Evaluating the performance of length and catch models in the Stock Synthesis framework

U.S. West Coast fisheries vary in data quantity and quality, which affect the amount of information used to inform stock status. Integrated stock assessment frameworks such as Stock Synthesis (SS; Methot and Wetzel 2015) allow for flexible approaches to data treatment. This ranges from a full complement of catch, abundance index and biological data (mainly, length and/or age compositions) to more limited data using only one of the above information inputs.

The Scientific and Statistical Committee (SSC) of the Pacific Fishery Management Council has established a stock assessment categorization system that reflects both data availability and model treatment. There are currently 3 broad assessment categories, each with an associated Allowable Biological Catch (ABC) buffer defined by model uncertainty, and a default uncertainty that increases as data availability in the category decreases. This ABC buffer defines the percent reduction from the overfishing limit (OFL).

Category 1 assessments estimate the overfishing limit (OFL) using a mixture of data types including total catch, abundance indices or surveys, and length and/or age composition data. On the other side of the data spectrum, category 3 ("data-limited") stocks rely on data-limited methods that currently only estimate the OFL using catch, life history parameters and a prior on relative stock abundance in a specified year of the time series (Dick and MacCall 2011; Cope 2013). Current data-limited methods for category 3 stocks do not include abundance indices or biological compositions that would inform stock status over time (Thorson and Cope 2015).

To date, approved category 2 ("data-moderate") assessments, that were specifically developed using data-moderate approaches for West Coast groundfish stocks, combined catch-based methods with an abundance index, excluding biological composition data (Cope et al. 2013; Wetzel and Punt 2015). These models may also include a prior on relative stock abundance that should be updated from the data (if not, then they are functionally a category 3 assessment). Data-moderate assessments currently rely on adaptive-importance resampling (Cope et al. 2013), which can be time intensive to run and demonstrate aberrant behavior when the upper scale of the population is difficult to resolve (Thorson and Cope 2019). Future modifications to this approach may be able to speed up application and look to avoid the problems with scale estimation (Monnahan et al. 2019).

Abundance indices, as suggested above, are not always informative about the population. In some cases, category 1 stock assessment results may show little change to the reference model outputs with the exclusion of the abundance index (Wetzel et al. 2017). These assessment sensitivity results observationally suggest cases where the category 1 stock assessment relies mainly on the catch and compositional data only, and thus offer evidence that catch and length only models can be adequate to inform management metrics. Length compositions are a key input to most stock assessments, as the information is easier to obtain than ages, and thus a main source of recruitment and spawning potential information. It is also typically believed many fisheries use length-selective gear, therefore lengths are essential to estimating gear selectivity and fishing intensity.

The amount of length composition needed, though, is not well-understood, thus catch and length only assessments are not currently explicitly recognized as a possible option for category 2 stock assessment application. The original data-limited methods review in 2011 declined to evaluate catch and length only models in order to focus on catch and index only models (PFMC 2011). Such models are specifically important considerations for many stocks that have catch and compositional data but lack a reliable index. Efforts to set catch limits for these stocks has subsequently applied category 3 methods, ignoring any information potentially included in available compositional data, thus leading to a large buffer between the OFL and ABC. Lastly, it may be of strategic interest to perform length-based category 2 assessments for stocks that are minor targets or are believed to be at high abundance, rather than spending limited resource capacity to perform a full assessment.

The following analyses look at the asymptotic behavior in models that use catch and length data in Stock Synthesis (SS), the predominant modelling framework for West Coast groundfishes. The first section uses simulation testing to evaluate the performance of length and catch models under a variety of life history and data permutations. The second section takes current benchmark stock assessments and removes data to explore how sensitive our current benchmarks are to the removal of different data types, and specifically to the amount of length composition data available. Both approaches provide unique ways to evaluate the use of different applications of length and catch models for consideration as an additional option as a category 2 stock assessment.

Section 1: Simulation testing stock assessment models with catch and length data

Merrill Rudd, Chantel Wetzel, and Jason Cope

Introduction

Data-limited assessment methods are often structured to require specific, limited inputs leading to high uncertainties in population estimates and requiring a large buffer between the estimated OFL and ABC. Data-limited assessment methods currently used for category 3 stocks typically require a prior on relative stock abundance in a specified year of the time series, which is a counter-intuitive input as relative stock abundance is a desired output from a stock assessment model (Dick and MacCall 2011; Cope 2013). Thorson and Cope (2015) demonstrated through Catch Curve Stock Reduction Analysis (CC-SRA) that at least one year of age composition could replace the need for the prior on relative stock abundance and inform stock status over time using information contained in the compositional data. Age data is resource-intensive to collect, and thus length data are much more common. Age and length composition data theoretically both hold information on recruitment and spawning potential information, but length compositions, in the absence of age data, often require assumptions about growth.

Rudd and Thorson (2018) considered the concepts used by CC-SRA in a more data-limited context. When total catch data cannot be accurately quantified, length composition and assumptions about biology are used to estimate spawning potential. However, most data-limited, length-based stock assessment methods assume recruitment is in equilibrium with the potential for biases if this assumption is violated. Rudd and Thorson (2018) developed a Length-based Integrated Mixed Effects (LIME) model that only required length composition and assumptions about biological parameters to estimate stock status over time, relaxing the assumption of equilibrium recruitment by treating recruitment as a random effect. The 'integrated' aspect of LIME allowed for the inclusion of catch and an abundance index, if available, as a way of estimating population scale with at least one year of catch data or including an index to inform stock status as monitoring programs improved. LIME was developed to run without catch data by estimating fishing mortality rates as fixed effects, assuming information about fishing mortality, recruitment, and selectivity are contained within the length compositions.

LIME was initially considered a potential option to assess more West Coast stocks with length and catch data only. However, LIME development was focused mainly on length-only scenarios, resulting in unreasonably long run times when including catch time series typical of West Coast stocks. Simulation-testing of LIME demonstrated that length data alone can theoretically inform recruitment, fishing mortality, and selectivity (Rudd and Thorson 2018). However, in order to adequately examine the long catch time series and length data scenarios typical of West Coast stocks, SS proved to be a more useful tool and already the predominant modeling framework for West Coast groundfishes.

This study provides a proof-of-concept that SS can be used with catch and length compositions to estimate stock status and catch limits. We considered various scenarios of life history, number of years of length composition, sample size of length composition, and recruitment variability under a fishing mortality trajectory representative of many U.S. West Coast groundfish stocks.

Methods

Operating model

We used SS as an operating model to simulate "true" populations and generate data, based on the approach developed in the R package *ss3sim* (Anderson et al. 2014). We simulated populations over 100 years. Scenarios varied in the number of years and sample size of length data, standard deviation of recruitment deviates, and the magnitude-- but not pattern-- of fishing mortality. Selectivity was constant over scenarios and time, using the double-normal selectivity function to represent logistic selectivity assuming a peak selectivity at 42 cm.

Life history scenarios

We considered four life history scenarios that varied in longevity (e.g., how long they live) and individual growth rate (e.g., how much of their lives are spent at their maximum size). The shorter-lived life history type had a natural mortality rate (M) of 0.18 for an approximate maximum age of 30 years, while the longer-lived life history type had a M of 0.09 for an approximate maximum age of 60 years (Hamel, 2015). We considered slower- and faster-growing options for the shorter- and longer-lived life history types by adjusting the von Bertalanffy growth coefficient (k). We chose values of k associated with reaching asymptotic length at 90% of maximum age (slower-growing) and 50% of maximum age (faster-growing). We confirmed that the M/k ratios were close to 1.0 for slower-growing life history types and close to 0.60 for faster-growing life history types, as these M/k ratios are representative for category 1 stocks assessed through 2018 (Appendix 1). We kept asymptotic length constant at 55 cm for all life history types, and assumed length at 50% maturity was equal to 66% of the asymptotic length (i.e., 36.3 cm). Parameter values used for each of the four life history types are available in Table 1 and growth curves are shown in Figure 1.

Fishing mortality scenario

After reviewing fishing mortality rate time series from U.S. west coast stocks, we identified a general pattern of low exploitation prior to World War II, then increases in exploitation rate after World War II until the 1980s or 1990s. After remaining at a high exploitation rate in the 1990s, the exploitation rate declines through the present (Figure 2).

We developed F scenarios based on a common ratio of F/F_{MSY}, then scaled F based on the F_{MSY} for each life history type. The F/F_{MSY} ratio remained relatively low for the first 25 years, increasing from 0.01 to 0.05, then increased more rapidly over the next 30 years from 0.05 to 2. The F/F_{MSY} ratio remained at 2 for five years, then decreased from 2 to 0.6 over the next 20 years, and remained at 0.6 for the last 20 years of the time series. The comparison of our simulated time series of F/F_{MSY} with those from West Coast stocks is shown in Figure 2.

We then obtained the input fishing mortality rate by multiplying the F/F_{MSY} by F_{MSY} specific to each life history scenario. F_{MSY} was calculated by finding the constant F value that maximizes long-term catch. For both longer-lived life histories the F_{MSY} was 0.08, whereas F_{MSY} was 0.17 for the shorter-lived, faster-growing life history type, and 0.16 for the shorter-lived, slower-growing life history type (Table 1).

The catch time series was calculated within the operating model based on the input fishing mortality rate time series and scale of the population.

Recruitment scenarios

Recruitment followed an underlying Beverton-Holt stock-recruit curve with steepness (h) equal to 0.7. The "low variability" recruitment scenario assumed lognormally-distributed recruitment deviates with mean 0 and standard deviation 0.4, representative of many West Coast species. The "high variability" recruitment scenario assumed a standard deviation of 0.8 (Appendix 1). Each simulation replicate varied in the recruitment deviates drawn from a normal distribution. The log of equilibrium recruitment was assumed constant at 10.0.

Data generation

The true population was determined based on the input fishing mortality time series (Figure 3, input recruitment deviations, and life history information. Simulation replicates varied by the input recruitment deviates; the time series of recruitment deviates for each simulation replicate was identical across life history types (Figure 4). True population values were calculated by running SS without calculating the Hessian matrix for standard deviations. A unique "true" population was associated with each life history type, recruitment variability scenario, and simulation replicate (Figure 5).

Each true population determined by life history type, recruitment variability, and simulation replicate was then subject to data availability scenario tests based on number of samples of length data annually and number of years of length data included in the model. We generated observation data from the operating model population by sampling from the expected data values available in the second section of the "data.ss_new" file.

To test the ability of SS to estimate key parameters of interest with catch and length data alone, we included “data-rich” and “perfect” scenarios. The “data-rich” scenario included an abundance index and age compositions with sample size of 100 ages per year, in addition to length composition that was known perfectly with input sample size of 1,000 over all 100 years. The “perfect” length data scenario excluded the index and age data and included only catch and the length composition data known perfectly over 100 years. These “data-rich” and “perfect” scenarios were used to make sure that any biases in scenarios with catch and length data only were due to the limitation of data types or recruitment variability, rather than structural inconsistencies between the observation and estimation models.

For all other length data scenarios of interest, we sampled from the length composition with either 200 samples (representing moderate sample size) or 50 samples (representing limited sample size) using a multinomial distribution. We tested an additional scenario where length data sample size decreased over time, specifying 200 samples prior to year 88 and 50 samples from year 88-100.

For each sample size scenario, we also explored the number of years of length data included in the model. From the sampled data, we considered the inclusion of the final a) 75 years, b) 20 years, c) 10 years, and d) 1 year of the sampled length data. We tested the decline in sample size over time only with the 20-year length data scenario. The approach of subsetting the number of years after the data were generated allowed us to directly compare the number of years of length data included in the model, rather than any stochasticity associated with re-sampling the length composition for each independent scenario.

For all scenarios, we assumed the catch was known without error based on the input fishing mortality time series.

Estimation model

We used SS 3.30.14 to conduct the simulation testing. Estimated parameters included the recruitment deviations, log of equilibrium recruitment log(R_0), and selectivity parameters governing the shape and peak of the left side of the double-normal selectivity curve. We fixed natural mortality, growth parameters, steepness, and the recruitment standard deviation to their true values for this set of model runs.

Recruitment estimation scenarios

The first year of estimated recruitment deviates was the maximum age subtracted from the first year of length data, starting in year 1 if length data is available before year 29. To determine the final year of estimated recruitment deviates, we identified the age associated with 5% selectivity for each life history type, and subtracted that age from the final year in the model. For example, the short-lived, slow-growing life history type reached 5% selectivity at age 3, so the final year of estimated recruitment deviates was 97 out of 100 (Table 1).

To prevent biased estimates of spawning stock biomass (SSB) in the early years of the time series, we allowed for the estimation of early recruitment deviates starting 30 years before the first year of removals by the fishery.

We used the iterative procedure developed by Methot and Taylor (2011) to account for bias adjustment in estimated recruitment deviates. We first ran SS calculating the Hessian matrix, then estimated the bias ramp parameters based on the model estimates from the first run. We then input the bias ramp parameter estimates and re-ran the model without calculating the Hessian matrix to speed up the simulation model runs. We saved both the unadjusted and bias-adjusted results to compare the bias-adjusted model estimates across scenarios. Next steps could estimate the Hessian matrix on the second run to explore the characterization of uncertainty of individual model runs for length data scenarios.

Performance metrics

We determined the convergence rate of each scenario defined by the maximum final parameter gradient less than 1.0 and the maximum likelihood estimate of $\log(R_0)$ less than 12.0 (e.g., failed model convergence that would likely be due to bad starting value selected via jitter rather than in inability to estimate the population size). From the converged runs, we calculated relative error, *(estimated - true)/true,* for key population parameters for each of 100 simulation replicates. We defined bias as the median relative error (MRE) and precision as the median absolute relative error (MARE; Ono et al., 2012).

Results

The general results reported in this section considered recruitment standard deviation of 0.8. Low variability (σ_R = 0.4) results are compared to the high variability scenario in this section with more detailed results in Appendix 2.

SS models converged at high rates across scenarios. The highest rates of non-convergence occurred for scenarios with a single year of length composition, particularly with only 50 samples of lengths per year (Table 2). One simulation replicate with declining sample size across the 20-year time series did not converge for the short-lived, fast-growing life history type because the maximum gradient for model parameters was greater than 1.0 (maximum final gradient = 1.43; Table 2).

The "data-rich" scenario, which included an abundance index, age composition, and perfectly-known length compositions over 100 years, confirmed that SS estimated unbiased key population parameters. The "perfect information" scenario, excluding the abundance index and age composition but the length composition was known perfectly for all 100 years, confirmed that SS estimated unbiased and precise key population parameters across 100 simulation replicates with a high amount of length data under low variability (Appendix 2). Bias increased in the "perfect information" scenario with high recruitment variability for the long-lived, fast-growing life history type estimates of the fraction of unfished biomass and the fast-growing life history types' estimation of the OFL (Table 3, Table 4). The unbiased data-rich scenarios under high recruitment variability and unbiased perfect information scenarios under low recruitment variability led us to assume that any breakdown in bias or precision under alternative sampling scenarios was due to the limited number of samples, number of years of length data included in the model, or higher recruitment variability.

The number of years and sample size of length data did not generally impact bias in key population parameters (Table 3, Table 4, Figure 6, Figure 7). Bias was also not necessarily affected by higher recruitment variability (Figure 8). Under high recruitment variability, some scenarios with perfect information or 75 years of length data were more biased than scenarios with fewer years of data (Table 3, Table 4). The long-lived, slow-growing life history type estimated bias greater than 15% in both the fraction of unfished biomass and OFL for scenarios with 50 samples of length data under high recruitment variability (Table 3, Table 4, Figure 7). Under low recruitment variability, bias was generally greater than 25% with a single year of length data with 50 samples across life history scenarios. An exception to this pattern was for the long-lived, slow-growing life history type, where the bias was less than 10% with a single year of length data and 50 samples, but was greater than 50% with a single year of length data with 100 samples (Appendix 2).

Precision decreased with the number of years of length data, sample size of the length data, and higher recruitment variability (Table 5, Table 6, Figure 6, Figure 7, Figure 8). A higher MARE indicates lower precision. The strongest increase in MARE occurred when paring down from 10 years to 1 year of length data (Table 5, Table 6). Under the high variability scenario with only one year of data, the relative error in fraction of unfished biomass was greater than 1.0 for 8-15% of simulation replicates with 200 samples and 14-23% of simulation replicates with 50 samples. In comparison, fewer than 10% of model runs had relative error greater than 1.0 with 50 samples per year for more than one year, and fewer than 2% of scenarios had relative error greater than 1.0 with 200 samples per year under high variability (Table 7). The proportion of simulation replicates with relative error greater than 1.0 decreased with low recruitment variability, but was still highest with only one year of length data (Appendix 2).

The decline in sample size from 200 samples to 50 samples over 20 years of length data generally performed similarly to the scenarios with a constant 50 samples per year in terms of both bias and precision. In some scenarios, the bias associated with the decline in sample size fell somewhere between the scenarios with a constant sample size of 200 and 50. However, bias was consistently less than 16% for all life history scenarios with declining sample size over 20 years of data under high variability (with the exception of the long-lived, slow-growing estimates of the OFL, where bias was 22%; Table 3, Table 4, Figure 9). Bias was less than 10% for all life history scenarios with declining sample size over 20 years of length data under low recruitment variability (Appendix 2).

The long-lived, slow-growing life history type under high recruitment variability was the only life history scenario to have consistent biases across the study. Under the high variability scenario, the long-lived, slow-growing life history type was biased greater than 18% with a low sample size of length data (Table 3, Table 4). With perfect length data, the fast-growing life history types had slightly higher bias than the slow-growing life history types under the high variability scenario. Under low recruitment variability, this pattern held true for the final year fraction unfished, while the longer-lived life histories had slightly higher bias in the OFL estimate than the slow-growing life history types (Appendix 2). Other than the bias associated with long-lived, slow-growing life history type with a low sample size, any patterns apparent from the perfect information scenario broke down with fewer years of length data and lower sample size.

The recruitment bias adjustment did indeed decrease bias in estimates of recruitment in the final year of data. However, bias did not necessarily decrease with bias adjustment in the fraction of unfished biomass and the OFL. Bias decreased for population parameters in the final year, besides recruitment, for only about half of the scenarios with bias adjustment.

Discussion

This simulation study demonstrated we can expect unbiased estimates of key population parameters on average when including only catch and length data in SS. The probability of an accurate parameter estimate (i.e., precision) for any given stock assessment generally increases with more years of length composition data, a higher sample size of length data, and for stocks with lower recruitment variability.

We found that models with catch and length data would be useful for stocks with both high and low recruitment variability. Higher recruitment variability led to higher MARE but not necessarily higher bias. For example, models were relatively unbiased with a single year of length data and high recruitment variability, while scenarios with a single year of length data were biased with low recruitment variability. One potential reason for this improvement may be that higher recruitment variability is associated with larger recruitment pulses (or lack thereof), which may provide sharper contrast in the length data thus making them more informative.

The only major difference in performance across life history scenarios was for the long-lived, slow-growing life history type under high recruitment variability. We were expecting models with catch and length data alone to be less informative for fast-growing species that reach their asymptotic length more quickly; however, this expected pattern did not play out very clearly. The long-lived, slow-growing life history type used in this simulation had an *M/k* ratio close to 1.0 (Appendix 1). However, the longer-lived rockfish on the U.S. West Coast have very low *M/k* ratios, and thus may be more similar to the long-lived, fast-growing life history type which generally performed well under both low and high recruitment variability with limited years and sample size of length data (Appendix 1).

The unbiased results are generally insensitive to the number of years and sample size of length compositions. On the other hand, uncertainty is more sensitive to the number of years and sample sizes of length data. We generally observed increased precision in estimates of key population parameters with more years of length data and higher sample sizes. Next steps should explore how bias and precision change between 2-10 years of length data to examine the minimum number of years of length data required, particularly for low recruitment variability scenarios. Next steps should also examine the uncertainty within model runs to assess whether more years or higher sample sizes of length data lead to higher certainty in key population parameters, which may have an effect on the buffer size required between the estimated OFL and ABC.

These simulations represent simplified versions of fish stocks with only a single fleet operating in a homogenous area. Length compositions must be representative of the entire fishery to accurately inform stock status, and this task becomes more complicated with multiple fleets and selectivities. Section 2 explores the performance of SS models with catch and length in relation to benchmark West Coast groundfish assessments with multiple fleets.

Tables

Table 1. Parameter values used to develop life history scenarios in the operating model.

		Life history			
Parameter	Description	Short-lived, slow-growing	Short-lived, fast-growing	Long-lived, slow-growing	Long-lived, fast-growing
Amax	Maximum age (years)	30	30	60	60
M	Natural mortality (1/year)	0.180	0.180	0.090	0.090
K	Von Bertalanffy growth coefficient (1/year)	0.170	0.300	0.088	0.150
L_{inf}	Asymptotic length (cm)	55.0	55.0	55.0	55.0
t_0	Length at age = 0 (cm)	-1	-1	-1	-1
L_{50}	Length at 50% maturity (cm)	36.3	36.3	36.3	36.3
M/k	Ratio of *M* to *k*	1.06	0.60	1.02	0.60
F_{MSY}	Fishing mortality expected to produce maximum sustainable yield (MSY)	0.16	0.17	0.08	0.08
H	Steepness	0.7	0.7	0.7	0.7
last_recdev	Last year of estimated recruitment deviates based on age at 5% selectivity	97	98	94	96

Table 2. Number of model runs that did not converge out of 100 simulation replicates with recruitment standard deviation equal to 0.8 with bias-adjusted recruitment, based on the criteria that the maximum final gradient must be less than 1.0 and the estimated log(R_0) less than 12.

Length data scenario		Life history scenario			
Sample size	Number of years	Short-lived, slow-growing	Short-lived, fast-growing	Long-lived, slow-growing	Long-lived, fast-growing
N = 1000 (perfectly specified, with index + ages)	100	0	0	0	0
N = 1000 (perfectly specified)	100	0	0	0	0
N = 200	75	0	0	0	0
N = 200	20	0	0	0	0
N = 200	10	0	0	0	0
N = 200	1	7	9	1	1
N = 50	75	0	0	0	0
N = 50	20	0	0	0	0
N = 50	10	0	0	0	0
N = 50	1	10	13	2	5
Decline from 200 samples (year $\leq$ 87) to 50 samples (year > 87)	20	0	1	0	0

Table 3. Median relative error (i.e., bias) of the final year fraction of unfished biomass across 100 simulation replicates by life history type and length data scenario with recruitment standard deviation = 0.8 with bias-adjusted recruitment. Cells are shaded green for bias less than 10%, yellow less than 20%, and red greater than 20%.

Length data scenario		Life history scenario			
Sample size	Number of years	Short-lived, slow-growing	Short-lived, fast-growing	Long-lived, slow-growing	Long-lived, fast-growing
N = 1000 (perfectly specified, with index + ages)	100	-0.0071	-0.0225	-0.0423	-0.0356
N = 1000 (perfectly specified)	100	0.0802	0.0834	0.0418	0.1065
N = 200	75	-0.1013	-0.1097	0.0127	0.0178
N = 200	20	-0.0308	-0.0743	0.0117	0.0628
N = 200	10	-0.0203	-0.0585	-0.0704	0.0136
N = 200	1	-0.0475	-0.1230	0.0408	0.0734
N = 50	75	0.0584	-0.0601	0.2844	0.0426
N = 50	20	0.0582	0.0236	0.2091	0.0047
N = 50	10	0.0713	0.1004	0.1939	-0.0572
N = 50	1	-0.0656	0.0552	0.1684	-0.1094
Decline from 200 samples (year $\leq$ 87) to 50 samples (year > 87)	20	0.0446	-0.1199	0.1558	-0.0116

Table 4. Median relative error (i.e., bias) of the overfishing limit (OFL) in the first forecast year across 100 simulation replicates by life history type and length data scenario with recruitment standard deviation = 0.8 with bias-adjusted recruitment. Cells are shaded green for bias less than 10%, yellow less than 20%, and red greater than 20%.

Length data scenario		Life history scenario			
Sample size	Number of years	Short-lived, slow-growing	Short-lived, fast-growing	Long-lived, slow-growing	Long-lived, fast-growing
N = 1000 (perfectly specified, with index + ages)	100	0.0087	-0.0020	-0.0052	-0.0020
N = 1000 (perfectly specified)	100	0.0937	0.1557	0.0980	0.1838
N = 200	75	-0.1024	-0.1359	0.0181	0.0377
N = 200	20	0.0079	-0.0402	0.0197	0.1010
N = 200	10	0.0063	0.0457	-0.0074	0.0133
N = 200	1	-0.0224	-0.0863	0.0370	0.0698
N = 50	75	0.0083	-0.1163	0.3280	0.0785
N = 50	20	0.0434	-0.0166	0.2858	0.1011
N = 50	10	0.0602	0.1192	0.2637	0.0409
N = 50	1	-0.0151	0.0005	0.1759	-0.1549
Decline from 200 samples (year $\leq$ 87) to 50 samples (year > 87)	20	0.0652	-0.0779	0.2215	0.0131

Table 5. Median absolute relative error (i.e., precision) of the final year fraction of unfished biomass across 100 simulation replicates by life history type and length data scenario with recruitment standard deviation = 0.8 with bias-adjusted recruitment. Cells are shaded green for precision less than 20%, yellow less than 30%, and red greater than 30%.

Length data scenario		Life history scenario			
Sample size	Number of years	Short-lived, slow-growing	Short-lived, fast-growing	Long-lived, slow-growing	Long-lived, fast-growing
N = 1000 (perfectly specified, with index + ages)	100	0.0833	0.0868	0.0798	0.0832
N = 1000 (perfectly specified)	100	0.1585	0.1984	0.1575	0.1808
N = 200	75	0.2618	0.2696	0.1738	0.1841
N = 200	20	0.2781	0.3131	0.1923	0.2106
N = 200	10	0.2717	0.3098	0.2227	0.2254
N = 200	1	0.4385	0.2785	0.5237	0.3626
N = 50	75	0.2525	0.2948	0.3049	0.2181
N = 50	20	0.2659	0.3339	0.2794	0.2452
N = 50	10	0.3270	0.3252	0.2690	0.3505
N = 50	1	0.4187	0.5167	0.4724	0.5198
Decline from 200 samples (year $\leq$ 87) to 50 samples (year > 87)	20	0.2460	0.3377	0.2203	0.2064

Table 6. Median absolute relative error (i.e., precision) of the overfishing limit (OFL) in the first forecast year across 100 simulation replicates by life history type and length data scenario with recruitment standard deviation = 0.8 with bias-adjusted recruitment. Cells are shaded green for bias less than 20%, yellow less than 30%, and red greater than 30%.

Length data scenario		Life history scenario			
Sample size	Number of years	Short-lived, slow-growing	Short-lived, fast-growing	Long-lived, slow-growing	Long-lived, fast-growing
N = 1000 (perfectly specified, with index + ages)	100	0.0288	0.0244	0.0236	0.0201
N = 1000 (perfectly specified)	100	0.1884	0.2679	0.1374	0.2346
N = 200	75	0.3191	0.3066	0.2222	0.2685
N = 200	20	0.3358	0.3233	0.2446	0.2950
N = 200	10	0.3534	0.3455	0.2741	0.2915
N = 200	1	0.4878	0.4126	0.5656	0.4210
N = 50	75	0.2880	0.3196	0.3280	0.2568
N = 50	20	0.2940	0.4212	0.3162	0.3676
N = 50	10	0.3806	0.4310	0.2909	0.4124
N = 50	1	0.5270	0.6147	0.5541	0.5985
Decline from 200 samples (year $\leq$ 87) to 50 samples (year > 87)	20	0.3308	0.4204	0.2760	0.2694

Table 7. Number of model runs out of 100 simulation replicates where the absolute relative error in final year fraction unfished biomass was greater than 1.0, given recruitment standard deviation 0.8 and bias-adjusted recruitment.

Length data scenario		Life history scenario			
Sample size	Number of years	Short-lived, slow-growing	Short-lived, fast-growing	Long-lived, slow-growing	Long-lived, fast-growing
N = 1000 (perfectly specified, with index + ages)	100	0	0	0	0
N = 1000 (perfectly specified)	100	1	1	0	1
N = 200	75	1	2	0	1
N = 200	20	1	2	1	1
N = 200	10	2	2	1	2
N = 200	1	10	8	15	8
N = 50	75	2	6	2	1
N = 50	20	5	7	4	1
N = 50	10	6	10	4	5
N = 50	1	23	15	17	14
Decline from 200 samples (year $\leq$ 87) to 50 samples (year > 87)	20	4	4	2	1

Figures

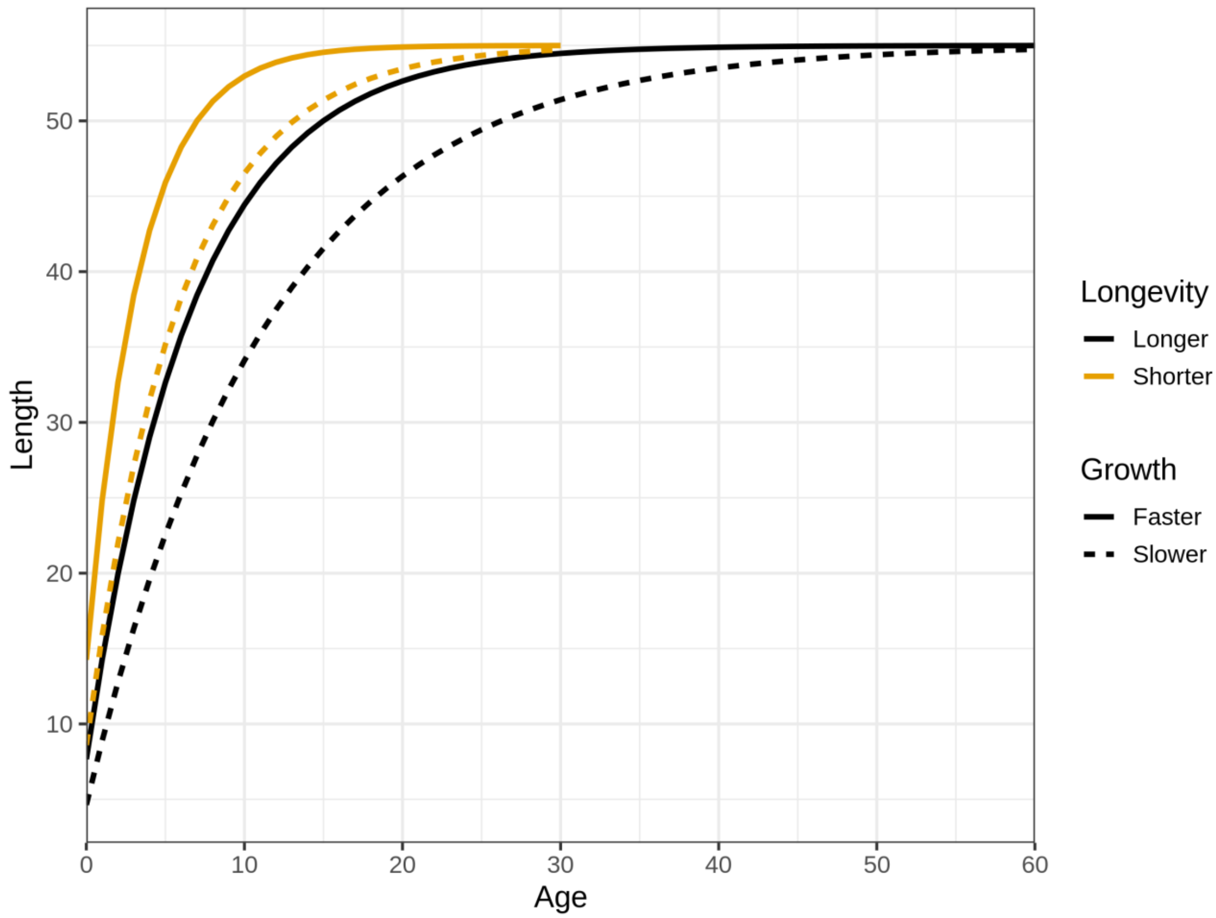

Figure 1. Length-at-age for each of the four life history types, varying in their longevity and growth rate, either reaching their maximum size at 50% (faster growth) or 90% (slower growth) of their maximum age (short = 30 years, long = 60 years).

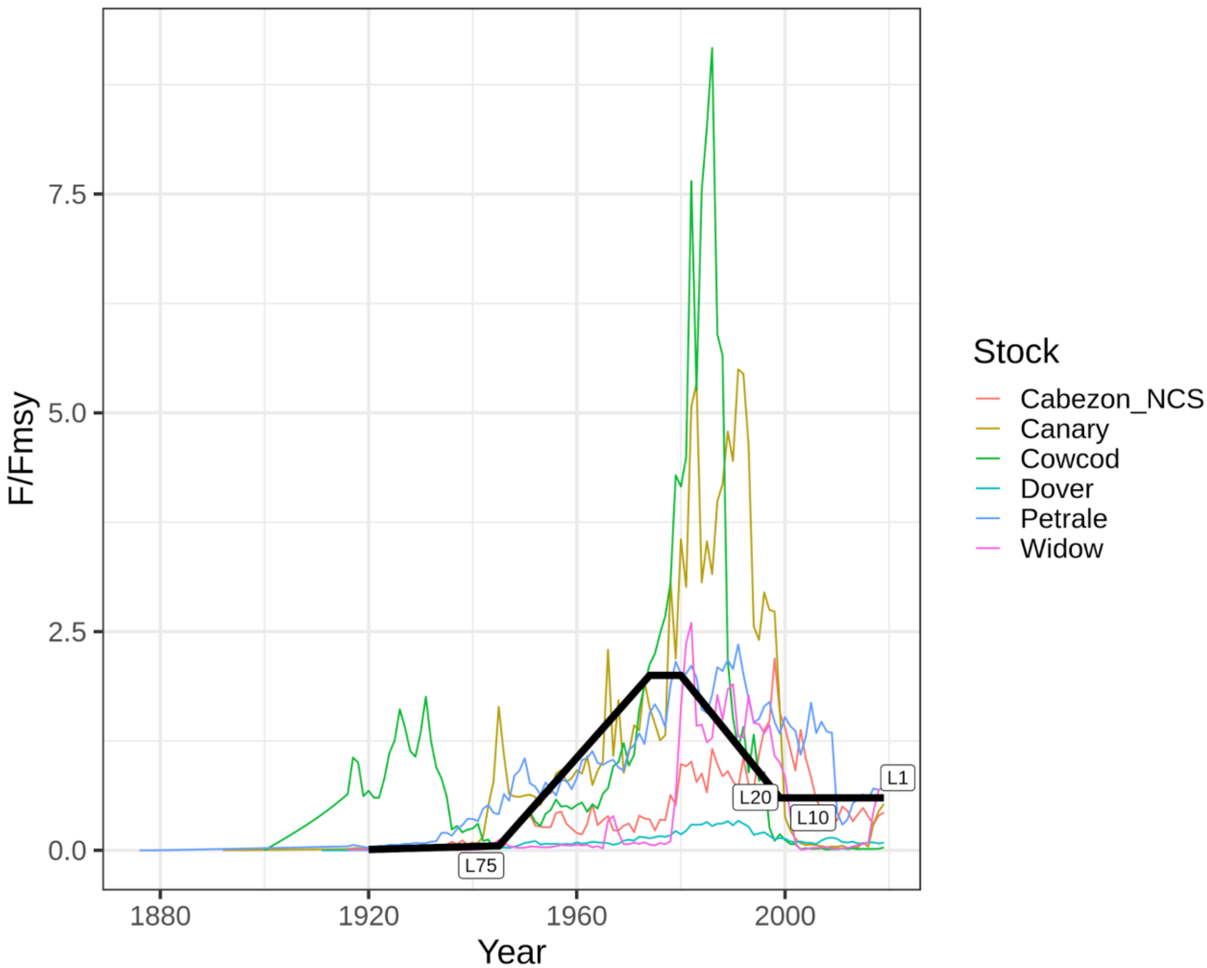

Figure 2. Ratio of estimated fishing mortality rate to F_{MSY} for six West Coast stocks, which informed the shape of our chosen F/F_{MSY} time series on which to base the simulated fishing mortality rate scenario. The simulated F/F_{MSY} series is labeled where the length data are included: for 75 years, 20, 10, and 1 year of data.

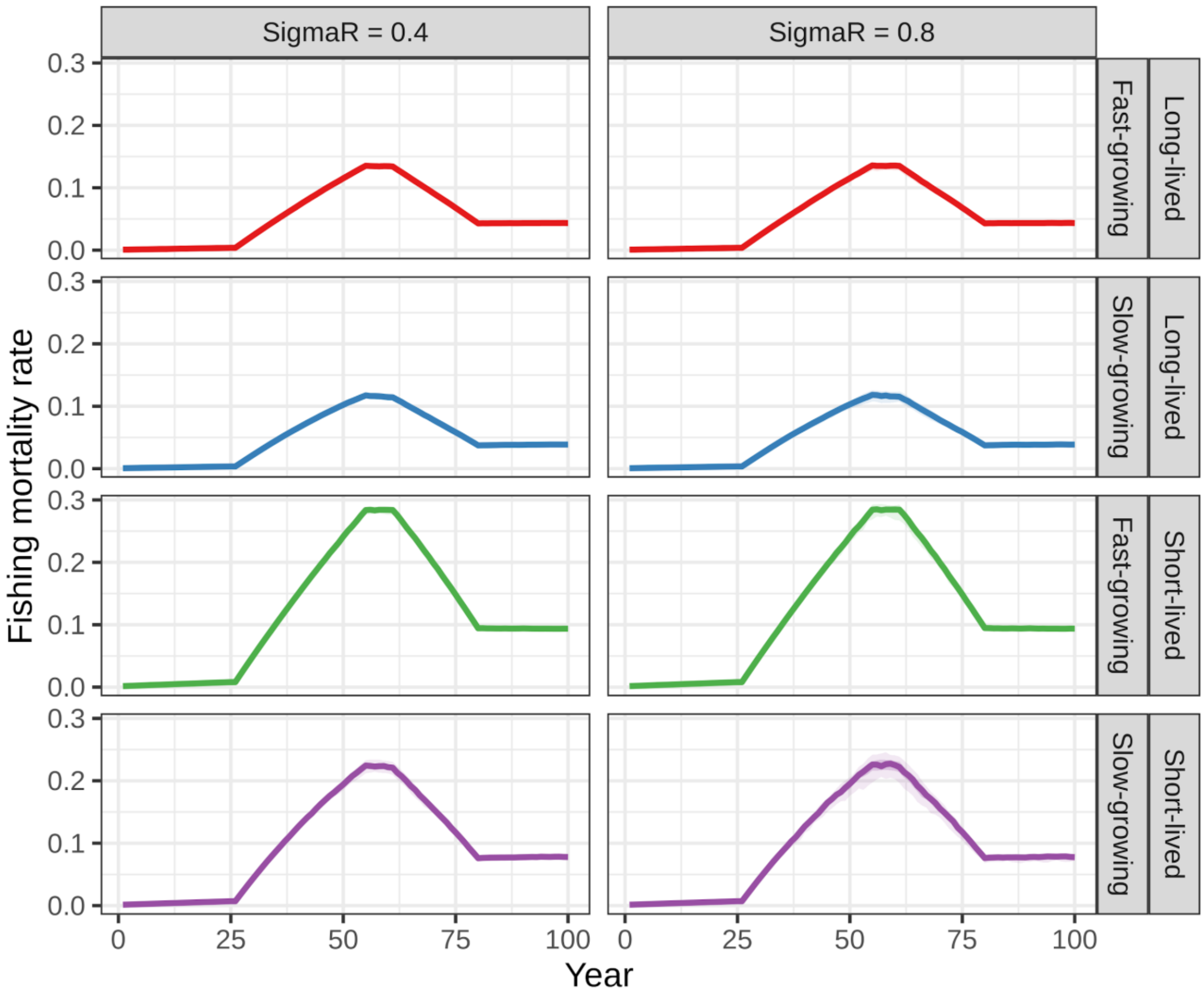

Figure 3. Comparison of the F trajectories for each life history type based on the F_{MSY} for each life history type (Table 1) and a shared pattern of F/F_{MSY} for each recruitment scenario. Solid lines show the median with shading showing the 90th quantiles of F which are based on the inputs including a standard error of 0.1.

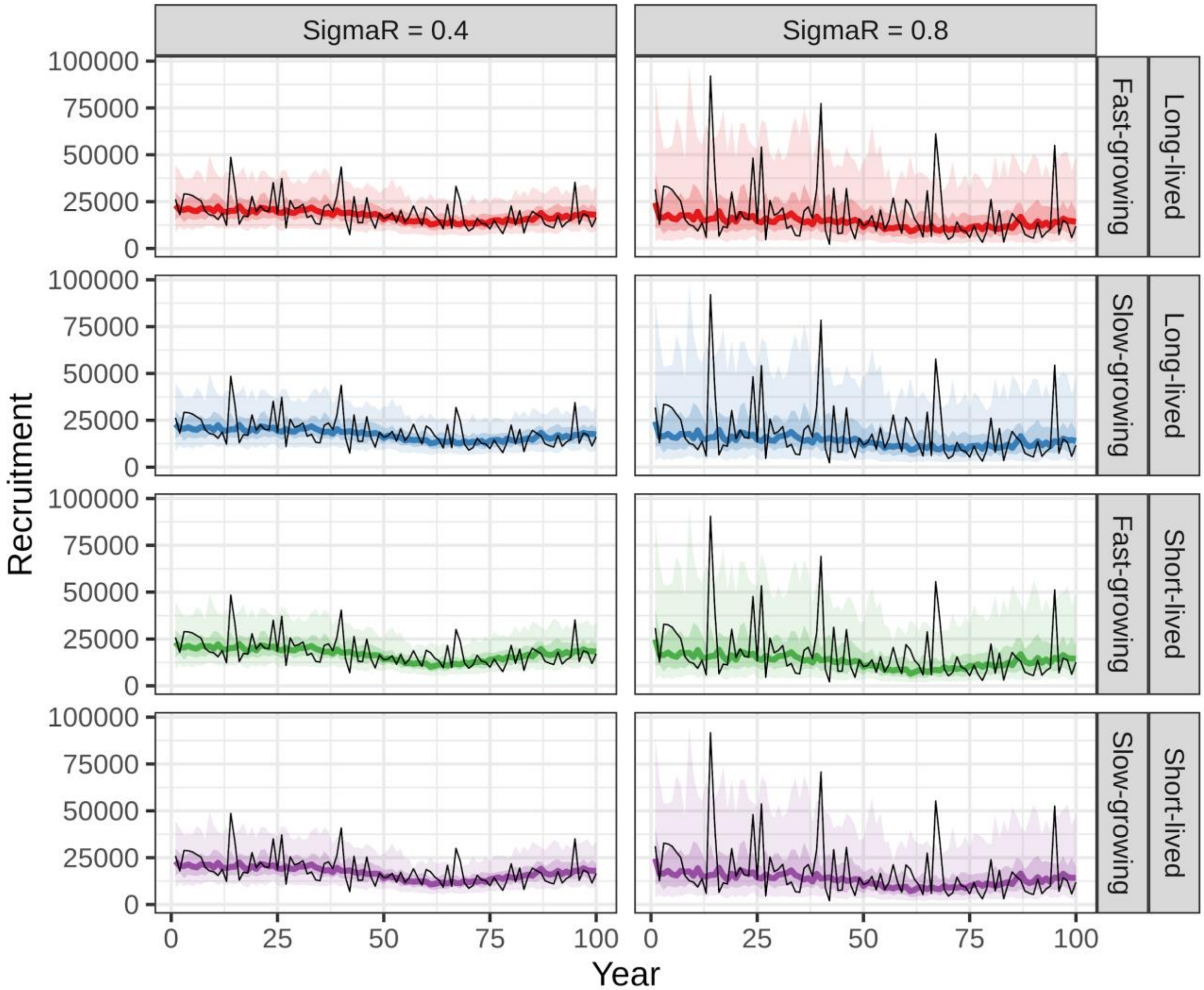

Figure 4. Comparison of the median, 50th quantiles, and 90th quantiles of recruitment with sigma R = 0.4 vs. 0.8 across 100 simulation replicates for each life history scenario. The single black lines show a single simulation replicate of recruitment.

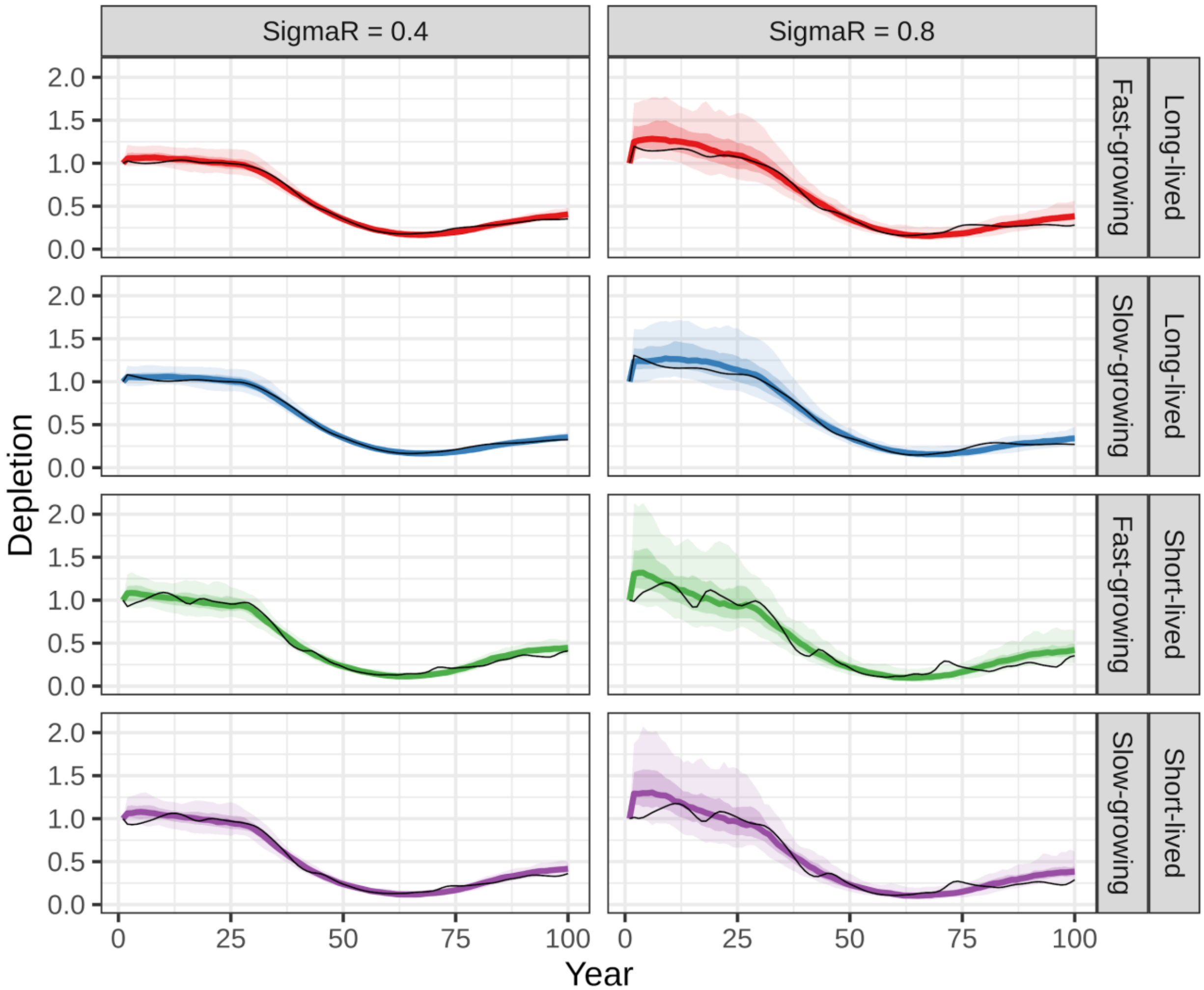

Figure 5. Comparison of the median, 50th quantiles, and 90th quantiles of fraction of unfished biomass with sigma R = 0.4 vs. 0.8 across 100 simulation replicates for each life history scenario. The single black lines show a single simulation replicate of depletion.

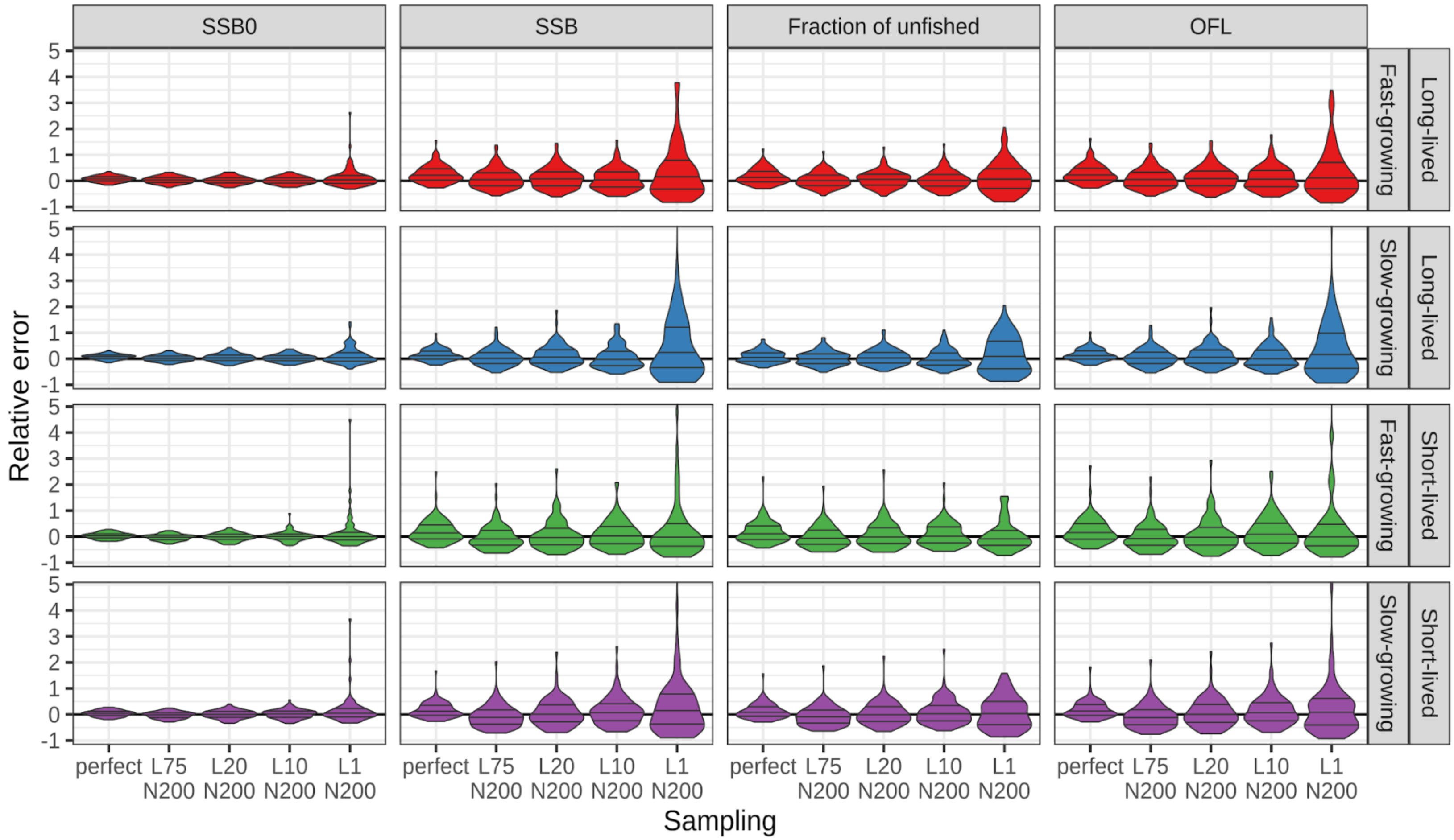

Figure 6. Relative error distributions for the unfished spawning biomass, final year spawning biomass, final year fraction unfished, and the OFL in the first forecast year across 100 simulation replicates with 200 samples of length data annually by each life history type and number of years of length compositions, when standard deviation in recruitment deviates was 0.8 and the model included bias adjustment.

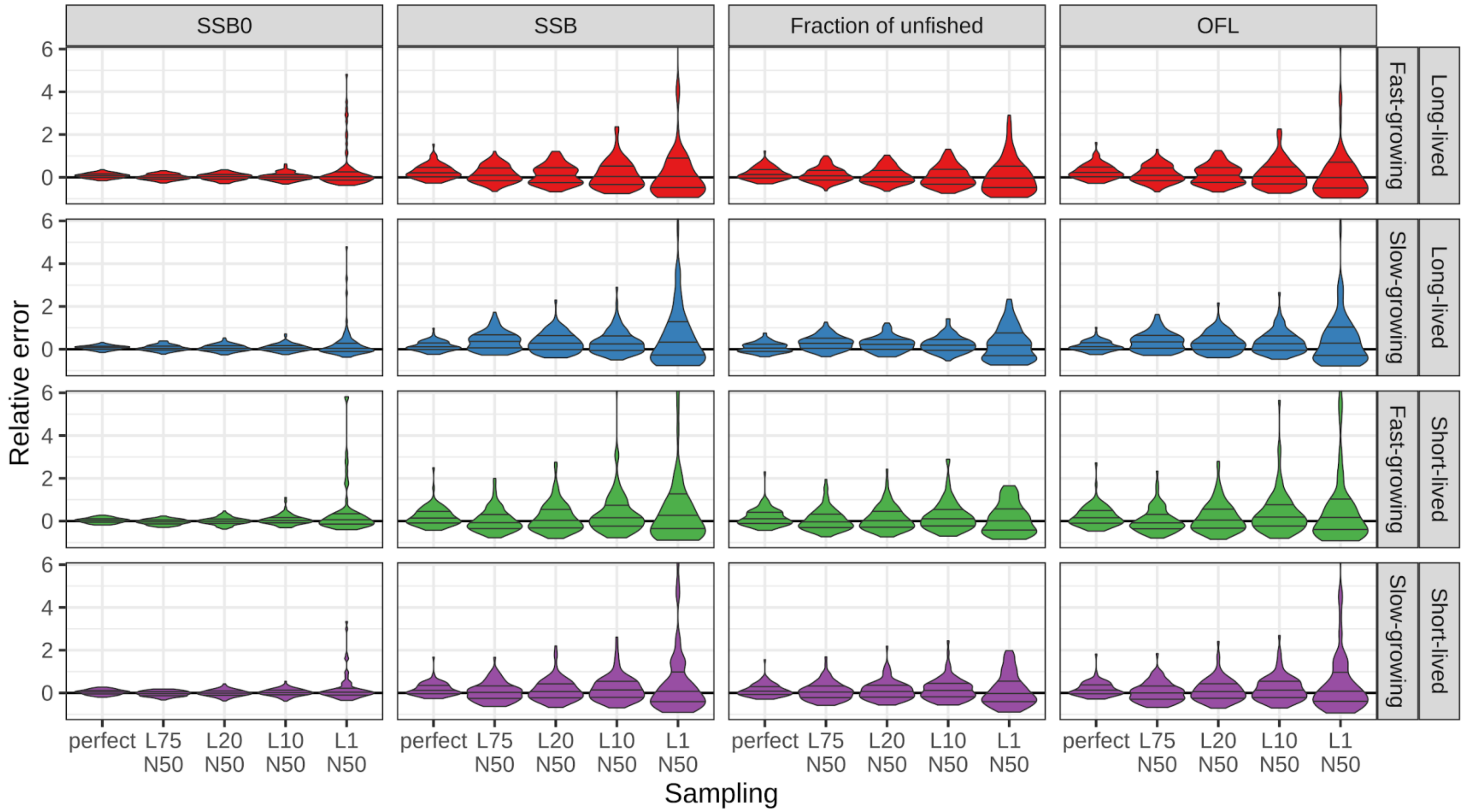

Figure 7. Relative error distributions for the unfished spawning biomass, final year spawning biomass, final year fraction unfished, and the OFL in the first forecast year across 100 simulation replicates with 50 samples of length data annually by each life history type and number of years of length compositions, when standard deviation in recruitment deviates was 0.8 and the model included bias adjustment.

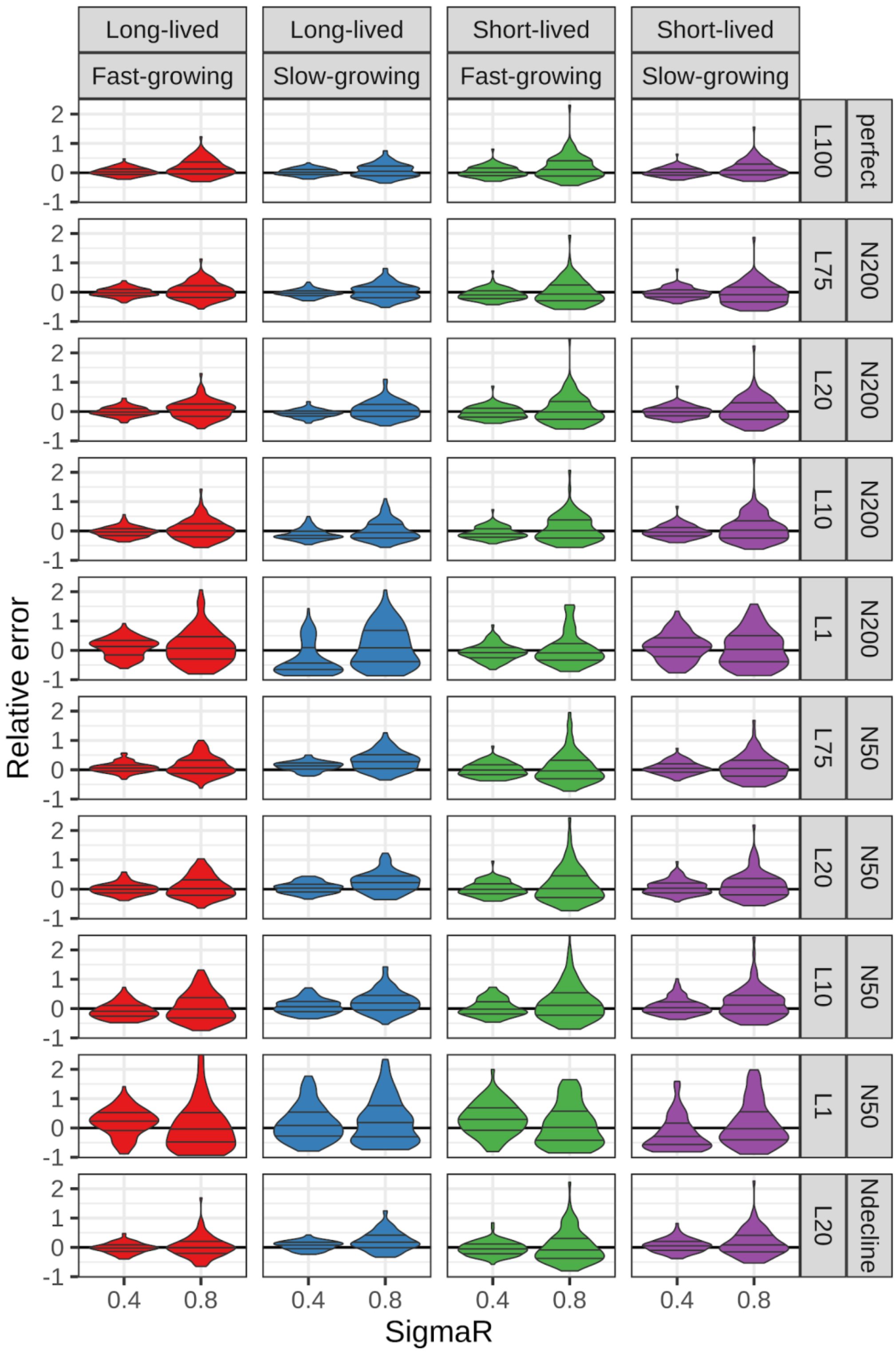

Figure 8. Relative error distribution for final year fraction of unfished biomass compared between sigma R of 0.4 and 0.8 across life history types and length data sampling scenarios for 100 simulation replicates.

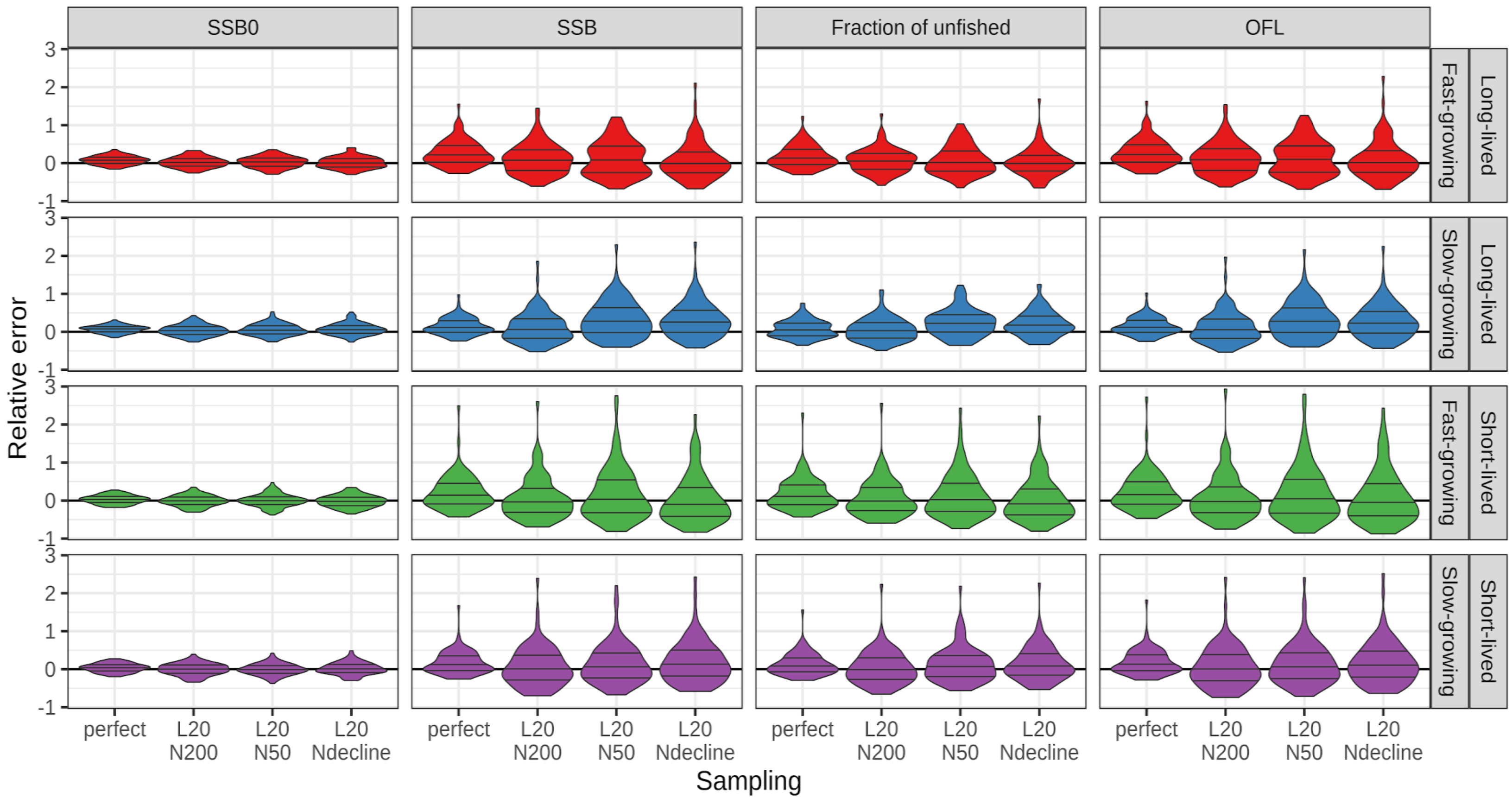

Figure 9. Relative error in the unfished spawning biomass, final year spawning biomass, final year fraction unfished, and the OFL in the first forecast year across 100 simulation replicates for each life history type and sample size with 20 years of length data compared with the perfect length data scenario, when standard deviation in recruitment deviates was 0.8 and the model included bias adjustment. 'Ndecline' indicates the scenario with 200 samples from years 81-87 and 50 samples from years 88-100.

Section 2: Peeling the stock assessment onion: Evaluating changes in model output under systematic data reduction in reference stock assessments, with special consideration of length plus catch models.

Jason Cope and Chantel Wetzel

Introduction

Simulation testing provides an experimental framework for controlled testing of complex systems. Section 1 applied this approach to test the potential of catch and length-based models in Stock Synthesis (SS) to provide robust model estimation of important management quantities. While a powerful method, it is limited in the number of scenarios and complexities that reasonably can be considered.

A different approach that trades modelling control for real world complexity is examining what happens to a full stock assessment model when data is removed. This is a commonly applied approach in stock assessment analysis to understand general likelihood component influence and potential contradictory information, but also provides an alternative way of measuring how data conditioning affects model outputs (Cope et al. 2015). Here we present 10 West Coast groundfish stocks with accepted assessments representing different ecologies, fishery fleet types, and data availabilities to look for patterns in the removal of data-- with a specific focus on the availability of length compositions. This will allow comparisons to the simulation results to see if those results hold in more complex "real" data and modelling scenarios while also offering an independent measure of the influence of length composition data on stock assessment outputs.

Material & Methods

A subset of West Coast groundfish stocks with existing full assessments were selected for data explorations (Table 2.1). The selected stock encompass a range of life histories (e.g., flatfish, roundfish, elasmobranch, rockfish), exploitation (e.g., recreational or commercial fisheries), and data availability (e.g., CPUE, fishery-independent indices, and length and age compositions). Each assessment also presents variable amounts of data quality and quantity within each data type, thus there is no ability to standardize data within the data scenarios.

Model treatments

To generate appropriate data scenario comparisons to the full assessments a number of steps were taken based on the data scenario. The archived assessment for each of the listed assessment years for each of the stocks in Table 2.1 were used as a starting point for analysis. A select group of the archived assessment was then converted to SS v.3.30.15 (Dover sole, longspine thornyhead, kelp greenling) for ease of exploration with the converted model compared to the original model to ensure similar estimates and model performance. All biological parameters were fixed across scenarios to limit the effects of aberrant parameter estimation, as it is possible to estimate those parameters outside the model, while allowing selectivity and recruitment deviations (if estimated) to remain estimated (parameters not typically estimated outside the models). Additionally, if present in the model, retention parameters were fixed at the MLE estimates in order to avoid variances in estimates of total mortality among runs. Next, each of the full assessments were run with the full Hessian and reweighted according to the Francis data weighting approach. This step was performed since the assessments ranged over a period of time when model weighting approaches were evolving. Additionally, since the scenarios were aimed to explore the sensitivity of the model to data, applying the appropriate data weighting within each scenario was considered essential. The reweighted full assessment model was termed the "reference" model.

Seven data scenarios were conducted relative to the reference model. Each of the scenarios and the steps applied in their creation were as follows:

1. Removal of all indices (" - Indices") - the likelihood contribution for all indices in the model were set equal to 0. The data remaining in the models were the catches, lengths, and ages that were available in the reference model.
2. Removal of all lengths (" - Lengths") - the likelihood contribution for all length data in the model were set equal to 0. The data remaining in the model were the catches, indices, and ages that were available in the reference model.
3. Removal of all ages (" - Ages") - the likelihood contribution for all age data in the model were set equal to 0. The data remaining in the model were the catches, indices, and lengths that were available in the reference model.
4. "Only Lengths" - the likelihood contribution for all indices and age data in the model were set equal to 0. The data remaining in the model were the catches and lengths that were available in the reference model.
5. "Lengths 20 Years" - the likelihood contribution for all indices and age data in the model were set equal to 0 and all length data prior to 20 years before the final model year were removed. The data remaining in the model were the catches from all years and length data from the last 20 years of the model. If the reference model had selectivity blocks applied to fleet selectivity (could be survey or fishery) that were outside the new length data range, those parameters were fixed at the reference model MLE estimate.
6. "Lengths 10 Years" - the same set-up as described above for the "Lengths 20 Years" scenario but lengths were reduced to the last 10 years of the model.

7. “Lengths 1 Year” - the same set-up as described above in the previous two length-based scenarios but where only the final year length data were retained.

Metrics

The performance of each data scenario was evaluated using the measure of relative error in four estimated quantities: 1) unfished spawning biomass, 2) spawning biomass in the final year, 3) fraction of unfished biomass in the final year, and 4) the first OFL value. The relative error was calculated as:

$$RE = \frac{E_s - E_f}{E_f}$$

where E_s is the estimate quantity from the data scenario s and E_f is the estimate from the full reference model f. The 95% confidence interval of the RE is also provided to indicate whether a given scenario would be found within the reference models estimated uncertainty. Those scenarios that do fall outside those bounds would be indicative of more significant departure from the reference model.

Results

There are several caveats to be mindful of when interpreting these results. Each data inclusion scenario is within an assessment that shows variable levels of consistent or inconsistent signal among data types, as well as how much each data type is weighted in the reference model. Each truncated length scenario represents a different amount of the total available length data. Additionally, the level of sampling in the most recent years of data is also highly variable among the stocks. Lastly, the model structure assumed across data scenarios (e.g., estimated vs. fixed parameters) likely do not fully reflect the decisions an assessment author may make when faced with the data remaining for a real world assessment. Results are structured first within species categories, as those often share common data issues, then general result patterns are provided and the relative error for unfished spawning biomass, final spawning biomass, final fraction of unfished, and the OFL are shown for all species in Figures 2.1 and 2.2.

Nearshore rockfishes

Gopher and Black-and-yellow rockfish

The gopher and black-and-yellow rockfish complex consists of two shallow nearshore demersal species that are a minor target for recreational and commercial fisheries. They are mostly taken by hook-and-line, and the live-fishery nature of the commercial fishery makes length collection a more suitable sampling option. The model covers the waters of California up to Cape Mendocino. The catch time series begins over 100 years ago, with abundance indices and length compositions beginning around 35 years prior to the end of the model (2019, Figure 2.3). Age compositions are very limited. Likelihood profiling indicated a weak but generally consistent signal in the length and age compositions, and to a lesser extent, in the indices (Monk and He 2019). The six indices of abundance in the model show stark contradictions in the information content for various model parameters (Monk and He 2019). So while there are several indices, they do not provide a consistent signal within the model. The reference model exhibits high uncertainty in biomass, and lower uncertainty in current relative stock status. Biologically, gopher and black-and-yellow rockfish would be more similar to the lower growing life history in the simulation study.

Removal of the length compositions demonstrated the largest effect on model outputs relative to the reference model, although that effect was minimal (Figure 2.4). Truncation of length data time series to 20 or 10 years of data show notable changes in terminal year biomass and current fraction of unfished (Figure 2.4). These changes tended to be higher biomass and subsequently more optimistic relative stocks status. Relative changes across model outputs and data scenarios were within the confidence intervals of the model, with the exception being the high positive relative error of the estimated OFL value (0.96) when all length compositions are removed (Figure 2.1).

China rockfish

China rockfish is a deeper nearshore demersal species that is a minor target for recreational and commercial fisheries. It is mostly taken by hook-and-line and the live-fishery nature of the commercial fishery makes length collection a more suitable sampling option. The northern model covers the waters of Washington state. The catch time series begins just over 50 years ago, with a single abundance index beginning in the early 1980s, and the majority of length and age composition data present during approximately the last 20 years of the model (Figure 2.5). Likelihood profiling indicated a weak but generally consistent signal in the index and length and age compositions (Dick et al. 2015). The reference model exhibits high uncertainty in biomass, but low uncertainty in current relative stock status. Biologically, China rockfish would be more similar to the faster growing life history in the simulation study. Recruitment deviations are not estimated in the reference model.

There is very little effect on model outputs with any of the data scenarios (Figure 2.6). All model outputs were within the confidence intervals of the reference model (Figure 2.1). The biggest deviation from the reference model was found in the estimate of OFL for the 1 year of length data scenario.

Black rockfish

Black rockfish is a mostly nearshore, pelagic schooling species that is a major recreational target. It therefore is limited in the net-based catches and is mostly taken by hook-and-line gear. The Washington state black rockfish stock assessment catch time series begins roughly 80 years ago, with the abundance indices, length and age composition beginning mostly around 40 years prior to the final model year (2015, Figure 2.7). Likelihood profiling indicated indices and length composition data show some agreement, though often in contradiction to the age composition data

(Cope et al. 2015). The reference model exhibits low uncertainty and very little retrospective patterns. Biologically, black rockfish would be more similar to the slower growing life history in the simulation study.

Removal of the length composition data demonstrated the largest effect on model outputs, especially on the population scale estimate (Figure 2.8). The removal of indices or ages had little effect in the estimate of initial spawning biomass and across the majority of the time series, but both data scenarios had departures in the estimates during the final years of the model resulting in estimates of a more depleted stock relative to the reference model (Figure 2.8). Truncation of lengths shows significant decreases in terminal year biomass and current relative stock status (Figure 2.8). Most data scenarios demonstrated lower terminal biomass and current relative stock status estimates compared to the small confidence intervals of the reference model (Figure 2.1). While the reference model was near the target relative biomass level, the length scenarios tended to be closer to the minimum stock size threshold (Figure 2.8).

Slope rockfish

Yelloweye rockfish

The yelloweye rockfish stock assessment is a two area model containing submodels for California and Oregon/Washington. Yelloweye rockfish typically inhabit deep rocky habitat, and are difficult to sample using trawl gear but are effectively sampled using hook-and-line gear. The large size of yelloweye rockfish has made them a target of recreational fisheries, though they were believed overfished for many years, and thus have been under strict harvest guidelines since the mid-2000s. The restrictions had also decreased the amount of data available to subsequent stock assessments.

The catch time series is about 130 years, with data sources starting around 40 years ago (Figure 2.9). Likelihood profiling indicated indices and length composition data were generally in agreement, though in opposition to age data. Sensitivity analysis shows the removal of lengths caused issues with estimating the initial biomass and current relative stock abundance (Gertseva and Cope 2017). Overall uncertainty in the model is relatively low. This low uncertainty in the asymptotic estimates may be due to the model being one sex, as female and male life history parameters are very similar. While natural mortality and recruitment compensation (i.e., steepness) parameters are fixed (a common approach for a west coast groundfish stock assessment), growth, recruitment and many selectivity parameters are estimated, providing ample space for uncertainty in parameter estimation if data lacked information. Biologically, yelloweye rockfish is more similar to the slower growing life history in the simulation study.

Removal of the length compositions demonstrated the largest effect on model outputs (Figure 2.10), as the model estimated very different stock abundance throughout the time series, indicating that the length data are a primary source of information in the reference model. Removal of indices or ages had little effect resulting in very similar spawning biomass trajectories and fraction of unfished over time (Figure 2.10). The truncation of length data had small and consistent effects on model output, though enough to be outside the small confidence intervals of the reference model (Figure 2.1).

Slope scorpaenids

Longspine thornyhead

Longspine thornyhead are a deep water species off the West Coast that are primarily targeted by commercial trawl fishing and are frequently sampled by fishery independent trawl surveys (Figure 2.11). The most recent assessment for longspine thornyhead was conducted in 2013 and accepted as a category 2 stock assessment because no age data were included due to the current inability to age this species.

Comparing the two data scenarios that either included catch plus index only data or catch and length data, the estimate when only length data were present were nearly identical to the reference model (Figures 2.1 and 2.12). This indicates that the indices in the reference model have little influence on the model estimates. Examining the model estimates when variable amounts of length data were used, the model performance in terms of stock status were similar to the reference model when either 20 or 10 years of data were used (Figures. 2.1 and 2.12). The relative error across data scenarios to the OFL estimates were well within the reference model 95% confidence interval (Figure 2.1).

Nearshore roundfishes

Cabezon

The northern California substock of cabezon was used in this example, with a range from Point Conception to the California-Oregon border. The catch time series is also very long, with length compositions starting in earnest 40 years ago, and an index of abundance that stretches back 60 years, terminating 25 years ago, though also with another index in the most recent years (Figure 2.13). Age composition data was limited. Likelihood profiling indicated indices and length composition data were generally in agreement, and sensitivity analysis shows the removal of either caused issues with estimating the final biomass and relative stock abundance (Cope et al. 2019). This is also reflected in the levels of asymptotic uncertainty in reference model biomass being highest for the final year. Biologically, cabezon would be more similar to the faster growing life history in the simulation study.

Removal of the length compositions demonstrated the largest effect on model outputs (Figure 2.14), as the model had a hard time converging without lengths, with the highest uncertainty in the final year stock abundance. Removal of indices or ages had little effect. But even one year of length data allowed the model to provide reasonable results (Figure 2.14). For scenarios lacking indices and ages, there is a linear trend downward in biomass and subsequently lower relative stock abundance as the time series of lengths decreases (Figure 2.2). While lower than the reference model, all results are within the reference confidence intervals for each metric under all length scenarios.

Kelp greenling

Kelp greenling is a nearshore species that experience both recreational and commercial exploitation and is not sampled by existing West Coast trawl surveys (Figure 2.15). The data available in the reference model consists primarily of fishery dependent catch-per-unit-effort (CPUE) indices, length, and age composition data after the year 2000. The reference model included three CPUE indices and when they were removed from the model (" - Indices") the estimated spawning biomass scale declined with the stock trajectory at the lower 95% confidence interval of the reference model (Figures 2.2 and 2.16). However, the relative stock trajectory was similar to the reference model until the end of the time series where the data scenario sharply declined. When all length data were dropped from the model the spawning biomass was lower than the reference model with changes in the pattern of the stock trajectory over time but estimated a similar unfished fraction at the end of the time series. The data scenario where the age data leaving only catch, CPUEs, and lengths resulted in the most similar stock estimates of stock sizes and unfished fraction (Figure 2.16).

The suite of scenarios examining the model performance relative to the reference model when only catch and length data were available were highly variable (Figures 2.2 and 2.16). The scenario that retained all length data had a similar trajectory post-1980 but then diverged from the reference model at the end of the time series. The difference in the recent year estimates indicates that the CPUE indices in the reference model have a large influence in recent year estimates that when removed the length data did not support. However, when only the last twenty years of lengths were used in the model the stock trajectory over time does not have the large peaks and valleys in the spawning biomass time series but estimates a similar stock status in recent years. Estimates of stock size, status, and the trajectories differed greatly from the reference model when only limited data were available (10 years or 1 year, Figure 2.16). The relative error of the estimated final spawning biomass and unfished fraction were well outside the reference model confidence interval for the 1 year data sensitivity (Figure 2.2).

The OFL estimates across each of the data scenario runs were quite variable for kelp greenling relative to the reference model (Figure 2.2) where the estimates were underestimated with approximately -0.50 bias for each of the no indices, no lengths, and length only runs. The OFL estimates were most similar to the reference model when the ages were removed or when only twenty years of lengths were used (Figure 2.2).

Lingcod

The lingcod north substock comprises the areas of Oregon and Washington. The catch time series is very long, with most abundance indices and length compositions starting in the early 1980s (Figure 2.15). Age composition data is available for the final 20-30 years of the model (Figure 2.15). Likelihood profiling indicated indices and length composition data show some agreement, though often in contradiction to the age composition data (Haltuch et al. 2018). The reference model exhibits the most uncertainty in the initial abundance, though the last 20 years of biomass also show increasing uncertainty. Biologically, lingcod, especially females, would be more similar to the slower growing life history in the simulation study. Recruitment deviations are estimated in the reference model.

Removal of the length compositions or indices demonstrated the largest, and very similar, effects on model outputs (Figure 2.16). Removal of ages had little effect. The length only models show divergence in models with varying degrees of data showing reduced absolute abundance with the most length data, and higher abundance with lower years of length data (Figure 2.16). The larger departures from the reference model occurred with less sampled years. All data scenarios demonstrated consistent estimates of fraction of unfished with even the one year of length data scenario resulting in an informative estimate in the final year. There was a linear trend upward in biomass and overall steady and slightly larger fraction of unfished as the time series of lengths decreased (Figure 2.2). The estimates were all within the reference confidence intervals for each metric under each length scenario.

Flatfish

Dover sole

Dover sole is primarily exploited commercial trawl gear of the West Coast. The co-occurrence of this species with sablefish, a highly valuable stock, along with its own marketability have resulted in a long exploitation history. The reference assessment has a large number of length and age composition data from both commercial fleets and survey fleets, with four fishery independent surveys that were relatively flat across the sampled years (1980 - 2010, Figure 2.19).

The model was relatively insensitive to the removal of the index data (“- Indices”, Figures 2.2 and 2.20) with only a small decline in the spawning biomass stock size across time. The reference model estimated a relatively stable spawning biomass time series with limited impacts to the stock size due to removals. The indices in the model were relatively flat across time, especially the most recent index from the West Coast Groundfish Bottom Trawl Survey (WCGBTS, 2003-2010), and the limited change in the model estimates when the indices were removed highlights the lack of information in these data. The data scenarios that removed either all the length (“- Lengths”) or the age data (“- Ages”) resulted in downward shifts in the estimated spawning biomass size but were similar to the reference model in terms of scale.

The data scenarios that explored using only catch and length data generally varied based on the amount of length data available. The scenario which included either all or 20 years of length data were comparable with the "- Ages" data scenario which used all the lengths and indices in the reference model (Figures 2.2 and 2.20). The scale of the population from these scenarios were lower than the reference model but resulted in similar population scale estimates. However, when a larger amount of length data were removed, either only 10 or 1 years of length data, the estimates varied to a greater extent from the reference model and in the 1 year scenario resulted in a stock status that was outside the 95% confidence interval from the reference model.

The relative error of the OFL estimate varied across data scenarios (Figure 2.2). The estimate of the OFL when the indices were removed was the closest to the reference model. The scenario that removed all the ages had a negative bias in the estimate of -0.17 and the scenario that used only catch and all lengths available had a similar, but slightly larger relative error of -0.22.

Elasmobranch

Big skate

The big skate assessment was done on a coastwide basis, and is an example of a stock with a long catch history, but mostly limited to data within the last 20 years (Figure 2.21). While the indices of abundance had small average slopes upward, the fits to the indices of abundance are mostly flat, indicating very little influence or information content. The age data also seem to be weakly informative, and contradictory to the signal in the index. Likelihood component analyses (Taylor et al. 2019) indicate length compositions to be the most informative data source. The estimates of biomass are highly largely uncertain. Biologically, big skate growth is slow to reach asymptotic size, thus having relatively more informative length compositions. Recruitment was not estimated in this model.

Removal of the length compositions demonstrated the largest effect on model outputs (Figure 2.22), though given the already large uncertainty in reference model biomass, was just within the confidence bound (Figure 2.2). Further examination of altering available length composition data when no indices or age data are included showed mostly conservative deviations from the reference model, with even 1 year of length data being informative of model scale and relative abundance (Figure 2.22). The OFL estimate demonstrated the largest deviation from the reference model.

General results

Some general results across all stocks:

- Removing length compositions from the assessment often caused large model deviations in outputs compared to removing other data sources.

- Models with only length compositions tended to provide informative outputs relative to the reference model, especially for relative stock status.
- In several instances there was a trend in the metric values with less length data (e.g., going up or down with less data), though the slope of that trend was not always consistent among species. The lack of length data most often led to lower estimates of biomass, fraction of unfished, and OFL compared to the reference model. Having either 1 or 10 years of length data most often lead to the most variable results.
- Having only one year of length data led to more conservative estimates of model output in 7 of 10 models. Only one of the higher estimates (kelp greenling) was outside the confidence intervals of the reference model.
- The terminal biomass and OFL tended to be the most sensitive model outputs.
- Most data scenarios fell within the confidence intervals of the reference model.

Discussion

Length composition data proved a critical input to a variety of west coast groundfish stock assessments. Length data were not just ancillary to other data types, as models reduced to only length and catch histories, including those with short times series (10 or less years) of length compositions retained much of the information of reference models. Reducing length compositions would often offer simplified views of past population dynamics, but could still provide informed estimates of the fraction of unfished in the final model years. The results here do not discourage the use of length and catch models as viable category 2 stock assessment candidates.

While it is most desirable to have all forms of data integrated and working together in a stock assessment, it is not unusual that different data types show weak and/or conflicting contributions of indices of abundance. Many nearshore species do not have scientifically-designed abundance indices available; the fishery-dependent catch-per-unit-effort time series that are available subsequently suffer from large uncertainty, and thus weak influence on model outputs. For years the standardized trawl survey index provided low contrast, and thus low information for many groundfish stocks. These types of stock assessments tend to behave similarly to length-only models, so there is a precedent for length-driven models to inform west coast fisheries management.

In instances where data sources are more influential, contradictory signals present real problems (Maunder et al. 2017). Data-weighting is an important, non-trivial aspect of developing reference stock assessment models, and there is no one way to do it correctly (Francis 2017; Punt 2017). Thus, decisions are necessary to resolve contradictory data. Down-weighting certain likelihood components in favor of others is common, but may instead mask important model mis-specifications (Maunder and Piner 2017; Wang and Maunder 2017). And the inclusion of multiple data types in an integrated model may cause additional challenges as data may have influence on unrelated model processes (Piner et al. 2016), thus arguing for the specification of model parameters outside the model. One example is establishing life history values such as natural mortality or growth external to the model in order to better establish selectivity parameters, and subsequently exploring model misspecification through sensitivity analyses. The decision to fix the life history parameters in these model comparisons is therefore not unrealistic, and also provided one level of experimental control in separating the effects of data exclusion rather than life history misspecification. This decision likely decreased the influence of age-- and possibly length-- composition to a certain extent, but was a trade-off to gain interpretability of results.

Length compositions do tend to offer a more conservative view of the population if the only source of the data, though the degree of this difference varies. The basic argument of including length compositions is that they provide information on length-based selectivity, fishing intensity, recruitment deviations and current stock status. When indices of abundance or age compositions are either unavailable or too resource intensive to process, length plus catch models show the capacity to provide suitable estimates of sustainable catch (defined as not being above the benchmark OFL estimate) when applying the category 2 stock assessment ABC buffer (assuming P*=0.45). In most cases, length-only models were more conservative than the reference model in all examined model outputs, decreasing the chance that such models will lead to overfishing in the short term.

The stripping back of stock assessments into less data does not also presume simpler models. Most of the model complexities were maintained across these data scenarios, and are likely not how one would specify a stock assessment model if truly faced with limited data. When data are sparse, parsimony is beneficial, as the estimation of numerous selectivity parameters with little data may complicate mode convergence. How this would influence the comparisons was not explored, though keeping model complexity high still resulted in reasonable results for the length only models. Likewise, other simplifications were made, such as fixing life history parameters to the reference model, thus possibly reducing the amount of deviance from the reference model.

Tables

Table 2.1: List of West Coast groundfish stock assessments evaluated. The Fleet column indicates the fleet structure in the reference model showing the number of fishing fleets with removals (fishery) and the number of survey fleets (independent and fishery-dependent) where the number inside parentheticals indicates the number of fishery dependent indices.

Species	Model Years	Fleets	Authors
Dover sole	1910-2010	3 fishery and 4 survey	Hicks and Wetzel 2013
big skate	1916-2018	4 fishery and 2 survey	Taylor et al. 2019
cabezon (NCS)	1916-2018	5 fishery and 2 (1) survey	Cope et al. 2019
lingcod (North)	1889-2016	4 (4) fishery and 4 survey	Haltuch et al. 2017
kelp greenling	1915-2014	5 fishery and 3 (3) survey	Berger et al. 2015
longspine thornyhead	1964-2012	1 fishery and 3 survey	Stephens and Taylor 2013
yelloweye rockfish	1889-2016	7 (3) fishery and 5 (2) survey	Gertseva and Cope 2017
black rockfish (WA)	1940-2017	3 fishery and 2 (2) survey	Cope et al. 2015
China rockfish (North)	1900-2014	3 (1) fishery	Dick et al. 2015
gopher and black and yellow rockfish	1916-2018	3 fishery and 7 (3) survey	Monk and He 2019

Figures

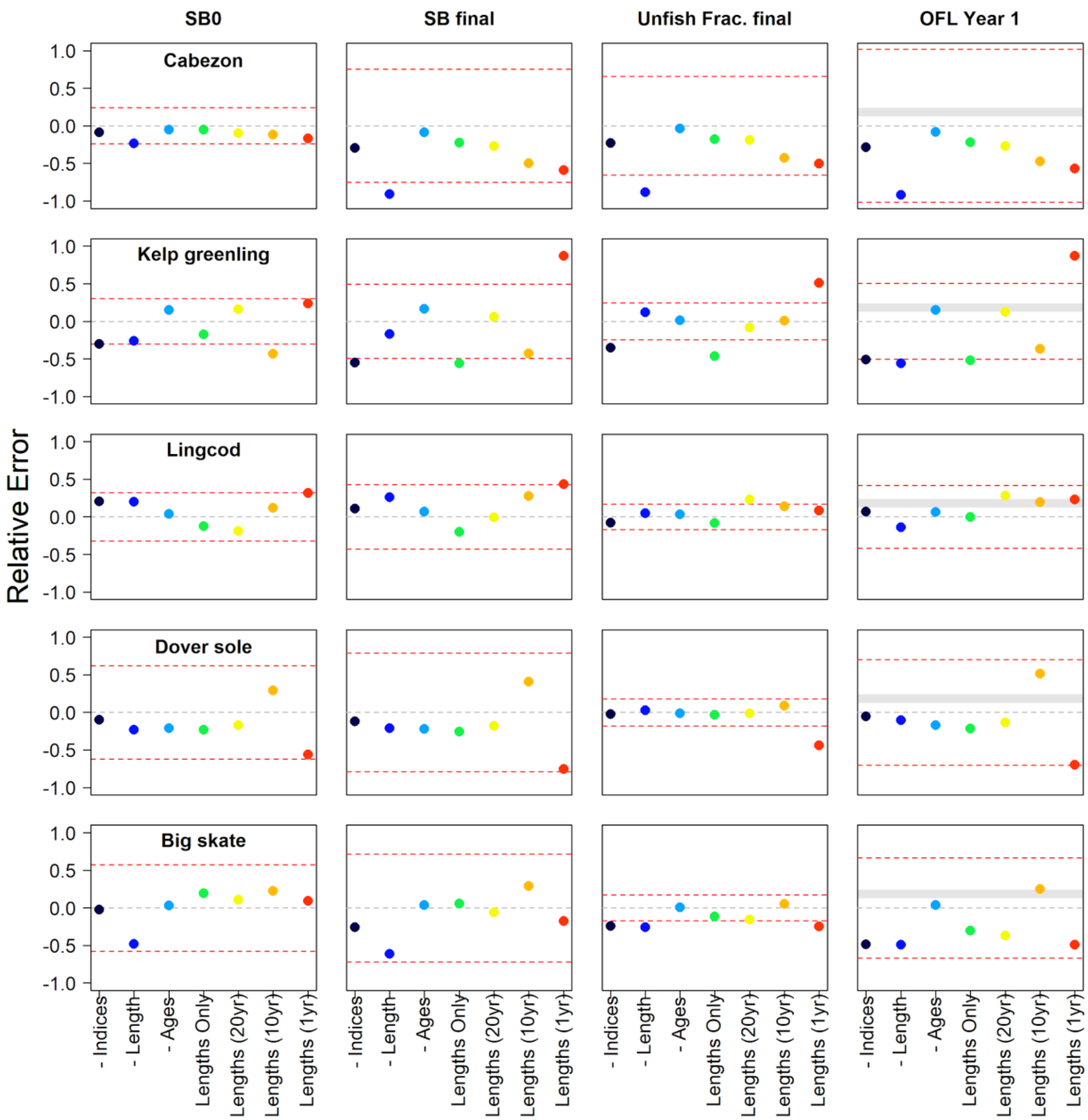

Figure 2.1. The relative error of unfished spawning biomass, final model year spawning biomass, final model year fraction of unfished, and the first overfishing level estimated from each data scenario compared against the reference model. The dashed grey line identifies the zero line and the dashed red lines identify the 95% confidence interval from the reference model for each of the estimated quantities. The grey banded area on the OFLs indicates the area between a category 2 sigma of 1.0 and either a P* value of 0.45 or 0.45 (buffer = 0.874 vs. 0.761) translated into relative error (0.126 - 0.238) where the resulting Acceptable Biological Catch if based on the estimated OFL would be greater than the OFL of the reference model.

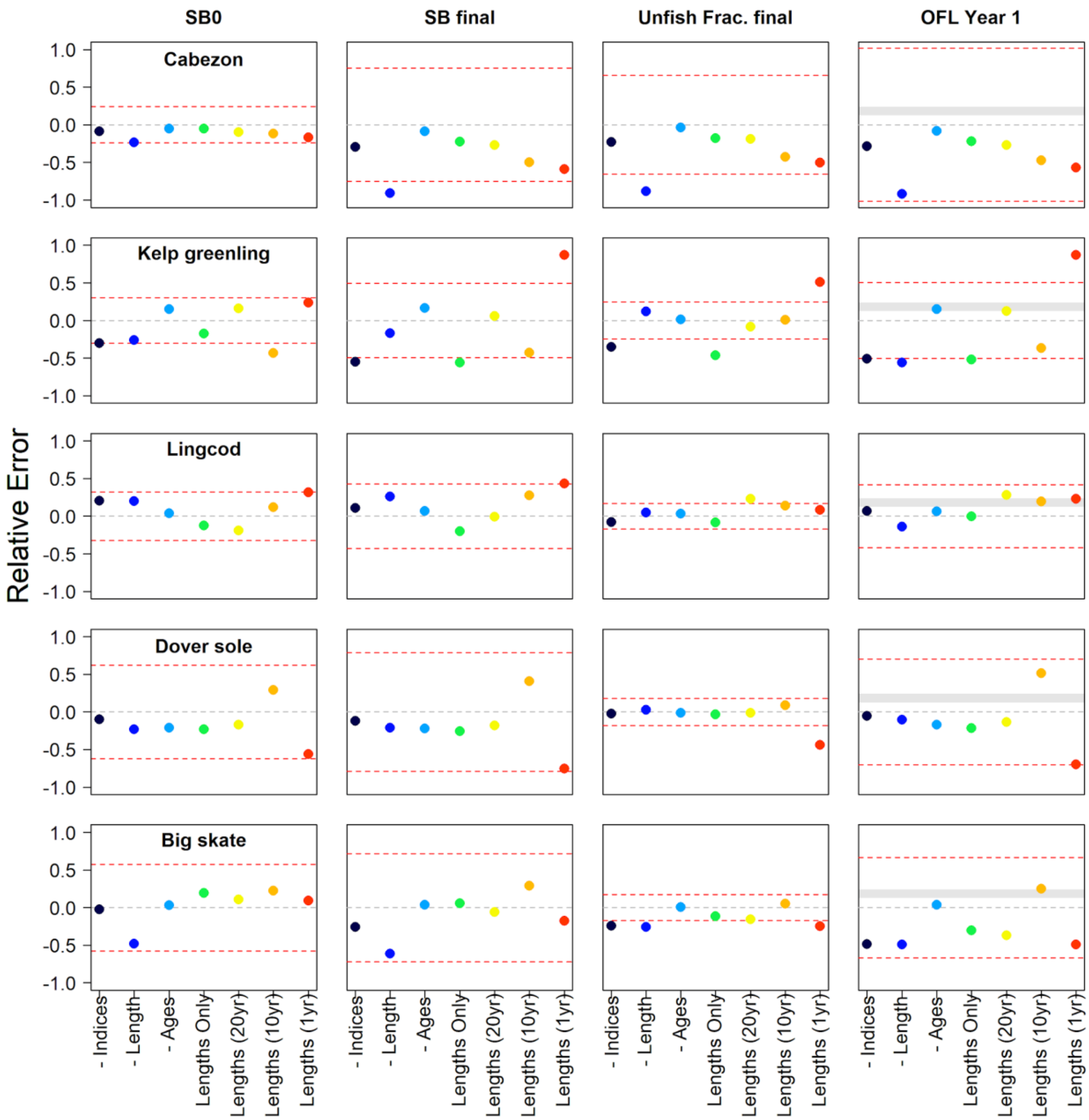

Figure 2.2. The relative error of unfished spawning biomass, final model year spawning biomass, final model year fraction of unfished, and the first overfishing level estimated from each data scenario compared against the reference model. The dashed grey line identifies the zero line and the dashed red lines identify the 95% confidence interval from the reference model for each of the estimated quantities. The grey banded area on the OFLs indicates the area between a category 2 sigma of 1.0 and either a P* value of 0.45 or 0.45 (buffer = 0.874 vs. 0.761) translated into relative error (0.126 - 0.238) where the resulting Acceptable Biological Catch if based on the estimated OFL would be greater than the OFL of the reference model.

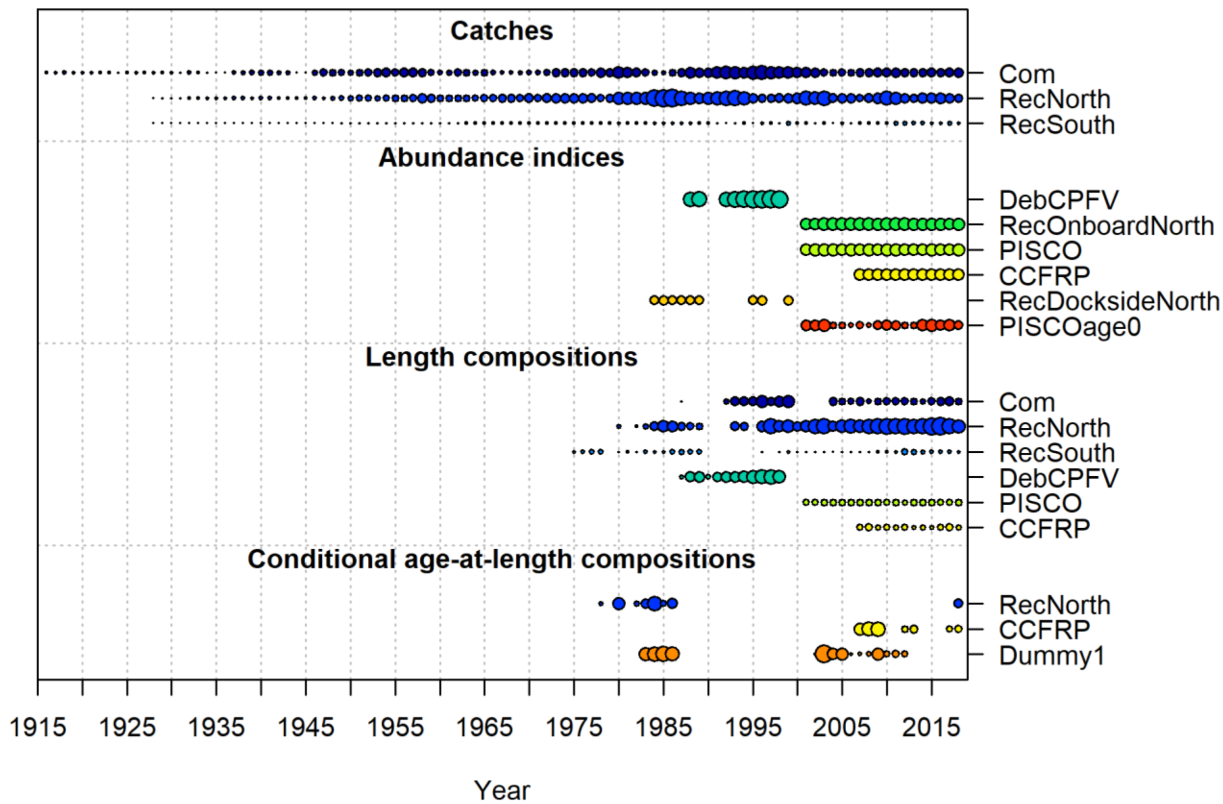

Figure 2.3. Summary of the data type, length of data time series, and quantity for gopher and black-and-yellow rockfish. The size of the bubble by year is based on the sample size (e.g., larger bubbles indicate higher sample sizes). The bubble size for the indices is equal to the inverse of the mean standard error for the survey.

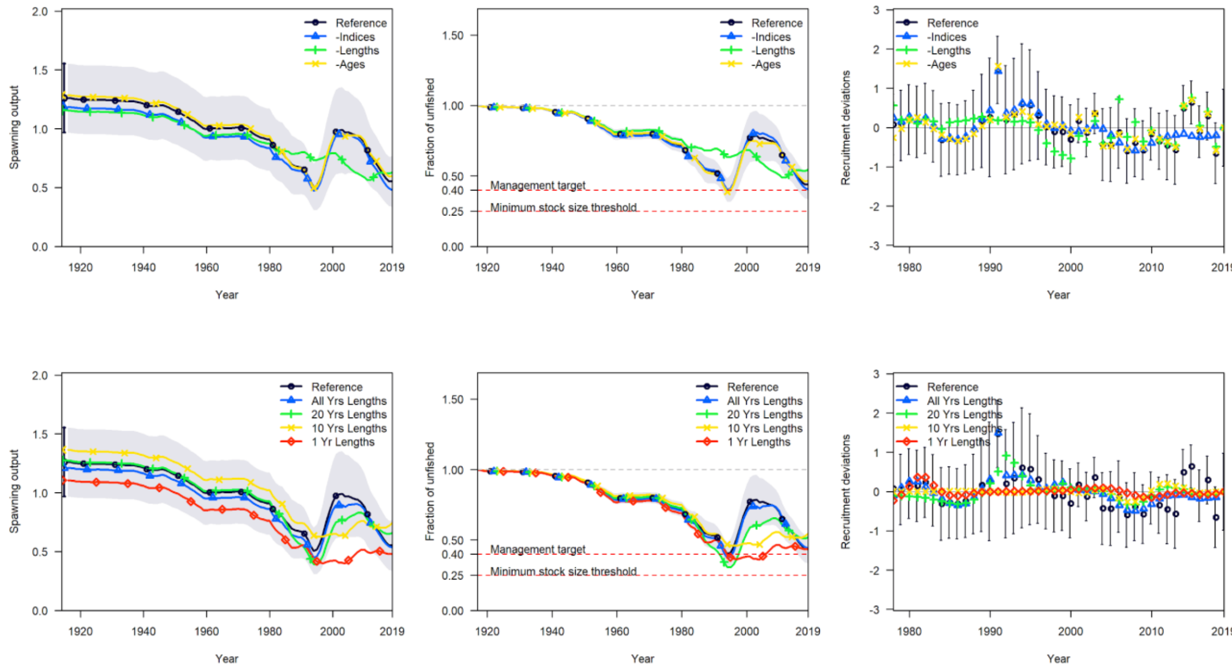

Figure 2.4. Comparison of model estimates of spawning biomass, the fraction of unfished, and annual recruitment deviations relative to the reference model (black) for gopher and black-and-yellow rockfish. The top row are comparisons between the reference model and when all indices (blue), all lengths (green), or all ages (yellow) are removed. The bottom row are comparisons between the reference model and when only catches and all lengths (blue), 20 years of lengths (green), 10 years of length (yellow), or only 1 year of lengths (red) are used in the model.

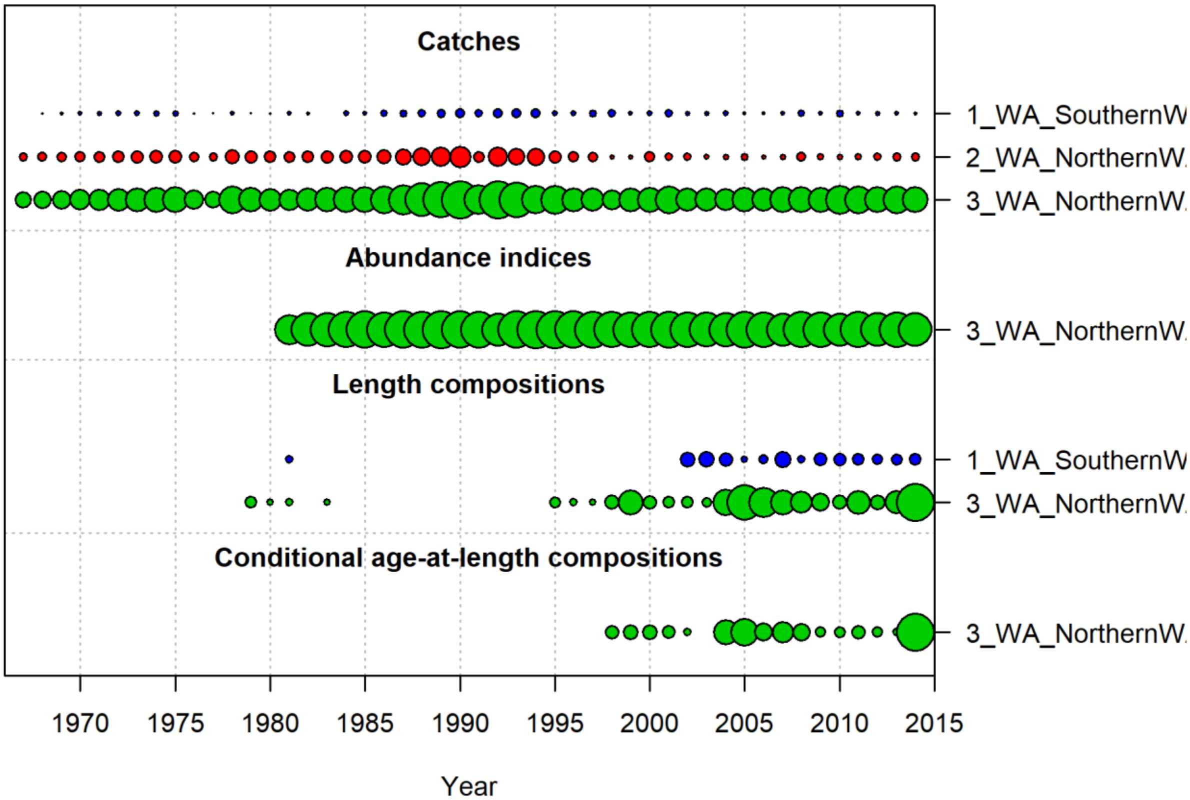

Figure 2.5. Summary of the data type, length of data time series, and quantity for China rockfish (North). The size of the bubble by year is based on the sample size (e.g., larger bubbles indicate higher sample sizes). The bubble size for the indices is equal to the inverse of the mean standard error for the survey.

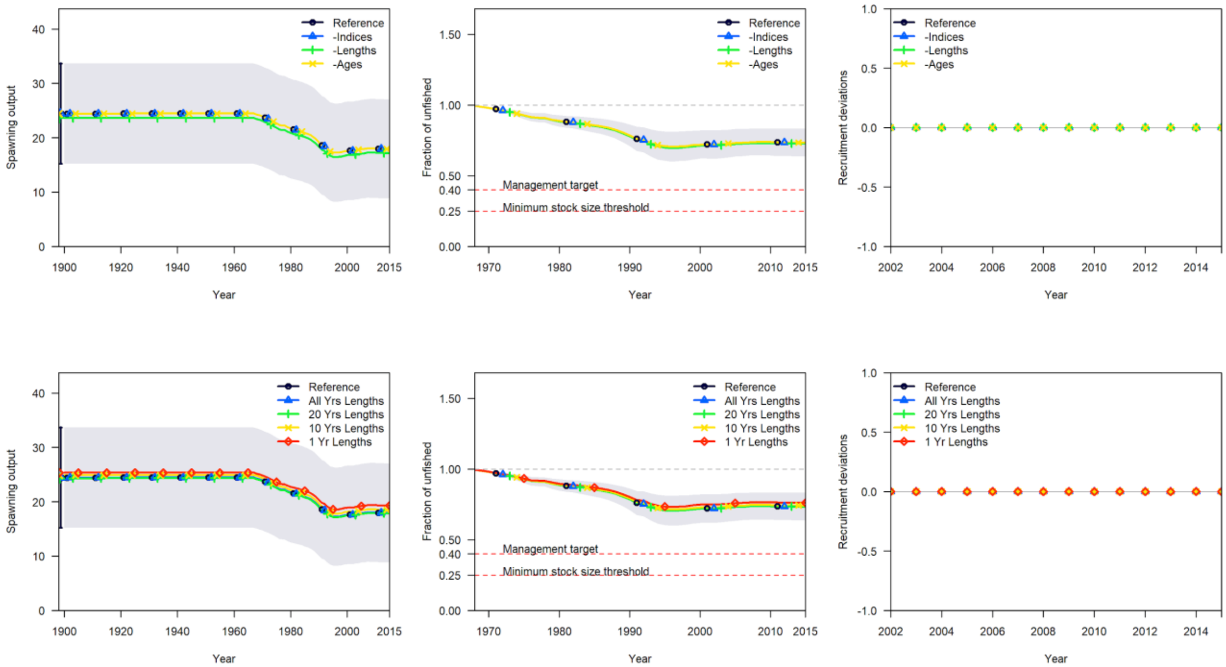

Figure 2.6. Comparison of model estimates of spawning biomass, the fraction of unfished, and annual recruitment deviations (not estimated in the model) relative to the reference model (black) for China rockfish. The top row are comparisons between the reference model and when all indices (blue), all lengths (green), or all ages (yellow) are removed. The bottom row are comparisons between the reference model and when only catches and all lengths (blue), 20 years of lengths (green), 10 years of length (yellow), or only 1 year of lengths (red) are used in the model.

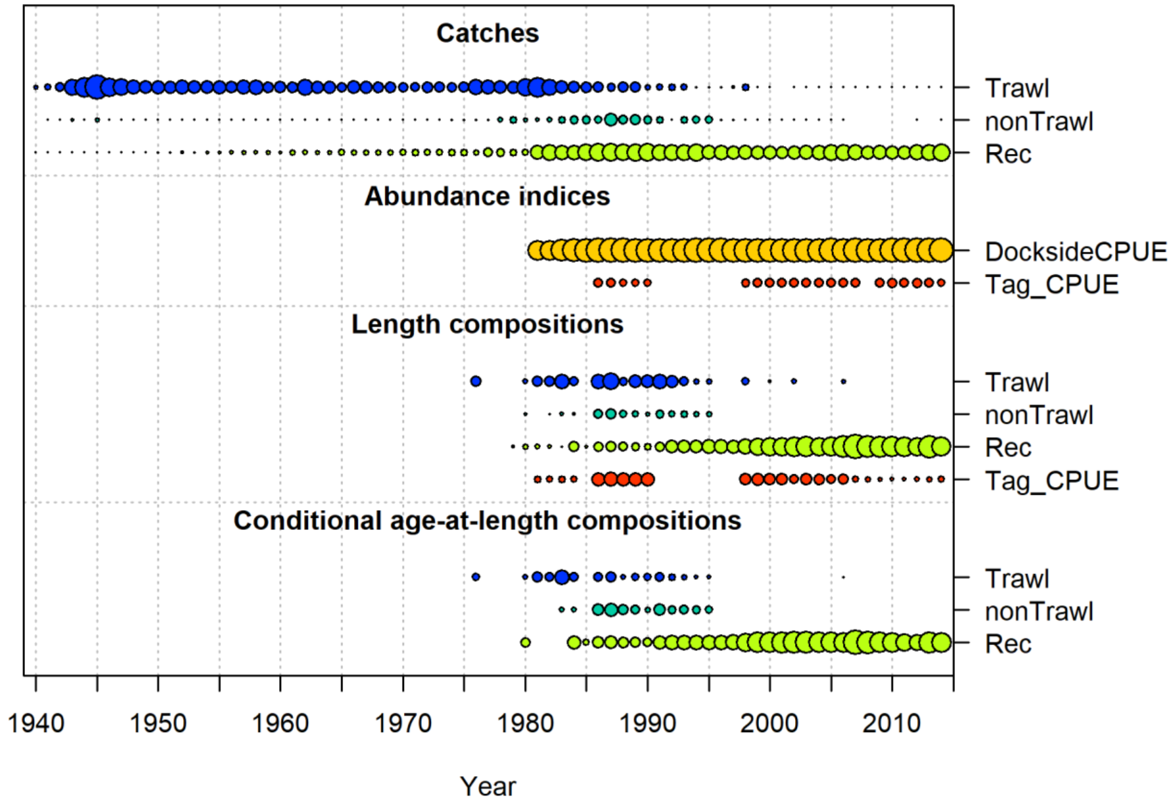

Figure 2.7. Summary of the data type, length of data time series, and quantity for black rockfish (WA). The size of the bubble by year is based on the sample size (e.g., larger bubbles indicate higher sample sizes). The bubble size for the indices is equal to the inverse of the mean standard error for the survey.

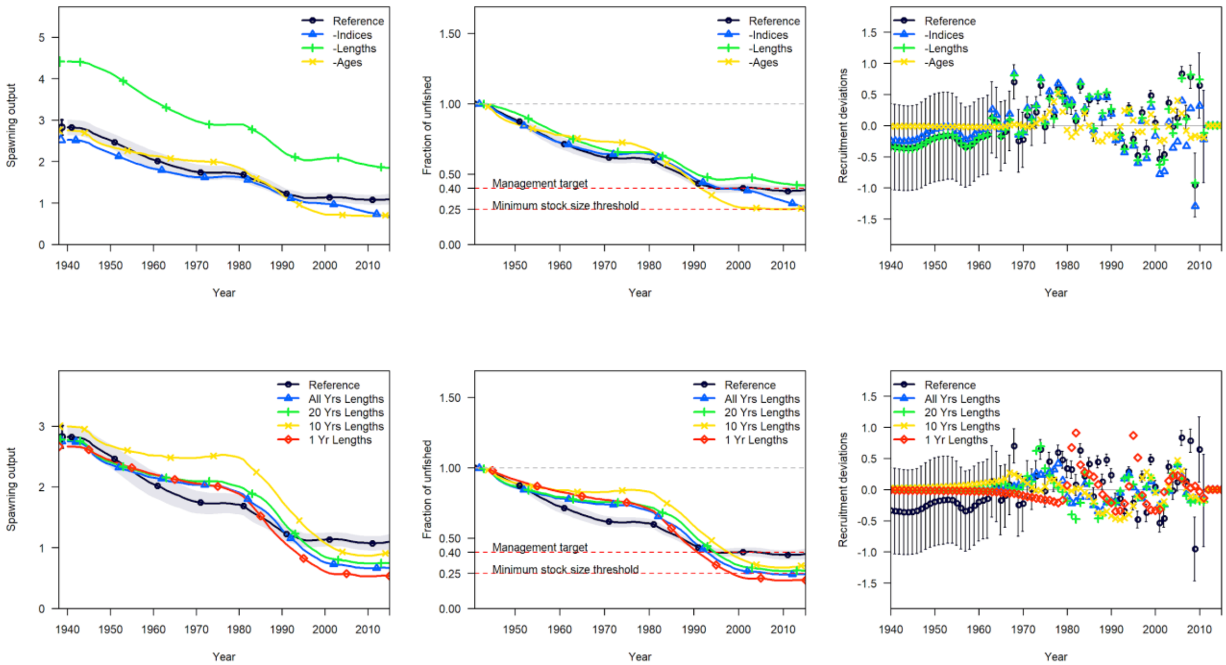

Figure 2.8. Comparison of model estimates of spawning biomass, the fraction of unfished, and annual recruitment deviations relative to the reference model (black) for black rockfish. The top row are comparisons between the reference model and when all indices (blue), all lengths (green), or all ages (yellow) are removed. The bottom row are comparisons between the reference model and when only catches and all lengths (blue), 20 years of lengths (green), 10 years of length (yellow), or only 1 year of lengths (red) are used in the model.

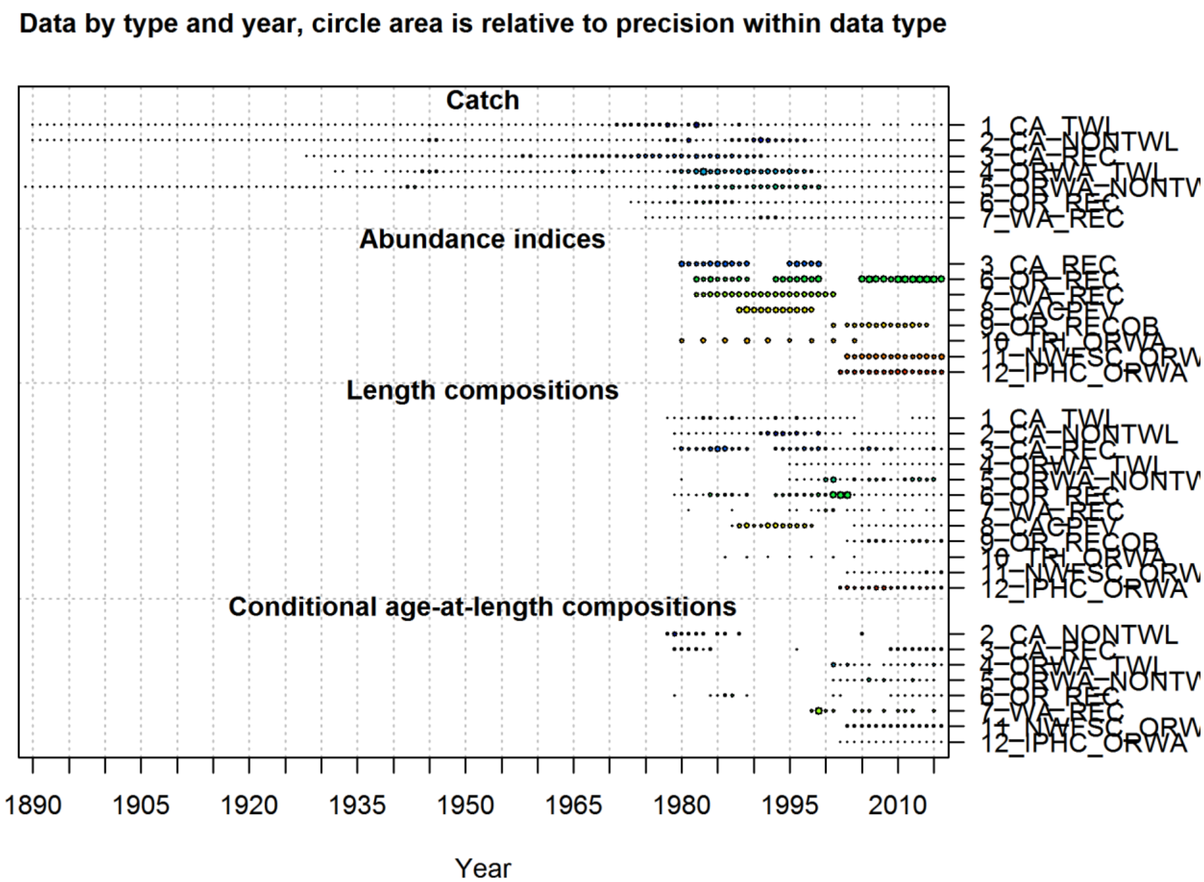

Figure 2.9. Summary of the data type, length of data time series, and quantity for yelloweye rockfish. The size of the bubble by year is based on the sample size (e.g., larger bubbles indicate higher sample sizes). The bubble size for the indices is equal to the inverse of the mean standard error for the survey.

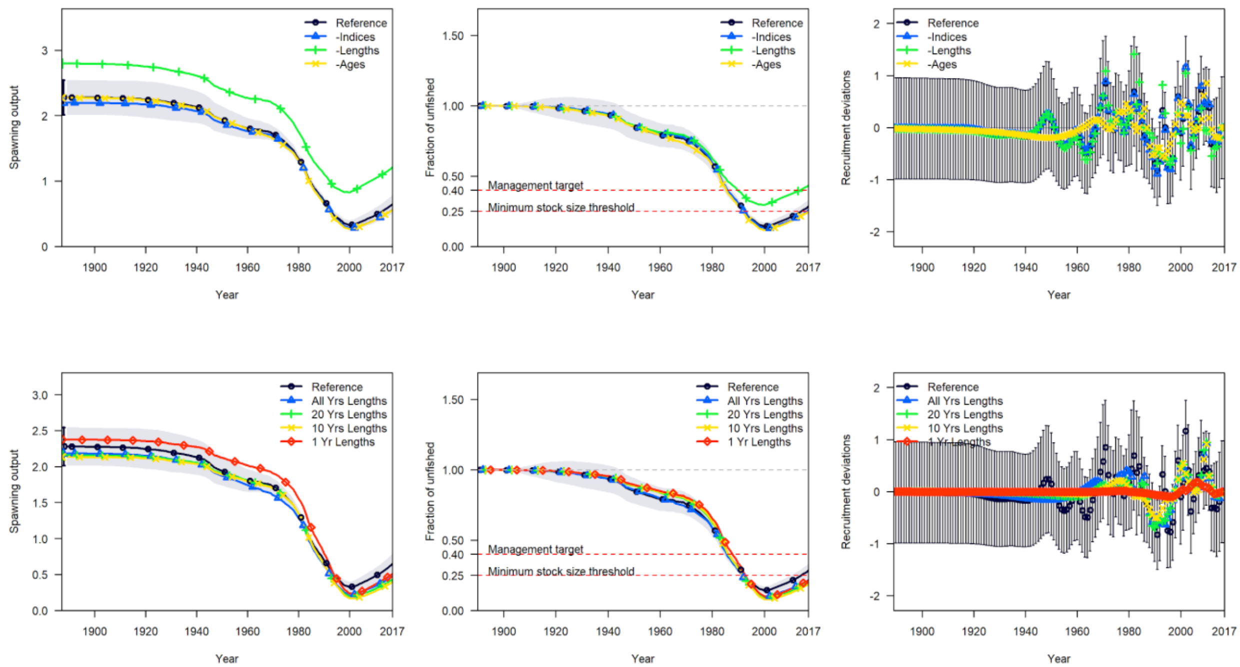

Figure 2.10. Comparison of model estimates of spawning biomass, the fraction of unfished, and annual recruitment deviations relative to the reference model (black) for yelloweye rockfish. The top row are comparisons between the reference model and when all indices (blue), all lengths (green), or all ages (yellow) are removed. The bottom row are comparisons between the reference model and when only catches and all lengths (blue), 20 years of lengths (green), 10 years of length (yellow), or only 1 year of lengths (red) are used in the model.

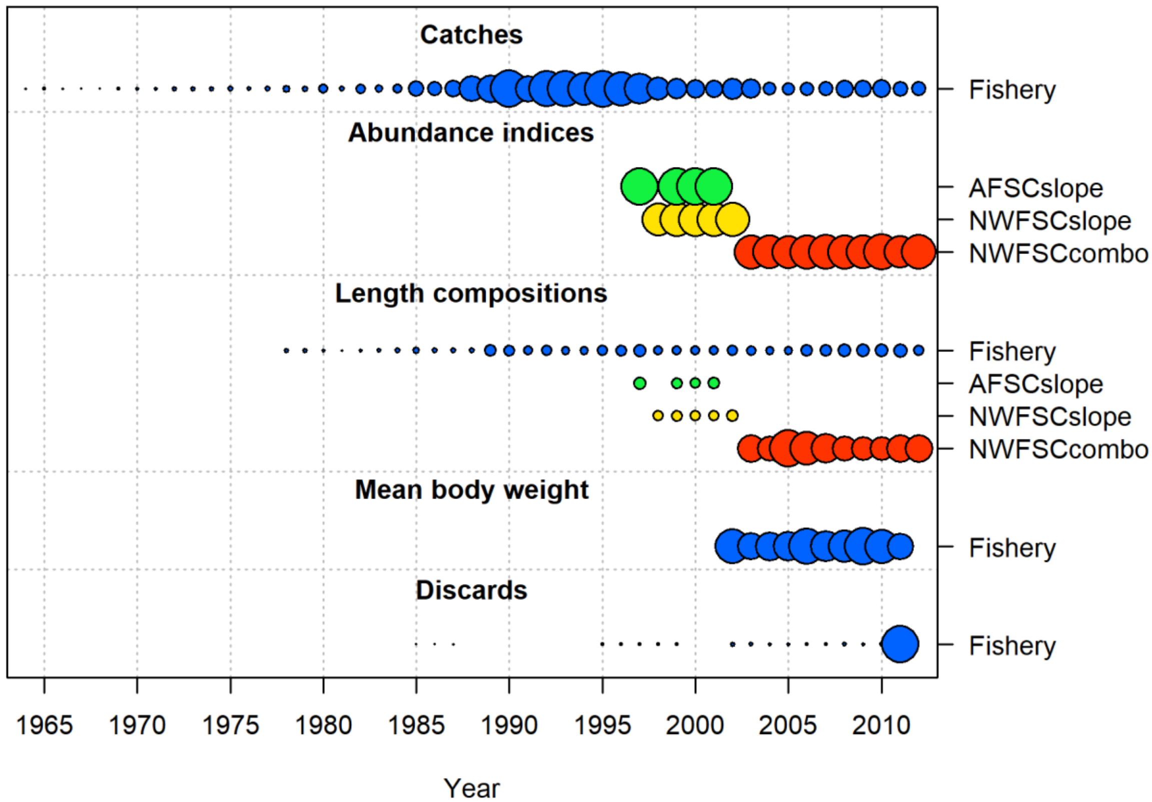

Figure 2.11. Summary of the data type, length of data time series, and quantity for longspine thornyhead. The size of the bubble by year is based on the sample size (e.g., larger bubbles indicate higher sample sizes). The bubble size for the indices is equal to the inverse of the mean standard error for the survey.

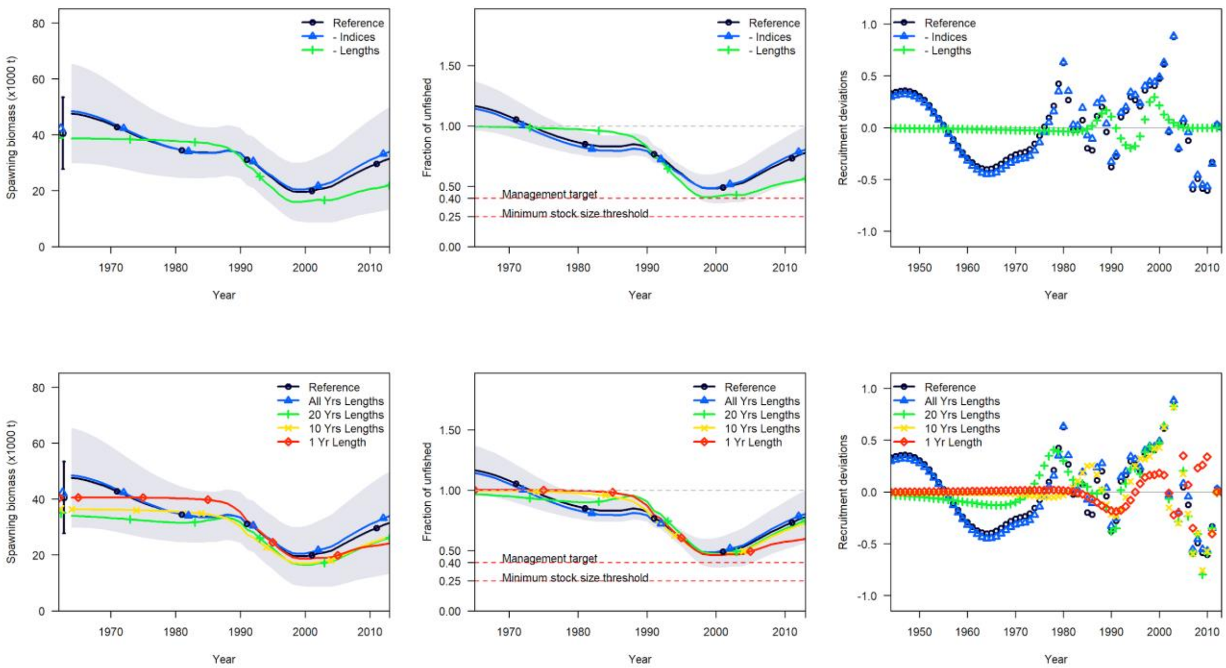

Figure 2.12. Comparison of model estimates of spawning biomass, the fraction of unfished, and annual recruitment deviations relative to the reference model (black) for longspine thornyhead. The top row are comparisons between the reference model and when all indices (blue) or all lengths (green). The bottom row are comparisons between the reference model and when only catches and all lengths (blue), 20 years of lengths (green), 10 years of length (yellow), or only 1 year of lengths (red) are used in the model.

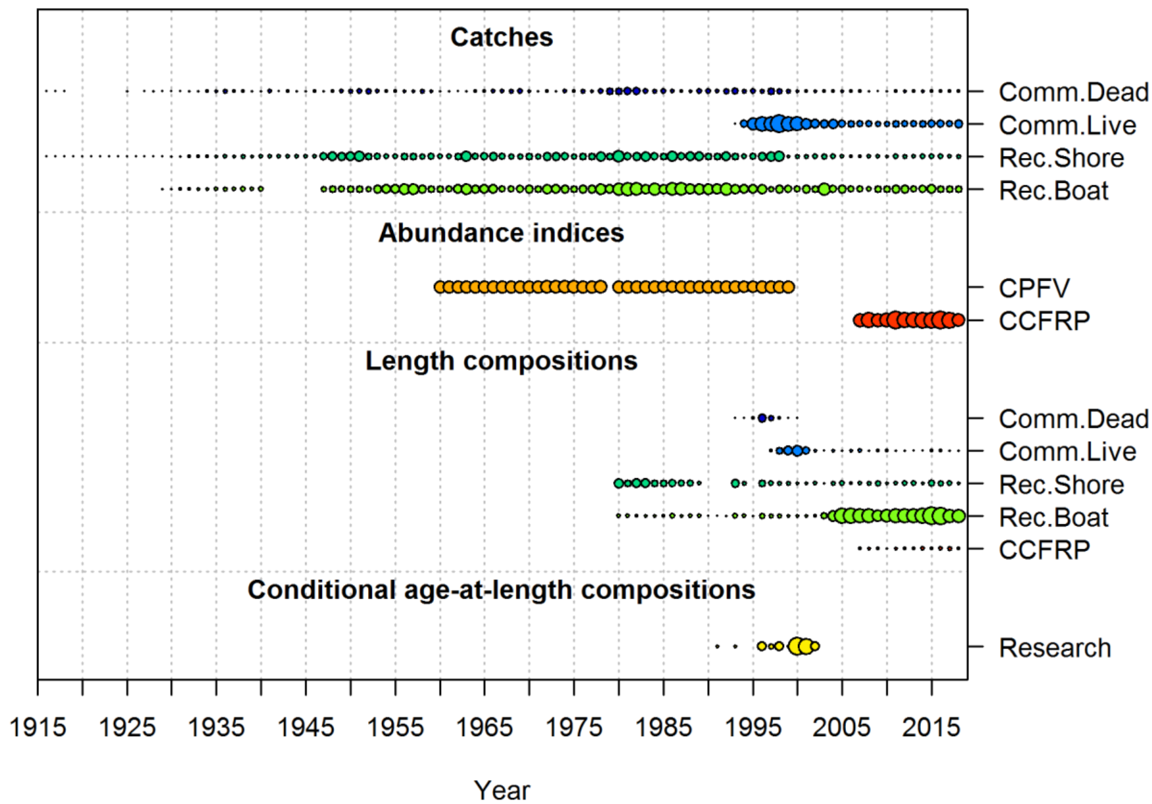

Figure 2.13. Summary of the data type, length of data time series, and quantity for cabezon (NCS). The size of the bubble by year is based on the sample size (e.g., larger bubbles indicate higher sample sizes). The bubble size for the indices is equal to the inverse of the mean standard error for the survey.

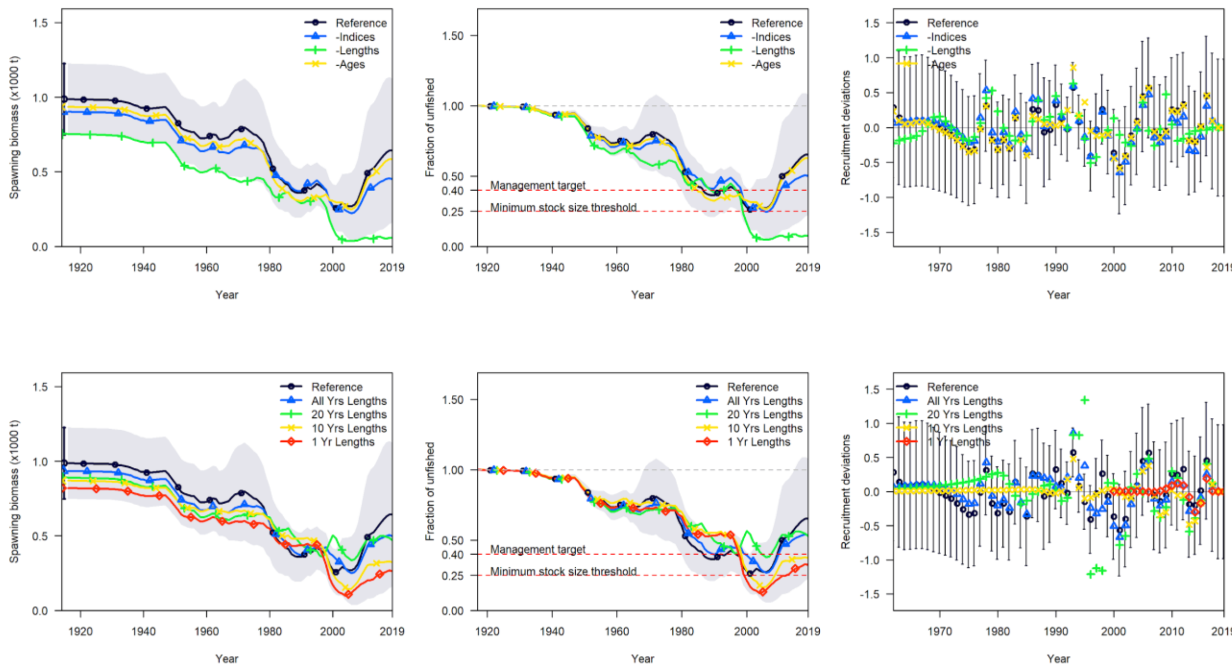

Figure 2.14. Comparison of model estimates of spawning biomass, the fraction of unfished, and annual recruitment deviations relative to the reference model (black) for cabezon (NCS). The top row are comparisons between the reference model and when all indices (blue), all lengths (green), or all ages (yellow) are removed. The bottom row are comparisons between the reference model and when only catches and all lengths (blue), 20 years of lengths (green), 10 years of length (yellow), or only 1 year of lengths (red) are used in the model.

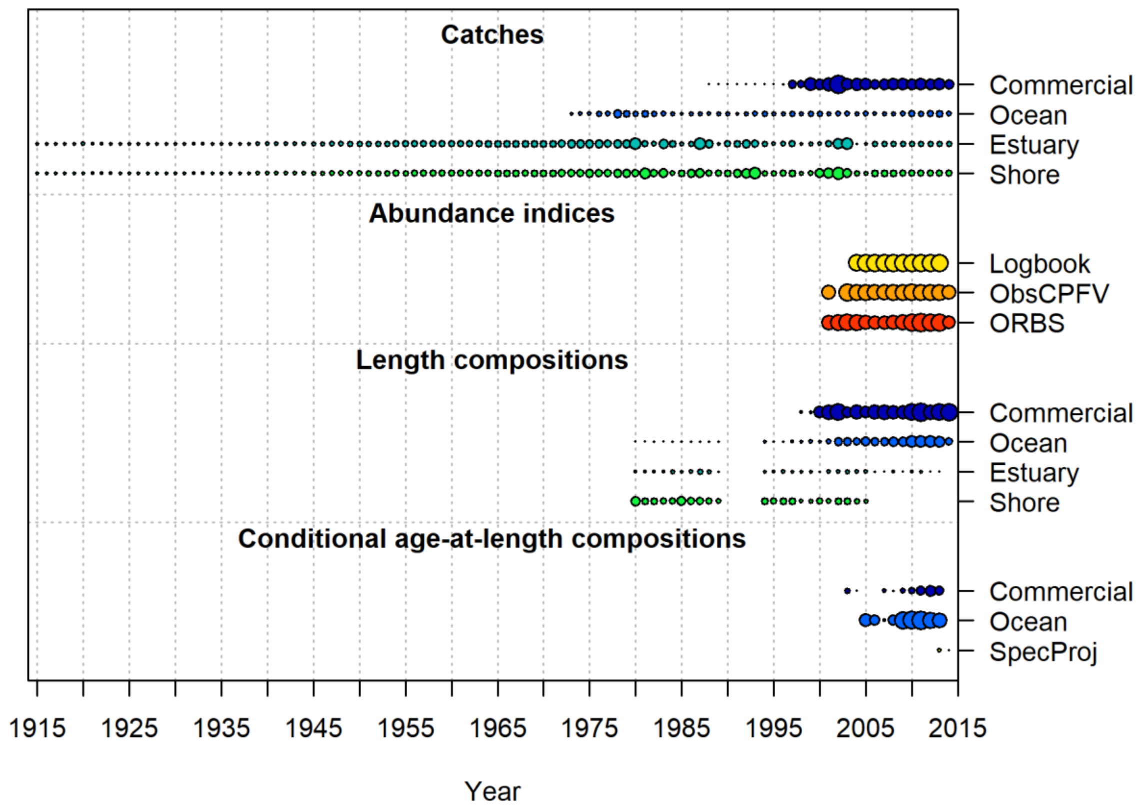

Figure 2.15. Summary of the data type, length of data time series, and quantity for kelp greenling. The size of the bubble by year is based on the sample size (e.g., larger bubbles indicate higher sample sizes). The bubble size for the indices is equal to the inverse of the mean standard error for the survey.

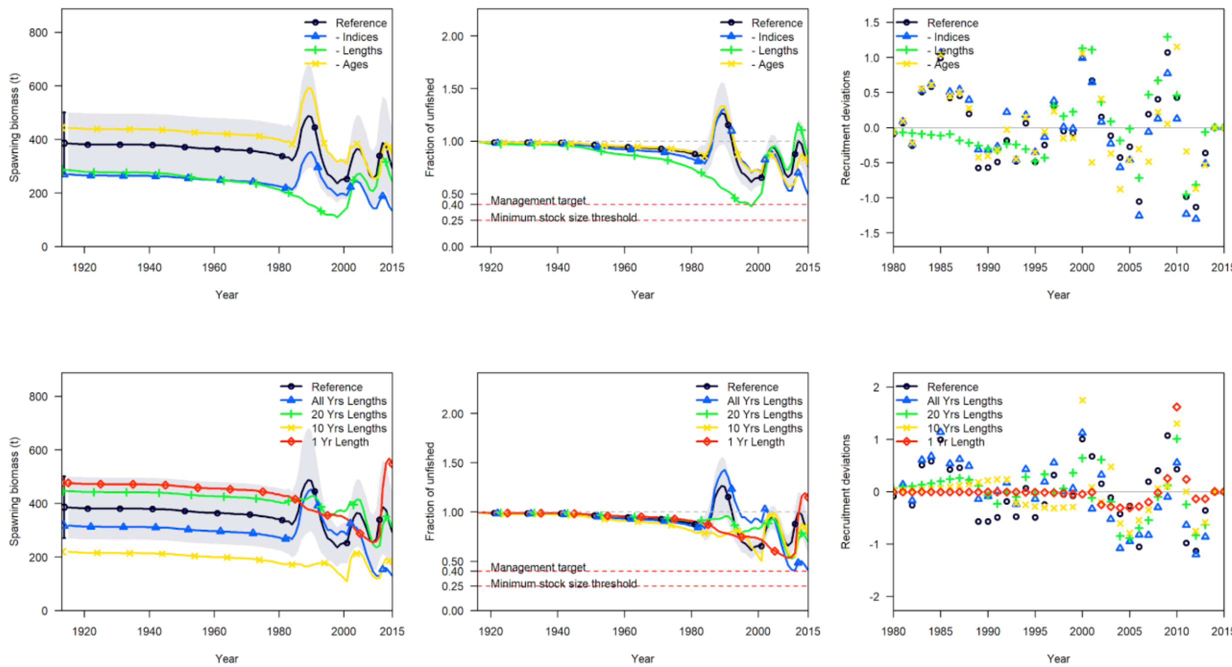

Figure 2.16. Comparison of model estimates of spawning biomass, the fraction of unfished, and annual recruitment deviations relative to the reference model (black) for kelp greenling. The top row are comparisons between the reference model and when all indices (blue), all lengths (green), or all ages (yellow) are removed. The bottom row are comparisons between the reference model and when only catches and all lengths (blue), 20 years of lengths (green), 10 years of length (yellow), or only 1 year of lengths (red) are used in the model.

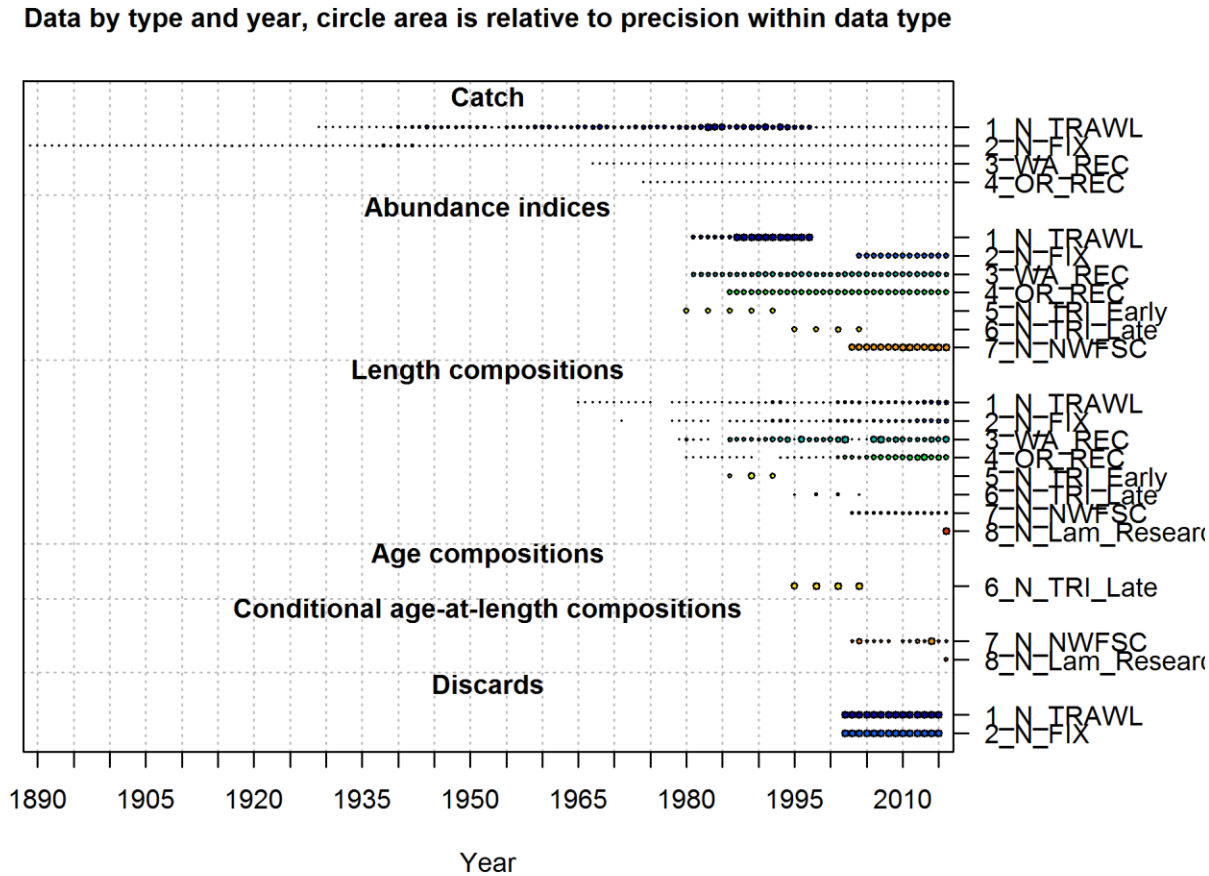

Figure 2.17. Summary of the data type, length of data time series, and quantity for lingcod (North). The size of the bubble by year is based on the sample size (e.g., larger bubbles indicate higher sample sizes). The bubble size for the indices is equal to the inverse of the mean standard error for the survey.

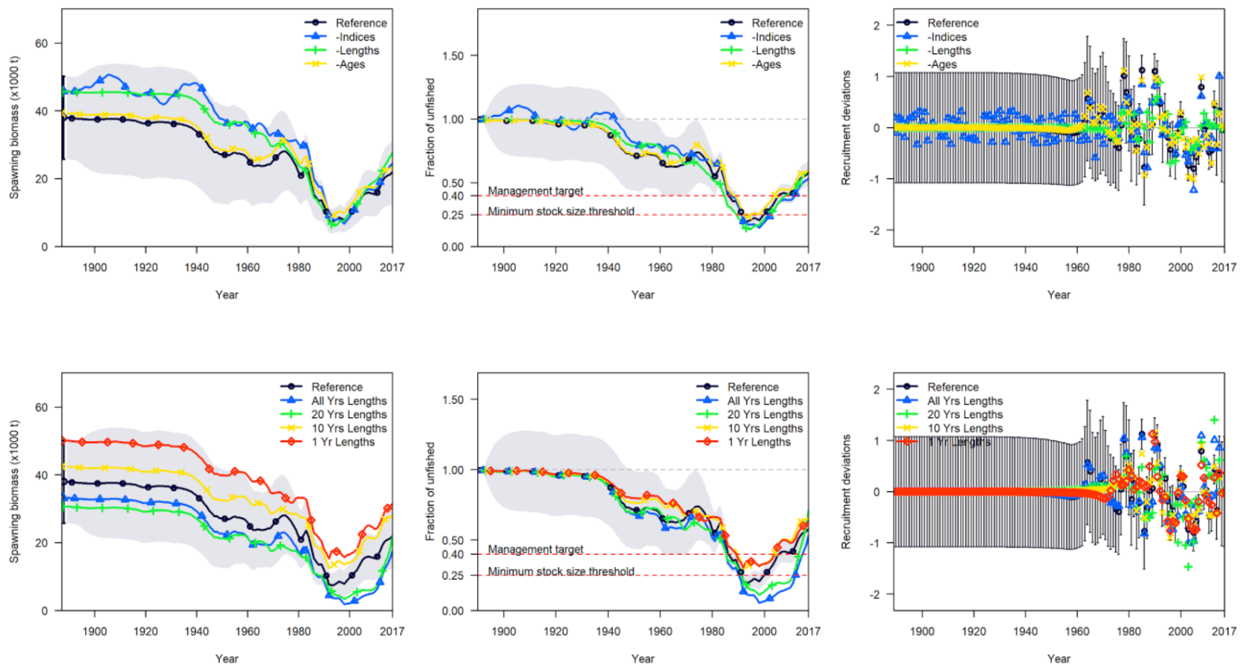

Figure 2.18. Comparison of model estimates of spawning biomass, the fraction of unfished, and annual recruitment deviations relative to the reference model (black) for lingcod (North). The top row are comparisons between the reference model and when all indices (blue), all lengths (green), or all ages (yellow) are removed. The bottom row are comparisons between the reference model and when only catches and all lengths (blue), 20 years of lengths (green), 10 years of length (yellow), or only 1 year of lengths (red) are used in the model.

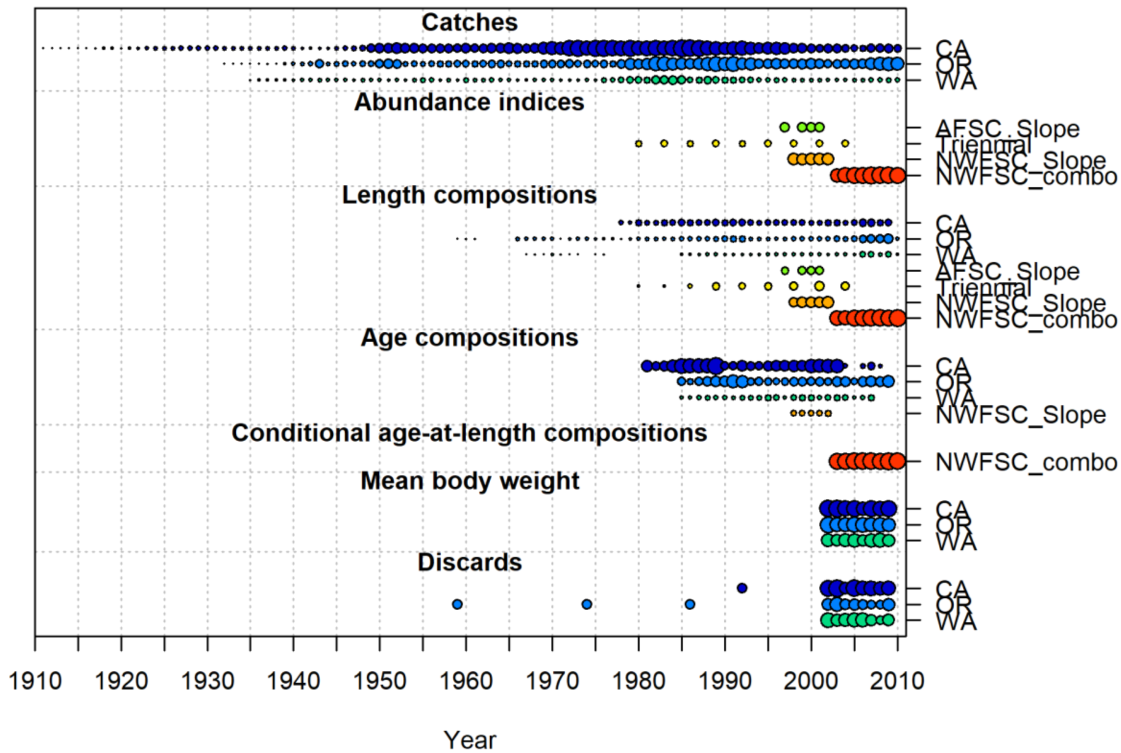

Figure 2.19. Summary of the data type, length of data time series, and quantity for Dover sole. The size of the bubble by year is based on the sample size (e.g., larger bubbles indicate higher sample sizes). The bubble size for the indices is equal to the inverse of the mean standard error for the survey.

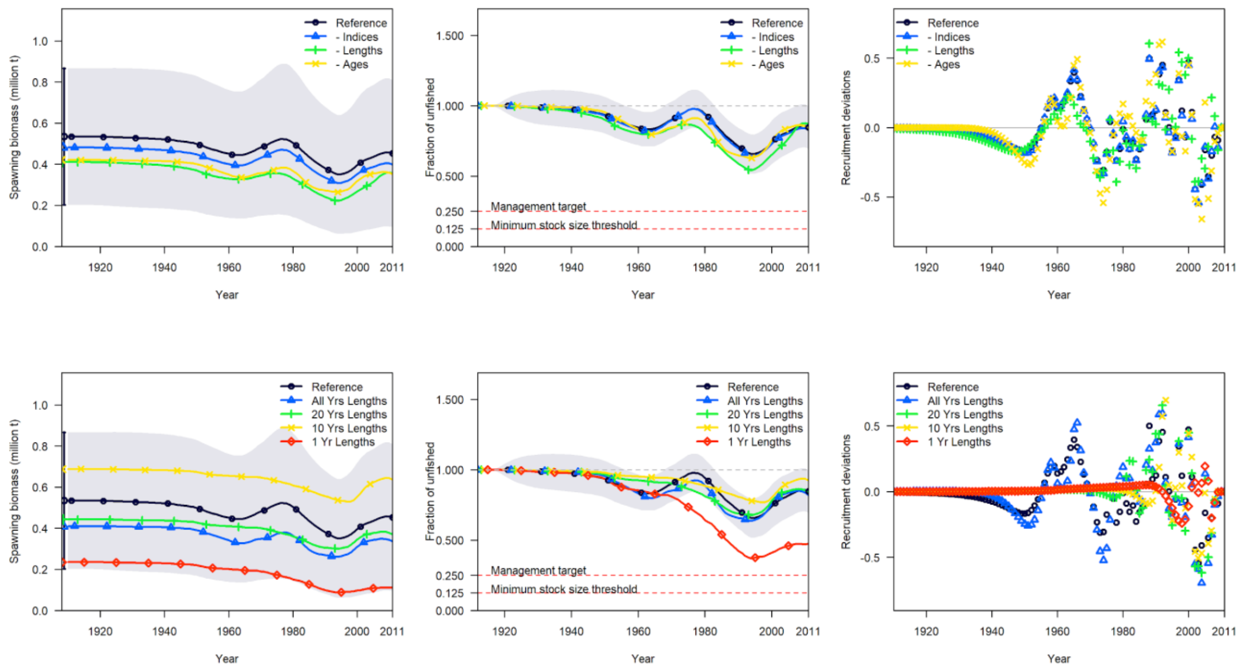

Figure 2.20. Comparison of model estimates of spawning biomass, the fraction of unfished, and annual recruitment deviations relative to the reference model (black) for Dover sole. The top row are comparisons between the reference model and when all indices (blue), all lengths (green), or all ages (yellow) are removed. The bottom row are comparisons between the reference model and when only catches and all lengths (blue), 20 years of lengths (green), 10 years of length (yellow), or only 1 year of lengths (red) are used in the model.

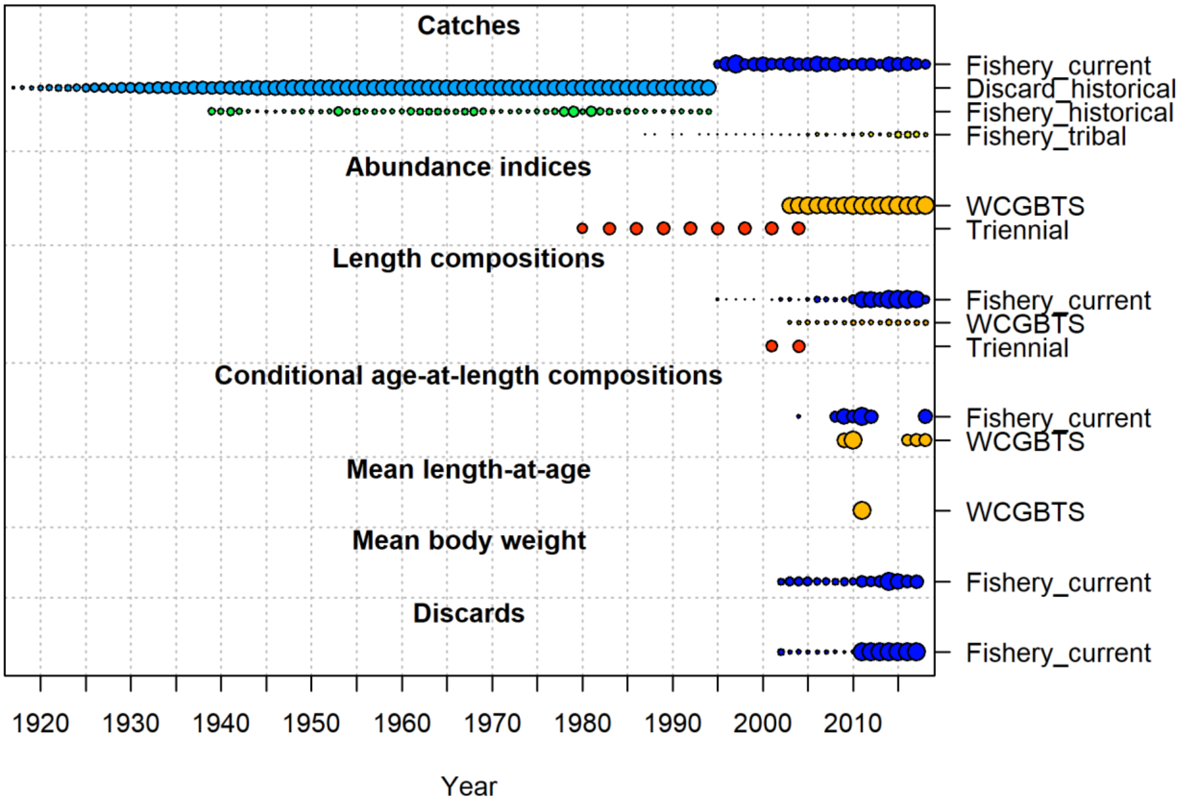

Figure 2.21. Summary of the data type, length of data time series, and quantity for big skate. The size of the bubble by year is based on the sample size (e.g., larger bubbles indicate higher sample sizes). The bubble size for the indices is equal to the inverse of the mean standard error for the survey.

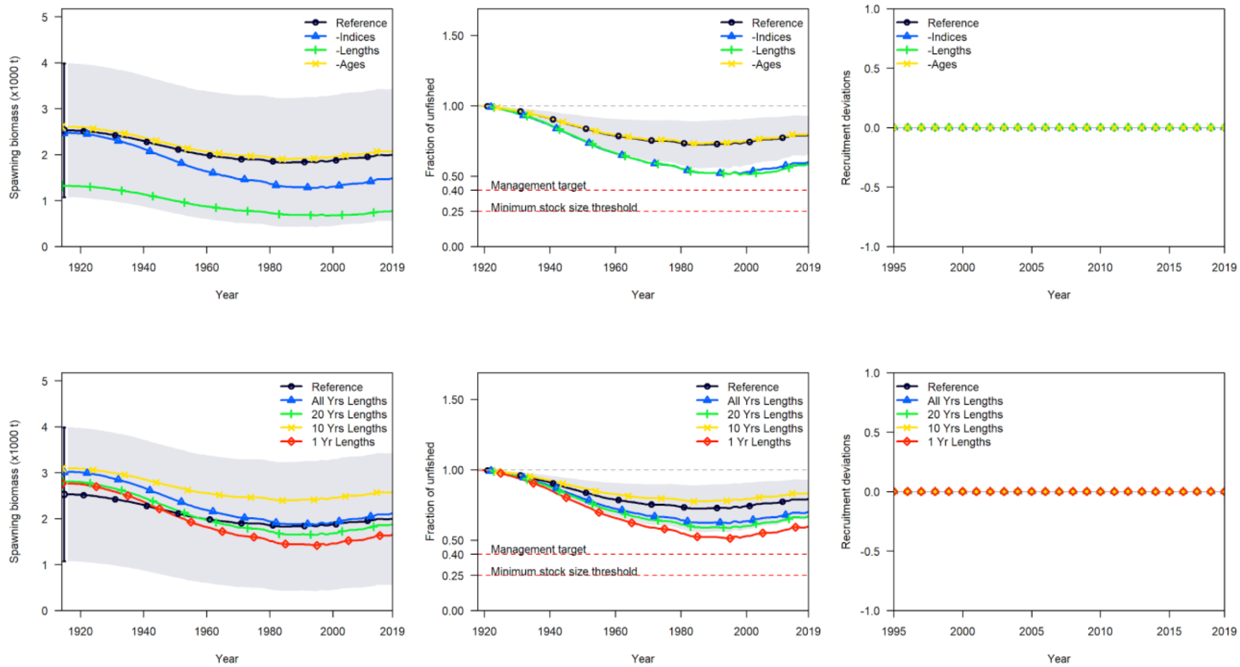

Figure 2.22. Comparison of model estimates of spawning biomass, the fraction of unfished, and annual recruitment deviations (not estimated in the model) relative to the reference model (black) for big skate. The top row are comparisons between the reference model and when all indices (blue), all lengths (green), or all ages (yellow) are removed. The bottom row are comparisons between the reference model and when only catches and all lengths (blue), 20 years of lengths (green), 10 years of length (yellow), or only 1 year of lengths (red) are used in the model.

Overall Conclusions

The ability to use only length data with a catch times series, assuming life history parameters can adequately be estimated outside the model, has been shown to provide informative, science-based estimates of important stock assessment outputs for fisheries management. Simulation testing across several data scenarios and life history specifications demonstrated generally unbiased results regardless of the time series or sample sizes explored. Data-limitations down to 1 year and 50 samples still provided useful model output. The major differences came in the level of imprecision, which grows consistently as the time series and sample size decrease. The approach of reducing the time series of length data in fishery management-approved stock assessments, despite not being able to control for other aspects of data quality throughout the time series, supported the findings of the simulation testing. Missing in these comparisons are repeated measures of a realized stock assessment (which is done with the 100 iterations of the simulations), but this approach does mimic the real approach to inform fisheries management.

For consideration of category 2 stock assessment designation, simulation testing suggests that even 10 years of length data is sufficient to qualify. While only 1 year of data is an extremely data-limited case, the level of imprecision and bias may better be considered a category 3 stock assessment. Stock assessments with somewhere between 1 and 10 years of length data may need further consideration as to the break between category 2 and 3.

References

Anderson, S.C. Monnahan, C.C., Johnson, K.F., Ono, K., Valero, J.L. 2014. ss3sim: An R package for fisheries stock assessment simulation with Stock Synthesis. PLoS ONE 9(4): e92725. https://doi.org/10.1371/journal.pone.0092725

Cope, J.M., 2013. Implementing a statistical catch-at-age model (Stock Synthesis) as a tool for deriving overfishing limits in data-limited situations. Fisheries Research 142, 3–14. https://doi.org/10.1016/j.fishres.2012.03.006

Cope, J.M., Sampson, D., Stephens, A., Key, M., Mirick, P.P., Stachura, M., Tsou, T.S. Weyland, P., Berger, A., Buell, T. Councill, E., Dick, E.J., Fenske, K.H., Monk, M., Rodomsky, B.T. 2016. Assessments of California, Oregon, and Washington stocks of black rockfish (*Sebastes melanops*) in 2015. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220.

Cope, J.M, Berger, A.M., Whitman, A.D. Budrick, J.E., Bosley, K.M., Tsou, T.S., Niles, C.B., Privitera-Johnson, K., Hillier, L.K., Hinton, K.E., Wilson, M.N. 2019. Assessing cabezon (*Scorpaenichthys marmoratus*) stocks in waters off of California and Oregon, with catch limit estimation for Washington state. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220.

Cope, J.M., Thorson, J.T., Wetzel, C.R., DeVore, J., 2015. Evaluating a prior on relative stock status using simplified age-structured models. Fisheries Research, Development, testing, and evaluation of data-poor assessment and fisheries management methods 171, 101–109. https://doi.org/10.1016/j.fishres.2014.07.018

Dick, E.J., MacCall, A.D., 2011. Depletion-Based Stock Reduction Analysis: A catch-based method for determining sustainable yields for data-poor fish stocks. Fisheries Research 110, 331–341. https://doi.org/10.1016/j.fishres.2011.05.007

Dick, E.J., Monk, M., Taylor, I., Haltuch, M., Tsou, T.S., Mirick, P. 2016. Status of China rockfish off the U.S. Pacific Coast in 2015. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220.

Francis, R.I.C.C. 2017. Revisiting data weighting in fisheries stock assessment models. Fisheries Research 192, 5–15. https://doi.org/10.1016/j.fishres.2016.06.006

Gertseva, V., Cope, J.M., 2017. Stock assessment of the yelloweye rockfish (*Sebastes ruberrimus*) in the state and federal waters off California, Oregon, and Washington. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220.

Haltuch, M.A., Wallace, J., Allen Akselrud, C., Nowlis, J., Barnett, L.A.K., Valero, J.L., Tsou, T.S., Lam, L. 2017. 2017 lingcod assessment. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220.

Hamel, O.S., 2015. A method for calculating a meta-analytical prior for the natural mortality rate using multiple life history correlates. ICES Journal of Marine Science. https://doi.org/10.1093/icesjms/fsu131

Hick, A. C., Wetzel, C.R. 2011. The status of Dover sole (*Microstomus pacificus*) along the U.S. West Coast in 2011. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220.

Maunder, M.N., Crone, P.R., Punt, A.E., Valero, J.L., Semmens, B.X., 2017. Data conflict and weighting, likelihood functions and process error. Fisheries Research 192, 1–4. https://doi.org/10.1016/j.fishres.2017.03.006

Maunder, M.N., Piner, K.R., 2017. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Fisheries Research 192, 16–27. https://doi.org/10.1016/j.fishres.2016.04.022

Methot, R.D., Taylor, I.G., 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Canadian Journal of Fisheries and Aquatic Sciences. https://doi.org/10.1139/f2011-092

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

Monk, M.H., He, X. 2019. The combined status of gopher (*Sebastes carnatus*) and black-and-yellow rockfishes (*Sebastes chrysomelas*) in U.S. wasters off California in 2019. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220.

Monnahan, C.C., Branch, T.A., Thorson, J.T., Stewart, I.J., Szuwalski, C.S. 2019. Overcoming long Bayesian run times in integrated fisheries stock assessments. ICES J Mar Sci. https://doi.org/10.1093/icesjms/fsz059

Ono, K., Punt, A.E., Rivot, E., 2012. Model performance analysis for Bayesian biomass dynamics models using bias, precision and reliability metrics. Fisheries Research. https://doi.org/10.1016/j.fishres.2012.02.022

Pacific Fishery Management Council. 2011. Assessment methods for data-poor stocks: Report of the review panel meeting. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220.

Piner, K.R., Lee, H.-H., Maunder, M.N., 2016. Evaluation of using random-at-length observations and an equilibrium approximation of the population age structure in fitting the von Bertalanffy growth function. Fisheries Research 180, 128–137. https://doi.org/10.1016/j.fishres.2015.05.024

Punt, A.E., 2017. Some insights into data weighting in integrated stock assessments. Fisheries Research 192, 52–65. https://doi.org/10.1016/j.fishres.2015.12.006

Rudd, M.B., Thorson, J.T. 2018. Accounting for variable recruitment and fishing mortality in length-based stock assessments for data-limited fisheries. Canadian Journal of Fisheries and Aquatic Sciences 75(7): 1019-1035.

Stephens, A., Taylor, I.G. 2013. Stock assessment and status of longspine thornyhead (*Sebastolobus altivelis*) off California, Oregon, Washington in 2013. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220.

Taylor, I.G., Gertseva, V., Stephens, A., Bizzarro J. 2019. Status of big skate (*Beringraja binoculata*) off the U.S. Pacific Coast in 2019. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220.

Thorson, J.T., Cope, J.M., 2015. Catch curve stock-reduction analysis: An alternative solution to the catch equations. Fisheries Research. https://doi.org/10.1016/j.fishres.2014.03.024

Thorson, J.T., Cope, J.M., 2017. Uniform, uninformed or misinformed?: The lingering challenge of minimally informative priors in data-limited Bayesian stock assessments. Fisheries Research 194, 164–172. https://doi.org/10.1016/j.fishres.2017.06.007

Wang, S.-P., Maunder, M.N., 2017. Is down-weighting composition data adequate for dealing with model misspecification, or do we need to fix the model? Fisheries Research 192, 41–51. https://doi.org/10.1016/j.fishres.2016.12.005

Wetzel, C.R., Punt, A.E., 2015. Evaluating the performance of data-moderate and catch-only assessment methods for U.S. west coast groundfish. Fisheries Research, Development, testing, and evaluation of data-poor assessment and fisheries management methods 171, 170–187. https://doi.org/10.1016/j.fishres.2015.06.005

Wetzel, C.R., Cronin-Fine, L., Johnson, K.F. 2017. Status of Pacific ocean perch (*Sebastes alutus*) along the U.S. west coast in 2017. Pacific Fishery Management Council, 7700 Ambassador Place NE, Suite 200, Portland, OR 97220.

Appendix 1. Category 1 life history parameters

Table A1. Summary of life history parameter values used in category 1 stocks through 2018.

Stock	M	k	M/k	Linf	Lmat	Lmat/Linf	SigmaR
Aurora rockfish	0.035	0.092	0.382	30.768	25.540	0.830	0.500
Black rockfish (CA)	0.181	0.152	1.192	54.541	43.690	0.801	0.500
Black rockfish (WA)	0.163	0.137	1.188	53.911	43.690	0.810	0.500
Bocaccio	0.178	0.226	0.786	67.340	37.700	0.560	1.000
Cabezon (ORS)	0.250	0.191	1.309	66.108	43.700	0.661	0.500
Cabezon (NCA)	0.250	0.149	1.674	60.194	34.600	0.575	0.500
Cabezon (SCA)	0.250	0.126	1.984	64.000	34.600	0.541	0.700
California Scorpionfish	0.235	0.292	0.806	31.886	18.000	0.565	0.600
Canary rockfish	0.088	0.129	0.684	60.038	40.500	0.675	0.500
Chilipepper	0.160	0.195	0.823	47.300	24.400	0.516	1.000
Darkblotched rockfish	0.054	0.198	0.273	42.662	34.590	0.811	0.750
Dover sole	0.117	0.150	0.779	47.792	35.000	0.732	0.350
Kelp greenling	0.360	0.261	1.377	36.439	29.340	0.805	0.650
Lingcod (N)	0.257	0.129	1.999	110.000	56.700	0.515	0.550
Lingcod (S)	0.257	0.129	1.996	93.562	52.300	0.559	0.750
Petrale sole	0.157	0.133	1.173	54.303	33.100	0.610	0.400

Pacific ocean perch	0.054	0.167	0.324	41.601	32.100	0.772	0.700
Sablefish	0.081	0.335	0.243	63.972	58.000	0.907	1.100
Splitnose rockfish	0.048	0.156	0.308	29.624	21.840	0.737	1.000
Widow rockfish	0.157	0.199	0.791	50.357	29.680	0.589	0.600
Yellowtail rockfish (N)	0.174	0.140	1.245	53.580	42.490	0.793	0.500
Yelloweye rockfish	0.044	0.065	0.676	63.454	42.071	0.663	0.500

Appendix 2. Simulations with sigma R = 0.4

Table A1. Number of model runs that did not converge out of 100 simulation replicates, based on the criteria that the maximum final gradient must be less than 1.0 and the estimated log(R_0) less than 12.

Length data scenario		Life history scenario			
Sample size	Number of years	Short-lived, slow-growing	Short-lived, fast-growing	Long-lived, slow-growing	Long-lived, fast-growing
N = 1000 (perfectly specified, with index + ages)	100	0	1	0	0
N = 1000 (perfectly specified)	100	0	1	0	0
N = 200	75	0	0	0	0
N = 200	20	0	0	0	0
N = 200	10	0	0	0	0
N = 200	1	0	0	0	2
N = 50	75	0	0	0	0
N = 50	20	0	0	0	0
N = 50	10	0	0	0	0
N = 50	1	5	7	4	1
Decline from 200 samples (year $\leq$ 87) to 50 samples (year > 87)	20	0	0	0	0

Table A2. Median relative error (i.e., bias) of the final year fraction of unfished biomass across 100 simulation replicates by life history type and length data scenario with recruitment standard deviation = 0.4 with bias-adjusted recruitment. Cells are shaded green for bias less than 10%, yellow less than 20%, and red greater than 20%.

Length data scenario		Life history scenario			
Sample size	Number of years	Short-lived, slow-growing	Short-lived, fast-growing	Long-lived, slow-growing	Long-lived, fast-growing
N = 1000 (perfectly specified, with index + ages)	100	-0.0050	-0.0190	-0.0170	-0.0143
N = 1000 (perfectly specified)	100	0.0047	0.0123	0.0090	0.0279
N = 200	75	-0.0418	-0.0947	-0.0364	-0.0359
N = 200	20	-0.0182	-0.0430	-0.0650	-0.0218
N = 200	10	-0.0485	-0.1006	-0.1650	-0.0307
N = 200	1	0.1038	-0.0841	-0.5292	0.1360
N = 50	75	0.0315	-0.0352	0.1314	0.0534
N = 50	20	0.0087	-0.0156	0.0435	0.0064
N = 50	10	0.0282	-0.0554	0.0308	-0.1208
N = 50	1	-0.4091	0.2789	0.0833	0.2465
Decline from 200 samples (year $\leq$ 87) to 50 samples (year > 87)	20	0.0542	-0.0531	0.0884	-0.0286

Table A3. Median relative error (i.e., bias) of the overfishing limit (OFL) in the first forecast year across 100 simulation replicates by life history type and length data scenario with recruitment standard deviation = 0.4 with bias-adjusted recruitment. Cells are shaded green for bias less than 10%, yellow less than 20%, and red greater than 20%.

Length data scenario		Life history scenario			
Sample size	Number of years	Short-lived, slow-growing	Short-lived, fast-growing	Long-lived, slow-growing	Long-lived, fast-growing
N = 1000 (perfectly specified, with index + ages)	100	0.0004	-0.0120	0.0067	-0.0040
N = 1000 (perfectly specified)	100	0.0175	0.0177	0.0428	0.0519
N = 200	75	-0.0626	-0.1005	-0.0473	-0.0238
N = 200	20	-0.0037	-0.0352	-0.0777	-0.0181
N = 200	10	-0.0394	-0.0649	-0.1267	-0.0359
N = 200	1	0.0853	-0.0866	-0.5128	0.1489
N = 50	75	0.0036	-0.0607	0.1432	0.0344
N = 50	20	0.0034	-0.0382	0.0227	0.0120
N = 50	10	0.0085	-0.0050	0.0592	-0.0849
N = 50	1	-0.3235	0.3406	0.0975	0.2689
Decline from 200 samples (year $\leq$ 87) to 50 samples (year > 87)	20	0.0611	-0.0198	0.0954	-0.0056

Table A4. Median absolute relative error (i.e., precision) of the final year fraction of unfished biomass across 100 simulation replicates by life history type and length data scenario with recruitment standard deviation = 0.4 with bias-adjusted recruitment. Cells are shaded green for MARE less than 20%, yellow less than 30%, and red greater than 30%.

Length data scenario		Life history scenario			
Sample size	Number of years	Short-lived, slow-growing	Short-lived, fast-growing	Long-lived, slow-growing	Long-lived, fast-growing
N = 1000 (perfectly specified, with index + ages)	100	0.0381	0.0461	0.0407	0.0368
N = 1000 (perfectly specified)	100	0.0991	0.1232	0.0815	0.0909
N = 200	75	0.1386	0.1510	0.0826	0.0998
N = 200	20	0.1338	0.1701	0.0937	0.1025
N = 200	10	0.1464	0.1602	0.1907	0.1218
N = 200	1	0.3324	0.1730	0.6166	0.2677
N = 50	75	0.1454	0.1670	0.1610	0.1034
N = 50	20	0.1575	0.1560	0.1237	0.1310
N = 50	10	0.1648	0.1909	0.1825	0.2145
N = 50	1	0.5330	0.3828	0.3730	0.3757
Decline from 200 samples (year ≤ 87) to 50 samples (year > 87)	20	0.1528	0.1566	0.1288	0.1018

Table A5. Median absolute relative error (i.e., precision) of the overfishing limit (OFL) in the first forecast year across 100 simulation replicates by life history type and length data scenario with recruitment standard deviation = 0.4 with bias-adjusted recruitment. Cells are shaded green for MARE less than 20%, yellow less than 30%, and red greater than 30%.

Length data scenario		Life history scenario			
Sample size	Number of years	Short-lived, slow-growing	Short-lived, fast-growing	Long-lived, slow-growing	Long-lived, fast-growing
N = 1000 (perfectly specified, with index + ages)	100	0.0311	0.0281	0.0288	0.0151
N = 1000 (perfectly specified)	100	0.1178	0.1428	0.0834	0.1087
N = 200	75	0.1560	0.1672	0.1063	0.1192
N = 200	20	0.1604	0.1874	0.1110	0.1311
N = 200	10	0.1572	0.1782	0.1700	0.1305
N = 200	1	0.3221	0.2154	0.6406	0.2985
N = 50	75	0.1648	0.2059	0.1651	0.1346
N = 50	20	0.1839	0.1796	0.1572	0.1709
N = 50	10	0.1912	0.2285	0.1556	0.2254
N = 50	1	0.5507	0.4932	0.3956	0.4491
Decline from 200 samples (year $\leq$ 87) to 50 samples (year > 87)	20	0.1761	0.2071	0.1452	0.1410

Table A6. Number of model runs out of 100 simulation replicates where the absolute relative error in final year fraction unfished biomass was greater than 1.0, given recruitment standard deviation 0.4.

Length data scenario		Life history scenario			
Sample size	Number of years	Short-lived, slow-growing	Short-lived, fast-growing	Long-lived, slow-growing	Long-lived, fast-growing
N = 1000 (perfectly specified, with index + ages)	100	0	0	0	0
N = 1000 (perfectly specified)	100	0	0	0	0
N = 200	75	0	0	0	
N = 200	20	0	0	0	0
N = 200	10	0	0	0	0
N = 200	1	3	0	2	2
N = 50	75	0	0	0	0
N = 50	20	0	0	0	0
N = 50	10	1	0	0	0
N = 50	1	7	13	14	5
Decline from 200 samples (year $\leq$ 87) to 50 samples (year > 87)	20	0	0	0	0