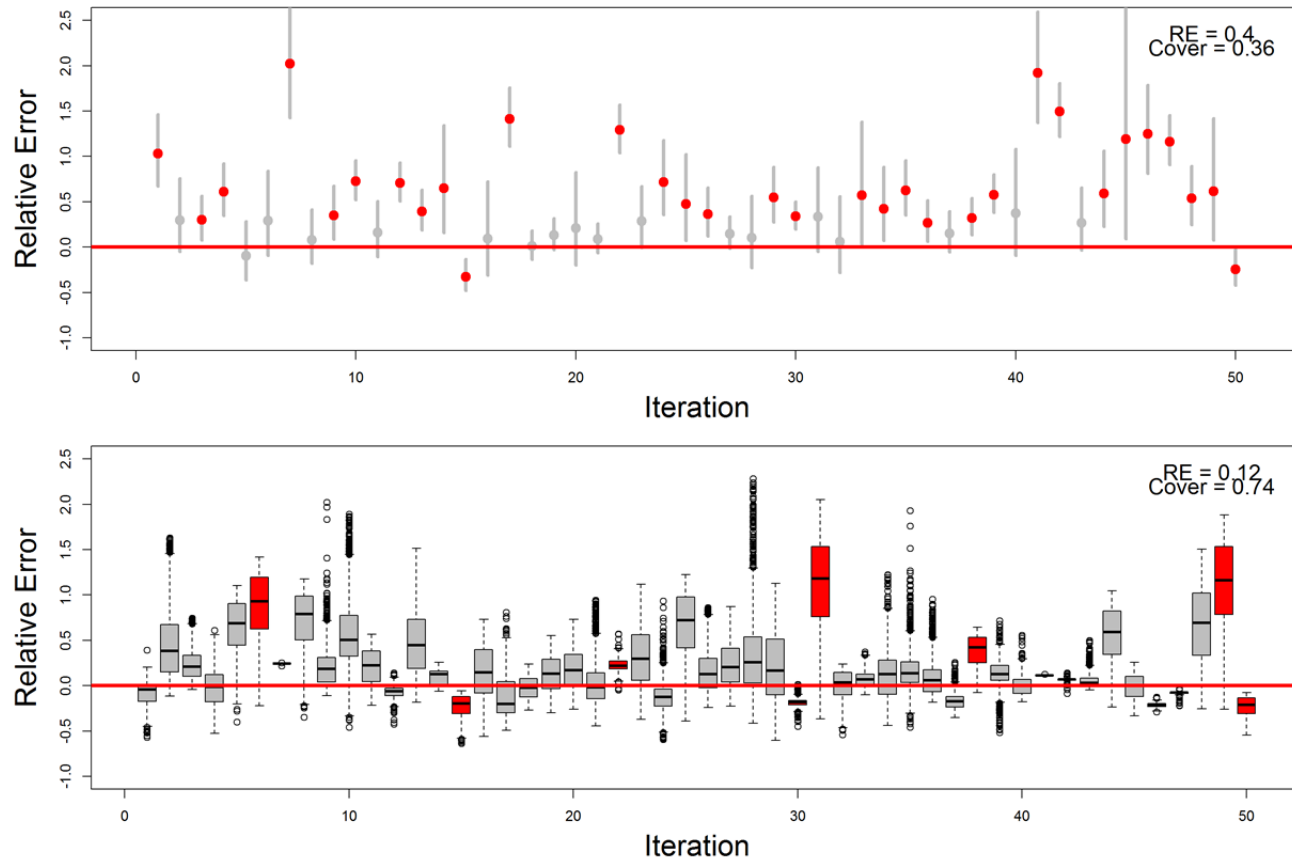
Figure 13. Comparison of the biomass-dynamics (top) and age-structured catch-MSY (bottom) methods to estimate MSY when both catch and an abundance index were available. The figures show relative error between the estimates of MSY and the true MSY across 50 iterations of generated data for the long-lived hake life history type, under endogenous fishing mortality pattern and constant recruitment, with recruitment standard deviation of 0.6. The median relative error across iterations is designated by RE in the upper-right corner. The gray boxplots (or points) indicate iterations where the true MSY was within the 95% credible intervals (proportion indicated by Cover in the upper-right corner). Red boxplots (or points) indicate iterations where the true MSY was not within the 95% credible intervals.