Fisheries Research 94 (2008) 290–303

Contents lists available at ScienceDirect

Fisheries Research

journal homepage: www.elsevier.com/locate/fishres

Evaluating alternative estimators of fishery management reference points

Melissa A. Haltuch[a,b,*], André E. Punt[b], Martin W. Dorn[c]

[a] Northwest Fisheries Science Center, National Marine Fisheries Service, 2725 Montlake Boulevard East, Seattle, WA 98112, USA
[b] School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, WA 98195-5020, USA
[c] Alaska Fisheries Science Center, National Marine Fisheries Service, 7600 Sand Point Way NE, Seattle, WA 98115, USA

ARTICLE INFO

Article history:
Received 17 October 2007
Received in revised form 24 January 2008
Accepted 28 January 2008

Keywords:
Simulation
Biomass reference points
Estimators
Groundfish

ABSTRACT

Fishery management plans for U.S. fisheries are required to specify status determination criteria (e.g. whether the stock is overfished and whether overfishing is occurring) and typically use harvest control rules to adjust target and limit fishing mortality and catch levels to prevent overfishing, achieve optimum yield and rebuild overfished stocks. The status determination criteria are based on the concept of the fishing mortality rate ($F_{\rm MSY}$) that maximizes long-term catch as the upper limit on the allowable rate of fishing and the associated $B_{\rm MSY}$, the spawning biomass which produces MSY, is the target for rebuilding of overfished stocks. In practice, proxies for the biological reference points $F_{\rm MSY}$ and $B_{\rm MSY}$ are often employed. Although several methods exist for estimating these quantities, it is unclear which performs best. Simulation is therefore used to evaluate alternative estimators for these quantities. These estimators differ in terms of whether a stock–recruitment relationship is estimated, and whether a prior based on Bayesian meta-analysis is used as a penalty on steepness, a critical parameter of the stock–recruitment relationship. The simulations consider three life histories: a long-lived unproductive rockfish, a moderately long-lived and productive flatfish, and a hake, which is also moderately long-lived and productive, but exhibits highly variable recruitment. Results indicate that estimator performance varies among reference points. However, estimators of B_0, the average spawning biomass in the absence of exploitation, and stock depletion based on a fitted stock–recruitment relationship generally perform best. B_0 is estimated either better (the rockfish and flatfish) or similarly (the hake) to stock depletion. Estimating $B_{\rm MSY}$ from the fit of the stock–recruitment relationship performed best for the rockfish and flatfish life histories; average recruitment estimators proved to be best for the flatfish and hake life histories. Proxy methods of calculating $B_{\rm MSY}$ generally performed relatively poor in comparison to the non-proxy measures. The performance of estimators of biological reference points was generally better for the rockfish and flatfish life histories, which were similar, than for the hake life history. Estimator performance was generally poorer in the presence of high recruitment variability.

1. Introduction

Control rules, pre-agreed strategies for making management decisions, used to determine catch limits, typically require estimates of at least one of the following: B_0, $B_{\rm MSY}$, or current spawning biomass relative to B_0 (stock depletion) (PFMC, 1998; Restrepo and Powers, 1999). Current spawning biomass is the spawning biomass at the end of the last year of the assessment period, $B_{\rm MSY}$ is the spawning biomass at which maximum sustainable yield (MSY) is achieved, and B_0 is defined as the equilibrium, average, spawning biomass in the absence of exploitation. Reliable estimation of biomass reference points is important because they directly impact fisheries management, typically via the setting of catch quotas. Biomass reference points should therefore be estimated in the most unbiased and precise manner possible to provide consistent scientific advice and maintain credibility with both managers and fishers. Meeting the goals of sustainable fisheries (such as maintaining fish stocks at or above $B_{\rm MSY}$) will be difficult if the estimators of these quantities are biased or imprecise. Understanding the properties of commonly used estimators for these three reference points should therefore result in better informed management decisions.

Two general approaches are used to estimate $B_{\rm MSY}$: (1) estimating it directly, and (2) setting it at some fraction of the estimate of B_0. The first approach is most appropriate when a stock has been fished historically such that obtaining data from a period during which the stock was lightly fished may be impossible. It is gener-

* Corresponding author at: Northwest Fisheries Science Center, National Marine Fisheries Service, 2725 Montlake Boulevard East, Seattle, WA 98112, USA. Tel.: +1 206 860 3480; fax: +1 206 860 6792.
E-mail address: Melissa.Haltuch@noaa.gov (M.A. Haltuch).

0165-7836/$ – see front matter
doi:10.1016/j.fishres.2008.01.008

ally believed that the ability to estimate B_0 in this case will be poor (NMFS, 2002). For example, fishery resources in the U.S. Northeast have been exploited for hundreds of years resulting, in some cases, in over-exploited stocks prior to the collection of more standardized landings and survey data (NMFS, 2002). For over-exploited stocks, estimates of B_{MSY} based directly on more recent landings and survey data may underestimate the true level of stock productivity, resulting in estimates of B_{MSY} that are too low, particularly if recruitment has declined due to lower spawning stock biomass (NMFS, 2002). However, recent average recruitment should be a good proxy for the recruitment at MSY for fully exploited stocks, stock sizes that are near or below B_{MSY}, which have sustained relatively stable levels of exploitation over extended periods of time because there are more observations of recruitment. The second approach to formulating biomass reference points is more appropriate for fisheries that have been established more recently and there are likely few, if any, observations of stock size near B_{MSY}. For such fisheries, estimating B_0 is more feasible so biomass reference points tend to be based on setting B_{MSY} as some fraction of B_0 (PFMC, 2006). The U.S. West Coast provides an example where many fish stocks were more recently exploited and high fishing pressure played a role in driving many stocks below B_{MSY} quickly enough that few recruitments near B_{MSY} were observed.

In the past, estimates of B_0 for U.S. West Coast groundfish species have been calculated by multiplying unfished, equilibrium, spawning biomass-per-recruit by the estimate of the average recruitment during a period of high stock biomass (PFMC, 2002). Recently, however, it has become more common to estimate B_0 using alternative methods such as the results from the fit of a stock assessment method with an integrated stock–recruitment relationship (PFMC, 2002; Methot and Stewart, 2005). Estimating B_0 in this way requires information about stock productivity, primarily captured within an assessment model by the steepness parameter of the stock–recruitment relationship, h, and recruitment deviations. In the integrated approach, the stock assessment method estimates one or both parameters of the stock–recruitment relationship in order to determine the central tendency of estimated recruitment as well as to provide the basis for calculating recruitment at B_0. The benefit of the integrated approach is that all of the information about the early average recruitments is used to estimate B_0 and other reference points, rather than just the recruitments during the first few years with data. An additional method of estimating B_0 is to calculate the average size of the stock, given the final parameter estimates from a stock assessment, during the model period in the absence of fishing, referred to as "Dynamic B_0" (MacCall et al., 1985). Dynamic B_0 provides an estimate of the size of the stock had fishing not occurred.

Conceptually, B_{MSY} is the product of the spawning biomass-per-recruit corresponding to F_{MSY} and R_{MSY} (the average recruitment corresponding to F_{MSY}) (Quinn and Deriso, 1999). F_{MSY} can be estimated by maximizing the product of yield-per-recruit and recruitment as a function of fishing mortality, which also depends on the estimate of h. Alternatively, B_{MSY} can be estimated by setting F_{MSY} to a proxy such as $F_{40\%}$ or the value of fishing mortality at which yield-per-recruit is maximized and setting R_{MSY} to the average recruitment over some subset of the years included in the stock assessment, similarly to fisheries in the New England and Alaska regions of the U.S. (NEFMC, 2003; NPFMC, 2006). The final alternative method for estimating B_{MSY} considered in this paper, which in the past has been standard practice for U.S. West Coast groundfish species, is to use a proxy for B_{MSY} based on B_0, typically $B_{\text{MSY}} \sim 0.4B_0$ (PFMC, 2006).

This paper uses Monte Carlo simulation to explore the performance of several alternative estimators for B_0, B_{MSY}, and current biomass relative to B_0. The impact of some of the key factors that might determine the performance of alternative estimators of these reference points is explored. Results are presented for three life histories based on groundfish species included in the Pacific Fishery Management Council's Groundfish Fishery Management Plan (PFMC, 2006), viz: petrale sole (*Eopsetta jordani*), canary rockfish (*Sebastes pinniger*), and Pacific hake (*Merluccius productus*). Examining three life history types, a flatfish, rockfish, and hake, allows the impact of biological parameters such as longevity, growth and maturity rates, recruitment variability, stock productivity, and age-based selectivity on estimator performance to be evaluated. Although the catch history and biological information for the life histories examined in this paper are similar to U.S. West Coast groundfish, these life histories are generalized enough for the conclusions of this study to be widely applicable to groundfish species elsewhere given a relatively complete catch history.

The generalized simulation framework applied in this paper allows the exploration of a variety of factors potentially impacting the ability of age-structured stock assessment methods to estimate biomass reference points. The results are intended to discern broad patterns and to identify best practices for calculating commonly used reference points in addition to identifying factors controlling the quality of the estimates. Due to the broad scope and generalized species life histories examined, the goal is to identify a small set of preferred estimators which might be further investigated rather than a single "best" estimator.

2. Methods

2.1. Simulation model

Simulation modeling to evaluate a statistical estimator consists of four basic steps (Punt, 2003a). First, an "operating model" is developed which defines the "true" state of the system for the purposes of the simulations; the operating model provides a population which can be sampled in a manner similar to the actual collection of field data. Next, the operating model is parameterized to address the questions of interest and a number of data sets are generated. Each alternative estimator for B_0, B_{MSY}, and current spawning biomass relative to B_0 is then applied to each data set. Finally, the outputs from these estimators are compared with the "true" values from the operating model.

The simulation model that forms the basis for the evaluation of alternative estimators is age-structured, assumes that catches are taken in the middle of the year after half of natural mortality, and relates the number of age-0 animals to spawning biomass by means of the Beverton–Holt stock–recruitment relationship. The basic population dynamics are given by the equation:

$$N_{y+1,a} = \begin{cases} N_{y+1,0} & \text{if } a = 0 \\ (N_{y,a-1}\,e^{-0.5M} - C_{y,a-1})e^{-0.5M} & \text{if } 1 \le a < x \\ ((N_{y,x-1} + N_{y,x})e^{-0.5M} - (C_{y,x-1} + C_{y,x}))e^{-0.5M} & \text{if } a = x \end{cases} \tag{1}$$

where $N_{y,a}$ is the number of fish of age a at the start of year y, M is the instantaneous rate of natural mortality, x is the maximum modeled age (rockfish = 50, flatfish = 15, and hake = 17), and $C_{y,a}$ is the catch of animals of age a in numbers during year y. The number of age-0 animals is given by

$$N_{y,0} = \frac{4hR_0\,\text{SB}_y}{(1-h)B_0 + (5h-1)\text{SB}_y}\,e^{\varepsilon_y - \sigma_R^2/2}, \quad \varepsilon_y \sim N(0, \sigma_R^2) \tag{2}$$

where SB_y is the spawning biomass at the start of year y, h is the "steepness" of the stock–recruitment relationship, R_0 is the average number of 0-year olds in the absence of fishing, ε_y is the normally distributed error about the stock–recruitment relationship, and σ_R^2 is the log-scale variance of the ε_y's. In this parameterization of the Beverton–Holt stock–recruit relationship, R_0 is a measure of the carrying capacity of the population, while steepness, defined as the expected proportional reduction in recruitment when the stock is reduced to 20% of unfished population size, measures the response of the population to fishing. The spawning biomass during year y is given by

$$\mathrm{SB}_y = \sum_{a=1}^{x} f_a w_a N_{y,a} \tag{3}$$

where f_a is the proportion mature at age a, and w_a is the mass of an animal of age a. The selectivity of both the survey and fishery on animals of age a, S_a, is assumed to be governed by a logistic curve:

$$S_a = \frac{1}{1+\exp[-(a-a_{50})/\delta]} \tag{4}$$

where a_{50} is the age-at-50% recruitment to the fishery (or to a survey which provides data for stock assessment purposes), and δ is the parameter which determines the width of the selectivity ogive. Survey and fishery selectivity are the same in the operating model, as might be expected if the survey was conducted using similar gear and in similar places to the fishery.

The mass of the catch during year y is given by

$$\tilde{C}_y = \sum_{a=0}^{x} w_a C_{y,a} \tag{5}$$

and the catch of animals of age a in numbers during year y is

$$C_{y,a} = F_y S_a N_{y,a}\,\mathrm{e}^{-0.5M} \tag{6}$$

where F_y is the fully selected exploitation rate during year y. F_y is calculated so that the simulation model-estimate of the catch, Eq. (5), matches the pre-specified catch exactly, i.e. catches are known without error.

The population is assumed to be at its pre-exploitation equilibrium biomass at the start of the first year (y = 1). The pre-exploitation age structure (in relative terms) is given by

$$\tilde{N}_a = \begin{cases} 1 & \text{if } a = 0 \\ \tilde{N}_{a-1}\,\mathrm{e}^{-M} & \text{if } 1 \le a < x \\ \tilde{N}_{x-1}\,\mathrm{e}^{-M}/(1-\mathrm{e}^{-M}) & \text{if } a = x \end{cases} \tag{7}$$

so that:

$$R_0 = \frac{B_0}{\sum_{a=1}^{x} f_a w_a \tilde{N}_a} \tag{8}$$

The impact of the assumption that the population was in deterministic equilibrium at the start of the first year considered in the operating model is removed by "burning in" the model for a period equal to twice the maximum age of the species under consideration prior to the first year for which there is catch. The population dynamics are governed by Eqs. (1)–(3) during the "burn-in", but F_y is assumed to be equal to zero for all years of the "burn-in" period. The initial age–structure at the start of the burn-in period is given by

$$N_{1,a} = R_0 \tilde{N}_a \tag{9}$$

The data typically available for stock assessments of the groundfish species off the U.S. West Coast (and many other groundfish fisheries worldwide) include catches, biological parameters, samples of the age-composition of the catches and information (indices and age-composition) from surveys. An observation model is used to generate samples from the population for use by the simulated stock assessment. The catches and biological parameters (natural mortality, mass-at-age, fecundity-at-age, and selectivity-at-age) are assumed to be known exactly. Survey catch rates are generated with log-normally distributed error, i.e.:

$$I_y = q\tilde{B}_y^{\mathrm{s}}\,\mathrm{e}^{\eta_y - \sigma_q^2/2}, \quad \eta_y \sim N(0, \sigma_q^2) \tag{10}$$

where I_y is the survey catch rate for year y, $\tilde{B}_y^{\mathrm{s}}$ is the biomass available to the survey in the middle of year y, q is the catchability coefficient, set equal to 1, η_y is the observation error in survey catchability, and σ_q is the standard deviation of the observation error in survey catchability. The value of $\tilde{B}_y^{\mathrm{s}}$ is given by:

$$\tilde{B}_y^{\mathrm{s}} = \sum_{a=0}^{x} w_a \tilde{S}_a N_{y,a}\,\mathrm{e}^{-0.5M} \tag{11}$$

where $\tilde{S}_a$ is the selectivity of the survey on fish of age a, assumed to be governed by a logistic curve (Eq. (4)).

The collection of survey and age-composition data generally coincides with the beginning of the catch time-series. Age-composition data for both the survey and catch for each year y are random multinomial samples of size $n_y^{\mathrm{c/s}}$ from the survey (s) or catch (c) for year y. Each age a is selected with the probability $P_{y,a}^{\mathrm{s/c}}$:

$$P_{y,a}^{\mathrm{c}} = \frac{C_{y,a}}{\sum_{a=1}^{x} C_{y,a}}; \qquad P_{y,a}^{\mathrm{s}} = \frac{\tilde{S}_a N_{y,a}\,\mathrm{e}^{-0.5M}}{\sum_{a=1}^{x} \tilde{S}_a N_{y,a}\,\mathrm{e}^{-0.5M}} \tag{12}$$

The true value of B_{MSY} is the spawning biomass at which the product of the relationship between recruitment as a function of fishing mortality and yield-per-recruit as a function of fishing mortality is maximized.

2.2. Estimator implementation

The estimators for B_0, B_{MSY}, and current spawning biomass relative to B_0 involve first applying a stock assessment method (the estimation model) to the data generated from the operating model and then using the results from the estimation model as the basis for estimating the three quantities.

The dynamics of the population in the estimation model are the same as those in the operating model, except that, for consistency with how assessments are typically conducted for U.S. West Coast groundfish, the population is assumed to be in deterministic equilibrium at the start of first year for which catches are available. The quantities estimated by the stock assessment method include B_0 (henceforth referred to as $\hat{B}_0$), the survey catchability coefficient (q), the time-trajectories of spawning biomass and recruitment, and the logistic selectivity functions for the survey and fishery. Catches, the variance of the error about the survey catch rates (σ_q), and the age-composition sample sizes ($\underline{n}^{\mathrm{c}}$ and $\underline{n}^{\mathrm{s}}$) are treated as known without error. The time-trajectory of fishing mortality is calculated from the catches. The objective function minimized to find the point estimates of the parameters of the estimation model can have up to four terms related to: (1) the survey index of relative abundance, (2) the fishery catch age-composition data, (3) the survey catch age-composition data, and (4) a penalty related to the stock and recruitment data. The contribution of the data (terms (1)–(3)) to the objective function is given by

$$-\ln L = \frac{1}{2\sigma_q^2}\sum_y \ln\left(\frac{I_y}{\hat{q}\hat{B}_y^{\mathrm{s}}}\right)^2 + \sum_y n_y^{\mathrm{s}} \sum_a P_{y,a}^{\mathrm{s}} \ln(\hat{P}_{y,a}^{\mathrm{s}}) + \sum_y n_y^{\mathrm{c}} \sum_a P_{y,a}^{\mathrm{c}} \ln(\hat{P}_{y,a}^{\mathrm{c}}) \tag{13}$$

where $\hat{B}_y^s$ is the model-estimate of survey biomass for year y, $\hat{q}$ is the estimate of the survey catchability coefficient, σ_q^2 is the variance associated with the observation error for the survey indices, and $\hat{P}_{y,a}^{s/c}$ is the model-estimate corresponding to $P_{y,a}^{s/c}$.

The contribution of recruitment deviations to the objective function is given by

$$-\ln P = m \ln \sigma_R + \frac{1}{2\sigma_R^2} \sum_{y=1}^{m} \varepsilon_y^2 \tag{14}$$

where σ_R is the standard deviation of recruitment (assumed to be known), and m is the number of years for which recruitment deviations are estimated. All of the stock assessment methods considered in the paper estimate the log deviations about the fitted stock–recruitment relationship and hence include the stock–recruitment penalty (Eq. (14)).

Three alternative methods of stock assessment are considered. All of these involve fitting a stock–recruitment relationship and hence estimating h:

(A) Fitting the model assuming that the stock was at B_0 at the start of the first year of the modeled period (abbreviation "$B0$").

(B) As for (A), except that an informative prior is placed on the logit of steepness. The mean of this prior is set equal to the true value of h and the variance is based on a Bayesian meta-analysis of West Coast rockfish and flatfish steepness values (Fig. 1). Since there is not a West Coast meta-analysis of gadid steepness values the variance from the flatfish prior was used for the hake life history. The contribution of this prior to the objective function is:

$$\ln \sigma + \frac{(\beta - \mu)^2}{2\sigma^2} \tag{15}$$

where β is the logit of steepness, and μ and σ are, respectively, the mean and standard deviation of the prior for the logit of steepness (abbreviation "$B0 + h$"). Stock assessment method (B) is included to investigate how estimator performance is improved with highly optimistic information about h (abbreviation "$B0 + h$").

(C) As for (A), except that the numbers of age-0 animals for one generation prior to the first year for which catch data are available are treated as estimable parameters, with an expectation based on an alternative value for B_0, B_1. Recruitment deviations during the time period in which catches occur relate to the stock–recruitment relationship parameterized by B_0 and h. A single vector of recruitment deviations is estimated for both the pre-catch and post-catch portion of the model. This effectively allows for a shift in mean recruitment at the beginning of the catch time-series from B_1 to B_0 conditions, eliminating the assumption that the stock is at unfished equilibrium at the start of the catch time-series. Stock assessment method (C) is based on the idea that the stock may not have been at its unfished equilibrium biomass, the long-term average biomass, at the start of the first year for which catches are available, for example, due to unrecorded catches or environmental variability (Dorn et al., 2006) (abbreviation "$B1 + B0$").

These three assessment methods lead to 13 alternative estimators for B_0 (Table 1):

(1) The equilibrium point of the stock–recruitment relationship (all three assessment methods).

(2) Average mean recruitment multiplied by the spawning biomass-per-recruit in the absence of exploitation (all three assessment methods). Two periods of average recruitment are evaluated:

(a) the entire period considered in the assessment;

(b) the first 10 years for which catch data are available.The second period used for calculating average recruitment captures the idea that the best estimate of average recruitment in an unfished state is the average of those recruitments before the stock was fished heavily. In actual applications, the selection of the years used for determining average recruitment may also account for the likely precision of the estimates of recruitment so that, although estimates of recruitment are available for the earliest years of exploitation, these may be ignored when estimating B_0 if there were few data to inform their magnitudes.

(3) The average value for $\hat{B}_0$ over the period for which catches are available, calculated by projecting the post-convergence estimates for the model parameters forward under zero fishing mortality ("Dynamic B_0"; MacCall et al., 1985) (all three assessment methods). This method calculates the average spawning biomass while holding all model parameter estimates fixed at

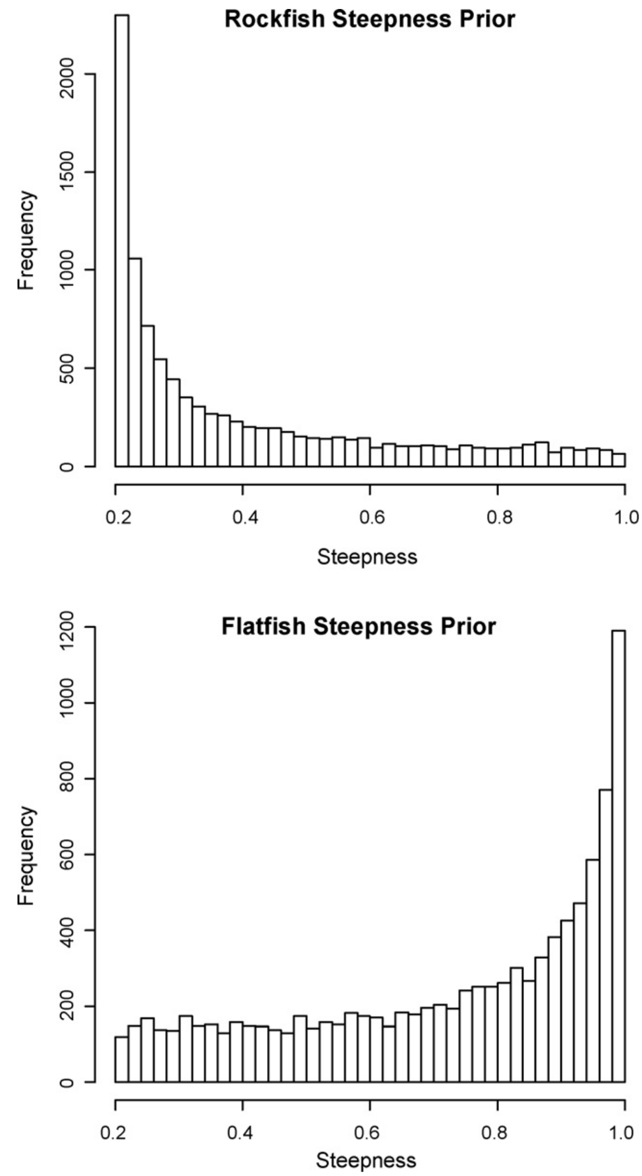

Fig. 1. Steepness priors used for estimation method $B0 + h$.

Table 1
Summary of the estimators considered in this paper

Source of estimator	B_0 (and B_1 for method (C))	B_{MSY}	Depletion (B_{last}/B_0)
Stock–recruitment relationship (B_{MSY} Method I)	4	4	4
Average recruitment (full assessment period; B_{MSY} Method II)	3	3	3
Average recruitment (first 10 years of the assessment period; B_{MSY} Method II)	3	3	3
Dynamic B_0	3	–	3
Average recruitment (most recent half of the assessment period; B_{MSY} Method III)	–	3	–
$B_{MSY} = 0.4B_0$ (B_{MSY} Method IV)	–	13	–
Total number of estimators	13	26	13

the values from the final stock assessment model and projecting the population forward without fishing.

(4) $\hat{B}_1$ (assessment method (C) only; for comparison with $\hat{B}_0$).

The 13 alternative estimates of B_0 lead to 13 corresponding estimates for current stock depletion, D, calculated as the ratio $D = B_{last}/B_0$, where B_{last} is the current spawning biomass (Table 1).

Four general methods can be used to estimate B_{MSY}. All except (IV), involve estimating $\hat{B}_{MSY}$ as $\tilde{B}_{MSY}\tilde{R}_{MSY}$ where $\tilde{B}_{MSY}$ is the spawning biomass-per-recruit (SBPR) corresponding to $\hat{F}_{MSY}$ (the estimate of F_{MSY}), and $\tilde{R}_{MSY}$ is the average recruitment corresponding to $\hat{F}_{MSY}$:

(I) Select $\hat{F}_{MSY}$ by maximizing the product of yield-per-recruit and recruitment as a function of fishing mortality; this method depends on the estimate of h.

(II) Calculate $\hat{B}_{MSY}$ setting $\hat{F}_{MSY}$ to the value of fishing mortality at which yield-per-recruit is maximized and set $\tilde{R}_{MSY}$ to the average recruitment over some subset of the years included in the assessment.

(III) Set $\hat{B}_{MSY}$ to 40% of the unfished spawning biomass-per-recruit given the average recruitment over the last half of the period of catches ($\hat{B}_{MSY} = 0.4\hat{\bar{R}}_{MSY}$ SBPR at $F = 0$).

(IV) Estimate $\hat{B}_{MSY}$ using the proxy, $\hat{B}_{MSY} = 0.4\hat{B}_0$; this proxy was selected because it is the default proxy for B_{MSY} used when applying control rules for West Coast groundfish species (PFMC, 2006).

The four general methods described above lead to 26 possible ways of estimating $\hat{B}_{MSY}$:

(1) Method (I) where the estimators for $\hat{B}_0$ that use a stock–recruitment relationship are used (all three assessment methods as well as the second estimate from assessment method (C), $\hat{B}_1$).

(2) Method (II) where the average recruitment is based on the estimates of recruitment for the whole assessment period (all three assessment methods).

(3) Method (II) where the average recruitment is based on the estimates of recruitment for the first 10 years of the assessment period (all three assessment methods).

(4) Method (III) where the average recruitment is based on the estimates of recruitment for most recent half of the assessment period (all three assessment methods).

(5) Method (IV) where all estimators and methods for estimating B_0 are used (13 estimates, one from each estimate of $\hat{B}_0$).

Table 2 summarizes the abbreviations used for the estimators considered in this paper.

2.3. Scenario parameterization

The values for the biological and fishery parameters (i.e. mass- and proportion mature-at-age, natural mortality, steepness, current stock depletion (i.e. current spawning biomass relative to B_0), and survey and fishery selectivity-at-age) were set to values similar to those from recent stock assessments for the three example species: canary rockfish, petrale sole, and Pacific hake (Helser et al., 2005; Lai et al., 2005; Methot and Stewart, 2005) (Fig. 2 and Table 3a). Table 3a does not list the value for B_0. Rather B_0 is calculated from the current depletion and the relative trend in catches. The current depletion, D, is specified as B_{last}/B_0, where B_{last} is the current spawning biomass. D equals the value in Table 3a if the population is projected forwards. Therefore, there is a single true value for B_0, B_{MSY}, and stock depletion for each simulation.

A range of plausible values were chosen for the remaining parameters thought to impact estimator performance (number of years of catches, extent of recruitment variability, extent of observation error, effective sample sizes for age-composition data, and number of years of age data and index length given a 50-year time-series of catches; Table 3b). A full factorial experiment was conducted for these parameters. Three values for the extent of recruitment variability were chosen. The middle value was set equal to the value used in the actual stock assessments and reasonable upper and lower bounds were chosen. Three values for the extent of observation error were chosen which should capture the true

Table 2
Notation used in plots and tables for the estimators

Estimator description	B_0	B_{MSY}	Depletion (B_{last}/B_0)
Stock–recruitment relationship, $\hat{B}_0$ equilibrium	SR	SR (I)	SR
Stock–recruitment relationship, $\hat{B}_1$ equilibrium	SR_{B1}	SR_{B1} (I)	SR_{B1}
Average recruitment during the whole period of catches ($\hat{B}_{MSY} = \bar{R}_{ALL}$ SBPR at F_{MAX})	$\bar{R}_{ALL}$	$\bar{R}_{ALL}$ (II)	$\bar{R}_{ALL}$
Average recruitment during the first 10 years of catches ($\hat{B}_{MSY} = \bar{R}_{F10}$ SBPR at F_{MAX})	$\bar{R}_{F10}$	$\bar{R}_{F10}$ (II)	$\bar{R}_{F10}$
Stock–recruitment relationship, $\hat{B}_0$ without catches	SR_{DynB0}	–	SR_{DynB0}
Stock–recruitment relationship, $\hat{B}_{MSY}$ proxy based on $0.4\hat{B}_0$ ($\hat{B}_{MSY} = 0.4\hat{B}_0$)		SR_{proxy} (IV)	
Stock–recruitment relationship, $\hat{B}_{MSY}$ proxy based on 40% of the dynamic $\hat{B}_0$ ($\hat{B}_{MSY} = 0.4$ Dynamic $\hat{B}_0$)		$SR_{DynB0.proxy}$ (IV)	
Stock–recruitment relationship, $\hat{B}_{MSY}$ proxy based on $0.4\hat{B}_1$ ($\hat{B}_{MSY} = 0.4\hat{B}_1$)		$SR_{B1.proxy}$ (IV)	
Average recruitment over the whole period of catches, $\hat{B}_{MSY} = 0.4\bar{R}_{ALL}$ SBPR at $F = 0$		$\bar{R}_{ALL.proxy}$ (IV)	
Average recruitment over the first 10 years of catches, $\hat{B}_{MSY} = 0.4\bar{R}_{F10}$ SBPR at $F = 0$		$\bar{R}_{F10.proxy}$ (IV)	
Average recruitment during the last half of the catch period, $\hat{B}_{MSY} = 0.4\bar{R}_{Lhalf}$ SBPR at $F = 0$		$\bar{R}_{Lhalf}$ (III)	

The annotations in the column for B_{MSY} and estimator description indicate which of the four methods outlined in the text is applied.

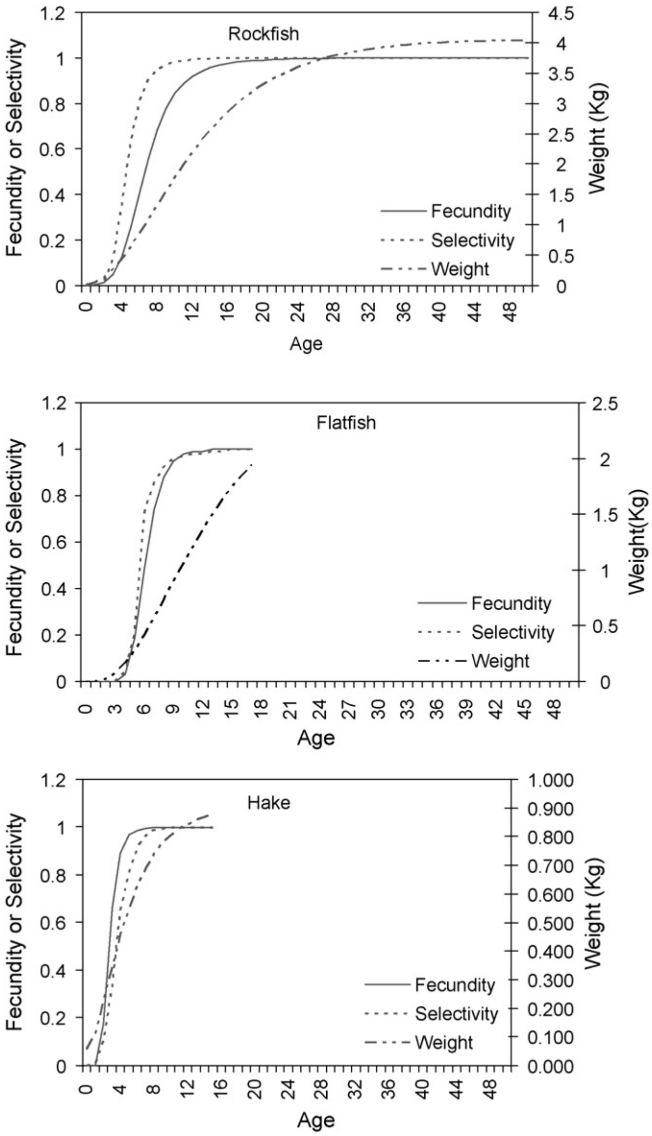

Fig. 2. Biological information used to parameterize the operating model.

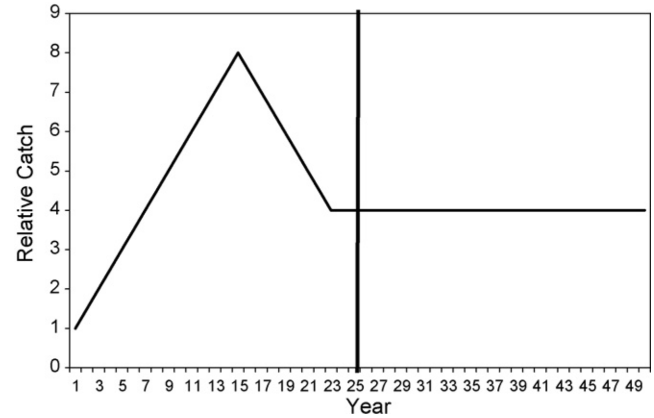

Fig. 3. The relative catch time-series used for the 25- and 50-year simulations. The vertical line indicates both the end of the 25-year catch time-series and the beginning of data collection for the trials with 50 years of catches and 25 years of survey data.

values from data collection programs for West Coast groundfish (Helser et al., 2005; Lai et al., 2005; Methot and Stewart, 2005). The combination of 25 years of index and age-composition data given a 50-year time-series of catches was chosen to approximate current data sets typically available. The other combinations of survey and catch data were chosen to approximate the types of data sets which might be available in the near future. The trend in catches is such that catches increase rapidly during the first 15 years, decrease to half the maximum catch during the next 8 years and then remain stable (Fig. 3).

There are 81 trials (based on the combinations of factors in Table 3) for each species life history and estimation method. Each treatment used the same set of random numbers so the differences among treatments could be investigated effectively. The total number of trials, accounting for the three species life histories and three assessment models, was 729. Each trial involved 100 replicates. Individual simulations are evaluated for convergence by determining if the eigenvalues of the Hessian matrix are positive, which indicates a minimum.

2.4. Performance evaluation

The relative errors (REs) and absolute relative errors (AREs) (Eqs. (16) and (17)) for B_0, B_{MSY}, stock depletion, and h for each estimation method, expressed as percentages, are calculated for each replicate (individual simulation) within each trial resulting in 100 REs and AREs for each trial, and estimator.

$$\mathrm{RE} = 100\left(\frac{E-T}{T}\right) \tag{16}$$

$$\mathrm{ARE} = 100\left|\frac{E-T}{T}\right| \tag{17}$$

where E denotes the estimated value of some quantity of interest and T denotes its true value.

Table 3
The parameters and factors considered for each species

Species	Rockfish	Flatfish	Hake
(a) Parameters which are the same for all simulations			
Depletion, B_{last}/B_0	0.1	0.3	0.35
Steepness, h	0.3	0.8	0.75
Natural mortality (year^{-1})	0.12	0.2	0.23
(b) Parameters which vary among the simulation scenarios			
Extent of recruitment variation, σ_R	0.2, 0.4, 0.6	0.3,0.5,0.7	0.9, 1.13, 1.3
Extent of observation error, σ_q	0.2, 0.4, 0.6	0.2, 0.4, 0.6	0.2, 0.4, 0.6
Age-composition sample size	10, 50, 200	10, 50, 200	10, 50, 200
Length of catch time-series, number of years	(25, 25)	(25, 25)	(25, 25)
of survey and age-composition data given 50	(50, 50)	(50, 50)	(50, 50)
years of catches[a]	(50, 25)[a]	(50, 25)[a]	(50, 25)[a]

[a] Age-composition and survey index data are available for the most recent 25 years.

Table 4
The MAREs for each species life history, across all trials and replicates, for each estimation method and B_0 estimator

Species	Estimation method	SR_{B1}	SR	$\overline{R}_{ALL}$	$\overline{R}_{F10}$	SR_{DynB0}
Rockfish	*B0*		**7.01**	*37.81*	9.56	**6.97**
	B1+B0	8.43	9.42	*37.87*	*10.08*	7.52
	B0+h		**7.01**	*37.80*	9.56	**6.96**
Flatfish	*B0*		9.77	*35.12*	11.46	**9.37**
	B1+B0	12.76	**7.75**	*43.73*	*14.14*	13.09
	B0+h		**9.76**	*35.01*	11.49	**9.42**
Hake	*B0*		39.33	*58.04*	**38.98**	**34.18**
	B1+B0	*50.73*	46.09	*67.09*	**38.27**	34.51
	B0+h		39.22	*58.23*	39.05	**34.50**

The estimators identified to be among the "best" 25% for each species life history are bolded and the single "best" estimator for each species is highlighted. The "worst" 25% of estimators for each species life history are italicized. Values that are neither bolded nor italicised indicate estimators that are neither "best" nor "worst".

The median of the absolute relative errors (MARE) is used as a single statistic to summarize performance across either replicates of an individual trial or replicates of multiple trials. The MARE represents a single value that summarizes an estimator's bias and imprecision, such that higher values indicate poor performance and values close to zero indicate good performance. This statistic has been used in several previous evaluations of estimator performance (e.g. Punt, 2003b; Wilberg and Bence, 2006).

For the purposes of these simulations, the "best" set of estimators for each species life history, relative to the set of estimators evaluated, was identified as those falling within the lowest 25% of the MAREs for each estimator when the MAREs were calculated across all replicates and trials for each species life history and estimator. Basing the MAREs on all of the simulations provides an overall measure of performance. The set of "best" estimators is based on the median because summary statistics based on the mean are sensitive to outliers from large errors or non-converged simulations. Conversely, the "worst" set of estimators, relative to the set of estimators evaluated, was identified as those whose MAREs across all replicates and trials fell within the highest 75% of MAREs for each species life history and estimator across all replicates and trials.

Following Dichmont et al. (2006) and Yin and Sampson (2004), the importance of the factors in Table 3b, and their first-order interactions, in determining estimator performance was explored by fitting a series of General Linear Models to the logarithms of the MAREs for each trial and life history type, separately for B_0, B_{MSY}, stock depletion, and h. The GLM models included: (a) models for

Table 5
As for Table 4, except the results are for estimators of depletion

Species	Estimation Method	SR_{B1}	SR	$\overline{R}_{ALL}$	$\overline{R}_{F10}$	SR_{DynB0}
Rockfish	*B0*		**18.54**	*62.69*	20.94	19.37
	B1+B0	**17.60**	**18.92**	*61.79*	21.89	*23.36*
	B0+h		**18.54**	*62.69*	20.94	19.36
Flatfish	*B0*		**14.57**	*36.49*	16.56	**15.89**
	B1+B0	22.52	*23.29*	*39.34*	19.17	22.47
	B0+h		**14.55**	*36.52*	16.52	**16.08**
Hake	*B0*		**40.49**	56.02	**49.99**	54.33
	B1+B0	*71.67*	*75.36*	*80.14*	*70.61*	65.73
	B0+h		**40.24**	56.12	**49.72**	58.40

Table 6
As for Table 4, except the results are for estimators of B_{MSY}

Species	Estimation Method	SR_{B1}	SR	$\overline{R}_{ALL}$	$\overline{R}_{F10}$	$SR_{B1.proxy}$	SR_{proxy}	$\overline{R}_{ALL.proxy}$	$\overline{R}_{F10.proxy}$	$\overline{R}_{Lhalf}$	$SR_{DynB0.proxy}$
Rockfish	*B0*		**9.68**	*74.82*	*63.06*		**12.53**	43.29	17.08	62.71	12.74
	B1+B0	**9.38**	22.40	*74.75*	*62.92*	**12.31**	14.02	*43.31*	16.58	42.32	**12.21**
	B0+h		**7.02**	*74.82*	*63.06*		**12.53**	43.28	17.08	62.71	12.74
Flatfish	*B0*		**39.07**	**41.47**	**40.05**		*67.73*	*67.80*	60.50	44.74	66.62
	B1+B0	44.18	**32.85**	42.50	**39.42**	66.56	61.82	*76.94*	*67.85*	56.21	*102.76*
	B0+h		**32.15**	41.49	**40.04**		*67.69*	*67.80*	60.36	44.60	66.60
Hake	*B0*		64.15	**41.78**	**42.47**		*75.56*	*72.71*	55.23	**47.38**	63.54
	B1+B 0	*73.24*	69.67	**47.35**	**41.23**	*73.56*	65.75	*98.30*	20.92	64.83	*166.33*
	B0+h		58.55	**41.87**	**42.38**		*75.80*	71.91	55.58	47.50	63.38

each individual factor in Table 3b on its own, (b) a model with all of the main factors in Table 3b, and (c) a model with main factors and first-order interactions selected using AIC (Burnham and Anderson, 2000). The purpose of the last analysis was to assess the importance (rather than statistical significance) of interactions among factors.

3. Results and discussion

3.1. Best and worst estimators

There is no single best or worst estimator across all estimated quantities, stock assessment methods, and species (Tables 4–7 and Figs. 4–6). However, the “best” estimators of B_0 and stock depletion are typically based on estimating the stock–recruitment relationship (bold entries in Tables 4–6). Using estimators based on the stock–recruitment relationship from the B_1+B_0 stock assessment method to estimate stock depletion for the hake and flatfish life histories is an exception to the above as these are amongst the “worst” set of estimators (Tables 4 and 5 and Figs. 4 and 5). Average recruitment-based estimators of B_0 and stock depletion generally fell into the “worst” set of estimators, except for the hake for which estimating stock depletion based on average recruitment in the first 10 years ($\bar{R}_{F10}$) was among the “best” estimators (Tables 4 and 5 and Figs. 4 and 5). The “best” estimators of B_{MSY} varied among life histories. The stock recruitment-based estimators, or proxies based on estimating B_{MSY} as $0.4\hat{B}_0$, were most robust, had the lowest MARE, for the rockfish (Table 6 and Fig. 4). Estimating B_{MSY} from the fit of the stock–recruitment relationship, as well as based on the average recruitment during the first 10 years of the catch history ($\bar{R}_{F10}$), was most robust, had the lowest MARE, for the flatfish (Table 6 and Fig. 5). The average recruitment estimators ($\bar{R}_{\mathrm{ALL}}$ or $\bar{R}_{F10}$) were most robust, had the lowest MARE, for the hake (Table 6 and Fig. 6). While basing reference points for B_0, stock depletion, and B_{MSY} on the

Table 7
As for Table 4, except the results are for each estimation method and steepness, *h*

Species	*B0*	*B1+B0*	*B0+h*
Rockfish	**12.43**	13.81	**0.00033**
Flatfish	**18.91**	24.91	**0.00025**
Hake	33.33	33.33	**0.00107**

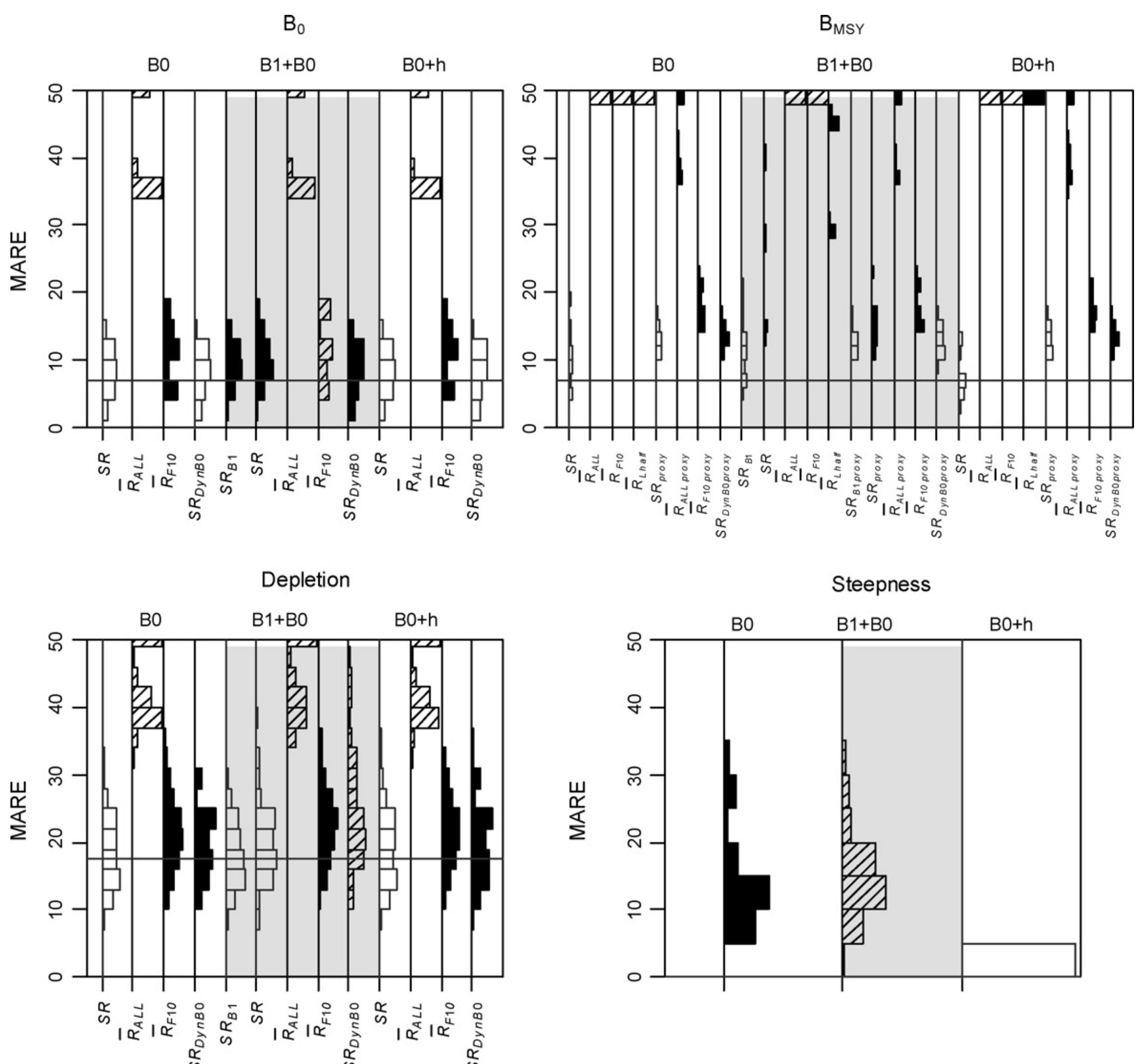

Fig. 4. Distribution of the MAREs for each estimator of B_0, B_{MSY}, current biomass relative to B_0, and stock–recruitment steepness (h) for each of the 81 trials for the rockfish. Labels across the top of each panel indicate the three methods of stock assessment. Labels across the bottom of each panel indicate each estimator (Table 2). The “best” 25% of estimators are graphed as white bars and the “worst” 25% of estimators are graphed as grey striped bars. The horizontal line across each panel indicates the median of the AREs, across all trials and replicates, for the single best estimator for the quantity concerned (lowest median ARE).

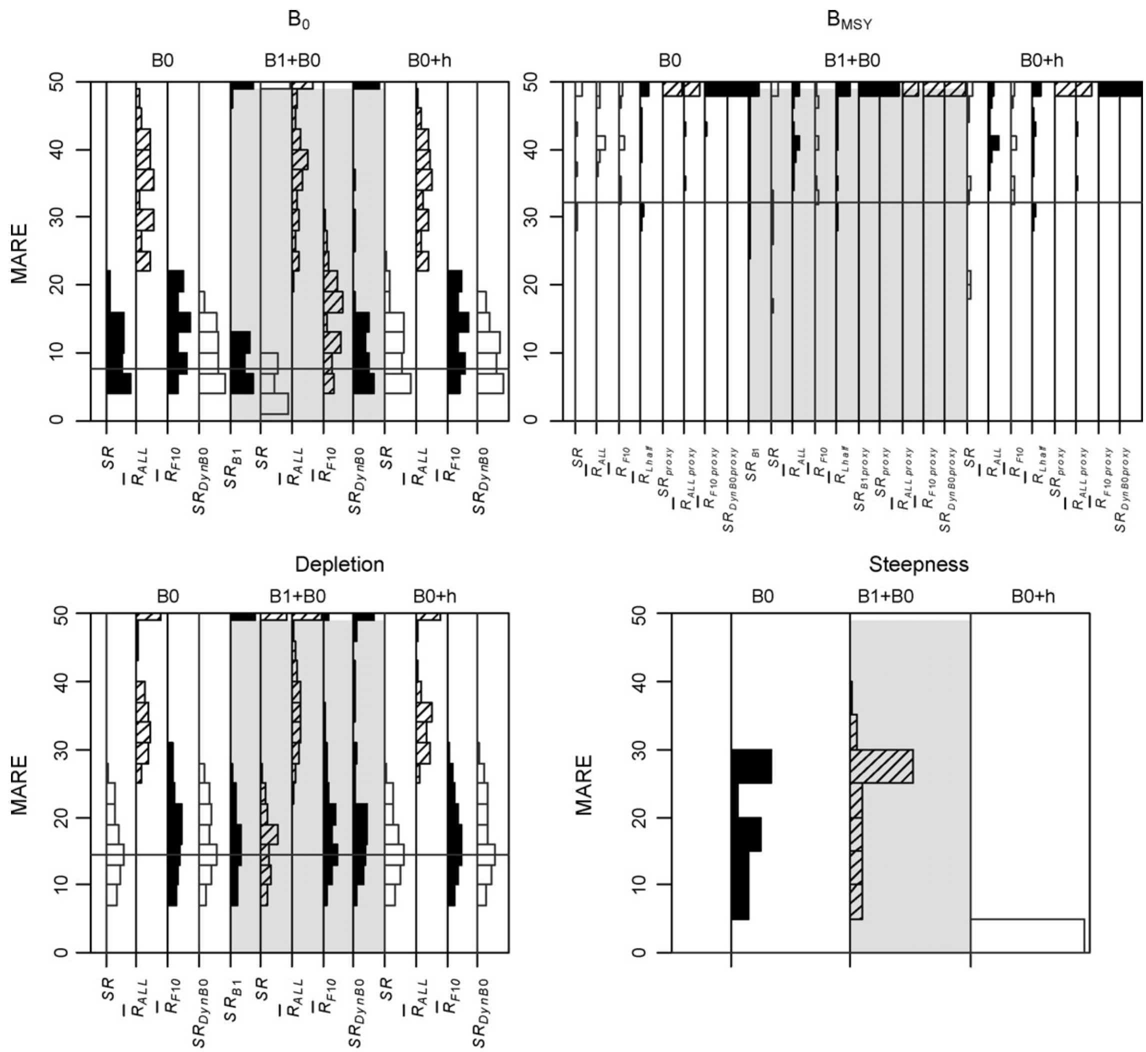

Fig. 5. As for Fig. 4, except that results are shown for the flatfish.

time-series of recruitments consistently led to poorer performance for the rockfish and flatfish, basing reference points for B_0, stock depletion, and B_{MSY} on the time-series of recruitments ($\bar{R}_{ALL}$ or $\bar{R}_{F10}$) consistently led to good performance for the hake (Tables 4–6). For the flatfish and hake, the B_{MSY} estimators based on assuming that B_{MSY} occurred at 40% of $\hat{B}_0$ were among the "worst" estimators, except for the rockfish (Table 6 and Figs. 4–6).

The catch time-series used in the simulations increases catches quickly at the beginning of the time-series (Fig. 3). Rapidly increasing catches might be expected to lead to poor performance of average recruitment-based estimators. In contrast, if there is a period of time during which fishing pressure is minimal and data are being collected then the performance of average recruitment estimators may improve, particularly those estimators based on the average recruitment during the period of light exploitation. However, there are likely few instances where monitoring starts prior to increasing levels of fishery exploitation. Data collection is often reactive, beginning during, or after, a fishery is exploiting a population at rather high levels (e.g. see Fig. 7 for actual exploitation history for canary rockfish). Use of recruitment estimates for year-classes spawned after the start of intense exploitation for defining average recruitment when estimating B_0, may lead to misleading results because expected recruitment may be decreasing due to a reduction in spawning biomass.

3.2. Best estimation method

Comparing results from the $B0$ and $B0+h$ stock assessment methods allows an evaluation of the impact of uncertainty in steepness on the ability to estimate other quantities of interest. The results for these two estimation methods are generally very similar (Tables 4–7). As expected, h is estimated best when a highly informative prior is included in the assessment ($B0+h$) (Table 7 and Figs. 4–6). In fact, the relative errors for $\hat{h}$ from the $B0+h$ stock assessment method are only slightly higher than zero. In contrast, the $B1+B0$ stock assessment method produces the poorest estimates of h for both the rockfish and the flatfish (Table 7 and Figs. 4 and 5), while the $B0$ and $B1+B0$ stock assessment methods perform similarly at estimating h for the hake (Table 7 and Fig. 6). By and large, the $B0+h$ stock assessment method performs similarly to the $B0$ stock assessment method suggesting that a good estimate of h does not impact the ability to estimate biomass reference points (Tables 4–6).

Comparing the performance of estimators based on the $B0$ and $B1+B0$ stock assessment methods addresses the question of the cost of estimating extra parameters when, in fact, this is not necessary. This is because, although there is recruitment variability prior to the first year with catches, the current set of simulations does not include unobserved physical forcing or unreported his-

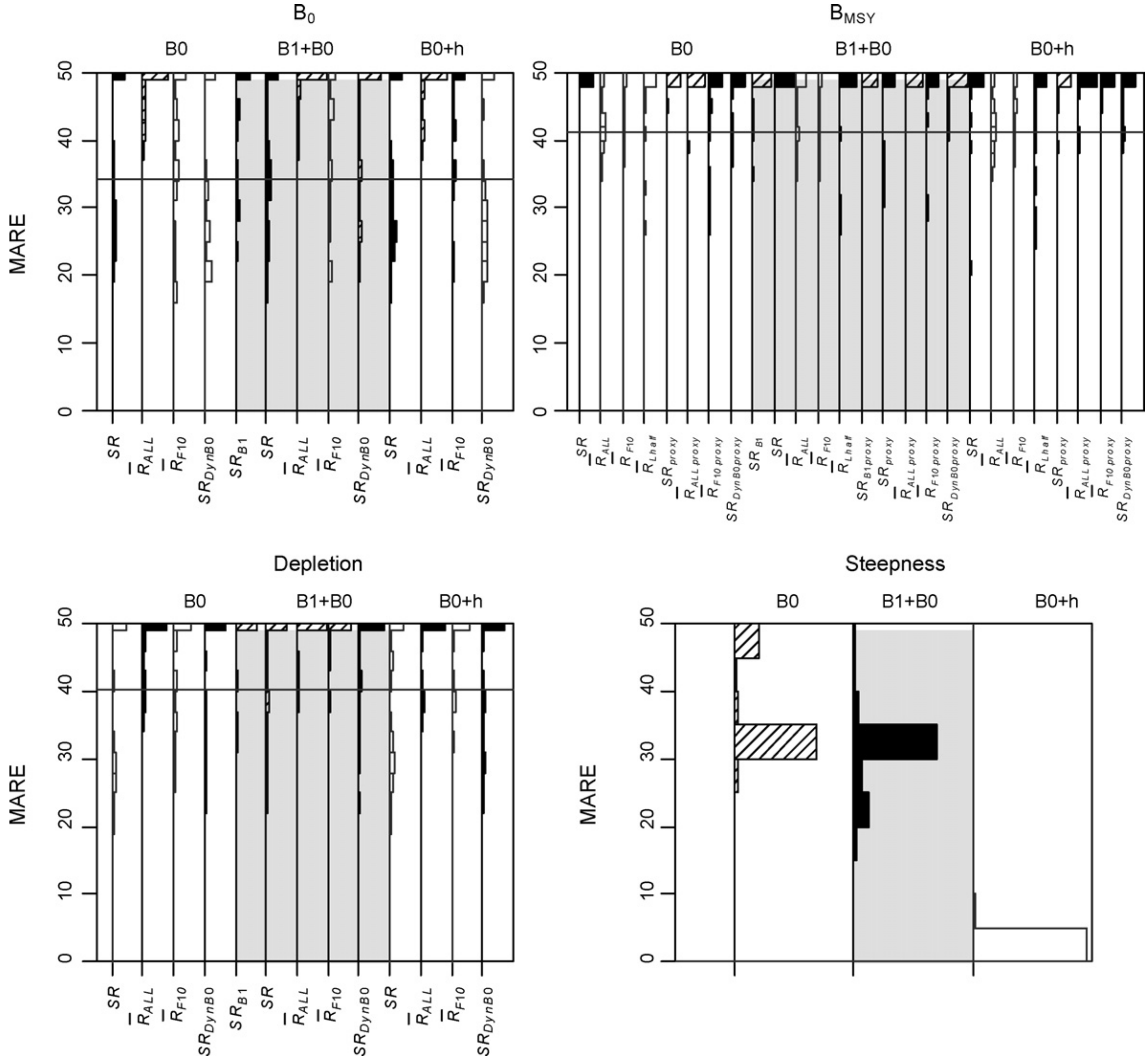

Fig. 6. As for Fig. 4, except that results are shown for the hake.

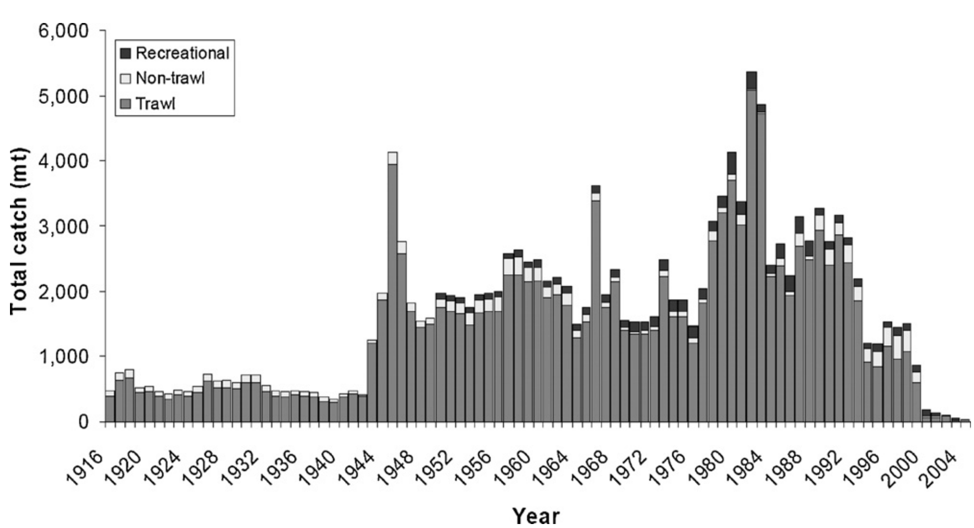

Fig. 7. Total landings of canary rockfish from 1916 through 2004 for each sector of the fishery. Reliable survey data are available beginning in 1980 (Methot and Stewart, 2005).

Table 8
The percentage of deviance explained by fitting GLMs for each factor, all main effects (Table 3b) and a model with main effects and first-order interactions to the MAREs for each estimation method, estimator, and species life history

Species	Estimated Quantity	Estimation Model	Estimator	σ_R	Length of the Catch Time Series	σ_q	Age-Composition Sample Size	Number of Years of Survey Data	All Main Factors	First Order Interactions
Rockfish	B_0	$B0$	SR	**95.50**	0.06	0.01	0.45	0.03	96.02	99.24
			SR_{DynB0}	**94.63**	0.01	0.00	2.01	0.03	96.73	99.091
		$B1+B0$	SR	**58.90**	8.23	0.05	6.70	3.53	88.50	98.90
	Depletion	$B0$	SR	15.44	0.15	**63.29**	0.64	5.86	88.62	96.16
		$B1+B0$	SR_{B1}	11.24	0.86	**57.70**	1.78	11.45	91.31	96.53
			SR	19.66	0.98	**44.58**	1.70	12.84	89.09	95.37
	B_{MSY}	$B0$	SR	**73.84**	12.67	0.01	3.73	2.25	90.34	99.51
			SR_{proxy}	**72.36**	13.28	0.03	7.26	5.13	93.19	97.51
		$B1+B0$	SR_{B1}	**57.90**	2.45	0.29	26.03	11.82	96.06	99.07
			$SR_{B1.PROXY}$	7.35	**33.12**	0.25	5.50	5.37	82.19	90.12
			$SR_{DynB0.PROXY}$	**68.76**	4.35	0.23	7.73	11.82	88.73	95.11
	h	$B0$		32.33	**53.01**	0.16	7.73	17.71	93.66	99.04
		$B1+B0$		32.65	**33.86**	1.43	21.28	14.94	90.44	97.9
Flatfish	B_0	$B0$	SR_{DynB0}	**87.68**	0.42	0.03	0.59	3.22	94.72	98.17
		$B1+B0$	SR	0.10	20.20	0.00	6.43	**25.35**	97.45	99.81
	Depletion	$B0$	SR	32.64	1.09	**47.97**	1.09	3.40	90.26	94.38
			SR_{DynB0}	**44.66**	6.29	34.12	2.13	0.21	91.13	94.05
	B_{MSY}	$B0$	SR	**77.72**	0.35	0.09	0.64	6.72	89.93	97.41
			$\bar{R}_{ALL}$	**87.55**	4.32	0.13	0.26	1.29	92.26	96.39
			$\bar{R}_{F10}$	**94.25**	0.28	0.32	0.17	0.26	95.82	98.14
		$B1+B0$	SR	2.66	1.54	0.71	0.74	**55.00**	91.76	98.25
			$\bar{R}_{F10}$	**86.93**	0.08	0.04	1.24	0.33	88.96	95.83
	h	$B0$		27.51	**49.66**	0.25	0.03	4.28	80.26	99.01
		$B1+B0$		6.54	12.39	0.27	0.60	**71.56**	79.64	98.03
Hake	B_0	$B0$	$\bar{R}_{F10}$	**31.20**	25.73	0.27	2.79	2.94	84.07	98.78
			SR_{DynB0}	0.27	18.42	0.08	5.34	**18.55**	79.62	98.17
		$B1+B0$	$\bar{R}_{F10}$	**51.89**	15.93	0.20	0.87	3.08	87.63	96.99
	Depletion	$B0$	SR	0.49	**42.92**	9.26	11.17	0.30	83.29	94.02
			$\bar{R}_{F10}$	8.86	**28.12**	7.13	21.28	0.87	69.33	95.32
	B_{MSY}	$B0$	$\bar{R}_{ALL}$	**31.51**	9.14	3.86	6.44	0.29	56.53	89.25
			$\bar{R}_{F10}$	**78.53**	1.20	0.31	6.97	2.07	88.07	96.00
			$\bar{R}_{Lhalf}$	**38.05**	0.00	0.03	31.15	2.15	70.91	90.31
		$B1+B0$	$\bar{R}_{ALL}$	6.54	**22.66**	6.16	12.05	5.50	58.94	89.56
			$\bar{R}_{F10}$	**33.86**	12.17	0.79	9.86	3.29	73.55	95.89
	h	$B0$		0.33	10.89	0.49	11.02	**20.99**	74.49	97.99
		$B1+B0$		0.02	8.95	10.39	4.78	**17.24**	66.67	88.84

Results are shown only for the "best" estimators (i.e. the highlighted estimators in Tables 4–7) and the $B0$ and $B1+B0$ stock assessment methods (results for $B0+h$ are similar to those for the $B0$ stock assessment method). The factor explaining the most deviance individually is highlighted.

torical catches. The differences between the results for the $B0$ and $B1+B0$ stock assessment methods are generally least for the rockfish life history. This is hardly surprising because the rockfish has the lowest variation in recruitment (Table 3b) so the impact of random recruitment variation before the start of the catch series is small for this species. The stock–recruitment estimator for B_0 based on the $B1+B0$ stock assessment method outperforms all other B_0 estimators for the rockfish life history (Table 4). However, the stock–recruitment estimator for stock depletion from the $B1+B0$ stock assessment method performed consistently worse than the other estimators of stock depletion for the flatfish and hake life histories (Table 5). Estimators of B_{MSY} from the $B1+B0$ stock assessment method are consistently among the best set of estimators for each life history (Table 6). In most cases, the performance of the other estimators based on $B1+B0$ stock assessment method have differences in the MARE of <5% compared to the same estimators from the $B0$ stock assessment method (Tables 4–6). It might be expected that the $B1+B0$ stock assessment method would perform better than the $B0$ stock assessment method for the hake because this species has the greatest extent of recruitment variation (Table 3b) and, in principle, the $B1+B0$ stock assessment method can account for historical variation in recruitment. However, the $B1+B0$ stock assessment method does not appear to outperform the other stock assessment methods for most estimators (Tables 4–6).

3.3. Importance of factors

Comparing the AIC-selected first-order interaction model with the full main effects model indicates that interactions were generally not needed to explain a high proportion of the deviance for the rockfish and flatfish life histories; main effects are sufficient (Table 8). The interaction model explains at least 20% more

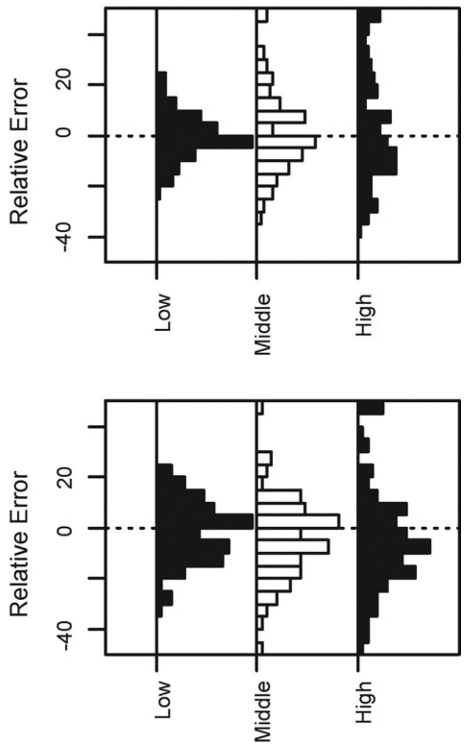

Fig. 8. Relative errors (RE) for the flatfish life history for B_0 (top panel) and stock depletion (bottom panel) for the estimators based on the stock–recruitment relationship and the $B0$ stock assessment method. The panels show the impact of the three levels of recruitment variability for a single set of high data quality (50-year time-series of catches with age data for all years and low observation error) trials.

deviance than the main effects model for the hake life history for nearly half of the estimators (Table 8). Interactions between factors dealing with recruitment variability, and data quality and quantity are largely responsible for explaining the increase in deviance explained for these cases, which suggests that the length over which recruitments are observed, as well as how well recruitments are observed, impacts the estimators, particularly when recruitment variation is high.

The factors in this study can be categorized as pertaining to data quality, data quantity, and process error. As a main effect, observation error (data quality; σ_q, Table 8) is generally unimportant, only impacting the estimation of stock depletion, which is clearly dependent upon having good data on stock size at the end of the time-series. Often data on recent stock size are not highly informative as recent recruitments have not been fully observed by the fishery or surveys. The age-composition sample size did not explain much of the deviation for any of the quantities (Table 8). Recruitment variability, the number of years of survey data, and the length of the time-series (data quantity) are commonly identified factors for the hake, which might be expected because the hake has high recruitment variation and not having data near the start of the time-series of catches may severely impact the ability to estimate B_0 (and hence the other biomass reference points). Accordingly, either the number of years with survey data or the length of the catch time-series was important in determining each stock assessment method's ability to estimate h. Recruitment variability is important for nearly all of the estimators (Table 8 and Fig. 8), which is consistent with the findings of Yin and Sampson (2004).

While these simulations do not directly investigate the impact of a fragmented catch history, it appears that having a complete catch history and a survey index with low levels of observation error are more important for estimating B_0 and stock depletion than having a good estimate of h. The dependence of estimates of B_0 and stock depletion on catch history has been commented on previously, for example, for flatfish (Lai et al., 2005; Sampson and Lee, 1999) where estimates of B_0 and stock depletion are sensitive to whether or not all of the historical catches are included in the assessment. An incomplete catch history can subject stock assessment results to a "shifting baseline" problem, whereby the true magnitude of stock declines are masked by data streams which begin after the stock is already depleted.

3.4. Life history types

The results in Tables 4–7 and Figs. 4–6 highlight the differences among estimators and species life history in terms of estimating the four quantities of interest. Trials for the hake generally show poor performance with large variances in comparison to the other two life histories. The large variances appear to be caused primarily by highly variable recruitment. Due to the noisy nature of the simulated data, the stock assessment methods frequently fail to converge for the hake while the same methods did not encounter this problem for the rockfish and the flatfish life histories (Fig. 9), once again illustrating the uninformative nature of the simulated data for the hake. In addition, for all life histories, the $B1+B0$ stock assessment method had more problems with convergence than the other stock assessment methods (Fig. 9).

3.5. Performance of estimated quantities

Contrary to the results of Yin and Sampson (2004) and Punt et al. (2002), which suggest that B_0 and B_{last} share similar levels of error which factor out in the calculation of stock depletion, B_0 was estimated either better (rockfish and flatfish) or similarly (the hake) to stock depletion (Tables 4 and 5). The enhanced ability to estimate B_0 suggests that the biomass at the end of the model time period is not well determined, most likely due to incomplete information about incoming recruitments, subjecting B_{last} to a higher level of error which does not factor out in the calculation of stock depletion. As expected, calculating B_{MSY} directly from the stock–recruitment relationship performs best for the life histories with lower recruitment variability (rockfish and flatfish) (Table 6). However, calculating B_{MSY} using average recruitment performs best for the hake life history (Table 6).

Generally all of the proxy methods of calculating B_{MSY} performed relatively poorly in comparison to the non-proxy measures (Table 6). However, the operating model values for B_{MSY}/B_0 are 0.448, 0.275, and 0.292 for the rockfish, flatfish, and hake, respectively. Since $B_{MSY} \neq 0.4B_0$, the relatively poor performance of the estimators based on the assumption $B_{MSY}/B_0=0.4$ is not unexpected. In this case the 40% proxy is not conservative enough for the rockfish and too conservative for the flatfish and hake. The even poorer performance of the 40% $\hat{B}_0$ proxy for the hake compared to the flatfish is due, in part, to poor convergence of more simulations for this life history (Fig. 9).

Myers et al. (1994) recommended against B_{MSY} proxies dependent upon estimating virgin biomass due to estimation problems and the inappropriateness of applying such a proxy universally, favouring instead a B_{MSY} proxy based on average recruitment. In contrast, Restrepo et al. (1998) recommended calculating B_{MSY} based on average recruitment times some pre-specified fraction (0.4 in the case of this paper). The results of this paper suggest that the $B_{MSY} \sim 0.4\hat{B}_0$ proxy used by the PFMC (PFMC, 2006) is outperformed by either the stock recruitment or average recruitment-based estimators, depending upon the life history (Table 6).

Myers et al. (1994) suggest that when virgin biomass is unknown it may be approximated by the product of average recruitment and SBPR at $F = 0$ (somewhat similar to the $\hat{\hat{R}}_{\mathrm{ALL}}$ and $\hat{\hat{R}}_{F10}$ estimators of B_0 in the current paper). Such an approach seems appropriate when stock productivity is high or when a stock is lightly exploited. Currently, the PFMC Groundfish Management Plan recommends, when possible, estimating B_0 by multiplying unfished spawning biomass-per-recruit by average recruitment during a period of high stock biomass (PFMC, 2006), similar to the $\hat{\hat{R}}_{F10}$ estimator in this paper. However, this approximation may be unrealistic because it assumes an absence of density-dependent changes in growth, survival, or age at maturity during the "fishing down" period (Restrepo et al., 1998). Generally this paper finds that it is better to estimate B_0 via the stock–recruitment relationship (Table 4).

Estimators of B_0 and stock depletion based on dynamic B_0 are intended to track changes in stock size due to causes other than fishing (MacCall et al., 1985; Sibert et al., 2006). While this paper does not explicitly include non-fishing impacts on the population dynamics, the high recruitment variability for the hake life history can be considered an approximation for non-fishing impacts. Both B_0 and stock depletion as estimated by dynamic B_0 perform within the best 25% of estimators on several occasions, but perform quite poorly when based on the $B1 + B0$ stock assessment method for the species with faster dynamics (the flatfish and the hake). The lack of difference in results between "static" and "dynamic" estimators (SR and $\mathrm{SR}_{\mathrm{DynB0}}$ in Tables 4 and 5) based on the $B0$ stock assessment method for the rockfish can be explained by lower variation in recruitment than for the other species. The $B1 + B0$ stock assessment method becomes more attractive, in principle, as recruitment variability increases because it theoretically has the ability to capture changes in stock size prior to the catch time-series. However, the performance of the dynamic B_0 estimator, in combination with the $B1 + B0$ stock assessment method, is in many cases among the worst estimators (Tables 4 and 5).

3.6. Climate considerations

This study is predicated on a stable stock–recruitment relationship and does not explicitly consider strongly autocorrelated recruitment deviations or more extreme shifts in the stock–recruitment relationship. The PFMC defines B_0 as the average stock size in the absence of exploitation (PFMC, 2006). Variation in average productivity over time due to, for example, regime shifts or directed climate change (e.g. as discussed in IPCC, 2001) imply the need to: (1) calculate average recruitments over the full range of environmental variability or (2) identify regime shifts in a timely fashion so that recruitment can be averaged during quasi-permanent ecosystem shifts (e.g. Dorn et al., 2006). How-

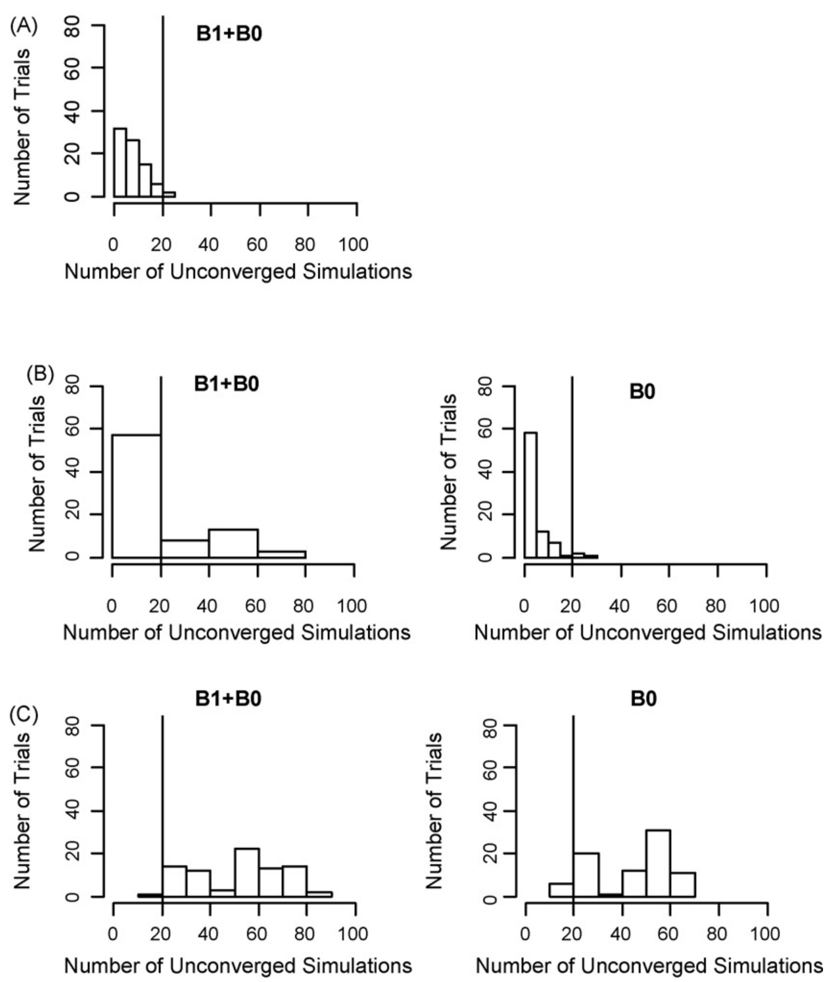

Fig. 9. The number of trials, out of the 81 different combinations of experimental factors for each life history (Table 3b) vs. the percent of the simulations trial that did not converge (out of a total of 100). For example, the first bar in Panel A indicates that there were about 30 trials which had between 0 and 5 individual unconverged simulations. Results are shown in panels A–C for the rockfish, flatfish, and hake life histories, respectively. The vertical line divides the trials into those with fewer and greater than 20% unconverged total simulations. Results for the $B0 + h$ stock assessment method are not markedly different from those for the $B0$ stock assessment method and are not shown. Panel A only shows the results for the $B1 + B0$ stock assessment method because the $B0$ stock assessment method did not have convergence problems.

ever, time-series of recruitment estimates are often far too short for this. Investigating variable environmental influences is particularly important because environmental forcing is translated in age-structured models as high recruitment variability, making estimation of biomass reference points needed for management more difficult. The impact of environmental forcing, which typically has some level of autocorrelation, provides an explanation of why stock–recruitment models with autocorrelated error structure often fit time-series of stock and recruitment data better that those without autocorrelation (Brodziak et al., 2001; Ianelli, 2002). Furthermore, Clark (1993) found that it was difficult to maintain target spawning biomass levels given the commonly accepted fishing policies of $F_{35\%}$–$F_{45\%}$ when recruitment is strongly correlated. The North Pacific Fishery Management Council calculates B_{MSY} using the average post-1977 recruitment when fishing at $F_{40\%}$ (essentially the $B1+B0$ estimator) due to an observed shift in long-term recruitment patterns, i.e. this approach attempts to address regime shift impact on the estimation of biomass reference points (NPFMC, 2006). The impact of long-term, low frequency, climate variability, and directional climate change on the performance of estimators is ignored in this study, but may be important, and should form the basis for additional work.

3.7. General discussion

This study provides a starting point for additional, more specific, simulation studies considering the effects of fishing on fish stock population dynamics. For example, the current experimental design assumes that the whole catch time-series is known and therefore results do not apply to cases where a large segment of catches are unobserved. The biological parameters are also treated as known without error, which could be estimated or varied in future studies. Likewise the extent of observation and process error are treated as known without error in the current simulations. In many cases, fishery selectivity also varies over time, effectively masking the recruitment signal. The trials in this study, which are most similar to many situations in the U.S., are those with 50 years of catch data and only 25 years of survey data. The results of this study also include trials with longer time-series of survey data suggesting that actual percent MAREs may be optimistic. However, the general patterns between estimators are expected to remain the same.

Acknowledgements

MAH was funded by a NMFS/Sea Grant Population Dynamics Fellowship (NMFS Grant #NA17RG1353, project E/I-5). AEP and MAH acknowledge funding through NMFS grant NA04NMF4550330. Thanks to I. Stewart, R. Methot, J. Cope, J. Bence, B. Clark and an anonymous reviewer whose comments greatly improved this manuscript.

References

Brodziak, J.K.T., Overholtz, W.J., Rago, P.J., 2001. Does spawning stock affect recruitment of New England groundfish? Can. J. Fish. Aquat. Sci. 58, 306–318.
Burnham, K.P., Anderson, D.R., 2000. Model Selection and Inference: A Practical Information-theoretic Approach. Springer-Verlag, New York.
Clark, W.G., 1993. The effect of recruitment variability on the choice of a target level of spawning biomass per recruit. In: Kruse, G., Eggers, D.M., Marasco, R.J., Pautzke, C., Quinn II, T.J. (Eds.), Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations. Alaska Sea Grant College Program Report No. 93-02. University of Alaska, Fairbanks, pp. 233–246.
Dichmont, C.M., Deng, A., Punt, A.E., Venables, W., Haddon, M., 2006. Management strategies for short-lived species: the case of Australia's northern prawn fishery. 3. Factors affecting management and estimation performance. Fish. Res. 82, 235–245.
Dorn, M.W., Aydin, K., Barbeaux, S., Guttormsen, M., Megrey, B., Spalinger, K., Wilkins, M., 2006. Assessment of walleye pollock in the Gulf of Alaska. In: 2006 North Pacific Marine Fishery Council Gulf of Alaska Stock Assessment and Fishery Evaluation Reports for 2007. North Pacific Fishery Management Council, 605 W 4th Avenue, Suite 306, Anchorage, AK 99501, USA, pp. 39–146 (Chapter 1).
Helser, T.E., Fleischer, G.W., Martell, S., Taylor, N., 2005. Stock Assessment of Pacific Hake (whiting) in U.S and Canadian Waters in 2004. Pacific Fishery Management Council, 7700 NE Ambassador Place, Suite 200, Portland, OR 97220, USA.
Ianelli, J.N., 2002. Simulation analyses testing the robustness of productivity determinations from west coast Pacific Ocean perch stock assessment data. N. Am. J. Fish. Manage. 22, 301–310.
IPCC, 2001. In: Houghton, J.T., Ding, Y., Griggs, D.J., Noguer, M., van der Linden, P.J., Xiaosu, D. (Eds.), Climate Change 2001: The Scientific Basis. Cambridge University Press, pp. 527–578.
Lai, H.-L., Haltuch, M.A., Punt, A.E., 2005. Stock assessment of petrale sole: 2004. Status of the Pacific Coast Groundfish Fishery Through 2005, Stock Assessment and Fishery Evaluation: Stock Assessments and Rebuilding Analyses, vol. 1. Pacific Fishery Management Council, 7700 NE Ambassador Place, Suite 100, Portland, OR 97220, USA.
MacCall, A.D., Klingbeil, R.A., Methot, R.D., 1985. Recent increased abundance and potential productivity of Pacific mackerel (*Scomber japonicus*). CalCOFI Rep. 26, 119–129.
Methot, R.D., Stewart, I.J., 2005. Status of U.S. canary rockfish resource in 2005. Status of the Pacific Coast Groundfish Fishery Through 2005, Stock Assessment and Fishery Evaluation: Stock Assessments and Rebuilding Analyses, vol. 1. Pacific Fishery Management Council, 2130 SW Fifth Avenue, Suite 224, Portland, OR 97201.
Myers, R.A., Rosenberg, A.A., Mace, P.M., Barrowman, N., Restrepo, V.R., 1994. In search of thresholds for recruitment overfishing. ICES J. Mar. Sci. 51 (2), 191–205.
NEFMC (New England Fishery Management Council), 2003. Final amendment 13 to the northeast multispecies fishery management plan including a final supplemental environmental impact statement and an initial regulatory flexibility analysis. New England Fishery Management Council, 50 Water Street, Mill 2, Newburyport, MA 01950.
NMFS (National Marine Fisheries Service), 2002. Final report of the working group on re-evaluation of biological reference points for New England groundfish. National Marine Fisheries Service, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02543, USA.
NPFMC (North Pacific Fishery Management Council), 2006. Fishery management plan for groundfish of the Bering Sea and Aleutian Islands management area. North Pacific Fishery Management Council, 605 W 4th Avenue, Suite 306, Anchorage, AK 99501, USA.
PFMC (Pacific Fishery Management Council), 1998. Final environmental assessment/regulatory impact review for amendment 11 to the Pacific coast groundfish fishery management plan. Pacific Fishery Management Council, 2130 SW Fifth Avenue, Suite 224, Portland, OR 97201, USA.
PFMC (Pacific Fishery Management Council), 2002. Scientific and statistical committee report on rebuilding analyses for bocaccio, canary rockfish, yelloweye rockfish, widow rockfish, and whiting. In: Pacific Fishery Management Council: Status of the Pacific Groundfish Fishery Through 2002 Stock Assessment and Fishery Evaluation, vol. I. Pacific Fishery Management Council, 2130 SW Fifth Avenue, Suite 224, Portland, OR 97201, USA.
PFMC (Pacific Fishery Management Council), 2006. Pacific coast groundfish fishery management plan for the California, Oregon and Washington groundfish fishery as amended through Amendment 19. Pacific Fishery Management Council, 7700 NE Ambassador Place, Suite 200, Portland, OR 97220, USA.
Punt, A.E., 2003a. Evaluating the efficacy of managing West Coast groundfish resources through simulations. Fish. Bull. 101, 860–873.
Punt, A.E., 2003b. The performance of a size-structured stock assessment method in the fact of spatial heterogeneity in growth. Fish. Res. 65, 391–409.
Punt, A.E., Smith, A.D.M., Cui, G.R., 2002. Evaluation of management tools for Australia's south east fishery. 2. How well can management quantities be estimated? Mar. Fresh. Res. 53, 631–644.
Quinn, T.J., Deriso, R.B., 1999. Quantitative Fish Dynamics. Oxford University Press, New York.
Restrepo, V.R., Powers, J.E., 1999. Precautionary control rules in US fisheries management: specification and performance. ICES J. Mar. Sci. 56, 846–852.
Restrepo, V.R., Thompson, G.G., Mace, P.M., Gabriel, W.L., Low, L.L., MacCall, A.D., Methot, R.D., Powers, J.E., Taylor, B.L., Wade, P.R., Witzig, J.F., 1998. Technical guidance on the use of precautionary approaches to implementing national standard 1 of the Magnuson-Stevens Fishery Conservation and Management Act. NOAA Tech. Memo NMFS-F/SPO-31.
Sampson, D.B., Lee, Y.W., 1999. An assessment of the stocks of petrale sole off Washington, Oregon and northern California in 1998. In: Status of the Pacific Coast Groundfish Fishery Through 1998 and Recommended Biological Catches for 1994: Stock Assessment and Fishery Evaluation. Pacific Fishery Management Council, 2130 SW Fifth Avenue, Suite 224, Portland, OR 97201.
Sibert, J., Hampton, J., Kleiber, P., Maunder, M., 2006. Biomass, size, and trophic status of top predators in the Pacific Ocean. Science 314, 1773–1776.
Wilberg, M.J., Bence, J.R., 2006. Performance of time-varying catchability estimators in statistical catch-at-age analysis. Can. J. Aquat. Fish. Sci. 63, 2275–2285.
Yin, Y., Sampson, D.B., 2004. Bias and precision of estimates from an age-structured stock assessment program in relation to stock and data characteristics. N. Am. J. Fish. Manage. 24, 865–879.