WORKSHOP ON GUIDELINES FOR MANAGEMENT STRATEGY EVALUATIONS (WKGMSE2)

VOLUME 1 | ISSUE 33

ICES SCIENTIFIC REPORTS

RAPPORTS SCIENTIFIQUES DU CIEM

ICES INTERNATIONAL COUNCIL FOR THE EXPLORATION OF THE SEA
CIEM CONSEIL INTERNATIONAL POUR L'EXPLORATION DE LA MER

International Council for the Exploration of the Sea
Conseil International pour l'Exploration de la Mer

H.C. Andersens Boulevard 44-46
DK-1553 Copenhagen V
Denmark
Telephone (+45) 33 38 67 00
Telefax (+45) 33 93 42 15
www.ices.dk
info@ices.dk

ISSN number: 2618-1371 I © 2019 International Council for the Exploration of the Sea

ICES Scientific Reports

Volume 1 | Issue 33

WORKSHOP ON GUIDELINES FOR MANAGEMENT STRATEGY EVALUATIONS (WKGMSE2)

Recommended format for purpose of citation:

ICES. 2019. Workshop on Guidelines for Management Strategy Evaluations (WKGMSE2). ICES Scientific Reports. 1:33. 162 pp. http://doi.org/10.17895/ices.pub.5331

Editor

Fernández Carmen

Authors

Michel Bertignac • Höskuldur Björnsson • Mollie Elisabeth Brooks • Thomas Brunel • Doug Butterworth Andrew Campbell • Santiago Cerviño • Jonathan Deroba • Bjarki Thor Elvarsson • Simon Fischer Dorleta Garcia • Daisuke Goto • Michael Gras • Niels Hintzen • Noél Holmgren • Daniel Howell • Quang Huynh • Ernesto Jardim • Laurence Kell • Christoph Konrad • Sarah B. M. Kraak • Allen R. Kronlund Polina Levontin • Colm Lordan • Hugo Mendes • Tanja Miethe • Iago Mosqueira • José De Oliveira Martin Pastoors • Cecilia Pinto • Claus Reedtz Sparrevohn • Marc Taylor

ICES CIEM International Council for the Exploration of the Sea Conseil International pour l'Exploration de la Mer

Contents

i Executive summary

The purpose of the meeting was to bring up to date the methodologies and technical specifications that should be incorporated in Management Strategy Evaluation (MSE) work in ICES. The workshop was tasked with reviewing recent methodological and practical MSE work conducted in ICES and around the world, as well as the guidelines provided by the 2013 ICES Workshop on Guidelines for Management Strategy Evaluations (WKGMSE). The Terms of Reference indicated that the revision should include all aspects involved in MSE, while paying specific attention to several issues that had been identified through ICES practice. The Terms of Reference also requested WKGMSE 2 to consider how best to disseminate the guidelines to experts within the ICES community and the need for training courses. The workshop addressed all its Terms of Reference.

The main results of the workshop are the revised MSE guidelines, as well as recommendations in relation to the ICES criterion for defining a management strategy as precautionary and in relation to the evaluation and advice on rebuilding strategies.

ii Expert group information

Expert group name	Workshop on Guidelines for Management Strategy Evaluations (WKGMSE2)
Expert group cycle	NA
Year cycle started	2019
Reporting year in cycle	1/1
Chair	Carmen Fernandez, Spain
Meeting venue and dates	4–8 February 2019, Joint Research Center EC, Ispra, Italy (34 participants)

1 Introduction

ICES regularly evaluates management strategies and gives advice on their performance. The "Study Group on Management Strategies" (SGMAS) prepared a set of guidelines in 2008 (ICES, 2013a), which were reviewed and updated by the "Workshop on Guidelines for Management Strategy Evaluations" (WKGMSE) in 2013 (ICES 2013b). After six years of experience with the 2013 guidelines, the ICES Advisory Committee (ACOM) noted the need for a new review of recent work and practices in ICES and elsewhere, taking into account the experiences had with the application of the 2013 guidelines. ACOM further requested the preparation of an up-to-date set of guidelines that should serve as reference for Management Strategy Evaluation (MSE) work in ICES over the next few years. In response, ICES set up the "second Workshop on Guidelines for Management Strategy Evaluations" (WKGMSE 2), with the ToRs provided below (Section 1.1).

The title of the workshop includes the term "management strategy", but in the European context (and generally worldwide) several different terms and inter-related concepts have been used by various actors to describe elements of pre-agreed actions for fisheries management, e.g. Harvest Control Rule, Multi-annual Management Plan, Harvest Strategy, Management Procedure, Management Strategy or Management Arrangement. In line with the terminology used most frequently in ICES in recent years, and the glossary recently proposed by the joint tuna RFMO meeting (tRFMO 2018), **this report uses the term "Management Strategy" to refer to the combination of monitoring, assessment, harvest control rule and management action designed to meet the stated objectives of a fishery**.

1.1 ICES Resolution and Terms of Reference

2018/2/FRSG27 **The second Workshop on guidelines for management strategy evaluations** [WKGMSE 2] chaired by Carmen Fernández (Spain) will meet from 4 – 8 February 2019 at the JRC, Ispra, Italy, to:

a) Review recent methodological and practical MSE work conducted in ICES and in other fora around the world. Based on the work of WKGMSE (2013) and this review, bring up to date the methodologies and technical specifications that should be incorporated in MSE work in ICES.

b) The methodological and technical revision should include all aspects involved in MSE, and pay specific attention to the following issues that have been identified through recent work in the ICES system:

   1. Evaluation of performance in the short-term versus the long-term, including treatment and interpretation of MSE projection results relative to forecasts from stock assessment models used to annually assess the resource;
   2. Appropriate range of scenarios to consider in the MSE and how to deal with outcomes from multiple scenarios, including "worst-case" scenarios;
   3. With reference to the work of WKGMSE (2013), review risk definition and computation in MSE;
   4. How to deal in the context of MSE with the broad range of models currently used for stock assessment in ICES (e.g. stock assessment models that include process error);
   5. Evaluate the efficiency and effectiveness of "short-cut" approaches versus "full-feedback" simulation incorporating annual stock assessment models in the MSE loop;

6. Presentation of MSE results e.g. properly describing the process, standardising outputs to present results, etc;
7. Review initiatives on the science side, including model developments, operating frameworks, etc. that could be incorporated in the ICES system.

c) Update the guidelines for MSE evaluations in ICES originally prepared by WKGMSE (2013).

d) Consider how to best disseminate the guidelines to experts within the ICES community and the need for training courses.

WKGMSE2 will report to ACOM by March 4 2019.

1.2 Approach to the ToRs

ToR a was addressed primarily through an evaluation of recent MSE work in ICES and elsewhere, and by a detailed review of the guidelines given in the WKGMSE 2013 workshop (ICES, 2013b). To facilitate this review, the scientists responsible for each MSE conducted in ICES since 2013 were requested to provide the filled MSE summary template for the corresponding MSE. The template was created by the WKGMSE 2013 workshop, with the intention that it should be filled for all MSE work done in ICES at the time the work was conducted. However, in almost all cases, the templates had to be filled now, since they had not been filled during the actual MSE work. The filled templates are incorporated in Annex 2 of this report, whereas Section 2 provides a summary of the findings from reviewing the templates.

The first two days of the WKGMSE 2 workshop were spent on presentations provided by participants on recent MSE work. Participants were requested to focus the presentations on aspects of relevance to the workshop's ToRs. A summary of the presentations is provided in Annex 3 of this report. After the round of presentations, participants split in two groups, one of which dealt with the methodological and technical revision of the guidelines and another one which dealt with aspects related to the conduct of MSE processes, including communication of MSE results.

The groups took into account the results from the presentations given, and their knowledge and recent experiences on MSE, to carefully review and bring up to date the contents of the MSE guidelines (provided in the WKGMSE 2013 report). The process of doing this was the workshop's approach to addressing ToRs b and c.

The updated simulation guidelines are presented in Section 3 of this report, and cover all technical aspects pertaining to the operating model and management procedure, including the important topic of validation of MSE assumptions and outputs. Section 4 examines aspects related to risk definition and computation, and the ICES criteria for defining management strategies as precautionary, with specific discussion of stocks that are in a rebuilding phase. Sections 3 and 4 together generally address ToRs b1 to b5 (although b2 was not addressed in full, particularly in what refers to "worst-case" scenarios). The recommendation from WKLIFE VIII that "MSE convergence diagnostics should be developed to determine if enough simulations have been run to result in stable performance statistics, and to avoid running more simulations than necessary." is partially addressed in Section 4.3.

Section 5 focuses on MSE in terms of process (how to conduct MSE processes, including communication) and, as part of that, Section 5.3 is closely related to ToR b6. Section 6 presents the MSE guidelines in condensed format and discusses MSE reporting, therefore addressing ToR c. Section 7 gives a summary and links to a range of useful MSE software.

ToR d was about disseminating the guidelines to experts within the ICES community and the need for training courses. Operationalising the guidelines within the MSE process and using it in drawing up the "protocol" for the MSE based on the condensed guidelines (Section 6.1) is described in Section 5.1. ICES also plans to publish a standalone guidelines document based on Section 6.1 of this report. The need for further software development will be followed up in an ICES workshop, WKMSEDEV, planned for later in 2019. There is a recognised issue with lack of expertise to run MSE within the ICES community. A convergence on fewer methods and associated training courses or materials are required for these methods.

2 Recent experience

In advance of the workshop experts who have carried out MSEs since 2013 were asked to fill the reporting template devised by WKGMSE in 2013 (ICES 2013b) covering most important aspects of recent management plan evaluations. These forms are attached as Annex 2 to the report. A brief summary of the findings is presented here.

2.1 Review of recent MSEs in ICES

Since 2013 ICES has carried out around 30 special requests that involved the evaluation of management strategies. These have covered 15 species and 24 stocks with 6 stocks being evaluated more than once. On average, there are around 6 such requests every year. The complexity of these requests has been increasing over time. There is also a clear trend to more prescriptive special requests over time (particularly when recent requests are compared with similar requests in the past). The time line involved in these special requests is often short (<1-year time frame).

The requests have been mainly top-down from the relevant management authorities, such as the EU, NEAFC, individual member countries of ICES or combinations of countries. In most cases, these special requests were already a product of consultations and negotiations, usually involving managers, scientists and industry stakeholders. In many cases, there were specific requests to evaluate certain elements of harvest control rules and report particular performance metrics.

A motivation to request ICES to evaluate management strategies is to establish if the strategy is consistent with the ICES precautionary approach. ICES provides advice based on the hierarchy set out in the introduction in the advice (ICES, 2018a). Once a management strategy has been classified as precautionary by ICES and has been implemented by the relevant management authorities, then future ICES advice will be based on the management strategy. This provides managers with stability and consistency with longer-term objectives.

The requests received by ICES typically ask to evaluate certain elements of harvest control rules without stating anything about the monitoring and assessment methodology to be used as the basis for application of the harvest control rule (although they sometimes make reference to the benchmark assessment). For stocks with assessment methods agreed at benchmarks, the ICES practice has been to use those agreed assessment methods for this purpose. A more comprehensive approach would also undertake an evaluation of a range of alternative monitoring and assessment methods as part of the MSE, and consider a broader range of biological scenarios than normally done in ICES, in order to achieve an in-depth robustness examination of alternative management strategies. This is discussed later in this report (Section 5.2).

Benchmark workshops in ICES agree the stock assessment method and reference points, and are organized when scientists consider that the current assessment method is not working satisfactorily or when new data or models become available. The ICES benchmarking process is open to stakeholder participants. In most cases, the MSE requests to ICES utilised the agreed ICES reference points or requested ICES to estimate potential reference points as part of the analysis. Planning of benchmark workshops should be coordinated with review clauses in existing or agreed management strategies.

Revision of assessment and reference points has triggered requests to evaluate updated management strategies (or ICES itself has deemed the changes of sufficient magnitude to require a re-evaluation of the existing management strategy). Given the workload currently associated with such evaluations, this has created serious challenges for ICES to be able to cope with the work and respond to such requests.

WKGMSE 2 recommends that, as far as possible, the benchmarks for stocks should be planned consistently with the evaluation period for management strategies.

According to the "Glossary of terms" included as Appendix 3 in the 2018 Joint tuna RFMO Management Strategy Evaluation Working Group Meeting (tRFMO, 2018), a Management Strategy Evaluation is defined as: "A process whereby the performances of alternative harvest strategies are tested and compared using stochastic simulations of stock and fishery dynamics against a set of performance statistics developed to quantify the attainment of management objectives". **The ICES approach to MSE is consistent with that definition**.

In practice, almost all simulation work in ICES was done at national institutes, or sometimes in cooperation between institutes. In most cases, this cooperation was formalized and organised in an ICES process. This usually involved some WebEx meetings and a workshop meeting to consolidate the results and produce draft advice. The draft workshop report is normally reviewed by two independent scientists, and these reviews are included as Annexes in the final report.

The final report is used as the evidence base for the ICES advice, which is drafted by an Advice Drafting Group and approved by ACOM. The ICES guidelines for expert groups outlines the various procedures and working practices (ICES, 2019). In general, data and methods are well documented in the expert group reports. However, the process in most cases is not fully transparent because the code and data used to carry out the evaluations are not usually publically available. WKGMSE 2 recommends that, in the future, code and data should be shared using TAF or GitHub repositories.

There were 19 different methods used to carry out these ICES evaluations and only 6 methods were used more than once. In a few cases, multiple software packages were used. Particular methods are often favoured by particular scientists or institutes. This indicates that there is very little standardization in frameworks to carry out these evaluations within the ICES community so far. This is a quality control issue since it leads to problems getting reviewers with sufficient knowledge of the tools to carry out sufficiently detailed reviews or to validate that the coding is correct. The diversity in methods also contributes to the lack of consistency in approach and outputs across different evaluations.

The workshop concluded that a standardisation of MSE tools is urgently needed, while recognising that some diversity in the toolbox will also be required. The platform used should be open source and address all the requirements outlined in the MSE guidelines. Furthermore, WKGMSE 2 recommends that the software used is thoroughly tested using reference stocks (either real or simulated). This could be a ToR for the ICES "Methods Working Group" (WGMG) who have previously proposed methods to do this testing (WGMG, 2004).

The majority of evaluations were for stocks with age-structured assessments, with a few evaluations (5) with length-structured assessments and even fewer (3) with survey-based or empirical evaluations.

The stock-recruitment relationship (SRR) was generally the most important consideration when conditioning the operating model. Most evaluations only used one SRR scenario. Segmented regression was used in 60% of the evaluations, and in some cases, multiple SRRs were combined. Quite often the SRRs were fit to truncated time series. Auto-correlation in recruitment was not applied in two thirds of the simulations. Lognormal was the most common form of stochasticity, with resampling from predictive distribution (EqSim method; Simmonds *et al.*, 2019; ICES, 2014a, 2015, 2016a) also often used.

Weights, maturities and selections were mostly just recent averages, with stochastic variability in some cases and density dependence in a few. Natural mortality was almost always a fixed vector from the assessment (either constant across all ages or age varying). The exception was where predation mortality or cannibalism (M2) was incorporated, where e.g. multispecies models provide estimates. Stochasticity is often only included for weights at age and not for maturity or natural mortality.

Initial numbers were always taken from the most recent assessment. In most cases, they were stochastic, with CVs of 0.2-0.3. The way the parameters of the distributions are derived is not always stated, but where it is, the inverse Hessian is a common source.

In most cases (17), a short-cut approach to the MSE (explained in Section 3 of this report) was used to avoid conducting a stock assessment within the simulation loop, compared to 9 cases which used a full MSE approach (Section 3 of this report), of which 4 applied both full and short-cut methods. Among those cases using the short-cut approach, there are some very good examples of how the various uncertainties were taken into account in the projections (e.g. ICES, 2018b).

2.2 MSE process outside ICES

Several participants in the WKGMSE 2 workshop presented MSE approaches that involved scoping meetings with managers, stakeholders and scientists and testing management strategies for robustness against different plausible operating models. This is broader than the MSEs that have been carried out within ICES. This "Strategic MSE" is characterized by the inclusion of all parties having an interest in the natural resource, i.e. industry, managers, eNGOs, scientists. These parties should participate from the initiation of the process until its end. The scientists facilitate the process through iterative feedback from the other parties. The outcome of an MSE (*sensu strictu*) is a Management Procedure, containing a harvest control rule, which is robust to plausible environmental and structural uncertainties. These uncertainties are tested through different operating models (structural uncertainties) and scenarios (environmental uncertainties).

ICES MSEs are usually limited to one operating model for the historical years, which is the agreed stock assessment model. The parameters of the operating model are used to stochastically project the fish stock with decided uncertainties. The proposed harvest control rules are evaluated using the operating model and an agreed range of scenarios. Often, variations in future recruitment levels are included in the MSE scenarios. Bias in the assessment is also sometimes entertained. Structural uncertainties, such as e.g. different levels of M, recruitment variances or variation in growth parameters are rarely, if ever, evaluated.

Further differences between MSEs carried out by other RFMOs and those carried out by ICES include harvest control mechanisms as well as aspects of ownership of the process. In addition to testing levels of fishing mortalities, possibilities exist to incorporate social, economic and ecosystem criteria within a harvest control rule to calculate a TAC. The increased participation of stakeholders within the MSE process generally leads to a feeling of collective ownership, often leading to a higher acceptance within the community.

3 Guidelines for simulation

3.1 Building blocks in simulation procedures and MSE terminology

This section provides a brief outline of the building blocks, with terminology as used in this report.

The term **Management Strategy** refers to the combination of monitoring, assessment, harvest control rule and management action designed to meet the stated objectives of a fishery (tRFMO 2018).

A **Management Strategy Evaluation (MSE)** simulation procedure is composed of the following blocks (Figure 3.1):

- An **operating model** represents a realization of the "real world" and includes:
  - A **biology and fishery model** capturing the underlying dynamics of the population and its exploitation.
  - An **observation model** that extracts, with error, information from the operating model that is used in the estimation model and decision process.
  - An **implementation model**, which translates the decided removals into actual removals from the real stock.
- A **management procedure** represents a perception of the "real world" through data and the decisions made on the basis of this perception. It includes:
  - An **estimation model**, that assesses stock status based on available information; this could include an assessment (or proxy for this) or an empirical approach (e.g. a biomass index or CPUE).
  - A **decision model**, in which a decision on removals (typically a TAC) is derived from the outcome of the estimation model. The decision model is often referred to as a "harvest rule" or "harvest control rule".

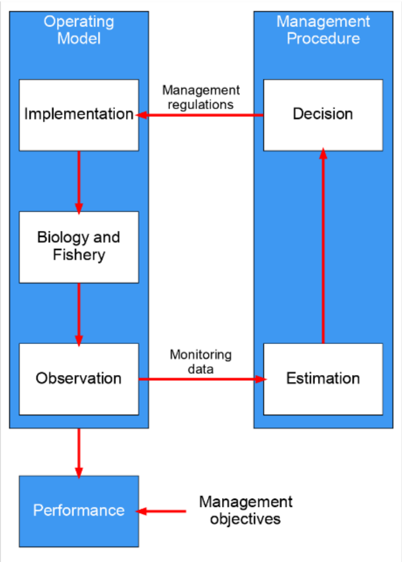

Figure 3.1. A conceptual overview of the MSE modelling process (Punt *et al.*, 2016).

The only communication between the operating model and management procedure should be through the data that the operating model passes to the management procedure, and the management regulation (e.g. TAC) that the management procedure passes back to the operating model. Furthermore, performance of the management procedure is evaluated through performance statistics, which are defined on the basis of management objectives (Figure 3.1).

In a simulation framework, these models constitute a loop, which is repeated for a number of years. Each sub-model has stochastic elements. The MSE performs a number of stochastically independent replicates; in this report, we use mainly the terminology "iteration" to refer to each of these independent replicates, in order to maintain consistency with previous terminology used for MSE in ICES.

It should be stressed that although MSE simulations are often carried out using long-term projections to study the behaviour of management strategies and to run populations into equilibrium, they are used to inform managers on what will likely happen in the short- to medium- term. MSE simulations are normally parameterized based on the current (or historically observed) ecosystem, biological and fishery state, and results are only valid under the conditions simulated in the operating models; hence, they should not be taken as long-term predictions.

Moreover, results from MSE evaluations should not be interpreted as if they were forecasts from stock assessment models used annually to assess the resource. An MSE simulates future data (generated via the observation model) and incorporates feedback control in the management procedure (i.e. the simulation takes into account that the harvest rule responds to the signals it receives about stock development from the estimation model). MSE can be used to identify harvest strategies that fulfil management objectives, while identifying, at the same time, trade-offs between different strategies and objectives. The reliability of MSE results is dependent on having properly characterised the existing uncertainty.

Each of the MSE building blocks is discussed in detail below.

3.2 Choice of model and modelling approach

The choice of model will depend on the experience of the analyst, but should be guided by the purpose of the simulation study.

One purpose may be to outline candidate management strategies for a stock with some, perhaps conflicting objectives, and to show trade-offs between objectives. For example, one may want to scan over a large range of harvest rule parameter options, and test for sensitivity to a variety of assumptions. This will require software that is fast, typically software without stock assessments in the model. Once a proposed harvest rule is reached, it can be further examined, with the same or other methods. At this second stage, a key issue is that the model reflects the biology of the stock and the fishery, and that the observation, estimation and decision models reproduce the procedure that will be implemented in practice. Much effort has to be put into validating the model conditioning, whereas the computing time (which can be very onerous in MSE) is a secondary consideration. The same applies if a single management procedure is presented for approval.

If the knowledge of the stock is limited, for example for stocks where assessments are not possible, the first task may be to develop rules that are likely to work for a type of stock that is similar to the stock in question. If so, a generic range of stock biology can be created, with little emphasis on getting all details 'correct'; the goal of the simulations will then be to find harvest rules that are likely to work irrespective of the unknown finer details.

3.3 Biology and fishery components of the Operating Model

The biological and fishery components of the operating model are intended to reflect the "true" dynamics of the stock productivity and exploitation pattern. Key elements of this are growth, recruitment, natural mortality and sexual maturation, as well as fisheries selectivity. The dynamics of these processes need either to be modelled or have their variability captured by the operating model and conditioned on available data and knowledge. Some important aspects of this are considered below.

During conditioning of the operating model, many of the parameter estimates are obtained by fitting to historical data within a stock assessment, although some parameters may be considered fixed. In combination with validation (Section 3.7), this ensures that the parameter values used in the projection period are consistent both with the available data and current understanding of the system.

Uncertainty estimates for parameter values within an operating model can be based on e.g. samples obtained from bootstrapping, Bayesian posterior distributions, or variance-covariance and MCMC approaches that can take into account several sets of parameter values and correlations between them.

The set of parameter values and uncertainties most consistent with current under-standing can be considered to constitute a baseline operating model. If current reference points have been determined previously using an EqSim analysis, then the main assumptions from this exercise can be used as guidance for the conditioning of the baseline operating model in the MSE (e.g. recruitment function, biological parameters, selectivity, etc.).

Additional key uncertainties in the conditioning process can be explored using a number of alternative operating models, which can be developed to evaluate the effects of deviations from the baseline model. This can include alternative assumptions, models, and error structures considered when selecting the uncertainties to include in the operating model (McAllister and Kirchner, 2002; Hill *et al.*, 2007), so that the robustness of the management strategies to such uncertainties can be evaluated. Alternative operating models could also be used to deal with potential data conflicts.

3.3.1 Initial population matrix:

The estimates of abundance for the start of the projection period can be obtained from the output of the most recent assessment either directly from the estimate of numbers in the final year or from the resampling methods described above. It is important to appropriately include information on the uncertainty in the initial state of the true stock being simulated, as it will influence the perception of risk in the short term.

The uncertainty associated with the youngest age classes in the initial population should be considered with care. Typically, assessments estimate a high uncertainty for younger ages due to lack of information on year class strength and direct use of this for deriving initial population in the MSE may lead to an unrealistically wide uncertainty range. In this situation, recruits could instead be drawn from the stock-recruit function for each iteration, or the CV could be reduced to the CV of the stock-recruit function.

The important consideration here is that the uncertainty in the initial state is considered and arguments are given for how this contributes to a plausible range of realities when incorporated in the simulation.

3.3.2 Recruitment

The minimum standard is a single stochastic stock-recruit model to reflect potential variability. It is recommended that modelled recruitment not be implemented stochastically around a fixed stock-recruit fit, but rather that the parametric stock-recruit fit should also be stochastic, such that recruitment is drawn stochastically from around a different stock-recruit curve at each simulation replicate. Accounting for temporal dynamics (e.g. autocorrelation, periodicity and occasional extreme values) is important, and metrics to show the appropriateness of the modelled dynamics to those historically observed should be presented (see examples below).

3.3.2.1 Choice of stock-recruit function

If a single stock-recruit model explains the data well over the full range of biomass covered by the simulation, it would be sufficient to continue on this basis. The stochastic component can be obtained through resampling residuals (with replacement) or using a fitted statistical distribution (truncated as necessary, e.g. log-normal). If resampling methods are used, care needs to be taken to ensure autocorrelation is included where appropriate.

The choice of stock-recruit model may be critical to the performance of the management strategy, even when the fit of different models to the historical data is almost equal. If the choice of stock-recruit model is uncertain, a simple single model approach would not be sufficient to capture the recruitment dynamics. In this case, a range of scenarios should be tested to cover a range of plausible possibilities by fitting alternative stock-recruit models and testing a range of management procedures under each circumstance. In particular, if there is a great deal of uncertainty in the slope of the stock-recruit relationship near the origin or in the recruitment at large stock biomasses, different options must be tested. If the MSE results are relatively insensitive to these choices, one model may be chosen for further work.

If, following this investigation, it is found that performance of the management strategies tested is critically dependent on the choice of stock-recruit or growth models, then multiple models with different parameters can be selected using, for example, the method of Michelsens and MacAlister (2004), as implemented in the EqSim software package. This method provides a formal way of including uncertainty in the form of the stock-recruit functional relationship, parameters and stochasticity in the evaluation. An alternative approach is to construct a separate operating model for each stock-recruit relationship under consideration.

3.3.2.2 Accounting for temporal dynamics.

If there are concerns that distributions around one or several stock-recruitment relationships are not stationary over time, i.e. that factors that influence recruitment in addition to the spawning biomass fluctuate beyond independent random variations, introducing autocorrelations may give an adequate representation of this fluctuation. If there are periodicities or trends, they could be included in the model; however, that implies predicting future fluctuations, which requires that such predictions are well justified.

An alternative approach is to specifically examine the robustness of the management strategy to such fluctuations, and require that it should perform adequately with a realistic range of future recruitment regimes. Such robustness testing may be done by inducing changes at fixed times, and examining the response.

There may be cases where externally-driven factors may be more important for recruitment than SSB. It is often not possible to include such externally driven processes in the MSE, but even in such cases, low SSB must begin to affect recruitment. A hockey-stick (i.e. segmented regression) model could then be considered as a way to capture the mean level of recruitment with suitable fluctuations. Careful consideration would be needed to define where to place the breakpoint of the hockey-stick.

Some stocks have exceptional year classes occurring with more or less regular intervals, so-called 'spasmodic' year classes. Such year classes may be included in the simulations. An example from the blue whiting MSE is given below (Figure 3.2). This diagnostic compares the cumulated distributions of the modelled recruitment and the observed recruitment in a period with occasional large year classes. This kind of plot is useful for determining if the probability of large year classes is appropriate, but does not inform about the intervals between such year classes.

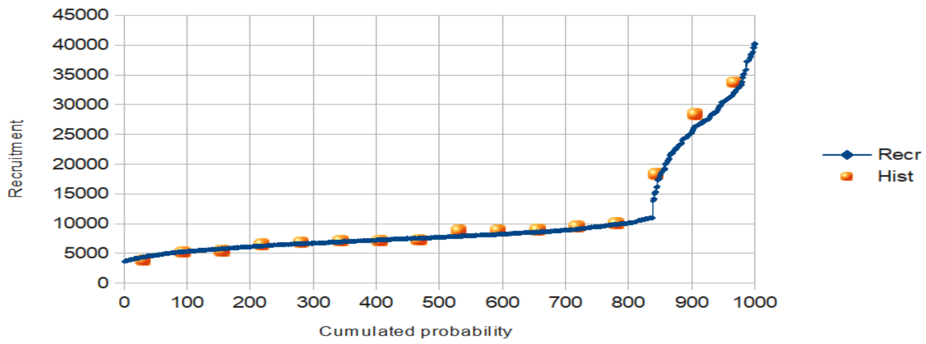

Figure 3.2. Cumulated distribution of simulated and observed stock recruit pairs. Blue whiting in a period with occasional large year classes.

3.3.2.3 Regime shifts

There are numerous plausible hypotheses relating environmental changes to changes in the parameters of a biological population. If it is likely that for example, growth or recruitment are dependent on environmental drivers, then a plausible range of scenarios should be considered when evaluating an MSE.

Punt *et al.* (2014) identify two approaches to consider climate and environmental uncertainty in an MSE. Firstly, they identified a "mechanistic approach" that estimates the relationship between the environment and elements of the population dynamics in the operating model in order to make predictions for population trends using outputs from global climate models. Secondly, they identified an "empirical approach" that examines possible broad scenarios of how biological parameters in the operating model (e.g. natural mortality, growth, recruitment) may change in the future due to environmental factors without explicitly identifying mechanisms. This study found that modifying the management procedure to include environmental factors does not much improve its ability to meet management objectives, if at all, and will only do so if mechanisms are well known and understood. They conclude that it is better to assess the robustness of management procedures against plausible broad forecasts of how biological parameters may change in future (be they using "mechanistic" or "empirical" approaches), rather than trying to specifically incorporate environmental factors in the management procedure.

The issue of regime shifts is related to the classic dilemma between having a long time series of data and a large dynamic range, versus considering a (fairly) constant ecosystem regime existing only for a shorter time. Due to the large variability of recruitment, a time series of say 20 years is considered a short time-series in the context of estimating stock-recruit parameters.

Questions that should be addressed when considering regime shifts include: can individual years be regarded as a regime shift, or is that better dealt with as noise? What about two years, three years etc.? Is there a minimum length in terms of the number of years for a regime? It is important to realise that a regime shift does not have to be sudden, but can also be gradual.

It should be noted that the time series do not have to be continuous. If there is a temporal anomaly, like the Gadoid Outburst for the North Sea, then it might or might not be appropriate to delete a time window, but not all data points before the end of such an event. However, when setting up robustness tests to regime shifts, it is probably better to fix the timing of the shift, and examine the performance in those years, rather than having the time as a stochastic variable, which would smear out the effect.

Regime shifts can be a result of fisheries management, e.g. for the Baltic Sea the high F on cod has driven the stock to a low level, and the sprat stock has increased simultaneously due to low predation from cod. Sprat in turn eat cod eggs and the cod stock-recruit relationship seems thus to be in a new regime. Thus, theoretically in this case, fisheries management could reverse the regime, if desired.

It is also worth considering that when a regime shift has been identified, whether it is then best to completely ignore data related to the period of anomaly, or whether some useful information can be extracted from e.g. the stock-recruit pairs prior to the regime shift.

The answers to these questions are not obvious. For the purpose of evaluating management strategies, one guideline may be that the strategy should work well under a plausible range of future productivity regimes, and that it should cope with the kind of changes in productivity regimes that have been encountered in the past. Furthermore, whatever decision is made, it should be properly justified and documented.

3.3.3 Fishery selection at length or age

Selectivity in the fishery can appear in several contexts in an MSE. Within an operating model, fishery selectivity appears in the implementation model (when a TAC or catch is translated into the actual removal of numbers at age, or at length, from the real stock) and can also be in the observation model (the generation of catch at age, or at length, data for input to the assessment). If the MSE includes an age- or length- structured stock assessment model, this model will also have to handle selectivity in some way. Selectivity can also be an input to the decision model within the management procedure (e.g. if conducting a short-term forecast is required, then assumptions about selectivity will be needed for the forecast year(s) and any intermediate year(s)). The selectivity will not necessarily be the same in all these contexts. It is noted that, in the observation model, uncertainty can be applied either to selections or to the ‘observed’ catches.

MSEs are generally run contingent on the current situation in terms of selection at age, and they are valid only under the assumed conditions. Some assessment models such as SS3, for example, are able to provide estimates of selectivity and associated uncertainties. The selectivity at age in the current assessment can be analysed for recent stable representative periods. As fishery dynamics change over time and it may be unlikely that such changes are reversed, more recent time periods should be considered (recent 3, 5 or 10 years) to be more applicable for the future years included in the MSE.

If trends in selectivity have been observed in past years, the assumption of the continuation of any trend will eventually lead to unrealistic selection profiles in a long-term simulation. However, if a mean is derived from a period exhibiting a trend then future values will differ from the most recent past, which also may not be realistic. Often, it is assumed that future values will continue as at the present by taking a mean over a selected period. Regardless of the approach taken, any choice should be justified and the implications outlined.

When estimates of changes and uncertainty in selectivity are not readily available, an approach to estimate uncertainty may be to use smoothed selectivity curves in catch curve analysis, and use catch curve prediction intervals to determine uncertainty in the estimation of selectivity.

Selectivity has a direct effect on estimates of yield and F_{MSY}. It is important to consider the sensitivity of the MSE results and MSY estimates to any proposed selection model and associated uncertainty.

3.3.4 Weight-at-length or age

Weights are important in several contexts within an MSE, as there are frequent conversions between abundances and biomasses within the individual models of the MSE framework. For example, the translation of numbers to biomass within the operating model to calculate an SSB of input to a stock-recruit model, and also possibly density dependence models. Since many management (decision) models are based on SSB, estimates of weight at age or length, and their uncertainty, lead to a range of possible management decisions.

It is therefore important that the weights used within the projection period of the MSE appropriately represent historically observed variability and any within/between cohort correlations. There is no universal recipe for implementing this in the simulations, but any choices should be justified, and the implications made clear.

A common approach is to assume a mean from a recent period (e.g. 3-5 years), which implies an assumption that future values will continue as the recent average. Stochasticity can be introduced by setting the weights in a future year equal to those from a randomly selected year in the recent past. In some cases, randomly selected blocks of years of variable length have been chosen as the basis for future weights with the blocks appended until a sufficiently long time series of weights is generated.

Trends in weight at age (or length) are frequently observed, often over significant time scales. However, it is not appropriate to consider that such trends will continue indefinitely, as eventually they will become unrealistic. Without a defensible hypothesis for declines or increases in weight that can be incorporated within the MSE, the robustness of the management procedure to the continuation or the reversal of observed trends is most appropriately explored via alternative operating models.

It is important to recognise that within the biological system, changes in fish weight are often correlated with changes in other parameters e.g. maturity.

Following changes in the exploitation pattern within a fishery, dynamics in weights at age may change over time. A recent study by Kraak *et al*. (submitted) shows that when fishing intensity or especially the selection pattern change, the length at age distribution in the surviving population changes, caused by the so-called Rosa Lee Phenomenon (Lee, 1912). This effect is larger in slow-growing species. If these changes in length at age (and thus weight at age) are not accounted for, biases in the calculations of B, SSB, and catch occur. In most cases these biases are negligible, but with substantial changes in selection pattern the biases can be up to ~30% and much higher for the discard fraction (or fraction below MCRS) of the catch. In such cases it is recommended to take these changes into account. The scripts developed by Kraak *et al.* (submitted) are available on https://github.com/sarahbmkraak/Rosa-Lee-paper.

As with selectivity, weight at age also has a direct effect on MSY level in terms of long-term yield and F_{MSY}. Exploring the sensitivities of the MSE to uncertainty around weight at age is important.

3.3.5 Natural mortality

3.3.5.1 Constant natural mortality

Natural mortality (M) in the operating model should be handled consistently in the past (years used for conditioning the operating model) and future (years over which management strategies are evaluated). If the stock assessment used to condition the operating model for past years uses a year-independent M, the same value of M should be applied in future years in the operating model (at least in the base case operating model). This is because the historical development of the stock fishing mortality, recruitment and biomass are correlated with the assumed natural mortality. The use of different natural mortalities for future years in the MSE would lead to inconsistencies between the assessment used to parameterize the MSE simulation and the forward projections. Sensitivity testing of the effect of a higher or lower natural mortality in future years is easy to carry out, but it is difficult to evaluate the results without a change in the historical values of M as well. Alternative natural mortality hypotheses are most appropriately explored with an alternative operating model configuration.

3.3.5.2 Time-varying natural mortality

When time variable Ms are used in the stock assessment (e.g. North Sea cod, North Sea herring) used to condition the operating model for past years, the estimates of M from the latest period (terminal year if smoothed M values are used, average over a suitable time period if not, or sample from a suitable time period) can be used in the MSE for future years in the short term. For longer-term simulations (and recovery scenarios) the effect of a variable M should be investigated, either as a part of a sensitivity analysis or modelled explicitly. In principle, sharp discontinuities should be avoided, as rapid changes may not have a scientific rationale.

Changes in M associated with a regime shift may not require a review of M in the historical period. However, in this situation it is important to include realistic coupling with e.g. stock-recruit and other biological processes that would likely accompany a regime shift.

3.3.5.3 Prey species (e.g. North Sea herring)

For typical prey species, the natural mortality is very variable over time and depends to a large extent on the biomass of predators, the abundance of the prey species itself and the availability of alternative prey species (functional feeding response). MSE simulations are typically single-species and, as such, changes in natural mortality cannot be estimated. However, the range of historical natural mortalities may be available from the stock assessment (and used there) which makes it possible to test the robustness of the proposed management strategy to the historically observed variability in M. This can be done by, for example, min-max scenarios or by bootstrapping from the observed distribution of natural mortalities over time. The historic period from which values should be sampled or bootstrapped requires consideration (e.g. from times with low or high predator stock biomasses).

3.3.5.4 Cannibalistic predators (e.g. cod)

Stomach contents of e.g. cod and whiting have shown that cannibalism is an important part of natural mortality for the younger individuals. Ignoring cannibalism within an MSE can lead to very different conclusions about performance of the management strategy (e.g. cod recovery in the North Sea; ICES 2014b) and, in such cases, cannibalism must be included in the MSE, at least for long-term simulations and recovery scenarios.

ICES WGSAM (2011) made a first approach to model predation mortality based on simple relationships between predation mortality and the biomass of predators. This approach can be applied based on the biomass of the species considered (e.g. cod) estimated in the MSE. It will also be possible to estimate the relationship between the partial predation mortality and the species itself, assuming a constant population of other predators. Such approach will deliver a simple relationship: $M_{age\,1} = a + b * SSB$, where SSB is the SSB of the cannibalistic species at the beginning of the year as calculated in the MSE, and a and b are parameters estimated from multispecies output.

However, when modelling cannibalism explicitly, it has to be ensured that cannibalistic effects are not doubled. For example, one could use a Ricker stock-recruitment relationship to already take into account cannibalistic effects. In this case, only cannibalistic effects on older age groups not covered by the stock-recruitment relationship should be modelled explicitly.

3.3.6 Maturity

When temporal estimates of maturity are available, a recent mean may suffice for short-term considerations and often forms the basis of the base case operating model. The majority of MSEs to date have assumed a time-invariant ogive. A plausible starting point is consistency with the approach taken in the estimation of long-term equilibrium reference points. However, if there is supporting data, it will be important to include variability representative of the historical period. Previous approaches and important points to consider include:

- Using a replica of the maturity ogive from the assessment, often an age dependent maturity fraction averaged over a recent (stable) period;
- Selection of a number of random length time blocks from a relevant historical period and appending them until a sufficiently long future time series is available;
- If there are historical trends, is it viable that they will continue/reverse, and is there an identifiable mechanism that can be included in the operating model?
- For stocks that exhibit spasmodic and highly variable recruitment, large year classes may exhibit a different maturity profile;
- Maturity is often correlated with other operating model parameters (e.g. natural mortality, weight-at-age) and these correlations should, where possible, be reflected in the future period.

3.3.7 Confounding between variables / correlated processes

Throughout this section, a number of individual parameters and processes have been discussed. However, it is rarely the case that they can be considered in isolation, either with respect to the historical period or other operating model parameters and processes.

Operating models are typically conditioned using existing historical data. However, observations can sometimes be explained by alternative processes (e.g. dome shaped selection and senescence) and it is necessary to consider if there exists sufficient information in the data to estimate all parameters.

When correlations can be reliably identified and are considered to be likely features of the true biological population (such as e.g. correlation between weight-at-age and maturity-at-age), they should be included within the operating model. If future values or estimates of variance are based on a period of historical data, then equivalent periods should be considered for related parameters (e.g. weights and maturities) and significant discontinuities avoided.

A number of variables may exhibit autocorrelation, the most commonly considered being recruitment. It is also possible that there may be correlation between ages that should be taken into account.

When confronted with uncertainty arising because of possible correlations and/or confounding of parameters in the estimation process, this can be dealt with via alternative operating models.

However, be aware when proposing alternative operating models that combinations of confounded parameters can lead to redundancy in the set of operating models. When there is more than one plausible explanation for observations, then propose alternative operating models to test robustness to each hypothesis.

3.3.8 Ecosystem, biological and technical interactions

A critical part of designing any MSE exercise is to identify early on, which key processes need to be included in the operating model(s), potentially in collaboration with stakeholders. As part of this process it is important to consider multispecies and ecosystem interactions as well as technical (mixed fisheries) interactions. These may be as, or more, important than uncertainties within the single species being considered. The key is not to include every possible process, but rather to identify which processes are sufficiently important that they should be included in the operating model or as robustness tests. Where processes are included in the existing assessments then they should be included in the operating models where possible, but other processes may also impact on the projected stock dynamics. Ecosystem effects include, for example, cannibalism, density-dependence in growth or maturation, variable predation mortality or environmental drivers of stock dynamics. Technical interactions will affect the implementation of the proposed management strategies (for example, for choke species by imposing a minimum catch level on a species bycaught in other fisheries). Where such processes are important drivers of stock and fleet dynamics over the expected life span of the management strategy, then these should be included in the operating models.

To incorporate ecosystem/multispecies interactions in MSE, one could employ full multispecies models operating models. For some management questions and ecosystems this may the preferred route. However, it is also possible to include many processes as add-ons to single species operating models (for example through density dependence). In a similar way, one could account for technical interactions through a multistock mixed fishery model, but it may also possible to capture the key interactions in a simpler manner. Imposing implementation error in the form of a minimum catch, or through noise replicating the variations in the other stocks would be examples of possible approaches.

The method used to incorporate these dynamics will depend on the particular driver(s) being modelled, the availability of existing models, the management objectives, and the time and resources available. Furthermore, one must decide if the process(es) need to be included in the base case operating model, as an alternative operating model, or as a robustness test.

3.4 Observation and estimation models

An MSE includes an observation model which generates observations for use in the management procedure (see Figure 3.1). Depending on the form of the decision model (harvest rule) in the management procedure, the observations are used either as input to a stock assessment model (for “model-based” harvest rules) or directly in the harvest rule (for “empirical” harvest rules). In all cases, the observations generated should have the same statistical properties as those arising in practice.

3.4.1 Model-based harvest rules

When the harvest rule in the management procedure is based on the results of a stock assessment (typically followed by a short-term forecast), one would, in principle, reproduce the stock assessment (and short-term forecast) within the MSE simulation. It is, however, recognized that there may be technical difficulties when attempting to do this (such as over-long computing times, convergence difficulties in the stock assessment, assessments not amenable to automation, etc.). As a consequence, two types of technical approaches have been used for model-based harvest rules, the so-called "full" MSE and the "shortcut approach", discussed below.

3.4.1.1 Full MSE

In a full MSE, the stock assessment and short-term forecast are conducted within the MSE simulation, replicating the procedure used in practice as closely as possible.

Thus, a full MSE includes an observation model, which generates observations that are used as input to the assessment model.

After observations have been generated, an assessment model with the same settings that will be used when implementing the management strategy in reality should be applied within the MSE simulation. This can be challenging, as noted above. For example, some assessment models may estimate a potentially large set of parameters over the time series of available stock and fishery monitoring data (e.g. as may occur in Stock Synthesis). Often the estimation may be over hundreds of parameters, which can require extensive calculations to obtain a solution, as well as requiring diagnosis of convergence and sometimes expert intervention to obtain a solution. Even when the model can be successfully fit to the observations, data simulated over closed-loop simulations can generate data realizations that lead to failed convergence, local minima, or circumstances that drive the simulated stock to conditions that cause unstable estimates (e.g. extremely low levels of estimated biomass that require coding solutions to situations like stock extinction or unrealistically high fishing mortality states) (Wiedenmann *et al.*, 2015).

3.4.1.2 Shortcut approach

This approach provides an alternative to the full MSE by substituting the observation and assessment models with a stochastic process that should deliver generally the same stock estimates and match the error structure found in the stock assessment. Therefore, when the shortcut approach is used, no observations are generated and no stock assessment is performed within the MSE simulation. In Figure 3.1, the estimation step is essentially replaced with an assessment emulator, which adds error to the quantities used in decision-making. In some cases the shortcut approach has also included an approximation to the short-term forecast used in actual practice (e.g. ICES 2018c); in some other cases the short-term forecast step has been simply ignored (rather than approximated) in the MSE, whereas in yet other cases the short-term forecast used in actual practice has been fully replicated within the MSE.

The shortcut approach can reduce the computation time and fragility of the full approach (e.g. when convergence of the stock assessment model is not straightforward in closed-loop simulations). The utility of the shortcut approach depends on how well the approximation mimics the stochastic behaviour of the stock assessment, including predictive performance, bias and correlation of errors, as discussed below in the validation section. The shortcut approach should be designed to deliver the full range of output needed for the management procedure.

3.4.1.3 Full MSE or shortcut approach: Pros and cons

Advantages of the shortcut approach compared to the full MSE are that it is faster, simpler, and more robust in certain circumstances. Shortcut MSEs require less time to run, which can facilitate stakeholder interaction when time to make decisions is important. They are simpler and require less advanced coding because they avoid the observation model needing to produce valid input data for the assessment and running the assessment model in an automated loop. However, it should be noted that producing an adequate approximation to the combined behaviour of the observation and assessment models is not an easy task. As already noted, the shortcut approach has the advantage of avoiding convergence issues, which may arise particularly for complex stock assessments that may require frequent tuning by stock assessment groups, making these assessments difficult to implement automatically in a full MSE (Wiedenmann *et al.,* 2015). It also avoids needing to ensure that the assessment model automatically produces output that can be used as input to the forecast under all circumstances; there may be unusual circumstances such as after a period with no catch that need special handling.

Full MSE is preferable to the shortcut approach when there is a need to evaluate if the management procedure can handle mismatches between the biological and assessment models. Examples of use are:

- Multispecies model mixed fisheries;
- Effect of different M in the biological and the assessment model;
- Looking for explanations of trends in retrospective patterns;
- Stocks consisting of a number of substocks.

The behaviour of some stock assessment models may change depending on the data coming in. For example, a series of catch levels associated with low Fs could cause the performance of some assessment models to deteriorate (e.g. for VPA-type assessment models). Including the assessment model in the simulation loop (i.e. a full MSE) would be able to capture this behaviour, which may not be easily captured or anticipated when using approaches that shortcut the assessment.

Along similar lines, Kraak *et al.* (2008) found that setting a high shrinkage level in XSA (a setting whereby fishing mortality in the most recent year should resemble the value obtained for previous years) led to the assessment model overestimating F when the true F was declining and the other way around when the true F was increasing. This behaviour appears as an emerging property of the simulation if a full MSE is implemented, but may be difficult to mimic, or may not even be anticipated, in a shortcut approach.

Another aspect to keep in mind is that a change in the assessment methodology may change the error structure in the assessment. Models such as XSA are set up to try to estimate change and be sensitive to recent changes in F. The move to F smoothing models, such as SAM, will give lower CVs but more autocorrelation in the assessment error.

Studies that have compared the results of full and shortcut approaches have shown that choices of the appropriate management policy can differ between approaches (ICES 2013b; Punt *et al.*, 2016). Therefore, if the shortcut approach is considered to be necessary in order to facilitate investigation of a range of plausible stock and fishery scenarios, then simulation studies as described by Weidenmann *et al.* (2015) can help to match the patterns of errors of the actual assessment model. Such an approach could provide evidence that the shortcut method provides an acceptable approximation to the behaviour of the actual stock assessment model.

Two examples of the comparison between a full MSE and a shortcut approach are given in Kell *et al.* (2005) and ICES (2008), where the shortcut approach used in both cases approximated the stock assessment and ignored the short-term forecast step. The first of these examples examined the effects (on stock biomass, yield and stability) of constraining interannual variation in TAC and found that, with the shortcut approach, expected yield and SSB converged rapidly towards equilibrium, whereas with a full MSE the dynamic behaviour of the stocks and fisheries could not be predicted from biological assumptions alone or from simulations based on a target fishing mortality (i.e. without feedback from the management strategy to the operating model). The second example used the EU and Norway management plans to compare a full MSE to a shortcut approach, and came to a similar conclusion. It found that the shortcut MSE led to one management plan being clearly favoured over the other in terms of a composite statistic reflecting both yield and resource risk, whereas this would not have been the case had a full MSE been performed. Differences were not as marked when the assessment was approximated but the short-term forecast step was performed rather than ignored.

An important message from these studies is that it is not advisable in MSE to ignore the short-term forecast step and time lag between final assessment year and the TAC year, when these are part of the management procedure that will be used in practice. Neither is it advisable to conduct a short-term forecast using different assumptions from those that will be used in actual practice. Although reproducing the assumptions made in the projection may be challenging for the programmer, for example with regard to future recruitments, weights and selections, it should not be considered a valid reason for unrealistic simplifications. These conclusions are relevant for any type of MSE, be it a full MSE or a shortcut approach to MSE.

We recommend that, if feasible, the full MSE approach be used and appropriate MSE software be chosen accordingly. The problem could be for ICES to find the people with the right set of skills to run the software, because institutes tend to have traditions regarding the software commonly used. In that case, ICES could request an expert from outside the stock area to help with the modelling. However, getting outsiders to work on MSEs for ICES may require the adoption of a general MSE glossary to facilitate communication.

3.4.1.4 Technical details of the full MSE approach

As already said, a full MSE should generate observations, such as survey indices or catch-at-age, for use as input data to the stock assessment model. The observations are generated from the biological and fishery components of the operating model, incorporating measurement error with the same statistical properties as the input data supplied to the assessment used in practice. One way to estimate these statistical properties is from the fit of the original assessment to the observed data series. For example, if a survey index at age is fitted to abundance, observation error can be implemented through the index catchability-at-age using the estimated variability or residuals. If there is evidence for correlation in the survey index, such as e.g. year effects resulting in correlation between ages, this should be reproduced in the observation model using appropriate covariance structures. In some cases, the covariance structure can be estimated by the assessment model (e.g. SAM). If that is not the case, it can be approximated by examining the model residuals. Similarly, observation error can be added to catches at age from the operating model.

Fisheries selectivity and biological parameters, such as weights-at-age, maturity, and natural mortalities, which enter in the management procedure (e.g. in stock assessment and/or short-term forecast) may differ from their true values in the operating model. Depending on the assumption for the operating model (Section 3.3) and in the actual current assessment, fisheries selectivities and biological parameters can be smoothed or observation error added.

The stock assessment model in the MSE should be the same one that will be used when the management strategy is implemented in reality. In ICES, this is usually the assessment model agreed in the most recent benchmark. In some cases it is not feasible to include the agreed assessment model in the MSE (e.g. long running time in TSA), and an alternative assessment model may instead be used in the MSE; however, it must be checked that the behaviour of the alternative assessment model is similar to that of the agreed assessment so that it provides an adequate approximation.

The assessment model outputs and other relevant quantities may then be used as input to a short-term forecast or applied as otherwise required by the decision model. If a short-term forecast is used in the harvest rule, then the forecast settings used in the MSE should be the same ones that will be used when the procedure is implemented in practice; often these will be the short-term forecast settings agreed at the most recent benchmark.

The MSE simulation should reproduce all parts of the harvest rule as closely as possible, potentially including a short-term forecast, calculating a 'primary' TAC, then adding constraints such as minimum or maximum TAC values and/or TAC stabilizing mechanisms in order to arrive at the 'final' TAC according to the management procedure.

3.4.1.5 Technical details of the shortcut approach

The main challenge here is to approximate the behaviour of the assessment model by adding structured noise to appropriate quantities from the operating model with specified distributions, and to ensure that this approximation is adequate. It is generally not sufficient to simply add unstructured random noise to quantities derived from the operating model.

In existing software, emulating an assessment is done at various levels of sophistication, for example by combining random year effects and age effects, and/or including autocorrelations to imitate retrospective errors.

Errors in N and F in assessment period, followed by short-term forecast:

In many cases, assessment errors are applied to the "true" stock numbers and fishing mortality-at-age coming from the operating model, thereby generating "assessed" or "estimated" values covering the assessment period (e.g. ICES, 2012 2014b, 2016b). This "assessed" stock status is then used as input to short-term forecasts. This approach aims to reproduce as closely as possible the procedure used by the ICES working group to formulate the TAC advice.

A common method to produce these errors consists in generating random deviations, typically assumed log-normally distributed, using a combination of age-specific errors and autocorrelated year effects (e.g. ICES, 2014b, 2016b). For the age-specific errors, sigma values can be taken from the assessment as the standard deviations of the log stock numbers and fishing mortality-at-age from the most recent real assessment (or alternatively an average of the last x years). The year effects are generated using a sigma and autocorrelation, which are adjusted "manually". This "calibration" of the noise can be conducted by searching for the sigma and autocorrelation values that would make the standard deviation of the log ratios between true and observed ('assessed') values (e.g. terminal year SSB), calculated over all the years and 100 iterations similar to that in historical data.

Alternatively, assessment errors may be obtained by running 10 years of retrospective analyses (for each simulation trial), computing the log-ratio of the true numbers-at-age and retrospective numbers-at-age for each peel of the retrospective analyses, and applying for each year in the simulations the errors from one of the peels, selected randomly amongst the 10 retrospective years (ICES, 2012).

Errors in SSB (or N) and F in the advice year:

An alternative approach, as implemented in the ICES EqSim and SimpSim softwares (e.g. ICES 2016b) consists in applying errors on the quantities used in the HCR, such as SSB and F_{bar} in the advice year (i.e. the year for which the management procedure provides a TAC).

Errors are simulated by applying directly on the "true" SSB and/or F in the advice year a log-normally distributed error, with CVs and time-autocorrelation estimated from a retrospective hindcast analysis. The methodology is that applied at the Joint ICES-MYFISH Workshop to consider the basis for F_{MSY} ranges for all stocks (WKMSYREF3, ICES 2014a). Conditioning on the observed time series of catches, the approach compares the F or SSB forecast in each historical assessment (subsequently peeling off more years) to the values estimated in the most recent assessment and measures the variability (SSB_{cv} and F_{cv}) and autocorrelation (SSB_{phi} and F_{phi}) of the prediction errors over time. As such, this method simultaneously takes assessment and short-term forecast error into account. The estimated CV and autocorrelation values can then be used in a two parameter error function applied directly to the SSB used in the HCR and the resulting target F according to the HCR.

A note of caution when applying errors on SSB (or stock numbers) and F is that, in reality, the estimates or short-term forecasts of these quantities are likely to be strongly correlated rather than independent. This is a technical aspect of the shortcut approach that requires further consideration.

3.4.2 Empirical harvest rules

In some cases the decision model does not require a stock assessment, e.g. if the harvest rule relies directly on data, such as a survey biomass index or e.g. the mean length of fish caught in the fishery. In this case, the data should be generated in a realistic way, using the same principles as when input data for an assessment are generated in the loop (including validation checks).

3.4.3 Generation of other data types

Other metrics may be required for management (e.g. environment metrics related to population dynamics) and evaluation of these could be conducted by either including mechanistic models linked to population dynamics (modelling change in climate or variables that might directly or indirectly impact the population dynamics) or following an empirical approach to evaluate the impact of climate change and environmental variation ("what if" scenarios).

3.5 Decision model

This component of the management procedure (Figure 3.1) uses the assessment results, or directly the generated data, to produce the management action to be taken in response to the perceived status of the stock and fishery, according to a pre-determined process. On many occasions, a harvest rule (or harvest control rule) will be applied to establish a level of removals (TAC) from the population. Common types of harvest rules are:

- F-regimes: TAC derived from F, TAC as a fraction of measured biomass, direct effort regulation;
- Catch regimes: permanent quotas plus protection rule;
- Escapement regimes: leave sufficient spawning biomass after harvest to prevent recruitment impairment.

The output from the decision model could include recommendations for:

- Total allowable catch (TAC) or effort (TAE);
- Area or seasonal closures;
- Mesh or hook size restrictions.

It can be useful to structure a harvest rule in terms of some components. This way of structuring a rule may promote modular programming, and it may be a convenient framework for discussing and designing a rule.

The harvest rule often includes several components applied in a sequential manner:

1. A mathematical rule that prescribes a ***'primary' TAC*** (or other management measure). For example, a translation of an exploitation rate into a TAC.
2. ***Stabilizing terms***, which modify the 'primary' TAC by constraining the change in TAC from year to year, perhaps with exceptions (such as may be applied e.g. if stock biomass is perceived to be low).

3. ***Other modifying terms***, for example a fixed maximum and/or minimum TAC.

Interannual quota flexibility (often called a "banking and borrowing" scheme), when allowed in a fishery, may be used in many different ways. As such, in the MSE context, it may be more naturally treated as part of the implementation model (in the next section) than of the decision model.

Management strategies are typically expressed as legal texts. For a scientific evaluation, it is essential that there are no ambiguities. The practical test is that it should be possible to program (i.e. to code) the harvest rule. Ambiguities revealed at this stage may require iterative discussion with managers.

The main elements of the decision model are described next.

The basis on which the harvest rule is applied is often the SSB estimate at some time, according to the most recent assessment. There are other potential measures (e.g. estimates of total stock biomass, survey index, estimates of recruitment, observed mean length or age, estimated biomass of other stocks...) which may be used alone or in combination, or applied under different conditions. The basis may come from an assessment and short-term forecast, but may also be derived directly from a survey or fishery data rather than from the outputs of an assessment model (i.e. an empirical management procedure).

The MSE simulation should take into account that the basis is an observed or estimated quantity rather than a true population value (e.g. the basis may be the estimated SSB, but not the true SSB in the operating model). If the basis includes some variable that may also impact the true stock (e.g. an environmental influence that affects the stock), the influence as seen by the decision model has to have uncertainty attached to it, and not be identical to the impact on the true stock.

The decision rule itself is a parametric function of the basis, and can be specified in many ways. A common type of rule is a steady exploitation rate if the spawning stock is estimated to be above some threshold, with a reduction in the exploitation rate as stock status declines (Figure 3.3). The parameters will typically be a "management" value for F ("*Fmgt*", e.g. F_{MSY}) and a trigger point based on the SSB, and the rule is *F = Fmgt*min(SSB/SSBtrigger,1)*. Other rules can have more parameters, for example a parameter indicating the slope of the decline in F below the SSB breakpoint. These control parameters should be chosen to give desired performance of the rule. They are conceptually different from reference points, which must be avoided (limit reference point) or achieved (target reference point). Although sometimes relevant, there is no need for a control parameter to be identical to a reference point such as Bpa.

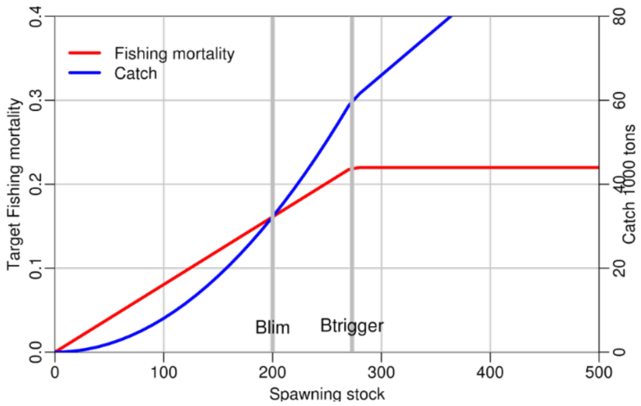

Figure 3.3. Common types of harvest rules.

The exploitation measure in the rule is most often a fishing mortality, but it can also be a harvest rate (i.e. the TAC is a fraction of the stock biomass), the TAC itself, or some effort measure, and it can be expressed in relative or absolute terms.

The translation mechanism typically converts fishing mortality to a TAC.

Both the basis and the translation mechanism may need stock numbers at some time after the last year in the assessment. If so, a short-term forecast is needed and should be implemented within the MSE simulation loop. The form of the harvest rule may sometimes require applying an iterative procedure, e.g. when the basis is the SSB forecast at the end of the TAC year; an example of this are escapement strategies which require solving for the fishing mortality that leads to the desired escapement at the end of the TAC year (e.g. NEA cod).

Stabilizers are often included in harvest rules. Their purpose is to limit the year-to-year variability in catches. The two most common stabilizers are ‘percentage rules’ and 'filter rules':

Percentage rule stabilizers limit year-to-year TAC changes to a specified percentage of the previous year’s TAC or realized catch. An exception is often made if the estimated stock falls below a certain threshold. Stabilizers can have important side-effects, which should be carefully considered. For example, they cause lags in TAC changes both when the stock is increasing, which may lead to foregone yield, and when the stock is declining, which may increase biological risk to the stock. Another feature of percentage stabilizers is that if the TAC gets drastically reduced in one year (perhaps because of a poor assessment that erroneously estimates a very low stock size) it may take a long time for it to get up again. It is also important to understand that, whereas this type of stabilizer tends to reduce TAC variability between consecutive years, it may increase the overall span of TACs over many years.

Another stabilizer type is a *filter rule* where the TAC is calculated as a weighted mean of the 'primary' TAC and the TAC (or catch) in *n* preceding or future years (e.g. a mean of the 'primary' TAC and the TAC from the previous year, or a mean of the 'primary' TAC and predicted future TACs). Formally, this is a simple low-pass filter. This type of stabilizer, when operating on past values (not predictions), follows change in the stock and may result in large changes following significant changes in stock size; it tends to reduce the overall span of TACs relative to percentage rule stabilizers.

In general, the parametric form of the harvest rule should avoid discontinuities, such as stepwise changes in exploitation rates (e.g. when the estimated stock biomass crosses some threshold), or very steep ramps. These forms can impose disruptive changes in TAC. Similar behaviour can be caused by stabilizers that are applied or removed at specified levels of estimated stock biomass. The trade-offs in management outcomes caused by harvest rule design, including stabilizers, should be evaluated as part of the simulation exercise and explicitly described to the managers and stakeholders.

Stabilizers may be considered in situations where a stock under rebuilding is estimated to have recovered to the point where it can move from a recovery program or moratorium to directed (e.g. F_{MSY}-based or HCR-based) fisheries. In these situations, simulated projections will be extrapolating beyond the recent data (in terms of stock size or fishing pressure) and, consequently, to states where the future stock response is particularly uncertain. As a result, it is important to be careful where simulations suggest that a stock could sustain a significantly higher fishing pressure than the stock has recently experienced. This is especially true for long-lived species (e.g. *Sebastes mentella*), which may recover very slowly from overfishing and where robustness testing to future productivity hypotheses is especially important. Two potential red flags are:

1. Where a proposed fishing pressure or catch level is significantly higher than that seen in the recent past.
2. Where proposed catches are at similar levels to those which have crashed the stock in the past.

For a medium- or long-lived stock, a potential approach to situations where the fishing pressure evaluated in simulation work to be sustainable for the stock is significantly higher than in the recent past, would be to phase in the increase in fishing pressure over several years, with the response of the stock monitored during this transition phase. Where the simulations suggest that catches at levels similar to those which previously crashed the stock would be sustainable, then these catches should not form the basis of advice unless robustness testing or additional evidence indicates that the stock can indeed sustain higher fishing pressure than in the past. In general, care should be taken with extrapolating into stock states beyond the recent data and either not experienced recently or in the historical data.

The duration of the decision is most often one year, but it can be longer (or shorter). Long intervals between decisions may be combined with gradual change of the TAC during the interval. This can be relevant in for example, rebuilding situations, where a very large reduction of the TAC is seen as necessary but hard to implement in a single year.

Potentially, harvest control rules may address more than one species at once, e.g. if mixed species advice is implemented according to set rules.

3.6 Implementation model

This is the step where the TAC derived from the harvest rule is converted to removals accounted for by the operating model. For an age-structured operating model, the TAC (or another measure derived in the decision model), needs to be converted to removals from the true stock in numbers at age. The selectivities and weight-at-age values needed for these calculations correspond to the true ones (i.e. those specified in the biological and fishery components of the operating model) and normally deviate from those assumed in the decision process.

An implementation model should account for the effects of differences between the intended pattern of removals derived from the harvest rule and the actual removals. Such differences can be caused by variable discarding practices, misreported catch, the implementation of different catch share management systems, bycatch in other fisheries not regulated by the TAC (for example industrial bycatch), or un-modelled fleet behaviour.

The extent to which assumptions shall be made about over-fishing (or under-fishing) of quotas is an open question that may have to be clarified with the managers. In some cases, set quotas have been consistently exceeded in the past, and the robustness of the rule to such persistent bias should be examined.

3.7 Validation

Simulation models are simplified descriptions of reality and can never replicate the real world exactly. Hence, validation of models is needed to ensure that the model describes the real system realistically enough for the intended purpose (Rykiel, 1996).

The absolute validation of ecosystem models is impossible because the ecosystems are open systems (Augusiak *et al.,* 2014). Nevertheless, we can gain confidence in the model through the application of the tools available for validation. The available tools are very diverse, from informal tools based on consultations with experts to formal tools based on mathematical methods like inference or induction. Balci (1997) provides an exhaustive list of the methods available to validate models. Alternative methods may also be helpful. Global sensitivity analysis (Saltelli *et al.*, 2008), for example, is a useful tool to validate models and it is a recommend by the European Commission in the implementation of impact assessment of management plans. This approach identifies the factors that have the highest impact on the output variance. In terms of validation, it is a useful tool to test if the model is really behaving as expected and if the range of scenarios defined is sufficiently broad.

Reality checks are also very important to increase confidence in the suitability and plausibility of the assumptions made in the MSE.

3.7.1 Biology and fishery model validation

Operating models are generally conditioned on data to ensure they are consistent with observations. If the future is intended to reflect past dynamics, as represented in the operating model, then validation needs to ensure this aim has been achieved. The following are examples of how this may be checked:

- Comparison of historical and simulated recruitment against SSB, check distributional form (e.g. via Q-Q and cumulative distribution plots; Figure 3.4), autocorrelation, and fluctuating and episodic recruitment.
- Ensuring that there are no unexpected discontinuities between the past and future dynamics in the operating model (Figure 3.5)
- Ensuring that the model can replicate the recent past by hindcast projections, i.e. runs where the operating model starting some years back in time and condition it to reproduce the historical development of the stock. The hindcast projection is then compared with the realised values of key statistics/input data.

The behaviour of the population model itself can be checked by running it forward with fixed values of F, for example, running the evaluation with zero F in future to check the behaviour is as expected.

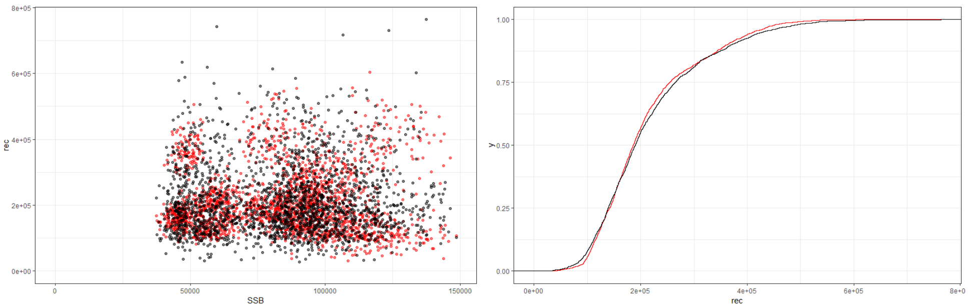

Figure 3.4. A comparison of historical (red) and future simulated (black) recruitment, plotting all stock-recruit pairs for 100 MSE iterations (left) and using an empirical cumulative distribution function (ecdf in R; right). These historical and future recruitment values are comparable because they are based on the same SSB values.

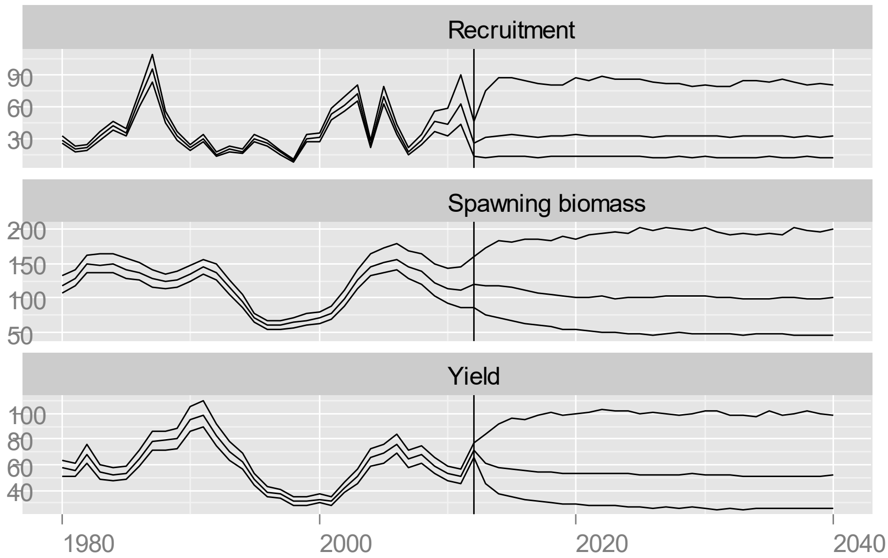

Figure 3.5. Example of historical assessment (assessment year is 2012) and future expectation of recruitment, SSB and yield, when future fishing mortality is kept similar to the average of that observed historically. The figure displays 5th, 50th and 95th percentiles

3.7.2 Observation model validation

Validating the performance of the observation model is essential to ensure the MSE provides a realistic evaluation of management procedure performance, whether running a full or shortcut MSE. There are no routine tests that can be universally recommended. The bullet points below could be worth considering.

- For a full MSE, justify the approach used to characterise noise in the input data, and validate e.g. by ensuring future noise is consistent with historically observed noise. Generating data that is as "messy" as historically observed when fitting them (e.g. due to overdispersion and model misspecification) is one of the most difficult steps in the MSE process.
- Since the shortcut approach combines the observation and estimation models in order to approximate the assessment model behaviour, careful validation is needed to ensure the approximation used is adequate (see next subsection).

3.7.3 Estimation model validation

In a full MSE, as long as the properties of the data are appropriately captured in the observation model, there is less concern about behaviour of the estimation model, given that the stock assessment model to be used in practice is implemented within the simulation loop. It should be remembered that the purpose of the MSE evaluation is to evaluate the performance of the management procedure as a whole, including both estimation and decision models.

Since the shortcut approach combines the observation and estimation models in order to approximate the assessment model behaviour, validation should ensure future assessment behaviour is consistent with that observed historically (Figure 3.6 provides one potential check). Furthermore, the justification for using an approximation to a management strategy may be examined by running a few simulations for the actual management strategy (i.e. running a few simulations of a full MSE) and the shortcut approximation using the same operating model set, and comparing the results to ascertain whether the approximation is adequate (Punt *et al.*, 2016).

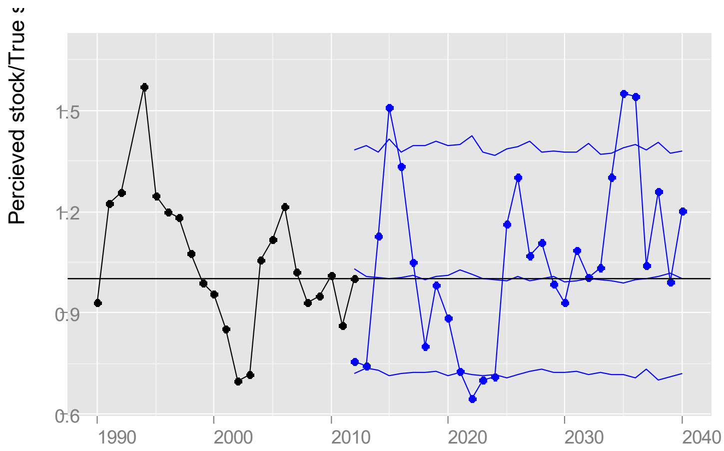

Figure 3.6. Example of the ratio of the perceived stock vs. the true stock biomass. The historical part of the plot (black line) is the based on empirical retrospective performance (ratio of contemporaneous estimates vs. the most recent assessment) upon which future assessment error is based (ratio of observation model biomass vs. the true biomass). The lines denote the 5th, 50th and 95th percentiles, and one iteration is shown as an example.

3.7.4 Decision model validation

A first practical test of any decision model is that it can be programmed (i.e. if a request for an evaluation of a decision rule is received, then one must be able to convert this decision rule into computer code). Further validation tests could include the following:

- Run the MSE with perfect knowledge, and compare this with the management decision model including observation and assessment error to check the impact of the errors. It may be that the management strategy is not precautionary even under perfect knowledge. This is also useful as a code check.

3.7.5 Implementation model validation

One could model fleet responses to management decisions, which may need to be validated based on historical observations.

3.8 Special considerations short-lived species

The specific life-history characteristics of these stocks often result in different types of management strategies (such as biomass escapement strategies) being applied to the fisheries on these stocks when compared to the types of strategies normally applied to fisheries catching medium- or long-lived species. The risk tolerance and the reference points developed for these stocks may also differ from their counterparts for medium- and long- lived stocks. There was no time at the WKGMSE 2 workshop to devote special attention to this situation, but information can be found in ICES (2013b) and, more recently, in ICES (2016c, 2017a, 2017b). The technical simulation guidelines for MSE for these stocks are no different from those presented in this report for other types of stocks.

3.9 Stocks with sparse information

When the information about the stock is too sparse to permit the usual procedure of assessment and prediction, management strategies may still be developed, but will normally have a different form from those applied when more information exists. Simulating such management strategies requires operating models which may have to be more generic and less stock-specific, and the strategies will have to be more robust to uncertainties than when more precise information is available. There was no time at the WKGMSE 2 workshop to devote attention to this situation, but the reader is referred to the reports of the ICES "Workshop on the Development of Assessments based on LIFE history traits", which provide extensive information and analysis of this situation; there are eight reports from this workshop so far, the latest report being from 2018 (ICES 2018d).

4 ICES Precautionary Approach Evaluation Criteria.

4.1 Sources of variability - what does risk cover

A criterion that must be considered when evaluating a management strategy is whether it conforms to the precautionary approach. This requires consideration of the probability of the stock biomass (typically SSB) falling below the limit biomass reference point (B_{lim}, the SSB value below which the stock is considered to have reduced reproductive capacity) when the management strategy is used. For a management strategy to be considered precautionary, ICES requires that this probability should not exceed 5% (see Technical Guidelines document "ICES criteria for defining multi-annual plans as precautionary", ICES 2016c).

The workshop noted that having a single absolute value of B_{lim} for each stock created problems when conducting an MSE, because different operating models are used to represent alternative resource dynamics, for example different natural mortality or stock-recruitment relationship, which means that the appropriate B_{lim} value may be different for different operating models. The workshop therefore suggests that relative measures, instead of absolute ones, should be considered when deriving operating model summary statistics from MSE results, although a single absolute value could still be used in the definition of the harvest control rule.

Another aspect that deserves consideration is that, when conducting an MSE, the value obtained for the probability that SSB < B_{lim} (also termed "risk"), depends on the assumptions made during the MSE, such as those concerning the operating model, assessment and implementation errors. To ensure the robustness of a management strategy it is important that the assumptions made in the MSE encompass the main sources of uncertainty about the resource dynamics. However, the variability of SSB outcomes, and thus the risk of being below B_{lim}, will vary according to the amount of uncertainty that has been assumed in the MSE. As such, it cannot be scientifically determined whether absolute risk is above or below 5%.

Globally, general practice considers that MSE provides answers about relative performance among alternative management strategies rather than absolute levels of performance for a particular management strategy (e.g. Butterworth *et al.,* 2010; Rochet and Rice, 2009). If the absolute criterion that $P(SSB<B_{lim}) \leq 5\%$ is to be retained by ICES, WKGMSE2 recommends that further attention be given to finding ways of ensuring consistency in the calculation of this statistic within and between stocks/management strategies. The statistic should be derived from simulations making comparable assumptions across the board by convention. Given different complexities for different stocks, any 'base case' operating model for calculating this absolute performance statistic would need to make basic assumptions that could be widely applied, such as e.g. simply using best estimates of biological parameters and incorporating variability on recruitment and starting numbers (from the benchmark assessment), with no implementation error. A procedure along these lines might best be implemented within some broad categories of situations (e.g. separately distinguishing short-lived species). On the other hand, to ensure within stock consistency, similar assumptions to those made when calculating reference points for the stock based on the benchmark assessment may be appropriate. The workshop considers that ICES should give further attention to these issues.

4.2 Definitions (percentage, time frame)

Separately from the issues discussed in the previous section, another relevant aspect is that the statement "the probability that SSB is below B_{lim}" can be interpreted in different ways, and different interpretations did indeed occur when ICES evaluated management plans in the past. This is important because, depending on the interpretation used, the requirement that this probability should not exceed 5% is more or less stringent. The previous ICES workshop on guidelines for MSE (ICES 2013b), made a proposal for a particular interpretation of this statement across the board ("Prob3", see below), which was subsequently adopted by ACOM (ICES, 2016c). This subsection provides an update to the text in the corresponding subsection of the ICES (2013b) report.

This report uses the wording "probability that SSB is below B_{lim}" to avoid confusion with other interpretations of risk (see ICES 2013b for further details).

A review of previous ICES practices undertaken by ICES (2013b) showed that three interpretations of "the probability that SSB is below B_{lim}" had been used in the past:

- $\boldsymbol{Prob1}$ = *average* probability that SSB is below B_{lim}, where the average (of the annual probabilities) is taken across ny years.
- $\boldsymbol{Prob2}$ = probability that SSB is below B_{lim} *at least once* during ny years.
- $\boldsymbol{Prob3}$ = *maximum* probability that SSB is below B_{lim}, where the maximum (of the annual probabilities) is taken over ny years.

These definitions imply that $Prob2 \geq Prob3 \geq Prob1$, so requiring that $Prob2 \leq 0.05$ is a more stringent condition than if this is required based on $Prob3$ or $Prob1$. It is clear from their definition that in a stationary situation (which generally occurs in the "long term", after the effect of the initial stock numbers has disappeared), $Prob3 = Prob1$ (note that this statement refers to the actual definitions of $Prob1$ and $Prob3$, and not to the way these quantities may be computed in the MSE simulation, which will be discussed in Section 4.3). By contrast, in a non-stationary situation (generally in the "short term", corresponding to the first few years in the simulation) $Prob3$ can be considerably larger than $Prob1$. $Prob2$ can also be considerably larger than $Prob3$ and $Prob1$, particularly for stocks with low temporal autocorrelation in SSB (as may be expected for short-lived species). This means that, all other things being equal, $Prob2$ may be expected to be higher for short-lived than for long-lived species. On the other hand, once a stock is below B_{lim}, it will generally take longer for it to recover if it is a long-lived species, but $Prob2$ does not take this into account as it is just focused on the probability of the stock being below B_{lim} at least once in the ny years period considered.

As already noted, currently ICES uses the *Prob3* ≤ 0.05 criterion as the basis for defining a multiannual plan as precautionary, although with exceptions made in cases requiring an initial recovery phase, or where a short-lived stock's natural variability (without fishing) exceeds the 5% threshold value (ICES 2016c).

MSE simulations normally consist of a non-stationary phase, with dependence on initial stock numbers (the "short term"), and a stationary phase, which is further into the future once the dependence on initial stock numbers has disappeared (the "long term"). In the short term, the distribution of SSB changes from year to year and, therefore, so does the probability that SSB is below B_{lim}. In this case, it is recommended that these probabilities are examined in each individual year, to get a good understanding of how the stock biomass is evolving over time, and that this examination is carried forward in time until the long-term stationary phase has been reached.

In particular, two forms of reporting should be used:

1. A plot showing the 5, 50 and 95 percentiles of the marginal distribution of SSB/B_{lim} in each year, together with a horizontal line at 1 (i.e. corresponding to $SSB = B_{lim}$; e.g. Figure 4.1). This allows seeing immediately from the graph whether the probability that SSB is below B_{lim} is bigger or smaller than 5% in each of the years. It also allows detecting possible trends in this probability and, potentially, picking up other factors that may be having an impact on it.

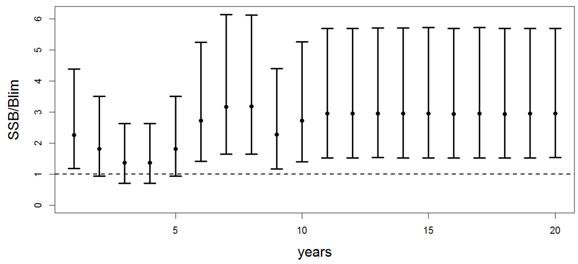

Figure 4.1. Distribution of SSB/B_{lim} per year (median, 5 and 95 percentiles). The horizontal line at 1 corresponds to $SSB = B_{lim}$

2. A table showing the probability that SSB is below B_{lim} in each of the years.

Table 4.1

Year	1	2	3	4	5	6	7	8	9	10	11 and onwards
P(SSB<B_{lim})	0.02	0.07	0.22	0.22	0.07	0.01	0.00	0.00	0.02	0.01	0.00

With this figure and table it is possible to gain a good understanding of how the stock biomass evolves over time in relation to B_{lim}. In this case, there is more than 5% probability that SSB is below B_{lim} in years 2, 4, and 5 of the simulation, whereas it is less than 5% in all other years, including in the long term.

When using multiple operating models, the figure and table should allow for easy comparison of their associated risk.

Table 4.2 presents the values of $Prob1$, $Prob2$ and $Prob3$ calculated over the 20 years, only over the first 10 years and only over the final 10 years. $Prob1$ and $Prob3$ can just be obtained from Table 4.1. This is not the case for $Prob2$, whose value depends on the amount of temporal autocorrelation in SSB. The $Prob2$ values shown in Table 4.2 are from an example with autocorrelation in SSB among years of 0.5. This shows the short term difference and long term similarity in *Prob1* and *Prob3* and the increase in *Prob2*.

Table 4.2

	Years 1-20	Years 1-10	Years 11-20
$Prob1$	0.03	0.06	0.00
$Prob2$	0.42	0.40	0.02
$Prob3$	0.22	0.22	0.00

4.3 Precision – iterations needed

The performance of management strategies may be assessed according to a variety of summary statistics (i.e. indicators) of interest, such as breaching limit reference points related to stock status, catch, and risk statistics. As MSEs are stochastic by nature, the number of iterations (i.e. independent replicates in the simulation) performed will affect the level of precision and the robustness with which the summary statistics are computed. An MSE should thus demonstrate that the number of iterations is sufficiently large in order to allow the performance of alternate management strategies to be compared and to provide robust estimates of risk.

Estimation of risk often deals with the tails of a given distribution (mainly P(SSB < B_{lim})), such that defining the number of iterations needed for its robust computation is likely to provide a stringent enough value for the number of iterations needed to robustly compute other summary statistics. Nevertheless, the stability of each indicator may need to be assessed independently. Furthermore, the life history of the species and the assumed variability in recruitment and other processes will affect the effective iteration number due to an influence on the degree of stochasticity. A general rule may be to increase the number of iterations until a given level of convergence is reached for an index's distribution quantiles (e.g. confidence intervals assessed via quantile regression). The following provides an example the influence of iteration number on risk estimation precision.

If $prob$ is the value of the probability that SSB is below B_{lim} obtained if an infinite amount of iterations could be performed (i.e. averaging the results from an infinite number of iterations), its value computed on the basis of $niter$ independent iterations has a distribution centred at $prob$ (except for $Prob3$, where this procedure is biased, as explained later), with standard deviation $\{prob * (1 - prob)/niter\}^{1/2}$. Therefore, the probability calculated on the basis of $niter$ iterations will be within the interval $prob \pm 1.96 * \{prob * (1 - prob)/niter\}^{1/2}$ in approximately 95% of the cases. This allows an approximate calculation of the number of iterations required to compute $prob$ with a certain precision. For $prob = 0.05$, the following table gives the intervals that result for different number of iterations:

Table 4.3. Distribution of $P(SSB < B_{lim})$ computed based on niter iterations, when prob=0.05 for an infinite number of iterations Table 4.3 implies that if $prob = 0.05$, then performing a simulation with $niter$ iterations and computing $P(SSB < B_{lim})$ based on the simulation produces a value which is within the interval presented in the table in approximately 95% of the cases. Therefore, if e.g. a simulation based on 500 iterations gives a value of $P(SSB < B_{lim})$ smaller than 0.03, one can be quite certain that $prob < 0.05$, whereas if it gives a value of $P(SSB < B_{lim})$ bigger than 0.07, one can give quite certain that $prob > 0.05$. However, if it gives a value between 0.03 and 0.07, it is unclear whether $prob$ is above or below 0.05. In that case, further precision can be obtained by increasing the number of iterations.

$niter$	2.5 percentile	97.5 percentile
100	0.01	0.09
250	0.02	0.08
500	0.03	0.07
1000	0.04	0.06
2000	0.04	0.06
5000	0.04	0.06
10 000	0.05	0.05

The intervals in Table 4.3 are directly applicable to annual values of $P(SSB < B_{lim})$ (for each individual year, considered separately from the other years) and $Prob2$.

The intervals in Table 4.3 can also be used as "safe" guidance for $Prob1$ computation, even though the intervals for $Prob1$ will typically be narrower than those given in Table 4.3 because in $Prob1$ an average is taken over several years, which increases precision (although the gain in precision decreases with increasing autocorrelation in SSB). A simple simulation exercise showed that in a stationary situation, the interval in Table 4.3 reduces to [0.04, 0.06] with $niter = 250$ when $Prob1$ is computed as a 10-year average, even under high autocorrelation in SSB (ϱ = 0.8).

The computation of $Prob3$ is less precise than Table 4.3 indicates, because, as $Prob3$ is the maximum of the annual values of $P(SSB < B_{lim})$, it amplifies the noise in the computed annual values, resulting in poor convergence of the computational approach, including bias in the sense that more often than not the computed value of $Prob3$ will be larger than the true value (i.e. the value obtained under an infinite number of iterations). The rest of this section explores computational aspects of the $Prob3$ calculation.

Figure 4.2, taken from MSE work for North Sea cod done in ICES after the WKGMSE 2 meeting ended, shows the poor numerical convergence properties and the bias of the $Prob3$ computation.

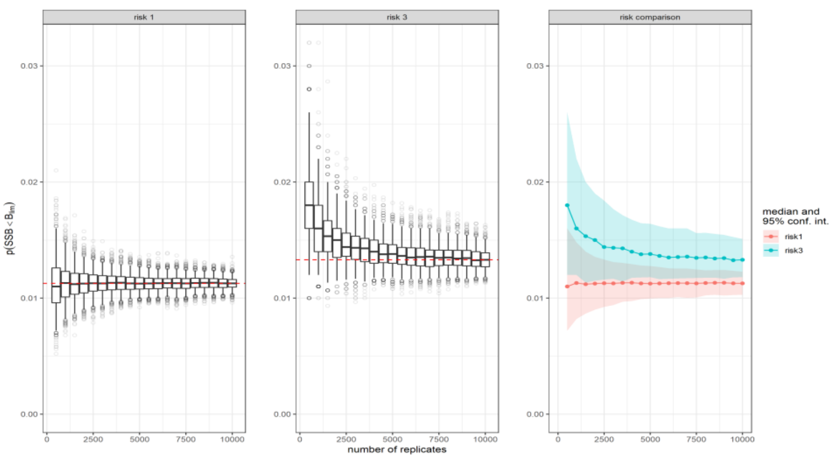

Figure 4.2. Computed values of Prob1 (left panel), Prob3 (center panel) and both Prob1 and Prob3 (right panel) versus number of iterations (or "replicates") in the MSE simulation. The Prob1 and Prob3 values shown in this figure were calculated for years 11-20 of the projection period.

The $Prob1$ and $Prob3$ values shown in Figure 4.2 were calculated over years 11-20 of the projection period. An examination of $P(SSB < B_{lim})$ by year suggested that those projection years were very near stationarity (see Figure 4.3, where the 10 years within vertical dashed lines correspond to those used in Figure 4.2).

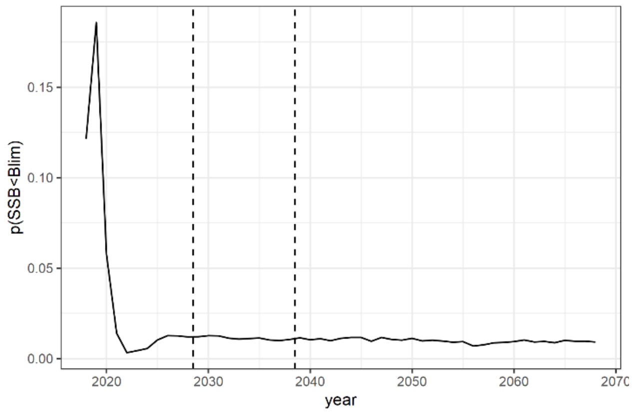

Figure 4.3. Annual values of P(SSB<B_{lim}) computed based on 10000 iterations.

As a further investigation of the issues related to the $Prob3$ computation, a controlled experiment corresponding to a perfectly stationary situation was constructed. In the experiment, the true value of $P(SSB < B_{lim})$ was 0.04 in every year, so the true $Prob1$ and $Prob3$ value, calculated over any number of years, was 0.04. Figure 4.4 shows the values of $Prob1$ and $Prob3$ computed over 10 years using different numbers of iterations. It is clear from the figure that convergence is much better for $Prob1$ than for $Prob3$.

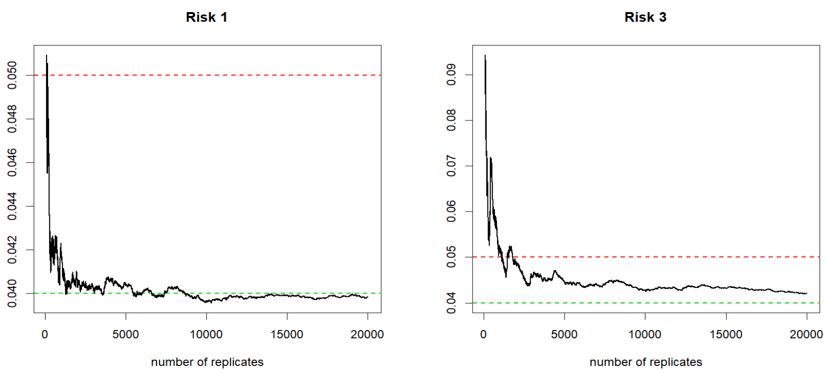

Figure 4.4. Computed values of Prob1 (left panel) and Prob3 (right panel) versus number of iterations (or "replicates") in a controlled experiment with perfect stationarity. Up to 20000 replicates were considered. The Prob1 and Prob3 values shown in this figure were calculated over 10 years and the true value is 0.04 for both Prob1 and Prob3. Note the different scales on the vertical axes.

Based on the investigations conducted, it is concluded that, in a stationary situation, given that $Prob3 = Prob1$ in that case, only $Prob1$ should be computed (because of the much better convergence of the algorithm to compute $Prob1$). In the MSE context, stationarity is generally reached for a period of years sufficiently far into the future. Therefore, a more computationally efficient approach for $Prob3$ over such a period of years (with stationarity) is to compute the average risk over those years, in exactly the same way $Prob1$ would be computed. This computational approach is valid because, as noted in Section 4.2, in a stationary situation $Prob3 = Prob1$.

It must, however, be kept in mind that stationarity in an MSE is not a given, and it should not be taken for granted that it will always occur in the long-term. There may be non-stationary aspects, such as trends or random walks, built in some MSEs that do not disappear even when simulating the long-term. So this is not a straightforward issue. Displaying the development of SSB relative to B_{lim} by year, as in e.g. Figures 4.1 and 4.3 and Table 4.1, is helpful for gaining insights and is recommended.

In the short term, where the situation is definitely non-stationary, it makes sense to consider annual $P(SSB < B_{lim})$ for each of the years, as indicated in Section 4.2 (see also Figure 4.3). When each year is seen in isolation, the intervals in Table 4.3 apply. However, when looking at the ensemble of ny years and then focusing on the worst year (i.e. $Prob3$) the situation is different. In computational terms, $Prob3$ is not just a direct average over the iterations; instead, an average over the iterations is computed for each year separately, and then a maximum taken over the ny years. To illustrate the effect of this, imagine that the true value of $P(SSB < B_{lim})$, i.e. the value obtained under an infinite amount of iterations in the MSE, is < 0.05 in all years and that $niter$ iterations are used in the actual computation. When a specific year y is considered, there is some probability that the computed value of $P(SSB < B_{lim})$ is bigger than 0.05 (just by chance), leading to the wrong conclusion for that particular year. Using the same amount of iterations ($niter$), it is intuitively clear that the probability of wrongly concluding that $Prob3 > 0.05$ increases when ny years are considered together and the focus is on the worst year. Intuitively, $Prob3$ computed based on $niter$ iterations is a *biased* estimator of the true value (i.e. the value obtained under an infinite number of iterations) and more often than not the computed value of $Prob3$ will be larger than the true value. The bias is bigger the bigger the number of years ny considered, the more similar the annual values of $P(SSB < B_{lim})$ in the different years, and the lower the autocorrelation in SSB.

Applying the following steps can help more quickly to reach a conclusion regarding whether the true value of $Prob3$ is above or below 0.05 for periods of years where the MSE is not stationary:

1. Start by computing $Prob3$, for the relevant period of years, using the number of iterations in Table 4.3.1.
2. If the computed $Prob3$ value is below the lower end of the interval in Table 4.3, then it may be concluded that $Prob3 < 0.05$ (given the bias in the $Prob3$ computation).
3. If the computed $Prob3$ value is above the upper end of the interval in Table 4.3, then compute $Prob1$ for the same period of years. If the computed $Prob1$ value is also above the upper end of the interval in Table 4.3, then it may be concluded that $Prob1 > 0.05$ and, therefore, the same holds for $Prob3$ (since $Prob3 \geq Prob1$, as noted in Section 4.2).
4. Otherwise no conclusion can be reached and the number of iterations should be increased until the value of $Prob3$ stabilizes in an area where conclusions can be drawn.

A different technical approach was applied by ICES (2017c) to deal with the poor numerical convergence properties of the $Prob3$ computation. In that case, the time-series of computed annual risks was smoothed by applying a smoothing spline with 14 degrees of freedom, before taking the maximum over the relevant period of years. For the long-term (stationary) period, the results of the smoothing approach were contrasted with those obtained calculating $Prob3$ as $Prob1$ for that period, which suggested that the process followed with the smoother on the annual risks did a reasonable job.

As already noted, a simple idea that can help with these computational difficulties, is to produce plots and tables such as those shown in Figures 4. 1 and 4.3 and Table 4.1, which can help to get an understanding of how risk develops over time. Some trends in risk over several years will likely be seen, which can help to “visually” smooth the risk values in some years.

Conclusions:

- For computing $Prob2$, $Prob1$ and $P(SSB < B_{lim})$ in a single specific year, the intervals in Table 4.3.1 can serve as guidance.
- In most cases, $Prob1$ requires fewer iterations than suggested in Table 4.3 (taking advantage of averaging over years, but the gain in precision decreases with increasing SSB autocorrelation).
- Computing $Prob3$ requires more iterations than suggested in Table 4.3 (potentially many more, as the computed $Prob3$ value can converge very slowly to the true value, i.e. the value obtained under an infinite number of iterations). Several ideas that may result in some efficiency gains in the computational approach were presented above.
- It is recommended that the relevant risk measure used in the analysis (for ICES stocks usually $Prob3$, as agreed by ACOM) be plotted against iteration number (i.e. compute $Prob3$ based on the first $iter$ iterations and plot it vs. $iter$, the iteration number), to get an understanding of how many iterations are required for the computation to stabilize in an area where conclusions can confidently be drawn.

4.4 ICES PA practice

Following a proposal from the previous ICES workshop on MSE guidelines (ICES 2013b), ACOM agreed to use $Prob3 \leq 0.05$ as the main criterion for defining multi-annual plans as precautionary, albeit with additional special considerations for short-lived stocks and stocks in a recovery phase. This is detailed in the Technical Guidelines document entitled “ICES criteria for defining multi-annual plans as precautionary” (ICES 2016c).

In particular, for recovery plans or initial recovery phases within long-term management strategies, the document states:

“If a stock’s SSB is currently below B_{lim}, it is not logical to expect that the probability of SSB < B_{lim} is ≤ 5% in all years. It seems more logical to judge a recovery plan (or an initial recovery phase within a long-term management plan) by its ability to deliver SSB recovery within an appropriate time frame. In such a case, the recovery plan would only be considered precautionary if the probability of SSB > B_{lim} in a pre-specified year is ≥95%. If the recovery plan constitutes an initial recovery phase within a long-term management plan, the requirement that the maximum probability of SSB < B_{lim} is ≤5%, where the maximum (of the annual probabilities) is taken over all years in the plan/strategy, should apply to the after-recovery long-term management plan.”

4.4.1 Rebuilding MSEs

WKGMSE 2 participants felt better guidance is needed on how to evaluate stock rebuilding plans, i.e. where the stock biomass falls below B_{lim}. Currently there are two standards which give rise to conflict, (i) the requirement that, for a multi-annual plan to be considered as precautionary, the probability that SSB > B_{lim} must be at least 95% for every year of its implementation period, and (ii) the current ICES advice rule, where ICES practice is to advise zero catch when SSB cannot be rebuilt above B_{lim} in the short term (1 year). Since a rebuilding plan starts from a non-precautionary situation by definition, it does not pass the precautionary standard.

The workshop recommends the stock's recovery phase should be evaluated in a longer term strategy and not in isolation. Having a longer-term perspective of management performance will allow the evaluation of the rebuilding phase and its impact on the long-term objectives. In such cases the precautionary requirement can be evaluated for the period after the recovery phase and different options to recover the stock can be evaluated as well.

With regards to catch advice when a stock is estimated to have biomass levels below B_{lim}, the workshop recommends that, during the evaluation of management options, the specification of this region of the harvest control rule should be better explored. This would include taking account that there is a high risk of recruitment impairment below B_{lim}. When stocks are already below B_{lim} a dialogue is required between scientists, as risk assessors, and policy makers, who have the competency for risk management. The trade-off between the level of acceptable risk and the time-frame to recovery above B_{lim} is a management decision. WKGMSE 2 recommends that a further workshop should be convened to develop guidelines on how to evaluate and advise on rebuilding strategies in the context of longer-term management strategies.

5 MSE process

Involving all the players (Advisory Councils (ACs), managers, policy makers and scientists) in the MSE process from the earliest stage is important to underpin the legitimacy and saliency of the result. The process should encourage representative participation from the stakeholders, which means not only actively striving for gender balance in workshops and meetings, but also aiming to ensure that all affected and interested parties are represented across relevant ethnic, cultural and social groups.

Dialogue should underpin the MSE process that ought to accommodate and respond to information that comes from the identified stakeholders, including the information in the form of local knowledge (e.g. MSC requirement). The stakeholders need to be kept informed of how the information they provided was used and, if it was not used, why that was the case. It is also important that there is a common and detailed understanding of what the request from managers actually means and what should be done by those scientists trying to answer the questions asked. Such clarity would lead to a more efficient evaluation workshop, as there would be no need to spend time debating the likely meaning of the request.

In this section two types of procedures for answering requests on management strategies are presented: one of them may be called "standard MSE process" and is similar to the process that has been applied in ICES in recent years; the other one is what we term "strategic MSE process".

5.1 Standard MSE process

The four steps below are recommended by WKGMSE2 after ICES has received a request from clients (Figure 5.1).

5.1.1 Initiating the process and scoping

The process will typically start when ICES receives a request from a client, to evaluate a management procedure, often involving a harvest control rule with different exploitations levels and management actions reference points. Often the basic question asked is to find the F_{target} and Btrigger combinations that are considered precautionary when used in the context of the requested harvest control rule (often this rule is one that aims at constant F_{target} exploitation, reduced as a linear function of stock biomass when this falls below $B_{trigger}$).

At this stage there would have been some, but varying, interactions between clients, stakeholders (mainly industry) and scientists.

When the ICES Secretariat receives the request, a scoping process is started where the chair of the process, reviewers and modellers are identified. The request and pro-posed process are posted on the ACOM forum for formal approval.

5.1.2 Clarification meeting and protocol

Prior to the start of technical work to answer the request, a clarification meeting should take place (currently in ICES clarification is normally done by correspondence). At this meeting, the clients who submitted the request should be present, together with scientists and the ICES Secretariat. The purpose of this meeting is to clarify the request in order to avoid later misunderstanding, and also potentially to add alternative aspects to the request. During the meeting, performance indicators will also be identified and agreed, as well as definition of short term, medium term and long-term aspects. The product of the meeting will be a protocol based on the condensed guidelines in Section 6.1, which should include the following elements:

1. An explanation of any unclear parts of the request;
2. Data available;
3. Methods description:
   Following the terminology in Figure 3.1.1:
   Operating Model, including components for: biology and fishery, observation generation, and implementation.
   Management Procedure, including components for: estimation, and decision (with explicit description of the form of the harvest control rule(s)).
4. A workplan for the simulations and model runs expected to be performed;
5. A list of performance indicators;
6. When the request includes testing ranges (e.g. F from 0 to 0.5), the steps should be specified;
7. Definition of short, medium and long term;
8. Explicit description of how the harvest control rule is applied in cases where the stock is below B_{lim};
9. Scenarios to be tested;
10. Validation checks to be carried out;
11. A plan for dissemination of the result (e.g. verbal presentations at Advisory Councils).

Once the protocol has been agreed at the clarification meeting, it should be sent to the reviewers for commenting. These comments should be considered, in case modifications are needed to address reviewer comments before the subsequent workshop. This ensures that reviewers are aware of the context, content and scope of the work at an early stage.

5.1.3 Workshop / Consultation process

The workshop will resemble a typical ICES MSE workshop. It will be preferably a physical meeting lasting approximately one week. The workshop should conclude with a larger consultation process where the scientists present preliminary results to stakeholders and clients. This will allow for further analysis to be carried out (e.g. additional scenarios) before the results and report are finalized.

5.1.4 Review and publication

The results of the workshop form the evidence base for the Advice Drafting Group (ADG) made up of ACOM members, alternates and nominated participants. The ADG prepares the advice, which is approved by ACOM, usually by correspondence and a WebEx meeting if required. Both the ADG and the ACOM approval WebEx are open to client and stakeholder observers.

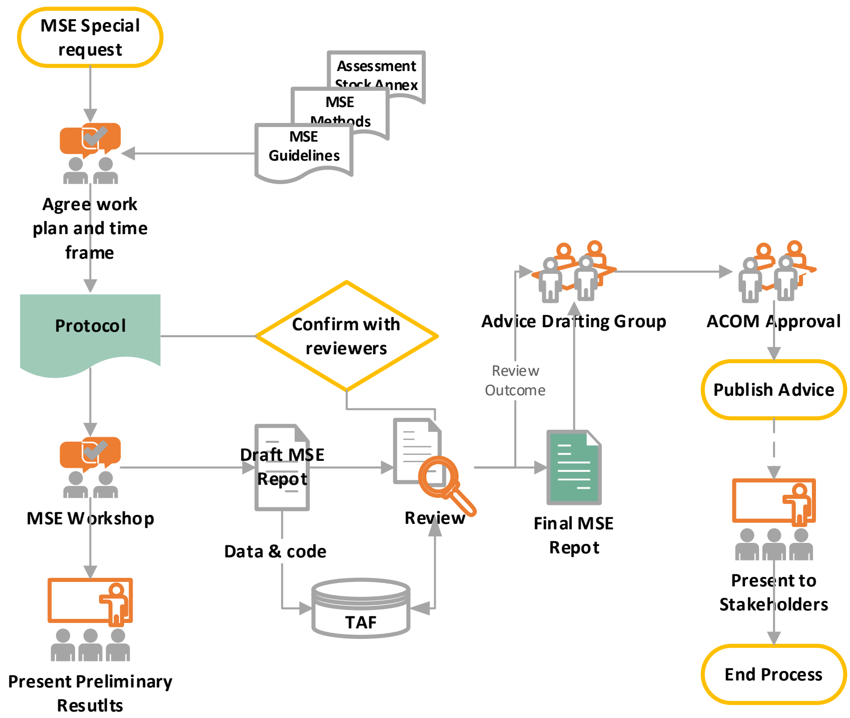

Figure 5.1. Workflow for standard MSE process.

5.2 Strategic MSE process

A more strategic MSE process could take place as part of a benchmark process. This type of process can be developed internally in ICES, and not necessarily linked to a special request for clients. The process could start with concerns raised during an update assessment about the performance of the assessment or management procedure. The start of the process could also be stakeholder driven, e.g. related to the Advisory Councils developing their own management strategy, or it could be linked to a management strategy being developed in collaboration between industry from several countries. This strategic MSE should include robustness testing of the underpinning assessment model or models to a range of proposed management procedures and, if relevant, take into account multi-species, ecosystem, social and/or economic objectives.

The setting of objectives and defining the types of tactical approaches to be considered is a role for the stakeholders (managers and industry and NGOs). In an iterative process, scientists can help express these objectives and tactics as rules which can be implemented in an MSE. It is a role of the scientists to provide the technical documentation that provides the evidence base for the decisions adopted in the management plan.

The main participants and the actions at each stage in the above process are:

Initiation (Mainly Decision Makers, but also ACs+ others)

- Begin discussions amongst all coastal states
- Scope the problem (Decision makers, Experts, ACs, Implementers)
  - Decide who is involved and what biological/environmental /social / economic / other aspects should / can be involved. Decide which part of the modelling approach is feasible interactively.
- Development process (Coordination – responsibility is the initiators)
  - Define Resources (Decision makers, Experts, (Implementers))
    - Time frame
    - Personnel resources
  - Set criteria and analytical aspects (Decision makers ACs (facilitator experts))
  - Carry out calculations (Experts, (Implementers) (All))
    - These needs to be transparent – but also needs to be quality checked
    - May not be possible interactively
  - Carry out evaluations (all)
  - Develop visualisations of the results
  - Communicate discuss results
    - All
- Iterate around the loop as required

5.3 Communication of MSE results

Communication of MSE results should reflect the needs of the various audiences, chief amongst these are other scientists, managers, NGOs and general public. A single method is not suitable for all of the audiences although improvements and standardisation are needed at all levels of technical details compared to the status quo. MSE results should be presented in a way that empowers a reader, for example, to think critically about the results in an appropriate context.

In order to create the appropriate context, a standardised approach to representing the main uncertainties accounted for in the process relative to a wide range of generic uncertainties should be considered. Further, a guide to relative reliability of the model should be provided, this should address data, knowledge basis as well as model-related issues. Specific concerns about data, knowledge or model validity should be clearly highlighted in an easy-to-find section of the MSE report.

The complexity of fisheries modelling makes it imperative to involve experts in communication, in particular visual communication, in the formulation of advice on how to communicate MSE results. Further, methods developed with experts, such as graphic designers, should be tested with the intended audiences, improved and then standardised across ICES.

Such standardisation might be achieved through a catalogue of examples accompanied by R code, or other graphics that can be easily reproduced and based on widely used platforms:

- Microsoft: https://powerbi.microsoft.com/en-us/
- R shiny: https://shiny.rstudio.com/gallery/
- Tableau: https://www.tableau.com/
- D3: https://github.com/d3/d3/wiki/Gallery

Increasingly, interactive visualisations via web applications are being used to let stakeholders explore MSE results. For example, North Atlantic Swordfish MSE had worked with graphic designers to create a shiny web application that communicates background uncertainty, reliability and results of MSE evaluations for a range of management procedures and operating model hypotheses: https://pl202.shinyapps.io/Swordfish_MSE_Vis/

The current visualisations used within ICES do not follow design-based principles, impeding communication between scientists as well as making the details of assessment unnecessarily difficult to understand for anyone else. This is not just a matter of transparency or accessibility, but of placing obstacles in a way of potential contributions by the various stakeholders in the MSE process.

As already discussed in Section 4.7 of this report, the credibility of simulation results can be demonstrated through diagnostic procedures and summary statistics. For example, comparison of the dynamics of recruitment, spawning biomass, and catch in the future projections and the historical period can help to evaluate whether the simulations produced reasonable dynamics (Figure 5.2). Although outliers in these time series could be features of either the harvest control rule or operating model conditioning, their identification would allow for further investigation of the simulation by the analyst.

Many diagnostics procedures are performed during a stock assessment that should also be done in the MSE simulation. For example, it should be confirmed that it is likely that that the assessment model has converged and that estimated parameter values are reasonable. In a scenario when many assessment runs are performed (depending on the length of the projection period and the number of simulation replicates) summary statistics that indicate good behaviour, such as the convergence rate and the rate of parameter estimates within boundaries, can provide a quick check of performance. Even with high convergence rate, additional summary statistics, such as the mean value of model residuals, can be reported to serve as indicators of assessment quality within the MSE. Looking at the behaviour of residuals can be helpful – patterns in residuals may indicate various problems.

Finally, succinct summary figures of performance metrics are needed to compare different harvest control rules, especially for managers. Risk tables are commonly used for their ease of comparison, and are organized in a large grid by harvest control rule and the performance metric within different time periods, e.g. short-term versus long-term (Figure 5.3). For each harvest control rule, a grid of performance metrics, e.g. yield or interannual catch variability, organized by F_{target} and $B_{trigger}$ is presented. Combinations of F_{target} and $B_{trigger}$ which are not sufficiently precautionary can be highlighted and removed from consideration. From those remaining, the combinations most favourable for the respective performance metric can be highlighted.

Risk tables are often only shown for the base operating model scenario. While alternative robustness scenarios should also be displayed, presentation can become lengthy. Interactive online apps such as Shiny are convenient for readers to toggle risk tables among different operating model scenarios.

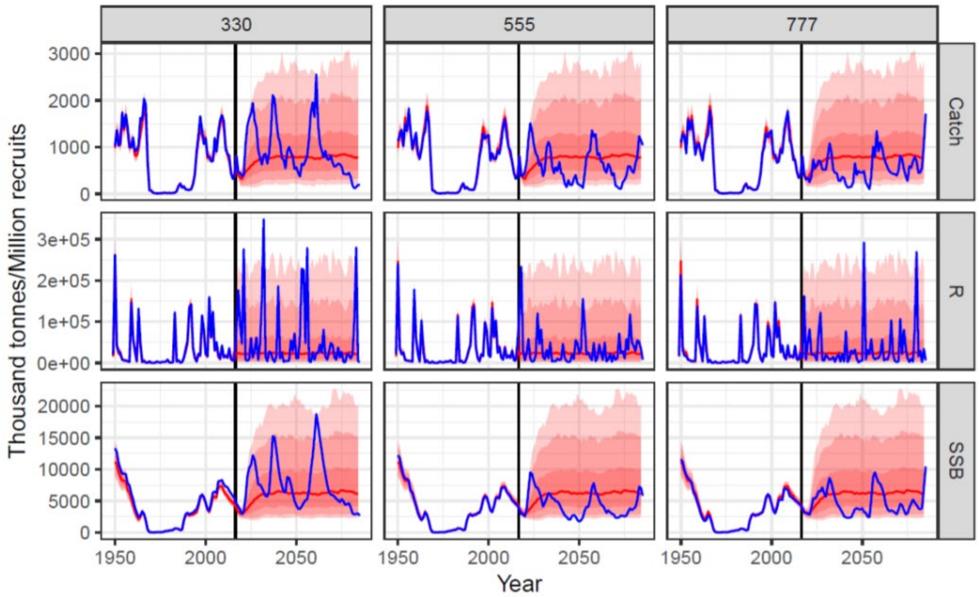

Figure 5.2. The realized catch, recruitment, and spawning biomass (rows) from three simulations (columns). Vertical, black line separates the historical assessment and the future projection. Individual panels also show the 5th, 25th, 50th, 75th, and 95th percentiles of their distribution. Figure was taken from the WKNSSHMSE summary advice sheet.

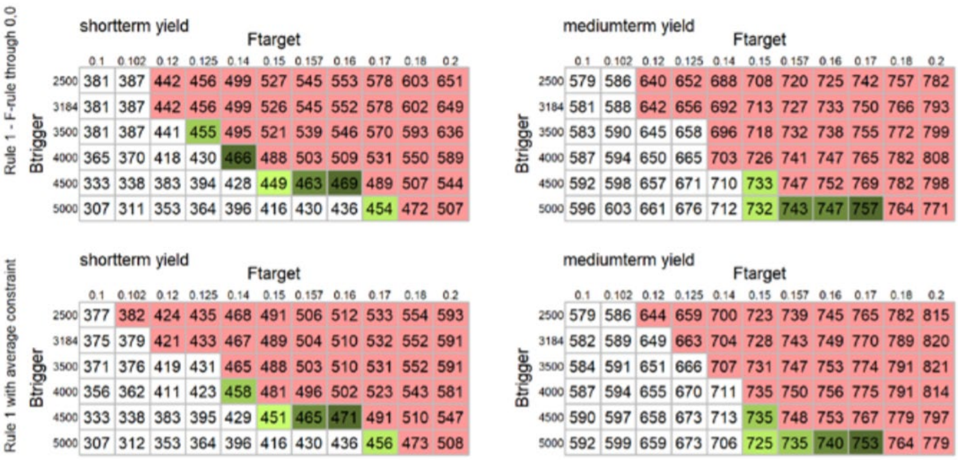

Figure 5.3. Short-term and medium-term median yield (columns) for two harvest control rules (rows). Each cell contains a grid of yield values organized by $B_{trigger}$ and F_{target} tested in the MSE. Red indicates combinations of $B_{trigger}$ and F_{target} which are not precautionary, and green indicates ≥95% of the maximum yield from remaining combinations. Figure was taken from the WKNSSHMSE summary advice sheet.

6 Condensed MSE guidelines and reporting requirements for studies done for ICES

6.1 Condensed MSE guidelines

The following table presents the ICES technical guidelines for MSE in a condensed format.

The guidelines are intended to guide the decisions based on best practice throughout the evaluation. Scientists involved in ICES MSE analyses are expected to consult these guidelines. They are also expected fill in a table, with exactly the same format as the table here, where each of the non-shaded boxes on the right column of the table is filled by explaining how the corresponding issue was addressed. Following or deviating from the guidelines should be appropriately motivated.

This table, together with the twin table filled by the scientists conducting the MSE, provides basic documentation and checkpoints that reviewers may consider, to ensure that the simulations cover a realistic range of future developments.

Condensed MSE guidelines	
Operating Model	
Biology and Fishery Model (Base Case)	
Basis for the Base Case	The Base Case corresponds to the representation of reality considered "most plausible" among the set of models considered in the study. For ICES stocks that have a stock assessment agreed in a benchmark, the default would be the stock assessment agreed at the benchmark. All settings for future years (recruitment, growth, M, maturity, fishery selectivity) should be congruent with the historical past, but reflect what is considered to be more likely for the upcoming period of application of the management strategy. No obvious disconnect should occur between recent past and near future. When variability in the above-mentioned variables has been observed in the past, it is desirable to account for it in the upcoming period. This refers not only to variance, but also to autocorrelation or time trends. However, recruitment is the variable that definitely needs to be modelled as stochastically varying from year to year.
Recruitment	Should be modelled as stochastic, varying from year to year
Growth	As described above
Natural mortality	As described above
Maturity	As described above
Fishery selectivity	As described above
Initial stock numbers	From the stock assessment agreed by ICES for the stock, including a plausible range of uncertainty. The variance-covariance matrix from the assessment should not be used blindly. Care should be taken not to over-amplify uncertainty at young ages, which can occur in some cases if uncertainty is taken directly from the final year in the assessment.

Technical interactions (mixed fisheries)	Where technical interactions are expected to be an important driver of stock and fleet dynamics, they should be included in the MSE, either through a full multi-stock mixed fisheries operating model or by approximating the processes within a single species operating model as appropriate.
Biological interactions	Where biological interactions are believed to be significant drivers of stock development, they should be included in the MSE, either through a full multi-species operating model or by approximating the processes within a single species operating model as appropriate.
Biology and Fishery Model (alternative dynamics)	
Alternative biology and fishery scenarios	Alternative biology and fishery dynamics, covering a range of plausible realities (taking the main sources of uncertainty into account), should be included in the evaluation.
Observation Model	
Simulation of input data for a stock assessment or for direct use in a harvest rule (e.g. for survey-based harvest rule)	All data generated in the MSE should be sufficiently noisy (including patterns and correlations) to provide realism for the evaluation. The data to be simulated should be fully specified and correspond to the data that will be used when the management strategy is applied in future years. For shortcut approaches to MSE (see under Estimation Model below) no data are simulated. The need for alternative observation scenarios in addition to the Base Case should be considered.
Implementation Model	
Implementation error	Should be included when relevant, such as when historical comparison indicates systematic deviations between management measure (TAC) and actual catch, provided this phenomenon is expected to continue in the future. An MSE scenario without implementation error should also always be run. The Base Case (i.e. with or without implementation error) would need to be decided in each situation as considered appropriate. The interpretation of performance statistics from the MSE can become difficult under implementation error.
Management Procedure	
Estimation Model	
If a full assessment is conducted in the MSE loop	The assessment model should be comparable to the one that will be used when the management strategy is applied in future years.
If a shortcut approach (instead of a full assessment) is used in the MSE loop	The properties of the resulting “shortcut assessment” should be comparable to those resulting from the routine assessment (the one that will be used when the management strategy is applied in future years) and, where relevant, short-term forecast, including retrospective errors and relevant autocorrelations. Where possible, this should be pre-tested.
Harvest rules requiring a stock assessment followed by a short-term forecast	The short-term forecast settings used in the MSE should be implemented in any future application of the harvest rule.
Decision Model (Harvest rule)	
Harvest rule design	Discontinuities or steep slopes in harvest rules should be avoided (clients should be made aware of this)

Harvest rules that include stabilizers	The performance of the management strategy without catch stabilizers should be examined. The effect of stabilizers should be carefully examined (e.g. if stabilizers are included or removed instantly, depending on SSB being above or below some $B_{trigger}$)
Duration of decisions	The MSE should use the duration of decisions specified in the management strategy (e.g. TAC annually or every 2 years)
Conditions for reevaluating the MSE in the future	Three main situations may be identified: • If there is a revision clause included in the management strategy • If the performance of the stock assessment used to apply the harvest rule (for model-based harvest rules) or the quality of the data used in the harvest rule (for empirical harvest rules) deteriorates substantially relative to what was assumed in the MSE • If the observed conditions of the stock and/or fishery depart considerably from what was assumed in the MSE. The range of plausible future outcomes under satisfactory performance of the management strategy should be documented and it should be possible to control how the stock and the decisions develop compared to the assumed range. Examples include the distribution and time course of recruitment, growth, maturity and selection, as well as adherence to the harvest rule.
Running the MSE simulation	
Number of iterations (independent replicates simulated in the MSE)	Should be sufficient to calculate performance statistics to a "suitable" degree of precision, in particular the ICES precautionary criterion *Prob3*. As initial guidance, use around 1000 iterations (see Section 4.3, of WKGMSE 2 report).
Projection time (number of future years included in the MSE)	As indicated by clients
Reporting outputs	It is preferable that graphs present percentiles of future trajectories (5%, 50%, 95%), which are easier to interpret than box-plots.
Validation checks (for different components of MSE simulation)	Validation should provide confirmation that the assumptions made in the MSE are plausible, realistic and consistent with available data and knowledge. Validation of the operating model should demonstrate that simulated and historical recruitment are comparable, and that there are no unexpected discontinuities between past and future dynamics. The behaviour of the operating model could also be checked against expectations by running it forward under fixed fishing mortality. If a full MSE is applied, validation of the observation model should demonstrate that that future noise is consistent with historically-observed noise. If a shortcut approach to MSE is applied, therefore combining the observation and estimation models in order to approximate assessment model behaviour, validation should demonstrate that the approximation used is adequate. The decision model should be programmable (i.e. it should be possible to convert requests into computer code), and should be checked with a perfect knowledge scenario (for code-checking and to check the decision model is plausible). The implementation model should be validated based on historical observations (e.g. validating fleet responses to management decisions).
Reference points	

Reference points used in the MSE	Reference points consistent with the operating model should be used in the evaluations of performance measures related to operating model variables (such as the probability of SSB falling below B_{lim}).
Performance statistics and precautionary criterion	
Performance statistics	As indicated by clients. Short-term, medium-term, and long-term (where the meaning of the time intervals is dependent on species biology and other features, e.g. economic) performance statistics typically relate to the following: • Yield (median and other relevant measures, e.g. 10th percentile or probability of yield dropping below some threshold) • Probability of SSB falling below B_{lim} • SSB (median and other relevant measures) • Indicator for year to year variability in SSB and yield • Realised ("real") F • If relevant (case-specific), probability of "perceived" (i.e. estimated) SSB dropping below MSY $B_{trigger}$ • "Perceived" (i.e. estimated) F relative to F_{MSY} • Time to recover from below B_{lim}
Risk type****	*Prob3*, and possibly others if found relevant (case-specific) or requested by clients, should be computed
Precautionary criterion	*Prob3* $\leq$ 5% over all years included in the management strategy (short and long terms) is the ICES criterion for considering a management strategy as precautionary (while including special considerations made for recovery plans or initial recovery phases within long-term management strategies, and for short-lived stocks that have more than 5% probability of falling below B_{lim} under natural conditions of no fishing).
Experiences and comments	
Use of ICES guidelines for MSE (WKGMSE2 2019)	The guidelines are intended to guide the decisions based on best practice throughout the evaluation. Following or deviating from the guidelines should be appropriately motivated.

****** Risk types (for a period of ny years):**

***Prob1* = average probability that SSB is below B_{lim}, where the average is taken across the ny years.**

***Prob2* = probability that SSB is below B_{lim} at least once during the ny years.**

***Prob3* = maximum probability that SSB is below B_{lim}, where the maximum of the annual probabilities is taken over the ny years.**

6.2 Reporting requirements

A number of specific outputs have been identified as required in the MSE reports to ICES.

A table, with exactly the same format as the table displayed in Section 6.1 of this report, should be presented in the MSE report, where each non-shaded box on the right column of the table should be filled explaining how the issue was addressed. Following or deviating from the guidelines should be appropriately motivated.

In addition to the condensed reporting format in the table, further details should be presented in the full MSE report.

The report should provide the technical details of the assumptions made for the MSE, in a clear and structured way, including the parameter values used in various parts of the MSE and a clear description of the range of scenarios tested.

Validation checks are very important to increase confidence in the suitability and plausibility of the assumptions made in the MSE. Ways of conducting validation checks were discussed in Sections 4.7 and 5.3 of this report and the outcomes of such checks should be included in the MSE report (mostly in graphical form).

Graphs showing percentiles (5%, 50%, and 95%) are much easier to interpret than box-plots and are, therefore, preferred for the MSE report.

Both tables and graphs should be displayed giving the ICES-agreed risk measure *Prob3*, and possibly others if found relevant (case-specific) or requested by clients.

Additional ideas for communication of MSE results have been discussed in Section 5.3 of this report, which should be consulted by the scientists doing the MSE work.

7 Software

7.1 General Comments

A number of software packages have been developed in recent years for the purpose of conducting management strategy evaluations (see Section 7.3). It is likely that one or more of these packages can either be used directly or (more likely) modified for the MSE in question. Given the general recommendation that MSE evaluation is not limited to a single approach, reuse of existing tools can result in significant timesavings and simplify validation, which can then be limited to new modules or procedures. When selecting from the available software packages consideration should be given to

- The underlying capabilities of the software in terms of the biological and fishery, observation, estimation, decision and implementation models already available and tested (see the accompanying documentation);
- Is the software readily modifiable for your needs?
- The programming language and operating system, and your experience with these;
- The availability of support. This can be available from the original author(s) and/or other users of the software;
- Are there any hardware/licensing issues?

7.2 Software Development and Quality Standards

In the event that a new application is to be built, or extensions and changes on an existing platform are necessary, a series of guidelines to ensure transparency, replicability and peer review, should be followed. A brief checklist is provided here, as a starting point, but a good number of complete guidelines are available for scientific software and some are linked below.

- Carry out the software development using a version control system, such as Github or Gitlab. ICES is using Github as its development platform and can provide developers with access to private repositories for ICES-related work.
- Use of a version control system will also simplify software releases. Make use of the existing mechanisms (e.g. git tags) to keep track of software versions made public, and what source code they refer to.
- Document as much as possible the code written for a particular analysis. For example, if using the R language, documentation for any function can be prepared using the R documentation standards.
- Write tests for elements of the analysis, like new functions, to check if they return the expected result, and that they fail when called with the wrong inputs.
- Assemble a replicable procedure, for example a script that checks software versions, and sets random number generation seeds, so that subsequent runs of the analysis return the same results.

Some recent references with a particular focus on scientific software development, are Wilson *et al.* (2014, 2017).

7.3 Available Software

The following is a list prepared by WKGMSE 2 based on the software that is most familiar to participants, but should not be taken to constitute a complete list of software available for MSE.

7.3.1 FLR

The FLR project has been developing over the last few years a series of packages in the R statistical language with the first objective of providing the necessary tools for the implementation of full-feedback MSE analysis of fishery systems (Kell *et al.* 2007). The packages closely follow R conventions in syntax and procedures, but extend the language to accommodate the data types and methods commonly used in fisheries science.

The development of FLR has followed from its start an open source model, in which the whole source code of the packages is freely available, discussions are carried out in an open mailing list, and users are encouraged to participate as much as possible in the development.

The current set of FLR packages includes all the basic elements necessary to assemble an MSE simulation for a single age-structured stock, including multiple fleets, spatial complexity, time steps of any length. Multi-species considerations can be currently incorporated at the technical level, by creating fleets that operate over multiple stocks, but no specific dynamics have been coded linking them at the biological level, such as predator-prey dynamics, or synchronized recruitment.

A key element in the FLR approach has been the development of a series of data structures, classes in R's S4 Object-Oriented Programming (OOP) system, that encapsulate the different elements in the fishery system under evaluation. A series of methods, in the OOP sense of functions that operate on individual classes, are then available to carry out a large range of operations, including manipulation, mathematical calculations, statistical summaries and estimates, plotting, etc. The OOP approach ensures data integrity by specifying a strict set of validity checks for each class. Code can thus be developed that carries out with confidence a large number of operations on various data elements.

A growing variety of stock assessment methods are available for incorporation in the management procedure section. From biomass dynamics models using a Pella-Tomlison formulation, to VPA-based methods, such as Separable VPA and XSA, and statistical catch and age methods, like FLa4a (http://flr-project.org/FLa4a) and FLSAM (http://flr-project.org/FLSAM). Tools exist for interfacing with existing stock assessment models coded in either C, C++, Fortran or ADMB.

The projection capabilities of FLR, currently being extended by the FLasher package, use Automatic Differentation to solve the population equation for a wide range of targets (SSB, catch, fishing mortality, effort, revenue). This allows the implementation of a large variety of harvest control rules in an efficient way.

A new package, mse (http://flr-project.org/mse), provides a unified but flexible infrastructure to run MSE evaluations.

The programming approach of the FLR system gives huge flexibility to the user, at the obvious cost of extra complexity and a steeper learning curve. Models and simulations of very different levels of complexity can be implemented in FLR, and extra elements can be added on to a common code base, with relatively little cost.

Recent examples of use of FLR in MSE analyses include Azevedo *et al.* (2017), Mosqueira (2018), and STECF (2015, 2016, 2018).

The FLR packages are under active development, with continuous improvement to the existing code, and a number of useful extensions being tested and released. Stable versions have been released sporadically, but the FLR project has now setup a system for automated testing and building of R packages that will allow continuous release of development versions of all packages, and two or three stable releases a year, following R's own development cycle.

7.3.2 HCS

HCS is a harvest rule simulation program of the 'short cut' type. The operating model is single species, age disaggregated with annual time steps. It has several options for obtaining initial numbers, including priming the stock with a fixed fishing mortality and random recruitments, weights and maturities. It has a wide range of options for recruitment variation, including periodic fluctuations, time trends, spasmodic recruitments and regime shifts. Growth and maturity can be density dependent. Natural mortality is fixed. The observation model generates 'assessed' stock numbers at age, backwards in time if needed, with algorithms intended to reproduce the influence of noise in input data to an assessment. The decision model imitates the normal process with projection through an intermediate year, and has a variety of options for decision basis and decision rules. The implementation model adds noise to catch numbers, thus altering the realized selection at age.

HCS is constructed to scan over numerous options for decision rules and for noise in the observation model. Each run with 1000 iterations for one set of options takes 10-30 seconds on a modern computer. The output is both detailed tables of annual means and fractiles of interest parameters for each option, and collecting tables giving the main interest parameters (Catch, SSB, TSB, Inter-annual variation of catches and risks) averaged over time periods. Risk is now being changed to *Prob3* (see Section 4.2 of this report), previously it was *Prob2*. A yield per recruit calculation, including stock-recruitment is also provided. Hence, it is specifically made to assist in the development phase of harvest rules, although it also is used for final evaluations, in particular in cases where including an assessment in the loop is out of reach.

HCS is distributed as open source software. It is still evolving, both in terms of improved algorithms and in terms of new harvest rules. Updated versions of HCS with manual, executable program, program code (Fortran77) and examples of input files can be downloaded from www.dwsk.net.

7.3.3 FPRESS

FPRESS (Fisheries Projections and Evaluation by Stochastic Simulation) is written and run in R and is designed to be easy to edit by end users to suit their requirements. The model is designed as a stochastic simulation tool for evaluating fisheries management strategies and developing management advice and was used in the evaluation of the Western Horse Mackerel and NEA Mackerel management plans.

FPRESS is as a population projection model with the following characteristics and limitations:

- Stochastic
- Single species
- Non-spatial
- Age-structured population
- Exponential mortality
- F or TAC controlled fishery
- Various recruitment models, and
- Various harvest control strategies

The coding structure used for FPRESS (open source, modular programming) means that the model can be readily adapted to incorporate specific recruitment models or harvest control rules.

The FPRESS operating model uses the standard single-species age-structured population with an exponential mortality model. It does not include any spatial elements or allow for mixed species interactions. Noise and bias can be added to the population vectors (initial numbers, weights, maturities, fishing and natural mortalities). These stochastic elements are implemented as multipliers for bias and random draws from an appropriate distribution for noise. Implementation errors are incorporated in a similar fashion via a CV andbias on F or TAC.

In addition to the operating model, FPRESS includes an observation (assessment) model where the stock assessment process can be simulated and a management and decision-making model will apply the prescribed harvest control rule. Both of these model elements can include stochastic behaviour via a prescribed noise and bias. In this way, it is possible to parameterize the effects of uncertainty in the stock assessment process and phenomena such as TAC non-compliance and data errors. The model (deliberately) avoids a complex "assessment feedback" model so that all bias and noise introduced in the assessment process can be qualitatively controlled.

FPRESS inputs are the stock and fishery parameter data with appropriate CV values. These values are often derived from recent stock assessments and studies of parameter accuracy. The model output is configurable and is saved as FLR FLQuant objects. In this way, the functionality offered by the FLR library can be used to explore the model output. Included in the F-PRESS model are a number of functions for graphing and analysing model output.

FPRESS can be configured to run on parallel processors and is a useful simulation tool for exploring multiple combinations of parameters within HCRs. Input options are specified in xml files and a full A full simulation audit trail is saved in a log file which includes the version number of each source code file, all simulation options (as specified in the simulation options file) and run statistics (start and finish times and any debug information written to the console) are recorded in a log file.

7.3.4 FLBEIA

FLBEIA is a generic tool to conduct Bio-Economic Impact Assessment of fisheries management strategies in a management strategy evaluation framework (Garcia *et al.,* 2017). FLBEIA can be categorized as a 'Models of Intermediate Complexity for Ecosystem assessments' (MICE, (Plagányi *et al.,* 2014)) which is focused on the fishing activity in a mutistock and multifleet context.

FLBEIA has been built using R- FLR packages (Kell *et al.,* 2007) and can be automatically beneficiated from the new developments in those packages. As any MSE algorithms it is formed by two main blocks the Operating Model (OM) and the Management Procedure (MP) (Figure 7.1). The OM has three components that interact among themselves, the stocks, the fleets and the covariates. In turn, the MP is divided in other three components, the observed data, the perceived stocks and the management advice.

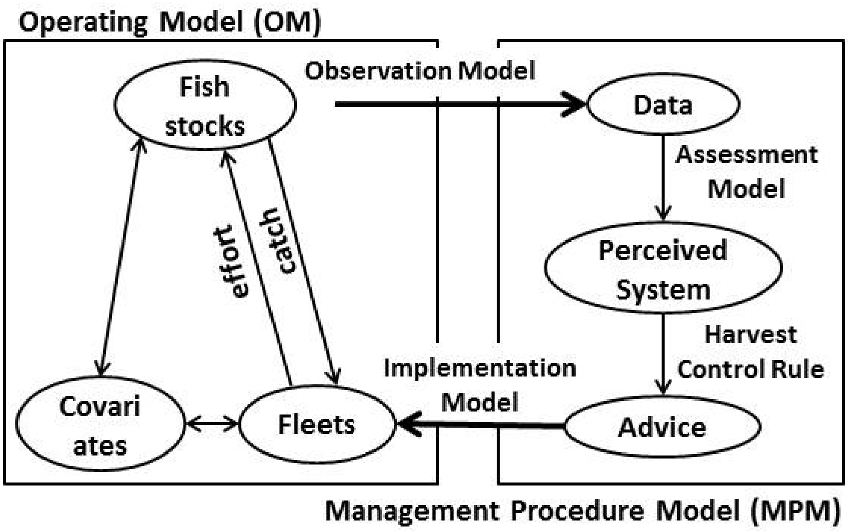

Figure 7. 1. Conceptual diagram of FLBEIA (taken from Garcia *et al.* (2017))

The stocks can be age or biomass structured. Trophic interactions have never been modelled in FLBEIA but it could be done. There is also a development version where Gadget (Begley and Howell, 2004) can be used as operating model. The activity of the fleet is divided in metiers and four processes are modelled. The short-term dynamics (total effort and its distribution along metiers), long term dynamics (entry-exit of new vessels in the fishery), price formation and catch production. The covariates can be used to store any variable not included in the stocks and fleet components.

The link between the OM and the MP is done through the observation model that generate de observed data. Two types of data can be generated, the stocks and the abundance indices. Any observable variable can be subject to observation error and the error is divided in two components, the aging error component and a multiplicative error. As the errors are introduced as input data they can be conditioned using any distribution, bootstrap or other analysis. The perceived population is generated using an assessment model. There is the possibility of using the 'short-cut' approach or any assessment model available in R/FLR. What is needed is a wrapper that generates the input and output of the model in the right shape. Wrappers are already available for SPiCT, XSA, SCA in Fla4a and FLSAM. The management advice is generated using a harvest control rule. Two types are available, model-free HCRs and model-based ones. The model-free HCRs use the abundance indices generated by the observation model and do not require to apply any assessment model. In turn, the model-based HCRs use the output of the short-cut approach of an assessment model to generate the advice.

The adaptative management advice based on catch can be accompanied by technical measures like changes in selectivity, implicitly simulated spatio-temporal closures or effort restrictions for example.

The stochasticity is introduced using montecarlo approximation and the iterations run in parallel. The results can be analysed and presented using the Shiny application available in the FLBEIAShiny package (https://github.com/flr/FLBEIAshiny).

The model is constructed in a modular way. The fishery system is discomposed in processes (recruitment, catch production, population growth) and several models are provided to simulate each of them. Alternatively, new models can be coded and call from the function with no extra coding.

The model documentation is extensive. There is a research paper describing the model (Garcia et al., 2017). A manual, which describes in detail all the models available is provided within the R library. And there is a set of dedicated tutorials in the FLR website http://www.flr-project.org/. The source code can be downloaded from github (https://github.com/flr/FLBEIA) and the compiled package from the FLR website (http://www.flr-project.org/). There is a support mailing list flbeia@azti.es.

7.3.5 Impact Assessment Model for fisheries (IAM)

The program IAM has been developed to carry out bio-economic integrated stochastic simulations of management decision rules. The program couples the biological dynamics of fish stocks with the economic dynamics. It is described in detail in Merzéréaud *et al.* (2011). It can be used to carry out impact assessment for management strategies and provide results on transition phases and cost benefit analysis. The fish population model is age structured, has yearly time steps and is spatially aggregated. The fishery model is multi species, multi fleet and multi-*métier*. The program has a modular structure to allow flexibility in the development as shown on Figure 7.2.

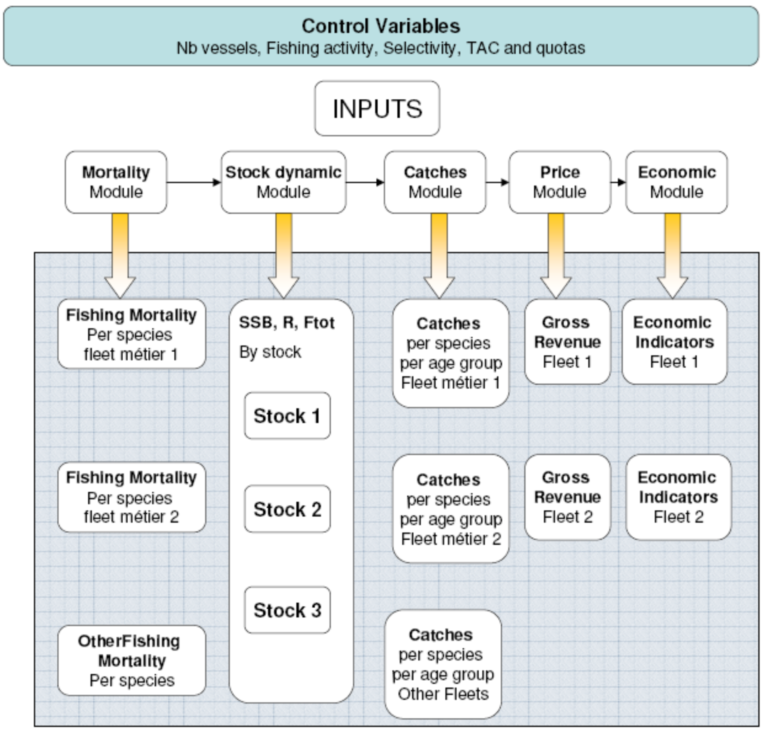

Figure 7.2. Simplified representation of the Impact Assessment bio-economic model for fisheries

The main characteristics of the model can be summarised as follows:

- Age structured, yearly time steps, spatially aggregated.
- Multi species, multi fleet and multi-*métier*
- Stochasticity (using bootstrapping).
- A mortality module splits fishing mortality between fleets according to *métier* by fleet based on landings proportion.
- Several kinds of market assumptions are possible:
  - constant price assumptions
  - price-quantities relationship
  - price-importations/exportations relationship

Economic dynamics such as fleet dynamics, catchability increase through investment or technical creeping, or short terms behaviours can been included.

Several assumptions concerning impacts of scenarios on gross revenue are possible including reallocation of effort assumptions.

It has a wide range of options for harvest rules including options for:

- Selection pattern
- Fishing activity (i.e. fishing time, number of operations)
- Number of vessels
- TACs

The results are presented in terms of several statistics:

- status of stocks (biomass, spawning biomass, fishing mortality, total catch)
- fleet performance (Total Gross Return, Total Gross Cash Flow of the fleet)
- individual performance by fleet (Mean Gross Return, mean Gross Cash Flow)
- total vessel number by fleet
- employment in the fishery
- crew salaries
- producer, consumer and state surplus variation ie rent (net present value)

The program can also be run for optimization (ex: rent maximization)

The program has been developed in R/C++ to allow easy handling, flexibility and performance. The core of the program has thus been coded in C++ and the interface uses R for data handling, for outputs and to produce graphs.

Parameterization is easy as the model can make direct use of outputs from assessment working groups (inputs for short-term prediction) and a limited number of indicators calculated from DCF data.

Recent examples of use of IAM in MSE analyses include Gourguet *et al.* (2013), Guillen *et al.* (2014), STECF (2015), Doyen *et al.* (2017), and Bellanger *et al.* (2018).

7.3.6 DLMtool/MSEtool

The DLMtool and MSEtool packages are written in R and designed for rapid testing of management procedures in full-feedback simulation. The packages are distributed through the CRAN repository, whose maintainers control the quality standards of widely-distributed R packages.

The operating model, termed "OMx" and contained in DLMtool, is highly-parametrized and flexible age-structured model designed to conveniently accommodate alternative configurations (Carruthers and Hordyk, 2018). For example, growth parameters such as the von Bertalanffy asymptotic length, like most parameters in the operating model, are uniformly distributed among simulation replicates by default. Alternative values, such as normally distributed values obtained from MCMC output of an assessment model, can be passed by the analyst into the operating model. More generally, alternative growth functions and time-varying processes for growth are accommodated in the operating model if the analyst provides an array that specifies the length by age, year, and simulation replicate. OMx is also spatially explicit, and models: (1) movement of stock biomass, and (2) the opening and closing of spatial areas to fishing. In the case of fishery spatial closures, it can be specified whether the fishing effort is reduced or re-distributed to the remaining open areas.

The different components of the MSE simulation are organized by R classes. Data and life history parameters for the management procedures are generated from the operating model and are placed in an object of class "Data". Biased data and parameters can be generated for the Data object in order to evaluate the effects of the bias on the performance of management procedures. Management procedures are functions of class "MP" that describe the algorithm and returns the management advice, i.e. TAC, effort limits, spatial closures, or size regulations, in an object of class "Rec". The operating model updates the stock based on the management action prescribed in the "Rec" object. When the simulation proceeds forward in time, the "Data" object is updated with new years of data.

DLMtool distributes many data-limited management procedures that have been proposed and evaluated, e.g. Geromont and Butterworth (2014), Carruthers *et al.* (2016), Jardim *et al.* (2015).

The simulation output is returned in an object of class "MSE."A suite of performance metrics are provided to return summary statistics. The standardized "MSE" class output can be used to write additional custom performance metrics.

The MSEtool package provides a framework for testing data-rich management procedures that combine assessment models and harvest control rules. MSEtool uses the OMx operating model in DLMtool as part of the full-feedback simulation loop. A suite of assessment models are distributed in the MSEtool package, such as the delay-difference model, surplus production model, and statistical catch-at-age model. State-space implementations of these models use Template Model Builder (TMB) for rapid estimation. Management procedures are built modularly by combining an assessment model (class "Assess") and harvest control rule (class "HCR") in a single function (of class "MP").

MSEtool also provides several diagnostic procedures, such as the convergence rate, for evaluating performance of the assessment models in the closed-feedback simulation. Standardized reporting of assessment output is available for each model fit for examining convergence issues. Class "Assess" functions return an object of class "Assessment" which contain estimates of biomass, fishing mortality, and reference points. For each assessment model, the "plot" function generates plots of these quantities in a HTML document. These plots can be useful for diagnostics for identifying goodness of fit and causes of model non-convergence. Finally, retrospective analysis of the assessment model over time in individual simulation runs can be used to characterize stability of estimates in historical biomass, fishing mortality, and reference points over subsequent assessments.

DLMtool has been used to identify and test data-limited management procedures for U.S. federally-managed stocks in the U.S. Atlantic (Newman *et al.*, 2014, McNamee *et al.*, 2016, Wiedenmann, 2016, SEDAR 2016a, SEDAR 2016b), California stocks (Hordyk *et al.*, 2017), and ICES Category 3-4 stocks (ICES 2018d).

Both DLMtool and MSEtool are currently actively developed, with extensive user manuals accompanying the software. Developmental versions are hosted on Github at https://www.github.com/DLMtool. Currently, users can write custom wrapper functions for their assessment model and harvest control rule of choice to be used in MSEtool. Future versions will distribute wrapper functions for more assessment models such as spict, SAM, and XSA/VPA methods.

8 References

Augusiak, J., Van den Brink, P. J., and Grimm, V. 2014. Merging validation and evaluation of ecological models to 'evaludation': A review of terminology and a practical approach. Ecological Modelling, 280: 117-128.

Azevedo, M., Mendes, H., Costas, G., Jardim, E., Mosqueira, I. and Scott, F. 2017. Long-Term Management Strategy for Southern Horse Mackerel (hom27.9a) – Management Strategy Evaluation. Relat.Cient.Téc. do IPMA no 19. 23p + Anexos.

Balci, O. 1997. Verification, validation and accreditation of simulation models. In Winter Simulation Conference (Vol. 1997, pp. 135-141).

Begley, J., and Howell, D. 2004. An overview of GADGET, the Globally applicable Area Disaggregated General Ecosystem Toolbox ICES CM, 2004/FF,13,16.

Bellanger M., C. Macher, M. Merzéréaud, O. Guyader and C. Le Grand. 2018. Investigating trade-offs in alternative catch share systems: an individual-based bio-economic model applied to the Bay of Biscay sole fishery. Canadian Journal of Fisheries and Aquatic Sciences 75 (10) 1663-1679. dx.doi.org/10.1139/cjfas-2017-0075

Butterworth, D.S., Bentley, N., De Oliveira, J.A., Donovan, G.P., Kell, L.T., Parma, A.M., Punt, A.E., Sainsbury, K.J., Smith, A.D. and Stokes, T.K., 2010. Purported flaws in management strategy evaluation: basic problems or misinterpretations? ICES Journal of Marine Science, 67(3), pp.567-574.

Carruthers, T.R. and Hordyk, A.R. 2018. The Data-Limited Methods Toolkit (DLMtool): An R package for informing management of data-limited populations. Methods in Ecology and Evolution, 1-8.

Carruthers, T. R., Kell, L. T., Butterworth, D. D. S., Maunder, M. N., Geromont, H. F., Walters, C., McAllister, M. K., Hillary, R., Levontin, P., Kitakado, T., and C. R. Davies. 2016. Performance review of simple management procedures. ICES Journal of Marine Science, 73: 464–482.

Doyen, L., Bene, C., Bertignac, M., Blanchard, F., Cisse, A., Dichmont, C., Gourguet, S., Guyader, O., Hardy, P.-Y., Jennings, S., Little, L. R., Macher, C., Mills, D.J., Noussair, A., Pascoe, S., Pereau, J.-C., Sanz, N., Schwarz, A.-M., Smith, T. and O. Thebaud. 2017. Ecoviability for ecosystem-based fisheries management. Fish And Fisheries, 18(6), 1056-1072 . http://doi.org/10.1111/faf.12224.

Garcia, D., Sánchez, S., Prellezo, R., Urtizberea, A., and Andrés, M. 2017. FLBEIA : A simulation model to conduct Bio-Economic evaluation of fisheries management strategies. SoftwareX, 6: 141-147.

Geromont, H. F. and Butterworth, D. S. 2014. Complex assessments or simple management procedures for efficient fisheries management: a comparative study. ICES Journal of Marine Science, 72: 262–274.

Gourguet, S., C. Macher, L. Doyen, O. Thébaud, M. Bertignac and O. Guyader. 2013. Managing mixed fisheries for bio-economic viability. Fisheries research. Volume 140, February 2013, Pages 46–62.

Guillen, J., Macher, C.,Merzéréaud, M, Guyader, O. and J. Boncoeur. 2014. Effects of the Share Remuneration System on Fisheries Management Targets and Rent Distribution. Marine Resource Economics. Volume 30, number 2.

Hill, S.L., Watters, G.M., Punt, A.E., McAllister, M.K., Quéré, C.L. and J. Turner. 2007 Model uncertainty in the ecosystem approach to fisheries, Fish Fish., 8: 315-336.

Hordyk, A., Newman, D., Carruthers, T., and L. Suatoni. 2017. Applying Management Strategy Evaluation to California Fisheries: Case Studies and Recommendations. Accessed at: https://www.datalimitedtoolkit.org/wp-content/uploads/2017/07/Applying-MSE-to-CA-Fisheries-Case-Studies-Recommendations.pdf (19 February 2019)

ICES. 2004. Report of the Working Group on Methods of Fish Stock Assessments (WGMG), 11-18 February 2004, Lisbon, Portugal. ICES CM 2004/D:03.

ICES. 2008. Report of the Working Group on Methods of Fish Stock Assessments (WGMG), 7-16 October 2008, Woods Hole, USA. ICES CM 2008/RMC:03.

ICES.2011. Report of the Working Group on Multispecies Assessment Methods (WGSAM), 10–14 October 2011, Woods Hole, USA. ICES CM 2011/SSGSUE:10. 229 pp.

ICES. 2012. Report of the Workshop for Revision for the North Sea Herring Long Term Management Plan, 1–2 October 2012, Copenhagen, Denmark. ICES ACOM:72. 110 pp.

ICES. 2013a. Report of the Study Group on Management Strategies (SGMAS), 17 - 21 November 2008, Lisbon, Portugal. ICES CM 2008/ACOM:24. 74 pp.

ICES. 2013b. Report of the Workshop on Guidelines for Management Strategy Evaluations (WKGMSE), 21–23 January 2013, ICES HQ, Copenhagen, Denmark. ICES CM 2013/ACOM:39. 128 pp.

ICES. 2014a. Report of the Workshop to consider reference points for all stocks (WKMSYREF2), 8-10 January 2014, ICES Headquarters, Copenhagen, Denmark. ICES CM 2014/ACOM:47. 91 pp.

ICES. 2014b. Interim Report of the Working Group on Multispecies Assessment Methods (WGSAM), 20–24 October 2014, London, UK. ICES CM 2014/SSGSUE:11. 104 pp.

ICES. 2014c. Report of the EU Workshop on the NEA Mackerel Long-term Management Plan (WKMACLTMP), 24–27 June and 17–19 November 2014, Copenhagen, Denmark. ICES CM 2014/ACOM:63. 120 pp.

ICES. 2015. Report of the Joint ICES–MYFISH Workshop to consider the basis for F_{MSY} ranges for all stocks (WKMSYREF3), 17–21 November 2014, Charlottenlund, Denmark. ICES CM 2014/ACOM:64. 156 pp.

ICES. 2016a. Report of the Workshop to consider F_{MSY} ranges for stocks in ICES categories 1 and 2 in Western Waters (WKMSYREF4), 13–16 October 2015, Brest, France. ICES CM 2015/ACOM:58. 187 pp.

ICES. 2016b. Report of the Workshop on Blue Whiting (*Micromesistius poutassou*) Long Term Management Strategy Evaluation (WKBWMS), 30 August 2016, ICES HQ, Copenhagen, Denmark. ICES CM 2016/ACOM:53. 104 pp.

ICES. 2016c. ICES criteria for defining multi-annual plans as precautionary. ICES Advice Technical Guidelines, Section 12.4.10. Published 16 December 2016. ICES Advice 2016, Book 12: 3pp.

ICES. 2017a. ICES fisheries management reference points for category 1 and 2 stocks. ICES Advice Technical Guidelines, Section 12.4.3.1. Published January 2017. ICES Advice 2017, Book 12: 19pp.

ICES. 2017b. Report of the Workshop to review the ICES advisory framework for short lived species, including detailed exploration of the use of escapement strategies and forecast methods (WKMSYREF5), 11 -15 September 2017, Capo Granitola, Sicily. ICES CM 2017/ACOM:46 A. 63 pp.

ICES. 2017c. Report of the Workshop on management strategy evaluation for the mackerel in subareas 1–7 and 14, and in divisions 8.a–e and 9.a (Northeast Atlantic) (WKMACMSE), 28–29 August 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:48. 212 pp.

ICES. 2018a. Advice basis. In Report of the ICES Advisory Committee, 2018. ICES Advice 2018, Book 1, Section 1.2. https://doi.org/10.17895/ices.pub.4503.

ICES. 2018b. Report of the Workshop for management strategy evaluation for Norway Pout (WKNPOUT). ICES WKNPOUT Report 2018 26-28 February 2018. ICES HQ, Copenhagen, Denmark. ICES CM 2018 / ACOM:38 Ref: WGNSSK. 96 pp.

ICES. 2018c. Report on the Long-term Management Strategy Evaluation for Northern Shrimp (*Pandalus borealis*) in Division 4.a East and Subdivision 20 (PandLTMS), October–November 2017, Massimiliano Cardinale, Carmen Fernandez, Ole Ritzau Eigaard and Guldborg Søvik. ICES CM 2017/ACOM:52. 182 pp.

ICES. 2018d. Report of the Eighth Workshop on the Development of Quantitative Assessment Methodologies based on LIFE-history traits, exploitation characteristics, and other relevant parameters for data-limited stocks (WKLIFE VIII), 8–12 October 2018, Lisbon, Portugal. ICES CM 2018/ACOM:40. 172 pp.

ICES. 2019. Guidelines for ICES groups. Version 2019-1. Internal document available at ICES website https://www.ices.dk/explore-us/who-we-are/Pages/Expert-Groups.aspx.

Jardim, E., Azevedo, M., and Brites, N.M. 2015. Harvest control rules for data limited stocks using length-based reference points and survey biomass indices. Fisheries Research, 171: 12-19.

Kraak, S. B. M., Buisman, F. C., Dickey-Collas, M., Poos, J. J., Pastoors, M. A., Smit, J. G. P., Oostenbrugge, J. A. E. van, and Daan, N. 2008. The effect of management choices on the sustainability and economic performance of a mixed fishery: a simulation study. ICES Journal of Marine Science, 65, 697-712.

Kraak, S.B.M., Haase, S., Minto, C. and Santos, J. 2019, submitted to ICES Journal of Marine Science. The Rosa Lee Phenomenon and its consequences for fisheries advice on changes in fishing mortality or gear selectivity.

Kell, L. T., Mosqueira, I., Grosjean, P., Fromentin, J.-M., Garcia, D., Hillary, R., Jardim, E., et al. 2007. FLR: an open-source framework for the evaluation and development of management strategies. ICES J. Mar. Sci., 64: 640-646.

Kell, L.T., Pastoors, M.A., Scott, R.D., Smith, M.T., Van Beek, F.A., O'Brien, C.M. and Pilling, G.M., 2005. Evaluation of multiple management objectives for Northeast Atlantic flatfish stocks: sustainability vs. stability of yield. ICES Journal of Marine Science, 62(6), pp.1104-1117.

Lee, R. M. 1912. An investigation into the methods of growth determination in fishes. Conseil Permanent International pour l'Exploration de la Mer, Publications de Circonstance 63, 35 pp.

McAllister, M.K. and C.H. Kirchner, 2002: Accounting for structural uncertainty to facilitate precautionary fishery management: illustration with Namibian orange roughy. In: Targets, Thresholds, and the Burden of Proof in Fisheries Management, M. Mangel (ed.), Bulletin of Marine Science, 70, pp. 499–540.

McNamee, J., Fay, G., and S. Cadrin. 2016. Data Limited Techniques for Tier 4 Stocks: An alternative approach to setting harvest control rules using closed loop simulations for management strategy evaluation. SEDAR46-RD-08. SEDAR, North Charleston, SC. 58 pages. Accessed at: http://sedarweb.org/docs/wsupp/S46_RD_08_Black%20Sea%20Bass_Mid-AtlSSC.pdf (19 February 2019)

Merzéréaud, M., Macher, C., Bertignac, C., Frésard, M., Le Grand, C., Guyader, O., Daurès, F. and S. Fifas. 2011. Description of the Impact Assessment bio-economic Model for fisheries management (IAM), Amure Electronic Publications, Working Papers Series D-29-2011, 19 p. Available at: http://www.umramure.fr/electro_doc_amure/D_29_2011.pdf.

Michielsens, C.G.J. and M.K. McAllister. 2004. A Bayesian hierarchical analysis of stock-recruit data: quantifying structural and parameter uncertainties. Can. J. Fish. Aquat. Sci. 61: 1032-1047.

Mosqueira, I. 2018. Update on the development of MSE analysis of candidate Management Procedures for Indian Ocean albacore. IOTC-2018-SC21-17.

Newman, D., Carruthers, T., MacCall, A., Porch, C., and L. Suatoni. 2014. Improving the Science and Management of Data-Limited Fisheries: An Evaluation of Current Methods and Recommended Approaches. NRDC report R-14-09B. Accessed at: https://www.nrdc.org/sites/default/files/improving-data-limited-fisheries-report.pdf (19 February 2019).

Plagányi, É. E., Punt, A. E., Hillary, R., Morello, E. B., Thébaud, O., Hutton, T., Pillans, R. D., et al. 2014. Multispecies fisheries management and conservation: tactical applications using models of intermediate complexity. Fish and Fisheries, 15: 1-22.

Punt, A.E., A'mar, T., Bond, N.A., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A., Haltuch, M.A., Hollowed, A.B. and C. Szuwalski. 2014. Fisheries management under climate and environmental uncertainty: control rules and performance simulation, *ICES Journal of Marine Science*, Volume 71, Issue 8, 1 October 2014, Pages 2208–2220, https://doi.org/10.1093/icesjms/fst057

Punt, A.E., Butterworth, D.S., de Moor, C.L., De Oliveira, J.A.A. and M. Haddon. 2016. Management strategy evaluation: best practices. Fish and Fisheries 17: 303-334.

Rochet, M-J, and Rice, J. C. 2009. Simulation-based management strategy evaluation: ignorance disguised as mathematics? ICES Journal of Marine Science, 66: 754–762.

Rykiel, E. J. 1996. Testing ecological models: the meaning of validation. Ecological Modelling, 90: 229-244.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. 2008. Global Sensitivity Analysis: The Primer, Wiley.

SEDAR (Southeast Data, Assessment, and Review). 2016a. Stock Assessment Report: Caribbean Data-Limited Species. SEDAR, North Charleston, SC. 373 pages. Accessed at: http://sedarweb.org/docs/sar/S46_final_SAR.pdf (19 February 2019)

SEDAR (Southeast Data, Assessment, and Review). 2016b. Stock Assessment Report: Gulf of Mexico Data-Limited Species. SEDAR, North Charleston, SC. 618 pages. Accessed at: http://sedarweb.org/docs/sar/SEDAR_49_SAR_report.pdf (19 February 2019)

Simmonds, E.J., Hjorleifsson, E., Fernandez, C. and C.P. Millar. 2019. EQSIM: Estimation of equilibrium reference points for fisheries. R package version 0.1.18. http://github.com/ices-tools-prod/msy

STECF. 2015. Multiannual management plans SWW and NWW (STECF-15-08). Publications Office of the European Union, Luxembourg, EUR 27406 EN, JRC 96964, 82 pp.

STECF. 2016. Multiannual plan for demersal fisheries in the Western Mediterranean (STECF-16-21); Publications Office of the European Union, Luxembourg; EUR 27758 EN; doi:10.2788/103428

STECF. 2018. Data gaps and Biomass Escapement Strategy for Adriatic anchovy and sardine (STECF-18-01). Publications Office of the European Union, Luxembourg. ISBN 978-92-79-79383-7, doi:10.2760/800034.

tRFMO. 2018. Report from the joint tuna RFMO Management Strategy Evaluation Working Group Meeting in Seattle, USA (June 13-15, 2018). http://www.tuna-org.org/Documents/tRFMO_MSE_2018_TEXT_final.pdf

Wiedenmann, J., Michael J. Wilberg, M.J., Sylvia, A., Miller, T.J. 2015. Autocorrelated error in stock assessment estimates: Implications for management strategy evaluation. Fisheries Research, 172: 325–334

Wiedenmann, J. 2016. Application of Data-Poor Harvest Control Rules to Atlantic Mackerel. SEDAR46-RD-09. SEDAR, North Charleston, SC. 53 pages. Accessed at: http://sedarweb.org/docs/wsupp/S46_RD_09_Mackerel_ABC_reportOpt.pdf (19 February 2019)

Wilson, G., Aruliah, D.A., Brown, C.T., Chue Hong, N.P., Davis, M., Guy, R.T., Haddock, S.H.D., Huff, K.D., Mitchell, I.M., Plumbley, M.D., Waugh, B., White, E.P., and P. Wilson. 2014. Best Practices for Scientific Computing. PLoS Biol 12(1): e1001745. https://doi.org/10.1371/journal.pbio.1001745

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., and T.K. Teal. 2017. Good enough practices in scientific computing. PLOS Computational Biology 13(6): e1005510. https://doi.org/10.1371/journal.pcbi.1005510

Annex 1: List of participants

Name	Email
Allen R. Kronlund	Allen.Kronlund@dfo-mpo.gc.ca
Andrew Campbell	andrew.campbell@marine.ie
Bjarki Thor Elvarsson	bjarki.elvarsson@hafogvatn.is
Carmen Fernandez (chair)	Carmen.Fernandez@ieo.es
Cecilia Pinto	Cecilia.PINTO@ec.europa.eu
Christoph Konrad	Christoph.KONRAD@ec.europa.eu
Claus Reedtz Sparrevohn	crs@pelagisk.dk
Colm Lordan	colm.lordan@ices.dk
Daisuke Goto	daisuke.goto@hi.no
Daniel Howell	daniel.howell@hi.no
David Miller	david.miller@ices.dk
Dorleta Garcia	dgarcia@azti.es
Doug Butterworth	doug.butterworth@uct.ac.za
Ernesto Jardim	ernesto.jardim@ec.europa.eu
Höskuldur Björnsson	hoskuldur.bjornsson@hafogvatn.is
Hugo Mendes	hmendes@ipma.pt
Iago Mosqueira	iago.mosqueira@ec.europa.eu
Jonathan Deroba	Jonathan.Deroba@noaa.gov
José De Oliveira	jose.deoliveira@cefas.co.uk
Laurence Kell	laurie@seaplusplus.co.uk
Marc Taylor	marc.taylor@thuenen.de
Martin Pastoors	mpastoors@pelagicfish.eu
Michael Gras	Michael.Gras@Marine.ie
Michel Bertignac	Michel.Bertignac@ifremer.fr
Mollie Elisabeth Brooks	molbr@aqua.dtu.dk
Niels Hintzen	niels.hintzen@wur.nl
Noél Holmgren	noel.holmgren@slu.se

Polina Levontin	polina.levontin02@imperial.ac.uk
Quang Huynh	q.huynh@oceans.ubc.ca
Santiago Cerviño	santiago.cervino@ieo.es
Sarah B. M. Kraak	sarah.kraak@thuenen.de
Simon Fischer	simon.fischer@cefas.co.uk
Tanja Miethe	t.miethe@marlab.ac.uk
Thomas Brunel	thomas.brunel@wur.nl

Annex 2: ICES Management Strategy Evaluations 2013 -2018

Table A2.1. Management Strategy Evaluations (MSEs) conducted by ICES since the previous WKGMSE meeting (ICES, 2013b).

Year of advice	Species	Stock	Report
2013	Blue whiting	whb.27.1-91214	ICES. 2013. ICES Ad Hoc Group on blue whiting management plan evaluations. Authors: David C. M. Miller and Dankert Skagen. ICES CM 2013/ACOM:76.
2013	Haddock	had.27.5a	Björnsson, H. 2013. Report of the evaluation of the Icelandic haddock management plan. ICES CM 2013/ACOM:59.
2013	Haddock	had.27.6b	ICES. 2013. Report of the Second Workshop to evaluate the EU–Russian proposal for the harvest control component of the management plan for Rockall haddock, 4–6 June 2013, ICES HQ, Copenhagen, Denmark. ICES CM 2013/ACOM:67.
2013	Herring	her.27.1-24a514a	ICES. 2013. Report of the Blue Whiting/Norwegian Spring-Spawning (Atlanto-Scandian) Herring Workshop (WKBWNSSH), 11–13 March 2013, Bergen, Norway. ICES CM 2013/ACOM:69. 88 pp.
2013	Horse mackerel	hom.27.2a4a5b6a7a-ce-k8	ICES. 2013. Report of the Workshop to evaluate the EU management plan for Western horse mackerel (WKWHMAC), 18–19 June 2013, ICES Headquarters, Copenhagen. ICES CM 2013/ACOM:59.
2013	Saithe	pok.27.5a	Hjörleifsson, E. and Björnsson, H. 2013. Report of the evaluation of the Icelandic saithe management plan. ICES CM 2013/ACOM:61.
2013	Sardine	pil.27.8c9a	ICES. 2013. Report of the Workshop to Evaluate the Management Plan for Iberian Sardine (WKSardineMP), 4–7 June 2013, Lisbon, Portugal. ICES CM 2013/ACOM:62.
2014	Horse mackerel	hom.27.3a4bc7d	ICES. 2014. Evaluation of a multi-annual plan including an index based HCR for North Sea horse mackerel, 17-18 June 2014, IJmuiden, the Netherlands. ICES CM 2014/ACOM:55.
*2014**	Redfish - beaked	reb.2127.dp	ICES. 2014. Workshop on Redfish Management Plan Evaluation (WKREDMP), 20–25 January 2014, ICES Headquarters. ICES CM 2014/ACOM:52
2014	Redfish - beaked	reb.27.1-2	ICES. 2014. Workshop on Redfish Management Plan Evaluation (WKREDMP), 20–25 January 2014, ICES Headquarters. ICES CM 2014/ACOM:52
*2014**	Redfish - golden	reg.27.561214	ICES. 2014. Workshop on Redfish Management Plan Evaluation (WKREDMP), 20–25 January 2014, ICES Headquarters. ICES CM 2014/ACOM:52
*2015**	Herring	her.27.3a47d	ICES. 2015. Report of the Workshop to evaluate the TAC calculation for herring in IIIa and management plan for herring in the North Sea (WKHerTAC), 13–16 January 2015, Copenhagen, Denmark. ICES CM 2015/ACOM:47. 141 pp.

Year of advice	Species	Stock	Report
2015	Mackerel	mac.27.nea	ICES. 2015. Report of the EU Workshop on the NEA Mackerel Long-term Management Plan (WKMACLTMP), 24–27 June and 17–19 November 2014, Copenhagen, Denmark. ICES CM 2014/ACOM:63.
2016	Blue whiting	whb.27.1-91214	ICES. 2016. Report of the Workshop on Blue Whiting Long Term Management Strategy Evaluation (WKBWMS), 30 August 2016ICES HQ, Copenhagen, Denmark. ICES CM 2016/ACOM:53
2016	Capelin	cap.27.1-2	ICES. 2016. Report of the second Workshop on Management Plan Evaluation on Northeast Arctic cod and haddock and Barents Sea capelin (WKNEAMP-2), 25–28 January 2016, Kirkenes, Norway. ICES CM 2016/ACOM:47. 76 pp.
2016	Cod	cod.27.1-2	ICES. 2016. Report of the second Workshop on Management Plan Evaluation on Northeast Arctic cod and haddock and Barents Sea capelin (WKNEAMP-2), 25–28 January 2016, Kirkenes, Norway. ICES CM 2016/ACOM:47. 76 pp.
2016	Cod	cod.27.22-24	ICES. 2016. Annex 11: Report of the Baltic Fisheries Assessment Working Group (WGBFAS), 12–19 April 2016, ICES Headquarters, Copenhagen, Denmark. ICES CM 2016/ACOM:11.
2016	Haddock	had.27.1-2	ICES. 2016. Report of the second Workshop on Management Plan Evaluation on Northeast Arctic cod and haddock and Barents Sea capelin (WKNEAMP-2), 25–28 January 2016, Kirkenes, Norway. ICES CM 2016/ACOM:47. 76 pp.
*2016**	Pandalus	pra.27.3a4a	ICES. 2016. Report of the Workshop on management strategy evaluation for the Pandalus in Subdivision 3.a.20 and Division 4.a East fishery (WKPANDMSE), 23–25 August 2016, Bergen, Norway. ICES CM 2016/ACOM:54.
2017	Herring	her.27.5a	ICES. 2017. Report of the Workshop on Evaluation of the Adopted Harvest Control Rules for Icelandic Summer-Spawning Herring, Ling and Tusk (WKICEMSE), 21–25 April 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:45. 49 pp.
2017	Ling	lin.27.5a	ICES. 2017. Report of the Workshop on Evaluation of the Adopted Harvest Control Rules for Icelandic Summer-Spawning Herring, Ling and Tusk (WKICEMSE), 21–25 April 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:45. 49 pp.
*2017**	Mackerel	mac.27.nea	ICES. 2017. Report of the Workshop on management strategy evaluation for the mackerel in subareas 1–7 and 14, and in divisions 8.a–e and 9.a (Northeast Atlantic) (WKMACMSE), 28–29 August 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:48. 210 pp.
2017	Pandalus	pra.27.3a4a	Cardinale, M., Fernandez, C., Eigaard, O.R., and Søvik, G. 2017. Report on the Long-term Management Strategy Evaluation for Northern Shrimp (Pandalus borealis) in Division 4.a East and Subdivision 20, October–November 2017. ICES CM 2017/ACOM:52. 185 pp.
2017	Sardine	pil.27.8c9a	ICES. 2017. ANNEX 11: Report of the Benchmark Workshop on Pelagic Stocks (WKPELA), 6–10 February 2017, Lisbon, Portugal. ICES CM 2017/ACOM:35. 294 pp.
2017	Tusk	usk.27.5a14	ICES. 2017. Report of the Workshop on Evaluation of the Adopted Harvest Control Rules for Icelandic Summer-Spawning Herring, Ling and Tusk (WKICEMSE), 21–25 April 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:45. 49 pp.

Year of advice	Species	Stock	Report
2018	Herring	her.27.1-24a514a	ICES. 2018. Report of the Workshop on a long-term management strategy for Norwegian Spring-spawning herring (WKNSSHMSE), 26-27 August 2018, Torshavn, Faroe Islands. ICES CM 2018/ACOM:53. 113pp.
2018	Herring	her.27.irls	ICES. 2018. Report on Celtic Sea Herring Management Plan Evaluation. Coming as Annex 8 in the report of the Herring Assessment Working Group for the Area South of 62°N (HAWG), scheduled to meet 12–20 March 2018 at ICES HQ, Denmark.
2018	Horse mackerel	hom.27.9a	ICES. 2018. Report on the Assessment of a Long-term Management Strategy for Southern Horse Mackerel (hom27.9a), 15–16 February 2018. Manuela Azevedo, Hugo Mendes, Gersom Costas, Ernesto Jardim, Iago Mosqueira, Finlay Scott (Authors.) ICES CM 2018/ACOM:42.
2018	Norway pout	nop.27.3a4	ICES. 2018. Report of the Workshop for management strategy evaluation for Norway Pout (WKNPOUT), 26–28 February 2018, ICES HQ, Copenhagen, Denmark. ICES CM 2018/ACOM:38. 96 pp.
2018	Redfish - beaked	reb.27.1-2	ICES. 2018. Report of the Workshop on the evaluation of harvest control rules for Sebastes mentella in ICES areas 1 and 2 (WKREBMSE). June–August 2018, by correspondence. ICES CM 2018/ACOM:52.

**Summary table not provided*

The following footnotes apply to the summary tables below:

* Knowledge base: This is the information that will be available about the state of the stock, in particular whether there is an assessment or not. If it is something else, please specify.

** Decision basis: This is the measure that determines the exploitation in the harvest rule. For example, SSB at the start of the TAC year, TSB in the last assessment year.

*** Comparison with ordinary assessment? This is to indicate whether there has been attempts to verify that the performance of the assessment in the model is similar to that experienced by the WG, for example with respect to retrospective problems and inconsistencies.

**** Risk types:

Risk1 = average probability that SSB is below B_{lim}, where the average is taken across the *ny* years.

Risk2 = probability that SSB is below B_{lim} at least once during the *ny* years.

Risk3 = maximum probability that SSB is below B_{lim}, where the maximum is taken over the *ny* years.

If your definition of risk does not fit any of these, please explain.

2013 Haddock had.27.5a

Björnsson, H. 2013. Report of the evaluation of the Icelandic haddock management plan. ICES CM 2013/ACOM:59.

Background		
Motive/ initiative/ background	Request to ICES from the government of Iceland. The goal was to adopt a management plan where the TAC next fishing year is 40% of the biomass of 45cm and larger in the beginning of the calendar year following the assessment year.	
Main objectives	Keep SSB > $B_{trigger}$ with > 95% probability while trying to have as much stability in catches as possible.	
Formal framework	ICES on request from the Icelandic Fishery Ministry of Fisheries.	
Who did the evaluation work	ADGISAHA, mainly Höskuldur Björnsson and Einar Hjörleifsson	
Method		
Software Name, brief outline include ref. or documentation	Assessment model called muppet with SSB-rec parameters and other parameters estimated. Replicas of the estimated parameters based on MCMC simulations are used in stochastic simulations with additional stochasticy in weight at age and assessment error added. Density depent growth modelled and selection of the fisheries is size based. Model is described in https://github.com/hoski/Muppet_HCR Details about the software can be found here: Björnsson, H. 2016. Working document on assessment model for Norwegian Spring Spawning Herring. WD 13 to Benchmark Workshop on Pelagic stocks (WKPELA), 29 February–4 March 2016, ICES Headquarters, Copenhagen, Denmark. Information also found at http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2013/ADHOC/IntroAndHad.pdf	
Type of stock	Medium life span, demersal, valuable	
Knowledge base *	Category 1 stock. Catch in numbers by age 1978-2011 and age disaggregated abundance indices from the groundfish survey in March 1985-2012 and the groundfish survey in October 1996-2011. Variability in recruitment is large. Growth is density dependent and size at recruitment depends on yearclass size. The surveys are very good indicators of yearclass size, already at age 1.	
Type of regulation	Tac is given for fishing year from September 1st in the assessment year until August 31st the following year.	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Hockey stick with SSB_{break}, R_{max} and σ estimated	Autocorrelated lognormal with ρ given but σ estimated.
Growth & maturity	Stock weights (from March survey) function of yearclass size and growth function of stock size. Catch weights and maturity function of stock weights.	Deviations from average based on an autocorrelated lognormal year factor.
Natural mortality	0.2	None
Selectivity	Size based. Implemented in the model as function of stock weights.	Autocorrelated lognormal with ρ=0.8 given but σ=0.2 .
Initial stock numbers	Assessment model. Assessment is done by an Adapt type model giving very	MCMC simulations from the

	similar stock size.	assessment model.
Decision basis **	TAC for the fishing year starting September 1st in the assessment year is 40% of the biomass of 45cm and larger fish in the beginning of the calendar year following the assessment year. Proportion of 45 cm and larger in the total biomass of a cohort is compiled by a logit function based on stock weights.	
Number of iterations	2 million mcmc simulations saving every 1000 simulation leading to 2000 parameters sets for stochastic simulations.	
Projection time	60 years.	
Observation and implementation models		
Type of noise	Assessment error autocorrelated lognormal with σ = 0.16 and ϱ = 0.6. σ applies to uncertainty in B3+ in the assessment year but increases through the assessment year and is approximately 0.22 for the biomass of 45cm and larger in the beginning of the calendar year following the assessment year. Both σ and ϱ are based on analytical retros, estimates of σ from assessment model are lower.	Starting value of the assessment error for each iteration is the ratio of the B3+ in the starting year and B3+ for each simulation, overestimation for replicas with low B3+ and vice versa.
Comparison with ordinary assessment?	The ordinary assessment is based on an Adapt type model that leads to very similar results as the HCR evaluation model that is also run every year.	
Projection: If yes - how?	The stock has to be projected through the assessment year to get the biomass in the year following the assessment year. The main problem here is to predict growth. Growth is modelled by following a cohort but using a yearfactor in growth with the average of last 2 years used as the yearfactor in the assessment year.	
Projection: Deviations from WG practice?		
Implementation	Perfect implementation assumed.	
Harvest rule		
Harvest rule design	*The annual total allowable catch (TAC) will be set by applying the following harvest control rule (HCR):* *1. When SSB_{y+1} is equal to or greater than 45 thousand tonnes (SSBtrigger),* $TAC_{y/y+1} = a \times B_{45cm+,y+1}$ *2. When SSB_{y+1} is below 45 thousand tonnes (SSBtrigger),* $TAC_{y/y+1} = a \times SSB_{y+1} / SSB_{trigger} \times B_{45cm+,y+1}$ *where y refers to the assessment year, y/y+1 the fishing year, B_{45cm+} to the biomass of 45cm and larger haddock, and 'a' to the target harvest rate. a is set to 0.4.*	
Stabilizers	None.	
Duration of decisions	Annual	
Revision clause	None but a reevaluation has been asked for in 2019.	
Presentation of results		
Interest parameters	• Average yield	

	• Interannual variability in yield. • 10^{th} percentile of yield. • Risk of SSB falling below Blim/$B_{trigger}$ • Short term risk. In 2013 risk in short term was higher than in the medium and long term.
Risk type and time interval	Type 2 risk. Maximum for any given year.
Precautionary risk level	e.g. 5% of risk type 2.
Experiences and comments	
Review, acceptance:	Accepted by ICES in 2013.
Use of ICES guidelines (WKGMSE 2013)	The guide lines were mostly followed. The main deviation was the use of $B_{trigger}$=Blim that was accompanied by lower harvest rate that would have been if $B_{trigger}$=B_{pa}. Purpose was to increase stability while the “cost” was lower harvest Rate.
Experiences and comments	Should have based Tac on biomass one year earlier. Prediction of growth has turned out to be somewhat problematic especially as it has to be hardwired into prespecified equation. Person knowing details about the stock could do reasonably well each year, better than done with prespecified equation.

2013 Haddock had.27.6b

ICES. 2013. Report of the Second Workshop to evaluate the EU–Russian proposal for the harvest control component of the management plan for Rockall haddock, 4–6 June 2013, ICES HQ, Copenhagen, Denmark. ICES CM 2013/ACOM:67.

Background		
Motive/ initiative/ background	NEAFC request for advice on a Rockall haddock HCR. Evaluate the proposals for HCRs of the management plan for Rockall haddock, for consistency with the precautionary approach and maximization of yield. The evaluation should also include re-consideration of reference points, evaluation of the proposed modification to the harvest control rule for recent low recruitment period. The HCRs was evaluated in 2013.	
Main objectives	Precautionary, maximization of yield	
Formal framework	ICES on request from NEAFC	
Who did the evaluation work	WKHADROCK-2	
Method		
Software Name, brief outline include ref. or documentation	The evaluation of the HCR was carried out using two different methods. One of the methods uses FLR, while the other method was done in an Excel spreadsheet (exploratory runs on the basis of that method were also made in the R) and simulates assessment errors	
Type of stock	Medium life span, demersal	
Knowledge base *	Analytic assessment (XSA), 1 survey.	
Type of regulation	TAC for EU fleets. Calculated using the MSY approch	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	The recruitment was simulated using the method of random numbers with actual observed historical numbers of recruitment. Values for bootstrapping are drawn from two periods: the first (“long-term recruitment”) was the full time–series of the assessment, while the second (“recent recruitment”) used the period 2003–2012 when recruitment has generally been low.	No
Growth & maturity	Mean values for the all period (1991-2012) were used for stock weights at ages and catch weights at ages. Maturity at age: 0 for ages 1-2 and 1 for ages 3-7+.	No
Natural mortality	Natural mortality at age: 0.2	No
Selectivity	Constant, mean for 2002-2012	No

Initial stock numbers	From assessment	No
Decision basis **	SSB in the TAC year	
Number of iterations	100 and 1000 for some scenarios	
Projection time	for 28 years	
Observation and implementation models		
Type of noise	The simulation of assessment errors were applied in Excel spreadsheet method. The initial assessment errors were calculated by retrospective analysis of XSA and were modelled using of random method with that observed historical values. The main uncertainty in the assessment and forecast is estimation of discards.	
*** Comparison with ordinary assessment?	No	
Projection: If yes - how?	Yes, on basis of 50 percentile with random values of recruitment. A TAC is applied in the intermediate year. Assessment errors simulated by random method.	
Projection: Deviations from WG practice?	A TAC constraint is applied in the intermediate year, this is consistent with assessment procedure.	
Implementation	Catch at age calculated from the perceived stock using the fishing mortality derived from the harvest control rule and the given exploitation pattern. This catch at age is then applied to the actual stock	
Harvest rule		
Harvest rule design	ICES advises that the proposed, using the existing reference points (Fpa=0.3), is not in accordance with the precautionary approach under the conditions of low recruitment that have prevailed since 2004. The proposed HCR implies that the stock will be below Blim with a high probability under those recruitment conditions.	
Stabilizers	The TAC should be changed compared with the previous year TAC according to the following formula: TACy = TACf + a * (TACy-1 - TACf) where TACy is the TAC that is to be set by the management plan, TACy-1 is the TAC that was fixed the previous year and TACf is the initial TAC calculated for year y, a – coefficient stabilizer	
Duration of decisions	Annual	
Revision clause	Revision asked for 2019 by NEAFC	
Presentation of results		
Interest parameters	Risk, Catch, Inter-annual catch variation, proportion of year when different parts of the HCRs apply, stock size	

**** Risk type and time interval	Type of risk is depended on the scenarios: recruitment period (recent low or medium recruitment) and on referents points (F=0.2 or 0.3)
Precautionary risk level	5% of risk type 1.
Experiences and comments	
Review, acceptance:	Accepted by ICES in 2013 with changes of Fpa. ICES advises that when SSB is greater than Bpa a maximum F value of 0.2 would be required for the HCR to be consistent with the precautionary approach even under a low recruitment regime.
Use of ICES guidelines (WKGMSE 2013)	The guidelines in the WKGMSE 2013 report were used. It was useful. Probability 3 should be used in accordance to guide but it should be mentioned that this type of criteria is dependent on number of years used in the simulation model. The more years used the higher value of Prob 3 parameters we get. It is problematic to base the conclusion on this criterion. The more appropriate criterion is Prob 1.
Experiences and comments	

2013 Herring her.27.1-24a514a

ICES. 2013. Report of the Blue Whiting/Norwegian Spring-Spawning (Atlanto-Scandian) Herring Workshop (WKBWNSSH), 11-13 March 2013, Bergen, Norway. ICES CM 2013/ACOM:69.

Background		
Motive/ initiative/ background	The Coastal States have in June 2012 submitted a request for ICES to re-evaluate NSSH long term management plan. Particularly the effect of implementing F_{MSY} = 0.15 estimated in WGWIDE 2010 and a strategy taking into account the recruitment in recent past were to be studied.	
Main objectives	Precautionary, stable and high catches	
Formal framework	ICES on request from The Coastal States	
Who did the evaluation work	WKBWNSSH 2013	
Method		
Software Name, brief outline include ref. or documentation	HCS-simulation programme, downloadable with documentation from (http://www.dwsk.net/Downloads.htm) Age structured operating model, no full assessment in the loop.	
Type of stock	Long life span, pelagic, straddling, very valuable	
Knowledge base *	Analytic assessment	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Beverton-Holt and spasmodic strong year classes with expected interval 8 yrs and range 2 years. The frequency distribution of recruitment and mean SSB fitted to SR pairs form 1950-2008, excluding the collapse period 1968-87.	Log-normal, CV adjusted to resemble that of the observed recruitment.
Growth & maturity	Weight in catch: average over 1997-2011 Weight in stock: average over 1998-2012 no density dependence in growth Maturity: maturity ogive for a normal year class	Normally distributed multiplicative noise with sigma of 0.1 for all ages in both weight at age in catch and stock Deterministic maturation probability
Natural mortality	For ages 0-2 M = 0.9, ages 3+ M = 0.15	No
Selectivity	Average 2007-2011 from WGWIDE 2012, smoothed for the older ages (10-15+)	No
Initial stock numbers	From assessment	Normally distributed with CV 0.3
Decision basis **	SSB in the TAC year	
Number of iterations	1000	
Projection time	98 years	

Observation and implementation models		
Type of noise	Year factor + age factor on stock numbers at age	CVs for age factor from 2012 assessment bootstraps.
*** Comparison with ordinary assessment?	Year factor scaled to give CV of SSB resembling the observed.	
Projection: If yes - how?	No.	
Projection: Deviations from WG practice?		
Implementation	First F given by the HCR is found based on the perceived SSB. Then a TAC is calculated, and this TAC is translated into catch numbers at age by a searching routine that finds the overall F leading to the TAC when applied to the true population, assuming the standard selection at age and the currently valid weights at age. • Then noise is added to the catch numbers at age as age dependent log-normally distributed random noise, with CVs specified individually for each age. • The realized total catch is the decided TAC multiplied by a log-normally distributed year factor. • The numbers caught are adjusted with a common factor to give a sum of products of the catches and catch weights equal to the realized total catch. • Realized fishing mortalities are derived from the resulting catches in numbers at age and the true stock numbers.	Recruitment survey CV = 0.2, age factors for uncertainty (implementations and observation) taken from TASACS 100 bootstraps, average over last 5 years. No bias, although the effect of bias has been studied and reported.
Harvest rule		
Harvest rule design	Seven rules were studied: The current management plan – option A (no amendment): F = 0.125 is used when SSB is above Btrigger (= Bpa = 5 000 000 t). When SSB is below Btrigger F decreases linearly down to 0.05 when SSB = Blim (= 2 500 000 t), below which F = 0.05 is applied. Modification according to option B (use FMSY): F = FMSY = 0.15 is used when SSB is above Btrigger. Otherwise the same as the current management plan (HCR 1).	

	Modification according to option C: F depends on recent recruitment when SSB is above Blim. If recent recruitmentis above the long term average F = 0.15 is used when SSB is above Btrigger. If Btrigger > SSB > Blim, F is taken from the linear curve between F = 0.15 at Btrigger and F = 0.05 at Blim, depending on the current level of SSB. If recent recruitment is below the long term average F = 0.125 is used when SSB is above Btrigger. If Btrigger > SSB > Blim, F is taken from the linear curve between F = 0.125 at Btrigger and F = 0.05 at Blim, depending on the current level of SSB. Otherwise the same as the current management plan. Zero catch at zero SSB, otherwise same as option A. F = 0.125 is used when SSB is above Btrigger. When SSB is below Btrigger F decreases linearly to 0 when SSB = 0. Zero catch at zero SSB modified according to option B. F = 0.15 is used when SSB is above Btrigger. When SSB is below Btrigger F decreases linearly to 0 when SSB = 0. Zero catch at zero SSB modified according to option C. When SSB is below Btrigger F decreases linearly to 0 when SSB = 0. F depends on recent recruitment: If recent recruitment is above the long term average F = 0.15 is used when SSB is above Btrigger. If SSB < Btrigger, F is taken from the linear curve between F = 0.15 at Btrigger and F = 0 at SSB = 0, depending on the current level of SSB. If recent recruitment is below the long term average F = 0.125 is used when SSB is above Btrigger. If SSB < Btrigger, F is taken from the linear curve between F = 0.125 at Btrigger and F = 0 at SSB = 0, depending on the current level of SSB. The current management plan, but with the modification that at high SSB a further linear increase in F is applied.
Stabilizers	No stabilizer in the basic runs. However, year-to-year TAC constraint was explored with the current management plan.
Duration of decisions	Annual
Revision clause	?
Presentation of results	
Interest parameters	• Average catch • Average F • Average SSB • Average of the interannual variability in TAC (TAC IAV) calculated as mean (over all years in the time frame and all bootstrap replicas) of the absolute inter-annual variation in percent: abs{[TAC(y)-TAC(y-1)]/[(TAC(y-1)+TAC(y)) /2]*100}. • The probability of the (true in the model) SSB falling below B_{lim} (risk type 3 as defined by WKGMSE 2013; the maximum probability that SSB is below B_{lim}, where the maximum (of the annual probabilities) is taken over the relevant years; and risk type 2 as defined by WKGMSE 2013; the probability that SSB is below B_{lim} at least once within the time period) •The probability of the (assumed in the model, i.e., including observation uncertainty) SSB falling below $B_{trigger}$ risk type 3 as defined by WKGMSE 2013; the maximum probability that SSB is below $B_{trigger}$, where the maximum (of the annual probabilities) is taken over the relevant years). All the performance indicators were estimated on short term (first 5 years), medium term (first 10 years and years 6 to 15), long term (first 50 years), and quasi-equilibrium (last 50 years of the 98 years simulation period).

**** Risk type and time interval	Risk type 3 as defined by WKGMSE 2013; the maximum probability that SSB is below B_{lim}, where the maximum (of the annual probabilities) is taken over the relevant years). For short, medium and long term and quasi-equilibrium (see definitions above). Risk type 2 as defined by WKGMSE 2013; the probability that SSB is below B_{lim} at least once within the time period. For short, medium and long term and quasi-equilibrium (see definitions above).
Precautionary risk level	5% of risk type 3.
Experiences and comments	
Review, acceptance:	The current management plan has been in effect since 2001.
Experiences and comments	Estimates of Blim vary considerably depending on the assessment year and time period of stock-recruitment pairs used. The current Blim 2.5 million tonnes is within confidence intervals of all performed estimations, and is not recommended to be changed at the moment, even though the analysis suggests that Blim might be somewhat higher. However, further investigation is needed, particularly when the assessment model in use for this stock will be applying data further back in time (currently only data from 1988 onwards is used in TASACS assessment). The different harvest control rule options investigated yield in general rather similar results. None of the HCR options is precautionary in the short term. However, this is not surprising considering that the stock is decreasing and no strong year classes have recruited since 2004. In the stationary equilibrium situation where the current conditions no longer influence stock dynamics (the last 50 years of the 98 year simulation period), all strategies are precautionary. Assessments show positive bias in the last two decades with average factor of 0.26, i.e., the stock tends to be overestimated. This kind of overestimation would lead to overestimation of SSB of about 25%.

2013 Horse mackerel hom.27.2a4a5b6a7a-ce-k8
ICES. 2013. Report of the Workshop to evaluate the EU management plan for Western horse mackerel (WKWHMAC), 18–19 June 2013, ICES Headquarters, Copenhagen. ICES CM 2013/ACOM:59.

Background		
Motive/ initiative/ background	In 2008 a management plan for WHM was proposed following collaboration between scientists and the Pelagic AC. During the initial review ICES raised concerns with 1. The robustnes of the plan to poor recruitment scenarios 2. The lack of an explicit protection rule in the HCR 3. The long term trajectory of SSB. For these reasons, the plan was judged to be precautionary in the short term only. In 2013 a workshop was held in ICES and tasked to 1. Evaluate the MP whether it is precautionary in the long as well as the short term and for consistency with the PA approach and maximization of yield 2. If not precautionary in the long term, identify reinforcements in the rule that would resolve this 3. Identify the appropriate TAC for 2013 in accordance with the revised rule	
Main objectives	Precautionarity, stability in catch	
Formal framework	Pelagic RAC initiative	
Who did the evaluation work	WKWHMAC 2013	
Method		
Software Name, brief outline include ref. or documentation	2 separate models: 1. F-PRESS (R) Codling, E. and Kelly,C. "F-PRESS: A stochastic simulation tool for developing fisheries management advice and evaluating management strategies".Irish Fisheries Investigations No17, Marine Institute 2006, ISSN 0578 7476. 2. Bespoke Fortran application	
Type of stock	Medium life span, pelagic, high value	
Knowledge base *	Analytic assessment	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	FPRESS – Segmented Regression fit to assessment SR pairs from 2008 assessment (excluding 1982). Spikes modelled separately. Fortran – Ricker fit to each stock replicate. Spikes modelled separately.	LN error with CV derived from residuals to fit. Spikes modelled using boxcar distribution. From iteration-specific fits. Spikes modelled using boxcar distribution.
Growth & maturity	Stock weights from sampling	Random draw of weight at age vector from 1998-2010.

		No change through simulation period
Natural mortality	0.15	No
Selectivity	FPRESS – Assessment Fortran – 1000 realisations generated from assessment output.	Normally distributed with 20% CV Iteration specific
Initial stock numbers	Assessment year 2004	LN error tuned to replicate uncertainty in SSB from the assessment
Decision basis **	Slope of the previous 3 egg survey egg abundance estimates (FPRESS and Fortran) SSB - constant proportion strategy (Fortran only)	
Number of iterations	FPRESS 10000, Fortran 500	
Projection time	FPRESS 40 years, Fortran 20 years	
Observation and implementation models		
Type of noise	Eggs - noise is applied to the modelled egg count via a process error (LN) component and an observation error (N) component (FPRESS and FORTRAN) SSB - (not actually required for the operation of the HCR), normal error distribution with cv=25% (FPRESS)	
*** Comparison with ordinary assessment?	NA	
Projection: If yes - how?	No	
Projection: Deviations from WG practice?	NA	
Implementation		
Harvest rule		
Harvest rule design	1) Hybrid slope design incorporating a fixed TAC proportion and a slope factor from the previous 3 egg surveys 2) Constant proportion of standing SSB	
Stabilizers	None, inherent in rule design	
Duration of decisions	Triennial (following egg survey)	
Revision clause	3 yearly	
Presentation of results		
Interest parameters	Short term 2011-2020, medium term 2021-2030 Yield (Mean, Median and Cumulative) TAC changes (number positive, negative, 10th, 50th and 90th percentiles) Risk types 1 and 2 Percentage of stock crashes SSB (graphical)	

**** Risk type and time interval	Risk types 1 and 2 reported
Precautionary risk level	5% Risk 1
Experiences and comments	
Review, acceptance:	Reviewed by ADGWHMAC (10-11/07/2013), main conclusions (no longer precautionary as originally configured) accepted and summarised in special request advice.
Use of ICES guidelines (WKGMSE 2013)	This exercise expands on the original evaluation as reported in WKGMSE2013. An additional model was used (the FORTRAN approach described above), more detailed stock-recruit modelling was undertaken and additional stochastic elements incorporated.
Experiences and comments	Final report remains in draft format. Outcome of process became somewhat irrelevant from difficulties arising from EU co-decision making which meant the HCR would not be enshrined in legislation.

2013 Sardine pil.27.8c9a

ICES. 2013. Report of the Workshop to Evaluate the Management Plan for Iberian Sardine (WKSardineMP), 4–7 June 2013, Lisbon, Portugal. ICES CM 2013/ACOM:62. 84 pp

Background	
Motive/ initiative/ background.	The multiannual MP proposal started as an initiative of the Portuguese Commission for the sardine fishery. This Commission has representatives from the main stake- holders of the Portuguese sardine fishery, the National Purse-seine Fishers Associa- tion (ANOPCERCO), the National Association of the Canning Industry (ANICP), the auctions managers' enterprise (DOCAPESCA), IPMA and is chaired by the Fish- eries Directorate (DGRM). The plan was endorsed by the Spanish and Portuguese governments and submitted to ICES for evaluation through the EU Commission. The EU Commission submitted the MSE request to ICES.
Main objectives	Precautionary, stable catches near MSY
Formal framework	ICES on request from EU
Who did the evaluation work	WKSardineMP2013
Method	
Software Name, brief outline include ref. or documentation	HCS, version 13_3 Age structured operating model, no assessment but options to imitate some of the effects of noise in the input data Manual and code can be downloaded from www.dwsk.net
Type of stock	Short–medium lifespan, pelagic, very valuable

Knowledge base *	Analytic assessment, approved but with some retrospective bias at present.	
Type of regulation	No EU TAC, national regulations include effort and catch limitations and area closures.	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Hockey stick, adjusted to imitate the distribution of recruitments 1993– 2010	Lognormal, truncated, CV from residuals
Growth & maturity	Weights-at-age: Average over 2009– 2011, no density-dependence. B1+ was used as a proxy for SSB	Random lognormal noise on weights-at-age.
Natural mortality	Gislason *et al*. formula	No
Selectivity	Selectivity estimate by the assess- ment as a fixed vector since 1991 (with flat selectivities at ages 3–5 = 1).	No, except deviations at the implementation step, with CVs for all ages similar to those in the assessment
Initial stock numbers	From assessment	Variances from assessment (inverse Hessian)
Decision basis **	B1+ at the start of the intermediate year	
Number of iterations	1000	
Projection time	30 years	
Observation and implementation models		
If assessment in the loop		
Input data		
***** Comparison with ordi- nary assessment?**		
Deviations from WG prac- tice?		
If no assessment in the loop		
Type of noise	Year factor + age factor in an auto- regressive model on stock numbers- at-age along year classes.	Both lognormal + auto-regressive model along year classes Age factor from CV estimates in assessment Year factor adapted to reproduce CV of SSB estimate in assessment

***** Comparison with ordi- nary assessment?**	Year factor scaled to give CV of SSB in year 0 as CV of SSB in assessment Cumulated distribution of recruitments was compared to historical recruitments	
Projection: If yes - how?	No.	
Projection: Deviations from WG practice?		
Implementation	TAC according to the rule translated into catch numbers-at-age.	Lognormally distributed error on catches in numbers CVs for all ages similar to those in the assessment, no bias
Harvest rule		
Harvest rule design	TAC rule with two breakpoints B_0 and $B_{trigger}$. TAC = 0 below B_0, fixed TAC above $B_{trigger}$, linear reduction between these points. Other alternatives were explored.	
Stabilizers	None in the proposed rule. Some options explored.	
Duration of decisions	Annual	
Revision clause	Not decided, suggest after three years.	
Presentation of results		
Interest parameters	Risk to B_{lim} and of stock collapse (crash), Catch (Mean and 10-50-90 percentiles), Interannu- al variation, Probability of recovery from below B_{lim} in **n** years, n=1,5	
****** Risk type and time interval**	Type 3 for years 21–30	
Precautionary risk level	To be decided. Proposed to be kept below 15%.	
Experiences and comments		
Review, acceptance:		
Experiences and comments	Recruitment has declined recently for unknown reasons, the SR function represents the re- cent past (since 1993)	

*** Knowledge base: This is the information that will be available about the state of the stock, in particu- lar whether there is an assessment or not. If it is something else, please specify.**

**** Decision basis: This is the measure that determines the exploitation in the harvest rule. For example, SSB at the start of the TAC year, TSB in the last assessment year,.**

***** Comparison with ordinary assessment? This is to indicate whether there has been attempts to verify that the that the performance of the assessment in the model is similar to that experienced by the WG, for example with respect to retrospective problems and inconsistencies.**

**** Risk types:

- **Risk1** = *average* probability that SSB is below B_{lim}, where the average is tak- en across the ny years.
- **Risk2** = probability that SSB is below B_{lim} *at least once* during the ny years.
- **Risk3** = *maximum* probability that SSB is below B_{lim}, where the maximum is taken over the ny years.

If your definition of risk does not fit any of these, please explain.

2013 Saithe pok.27.5a

Hjörleifsson, E. and Björnsson, H. 2013. Report of the evaluation of the Icelandic saithe management plan. ICES CM 2013/ACOM:61.

Background		
Motive/ initiative/ background	Request to ICES from the government of Iceland. The goal was to adopt a management plan where the TAC next fishing year is the average of TAC last fishing year and 20% of biomass 4+ in the beginning of the assessment year. The rule is the same as for Icelandic cod. The advice does not depend on the selection pattern of the fleet.	
Main objectives	Keep SSB > $B_{trigger}$ with > 95% probability while trying to have as much stability in catches as possible.	
Formal framework	ICES on request from the Icelandic Fishery Ministry	
Who did the evaluation work	ADGISAHA, mainly Einar Hjörleifsson and Höskuldur Björnsson.	
Method		
Software Name, brief outline include ref. or documentation	Assessment model called muppet with SSB-rec parameters and other parameters estimated. Replicas of the estimated parameters based on MCMC simulations are used in stochastic simulations with additional stochasticity in weight at age and assessment error added. Model available at https://github.com/hoski/Muppet_HCR Details about the software can be found here: Björnsson, H. 2016. Working document on assessment model for Norwegian Spring Spawning Herring. WD 13 to Benchmark Workshop on Pelagic stocks (WKPELA), 29 February–4 March 2016, ICES Headquarters, Copenhagen, Denmark. Information also available at http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2013/ADHOC/IntroAndHad.pdf	
Type of stock	Medium life span, pelagic and demersal, straddling, schooling, valuable	
Knowledge base *	Category 1 stock. Catch in numbers by age 1978-2011 and age disaggregated abundance indices from ICEGFS 1985-2012. Indices are noisy as is usual for this species due to its schooling semi-pelagic behavior.	
Type of regulation	Tac is given for a fishing year from September 1st in the assessment year until August 31st the following year.	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Hockey stick with SSB_{break}, R_{max} and σ estimated	Autocorrelated lognormal with ρ given but σ estimated.
Growth & maturity	Average weights over 2009-2011 (low values), no density dependence. Maturity 2011 generated by a gam model based on survey data.	Deviations from average based on a autocorrelated lognormal year factor. Maturity at age fixed.
Natural mortality	0.2	None
Selectivity	Estimated separately for 3 periods. The selection from the last period (2004-2011) used in stochastic simulations. Advice (proportion of biomass 4+) not depending on selectivity.	None.
Initial stock numbers	Assessment model.	MCMC simulations of the assessment model.

Decision basis **	Biomass of age 4 using catch weights in the beginning of the assessment year is the basis of TAC for the fishing year starting September 1st in the assessment years.	
Number of iterations	2 million mcmc simulations saving every 1000th simulation leading to 2000 sets of parameters for stochastic simulations.	
Projection time	60 years.	
Observation and implementation models		
Type of noise	Assessment error , autocorrelated lognormal with σ = 0.2 and ϱ = 0.5. σ applies to uncertainty in B4+ in the assessment year that is the basis for advice. Basis for σ output from assessment model. Basis for ϱ retrospective runs. Bias is the lognormal bias of 0.02.	Starting value of the assessment error for each iteration is the ratio of the B4+ in the starting year and B4+ for each simulation, overestimation for replicas with low B4+ and vice versa.
Comparison with ordinary assessment?	The ordinary assessment is based on the same model so similarity is expected.	
Projection: If yes - how?	Projection is not used to generate annual advice. The simulations do therefore cover that step of the process perfectly.	
Projection: Deviations from WG practice?	No projection needed.	
Implementation	Perfect implementation assumed.	
Harvest rule		
Harvest rule design	1. *When spawning stock biomass in the assessment year (SSB_y) is equal to or greater than 65 thousand tonnes ($SSB_{trigger}$),* $TAC_{y/y+1} = (a \times B_{4+,y} + TAC_y) / 2$ 2. *When SBB_y is below 65 thousand tonnes ($SSB_{trigger}$),* $TAC_{y/y+1} = a \times SSB/SSB_{trigger} \times B_{4+,y}$ *where y refers to the assessment year, y/y+1 the fishing year, $B_{4+,y}$ to the biomass of 4-year and older saithe in the assessment year, and 'a' to the target harvest rate. a is set to 0.2.*	
Stabilizers	Last fishing years advice weights 50%.	
Duration of decisions	Annual	
Revision clause	None but a reevaluation has been asked for in 2019.	
Presentation of results		
Interest parameters	• Average yield • Interannual variability in yield. • 10th percentile of yield. • Risk of SSB falling below Blim • Short term risk. Risk in short term was lower than in medium and long term.	

**** Risk type and time interval	See definitions below
Precautionary risk level	e.g. 5% of risk type 2.
Experiences and comments	
Review, acceptance:	Accepted by ICES in 2013
Use of ICES guidelines (WKGMSE 2013)	The guide lines were mostly followed. The main deviation was the use of Btrigger=Blim that was accompanied by lower harvest rate that would have been if $B_{trigger}$=B_{pa}. The purpose was to increase stability while the "cost" was lower harvest rate. Later B_{lim} was changed so the current trigger point is close to B_{pa}
Experiences and comments	History shows that saithe is a difficult species to catch and the composition of the fleet in recent years has made it less saithe fleet (increased longlining a gear hardly catching saithe) . Perhaps a factor that should be taken into account to reduce harvest rate from what biological considerations give.

2013 Blue whiting whb.27.1-91214

ICES. 2013. Report of the Blue Whiting/Norwegian Spring-Spawning (AtlantoScandian) Herring Workshop (WKBWNSSH), 11–13 March 2013, Bergen, Norway. ICES CM 2013/ACOM:69. 88 pp

Background		
Motive/ initiaitve/ background.	The industry was not satisfied with current unpredictable quotas, and developed a proposed management plan. NEAFC requested ICES to examine the current MP and to examine the proposed HCR further.	
Main objectives	Precautionary, stable catches near MSY, ability to handle changes between productivity regimes (recruitment)	
Formal framework	ICES on request from NEAFC	
Who did the evaluation work	WKBWNSSH 2013	
Method		
Software Name, brief outline include ref. or documentation	HCS (Harvest Control rule Simulator) software v13.2 (http://www.dwsk.net/). Additional runs done using R and FLR Assessment model in AD model builder (SAM - state space model). Age structured operating model, with catches at age, assessment error estimated (some full feedback runs) Unpublished, undocumented, code available on request.	
Type of stock	Medium life span, pelagic, low value	
Knowledge base *	Analytic assessment	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Spasmodic peaks from average recruitment, two 'regimes' (high and low)	Log-normal, CV from residuals
Growth & maturity	Average over last three years, density dependence also examined	Yes, small CV
Natural mortality	Constant estimate (0.2)	No
Selectivity	Average F at age over last 3 years in 2012 assessment, scaled to mean 3-7.	Yes, small CV
Initial stock numbers	From assessment	According to observed distribution in the final year
Decision basis **	SSB (or TSB) in the TAC year	
Number of iterations	1000, fewer for full feedback runs	
Projection time	30 years	
Observation and implementation models		
If assessment in the loop		
Input data	Catches + 1 surveys (full feedback runs)	Catches and surveys: Log normal, CV from assessment residuals
*** Comparison with ordinary assessment?	No	
Deviations from WG practice?	No	
If no assessment in the loop	*Below is just an example of how this could be presented if there was no assessment in the loop*	

Type of noise	Year factor + age factor on stock numbers at age	Both log-normal + auto-regressive model along year classes Age factor from CV estimates in assessment Year factor adapted to reproduce CV of SSB estimate in assessment
*** Comparison with ordinary assessment?	Year factor scaled to give CV of SSB in year 10 as CV of SSB in assessment	
Projection: If yes - how?	Yes, deterministic with recruitment according to deterministic SR function, assuming TAC as decided, through the intermediate year and the TAC year	
Projection: Deviations from WG practice?	Yes, WG runs a stochastic projection tied to the SAM model	
Implementation	Catches in numbers at age from projection according to the rule.	Log-normally distributed error, CV 10% for each age SOP correct for total catch, no bias.
Harvest rule		
Harvest rule design	Current MP: F-rule with two breakpoints on SSB: If SSB < B1, F = F_{low} If B1<BBB<B2: linear slope between F_{low} and F_{std} If SSB > B2: F = F_{std} Proposed HCR: TAC/HR/F-rule with four breakpoints on SSB/TSB: If SSB < B1, TAC/HR/F = TAC/HR/F_{low} If B1<BBB<B2: linear slope between TAC/HR/F_{low} and TAC/HR/F_{std} If B2<BBB<B3: TAC/HR/F = TAC/HR/F_{std} If B3<BBB<B4: linear slope between TAC/HR/F_{std} and TAC/HR/F_{high} If SSB > B4: TAC/HR/F = TAC/HR/F_{high}	
Stabilizers	Two HCR variants tested: 1. TAC set as average of what it would be for the next three years using the HCR. 2. TAC set as average of previous TAC and HCR advised TAC	
Duration of decisions	Annual	
Revision clause	None (no single MP/HCR advised, work provided list of options)	
Presentation of results		
Interest parameters	Maximum annual probability of SSB<Blim and SSB<$B_{trigger}$, SSB, F, Catch, Inter-annual variation	
**** Risk type and time interval	Type 3	
Precautionary risk level	5%	
Experiences and comments		
Review, acceptance:	Accepted by review group, implemented from 2012 onward. The Blim is provisional, but accepted for the present purpose	
Experiences and comments		

*** Knowledge base: This is the information that will be available about the state of the stock, in particular whether there is an assessment or not. If it is something else, please specify.**
**** Decision basis: This is the measure that determines the exploitation in the harvest rule. For example, SSB at the start of the TAC year, TSB in the last assessment year,.**
***** Comparison with ordinary assessment? This is to indicate whether there has been attempts to verify that the that the performance of the assessment in the model is similar to that experienced by the WG, for example with respect to retrospective problems and inconsistencies.**
****** Risk types:**
Risk1 = *average* probability that SSB is below Blim, where the average is taken across the ny years.
Risk2 = probability that SSB is below Blim *at least once* during the ny years.
Risk3 = *maximum* probability that SSB is below Blim, where the maximum is taken over the ny years.
If your definition of risk does not fit any of these, please explain.

2014 Horse mackerel hom.27.3a4bc7d

ICES. 2014. Evaluation of a multi-annual plan including an index based HCR for North Sea horse mackerel, 17-18 June 2014, IJmuiden, the Netherlands. ICES CM 2014/ACOM:55.

Background		
Motive/ initiative/ background	North Sea horse mackerel was classified by ICES as a data-limited stock (DLS) in category 5 with only landings data available in 2012. The main stakeholder in the fishery - the Dutch Pelagic Freezer-trawler Association (PFA) - sought collaboration with IMARES to develop a multi-annual plan for this stock. ICES was asked by the Netherlands to evaluate the proposed HCR.	
Main objectives	Develop an assessment and MP to ensure sustainable management of the stock, as well as identify how improvement of the knowledge base to underpin scientific advice could be improved	
Formal framework	Request to ICES from the Netherlands	
Who did the evaluation work	IMARES, NL (Aukje Coers and David Miller)	
Method		
Software Name, brief outline include ref. or documentation	The stock assessment models were fit in ADMB (Fournier et al 2012) and the projections were carried out in R (v3. 1; R Development Core Team 2014). All code, data and additional sources for checking, validating and evaluation are freely available upon request.	
Type of stock	(semi-) pelagic	
Knowledge base *	An index of abundance was developed using IBTS data. Exploratory assessments (XSA, custom SCA) were fit to available data.	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Two functions: - Segmented regression with breakpoint fixed at Bloss - Segmented regression and Ricker in equal proportions	Yes, Recruitments are re-sampled from their predictive distribution which is based on parametric models fitted to the timeseries provided
Growth & maturity	Maturity: fixed curve over time	Maturity: No
Natural mortality	Two types: - Fixed value for all ages and years (from western HM) - Fixed over years but varying over ages (from southern HM)	No
Selectivity	Fixed curves from candidate assessment models used.	No
Initial stock numbers	Six different candidate assessments were used to condition the Oms in the reference set (differing in terms of the method used to calculate the biomass index, the weighting applied to each source of data in the model (catch vs index) and the choice of natural mortality)	Yes, MCMC analyses were run to account for uncertainty in model estimates and create alternative valid starting points
Decision basis **	Index value (trend rule utilising an index target value – see below)	

Number of iterations	100	
Projection time	8 years (Given the current poor state of the stock, short-term considerations outweigh potential long-term performance.)	
Observation and implementation models		
Type of noise	Index values generated using the estimated catchability of the index used (q) in the HCR.	Future residuals are sampled form a random lognormal distribution with mean and standard deviation estimated from the estimated model residuals.
*** Comparison with ordinary assessment?	No accepted assessment existed.	
Projection: If yes - how?	No forecast (In order to set a management measure for year y, survey data will up to year y-2 was used, with the assessment itself carried out in year y-1)	
Projection: Deviations from WG practice?	N/A	
Implementation	No implementation error	
Harvest rule		
Harvest rule design	Candidate HCRs follow a simple trend based rule: $TAC_{y+1} = TAC_y \times \left(min\left(\frac{I_{rec}}{I_{trig}},1\right) + \lambda Slp\right)$ Where: TAC = Total Allowable Catch; y = assessment year; λ = slope multiplier; Slp = slope of the log-linear regression for the last x years of the survey index; Irec = recent survey index = average of index values for the last x years up to year y-1; Itrig = survey index trigger value = I2012 The ICES DLS 2:3 rule was also tested, and zero fishing	
Stabilizers	No, incorporated in the lambda parameter of the HCR. Change limits apply for the ICES DLS 2:3 rule	
Duration of decisions	Annual	
Revision clause	N/A	
Presentation of results		
Interest parameters	three potential measures of recovery were examined: • 1. P(SSB50%,2020>SSB50%,2012) : Median SSB in 2020 should be above median SSB in 2012. • 2. P(SSB5%,2020>SSB5%,2012) : The 5th percentile of SSB in 2020 should be above the 5th percentile of SSB in 2012. • 3. P(SSB5%,2020>SSB50%,2012) : The 5th percentile of SSB in 2020 should be above the median SSB in 2012. Landings: • Annual values from 2014–2018 • Average values in the short term (2014–2018) and a ten year medium-term period following the initial three year constant TAC (2017–2026) Spawner stock biomass: • SSB in 2015 (initial target year for FMSY) • SSB 2020 (secondary target year for FMSY) • Median SSB2020/SSB201 (values greater than one indicating recovery). • Probability of recovery of the stock biomass above SSB2012 in 2020 Fishing mortality • F in 2015 (initial target year for FMSY) • F 2020 (secondary target year for FMSY) • F/FMSY, F/F35 (candidate Fmsy values) and F/Fmean (the mean F over the timeseries)	
**** Risk type and time	N/A – alternative measures of 'recovery' comparing SSB 202 with SSB 2012	

interval	were used.
Precautionary risk level	N/A
Experiences and comments	
Review, acceptance:	ICES considered that none of the options examined were in accordance with the precautionary approach given the poor status of the stock. The survey index developed has been used to provide DLS advice since then.
Use of ICES guidelines (WKGMSE 2013)	Not used.
Experiences and comments	A lot of work was put into developing the knowledge base for the stock, but given its very poor condition, any non-zero catches would be unlikely to be precautionary. While no new MP was accpeted on the basis of this work, the knowledge gained about the stock (and the index dveloped) have been useful in its subsequent management.

2014 Redfish - beaked reb.27.1-2

ICES. 2014. Workshop on Redfish Management Plan Evaluation (WKREDMP), 20–25 January 2014, ICES Headquarters. ICES CM 2014/ACOM:52

Background		
Motive/ initiative/ background	The HCR evaluation was initiated by two requests: Firstly, NEAFC requested exploration of possible long-term management plan options for the stock, complying to the precautionary approach and the principle of MSY. This request suggested no specific scenarios. Secondly, the Joint Norwegian-Russian Fisheries Commission (JNRFC) requested evaluation of HCR-rules with F_{MSY} approximated by $F_{0.1}$ as well as ¾ $F_{0.1}$ and $^4/_3$ $F_{0.1}$.and with $B_{trigger}$ at B_{MSY} and ¾ B_{MSY}.	
Main objectives	Establish HCRs for beaked redfish that comply to precautionary and MSY principles and specify actions to be taken if SSB falls below productive levels.	
Formal framework	ICES on requests from JNRFC and NEAFC	
Who did the evaluation work	Bjarte Bogstad, Daniel Howell, Kjell Nedreaas, Benjamin Planque	
Method		
Software Name, brief outline include ref. or documentation	PROST is a software for single-fleet, single-area, long-term stochastic simulations (ICES 2006, Åsnes & Bogstad 2014). It was modified as needed for the features in the HCR scenarios. Each operating model/HCR combination was simulated 10 000 times.	
Type of stock	Long lived (>30 years), demersal/pelagic, valuable	
Knowledge base *	Assessment with statistical catch at age	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Two scenarios, a hockey-stick with a fixed breakpoint at 132 million and a cyclic model based on the recruitment variation in 1992-2012.	Normally distributed error term ε
Growth & maturity	Weight- and Maturity-at-age are modelled values taken from Anon (2009) and are assumed constant. For the 19+ group weight at age was set to 700 g and it was assumed 100% mature.	No
Natural mortality	M=0.05	No
Selectivity	Three age-based selection curves for demersal, pelagic and for the total fleet average 2008-2012. The latter is the default.	No
Initial stock numbers	From latest assessment The multiplier for the total stock compared to the stock accessible by bottom trawl during the ecosystem survey is defined as q=1/3.5.	Uncertainty was introduced into the numbers-at-age for 2013 and was higher for the most recent year classes. CV=0.2 on a log scale for years prior to 2004, 0.3 for 2004-2011 and 0.4 for 2012-2013.
Decision basis **	Probability of SSB falling below 800 kt or 400 kt for the periods 2014–2018, 2014–2023 and 2024–2063. Mean Yield for individual years from 2014 to 2018, as well as the periods	

	2014–2018, 2014–2023 and 2014–2063. Mean realised F and SSB for 2014-2063. Mean Interannual variation in yield for the period 2014–2063.	
Number of iterations	10 000 per HCR scenario	
Projection time	50 years	
Observation and implementation models		
Type of noise	Single "assessment error" term set to CV=0.2 on log scale for all age groups and years.	The assessment error for any given year was uncorrelated between age groups.
*** Comparison with ordinary assessment?	There was no annual assessment run in the projection.	
Projection: If yes - how?	Projection to equilibrium (1000 years)	
Projection: Deviations from WG practice?	• M= randomly sampled 0.03 to 0.07 • Weight-at-age= mean of 1990–2012 • Maturity ogive= mean of 1990–2012 • selectivity= mean of the demersal and the pelagic fleet	
Implementation	No implementation error included	
Harvest rule		
Harvest rule design	$F=F_{0.1}$, $^4/_3*F_{0.1}$, $^3/_4*F_{0.1}$, 0.052, 0.02925, 0 $B_{trigger}/B_{stop}$=800/400, 800/200, 800/0, 400/200, 400/100, 400/0 kt for q=1/3 and q=1/3.5 as well as 800/0 and 400/0 for cyclic recruitment with q=1/3.5	
Stabilizers	Five-year averaging and/or max. 20% annual TAC variation for hockey-stick as well as cyclic recruitment	
Duration of decisions	Annual	
Revision clause	None implicit, but recommendation for revision after pelagic survey in 2016	
Presentation of results		
Interest parameters	Risk, catch, F and SSB for short (2014-2018), medium (2014-2023) and long (2024-2063) terms.	
**** Risk type and time interval	As Risk2 but with $B_{trigger}$ rather than B_{lim}	
Precautionary risk level	5% of SSB to fall below 400 kt in the next 50 years	
Experiences and comments		
Review, acceptance:		
Use of ICES guidelines (WKGMSE 2013)	Guidelines were taken note of. However, only some of the sources of uncertainty in simulations were taken into account but minimum standards for simulations were followed. The MSE used the "short-cut" approach by using an uncertainty level rather than doing full annual assessments within the simulation.	
Experiences and comments		

2015 Mackerel mac.27.nea

ICES. 2015. Report of the EU Workshop on the NEA Mackerel Long-term Management Plan (WKMACLTMP), 24–27 June and 17–19 November 2014, Copenhagen, Denmark. ICES CM 2014/ACOM:63.

Background		
Motive/ initiative/ background	COASTAL STATES REQUEST TO ICES ON THE LONG TERM MANAGEMENT PLAN FOR MACKEREL In order for the Parties to develop a revised management plan for mackerel on which to base the appropriate fishing levels in the years 2015 to 2018, ICES is requested by September 2014 to: 1. Evaluate new biological reference points for the North East Atlantic mackerel stock based on the revised (WKPELA 2014) mackerel assessment method. 2. Evaluate the alternative fishing mortalities corresponding to Fmsy, 0.20, 0.25, 0.30 and 0.35 for appropriate age groups as defined by ICES. 3. Each alternative should be assessed in relation to how it performs with respect to stock development in the short, medium and the long term and the level of uncertainty in the stock assessment, inter annual TAC variability, long term yield, as well as in relation to the precautionary approach. Each alternative shall be evaluated with an annual quota flexibility of 10%. Each alternative shall also be assessed with a stability clause where the TAC shall not deviate by more than 20% from the TAC of the preceding year, but the F shall not deviate by more than 10% from the target F.	
Main objectives	Estimation of reference points Evaluation of managements strategies (range of management reference points, with and without TAC and F variation limit and annual quota flexibility	
Formal framework	EU Workshop on the NEA Mackerel Long-term Management Plan 2014 (WKMACLTMP)	
Who did the evaluation work	Thomas Brunel, Einar Hjörleifsson, Agurtzane Urtizberea	
Method		
Software Name, brief outline include ref. or documentation	Simulation tool developped in R/FLR	
Type of stock	pelagic, straddling, very valuable	
Knowledge base *	Stock assessment from WGWIDE2014 (SAM model)	
Type of regulation	TAC.	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Based on resampled SR pairs (1000 replicates) for the period 1990-2012 Combination of SR models (Beverton	Modelled using log-normal distributed deviations, folowing a AR1 process

	and Holt 22%; Segmented regression 33%; and Ricker 45%) with probabilities of the three model forms were calculated based on the like-lihoods of the three sets of 1000 model fits, as described in Simmonds et al. (2011).	Model parameters (shape, variance, autocorrelation) estimated for each SR pair (i.e. replicate specific).
Growth & maturity	Both growth and maturity showed strong trend in the recent years. 2 scenarios were modelled in the simulation: - Permanent changes : future mean of W@age and pmat@age is taken as mean of the last 3 years - Reversble : future mean (after a transition period) egual to long term mean in the historical data	Modelled as AR1 processes with mean derived as explained and variance and autocorrelation estimated over the entire historical period
	Same procedure applied for proportion of M and F before spawning	
Natural mortality	Same as assessment (0.15)	Constant and age invariant
Selectivity	Resampling blocks of year in the past (iteration specific), revert the years or not, and use it in the future	
Initial stock numbers	Function monteCarloStock (FLSAM library)	Resamples a set of parameters and states from the variance – covariance matrix from the SAM assessment
Decision basis **	SSB at spawning time	
Number of iterations	1000	
Projection time	40 years	
Observation and implementation models		
Type of noise	Short cut approach (applying annual deviations on the N and F from the operating model)	Log normally distributed deviations with : - One component (age and year specific) with a variance based on the assessment CV - A second component (assessment year specific) modelled as a AR1 process with variance and autocorrelation fixed ad hoc in order to obtain the overall level of variance and autorcorrelation in the assessment errors on SSB and Fbar in the advice year Those levels were estimated based on a retrospective performance of an assessment conducted with a separable model, and not the official SAM assessment. reasons

		for that were : - Impossible to conduct a retrospective analysis with the SAM assessment at the time - It was believed that part of the SAM uncertainty was in the process error and that retrospecive from SAM give a too optimistic view of assessment performance
*** Comparison with ordinary assessment?	Simulations were run imposing a time series of future Fbar (generated using fourrier surrogates method, i.e. this similiar characristics as historic Fbar). Levels of uncertainty and autorocrrelation in SSB and Fbar for the advice year were compared to desired levels	
Projection: If yes - how?	Short term forecast implemented as in the WGWIDE	
Projection: Deviations from WG practice?	none	
Implementation	TAC implemented without error or bias	
Harvest rule		
Harvest rule design	Hockey stick shape with Fbar=Ftarget if SSB advice year forecasted to be above Btrigger, and Fbar reduced proportionnally to SSB otherwise Also tested (no requested) a harvest rate rule	
Stabilizers	Variation of TAC from 1 year to the next <20%, Deviation of Fbar from Ftarget <10%	
Duration of decisions	Annual	
Revision clause	none	
Presentation of results		
Interest parameters	Short term (2014-2018), medium term (2019-2028) and long term (2028-2052): • Average SSB - Mean Fbar - Mean individual weight (mean of W@age weight by N@age) • Average yield • Indicator for year to year variability in TAC • Risk of SSB falling below Blim (type 1 ,2 ,3)	

**** Risk type and time interval	Long term risk 3
Precautionary risk level	5% of risk type 3.
Experiences and comments	
Review, acceptance:	Reviewer attended all meetings and input was received throughout the process
Use of ICES guidelines (WKGMSE 2013)	Definition of risk Choise of the appropriate number of iteration Assumption to be made for the future mean of biological vectors
Experiences and comments	<u>MSE on SAM assessment (short cut approach):</u> How to deal with the process error? Can we use the actual CV on Ns and Fs for future assessment errors, or is part of the uncertainty actually also in the process error of the historical assessment? should we model the process error in the operating model? <u>Trends in growth and maturation:</u> A density dependent growth model (von Bertalanffy based) was developped to model the link between future growth and future stock size. However it was felt that there was not enough biological evidence for density dependence (and lack of research on other drivers of growth), and this model was not used in the end. <u>Lack of interaction with the stakeholder :</u> It would have been useful to have a formal process to interact with the stakeholders. Some desicions on small details of the implementation of the MP were made assuming what was meant by the manageres (e.g. what to do below Blim, when constraints on TAC variation should be lifted and applied again)...

2016 Capelin cap.27.1-2

ICES. 2016. Report of the second Workshop on Management Plan Evaluation on Northeast Arctic cod and haddock and Barents Sea capelin (WKNEAMP-2), 25–28 January 2016, Kirkenes, Norway. ICES CM 2016/ACOM:47. 76 pp.

Background		
Motive/ initiative/ background	Managers (Joint Norwegian-Russian Fisheries Commission, JNRFC) at its 45th session in October 2015, suggested a number of alternative harvest control rules (HCRs) for Barents Sea. The HCRs were evaluated in 2016 and of the rules evaluated only the rule already in use was found to be precautionary.	
Main objectives	Precautionary	
Formal framework	ICES on request from JNRFC	
Who did the evaluation work	WKNEAMP-2	
Method		
Software Name, brief outline include ref. or documentation	Assessment model used: CapTool (Excel spreadsheet using @risk add-on). Simulates stock development 6 months ahead (October-April) taking into account uncertainty in initial stock estimate (CV=0.2) as well as in maturation and natural mortality. Changing the risk levels as indicated in the request (to 90, 85 or 80%) means that the resulting HCR is not precautionary in the ICES sense if the B_{lim} of 200 000 tonnes is maintained. Thus, there was not a need to make an operating model to evaluate those HCRs, but some considerations made during WKNEAMP-2 are given below: Changing the risk criterion is equivalent to change the B_{lim} and maintain the 5% probability criterion. Changing the risk and maintaining B_{lim} suggests that the reproduction dynamics of the capelin stock is unchanged while changing B_{lim} should be based on information on capelin reproduction dynamics. WKNEAMP-2 concluded that there is no basis on which to revise the B_{lim} value. However, there are large uncertainties in the calculation of SSB, many historical SSB values are very low and further research on stock-recruitment relationships and reference points is required.	
Type of stock	Short-lived, pelagic, moderately valuable, semelparous	
Knowledge base *	Assessment based on annual survey which is considered an absolute estimate. Length-based maturation model used to split out maturing part. Stock predicted 6 months ahead (survey time-spawning time) taking into account predation by cod estimated from stomach sampling	
Type of regulation	TAC. Fishery only on mature (pre-spawning) capelin in January-March.	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment		
Growth & maturity		
Natural mortality		
Selectivity		

Initial stock numbers		
Decision basis **	SSB at spawning time (1 April)	
Number of iterations		
Projection time		
Observation and implementation models		
Type of noise		
*** Comparison with ordinary assessment?		
Projection: If yes - how?		
Projection: Deviations from WG practice?		
Implementation		
Harvest rule		
Harvest rule design	TAC set so that there is 95% probability of SSB > 200 000 t (target escapement strategy)	
Stabilizers	None	
Duration of decisions	Annual	
Revision clause	No clause, but revision asked for by 2021 by JNRFC	
Presentation of results		
Interest parameters		
**** Risk type and time interval	See definitions below	
Precautionary risk level	e.g. 5% of risk type 3.	
Experiences and comments		
Review, acceptance:		
Use of ICES guidelines (WKGMSE 2013)	Please indicate whether or not you specifically took note of the guidelines in the WKGMSE 2013 report. If so, - Were they useful?, - Where did you deviated from them and, if so, how did you deviate and why?	
Experiences and comments		

2016 Cod cod.27.1-2

ICES. 2016. Report of the second Workshop on Management Plan Evaluation on Northeast Arctic cod and haddock and Barents Sea capelin (WKNEAMP-2), 25–28 January 2016, Kirkenes, Norway. ICES CM 2016/ACOM:47. 76 pp.

Background		
Motive/ initiative/ background	Managers (Joint Norwegian-Russian Fisheries Commission, JNRFC) at its 45th session in October 2015, suggested a number of alternative harvest control rules (HCRs) for Northeast Arctic cod. The HCRs were evaluated in 2016 and found to be precautionary. One of them was chosen at the 2016 JNRFC meeting.	
Main objectives	Precautionary, stable catches near MSY	
Formal framework	ICES on request from JNRFC	
Who did the evaluation work	WKNEAMP-2	
Method		
Software Name, brief outline include ref. or documentation	All ad-hoc software, writen in Excel's Visual Basic. The model is similar to the previously used software, PROST, and is called new Prost or NE_PROST . It is realized in Excel. Excel sheets are used as source of input data and to print out results of calculations from simulation models. Program code is realized as macros written in Visual Basic. The program is open for reading and changing. Some Excel sheets are used to calculate all processes in a "traditional" way by Excel formulas to check if the program calculates things correctly. NE_PROST software is available on the WKNEAMP-2 SharePoint site.	
Type of stock	Medium life span, demersal, very valuable	
Knowledge base *	Analytic assessment (SAM), 4 surveys. Cannibalism included in assessment.	
Type of regulation	TAC. Calculated using the harvest control rule and assuming F status quo (F in last data year) in the intermediate year	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Hockey-stick with cyclic term, fitted to S/R time series for year-classes 1946-2014 Log-normal, CV from residuals	Hockey-stick with cyclic term, fitted to S/R time series for year-classes 1946-2001 Log-normal, CV from residuals
Growth & maturity	Weight in stock a function of total stock biomass in previous years for ages 6-13, with upper and lower limits Weight in catch a function of weight in stock for ages 3-8, WEST=WECA for ages 9-13+ Maturity at age a function of total stock biomass in previous years with upper and lower limits	No
Natural mortality	Constant M1 + Cannibalism M2 for ages 3-5 as a function of abundance of R (N at age 3) and predators (SB 6+)	No
Selectivity	Constant, average for recent period	No
Initial stock numbers	From assessment	No

Decision basis **	SSB in the TAC year
Number of iterations	10 000 simulations
Projection time	for 100 years
Observation and implementation models	
Type of noise	Implementation errors as Log normal distribution derived from historical data, Assessment errors as normal distribution. Age dependent, no correlation between age groups.
*** Comparison with ordinary assessment?	No
Projection: If yes - how?	Yes, deterministic with "known" recruitment. A TAC constraint is applied in the intermediate year. Assessment errors simulated as normaly distributed.
Projection: Deviations from WG practice?	A TAC constraint is applied in the intermediate year, this is not consistent with assessment procedure which assumes F status quo in intermediate year.
Implementation	Catch at age calculated from the perceived stock using the fishing mortality derived from the harvest control rule and the given exploitation pattern. This catch at age is then applied to the actual stock
Harvest rule	
Harvest rule design	The target level of exploitation is calculated according to the spawning-stock biomass (SSB) in the first year of the forecast as follows: - if SSB < Bpa, then Ftr = SSB / Bpa × FMSY; - if Bpa ≤ SSB ≤ 2×Bpa, then Ftr = FMSY; - if 2 × Bpa < SSB < 3 × Bpa, then Ftr = FMSY × (1 + 0.5 × (SSB – 2 × Bpa) / Bpa); - if SSB ≥ 3 × Bpa, then Ftr = 1.5 × FMSY; where FMSY = 0.40 and Bpa = 460 000 tonnes.
Stabilizers	If the spawning-stock biomass in the present year, the previous year, and each of the three years of prediction is above B_{pa}, the TAC should not be changed by more than ±20% compared with the previous year's TAC. In this case, F_{tr} should however not be below 0.30.
Duration of decisions	Annual
Revision clause	No clause, but revision asked for by 2021 by JNRFC
Presentation of results	
Interest parameters	Risk, Catch, Inter-annual catch variation, proportion of year when different parts of the HCRs apply, stock size
**** Risk type and time interval	interval Type 1 for years 21-100, to avoid initial transients. Type 3 tested and concluded to be not appropriate for this case.
Precautionary risk level	5% of risk type 1.
Experiences and comments	
Review, acceptance:	Accepted by ICES in 2016
Use of ICES guidelines (WKGMSE 2013)	The guidelines in the WKGMSE 2013 report were used. It was useful. Probability 3 should be used in accordance to guide but it should be mentioned that this type of criteria is dependent on number of years used in the simulation model. The more years used the higher value of Prob 3 parameters we get. It is problematic to base the conclusion on this criterion. The more appropriate criterion is Prob 1.
Experiences and comments	

2016 Cod cod.27.22-24

ICES. 2016. Annex 11: Report of the Baltic Fisheries Assessment Working Group (WGBFAS), 12–19 April 2016, ICES Headquarters, Copenhagen, Denmark. ICES CM 2016/ACOM:11.

Background		
Motive/ initiative/ background	The EU Commission was concerned that the biomass at the beginning of 2019 that remains (substantially) below Blim will risk continuing to harm the stock e.g. through impaired recruitment, and the request should clarify the extent to which such a risk exists, even when a small increase in biomass is predicted.	
Main objectives	Forecast the likely medium-term consequences of reductions in fishing mortality rates of the commercial and/or the recreational fisheries.	
Formal framework	Request to ICES from EU	
Who did the evaluation work	Ad hoc group (mainly DTU Aqua) for preparing the medium term forecasts.	
Method		
Software Name, brief outline include ref. or documentation	A two-fleet software tool was developed in R to conduct the requested stochastic projections. The two fleets represent the commercial and the recreational fisheries. Each fleet has its own exploitation pattern and catch mean weight-at-age, both of which are assumed to remain constant throughout the projection period.	
Type of stock	Long life span	
Knowledge base *	Analytic assessment (Cat 1)	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Recruitment were drawn stochastically from the recruitment estimates for the most recent 11 years, assuming linearly reduced recruitment when SSB is below Bloss	Yes
Growth & maturity	Constant, derived from average historical values	
Natural mortality	As above	
Selectivity	Constant, by fleet	
Initial stock numbers	Assessment results	Yes, derived from terminal N and uncertainties (log-normal dist errors on N)
Decision basis **	25 scenarios specified by the request. Mix of F reductions until F_{MSY} with the same and different reduction by the two fleets.	
Number of iterations	1000	
Projection time	Medium term forecast, 2017-2027	
Observation and implementation models		
Type of noise	Observation and implementation noise	

	are not taken into account	
*** Comparison with ordinary assessment?	no	
Projection: If yes - how?	Two-fleet (Commercial and recreational fisheries) projections	
Projection: Deviations from WG practice?	Yes	
Implementation		
Harvest rule		
Harvest rule design	25 scenarios specified by the request. Mix of F reductions until FMSY with the same or different reduction by the two fleets.	
Stabilizers		
Duration of decisions		
Revision clause		
Presentation of results		
Interest parameters	Short to medium term (2017-2027) • SSB (5th, 25th, 50th, 75th, and 95th percentiles) • F by fleet • Yield by fleet (5th, 25th, 50th, 75th, and 95th percentiles) • Recruitment (5th, 25th, 50th, 75th, and 95th percentiles)	
**** Risk type and time interval	Risk by year from 5th percentile of SSB	
Precautionary risk level		
Experiences and comments		
Review, acceptance:		
Use of ICES guidelines (WKGMSE 2013)	This request is for likely rebuilding scenarios given the very low SSB and recent low recruitments and is as such not a traditional MSE. It was answered by a medium-term two-fleet (commercial and recreational fisheries) forecast with noise on the initial stock numbers and recruitments in all forecast years.	
Experiences and comments	With the assumption that recruitments will remain low (as observed in 2006-2016) it was shown that rebuilding would take long time and require drastic cut in F. The very high recruitment in 2017 was not anticipated (actually higher than the 95th percentile of the assumed recruitments) in the forecast. ICES now (2018) considers that SSB will be above MSY Btrigger in 2020.	

2016 Haddock had.27.1-2

ICES. 2016. Report of the second Workshop on Management Plan Evaluation on Northeast Arctic cod and haddock and Barents Sea capelin (WKNEAMP-2), 25–28 January 2016, Kirkenes, Norway. ICES CM 2016/ACOM:47. 76 pp.

Background		
Motive/ initiative/ background	Managers (Joint Norwegian-Russian Fisheries Commission, JNRFC) at its 45th session in October 2015, suggested a number of alternative harvest control rules (HCRs) for Northeast Arctic haddock. The HCRs was evaluated in 2016 and found to be precautionary. One of them were choosen at the next Commissions meeting for practical using.	
Main objectives	Precautionary, stable catches near MSY	
Formal framework	ICES on request from JNRFC	
Who did the evaluation work	WKNEAMP-2	
Method		
Software Name, brief outline include ref. or documentation	Al ad-hoc software, writen in Excel's Visual Basic. The model is similar to the previously used software, PROST, and is called new Prost or NE_PROST . It is realized in Excel. Excel sheets are used as source of input data and to print out results of calculations from simulation models. Program code is realized as macros written in Visual Basic. The program is open for reading and changing. Some Excel sheets are used to calculate all processes in a "traditional" way by Excel formulas to check if the program calculates things correctly. NE_PROST software is available on the WKNEAMP-2 SharePoint site.	
Type of stock	Medium life span, demersal, very valuable	
Knowledge base *	Analytic assessment (SAM), 3 surveys. Consumption from cod included in assessment.	
Type of regulation	TAC. Calculated using the harvest control rule and assuming F status quo (F in last data year) in the intermediate year	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Segmented regression Ry=f(SSBy-3)+ε with autocorrelation, fitted to S/R time series for year-classes 1950-2011 with cap limit	Log-normal, Autocorrelation, Residuals drawn from obser-vations or parametric +/- 2*sigma
Growth & maturity	Weight in stock a function of total stock biomass in previous years for ages 3-11+, with upper and lower limits Weight in catch a function of weight in stock for ages 3-8, WEST=WECA for ages 9-13+ Maturity at age a function of weight at age in stock the same years with upper	No

	and lower limits	
Natural mortality	Constant (Mean 1984-2014 M2 for ages 3-6) 0.2 for ages 7+	No
Selectivity	Constant, mean for 1995-2004	No
Initial stock numbers	From assessment	No
Decision basis **	SSB in the TAC year	
Number of iterations	5 000 simulations	
Projection time	for 100 years	
Observation and implementation models		
Type of noise	Implementation errors as Log normal distribution derived from historical data, Assessment errors as normal distribution. with CV = 0.25 Age dependent, no correlation between age groups.	
*** Comparison with ordinary assessment?	No	
Projection: If yes - how?	Yes, deterministic with “known” recruitment. A TAC is applied in the intermediate year. Assessment errors simulated as normaly distrinuted.	
Projection: Deviations from WG practice?	A TAC constraint is applied in the intermediate year, this is consistent with assessment procedure.	
Implementation	Catch at age calculated from the perceived stock using the fishing mortality derived from the harvest control rule and the given exploitation pattern. This catch at age is then applied to the actual stock	
Harvest rule		
Harvest rule design	If the spawning stock falls below Bpa TAC for the next year will be set at level corresponding to F_{msy}=0.35 If the spawning stock falls below Bpa, the procedure for establishing TAC should be based on a fishing mortality that is linearly reduced from Fmsy at Bpa to F= 0 at SSB equal to zero. At SSB-levels below Bpa in any of the operational years (current year and a year ahead) there should be no limitations on the year-to-year variations in TAC.	
Stabilizers	The TAC should not be changed by more than +/- 25% compared with the previous year TAC.	
Duration of decisions	Annual	
Revision clause	No clause, but revision asked for by 2021 by JNRFC	
Presentation of results		
Interest parameters	Risk, Catch, Inter-annual catch variation, proportion of year when different parts of the HCRs apply, stock size	
**** Risk type and time interval	interval Type 1 for years 21-100, to avoid initial transients. Type 3 tested and concluded to be not appropriate for this case.	
Precautionary risk level	5% of risk type 1.	
Experiences and comments		
Review, acceptance:	Accepted by ICES in 2016	
Use of ICES guidelines	The guidelines in the WKGMSE 2013 report were used. It was useful.	

(WKGMSE 2013)	Probability 3 should be used in accordance to guide but it should be mentioned that this type of criteria is dependent on number of years used in the simulation model. The more years used the higher value of Prob 3 parameters we get. It is problematic to base the conclusion on this criterion. The more appropriate criterion is Prob 1.
Experiences and comments	

2016 Blue whiting cap.27.1-2

ICES. 2016. Report of the Workshop on Blue Whiting (Micromesistius poutassou) Long Term Management Strategy Evaluation (WKBWMS), 30 August 2016, ICES HQ, Copenhagen, Denmark. ICES CM 2016/ACOM:53. 104 pp.

Background		
Motive/ initiaitve/ background.	No agreed management plan has been in place for blue whiting in recent years. NEAFC requested ICES to evaluate a proposed long term manangement strategy.	
Main objectives	Precautionary (high likelihood of maintaining the stock above Blim), stable high catches.	
Formal framework	ICES on request from NEAFC	
Who did the evaluation work	WKBWMSE 2016	
Method		
Software Name, brief outline include ref. or documentation	Conditioned on an assessment model done using SAM on stockassessment.org Two simualtion models were used: An adapted veriosn of the EQSIM software R and FLR (SimpSIM) HCS (Harvest Control rule Simulator) software v15.1 (http://www.dwsk.net/). Both are age structured operating models, with catches at age, assessment error in cluded in numbers at age (HCS) or as advice error on SSB and F (SimpSIM) Code available on request.	
Type of stock	Medium life span, pelagic, low value	
Knowledge base *	Analytic assessment	
Type of regulation	TAC, in recent years unilateral quotas have been set (i.e. no agreed overall TAC).	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Segmented regression with Log-normal, CV from residuals, in HCS spasmodic a breakpoint at Bloss. Two peaks from average recruitment scenarios considered (high and low)	
Growth & maturity	Average over last five years	No
Natural mortality	Constant estimate (0.2)	No
Selectivity	Average F at age over last 5 years in IBPBLW 2016 assessment, scaled to mean 3-7.	No
Initial stock numbers	From assessment	HCS: according to observed distribution in the final year and estimate advice uncertainty SimpSIM: No
Decision basis **	SSB in the TAC year	
Number of iterations	1000	
Projection time	100 years	
Observation and implementation models		
If assessment in the loop		
Input data	N/A	
*** Comparison with ordinary assessment?	N/A	
Deviations from WG practice?	No	
If no assessment in the loop	Below is just an example of how this could be presented if there was no assessment in the loop	
Type of noise	SimpSIM: historic advice error on F and SSB HCS: Year factor + age factor on stock numbers at age	HCS: Both log-normal + auto-regressive model along year classes, age factor from CV estimates in assessment, year factor adapted to reproduce CV of SSB advice
*** Comparison with ordinary assessment?	Advice uncertianty is larger than model parametric uncertainty. The histroic advice error for the last ten years was used.	

Projection: If yes - how?	SimpSIM: N/A. Advice error was estimated for the forecast, nto the assessment year. HCS: Yes, deterministic with recruitment according to deterministic SR function, assuming TAC as decided, through the intermediate year and the TAC year
Projection: Deviations from WG practice?	Yes, WG runs a stochastic projection tied to the SAM model (negligible difference)
Implementation	Catches in numbers at age from projection according to the rule. No
Harvest rule	
Harvest rule design	Proposed HCR: F-rule with two breakpoints on SSB: If SSB < Blim, F = 0.05 If Blim<SSB<Bpa: linear slope between 0.05 and Fmsy(=0.32) If SSB > Bpa: F = FMSY
Stabilizers	20% TAC change limits (only applied above Bpa in the HCR, but scenarios of above Blim, never or always were also run)
Duration of decisions	Annual
Revision clause	LTMP specifies a revision in no more than 5 years time
Presentation of results	
Interest parameters	Maximum annual probability of SSB<Blim, Catch, Inter-annual variation
**** Risk type and time interval	Type 3
Precautionary risk level	5%
Experiences and comments	
Review, acceptance:	Unknown at time of evaluation.
Experiences and comments	Running two simulation models was useful in error-checking each of them Model results were not identical, but agree on key considerations (i.e. P(SSB<Blim)<5%)

*** Knowledge base: This is the information that will be available about the state of the stock, in particular whether there is an assessment or not. If it is something else, please specify.**

**** Decision basis: This is the measure that determines the exploitation in the harvest rule. For example, SSB at the start of the TAC year, TSB in the last assessment year,.**

***** Comparison with ordinary assessment? This is to indicate whether there has been attempts to verify that the that the performance of the assessment in the model is similar to that experienced by the WG, for example with respect to retrospective problems and inconsistencies.**

****** Risk types:**

Risk1 = *average* probability that SSB is below Blim, where the average is taken across the ny years.

Risk2 = probability that SSB is below Blim *at least once* during the ny years.

Risk3 = *maximum* probability that SSB is below Blim, where the maximum is taken over the ny years.

If your definition of risk does not fit any of these, please explain.

2017 Herring her.27.5a

ICES. 2017. Report of the Workshop on Evaluation of the Adopted Harvest Control Rules for Icelandic Summer-Spawning Herring, Ling and Tusk (WKICEMSE), 21–25 April 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:45. 49 pp.

Background		
Motive/ initiative/ background	Request to ICES from managers (Government of Iceland) because of adopting of a management plan for the stock.	
Main objectives	Evaluate several management strategies for Icelandic summer spawning herring to maintain the exploitation rate at the rate which is consistent with the precautionary approach and that generates maximum sustainable yield (MSY) in the long term.	
Formal framework	ICES on request from the Icelandic Fishery Ministry	
Who did the evaluation work	WKICEMSE	
Method		
Software Name, brief outline include ref. or documentation	The model is called **muppet**, and is a statistical catch-at-age model written by Höskuldur Björnsson and has also been used for MSE of Icelandic cod, Icelandic haddock and NEA-mackerel. Details about the software can be found here: Björnsson, H. 2016. Working document on assessment model for Norwegian Spring Spawning Herring. WD 13 to Benchmark Workshop on Pelagic stocks (WKPELA), 29 February–4 March 2016, ICES Headquarters, Copenhagen, Denmark. Software found at https://github.com/hoski/Muppet_HCR	
Type of stock	Long life span (age 4-15 typically in the fishery), pelagic, distribution limited to the Icelandic continental shelf, very valuable	
Knowledge base *	Category 1 stock, with analytical assessment where the main input data are catch-at-age and age disaggregate acoustic abundance indices. The stock was at high levels until around late 2000s but since then a substantial reduction has taken place despite a low fishing mortality. The reduction is consequence of mortality induced by Ichthyophonus outbreak in the stock in 2009-2011 and 2016-2018 in addition to small year classes entering the stock since around 2005, particularly the 2011-2014 year classes. Estimates of infection mortality are used in the stock assessment, as well as in the MSE.	
Type of regulation	Until 2016, the TAC was given on basis of an Fmsy but from 2017 and onwards on basis of a management plan.	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	A stock–recruitment relationship for this herring stock is relatively well defined. Future recruitment in the MSE is simulated by a hockey-stick stock–recruitment function with random annual deviations.	Future recruitment was based on a CV of approximately 0.5

Growth & maturity	Weight-at-age are derived from catch samples, while maturity ogive is fixed. Density-dependence was not noticed in the data and, therefore, not implemented in the simulations.	Average weights of last 20 years are multiplied by a lognormal year factor with σ=0.1 and ρ=0.7.
Natural mortality	Fixed at 0.10, but additional age dependent mortality is added for some years because of the *Ichthyophonus* infection -and the effects evaluated. The final simulation assumed a additional infection mortality in 3 years out of 20, on average, and resulted in average M of 0.124.	No, execept during future icthyophonus epidemics when infection M estimated from the 2009 – 2011 epidemic is added. This value is estimated so different value is used in each iteration.
Selectivity	The selectivity is estimated by the model but the HCRs tested were based on biomass in the beginning of the assessment year so no selection pattern needs to be specified for those HCRs to generate an advice.	No
Initial stock numbers	In coherence with the analytical assessment providing the advice, the simulations were based on assessment with catch-at-age matrix going back to 1947 and tuned with acoustical survey data on number-at-age going back to 1987. These gave the initial stock numbers, which were similat to what the regularly used assessment tool (NFT-Adapt) gave.	MCMC simulations from the assessment model.
Decision basis **	Biomass of age-4+ in the beginning of the assessment year (1 January) while the fishing year starts 1st September the same year.	
Number of iterations	2 million mcmc simulations saving every 1000 simulation leading to 2000 parameters sets for stochastic simulations for each of the five HCRs tested	
Projection time	2016-2085	
Observation and implementation models		
Type of noise	Assessment error was added to the simulations on basis of formal examination. The total bias was set 0.15 on log scale, which is 18% on ordinary scale. In addition to the bias, a stochastic lognormal error with rho=0.7 and σ=0.25 was used. The lognormal error is unbiased on log scale but has a bias of $e^{0.5\times0.25(^\wedge2)}$ =1.03 on ordinary scale.	Starting value of the assessment error for each iteration is the ratio of the SSV in the starting year and SSB for each simulation, overestimation for replicas with low SSBand vice versa.
*** Comparison with ordinary assessment?	The performance of the model used in the MSE gave very similar stock trends as the model used by the WG (NFT-adapt). Moreover, the retrospective pattern was more or less the same.	
Projection: If yes - how?	Advice based on biomass in the beginning of the assessment year so no projection is needed.	
Projection: Deviations from WG practice?		
Implementation	No implementation error included	

Harvest rule	
Harvest rule design	The HCR tested and adopted by the Icelandic government, is to fish 15% of biomass of age 4+ as estimated in the beginning of the assessment year, where MGT B trigger is 200 thousands tonnes.
Stabilizers	None
Duration of decisions	Annual
Revision clause	There is no clause for when the MP should be revised, beyond the normal period of ~5 years. It was mentioned in the report that changes in development or knowledge on the *Ichthyophonus* outbreak in the stock might call for revision.
Presentation of results	
Interest parameters	Short term (2017-2021), medium term (2022-2031) and long term (2032-2060): • Average SSB • Average yield (median) • Indicator for year to year variability in yield (standard dev.) • Risk of SSB falling below Blim Risk in the short term turned out to be highest.
**** Risk type and time interval	The constrain was HCRs with P(SSB(y) < Blim) ≤ 5%, on an annual basis, for all years in the short, medium and long terms (i.e. Risk type 3).
Precautionary risk level	e.g. 5% of risk type 3.
Experiences and comments	
Review, acceptance:	Accepted by ICES in 2017.
Use of ICES guidelines (WKGMSE 2013)	The guidelines were used where needed, and they came us through this work. The issue with the *Ichthyophonus* infection in the stock required some unconventional action, obviously not included in the guidelines.
Experiences and comments	

2017 Ling lin.27.5a

ICES. 2017. Report of the Workshop on Evaluation of the Adopted Harvest Control Rules for Icelandic Summer-Spawning Herring, Ling and Tusk (WKICEMSE), 21–25 April 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:45. 49 pp.

Background		
Motive/ initiative/ background	On 22 December 2016, ICES received the following request from Iceland: *The Government of Iceland is in the process of formally adopting management plans for Icelandic summer spawning herring (5a), ling (5a) and tusk (5a14):The management strategy for Icelandic summer spawning herring, ling and tusk is to maintain the exploitation rate at the rate which is consistent with the precautionary approach and that generates maximum sustainable yield (MSY) in the long term.* *A part of the management plan is the adoption of harvest control rules (HCR) for the three stocks for setting annual total allowable catch (TAC). The HCR adopted should be precautionary and in accordance with the ICES MSY approach.*	
Main objectives	Further in the request: *The Government of Iceland requests ICES to evaluate whether these harvest control rules are in accordance with its objectives, given current ICES definition of reference points or any re-evaluation of those points that may occur in the process. For ling and tusk the evaluation should also include review of input data and the applied assessment methodology (Benchmark).*	
Formal framework	ICES on request from Iceland	
Who did the evaluation work	An ad hoc group of managers, stakeholders, and scientists from the Marine and Freshwater Research Institute (MFRI), initiated by the Icelandic Ministry of Industries and Innovation in the summer of 2016 (mainly coordinated by B. Elvarsson).	
Method		
Software Name, brief outline include ref. or documentation	Gadget is a shorthand for the "Globally applicable Area Disaggregated General Ecosystem Toolbox", which is a statistical model of marine ecosystems (previously known as BORMICON (Stefánsson and Pálsson 1997) and Fleksibest Frøysa et al. (2002)). Gadget is an age-length structured forward-simulation modelling framework, where models can be coupled with an extensive set of data comparison and optimisation routines. Processes are generally modelled as dependent on length, but age is tracked in the models, and data can be compared on either a length and/or age scale. The framework allows for the creation of multi-area, multi-fleet models, capable of including predation and mixed fisheries issues, however it can also be used on a single species basis. Gadget models can be both very data- and computationally- intensive, with optimisation in particular taking a large amount of time.	
Type of stock	Demersal and has a medium life span and medium. The ling stock is managed as a single stock only within Iceland (5a).	
Knowledge base *	The request from Iceland to evaluate the harvest control rule also included benchmarks for the species, which were approved as category 1 analytical stock assessments with indices from an annual surveys, as well as length, age, and maturity distributions from survey and commercial samples. Catches were removed directly.	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)

Recruitment	Hockey-stick, series estimated during base model conditioning and breakpoint set to Bloss = Bpa (lowest observed SSB)	5-year block bootstrap
Growth & maturity	Von Bertalanffy growth and logistic maturity function, estimated during base model conditioning. Weights were generated from a fixed length-weight relationship estimated from all years external to the base model.	Spatial bootstrap used to generate error in estimated base model parameters. Refer to Elvarsson et al. (2014) and Lentin (2017) for further implementation details.
Natural mortality	Set to 0.15	No
Selectivity	Separate for survey and commercial stocks, logistic or dome-shaped.	Spatial bootstrap used to generate error in estimated base model parameters. Refer to Elvarsson et al. (2014) and Lentin (2017) for further implementation details.
Initial stock numbers	Estimated as parameters	Spatial bootstrap used to generate error in estimated base model parameters. Refer to Elvarsson et al. (2014,2018) and Lentin (2017) for further implementation details.
Decision basis **	A reference biomass estimated from the last assessment year (75+ cm for ling) is multiplied by a harvest rate to generate the following year's TAC, as long as SSB does not fall below Btrig, which is set to Bpa.	
Number of iterations	101,000	
Projection time	300 years	
Observation and implementation models		
Type of noise	Empirical	Spatial bootstrap used to generate error in estimated base model parameters. Refer to Elvarsson et al. (2014) and Lentin (2017) for further implementation details.
*** Comparison with ordinary assessment?	Because the benchmark and MSE were conducted simultaneously, there are very few years to compare with an ordinary assessment based on that benchmark. In comparison to the previous models used for ling, there was a consistent downward revision in biomass with consecutive years in retrospective plots. This pattern was likely a result of discounting a few earlier years with high recruitment and the importance of age data that are more available and informative toward the end of the time series. However, the state of the stock was not altered greatly: changes were within bootstrap interquantile ranges of the MSE.	
Projection: If yes - how?	Yes. Deterministic filling of the current year's TAC (2 quarters), followed by stochastic projections. The spatial bootstrap and recruitment block bootstrap generated stochasticity in population dynamics. Assessment error and autocorrelation (CV = 0.2, rho = 0.8) were included. TACs were filled exactly.	
Projection: Deviations from WG practice?	Likely not.	
Implementation	A range of 101 harvest rates were tested (0 – 1 in steps of 0.01) with 1000 stochastic projections each.	For each harvest rate, 100 bootstrap replicates were used, each with 10 runs containing recruitment and assessment error.

Harvest rule	
Harvest rule design	The TAC in year y is the reference biomass (75+ cm) observed in the previous assessment year (y-1) multiplied by a harvest rate scalar. The scalar was proposed to be 0.18, which was lower than that obtained at MSY but had the dual benefit of yielding catches close to MSY and maintaining high biomass levels. Btrigger was set to Bpa, thereby scaling the harvest rate with biomass level relative to Btrigger under this level.
Stabilizers	None
Duration of decisions	Annual
Revision clause	No clause but suggested to re-evaluate in 5 years.
Presentation of results	
Interest parameters	Short term (2019-2023), medium term (2024-2033) and long term (2034-2053): • Median and 90% interquartile ranges of catch • Median and 90% interquartile ranges of SSB • Median and 90% interquartile ranges of realized harvest rates • Risk of SSB falling below Blim
**** Risk type and time interval	Risk 2
Precautionary risk level	Risk 2
Experiences and comments	
Review, acceptance:	Accepted
Use of ICES guidelines (WKGMSE 2013)	We were not aware of the WKGMSE 2013 report so it was not used. However, guidelines were used from the ICES Technical Advice 2017, chapter 12.4.3.1 on ICES fisheries management reference points for category 1 & 2 stocks. We did not deviate from these guidelines.
Experiences and comments	

2017 Sardine pil.27.8c9a

ICES. 2017. Report of the Benchmark Workshop on Pelagic Stocks, 6–10 February 2017, Lisbon, Portugal. ICES CM 2017/ACOM:35. 278 pp.

Background		
Motive/initiative/background	Re-evaluate whether the Portuguese-Spanish sardine fishery management plan remains precautionary taking into account the new agreed analytical assessment method and potential new biological reference points from the benchmark assessment (WKPELA 2017). In addition, it should be evaluated whether the plan remains precautionary when adding the following condition to the original plan, as requested by the EU to ICES: "In cases where applying the plan results in catches of less than 50% of catches in the previous year, then ICES catch advice on a precautionary basis should apply."	
Main objectives	High probability of recovery to above Blim = 337 448 tons in 5 years' time frame, precautionary and stable catches after recovery.	
Formal framework	ICES on request from EU	
Who did the evaluation work	WKPELA 2017 Workshop to evaluate the management plan for Iberian sardine (WKEMPIS)	
Method		
Software	HCS_15_1 ("Harvest Control Simulation")	
Name, brief outline	Age-structured operating model, "short cut" type.	
Reference or documentation	Unpublished, documented in Skagen 2015, code available from www.dwsk.net	
Type of stock	Short-medium life span, pelagic, very valuable	
Knowledge base*	Analytic assessment (Stock Synthesis 3, v. 3.24AB), annual catches-at- age, annual spring acoustic survey, triennal DEPM survey	
Type of regulation	Catch and effort limitations agreed between Portugal and Spain	
Operating model conditioning		
	Function, source of data	Stochastic? – how (distribution, source of variability)
Recruitment	Hockey-stick fitted to SR pairs 1993–2015	Log-normal, CV = 0.49 from residuals
Growth & maturity	Average over 2010–2015, no density dependence. B1+ was used as a proxy for SSB.	Random lognormal noise on weights-at-age.
Natural mortality	M-at-age 0–6+ = 0.98, 0.61, 0.47, 0.40, 0.36, 0.35, 0.32 (Gislason *et al.*, 2010 formula)	No
Selectivity	F-at-age estimate in the benchmark assessment , scaled to the mean of ages 2–5 (selectivity fixed from 2006 to 2015, flat at ages 3–5)	No, except deviations at the implementation step, with CVs for all ages similar to those in the assessment
Initial stock numbers	From assessment, 0–group abundance replaced by the geometric mean of 0–group abundance in 2006–2015	According to the variance–covariance matrix from the assessment (inverse Hessian)
Decision basis **	B1+ at the start of the intermediate year	

Number of iterations	1000	
Projection time	30 years	
Observation and implementation models		
With assessment		
Input data		
Comparison with ordinary assessment? ***		
Deviations from EG practice?		
No assessment	*(example of how to present this when there is no assessment)*	
Type of noise	Year factor + age factor in an auto-regressive model on stock numbers-at-age along year classes.	Both log-normal and auto-regressive model among year classes Age factor from CV estimates in the assessment Year factor adapted to reproduce CV of the SSB estimate in the assessment
Comparison with ordinary assessment? ***	Year factor scaled to give CV of SSB in year 10 as CV of SSB in the assessment	
Projection: If yes, how?	Yes, deterministic with recruitment according to deterministic SR function, with provisional official catches in 2016 = 22 700 tons reported by WK members	
Projection: Deviations from EG practice?	TAC constraint in projections, EG uses F_{sq}	
Implementation	Catches in numbers-at-age from projection according to the rule	Log-normally distributed error, CV 10%, no bias
Harvest rule		
Harvest rule design	Catch rule with two breakpoints B0 and Btrigger. Catch = 0 below B0, fixed catch above Btrigger, linear reduction between these points.	
Stabilizers	None in the proposed rule.	
Duration of decisions	Annual	
Revision clause		
Presentation of results		
Interest parameters	Risk to Blim and to Blow=rounded geometric mean SSB in 2012–2016= 132 thousand tons (very close but below B0 in the HR), Catch (Mean and 10–50–90 percentiles), Inter-annual variation, P (B1+>Blim) > 0.50 and P (B1+>Blim) > 0.95, probability that the biomass increase from the recent low =Blow	
Risk type and time interval****	Type 3	
Precautionary risk level	5%	
Experiences and comments		
Review, acceptance		
Experiences and comments		

2017 Pandalus pra.27.3a4a

Cardinale, M., Fernandez, C., Eigaard, O.R., and Søvik, G. 2017. Report on the Long-term Management Strategy Evaluation for Northern Shrimp (Pandalus borealis) in Division 4.a East and Subdivision 20, October–November 2017. ICES CM 2017/ACOM:52. 185 pp.

Background		
Motive/ initiative/ background	ICES had suggested that rather than issuing advice for the following year in September, the advice could be delivered in March, just two months after the survey, in order to set a TAC for the same year based on that ICES advice.	
Main objectives	ICES was asked to evaluate and estimate the optimum combination of Ftarget and Btrigger and to perform separate evaluations for different TAC years (from May or from January).	
Formal framework	EU-Norway request.	
Who did the evaluation work	Members of the Joint NAFO/ICES Pandalus Assessment Working Group (NIPAG), led by Carmen Fernández (not a NIPAG member)	
Method		
Software Name, brief outline include ref. or documentation	The MSE was coded in R. This produced an Operating Model (OM) that follows the population dynamics of the agreed stock assessment (with SS3). The R code is available on the NIPAG SharePoint.	
Type of stock	Crustacean	
Knowledge base *	The basis for the MSE is the stock assessment conducted with SS3	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Segmented regression with autocorrelation (0.3). Recruitment scenarios: (1) Historic (full time-series), and (2) Low (corresponding to the recruitment observed during 2007– 2016)	Yes, Recruitments are re-sampled from their predictive distribution which is based on parametric models fitted to the timeseries provided
Growth & maturity	Growth: Length-at-age is modelled using the von Bertalanffy model. A fixed weight-at-length curve is assumed. Maturity: maturity-at-length curve	Growth: variability around the von Bertalanffy curve using CVyoung and CVold. The five growth parameters are assumed constant through time. Maturity: curve assumed fixed and perfectly known
Natural mortality	M is set at 0.75 (as in the stock assessment)	No, constant through time
Selectivity	Single fishing fleet with logistic length-based selectivity	No, constant through time
Initial stock numbers	From the most recent SS3 assessment.	Yes, 10 000 draws were generated from a multivariate normal distribution of the parameters estimated by the SS3 assessment,

		using the point estimates as the mean of the distribution and the inverse Hessian matrix as the variance-covariance matrix. Each of the 10 000 parameter sets were applied in the population dynamics model assumed in the SS3 stock assessment, and all abundance and fishing mortalities-at-age were reconstructed for all years and quarters in the stock assessment, up to the start of year 2017
Decision basis **	SSB at the start of the TAC year	
Number of iterations	10 000	
Projection time	100 years	
Observation and implementation models		
Type of noise	Assessment / advice error is represented by selecting appropriate parameters for the distribution of ln(Fforecast) – ln(Ftrue), where Fforecast is the F that the forecast aims to apply and Ftrue is the F that then occurs in reality (WKMSYREF approach).	
*** Comparison with ordinary assessment?	Not full-feedback. Analyses were done using the ordinary assessment to determine advice error.	
Projection: If yes - how?	Yes, depending on TAC year	
Projection: Deviations from WG practice?	Followed WG practice and explored alternatives depnding on the TAC year.	
Implementation	No implementation error	
Harvest rule		
Harvest rule design	Single breakpoint HCR (target F vs SSB at start of TAC year), reducing linearly to zero	
Stabilizers	No. Interannual quota-flexibility (10%) examined.	
Duration of decisions	Annual	
Revision clause	No (was evaluating options rather than a specific existing MP)	
Presentation of results		
Interest parameters	Four time periods: ST (short term): 2018–2022 MT (medium term): 2023–2032 LT (long term): 2033–2042 LT2 (long-term equilibrium): 2043–2116 Performance statistic considered: • Risk of SSB<Blim • SSB distribution • Yield distribution • Distribution of interannual variability of yield • Distribution of the Fbar intended when applying the HCR and Fbar real	
**** Risk type and time interval	• Risk3 (with the probability expressed as a percentage) in ST, MT, LT and LT2 (for LT2, it is calculated as Risk1). • Risk2 (with the probability expressed as a percentage) in ST, MT and LT • Risk2>once (with the probability expressed as a percentage) in ST, MT and LT	

	Note: Risk2>once was added for this evaluation. For a given year y, Risk>once(y) is defined as the probability that SSB<Blim at least twice over the years y to y+19. Risk2>once is then defined as the maximum of Risk>once(y), where the maximum is taken over the years y included in the stated time period (e.g. ST, MT…)
Precautionary risk level	5% of risk type 3 and 2.
Experiences and comments	
Review, acceptance:	RGPandLTMS reviewed the work and considered it a good basis for the ICES advice
Use of ICES guidelines (WKGMSE 2013)	Used and considered useful
Experiences and comments	

2017 Tusk usk.27.5a14

ICES. 2017. Report of the Workshop on Evaluation of the Adopted Harvest Control Rules for Icelandic Summer-Spawning Herring, Ling and Tusk (WKICEMSE), 21–25 April 2017, Copenhagen, Denmark. ICES CM 2017/ACOM:45. 49 pp.

Background		
Motive/ initiative/ background	On 22 December 2016, ICES received the following request from Iceland: *The Government of Iceland is in the process of formally adopting management plans for Icelandic summer spawning herring (5a), ling (5a) and tusk (5a14):The management strategy for Icelandic summer spawning herring, ling and tusk is to maintain the exploitation rate at the rate which is consistent with the precautionary approach and that generates maximum sustainable yield (MSY) in the long term.* *A part of the management plan is the adoption of harvest control rules (HCR) for the three stocks for setting annual total allowable catch (TAC). The HCR adopted should be precautionary and in accordance with the ICES MSY approach.*	
Main objectives	Further in the request: *The Government of Iceland requests ICES to evaluate whether these harvest control rules are in accordance with its objectives, given current ICES definition of reference points or any re-evaluation of those points that may occur in the process. For ling and tusk the evaluation should also include review of input data and the applied assessment methodology (Benchmark).*	
Formal framework	ICES on request from Iceland	
Who did the evaluation work	An ad hoc group of managers, stakeholders, and scientists from the Marine and Freshwater Research Institute (MFRI), initiated by the Icelandic Ministry of Industries and Innovation in the summer of 2016 (mainly coordinated by B. Elvarsson).	
Method		
Software Name, brief outline include ref. or documentation	Gadget is a shorthand for the "Globally applicable Area Disaggregated General Ecosystem Toolbox", which is a statistical model of marine ecosystems (previously known as BORMICON (Stefánsson and Pálsson 1997) and Fleksibest Frøysa et al. (2002)). Gadget is an age-length structured forward-simulation modelling framework, where models can be coupled with an extensive set of data comparison and optimisation routines. Processes are generally modelled as dependent on length, but age is tracked in the models, and data can be compared on either a length and/or age scale. The framework allows for the creation of multi-area, multi-fleet models, capable of including predation and mixed fisheries issues, however it can also be used on a single species basis. Gadget models can be both very data- and computationally- intensive, with optimisation in particular taking a large amount of time.	
Type of stock	Both stocks are demersal and have a medium life span and medium value. Tusk is considered to be a shared stock with Greenland (5a and 14), but landings have been historically very low in Greenland.	
Knowledge base *	The request from Iceland to evaluate the harvest control rule also included benchmarks for the species, which were approved as category 1 analytical stock assessments with indices from an annual surveys, as well as length, age, and maturity distributions from survey and commercial samples. Catches were removed directly.	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)

Recruitment	Hockey-stick, series estimated during base model conditioning and breakpoint set to Bloss = Bpa (lowest observed SSB)	5-year block bootstrap
Growth & maturity	Von Bertalanffy growth and logistic maturity function, estimated during base model conditioning. Weights were generated from a fixed length-weight relationship estimated from all years external to the base model.	Spatial bootstrap used to generate error in estimated base model parameters. Refer to Elvarsson et al. (2014) and Lentin (2017) for further implementation details.
Natural mortality	Set to 0.15	No
Selectivity	Separate for survey and commercial stocks, logistic or dome-shaped.	Spatial bootstrap used to generate error in estimated base model parameters. Refer to Elvarsson et al. (2014) and Lentin (2017) for further implementation details.
Initial stock numbers	Estimated as parameters	Spatial bootstrap used to generate error in estimated base model parameters. Refer to Elvarsson et al. (2014,2018) and Lentin (2017) for further implementation details.
Decision basis **	A reference biomass estimated from the last assessment year (40+ cm for tusk) is multiplied by a harvest rate to generate the following year's TAC, as long as SSB does not fall below Btrig, which is set to Bpa.	
Number of iterations	101,000	
Projection time	300 years	
Observation and implementation models		
Type of noise	Empirical	Spatial bootstrap used to generate error in estimated base model parameters. Refer to Elvarsson et al. (2014) and Lentin (2017) for further implementation details.
*** Comparison with ordinary assessment?	Because the benchmark and MSE were conducted simultaneously, there are very few years to compare with an ordinary assessment based on that benchmark. In comparison to the previous models used for tusk, there was a consistent downward revision in biomass with consecutive years in retrospective plots. This pattern was likely a result of discounting a few earlier years with high recruitment and the importance of age data that are more available and informative toward the end of the time series. However, the state of the stock was not altered greatly: changes were within bootstrap interquantile ranges of the MSE.	
Projection: If yes - how?	Yes. Deterministic filling of the current year's TAC (2 quarters), followed by stochastic projections. The spatial bootstrap and recruitment block bootstrap generated stochasticity in population dynamics. Assessment error and autocorrelation (CV = 0.2, rho = 0.8) were included. TACs were filled exactly.	
Projection: Deviations from WG practice?	Likely not.	
Implementation	A range of 101 harvest rates were tested (0 – 1 in steps of 0.01) with 1000 stochastic projections each.	For each harvest rate, 100 bootstrap replicates were used, each with 10 runs containing recruitment and assessment error.

Harvest rule	
Harvest rule design	The TAC in year y is the reference biomass (40+ cm) observed in the previous assessment year (y-1) multiplied by a harvest rate scalar. The scalar was proposed to be 0.13, which was lower than that obtained at MSY but had the dual benefit of yielding catches close to MSY and maintaining high biomass levels. Btrigger was set to Bpa, thereby scaling the harvest rate with biomass level relative to Btrigger under this level.
Stabilizers	None
Duration of decisions	Annual
Revision clause	No clause but suggested to re-evaluate in 5 years.
Presentation of results	
Interest parameters	Short term (2019-2023), medium term (2024-2033) and long term (2034-2053): • Median and 90% interquartile ranges of catch • Median and 90% interquartile ranges of SSB • Median and 90% interquartile ranges of realized harvest rates • Risk of SSB falling below Blim
**** Risk type and time interval	Risk 2
Precautionary risk level	Risk 2
Experiences and comments	
Review, acceptance:	Accepted
Use of ICES guidelines (WKGMSE 2013)	We were not aware of the WKGMSE 2013 report so it was not used. However, guidelines were used from the ICES Technical Advice 2017, chapter 12.4.3.1 on ICES fisheries management reference points for category 1 & 2 stocks. We did not deviate from these guidelines.
Experiences and comments	

2018 Herring 27.1-24a514a

ICES. 2018. Report of the Workshop on a long-term management strategy for Norwegian Spring-spawning herring (WKNSSHMSE), 26-27 August 2018, Torshavn, Faroe Islands. ICES CM 2018/ACOM: 53. 108 pp.

Background		
Motive/ initiative/ background	NEAFC, on behalf of the Coastal States have in May 2018 submitted a request for ICES to evaluate options for NSSH long term management plan. This followed on from the advice on the revision of NSSH reference points issues in the beginning off 2018 (WKNSSHREF).	
Main objectives	The objective is to ensure harvest of the stock within safe biological limits.	
Formal framework	ICES on request from NEAFC.	
Who did the evaluation work	WKNSSHMSE 2018	
Method		
Software Name, brief outline include ref. or documentation	XSAM based simulation framework. Age structured operating model, no full assessment in the loop.	
Type of stock	Long life span, pelagic, straddling, very valuable	
Knowledge base *	Analytic assessment	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Beverton-Holt, Ricker and segmented regression SRRs, with lowest AIC based on 5000 resamples of pairs of stock recruitment (SSB-Age2) from1950 onwards, including the collapse period 1968-87. Includes 1st order dependency in residuals.	Log-normal
Growth & maturity	Weight in catch: resampled from 1988-2016 Weight in stock: resampled from 1988-2016 no density dependence in growth Maturity: maturity ogive for a normal year class	Resampling from past values
Natural mortality	For age 2 M = 0.9, ages 3+ M = 0.15	No
Selectivity	As estimated by XSAM using data 1988-2017 (i.e. exploitation pattern follows the same model).	Yes
Initial stock numbers	From assessment	Obtained from the assessment model fit: provides the approximated simultaneous distribution of all parameters and stock sizes such that initial values can be sampled from this approximated distribution.
Decision basis **	SSB or Bref (4+ biomass) in the TAC year	
Number of iterations	3000	
Projection time	35 years	
Observation and implementation models		
Type of noise	CVs and correlations among the estimated and predicted values is accounted for. The F	Yes

	multiplier will be affected by the error in the weighting factors $w^F_{a,y+1}$ and selection pattern $s_{a,y+1}$. Finally, the TAC will be affected by the projected $N_{a,y+1}$ which gives $C_{a,y+1}$ in addition to the weight at age in the prediction	
*** Comparison with ordinary assessment?	Based on ordinary assessment.	
Projection: If yes - how?	No STF conducted (not full feedback).	
Projection: Deviations from WG practice?	N/A	
Implementation	First F given by the HCR is found based on the perceived SSB. Then a TAC is calculated, and this TAC is translated into catch numbers at age, accounting for the selection at age and weights at age. i.e. prediction error is accounted for, but no implementation error is assumed	
Harvest rule		
Harvest rule design	Four rules were studied, with different parameterisations (see request).	
Stabilizers	Two catch stabilising mechanisms were requested: 1. 20% down / 25% up restrictions 2. TAC = mean of currrent TAC and HR TAC	
Duration of decisions	Annual	
Revision clause	No clause for when the MP should be revised.	
Presentation of results		
Interest parameters	Short term (2019-2023), medium term (2024-2033) and long term (2034-2053): • Average SSB • Average yield • Indicator for year to year variability in SSB and yield • Risk of SSB falling below Blim	
**** Risk type and time interval	Risk type 3 as defined by WKGMSE 2013; the maximum probability that SSB is below B_{lim}, where the maximum (of the annual probabilities) is taken over the relevant years). For short, medium and long term and quasi-equilibrium (see definitions above).	
Precautionary risk level	5% of risk type 3.	
Experiences and comments		
Review, acceptance:	The current management plan has been in effect since 2001.	
Experiences and comments		

2018 Herring her.27.irls

ICES. 2018. Report on Celtic Sea Herring Management Plan Evaluation. Coming as Annex 8 in the report of the Herring Assessment Working Group for the Area South of 62°N (HAWG), scheduled to meet 12–20 March 2018 at ICES HQ, Denmark.

Background		
Motive/ initiative/ background	The EU requested ICES to assess the long-term management strategy (plan) for Celtic Sea herring as follows: ”ICES is requested to assess whether the plan is still precautionary. ICES is furthermore asked, in analysing the elements of the plan, as to ensure that the stock is fished and maintained at levels which can produce maximum sustainable yield, as soon as possible and at the latest in 2020. Should the proposed plan include elements that are in contradiction with ensuring that the stock is fished and maintained, also in the future, at levels which can produce maximum sustainable yield, ICES is requested to comment specifically on such elements, and their consequences for ensuring MSY.”	
Main objectives	Evaluate whether the long term management plan is still in accordance with the precautionary approachand MSY approach.	
Formal framework	ICES on request from the EU.	
Who did the evaluation work	Evaluation work and simulations done by staff of the Marine Institute, Ireland.	
Method		
Software Name, brief outline include ref. or documentation	Evaluations of the long term management plan for Celtic Sea Herring in 2018 were performed using HCS10_3 which is a general purpose program for stochastic simulation of management decision rules. A projection is made through the intermediate year to obtain the stock abundance at the start of the TAC year. HCS mimics the advisory process without running actual assessments as part of the simulations. Instead, observation errors are specified as distributions and carried forward in predictions to get the numbers that are basis for management decisions. Options for implementation error and bias are also available (Skagen, 2010).	
Type of stock	Moderate longevity (6+ years), small scale fishery regionally important	
Knowledge base *	Category 1 – Full analytical assessment and forecast. One acoustic survey used for tuning.	
Type of regulation	TAC only, closed area, with small sentinel fishery within.	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution,

		source of variability)
Recruitment	Segmented regression (period 1970-last assessment year -2)	Lognormal – CV from residuals
Growth & maturity	Growth: mean of the last three years; no density dependence Maturity: constant annual ogives as used in the assessments of the stocks	No
Natural mortality	Derived from the average annual *M* values from the 2011 SMS key run (period 1974 – 2010) (ICES, 2015)	No
Selectivity	Based on the assessment output (time invariant).	No
Initial stock numbers	The latest assessment (ICES, 2018) was used for conditioning the simulations with 2017 as the start year.	No
Decision basis **	TAC based on F in the TAC year.	
Number of iterations	1000	
Projection time	20	
Observation and implementation models		
Type of noise	A range of errors and biases assumed on implementation and observation (assessment), taken from observed values in assessments and management performance.	
*** Comparison with ordinary assessment?	No assessment in the simulation loop.	
Projection: If yes - how?	The projection is deterministic, starting with the observed stock at 1. Jan in the last assessment year, with specified assumptions for catches or fishing mortalities. Weights and maturities are the true values from the last update, representing year *y1*. Recruitments are according to the deterministic stock recruit function, with no periodic or spasmodic variation. That includes year classes that have been born, but are not yet recruited. Hence, strong or weak future year classes are not known.	
Projection: Deviations from WG practice?	WG uses 3 year averages for mean weights. Recruitment used in the forecast is calculated by fitting a segmented regression to the stock recruit data using Julio's algorithm. Recruitment is set as the plateau recruitment from this fit if the SSB is above the changepoint in forecast year -2. If SSB is below the changepoint, then Recruitment (forecast year) = plateau recruitment *(SSB forecast year -2) / SSB Changepoint)	
Implementation	Initial numbers projected forward, drawing from SR relationship, with some truncation. TAC set based on F in the TAC year.	

Harvest rule	
Harvest rule design	The management plan specifies the following elements • $B_{Trigger}$ = 61,000 • Target F = 0.23 • Percentage constraint = 30% • B_{lim} = 34,000 Target F =0.23 when SSB is above $B_{trigger}$. When the SSB is below $B_{trigger}$ the TAC is set according to an F SSB * 0.23 / 61,000 A 30% constraint on TAC change applies when the stock is above B_{lim}. HCS10_3 was run to examine these specific elements of the plan and investigate if the plan is precautionary and MSY compliant.
Stabilizers	30% constraint on TAC change when the stock is above B_{lim}
Duration of decisions	Annual
Revision clause	Yes, every three years the plan will be avaluated.
Presentation of results	
Interest parameters	Risk, SSB, F, Landings TAC, annual TAC change.
**** Risk type and time interval	Type 2. P(SSB<Blim) in any year of the 20 forward-simulated years
Precautionary risk level	5% of risk type 2.
Experiences and comments	
Review, acceptance:	ICES advises that the harvest control rule in the long-term management plan for Celtic Sea herring is no longer consistent with the precautionary approach. The management plan results in a greater than 5% probability of the stock falling below Blim in several years throughout the simulated period.
Use of ICES guidelines (WKGMSE 2013)	This evaluation carried out in 2018 uses the same approach as previously in terms of the software and operating model, but is informed by the results of a recent benchmark workshop (ICES, 2018).
Experiences and comments	Previous evaluations of this plan were carried out in 2014 and 2015 (ICES, 2014, 2015) in response to changes in assessment settings or methodology. Additional elements of the plan were not evaluated in this MSE. Neither the stock assessment nor the current MSE simulation framework incorporates spatial data. It is therefore not possible to assess the impact of an area closure on the development of the stock. The plan indicates that additional measures will be taken should the stock falls below B_{lim}. However, such measures are not quantified within the plan and thus could not be tested. **Conclusion to the MSE:** The HCR in the long term management plan was considered to be no longer precautionary. The management plan results in a greater than 5% probability of the stock fall-

	ing below B_{lim} in several years throughout the simulated period. The simulations indicate the management plan cannot ensure that the stock is fished and maintained at levels which can produce maximum sustainable yield as soon as or by 2020. However, simulations indicate that median SSB reaches MSY $B_{trigger}$ in 2020. ICES considers that the stability clause in the management plan that constrains TAC changes from year to year to ±30% can imply fishing mortalities greater than F_{MSY}, which in turn may result in a greater than 5% probability of the stock falling below B_{lim}. This may also prevent the stock being fished and maintained at levels which can produce maximum sustainable yield, in any given year.

2018 Horse mackerel hom.27.9a

ICES. 2018. Report on the Assessment of a Long-term Management Strategy for Southern Horse Mackerel (hom27.9a), 15–16 February 2018. Manuela Azevedo, Hugo Mendes, Gersom Costas, Ernesto Jardim, Iago Mosqueira, Finlay Scott (Authors.) ICES CM 2018/ACOM:42.

Background		
Motive/initiative/background	The LTMS was proposed for this stock by initiative of the Pelagic Advisory Council (PelAC) in a collaborative work between scientists from IPMA and IEO and Portuguese and Spanish stakeholders from the South Western Waters Advisory Council (SWWAC). The stock has no management plan and is currently above MSY $B_{trigger}$ and exploited below F_{MSY}.	
Main objectives	Evaluate whether the plan is in accordance with the precautionary approach and MSY approach.	
Formal framework	Request from PELAC to European Commission.	
Who did the simulations work	Scientists from IPMA, IEO, JRC.	
Method		
Software	Stock assessment model (*sca*) and MSE framework implemented in R using the FLR packages (FLCore, FLa4a, FLash).	
Name, brief outline	Age-structured operating model and assessment with catches-at-age and one survey (IBTS) included in the loop. Survey indices used as input to the assessments in the simulations were generated from the "true" population on the basis of estimated catchability-at-age (from the *sca* model) with error coefficients log-normally distributed to simulate observation error. Catch-at-age from the perceived stock is assumed known and without implementation error.	
Reference or documentation	Documentation for the stock assessment model and MSE framework in Jardim, *et al.* (2017). Code available upon request.	
Type of stock	Medium life span (11+), pelagic/demersal, medium value, regionally important.	
Knowledge base	ICES category 1 stock.	
Type of regulation	TAC based on F in the TAC year.	
Operating model conditioning	**Function, source of data**	**Stochastic? – how (distribution, source of variability)**
Recruitment	Hockey-stick model (Azevedo *et al.*, 2016)	Log-normal (μ=0, σ=0.6), autocorrelated in time (ϕ_1=0.8).
Growth & maturity	As in last assessment (WGHANSA,2017)	No significant trends in historical weight-at-age. No indications of density-dependent growth.
Natural mortality	As in last assessment (WGHANSA,2017)	No. Natural mortality is age dependent and time invariant.
Selectivity	F-at-age as in latest 2012-2016 selectivity block reviewed in 2017 assessment/benchmark	No. The recent exploitation pattern of increased selectivity of young ages and decreased selectivity of older ages reflected in simulations.
Initial stock numbers	Population vector from *sca* model mimicking AMISH assessment	Almost identical to AMISH model.
Decision basis	SSB at spawning time in the TAC advice year	

Number of populations	200	
Projection time	2017-2080; 64 years	
Observation and implementation models		
With assessment		
Input data	Catches and one survey	Survey: error coefficients log-normally distributed to simulate observation error.
Comparison with ordinary assessment?	Yes	*sca* model is used to condition the simulation framework using the same setting as the AMISH model. Comparisons in several parameters including CV´s, retrospective patterns.
Projection	Yes. STF conducted in the management procedure	
Deviations from WG practice?	No	Changes from WG practice were only applied in a range of robustness/sensitivity tests.
Harvest rule		
Harvest rule design	i) If SSB ≥ $B_{trigger}$, F = F_{MSY} ii) If B_{lim} < SSB < $B_{trigger}$, F = $F_{by\text{-}catch}$ + [(F_{MSY} - $F_{by\text{-}catch}$) x (SSB - B_{lim})/($B_{trigger}$ - B_{lim})] iii) If SSB ≤ B_{lim} , F = $F_{by\text{-}catch}$	
Stabilizers	TAC shall not deviate more than 15% from the TAC the year before.	
Duration of decisions	Annual.	
Revision clause	After 5 years.	
Presentation of results		
Interest parameters	Short term (2017-2027) Long term (2070-2080) SSB risk analysis (B_{lim} and $B_{trigger}$), median catch, median fishing mortality, Interannual catch variability.	
Risk type and time interval	Type 3, over entire simulated period (2017-2080).	
Precautionary risk level	5%	
Experiences and comments		
Review, acceptance	The Long-Term Management Strategy (LTMS) was considered precautionary with a HCR that can lead to catches around MSY. The LTMS was not used in the 2018 TAC advice following the client (EU) request. Review and evaluation available in: http://ices.dk/sites/pub/Publication%20Reports/Advice/2018/Special_requests/eu.2018.05.pdf http://www.ices.dk/sites/pub/Publication%20Reports/Expert%20Group%20Report/acom/2018/AD%20HOC/Ad_Hoc_Report_hom27_9a_2018.pdf	

Use of ICES guidelines (WKGMSE, 2013) Experience and comments	Yes. In general, the background for the methodology (Management Procedure and risk type analysis) and presentation of results (indicators and timeframe for analysis) followed the WKGMSE guidelines.

2018 Norway pout nop.27.3a4

ICES. 2018. Report of the Workshop for management strategy evaluation for Norway Pout (WKNPOUT), 26–28 February 2018, ICES HQ, Copenhagen, Denmark. ICES CM 2018/ACOM:38. 96 pp.

Background		
Motive/ initiative/ background	The industry would like to have a minimum TAC each year.	
Main objectives	Find combinations of Fcap, TACmin, TACmax that are precautionary (i.e. risk<0.05).	
Formal framework	ICES on request from EU/Norway	
Who did the evaluation work	WKNPOUT 2018	
Method		
Software Name, brief outline include ref. or documentation	MSE used specialized code, written in R. SEASAM assessment model and forecast in TMB/R.	
Type of stock	short lived small pelagic. This stock is not used for human consumption.	
Knowledge base *	In reality, there is an annul assessment done using SEASAM	
Type of regulation	annual TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	distribution of estimates of all years from recent assessment model	Sampled estimates of past recruitment vary among trials. Resample from estimates each year if SSB>Blim.
Growth & maturity	resampled from a single past year	varies among trials
Natural mortality	resampled from a single past year	varies among trials
Selectivity	resampled from a single past year	varies among trials
Initial stock numbers	random sample from predictions from most recent assessment model	varies among trials

Decision basis **	Estimates of stock numbers at age and exploitation pattern are used in a stochastic forecast to attempt to achieve a desired SSB escapement after the TAC year.	
Number of ~~iterations~~ simulation trials	1000	
Projection time	35 years	
Observation and implementation models		
Type of noise	Biological parameters are known perfectly without observation error. Process error (as estimated by SEASAM) affects total mortality.	Yes. Every year, new process errors are drawn from the distribution and applied to the true population.
*** Comparison with ordinary assessment?	no	
~~Projection~~: Forecast: If yes - how?	A stochastic forecast to target an escaped SSB is part of the MP. As with all SEASAM and SAM forecasts, it does not match the standard definition of a stochastic forecast. As input to the forecast, we used the true biological parameters, as well as stock numbers and exploitation patterns that included observation error. Stock numbers and exploitation pattern estimates came from an assessment emulator rather than a real assessment (i.e. short-cut MSE) because SEASAM was assumed to be too fragile to run in a loop.	
~~Projection~~: Forecast: Deviations from WG practice?	There were no deviations from the real management procedure.	
Implementation	The maximum Fbar applied to the stock in the OM is limited by the maximum Fbar estimated from the most recent assessment (i.e. Fhistorical)	No. Fhistorical is the same in every year and every simulation trial.
Harvest rule		
Harvest rule design	Fcap puts an upper limit on escapement strategy. TACmin and TACmax override other parts of the HCR.	
Stabilizers	no	
Duration of decisions	annual TAC	
Revision clause	no	
Presentation of results		

~~Interest parameters~~ Performance statistics	• Average SSB (mean and median) • Average yield (mean and median) • Fbar (mean and median) • Risk 1 (Short term (2018-2022) and long term (2023–2037)) • Risk 3 (Short term (2018-2022) and long term (2023–2037)) • probability of giving TACmin • probability of giving TACmax • Mean change in TAC from one year to the next • Mean change in relative TAC from one year to the next • Average number of years it takes to rebuild SSB to above Blim
**** Risk type and time interval	risk type 1 in the long term (2023–2037)
Precautionary risk level	5%
Experiences and comments	
Review, acceptance:	yes
Use of ICES guidelines (WKGMSE 2013)	No. I attempted to read the guidelines, but found them to be poorly written, so I read the published literature instead (e.g. Punt et al. 2014) and talked to my colleagues.
Experiences and comments	

2018 Redfish - beaked reb.27.1-2

ICES. 2018. Report of the Workshop on the evaluation of harvest control rules for Sebastes mentella in ICES areas 1 and 2 (WKREBMSE). June–August 2018, by correspondence. ICES CM 2018/ACOM:52.

Background		
Motive/ initiative/ background	Norway and Russia jointly requested the evaluation of harvest control rules (HCRs) with F19+=0.06, 0.08 and 0.10 as well as trigger points of 450, 600 and 800 kt and all possible combinations (9) thereof. Additionally for the $B_{trigger}$=450 kt with F19+=0.08 option a clause to reduce F by 50% if the average strength at age 2 for the year classes which are 3-12 years old in the first year for which the TAC advice is given, is below 100 million individuals. Due to changes in reference points at the preceding benchmark (WKREDFISH) an additional B_{pa}=315 kt was introduced. Limitations in the software prevented to test the reduction of F at low recruitment. Instead, the HCR ($B_{trigger}$=450 kt F19+=0.08 & 0.06) was tested for the period 1992-2018 that contains 8 consecutive years of low recruitment around 2000, with numbers, weight and maturity-at-age as well as maturity and recruitment from the assessment.	
Main objectives	Establish HCRs for beaked redfish that comply to precautionary and MSY principles and specify actions to be taken if SSB falls below productive levels.	
Formal framework	ICES on request from Norway and Russia	
Who did the evaluation work	Bjarte Bogstad, Daniel Howell, Daisuke Goto, Hannes Höffle	
Method		
Software Name, brief outline include ref. or documentation	PROST is a software for single-fleet, single-area, long-term stochastic simulations (ICES 2006, Åsnes & Bogstad 2014) already used for the evaluation of HCRs for this stock in 2014 (WKREDMP 2014). Each operating model/HCR combination was simulated 10 000 times.	
Type of stock	Long lived (>30 years), demersal/pelagic, valuable	
Knowledge base *	Assessment with statistical catch at age model	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	2016 & 2017: Linear regression of recruitment at age 2 against the 0-group and winter (5-9 cm class) survey indices. Later years: Log-normally distributed, SSB independent recruitment function $R=\alpha e^{\varepsilon}$	Normally distributed error term ε
Growth & maturity	Weight- and Maturity-at-age are modelled values taken from the latest assessment (Tables 6.7 and 6.19, AFWG 2018) and are assumed constant.	No
Natural mortality	M=0.05	No
Selectivity	Selectivity at age is based on the average total fishing mortality for 2015-2017, assuming that selectivity in each fishery and the pelagic/demersal ratio will continue into the future as they are.	No
Initial stock numbers	From latest assessment and projected	Uncertainty was introduced into

	throught the intermediate year of 2018 assuming the same F as in 2017.	the numbers-at-age with a CV=0.3 for year classes 2016 and 2017; CV=0.2 for earlier ones.
Decision basis **	Probability of SSB falling below 600, 400 or 315 kt for the periods 2019–2023, 2019–2028 and 2019–2063. Yield for individual years from 2019 to 2023, as well as its mean for the periods 2019–2023 and 2019–2068. Mean realised F for 2019–2023, 2019–2028 and 2019–2063. Mean Interannual TAC variability 2014–2063.	
Number of iterations	10 000 per HCR scenario	
Projection time	50 years	
Observation and implementation models		
Type of noise	Single "assessment error" term set to CV=0.2 on log scale for all age groups and years.	The assessment error for any given year was uncorrelated between age groups.
*** Comparison with ordinary assessment?	There was no annual assessment run in the projection.	
Projection: If yes - how?	Projection 100 years into the future	
Projection: Deviations from WG practice?	F=0.08, $B_{trigger}$=450 kt	
Implementation	No implementation error included	No
Harvest rule		
Harvest rule design	F=0.06, 0.08, 0.10 $B_{trigger}$=800, 600, 450, 315 kt	
Stabilizers	TAC capped at 50 kt	
Duration of decisions	Annual	
Revision clause	None implicit, but recommendation for revision in 2023	
Presentation of results		
Interest parameters	Risk, yield, realised F and SSB for short (2019-2023), medium (2019-2028) and long (2019-2063) terms.	
**** Risk type and time interval	As Risk2 but with $B_{trigger}$ rather than B_{lim}	
Precautionary risk level	5% of SSB to fall below $B_{trigger}$ in the next 50 years	
Experiences and comments		
Review, acceptance:		
Use of ICES guidelines (WKGMSE 2013)	Guidelines were taken note of. However, only some of the sources of uncertainty in simulations were taken into account but minimum standards for simulations were followed. The MSE used the "short-cut" approach by using an uncertainty level rather than doing full annual assessments within the simulation.	
Experiences and comments		

Annex 3: Summary of presentations at WKGMSE2

Notes for the following presentations are given below, focussing on how they related to the TORs of the workshop.

#	Presenter(s)	Presentation title
1	Colm Lordan and David Miller	Management Strategy Evaluations Experiences from ICES since 2013
2	Doug Butterworth (Andre Punt)	What makes an MP and MP and an MSE an MSE?
3	Doug Butterworth	More successful MSE/MP Implementation – What's needed?
4	Laurie Kell, Doug Butterworth, and Rob Kronlund	Experiences from tuna RFMOs
5	Rob Kronlund	Application of Management Strategy Evaluation to Pacific Herring
6	Jon Deroba	Relevant highlights from a US Atlantic herring MSE and other US experiences
7	Iago Mosqueira (+JRC)	Making MSE algorithms 'user friendly': the a4a standard MSE
8	Martin Pastoors	Reflections on the role of ICES and MSE in fisheries management
9	Martin Pastoors (Cox *et al.)*	Evaluation of potential rebuilding strategies for the Western Horse Mackerel (*Trachurus trachurus*)
10	Martin Pastoors	Evaluation of alternative harvest control rules for Western Baltic Herring
11	Bjarki Elvarsson and Höskuldur Bjornsson	Modelling stochasticity in Harvest Control Rule evaluations (Recent MSE experiences and approaches in Iceland)
12	Andy Campbell and Michaël Gras	MSE Experiences: NEA Mackerel, Western Horse Mackerel, and Celtic Sea Herring
13	José De Oliveira	EU-Norway request to ICES to evaluate Management Strategies for: cod, haddock, whiting, saithe and herring
14	Daniel Howell	Common sense in MSEs
15	Daniel Howell	Ecosystem/Multispecies considerations in MSEs
16	Sarah Kraak	The Rosa Lee Phenomenon and its consequences for fisheries advice on changes in F (and other personal experiences with MSE, focused on certain aspects of the workshop's ToRs)
17	Hugo Mendes (+Manuela Azevedo)	Evaluation of performance of a full-feedback versus short-cut MSE approach: Southern Horse Mackerel (hom27.9a) case study
18	Mollie Brooks	Sprat MSE: Full vs Short-cut
19	Santiago Cerviño	Convergence in Management Strategy Evaluation

#	Presenter(s)	Presentation title
20	Dorleta García	Global Sensitivity Analysis of a multi-stock and multi-fleet MSE implementation
21	Quang Huynh (Carruthers + Hordyk)	Using management strategy evaluation to establish indicators of changing fisheries: A power test for alternative operating models
22	Quang Huynh	DLMtool / MSEtool: Software packages for management strategy evaluation
23	Polina Levontin and Laurie Kell	Visualisation
24	Iago Mosqueira (+JRC)	Overview of evaluations of multiannual plans (MAPs) carried out by STECF and current work being done by the Indian Ocean Tuna Commission (IOTC)

Presentation 1
Title: Management Strategy Evaluations Experiences from ICES since 2013

Presenters: Colm Lordan and David Miller

a. Review recent methodological and practical MSE work conducted in ICES and in other fora around the world…	The review by Lordan and Miller provides an overview of the MSE practice within ICES and is based on questionnaires sent out to scientists responsible for MSEs. There were 30 MSEs carried out between 2013 and 2018 (on average 5 per year) using 19 different methods (only 6 methods were used more than once). This shows that there is very little standardization in MSEs so far. In most cases (16), a short-cut method was used, compared to 5 with full-feedback methods and the rest with other combinations. Respondents indicated that the ICES MSE guidelines had been used and were considered useful guidance. However, MSE reports and presentation of MSE results were very heterogeneous. Communication on MSE results with managers has been challenging partly because of uninformative complexity. There is a need to develop guidelines for rebuilding plans.
b.2 Appropriate range of scenarios to consider in the MSE and how to deal with outcomes from multiple scenarios, including "worst-case" scenarios;	Half of the simulations considered only one recruitment scenario (though in some cases multiple SRRs were combined). Auto-correlation in recruitment was applied in one thirds of the simulations.
b.3 With reference to the work of WKGMSE (2013), review risk definition and computation in MSE;	Most MSEs consider P($SSB<B_{lim}$). Both Risk2 and Risk3 commonly used. In a few cases both were examined or Risk1 was also used for comparison. It was noted that Risk3 depends on time period length. More consistency on the definition of risk used is needed, with more clear guidance on which risk definition is appropriate in different situations.
b.6 Presentation of MSE results e.g. properly describing the process, standardising outputs to present results, etc.;	MSE reports and presentation of MSE results were very heterogeneous.
b.7 Review initiatives on the science side, including model developments, operating frameworks, etc. that could be incorporated in the ICES system.	Future MSE requirements – spatio-temporal, technical interactions, ecosystem components etc.

Presentation 2
Title: What makes an MP and MP and an MSE an MSE?

Presenters: Doug Butterworth (Andre Punt)

Relates to ToR elements: b5, b6, c

(b5) Evaluate the efficiency and effectiveness of "short-cut" approaches versus "full-feedback" simulation incorporating annual stock assessment models in the MSE loop;

(b6) Presentation of MSE results e.g. properly describing the process, standardising outputs to present results, etc;

(c) Update the guidelines for MSE evaluations in ICES originally prepared by WKGMSE (2013). (clarifying terminology)

<u>*Description*</u>

Uses of Management Strategy Evaluation

a) development of the management strategy for a particular fishery;

b) evaluation of generic management strategies, but experience has shown that the need for context-specific detail often arises quickly when applying generic approaches; and

c) identification of management strategies that will not work and should therefore be eliminated from further consideration.

- Regardless of the MSE usage category, the ultimate purpose of the MSE exercise is to develop management procedure(s) that provide acceptable trade-offs amongst management outcomes while being robust to uncertainties.
- The generic approach can be useful to understand feasibility, i.e., if a perfect information scenario fails to provide acceptable performance, then that suggests the need for more data.
- Other uses include the "value of information" analysis, e.g., what is the value (improvement in meeting the objective of maximizing catch without increasing resource risk) of conducting a survey that informs catch adjustments in response to abundance index? Relating the cost of additional information to the benefits (e.g., increased catch or reduced likelihood of fishery closures) provides evidence to support further data collection.

Empirical vs model-based strategies

a) Model-based strategy: one which includes a population dynamics model in the MP and hence incorporates explicit "estimation" and "harvest control rule" component.

b) Empirical strategy: one based on data collected directly from the fishery where the survey and/or fishery-dependent data are not analysed using a population dynamics model, though they may be pre-preprocessed (e.g. CPUE standardization).

- Empirical management strategies may provide similar performance to model-based ones, but the empirical MPs can be easier to explain and simulation test.
- Stakeholder engagement is a key benefit of empirical strategies since they can readily understand them and possibly calculate the catch limit themselves.

The Good, the Bad and the Ugly

a) *Good management strategies* fully document which data will be collected, how those data will be analysed, and the control rules/formulae used to produce a catch (effort) limit. Good management strategies are relatively rare.

b) *Ugly management strategies* hardly ever simulation test the assessment method exactly, often using an approximation of the actual assessment method in the strategy for pragmatic reasons. The issue here is how well does the proxy assessment method have to approximate the actual assessment method? MSE can be used in this situation, but evaluation is more complex.

c) *Bad management strategies* may involve assuming that management has perfect information, inclusion of impractical strategies and testing an assessment method in the strategy that differs from the actual method applied to the real fishery (i.e., test a strategy that includes a production model when the actual assessment in the strategy is an age-structured model). MSE is not appropriate in these situations.

- Since the purpose of MSE is to identify robust MPs, this means that vagueness about the assessment method in the MP (e.g., stating that a production model will be used without giving details) necessitates that a proposed HCR needs to be tested in conjunction with a large number of different assessment methods - in effect a claim to robustness cannot be made unless the MP intended for use is fully tested.

Summary

- MSE is an approach that can be used to assess management strategies that are (i) well (but not necessarily perfectly) specified, and (b) feasible to implement.
- A MP is a fully-specified management strategy (i.e., including specification of data to be collected, data analysis and HCR) and hence is more straightforward to evaluate and is the most pure of the types of management strategy to which MSE can be applied.
- The more incompletely the management strategy is specified, the more complex and lengthy the evaluation process will be and (likely) the more conservative the strategy needs to be given a pre-specified risk criterion.

Presentation 3
Title: More successful MSE/MP Implementation – What's needed?

Presenter: Doug Butterworth

Relates to ToR elements: b6, d

(b6) Presentation of MSE results e.g. properly describing the process, standardising outputs to present results, etc;

(d) Consider how to best disseminate the guidelines to experts within the ICES community and the need for training courses.

<u>*Description*</u>

Summarises *Shana K. Miller, Alejandro Anganuzzi, Doug S. Butterworth, Campbell R. Davies, Greg P. Donovan, Amanda Nickson, Rebecca A. Rademeyer, and Victor Restrepo. Improving communication: the key to more effective MSE processes. Can. J. Fish. Aquat. Sci. 00: 1–14 (0000) dx.doi.org/10.1139/cjfas-2018-0134.*

Context

- Consider the situation of RFMOs responsible for management in international waters:
  a) Many member countries with very different levels of ability and experience amongst scientists and fishery managers.
  b) Generally they operate by consensus, which is difficult to achieve unless "no change" (to catch limits) – this has led to poor stock management performance.
  c) Commissioners (and managers) are typically more adept at negotiating shared catch limits and are less comfortable with scientific discussions.
- The result has been generally poor management performance, so an attempt is being made to improve performance at RFMOs by implementing a management procedures (MP) approach where member states agree to the data inputs and harvest calculation formulae following simulation-testing performance.

Problems

- Major impediments to improving MSE/MP implementation were identified by the PEW Foundation who would like to see RFMOs adopt clearer rules for setting catch limits to discourage the common practice of decisions for no change).
  a) Scientists are seen as a major stumbling blocks to progress.
  b) Lack of commonality in explaining the MSE process and results, and diversity in the material provided to stakeholders was cited as an important problem, i.e., there is a need for standardization.
  c) Many scientists are themselves not well up on the concepts (leading to incorrect communication).
  d) The main underlying problem is that scientists are not good at communicating with a diverse group of fishery managers and stakeholders (i.e., they need to better design communication for the audience).
- The PEW Organization launched an initiative in 2017 to attempt to improve this communication. The paper describes the main outcomes of a workshop in early 2017 organised as part of this process, at which Andre Punt gave the presentation summarized above.

Examples of RFMO operation

- The basic RFMO decision structure includes (i) *scientific sub-committees* (scientists who do analytic work), (ii) a *scientific committee* (scientific review, development of recommendations), (iii) an *intermediary group of some form* (mixed membership including scientists, managers, stakeholders), and (iv) *commission* (usually made up of senior civil servants, where decisions are made).
- Recently, revision of the NAFO Greenland halibut MP was achieved quickly in 7 months with 6 meetings:
  a) Two scientific committee meetings were conducted, and four meetings of a "risk-based MSE group" that included national delegates consisting of scientists, managers and industry members);
  b) Following a presentation summarizing the process, the commission agreed to adopt the revised MP recommended by the "risk-based management strategy group" without further discussion.
  c) This expedited successful result was mainly the consequence of increased clarity of communication at the level of the intermediary "risk-based management strategy group".

Towards Solutions

- More focus on intermediary groups and their scientist-stakeholder interaction is the greatest need for effective communication.
- Improved visual communication tools for presentation of results.

Presentation 4
Title: Experiences from tuna RFMOs

Presenters: Laurie Kell, Doug Butterworth, and Rob Kronlund

The talk reported on the tuna RFMO Management Strategy Evaluation Technical Working Group meeting that convened 13-15 June 2018. The aims of this group are: to review the literature and experiences of tRFMOs in relation to MSE; provide guidance for developing MSE and operating models (OMs); and collaborate on developing the modelling framework. The group therefore covered a number of topics, including a review of the MSE process across the tuna RFMOs, dialogue with stakeholders, conditioning Operating Models (including accounting for multispecies and spatial/stock structure aspects), and provisions for exceptional circumstances. The meeting also discussed computational aspects, the dissemination of results, including tools to facilitate this (such as the vocabulary used), platforms for communication and dissemination (such as websites, github, etc.) and visualisation tools.

Among the recommendations was that a successful and efficient MSE technical process should not be assigned to a single individual, but should instead be iterative and involve a core group of experts that regularly communicate with scientists, managers and other stakeholders, and that these stakeholders be identified early in the process, with their role and input into the process clarified. The group also recommended capacity-building workshops for managers to facilitate better targeted information-sharing, and an early specification of how the review process of the MSE would take place, stressing the need for transparency. MSC (or similar) certification was recognised as a key motivator for the fishing industry, so there was also a recommendation for dialogue with MSC (e.g. through a workshop) to discuss their criteria for certification in an MSE context. Regarding the conditioning of operating models, the group recognised the value of limiting their number to adequately address the key uncertainties, but stressed that this limitation should not be taken so far that MPs require constant re-testing due to the range of plausible scenarios (i.e. those considered sufficiently consistent with historic data) not being sufficiently broad. The group also stressed that sufficient consideration be given to stock structure, including the provision of necessary data to allow spatially-structured operating models if found to be important, and where multispecies considerations are important, that initial operating model development focus on technical interactions rather than progressing straight to fully multispecies operating models. Full documentation of methods was stressed, including the mathematical specification of code. The use of tools such as Github for dissemination was recommended, and also that visualisation approaches for presenting results be tested with various focus groups to check suitability.

Since it focussed on various aspects of MSE development, communication of results and stakeholder engagement, the tRFMO MSE WG can be viewed as a parallel process to WKGMSE2, but in the tuna RFMO context. There was therefore a strong link to almost all of the TORs for WKGMSE2, and in particular to those TORs covered under sections 4 and 5 of the report, although the discussions in this group tended to be at a higher level than detailed technical specifications (e.g. how to model recruitment). Readers interested in the ICES guidelines for conducting MSEs are therefore encouraged to read the report from this tuna RFMO meeting.

Presentation 5
Title: Application of Management Strategy Evaluation to Pacific Herring

Presenter: Rob Kronlund

Relates to ToR elements: (a), (b6)

(a) Review recent methodological and practical MSE work conducted in ICES and in other fora around the world.

(b6) Presentation of MSE results e.g. properly describing the process, standardising outputs to present results, etc;

<u>*Description*</u>

Context

- Five major stocks of Pacific Herring in British Columbia are fished for roe and spawn on kelp (eggs deposited on kelp fronds) in late winter or early spring.
- Controversy related to both the stock assessment and management system led to a decision to revise the Pacific Herring management system, including conducting Management Strategy Evaluation (MSE) processes for the five major stocks in British Columbia:
  a) The fisheries may occur in populated areas (e.g., Strait of Georgia) and are highly visible due to obvious signs of spawning and presence of herring predators (e.g., marine mammals and seabirds);

b) Allocation conflicts and concerns about ecosystem effects of fisheries on forage species from academia and the public have created controversy;

c) Scientific debate about the Bayesian statistical catch-at-age model used to assess the stocks (survey catchability) and the costs of the coast-wide, annual spawn (egg deposition) survey have contributed to the controversy; and

d) Low estimated spawning biomass has led to prolonged fishery closures in 3 of the 5 major stock areas.

- For the first MSE cycle, the West Coast Vancouver Island (WCVI) and Strait of Georgia (SOG) stocks were selected because of investment in participant engagement since 2015 and contrasting stock histories (i.e., WCVI stock closed since 2006 due to low estimated spawning biomass, SOG stock near historical high spawning biomass).

MSE Process

- A set of measurable objectives for the stocks and fisheries was identified based on (i) Canadian fisheries policy and (ii) engagement with fishery managers, Indigenous peoples since 2015, and the fishing industry.
- Only some of the objectives could be addressed for this MSE cycle (spatial objectives require a spatial operating model which is currently not available but planned for development):

a) Avoid the **limit reference point** (LRP) of 0.30 B_0 (equilibrium unfished biomass) with high probability over three herring generations, where "high probability" is defined as 75-95% by policy.

b) Maintain spawning stock biomass above the Upper Stock Reference (USR) of 0.60 B_0, with 50% probability over three herring generations (*in fact 3 different USR choices were evalua*ted).

c) Maintain spawning stock biomass at or above a **target reference point** biomass level of 0.75B_0 with 75% probability over three herring generations (WCVI stock only).

- A decision was made in 2018 to change the management of the WCVI and SOG stocks from the model-oriented “best” assessment approach to a management-oriented MSE process for the 2019 fishery; the other three major stocks will continue to be managed based on the “best” assessment approach until the first cycle of MSE processes can be completed.
- Estimated population dynamics for both the SOG and WCVI stocks are dominated by time-varying natural mortality.
- Operating model scenarios based on the current statistical catch-at-age model used for assessment varied in their treatment of natural mortality; operating models were conditioned to observed survey indices, proportion-at-age and catches by fishery (annual observed weight-at-age and a maturity-at-age schedule are fixed inputs).
- Three natural mortality scenarios were considered:

1. Time-varying natural mortality in the past with (i) future natural mortality rates trending (random walk process) from $\boldsymbol{M_{2017}}$ to fluctuate around long-term **average** $\boldsymbol{M_{1951}}$**-**$\boldsymbol{M_{2017}}$, and (ii) **pulse natural mortality** events of 1.5M_t occurring at random 6% of the time when spawning biomass is < 0.3B_0 (some justification based on Allee effects).
2. Time-varying natural mortality in the past with future natural mortality rates (random walk) trending from $\boldsymbol{M_{2017}}$ to fluctuate around **10-year** $\boldsymbol{M_{2008}}$**-**$\boldsymbol{M_{2017}}$ average.
3. Constant natural mortality (estimated) in the past and future at the long-term 1951-2017 estimate.

- Management procedures (MPs) consisted of the future data simulated by the OM, the statistical catch-at-age model, and a range of harvest control rules that varied (i) harvest rate (10% or 20%), (ii) two spawning biomass operational control points where harvest rate is adjusted (specified as multipliers of estimated unfished biomass), and (iii) whether a catch cap is applied.

- Operational control points (OCPs) were not required to align with the limit reference point and upper stock reference point. The reduction in harvest rate and catch caps both served to mitigate the effects of large positive assessment errors diagnosed from simulation results.
- For the SOG stock, all MPs were able to meet objective 1 and most were able to meet objective 2 for most choices of upper stock reference points. A MP with a 20% harvest rate, OCPs at 0.3 and 0.6 of the estimated unfished biomass, and a catch cap of 30,000 t was adopted.
- For the WCVI stock, no MP was able to meet objectives across all scenarios. A MP with a 10% harvest rate and 2,000 t catch cap could meet Objectives 1 and 2 as long as future *M* was less than the 2008-2017 average. When the 2,000 catch cap was added, realized harvest rates well below 20% and often below 10% which served to mitigate assessment errors. The WCVI stock remains closed for 2019.
- For the WCVI stock, a MP with a 20% harvest rate, OCPs at 0.5 and 0.6 of the estimated unfished biomass, and a catch cap of 2,000 t will be tested after adding the constraint that the estimated biomass must be above the lower control point for *X* years before opening the fishery. Such procedures could ensure defensibility of on-going management advice while safeguarding against heavy depletion in the short-term as the further strategic MSE work progresses.

Lessons Learned

1. *Engagement*: Increased by (i) ongoing communication with Indigenous groups that began in 2015 to identify objectives, and (ii) frequent interactions between scientists and fishery managers throughout the process.
2. *Stock Assessment Model Properties*: Simulation showed large positive assessment errors can occur that can result in at least short-term harvest rates higher than intended.
3. *Harvest Control Rule Design*: Harvest rate had the highest impact on outcomes by reducing the impact of positive assessment errors, followed by catch caps which are a model-free way to mitigate assessment errors. Operational control points in the HCR had a lesser effect in the presence of a lower target harvest rate and catch caps because, in particular, the lower control point was reached less frequently. The OCPs are not required to match reference points. The HCR and assessment model in the MP interact in ways that cannot be predicted in advance of simulation results.
4. *Integration of Science and Management*: Increased meetings between scientists and fishery managers relative to past practice, co-development of presentations and communication documents by scientists and managers, and meetings directly with decision-makers resulted in clear direction from decision-makers in the development of management recommendations.

Presentation 6
Title: Relevant highlights from a US Atlantic herring MSE and other US experiences

Presenter: Jon Deroba

Summary

Managers for U.S. Atlantic herring requested an MSE to identify potential harvest control rules for management. It was also requested that operating model included ecosystem dynamics, namely, the role of herring as a forage (prey) species. The effects of herring abundance on tuna growth, tern reproductive success, and dogfish natural mortality, were specified based on data and hypotheses. Operating models included a range of uncertainty in herring life history and assessment bias, although spatial dynamics were not considered due to time and modeling constraints.

The experiences of this MSE process highlighted the need to (1) educate stakeholders on MSE concepts, (2) communicate the results of the MSE in terms relevant to stakeholders, and (3) provide flexibility and broad types of figures used to display results of the MSE. While stakeholder engagement took place over two workshops, approximately only half of participants were present at both workshops. As a result, time was spent at both workshops spent time to explain the MSE process. Future educational workshops

should be provided separately from the stakeholder engagement workshops that are used to identify objectives and concerns.

The MSE process brings people of different backgrounds and motivations together, and communication needs to be clear to explain concepts between groups. For example, perceptions about herring abundance differ between different stakeholders. While conservationists are concerned with relative stock abundance (compared to reference points), anglers are interested in future catch in absolute magnitude.

Finally, no single figure could provide all information relevant to all parties. Schematic and cartoon diagrams can be used to describe the conditioning of operating models and selection of harvest control rules, while boxplots and scatterplots could be used to summarize the results of simulations. Figures have a limit on the amount of information that can displayed. While all performance metrics from many candidate harvest control rules can be shown in a single figure, fewer performance metrics should be shown if confidence intervals are included.

How the presentation addresses the Terms of Reference:

ToR (b2) Appropriate range of scenarios to consider in the MSE and how to deal with outcomes from multiple scenarios, including “worst-case” scenarios;

The MSE explored typical uncertainties in stock parameters, particularly natural mortality, steepness, and growth. Spatial dynamics were not incorporated due to time constraints.

ToR (b5) Evaluate the efficiency and effectiveness of "short-cut" approaches versus “full-feedback” simulation incorporating annual stock assessment models in the MSE loop;

This MSE used the short-cut approach, although there was no comparison between the short-cut and full-feedback approaches. An analysis of whether the short-cut approach is acceptable would guide the decision of the approach to be used in future MSEs.

ToR (b6) Presentation of MSE results e.g. properly describing the process, standardising outputs to present results, etc;

This presentation included a variety of figures used to display MSE results. However, presentation of MSE results also include verbal aspects, such as communication between participants, and educational aspects, such as explanation of MSE to participants that are unfamiliar with the concept.

Presentation 7
Title: Making MSE algorithms 'user friendly': the a4a standard MSE

Presenter: Iago Mosqueira

In relation to TOR b.6 Presentation of MSE results e.g. properly describing the process, standardising outputs to present results, etc;

The mse FLR/a4a package (http://www.flr-project.org/mse/) contains methods for calculating performance indicators from any of the simulation outputs, and over any specified time period. Indicators are simply defined as formulas and/or functions in the R language. The package contains as an example set of indicators those currently in use at the IOTC.

The computed indicators, together with times series of a number of OM metrics (SSB, F, ...) can be used to produce graphical output through the use of the mseviz FLR/a4a package. Again, the standard set of plots curently being applied in IOTC is currently available, together with a number of graphical displays that have been used in various STECF MSE reports.

In relation to TOR b.7 Review initiatives on the science side, including model developments, operating frameworks, etc. that could be incorporated in the ICES system.

The EC JRC 'Assessment for All' (A4A) initiative aims to provide a toolbox to develop, test, and distribute methods for stock assessments and MSEs. An extension of the FLR platform has been developed that provides a flexible but simplified code base to develop and execute MSE simulations. The system, mostly contained in the mse FLR/a4a package (http://www.flr-project.org/mse/), presents a series of S4 classes to represent the various elements in a simulation of the fishery system (e.g. OM) and methods to apply a wide range of possible altrnatve for any of the MSE steps (e.g. HCR, SA, OEM). The system has been designed to make implementing alternative elements as simple as possible. For example, a function to allow the SPICT stock assessment model to be used as part of an MP was coded up in 16 lines of code.

A tracking system has been implemented that informs the user of what each step in the process is doing, for example how implementation error has altered the effected fishing mortality on the OM from that decided by the HCR. This is particularly useful when multiple simulations are being run in an HPC setting.

This platform has been used to develop a number of MSE simulations, for example the STECF work on Multiannual plans for Adriatic pelagics, and is currently being applied, and extended, for ICES' work on management plans for North sea stocks, and for IOTC work on management procedures for Indiacn Ocean albacore tuna and swordfish.

Presentation 8
Title: Reflections on the role of ICES and MSE in fisheries management

Presenter: Martin Pastoors

a. Review recent methodological and practical MSE work conducted in ICES and in other fora around the world…	The presentation starts from the basis of the type of requests that are received by ICES to carry out MSE for different stocks. The requests often ask for developing specific elements of HCRs that will be consistent with the PA and MSY approaches. This can be interpreted as methods to evaluate potential effects of management actions (robustness) or as validating the content of management plans (certifying). An example is presented on the interactions between ICES advice and decisionmaking for North Sea herring that showed very frequent interactions between advice and policy where requests for evaluation of management plans are being triggered by changes in the stock perceptions (e.g. through benchmarks). Because of the delays between request and answer (1-3 years), this could lead to management actions that are no longer guided by management plans. Since the ICES advice on management plan requests often boil down to a table with risks to Blim (defining the possible space to operate in), it is argued that more emphasis should be devoted to standardizing the methodologies to generate those tables/metrics. An example was presented for WBSS herring (see WD xx). This could perhaps be developed in a similar way as EQSIM, where standardized methods have been implemented for reference points. Ideally the standardized MSE tools should be useable in the expert groups as a quick check on consistency of reference points and stock size and as the basis for requests to evaluate management plans.
b.5 Evaluate the efficiency and effectiveness of "short-cut" approaches versus "full-feedback" simulation…	Short-cut approaches – preferably standardized – are probably the most relevant tools for answering the type of requests received by ICES.

Presentation 9
Title: Evaluation of potential rebuilding strategies for the Western Horse Mackerel (*Trachurus trachurus*)

Presenters: Martin Pastoors (S.P. Cox, A.J. Benson, B. Doherty, S. Johnson)

a. Review recent methodological and practical MSE work conducted in ICES and in other fora around the world…	This working document presents an initial MSE carried out for Western Horse Mackerel evaluating potential harvest rules that can be applied over the full range of stock sizes (including below Blim). The method is based on a full feedback model where the operating model is conditioned on the Stock Synthesis model that is used for western horse mackerel. The assessment model in the management procedure is a statistical catch at age model. This preliminary evaluation did not include uncertainty in the starting conditions. Key findings are that reference points could be treated as absolute or relative, where in this case relative reference points would probably be preferred. It also showed that rebuilding could be achieved from below Blim when appropriate measures are implemented in the HCR. If immediate closures are implemented when the stock goes below Blim, this appeared to result in lower capacity to assess the state of the stock in the following years.
b.6 Presentation of MSE results e.g. properly describing the process, standardising outputs to present results, etc;	The approach explicitly recognizes both biomass and fishery objectives. Simulated outcomes under alternative rebuilding plans defined by alternative harvest control rules can be used to examine potential trade-offs among stock rebuilding and fishery performance objectives in both the short- and long-term.

Presentation 10
Title: Evaluation of alternative harvest control rules for Western Baltic Herring

Presenter: Martin Pastoors

a. Review recent methodological and practical MSE work conducted in ICES and in other fora around the world…	This work is a demonstration of a simple MSE using the FLR framework and conditioned on the most recent assessment of WBSS herring. It provides an exploration of a number of modifications to the standard ICES MSY rule, by exploring the consequences of different options for fishing when the stock is below B_{lim}. All reference points are taken directly from the most recent stock assessment. The paper concludes that when the stock recruitment curve, that was used to estimate the reference points, is also used in the MSE, a rapid recovery of the stock to above B_{lim} and MSY $B_{trigger}$ could be expected when applying a F_{target} of 0.31 (F_{MSY}). In that case there is litle difference in recovery time between the scenario where F is set to zero below B_{lim} or were F is gradually reduced to zero below B_{lim}. However, if only the most recent recruitment is used in the MSE, the F_{target} will not allow for rebuilding in any of the scenarios. It is concluded that revision of reference points should preferably be accompanied by a simple and standardized MSE to review whether the reference points are consistent with the assumed productivity of the stock.
b.1. Evaluation of performance in the short-term versus the long-term, including comparison of MSE projection results relative to assessment forecasts…	The evaluation compares the results of the MSE with the advice provided by ICES for WBSS herring.
b.2 Appropriate range of scenarios to consider in the MSE and how to deal with outcomes from multiple scenarios, including “worst-case” scenarios;	Using the most recent recruitment can be seen as a worst case scenario
b.5 Evaluate the efficiency and effectiveness of "short-cut" approaches versus “full-feedback” simulation…	Short-cut methods allow for quick inspection of consequences of combination of reference points and productivity of the stock. Standardization of short-cut methods (similar to EQSIM) would be required.

b.6 Presentation of MSE results e.g. properly describing the process, standardising outputs to present results, etc.;	The document is based on Rmarkdown. All code, text and visualizations are embedded in one document.

Presentation 11
Title: Modelling stochasticity in Harvest Control Rule evaluations (Recent MSE experiences and approaches in Iceland)

Presenters: Bjarki Elvarsson and Höskuldur Bjornsson

Contributing to ToR a and b

Summary:

For a number of Icelandic stocks (cod, haddock, saithe, ling, tusk and herring) MSEs were performed using short-cut approaches within the Gadget (processes size based) and MUPPET (only age-based) frameworks. In these short-cut approaches the candidate HCRs were evaluated on the basis of stochastic projections from the assessment model in combination with simulated assessment uncertainty.

In Icelandic waters the common form for harvest control rules (HCR) is in terms of harvest rate (HR) on a particular reference biomass (e.g. Icelandic haddock where TAC = HR x B_{45cm+}). Biomass rules are preferred over average F rule as biomass rules are easier to understand and communicate to stakeholders. In addition the biomass rules are less sensitive to changes in selection patterns, while F rules can be affected by recent changes in the fishery. The reference biomass is ideally selected based on a proxy of the exploitable biomass but should constant between years. To ensure stability in TAC, a combination of low $B_{trigger}$ and low F/HR is preferred. If high assessment error can be assumed, fixed but rather low harvest rate can be expected to perform better in terms of stability.

Operating models are generally based on assessment models and conditioned using either bootstrap or MCMC sampling. Process error in the projection is generally dominated by SSB recruitment relationship. Species specific issues such as density dependence in growth (Icelandic haddock) can be included if needed. Assessment uncertainty in the projections can be estimated using analytical assessment retrospectives with a number of peels between 10 to 15. However, short survey time series often limit the quality of estimates for assessment uncertainty. Potential solution to this problem could be to run the retrospective based on fixed survey catchability.

Presentation 12
Title: MSE Experiences: NEA Mackerel, Western Horse Mackerel, and Celtic Sea Herring

Presenters: Andy Campbell and Michaël Gras

Presentation on various “policy” and “technical” aspects encountered in MSE work for ICES

The talk presented an overview of stock and management histories for three pelagic stocks: Northeast Atlantic Mackerel (NEAM), Celtic Sea Herring (CSH), and Western Horse Mackerel (WHOM). All cases show that current stock states are somewhat depleted with SSB levels at or below MSYBtrigger reference points. Management has been somewhat ineffective in implementing management plans; the only stock that has followed the management plan is the CSH. All three stocks show little evidence of a SRR, while simultaneously recognizing that recruitment dynamics dominate the production levels.

The main issues for MSE included: 1) parameterization of assessment uncertainty, especially for the evaluation of short-term risk; 2) Recruitment spikes "dominate everything" in WHOM, and their implementation needs to be considered carefully; 3) Poor estimation of initial numbers for young ages requires a robust approach to describe uncertainty; 4) Different approaches for model validation include "sanity checks" for realism (e.g. under F=0), multiple model comparisons, and comparisons to historical data (recruitment, implementation of assessment error); and 5) How to evaluate the justification for using an alternative assessment model (XSA vs ASAP) (especially in cases where the assessment method has been explicitly defined by a benchmark).

The experiences highlighted several "take-home" messages for future work, such as the importance of stakeholder involvement in the MSE process, and the value of simple MSEs / HCRs, and the benefit of using understandable summary statistics that facilitate communication (e.g. long-term yields/SSBs may be less meaningful to managers than more direct indices of catch). It was recommended that we need to develop strategies for situations where the stock drops below Blim in order to avoid complete fishing closures and, subsequently, a reactionary triggering of the definition of a new Management Plan (via MSE). Additionally, the authors stressed the need for improved documentation and training for MSE-related methods and tools.

Presentation 13
Title: EU-Norway request to ICES to evaluate Management Strategies for: cod, haddock, whiting, saithe and herring

Presenter: José De Oliveira

Work currently underway to respond to a request from EU-Norway to ICES on the evaluation of long term management strategies for 5 jointly managed North Sea stocks was described. The work programme is not yet complete and has a very limited time frame (4 months), given the complexity of the request which includes a number of candidate harvest rules along with options for TAC change constraints and banking and borrowing. The time limitation has impacted the range of simulations that can be conducted.

A full feedback approach, mimicking the annual (SAM) assessment and short term forecast cycle has been adopted and operationalized using high performance computing. The baseline operating model for each stock is based on the most recent (SAM) stock assessment with the exception of Haddock for which the TSA stock assessment has been replaced with SAM, following acceptance testing. Optimisation of the HCRs, involving the identification of F_{target} and $B_{trigger}$ values that maximize yield whilst maintaining the maximum annual risk (type 3) below 5% will be identified for the baseline operating models only. Defined alternative operating models for each stock will then be used to conduct sensitivity tests. Additional work will be required to develop methods to incorporate these results into the overall conclusions.

Initial testing of long term simulations with zero target F indicates that a 20-year simulation is likely sufficient for the testing for the rules. A random sample from 10000 model iterations shows 1000 simulations to be appropriate for calculation of long term statistics.

Initial results for Cod imply an optimal target F (0.37) significantly higher than the F_{MSY} value of 0.31. Implementing TAC stability along with the most extreme banking and borrowing scheme raises IAV from 19% to 31%. It was recognised that further input from stakeholders during the process would be useful in terms of identifying the optimal HCRs.

Contribution of WKNSMSE work to ToRs

The above exercise is not yet complete and additional runs and investigations will be carried out.

(a) Review recent methodological and practical MSE work conducted in ICES and in other fora around the world.

This exercise can be considered a good example of a full feedback MSE and illustrates what is currently possible, albeit with extreme time challenges and large computational requirements

(b) The methodological and technical revision should include all aspects involved in MSE, and pay specific attention to the following issues that have been identified through recent work in the ICES system:

(b1) Evaluation of performance in the short term versus the long term, including treatment and interpretation of MSE projection results relative to forecasts from stock assessment models used to annually assess the resource (b2) Appropriate range of scenarios to consider in the MSE and how to deal with outcomes from multiple scenarios, including "worst-case" scenarios.

This is under development. Still in the process of conducting baseline scenarios.

(b3) With reference to the work of WKGMSE (2013), review risk definition and computation in MSE

Type 3 used here

(b4) How to deal in the context of MSE with the broad range of models currently used for stock assessment in ICES

SAM OM and assessment model in the loop. Replaces TSA for Haddock (tests/checks?)

(b5) Evaluate the efficiency and effectiveness of short-cut approach versus full-feedback simulation incorporating annual stock assessment models in the MSE loop

No shortcut approach run. This is a full feedback exercise. Computational challenges necessitate familiarity with high performance computing.

(b6) Presentation of MSE results e.g. properly describing the process, standardising outputs to present results etc.

Initial results presented to WKGMSE2 in colour-coded tabular format

(b7) Review initiatives on the science side, including model developments, operating frameworks etc. that could be incorporated in the ICES system

Use of high performance computing enables a large number of simulations to be conducted

(c) Update original guidelines

Presentation 14
Title: Common sense in MSEs

Presenter: Daniel Howell

Any MSE exercise is only as good as the models that were used in the simulations. In addition to the quantifiable uncertainties within the simulations, unquantifiable uncertainties arise from mis-specification of biological processes and variability or observation noise, or from unpredictable future changes in these factors. Provided that the forecasts are relatively similar (in terms of stock size, fishing mortality and so on) to those seen in the recent tuning data then there is reason to believe that simulations may be reasonably accurate. However, when the simulations involve stock sizes, catch levels or fishing mortalities beyond those seen in the recent past they move into the realm of extrapolating beyond the recent data, and the risk of unquantified errors increases. Such situations may arise where a stock moves from a period of recovery strategy of fisheries moratorium into Fmsy-based fisheries.

As a result, it is important to be careful when the MSE simulations suggest that a stock could sustain a significantly higher fishing pressure than it has been subjected to in the recent past. This is especially true for long-lived species, which recover very slowly from overfishing. Two concrete measures are suggested:

1. Where a proposed F or catch levels is significantly higher than that seen in the recent past, the actual quotas should be increased over a period of years to the proposed level and the stock response should be monitored. The longer lived the stock, the more caution should be exercised
2. Where proposed catches are similar to those which have crashed the stock in the past then extreme caution should be employed. Quotas should never be set at levels which previously led to major depletion of the stock without solid evidence (beyond the simulation studies) that something has changed in the ecosystem to support such catches.

In general, one shoud never follow directives blindly, do not disengage the brain while running MSEs.

Presentation 15
Title: Ecosystem/Multispecies considerations in MSEs

Presenter: Daniel Howell

A key part of designing any MSE is to identify which key processes need to be included in the operating model(s). As part of this process it is important to consider multispecies and ecosystem interactions, and identify which processes are sufficiently important that they need to be included in the operating model. Where such processes are important drivers of stock dynamics at the stock sizes expected over the life span of the HCR then these should be included in the OMs, but it is obviously not possible to include everything. It is possible to use full multispecies models as OMs within a MSE, and code has been developed to allow Gadget models to be used in this was with FLR. However, it is also possible to include many processes as add-ons to single species models (for example through density dependence). The choice between these approaches will depend on the particular driver being modelled, the availability of existing models, and the time and resources available.

Presentation 16
Title: The Rosa Lee Phenomenon and its consequences for fisheries advice on changes in F (and other personal experiences with MSE, focused on certain aspects of the workshop's ToRs)

Presenter: Sarah Kraak

Rosa Lee Phenomenon

Some results from a submitted paper (Kraak et al., submitted) were presented.

The Rosa Lee phenomenon is based on inherent variability in growth rate. Size selective fishing increases the probability of fast growing fish dying sooner. This results in a Von Bertalanffy L_∞ which is lower with fishing than without fishing. This effect is particular strong when F is high, and should be accounted for in forecasts/simulations.

Even though the Rosa Lee paper dates back from 1912, there are few studies taking this into account in forecasts. It was noted that the Gadget model takes the Rosa Lee effect account, and this was explored in a paper in 2006 for NEA Cod (Kvamme and Bogstad, 2006).

A length and age based population simulation model tracking numbers of individuals in length bins within age bins was used to illustrate the Rosa Lee effect. The results showed that changes in length at age should be taken into account when simulations involve changes in selection pattern, and probably also when changes in F are large, especially with slow-growing species.

Kvamme, C. & Bogstad, B. (2006). The effect of including length structure in yield-per-recruit estimates for northeast Artic cod. ICES Journal of Marine Science. 64. 10.1093/icesjms/fsl027.

Regarding the probability of <5% below Blim

Rochet & Rice (2009) warn that tails of distributions (e.g. for risk) are inherently very poorly estimated. In response, Butterworth et al. (2010) noted that MSEs are for robustness testing, comparing relative performance among management strategies but not for absolute values. However, MSEs in ICES use a 5% absolute risk criterion to determine if HCRs are precautionary or not.

Implementation error in mixed fisheries

It may not make sense to assume perfect implementation for stocks caught in mixed fisheries. We should not run HCRs in MSEs knowing that they could not be perfectly implemented in reality: in mixed fisheries catches are often driven by TACs/effort on other species rather than by the species own TAC.

Short-cut vs full feedback

Kraak et al. (2008) noted that shrinkage in assessments can cause a lag in the perception of the stock size and F. Such biases can change sign in a cyclical manner. This behaviour would not be taken into account using short-cut approaches for including assessment uncertainty unless it is explicitly taken into account.

Take Home Messages linked to ToRs

b. The methodological and technical revision should include all aspects involved in MSE, and pay specific attention to the following issues…	Take the Rosa Lee phenomena into account when the Management Procedure involves changes in selectivity and large changes in F especially for slow growing species.
b.2 Appropriate range of scenarios to consider in the MSE and how to deal with outcomes from multiple scenarios, including “worst-case” scenarios;	MSEs for demersal mixed fisheries should consider implementation error in the HCR due to mixed fisheries interactions constraining the realized F.
b.3 With reference to the work of WKGMSE (2013), review risk definition and computation in MSE;	In the ICES context the MSE is judged to be precautionary according to the 5% risk level (ICES, 2016). This is different to other fora where robustness testing with multiple OMs is common. The MSE should be setup to ensure that the tails of the distribution are sufficiently well estimated. Absolute levels of risk are conditional on the uncertainties you are taking into account. If you are pretending that you know the absolute risk level then you are misleading people.
b.5 Evaluate the efficiency and effectiveness of "short-cut" approaches versus “full-feedback” simulation…	The full feedback approach may highlight lags and biases generated by the assessment model which may not be taken into account in the short-cut approach.

Presentation 17
Title: Evaluation of performance of a full-feedback versus short-cut MSE approach: Southern Horse Mackerel (hom27.9a) case study

Presenters: Hugo Mendes (+Manuela Azevedo)

Compares a Full MSE approach used in the MP for the southern horse mackerel and two examples of a short MSE. The OM are set up the same way and the two short-cut approaches mimic the assessment with:

i) Observation error simulated by applying directly on the "true" SSB a log-normal distributed error using the mean CV´s from the assessment model. Advice error is simulated by applying log-normal distributed error on the F in the advice year using the observed CV of F.

ii) Observation error applied to the stock numbers-at-age (excluding age 0) using a log-normal distributed error with CV´s from the assessment model and advice error as above i).

The scenarios were tested under a HCR with four different catch stability mechanisms and under two productivity scenarios regimes; unchanged (from the observed historical series) productivity and low productivity

Results of all MSE methods are very similar in the scenario assuming unchanged productivity in the future.

With low productivity in the future, the methods differ in some way that is difficult to explain, but seems to be related with the lower CV of F does generally lead to lowest F and highest SSB while the full MSE does give the lowest SSB and highest F. All lead to similar catches that is in line with productivity of the stock. In this scenario of low productivity, conclusions on precautionary considerations can be different over the MSE methods (using error type 1 and 3).

Difficult to infer the reasons behind these difference unless we know exactly how the MSE methods are behaving in the scenarios (e.g. assessment model seems to fail with strong catch stabilizers/low productivity) but there is a clear discrepancy between the results for unchanged productivity and reduced productivity. Moreover, depending on the the catch stability mechanisms there are constrasting results between the short and full MSE method.

Presentation 18
Title: Sprat MSE: Full vs Short-cut

Presenter: Mollie Brooks

Methods

Mollie Brooks presented an MSE conducted on North Sea sprat (a short-lived pelagic species), with 4 seasons, and the assessment year starting 1 July. Because of the seasonality, Mollie wrote her own code. The HCR is an escapement strategy with TAC based on a deterministic forecast and Fcap. A hockey-stick SR relation was used with smoothed residuals (not a log-normal distribution).The baseline MSE was run for 35 years with 1000 replicates. The biological parameters (M, weight, maturity) were sampled from the past; the initial N was a random sample from the last stock assessment, as was the exploitation pattern E. These were all constant through time within replicates but varied among replicates. The "full" MSE used an observation simulator and an SMS assessment; the observation simulator produced catch and three surveys using multivariate lognormal error from the residuals of the last assessment. The "short-cut" MSE was based on adding multivariate normal error with mean 0 and var-cov from the last assessment to the true values of log N and log E.

Results

Visual comparison of the "full" and "short-cut" MSEs results look similar, including worm plots, yearly risk estimates, time series with confidence intervals, the relation between the true and the estimated stock numbers, and the relation between the true and the estimated Fbar. Estimates of F separated by age and quarter showed that there were slightly larger differences between the true and estimated values in the "full" MSE. Performance statistics (2032-2052) (risk1, prob low TAC, mean TAC, median TAC, atFhist) are quite similar. Risk1 estimates differed by less than 0.01. Both MSEs take < 1 hour to run. The complexity of hardware is an issue with the "full" MSE, because it requires different directories in which to run ADMB.

Comments from the group

Laurie Kell: Look at lags and biases, cross-correlations, autocorrelations. Look at them as an engineer, from a control point of view, like frequency rather than time series.

Daniel Howell: Look at what caused the convergence failure for some of the replicates.

Presentation 19
Title: Convergence in Management Strategy Evaluation

Presenter: Santiago Cerviño

The variability in MSE results in general is due to statistical and simulation variability. The statistical variability is derived from the variability introduced in the system (the natural variability and the uncertainty in the management procedure) and cannot be reduced. However, the simulation variability is derived from a lack of convergence of the model.

In the presentations done by the participants in the first two days only the presentation by Jose de Oliveira did something avoid convergence.

In theory the exact values would be only obtained if the simulations would be run for infinity replications, but as in practice this is not possible the problem is to determine an adequate number of replicated. An adequate number of iterations is the one that avoids leading to wrong conclusions.

Some tools than can help to assess the convergence like cumulative plots and histograms were presented.

The convergence is percentile dependent, so convergence in the central values of the distributions is reached with fewer replicated than for the tails.

Key Points:

Rationale about the number of iterations chosen should be presented in MSE reports since this number can affect the MSE conclusions.

There is a method to predict the iterations needed to reach convergence once you have run a certain number of iterations (LK)

Identify one or two critical scenarios where convergence is an issue, explore in depth this scenario to set the minimum iterations needed. And apply these to all the scenarios.

Precautionary criteria is a critical issue with limited iteration. Pay special attention to convergence!

Presentation 20
Title: Global Sensitivity Analysis of a multi-stock and multi-fleet MSE implementation

Presenter: Dorleta García

The presentation mostly relates to ToR b2 and b7.

Keypoints:

- Global sensitivity analyses (GSA) is recommended by the EC when running impact assessments. However, it has never been rarely used in MSE implementations.
- GSA is a valuable tool in the validation of simulation models.
- Specifically, GSA can be used to:
    - Identify the most important input factors.
    - Identify the factors that can be fixed in the simulation without impacting in the results.
    - To direct the modelling and conditioning effort.
    - GSA in combination with MSE can be used to assess the impact of increasing the knowledge in the system, through further sampling or research.
    - Gain a better knowledge of the performance of model.
    - Identifying coding or conditioning errors.

Summary

Sensitivity Analysis is the study of how the uncertainty in the model output can be described through the uncertainty in the model input. The sensitivity analysis in fisheries management simulation models are usually carried out changing the value of an specific parameter, for example testing different values of natural mortality. However, these type of test ignore the interaction between factors which are usual in highly nonlinear models such as MSE models. GSA goes further and studies the uncertainty in the whole domain of the input parameters including their interactions (Saltelli et al., 2008; Pianosi et al., 2016). Furthermore, it is recommended by the EC in the implementation of impact assessmen (EPA, 2009).

A GSA implementation was presented using an MSE application of FLBEIA to demersal mixed-fisheries operating in Iberian waters. After an efficient conditioning of the model the number of input effective factors in the GSA were 135. The GSA was conducted in two steps, first an screening method was applied to select the most important input factors (56) and then the output variance was decomposed as a function of those input factors. The variance, especially in the long term, was driven by the interaction between factors. The impact of most of the economic and technical factors in the output variance was marginal. In the long term, the impact of initial abundance decreases and that of errors in the assessment increases. Effort Share of trawlers was the factor with the highest impact on the results. Among the observation errors aging error was the one with the highest impact on the results. In the MP a short-cut approach was used and the error in the abundance of some stocks had also a significant impact on the results. Maturity of most of the stocks was discarded by the screening method and in the ones left had a low impact on the output variance according to the variance decomposition method. The impact of the uncertainty in recruitment was lower than expected and factors like weight at age and natural mortality had a bigger impact. The results were stock dependent and depend on the role of the stock in the fishery. (target-bycatch, mixed or single fishery) and the models used to describe the stock dynamics.

Presentation 21

Title: Using management strategy evaluation to establish indicators of changing fisheries: A power test for alternative operating models

Presenters: Quang Huynh (Carruthers + Hordyk)

A statistical power tool was presented to evaluate the performance of alternative OM if future data falls outside the confidence intervals of an established OM. These “outliers” can be caused by changes on biological/productivity parameters.

The power test is based on the Mahalanobis distance which describes the distance of a multivariate data from a distribution. The null distribution of the Mahalanobis distance is generated from the data in the established OM. Once the critical value is established, the power can be calculated from the predicted data in the alternative OM. Statistical power generally increases with time and larger Null/alternative differences.

Following a small discussion, it was mentioned that the use of alternative operating models should be tested in the MSE process.

Presentation 22

Title: DLMtool / MSEtool: Software packages for management strategy evaluation

Presenter: Quang Huynh

MSE tool is a software package written for the R environment that expands the DLMtool for performing MSE on data-rich stocks (e.g. ICES category 1) using assessment models, including state-space models and a combination of HCR functions. The tool is also able to model spatially-explicit MSE.

The package uses TMB computation for the assessment models which is much faster than the “standard” ADMB. The tool produces standard assessment reporting and diagnostics for estimates of spawning stock biomass, fishing mortality, recruitment, and retrospective analyses. Additional tools evaluate convergence and retrospective patterns of the assessment model in full-feedback simulation.

Presentation 23

Title: Visualisation

Presenters: Polina Levontin and Laurie Kell

An example of a web based "Visualisation" shinny-app was presented, http://www.seaplusplus.co.uk/shiny-app-for-mse/. The app is intended to enhance communication throughout the MSE process. Although the tool focused on the main sources of uncertainty as represented by the Operating Model (OM), model validation and the results, it is intended to involve all the players (e.g. RFMOs, managers, implementers and scientists) in the MSE process from the earliest stage. This work is mainly relevant to ToRs B6 and Section 5 of the Report.

The app included diagnostics related to model validation, e.g. convergence, and goodness of fit tests, and scorings for data and models. The tool was initially developed to help scientists to condition the OM but can be used to present different sources of information depending on the target audience. For example while scientist may need to understand the details of how the MSE was developed, simpler tools are required for communication with stakeholders about the performance of the alternative management procedures. Although the initial development was driven by the technical aspects the eventual audience when presenting MSE results will be mainly managers and stakeholders. Stakeholders may be less interested in all the sources of uncertainty. It was thought important to involvement stakeholders in the MSE process from the beginning, which the tool allows and to include objectives such as politics, economics to engage stakeholders.

Presentation 24

Title: Overview of evaluations of multiannual plans (MAPs) carried out by STECF and current work being done by the Indian Ocean Tuna Commission (IOTC)

Presenters: Iago Mosqueira

An overview of evaluations of multiannual plans (MAPs) carried out by STECF EWGs between 2012 and 2018, at the request of DG MARE, showed the range of approaches that were taken for each of the fisheries under analysis. Standard single-stock OMs, were generally constructed around the stock assessment model being used to provide advice, together with a number of sources of uncertainty. These models were extended and complemented in most cases by other analysis that informed on, for example:

- the likely effect on the ecosystem and associated species (e.g. use of ecosystem models on the North Sea MAP evaluation, STECF-15-04)
- economic effects on multiple fleets operating on multiple species (e.g. SWW and NWW MAP evaluation, STECF-15-08)
- consideration of the biological interactions among stocks (e.g. multispecies plans for the Baltic, STECF-12-06)
- identify the dependency of individual fleets in multi-species fisheries to particular species and the regulations affecting them (e.g. SWW and NWW MAP evaluation, STECF-15-08)
- effect of the levels of disparities between fleet effort and fishing mortality on the ability of an effort-based management regime (e.g. demersal fisheries in the Western Mediterranean, STECF-16-21)

An important addition in any STECF-conducted analyses has been the attempts at calculating various economic indicators that evaluate the changes in economic performance that could be expected on the affected fleets. A range of methods have been used for this, very often limited by the availability of data.

A quick overview of the work being currently carried out for the development of management procedures for stocks under the jurisdiction of the Indian Ocean Tuna Commission (IOTC) highlighted a number of issues of interest. A process of dialogue between scientists and managers led to the adoption of guidelines for the presentation of MSE results (IOTC, 2018) that reflect both what scientists think managers should consider when evaluating MPs, as well as what managers consider they need to assess their usefulness. These guidelines included both a series of performance indicators and a set of graphical outputs used to summarize them.

Annex 4: Working Documents presented at the workshop

MSE Convergence diagnostics WKMSE2

Santiago Cerviño

(see Annex 3, presentation 19 notes)

Reference points

Höskuldur Björnsson

An examination of requirements for Fmsy in various international agreements, and how these relate to the ICES definitions of reference points.

Evaluation of alternative harvest control rules for Western Baltic Herring

Martin Pastoors

(see Annex 3, presentation 10 notes)