Managing fisheries for catch and conservation: should harvesting be adjusted to account for perceived biases in fish stock assessments?

by

Adam Gerrit John Keizer

B.Sc., Simon Fraser University, 2007

Research Project Submitted in Partial Fulfillment of the
Requirements for the Degree of
Master of Resource Management

Report No. 614

in the
School of Resource and Environmental Management
Faculty of Environment

SIMON FRASER UNIVERSITY

Spring 2015

Approval

Name:	**Adam Gerrit John Keizer**
Degree:	**Master of Resource Management**
Title:	***Managing fisheries for catch and conservation: should harvesting be adjusted to account for perceived biases in fish stock assessments?***
Report No:	**614**
Examining Committee:	**Chair:** Michelle Jones Ph.D. candidate in Resource and Environmental Management

Dr. Sean Cox
Senior Supervisor
Associate Professor

Dr. Robyn Forrest
Supervisor
Research Scientist,
Fisheries and Oceans Canada

Date Defended/Approved: January 8, 2015

Partial Copyright Licence

Abstract

Information from stock assessment models is uncertain, and biased information can result in a fishery failing to meet its objectives. A retrospective pattern is a measure of uncertainty produced by a retrospective analysis that indicates that the data are inconsistent with assumptions made in the stock assessment model specification. Decision-makers may make ad hoc adjustments to harvest advice based on information from a retrospective analysis, however the ability of a fishery to meet conservation and yield objectives as a result of such an adjustment are often unknown. I use feedback simulation modelling to evaluate the performance of alternative harvest control rules and adjustments in the presence of retrospective patterns. My study found no conservation benefit to using a retrospective adjustment derived from recent historic retrospective patterns, and that the adjustment comes at a cost to catch stability.

Keywords: feedback simulation; retrospective; fisheries management procedures; stock assessment models; management strategy evaluation; uncertainty

Acknowledgements

I would like to thank my senior supervisor, Sean Cox, for his guidance, and Robyn Forrest for her support and feedback on my research project and professional development. I thank the many contributors to *mseR*, software that made my research process considerably more interesting. I appreciate the financial support provided by the Natural Sciences and Engineering Research Council of Canada, Fisheries and Oceans Canada, and Simon Fraser University.

I am grateful to my classmates and instructors in the School of Resource and Environmental Management, and in particular to my friends in the Fisheries Research Group, who made my education a challenging and rewarding experience. To my family, who carried more than their share, I am forever grateful. Thank you, Angeline, Quinn, and Adalyn.

Table of Contents

List of Tables

List of Figures

1. Introduction

Fishery decision-makers use information from fishery stock assessment models to balance trade-offs between conflicting objectives of harvest and conservation. However, information from stock assessment models is uncertain, and biased information can result in a fishery failing to meet its objectives (Hutchings and Myers, 1994; Walters and Maguire, 1996; National Research Council (NRC), 1998). With complex systems such as fisheries, stock assessment models require simplifying assumptions because underlying processes and the value of parameters that describe them are usually inestimable with available data (e.g., the rate of natural mortality and how it changes over time). Fishery-dependent and independent observations of fish abundance and biological characteristics are also often uncertain and may be biased. In addition, stock assessment model outputs are sensitive to assumptions about parameters and their relationships to data, and may produce stock assessment advice that is biased. These types of uncertainty in stock assessment advice contribute to some of the greatest challenges faced by fishery decision-makers.

One approach for characterizing uncertainty is to quantify changes in our understanding of uncertainty over time. A retrospective analysis is the process of examining the consistency of parameter and state variable estimates over time as data accumulate in a stock assessment model (Rivard and Foy, 1987; Sinclair et al., 1991; Mohn, 1999; NRC, 1998). For example, as additional data are added to a stock assessment model over time, estimates of biomass are recalculated for each model year. Whereas one might expect that revised estimates of biomass for a given year remain unchanged year after year, revised estimates may in fact be systematically biased relative to previous estimates (Figure 1). This systematic difference among estimates is known as a retrospective pattern (Mohn, 1999). A retrospective pattern is a measure of uncertainty that indicates that the data are inconsistent with assumptions made in the stock assessment model specification (NRC, 1998). An indication of model misspecification, such as a retrospective pattern, is useful because traditional measures of uncertainty, such as standard error, are based on the assumption that the model is specified correctly. Thus conventional measures may underestimate uncertainty relative to a retrospective analysis (Parma, 1993; NRC, 1998).

Identifying uncertainty is important, but what to do with the information generated through a retrospective analysis is not always clear. The specific causes of a retrospective pattern are often unknown and difficult to correct (Legault, 2009; Mohn, 1999). A retrospective pattern may present itself through an among-model analysis (comparing estimates over time from distinct modelling approaches), or through a within-model analysis. While an among-model analysis can be illustrative of uncertainty, the patterns cannot be assessed or ameliorated given that they are a function of the analyst and thus are unpredictable. The causes of within-model retrospective patterns are numerous, including observation error, process error, and the misspecification of state dynamics (NRC, 1998; Mohn, 1999; Legault, 2009). Despite the uncertain causes of within-model retrospective patterns and their relationship to stock assessment model biases, some assessments attempt to "correct" retrospective patterns so that harvest advice is consistent from year to year. The 1994 assessment of Silver Hake (*Merluccius billnearis*) on the Scotian Shelf off eastern Canada presented a retrospective pattern that consistently underestimated *F* by 40%-60%. The cause of the pattern is unknown, but projected population numbers from the assessment model were reduced by 40% (Showell and Bourbonnais, 1994). A retrospective pattern in the stock assessment for Georges Bank Yellowtail Flounder (*Limanda ferruginea*), believed to result from changing catchability, was reduced by splitting the survey into two time series (Stone and Legault, 2005; Northeast Fisheries Science Center (NEFSC), 2008(a)). Such splitting of fishery-independent survey data has since been used to address retrospective patterns in several groundfish stock assessments along the northeast U.S. Atlantic coast (NEFSC, 2008(a); NEFSC, 2012). If splitting the survey time series did not ameliorate a retrospective pattern, numbers-at-age in the assessment's terminal year were adjusted by the recent average retrospective pattern (Stone and Legault, 2005; NEFSC, 2008(b); NEFSC, 2012). Since 1994, three distinct assessment models for Pacific Halibut (*Hippoglossus stenolepis*) have presented retrospective patterns that consistently under- or overestimated biomass (Parma, 1993; Clark, 2003; Stewart and

Martell, 2013). In all three instances a change in the parameterization of selectivity was used to reduce the retrospective pattern (Stewart and Martell, 2013). Stewart and Martell (2013) demonstrated that the most recent retrospective pattern was created by decreasing length-at-age where initial recruitment was underestimated, but subsequently increased as the fish grew and were fully selected (Stewart and Martell, 2013). Before changing the parameterization of selectivity, Valero (2011) proposed adjusting the population abundance-at-age of Pacific Halibut in a manner similar to NEFSC (2008(b), 2012) and Showell and Bourbonnais (1994). These assessments have attempted to address retrospective patterns by interpreting the systematic bias as deviations from the best estimate of true biomass. A measure of this deviation is then used to directly adjust the estimated biomass before catch advice is calculated. Typically such ad hoc adjustments are made to avoid the risks associated with failing to meet harvest and conservation objectives. However, efforts to account for the effects of retrospective patterns in this way often only address aesthetic aspects of a pattern and may not reduce overall risk of failing to meet fishery objectives (NEFSC, 2008(b); Legault, 2009). For example, adjusted biomass estimates may not be closer to the true biomass, or the assessment model may be biased in other ways despite good diagnostics for the revised model (Legault, 2009). NRC (1998) suggests downward biomass adjustments are precautionary to protect against over harvest, but cautions that upward adjustments to compensate for a negative retrospective pattern are risky because retrospective patterns can change direction without notice. In reality, decision-makers balance trade-offs of harvest and conservation and are likely to consider adjustments in both directions. In the presence of a retrospective pattern, can and how should harvest advice be adjusted to address model misspecification? If an adjustment is made to harvest advice, how do the risks to conservation and yield change?

I use feedback simulation modelling to investigate the characteristics of retrospective statistics and to evaluate the performance of alternative management strategies in the presence of retrospective patterns. A feedback simulation model has three components: (1) a *system*, an operating model that represents the assumed “true” fish stock population dynamics; (2) a *sensor*, a model that provides the source of data about the stock (e.g., a simple survey index value, a complex stock assessment model); and (3) a *controller*, a harvest control model that applies management actions (e.g., harvest quota) to the operating model based on the output of the *sensor* (Figure 2). Assumptions about operating model structure, parameter values, and natural variability, are combined to create scenarios within the *system*. The components of the *sensor* and the *controller* are the *management strategy*. To evaluate the performance of a management strategy in a scenario, the output of the *controller* is applied to the *system* and the process is repeated over a number of projected years, and the entire process is replicated a number of times to generate a distribution of outcomes.

A feedback system that simulates the application of a management strategy to a resource is essential to evaluate management strategies. For example, in a system without a feedback loop, the controller could set a fixed catch amount to be caught each year, based on current stock productivity. However, if stock productivity decreases, the fixed harvest amount would result in a greater proportional catch of the total stock than desired. Without a sensor to monitor the stock and its response to the management strategy, it is impossible to evaluate, and where necessary, adjust the management strategy. Among fisheries science literature, control theory is generally discussed in the context of management strategy evaluation (MSE). The advantages of MSE are its ability to evaluate risk and lay bare the tradeoffs between alternative management strategies (Smith, 1993; Butterworth, 2007), as well as to provide an effective framework for decision-makers, stakeholders, and analysts to consider and implement management strategies (Butterworth, 2007; Cox and Kronlund, 2008; Punt et al., 2014). Management strategy evaluation does not prescribe the optimal management strategy, rather it requires participants to be clear about fishery objectives and then enables participants to identify a management strategy that best meets the fishery objectives. While the methods I employ in this study follow the same structure as an MSE would, I do not work with decision-makers and stakeholders to define specific objectives. Instead I define and evaluate high-level goals about conservation and fishery yield. With this distinction I define my study as "feedback simulation modelling," rather than a stakeholder-driven MSE.

Using existing retrospective statistics as a guide (Parma, 1993; Mohn, 1999; NEFSC, 2008(a); NEFSC, 2012), I examined a retrospective statistic that measures the consistency of historical state variable estimates relative to the current best estimates.

This statistic is also used to determine a retrospective adjustment. I used simulations to evaluate a set of management strategies, specifically combinations of harvest control rules (HCRs) and retrospective adjustments, against conservation and yield outcomes over a range of scenarios. My use of feedback simulation is a novel approach to evaluating harvest control rules and retrospective adjustments in the presence of retrospective patterns.

2. Methods

2.1. Model Implementation

I modified the computer software package, *mseR, Management Strategy Evaluation in R* (Kronlund et al., 2012), to implement feedback simulations designed to examine the performance of HCRs and retrospective adjustments over a range of operating model scenarios. The operating model is an age-structured model with a Beverton-Holt stock-recruitment relationship (Beverton and Holt, 1957), parameterized represent a haddock-type species (Table 1). Non-stationarity of biological parameters (e.g., natural mortality) is one cause of retrospective patterns (Mohn, 1999). For example, if true natural mortality (*M*) is increasing over time but the assessment model assumes it is fixed, the assessment model may overestimate biomass. To create retrospective patterns in my modeled assessments, I parameterized natural mortality in the operating model with a random walk in a constant (scenario S1), negative (scenario S2), or positive (scenario S3) trend (Table 1). In the feedback simulation model, the observation model provides the assessment model with a log-normally distributed index of exploitable biomass, and a vector of catch-age proportions in each year. The assessment model is a statistical catch-age model with a structure similar to the operating model, except for two main differences: (i) in the assessment model, recruitment process errors are uncorrelated; and (ii) catch is taken at the beginning of each year instead of continuously, which greatly speeds up calculations.

I used two HCRs: (1) a "constant *F*" rule, where the intended fishing mortality rate is constant regardless of stock status; and (2) a "variable *F*" rule, where the intended fishing mortality rate is dependent on the stock status, as defined by two reference points. The reference points are defined as a fraction of B_{MSY}, and for the purposed of my study I used the default values defined by Fisheries and Oceans

Canada's (2009) "A Fishery Decision-Making Framework Incorporating the Precautionary Approach."

$$F_{n|n} = F_{MSY} \tag{1}$$

$$F_{n|n} = \begin{cases} 0, & \hat{B}_{n|n} < 0.4B_{MSY} \\ F_{MSY}\left(\frac{\hat{B}_{n|n} - 0.4B_{MSY}}{0.4B_{MSY}}\right), & 0.4B_{MSY} \leq \hat{B}_{n|n} < 0.8B_{MSY} \\ F_{MSY}, & \hat{B}_{n|n} \geq 0.8B_{MSY} \end{cases} \tag{2}$$

Where $\hat{B}_{n|n}$ is the biomass estimate from the assessment model at the current time step, n. The two harvest control rules are shown in Figure 3.

Feedback simulation modelling enables one to examine the performance of constant *F* and variable *F* HCRs where both the estimated biomass and the reference points in the HCR are affected by a retrospective pattern. I evaluated how the HCRs performed when a retrospective adjustment was applied to $\hat{B}_{n|n}$. The combinations of possible HCRs and retrospective adjustments produced four management strategies (Table 2) that were evaluated under each of the three operating model scenarios (Table 1). The 12 simulations were each run for 60 years, while the management strategy and feedback simulation start at year 35. During the historical tuning period from *t* = 1, 2 ... 34, the operating model constructs the stock with the initialization parameters provided, as well as vectors of recruitment error and fishing mortality. The catch-age assessment model is informed by observations of abundance and age composition collected during the tuning period and throughout the management strategy period. For each scenario, the entire feedback loop was repeated for 25 replicates.

2.2. Retrospective statistics

I assessed the ability of a retrospective statistic to measure retrospective patterns. Although retrospective "pattern" and "bias" are occasionally used interchangeably in the literature, I make the distinction that a "pattern" is a systematic difference among estimates (as defined by Mohn, 1999), whereas a "bias" is a difference between estimates and true population parameters. While the intended use of retrospective statistics is to quantify and ameliorate an assumed assessment model bias, retrospective statistics are only able to quantify retrospective patterns (i.e., differences between biomass estimates from year to year) because true bias (i.e., differences between biomass estimates and true biomass) cannot be measured empirically. A key advantage to the application of feedback simulation models is they enable characterization of true bias, at least within the feedback simulation framework.

I begin by defining some key terminology in retrospective analysis. A retrospective analysis is performed by sequentially adding one year of data to a stock assessment model. The reference model is defined as the assessment model estimate with the longest and most recent dataset, representing the "best" estimate obtained with the most data. Historical assessment estimates are the terminal estimates calculated in the time steps previous to that of the reference model (Figure 1). To describe the retrospective statistics, I use the following notation:

Current estimate	$x_{n\,|\,n}$	The estimate in the current time step, n, from the reference model that spans time=1 to time= n.
Reference estimate	$x_{t\,|\,n}$	The estimate at time t from the reference model that spans time=1 to time= n. n is the current time step.
Terminal estimate	$x_{t\,|t}$	The estimate at time t from the model that spans time=1 to time=t.
Retrospective duration	i ... $(n-1)$	This is the duration over which a retrospective statistic is calculated, where i is the first time step that a retrospective analysis is completed. Often the total length of the retrospective analysis (i through $n-1$) is fixed. As a result, as data are added with each time step, and the value of n and i each increase by 1.

For all retrospective measurements I initially used a five year-long retrospective duration, with the first measurement occurring five years after implementation of a management strategy. There is no theoretical basis for determining the length of a retrospective duration, so a length of five years was selected based on the length of time available for the analysis (25 years), and to allow multiple cohorts to recruit and pass through the fishery (life history characteristics include age at 50% maturity = 3 years, 10+ age-classes). With a five year retrospective duration the first retrospective adjustment occurs five years after the start of the management strategy, year 40. In a sensitivity analysis I also examined retrospective durations that were three and seven years long.

2.3. Sum of Relative Errors

The sum of relative errors (SRE) is a retrospective statistic that measures the differences between terminal estimates and a reference estimate. The SRE is defined as:

$$\mathrm{SRE} = \sum_{t=i}^{n-1} \frac{(x_{t\,|\,t} - x_{t\,|\,n})}{x_{t\,|\,n}} \tag{3}$$

The SRE is similar to the measurement proposed by Mohn (1999). Terminal estimates that are consistently greater than the reference estimate will result in a large, positive SRE. Terminal estimates that are consistently less than the reference will result in a large, negative SRE. Even if they are large, terminal estimates that have inconsistent direction or error relative to the reference estimate will result in a small SRE.

2.4. Retrospective adjustments

A retrospective adjustment is a post hoc alteration of parameter or state variable estimates intended to correct the assumed bias, as measured by a retrospective statistic. Assuming that the direction and magnitude of a retrospective pattern continues into the future, a simple retrospective adjustment is to adjust the assessment model estimate by the same amount of the measured retrospective pattern as is observed in recent assessments (NRC, 1998). This type of adjustment has been used or suggested in stock assessments for fisheries on the east and west coasts of North America (Showell and Bourbonnais, 1994; NEFSC, 2008(a); NEFSC, 2012; Valero, 2011).

In this study, I applied a retrospective adjustment based on the recent mean retrospective pattern that adjusted $\hat{B}_{n|n}$ prior to applying the HCR (equation 1 or equation 2). Specifically,

$$\hat{B}'_{n|n} = \frac{\hat{B}_{n\,|\,n}}{1+\overline{\left(\frac{(\hat{B}_{t\,|\,t}-\hat{B}_{t\,|\,n})}{\hat{B}_{t\,|\,n}}\right)}} \tag{4}$$

Where the numerator, $\hat{B}_{n\,|\,n}$, is the current estimated exploitable biomass from the reference model. The denominator is the mean relative error of the terminal models' estimates relative to the reference model. The denominator is the SRE averaged over the length of the retrospective duration. For example, if terminal estimates were overestimating biomass by an average of 10% relative to the reference model each year over the past 5 years, the most recent reference model biomass estimate, $\hat{B}_{n\,|\,n}$, would be reduced by 10%. The result is $\hat{B}'_{n|n}$, the adjusted biomass estimate, which is used in calculating a catch limit at each time step for which the adjustment is applied. This adjustment is applied only in the HCR, and does not remain for subsequent retrospective analyses. For example, if the estimated biomass was decreased by 10% as a result of the adjustment, subsequent years' estimates and retrospective analyses consider the original estimate, not the reduced estimate.

2.5. Management strategy performance

Management strategies were evaluated with three measures: final depletion, average catch, and average absolute variation (AAV). These three performance statistics were used to evaluate management strategies against three broad goals: minimizing the probability of depleting the stock, maximizing average catch, and minimizing catch variability (Rademeyer et al., 2007). Final depletion (equation 5) is the median operating model biomass at a given year across all replicates in a simulation, relative to the unfished operating model biomass. Average catch is the average catch across all replicates in a simulation. Average absolute variation (equation 6) is the amount by which catch varies annually over a given time period, relative to total catch over the same time period in a simulation (Punt and Smith, 1999).

$$Final\ Depletion = \tilde{B}_{\mathrm{t}} / B_0 \tag{5}$$

$$AAV = \frac{\sum_{40}^{60} |C_t - C_{t-1}|}{\sum_{40}^{60} C_t} \tag{6}$$

2.6. Closed-loop simulation evaluations

I measured retrospective patterns and evaluated management strategies under each of the three operating model scenarios in the following manner:

(i) Initialize operating model scenarios for a 34 year time period where *M* is parameterized with a constant, negative, or positive trend, resulting in three scenarios (Table 1).

(ii) Define four management strategies with combinations of HCRs (constant *F*, variable *F*) and a retrospective adjustment (Table 2).

(iii) The operating model and fishery advance in one year increments from *t* = 35:60. The assessment model is provided with a survey index of abundance, and survey and fishery catch-age composition to generate $\hat{B}$, $\hat{F}_{MSY}$, and $\hat{B}_{MSY}$.

(iv) As defined by the management strategy, calculate a catch limit for the pending year with the information provided in (iii). Where the management strategy dictates, the catch limit is first modified by the retrospective adjustment in year 40 based on the previous 5 years' retrospective pattern.

(v) Update the operating model based on the harvest that occurred as defined by the management strategy.

(vi) Repeat (iii) - (v) to year 60. Calculate performance statistics based on final depletion, average catch, and AAV.

(vii) Repeat (i) - (vi) for 25 replicates and calculate summary performance statistics.

3. Results

3.1. Retrospective patterns

I created retrospective patterns where the terminal estimates of biomass were systematically different from the reference estimates of biomass. In simulations where the operating model was parameterized with a constant trend in *M* (S1, Table 1), terminal spawning biomass estimates were generally not systematically different from the reference estimates of biomass (Figure 4). Simulations in scenario S1 produced variable $\hat{B}_{MSY}$ and $\hat{F}_{MSY}$ estimates that decreased over time (Figure 5). Where the operating model was parameterized with a negative trend in *M* (S2, Table 1), terminal spawning biomass estimates were generally less than reference model estimates (Figure 4). Simulations in scenario S2 produced relatively stable estimates of B_{MSY} and F_{MSY} over time (Figure 5). The assessment model estimated the trend in *M* reasonably well, but the assessment model bias tended to overestimate *M*, stock-recruitment steepness, and B_0 (Figure 7). Where the operating model was parameterized with a positive trend in *M* (S3, Table 1), terminal spawning biomass estimates were generally larger than reference model estimates (Figure 4). Simulations in scenario S3 produced variable $\hat{B}_{MSY}$ and $\hat{F}_{MSY}$ estimates that decreased over time (Figure 5). The assessment model generally estimated a constant, rather than positive, trend in $\hat{M}$ and often underestimated *M* and B_0 (Figure 8). In general, the assessment model performed reasonably well with few model convergence errors across all simulation runs.

While retrospective patterns are visually apparent (Figure 4), measures of the patterns using the SRE suggest that the patterns are less definitive. The median SRE, calculated at each time step between year 35 and year 60, across all replicates are: S1 = 0.323, S2 = -0.330, and S3 = 1.17. The median SRE for S1 suggests the assessment model has unexpectedly created a retrospective pattern where terminal biomass is consistently overestimated. The median SRE for S2 suggests the assessment model has created a retrospective pattern where terminal biomass is consistently underestimated, as hypothesized. The magnitude of this pattern is similar to that created in S1. The median SRE for S3 suggests the assessment model has created a retrospective pattern where biomass is consistently overestimated, as hypothesized. The magnitude of this pattern is almost four times greater than the patterns created in S1 and S2. While these values demonstrate a difference between scenarios, the retrospective statistics varied considerably among replicates of simulation runs, often so much so that patterns produced by the assessment model are indistinguishable among scenarios (Table 3). Within scenarios the SRE for different management strategies were also indistinguishable.

3.2. Management strategy performance

I compared the performance of constant *F* and variable *F* HCRs without applying the retrospective adjustment (MS1 vs. MS2). In all simulations the constant *F* HCR performed comparably to the variable *F* HCR in terms of final depletion and average catch (Figure 9). The variable *F* HCR resulted in a slightly larger median AAV for S1 and S2, and a noticeably larger median AAV for S3 (Figure 9). I evaluated how all HCRs performed when a retrospective adjustment was applied to $\hat{B}_{n\,|\,n}$. Figure 10 illustrates the retrospective adjustment applied to a pattern taken from a representative replicate in MS3. Retrospective adjustments varied considerably within simulations, both between replicates for a given time step and over time (Figure 11). The application of the retrospective adjustment increased median AAV in both management strategies (MS3, MS4) and magnified the differences in median AAV between the constant *F* and variable *F* HCRs, but did not result in an appreciable change to final depletion or average catch (Figure 9).

I examined the sensitivity of the management strategies' performance to the retrospective duration, and assumed observation error (catch aging error and survey sample variance). The management strategies' median final depletion and median average catch performance were not very sensitive to alternative retrospective durations of three and seven years (Figure 12). The median AAV increased as the retrospective duration decreased, where shorter durations did not have the same smoothing effect as longer time periods. The management strategies' median final depletion and median average catch performance were also not very sensitive to higher or lower levels of observation error. The median AAV increased noticeably as observation error increased with greater variability in model estimates of biomass (Figure 13).

4. Discussion

I evaluated the performance of alternative management strategies in the presence of retrospective patterns and found that an adjustment based on SRE was of little consequence to conservation objectives. No HCR was consistently more robust than another when measured against median final depletion and median average catch. The variable *F* HCR resulted in greater catch variability than the constant *F* HCR, which is unsurprising given how target *F* changes in response to biomass estimates near reference points. The retrospective adjustment resulted in greater catch variability.

For simulations with a constant trend in *M* (S1) the estimated biomass was generally greater than $0.8\hat{B}_{MSY}$, so as a result the target harvest rate for the constant *F* and variable *F* HCRs were both usually $\hat{F}_{MSY}$. Given similar target harvest rates, the performance of these HCRs was indistinguishable for median final depletion and median average catch, while median AAV was slightly greater after accounting for the few instances where the variable *F* HCR target harvest rate was less than $\hat{F}_{MSY}$. Unsurprisingly the similarity in performance between these HCRs did not change when the management strategies were modified by the retrospective adjustment. The constant trend in *M* scenario resulted in a retrospective adjustment that did not consistently increase or decrease estimated biomass and subsequent target harvest rates. As a result median final depletion and median average catch were indistinguishable between adjusted and unadjusted management strategies, while median AAV increased slightly because of (i) changes in estimated reference points, e.g., $\hat{B}_{MSY}$, $\hat{F}_{MSY}$ (Figure 5) and subsequent catch, and (ii) for the variable *F* HCR, changes in the target harvest rate when the estimated biomass fell below $0.8\hat{B}_{MSY}$ more frequently. It appears that the retrospective adjustment affected the stock enough to create a response that changed the estimated reference points and increase catch variability, but the adjustment was too small and inconsistent between replicates to move final depletion or catch in one direction.

Retrospective adjustments that have been proposed for real-world fish stock assessments have most commonly been proposed in situations where terminal biomass estimates consistently overestimate biomass (Showell and Bourbonnais, 1994; NRC, 1998; Stone and Legault, 2005; NEFSC, 2008(a); NEFSC, 2012), as occurred in my scenario with a positive trend in *M* (S3). In these simulations median final depletion and median average catch were less than half what they are in the baseline S1 simulations, while the median AAV increased more then 25% in S3 from S1. As in the S1 simulations, the variable *F* HCR median AAV was greater than the constant *F* HCR median AAV, however the difference between the two was greater than the difference observed in the S1 simulations. The estimated biomass in S3 was frequently less than $0.8\hat{B}_{MSY}$, which is the reason that the variable *F* HCR median AAV was greater than S1, and that the difference between the variable *F* HCR and constant *F* HCR median AAV was more pronounced in S3. The differences between variable *F* HCR and constant *F* HCR that exist for AAV are amplified by the retrospective adjustment. Despite applying retrospective adjustments to reduce estimated biomass, with the intention of reducing harvest to meet conservations objectives, there is little change to median final depletion or median average catch. In both the constant *F* and variable *F* HCRs $\hat{B}$ was reduced by the adjustment (Figure 11), but the predicted effect of reduced catch limits was not consistently realized as changing reference point estimates (Figure 5) confounded the effect of the adjustment. It appears that the adjustment resulted in an initial small decrease in catch and increase in biomass, but catch then increased and biomass decreased relative to unadjusted management strategies for a short period of time before the adjusted and unadjusted management strategies became indistinguishable. In these simulations *M*, and subsequently B_0 and B_{MSY}, were underestimated. This underestimation resulted in a less conservative variable *F* HCR as the $0.8\hat{B}_{MSY}$ stock reference point was reduced relative to the "true" value. This biased reference point offset conservation benefits that may have resulted from the retrospective adjustment. F_{MSY} was also underestimated, which made the HCR more conservative, but the underestimation was less consistent than the bias in $\hat{B}_{MSY}$ and affected the catch limit to a lesser degree. Moreover, it appears that the effect of the adjustment was muted by the noise of errors across the simulation replicates. The combined effect of the adjustment, biased reference points, and variability between replicates resulted in a median final depletion and median average catch that was similar between adjusted and unadjusted management strategies.

NRC (1998) suggests that retrospective adjustments in the presence of historical overestimation are ad hoc but precautionary, however, NRC does not recommend applying adjustments when faced with historical underestimation. NRC (1998) states that the risk from applying an adjustment is too great because the retrospective pattern can change from an under- to overestimation without warning. I examined management strategies in a scenario with underestimation of biomass that resulted from a negative trend in *M* (S2). These simulations enabled me to evaluate forgone harvest opportunity that may exist from unadjusted retrospective patterns. In these simulations median final depletion doubled and median average catch was 65% greater than the S1 baseline, while the median AAV decreased slightly in S2 from S1. Between adjusted and unadjusted management strategies, the only difference was a small increase in median AAV. In these simulations *M*, and subsequently B_0 and B_{MSY}, were overestimated. This overestimation resulted in a more conservative variable *F* HCR as the $0.8\hat{B}_{MSY}$ increased relative to the "true" value, however the changes in $\hat{B}_{MSY}$ over time were less than what was observed in the other scenarios The bias in these reference points offset any harvest benefits that may have resulted from the retrospective adjustment. Bias in estimated reference points and the "ramp" effect of the variable *F* HCR affected the AAV, although to a lesser degree in S2 than in the other scenarios. With a large fish stock relative to $0.8\hat{B}_{MSY}$, the "ramp" effect of the variable *F* HCR occurred infrequently which explains the similar median AAV values of the two HCRs. In the S2 simulations the limited effect of the retrospective adjustment also appears to be affected by the magnitude of the retrospective pattern. In most of the replicates, biomass rose rapidly and then fell resulting in a relativity small retrospective pattern and adjustment. The retrospective adjustment, based on recent historical retrospective patterns, lagged behind the fish stock assessment when the pattern changed direction quickly, resulting in larger maxima and smaller minima of the adjusted $\hat{B}$ and subsequent catch limits, which in turn increased AAV.

Retrospective patterns indicate that the data are inconsistent with assumptions made in the model. The inconsistency examined here is a non-stationarity of *M*, although other causes of retrospective patterns may include unreported catch, changing selectivity, or the amount of information in the available data. Mechanisms for non-stationarity of *M* as a result of environmental change are diverse, ranging from life history change (early maturation and senescence), parasite-induced mortality (either directly, or from increased predation susceptibility), changing prey availability, and changing predation. While some have attempted to identify specific mechanisms and their level of influence on the non-stationarity of *M* (e.g., Swain et al., 2011), the relationship between *M* and these factors is complex. Moreover the ability to achieve fishery objectives does not improve if a management strategy is modified to include environmental factors unless the relationship between the factors and system are well known (Punt et al., 2013). Given the challenges in identifying the specific cause of non-stationarity of *M*, or even if non-stationarity of *M* is the cause of a retrospective pattern, a retrospective pattern can be difficult to correct in an assessment model. In this context my research used non-stationarity of *M* to create a retrospective pattern and applied an ad hoc retrospective adjustment, rather than attempt to resolve the non-stationarity issue.

Retrospective adjustments are one way to use the information from a retrospective analysis to avoid the risks associated with failing to meet harvest and conservation objectives. Retrospective adjustments are not an approach to "correct" an assessment given that adjustments address aesthetic issues, rather than addressing the underlying cause of the pattern. Given that the decision to adjust an estimate is about managing risk, the adjustment procedure must be evaluated, and the decision of how and when to apply an adjustment must be made in concert with fishery decision-makers. Similar types of ad hoc adjustments have been used in fisheries management decisions. For example, the International Pacific Halibut Commission applied a "Slow Up-Fast Down" (SUFD) rule to the harvest policy with the intent of slowing annual catch limit changes that resulted from both biological and assessment methodological changes (Hare, 2010). The SUFD policy was not formalized until after it had already been used for a number of years. While a formalization of the SUFD policy aided decision-makers in making consistent decisions, a lack of prior simulation testing meant that the policy was being applied without being evaluated. Simulation studies later found that the approach produced harvest advice that was not responsive enough over short term to changes in stock status (Hare, 2010), and that the harvest policy could have led to fishery objectives not being met. Ad hoc adjustments are an important tool that provide decision-makers with the ability manage risk while trying to achieve harvest and conservation objectives, however such tools should be formalized and evaluated to understand how they affect risk.

My study found no conservation benefit to using an SRE-based retrospective adjustment, and that the adjustment comes at a cost to catch stability. The lack of conservation benefit exists because the adjustment is too small, and it lags behind the assessment model's perceived bias. This is in part a function of the SRE-based adjustment that adjusts based on the recent historical retrospective pattern, but also because of the inconsistency of the pattern itself. The magnitude of the retrospective pattern varied between replicates and, combined with the adjustment's lag, the performance of the adjustment with respect to final depletion is small over the long term. When performance is measured over a three year period where a consistent retrospective pattern is present across most replicates, the affect of the adjustment on depletion is more noticeable but still small. This is because the size of the adjustment (median values of $0.85\hat{B}$ and $0.94\hat{B}$ for the positive and negative trend in *M* scenarios, respectively) and the resulting catch is small relative to B_0. Given the retrospective pattern based on a non-stationarity in *M*, and the relatively small retrospective adjustments that were produced by the SRE-based retrospective adjustment, adjusting harvest to account for a perceived bias is unwarranted. While my adjustment did not improve the performance of the management strategies, I acknowledge several caveats. The way in which analysts and decision-makers would respond to a retrospective pattern over the long term differs from the short term. Over the long term, like the 25 years my assessment model ran, it is unlikely that the same retrospective pattern-producing stock assessment would be used to provide harvest advice. I also acknowledge that this conclusion is driven by choices made about how to parameterize the operating model (e.g., the use of a haddock-type fish stock, the type of non-stationarity), the structure of the assessment model and the information provided to it (e.g., priors on *h*, *M*), and the types of management strategies applied to the fish stock. Additional research should explore the short term use of a retrospective adjustment with alternative representative life history types. For example, given the influence of parameter and reference point bias in my results, the consequences of a retrospective adjustment may differ for a short-lived species like herring (*Culpea spp.*), with different assumptions about process and parameter uncertainty. Additional research should also explore retrospective adjustments applied to retrospective patterns generated from different types of model uncertainty. My retrospective pattern was generated by a non-stationarity in *M*, but retrospective patterns also arise from limited information in the available data. For example, a retrospective pattern observed in the 2013 Pacific Hake (*Merluccius productus*) stock assessment is believed to be the result of an error in recruitment estimates where cohorts have only been observed a few times (Joint Technical Committee, 2013).

While retrospective adjustments are one way to use the information from a retrospective analysis, it is insufficient to use only a retrospective adjustment when presented with a retrospective pattern. In scenarios such as those that I have examined, an SRE-based retrospective adjustment is unwarranted. Despite the difficulties, effort should be focused on resolving the root cause of a retrospective pattern, rather than addressing the symptoms. Using feedback simulation to evaluate alternative management strategies that consider different retrospective adjustments, changing data availability (e.g., survey frequency, aging data), and fisheries controls (e.g., frequency with which catch limits change) may also provide insightful results for management agencies with limited or reduced resources. A feedback simulation approach also helps to identify future research and monitoring needs when prioritizing science and management resources (Marasco et al., 2007). The strength of examining retrospective patterns and retrospective adjustments in feedback simulations is that this process offers a framework to explicitly account for uncertainty while providing fishery decision-makers with information to evaluate the trade-offs and identify management strategies that are robust to that uncertainty.

Tables and Figures

Table 1. Feedback simulation model parameterization. The operating model is an age-structured model with a Beverton-Holt stock-recruitment relationship (Beverton and Holt, 1957), parameterized to represent a generic haddock-type species. Three scenarios are created in the operating model by parameterizing the instantaneous natural mortality rate (*M*) with constant, negative, and positive trends. The observation model provides the assessment model with a log-normally distributed index of exploitable biomass, and a vector of catch-age proportions in each year. The assessment model is a statistical catch-age model with priors on *M* and the recruitment function steepness (*h*).

Simulation Component	**Parameter Value**
Operating Model, Population Size and Productivity, and Reference Points	
Unfished spawning biomass (B_0)	75 ('000 tonnes)
Recruitment function steepness (*h*)	0.7
Recruitment lag-1 autocorrelation	0
Standard error of log-recruitment	0.7
Total number of years in simulation	60 years
First year of management strategy	35 years
Maximum Sustained Yield (MSY)	5,147 tonnes
Spawning biomass at MSY (B_{msy})	24,970 tonnes
Target removal rate at MSY (F_{msy})	0.2258
Operating Model: Life History	
Age at 50% maturity	3 years
Age at 95% maturity	4 years
Age at 50% fishery selectivity	2 years
Age at 95% fishery selectivity	2.5 years
Age at 50% survey selectivity	2 years
Age at 95% survey selectivity	2.5 years
Survey catchability coefficient (*q*)	1
Length at age 1	10 cm

Simulation Component	**Parameter Value**
Asymptotic length (L_{inf})	65 cm
Standard error of length-at-age	0 cm
Von Bertalanffy *K*	0.35
Allometric a	$9.26*10^{-6}$
Allometric b	3.12
Number of age classes	10
Instantaneous natural mortality rate (*M*)	0.3
Standard error of instantaneous natural mortality rate (*M*)	0.1
Operating Model: Scenario Conditions	
S1: Constant trend in *M*	
Instantaneous natural mortality rate ($M_{t=0}$) mean	0.3 / year
Instantaneous natural mortality rate ($M_{t=60}$) mean	0.3 / year
S2: Negative trend in *M*	
Instantaneous natural mortality rate ($M_{t=0}$) mean	0.3 / year
Instantaneous natural mortality rate ($M_{t=60}$) mean	0.2 / year
S3: Positive trend in *M*	
Instantaneous natural mortality rate ($M_{t=0}$) mean	0.2 / year
Instantaneous natural mortality rate ($M_{t=60}$) mean	0.3 / year
Observation Model: Survey and Fisheries Ages Data	
Survey age proportion observation standard error	0.3
Fishery age proportion observation standard error	0.3
Survey CV mean	0.1
Survey CV standard deviation	0.1
Assessment Model: Statistical Catch-at-Age Model Priors	
Instantaneous natural mortality rate (*M*) mean	0.3 / year
Instantaneous natural mortality rate (*M*) standard deviation	0.1 / year
Recruitment function steepness (*h*) mean	0.7
Recruitment function steepness (*h*) standard deviation	0.1
Harvest Control Rule	
Upper stock reference point [for variable *F* only]	$0.8F_{msy}$
Lower stock reference point [for variable *F* only]	$0.4F_{msy}$

Table 2. Management strategies. I used feedback simulation modelling to examine the performance of constant *F* and variable *F* harvest control rules (HCRs), in combination with a retrospective adjustment, in the presence of a retrospective pattern. The combinations of possible HCRs and retrospective adjustments produced four management strategies that were evaluated under each of the three operating model scenarios.

Management strategy	Retrospective adjustment	Harvest control rule
MS1	No	Constant F
MS2		Variable F
MS3	Yes	Constant F
MS4		Variable F

Table 3. Summary statistics of sum of relative error (SRE). The SRE is a retrospective statistic that measures the differences between terminal estimates and a reference estimate. The SRE was measured in each time step between year 35 and year 60, and summary statistics were compiled across all replicates within each scenario and management strategy. SRE values greater than 0 indicate terminal estimates of biomass are consistently greater than then reference model estimates. SRE values less than 0 indicate terminal estimates of biomass are consistently less than then reference model estimates.

Scenario and Management Strategy	**Minimum**	**1st Quartile**	**Median**	**3rd Quartile**	**Maximum**
S1: Constant trend in *M*					
MS1: Constant *F* + no Adjustment	-1.43	-0.157	0.317	0.957	6.31
MS2: Variable *F* + no Adjustment	-1.37	-0.156	0.324	0.944	4.80
MS3: Constant *F* + Adjustment	-1.33	-0.158	0.333	0.941	4.63
MS4: Variable *F* + Adjustment	-1.33	-0.158	0.322	0.940	4.45
S2: Negative trend in *M*					
MS1: Constant *F* + no Adjustment	-1.34	-0.565	-0.291	0.054	1.91
MS2: Variable *F* + no Adjustment	-1.34	-0.565	-0.276	0.061	1.91
MS3: Constant *F* + Adjustment	-1.45	-0.609	-0.302	0.054	1.82
MS4: Variable *F* + Adjustment	-1.45	-0.609	-0.302	0.054	1.82
S3: Positive trend in *M*					
MS1: Constant *F* + no Adjustment	-2.18	0.633	1.33	2.24	61.4
MS2: Variable *F* + no Adjustment	-2.19	0.582	1.24	2.11	12.5
MS3: Constant *F* + Adjustment	-2.18	0.581	1.21	2.10	12.5
MS4: Variable *F* + Adjustment	-2.19	0.536	1.12	2.05	12.5

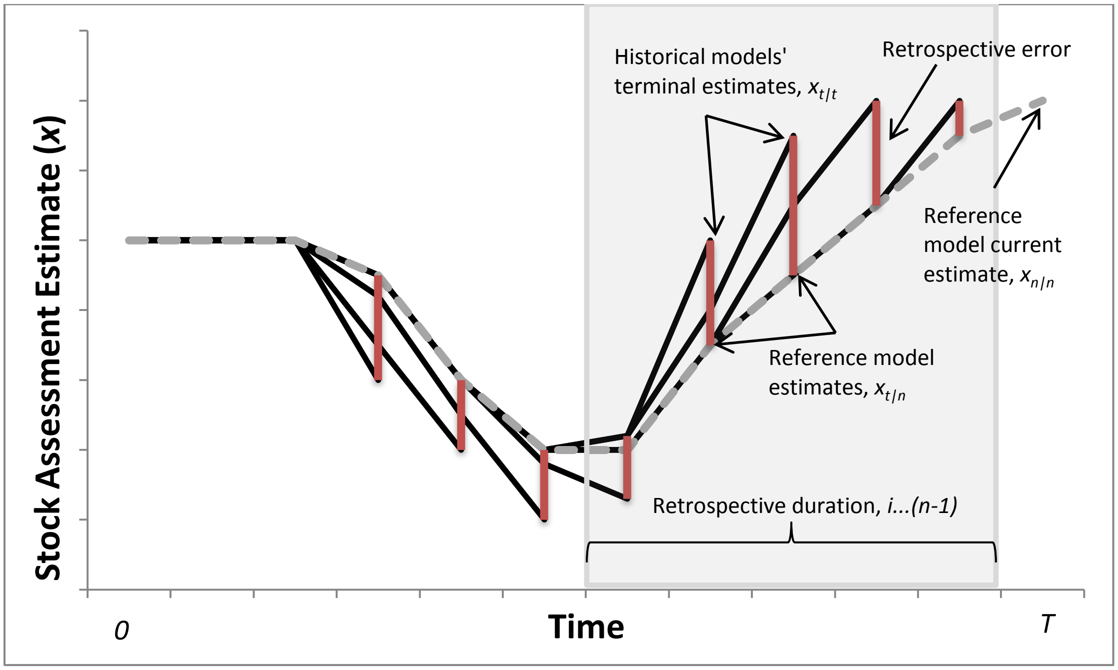

Figure 1. A retrospective pattern. The reference model is represented as the dashed line with biomass estimates from time = 0 to time = *T*. The historical assessment estimates are represented by the solid lines. Retrospective error, in red, is a measure of the difference between historical assessments' terminal estimates and the reference model's estimate at the same time step. The retrospective duration, the period of time over which a retrospective analysis is completed, is represented by the shaded area. As more data are added over time, the reference model spans a longer period of time and additional historical assessment estimates are made. The retrospective duration is a fixed length, so as newer historical estimates become available, older historical estimates are dropped and the retrospective duration's "window" moves forward.

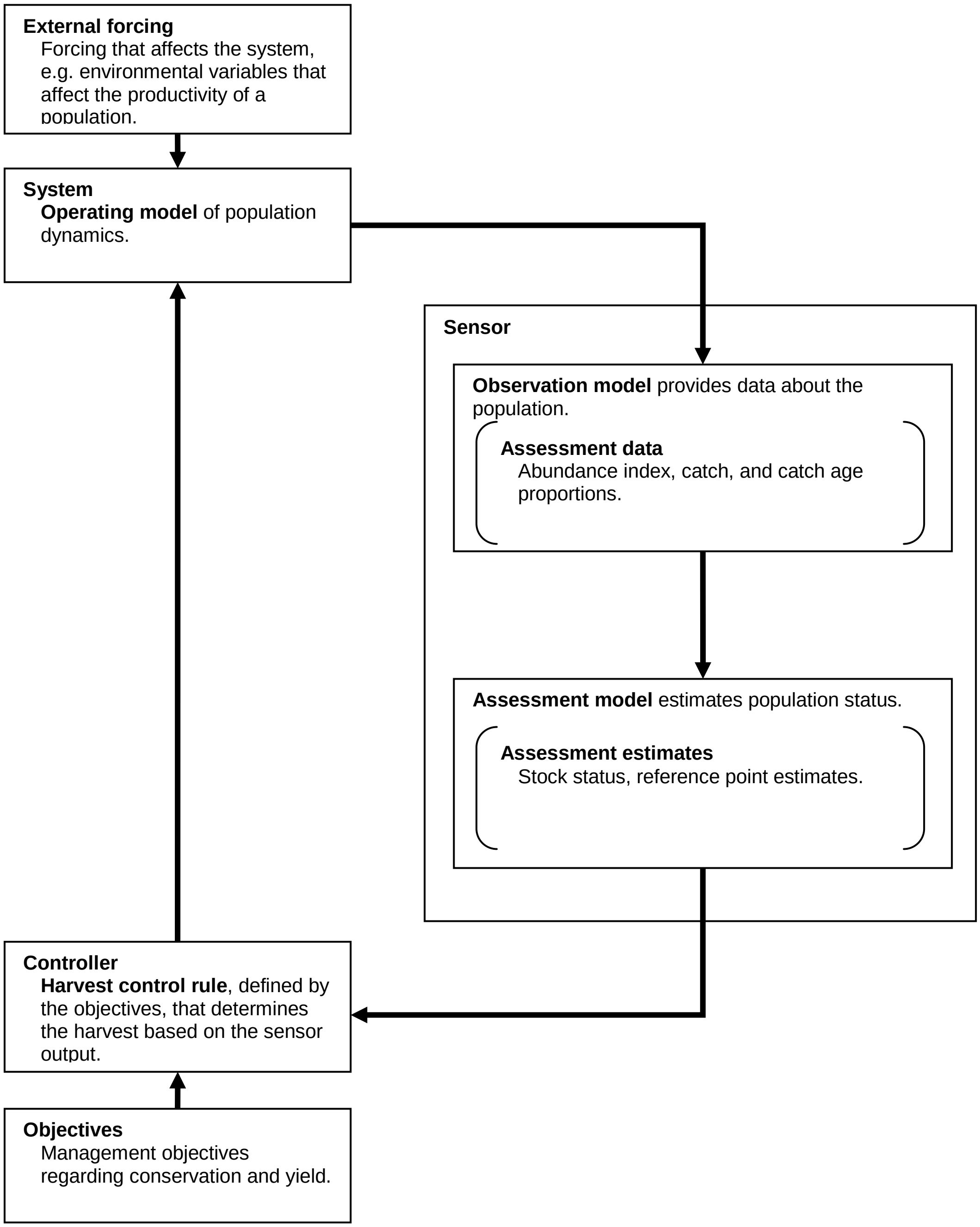

Figure 2. A feedback simulation model has three components: (1) a *system*, (2) a *sensor*, and (3) a *controller*. Assumptions about operating model structure, parameter values, and natural variability are combined to create scenarios within the *system*. The components of the *sensor* and the *controller* are the *management strategy*.

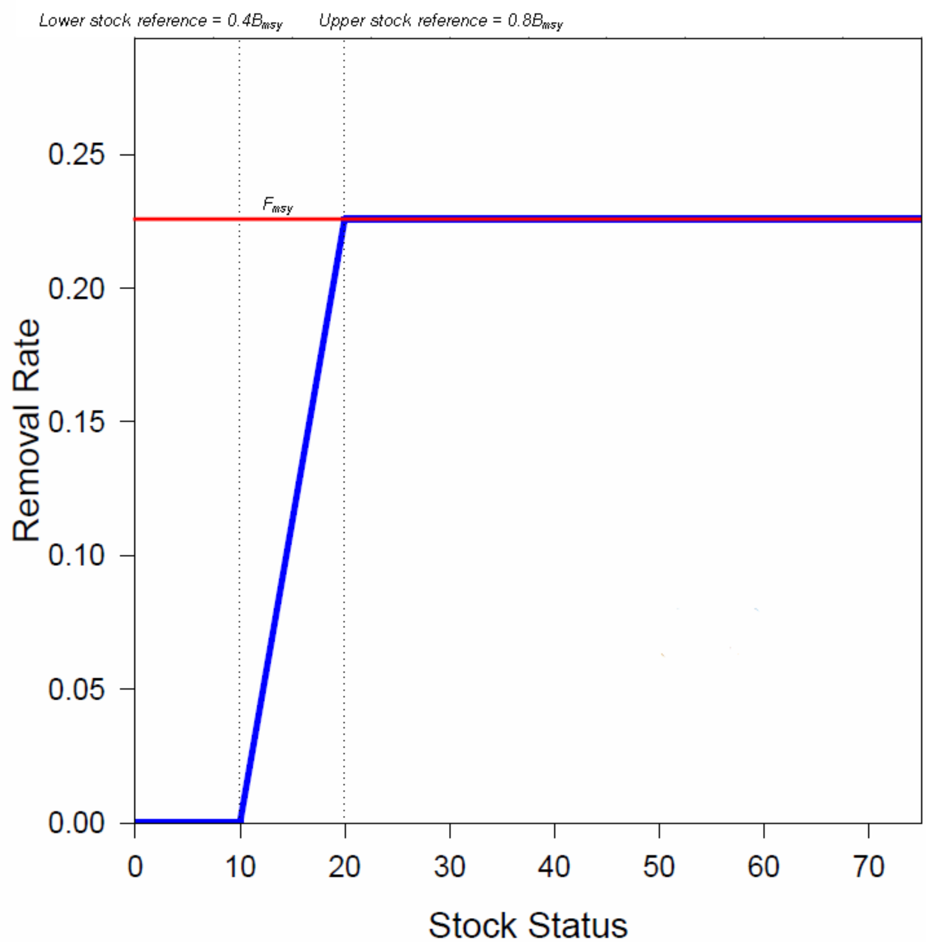

Figure 3. Example constant *F* (red) and variable *F* (blue) harvest control rules used in the management strategies. The constant *F* harvest control rule defines a level of harvest at time, *t*, where intended fishing mortality is F_{MSY}.The variable *F* harvest control rule defines a level of harvest at time, *t*, where intended fishing mortality is F_{MSY} when the fish stock is greater than or equal to the upper stock reference point ($0.8B_{MSY}$). As the fish stock declines, the intend harvest rate declines linearly until the stock is equal the lower stock reference point ($0.4B_{MSY}$). If the stock is less than the lower stock reference point, the intended harvest rate is 0.

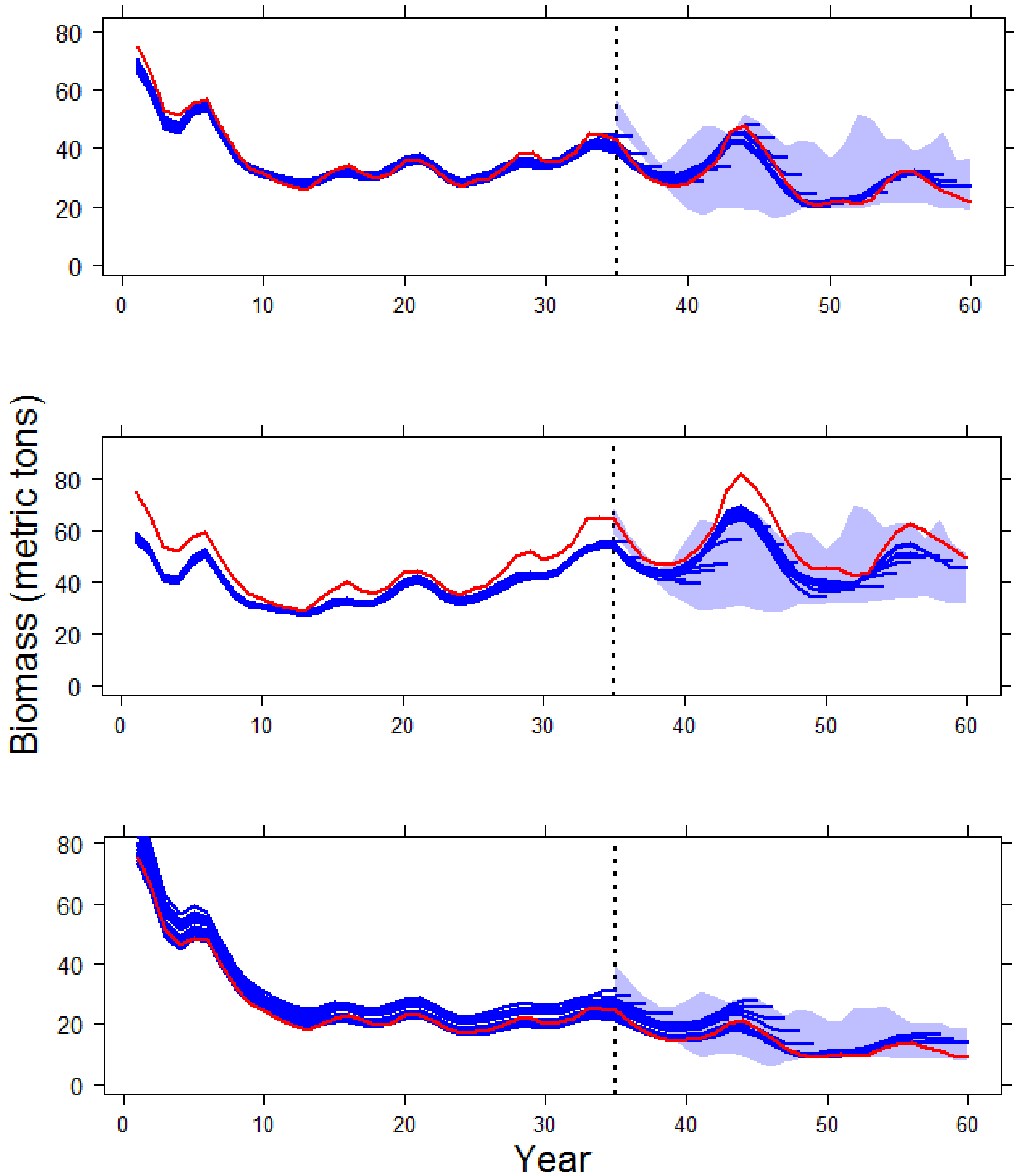

Figure 4. Retrospective patterns of estimated biomass created by parameterizing natural mortality (*M*) in the operating model with a random walk in a constant (S1, top), negative (S2), and positive (S3, bottom) trend. The shaded region is bounded by the 5th and 95th percentiles of terminal biomass estimates from all replicates. The lines are spawning biomass estimates from an individual representative replicate taken from management strategy MS1. The red line is the "true" biomass from the operating model.

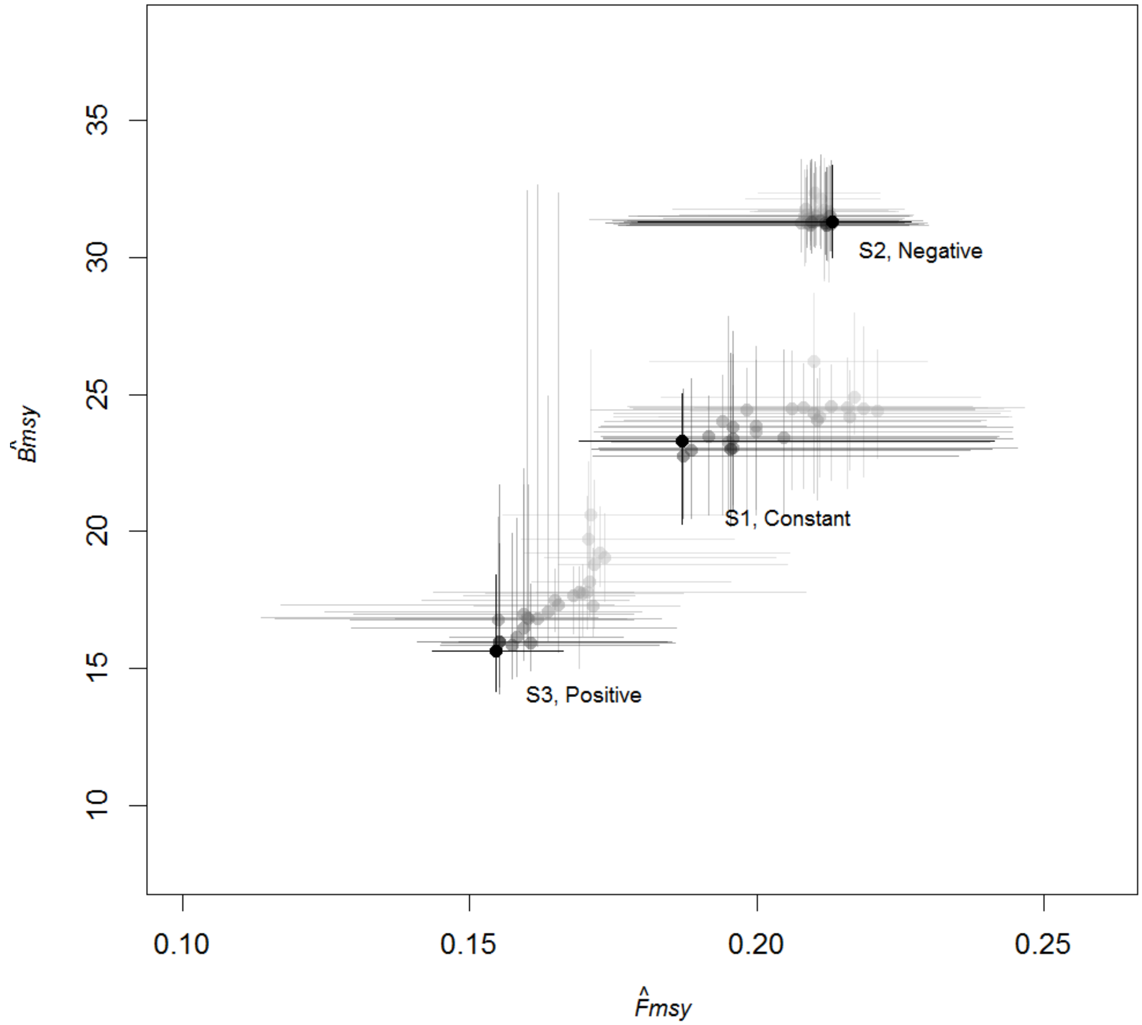

Figure 5. $\hat{F}_{MSY}$ and $\hat{B}_{MSY}$ over time, by scenario. Points represent the median value across all replicates and management strategies of a given scenario. Lines encompass the 5th and 95th percentiles of the estimates. The statistics are calculated for each time step between year 35 (light grey) and year 60 (black). $\hat{F}_{MSY}$ and $\hat{B}_{MSY}$ decrease over time for the scenarios with a positive (S3) and constant (S1) trend in natural mortality (*M*). $\hat{F}_{MSY}$ and $\hat{B}_{MSY}$ are relatively stable for the scenario with a negative trend (S2) in *M*.

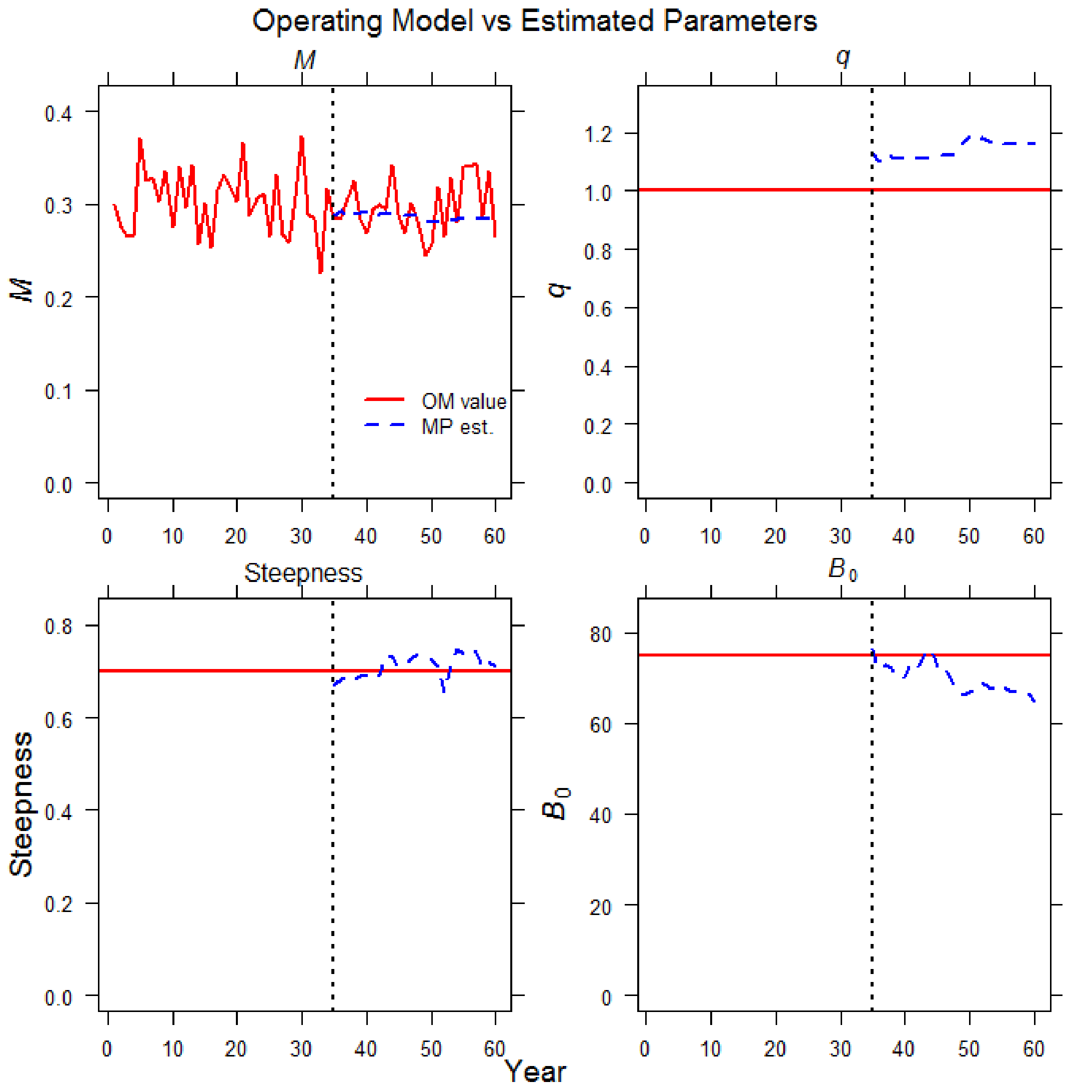

Figure 6. Operating model (OM) parameter values and parameters estimated from the assessment model (MP) for scenario S1, taken from a representative replicate in management strategy MS3. In scenario S1 natural mortality (*M*) was parameterized with a random walk and a constant trend. The assessment model estimated *M* and steepness reasonably well, while survey catchability ($\hat{q}$) was overestimated and $\hat{B}_0$ was estimated with a negative trend.

.

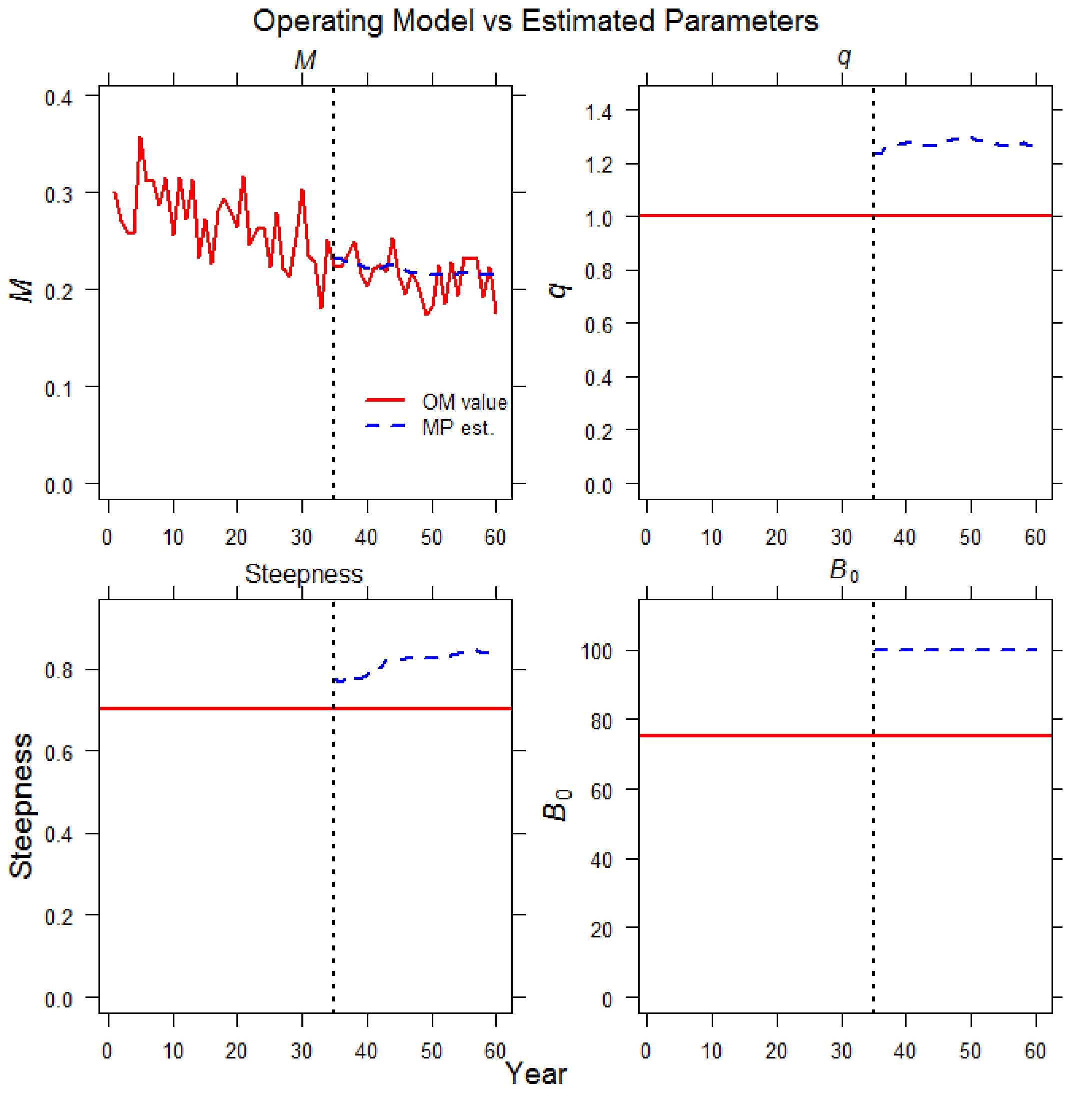

Figure 7. Operating model (OM) parameter values and parameters estimated from the assessment model (MP) for scenario S2, taken from a representative replicate in management strategy MS3. In scenario S2 natural mortality (*M*) was parameterized with a random walk and a negative trend. The assessment model estimated the trend in *M* reasonably well, but the assessment model bias tended to overestimate *M*, survey catchability ($\hat{q}$), steepness, and $\hat{B}_0$.

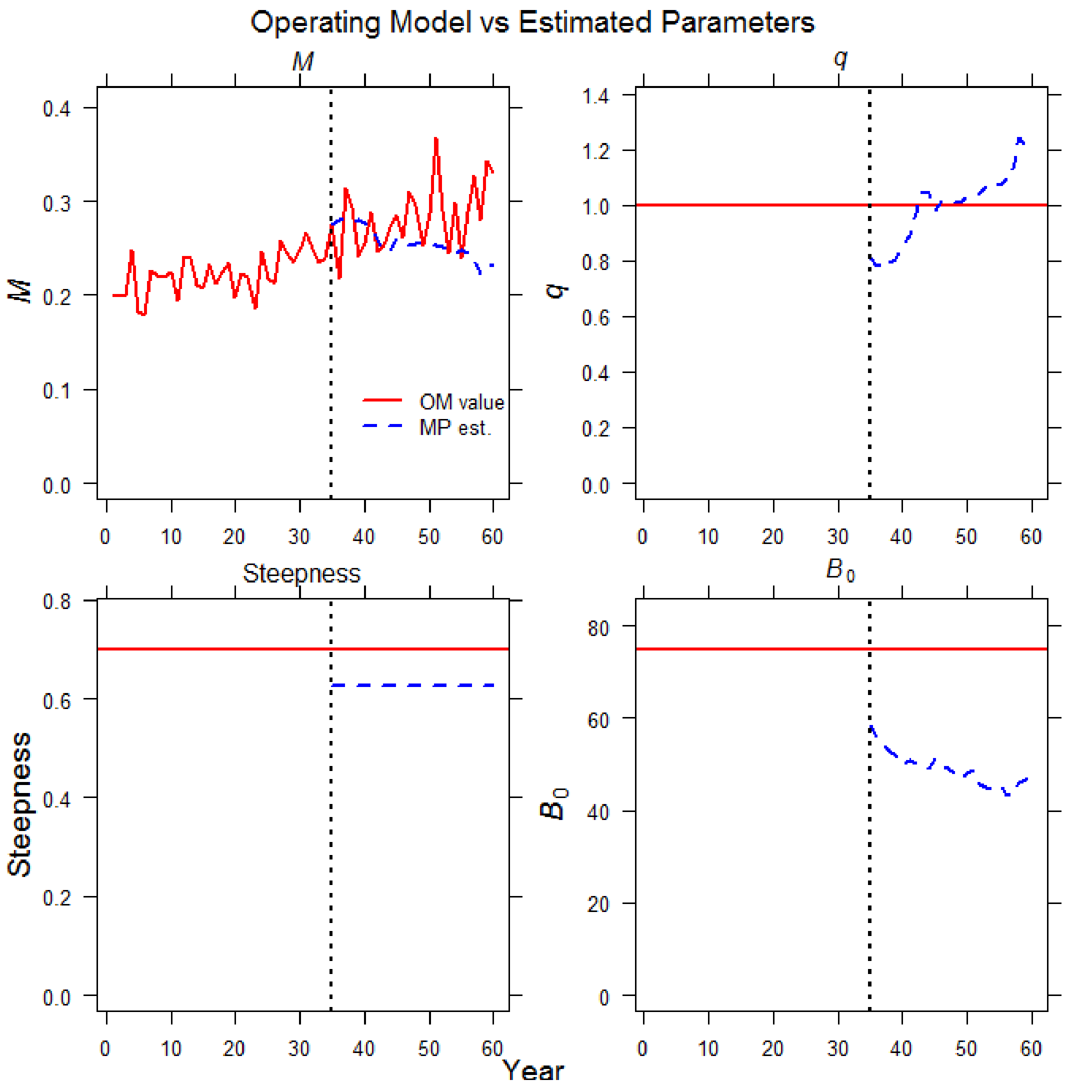

Figure 8. Operating model (OM) parameter values and parameters estimated from the assessment model (MP) for scenario S3, taken from a representative replicate in management strategy MS3. In scenario S3 natural mortality (*M*) was parameterized with a random walk and a positive trend. The assessment model bias tended to underestimate *M*, steepness, and $\hat{B}_0$, while survey catchability ($\hat{q}$) was estimated with a positive trend.

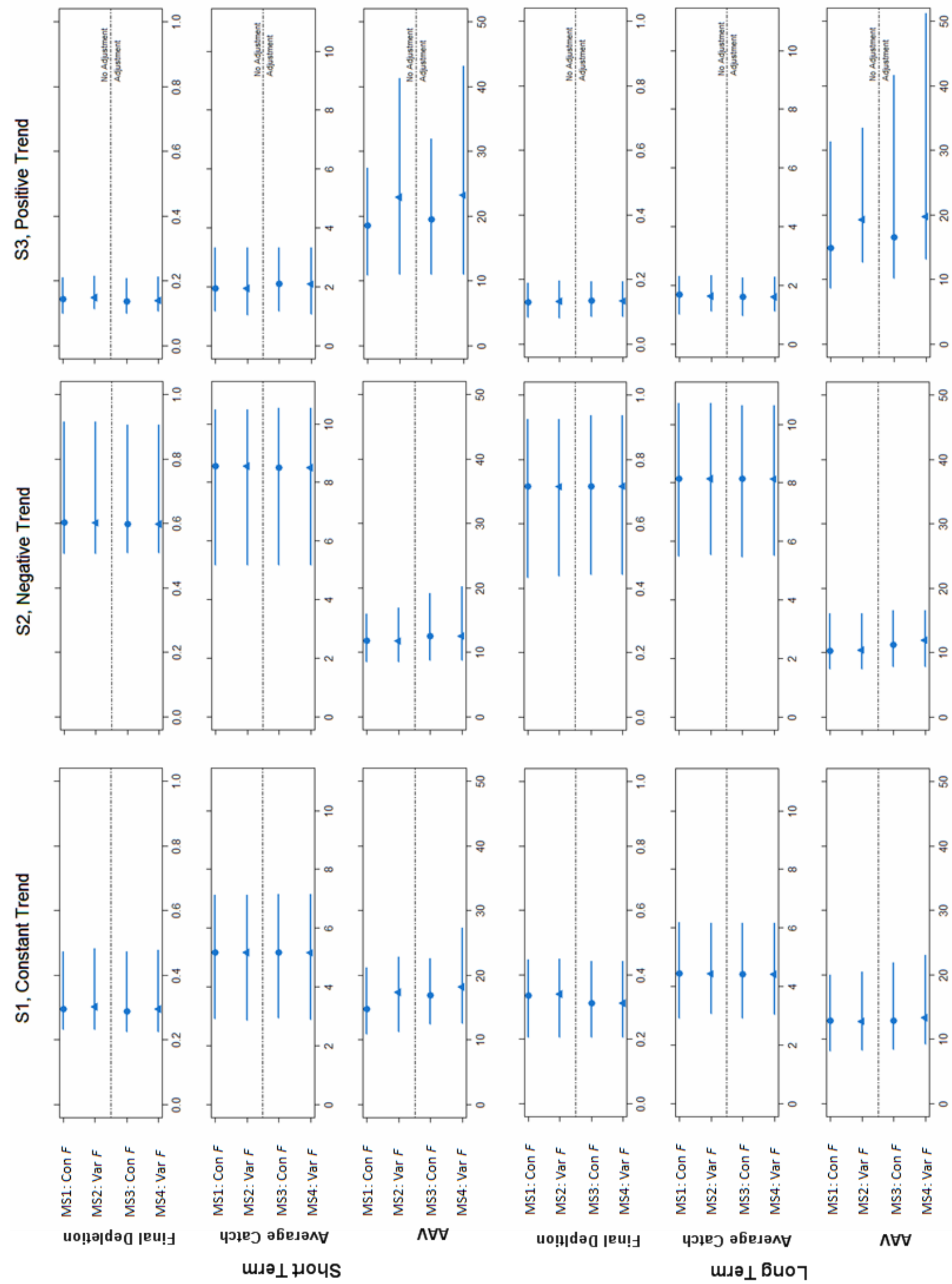

Figure 9. Performance of management strategies without (MS1, MS2) and with (MS3, MS4) retrospective adjustments in scenarios with a constant (S1), negative (S2), and positive (S3) trend in natural morality. Performance is measured by final depletion, average catch, and average absolute variation (AAV). Performance is measured over the short term, defined as year 40 through 49, and over the long term, defined as year 50 through 59. The median values for the constant *F* HCR are identified by circles, and triangles are used for variable *F* HCR.

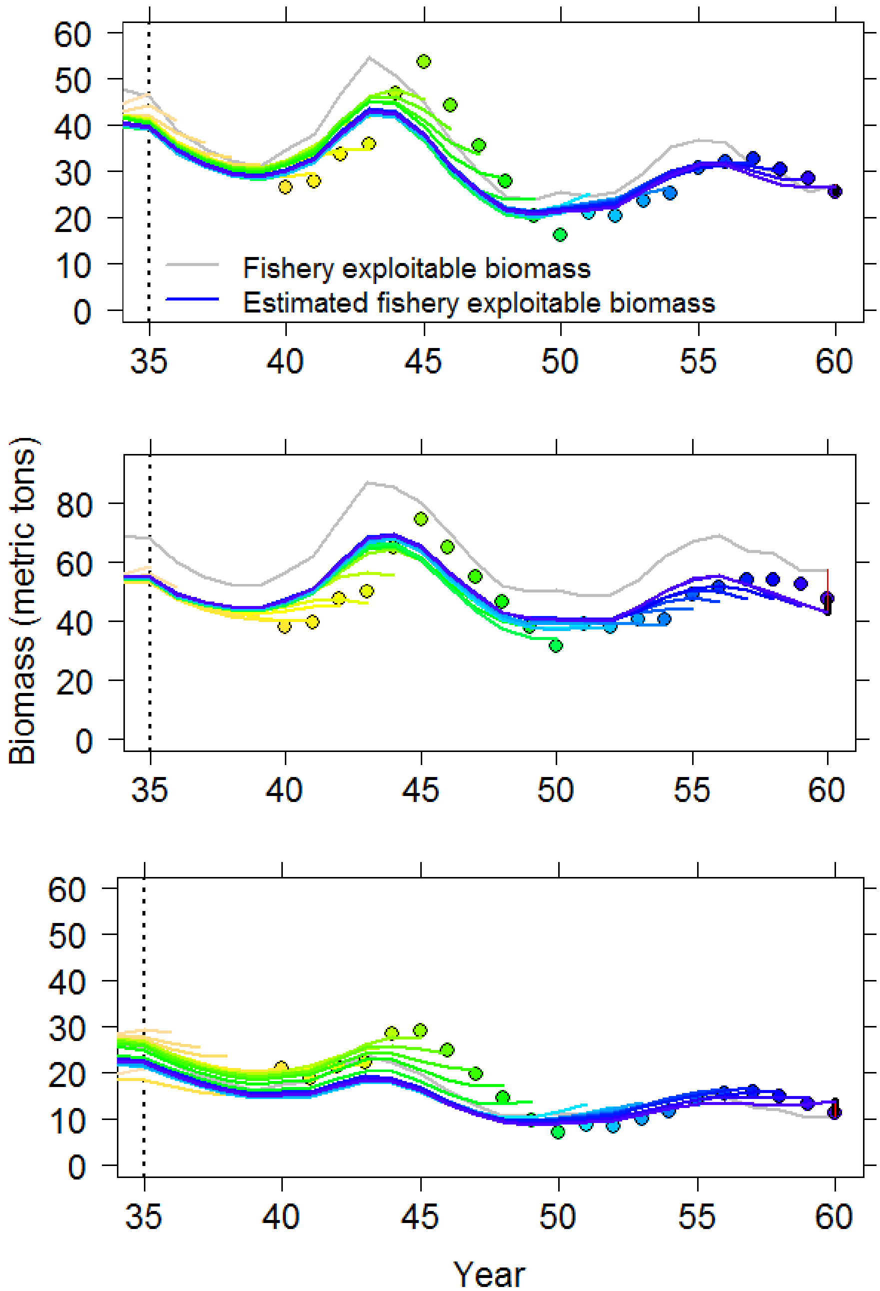

Figure 10. Detail of retrospective estimates and biomass adjustments for scenarios with a constant (S1, top), negative (S2), and positive (S3, bottom) trend in natural mortality. The adjusted biomass used in the harvest control rule is represented by the point estimates and correspond to the matching coloured assessment model biomass estimates. At year 60, the red line represents the true bias of a terminal estimate, and the black line represents the magnitude of the retrospective adjustment. Note the different y-axis scales. These assessments are taken from representative replicates in management strategy MS3.

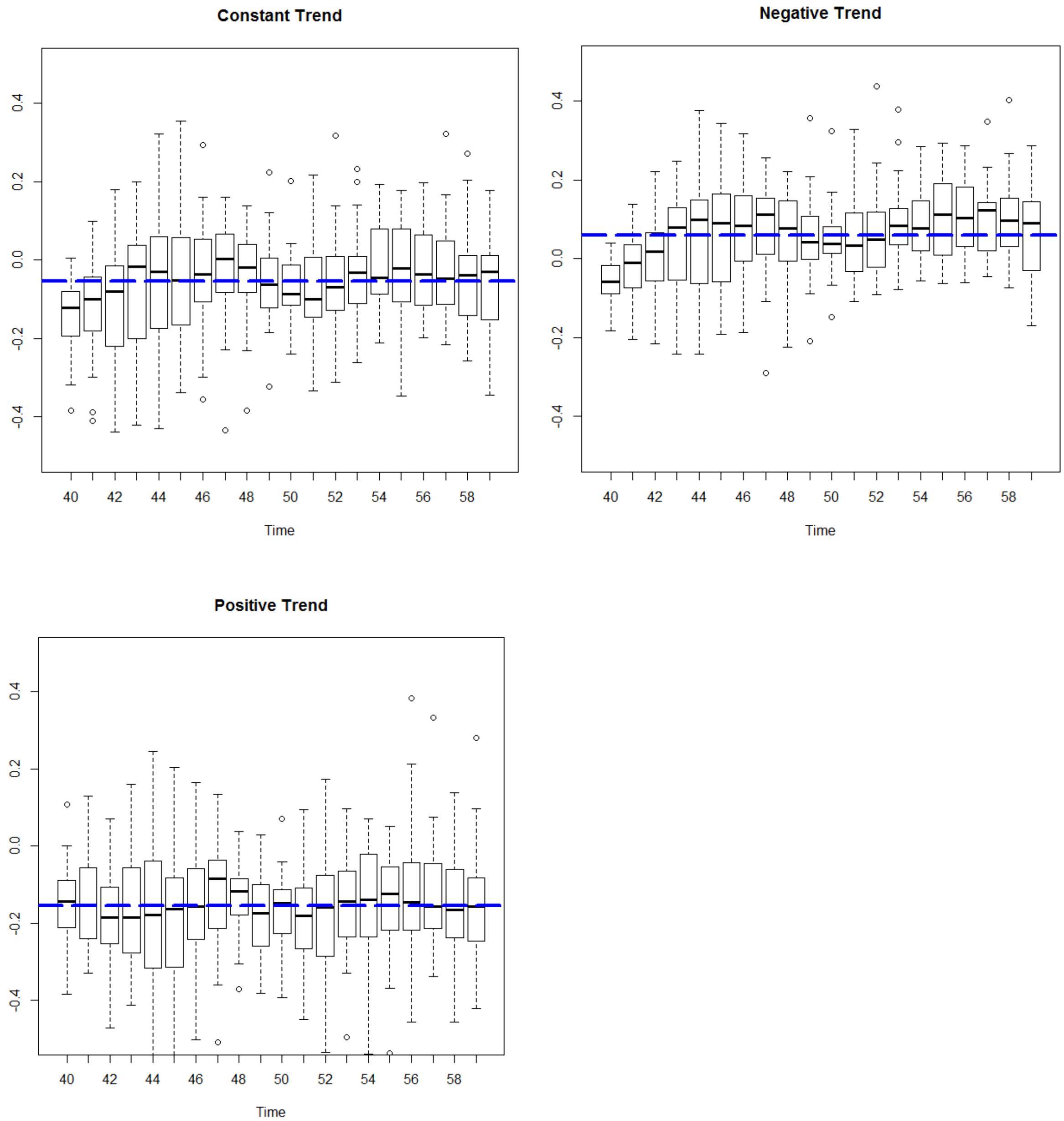

Figure 11. The difference between estimated biomass and the respective-adjusted biomass, relative to estimated biomass. The blue dashed line represents the median value across all replicates from year 40 though 59 for a management strategy with a variable *F* harvest control rule and a retrospective adjustment (MS4). A value less than 0 indicates biomass estimates have been reduced by the retrospective adjustment. A value greater than 0 indicates biomass estimates have been increased by the retrospective adjustment.

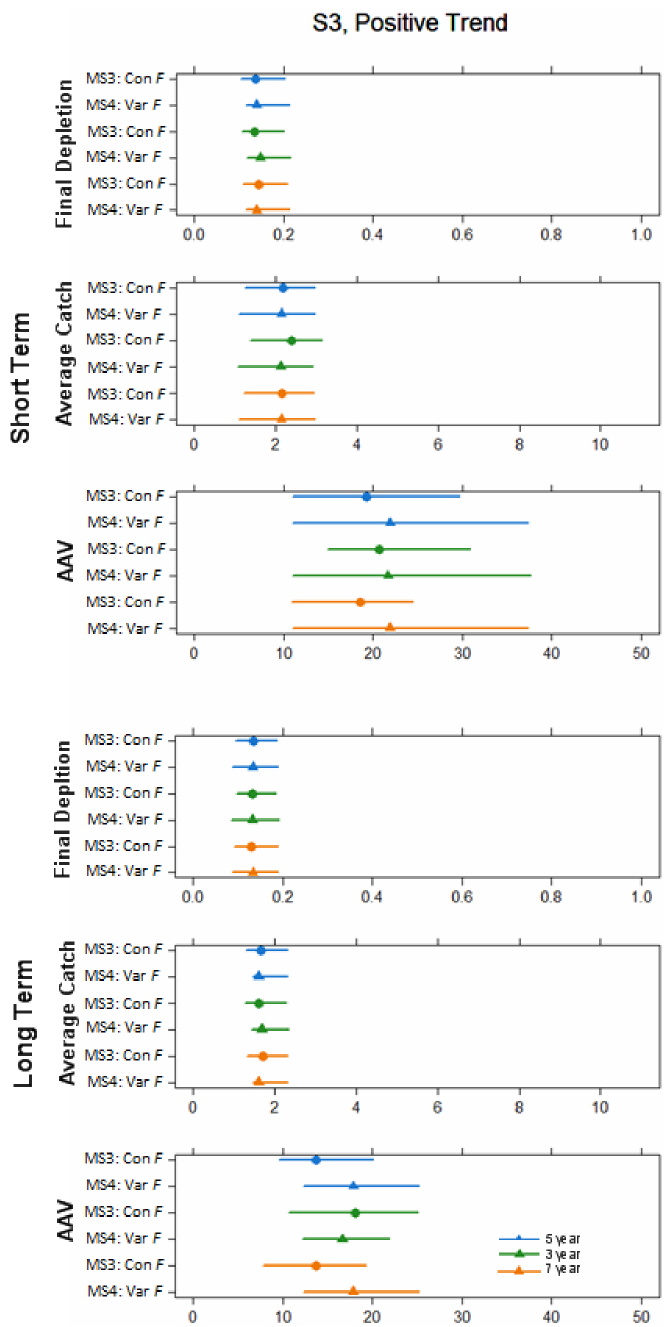

Figure 12. Management strategies' (MS3, MS4) performance sensitivity to the retrospective duration. Performance in the scenario with a positive trend in natural mortality (S3) is measured by final depletion, average catch, and average absolute variation (AAV). Performance is measured over the short term, defined as year 40 through 49, and over the long term, defined as year 50 through 59. The median values for the constant *F* HCR are identified by circles, and triangles are used for variable *F* HCR. Retrospective duration of 5 (top), 3, and 7 (bottom) years were used.

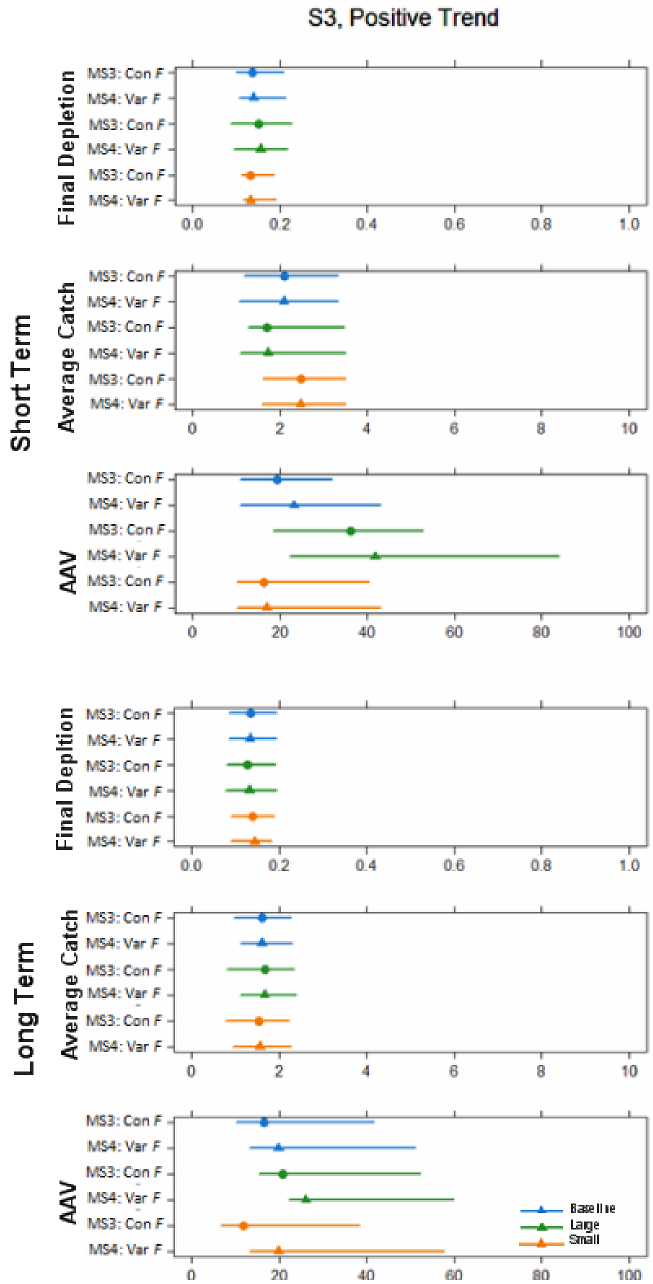

Figure 13. Management strategies' (MS3, MS4) performance sensitivity to the assumed level of observation error in the survey. Performance in the scenario with a positive trend in natural mortality (S3) is measured by final depletion, average catch, and average absolute variation (AAV). Performance is measured over the short term, defined as year 40 through 49, and over the long term, defined as year 50 through 59. The median values for the constant *F* HCR are identified by circles, and triangles are used for variable *F* HCR.

References

Beverton, R. J. H, S. J. Holt. 1957. *On the dynamics of exploited fish populations*. Vol. ser. 2, v. 19. London: H.M. Stationery Off.

Butterworth, D. S. 2007. Why a Management Procedure Approach? Some Positives and Negatives. ICES Journal of Marine Science 64 (4): 613–617.

Clark, W. G. 2003. A Model for the World: 80 Years of Model Development and Application at the International Pacific Halibut Commission. Natural Resource Modelling 16(4): 491–503.

Cox, S. P., and A. R. Kronlund. 2008. Practical Stakeholder-driven Harvest Policies for Groundfish Fisheries in British Columbia, Canada. Fisheries Research 94 (3): 224–237.

Fisheries and Oceans Canada. A Fishery Decision-Making Framework Incorporating the Precautionary Approach. Last modified: March 23, 2009. http://www.dfo-mpo.gc.ca/fm-gp/peches-fisheries/fish-ren-peche/sff-cpd/precaution-eng.htm

Hare, S.R. 2010. Potential modifications to the IPHC harvest policy. International Pacific Halibut Commission Report of Assessment and Research Activities 2010: 117-200.

Hutchings, J. A., and R. A. Myers. 1994. What Can Be Learned from the Collapse of a Renewable Resource? Atlantic Cod, *Gadus Morhua*, of Newfoundland and Labrador. Canadian Journal of Fisheries and Aquatic Sciences 51: 2126–2146.

Joint Technical Committee. 2013. Status of the Pacific Hake (Whiting) Stock in U.S. and Canadian Waters in 2013 International Joint Technical Committee for Pacific Hake. Prepared for the Joint U.S.-Canada Pacific hake treaty process. 190p.

Kronlund, A. R., S. P. Cox, and J. S. Cleary. 2012. Management Strategy Evaluation in R (*mseR*): user’s guide and simulation exercises. Canadian Technical Report of Fisheries and Aquatic Sciences 3001: vii + 52 p.

Legault C.M., Chair. 2009. Report of the Retrospective Working Group, January 14-16, 2008, Woods Hole, Massachusetts. US Department of Commerce, Northeast Fisheries Science Center Ref Doc. 09-01; 30 p.

Marasco, R. J., D. Goodman, C. B. Grimes, P. W. Lawson, A. E. Punt, and T. J. Quinn II. 2007. Ecosystem-Based Fisheries Management: Some Practical Suggestions. Canadian Journal of Fisheries and Aquatic Sciences 64 (6): 928–39.

Mohn, R. 1999. The Retrospective Problem in Sequential Population Analysis : An Investigation Using Cod Fishery and Simulated Data. ICES Journal of Marine Science 56 (4): 473–488.

National Research Council, Committee on Fish Stock Assessment Methods. 1998. *Improving Fish Stock Assessments*. Washington, DC: National Academy Press.

Northeast Fisheries Science Center. 2008 (a). Assessment of 19 Northeast Groundfish Stocks through 2007: Report of the 3rd Groundfish Assessment Review Meeting (GARM III), Northeast Fisheries Science Center, Woods Hole, Massachusetts, August 4-8, 2008. US Department of Commerce, Northeast Fisheries Science Center Ref Doc. 08-15; 884 p + xvii.

Northeast Fisheries Science Center. 2008 (b). Appendix to the Report of the 3rd Groundfish Assessment Review Meeting (GARM III): Assessment of 19 Northeast Groundfish Stocks through 2007, Northeast Fisheries Science Center, Woods Hole, Massachusetts, August 4-8, 2008. US Department of Commerce, Northeast Fisheries Science Center Ref Doc. 08-16; 1056 p.

Northeast Fisheries Science Center. 2012. Assessment or Data Updates of 13 Northeast Groundfish Stocks through 2010. US Department of Commerce, Northeast Fisheries Science Center Ref Doc.12-06; 789 p.

Parma, A. M. 1993. Retrospective Catch-at-Age Analysis of Pacific Halibut: Implications on Assessments Harvesting Policies. In Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations, 247–289.

Punt, A. E., and A. D. M. Smith. 1999. Harvest Strategy Evaluation for the Eastern Stock of Gemfish (Rexea Solandri). ICES Journal of Marine Science 56 (6): 860–875.

Punt, A. E., T. A'mar, N. A. Bond, D. S. Butterworth, C. L. de Moor, J. A. A. De Oliveira, M. A. Haltuch, A. B. Hollowed, and C. Szuwalski. 2013. Fisheries Management Under Climate and Environmental Uncertainty: Control Rules and Performance Simulation. ICES Journal of Marine Science 71 (8): 2208-2220.

Punt, A. E., D. S. Butterworth, C.L. de Moor, J. A. A. De Oliveira, and M. Haddon. 2014. Management Strategy Evaluation: Best Practices. Fish and Fisheries. doi:10.1111/faf.12104.

Rademeyer, R.A., É. A. Plagányi, and D. S. Butterworth. 2007. Tips and Tricks in Designing Management Procedures. ICES Journal of Marine Science 64 (4): 618–625.

Rivard, D., and M. G. Foy. 1987. An Analysis of Errors in Catch Projections for Canadian Atlantic Fish Stocks. Canadian Journal of Fisheries and Aquatic Sciences 44 (5): 967–981.

Showell, M.A., and M.C. Bourbonnais. 1994. Status of the Scotian shelf silver hake populations in 1993 with projections to 1995. Northwest Atlantic Fisheries Organization SCR Doc. 94/32.

Sinclair, A. F., D. Gascon, R. O'Boyle, D. Rivard, and S. Gavaris. 1991. Consistency of Some Northwest Atlantic Groundfish Stock Assessments. Northwest Atlantic Fisheries Organization Scientific Council Studies 16: 59-77.

Smith, A.D.M. 1993. Management Strategy Evaluation - The Light on The Hill. *In* Population Dynamics for Fisheries Management, ed. D.A. Hancock, 249–253. Canberra: Australian Government Publishing Service.

Stewart, I. J., and S. J. D. Martell. 2014. A Historical Review of Selectivity Approaches and Retrospective Patterns in the Pacific Halibut Stock Assessment. *Fisheries Research* 158 (October). Elsevier B.V.: 40–49.

Stone, H.H., and C.M. Legault. 2005. Stock Assessment of Georges Bank (5Zhjmn) Yellowtail Flounder for 2005. TRAC Reference Document - 2005/04. 83p.

Swain, D. P., H.P. Benoît, M.O. Hammill, G. McClelland, and É. Aubry. 2011. Alternative hypotheses for causes of the elevated natural mortality of cod (*Gadus morhua*) in the southern Gulf of St. Lawrence: the weight of evidence. DFO Canadian Science Advisory Secretariat Sec. Res. Doc. 2011/036. iv + 33 p.

Valero, J. L. 2011. Harvest Policy Considerations on Retrospective Bias and Biomass Projections. International Pacific Halibut Commission Report of Assessment and Research Activities 2011: 311-330.

Walters, C., and J. Maguire. 1996. Lessons for Stock Assessment from the Northern Cod Collapse. Reviews in Fish Biology and Fisheries 6 (2): 125–137.