Best practices for modeling time-varying selectivity

Steven Martell and Ian Stewart
International Pacific Halibut Commission
2320 W Commodore Way, Suite 300,
Seattle WA, 98199-1287

2013-01-10

Abstract

Changes in the observed size- or age-composition of commercial catch can occur for a variety of reasons including: market demand, availability, temporal changes in growth, time-area closures, regulations, or change in fishing practice, to name but a few. Two common approaches for dealing with time-varying selectivity in assessment models are the use of discrete time-blocks associated with an epoch in the history of the fishery, or the use of penalized random walk models for parametric or non-parametric selectivity curves. Time block periods, or penalty weights associated with time-varying selectivity parameters, are subjective and often developed on an ad hoc basis. A factorial simulation-estimation experiment, with discrete or continuous changes in selectivity, is conducted to determine the best practices for modeling time-varying selectivity in fisheries stock assessments. Both the statistical properties of the assessment model and the policy implications of choosing the wrong model are taken into consideration.

Introduction

There are many reasons why fisheries selectivity may vary over time and the impact of ignoring changes in selectivity in age- or size-structured stock assessment models leads to biased estimates of abundance and mortality rates. Moreover, not accounting for changes in selectivity can lead to extremely optimistic projections in stock abundance (e.g., 2J3KL cod stocks, Walters and Maguire, 1996).

Many statistical catch-age models assume age-based selectivity whe in fact the underlying harvesting process is size-based. Changes in size-at-age associated with changes growth rates can have serious implications on the interpretation of age-based selectivity. Changing to length-based selectivity and using empirical length-at-age data can resolve some of the model misspecification; however, ontogentic movement of fish can also lead to changes in age-based selectivity when the distribution of fishing effort, or fish distribution relative to effort, changes over time. Recently the International Pacific Halibut Commission (IPHC) changed from using time-invariant age-based selectivity to time-varying age-based selectivity to account for both ontogeny and the changes in the relative stock distribution (Stewart et al., 2012).

Currently there are two general approaches for incorporating time-varying selectivity in stock assessment models; 1) the use of discrete time-blocks, and 2) continuous penalized random walk approach. The use of discrete time-blocks should be done *a priori*, where the specified time blocks represent periods consistent fishing practice, and a new block is specified when significant changes in fishing practice occur that may result in changes in selectivity. In practice, however, the time-blocks are also implemented *post hoc* to justify residual patterns in age- or size-composition data. To some, this practice seems rather subjective, and it is. Another discrete approach is to decompose the fisheries catch statistics into specific time periods that correspond to major transitions in fishing practice. For example, the BC herring fishery prior to 1970 was largely a reduction fishery where herring were harvested during the winter months using purse seines. After the collapse of the fishery in 1969, the fishery re-opened as a gill-net fishery targeting older sexually mature female herring for valuable roe. This change in fishing practice led to a significant change in the selectivity of the fishing gear.

The alternative approach is to allow for continuous changes in selectivity and model estimated selectivity parameters as a penalized random walk. In this case, specification of the variance parameter in how quickly selectivity is allowed to change is also somewhat subjective. It should also be noted that the choice of a time-invariant selectivity is also a subjective structural assumption of the assessment model, and this choice can also greatly influence model results, estimates of reference points, and result in bias forecasts.

Changes in fisheries selectivity also has implications for reference points based on maximum sustainable yield (MSY, Beverton and Holt, 1993). Trends towards catching smaller fish result in reductions in the harvest rate that would achieve MSY; therefore, it is important to account for changes in selectivity (and the associated uncertainty) when developing harvest policy for any given stock.

The over-arching objective is to determine if it is safer to assume more structural complexity in selectivity when the data are in fact simple and is it safer to assume simple structural complexity in selectivity when the data come from a fishery with dynamic changes in selectivity. In this paper, we conduct a series of simulation experiments using a factorial design with fixed selectivity, discrete changes in selectivity, and continuous changes in selectivity and compare statistical fit, retrospective bias, and estimated policy parameters using simulated data. We also explore the use of two-dimensional interpolation methods to reduce the number of estimated latent variables when selectivity is assumed to vary over time.

Methods

Simulated data were generated from an age-structured simulation model based on the 2010 Pacific hake assessment. Simulated data were based on 3 alternative selectivity scenarios: (1) constant over time, (2) selectivity changes at 4 specific blocked time-periods, and (3) that selectivity changes continuously over time where the commercial fishery targets the most abundant cohort in each year. First we describe the model structure used to simulate data and estimate model parameters, followed by a description of the MSY-based reference points, and lastly the detailed description of the various scenario combinations explored.

Model description

A statistical catch-age model was used to both generate simulated data sets and estimate model parameters based on simulated data. These simulation-estimation experiments were based on actual data from the Pacific hake fishery from 1977 to 2009, using the historical catch time series from US and Canada combined and the empirical weight-at-age data from this fishery (Martell, 2009). The model was written in AD Model Builder (Fournier et al., 2011) and all model code and data are available from a code repository (see CAPAM branch at `https://github.com/smartell/iSCAM`).

Input data for the model consist of historical removals along with age-composition information and empirical weight-at-age data from the commercial fishery. In addition to the commercial data, a fisheries independent survey also exists and includes a relative index of abundance and age-composition information. The actual acoustic survey for Pacific hake historically occurred every 3 years prior to 1995, then every two years, and since 2011 has occurred every year. For the simulation-estimation experiments we fisheries independent abundance and age-composition information exist for all years.

Parameters for the simulation-estimation experiments were based on the maximum likelihood estimates of the initial numbers-at-age and annual recruitment deviations from the actual assessment by (Martell, 2009). The annual relative abundance data was assumed to be proportional to the available biomass and to have log-normal measurement errors:

$$I_t = q e^{\sigma_1 \epsilon_t - 0.5\sigma_1^2} \sum_a \nu_a N_{a,t} W_a \tag{1}$$

where the random deviate is $\epsilon \sim N(0,1)$, σ is the standard error, ν_a is the age-specific proportion that this selected by the acoustic sampling gear, $N_{a,t}$ is the numbers-at-age, and W_a is the average weight-at-age during the survey. For simplicity the scaling parameter $q = 1$.

Age-composition data for both commercial and survey samples were randomly drawn from a multivariate distribution with a probability of $p_{a,t}$ of sampling an age-a fish in a given year t. The age-proportion samples must sum to 1 in each year, and random samples were based on the the following:

$$\begin{aligned} x_{a,t} &= \ln(\hat{p}_{a,t}) + \sigma_2 \epsilon_{a,t} - \frac{1}{A}\left[\sum_a \ln(\hat{p}_{a,t}) + \sigma_2 \epsilon_{a,t}\right], \\ p_{a,t} &= \frac{e^{x_{a,t}}}{\sum_a e^{x_{a,t}}} \end{aligned} \tag{2}$$

where $\epsilon_{a,t}$ is a standard random normal deviate, σ_2 is the standard error, $\hat{p}$ is the expectation of the proportion-at-age in year t in the sampled catch.

True parameter values used in the simulation model are listed in Table 1. Annual fishing mortality rates were conditioned on the observed catch from the Pacific hake fishery and it was assumed that both natural mortality and fishing mortality occur simultaneously. Simulated age-specific fishing mortality rates were based on the annual age-specific selectivity which differs among three alternative simulation scenarios (see description in the Scenarios subsection).

Table 1: Parameters used for simulation model in the integrated statistical catch-age model.

Description	**Symbol**	**Value**
Unfished age-1 recruits	R_o	3.353
Steepness (Beverton-Holt)	h	0.727
Natural mortality rate	M	0.230
Average age-1 recruitment	$\bar{R}$	1.300
Initial recruitment	$\dot{R}$	0.428
Survey standard deviation	σ_1	0.300
Standard deviation in recruitment	σ_R	1.120
Age at 50% selectivity in survey	$\hat{a}$	2.500
Std in 50% selectivity in survey	$\hat{g}$	0.500
Std in age-sampling error	σ_2	0.300

Parameter estimation

Model parameters were estimated using maximum likelihood methods where the objective function includes additional penalties to constrain the shape of the selectivity curve and how much its allowed to vary over time (Table 2). There are 6 major components to the objective function that is being minimized: (1) the likelihood of the observed catch (T2.5), (2) the likelihood of the relative abundance index (T2.6), (3) the likelihood of the age-composition information (T2.7), (4) the likelihood of the stock-recruitment data to estimate steepness and unfished age-1 recruits (T2.8), (5) prior densities in negative log space for estimated model parameters(T2.9), and (6) penalties and constraints for selectivity coefficients (T2.10).

The observed catch data are assumed to have a lognormal error structure and the standard deviation in the residuals between observed and predicted log catch is fixed at 0.0707 for all years. The likelihood for the relative abundance data is assumed to have lognormal errors and the variance of the residuals is an estimated parameter. Note that the conditional maximum likelihood estimate for q is used in the likelihood calculation (Walters and Ludwig, 1994), and we use a weak informative prior of $\ln(q) \sim N(0, 0.75)$ for the derived value of

Table 2: Calculations for the various components of the objective function ($f(\Theta)$) that is being minimized in the integrated statistical catch age model.

Residuals

$$w_t = \ln(\hat{C}_t) - \ln(C_t) \tag{T2.1}$$

$$z_t = \ln(I_t) - \ln(B_t) - \frac{1}{I}\sum_{t \in I}\left[\ln(I_t) - \ln(B_t)\right] \tag{T2.2}$$

$$\eta_{t,a} = \ln(\hat{p}_{a,t}) - \ln(p_{a,t}) - \frac{1}{A}\sum_{a=1}^{A}[\ln(\hat{p}_{a,t}) - \ln(p_{a,t})] \tag{T2.3}$$

$$\delta_t = \ln(N_{1,t}) - \ln(f(R_o, h, B_{t-1})) \quad \text{for } t > 1 \tag{T2.4}$$

Negative loglikelihoods

$$\ell(C) = T[\ln(\sigma_C) + 0.5\ln(2\pi)] + \sum_{t=1}^{T}\frac{w_t^2}{2\sigma_C^2} \tag{T2.5}$$

$$\ell(I) = I[\ln(\sigma_1) + 0.5\ln(2\pi)] + \sum_{t \in I}\frac{z_t^2}{2\sigma_C^2} \tag{T2.6}$$

$$\ell(P) = (A-1)T\ln\left(\frac{1}{(A-1)T}\sum_{a \in p_{a,t}}\sum_{t \in p_{a,t}}\eta_{t,a}^2\right) \tag{T2.7}$$

$$\ell(R) = (T-1)[\ln(\sigma_R) + 0.5\ln(2\pi)] + \sum_{t=2}^{T}\frac{\delta_t^2}{2\sigma_R^2} \tag{T2.8}$$

$$p(\Theta) = R_o \propto U(-5,15) + h \propto \beta(3,2) + \bar{R} \propto U(-5,15) + \dot{R} \propto U(-5,15) \tag{T2.9}$$

$$\begin{aligned} \mathrm{P} = \quad & \lambda_k^{(1)}\sum_{a=3}^{A-1}(v_{a,t} - 2v_{a-1,t} + v_{a-2,t})^2 \\ & + \lambda^{(2)}\sum_{A=1}^{A-1}\begin{cases}(v_{a,t} - v_{a+1,t})^2 & \text{if} \quad v_{a,t} > v_{a+1,t} \\ 0 & \text{if} \quad v_{a,t} \leq v_{a+1,t}\end{cases} \\ & + \lambda_k^{(3)}\sum_{t-3}^{T}(v_{a,t} - 2v_{a,t-1} + v_{a,t-2})^2 \end{aligned} \tag{T2.10}$$

Objective function

$$f(\Theta) = \ell(C) + \ell(I) + \ell(P) + \ell(R) + p(\Theta) + \mathrm{P} \tag{T2.11}$$

---

q in our simulation studies to stabilize the scaling parameters in Monte Carlo trials. Test with this weak informative prior and a uniform prior on the true Pacific hake data yielded identical MLE estimates.

The likelihood for age-composition information collected from commercial fisheries and the fisheries independent survey was assumed to come from a multivariate logistic distribution, and these data were weighted by the conditional maximum likelihood of the variance. Residual difference between observed $(\hat{p}_{a,t})$ and predicted $(p_{a,t})$ age-proportions were calculated using (T2.3) with the constraint that $\sum_a \eta_{a,t} = 0$. The advantage of this approach over a multinomial likelihood with a fixed effective sample size, is that the age-composition data are weighted appropriately conditional on the model structure (Schnute and Richards, 1995). An important point to note about the calculation of the age-composition residuals in (T2.3) is that the function is undefined if $\hat{p}_{a,t} = 0$. The addition of a small constant to both the observed and predicted proportions seems like a reasonable solution; however, in cases where year-classes are extremely weak and only partially selected by the fishing gear, the assumed value of the constant can influence the overall result. To avoid this problem, we alter the definition of an age-class in years where the observed proportion-at-age is 0 and pool this cohort into the adjacent age-class. In our simulation testing, this grouping of age classes was much more robust for parameter estimation and did not appear to produce any significant biases in comparison to methods that just add a small constant. Similar results were also obtained by Richards et al. (1997).

Annual age-1 recruitment was estimated via a mean recruitment value and a vector of deviates that were constrained to sum to 0. The integrated statistical catch age-model also jointly estimates the parameters of the resulting stock recruitment relationship given estimates of annual age-1 recruits and the resulting spawning stock biomass. Residual deviations between annual recruitment and recruitment based on a Beverton-Holt stock recruitment model were calculated using (T2.4), and the unfished age-1 recruits (R_o) and steepness parameters (h) were jointly estimated based on the negative log likelihood (T2.8). The variance parameter for recruitment deviations σ_R^2 was also estimated from the data.

Uniform priors were assumed for all the estimated parameters with the exception of an informative beta distribution for the steepness parameter in the interval 0.2–1.0, and informative priors for the variance parameters. The expected value for the steepness prior was set at 0.6 with a standard deviation of 0.161. For the variance parameters, we adopted a variance partitioning approach for estimating observation and process error variance. The estimated quantities consist of the total precision φ^2 and the proportion of the total variance that is associated with the observation errors ρ and the total variance is partitioned as:

$$\begin{aligned} \sigma_1^2 &= \rho/\varphi^2 \\ \sigma_R^2 &= (1-\rho)/\varphi^2 \end{aligned} \tag{3}$$

The advantage of this approach over directly estimating σ_1^2 and σ_R^2 directly is increased numerical stability. For the total precision an informative Gamma distribution was used as the prior $\varphi^2 \sim \Gamma(14.87652, 20.0)$ and a beta prior for the variance ratio $\rho \sim \beta(5.76465, 80.3464)$.

A likelihood penalty for the selectivity parameters is defined in (T2.10). Note that in (T2.10) there are three terms, the first of which is a penalty on the second differences between the age-specific coefficients to ensure a smooth ascending–descending pattern. The second term is a penalty on the amount of dome-shaped selectivity (often necessary when jointly estimating natural mortality rates). The third term is a second difference penalty on how age-specific coefficients vary over time. In each case the user must specify the relative weights λ that each of these penalties. For example, and infinitely large value of $\lambda^{(3)}$ would imply that selectivity coefficients are invariant over time. For the simulation-estimation experiment, values for $\lambda^{(1)}$ and $\lambda^{(2)}$ were set at 12.5, which corresponds to a coefficient of variation of roughly 0.20. The value of $\lambda^{(3)}$ was set at 1.0, which is equivalent to a CV of $\sqrt{0.5}$.

Reference points

Reference points based on long-term maximum sustainable yield (MSY-based reference points) were calculated assuming steady-state conditions. It was assumed that removals from the fishery independent survey were negligible. The fishing mortality rate that produced the maximum sustainable yield was determined by setting the derivative of the catch equation to 0 and solving for F_{MSY}. MSY was was subsequently determined by calculating the steady-state catch using F_{MSY}. Similarly B_{MSY} was determined by calculating the steady-state spawning biomass under a fishing mortality rate of F_{MSY}. Detailed descriptions of the steady state calculations for MSY-based reference points can be found in Martell et al. (2008).

All MSY-based reference points were based on the estimated selectivity value in the terminal year of the assessment. In cases where selectivity is assumed to remain constant over time, the estimated MSY-based reference points vary with minor updates to population parameters as the time series increases in length. However in cases where selectivity is assumed to vary over time, MSY-based reference points become highly uncertain as the uncertainty in selectivity in the terminal year is a function of how much selectivity is allowed to vary.

Scenarios

Three alternative datasets were generated with the simulation model using: (1) fixed length-based selectivity based on an asymptotic logistic function and empirical length-at-age data from commercial samples, (2) four discrete time blocks where the same asymptotic length-based function changes in 1986, 1999, and 2001, and (3) continuous changes in selectivity each year where the fishery targets cohorts based on Ideal Free Distribution (IFD). For the IFD selectivity model, the age-specific selectivity coefficients were based on age-specific biomass that is vulnerable to the fishing gear. Given a vector of selectivity coefficients v_a (based on the same length-based logistic function used in the other scenarios), the age-specific selectivity coefficients each year were based on the relative biomass-at-age b_a in a given year, and rescaled such that the mean of the vector is equal to 0 in log-space:

$$\omega_a = \ln(v_a) + 0.25\ln(b_a) - \frac{1}{A}\sum_{a=1}^{A}\left[\ln(v_a) + 0.25\ln(b_a)\right] \tag{4}$$

The coefficient of 0.25 is an arbitrary scaling of the biomass-at-age that would relate price premiums to larger size fish. The larger the price premium the less dome-shaped the selectivity curve would be because there is a financial incentive to target larger more valuable fish that are less abundant. In any case, the unique feature of (4) is that it allows for modal and multi-modal selectivity curves based on the relative abundance of each cohort (Figure 1).

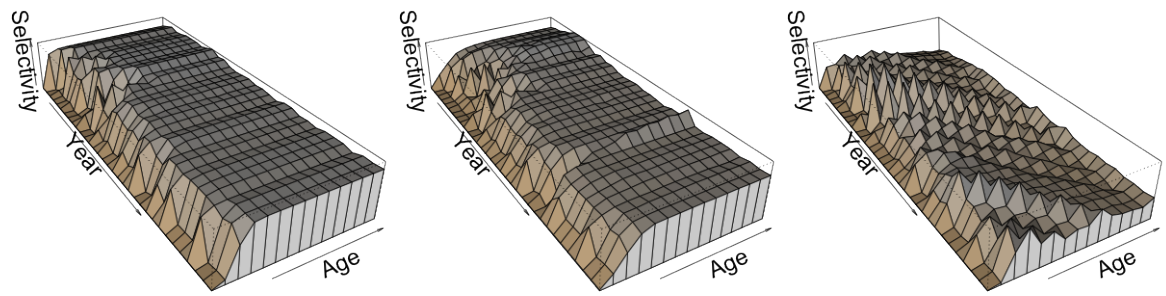

Figure 1: True selectivity curves used to generate simulated data sets for scenario 1 (left), scenario 2 (middle), scenario 3 (right).

Four alternative selectivity states were assumed in the assessment model (denoted by the letters a-d). For the fixed scenario, model (a), the probability of catching an individual of a given length was constant over time, and an estimated 7 equally spaced selectivity coefficients (or knots) starting at the length at age-1 and ending at age-15 length and interpolated between these knots using a cubic spline function. An additional penalty was added to the objective function to ensure smooth changes between ages and limit the amount of dome-shaped in the selectivity coefficients; values for $\lambda^{(1)}$ and $\lambda^{(2)}$ were both set at 12.5 which is roughly equivalent to a 20% coefficient of variation. The resulting age-based selectivities change with changes in the empirical length-at-age data over time.

The same selectivity penalty weights were also applied to model (b), which estimates the same 7 spline knots for four time blocks, one of which is only 2 years in length (1999 & 2000). In this case there are a total of 28 selectivity coefficients being estimated. Here we assume the timing of the discrete changes in selectivity corresponds to some known event and the years in which selectivity changes was correctly specified.

For model (c) the same penalized 7 spline knots are estimated for each year based on age (not length), resulting in a total of 231 estimated age-based selectivity parameters representing 495 age-year combinations. An additional penalty weight of $\lambda^{(3)} = 1.0$ (CV=0.707) was added to the time varying selectivity option to constrain how rapidly age-specific selectivity coefficients can change over time. Near identical results were obtained with $\lambda^{(3)} = 0.0$, but estimation convergence in Monte Carlo trials was less likely with $\lambda^{(3)} = 0.0$ due to potential confounding in the terminal year selectivity and terminal year fishing mortality rate.

An alternative to the annual time-varying selectivity is to interpolate in 2-dimensions where a series of knots for both age and year are the estimated parameters, and the age-year selectivity coefficients are interpolated using a 2-dimensional bicubic spline. For comparison with model (c), model (d) is based on estimating 7-age based knots and 12-time based knots for a total of 84 estimated selectivity coefficients. In this case the same penalty weights for smoothing, dome-shaped and time-varying changes in age-based selectivity were the same as that used for model (c).

The model permutations and combinations are summarized in Table 3, along with the total number of estimated and implied parameters for each of the four assessment models. The additional 67 implied parameters correspond to the use of conditional maximum likelihood estimates for the survey scaling parameter q and 66 additional variance terms for the measurement error in the survey and commercial age-composition data. Unless otherwise noted, figures and additional tables we maintain the same layout as Table 3 where the true states of selectivity are in the rows, and assumed states in the assessment models are represented in the columns.

Table 3: List of model scenarios and labels associated with each scenario explored. For example, scenario 2a is based on simulated data with a fixed selectivity curve, but assumes 3 discrete time blocks in the assessment model.

	Assumed selectivity states			
True states	**Fixed (a)**	**Discrete (b)**	**Continuous (c)**	**Bicubic spline (d)**
Fixed (1)	1a	1b	1c	1d
Discrete (2)	2a	2b	2c	2d
Continuous (3)	3a	3b	3c	3d
No. of parameters	95	116	319	172
Implied parameters	67	67	67	67
Total	162	183	386	239

We examine three criterion for choosing the appropriate selectivity parameterization, taking into consideration that the appropriate structural assumption may not be known. The first criterion compares how well each model configuration explains simulated data (statistical fit), examines the root mean square errors of the residuals, and uses Akaike Information

Criterion as a basis for contrasting overall performance. Retrospective performance is the second criterion we examine, as retrospective biases in stock assessments can potentially lead to severe over-exploitation, or under utilization of the resources depending on the direction of bias. Finally, we examine the performance of estimatining reference points using Monte Carlo trials.

Results

Statisical fit

Statisitics summarizing how well each model fits a single realization of simulated data is based on the overall objective function value , Deviance Information Criterion (AIC) and the Root Mean Square Error (RMSE) for survey index residuals, recruitment deviations and age-composition data from both commercial and survey samples (Table 4). Under conditions in which the true selectivity does not change over time, similar fits to the relative abundance index (see Survey RMSE in Table 4) were obtained regardless if commercial selectivity was assumed to be constant or varies over time. Allowing for additional structure in the selectivity coefficients over time resulted in decreases in the RMSE from 0.28 to 0.20 for the commercial age-composition information (Table 4).

We do not recommend using model selection criterion, such as DIC, as the only tool to choose the most appropriate model, but for statistica comparison we provide DIC and ΔDIC values to give a sense of the relative differences between the various assessment models for each simulation case. In the cases examined here (Table 4, DIC always favors the most structurally complex model (model c or d) with the largest number of estimated selectivity parameters. Model c also always results in over-fitting the age-composition information; the RMSE values for the commercial age-composition data were significantly less than the true value of 0.3 used in generating the age-composition data regardless of what the true underlying selectivtiy model was.

The effective number of estimated parameters is based on the difference between the expectation of the deviance and the deviance based on the expectation of the parameter values. The larger the effective number of parameters is a measure of how easy the model fits the data. In all cases the effective number of parameters was equal to or greater than the number of parameters estimated in the model. In the case of model (c) the effective number of parameters ($\approx$345) much greater than the 317 that were actually estimated in the ADMB code. However, it should aslo be noted that many of variance parameters and scaling parameters are based on the conditional maximum likelihood estimates, rather than explicity estimating them inside the model code. For example, the scaling parameter q in relative abundance index is based on the conditional maximum likelihood estimate (see Walters and Ludwig, 1994). The age-composition residuals are weighted by the conditional maximum likelihood estimates of the variance (T2.7). This parameterization implies an additional 67 parameters and hence the effective number of parameters is much less.

For the cases where the true selectivity is based on a fixed logistic function, the most appropriate model based on DIC is model (d) where a bicubic spline is used to model selectivity. The RMSE terms for model (d) are less than the true underlying values that were used to generate the data, which is of no surprice when additional flexibility in selectivity can accomodate some of the residual variance in age-composition in the form of minor changes in selectivity (Table 4). This pattern of explaining the residual variation is virtually the same regardless of what the true underlying selectivity pattern is.

In the case where the true selectivity changes discretely over three time blocks, assessment models with time-varying selectivity (c) and (d) appear to fit the data better that models that assumed fixed selectivity (a), or even the discrete changes in selectivity (see ΔDIC values in Table 4). For this particular case, it's better to allow for continous changes in selectivity than to assume fixed values.

In the case where the true selectivity changes annually, based on relative cohort abundance, the model results were a bit more surprising. The expectation would be that assuming fixed selectivity would perform less well than allowing for time-varying selectivity. Based on the ΔDIC values obtained in Table 4 there is very little differnce between models that allow for continous changes in selectivity (models c and d) and fixed selectivity (model a). There was less weight for the model that allows for discrete-block changes in selectivity (model b).

Retrospective performance

Two useful graphical tools for examining retrspective problems in stock assessment model are referred to here as spagetti plots (Figure 2), and squid plots (Figure 3), respectively. In the spagetti plots, successive estimates of spawning stock biomass based on sequentially removing the terminal year of data are overlayed on each panel. In addition to the estimated spawning biomass in Fig. 2, the true spawning biomass that was used to generate simulated data is also shown for reference. In the squid plots, successive biases of estimated spawning biomass relative to the spawning biomass in the terminal year of data are overlaid. In Fig. 4 the percent bias is shown as the relative difference between the true values; hence a 100% bias implies that the stock size is over-estimated by a factor of 2. In addition Fig. 4 shows the retrospective bias relative to the true spawning biomass that was used to generate the data.

For the simulated data based on fixed selectivity, the qualitative pattern in retrospective bias was similar for all four alternative selectivity scenarios (Figure 2, 1a, ..., 1d). The maximum retrospective bias observed was roughly 60% over the true spawning biomass (Figure 4). The worst performing models were the cases with continuous changes in selectivity over time (1c and 1d), with maximum estimates of retrospective bias relative to the terminal values approaching 25% for the bicubic spline model (Figure 3).

For cases based on simulated data with discrete time blocks in selectivity, the least amount of bias was observed in the case where the correct model was specified (Figure 3, 2b). Assuming fixed selectivity or continuous changes in selectivity resulted in significantly more retrospecitve bias. Assuming fixed selectivity also resulted in further departures from the true spawning biomass in the initial years.

In the cases where the true selectivity varies over time, the retrospective performance was least biased for the models that assume time-varying selectivity (Figure 3, 3c and 3d). Retrospective peformance is much worse if the selectivity is assumed to be constant, or change in a series of blocks, when the real underlying process is continuous change in selectivity (Figure 4, 3a and 3b).

Estimated reference points

A series of Monte Carlo trials (with 40 unique simulated data sets) were fitted with the four alternative selectivity assumptions to examine the distribution and estimation performance of reference points and how well the model fit various data components.

The impacts of model sepcification on the estimates of MSY-based reference points are summarized using a series of box-plots (Figure 5) based on Monte Carlo trials. In each box plot the $\ln_2$ ratios of the estimated versus true values are shown where the median bias is based on the solid bar.

In the case where the true model is based on fixed selectivity, estimated reference points are relatively unbiased, however the precision of the estimates decreases with increasing model complexity (Figure 5). In this case assuming a more complex time-varying selectivity patern may not result in severely biased estimates of reference points, however, uncertainty in estimated reference points increases.

In cases where the true model is based on time-blocks with discrete changes in selectivity, estimated F_{MSY} reference points are more likely to be precisely biased when the assumed selectivity is fixed is based on time-blocks (Figure 5, models a and b). Similar patterns were also observed for estimates of MSY. Spawning biomass reference points (B_{MSY} and B_o) were less sensitive to the assumed model structure.

For the case where the true model involves continuous changes in selectivity, estimates of F_{MSY} are biased upwards when selectivity is assumed constant over time. All other reference piont are much less biased if selectivity is allowed to vary over time continuously, or time-blocks.

The distribution of the RMSE values from the Monte Carlo trials displayed pretty consistent patters over each alternative dataset (Figure 6). The residual fit to the survey abundance index was very similar across all simulated datasets and all alternative assessment models. The true coefficient of variation used in simulating the true data was fixed at 0.3, and accounting for the bias correction in the lognormal errors, the theoretical RMSE in the residuals should be approximately 0.255. The standard deviation in the simulated recruitment residuals was fixed at 1.12 and the distribution of RMSE values is very similar for the fixed selectivity and time-block selectivity simulated datasets (Figure 6). In the case where the simulated data was based on continouse changes in selectivity, the RMSE values for the recruitment deviations was slightly higher than the true value, presumably to allow for more variation in recruitment that could be explained by annual changes in selectivity.

The pattern of RMSE values for the commercial age-composition residuals was consistent across simulated datasets, where largest RMSE values were always observed with fixed selectivity (Figure 6, model a), and the smallest under the annual time-varying selectivity (Figure 6, model c). Models with intermediate complexity were able to explain more residual variation in the comercial age-composition(models b and d, respectively).

Simulation performance

To qualtitatively evaluate how each of the four alternative selectivity models would perform if the analyst was nïave about the underlying true selectivity, we use a simple rank order system (Table 5). For example, if DIC was the statistica basis for choosing the most appropriate selectivity model then based on the simulation results in Tab. 4 when the true underlying model is based on fixed selectivity, the rank order of DIC values (low to high) is model d, a, b, and c. If the underlying model is not known, the DIC criterion favors model d (the bicubic spline) the most, and model a (fixed selectivity) the second most. Based on the total Root Mean Squared Error (RMSE) of the residuals

References

Beverton, R. J. and Holt, S. J. (1993). *On the dynamics of exploited fish populations*, volume 11. Springer.

Fournier, D., Skaug, H., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M., Nielsen, A., and Sibert, J. (2011). Ad model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models.

Martell, S. (2009). Assessment and management advice for pacific hake in u.s. and canadian waters in 2009. *DFO Can. Sci. Advis. Sec. Res. Doc.*, 2009/021:iv+54p.

Martell, S. J. D., Pine, W. E., and Walters, C. J. (2008). Parameterizing age-structured models from a fisheries management perspective. *Can. J. Fish. Aquat. Sci.*, 65:1586–1600.

Richards, L., Schnute, J., and Olsen, N. (1997). Visualizing catch-age analysis: a case study. *Canadian Journal of Fisheries and Aquatic Sciences*, 54(7):1646–1658.

Schnute, J. and Richards, L. (1995). The influence of error on population estimates from catch-age models. *Canadian Journal of Fisheries and Aquatic Sciences*, 52(10):2063–2077.

Stewart, I. J., Leaman, B., Martell, S., and Webster, R. A. (2012). Assessment of the pacific halibut stock at the end of 2012. In *International Pacific Halibut Commission Eighty-ninth Annual Meeting*.

Walters, C. and Ludwig, D. (1994). Calculation of Bayes posterior probability distributions for key population parameters. *Canadian Journal of Fisheries and Aquatic Sciences*, 51(3):713–722.

Walters, C. and Maguire, J. (1996). Lessons for stock assessment from the northern cod collapse. *Reviews in Fish Biology and Fisheries*, 6(2):125–137.

Table 4: Statistical performance based on the objective function value, effective number of estimated parameters, DIC, ΔDIC, Root Mean Square Error (RMSE) in recruitment deviations, survey abundance residuals, and the age-composition residuals for models fit to fixed, discrete time blocks and continous changes in commercial selectivity.

	Fixed (a)	**Discrete (b)**	**Continuous (c)**	**Bicubic spline (d)**
True selectivity is fixed				
Objective function	-1452.78	-1461.06	-1695.12	-1538.09
Eff. No. parameters	96	110	344	175
DIC	-2712.84	-2701.20	-2696.89	-2723.12
ΔDIC	10.28	21.92	26.23	0.00
Recruitment RMSE	1.03	1.03	1.03	1.03
Survey RMSE	0.25	0.26	0.25	0.25
Commercial RMSE	0.28	0.28	0.20	0.26
Survey age RMSE	0.26	0.26	0.27	0.26
True selectivity has 3 time blocks				
Objective function	-1340.49	-1428.81	-1689.57	-1516.78
Eff. No. parameters	96	109	345	175
DIC	-2487.96	-2638.03	-2683.48	-2680.78
ΔDIC	195.52	45.45	0.00	2.70
Recruitment RMSE	1.04	1.04	1.03	1.03
Survey RMSE	0.25	0.26	0.25	0.25
Commercial RMSE	0.32	0.29	0.20	0.26
Survey age RMSE	0.26	0.26	0.27	0.26
True selectivity changes annually				
Objective function	-1293.30	-1304.14	-1543.59	-1376.39
Eff. No. parameters	95	109	344	177
DIC	-2394.61	-2387.88	-2393.28	-2396.33
ΔDIC	1.72	8.45	3.05	0.00
Recruitment RMSE	1.13	1.13	1.12	1.12
Survey RMSE	0.26	0.27	0.27	0.27
Commercial RMSE	0.30	0.30	0.21	0.28
Survey age RMSE	0.33	0.34	0.33	0.32

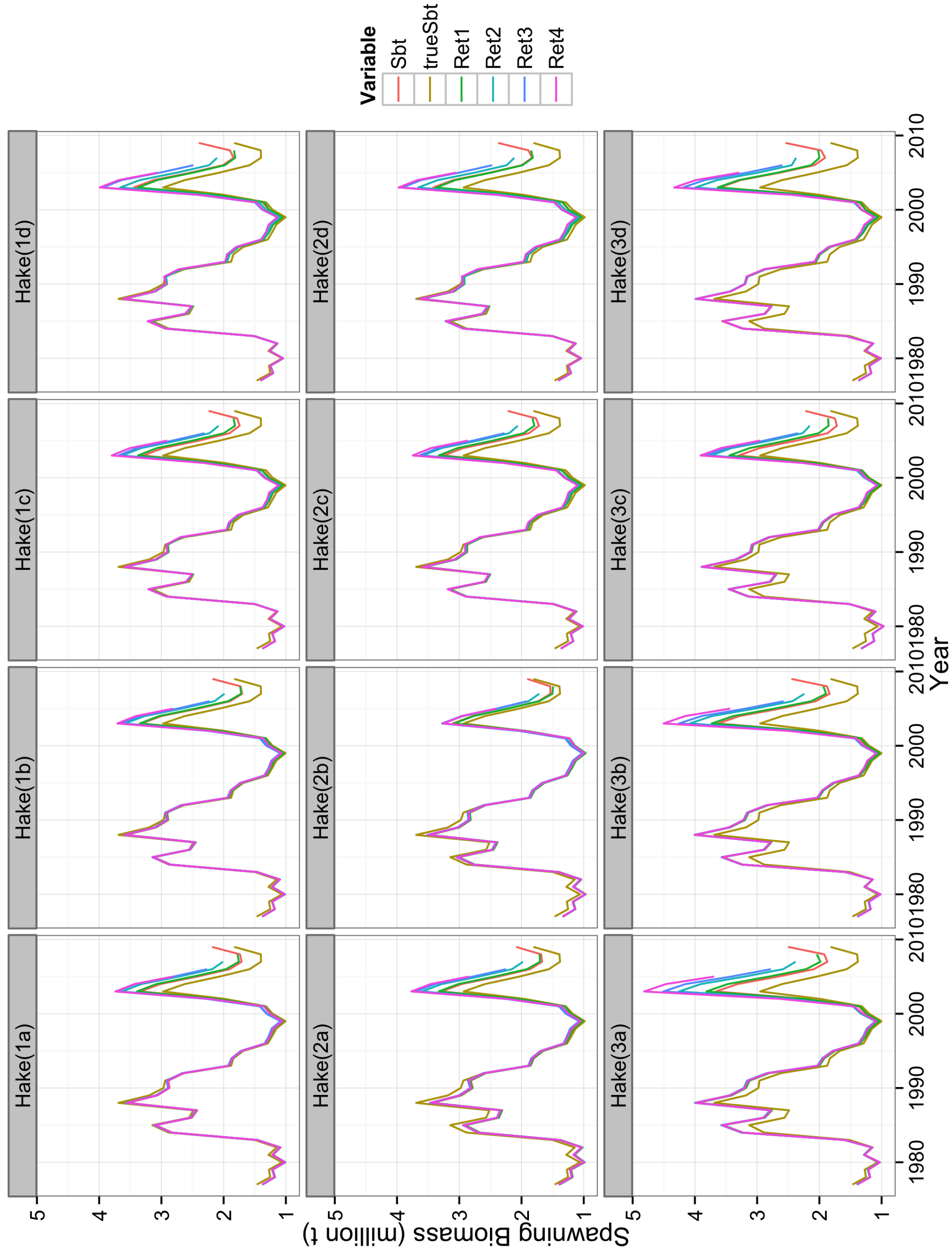

Figure 2: Retrospective estimates of spawning biomass for simulated Pacific hake populations where 4 years of data was sequentially removed from. The true spawning biomass used to simulated the data is included for reference.

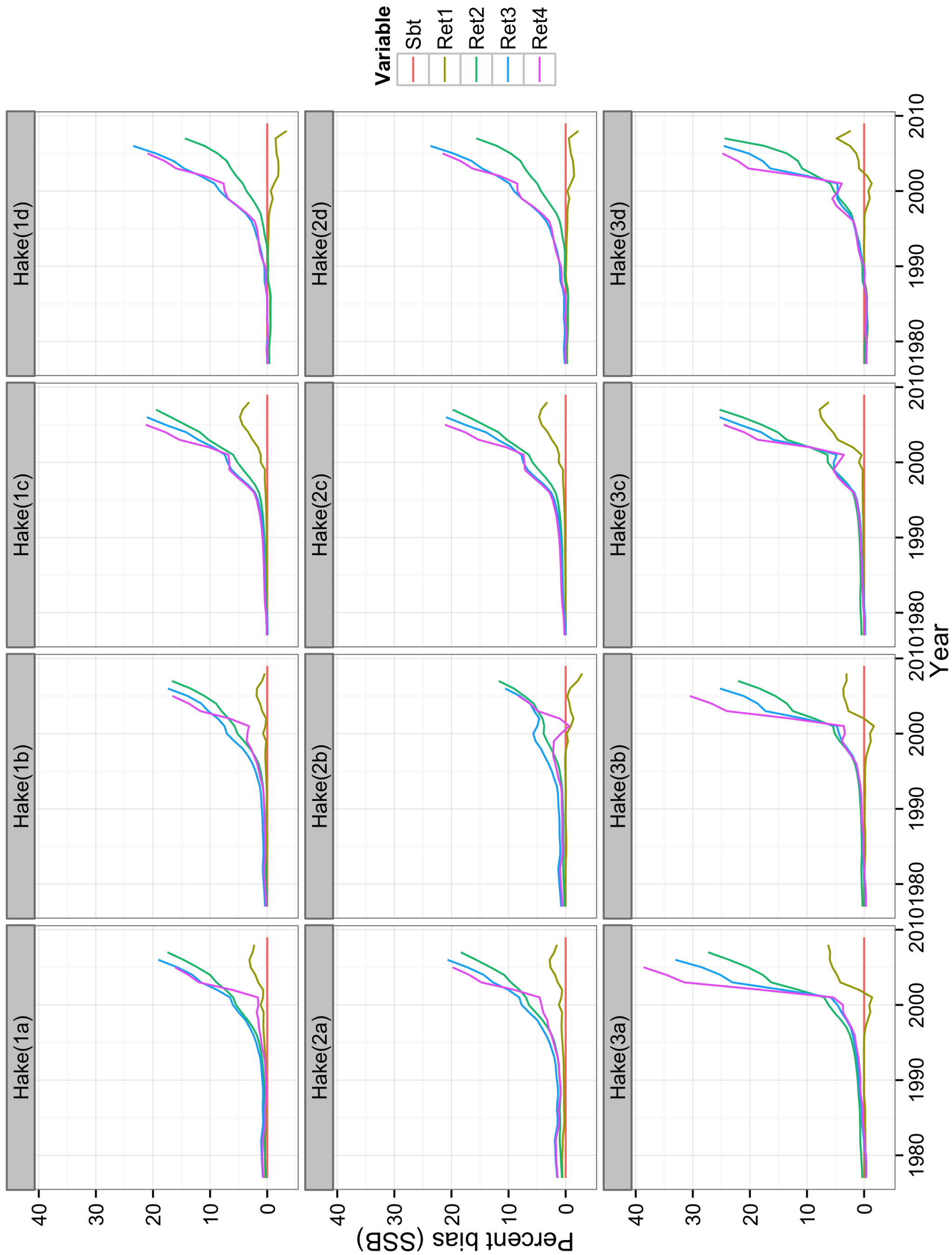

Figure 3: Retrospective estimates of bias in spawning biomass relative to spawning biomass estimated with all available data. These biases are based on the same spawning biomass trajectories in Figure 2.

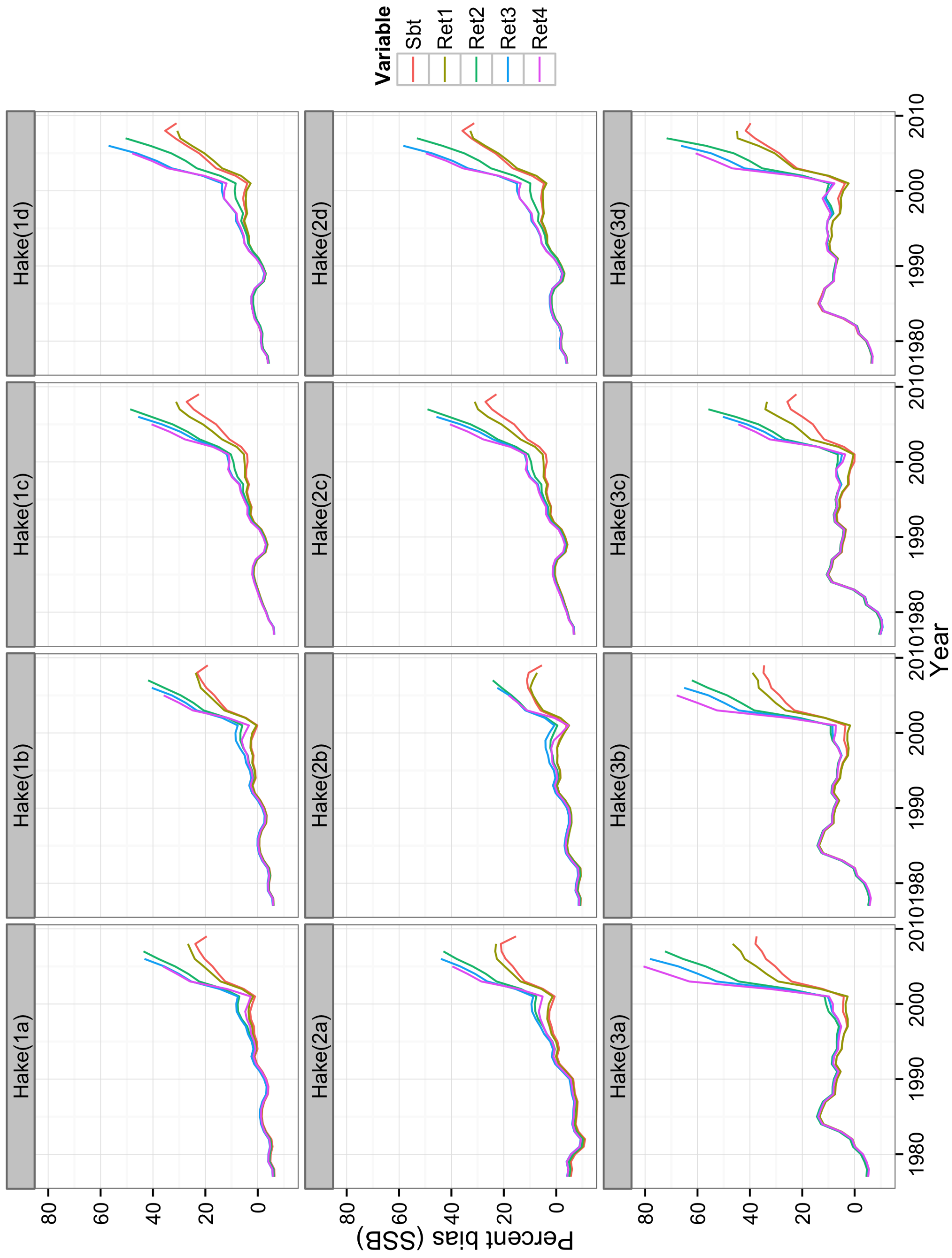

Figure 4: Retrospective estimates of bias in spawning biomass relative to the true spawning biomass used to simulated the data. These biases are based on the same spawning biomass trajectories in Figure 2.

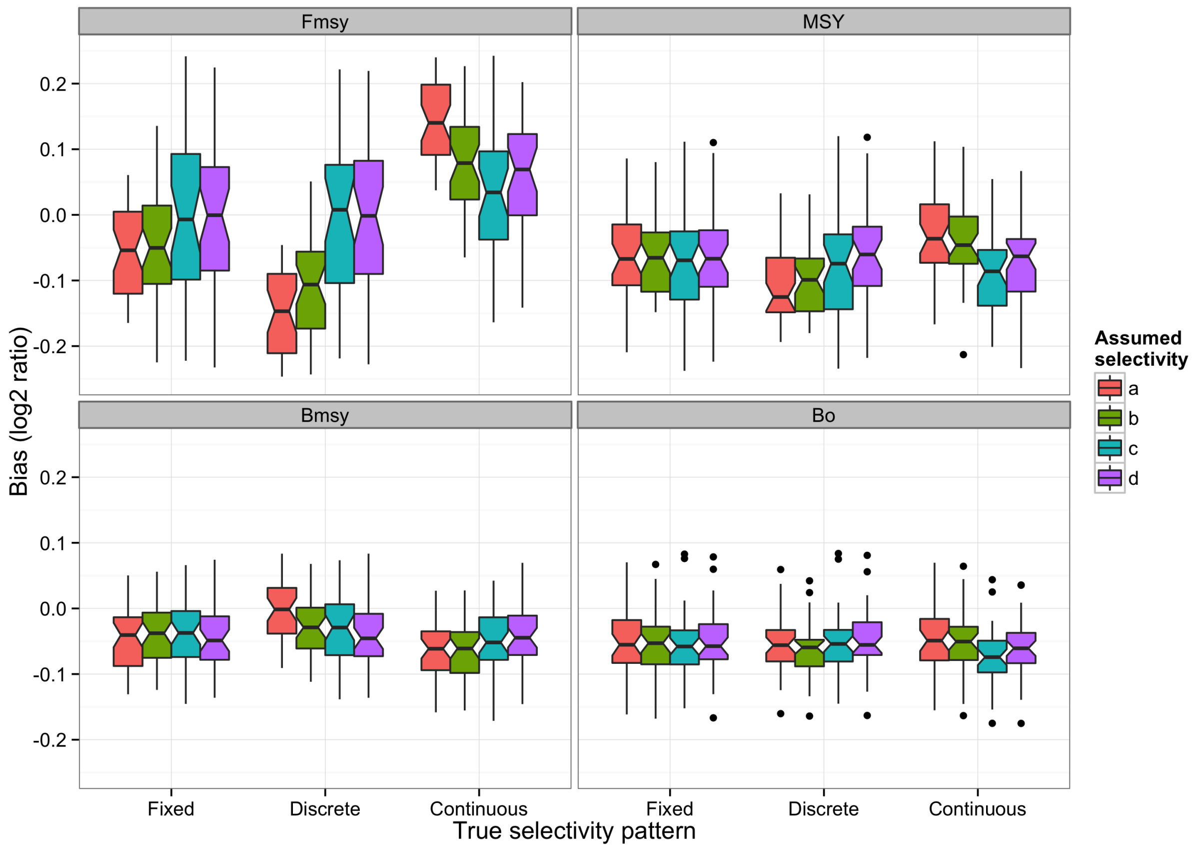

Figure 5: Estiamtes of precision and bias for fishing mortality rate reference points (F_{MSY}), maximum sustainable yield (MSY), spawning biomass as MSY (B_{MSY}) and the unfished spawning biomass (B_o) based on Monte Carlo trials using data simulated from fixed, discrete blocks and continuous changes in selectivity.

Table 5: Ranking of model based on Deviance Information Criterion, RMSE, retrospective bias and bias in the estimates of F_{MSY} based on Monte Carlo trials. Each column ranks the assumed selectivity model from most likely (left) to least likely (right) for simualtion case study. The top-two ranks represent the most and second most frequently selected model.

Criterion	**Fixed**	**Discrete**	**Continuous**	**Top-two ranks**
DIC from Tab. 4	d,a,b,c	c,d,b,a	d,a,c,b	d,a
RMSE	c,d,a,b	c,d,b,a	c,d,a,b	c,d
Retrospective	d,b,a,c	d,a,c,b	a,b,d,c	d,b
F_{MSY}	c,d,a,b	c,d,a,b	a,b,d,c	c,d

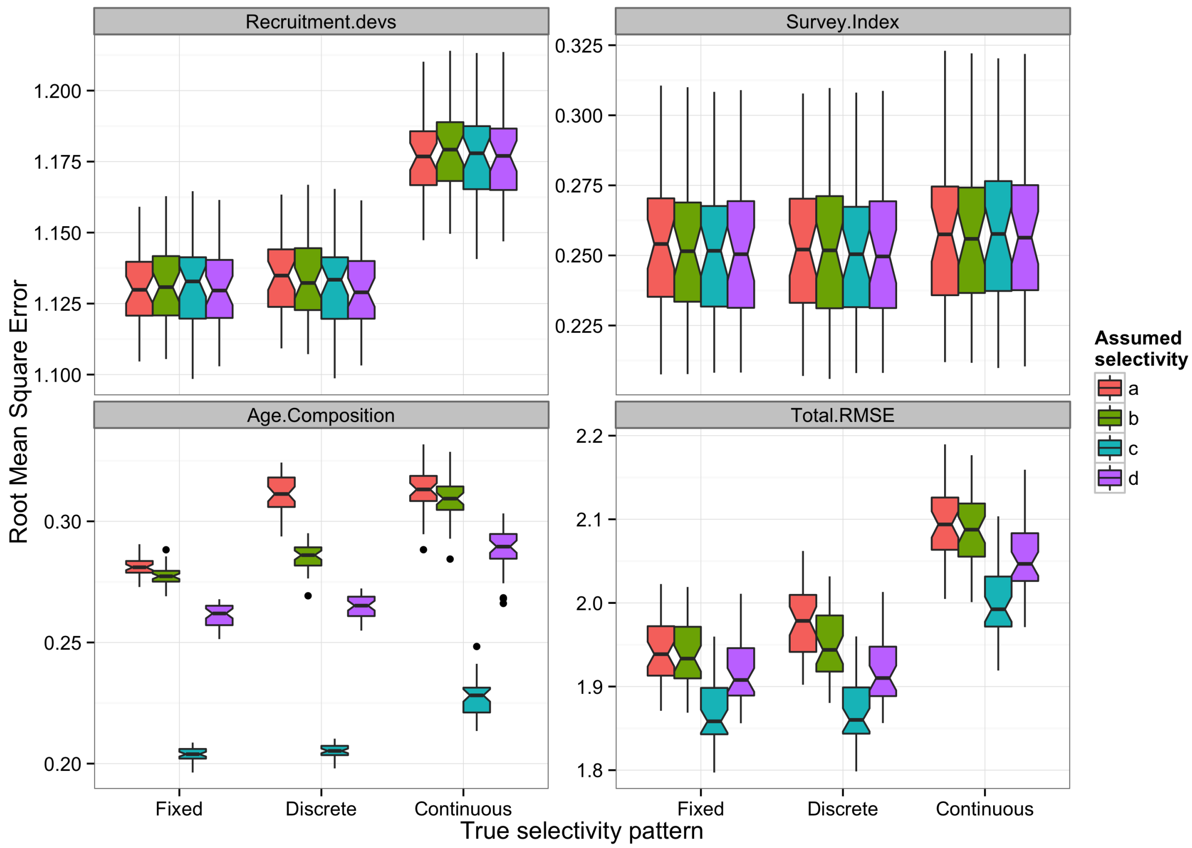

Figure 6: Distribution of Root Mean Squared Error values for the recruitment deviations, survey residuals, commercial age-composition residuals and the total RMSE from all data based on 40 Monte Carlo trials.