Fisheries Research 134–136 (2012) 82–94

Contents lists available at SciVerse ScienceDirect

Fisheries Research

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

Identifying research priorities for management under uncertainty: The estimation ability of the stock assessment method used for eastern Bering Sea snow crab (*Chionoecetes opilio*)

Cody Szuwalski*, André E. Punt

School of Aquatic and Fishery Sciences, Box 355020, University of Washington, Seattle, WA 98195-5020, USA

ARTICLE INFO

Article history:
Received 16 February 2012
Received in revised form 3 July 2012
Accepted 2 August 2012

Keywords:
Chionoecetes opilio
Eastern Bering Sea
Estimation ability
Simulation
Stock assessment

ABSTRACT

The fishery for snow crab in the eastern Bering Sea was declared overfished in 1999 and had by 2010 failed to rebuild in the time specified. A key component of the rebuilding plan for snow crab was the stock assessment on which management advice was based. In common with all stock assessments, there are numerous potential sources of uncertainty, but it is not clear which uncertainty is most influential, and hence how research should be focused to best reduce uncertainty. A simulation framework is used to explore the estimation ability of the assessment method when its assumptions are violated. The focus of this evaluation is on how well management quantities such as mature male biomass and the overfishing level can be estimated. Three categories of uncertainty are considered: mis-specification of pre-specified parameters, data quantity and/or quality, and assumptions in the assessment related to fishing mortality. Management quantities are estimated reasonably well when all assumptions of the stock assessment method are correct. Additional data on growth and natural mortality would be most beneficial. However, penalties on fishing mortality introduce bias into estimates, and survey selectivity patterns are poorly determined by the estimation method.

1. Introduction

Stock assessments aim to use the best available science to provide estimates of the biomass of a target species and then calculate acceptable removals based on a chosen harvest control rule. However, the best available science is often incomplete and/or uncertain. It is necessary to prioritize which research should be conducted when there are multiple uncertainties in a system and limited resources to address these uncertainties. Research priorities are often enumerated anecdotally (e.g. Jennings and Zigler, 2000), but cost-benefit analyses that evaluate the expected returns of a specific research project (e.g. a fishery-independent survey) are also found in the literature (e.g. McDonald et al., 1997; Powers and Restrepo, 1993). Simulation can be used to explore the impact of imperfect knowledge regarding the values of population parameters on the ability to estimate quantities used in management when experimental studies to determine the true parameter values are absent, infeasible and/or undesirable (Butterworth and Punt, 1999; Punt, 2003b). In a simulation framework, perfectly known (or 'true') simulated populations are produced and analyzed using a stock assessment. The impact of incorrect assumptions can be quantified by evaluating the degree to which the ability to estimate the management quantities is affected. This approach differs from varying the values of the parameters of a stock assessment in a set of sensitivity tests because most of the parameters of a stock assessment are not specified by the analyst but are rather estimated by fitting a population dynamics model to a data set. Although simulation studies should be conducted for all modeling endeavors, they are often not. An exercise using simulations to identify research priorities when there are multiple sources of uncertainty is presented here using the snow crab (*Chionoecetes opilio*) fishery in the eastern Bering Sea (EBS) as an example.

The U.S. domestic fishery for EBS snow crab is all-male and historically productive, with a maximum estimated biomass of over 680,000 mt in the early 1990s (Turnock and Rugolo, 2009). However, both biomass and catch have been quite variable since the start of the domestic fishery, and EBS snow crab was declared overfished in 1999 when it was assessed to have dropped below its Minimum Stock Size Threshold (MSST) (Turnock and Rugolo, 2011). As required under the U.S. Magnuson-Stevens Fishery Management and Conservation Act, a plan was developed to rebuild the stock to the target biomass (a proxy for B_{MSY}, defined as the biomass corresponding to a fishing mortality of $F_{35\%}$, the fishing mortality rate which reduces spawning biomass-per-recruit to 35% of its unfished level) within 10 years (NPFMC/NMFS, 2000). Stock assessment estimates of spawning biomass (and later mature male biomass, MMB)

* Corresponding author. Tel.: +1 785 979 4453; fax: +1 206 685 7471.
E-mail addresses: szuwalsk@uw.edu, c.s.szuwalski@gmail.com (C. Szuwalski).

0165-7836/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.fishres.2012.08.007

were used to calculate levels of fishing mortality that would return the stock to the target biomass and these levels of fishing mortality were used to set catch quotas. This plan was, however, declared a failure in September 2009 (Turnock and Rugolo, 2009), and the inability to achieve management goals calls into question the efficacy of the management strategy used for this stock and also the stock assessment on which management advice has been based. (The stock was declared rebuilt in 2011, but this was longer than the specified time frame of 10 years.)

The EBS snow crab stock is managed jointly by the North Pacific Fishery Management Council (NPFMC) and the state of Alaska. Each year, data from the directed pot fishery, the National Marine Fisheries Service (NMFS) summer trawl survey and bycatch data from the groundfish trawl fishery are used to estimate trends in fishing mortality, numbers-at-length, and MMB using a length-structured population dynamics model (Turnock and Rugolo, 2009). The estimate of MMB is used in conjunction with a catch control rule (Eqs. (1)–(3)) to determine the overfishing level (OFL) for the coming year (NPFMC/NMFS, 2007). The OFL is used to determine an Acceptable Biological Catch (ABC), which is less than the OFL to account for scientific uncertainty. A total allowable catch (TAC), to be divided among quota holders, is finally determined by the State of Alaska and must be less than or equal to the ABC.

Stock status level

$$\frac{B_{current}}{B_{35\%}} > 1 \qquad F_{\mathrm{OFL}} = F_{35\%} \tag{1}$$

$$\beta < \frac{B_{current}}{B_{35\%}} \leq 1 \quad F_{\mathrm{OFL}} = F_{35\%} \times \frac{(B_{current}/B_{35\%}) - \alpha}{1 - \alpha} \tag{2}$$

$$\frac{B_{current}}{B_{35\%}} \leq \beta \qquad \text{Directed fishery} \quad F = 0 \tag{3}$$

where α determines the fishing mortality used to compute the OFL, F_{OFL}, when biomass decreases to $\beta \times B_{\mathrm{MSY}}$ and β determines the threshold level of biomass at or below which directed fishing is prohibited. $B_{current}$ is the estimated MMB for the current year and $B_{35\%}$ is a proxy for the biomass at which maximum sustainable yield is achieved, B_{MSY}.

The currently used catch control rules cannot be expected to perform adequately without a reasonably accurate and precise estimate for $F_{35\%}$, the time-series for MMB and the B_{MSY} proxy. The current stock assessment method consists of a length-based population dynamics model that accounts for sex, shell condition and maturity state, with an associated method for parameter estimation (Turnock and Rugolo, 2009). The processes modeled (e.g. growth, mortality, selectivity, catchability) in the assessment have considerable associated uncertainty because of a lack of data on snow crab biology and demographics. It is prudent to examine how well one would expect to be able to estimate the parameters used to predict quantities important to management given current levels of uncertainty around the parameters governing the biology, exploitation and monitoring of the EBS snow crab resource. A better understanding of the relative magnitude of impact on estimation performance caused by uncertainty will allow for appropriate allocation of finite management, research and monitoring resources.

Length-based stock assessments are used world-wide to model the population dynamics of stocks for which aging is difficult (e.g. American lobster in Maine (e.g. Chen et al., 2005), rock lobster in New Zealand and Australia (e.g. Hobday and Punt, 2001) and king crab in Alaska (e.g. Zheng and Siddeek, 2010)). In addition to guiding future research directions for snow crab management, examining the impact of uncertainty on the estimation ability of size-based methods may identify general conclusions that could be useful for assessments of other stocks with size-based assessment methods.

Three categories of uncertainty are considered in this analysis: mis-specification of parameters, the amount or quality of data, and assumptions in the assessment related to fishing mortality. Questions related to mis-specification are: (1) what is the impact of uncertainty in growth on the ability of the EBS snow crab assessment to estimate MMB and other management-related quantities, (2) how does the assessment method perform when catchability and selectivity for the NMFS survey differ from the currently estimated values, and (3) how does assuming natural mortality at an incorrect value affect estimation ability? Changes in assumptions about quality and quantity of data will be used to answer the questions: (1) what are the benefits of additional growth data, and (2) how does the quality of survey data affect estimation ability? The final question (which addresses how the assessment is specified) examines how well assessment method is able to identify an overfished stock.

These questions (see below for further detail and rationale for their selection) have been selected because they are either contentious topics of discussion at management meetings between the fishing industry and assessment scientists (e.g. survey selectivity and natural mortality), there are few data available on the process (e.g. growth and natural mortality), the process has a large bearing on the ability of the management system to reach its goals (e.g. assumptions about fishing mortality), or because the process is one controllable by management (e.g. data quality in the survey). There are, of course, other questions (such as the values assigned to the factors used to weight different data sources) which could be addressed; those considered in the paper were selected because there is no straightforward method to explore them within the assessment process directly (e.g. sensitivity of model outputs to values of weighting factors). Answering these questions will aid in the process of identifying research priorities, be they experimental studies to more accurately determine parameter values, increased survey effort, or revision of the stock assessment model.

2. Methods

2.1. Overview

Briefly, the steps used to conduct the simulation study to identify research priorities here are: (1) identify processes with uncertainty that may influence estimation ability, (2) identify possible values for the quantities or parameters involved from the available literature, (3) develop a simulation framework, and (4) rank processes according to relative impact on the ability to estimate a quantity useful in management (e.g. the OFL). The word "process" is used loosely in this context. "Process" can mean not only the biological processes modeled in the assessment, but also the activities related to management (e.g. data collection or designation of allowable fishing mortality).

The processes to be examined in this study are those involved in the stated questions of interest. The Monte Carlo simulation framework used to address these questions is commonly used in the literature (e.g. Punt et al., 2002; Chen and Wilson, 2002; Kanaiwa et al., 2005; see Fig. 1 for an overview) and is recommended over sensitivity testing (Hilborn and Walters, 1992; NRC, 1998). This simulation framework has two main components: an operating model and an estimation model. The operating model provides the 'truth' for a given simulation (which can change depending on the question being addressed) and is based on the population dynamics model that currently forms the basis for the stock assessment (see Section A of Supplementary Material for details). Virtual populations are generated using the operating model, and the data necessary for assessment purposes are 'collected' from these populations with observation error. The data used for assessment purposes are numbers and length frequencies of crab caught in the directed pot fishery, the trawl fishery (as bycatch) and the NMFS summer survey.

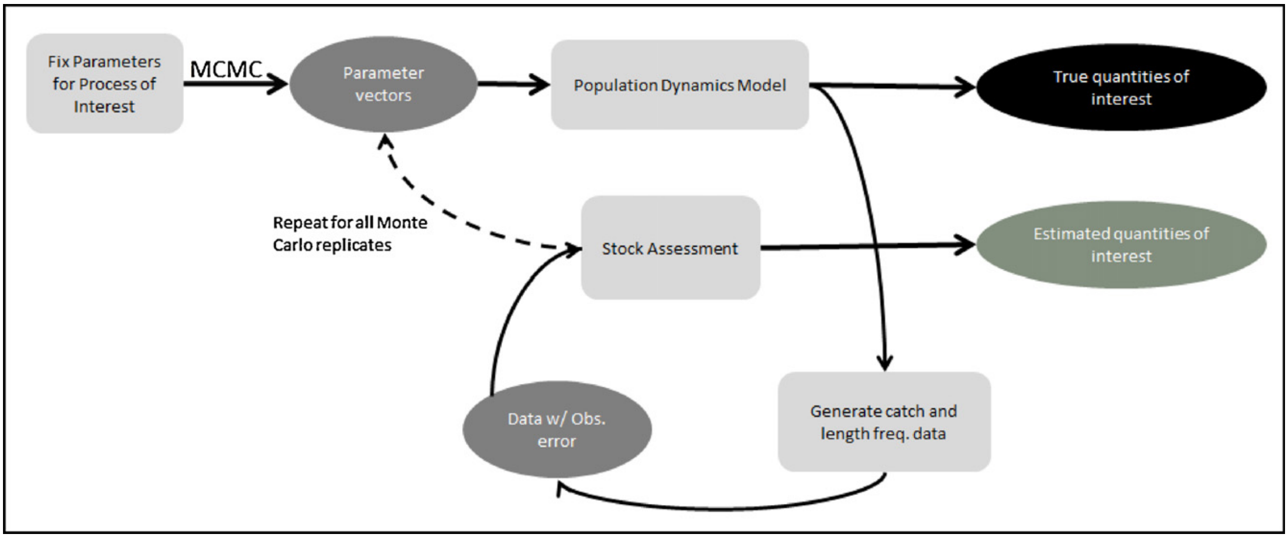

Fig. 1. Flow of information in assessing the estimation ability of a stock assessment method.

The estimation model produces parameter estimates using the data 'collected' from the operating model, and represents a manager's perception of the status and productivity of the stock (i.e. produces estimates of mature male biomass and other management related quantities on which management advice is provided). In this study, the estimation model (see Appendix B of Supplementary Material for the objective function used for parameter estimation) is very similar to the actual assessment method used to provide management advice for the EBS snow crab stock. The performance of the estimation model is evaluated in terms of the relative errors (REs) (Eq. (4)) of the estimates of key derived quantities (i.e. mature male biomass (MMB), the fishing mortality that will be used to compute the OFL (F_{OFL}), the fishing mortality that will result in a MMB that is 35% of virgin MMB ($F_{35\%}$), the MMB corresponding to the proxy for maximum sustainable yield (B_{MSY}) (i.e. the product of the MMB-per-recruit corresponding to $F_{35\%}$ multiplied by average recruitment from 1978 to 2009), and the overfishing level for the coming fishing year (OFL)). The full set of quantities (rather than just MMB or the OFL) are included to identify how error propagates through the estimation method. Moreover, these are all quantities which are reported to the decision makers.

$$E_t^{i,j} = \frac{\widehat{Q}_t^{i,j} - Q_t^{i,j}}{Q_t^{i,j}} \qquad (4)$$

where $E_t^{i,j}$ is the relative error for quantity i during year t for simulation j, $Q_t^{i,j}$ is the true (i.e. based on the operating model) value for quantity i during year t for simulation j, and $\widehat{Q}_t^{i,j}$ is the estimate of quantity i during year t for simulation j. The ratio of MMB to B_{MSY} is calculated for each scenario to explore how uncertainty in the processes of interest affects the ability to estimate relative versus absolute biomass. Relative errors and inter-simulation intervals for a given management quantity are presented in a common format, i.e. "median RE% [5% bound, 95% bound]"; in some instances, the lower and upper bounds are not presented. Abbreviations used for the various scenarios are listed in Table 1.

A different parameterization of the operating model is specified to address each of the questions. For each question, a population process is identified, and the parameters associated with that process are pre-specified. The parameters are pre-specified so that any estimation errors are the consequences of incorrect assumptions related to the process and not variation in the process. One hundred vectors of the remaining estimable parameters are then sampled from the posterior distribution obtained by fitting the operating model to the actual data for EBS snow crab using a Markov chain Monte Carlo (MCMC) algorithm given those pre-specified values. This involved conducting 1,000,000 cycles of the MCMC algorithm, implementing a 10% burn-in and selecting a thinning ratio so that 100 vectors are selected within the chain. Generating parameter vectors in this manner allows a parameter space that is consistent with both the actual fishery data and the question of interest to be represented (similar to Kanaiwa et al., 2005). Several diagnostic statistics (e.g. checking for lack of autocorrelation and calculating Geweke statistics) were used to check for evidence of convergence of the MCMC algorithm.

Each parameter vector is then used to produce a virtual population from which the data necessary for assessment purposes are generated. The estimation model differs from the operating model only in the assumptions about the process under examination and is then applied to the 'collected' data from each Monte Carlo replicate. This process is repeated for each parameter vector;

Table 1
Median absolute relative errors (MAREs) and mean absolute relative errors (MeAREs) for the OFL in descending order for each scenario considered (left two columns) with average relative errors (RE) and 90% inter-simulation intervals for RE. Values for all MAREs and MeAREs, except those for the reference case, are reported as the difference between the MARE or MeARE for a given scenario and that for the reference case (i.e. if there was no change in MARE for a given scenario, the value in the MARE column would be zero for that scenario).

Scenario	MARE	MeARE	RE	5%	95%
Reference	22%	23%	1%	−38%	52%
	(± reference)			(actual values)	
Mis-specification of parameters					
M − 50%	23%	19%	−45%	−68%	−10%
M + 50%	7%	14%	29%	−14%	91%
Growth-14	4%	12%	18%	−31%	93%
Sel era 3	3%	3%	−16%	−54%	36%
Sel all era	2%	3%	−17%	−52%	32%
Data quality/quantity					
CV EffN+	4%	9%	−1%	−46%	83%
CV+	2%	2%	1%	−40%	58%
CV EffN−	0%	3%	−3%	−45%	61%
CV−	−2%	−1%	−1%	−33%	49%
Growth 25	2%	6%	10%	−34%	81%
Growth 100	1%	3%	7%	−32%	59%
Assessment assumptions					
$F \times 2.0$	7%	11%	12%	−40%	83%
$F \times 0.5$	−2%	−1%	−1%	−33%	48%

the resulting 100 estimated series of MMB and the other quantities of management importance are compared to the true series from the operating model, and measures of bias and precision are obtained.

The following sections outline how the operating model is specified to address each question. The operating models are all variants of a reference operating model which matches the specifications of the 2009 assessment.

2.2. Mis-specification of parameters

2.2.1. Survey selectivity

Tows conducted by the Bering Sea Fisheries Research Foundation (BSFRF) suggest that survey selectivity increases slowly and smoothly with size and asymptotes to values less than one (Somerton et al., 2010) and that survey catchability, q (the probability of capture for the most selected crab) is much lower than currently estimated. In contrast, the currently estimated survey selectivity pattern is almost knife-edged with catchability close to 1. The ability to estimate survey selectivity within the stock assessment is examined by conducting simulations for scenarios in which survey selectivity for the most-recent group of years over which survey selectivity is assumed constant[1] (1989 to present) is set to the curve estimated by Somerton et al. (2010) and in which selectivity in all eras is set to Somerton's curve. There are three "eras" (referred to as "era 1", "era 2" and "era 3") which correspond to 1978–1981, 1982–1988 and 1989 to present, respectively. In both scenarios, the estimation method estimates all selectivity parameters for all eras.

2.2.2. Natural mortality

Natural mortality (M) for mature males is specified in the assessment as 0.23 year^{-1} (Turnock and Rugolo, 2009) based on the approach of Hoenig (1983) which extrapolates natural mortality from maximum observed age. Although M for snow crab has been the subject of several studies (Somerton, 1981; Zheng, 2003), all of them have provided markedly different estimates (0.13–1.07 year^{-1}), although the range of M estimates is considerably smaller when the assumption of a terminal molt is taken into account. The impact of incorrectly specifying natural mortality is assessed by fixing the true (i.e. operating model) value for natural mortality at 0.23 year^{-1} and then applying the estimation model when M is assumed to be ±50% of the true value. These scenarios will be referred to as "M + 50%" and "M − 50%" hereafter.

2.3. Amount of data quantity or quality

The amount and quality of data available to the stock assessment method has the potential to impact estimation ability. Each of the data sources for the assessment have associated measures of uncertainty. A change in the assumptions about the quantity or quality of data for any of these sources would change the associated uncertainty, and this would impact the uncertainty around model estimates. In particular, data associated with the survey (e.g. coefficients of variation around area-swept estimates of biomass and length frequencies) and growth (i.e. the number of observations used to develop the growth curve) appear to have the highest associated uncertainty. These two processes are the focus of the simulations directed at determining the impact of the quality of data on estimation ability.

[1] Such a group of years will henceforth be referred to as an 'era'. Eras are selected given changes to the gear used by the survey.

2.3.1. Survey data

The coefficients of variation, CV, around the survey data were increased and decreased by 0.08 in turn when generating data for assessment purposes to explore the relationship between model estimates and the precision of the survey data while the effective sample size used for the survey length data was held constant (abbreviated "CV+" and "CV−", respectively). The fixed amount 0.08 was chosen because the smallest CV around the survey estimates is 0.09. Analyses were also conducted in which the effective sample sizes used to determine survey length frequencies were halved and doubled (from 200 to 100 and 400, and abbreviated "CV EffN−" and "CV EffN+", respectively).

2.3.2. Growth

One of the key components of a length-based model is the size-transition matrix that describes how crab grow each year after molting as a function of its current size. Until the 2010 assessment, growth was estimated outside of the assessment method and growth parameters pre-specified in the assessment (the assessment authors are currently experimenting with different ways of estimating growth). The size-transition matrix used in the 2009 assessment was based on only 14 observations of pre- and post-molt lengths taken from crabs collected during a NMFS survey (Rugulo, NMFS, unpublished) and the 2010 assessment used inferences from these 14 observations as the basis for a prior on growth rates. Error in the size-transition matrix has been demonstrated to lead to poor estimates of management-related quantities when those quantities are estimated using a length-based method of stock assessment (Punt, 2003a).

The impact of using only 14 data points to determine the size-transition matrix is assessed by randomly generating 14 pre- and post-molt lengths using Eqs. (A.6)–(A.8) for each Monte Carlo replicate. These 14 data points are representative of the entire population (i.e. not only large males are sampled). The growth observations provide the data used to estimate the parameters of the size-transition matrix (intercept and slope of the growth equation and shape and scale parameter of the gamma distribution that describes the variance in growth increments; Eqs. (A7a) and (A7b) of Supplementary Material). The estimation method underlying each Monte Carlo replicate therefore has a unique size-transition matrix. Essentially, a virtual mark-recapture study is performed for each Monte Carlo replicate to define the growth curve to be used in the estimation model. Scenarios in which more length observations (25 and 100) are used to define the size-transition matrix are used to assess the utility of additional growth data.

2.4. Fishing mortality

It is important that an assessment method is able to detect when a stock is overfished given the need to achieve management goals. To evaluate the estimation model's ability to detect changes in fishing mortality, average fishing mortalities (catch, discard and trawl) are fixed at the values estimated from the current assessment method and vectors for the remaining estimable parameters are drawn from the posterior of the resulting model using MCMC. Average fishing mortality is then multiplied by a constant (i.e. 0.5, 2) from 1995 onward, data for assessment purposes are generated, and the estimation model is applied to the generated data. The method's ability to detect an overfished stock is summarized by the fraction of times the stock is correctly assessed to be above and/or below the true MSST (0.5 $B_{35\%}$). The stock assessment utilizes penalties on deviations around average fishing mortality in the objective function to constrain the inter-annual variation and overall magnitude of estimated fishing mortality. Analyses were conducted in which the penalties were eliminated, quartered, halved and

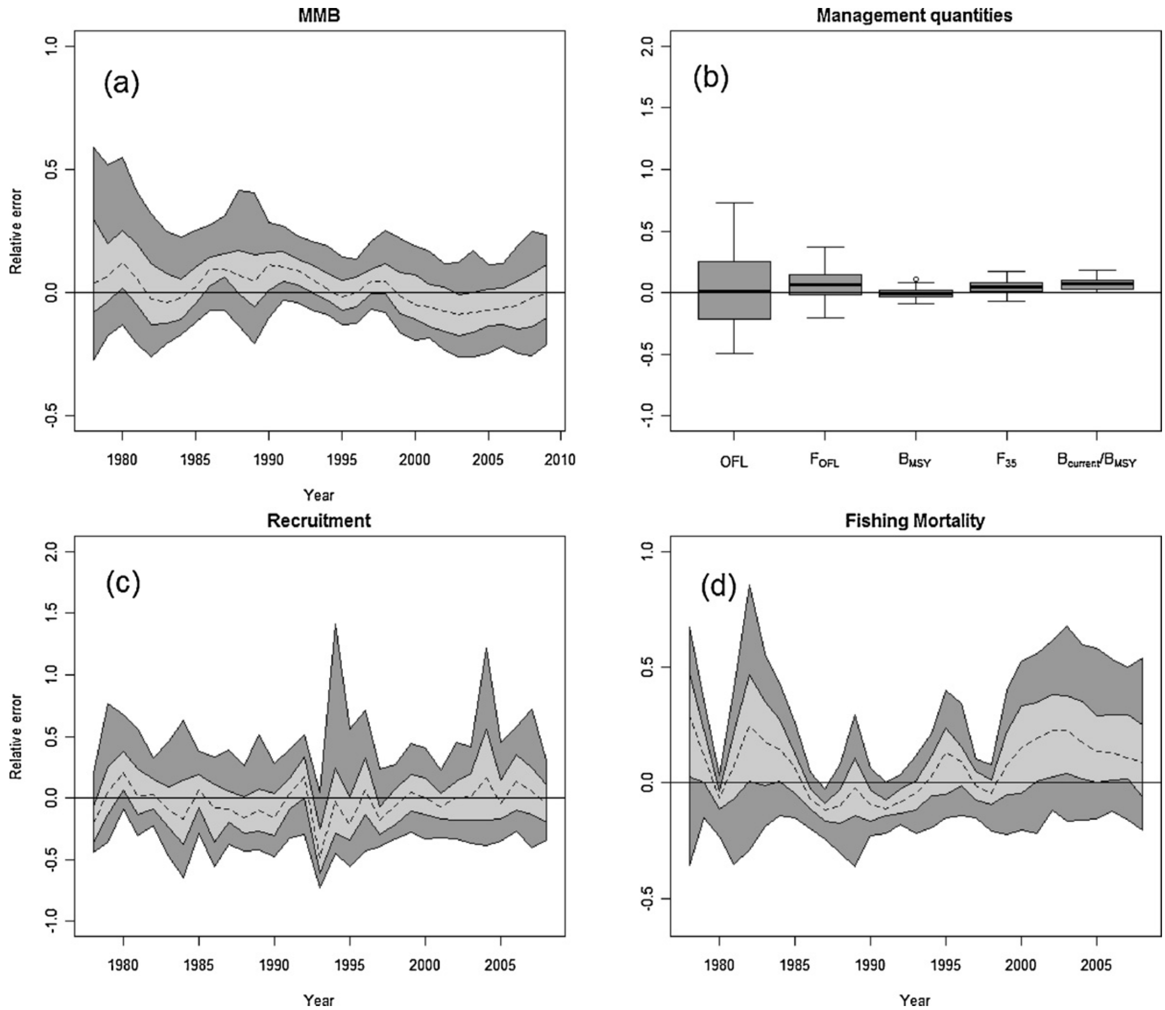

Fig. 2. Relative error distributions for mature male biomass, recruitment and fishing mortality in the directed pot fishery for the reference scenario (median: dashed line; 5th and 95th quantiles: dark gray; 25th and 75th quantiles light gray) (a, c, and d). Boxplots of relative error for key management quantities (b).

doubled to assess their impact on parameter estimation and the ability of the assessment method to identify an overfished stock.

3. Results

3.1. Reference scenario

Results are first shown for the 'reference' scenario in which the assumptions of the stock assessment are not violated. This scenario provides an impression of how well this method can be expected to perform in principle. For this scenario, *M* and the size-transition matrix are assumed to be known exactly, and data are generated based on sampling standard deviations and effective sample sizes which mimic those assumed when actually applying the stock assessment.

Mature male biomass (MMB) was relatively well estimated with an average median RE of 1% [−16%, 26%] over all years (Fig. 2a). Variability in estimated MMB translates to a range relative errors for the OFL (1% [−38%, 52%]), F_{OFL} (6% [−16%, 26%]), and B_{MSY} (0% [−6%, 6%]) (Fig. 2b). Recruitment was slightly negatively biased on average and relatively uncertain (−4% [−39%, 49%], Fig. 2c). On average, median estimated fishing mortality was positively biased by 7% [−20%, 35%] (Fig. 2d).

Estimated selectivity parameters describing the length at which 95% (Sel 95) and 50% (Sel 50) of the fish were selected by the survey gear in the first era (1978–81) were estimated poorest of all eras (median relative errors of −10% for Sel 95 and 9% for Sel 50). Sel 95 in the second era (1982–88) was slightly biased, but Sel 50 and survey catchability were well estimated (−4% for Sel 95, 0% for Sel 50 and 0% for *q* (catchability)). Selectivity for era 3 (1989 to present) was well estimated, with a median RE of 0 for all quantities and small inter-simulation intervals (Fig. 3a).

3.2. Mis-specification of parameters

3.2.1. Natural mortality

Simulations in which natural mortality, *M*, differs from the true value by ±50% had the largest increases in bias and loss of precision in management quantities of all scenarios considered (Table 1). When *M* is negatively biased, all management quantities are negatively biased, except MMB and B_{MSY}, which are positively biased (Fig. 4a–c). The opposite trend is apparent when *M* is positively biased (Fig. 4d–f).

3.2.2. Survey selectivity

Fixing survey selectivity in the operating model at Somerton's estimated curve for 1989 to present (referred to hereafter as the "Sel era 3" scenario) resulted in large biases in management quantities; MMB was negatively biased (−14%) (Fig. 5a), as was the OFL (−16%) [−54%, 36%] (Fig. 5b) and recruitment was noisier to compensate (Fig. 5c). Fixing survey selectivity in the operating model for all eras at Somerton's curve (referred to hereafter as the "Sel all era"

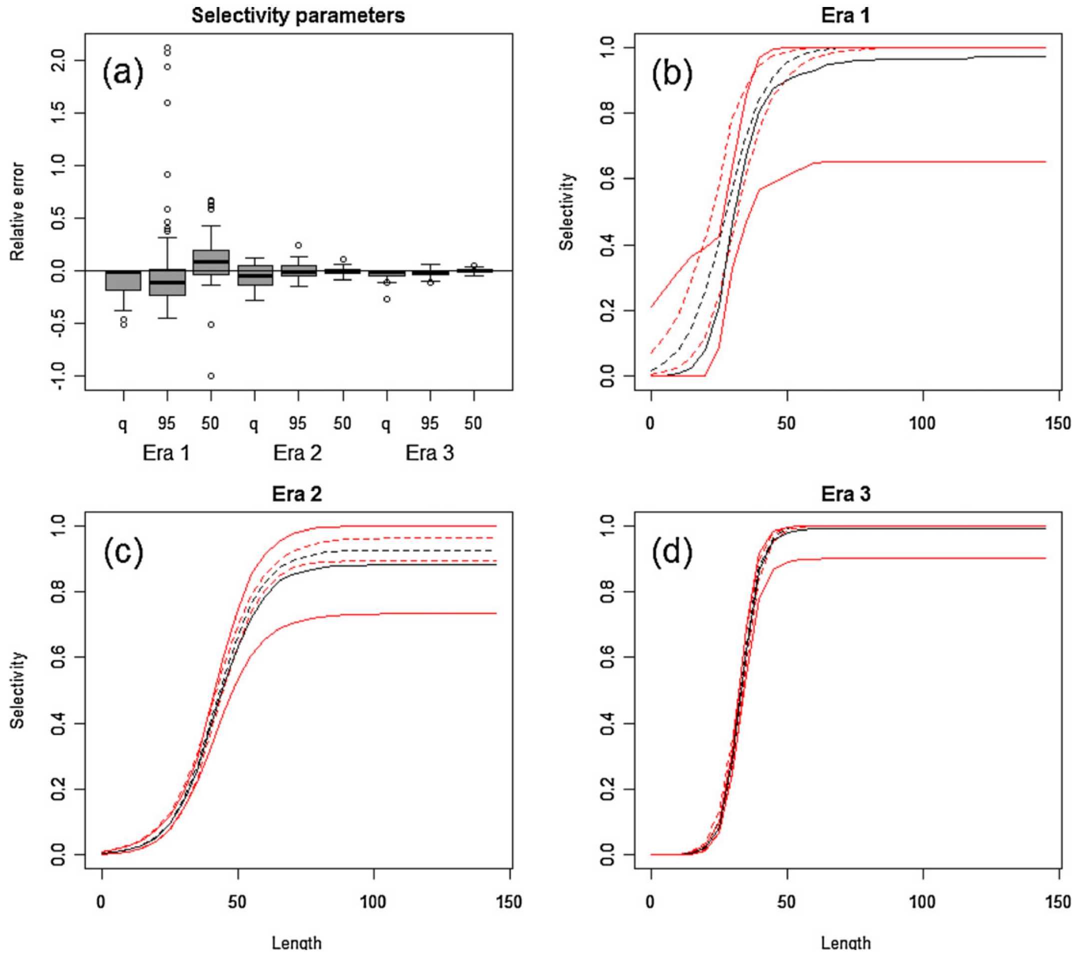

Fig. 3. Relative error distributions for the parameters determining survey selectivity (a). True and estimated survey selectivity curves for eras 1–3 (b–d). Dashed lines are the true curves and solid lines are the estimated curves, with the outer two dashed and solid lines respectively representing the 90% inter-simulation intervals for the true and estimated selectivity curves. The results in this figure are for the reference scenario.

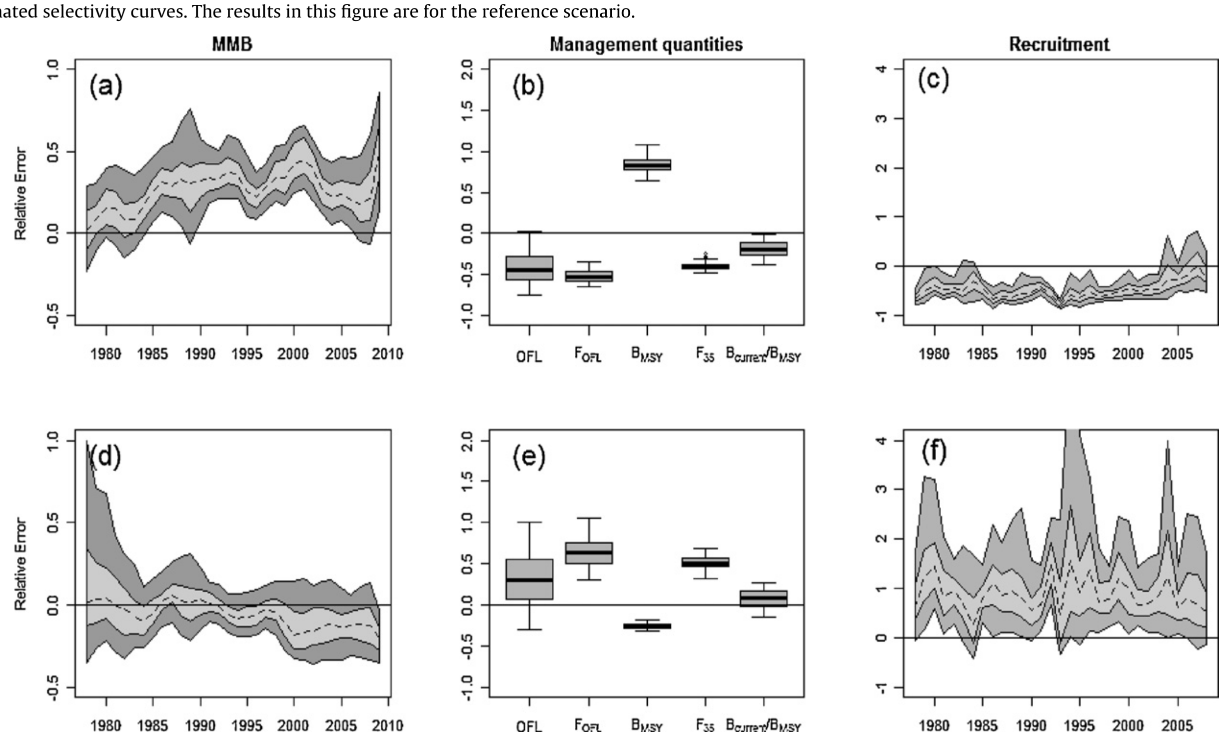

Fig. 4. Relative error in MMB (a and d) and recruitment (c and f) with 90% inter-simulation intervals in dark gray and 50% inter-simulation intervals in light gray, and management quantities (b and e). Panels a–c correspond to a scenario in which natural mortality is assumed to be 50% less than the truth in the estimation model and panels d–f to a scenarios in which natural mortality is assumed to be 50% higher.

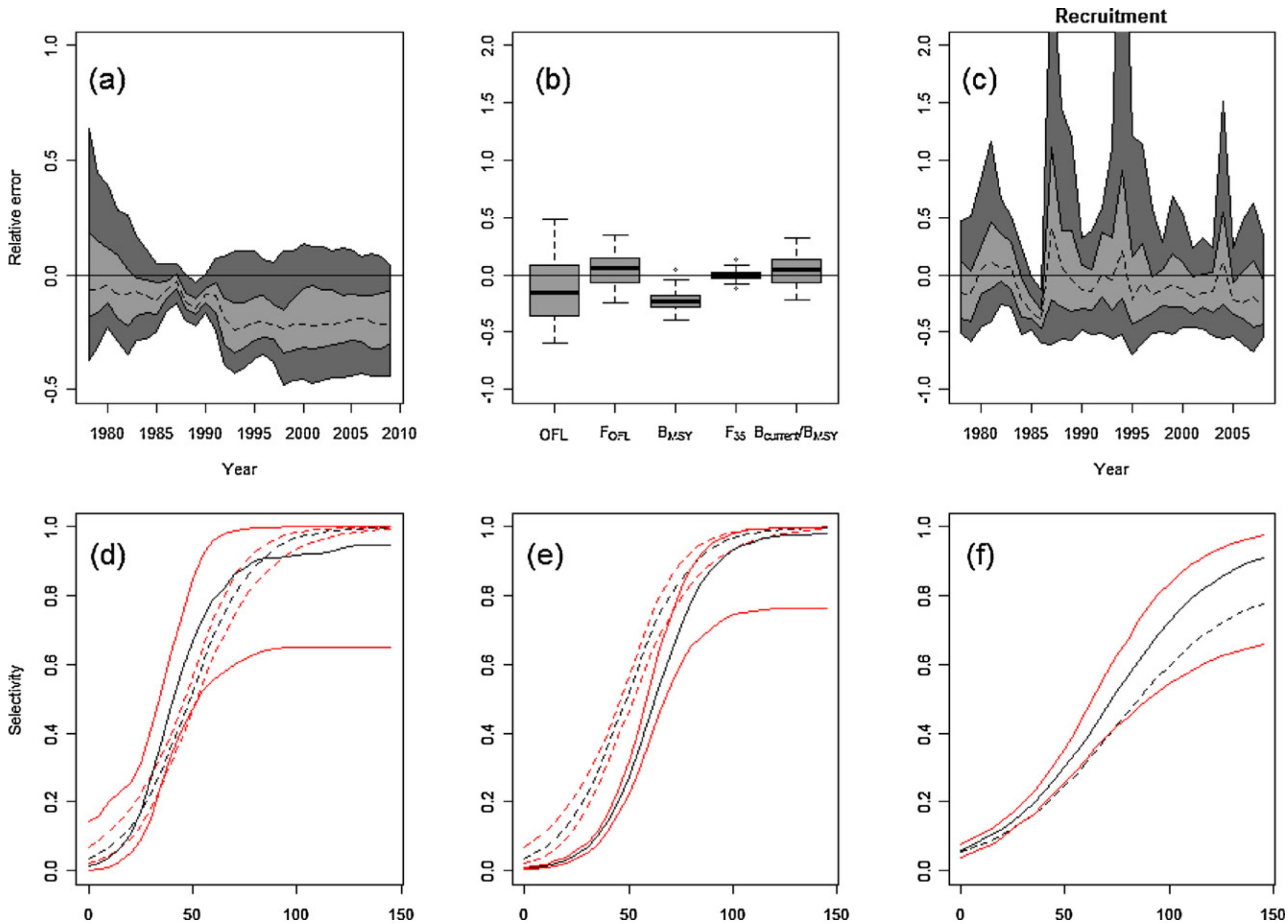

Fig. 5. Relative error in MMB and recruitment with 90% inter-simulation intervals in dark gray and 50% inter-simulation intervals in light gray (a and c) and management quantities (b). Panels d–f are selectivity curves in era 1–3, respectively. Dashed lines are true survey selectivity curves from the operating model and solid lines are estimated selectivity curves. Outer lines for each type (dashed/solid) are 90% inter-simulation intervals. The results in this figure are for the scenario in which survey selectivity is fixed at Somerton's curve for era 3 (1989 to present).

scenario) led to a slightly larger impact on bias in MMB (−19%) and the OFL (−17%) (results not shown because of similarity to the "Sel era 3" scenario).

The slope of the survey selectivity curve for era 1 was estimated reasonably well for the "Sel era 3" scenario (Fig. 5d). The estimated survey selectivity curve for era 2 was beneath the true curve (Fig. 5e), and the median estimated selectivity curve for era 3 was above the true curve (Fig. 5f) for this scenario. The median estimated curve for eras 1, 2 and 3 are respectively always beneath the true curve at larger sizes, reasonably close to the true curve and above the true curve for the "Sel all era" scenario (results not shown).

3.3. Amount of data quantity or quality

3.3.1. Growth

Estimating growth for each Monte Carlo replicate based on 14 observations of pre- and post-molt lengths (instead of assuming perfect knowledge of the growth curve as in the reference scenario) led, as expected, to estimates of MMB with a wider range of relative errors and a positive bias on average (e.g. 8% [−23%, 106%]) ("Growth-14"; Fig. 6a). Increasing the number of observations used to estimate growth to 25 ("Growth-25") and 100 ("Growth-100") observations resulted in narrower relative error distributions and a decrease in bias in MMB (25 observations: 7% [−23%, 56%], 100 observations: 3% [−19%, 35%]) (Fig. 6a). The trend of decreasing ranges of relative error for MMB was mirrored in the estimates for the parameters for the growth equation, OFL, B_{MSY} and F_{OFL}. The estimates of B_{MSY}, OFL and F_{OFL} were positively biased when there was uncertainty associated with growth (3%, 18% and 10% respectively for the 14-tag scenario; Fig. 6b–d), but these biases reduced with added data points for growth.

3.3.2. Survey data

The biases of the estimates of the management quantities are not sensitive to the CVs for the survey biomass estimates, although inter-simulation variation was slightly less for lower CVs. However, halving the effective sample sizes for the survey length-composition data when also increasing the CVs for survey estimates ("CV EffN−") had a negative impact on estimation ability (Fig. 7a and b). The inter-simulation intervals of all management quantities were smaller when the effective sample sizes were doubled and the CVs decreased ("CV EffN+") than when the effective sample sizes were halved and the CVs increased. However, estimates for the "CV EffN−" scenario had a larger bias and loss of precision than the estimates from the scenario in which only the CV was increased (Fig. 7a and b). Simply put, better survey data leads to better estimates of MMB and management quantities, with length frequency data being more important than the survey CV.

3.4. Assessment assumptions – fishing mortality

Operating models were constructed by multiplying average fishing mortality (F_{mort}) by 0.5 ("$F \times 0.5$" in Table 1), 1 and 2.0 ("$F \times 2.0$" in Table 1) from 1995 onwards. The estimation model was

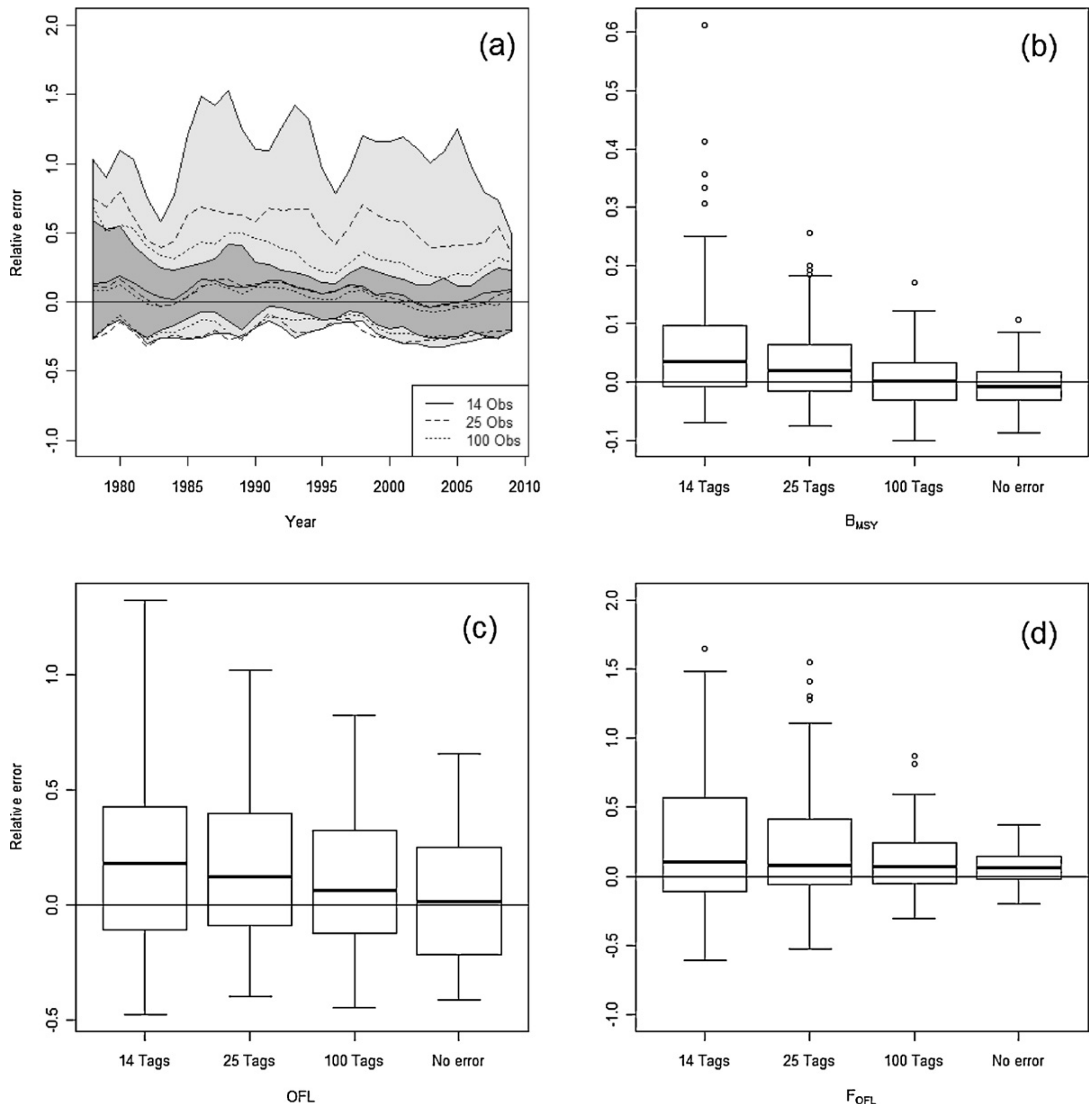

Fig. 6. Median relative error in MMB for scenarios with different amounts of data used to estimate the parameters of the growth equation (a). The dark gray area is the 5th/95th quantiles for the reference scenario. The light gray area is the 5th/95th quantiles for the 14 observation scenario. Outermost dashed upper and lower lines are the 95th/5th quantiles for the 25 and 100 tag scenarios. The solid, dashed and dotted lines in the center of the dark gray area are the medians for the 14, 25 and 100 observation scenarios. Relative error distributions for the management quantities (B_{MSY} (b), OFL (c), F_{OFL} (d)) for increasing amounts of growth information.

then applied to the resulting data sets under the assumption that average fishing mortality had not changed. A decrease in absolute bias in MMB was seen with a decrease in the multiplier for average F_{mort} from 1995 ($0.5 \times F_{mort}$: 0%[−17%, 27%], $1 \times F_{mort}$: 1%[−17%, 26%], $2 \times F_{mort}$: 5%[−13%, 28%]). Larger penalties applied to the fishing mortality deviations resulted in larger biases and narrower ranges of relative errors around MMB, while less severe penalties led to lower biases and less precision, with the scenario eliminating penalties altogether leading to the least bias in the management quantities (Figs. 8 and 9).

Biases in management quantities translated to average false positive rates (cases in which the estimation model reports the population as overfished ($B_{current} < 0.5B_{MSY}$), but the true population is not overfished) of 0.3%, 5.5%, and 1.2%, and average false negative rates (incorrectly identified overfished stocks as 'healthy') of 0%, 0.75%, and 5.4% for the 0.5 F_{mort}, 1 F_{mort}, and 2 F_{mort} cases, respectively (Fig. 10a–c). In the $1 \times F_{mort}$ scenario (which is very similar to the reference case), more severe penalties result in lower false negative rates, but a large increase in the number of false positives (Fig. 10b and e). The incidence of a false negative in the first year in which the stock is actually overfished in the $2 \times F_{mort}$ scenario dropped from nearly 100% under the current penalties (Fig. 10c) to approximately 25% when no penalties were applied (Fig. 10f).

4. Discussion

The suite of estimated management quantities and parameters presented are useful for diagnosing the causes of biases and imprecision, but the overfishing level (OFL) is ultimately the quantity that is used in management decisions. An idea of the assumptions in the stock assessment that, when violated, have the largest potential impacts on the management of the fishery, and should hence be the focus for additional research, can consequently be formed through comparison of the relative errors in the OFL among scenarios. The ratio of the current biomass to B_{MSY} is also important because it defines whether a stock is considered to be overfished according to the US Magnusson-Stevens Act. B/B_{MSY} is used when calculating the OFL, so uncertainty in B/B_{MSY} is also reflected in

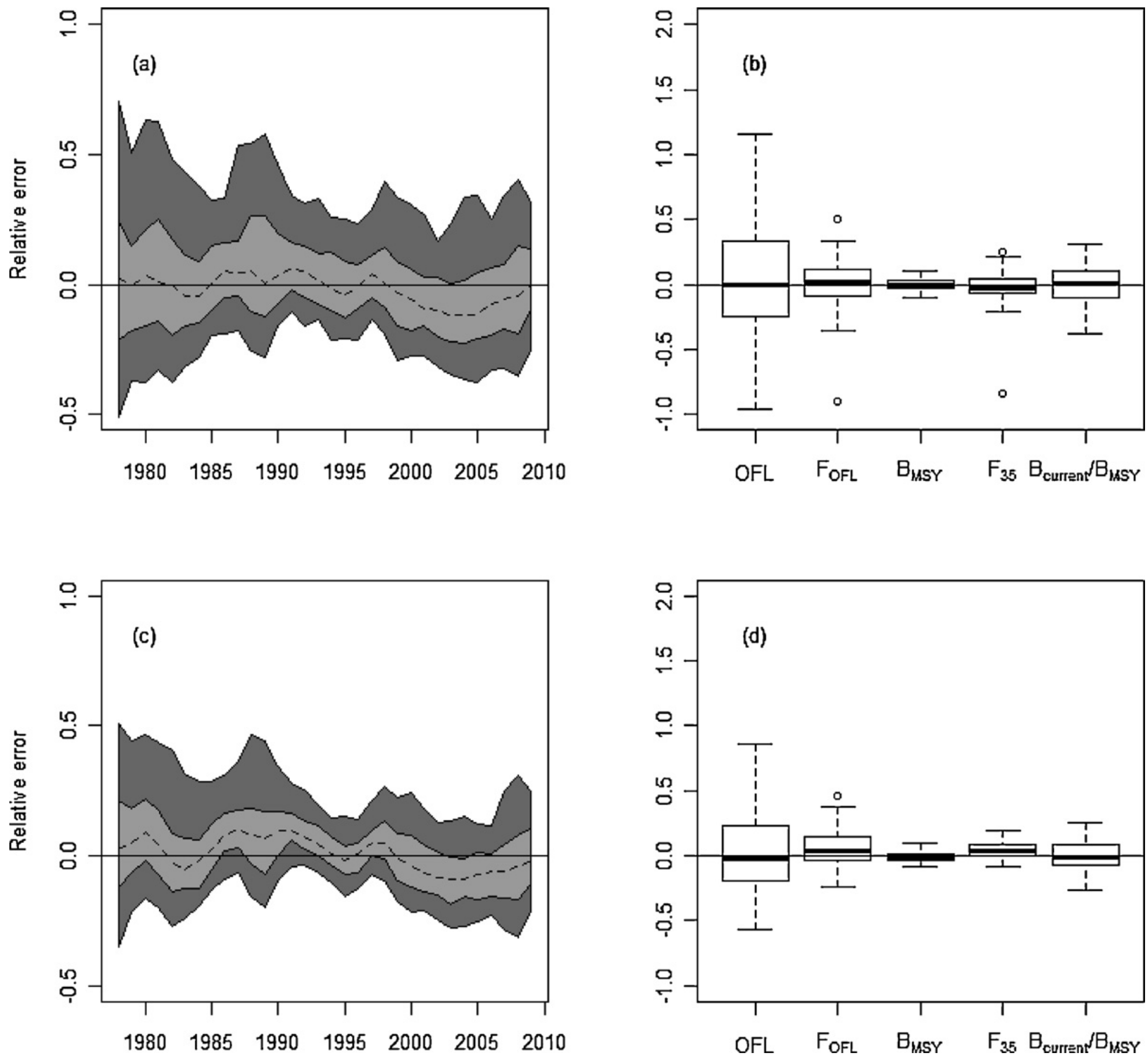

Fig. 7. Relative error distributions for mature male biomass (a and c) and management quantities (b and d) for scenarios in which survey coefficients of variation were increased and effective sample sizes (EffN) were decreased (top row; "CV EffN−") and survey CVs were decreased by 0.08 and EffN was doubled (bottom row; "CV EffN+") when generating data for assessment from the operating model.

the OFL. Smaller biases and losses of precision and less difference among scenarios appeared for B/B_{MSY} than for the OFL. Hence, in spite of the importance of B/B_{MSY} for management, discussion will focus on the OFL.

The OFL is reasonably close to unbiased (median RE = 1%) when all of the assumptions are correct (Table 1). The OFL is estimated relatively imprecisely, but this is expected because the OFL is a function of several estimated quantities (F_{OFL}, current biomass, B/B_{MSY}). Processes ranked in order of decreasing median absolute RE (MARE) for the OFL when the associated assumptions were violated as described in this study are listed in Table 1. Rankings may have been different had the assumptions been violated differently, but Table 1 still provides a way to synthesize a large number of effects. The results from each category of question asked will now be considered briefly with a view toward identifying possible research priorities.

4.1. Parameter estimation

Uncertainty in natural mortality, M, led to the largest changes in the MARE for the OFL. The OFL is positively biased when the assumed value for M exceeds to the true value ("M + 50%", Table 1) even though this results in a negatively biased estimates of MMB. This result arises because of the interplay and relative magnitude of bias in recruitment, current biomass, B_{MSY}, and $F_{35\%}$. When natural mortality is assumed to be higher than it really is, spawning biomass-per-recruit is negatively biased. This negative bias translates to a positive bias in $F_{35\%}$, which, in this case, was greater than the negative bias in MMB. Attempts to better incorporate this uncertainty into the assessment process and further studies to reduce uncertainty in M should be undertaken given the practice of specifying M rather than estimating it.

Incorporating the uncertainty associated with basing the size-transition matrix on a small sample resulted in a 4% increase to the MARE ("Growth-14"; Table 1). With only 100 samples ("Growth-100"; Table 1), the median RE for the OFL decreased from 18% to 7% (and the change in MARE from the reference scenario decreased to 1%). The precision of the OFL also began to approach that for the reference scenario. Considering the potential improvements in bias and precision, more accurately specifying the growth curve using an increased number of pre- and post-molt observations may be an efficient way of increasing confidence in the estimates produced by the stock assessment. Simulation could be useful in the experimental design for additional samples. Collection of data to better determine the growth parameters for snow crab began in 2011 in the EBS by the BSFRF.

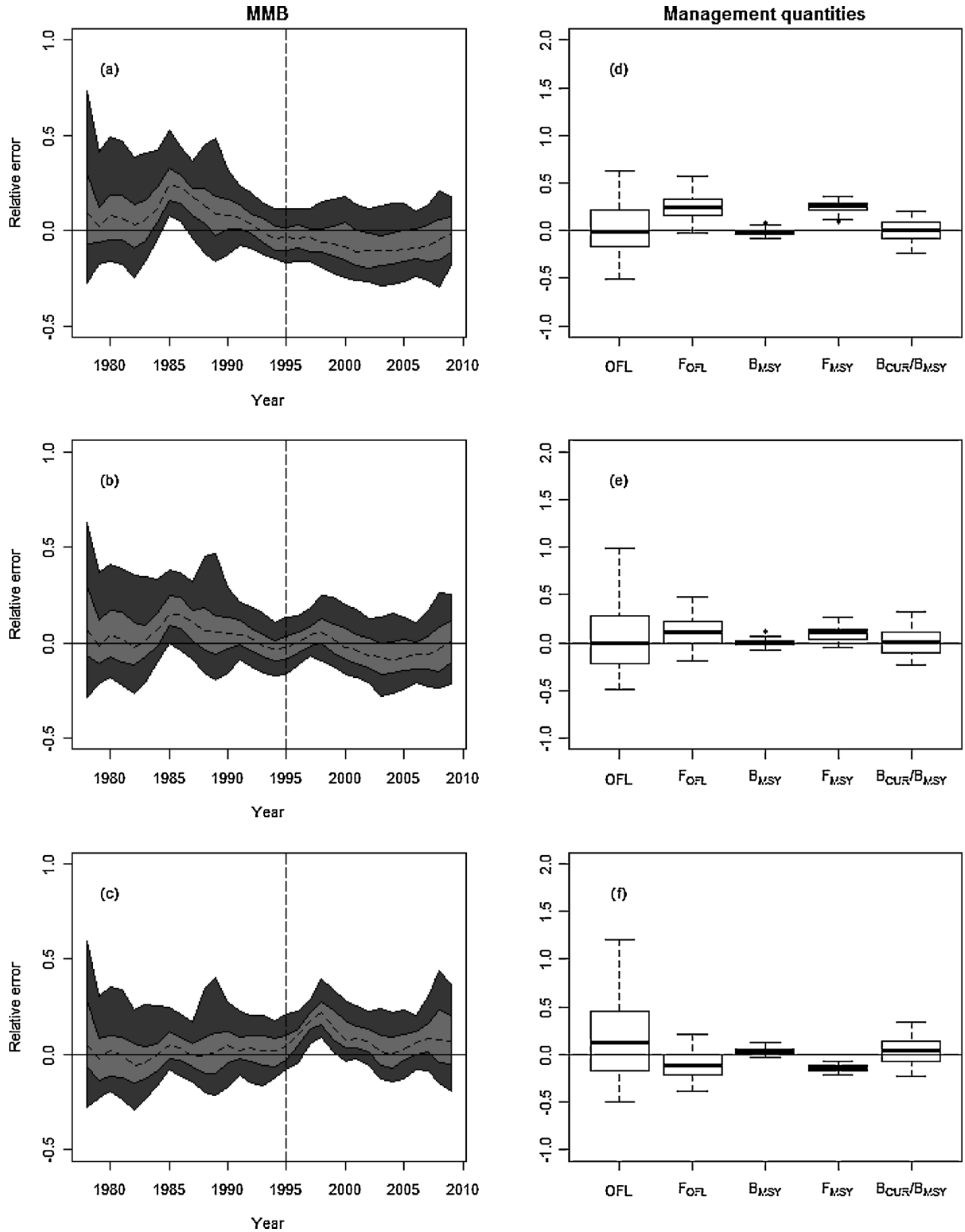

Fig. 8. Relative error distributions for mature male biomass (MMB) and management quantities under scenarios in which average F_{mort} was multiplied by 0.5 (a and d), 1 (b and e), and 2 (c and f) starting in 1995. Quantities are estimated under the original fishing mortality deviation penalties.

The parameter estimation scenario in which selectivity was fixed at Somerton's curve during 1989-present ("Sel era 3" in Table 1) led to large differences in the MARE of the OFL compared to the reference scenario. The observed differences between the estimated and true selectivity curves can explain the difference in MARE. Selectivity in the final era is overestimated, resulting in negatively biased estimates of MMB, which in turn result in negatively biased OFLs. A similar pattern is seen in the scenario in which the survey selectivity curves for all eras are set to Somerton's curve ("Sel All Era"). Selectivity for the first era is underestimated, that for the second era is relatively well estimated and that for the third era is overestimated. Although a suitable explanation for this pattern is not immediately forthcoming, the fact that the assessment method is unable to recognize shifts in survey selectivity emphasizes the value of estimating survey selectivity using field data (to the extent that this possible). Starting in 2011, additional BSFRF data collected from side-by-side towing experiments with the NMFS survey using a nephrops trawl (which is assumed to have a survey catchability of

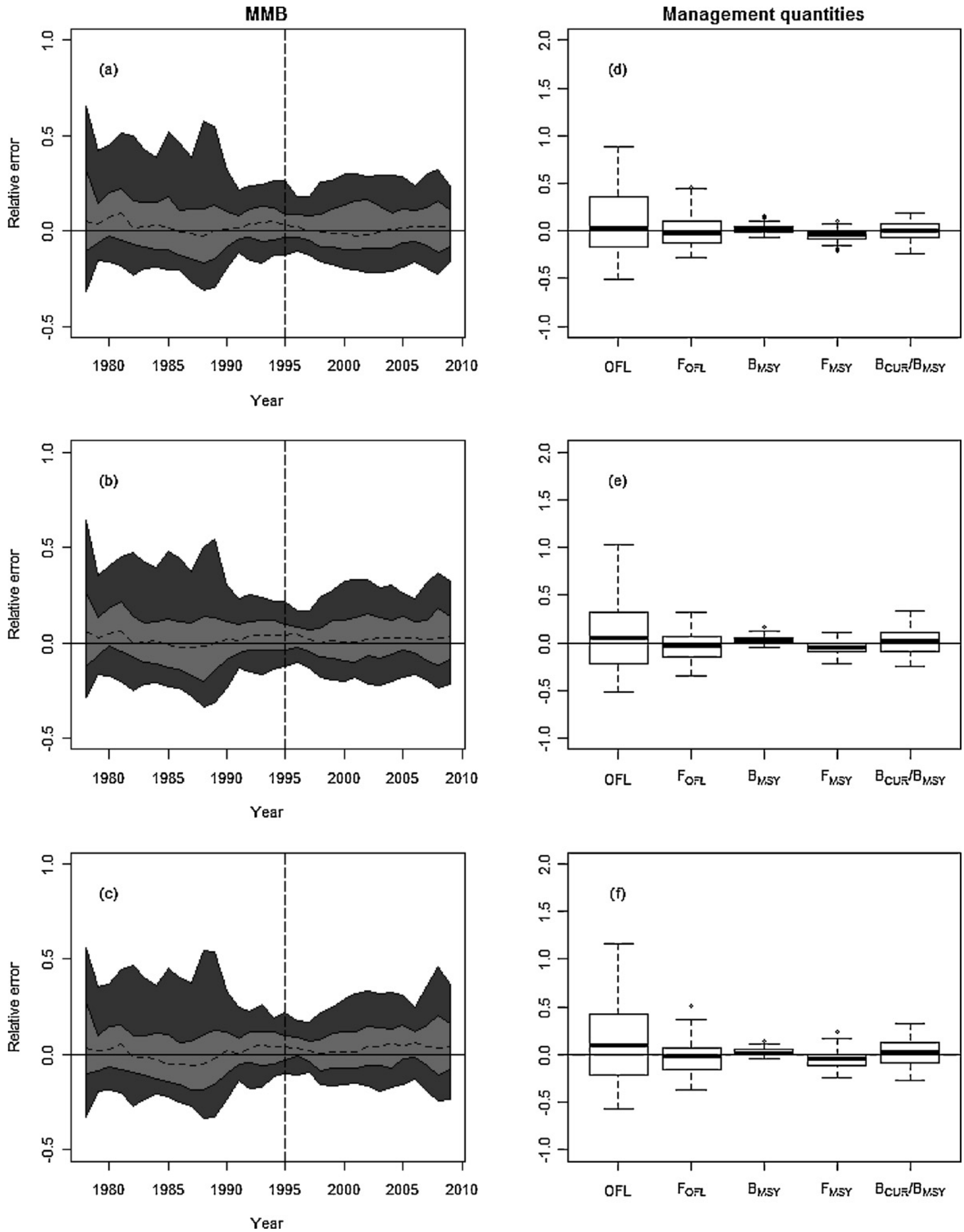

Fig. 9. Relative error distributions for mature male biomass (MMB) and management quantities under scenarios in which average F_{mort} in the operating model was multiplied by 0.5 (a and d), 1 (b and e), and 2 (c and f) starting in 1995. Quantities are estimated with no penalties on fishing mortality deviations.

1) and the standard NMFS nets (Somerton et al., 2010) were used to inform survey selectivity and catchability in the stock assessment.

4.2. Data quality

The benefits of additional growth data have already been mentioned and have the potential to improve model estimates greatly. Assuming the CV and effective sample sizes from the survey are better than they actually are leads to much larger variance in the estimated OFLs (resulting in a large MARE), but the median bias is small. The other survey data scenarios led to very little change in MARE from the reference scenario.

4.3. Assessment assumptions

The scenario in which average fishing mortality was multiplied by 2 from 1995 ("$F \times 2.0$" in Table 1) led to the largest change

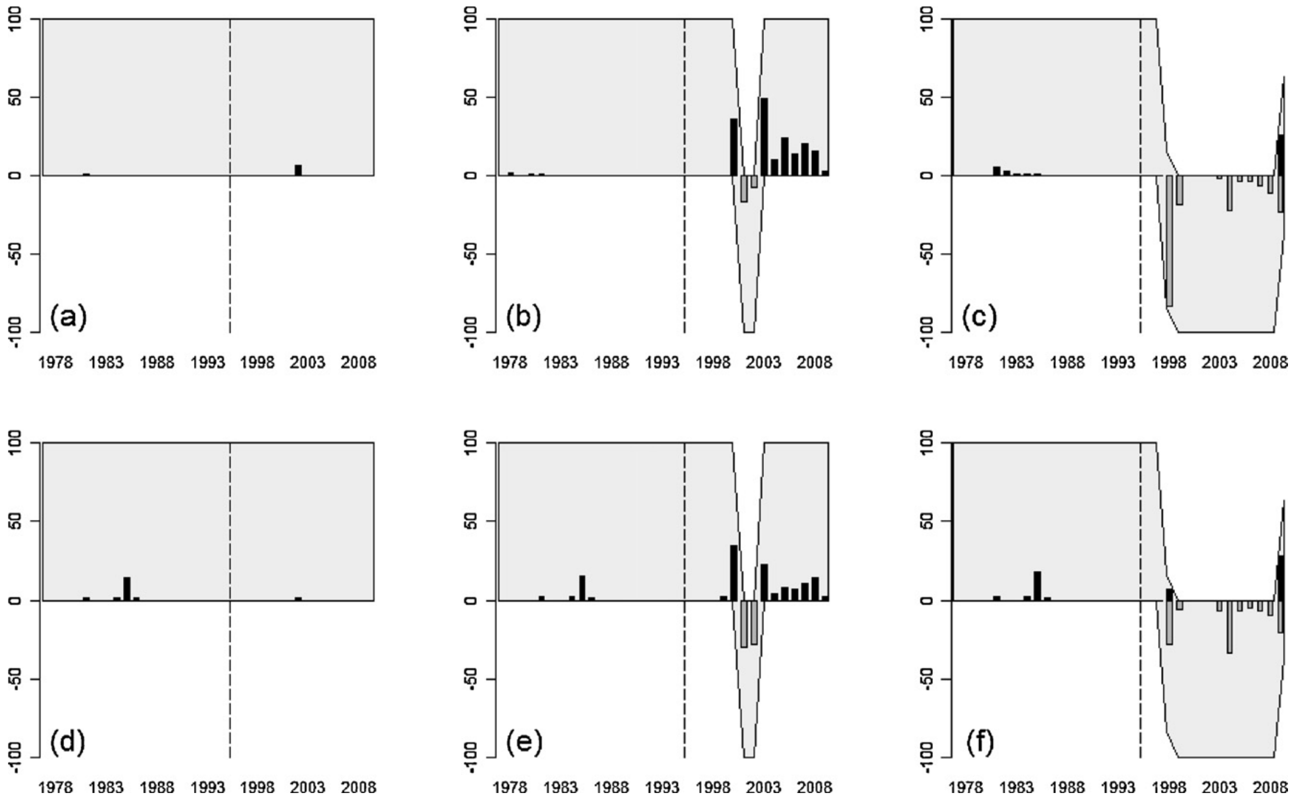

Fig. 10. Evaluation of the ability to correctly identify overfished status under scenarios in which average F_{mort} in the operating model is multiplied by 0.5 (a and d), 1 (b and e) and 2 (c and f) from 1995 forward. Gray background shading indicates the number of simulations in which the true population was overfished (those below the "0" line) or not overfished (those above the "0" line). Black bars above the "0" line indicate the number of times a 'healthy' stock was identified as being overfished and gray bars beneath the "0" line indicate the number of times an overfished stock was identified as healthy. Panels a–c show results for scenarios in which the original penalties are applied and panels d–f show results for scenarios in which penalties are eliminated.

in the MARE for the OFL compared to the reference scenario. Although this scenario is not likely in reality, it is instructive in evaluating the ability of the assessment method to detect overfished stocks. In the first year in which the simulated stock was overfished in this scenario, MMB was much less than MSST, but the assessment method was unable to identify a change in overfished status with penalties on fishing mortality in place. Detecting that MMB is less than MSST soon after MMB drops below MSST is very difficult when penalties are placed on changes on deviations in fishing mortality because the assessment method does not allow the large changes in fishing mortality to match the change in observed abundance. Consequently, penalties on fishing mortality can make achieving management goals more difficult when the stock is well below MSST. However, if MMB is just below the MSST on average ("$F \times 1.0$" scenario; Fig. 10b and e), fishing mortality penalties reduce the number of overfished stocks that are assessed as healthy. Hence, there is a trade-off between incorrectly identified overfished and incorrectly identified 'healthy' stocks, with more 'healthy' stocks being identified as overfished (particularly in the years after the stock recovers to above MSST) with increased penalties.

Although decreasing (and ultimately eliminating) penalties generally resulted in the least amount of bias in management quantities for all fishing mortality scenarios, estimates varied more among simulations when the penalties were weaker. This is because penalties add artificial certainty around the estimates of deviations in fishing mortality (which then propagates through to the management quantities). Consequently, the selection of penalties on fishing mortality should be made after considering the associated risk, the history and variability of fishing pressure, and perceptions regarding biomass in relation to the reference biomass.

5. Conclusions and caveats

Considering the biases in OFL across scenarios and the existing body of data related to the processes considered here, future field research directed at decreasing uncertainty around growth and natural mortality would be most valuable to improve the estimation ability of the model. More accurately specifying the growth curve using an increased number of pre- and post-molt observations may be an efficient way of increasing confidence in the estimates produced by the stock assessment (the BSFRF has since begun collection of immature, molting crab). Results from the selectivity scenarios suggest that time spent incorporating all data available on selectivity into the stock assessment method (e.g. Somerton et al., 2010, which is now used) could have large impacts on estimated management quantities.

This study has focused on EBS snow crab, but the results have implications for other assessments based on length-structured models. Our results support Punt (2003a) in demonstrating an incorrectly specified size-transition matrix (from small sample sizes or spatial heterogeneity in growth) can have large impacts on estimation ability. Potentially less specific to size-based models, penalties placed on deviations in fishing mortality were found to introduce bias and undue precision to management quantities and should therefore be carefully considered before implementation. Finally, relatively large relative errors around the OFL in the reference scenario emphasize the importance of precautionary management strategies.

Although our results are expressed in terms of how well quantities of management interest are estimated, this does not directly determine the implications on the ability to satisfy management goals. These implications can, however, be examined using Management Strategy Evaluation, MSE (e.g. Punt and Smith, 1999; Smith et al., 1999; A'Mar et al., 2009a,b) and efforts are currently underway to apply MSE to the EBS snow crab fishery.

Acknowledgements

We would like to thank the North Pacific Research Board for funding this project, Jack Turnock for providing his stock assessment code and two anonymous reviewers for comments that improved this paper. This is NPRB publication 364.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fishres.2012.08.007.

References

A'mar, Z.T., Punt, A.E., Dorn, M.W., 2009a. The impact of regime shifts on the performance of management strategies for the Gulf of Alaska walleye pollock (*Theragra chalcogramma*) fishery. Can. J. Fish. Aquat. Sci. 66, 2222–2242.

A'mar, Z.T., Punt, A.E., Dorn, M.W., 2009b. The evaluation of two management strategies for the Gulf of Alaska walleye Pollock fishery under climate change. ICES J. Mar. Sci. 66, 1614–1632.

Butterworth, D.S., Punt, A.E., 1999. Experiences in the evaluation and implementation of management procedures. ICES J. Mar. Sci. 56, 985–998.

Chen, Y., Wilson, C., 2002. A simulation study to evaluate impacts of uncertainty on the assessment of American lobster fishery in the Gulf of Maine. Can. J. Fish. Aquat. Sci. 59, 1394–1403.

Chen, Y., Kanaiwa, M., Wilson, C., 2005. Developing and evaluation a size-structure stock assessment model for the American lobster, *Homarus americanus*, fishery. N. Z. J. Mar. Fresh. 39, 645–660.

Hilborn, R., Walters, C.J., 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics, and Uncertainty. Chapman Hall, New York.

Hobday, D., Punt, A.E., 2001. Size-structured population modelling and risk assessment of the Victorian southern rock lobster, *Jasus edwardsii*, fishery. Mar. Freshwater Res. 52, 1495–1507.

Hoenig, J., 1983. Empirical use of longevity data to estimate mortality rates. Fish. Bull. 82, 898–903.

Jennings, C.A., Zigler, S.J., 2000. Ecology and biology of paddlefish in North America: historical perspectives, management approaches, and research priorities. Rev. Fish Biol. Fish. 10, 167–181.

Kanaiwa, M., Chen, Y., Hunter, M., 2005. Assessing a stock assessment framework for the green sea urchin *Strongylocentrotus drobachiensis* fishery in Maine, USA. Fish. Res. 74, 96–115.

McDonald, A.D., Smith, A.D., Punt, A.E., Tuck, G.N., Davidson, A.J., 1997. Empirical evaluation of expected returns from research on stock structure for determination of total allowable catch. Nat. Res. Mod. 10, 3–29.

NPFMC (North Pacific Fishery Management Council)/NMFS (National Marine Fisheries Service), 2000. Environmental Assessment for the Bering Sea Snow Crab Rebuilding Plan. Amendment 14. North Pacific Fishery Management Council, 605 W, 4th Ave., Suite 306, Anchorage, AK 99501, USA or www.dakr.noaa.gov.

NPFMC (North Pacific Fishery Management Council)/NMFS (National Marine Fisheries Service), 2007. Environmental Assessment for Amendment 24. Overfishing Definitions for Bering Sea and Aleutian Islands King and Tanner Crab Stocks. North Pacific Fishery Management Council, 605 W, 4th Ave, Suite 306, Anchorage, AK 99501, USA or www.dakr.noaa.gov.

NRC (National Research Council), 1998. Improving Fish Stock Assessments. Committee on Fish Stock Assessment Methods, Ocean Studies Board, Commission on Geosciences, Environment and Resources. National Academy Press, Washington, DC.

Powers, J.E., Restrepo, V.R., 1993. Evaluation of stock assessment research for Gulf of Mexico king mackerel: benefits and costs to management. N. Am. J. Fish. Manage. 13 (1), 15–26.

Punt, A.E., 2003a. The performance of a size-structured stock assessment method in the face of spatial heterogeneity in growth. Fish. Res. 65, 391–409.

Punt, A.E., 2003b. Evaluating the efficacy of managing West Coast groundfish resources through simulations. Fish. Bull. 101, 860–873.

Punt, A.E., Smith, A.D.M., 1999. Harvest strategy evaluation for the eastern stock of gemfish (*Rexea solandri*). ICES J. Mar. Sci. 56, 860–875.

Punt, A.E., Smith, A.D.M., Cui, G., 2002. Evaluation of management tools for Australia's south-east fishery, 2. How well can management quantities be estimated? Mar. Freshwater Res. 53, 631–644.

Smith, A.D.M., Sainsbury, K.J., Stevens, R.A., 1999. Implementing effective fisheries management systems – management strategy evaluation and the Australian partnership approach. ICES J. Mar. Sci. 56, 967–979.

Somerton, D.A., 1981. Life history and population dynamics of two species of Tanner crab, *Chionoecetes bairdi* and *C. opilio*, in the eastern Bering Sea with implications for the management of the commercial harvest. Ph.D. Dissertation. University of Washington, Seattle, p. 220.

Somerton, D., Wenberg, K., Goodman, S., 2010. Review of the research to estimate snow crab selectivity by the NMFS trawl survey. Report to NPFMC, January 2010.

Turnock, B.J., Rugolo, L.J., 2009. Stock assessment of eastern Bering Sea snow crab. In: NPFMC (North Pacific Fishery Management Council), Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Region: 2009 Crab SAFE, September 2009. North Pacific Fishery Management Council, 605 W, 4th Ave., Suite 306, Anchorage, AK 99501, USA.

Turnock, B.J., Rugolo, L.J., 2011. Stock assessment of eastern Bering Sea snow crab. In: NPFMC (North Pacific Fishery Management Council). Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Region: 2011 Crab SAFE, September 2011. North Pacific Fishery Management Council, 605 W, 4th Ave., Suite 306, Anchorage, AK 99501, USA.

Zheng, J., 2003. Uncertainties of natural mortality estimates for eastern Bering Sea snow crab, *Chionoecetes opilio*. Fish. Res. 65, 411–425.

Zheng, J., Siddeek, M.S.M., 2010. Bristol Bay red king crab assessment in spring 2010. In: NPFMC (North Pacific Fishery Management Council). Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Region: 2010 Crab SAFE, September 2010. North Pacific Fishery Management Council, 605 W, 4th Ave., Suite 306, Anchorage, AK 99501, USA.