Northwest Atlantic Fisheries Organization

Serial. No. N5225 **NAFO SCR Doc. 08/25**

SCIENTIFIC COUNCIL MEETING – June 2008

Management strategy evaluation for Greenland halibut (*Reinhardtius hippoglossoides*) in NAFO Subarea 2 and Divisions 3LKMNO

by

David C. M. Miller[1], Peter A. Shelton[1], Brian P. Healey[1], William B. Brodie[1], M. Joanne Morgan[1], Doug S. Butterworth[2], Ricardo Alpoim[3], Diana González[4], Fernando González[4], Carmen Fernandez[4], James Ianelli[5], Jean-Claude Mahé[6], Iago Mosqueira[7], Robert Scott[8] and Antonio Vazquez[9]

[1] Science Branch, Department of Fisheries and Oceans, Northwest Atlantic Fisheries Center, PO Box 5667, St John's, Newfoundland and Labrador, A1C 5X1, Canada
[2] Marine Resource Assessment and Management Group (MARAM), Department of Mathematics and Applied Mathematics, University of Cape Town, Rondebosch 7701, South Africa
[3] Instituto Nacional de Recursos Biológicos (INRB/IPIMAR), Av. de Brasilia, 1449-006 Lisbon, Portugal
[4] Instituto Español de Oceanografia, Cabo Estay, Canido, Vigo 36200, Spain
[5] Alaska Fish. Science Center, REFM Div., 7600 Sand Point Way NE, Seattle, WA 98115-6349, USA
[6] IFREMER, Station de Lorient, 8, Rue François Toullec, 56100 Lorient, France
[7] AZTI Tecnalia, Marine Research Division, Txatxarramendi ugartea z/g, 48395 Sukarrieta, Bizkaia, Spain (current address: Cefas, Lowestoft Laboratory, Pakefield Road, Lowestoft, Suffolk NR33 OHT, UK)
[8] Cefas, Lowestoft Laboratory, Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK
[9] Instituto de Investigaciones Marinas, Eduardo Cabello 6, 36208 Vigo, Spain

Abstract

A rebuilding plan for the 2+3KLMNO Greenland halibut stock developed by NAFO Fisheries Commission has been in effect since 2004. Under the plan, *ad hoc* TAC reduction steps were specified to 2007. The most recent assessment of this stock indicates that the rebuilding plan has been ineffective in initiating any recovery. Fishing mortality is still at high levels and spawner biomass has remained at very low levels. Management Strategy Evaluation provides a way of examining the performance of candidate management strategies with respect to rebuilding the stock. In particular, it allows the robustness of these strategies to be considered relative to alternative operating models of the "real world", for example the nature of the stock-recruit function for this stock. A preliminary MSE presented at the 2007 NAFO Scientific Council meeting has been updated following the NAFO SC Study Group on Rebuilding Strategies for Greenland Halibut meeting held in Vigo in February 2008. The analysis is carried out in FLR (Fisheries Libraries in R) environment, an open source framework for the development and evaluation of management strategies. A reference set of 20 possible operating models is presented, four of which are examined further. Results are presented in detail for five potential management strategies tested on these four operating models. A suite of performance statistics for this stock are used to assess these management strategies. The results of this MSE exercise are evaluated in the context of the NAFO approach to fisheries management and the potential for further progress with regard to the application of MSE on this stock in general is considered.

Key words: management strategy evaluation, operating model, fisheries management, management objectives, performance measures, stock-recruit, simulation, risk

Introduction

The distribution of Greenland halibut in the Northwest Atlantic is continuous along the continental slope within NAFO Subareas 0, 1, 2 and 3 (Bowering and Brodie, 1995). An argument could be made to treat this entire distributional range as a single unit for stock assessment and management purposes. This is strengthened by the theory that most of the spawning appears to occur in the Davis Strait area (Bowering and Brodie, 1995). Smaller Greenland halibut fisheries also occur in the Gulf of St. Lawrence, in fjords off West Greenland and in Cumberland Sound, and may be on separate stocks. This paper introduces management strategy evaluation (MSE) for the 2+3KLMNO management unit, but could be expanded to include Subarea 0 and 1 if these are considered to be part of the same management unit at some point in the future. Alternatively, management strategies could be evaluated for robustness to population fragmentation imposed by management unit boundaries.

The fishery for Greenland halibut in eastern Canada goes back to 1857 (Bowering and Brodie, 1995). Landings fluctuated between 250 and 1 000 t annually between 1916 and the early 1960s. The early fishery was conducted by longline and restricted to the deep channels in the bays. With the increase in factory freezer trawler activity in the offshore and the introduction of synthetic gillnets in the inshore, the nominal catches increased through the 1960s reaching a peak of about 37 000 t by 1969. Given that the stock was not under quota management and that much of the catch was processed at sea by factory freezer trawlers and factory mother ships, such as the Professor Baranov, capable of processing and freezing up to 100 tons of groundfish per day, the actual magnitude of fish dying from fishing operations must be assumed to be poorly known. It should also be noted that, while these nominal catches may seem modest compared to the equivalent data for northern cod over the same period, Greenland halibut is a slow growing, late-maturing, long-lived, deep-water species. Depending on the size of the stock and the actual catches, it is possible that fishing mortality was not sustainable during the late 1960s and early 1970s. This would be, consistent with the heavy overfishing occurring on other, better studied, groundfish stocks in the area at the time, such as cod, American plaice and yellowtail flounder. It is also clear that Greenland halibut was systematically overfished by gillnets in the deepwater bays around eastern Newfoundland in the late 1960s and early 1970s leading to local depletions. This fishery eventually moved offshore in the 1980s, first to the mid-shore deepwater channels in Divs. 2J and 3K, and eventually to the deepwater slopes in SA 2+3.

Greenland halibut stock in Subarea 2 and Div. 3KL first came under TAC management by Canada in 1974. The TAC in this portion of the stock area increased from 35 000 t in 1980 to 55 000 t in 1981-84, 75 000 t in 1985, and 100 000 t in 1986-89. These increases in TACs were the result of research vessel survey estimates of stock biomass (in excess of 400 000 t) which indicated presence of both high levels of fishable biomass as well as prospects of several better than average recruiting year-classes. The TACs were intended to apply to the entire stock area, and not just the portion in Canadian waters. After observing an estimated reduction in stock biomass from the late 1970s to the late 1980s in Subarea 2 and Div. 3KL of about 50%, the TAC was reduced to 50 000 t in 1990 and this level was maintained to 1993 despite further substantive declines in stock size throughout the normal range of observed historical stock distribution. The late 1980s-early 1990s coincided with a period of increased fishing effort in the regulatory area with the influx of about 40 Spanish factory freezer vessels displaced from Namibia. There was an increase in the violation of NAFO regulations over this period, including under-reporting and the use of illegal small mesh trawls. Although Scientific Council, in its deliberations during June 1993, could not advise on an appropriate catch level for 1994, the TAC was reduced to 25 000 tons by Canada in Subarea 2 and Divisions 3KL in consideration of low levels of stock size estimated for this area. It was intended that this TAC should include all catches in Subarea 2 and 3 for conservation purposes. Nevertheless, catches in the NAFO Regulatory area continued unregulated. In 1994, management of Greenland halibut in Subarea 2 and Div. 3KLMNO (note the change to include Div. 3MNO) became the responsibility of NAFO Fisheries Commission which imposed a TAC of 27 000 t for 1995. This level was maintained for 1996 and was proportioned throughout the management area in an attempt to reduce high concentrations of effort in localized areas. By 2003 the TAC had increased to 42 000 t.

The 2003 NAFO assessment (Darby *et al*., 2003) of Greenland halibut was a landmark event. There was a major downward revision relative to previous recent assessments which had indicated a growing stock. This downward revision was based on an additional year of data and a change in the XSA formulation. Instead of a growing stock, it estimated that the exploitable biomass had decreased to the lowest level in the recorded time series and that fishing mortality was increasing and was double the level of $F_{0.1}$ (Darby *et al*., 2003). Projections showed that a continuation of the catch at the prevailing level would rapidly collapse the stock, and SC advised a reduction in TAC to 16 000 t in 2004. NAFO Fisheries Commission responded by putting in place a fifteen year rebuilding plan with the objective of attaining a target of 140 000 tons exploitable (5+) biomass by the beginning of 2019. TAC levels were set for 2004-2007 at 20 kt, 19 kt, 18.5 kt and 16 kt, respectively. The intention was that subsequent TACs would depend on rebuilding progress, but with a 15% cap on any year-to-year change. Thus 2007 was the last year of specified TAC reductions and the TAC was kept at 16 kt in 2008. Although the TAC reduction steps were agreed to by FC, estimated catches have exceeded TACs up to about 40%, fishing mortality has remained high, and biomass remains at historic low levels (Healy and Mahé, 2006, 2007, 2008). The 2007 assessment found that average fishing mortality (ages 5-10) for 2006 was 0.59, over two times the F_{max} level (0.26) and four times the $F_{0.1}$ level (0.14). Violations of NAFO regulations have continued, including Canadian fishery officer reports of miss-identification of catch from the NAFO regulatory area as Hatton Bank Greenland halibut (i.e. from the Northeast Atlantic) and various other forms of under-reporting. Shelton (2005a, 2005b) criticised the Fisheries Commission rebuilding plan for having no scientific basis and for not being submitted for scientific peer review. Shelton (2005a) concluded that the plan was considerably less cautious than one which would be specified under a Precautionary Approach. In addition, he found that the rebuilding was not robust to retrospective error in estimates of recruitment nor was it robust to alternative assessment methods. Shelton (2005b) concluded that, in order to be compliant with the Precautionary Approach, fishing mortality should be immediately reduced to below $F_{0.1}$. A variable-F rebuilding plan was described and subject to preliminary simulation testing in which fishing mortality is initially set at $0.5\mathrm{x}F_{0.1}$, but increases to $F_{0.1}$ as the stock rebuilds. This strategy outperformed alternative constant-F strategies, such as F_{max}, $F_{0.1}$ and F_{sq}, in simulations.

The Precautionary Approach was adopted by NAFO in 2004 and in principle by Canada when she ratified the 1995 UN Fish Stocks Agreement in 1999, but implementation has lagged (Shelton, 2007). Reference points have been suggested for some stocks. For example, B_{lim} for Divs. 3LNO yellowtail flounder is considered to be 30% of B_{msy} and advice is provided on the basis of a $2/3F_{\mathrm{msy}}$ harvest control rule. Although not officially adopted by NAFO FC, this advice has been followed in the recent past and the stock could be considered to be sustainably managed at the present time. However, the yellowtail management strategy does not specify how fishing mortality will change should the stock begin to decline. For stocks for which there is an analytical assessment, NAFO Fisheries Commission makes the following standing requests for information from Scientific Council:

> *As general reference points, the implications of fishing at $F_{0.1}$ and $F_{current}$ in the next year and subsequent years should be evaluated. When spawner biomass reference points have been identified, short and medium term consequences should be identified for various exploitation rates (including no fishing) in terms of yield, stability in yield from year to year, and the risk or probability of maintaining the stock within, or moving it to, the Safe Zone. In order to consider the balance between risks and yield levels, each management strategy evaluation should provide risks and yields associated with various harvesting options in relation to B_{lim}, and F_{lim} and target F reference points selected by managers. Also, the present stock size and spawning stock size should be described in relation to those observed historically and those expected in the longer term under this range of options. Spawning stock biomass levels considered necessary for maintenance of sustained recruitment should be recommended for each stock. In those cases where present spawning stock size is a matter of scientific concern in relation to the continuing reproductive potential of the stock, management options should be offered that specifically respond to such concerns.*

The Fisheries Commission request for advice from Scientific Council outlined above is consistent with a PA approach, but what is lacking is the adoption of a predetermined management strategy comprising harvest control rules which can be demonstrated to have the desired performance, including the degree of risk-averseness required by the PA. Instead, FC makes management decisions on a largely *ad hoc* basis without clear logical links to the scientific advice that is provided. Although the current rebuilding plan for SA 2+3KLMNO Greenland halibut provides some structure, it is not science-based and was not simulation-tested to predetermine likely performance prior to implementation. Consequently, the present FC rebuilding strategy contradicts most of the principles for sustainable fisheries and the Precautionary Approach espoused above by NAFO. Fishing mortality continues to be high and spawner biomass continues to be extremely low under the rebuilding plan. We propose that in the future NAFO fisheries management should apply a science-based and simulation-tested prescribed harvest control rule based management strategy that can reasonably be expected to perform in such a way that the risk of the stock falling outside the Safe Zone is relatively low and, for a depleted stock such as Greenland halibut, the probability of rebuilding to the safe zone within a prescribed period of time is high.

Management strategy evaluation (MSE) is based on an approach of evaluating models through simulation before using them as a basis for decision-making in the real world. This approach gained increased prominence though the evaluation of management procedures by the International Whaling Commission in the 1980s and is described in Kirkwood and Smith (1996) and more recently in ICES by Kell *et al*. (2007) in the context of a new stock assessment environment in R-code called FLR (Fisheries Library in R; see www.flr-project.org). We work within the FLR environment in our MSE study of Greenland halibut. In 2007 a preliminary MSE for this stock was presented (Miller *et al.*, 2007). Based on these preliminary results NAFO SC struck a study group (NAFO Study Group for Rebuilding Strategies for Greenland halibut) to further develop this approach. A Wiki was established to garner input and the Study Group met in Vigo in February 2008 to review progress and to suggest further development. This meeting was attended by scientists from NAFO member countries, fishing industry representatives from Canada and the EU, fisheries managers and invited independent experts.

Based on the input prior to the study group meeting obtained through the Wiki and suggested changes and improvements made at the Study Group meeting in Vigo (NAFO SC, 2008) , the management strategy evaluation reported in Miller *et al.* (2007) has been substantially revised and improved. The present research document updates this progress. Further, relatively minor revision is anticipated with a view to applying the management strategy evaluation in provision of scientific advice on rebuilding options made to Fisheries Commission by Scientific Council in 2009.

Conceptual framework

The conceptual framework for Management Strategy Evaluation (MSE), adapted from Kell *et al.* (2007), comprises an “operating model” and a “management strategy/procedure” (Fig. 1). In this approach the operating model (OM) is constructed to simulate the fish stock and the fishery, and is conditioned on the available data in order to be a realistic representation. The operating model represents the “true” system and incorporates biological processes that make up the stock dynamics and fishery processes that result in the capture of fish. Conditioning of the operating model requires the estimation of parameters consistent with the data and hypotheses about how these were generated. These govern processes such as recruitment, growth, maturation and mortality with respect to stock dynamics, and selectivity at age with respect to the fishery. A perception of the operating model is generated by sampling “observed data”, with error, from the operating model. An assessment model can be fitted to this observed data. The performance of a management strategy (MS) can be evaluated by applying it to this perception of the “Real World” – either directly to the observed data or to model estimates based on the observed data. Performance Statistics (PSs) evaluate how well a particular management strategy is performing relative to other candidate strategies across a range of conservation and fishery related performance measures when applied to the simulated “real” stock. Application of the MS is repeated many times over some management time horizon to sample from various sources of uncertainty. The alternative MSs are evaluated through the generation of distributions for the PSs. Strategies that are robust to the uncertainty in terms of being risk-averse with respect to falling into NAFO PA Zones 2-5 (NAFO SC, 2003), i.e. which keep the stock in the “Safe Zone” with high probability, would be favoured under a PA approach. Note that specific PA reference points have not yet been accepted by SC for this stock. Given that there may be a number of major hypotheses regarding the biology of the stock, for example, the appropriate recruitment function, a reference set of operating models is required that capture the full range of hypotheses, rather than only a single operating model.

The key elements of the MSE approach (Smith *et al.*, 1999) include specifying management objectives, performance measures, and alternative management strategies. This is followed by simulation evaluation of alternative management strategy performance statistics, and presentation of results to decision makers.

Guidance on management objectives is necessary before any evaluation of potential management strategies can be undertaken. General objectives, common to most fisheries stocks include a low risk of depletion of the stock while maintaining a reasonably high average annual catch (yield) and maximizing the stability of catches year to year. Clearly this will involve an evaluation of trade-offs. It is also preferable to have a management strategy that is robust to uncertainties in the dynamics of the population. With regards to the NAFO PA framework, a good management strategy should have a high probability of maintaining the stock within, or moving it to, the Safe Zone (i.e. low probability of $F>F_{\mathrm{lim}}$ or $B<B_{\mathrm{lim}}$). These objectives can be encompassed in the form of performance statistics. Performance statistics need to be quantifiable measures that can be used to evaluate how well a particular

management strategy is performing, relative to other strategies. They should relate easily to the fishery and be meaningful to stakeholders and managers. They need to address both fishery related objectives and those that are stock-conservation related. In some cases tolerances can be predefined such as an unacceptable risk of falling outside a pre-specified limit reference point value. Other performance statistics may be more useful in evaluating the inevitable tradeoffs between measures such as mean annual catch and variation in annual catch. Trade-offs need only be evaluated for MSs that provide performance statistics that fall within specified risk tolerances.

To be consistent with the NAFO PA framework, management strategies under consideration should be suitably risk-averse and involve a decrease in *F* with decreasing biomass. The past management of this stock has resulted in the exact opposite outcome – fishing mortality has tended to increase with decreasing stock size. Risk-prone management of this kind is contrary to the PA. In addition to stock conservation concerns, in order to be practical for the fisheries industry, strategies that do not lead to large fluctuations in catch and *F* from year to year are preferred. Management strategies need not depend only on estimates from models fit to available data. Model-free strategies, using the index survey data directly can also be applied. Experience in other fisheries where MSE has been applied has indicated that model-free management strategies may have a higher degree of acceptability to fisheries managers than those based on model estimates while performing as well or better than model-based strategies (Butterworth, in press).

To evaluate alternative MSs, they are applied to data generated from a range of operating models constructed to simulate the fish stock and the fishery. Each operating model is conditioned on available data and various hypotheses regarding the true dynamics of the stock. The operating model represents the Real World over the duration of the simulation and management strategies are applied based on data taken from this simulated Real World. A lot of uncertainty usually exists around the dynamics of fish stocks, so instead of a single representation of 'reality" it is advisable to consider many options encompassing most of the possibilities to deal with this uncertainty. A group of operating models each conditioned on different data or based on an alternative hypothesis of the stock dynamics or future trends in the fishery, is referred to as a reference set of operating models. In order to carry out a full management strategy evaluation, one needs to examine each strategy under each operating model in the reference set. It is recommended that in the order of 100 simulations are run for each management strategy for each operating model (Rademeyer *et al*. 2007).

Conditioning of operating models requires consideration of the past system and initial starting point of the population, biological parameters of the stock (stock-recruit curve, maturity and weight/growth), behavior of the fishery/fleet(s) and the level of uncertainty/error in the observation of the system and estimation of stock parameters.

Deterministic (i.e. no error) simulations are useful for understanding the basic behaviour of each OM. However, to fully evaluate a management strategy, tests need to be carried out over a number of stochastic simulations. Stochastic MSE simulations can be run under different scenarios of uncertainty/error relating to the "Real World" stock and the perception of it. We have adopted the three step "POM" approach to management strategy evaluation described in the ICES COMFIE Report (ICES, 1997). P = Process error (this captures variation in the "Real World" e.g. variation in growth, maturation, recruitment, mortality, selectivity etc.); O = Observation error in the perception of the Real World such as the survey tuning indices (note: at this point no Observation error in commercial catch is considered); M = Model error (error associated with the XSA estimates, or any other stock assessment model that may be used, of population size and fishing mortality that would occur even if the model were given perfect data about the Real World). In Fig. 1, P-level analysis is considered by including Process error but applying the management strategy to the fishery as if there were perfect information (i.e. no Observation error and no Model estimation error). PO-level analysis includes Process error and Observation error, but no Model error. Indices of stock abundance are created by using index residuals from the initial XSA, used to create the population, to add error to the actual population numbers. POM-level analysis includes all three sources of error. The indices from the population (with Observation error) are used together with catch information in an XSA (Model error) to obtain a perceived view of the true population. Using a model such as XSA cancels out observation error to an extent by 'smoothing' over the error from each of the three indices used. However, further error can result from model biases and the assumptions used in the model (e.g. *M*, shrinkage of *F* in the final years etc.).

Management strategies that perform poorly under P, or PO levels of error are unlikely to perform well under the full POM error conditions. We initially evaluated the harvesting strategies sequentially, under P, PO and the full POM structures. Although this step-wise introduction of error was found to be useful, in practice, running prospective

MSs under the full POM error was found to be more expedient. The ultimate test of an MS's performance is whether or not it is successful, in terms of achieving the management objectives as measured by performance statistics, despite uncertainty caused by process, observation and model error (i.e. simulated "real world" conditions).

Management implementation error, including TAC-overruns, is a serious problem with respect to many stocks, including 2+3KLMNO Greenland halibut. It may be considered necessary to incorporate some form of implementation error in the MSE simulation process. This could be done by adding a proportional overrun. However, TAC overrun is likely to vary depending on the availability of fish to the fishery and the level of TAC. Hence, predicting the likely implementation error is a difficult undertaking. In addition to this, incorporating implementation error makes it difficult to determine the true success of a management strategy – is the failure of a strategy due to inherent shortcomings or to the non-compliance in the fishery? Implementation error could be considered to be more a problem for managers to deal with. For example, if managers anticipate a certain level of TAC overrun, then it would be logical to allocate lower TACs than those determined by the MS in anticipation that a certain overrun would occur.

Results of MSE simulations need to be presented in a clear and concise manner that managers/decision makers/stakeholders can easily interpret the outcome. When there is a large amount of uncertainty, the reference set of operating models can easily become excessive. It is important that an MSE is well planned so that only the most likely and influential sources of uncertainty are fully explored to keep the reference set manageable. The reference set may be reduced by examining deterministic (error-free runs) and eliminating OMs which essentially duplicate the behaviour of other OMs. Expert opinion may also be required to weight the various scenarios in order to determine the overall performance of each management strategy in the various models within the reference set. Alternatively, it may be best to rather identify which MSs satisfy pre-specified performance criteria and present a set of potential management strategies to fisheries managers from which to choose. The robustness of management strategies to uncertainty, as well as their ability to achieve key management objectives while maintaining a healthy stock, needs to be presented in clear understandable manner so that the appropriate management strategy can be selected.

MSE Application to Greenland halibut in NAFO Subarea 2 and Divisions 3LKMNO

The current application of the Greenland halibut MSE is based on the structure from the initial MSE presented at the 2007 NAFO SC meeting (Miller *et al.*, 2007). It has been updated substantially and includes a number of the recommendations made through a Wiki website and from the NAFO SG on Rebuilding Strategies of Greenland halibut which met in Vigo in February 2008 (NAFO SC, 2008), in particular:

1. Further investigation of operating models is required – particularly with regards to conditioning. It was recommended that commercial CPUE be incorporated in the conditioning of one or more operating models (to create a historical and current version of the stock more in line with industry perceptions).
2. An asymptotic PR scenario for ages 14+ should be considered in addition to dome-shaped scenarios.
3. While Management strategies investigated were generally acceptable, some minor modifications could be explored. In particular it was expressed by industry to the Study Group that stability in TACs is important and that annual adjustments should be as small as possible.

Following the conceptual framework given above, the current MSE application for the Greenland halibut stock is now described in more detail. The Greenland halibut MSE takes into account historical uncertainty in the form of Observation error through an XSA bootstrap procedure (Miller and Shelton, 2007), giving a distribution of initial population sizes and age compositions. Process error (variation in weights at age, proportions mature at age, partial recruitment at age and number of recruits) was accounted for by running the simulation for each management strategy out to year 2030 a total of 100 times. No management implementation error (i.e. TAC over/under-runs) was considered (i.e. if the strategy indicates a TAC of 30,000 t then this value is applied in the evaluation). It is thus assumed that managers will deduct expected TAC overruns from proposed TACs derived from application of a particular harvest control rule, based on recent TAC overrun information, before allocating quotas.

Management objectives

The immediate management objectives for 2+3KLMNO Greenland halibut are with respect to the rebuilding plan laid out by NAFO Fisheries Commission in 2003 (NAFO FC, 2003): to rebuild the stock to 140 000t exploitable (5+) biomass (the mean biomass from 1975-1999) by the end of 2018 while minimising the annual reduction in catch (no more than 15% TAC change from year to year). It is assumed that rebuilding to B_{msy} represents an unstated long-term objective once the "milestone" target of 140kt (or equivalent estimate) is achieved. Clearly tradeoffs between catch and rate of rebuilding to the FC target and B_{msy} will be a consideration. In order to allow for B_{msy} as an objective, the time horizon of the MSE is extended 2030.

Performance statistics

The results of the stochastic MSE simulations can be divided into two broad categories: descriptive statistics and performance measures. The former includes results that are produced to develop an understanding of the dynamics of the simulated stock and the fishery. The latter are measures that can be used to assess the relative merits of candidate MSs, either how they perform with regards to critical performance criteria or what trade-offs in performance they represent.

In the current MSE, descriptive statistics presented are annual exploitable (5+) biomass, spawner stock biomass, recruitment (age 1), mean F (ages 5-10), annual catch and the mean exploitable age of the stock (weighted mean age of all fish five and older). These statistics are presented by OM as time series plots for each MS.

The performance statistics (PSs) suggested include those presented at the Vigo SG meeting, which were found to be generally acceptable by industry and managers, and a few additional long term measures to account for NAFO obligations under the 1995 UN Fish Stocks Agreement. These have been divided up into industry and conservation focused measures, pertaining to three different reference years – beginning of 2010, 2019 and 2030 i.e. short term, rebuilding plan duration and long term.

Industry-focused statistics (calculated on the annual values from the start of the MS implementation up to the beginning of the reference year):

1. Avg. Catch – the average annual catch/yield.
2. Catch/MSY – the average annual catch as a fraction of MSY.
3. AAV in Catch – the average absolute annual variation in TAC (%):

$$AAV_y = 100 \times \frac{\sum_{i=y'}^{y} \left| 1 - \frac{C_i}{C_{i-1}} \right|}{y - (y' - 1)} \tag{1}$$

Where: C_y = commercial catch in year y
y' = the first year when the catch was determined by the MS (2008)

4. CV on F – the coefficient of variation in F.
5. Mean Age of Catch – weighted mean age in the catch.

Conservation-focused statistics:

6. Recov. Ratio – the ratio of exploitable (5+) biomass to the mean exploitable biomass from 1975-1999 (i.e. the rebuilding plan target biomass)
7. B/B_{msy} – biomass as a fraction of B_{msy}.
8. F/F_{msy} – fishing mortality as a fraction of F_{msy}.

These statistics, with the exception of Mean Age of Catch, are presented as box and whisker plots (medians and percentile ranges of replicates).

Should PA reference points be developed and accepted for this stock by SC, then the probability of falling outside these values could also be included as PA-relevant performance measures. However, B as a fraction of B_{msy} and F as a fraction of F_{msy} do provide indications of overfishing that can be interpreted in a PA context.

Alternative management strategies

All of the strategies considered in this MSE set Total Allowable Catches (TACs) for year y+1, in year y, using data up to year y-1 (as is done in practice). To implement strategies which are F-based, the stock is projected forward to the beginning of year y+1 using available data and the TAC set the previous year. The F value is then converted to a TAC using the projected numbers at age at year $y+1$ and three-year geometric means of commercial selectivity (PR) and weight at age. For both deterministic and stochastic simulations, the TAC is caught exactly, unless there is not enough exploitable biomass to support such a TAC, in which case an F_{cap} value of approximately 1.73 is applied (this equates to approximately 85% of available fish being caught).

Four constant catch scenarios (0, 8k, 16k and 32k t) were used in initial deterministic runs to evaluate how the various OMs behave under different levels of fishing pressure. Constant catch strategies are not considered as potential candidates for management of the fishery and were only applied for diagnostic purposes.

Five management strategies were considered in stochastic analyses:

(i) F status quo strategy (F_{sq})

F-based strategy. The stock is fished at the same fishing mortality as in the previous year. i.e. in each year y, F from the previous year, F_{y-1}, is converted to a TAC for year y+1, based on stock projections to the start of year y+1. This is recalculated each year, so F will vary. Given the current high level of F, this is a heavy fishing strategy.

(ii) Precautionary Approach strategy (PA)

F-based strategy. This is a simplified PA implementation based on the breakpoint in segmented regression as a reference point. In this case, the value of F is determined depending on how current SSB relates to *beta* (β), the breakpoint in the segmented regression curve (Fig. 2):

if SSB> β then F=$F_{0.1}$, else if SSB<0.5* β then F=0.5*$F_{0.1}$, else F=(SSB/ β)*$F_{0.1}$

(iii) Model-free, index-based TAC adjustment strategy (ModFree)

TAC-based strategy. A simple TAC adjustment strategy that uses the change in perceived status of the stock (from research surveys) to adjust the TAC accordingly:

$$\mathrm{TAC_y} = \mathrm{TAC_{y\text{-}1}} \times (1+\lambda \times slope) \tag{2}$$

Where: *slope* = average slope of log-linear regression lines fit to the last five years of each index (equally weighted)
λ = an adjustment variable to ensure that the relative change in TAC is greater than the perceived relative decline in stock size (i.e. $\lambda > 1$, therefore allowing the strategy to halt the decline of the stock size through positive feedback).

Various λ values >1 were tested in deterministic simulations and a value of 1.25 was selected (allows for adequate adjustment of the TAC in the case of a declining stock without having excessively large fluctuations from year to year). In addition to this, a constraint was made limiting the new TAC to a minimum of 25% of the previous TAC (to prevent setting negative TACs in the case of extremely steep stock declines). A more risk adverse strategy could be achieved by different values of λ depending on the value of the *slope*. A value greater than 1 may be required in the case of a perceived decline in stock size (*slope*<0) but this could hamper stock recovery in the case of a perceived increase in stock size (*slope*>0). Using λ<1 when *slope*>0 could allow for better recovery of the stock, though this would need to be examined further.

(iv) Fisheries Commission Rebuilding Plan Model-based TAC adjustment strategy (FCMod)

TAC-based strategy. This strategy was designed to comply as closely as possible with the constraints laid out by the FC rebuilding plan i.e. stability for fishery is considered important therefore no large TAC changes are allowed. The basic strategy is the same as the model-free strategy except this is a model-based strategy where:

a) *slope* is the slope of log-linear regression line fit to the last five years of exploitable (5+) biomass according to the latest XSA assessment (years y-4 to y-1 from the XSA and year y projected based on the previous years TAC – done automatically in the XSA).
b) $\lambda = 1.5$
c) The TAC from 2008 onwards shall not be set at levels beyond 15% less or greater than the TAC of the preceding year.
d) TACs are only changed every second year.

Note that, while this strategy attempts to address some of the aspects of the FC rebuilding plan, the FC plan specifies arbitrary *ad hoc* TAC reduction steps and does not specify a feedback harvest control rule of the kind explored here.

(v) Half $F_{0.1}$ strategy (HalfF0.1)

This strategy was added because of the poor rebuilding prospects experienced under the other strategies. Under this fixed F strategy, fishing mortality is immediately reduced to the $0.5^*F_{0.1}$ and retained at this level. Preliminary checks showed that in certain simulations using this MS the XSA had difficulty fitting to the available data. In some years catch would drop to very near 0kt because of very low F. XSA is incapable of dealing with years of 0kt or close to 0kt catch. Hence, if the final year (y-1) in the catch time series gets too low, XSA predicts a near zero population in year y. This obviously leads to a near 0kt TAC being set for year y+1. This in turn leads to the XSA yielding a near zero population in the year y+2 (when data up to year y+1 are used). Since a non-zero TAC would have been set for year y in year y-1, a reasonable estimate of stock size would be made for year y+1 (when data up to year y is used). This causes the perceived view of the population from the XSA to alternate between normal and extremely low population numbers from that point onwards. This problem could potentially be solved by setting a minimum catch level or provide a rule in the MS for how to deal with cases when the catch level is too low to fit an XSA to the data. For the current experiment, simulations in which this error occurred were replaced with ones in which it didn't. This may potentially lead to certain biases, but these are not believed to be sufficient to disregard the results.

A number of other 'rebuilding plan' strategies were also considered. These were designed to get the population to reach the rebuilding plan target either by a specific time or in equilibrium. However, technical difficulties with the versions of R and FLR libraries being run have hampered the successful implementation of these strategies thus far. The future use of such strategies is discussed later.

Operating models

There is substantial uncertainty around the dynamics and current state of the 2+3KLMNO Greenland halibut stock. To handle this uncertainty, we specified a reference set of 20 operating models (OMs) covering a broad range of possible "realities". Thus far, we have only analyzed the performance of MSs against four of these OMs in the MSE. Descriptive results (biological reference points) for the other OMs are presented in summary form and deterministic projection results are available from the authors on request. The four OMs to which the MSE is applied were selected on the basis that each represented only one change from a base OM conditioned on the 2007 assessment of the stock. While it may be possible to reduce the reference set to less than 20 OMs, this needs to be done in an objective manner, for example on the basis that two OMs are essentially equivalent and perform in an almost identical manner when the same management strategy is applied. Given that only four OMs were selected in the current analysis, the results cannot be considered an adequate evaluation of possible management strategies for Greenland halibut management at this stage and considerable further work is required. However, the present study lays the basis for an operational MSE should the desire to adopt a predetermined MS approach encompassing harvest control rules be expressed by FC, and further stochastic runs under different OMs should be reasonably straightforward, albeit time consuming.

Conditioning of Operating Models (creating the reference set)

The full reference set of 20 operating models is presented in Table 1. These operating models are distinguished by: starting point (historical numbers at age arising from the indices chosen for the assessment), stock-recruit function, M and the shape in commercial PR (selectivity) after age 13. These aspects are described in more detail below.

Tuning indices

The assessment of 2+3KLMNO Greenland halibut is based on XSA. Three research vessel survey series of age disaggregated abundance indices (mean numbers per tow, MNPT) are used to tune the XSA:

1. EU 3M - a European Union summer survey in Div. 3M from 1995-2006 ages 1-12 (González Troncoso *et al.*, 2006).
2. Can 2J+3K autumn survey, Campelen trawl data from 1996-2006, ages 1 to 13 (Healey, 2007).
3. Can 3LNO spring survey, Campelen trawl data from 1996-2005 (no 2006 survey), ages 1 to 8 (Healey, 2007).

These were used in the XSA assessment performed for the stock in 2007.

The Vigo Study Group meeting recommended that commercial CPUE indices should also be considered in tuning the XSA under the assumption that this would create a 'healthier' and more productive stock, in line with industry views. Consequently, two commercial CPUE indices were developed using Canadian and Spanish commercial fishery data (Fig. 3). Because XSA only uses age-disaggregated indices the Canadian Otter Trawl CPUE index (1990-2006) and Spanish Commercial CPUE index (1992-2006) were age-disaggregated using proportions at age (4-13) from the catch data:

$$I_{a,y}^{CPUE} = CPUE_y \times \frac{C_{a,y}}{C_y} \tag{3}$$

These indices were then used together with the Canadian Fall RV index (1996-2006, ages 1-11 – older ages of this index have large residuals in the current XSA and were therefore left out to let the CPUE indices have a greater effect on the estimation of numbers at these ages) to calibrate the XSA and thereby create an alternative view of the current state of the resource that is more consistent with industry perceptions i.e. slightly higher exploitable biomass, and higher levels of recruitment in recent years (stronger cohorts) resulting in a more productive stock. While this method created higher levels of recruitment in recent years as well as lower levels of *F*, the increase in the estimation of exploitable biomass was minor and SSB actually is perceived to be lower (Fig. 4).

To account for uncertainty around model fitting we have randomly resampled (bootstrapped) residuals with replacement from the 'best fit' XSA (within age and index), generating new pseudo-abundance indices to which the XSA was refitted for each individual simulation. This method is fully described in Miller and Shelton (2007).

Biological inputs to the operating model

Age structure

The current XSA structure used to assess this stock includes ages 1 to 13 with a 14+ plus age group. Given that Greenland halibut mature at an old age (>10) and are slow growing, it is likely that they live well beyond age 14. Fish have been caught that are older than 25 years but the apparent poor selectivity of fishery and survey gears for older fish means that an accurate estimate of the maximum age of Greenland halibut cannot be obtained. In addition, it is difficult to get reasonable weight at age data for any ages older than about age 18. Given that almost all of the fish are mature by age 20, it is considered that creating a "true" population in the OM that is age disaggregated up to age 20 should adequately capture the actual age structure of the stock (age 20+, a plus group).

XSA methodology does not include a dynamic plus group. Numbers in the plus group are predicted based on the annual catch in the plus group and the assumption that *F*-at-age for the plus group is the same as the *F*-at-age for the last age before the plus group. This method was considered unreasonable for the Greenland halibut stock given that survivorship to the plus group is high under reduced fishing mortality, and was replaced by an *F*-based dynamic pool method. The plus group numbers for each year were then expanded out to age 20 (a plus group) based on assumption about the PR (selectivity to the commercial fishery) for the older ages (see below).

Growth

For years 1975 to 2006, weight at age data (up to age 13) were taken from the inputs to the XSA (based on commercial catch data) in Healey and Mahé (2007). Some length data for fish older than age 13 are available from research surveys. These were converted to weight using the length-weight relationship in Gundersen and Brodie (1999). Incorporating these data, means for ages and simple rates of change from one age to the next, weight at age data for each year was extended to age 20.

Weights for projected years were resampled by year (all ages), from the period 1987 to 2006, because of a notable change in weights post-1986.

Maturation

Parameter estimates for the slope (age effect) and intercept from models of maturity at age by cohort were taken for 1966-1994 cohorts from Morgan and Rideout (2007) and used to produce estimates of proportion mature at age (μ):

$$\mu = \left(\frac{1}{1+\exp(-\eta)} \right) = proportion\,mature \tag{4}$$

where: $\eta = \tau + \gamma A$, τ is an intercept, γ age effect, A is age.

The maturity-at-age matrix for the true population was constructed by using the equation above for all of the cohorts. For the 1966 to 1994 cohorts, parameters (slope and intercept) were taken from the fits of the model to actual data from each cohort. For cohorts before 1966, slope and intercept pairs were randomly resampled from pre-1977 values i.e. before a significant increase in slope values occurred. For cohorts after 1993, slope and intercept pairs were randomly resampled from post-1976 values (excluding 1980 -outlier, and 1984, 1987 -poor fits).

Natural mortality

Assuming von Bertalanffy growth parameters for the best fit to survey length at age data of L_{inf} = 220cm, K = 0.33, age at 50% maturity =13 and length at 50% maturity = 75 cm, it can be concluded, based on Beverton-Holt life history invariants (e.g. Jensen, 1996), that the appropriate value for natural mortality (M) is closer to 0.1 than the currently used value of 0.2. Thus, F_{msy} would be around 10% of the biomass, following the rule-of-thumb that $F_{\text{msy}} \approx M$. Two M values were examined: 0.1 and 0.2 to account for uncertainty in M.

Stock-recruit relationship

Given the large degree of uncertainty, and the importance of the stock-recruit relationship in MSE simulations, it is necessary to consider a number of possible stock-recruit models to ensure potential management strategies are robust to this major source of uncertainty. Alternative models applied in the MSE are described below, although we do not advocate any of these as being strongly supported by the available data. Best fits were all calculated by minimising the log residual sums of squares (SS). In all cases SSB in each year (as calculated in equation 4) and recruitment data (n at age 1) data were obtained from the bootstrapped XSA at the start of each repetition of the operating model (i.e. based on the data for years 1975 to 2006).

$$SSB = \sum_{a=1}^{14+} n_a \times w_a \times m_a \tag{5}$$

Where: n_a , w_a and m_a are numbers, weight and maturity at age a, respectively.

1. Segmented Regression

Segmented regression provides a simple description of the stock-recruit data for Greenland halibut with constant recruitment above the breakpoint and recruitment declining linearly to zero below the breakpoint. It thus defined a recruitment-overfishing threshold. Estimates of current spawning stock biomass fall below the breakpoint indicating recruitment-overfishing is occurring.

2. Depleted segmented regression.

We constructed a constrained segmented regression model to have a maximum recruitment equal to the maximum observed recruitment, and a slope that is the best fit (SS) line through the origin. This is consistent with a stock that has a large maximum recruitment and that has been severely recruitment-overfished.

3. Ricker

Under a Ricker model current spawning stock size would be consistent with a stock that has been somewhat overfished. This relationship provides the best fit to the latest assessment data, but indicates strong compensation at higher stock sizes which requires biological justification, for example cannibalism by adults on pre-recruits (Morgan et al., 2008).

4. Beverton and Holt

This model has the advantage relative to segmented regression in that the change in recruitment with stock size is a smooth function and thus abrupt changes in yield do not occur with changing fishing mortality when a MS is applied. This can be an advantage in evaluating management strategy performance.

For all stock-recruit models, a bootstrapping approach is applied in order to introduce process error. For each year a residual is randomly resampled with replacement from the set of model log residuals and the model predicted recruitment was either multiplied or divided by the exponent of the residual (i.e. added or subtracted on the log scale).

Note that only S-R models 1 and 2 have been applied within OMs used to evaluate management strategies thus far.

Population projections

Beginning at the initial starting point, numbers at age are projected using the basic equation for updating population size (Equation 4). Natural mortality (*M*) and partial recruitment (*PR*) are specified by the operating model, while fishing mortality (*F*) depends on the harvest control rule (HCR) defined by the management strategy being evaluated. Recruitment (numbers at age 1) is determined by the stock-recruit model applied in the operating model.

$$N_{a+1,y+1} = N_{a,y} e^{-(M + F_y \times PR_{a,y})} \qquad (6)$$

and for the plusgroup (20+):

$$N_{20+,y+1} = (N_{20,y} + N_{20+,y}) e^{-(M + F_y \times PR_{20,y})} \qquad (7)$$

Where: $N_{a,y}$ = numbers at age a in year y,
M = natural mortality constant across all ages and years,
F_y = fishing mortality in year y,
$PR_{a,y}$ = partial recruitment (selectivity) at age a.

Fisheries/Fleet dynamics

Selectivity/Partial recruitment

Partial recruitment (PR) for this stock is calculated by dividing each F-at-age by the mean F for ages 5-10. For the years 1975-2006 this is done for each individual year based on the XSA assessment and available catch data. This varies from run to run because the non-parametric bootstrap replicates of the XSA, used to condition the operating model, differ.

There are two aspects of PR that need to be considered. Firstly, because these PRs are based on the XSA, there are only values estimated for ages 1-13. For the true population (likely to extend to age 30 but modeled in the OMs with a 20+ group) the PR curve can either be assumed to be asymptotic (or 'flat-topped' function i.e. PR for ages 14-20+ = PR for age 13) or dome-shaped (i.e. decreasing after age 13). Few fish older than 20 are caught by the fishery, so dropping PRs down to near 0 for the age 20 plusgroup would concur with this. However, this could give rise to a large 'cryptic biomass' in the projections, especially given that *M* is constant with age. Secondly, future trends in fishery selectivity at age may not be easy to predict. PR patterns for each year going into the future for the simulation are thus simply resampled from the recent period (1996 to 2004). It should be noted that PR patterns may change as the age structure and abundance of the stock changes. Also, potential gear changes (e.g. reduction in net mesh size) could change selectivity. These refinements could be built into future versions of the OMs should analyses be presented in support of such relationships.

Implementation error

Since the commencement of the rebuilding plan, reported catches have exceeded specified TACs by between 20 and 40%. This is a substantial overrun, given the current low biomass of the stock. The real overrun may potentially be even greater than this but is very difficult to estimate. However, no overrun has been included in this application. Implementation error cannot be simply modeled and assuming TAC overruns would make it more difficult to accurately gauge the relative success of each management strategy. Hence, it is assumed that managers should take implementation error into account in setting TACs prescribed by a management strategy, based on the results of the MSE simulations (i.e. post- rather than pre-analysis adjustments are necessary).

POM error

For all stochastic runs, a full POM error structure is examined (PO is applied in the case of the evaluation of model-free MSs). Uncertainty in commercial catch estimates is not considered at this stage given lack of information, although ideally this should be taken into account.

Comparison of Biological Reference points across Operating Models

Although only four of the 20 specified OMs are evaluated for management strategy performance at this stage, biological reference points for all 20 are briefly described in order to illustrate how they differ (Table 2). The OMs are characterized by a combination of XSA tuning indices, assumptions regarding *M*, characteristics of the stock-recruit function, and the shape of the partial recruitment function for older ages. The choice of natural mortality has a substantial impact. A lower level of *M* reduces the estimate of the current 5+ biomass (B_current) as well as the recovery target (mean biomass for 1975-1999) across all OMs. The ratio of current biomass to target biomass is somewhat higher at lower *M* (B/Target). The equilibrium biomass at F=0 (B_F0) is substantially higher for a lower value of *M*. The degree of current stock depletion from the estimated unexploited biomass is greater for the lower *M* cases. Note that depletion is estimated to be considerable in all OMs with the exception of the OMs based on the Ricker stock-recruit model. The biomass corresponding to MSY (B_Fmsy) is higher for the low *M* cases and the ratio of current biomass to B_{msy} (B/Bmsy) is lower for the low *M* cases. Current yield as fraction of MSY (Yield/MSY) is lower for the low *M* cases whereas the ratio of *F* to F_{msy} (F/Fmsy) is higher.

The choice of the shape of the partial recruitment function has a relatively small impact on the OMs. Domed shaped PRs give higher values for target biomass (Btarget) and lower values for biomass at F_{msy} (B_Fmsy). Ratio of current yield to MSY (Yield/MSY) is higher for the domed PR cases while the ratio of *F* to F_{msy} is lower (F/Fmsy).

The choice of the stock-recruit function has varying impacts on the OMs. The "depleted segmented regression model" (SR=2) results in a lower ratio of current biomass to unexploited biomass and biomass at MSY (B/B0 and B/Bmsy). This is by design. The ratio of current yield to MSY is also lower (Yield/MSY). The Ricker stock-recruit model has the counter-intuitive outcome that the ratio of current biomass to unexploited biomass (B/B0) is close to or greater than 1, depending on the tuning indices. This is a result of the high level of density dependence and the late age at maturation. The Beverton-Holt stock recruit model results in behaviour that is more similar to that obtained under the segmented-regression OMs as might be expected. However, the functions against *B* and *F* change smoothly, rather than the abrupt changes that occur under the segmented regression OMs. This could be and advantage in evaluating performance statistics.

A comparison of those OMs conditioned on an XSA tuned only with research vessel survey data and those OMs tuned with a combination of research vessel survey data and commercial catch rate data shows that current biomass is closer to target for those OMs conditioned on XSA tuned with both commercial and research data.

For the four OMs selected for MSE, OM4 has a lower Bcurrent and Btarget compared with the other 3 OMs but higher B_Fmsy. OM6 has a higher Bcurrent and a lower lower current F/Fmsy compared with the other OMs. OM10 has a much higher B_F0 compared with the other OMs. Sissenwine-Shepherd plots for these four OMs are illustrated graphically in Figs. 5 – 8. These include plots of yield per recruit (YPR) and spawning stock biomass per recruit (SSBPR) vs. *F*, recruitment (R) vs spawning stock biomass (SSB), equilibrium yield vs. *F* and equilibrium yield vs. 5+ biomass. Biological reference points superimposed on these plots are $F_{0.1}$, $F_{\max}$, F_{current}, F_{rp} (the *F* that will result in equilibrium 5+ biomass at the rebuilding target), SSB0 (equilibrium spawner biomass at F=0), 20%SSB0 (spawner biomass at 20% of the unexploited SSB and SSBmsy (SSB at MSY). In all four OMs, current SSB is far below SSB0. Recruitment data points are clustered around the breakpoint and the descending limb of the segmented S-R relationship. By design, for OM10 all the recruitment data points fall on the descending limb. Plots of equilibrium yield vs. *F* show that as *F* increases there is an abrupt collapse in equilibrium yield at point where the replacement line in the S-R plot has a slope that exceeds the slope associated with the break-point. In comparison, the decline in yield at high *F* would be more gradual under a Beverton-Holt curve. In thee of the four OMs, F_{current} exceeds F_{crash}, the exception being OM6 where F_{crash} is slightly higher than F_{current}.

Deterministic Constant Catch results (0, 8k, 16k, 32k)

Deterministic results under constant catch strategies were evaluated for OMs 2, 4, 6 and 10, representing a range of possible stock dynamics and starting points. To illustrate the deterministic results for the four OMs under the four constant TAC options, the time series for the recovery ratio, mean *F*, SSB and recruitment are plotted over time to 2030 for each TAC option (Figs. 9-12). Under zero TAC, exploitable biomass increases most rapidly for OM4 and reaches a higher level by 2019 despite the fact that the maximum recruitment level is the lowest of the 4 OMs. This illustrates the major impact that the assumption about *M* has on this stock. After starting off somewhat slower, OM10, which has the highest level of maximum recruitment, eventually overtakes the OMs 2 and 6 in terms recovery of both exploitable biomass and SSB. Under a constant TAC of 16kt the differences in the behaviours of the 4 OMs becomes very apparent (Fig. 11). While initially recovering more rapidly than the other OMs, OM6 is overtaken by OM4 after 2019. Fishing mortality rises to a very high level under OM10, and, although there is a small recovery around 2019 as a result of transient population effects, the population goes extinct before 2030. Under a 32kt constant TAC recruitment overfishing is rapid in all four OMs and the collapsing trajectories are essentially similar, occurring under very high levels of *F*.

In the absence of fishing, the population increases under all four OMs. By 2010 the biomass reached is just below the target for OM2 and OM10, but exceeds the target for OM4 and OM6. Thereafter the biomass is well above the target level in all OMs. Even a small constant TAC of 8kt results in a decreases recovery prospects slightly compared to zero catch. Over the rebuilding plan period (i.e. to beginning of 2019) recovery to the rebuilding target occurs for all four OMs for a constant TAC of 8kt and for OMs 4 and 6 for a TAC of 16kt. However, under a TAC of 32kt all four OMs show that biomass will not meet the recovery target and that the average catch possible under the fishing mortality cap of 1.73 is substantially less than the TAC.

Stochastic results under different management strategies

Figs. 13-16 plot the descriptive statistics of each MS for OM 2, 4, 6 and 10, respectively from the stochastic simulation results. The statistics show that each MS results in different future trajectories of the stock. Notable differences in *F* translate into a range of catch values that in turn impact upon the recovery or decline of the stock, SSB and recruitment. The high *F* values for the Fsq strategy crash the stock in most cases. The low *F* strategies (PA and HalfF01) have the best recovery, both exploitable and spawner biomass rapidly increasing from the period of implementation.

These patterns are very similar between OMs, with the exception of OM4 (M=0.1). Better rates of recovery of exploitable biomass and SSB occur under this scenario of lower natural mortality, which leads to decreased *F* and an increase in the mean age of the population. SSB increases substantially for all OMs after 2007 (the start of the simulation) for any strategy that keeps the level of *F* within the range witnessed in the modeled part of the time series. This significant increase in SSB is in most cases, with the exception of OM4, not matched by a similar increase in the average age of the population, possibly due to strong recruitment leading to many five year old fish dominating the exploitable biomass. For OMs with the standard segmented regression stock-recruit curve, recruitment levels rapidly increase to the maximum level of the SRR curve in the absence of very high *F*. In OM10, with the depleted segmented regression curve, the recovery in recruitment is slightly less rapid or prolonged.

The industry-focused performance statistics of the five MSs for each OM are shown in Figs. 17-19. In the short term (2010, Fig 17) there is very little difference between OMs. The Fsq strategy allows the greatest Avg. Catch (>20kt), while the ModFree and FCMod strategy maintain catch levels only slightly lower than recent observed catches. The PA and HalfF01 strategies result in substantial decreases in catch to around 5kt. All observed catches fall below the MSY level because in all cases the stock is below the optimum level and recovery is required before MSY can be achieved. OMs 4 and 10 would both lead to greater MSY in the long term, due to lower *M* and greater potential recruitment respectively, and hence Catch/MSY ratios are lower in both of these OMs. AAV in Catch is only high for the PA and HalfF01 strategies as these two produce a substantial initial decrease in *F* and hence catch. This is also reflected in these two have slightly higher CVs on *F*.

Over the length of the rebuilding plan period (2019, Fig 18) and into the longer term (2030, Fig 19), the Avg. Catch allowed by the Fsq strategy decreases steadily as the stock declines, reaching a low of 5kt, the lowest of any of the MSs examined. This decrease also results in a large AAV in Catch (>20%). The ModFree and FCMod strategies maintain similar catch levels into the long term keeping AAV in catch low, particularly for the FCMod strategy which only changes the TAC every two years. The PA and HalfF01 strategies produce increasingly higher levels of catch as the stock recovers, with the PA strategy having the greatest Avg. Catch in the long term (>20kt). This however coincides with a high AAV in Catch, particularly for the HalfF01 strategy. The mean age of the catch (Fig. 20) fluctuates more under low *F* strategies than observed in the historically modeled part of the time series. Not surprisingly, the average age caught increases most in OM4, where the lower *M* leads to greater survival to the older ages. In all the other OMs, mean age remains between 6 and 9 years, covering the peak selectivity of the fishery but below the age of 50% maturation.

Figs. 21 and 22 show the stock performance statistics of the five MSs for each OM. The choice of MS has little effect on the stock in the short term (two years), so only the 2019 and 2030 results are shown. Because of the higher catches in the short term the Fsq strategy eventually leads to a stock collapse and hence the reduction in Recov. Ratio and B/Bmsy, as well as the considerable reduction in average catch in the long term. Fishing mortality under the Fsq is the highest compared to F_{msy}. Only the HalfF01 and PA strategies lead to substantial rebuilding in stock size in the long term. But while both achieve a B/Bmsy ratio of greater than 1, only HalfF01 achieves the rebuilding target by the end of the rebuilding plan across all OMs. While the ModFree and FCMod strategies are unlikely to collapse the stock, by keeping catches high, neither lead to any real recovery of the stock either. Recov. Ratio and B/Bmsy show the same patterns for all OMs (both are ratios on exploitable biomass). However, OMs 4 and 10 have greater B_{MSY} levels and hence more recovery is required for stocks to reach B_{MSY} in these OMs. The quicker recovery observed in the descriptive stats for OM4 (Fig. 14) is evident, with all MSs except Fsq exceeding the rebuilding target by 2019.

Worm plots, of time series of each replicate, are shown for the FCMod (Fig. 23) and PA (Fig. 24) strategies in OM2. The big range of Recov. Ratios for the FCMod strategy can be seen in the great variability in individual replicate trajectories of exploitable biomass (Figs. 21 and 22). The perceived view of the stock (i.e. the XSA estimate of exploitable biomass for each year) is slightly higher than the true population in the short term because shrinkage of *F* results in a lower estimated *F* in the model than actually occurs in reality in the first few years. This leads to higher TAC being set than the HCR would specify if the true population was known without error. In particular a few replicates are wrongly perceived to have recovered substantially leading to steady increases in catch and the subsequent collapse after 2019. In contrast, the PA replicates have a much tighter distribution. The perceived view, which exaggerates the fluctuations in the true population, initially underestimates stock size; hence the very low catches in the short term. This allows the stock to recover rapidly and in turn leads to increased catch.

Discussion and conclusions

The current reference set of 20 operating models covers a fairly wide range of possible “realities”. This reflects the fact that there is considerable uncertainty regarding the biology and population dynamics of the 2+3KLMNO Greenland halibut stock. The operating models are all conditioned on an XSA assessment in which the tuning indices and assumptions regarding *M* and PR are varied. XSA estimates are combined with further assumptions to describe the underlying stock-recruit function as well as the selectivity of older fish in the population. Further operating models could be considered that are not conditioned on the XSA. For example, should the basis for the assessment change in the future (e.g. to a statistical catch at age analysis), then alternative forms of conditioning could be considered, although this would not be essential unless perceptions regarding the biology and population dynamics of the stock or the impact of the fishery changes (i.e. the OMs do not have to be conditioned with the same model that is used in the assessment).

A broad range of OMs have been specified in this study although only four have been applied in the MSE so far. While it is not possible to capture every eventuality in the reference set of operating models, some aspects may require further attention. For example, it is possible that movement of fish between this stock and the stock in Area 0 to the north takes place. The OM could be expanded to take migration into account. However, in the absence of an analysis of fish movements between the two areas, this would be highly speculative. Another source of uncertainty is how increased survival from lower fishing mortality under a rebuilding strategy will influence the dynamics of the stock, both in terms of generating recruits and in terms of the selectivity of the commercial fishery. Under assumptions regarding constant *M*=0.1 or even 0.2, there would be substantially more biomass in the older ages in a rebuilding and rebuilt stock (Figs 13-16). It is possible that *M* on older fish increases with age which would reduce this “cryptic” biomass, but there are no analyses to address this. The current assessment (Healey and Mahé, 2008) reflects a stock that has very low spawner biomass and low recruitment (Morgan et al., 2008), consistent with severe recruitment-overfishing. There is also some indication of reduced recruitment at higher biomass within the range of historical XSA estimates, possibly indicating that cannibalism by adults on pre-recruits could become an important factor in a rebuilding stock (i.e. that a Ricker stock-recruit model may pertain).

The coding structure we currently have in place allows for relatively easy incorporation of new OMs. Although further work on OMs should take place as scientific knowledge regarding this stock improves, the current OMs in the reference set comprise a fairly broad range and should be considered for ongoing evaluation of MSs. Agreement on the reference set of operating models is key aspect of the MSE process. A common problem experienced on a declining or declined stock, when commercial concentrations of fish remain accessible in a few prime locations, is a difference in perception between the fishing industry and the stock assessment scientists regarding the status of the stock and the expected outcome of different TAC options. While the development of the OMs is essentially in the domain of data-driven science, we have attempted to address this difference in opinion by generating commercial CPUE series and applying them in conditioning four of the OMs in the reference set. The results of one of these OMs are examined in detail. It should be noted, however, that NAFO Scientific Council has not accepted the commercial CPUE data in terms of potential tuning indices for the XSA assessment thus far.

Evaluating the results from a large number of OMs can be overwhelming. We have adopted a reference set of 20 OMs, 16 of which consist of all combinations of four factors, each with two levels: tuning indices, natural mortality, shape of PR for ages older than 13 and stock-recruit model (from within the two forms of segmented regression). The other four consist of two OMs, one with *M*=0.1 and the other with *M*=0.2, using a Ricker stock recruit curve and two with a Beverton-Holt stock recruit curve. The four OMs that we have thus far applied only deal with the current assessment view and one OM for a different value for each of the three factors corresponding to tuning indices, natural mortality and stock-recruit model (segment regression and depleted segmented regression only). This allows us to examine how different values for these factors impact on future projections of the stock under the different MSs. Performance statistics can be made available in the future for all OMs, as required by the decision-making process, should a decision be made by FC to apply a prescribed harvest control rule for this stock.

Given an acceptable range of OMs, the next step is consideration of alternative MSs. Suggestions for MSs are mainly the domain of fisheries managers in consultation with fishing industry and other stakeholders. These do, however, have to be fully specified in order to be properly evaluated within the MSE approach. The current FC rebuilding plan for Greenland halibut is incompletely specified because it does not describe how the TAC will vary beyond 2007. Further, the TAC reduction steps over the period 2004-2007 are arbitrary and do not follow any prescribed harvest control rule. Also, there is no prescribed rule for what should happen after 2007: Subsequent TAC levels "*may be adjusted by the Scientific Council advice*" but "*shall not be set at levels beyond 15% less or greater than the TAC of the preceding year*". In the absence of a pre-specified rule, the 2008 TAC was set by FC at the same level as 2007 despite lack of any recovery of the stock. The only structure provided by the rebuilding plan beyond the arbitrary TAC reduction steps to 2007 is the 15% cap on any future TAC adjustments.

There are a number of potential strategies that could be considered for rebuilding plans that have well specified goals (e.g. relative biomass targets, rate of recovery etc.). These involve determining each year either a fixed TAC or fixed F value that would lead to rebuilding the stock to the specified target in the specified time period. Also, if the time period is not critical, then a strategy that reduces F to the F value that corresponds to the target biomass in equilibrium could be considered. These F or TAC values would have to be calculated iteratively in a deterministic manner (i.e. assuming no process error). In stochastic simulations the values would have to be recalculated each year in the simulation as more data becomes available in the perceived world. Essentially these strategies represent two methods of achieving the target biomass exactly at the target year and therefore they could be very useful for assessing trade-offs that would be involved if the rebuilding plan was strictly adhered to. We attempted to implement strategies such as these but encountered some technical difficulties in the programming of the strategies. The current MSE FLR code was developed before the full release of FLR v2, and the methods required to implement these strategies effectively are only available in the newer package. Instead of pursuing these rebuilding plan strategies further, we focused our attention on the five proposed 'reactive' (feedback control) management strategies. Strategies aimed at achieving the rebuilding target by a particular date would also require "course adjustments" to be made as the perception regarding the real world changes from one assessment to the next.

A number of descriptive statistics, allowing an understanding of the stock dynamics under different OM-MS combinations can be developed by scientists. However, performance criteria are largely the domain of the stakeholders, both the fishing industry and representatives of broader segments of society, such as ENGOs. This process needs to be informed by national and RFMO policies on sustainable fisheries, and guided by legally binding agreements such as UNFSA and voluntary undertakings such as the FAO Code of Conduct for Responsible Fisheries. We have suggested a range of both industry-focused and conservation-focused performance statistics based on input from industry, scientists and fisheries managers. Further consultation is required with ENGOs to determine whether additional performance measures would be useful.

Comparing OM2 (Fig 13) and OM4 (Fig. 14), the projections of the stock are highly sensitive to M. This could in part be due to an interaction with PR. PR in the simulated part of the time series is resampled from past years and hence does not adapt to the changing age structure of the population. Average age increases notably by the end of the rebuilding plan for all MSs, creating a large biomass of fish that are less accessible to the fishery (due to lower PR at older ages), leading to sustained high, increasing exploitable biomass levels, large SSB biomass and consequently strong recruitment feeding into the stock. These factors all contribute to a more resilient stock than one with a higher M. The current annual assessment assumes M=0.2, while biological attributes suggest that it may be lower for this stock. It seems resolving this issue should be a high priority for the management of this stock.

The alternative depleted segmented regression SRR curve (OM10) had less of an effect on simulations than M, but it did allow for greater levels of recovery under low F, as would be expected from higher potential recruitment levels (Fig. 16). The CPUE indices in the XSA (OM6, Fig. 15) had very little effect on projections of the stock, suggesting that starting point could be less influential than stock dynamics. Alternatively, this method of creating a more "optimistic" scenario is not very effective. This could be due to the historically high F values over the period to which the XSA model is fit. This would mean that the fit is dictated by the catches, and only notable differences in indices are likely to have any great effect on the resulting perception of the stock. The CPUE XSA did not produce a great improvement in recent historical exploitable or spawner biomass, and, while F is lower and Recruitment is higher in later years, these values obviously fail to result in any long term resilience of the stock in projections. One possibility would be to examine the residual patterns of the indices from the current XSA fit. Selecting "optimistic periods" of these indices and using only these to fit an XSA, the assessment could possibly be 'tweaked' to give a more positive realization of the recent historical period. Another possibility, given the substantial retrospective pattern in the most recent XSA assessment (Healey and Mahé, 2008), would be to only fit the XSA to the data up to 2003 (when the rebuilding plan was implemented). The 2003 assessment showed the stock to be in a very poor condition with low exploitable biomass, SSB, recruitment and high F. If the numbers at age matrix from this assessment was projected forward using the known catches to date, this would create a substantially more negative view of the stock than the view from the most recent XSA assessment, which has seen a year on year increase in exploitable biomass and recruitment in the retrospective analysis. These two alternate views of the stock status (one a stock in trouble, the other a stock with low, but not critical, population levels) would provide two very different starting points from which to test alternative management strategies.

The descriptive statistics also show that in practice, the Fsq strategy leads to high *F* values. TACs in this strategy are set based on projections of stock size in the year after the XSA assessment is done. Shrinkage of *F* in the XSA leads to an underestimation of an increasing stock and conversely an overestimation of a declining stock. The stock declines in the initial simulation period, but because the TAC is based on the perceived view, which shows less decline, the actual *F* inflicted on the stock is higher than it should be. In reality, the combination of the XSA perception of the stock, shrinkage within the XSA and the two year lag between data (up to year y-1 used to fit the model) and the setting of the TAC (based on projections to year y+1) makes finding robust HCRs difficult. Generally XSA with shrinkage applied does not cope well with a rapidly increasing or decreasing stock. There are also other problems with the use of XSA in the simulations, such as the low catch problem found with the HalfF01 strategy and potential scale problems encountered when populations reach particularly high levels. In a few simulations, when the perceived population reached the 1 million ton level, estimates sometimes skyrocketed from one year to the next, reaching extremely high, unrealistic population sizes for some unknown reason. This in turn led to extremely high TACs being set, which then crashed the true population. The exact reasons for such problems are not fully understood at this point, but these are unlikely to pose significant problems in reality. The simulations run for over twenty years without checking that the results of the XSA are valid and realistic. By incorporating some sort of "sensibility rule" into the MS, an alternative method for setting the TAC can be specified in years when the XSA assessment produces an unrealistic result (i.e. near zero population or a substantial increase from the previous years' estimate). In practice, these issues would be identified and taken into account each year when setting TACs.

Evaluation of the performance of the alternative MSs across the reference set of operating models based on the results from the stochastic runs is not a simple task given the high dimensionality of the process. Box and whisker plots and sample trajectories provide some visual guidance. In theory it would be possible to develop some overall weighted sum of performance measures and choose the MS that has the highest score across all OMs. This "find the winner" approach is, however, not advocated as an initial step. Instead, we suggest that the approach of "Satisficing" be adopted. This is a decision-making strategy which attempts to meet criteria for adequacy, rather than to identify an optimal solution. Under this approach, FC would be presented with a number of strategies that perform adequately with respect to the performance measures (tolerances and risks), allowing some flexibility in the decision-making process. Clearly, MSs that don't meet PA risk tolerances and that have low probability of rebuilding to the intermediate recovery target in the short term and to the vicinity of B_{msy} in the longer term should not be considered "adequate" and should not be considered further.

The performance measures presented contain both satisficing measures and trade-offs. There are only a few satisficing constraints due to the lack of targets and industry requirements specified in the rebuilding plan. Likewise, risk levels have not been specified for determining the suitability of MSs with regards to these satisficing tolerances. Further input from decision-makers will therefore be required specify clearly the conditions to be used to decide whether or not MSs can be deemed suitable. At present, a recovery ratio greater than or equal to 1 by 2019 and an annual average variation in TAC of less than 15% are the only rebuilding plan tolerances levels that have been suggested. B/B_{msy} and F/F_{msy} have been added based on the need for fisheries management to meet the UNFSA criteria with respect to the PA. Suitable risk levels still need to be specified by decision makers. The remaining performance measures, average catch, catch/MSY and the CV on *F*, can provide an indication of trade-offs in performance, and can be used to judge the relative merits of candidate MS that are judged to be adequate based on first satisfying the prescribed risk tolerances.

The Fsq strategy is clearly not a viable way to manage this stock (although it has in the past been resorted to by NAFO FC), failing to achieve the rebuilding target, long term sustainability and, ultimately, leading to large annual downward changes in TAC. Both the model-free and FC model-based results have varied results. While the concept of MSE is relatively new to NAFO, experience elsewhere has indicated that model-free management strategies may have a higher degree of acceptability in fisheries managers than those based estimates from models (Butterworth, in press). This is mainly related to simplicity and transparency. These strategies may also perform as well or better than strategies based on model estimates. However, this current analysis shows that the large amount of observation error associated with this stock can cause incorrect adjustment of TACs when applying the model-free strategy, leading to a large variability in possible recovery outcomes. Likewise in the FCMod strategy, the XSA perception of stock each year leads to incorrect adjustments of TACs because of large error in the XSA estimates. This is exacerbated by the less frequent adjusting of TACs (i.e. adding to the reaction lag) in this strategy, hence the even broader range of possible outcomes. The failure of these two strategies to allow reasonably recovery of the stock in both the rebuilding plan period and the long term, albeit primarily due to error in the perception of the stock, deems them unsuitable for the management of the stock, despite the consistency in TAC they allow from year to year.

In the present evaluation, only HalfF01 achieved the desired exploitable biomass target level by the end of the rebuilding plan in all four OMs. After HalfF0.1, the PA strategy came closest to allowing adequate recovery in the rebuilding plan period. With this strategy, the rebuilding plan target of exploitable biomass in 2019 is achieved with about a 25% probability for OM2, 100% probability for OM4, 25% probability for OM6 and a 75% probability for OM10. In the longer term both these strategies had a very good likelihood of rebuilding the stock to B_{MSY} (100% in OMs 2,6 and 10, and just short of 75% in OM4). On the negative side, both the HalfF01 and PA strategies resulted in a large absolute annual variation in catch and a high CV on F (due to initial large decreases in TAC and subsequent large increases in TAC as the stock recovered). If the rebuilding plan tolerances need to be met with reasonably high probability, then only HalfF01is a viable strategy. However, if long term recovery of the stock is the primary aim rather than medium term, then both the PA and HalfF01 strategy would be suitable, and the trade-offs would need to be evaluated to choose a preferred MS. Over all three time periods examined, the PA strategy has a higher average catch, closer to MSY, with a lower annual average variation in catch (although still outside the rebuilding plan tolerance) and CV on F. In fact, the PA strategy yields the highest catches in the long term when taking all OMs into account. Hence, if the rebuilding plan date target is seen as less of a constraint and more of a guideline, the PA strategy is clearly the most favourable MS across the present set of four OMs.

Management strategy evaluation provides a powerful new tool for NAFO to consider in the management of stocks in the regulatory area. There is increasing international pressure on RFMOs to manage fish stocks in a sustainable manner consistent with the UNFSA and the FAO Code of Conduct for Responsible Fisheries. This requires adopting a PA-compliant management strategy that encompasses harvest control rules that can be demonstrated to have a high probability of rebuilding depleted stocks and a low probability of causing serious harm. Despite these pressures there has been reluctance by some RFMOs to move from current largely *ad hoc* approaches to a prescribed management strategy containing explicit harvest control rules. 2+3KLMNO Greenland halibut provides an example. Part of the reason could be that current methods of stock assessment may be considered too uncertain to be used as a basis for triggering a harvest control rule that will have very certain social and economic impacts, and that a subjective or negotiated approach will be more reliable. Unfortunately such an approach will always be suboptimal in terms of stock rebuilding and long-term sustainability. MSE provides a comprehensive means of capturing uncertainty and of evaluating alternative harvest control rules to ascertain their robustness to uncertainty by seeing whether they perform acceptably with regard to meeting tolerances within acceptable risk levels. Considerable flexibility still exists in the decision-making process in terms deciding on tolerances and risk levels and then selecting from among potentially suitable management strategies based on trade-offs between performance statistics. There is considerable advantage to adopting an MSE approach. The harvest control rule that provides the TAC is explicit and transparent, as are the tolerances and acceptable risk levels for performance statistics. The tradeoffs that were taken into account in choosing from among acceptable management strategies can be clearly shown, thereby increasing transparency. In contrast, a TAC negotiated behind closed doors is becoming increasingly difficult to justify as society becomes more aware of the dangers inherent in overfishing and the imperative that fish stocks be sustainably managed. It also provides no feedback to the scientific assessment process on how to be more effective in the future in terms of providing advice.

While there remains some work to be done in order to finalize a reference set of OMs suitable to all parties concerned, agree on a set of management strategies to evaluate and determine tolerances and risk levels, it is anticipated that and MSE for 2+3KLMNO Greenland halibut could be finalized for review by SC in June 2009 and be provided to FC prior to September 2009 as a basis for evaluating management options for this stock for 2010. However, further development would require the backing of the FC and in particular would require a commitment to base management decisions on a prescribed HCR that can be shown to adequately meet pre-specified risk tolerances. Further implementation of MSE for the 2+3KLMNO Greenland halibut stock is best achieved through the continuation of a working group approach, harnessing expertise to develop the operating model, apply alternative strategies and critically evaluate the inputs, outputs and performance measures. The approach will require some discussion between members of SC and FC prior to September 2009 if a properly evaluated rebuilding strategy is to be implemented in the management of this stock in 2010.

References

BOWERING, W.R and BRODIE, W.B. 1995. Greenland halibut (*Rheinhardtius hippoglossoides*). A review of the dynamics of its distribution and fisheries off eastern Canada and Greenland. In A.G. Hopper (ed.), Deep-Water Fisheries of the North Atlantic Oceanic Slope, 113-160. Kluwer Academic Publishers, Netherlands.

BUTTERWORTH, D.S. (in press). Some lessons from implementing management procedures, in Fisheries for global welfare and environment, memorial book for the 5th World Fisheries Congress in Japan, October, 2008, Yokohama. TERRAPUB: Tokyo, Japan.

DARBY, C.D., W.R. BOWERING, and J.-C. MAHÉ. 2003. An Assessment of Stock Status of the Greenland Halibut Resource in NAFO Subarea 2 and Divisions 3KLMNO Based on Extended Survivors Analysis with Short and Medium-term projections of Future Stock Development. *NAFO SCR Doc*., No. 03/64, Ser. No. N4883.

GONZÁLEZ TRONCOSO, D., J.M. CASAS, and F. SABORIDO-REY. 2006. Results from bottom trawl survey on Flemish Cap of July-August 2005. *NAFO SCR Doc*., No. 16, Ser. No. N5231.

GUNDERSEN, A.C. and BRODIE, W.B. 1999. Length-weight relationship for Greenland halibut (*Reinhardtius hippoglossoides*) in NAFO divisions 2GHJ 3KLMNO 1990-1997. *NAFO SCR Doc.*, No. 99/31, Ser. No. N4087, 21p.

HEALEY, B.P. 2007. Greenland halibut (*Reinhardtius hippoglossoides*) in NAFO Subarea 2 and Divisions 3LKMNO: stock trends based on annual Canadian research vessel survey results during 1978-2006. *NAFO SCR Doc.*, No. 07/45, Ser. No. N 5397.

HEALEY, B.P. and J.-C. MAHÉ. 2006. An Assessment of Greenland Halibut (*Reinhardtius hippoglossoides*) in NAFO Subarea 2 and Divisions 3KLMNO. *NAFO SCR Doc.,* No. 06/51, Ser. No. N5281.

HEALEY, B.P. and J.-C. MAHÉ. 2007. An Assessment of Greenland Halibut (*Reinhardtius hippoglossoides*) in NAFO Subarea 2 and Divisions 3KLMNO. *NAFO SCR Doc.,* No. 07/53, Ser. No. N5405.

HEALEY, B.P. and J.-C. MAHÉ. 2008. An Assessment of Greenland Halibut (*Reinhardtius hippoglossoides*) in NAFO Subarea 2 and Divisions 3KLMNO. *NAFO SCR Doc.,* No. 08/48, Ser. No. N5550.

ICES 1997. Report of the Comprehensive Fishery Evaluation Working Group (COMFIE). ICES CM 1997/Assess: 15.

JENSEN, A..J. 1996. Beverton and Holt life history invariants result from optimal trade-off of reproduction and survival. *Can. J. Fish. Aquat. Sci.* **53**: 820-822.

KELL, L.T., I. MOSQUEIRA, P. GROSJEAN, J-M. FROMENTIN, D. GARCIA, R. HILLARY, E. JARDIM, S. MARDLE, M.A. PASTOORS, J.J. POOS, F. SCOTT and R.D. SCOTT. 2007. FLR: an open-source framework for the evaluation and development of management strategies. *ICES J. Mar. Sci.* **64**(**4**): 640-646.

KIRKWOOD, G.P., and SMITH, A.D.M. 1996. Assessing the precautionary nature of fisheries management strategies. *FAO Fisheries Technical Paper*, 350(2). 210 pp.

MILLER, D.C.M. and P.A. SHELTON. 2007. A nonparametric bootstrap of the 2006 XSA assessment for Greenland Halibut (*Reinhardtius hippoglossoides*) in NAFO Subarea 2 + Divisions 3KLMNO using Fisheries Libraries in R (FLR). *NAFO SCR Doc*., No. 07/59, Ser. No. N5411.

MILLER, D.C.M., P.A. SHELTON, B.P. HEALEY, M.J. MORGAN and W.B. BRODIE. 2007. Management strategy evaluation for Greenland halibut (*Reinhardtius hippoglossoides*) in NAFO Subarea 2 and Divisions 3LKMNO. *NAFO SCR Doc*., No. 07/58, Ser. No. N5410.

MORGAN, M.J., P.A. SHELTON, D.C.M. MILLER and B.P. HEALEY. 2008. Is there any evidence of a stock recruit relationship for Greenland halibut in Subarea 2 + Div. 3KLMNO? *NAFO SCR Doc*., No. 08/46, Ser. No. N5548.

MORGAN, M.J. and RIDEOUT, R.M. 2007. An update of maturity estimates for Greenland halibut in NAFO Div. 2J+3K. *NAFO SCR Doc*., No. 07/23, Ser. No. N5374.

NAFO FC. 2003. NAFO Rebuilding Plan for the stock of Greenland Halibut in Subarea 2 and Divisions 3KLMNO. NAFO FC Doc 03/13, Ser. No. N4904.

NAFO SC. 2003. Proposed NAFO Precautionary Approach Framework from Scientific Council. *NAFO SCR Doc.* 03/23.

NAFO SC. 2008. Report of NAFO Scientific Council Study Group on Rebuilding Strategies for Greenland halibut. *NAFO SCS Doc*. 08/13, Ser. No N5518.

RADEMEYER, R.A., É.E. PLAGÁNYI and D.S. BUTTERWORTH. 2007. Tips and Tricks in Designing Management Procedures. *ICES J. Mar. Sci.* **64**(**4**): 618-25.

SHELTON, P.A. 2005a. Does the rebuilding plan for Greenland halibut in Subarea 2 and Divisions 3KLMNO have a scientific basis and is it on track? *NAFO SCR Doc*., No. 05/10, Ser. No. 5089.

SHELTON, P.A. 2005b. A PA-compliant rebuilding plan for Subarea 2+ Divisions 3KLMNO Greenland halibut based on the 2005 NAFO assessment. *NAFO SCR Doc*., No. 05/71, Ser. No. 5165.

SHELTON, P.A. 2007. The weakening role of science in the management of groundfish stocks off the east coast of Canada. *ICES J. Mar. Sci.* **64**: 723-729.

SMITH, A.D.M., SAINSBURY, K.J., and STEVENS, R.A. 1999. Implementing effective fisheries-management systems – management strategy evaluation and the Australian partnership approach. *ICES J. Mar. Sci.* **56**: 967-979.

Table 1. The reference set of Operating Models considered for the Greenland halibut 2J 3KLMNO stock in the MSE analysis.

Ref. Sub-set	OM #	Starting Pt.	Stock-Recruit	*M*	PR decline
"Current" Below B_{MSY} Poor current state	1	Current indices	Seg Reg	0.2	Domed
	2*	Current indices	Seg Reg	0.2	Asymptotic
	3	Current indices	Seg Reg	0.1	Domed
	4*	Current indices	Seg Reg	**0.1**	Asymptotic
"More Optimisitc" Near B_{MSY} Healthy current state	5	Catch Rate indices	Seg Reg	0.2	Domed
	6*	**Catch Rate indices**	Seg Reg	0.2	Asymptotic
	7	Catch Rate indices	Seg Reg	0.1	Domed
	8	Catch Rate indices	Seg Reg	0.1	Asymptotic
"Less Optimistic" Far below B_{MSY} Poor current state	9	Current indices	Depleted Seg Reg	0.2	Domed
	10*	Current indices	**Depleted Seg Reg**	0.2	Asymptotic
	11	Current indices	Depleted Seg Reg	0.1	Domed
	12	Current indices	Depleted Seg Reg	0.1	Asymptotic
"Best case" Far below B_{MSY} Healthy current state	13	Catch Rate indices	Depleted Seg Reg	0.2	Domed
	14	Catch Rate indices	Depleted Seg Reg	0.2	Asymptotic
	15	Catch Rate indices	Depleted Seg Reg	0.1	Domed
	16	Catch Rate indices	Depleted Seg Reg	0.1	Asymptotic
Alternative SRRs	17	Current indices	Ricker	0.2	Asymptotic
	18	Catch Rate indices	Ricker	0.2	Asymptotic
	19	Current indices	Bev Holt	0.2	Asymptotic
	20	Catch Rate indices	Bev Holt	0.2	Asymptotic

* = examined in stochastic simulations.

Table 2. Summary of biological characteristics and reference points of the 20 OMs constituting the suggested reference set.

OM	M	PR	SR	B current	B target	B/Target	B_F0	B/B0	B_Fmsy	B/Bmsy	Yield/MSY	F/Fmsy
1	0.20	2	1	76660	140474	0.55	579107	0.13	139669	0.55	0.81	1.98
2	0.20	1	1	76529	139957	0.55	579107	0.13	154644	0.49	0.79	2.31
3	0.10	2	1	64104	109566	0.59	1265910	0.05	277609	0.23	0.66	3.58
4	0.10	1	1	63972	109031	0.59	1265910	0.05	343883	0.19	0.58	5.10
5	0.20	2	1	87221	133955	0.65	583870	0.15	157256	0.55	0.77	1.65
6	0.20	1	1	87149	133595	0.65	583870	0.15	174231	0.50	0.75	1.95
7	0.10	2	1	72147	106037	0.68	1201201	0.06	278294	0.26	0.65	2.96
8	0.10	1	1	72076	105662	0.68	1201201	0.06	341186	0.21	0.57	4.28
9	0.20	2	2	76660	140474	0.55	901068	0.09	217319	0.35	0.52	1.98
10	0.20	1	2	76529	139957	0.55	901068	0.08	240620	0.32	0.51	2.31
11	0.10	2	2	64104	109566	0.59	1967894	0.03	431551	0.15	0.42	3.58
12	0.10	1	2	63972	109031	0.59	1967894	0.03	534576	0.12	0.37	5.10
13	0.20	2	2	87221	133955	0.65	893139	0.10	240552	0.36	0.50	1.65
14	0.20	1	2	87149	133595	0.65	893139	0.10	266519	0.33	0.49	1.95
15	0.10	2	2	72147	106037	0.68	1960635	0.04	454240	0.16	0.40	2.96
16	0.10	1	2	72076	105662	0.68	1960635	0.04	556894	0.13	0.35	4.28
17	0.20	1	3	76529	139957	0.55	79592	0.96	137453	0.56	0.68	1.60
18	0.20	1	3	87149	133595	0.65	55103	1.58	138049	0.63	0.65	1.08
19	0.20	1	4	76529	139957	0.55	961139	0.08	309719	0.25	0.52	3.18
20	0.20	1	4	87149	133595	0.65	948399	0.09	322013	0.27	0.49	2.48

B = exploitable (5+) biomass

PR=2 implies domed PR, PR=1 implies asymptotic PR;

SR=1 implies segmented regression stock-recruit model, SR=2 implies depleted segmented regression model, SR=3 implies Ricker model and SR=4 implies Beverton-Holt model.

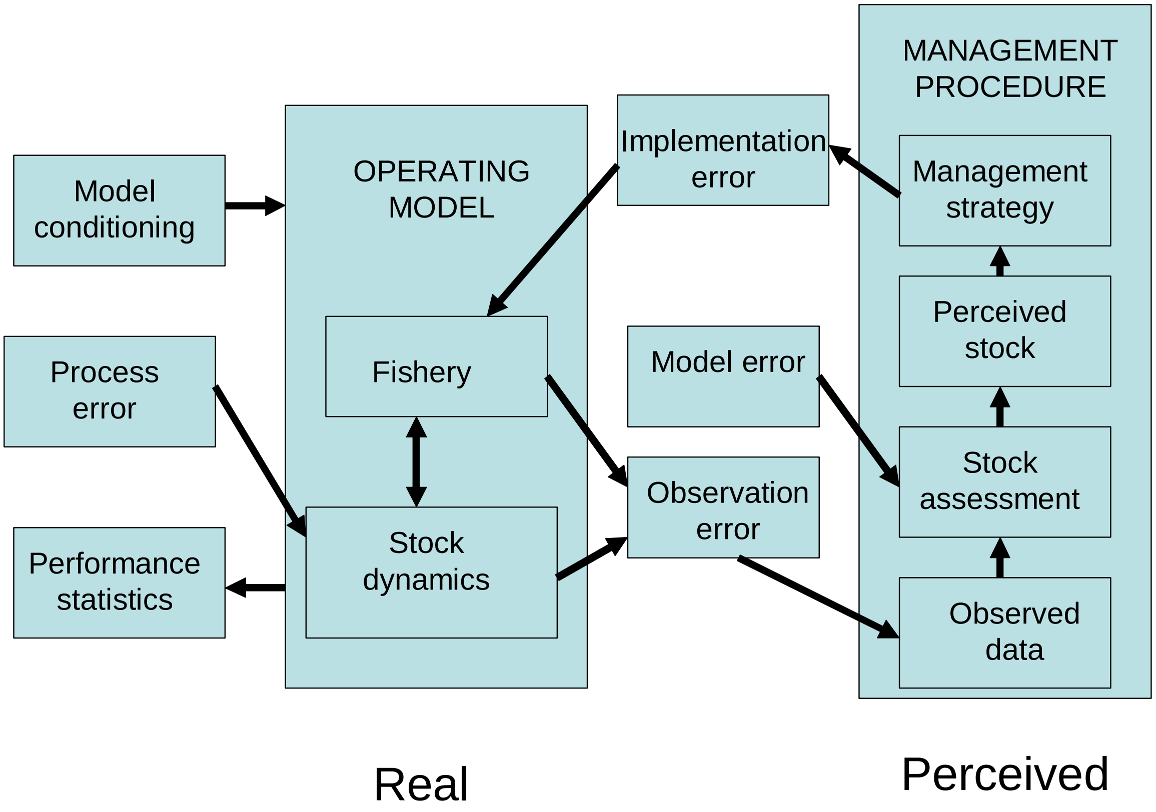

Fig. 1. Conceptual framework for Management Strategy Evaluation (MSE), adapted from Kell *et al.* (2007). The simulated “real” world is captured by the operating model and may contain process error (stochastic simulations) or not (deterministic simulations). The management strategy is applied to the “perceived world” which may only be known with error, either observation error alone, or observation and model error.

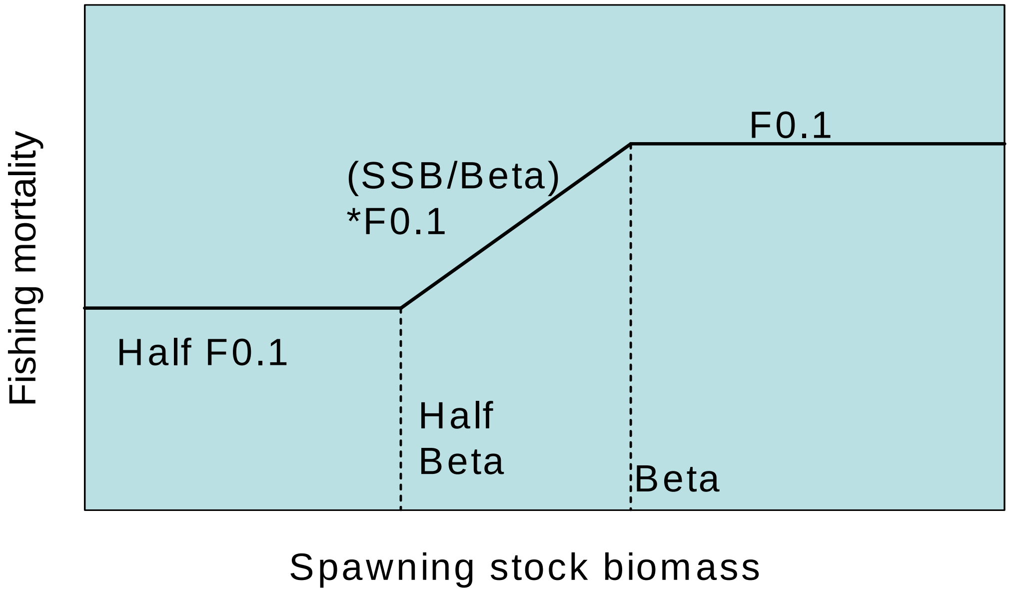

Fig. 2. A graphical representation of the XSA-based PA variable *F* Management Strategy.

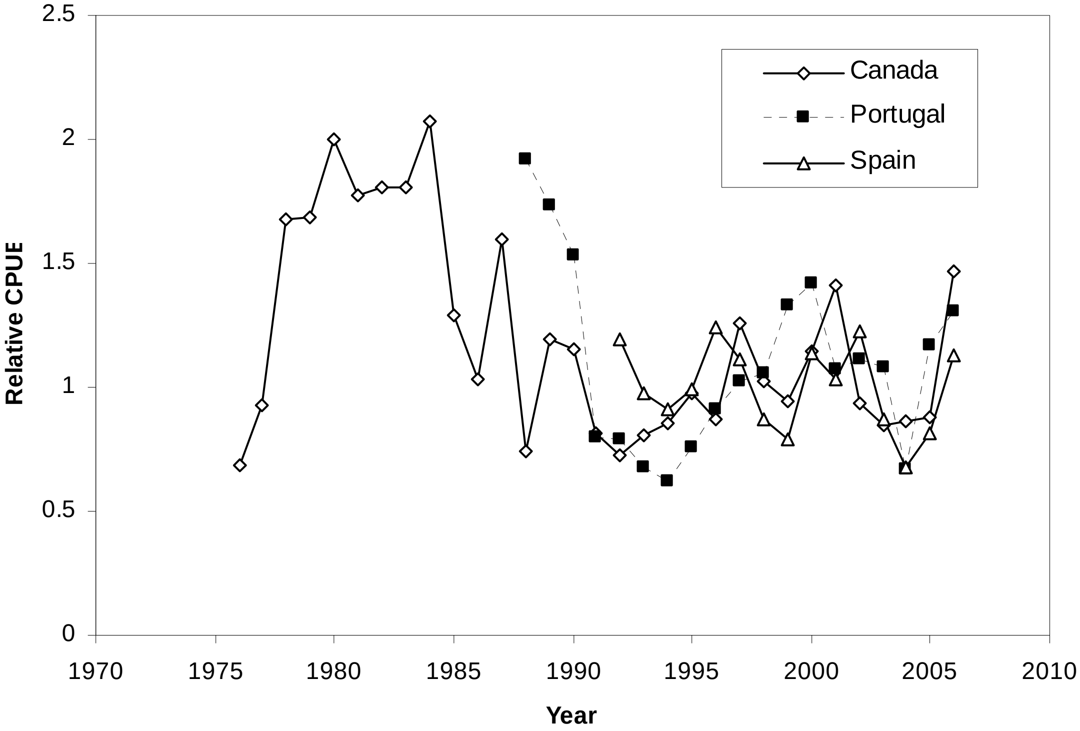

Fig. 3. Commercial CPUE indices for the Canadian, Portuguese and Spanish fleets. CPUEs have been scaled to the 1992-2006 average for each index. Only the Canadian and Spanish CPUE series were age-dissaggreated and used in the CPUE XSA.

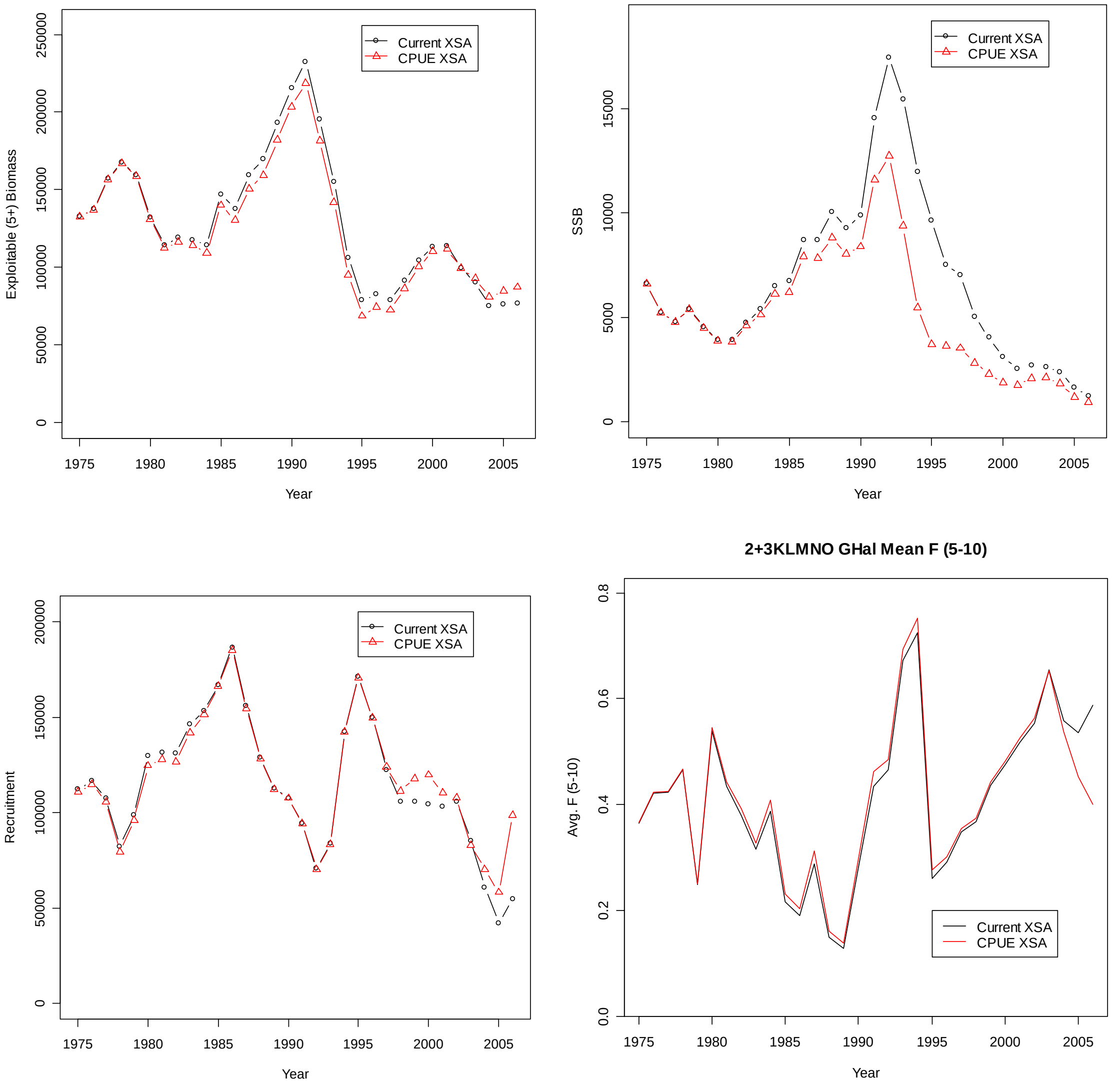

Fig. 4. Comparisons of stock status between the current NAFO XSA (black) and the XSA utilizing CPUE indices (red). Clockwise from top-left: Exploitable (5+) biomass; spawner stock biomass; recruitment (age 1); and mean *F* (ages 5-10).

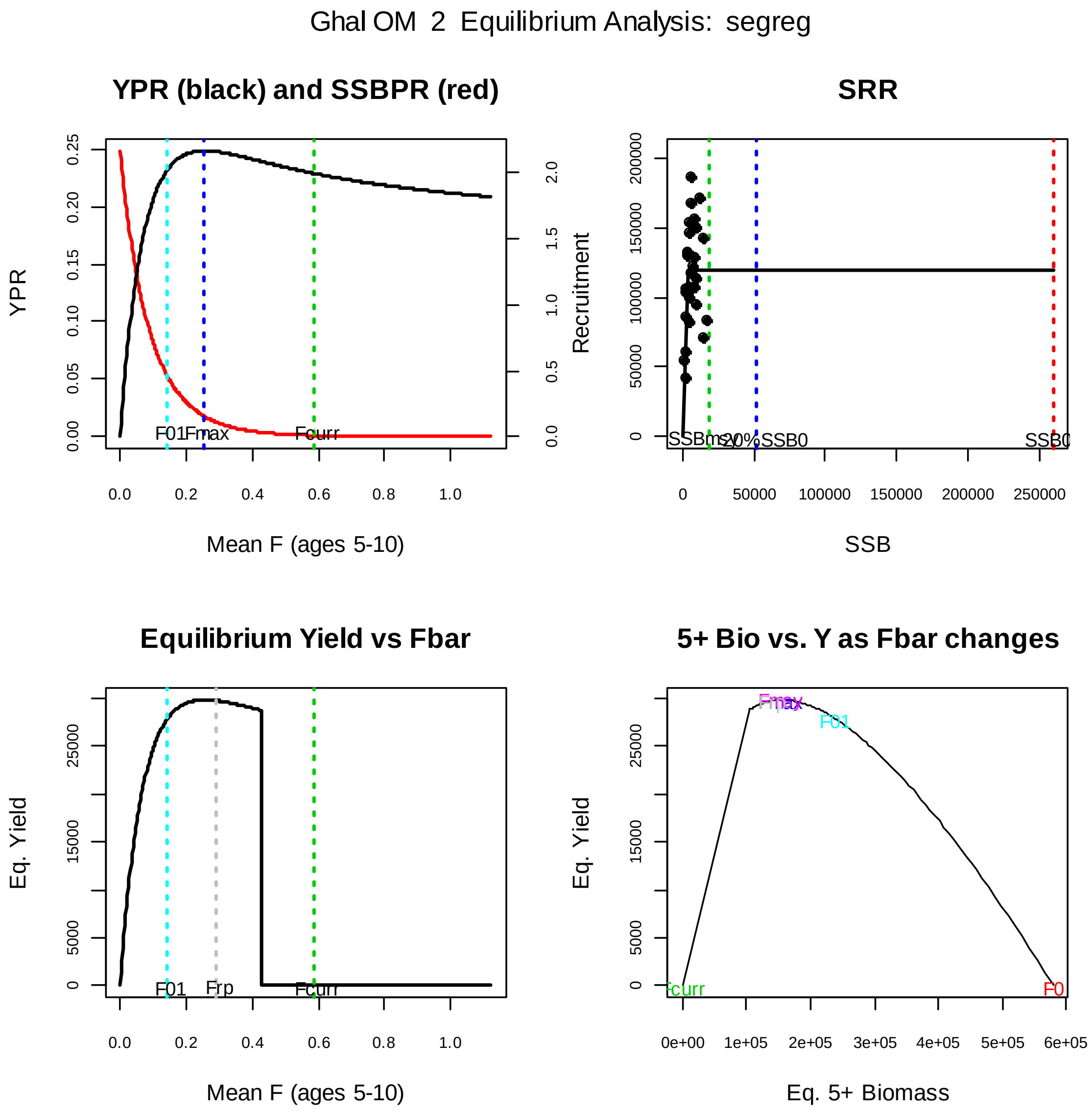

Fig. 5. Biological characteristics of the stock for Operating Model 2.

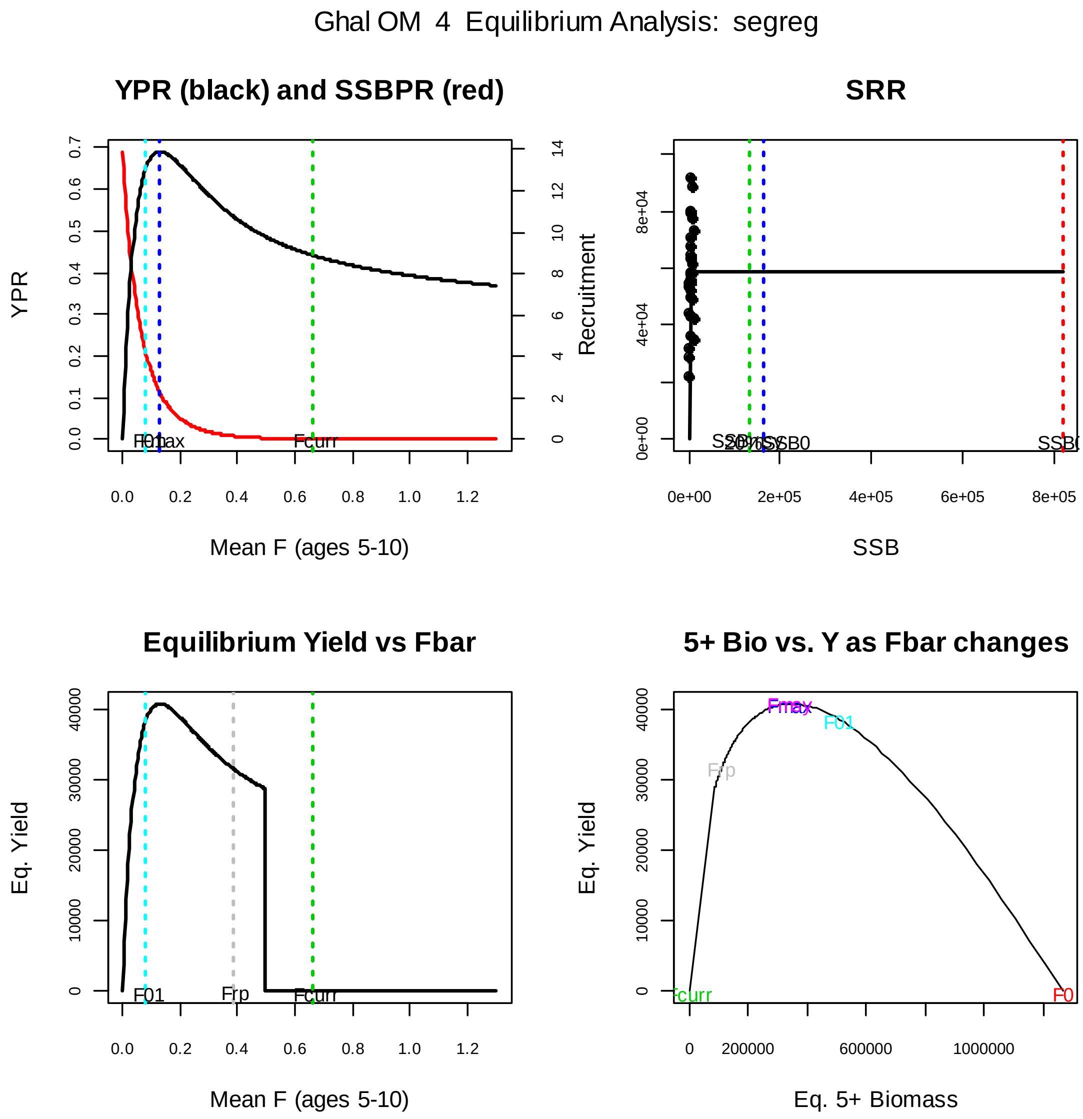

Fig. 6. Biological characteristics of the stock for Operating Model 4.

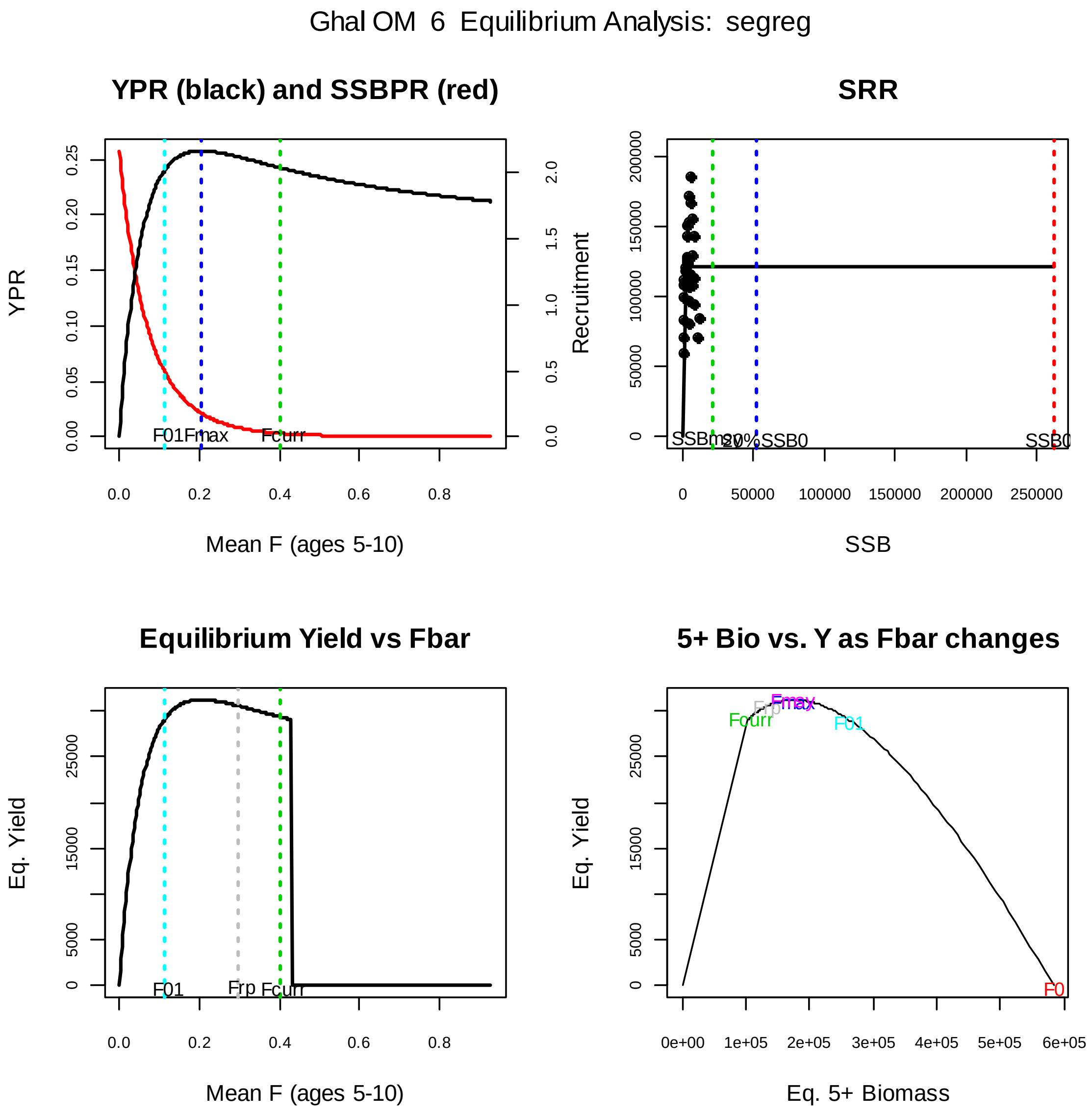

Fig. 7. Biological characteristics of the stock for Operating Model 6.

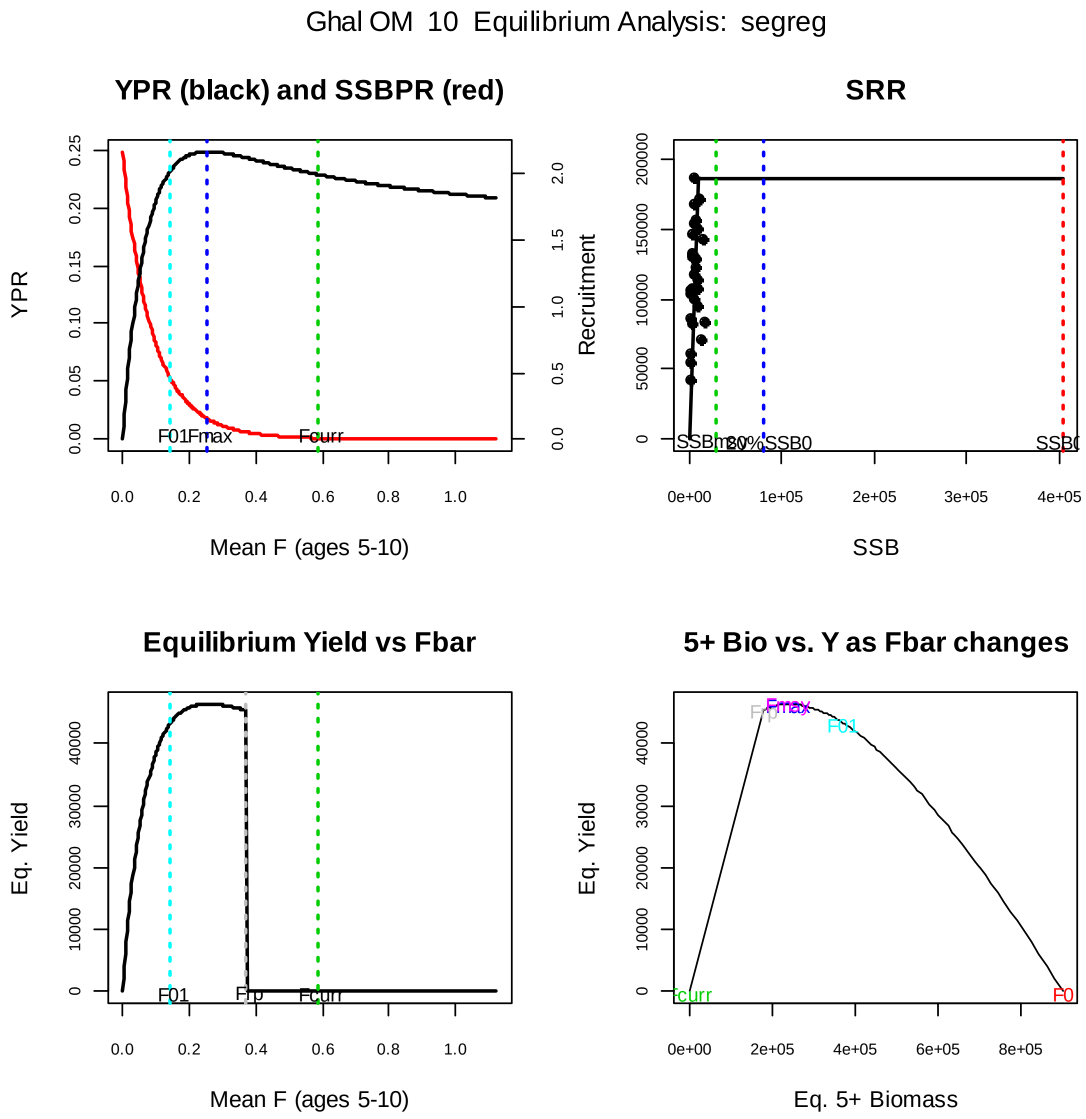

Fig. 8. Biological characteristics of the stock for Operating Model 10.

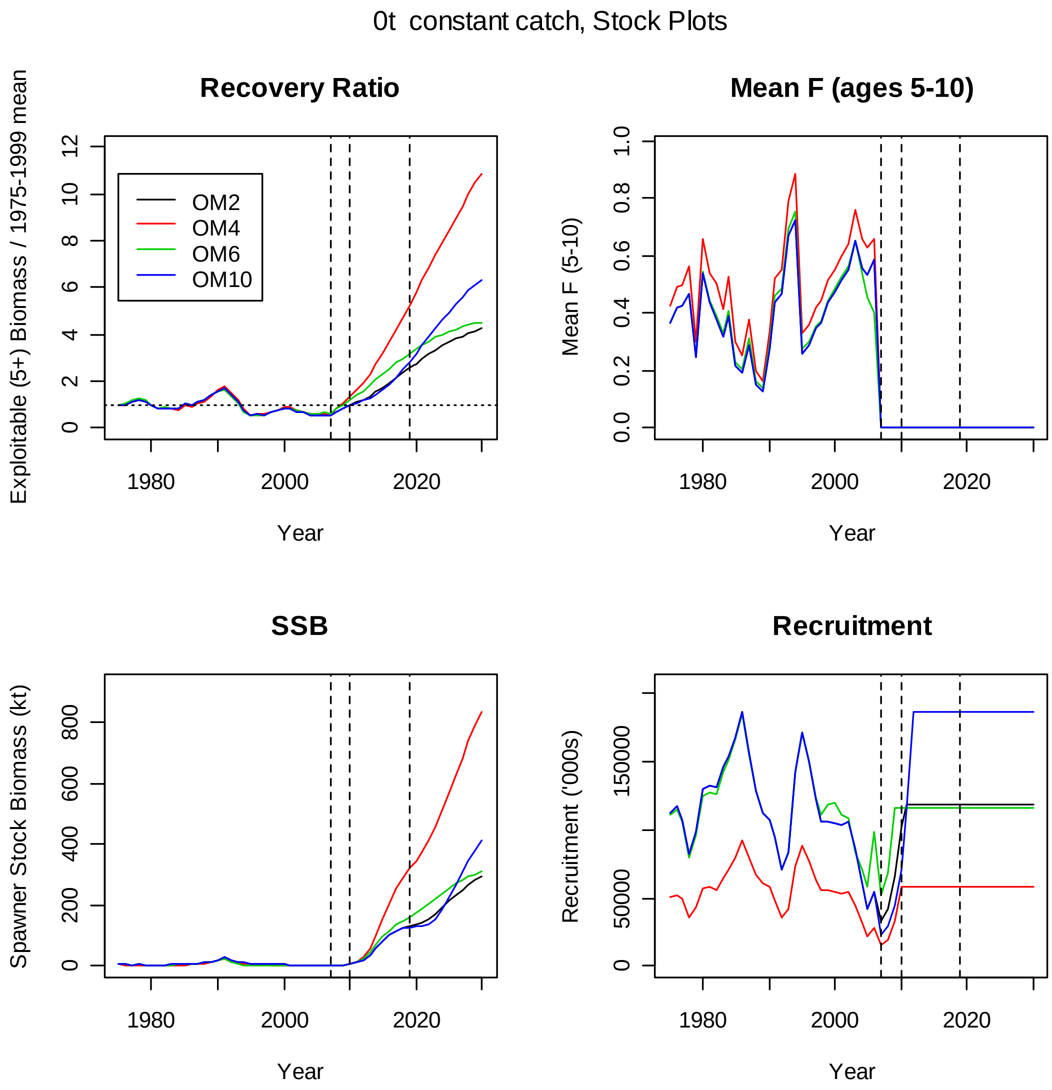

Fig. 9. Plots of time series of various stock trends for OMs 2, 4, 6 and 10 from deterministic projections using a 0t constant catch strategy.

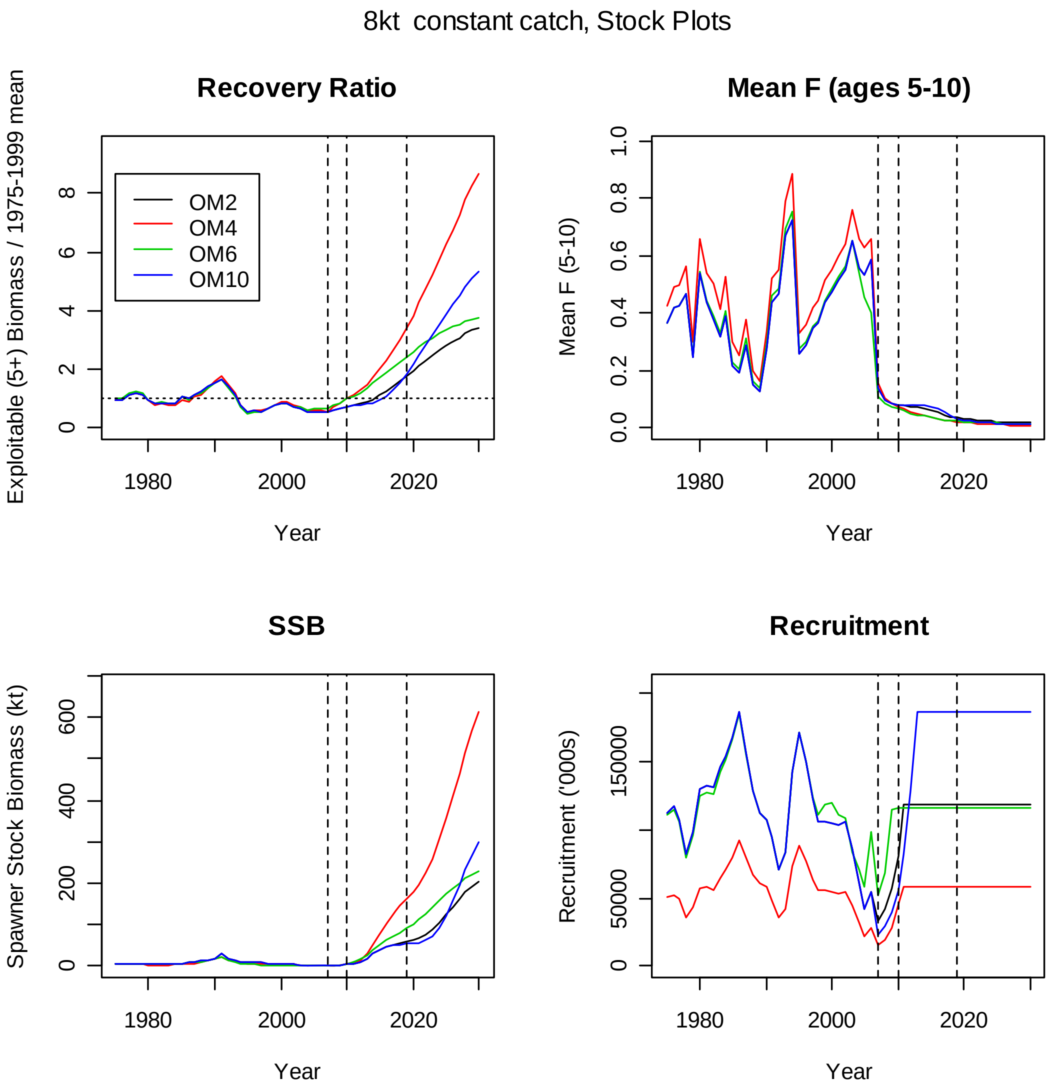

Fig. 10. Plots of time series of various stock trends for OMs 2, 4, 6 and 10 from deterministic projections using an 8 kt constant catch strategy.

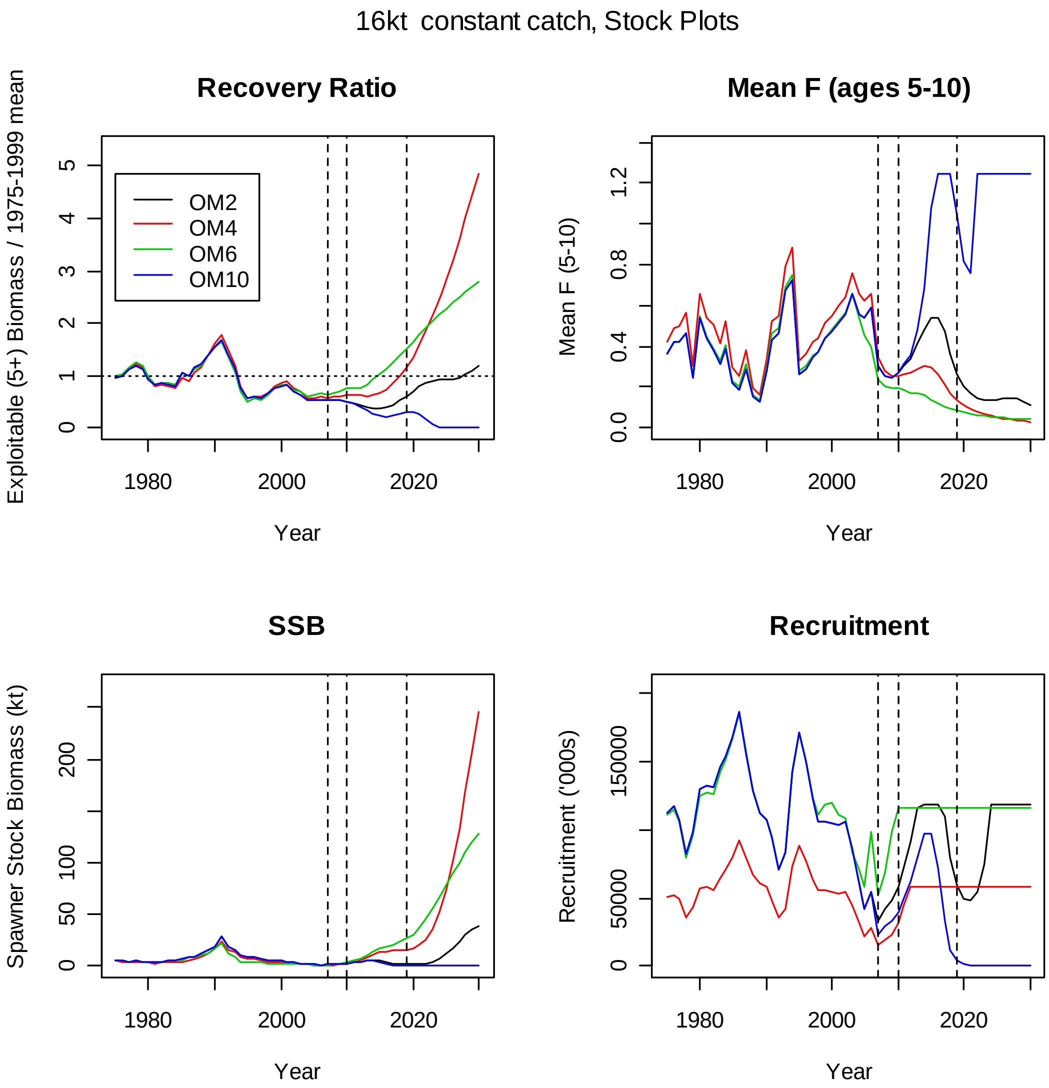

Fig. 11. Plots of time series of various stock trends for OMs 2, 4, 6 and 10 from deterministic projections using a 16 kt constant catch strategy.

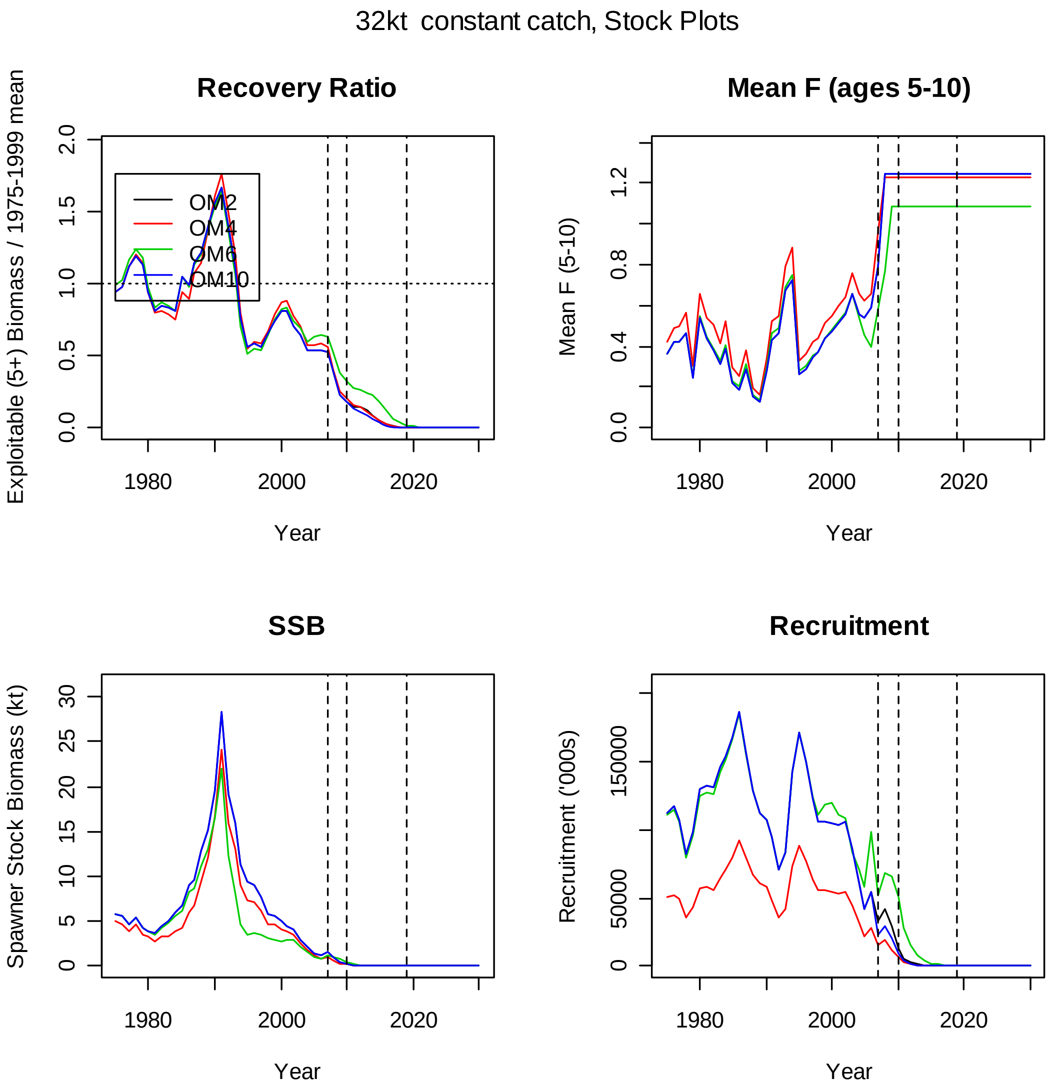

Fig. 12. Plots of time series of various stock trends for OMs 2, 4, 6 and 10 from deterministic projections using a 32 kt constant catch strategy.

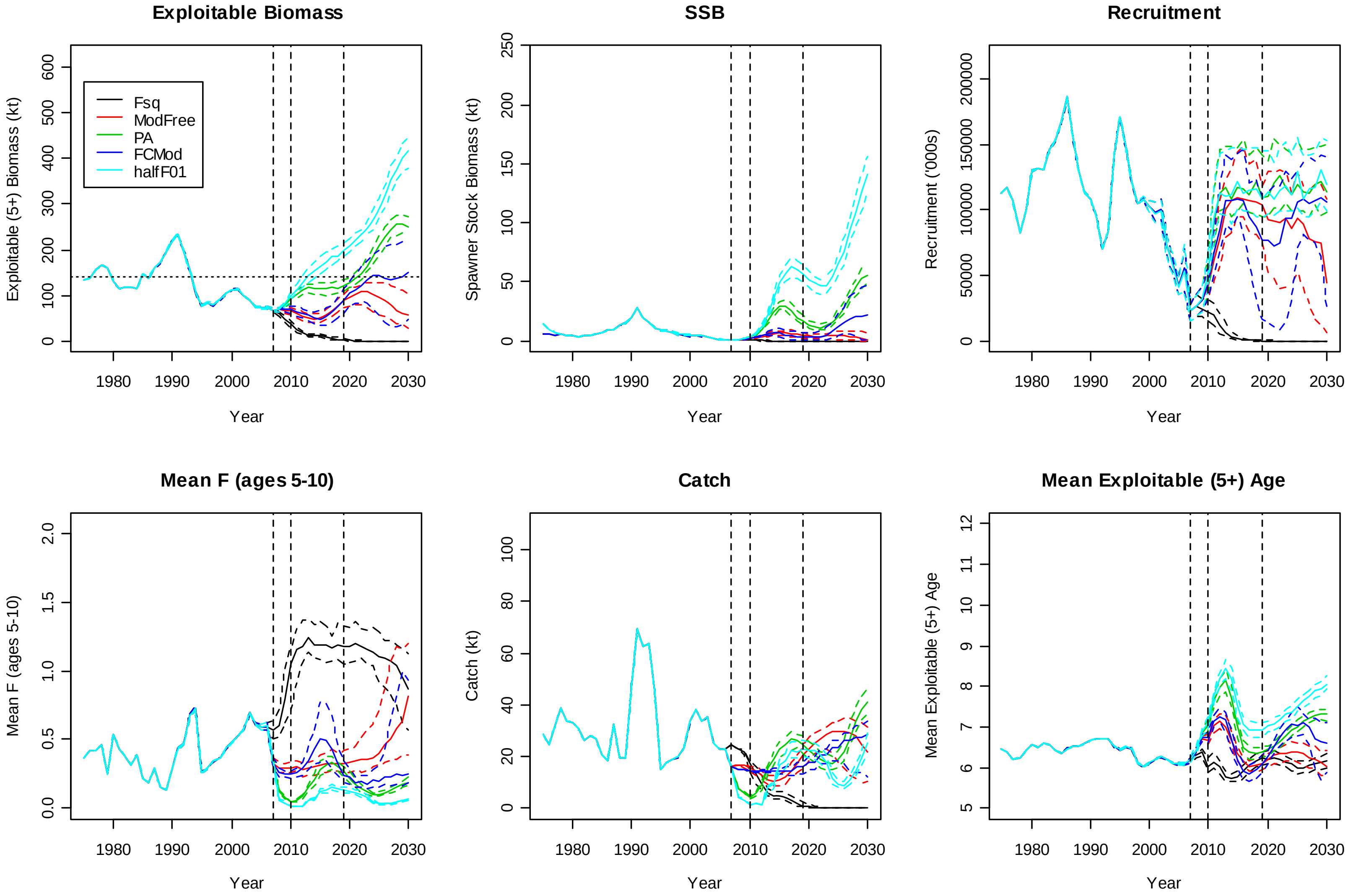

Fig. 13. Descriptive statistics from the stochastic simulations (100 runs) for the five Management Strategies in Operating Model 2 (solid lines represent the median value, dashed lines show the 25 and 75 percentiles).

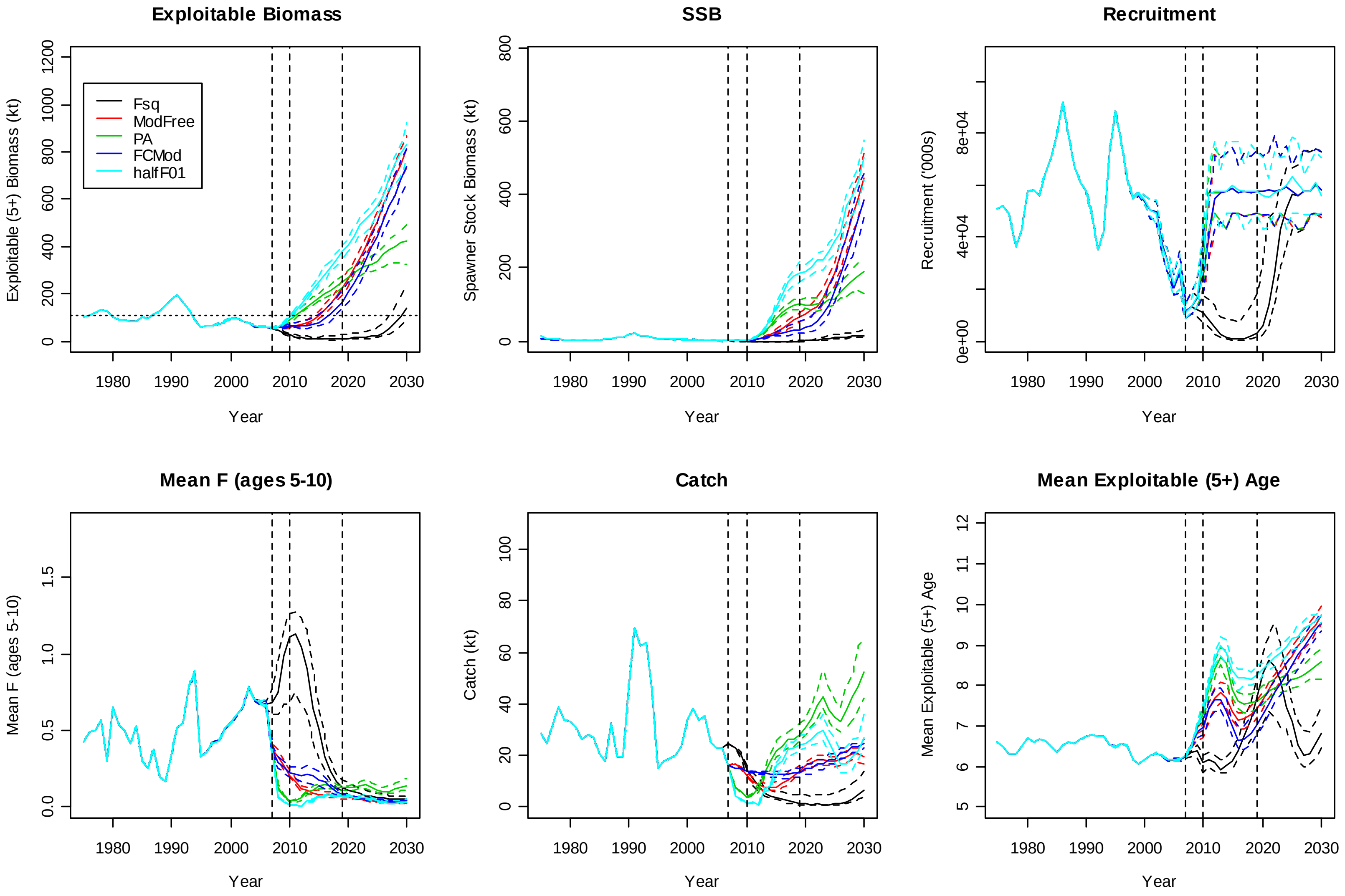

Fig. 14. Descriptive statistics from the stochastic simulations (100 runs) for the five Management Strategies in Operating Model 4 (solid lines represent the median value, dashed lines show the 25 and 75 percentiles).

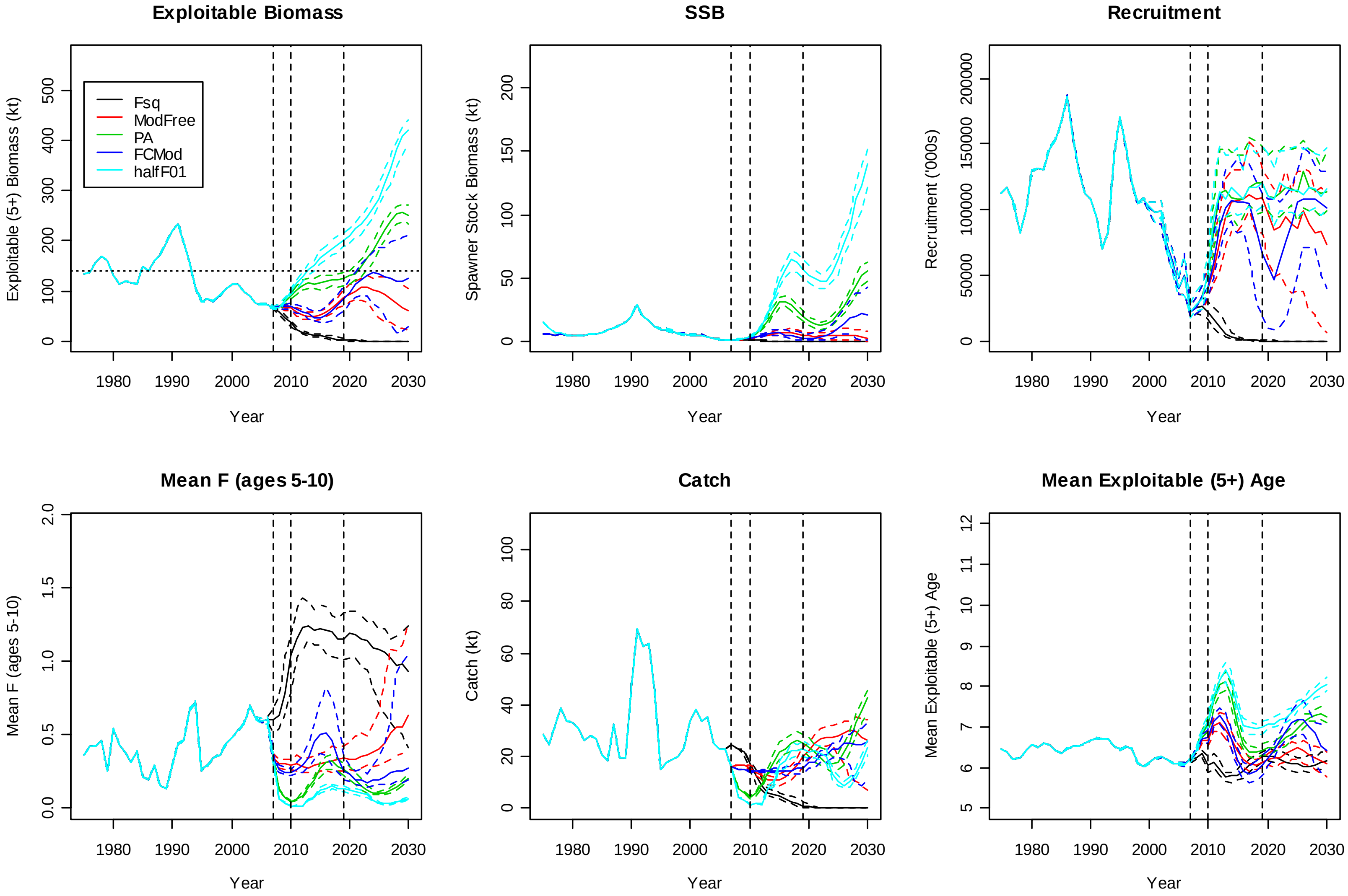

Fig. 15. Descriptive statistics from the stochastic simulations (100 runs) for the five Management Strategies in Operating Model 6 (solid lines represent the median value, dashed lines show the 25 and 75 percentiles).

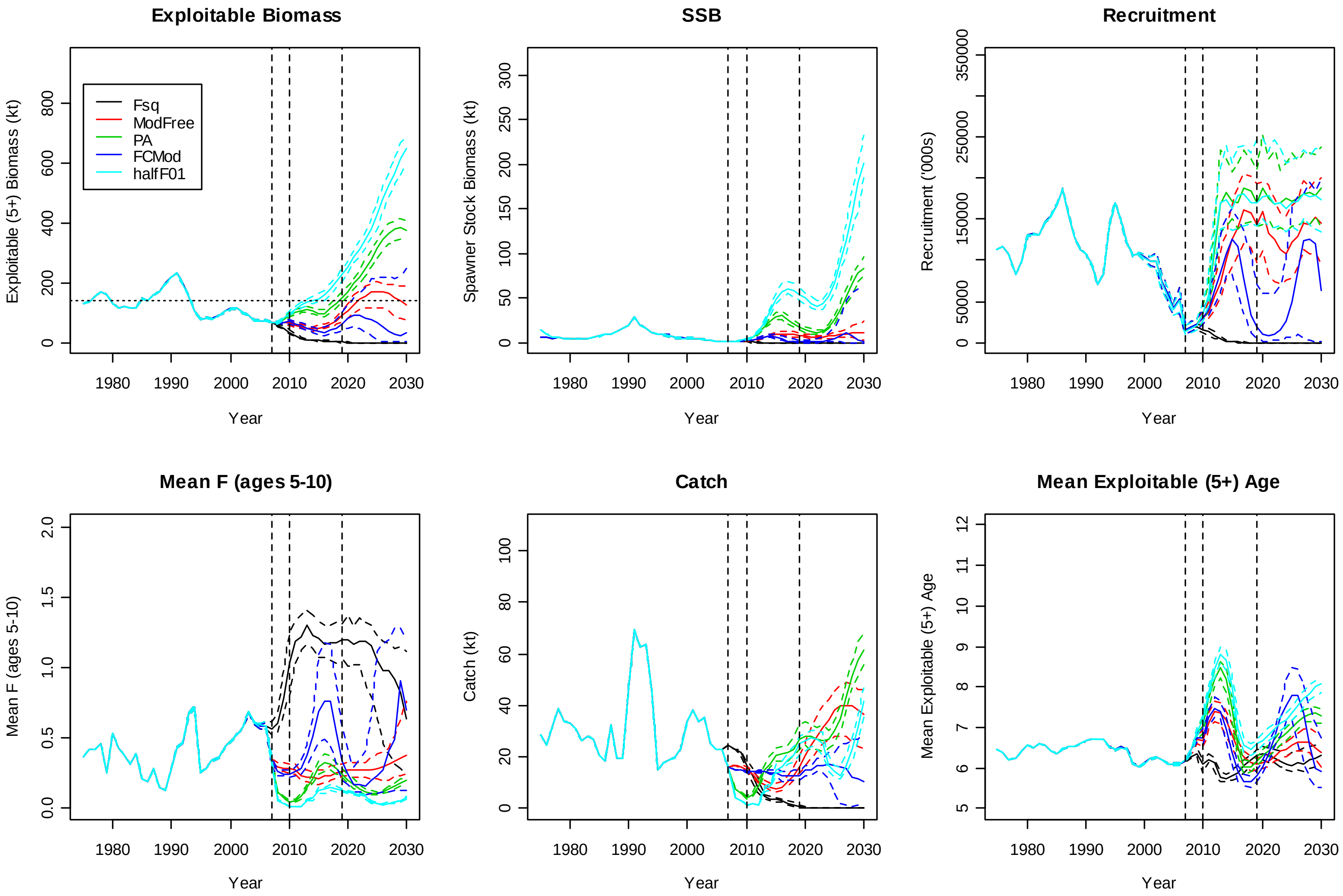

Fig. 16. Descriptive statistics from the stochastic simulations (100 runs) for the five Management Strategies in Operating Model 10 (solid lines represent the median value, dashed lines show the 25 and 75 percentiles).

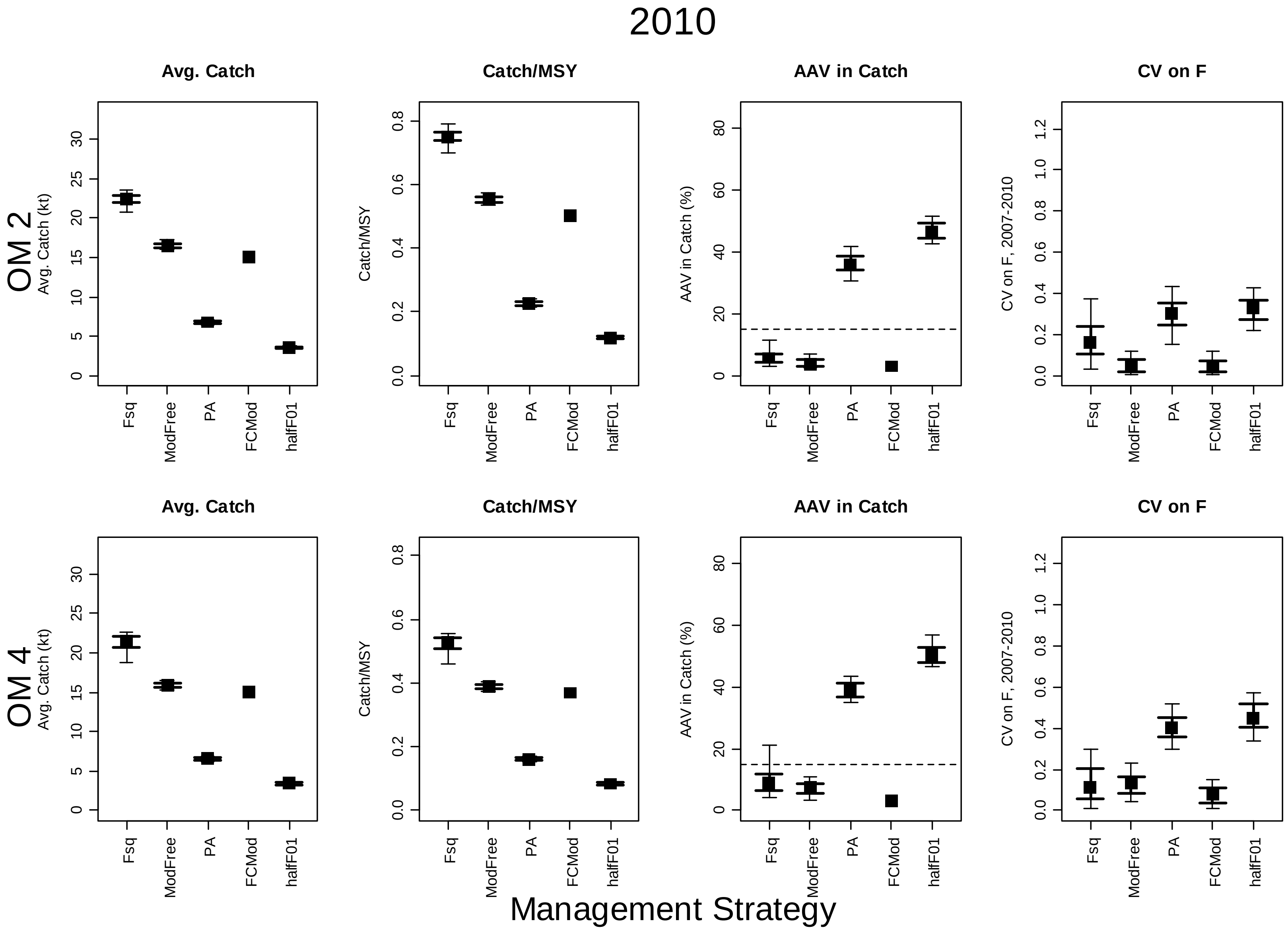

Fig. 17. Fishery performance statistics for the five Management Strategies, short term – 2010 (Operating Models 2 and 4). Points represent the medians and whiskers show the 5, 25, 75 and 95 percentiles.

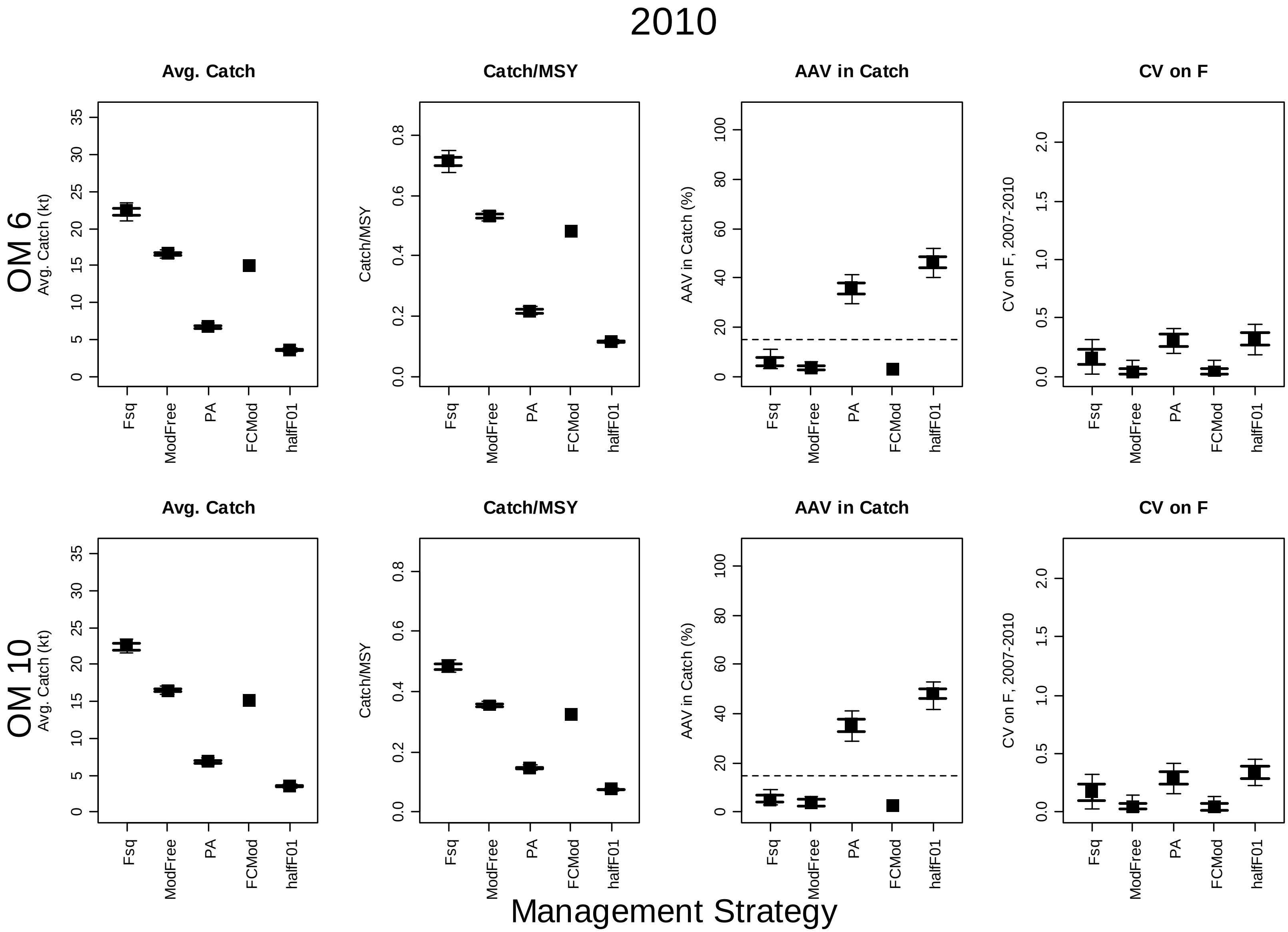

Fig. 17 cont.. (Operating Models 6 and 10).

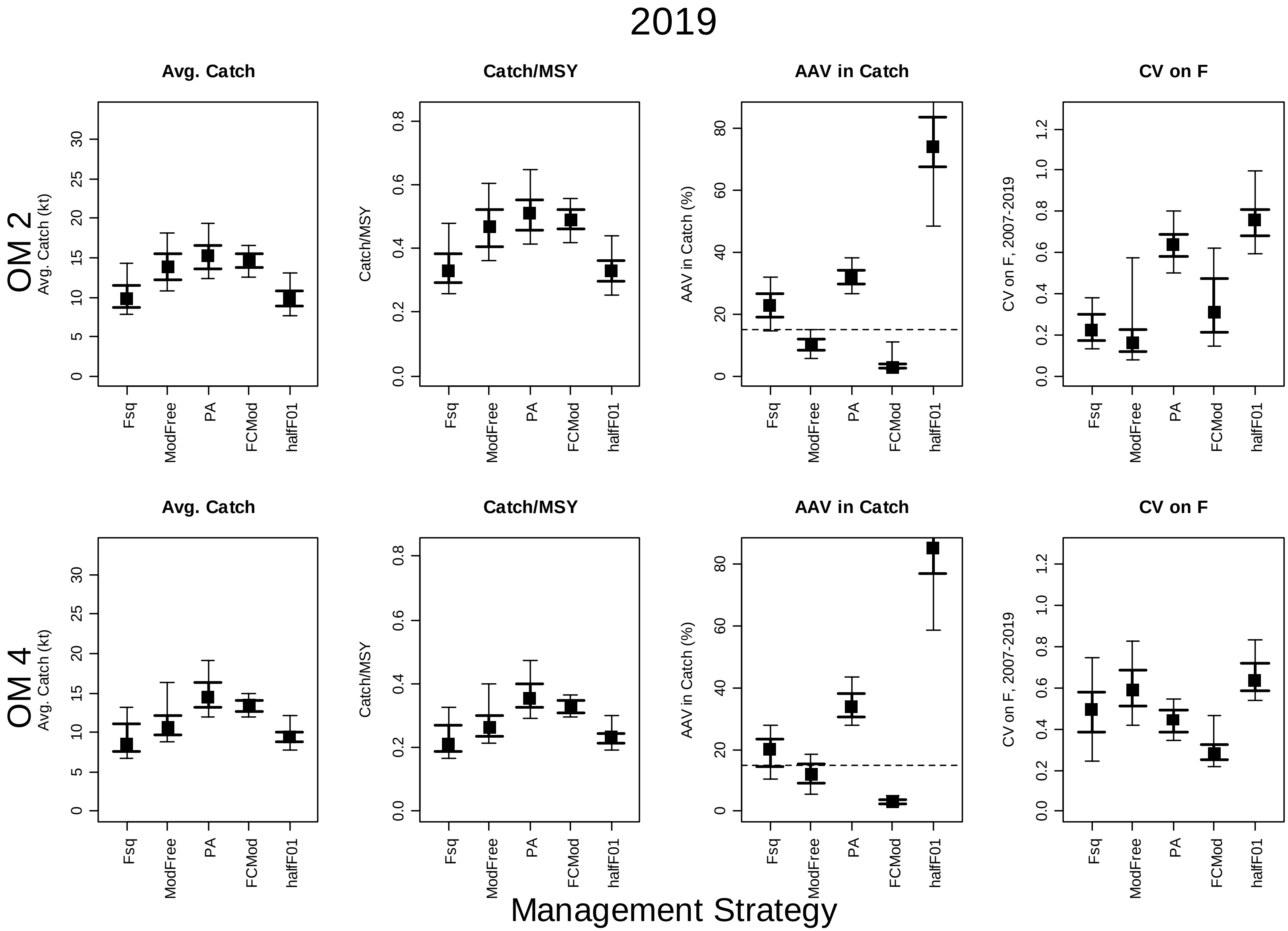

Fig. 18. Fishery performance statistics for the five Management Strategies, rebuilding plan period – 2019 (Operating Models 2 and 4). Points represent the medians and whiskers show the 5, 25, 75 and 95 percentiles.

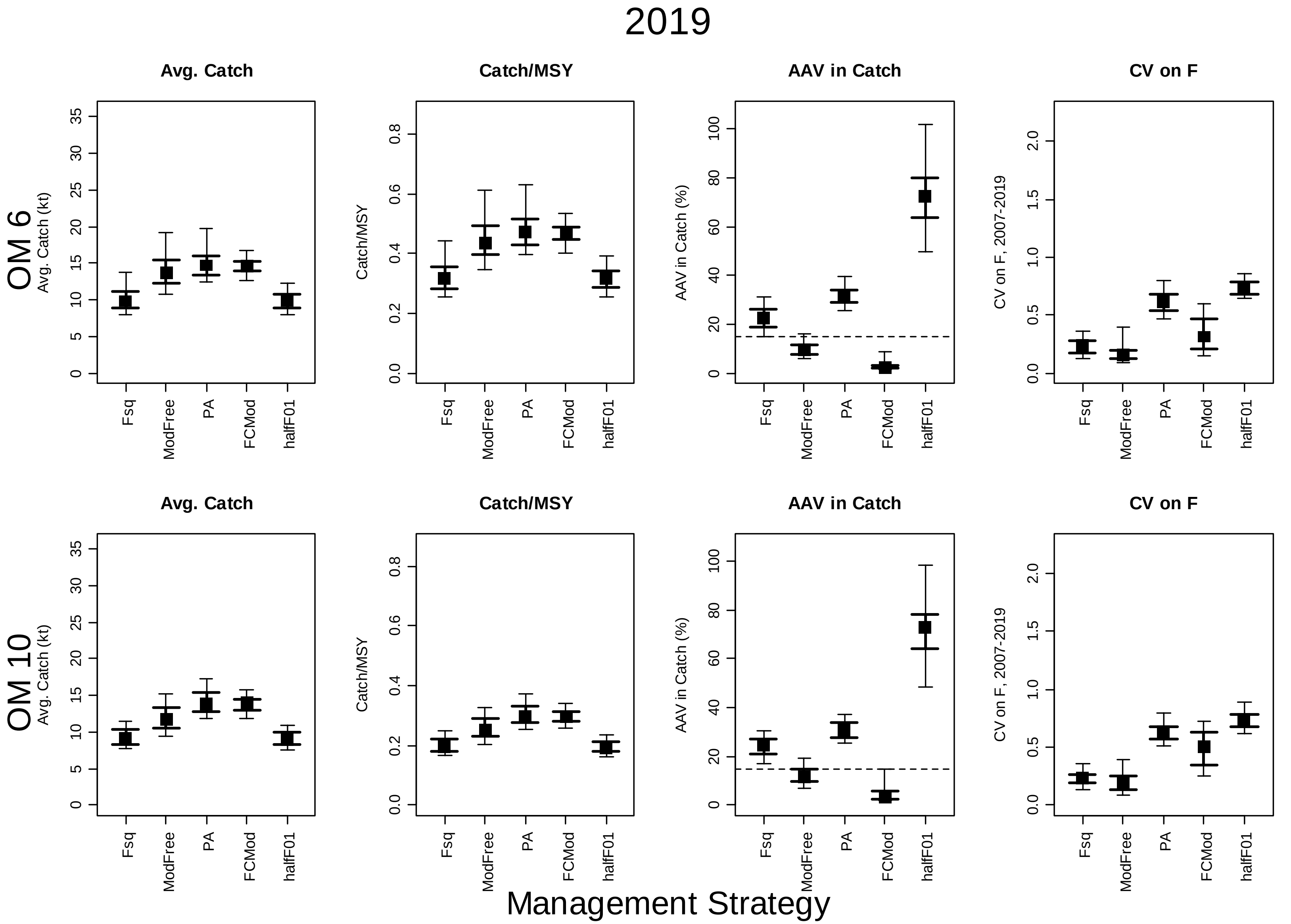

Fig. 18 cont.. (Operating Models 6 and 10).

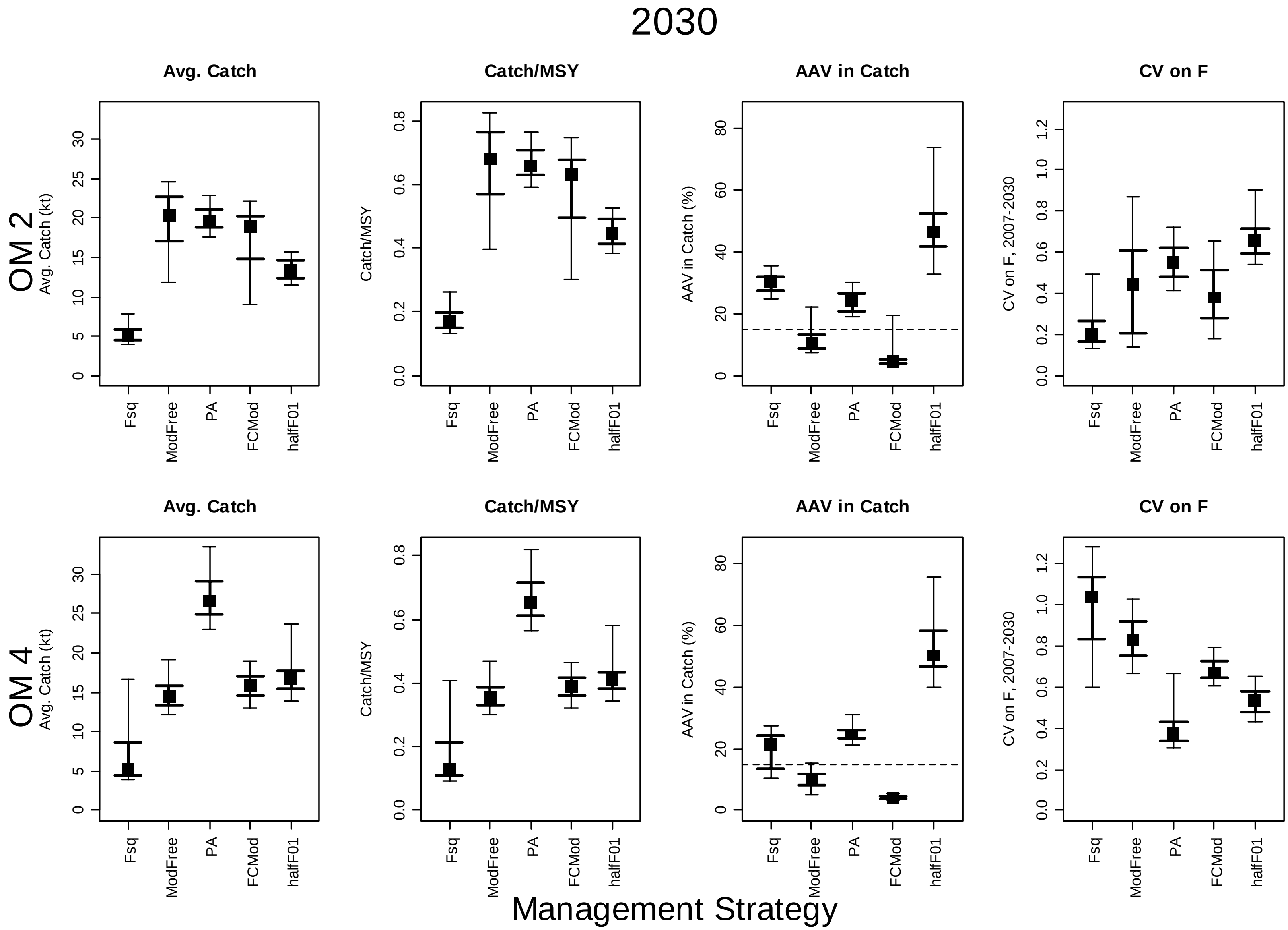

Fig. 19. Fishery performance statistics for the five Management Strategies, long term – 2030 (Operating Models 2 and 4). Points represent the medians and whiskers show the 5, 25, 75 and 95 percentiles.

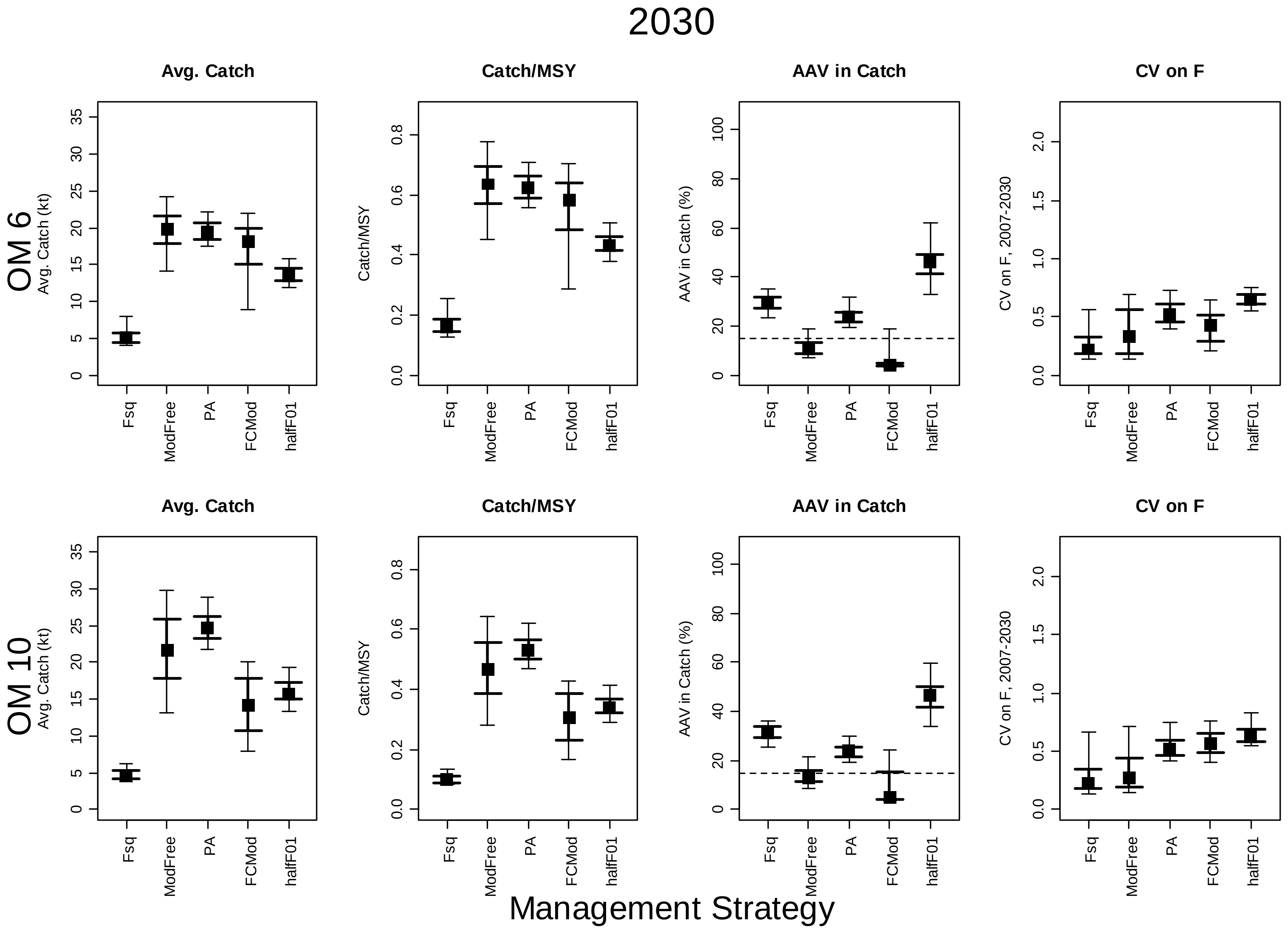

Fig. 19 cont. (Operating Models 6 and 10).

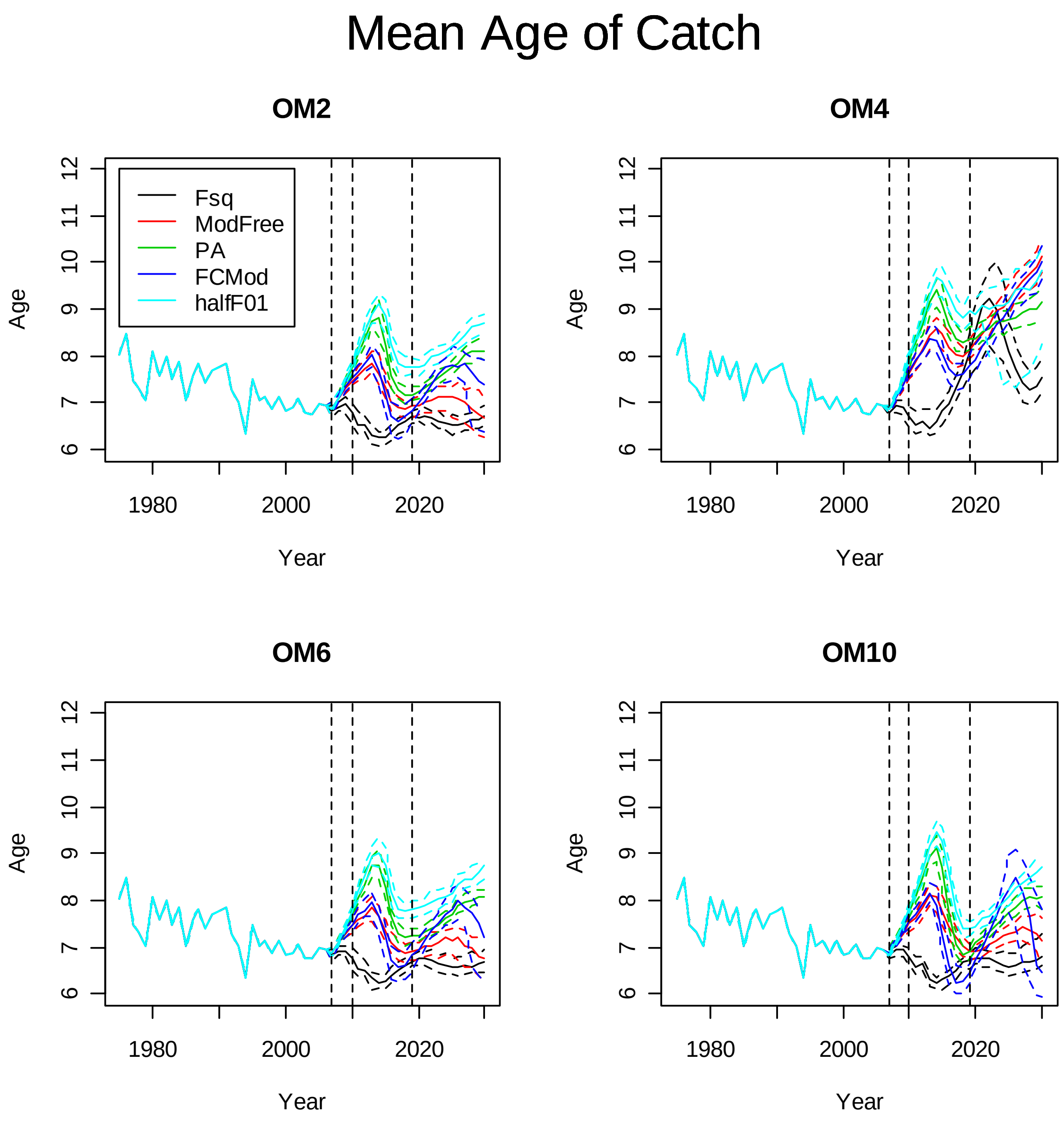

Fig 20. Mean weighted age in the annual catch for each Management Strategy in each Operating Model. Points represent the medians and whiskers show the 5, 25, 75 and 95 percentiles.

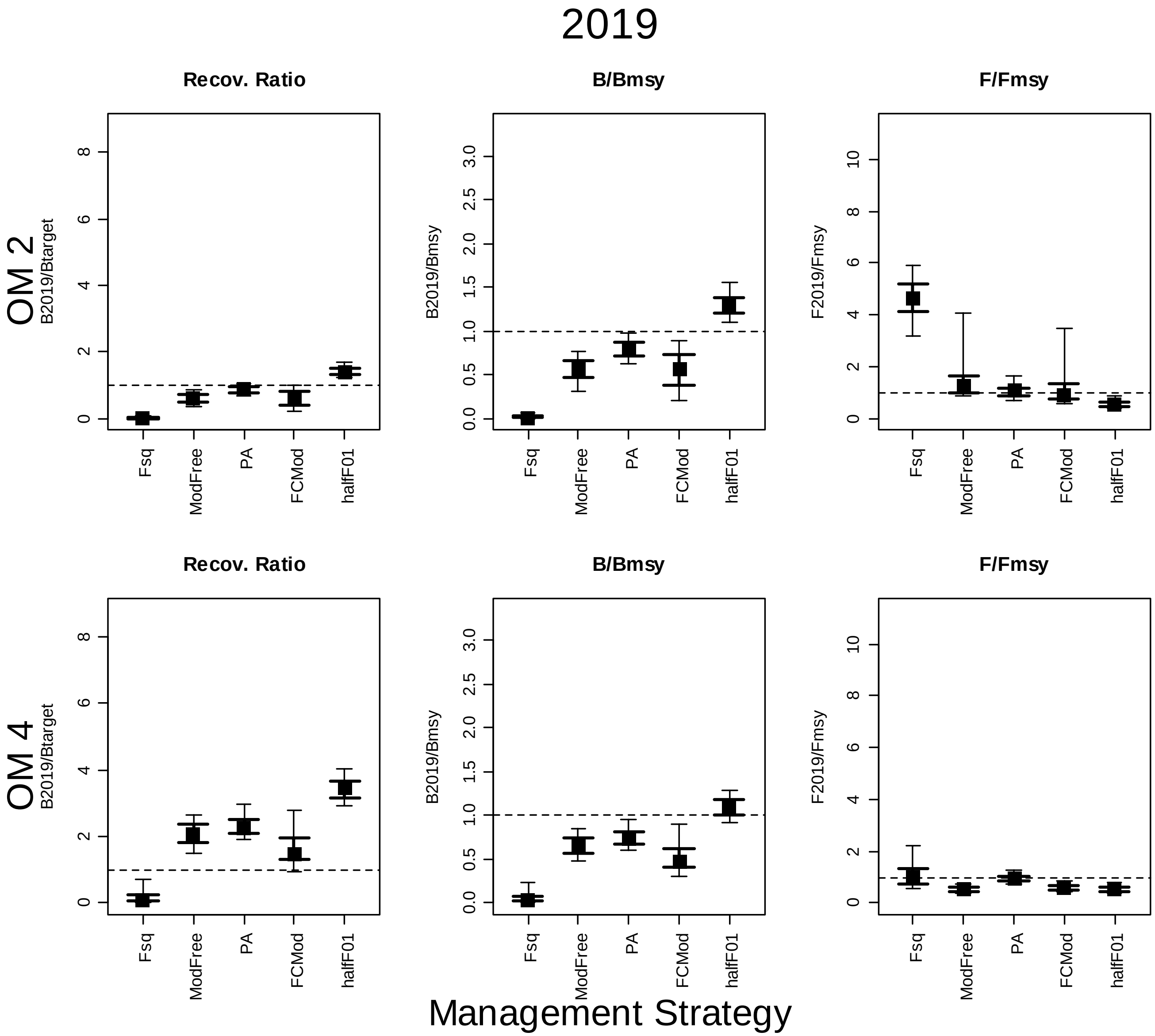

Fig 21. Stock performance statistics for the five Management Strategies, rebuilding plan period – 2019 (Operating Models 2 and 4). Points represent the medians and whiskers show the 5, 25, 75 and 95 percentiles.

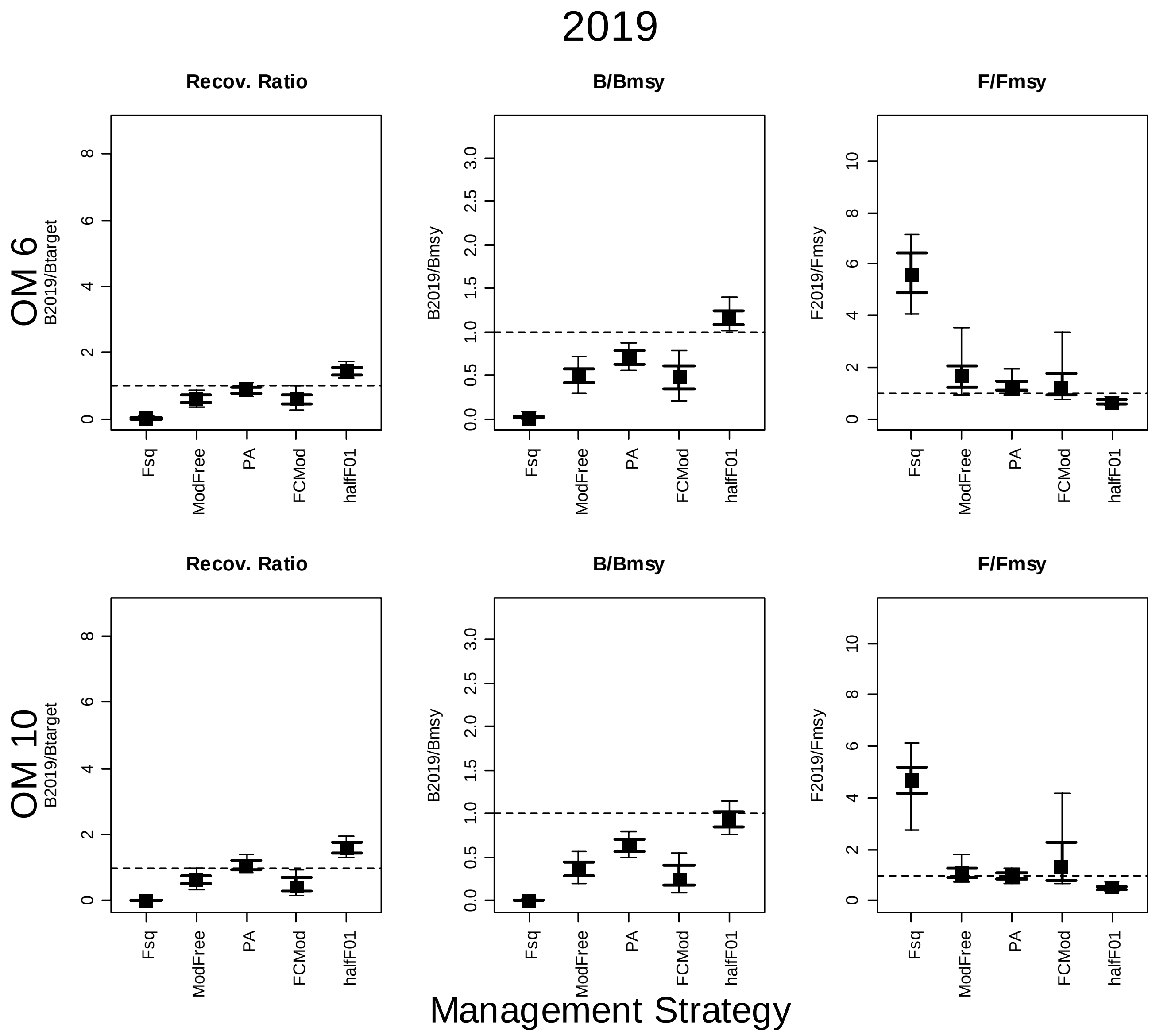

Fig. 21 cont.. (Operating Models 6 and 10).

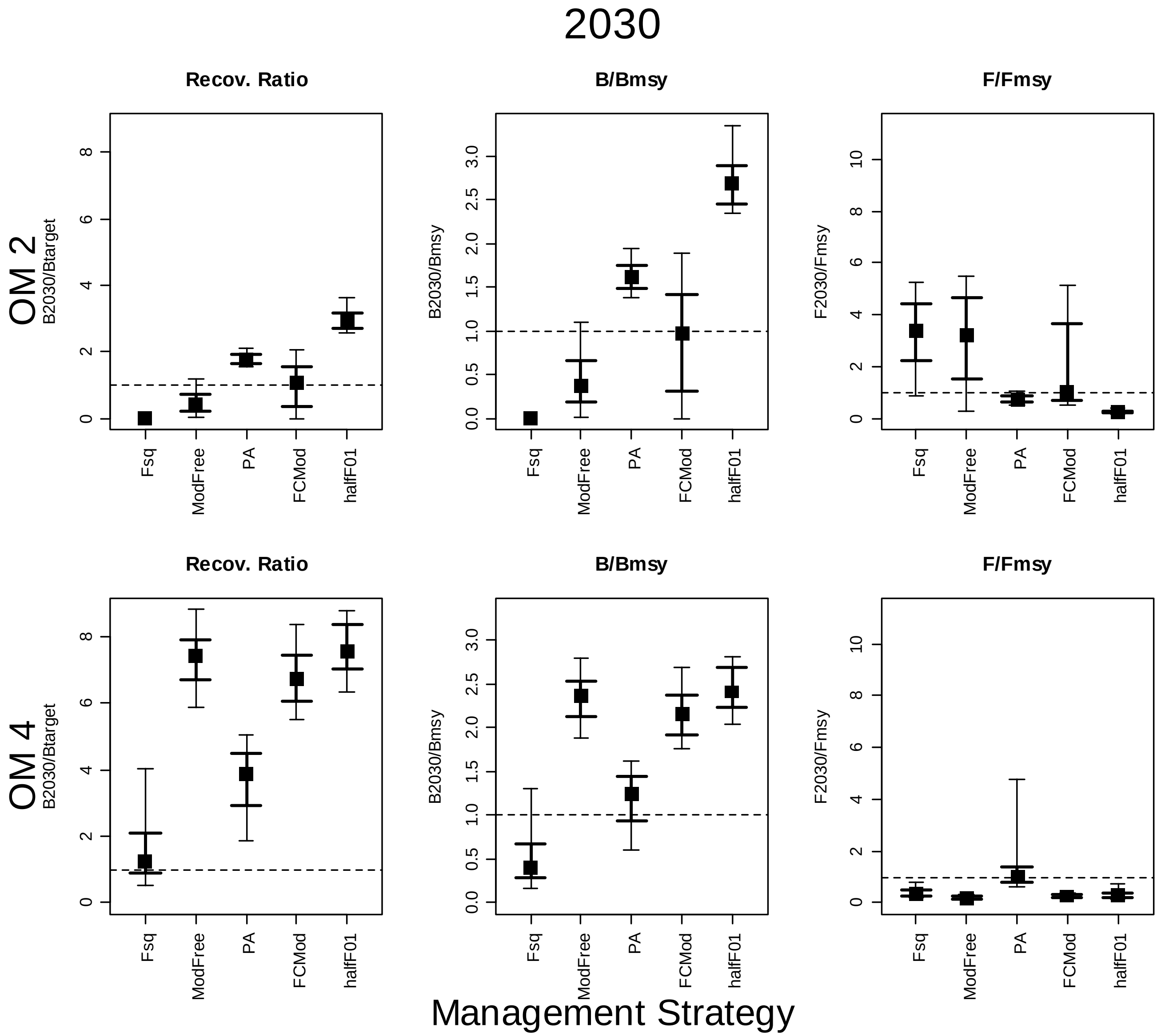

Fig 22. Stock performance statistics for the five Management Strategies, long term – 2030 (Operating Models 2 and 4). Points represent the medians and whiskers show the 5, 25, 75 and 95 percentiles.

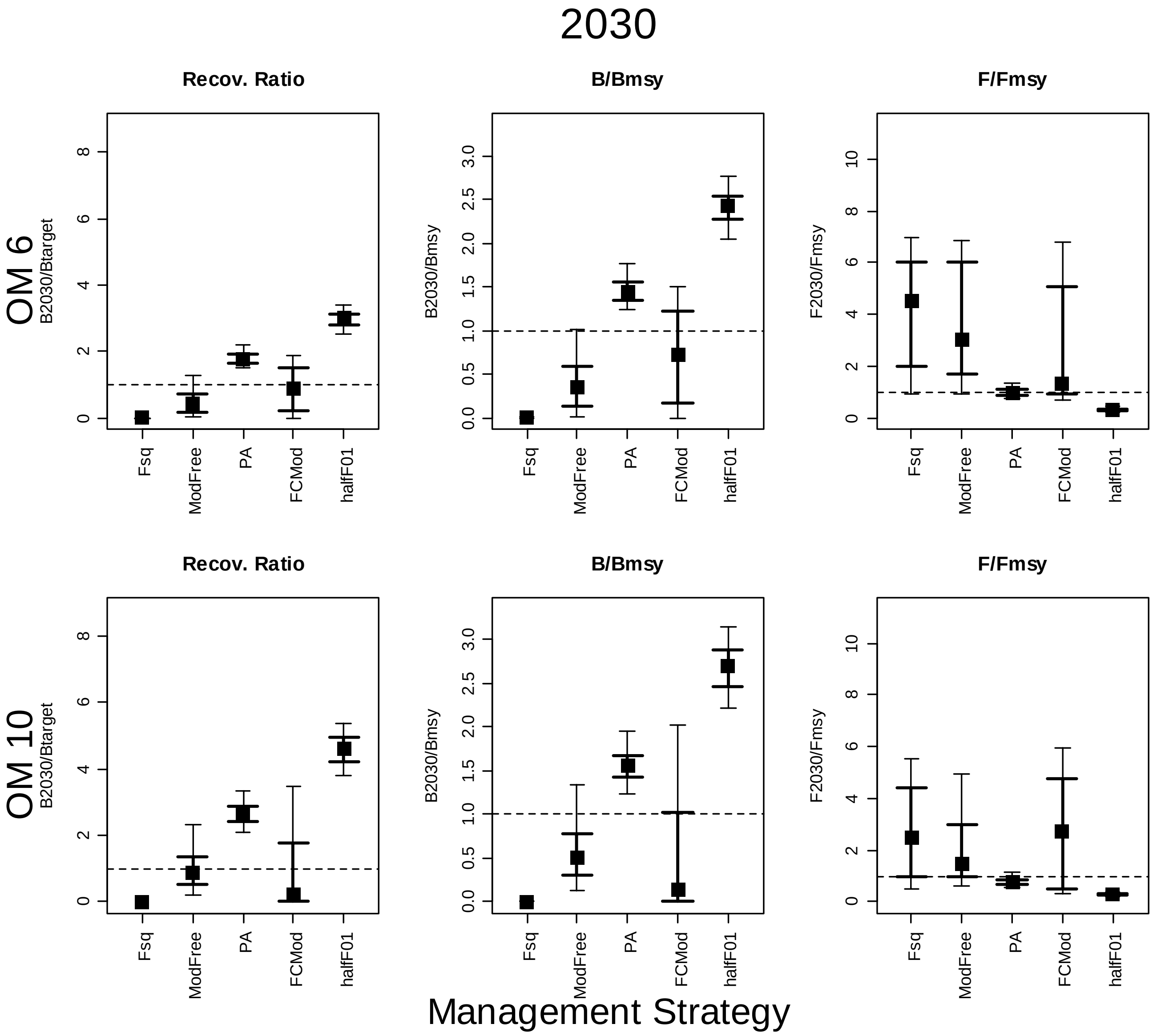

Fig. 22 cont.. (Operating Models 6 and 10).

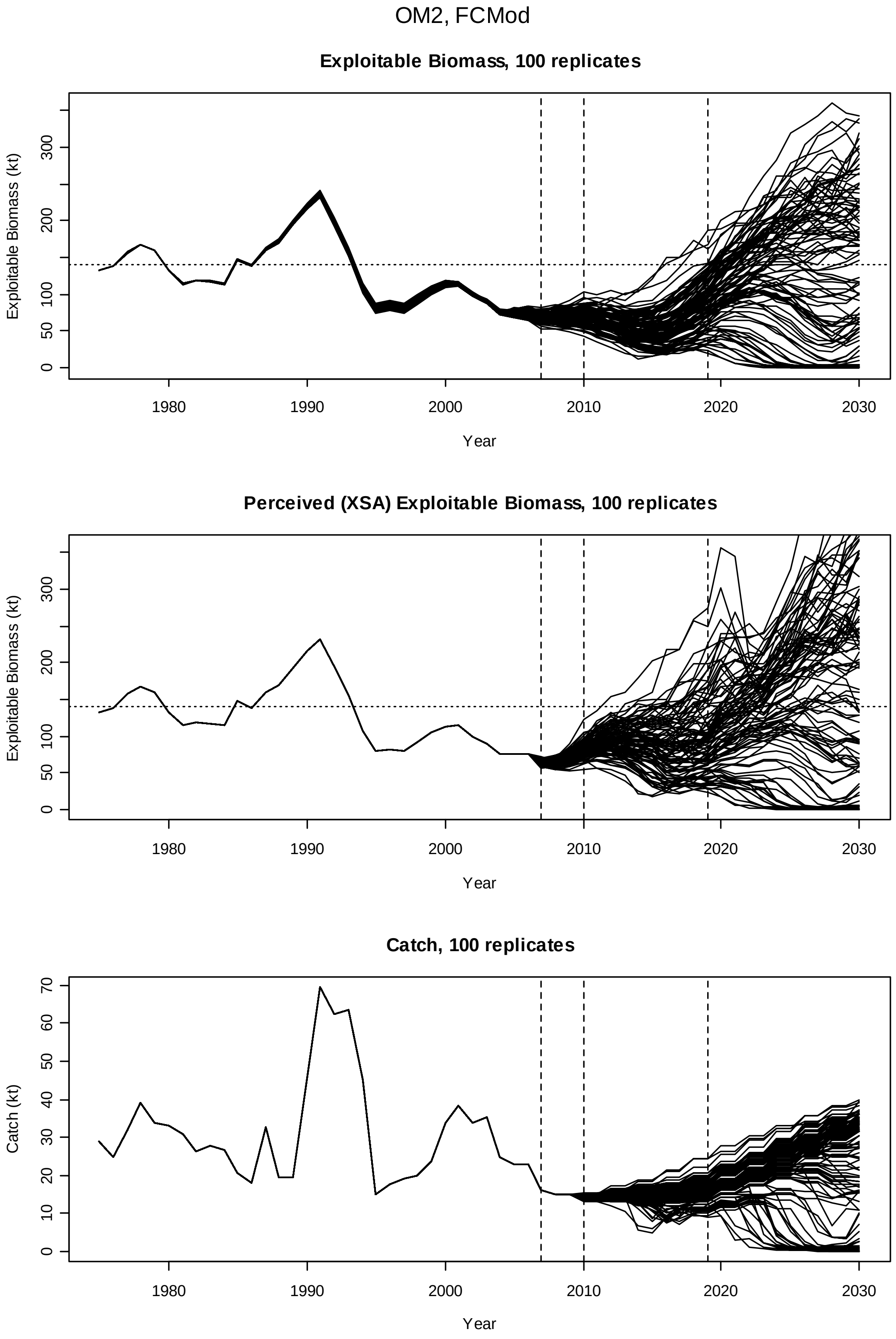

Fig 23. Plots of each individual replicate (*n*=100) from the stochastic simulations (worm plots) for the FCMod Management Strategy in Operating Model 2. The top panel shows the true population exploitable biomass through time, the middle panel shows how this was perceived by the XSA, and the bottom panel shows the catch.

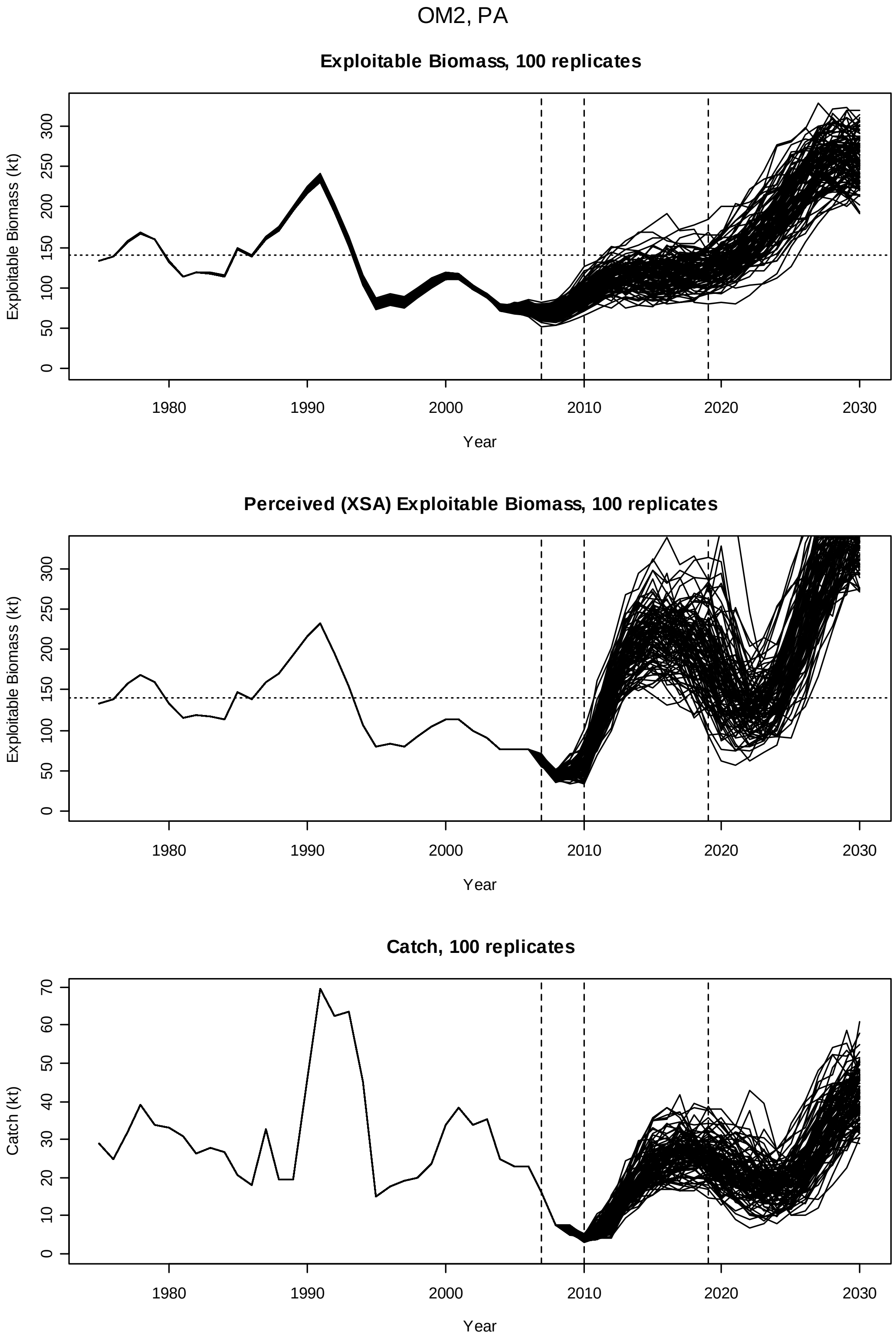

Fig 24. Plots of each individual replicate (*n*=100) from the stochastic simulations (worm plots) for the PA Management Strategy in Operating Model 2. The top panel shows the true population exploitable biomass through time, the middle panel shows how this was perceived by the XSA, and the bottom panel shows the catch.