ICES WKGMSE REPORT 2013

ICES ADVISORY COMMITTEE

ICES CM 2013 ACOM 39

REF. ACOM

Report of the Workshop on Guidelines for Management Strategy Evaluations (WKGMSE)

21 – 23 January 2013

ICES HQ, Copenhagen, Denmark

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

Recommended format for purposes of citation:

ICES. 2013. Report of the Workshop on Guidelines for Management Strategy Evaluations (WKGMSE) , 21 - 23 January 2013, ICES HQ, Copenhagen, Denmark. https://doi.org/10.17895/ices.pub.5304

Contents

Executive summary

The Workshop on Guidelines for Management Strategy Evaluations (WKGMSE) met 21-23 January in Copenhagen Denmark, The meeting was chaired by Dankert Skagen and John Simmonds, with 19 participants from 10 nations.

The purpose of the meeting was to review and bring up to date the methodologies and technical specifications that should be incorporated in MSE. The workshop also considered appropriate risk definitions for MSE, taking into account practices in ICES and elsewhere and developed an updated set of guidelines for MSE evaluations in ICES.

In order to review the methodologies and standards used in the past a summary template was prepared and circulated to participants. Eighteen MSEs were summarised using the template and reviewed at the workshop, these template tables are annexed to the report. Based on these reviews and the subsequent discussion the guidelines from SGMAS 2008 were revised at the workshop. A report evaluating the historic use of precautionary criteria used by ICES was prepared in advance of the meeting. This report is annexed to the report. The different precautionary criteria used for different MSEs were compared and following this the workshop recommended revised criteria that are consistent with the ICES precautionary approach for stocks not subject to MSEs based on Blim and the 95% biomass buffer Bpa. The workshop also included consideration of short lived species where the stocks may have greater than 5% probability of being below Blim with zero fishery.

The report describes first the review of past MSE work in Section 3 and then consideration of ICES standards for precautionary approach in Section 4. Based on these considerations revised guidelines for modelling and brief standards for reporting are provided, including a revised version of the reporting template to summarise the work.

The main results of the workshop are the revised guidelines and recommendations for revision of ICES precautionary criteria for management plans.

1 Introduction.

ICES regularly evaluates harvest control rules in management plans and gives advice on their performance. SGMAS prepared a set of guidelines in 2008 (ICES 2008), but these have not been updated and substantial experience has accumulated in the intervening years. ACOM has noted the need to review recent work and practices in ICES and elsewhere, and prepare an up-to-date set of guidelines that would serve as reference for MSE in ICES. In October 2012 ICES passed a resolution a provided ToR which are given below in Section 1.3

1.1 Background

SGMAS was created in 2005 to provide guidelines for evaluating management strategies in general and harvest control rules in particular. The incentive was the growing numbers of requests for evaluating such rules and the unclear standards for such evaluations. The SGMAS report from 2006 provides such guidelines. A further meeting was held in 2007 to summarize experience and to broaden the scope towards assisting in the development of rules rather than just evaluating proposed rules. This led to suggestions for improving the dialogue processes with managers and stakeholders some of which have been applied in the development of several plans. In 2008 SGMAS reviewed plans to date and provided updated guidelines. In 2009 ICES and STECF held a joint meeting WKOMSE and briefly reviewed progress and approaches. This has led to a number of plans being evaluated and reviewed in joint ICES STECF meetings. This meeting draws primarily for reviews by ICES but includes relevant experience from those involved in STECF as well.

1.2 ICES Resolution and Terms of Reference

2012/2/ACOM39 The **Workshop on guidelines for management strategy evaluations** [WKGMSE] will meet 21–23 January 2013 at ICES HQ, Copenhagen, chaired by John Simmonds, UK and Dankert Skagen, Norway, to:

a) With reference to the work of SGMAS (particularly the 2008 report, section 5) and WKOMSE, review and bring up to date the methodologies and technical specifications that should be incorporated in MSE.
b) Consider appropriate risk definitions for MSE, taking into account practices in ICES and elsewhere and other relevant aspects (e.g. short-lived versus long-lived species).
c) Develop a set of guidelines for MSE evaluations in ICES and prepare a document with these guidelines. This will be a living document that will serve as reference for MSE in ICES.

WKGMSE will report to ACOM by February 20, 2013. A preliminary report should be available for WKMSYREF, to be held following WKGMSE.

1.3 Approach to the ToRs

ToR a was addressed primarily through an evaluation of recent plans and a review of the guidelines given in SGMAS 2008. To facilitate this review a template to describe the elements of recent plans was prepared in advance of the workshop. This was circulated among participants and a total of 18 plans that had been evaluated since

2008 were documented. The workshop was organized with an initial session to review this work and draw out the main similarities and differences of approach. The completed evaluation sheets are given in Section 2 below with a brief summary of the conclusions.

In order to carry out ToR b an evaluation of the way the Precautionary Approach had been interpreted among these 18 plans was carried out in advance of the meeting and the results were presented. Section 3 presents a summary of this analysis, a more complete review is attached as Annex 2. Section 3 also contains the recommendations for PA resulting from the discussions. These criteria would need to be endorsed by ICES before they become policy.

ToR c (Section 4) was addressed through substantial extension of the guidelines taken from SGMAS 2008. Section 4 provides standards and advice for conducting MSEs. This is split into main sections dealing with the operating model and its biological basis, including variability in the fishery and the observation model and how to drive suitable errors. It is recognized that the level of complexity must necessarily be case specific and related to the resources available. However, the template is recommended as a good way to give a checklist of what is considered and to record the approaches chosen. Section 5 provides a brief description of the overall process of developing a plan with some guidance for the roles and responsibilities of the different participants. While every case is different this is intended to draw attention to the activities involved and to indicate who might be tasked with the different aspects.

Section 6 provides guidelines for reporting, including the template for use with future plans.

Section 7 gives a summary and links to a range of useful software.

2 Recent experience

Participants were asked to fill in a reporting template covering some important aspects of recent management plan evaluations. These forms are attached as Annex 1 to the report. Here, we give a brief summary of the results.

The initiative to develop a management plan mostly came from managers, but in some cases from the industry. In practice, the communication between industry and management may be tighter that this, but there seems to be a range from bottom-up processes (e.g. Celtic sea herring) to top-down (most EU-Norway shared stocks). In only one case (Barents sea capelin) the initiative apparently came from science.

The formal process was mostly a request from competent management bodies to ICES, but for some stocks such as sole in the Bay of Biscay STECF constituted the formal evaluation body.

In practice, almost all simulation work was done at national institutes, or sometimes in cooperation between institutes. In many cases, the cooperation was formalized and supervised by an ICES or STECF workshop. This illustrates that the effort associated with developing and evaluating a management plan is well beyond the scope of a brief meeting or single workshop. A formal workshop is sometimes useful to consolidate the work, however, and present it for final approval by e.g. ICES.

The software used for simulations varied considerably. FLR was used as the main tool for 4 of the 18 stocks presented, HCS for 2, for the others, software was developed *ad hoc* specifically or the purpose, but often applied subsequently to neighbouring stocks. Examples are PROST, which was developed for NEA Cod, and used subsequently for NEA haddock and saithe, and the ADMB/R software developed for Icelandic cod that was subsequently used for Icelandic haddock and saithe.

The reason for choosing the software was not asked for specifically, but the impression is that institutional experience and investments in software are important factors. This is not surprising, but may be a matter of concern if there are very different solutions to common problems in the various programs, and they rarely get compared. In some cases, like Barents Sea capelin and BoB anchovy, it was quite necessary to develop software to accommodate specific needs, but in others, it might be worth requesting a clearer justification for the choice of simulation tool. In some cases multiple software packages were run and this did find minor issues within some packages.

When conditioning the operating model, most studies have paid a good deal attention to the recruitment, with different solutions in each case. Weights, maturities and selections are mostly just recent averages, with stochastic variability in some cases and density dependence in a few. Natural mortality is always constant. In cases where it can vary in the assessment, a recent average is used. Initial numbers are always taken from the most recent assessment. In most cases, it is stochastic, though in one case 25, 50 and 75 percentiles were used. The way the parameters of the distributions are derived is not always stated, but where it is, the inverse Hessian is a common source. There are some examples (NE Arctic stocks) where simulations have been done with fishing mortalities at historical levels, to verify that the model reproduces the level of stock abundance seen historically.

Of all the software used, all tools except FLR use the 'short cut' approach rather than doing a full assessment within the simulation loop. Hence, only 4 out of 18 evaluations used a full assessment. Obviously, ICES is willing to accept evaluations with the short cut option. However, the way this option is practiced varies a good deal, and there may be a case for further investigations on how to best imitate an assessment-prediction procedure.

When doing an assessment within the loop, apparently a log-normal error is assumed on the surveys that go into the assessment, with sigma of 0.2 - 0.3, while catches are often without error. In several cases, XSA was used in the loop as a substitute because the assessment done by the WG could not be included in the simulation software, and in some cases, different input data were used. Verifying that the assessment performs in line with the WG assessments does not seem to be common practice.

With the 'short cut' approach, the error is mostly a combination of an age factor and a year factor (or only a year factor if the decision is based on a biomass without projection). In some cases, the year factor has been calibrated to reproduce the CV of the biomass in the assessment. Projecting the stock forward in the decision model is always done where needed, but sometimes with assumptions that differ from those of the WG. Implementation error has only been included in a few cases, but sensitivity to implementation bias has been explored in some cases where that was a concern.

Most of the rules are F-rules, but there are examples of harvest rate rules, TAC rules and escapement rules. A percentage rule has been included to stabilize catches in most cases. The problem of getting trapped by low TACs has been solved in various ways. In Iceland, a filter rule is used instead of a percentage rule, and seems to work well.

Both risk type 1 and type 2 (see Section 3 for definitions) have been used, although risk type 1 is most common. In rebuilding situation, the probability of rebuilding the stock to a certain level within a given time frame has been the criteria for acceptance.

In summary, there has been a diversity of solutions and practices, to a large extent depending on the institution that did the simulations. That is not necessarily bad, but some minimum standards may be desirable. This is further discussed later in the report.

3 ICES Precautionary Approach Evaluation Criteria.

3.1 Sources of variability – what does risk cover

A criterion that must be considered in the evaluation of a harvest control rule (HCR) for a management plan is whether it is in conformity with the precautionary approach. This requires consideration of the probability of the stock biomass (typically SSB) being below the limit biomass reference point (B_{lim}) when the HCR is used. For an HCR to be considered precautionary, it is usual to request that this probability should not exceed 5%.

When conducting an MSE, the value obtained for the probability that SSB is below B_{lim} can depend strongly on assumptions made during the MSE, such as those concerning the operating model, assessment and implementation errors. It is therefore very important that the assumptions made in the MSE are realistic and encompass the range of situations considered plausible in reality. Section 4 of this report provides guidelines in this respect.

3.2 Definitions (percentage, time frame)

There are alternative ways in which the statement "the probability that SSB is below B_{lim}" can be interpreted and different interpretations have actually been applied when management plans have been evaluated in the past by ICES. The issue is important because, depending on the interpretation used, the request that this probability should not exceed 5% is more or less stringent. The working document by Fernández (WD1 in Annex 2) explains this in detail and a summary is provided here (noting that instead of "risk", which is the wording employed in WD1, this report uses the wording "probability that SSB is below B_{lim}" to avoid confusion with other interpretations of risk).

A review of ICES practices (see e.g. section 2 of this report and section 6 of Annex 2) shows that three interpretations have been used in the past:

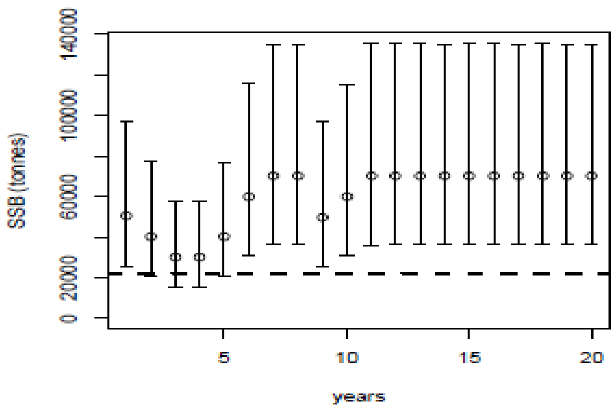
- $\boldsymbol{Prob1}$ = *average* probability that SSB is below B_{lim}, where the average (of the annual probabilities) is taken across ny years.
- $\boldsymbol{Prob2}$ = probability that SSB is below B_{lim} *at least once* during ny years.
- $\boldsymbol{Prob3}$ = *maximum* probability that SSB is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

Annex 2 shows that $Prob2 \geq Prob3 \geq Prob1$, so requiring that $Prob2 < 0.05$ 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 (generally in the "long term", after the effect of the initial stock numbers has disappeared), $Prob3 = Prob1$, although 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 time 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.

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 in each year, together with a horizontal line indicating where B_{lim} is. 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.

1.



2.

3.

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

Table 3.2.1											
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}. There is more than 5% probability that SSB is below B_{lim} in years 2, 4 ad 5 of the simulation, whereas it is less than 5% in all other years, including in the long term.

Table 3.2.2 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 3.2.1. This is not the case for $Prob2$, whose value depends on the amount of time autocorrelation in SSB. The $Prob2$ values shown in Table 3.2.2 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 3.2.2			
	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

3.3 Precision – iterations needed

MSEs perform stochastic simulation for a period of ny future years, based on a number $niter$ of independent iterations (sometimes also called replications, realisations, etc). Population, catch, risk statistics, and many other quantities of potential interest, are used to summarise performance of the MP over the ny year period. These statistics (including probabilities) are calculated based on the $niter$ independent iterations. Depending on how the simulation is set up (e.g. how assessment errors are dealt with or how it is programmed), carrying out a large number of iterations can be very time consuming. Sometimes in the past, as few as $niter = 50$ iterations have been used, though such a small number is unusual.

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 3.3.1 Distribution of $P(SSB < B_{lim})$ computed based on $niter$ iterations, when $prob = 0.05$ ($prob$ is the value of $P(SSB < B_{lim})$ obtained if an infinite amount of iterations could be performed)		
$niter$	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

Table 3.3.1 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.

The intervals in Table 3.3.1 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 3.3.1 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 3.3.1 because in $Prob1$ an average is taken over several years, which increases precision (although the gain in precision is less the more auto correlated SSB is). A simple simulation exercise showed that in a stationary situation, the interval in Table 3.3.1 reduces to [0.04, 0.06] already with $niter = 250$, when $Prob1$ is computed as a 10-year average, even under high autocorrelation in SSB (such as 0.8).

On the other hand, the computation of $Prob3$ is less precise than Table 3.3.1 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. In the stationary situation, given that $Prob3 = Prob1$, only $Prob1$ should be computed (because of the much better convergence of the algorithm to compute $Prob1$).

In the short term, where the situation is non-stationary, it makes sense to consider annual $P(SSB < B_{lim})$ for each of the years, as indicated in Section 3.2. When each year is seen in isolation, the intervals in Table 3.3.1 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, and then a maximum taken over the ny years. To illustrate the effect of this, imagine that $P(SSB < B_{lim})$ (based on an infinite amount of iterations) is < 0.05 in all years and that $niter$ iterations are used in the 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 a wrong conclusion for that particular year. Using the same amount of iterations, it is intuitively clear that the probability of reaching wrongly the conclusion 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 value that would be obtained if an infinite number of iterations could be performed (more often than not the computed value of $Prob3$ will be too large). The bias is stronger 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 less time auto correlated SSB is.

Conclusions:

- For $Prob2$, $Prob1$ and $P(SSB < B_{lim})$ in a specific year y, the intervals in Table 3.3.1 can serve as guidance.
- In most cases, $Prob1$ requires fewer iterations than suggested in Table 3.3.1 (taking advantage of averaging over years, but the gain in precision is less the more auto correlated SSB is).
- Computing $Prob3$ requires more iterations than suggested in Table 3.3.1 (potentially many more, as the computed value can converge very slowly) and the same holds for computing $P(SSB < B_{lim})$ for each of ny years and then focusing on the highest of these probabilities (since this is equivalent to computing $Prob3$). In the stationary situation, $Prob3 = Prob1$ and only $Prob1$

should be computed. In the non-stationary situation (i.e. short term), the following "solution" could be adopted for $Prob3$ computation:

1. Start by computing $Prob3$ based on the number of iterations in Table 3.3.1
2. If the computed $Prob3$ value is below the lower end of the interval in Table 3.3.1, then it may be concluded that $Prob3 < 0.05$ (given the bias in the $Prob3$ computation).
3. Otherwise compute $Prob1$ and $Prob2$ for the same range of years as $Prob3$.
   (3a) If the computed $Prob1$ value is above the upper end of the interval in Table 3.3.1, then it may be concluded that $Prob1 > 0.05$ (and the same, therefore, holds for $Prob3$).
   (3b) If the computed $Prob2$ value is below the lower end of the interval in Table 3.3.1, then it may be concluded that $Prob2 < 0.05$ (and the same, therefore, holds for $Prob3$).
   (3c) Otherwise no conclusion can be reached regarding $Prob3$. In this case, the number of iterations should be increased until the value of $Prob3$ stabilizes in an area where conclusions can be drawn.

- It is recommended that the relevant measure used in the analysis ($Prob1, Prob2$ or $Prob3$) be plotted against iteration number as follows: compute the relevant risk measure based on the first $iter$ iterations and plot it versus $iter$ (iteration number), to get an understanding of how long it takes for it to stabilize in an area where conclusions can confidently be drawn.
4.

3.4 Considerations with respect to MSY

In the development of management plans using the approaches defined here the evaluations should include information that is useful in setting values for MSY. For example, a harvest control rule based on a long term F strategy with reductions in F under some circumstances may deliver yields that are maximized and sustainable in the long term. Thus the evaluation can estimate F_{msy} and related ranges of biomass needed in the ICES MSY approach. Such targets will be similar to the management plans that aim at high long-term yield, although F_{msy} may be expected to be slightly higher than the F_{target} in the management plan if the management plan includes a term to stabilize catch and or significant observation error. In such cases, the group carrying out the MSE should evaluate the method and, if acceptable, apply it to recommend revised values for use in the ICES MSY approach.

3.5 Revision of reference and limit points

In developing the MSE parameters consideration needs to be given to other parameters used in management, such as B_{pa}, B_{lim}, F_{pa} and F_{lim}. In developing the parameterization of the model in the MSE it is quite likely that values of these parameters are implicit given the data and model choices, for example B_{lim} and F_{lim} can be obtained from the S-R model parameterization (See ICES 1998 report on Precautionary Approach). In this case, these should be compared to ICES limit reference points and, if considered appropriate, modified values proposed. In this context, if the stock being modelled has experienced little fishery dynamics, then it may be difficult to define B_{lim} and B_{pa}, particularly if B_{lim} is based on B_{loss}. In this case the group should carry out an evaluation of B_{lim} and F_{lim} in the context of similar stocks and evaluate if these values can be inferred better from external data (See WKFRAMEII report, ICES 2011, for example). If more suitable alternative PA reference and limit points coherent with the MSE can be estimated then these should be proposed along with the MSE.

3.6 Recommendation for ICES PA practice

ICES is explicitly required to evaluate management plans as conforming to the precautionary approach or that the objectives of the plan are consistent with MSY. The Precautionary Approach of ICES uses B_{pa} and B_{lim} to define precautionary management, which implies 5% probability of SSB<B_{lim}. Each year SSB is estimated (from an assessment) and if found to be < B_{pa} some remedial action would be proposed. Under management plans, requiring that Prob3 < 5% to consider the plan as precautionary is closely analogous to this approach. Each year in the simulation, both in the short and long term, is examined and action occurs if necessary. It is perhaps important to note that Prob1 = Prob3 in the long term stable or stationary situation. Prob3 is preferred over Prob1 for considering a plan as precautionary because it allows for both recovery periods and long term stationarity and may be applicable in systems with regime shift. However, the use of Prob3 < 5% (as opposed to e.g. the stricter Prob2 < 5%) implies that SSB goes below B_{lim} for stocks where F_{msy} is close to F_{pa}, so for these stocks checking that the management plan delivers recovery from below B_{lim} must be demonstrated. It is proposed that this should be done following the procedure carried out in the evaluations of North Sea sole and plaice (Coers *et al*, 2012 and Simmonds *et al*, 2010) where recruitment is reduced in the simulation until the stock declines to B_{lim} and then this scenario is continued and it is checked that SSB recovers above B_{lim} under the plan without additional intervention.

This approach for considering a management plan as precautionary based on Prob3 < 5% is pragmatic and does not imply revising ICES endorsement for any existing plans. Nevertheless, this precautionary criterion implies an implicit understanding that although SSB < B_{lim} should generally be avoided, going below it is not catastrophic and can be expected on occasions. Should managers require higher probabilities of maintaining SSB > B_{lim} this should be specified as part of their request to ICES for the evaluation. WKGMSE regards this choice of precautionary criterion to be compatible with historic classification of plans and, thus, historic classifications do not need to be revised.

A recovery plan (or an initial recovery phase within a long-term management plan) cannot be judged using the same criterion for precautionarity. If a stock's SSB is currently below B_{lim}, it is not logical to expect that P(SSB < B_{lim}) < 5% in all years of the simulation, including the initial recovery phase. It seems more logical to judge a recovery plan (or an initial recovery phase within a long-term management plan) according to its ability to deliver SSB recovery within a certain time frame that is appropriate for that stock (e.g. for a stock with around 5-10 cohorts in the fishery 5 years from the start of the plan). In that case, the requirement for considering the recovery plan as precautionary would be that the probability of SSB > B_{lim} in a prespecified year is ≥ 95%. If the recovery plan constitutes an initial recovery phase within a long-term management plan, the usual evaluation procedure and standards (including the requirement that Prob3 < 0.05) should be applied to the after-recovery long-term management plan. For a plan with only a recovery state the evaluation should state if the recovery plan is or is not expected to be precautionary once the stock has recovered above B_{lim} with a 95% probability.

It is recognised that some short lived stocks can go below B_{lim} naturally under conditions of zero fishery. Such stocks can be considered for precautionary management under a slightly amended approach. We define a factor 'a' which might initially be set to the value two. Stocks that are considered differently are those for which the probability of SSB < B_{lim} is ≥ 5% with F=0. For such a stock, a management plan could be considered as precautionary provided this probability does not increase by more than 'a' times under the management plan, where 'a' is an arbitrary number. Currently 'a' might be implemented as 2 but the effect of this number needs to be explored further. This regime implies a zero catch as part of the plan when the stock approaches or goes below B_{lim}.

Increasingly, ICES is requested also to examine consistency with MSY as part of the management plan evaluation. One option would be to examine the "real F" values that the management plan produces during a range of years in the simulation (e.g. the first 20 years in the simulation or another range of years considered appropriate) and to categorise the plan as MSY-consistent if there is less than 50% probability that "real F" exceeds Fmsy for the ensemble of years. This does not mean requiring that the condition holds for each and every year. Depending on the design of the harvest control rule, it would be possible for "real F" to be above Fmsy in some years with a high probability and below Fmsy in other years, and the plan would still be considered MSY-consistent if the less than 50% condition holds when the ensemble of years is considered together.

4 4 Guidelines for simulation

4.1 Building blocks in simulation procedures

This section is a brief outline of the building blocks, with terminology as used in this report. Briefly, a simulation procedure is composed of:

- An **operating model**, which represents a realization of a biological model for the 'real world' that shall be examined.
- An **observation model** that extracts, with error, information from the operating model that is used in the decision process.
- A **decision model**, in which a decision on removals (typically a TAC) is derived from the outcome of the observation model.
- An **implementation model**, which translates the decided removals into actual removals from the real stock.

In a simulation framework, these models constitute a loop, which is repeated for a number of years. Each sub-model has stochastic elements. Each of these steps is discussed in detail in the following.

4.2 Choice of model and modelling approach.

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

One purpose may be to outline candidate plans for a stock with some, perhaps conflicting objectives, and to show trade-offs between objectives. If so, one may want to scan over a large range of rule parameter options, and test for sensitivity to a variety of assumptions. This will require software that is fast, typically software without assessments in the observation model.

Once a proposed rule is reached, it can be further examined, with the same or other methods. At this stage, a key issue is that the operating model reflects the biology of the stock and the observation model reproduces the actual assessment as far as at all possible. The computing time is of minor importance, but much effort has to be put into validating the model conditioning. The same applies if a single rule is presented for approval.

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

4.3 Operating model (true biology)

4.3.1The biological operating model is intended to reflect the "true" dynamics of the stock productivity. Key elements of this are growth, recruitment, natural mortality and sexual maturation. The dynamics of these processes need either to be modelled or have their variability captured by the operating model. This process called conditioning is fundamental to the simulation, and should be addressed completely before final simulations are run to test proposed harvest rules. Some important aspects of this are considered below:

In general, most of the parameters of an operating model are obtained by fitting them to historical data using frequentist or Bayesian methods. This "conditioning" process ensures that the parameter values used in the projection period are consistent both with the available data and how the system is understood.

Uncertainty in the values of the parameters (i.e. usually observations; sampling and measurement error) of the operating model is usually based on samples obtained using bootstrapping, from Bayesian posterior distributions, likelihood maximisation in a frequentist approach and taking into account several sets of parameter values in each alternative operating model specification.

However, alternative assumptions, models, and error structures need also to be considered when selecting the uncertainties to include in an operating model (McAllister and Kirchner, 2002; Hill et al., 2007), so that the developed management strategies are robust to errors in the model structure. The process of selecting which alternative structural models to include in an MSE study begins with defining the plausible range of hypotheses and the parameter values that are to be used in the operating model. Defining alternative hypotheses and scenarios, as well as assessing their plausibility, can be obviously a difficult task.

4.3.1 Initial population vector:

In some cases this has been implemented as simply taking the final population vector from the most recent robust assessment (e.g. Norway Pout). However the initial population vector will influence the perception of risk in the short term. Therefore it is important to appropriately include information on the uncertainty in the initial state of the true stock being simulated. Using the input vector of the most recent assessment forecast and applying the estimation uncertainty (at age) from the assessment to the values has been applied in the case of NEA mackerel to reduce this sensitivity. Or in cases where the assessment is not very robust (e.g. western horse mackerel) a recently converged population vector from the assessment was used and a cv applied to this vector representing the assessment precision. In terms of a sensitivity analysis a range of scenarios of population vectors could be chosen as the initial values, to check for e.g. efficacy of the HCR to; a depleted stock state, or controlling exploitation rate on a declining stock.

Of specific interest is the youngest year classes in the starting vector. Often these are particularly uncertain and the CVs from the assessment may imply more uncertainty that the intrinsic variability represented by stochastic draws from the S-R function (see 4.3.2). In such a situation use of the assessment CVs directly is not recommended, recruits could be drawn from S-R function for each iteration or the CV reduced to the CV of the S-R 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.

4.3.2 Recruitment

In the 2008 SGMAS report the following was considered: A minimum standard is a single stochastic stock recruit model to reflect potential variability. It is recommended that modelled recruitment not be implemented stochastically from a fixed S/R fit, but rather that the parametric fit should be stochastic such that for e.g. recruitment is drawn from around a different mean at each iteration. (in the case of a hockey stick model). Accounting for temporal dynamics (eg. 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).

4.3.2.1 Choice of stock-recruit function

If a single S/R 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 bootstrap of residuals or use of a fitted statistical distribution (truncated as necessary). If bootstrap methods are used care needs to be taken to ensure autocorrelation is included.

The choice of stock-recruit model may be critical to the performance of the rule, even when the fit of different models to the historical data is almost equal. If the choice of S/R 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 S/R models and testing a range of HCRs under each circumstance. In particular if there is a great deal of uncertainty in the slope of the S/R relationship near the origin or in the recruitment at large stock biomasses, different options must be tested. If the HCR results are relatively insensitive to these choices one model may be chosen for further work.

If following this investigation it is found that the performance of the HCRs being tested are critically dependent on the choice of S/R or growth models, then multiple models with different parameters can be selected using for example the method of Michelsens and MacAlister (2004) and described in the NEA Mackerel evaluation (ICES 2008). This method provides a formal way of including uncertainty in the form of the S/R functional relationship, parameters and stochasticity in the evaluation. Figures 4.3.1-2 shows an example of NEA mackerel.

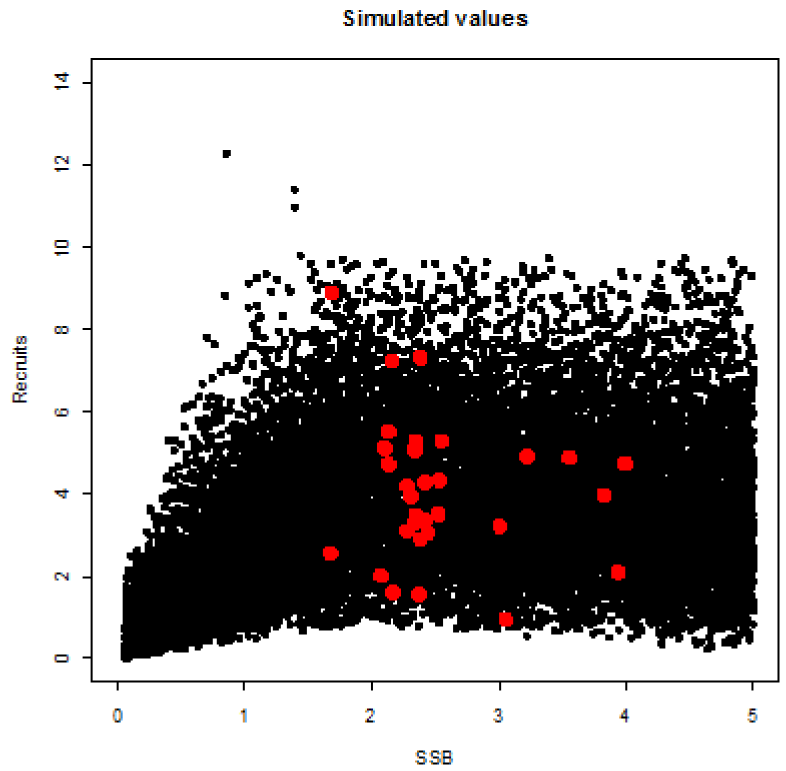


Figure 4.3.1. Simulated and observed stock recruit pairs, where the simulated pairs were drawn from multiple stock-recruit relations. Example of NEA mackerel showing comparison of observed (red) and simulated (black) recruitment for a) SSB from 100,000 to 5M tonnes SSB,

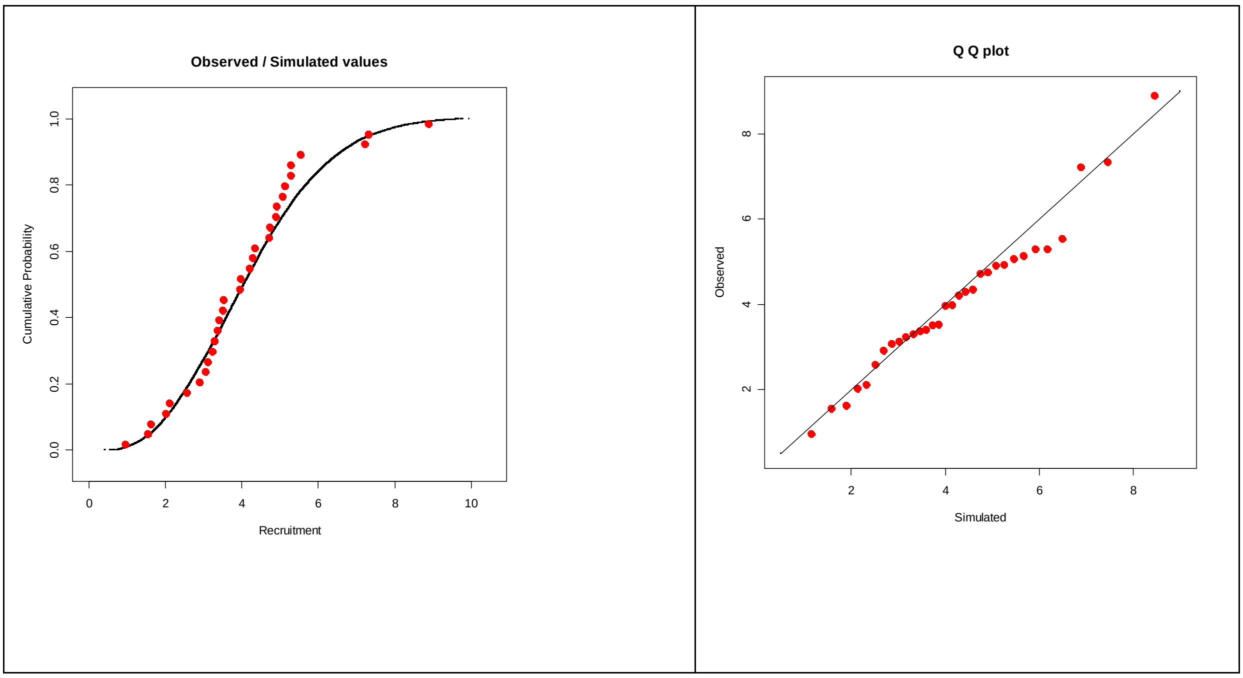


Figure 4.3.2. Example of NEA mackerel showing cumulative probability distributions of observed and simulated values for observed SSB (left) and Q Q plot of observed and simulated values for observed SSB (right) . Simulated values were derived from 1000 models with Hockey-stick and Ricker functional forms and Normal or Log Normal stochastic deviation

4.3.2.2 Accounting for temporal dynamics.

The general problem will be 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, are fluctuating beyond independent random variations.

In some cases, introducing autocorrelations may give an adequate representation of this fluctuation. In other cases, in particular if there are periodicities or trends, such dynamics may be included in the stock-recruit function parameters. However, that implies predicting future fluctuations, which requires that such predictions are well justified.

The alternative would be to specifically examine the robustness of the rule to such fluctuations, and require that the rule should function with a realistic range of future recruitment regimes. Such robustness testing may be done by inducing changes at fixed times, and examine the response.

An additional aspect that requires careful consideration is that with externally driven recruitment fluctuations, the historical stock-recruit data to a greater or lesser extent will reflect the SSB as a function of the previous recruitments, which will make the estimates of stock-recruit parameter values invalid. Testing the correlation between SSB and past recruitments may provide some warning.

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 4.3.3) 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 to get the probability of large year classes right, but does not inform about the intervals between such year classes.

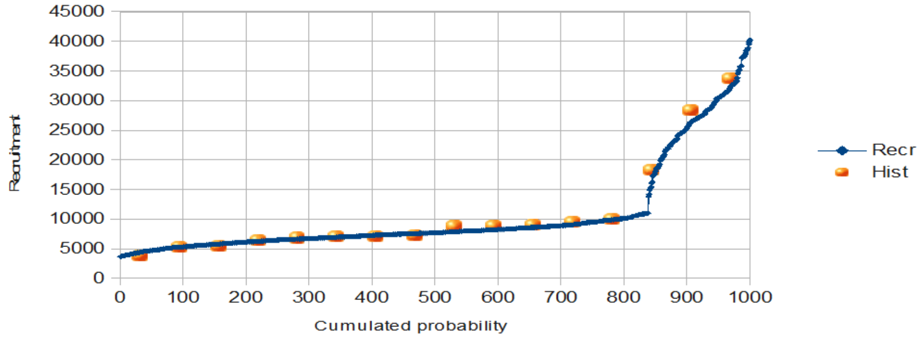


Figure 4.3.3. Cumulated distribution of simulated and observed stock recruit pairs. Blue whiting in a period with occasional large year classes.

4.3.2.3 Regime shifts (RS)

If it is likely that growth or recruitment are dependent on environmental drivers then a plausible range of possible scenarios should be included. If climate models with forecasts are available, then stochastic variability due to environmental drivers could be included in the growth or recruitment models. If climate models, without being able to provide forecasts, indicate that major shifts in stock productivity, through carrying capacity, reproductive capacity or growth may occur, such alternatives should be included as robustness tests.

However philosophically it might be fruitful to consider the following question: How can we sensibly identify ecosystem parameters of importance for a particular fish stock regarding RS, when we have no clue on which parameters that are influencing recruitment variability (except SSB) - are we introducing an inconsistency in our system by considering RS?

This issue of RS 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 a short time series in the context of estimating S-R parameters.

Questions that should be addressed when considering regime shifts include: Can individual years be regarded as a RS or is that better dealt with as noise? What about two years, three years etc.? Is there a minimum length in terms of 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 is also important to realise 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 and 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.

RS 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 S/R seems thus to be in a new Regime. Thus, theoretically fisheries management can in this case turn the regime back if wanted.

It is also worth considering that when a RS has been identified, is it then best to completely ignore data related to the anomaly period or can some useful information be extracted from e.g. the S-R prior to the RS?

The answers to these questions are not obvious. For the purpose of evaluating management plans, one guideline may be that the plan 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.

4.3.3 Selection and weight at size

Selectivity in the fishery appears in several contexts in a simulation, and should not necessarily be the same in all contexts:

- When generating catches that are input to assessments in the observation model
- When translating a decided TAC to removals in numbers at age in the implementation model
- When deriving catches at age in the decision model

Weights at age also appear in several places:

- Translating numbers to biomasses in the operating model, which propagates to stock recruit functions and possibly to density dependence models
- Translating numbers to biomasses in the observation model, which typically is needed for providing a decision basis
- Translating catches in numbers to TACs in the decision model
- Translating TACs to catches in numbers in the implementation model

Thus, uncertainty will contribute to the range of true stock scenarios, and to errors in decisions. The selections and weights should at least in principle not be the same in the true world (operating and implementation models) as in the decision-makers world. In the observation model, the uncertainty can be applied directly to weights and selections, or to the 'observed' catches and biomasses.

Trends in historical weights at age and selections are common. If such trends are continued in the future as linear trends, the values will sooner or later become unrealistic. If just the mean is taken over a period with a trend, the values in the future will be assumed to be different from the recent past, which may not be realistic. Often a mean over a recent period is applied (Icelandic cod is one example), which implies that it is assumed that trends are broken and values will continue at the present level. Again, there is no universal recipe, but the choice should be justified, and the implications made clear.

Uncertainty in selectivity at age and weight at age might have a large impact on the outcome of an MSE evaluation. Particularly, weight at age affects directly the estimation of the SSB. The variation in weight at age is commonly linked to both density-dependent (i.e. intra or inter species effects) and density independent processes (i.e. environmental effect) but also to interactions with other species in the ecosystem (i.e. ecosystem effect or links with other component of the ecosystem). Thus, it is important that uncertainty in weight at age reflects the observed historical uncertainty (observation uncertainty) but also any other known process which might affect growth during the projection period (i.e. process uncertainty). It is also important to stress that processes uncertainty might be caused not only by temporal but also by spatial variability in the dynamic of the population.

MSE are generally run contingent to the current situation in terms of selection at age and they are valid only under the assumed conditions. Both selectivity at age and weight at age also have a direct effect on MSY level in terms of long term yield and on the level of F associated to MSY. Thus, exploring the sensitivities of the MSE to uncertainty around selectivity at age and weight at age is important, along with study to better understand behavior of the species and the fleets (or their interaction) as related to selectivity. Some assessment models such as SS3 for example are able to provide estimates of selectivity and associated uncertainties. When such estimates are not readily available, a way to estimate uncertainty in selectivity could be to use smoothed selectivity curves in catch curve analysis, and use catch curve prediction intervals to determine uncertainty in the estimation of selectivity. However, also investigating the sensitivity of the MSY estimation, in terms of absolute level of catches, to the selectivity at age is essential as MSY is directly dependent on selectivity at age. Also, selectivity is in theory directly affects the structure of the population at the equilibrium and thus has direct implications on the interaction of a given species to the rest of the ecosystem through both top down (i.e. predation) and bottom up mechanisms (e.g. sensitivity of recruitment to climate changes mediated by the population structure).

4.3.4 Natural mortality

4.3.4.1 Constant natural mortality used in the assessment

In most assessments a year independent natural mortality (M) is used. This natural mortality has to be chosen also for the MSE simulation as the historical F and biomasses are linked to the chosen M. Using alternative values of natural mortalities in 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 M in the projections is easy to make, but it is difficult to evaluate the results without a change in the historical values of M as well. Such an exercise could be done as part of the assessment benchmark, but is not mandatory as part of a MSE.

4.3.4.2 Time variant natural mortalities used in the assessment

When time variable M are used in the assessment (e.g., North Sea cod, North Sea Herring) the estimates from the latest period (terminal year if smoothed values are used, average over a suitable time period if not) can be used in MSE for short term evaluations. For longer-term simulations (and recovery scenarios) the effect of a variable M has to be investigated, either as a part of a sensitivity analysis or modelled explicitly.

4.3.4.3 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 preys (functional feeding response). MSE simulations do normally just provide information on one particular species such that the changes in M cannot be estimated. The range of historically natural mortalities is available from the assessment (and used there) which makes it possible to test the robustness of the HCR to the observed variability in M. This can be done by e.g., min-max scenarios or by bootstrapping from the observed distribution of natural mortalities over time. It has to be decided from what historic time period values should be tested or bootstrapped (e.g., from times with low or high predator stock biomasses).

4.3.4.4 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 in MSE can lead to very different conclusions about the performance of the HCR (e.g. cod recovery in the North Sea; ICES 2004)) and cannibalism must be included in the MSE, at least for long term simulations and recovery scenarios.

ICES WGSAM (2011) has made a first approach to model predation mortality based on simple relationships between predation mortality and the biomass of predators. This approach can be used as it is, however with the biomass of the species considered (e.g. cod) estimated in the MSE. It will also be possible to estimate the relation between the partial predation mortality and the species itself, assuming a constant population of other predators. Such approach will deliver a simple relation: $M_{age\ 1} = a + b * SSB$, 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. Only cannibalistic effects on older age groups not covered by the stock-recruitment relationship should be modelled explicitly in this case.

4.3.5 Modelling ecosystem effects on the stock

The ecosystem can influence stocks in many different ways. Environmental factors influence recruitment success, food availability, growth, maturation, the spatial distribution of stocks, predator-prey relationships, just to name a few. This makes the prediction of ecosystem effects very difficult if not impossible. Some ecosystem effects have been explicitly included in assessments (e.g., predation mortalities, SST dependent recruitment for Baltic sprat) and should be included in the MSE by default.

Although MSE simulations are often carried out using long-term projections to study the behaviour of HCRs and to run populations into equilibrium, they are also used to inform managers about what will likely happen in the short- to medium term. MSE simulations are parameterized based on the current (or historically observed) ecosystem state and results are only valid under the assumption that the current (or historic) state will prevail in the future. They should not be used in the sense of long-term ***predictions*** as it is impossible to predict e.g., regime shifts.

There are two options to cope with this situation:

1. Management plans have to be re-evaluated every few years. Before each evaluation it has to be analysed whether the ecosystem and so e.g., recruitment dynamics or weight at age in the stock is different to what was observed in the evaluations carried out before. The parameterization has to be adapted accordingly.

2. If relationships between specific environmental factors, ecosystem components and fish stocks are known, sensitivity analyses to test the robustness of HCRs can be carried out or relationships can be modelled directly in the MSE where possible. However, it has to be also decided whether a relationship observed in the past can be

expected to hold in the future. An overview of ecosystem states and their potential effects on fish stocks may be found in the report from the Workshop on Ecosystem Overviews (WKECOVER 2013). Also reports of integrated assessment working groups (e.g., WGINOSE 2012, WGIAB 2012) provide useful information.

4.3.6 Modelling Indicators required under MSFD (Marine Strategy Framework Directive)

Descriptor 3 for determining Good Environmental Status (GES) under the MSFD was defined as "Populations of all commercially exploited fish and shellfish are within safe biological limits, exhibiting a population age and size distribution that is indicative of a healthy stock" (Directive 2008/56/EC, Annex I). In MSE evaluations described here it may be necessary or interesting to indicate the effect of different options on MSFD descriptor 3.

In the Commission Decision 2010/477/EU three criteria including methodological standards were described for descriptor 3. The three criteria and associated indicators are:

Criterion 3.1 Level of pressure of the fishing activity

- Primary indicator:

  Indicator 3.1.1 Fishing mortality (F)

- Secondary indicator (if analytical assessments yielding values for F are not available):

  Indicator 3.1.2 Ratio between catch and biomass index (hereinafter 'catch/biomass ratio')

Criterion 3.2 Reproductive capacity of the stock

- Primary indicator:

  Indicator 3.2.1 Spawning Stock Biomass (SSB)

- Secondary indicator (if analytical assessments yielding values for SSB are not available):

  Indicator 3.2.2 Biomass indices

Criterion 3.3 Population age and size distribution

- Primary indicator: Indicator

  3.3.1 Proportion of fish larger than the mean size of first sexual maturation

- Primary indicator:

  Indicator 3.3.2 Mean maximum length across all species found in research vessel surveys

- Primary indicator:

  Indicator 3.3.3 95% percentile of the fish length distribution observed in research vessel surveys

- Secondary indicator:

  Indicator 3.3.4 Size at first sexual maturation, which may reflect the extent of undesirable genetic effects of exploitation

Both criterion 3.1 and 3.2 and both theirs indicators are normally model outputs under most of the MSE simulations so both true and observed values can be output. For Criterion 3.3 for some MSE models which include simulations at length the primary indicators 3.3.1, 3.3.2 and 3.3.3 can be modelled directly. For age based models without length addition of growth parameters and some variability can be used to give plausible length distributions. The secondary indicator 3.3.4 might be calculable but it is considered that's the results would not indicate the response being examined and would require extensive model development to give any result. Also any model aimed at informing on this would be driven directly by the model assumptions and it may not be particularly informative in this context.

4.4 Observation model (assessment- basis for decisions)

4.4.1 Assessment or short-cut: Pros and cons

When performing MSEs of proposed management plans, where the management plan relies on the application of an assessment model coupled with a short-term forecast and a Harvest Control Rule (HCR) (jointly referred to here as a management decision model) in order to set a TAC, an approach that is commonly used is to approximate the management plan for the purposes of the evaluation. This approximation typically takes the form of simulating the behaviour of the assessment model by generating values directly from the operating model (the underlying "truth") with statistical characteristics (e.g. variance, bias and autocorrelation) that is assumed to reflect the behaviour of the assessment model. This is referred to as the 'short cut approach' as opposed to a "full" MSE (Section 4.4.2). A further approximation is to ignore the short-term forecast required for the year following the final assessment data year but preceding the year for which a TAC is needed, known as the intermediate year, even when such a short-term forecast is performed in practice. Short-term forecast assumptions can differ markedly from the operating model, with potentially serious consequences for the performance of HCRs being evaluated. These consequences could remain hidden if the intermediate year lag is ignored when conducting a MSE, and the approximation in these circumstances can produce a different perception of how the HCR impacts the underlying "true" population.

Two examples of the comparison between a "full" MSE and a "short-cut" MSE (one where both of the above-mentioned approximations are made) are given in Kell *et al.* (2005) and ICES (2008). The first of these examined the effects (on stock biomass, yield and stability) of constraining interannual variation in TACs, and found that when ignoring both the assessment model and the short-term forecast, expected yield and SSB converged rapidly on the equilibrium yields, whereas when these were both included, 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 decision model to the operating model). The second study used the EU and Norway management plans considered by AGCREMP (ICES 2009) to compare a full MSE to one that ignores both the assessment and the short-term forecast, and came to a similar conclusion. It found that the short-cut MSE lead 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 only the assessment was ignored.

A further advantage of including an assessment model in a simulation loop is that the behaviour of some 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), and this behaviour may not be easily captured or anticipated when using approaches that short-cut the assessment.

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

It is recognized that, there may be computational difficulties when trying to include assessment models within an MSE that may warrant approximating the behaviour of these assessment models (over-long computer time, convergence difficulties, assessment models not amenable to automation, etc.); however, an important message from the above studies is that lags and assumptions made when applying the HCR to derive a TAC in practice cannot be ignored in the evaluation.

4.4.2 Assessment in the loop

A key feature of input data for an assessment in the management decision model is that they should have the same statistical properties as the input data that are 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 $I_{y,a}$ is fitted to abundance assuming a lognormal error distribution:

$$\ln I_{y,a} = \ln q_a + \ln N_{y,a} + \varepsilon \qquad \text{where} \quad \ln I_{y,a} = \ln q_a + \ln N_{y,a} + \varepsilon$$

then the values of q_a and σ_a are estimated (and if there is evidence for auto-correlation in a particular set of residuals, the extent of this should also be estimated). These estimates (including auto-correlation, if present) are used to provide a link between the operating model (from which $N_{y,a}$ is taken) and the management decision model (to which $I_{y,a}$ is supplied).

Model uncertainty (related to conditioning the operating model), should include the uncertainty in the parameters used to generate the input data and is discussed elsewhere, but in brief, such uncertainty can be included by, for example, bootstrapping the original model fit on the basis of observation equations, such as the one above, or by using the variance-covariance matrix from the original model fit, taking care that, for example, uncertainty at the youngest ages is consistent with the uncertainty coming from the stock-recruit relationship.

4.4.3 The "short-cut" approach

Whereas the challenge in the case of an assessment used in the loop was to ensure that the statistical properties of the input data matched those of data used in practice, the 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.

Ignoring the short-term forecast is not acceptable unless the management rule does not require that. Reproducing the assumptions made in the projection may be a challenge for the programmer, for example with regard to future recruitments, weights and selections, but that should not be an excuse for unrealistic simplifications.

In existing software, imitating 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. Usually, the stock numbers at age have to be generated, in order to enable imitation of the projection as practiced in management advice.

4.4.4 Validation

Validating the performance of the observation model is essential to ensure a realistic evaluation of management procedure performance, whether running a full or short-cut MSE

The field of reality checks of assessments is one where further development should be encouraged. There are no routine tests that can be universally recommended. The bullet points below could be worth considering.

- The behaviour of the assessment model in a simulation setting should match the behaviour of the assessment model in practice. In this regard, a useful check is to confirm that the statistical properties (e.g. bias, variance and auto-correlation) of a metric such as Mohn's rho, as calculated for the assessment model in practice, matches those for the same metric calculated for the assessment model as applied in the simulations (e.g. at the end of the projection period). A useful visualization plot may be to include the historical assessment error on some key metrics with that modeled in the future (Figure 4.4.1).
- Run the evaluation with zero F in future to check the behaviour of the population model
- Run the management decision model with perfect knowledge, and compare this with the management decision model with assessment error included to check the impact of this assessment error. It may be that the management plan is not precautionary even under perfect knowledge. This is also useful as a code check.
- Justify the approach used to characterise either noise in the input data (for full MSE) or parameters used when approximating the assessment model (for short-cut MSE) by making use of "reality" checks (ensure future noise is consistent with historically observed noise).
- Run the model starting some years back in time and condition it to reproduce the historical development of the stock (i.e. catches and recruitments in particular). Then compare the assessment errors by the model with the actual assessment performance, in particular with respect to retrospective errors.
- Run the evaluation by forcing Fs to be in the range of Fs experienced historical in order to check how the properties of the assessment model in the loop compares with it's historical behaviour in practice (Figure 4.4.2).

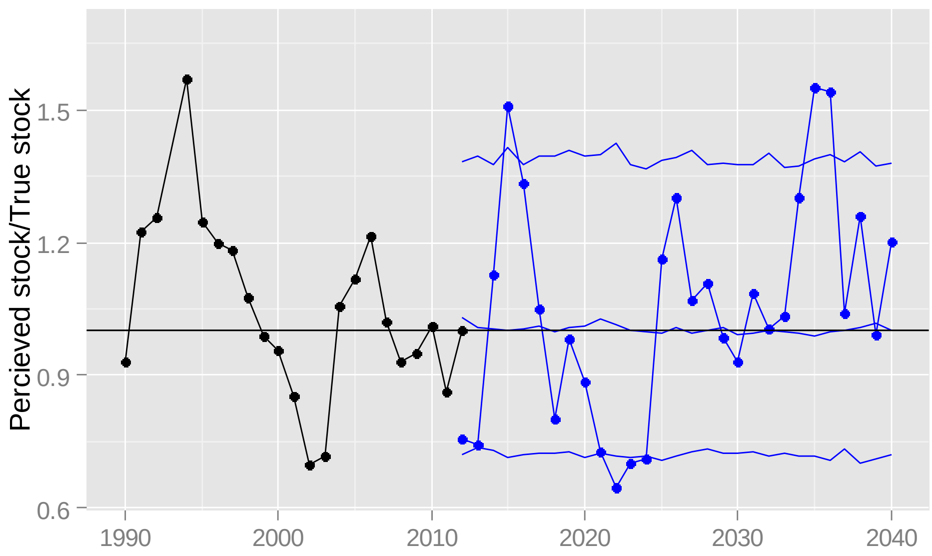


Figure 4.4.1. Example of the ratio of the perceived stock vs the true stock reference biomass. The historical part of the plot (black line) is the based on empirical retrospective performance (ratio of contemporaneous estimates vs. the most recent assessment) upon which future assessment error is based (ratio of observation model biomass vs. the true biomass). The line note the 5th, 50th and 95th percentile with one iteration shown as an example.

.

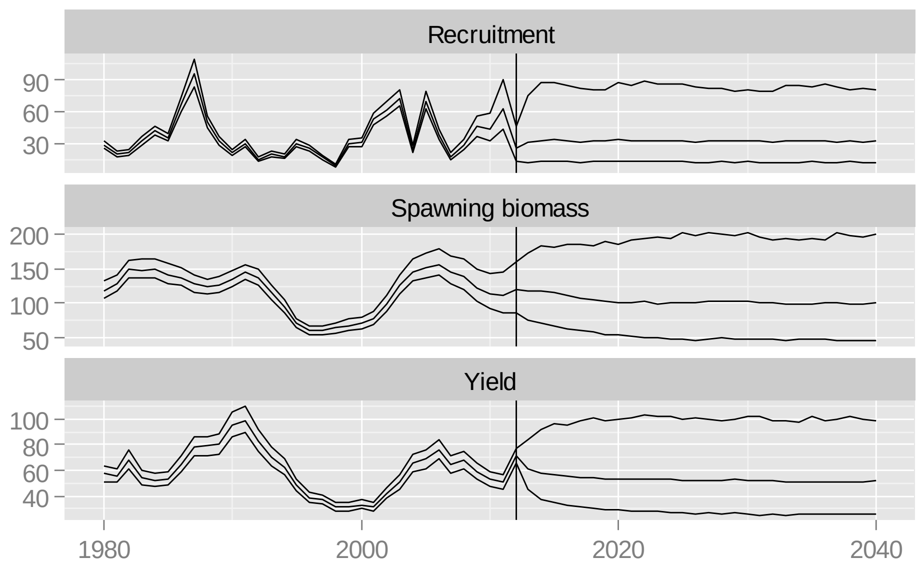


Figure 4.4.2. Example of historical assessment (assessment year is 2012) and future expectation of recruitment, SSB and yield when future fishing mortality is kept similar to the average of that observed historically. Shown are 5th, 50th and 95th percentiles

4.4.5 Generating other data for decisions (survey results, environmental impact, etc.)

In some cases the management decision model does not require an assessment (e.g. in cases where the HCR relies directly on input data (such as a survey biomass index), in which case input data should be generated in the same way as when input data are generated for an assessment in the loop (with the "reality" checks that go with that). 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).

4.5 Decision model

This component uses the assessment results to derive a decision on removals from the perceived status of the stock and fishery in a pre-determined process. On many occasions, a harvest control rule will be used (a recovery plan is regarded as a special case of a harvest control rule). These rules represent pre-agreed actions taken conditionally on quantitative comparisons between indicators of the status of the stock.

For example, current ICES harvest control rules generally fall into the following categories:

• F-regimes: direct effort regulation, TACs derived from F, TAC = fraction of measured biomass.

• Catch regimes: permanent quotas plus protection rule.

• Escapement regimes: leave enough for spawning but take the rest.

The output from the decision model could include recommendations for:

• TAC;

• Allowable effort;

• Closed areas;

• Mesh size regulations, although of limited use in hook and line fishery

It can be convenient 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 decision process has some typical elements, that are applied in sequence in a simulation program:

1. A basic rule, that prescribes a ***'primary' TAC*** (or other regulation) through the steps 1-3 below.
2. ***Stabilizing terms***, which modify the 'primary' TAC by constraining the change in TAC from year to year, perhaps with exceptions.
3. ***Other modifying terms***, for example maximum and/or minimum TAC.

The decision process in each of these steps can be structured as follows:

1. ***A decision basis***. This is the information that goes into the rule
2. ***A decision rule that*** sets a ***measure of exploitation*** as a function of the basis.
3. If needed, a ***translation*** mechanism, which translates the measure of exploitation into operational measures, for example a TAC.

Management rules are typically expressed as legal texts. For a scientific evaluation, it is essential that there are no ambiguities. The practical test is that the rule can be programmed. Ambiguities may not become apparent until at this stage, which hence may require some iterative procedures with the managers.

The basis typically is the SSB at some time according to the most recent assessment. There are however myriads of other potential measures (TSB, survey index, estimates of recruitment, mean length or age, biomass of other stocks, ...) used alone or in combination or applied under different conditions (e.g. exploitation rate applied being a function of recruitment).

The basis may come from an assessment (or a proxy for it) in the observation model, but may also represent a biomass measured in a survey or other measure. If so, a link between basis and true stock has to be included with realistic uncertainty. If the basis includes environmental influences that also may impact the true 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 rule itself is a parametric function of the basis, and can of course be formulated in many ways. The most common type is a steady exploitation if the stock is in a satisfactory shape with a reduction it if there are indications that the stock productivity is reduced. The parameter will typically be a standard value for F, and a breakpoint in SSB, and the rule is $F = min(StdF*SSB/breakpt,StdF)$. Other rules can have more parameters and other kinds of parameters, for example one indicating the slope of the decline in F below the breakpoint. These parameters and their values should be decided to give optimal performance of the rule, and are conceptually different from reference points. Although sometimes relevant, there is no need for a breakpoint to be identical to Bpa, for example.

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

The translation mechanism typically is to convert a fishing mortality to a TAC. That is normally done by projecting the vector of perceived stock numbers at age through the TAC year with the prescribed F or some other assumptions, and derive the catch according to that. Other exploitation measures will need other ways of translation, or no translation at all. For example, if the exploitation measure is a harvest rate, the TAC is obtained by simply multiplying the perceived stock biomass with a factor. If the exploitation measure is the TAC itself, no translation is needed.

Both the basis and the translation mechanism may need stock numbers at some time after the last assessment. If so, a projection step is needed. The form of the harvest rule may also necessitate some iterative procedures for example if the decision is based on SSB at a time when it is influenced by the decided removals.

Stabilizers are often included in proposed harvest rules. The purpose is primarily to avoid drastic changes in the TACs due to changes in the perceived stock status, perhaps due to assessment uncertainty. The two most common stabilizers are 'percentage rules' and 'filter rules':

The percentage rule is that the TAC shall not deviate more than a certain percentage from the previous TAC. Hence, the rule comes into effect only if the primary TAC deviates more that. Such rules often have an exception if the stock falls below a certain limit. Experience has shown that percentage stabilizers can lead to the paradox that if the TAC gets drastically reduced one year (perhaps because of a poor assessment) it takes a long time to get it up again. Likewise, if the stock productivity improves, for example because of some exceptional year classes, it takes long to increase the TAC, and when the productivity returns to normal, it takes a long time to get the TAC back to normal again. Hence, if there is periods with high and low productivity, the response can be too small and come too late. In such cases it is important that these side-effects of stabilizers are carefully examined and explicitly described to the stakeholders. It is also important to understand that this type of stabilizer tends to reduce year on year variability but may increase the overall span of TACs over many years.

Another stabilizer is a *'filter rule'*, e.g. where the final TAC is set as a weighted mean of the 'primary' TAC and the TAC the year before. or a mean of the 'primary' TAC and predicted future TACs. Formally, this is a simple low-pass filter. Rules where the TAC is a function of some estimates of the dicision basis over some number of years may also have a stabilizing effect. This type of stabilizer when operating on past values (not predictions) follows change and may result in large changes following significant changes in stock size, it tends to reduce the overall span of TACs.

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 e.g. rebuilding situations, where a drastic reduction of the TAC seems necessary, but it is hard to implement the whole reduction in one 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. Alternatively, taking mixed species fisheries into account could be part of the decision-making process.

As noted in Section 4.4, the conditioning of the decision model should mimic the annual decision-making process. If a projection is needed to convert an F to a TAC, the input to the projection should mimic the process that is normally done in a Working Group. For example, the constraint on catches (F or TAC constraint) in the intermediate year should be the same. In cases were assumptions about incoming recruitment are based on historical recruitments, this may necessitate assessment estimates of the recruitment back in time, which should be updated each year. In the short-cut approach, running a VPA backwards from the perceived terminal stock may be done to obtain estimates of historical recruitments. In some instances it is not possible to fully imitate the decision process, and simpler procedures may be considered. For example, if there is a deterministic component in the stock-recruit model in the operating model, that recruitment may be used in projections. However, doing so, the impact of such simplifications should be examined as far as possible, which in this example would be to examine the sensitivity of the actual removals in the implementation model to divergence between assumed recruitments and the real ones. If the incoming year classes contribute strongly to the subsequent catch, more realistic alternatives should be considered.

4.6 Implementation model

This is the step where the decided TAC is converted to real removals seen by the operating model. In practice, a TAC or other decisions have to be converted to removals in terms of numbers at age. The selections and weights needed in this calculation will deviate from those assumed in the decision process. Random elements may be introduced directly on these, or indirectly by adding random terms to the derived catch numbers.

To what extent 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. On one hand, one would not like to see a rule that breaks down once actual catches deviate slightly from the derived TACs. In some cases, quotas have been consistently exceeded in the past, and the tolerance of the rule to such over-fishing should be examined. On the other hand, one may argue that enforcement is a managers responsibility, science can show how the stock can be expected to develop if managers implement the rule that we investigate.

4.7 Stocks with sparse information

When the information about the stock is too sparse to permit the usual procedure of assessment and prediction, harvest rules may still be developed, but with a different form and with stronger limitations. Simulating such rules requires an operating model which may have to be more generic and less stock specific, and the rules will have to be more robust to uncertainties than when more precise information is available. Often, the rule can just set a TAC that appears to be safe, with a clause to alter it according to some indicators of trends in stock abundance or productivity.

Setting up a simulation for such stocks is not trivial, and outlining realistic options requires careful considerations. Quite often, life history parameters will be available, which, together with assumptions about selectivity and natural mortality will allow yield per recruit calculations. Recruitment is more problematic, but some indications of the likely level can be obtained by combining historic catches with yield per recruit. Then, setting up a simple operating model should be within reach. Using that, the sensitivity to variability in recruitment and growth may be explored for e.g. TAC rules. WKPOOR2 (ICES, 2009)) provided some examples. Indicators of altered stock abundance and/or productivity may be for example be survey data, CPUE data, area distribution of the fishery, information about depleted fishing grounds or perhaps even size distributions of the catches. Deriving such data from the operating model is not straight-forward, and the evaluations will often involve extensive sensitivity testing.

Stocks with sparse information is not a homogenous group, and at present, precise guidelines cannot be given. Below is one example of how the problem was approached.

4.7.1 Western Horse Mackerel

A management plan for the Western Horse Mackerel stock was proposed, refined and agreed by stakeholders in 2006-7 and was implemented in 2008. At the time, industry stakeholders were dissatisfied with a frequently changing quota and had little faith in the assessment process. The assessment model was under development and the results were considered to be exploratory by the working group.

Western Horse Mackerel suffers from a lack of fishery independent data – the only available index is an egg count from the triennial mackerel egg survey. Questionable catch data (in the past), a mismatch between the advice and management areas (up to 2009) and fisheries not covered by the TAC add to the uncertainty.

In the absence of a precise assessment and an independent estimate of SSB, the HCR is based on a hybrid rule which comprises a fixed TAC element (*TAC_{ref}*) and a variable element (*sl*) derived from the slope of the straight-line fit to the last 3 egg surveys (Roel and De Oliveira, 2007). The fixed TAC element was based on an equilibrium yield at F0.1?). The HCR sets the TAC for a period of 3 years using the equation

$$TAC_{y-y+2} = 1.07\left[\frac{TAC_{ref}}{2} + \frac{TAC_{y-3}sl}{2}\right]$$

The 2006 assessment was used to provide initialisation vectors for the MSE exercise (see 4.3.1) and the stock-recruit pairs from which the recruit relationship was derived. The population vectors from 2004 from the assessment were used since those from the terminal year and 2005 were more uncertain. A CV of 25% was applied at each age to the initial population numbers. A hockey-stock relationship was fitted to the SR pairs, disregarding the extremely large 1982 year class. The associated CV was derived from the residuals and the error was applied log-normally during simulation.

The observation model calculates an observed egg count from the operating model SSB incorporating process and observation error components. The observation error for SSB was considered to be 25% and this was applied prior to the risk calculation for the biomass limit (Probability SSBy<SSB1982).

4.8 Special considerations short lived species

4.8.1 Strategies with Biomass escapement criteria

For most short-lived stocks, the ICES MSY framework is aimed at achieving a target escapement MSY $B_{escapement}$, which is the amount of biomass left to spawn after the fishery has taken place), which is robust against low SSB and recruitment failure if recruitment is uncertain. The catch corresponds to the stock biomass in excess of the target escapement. No catch should be allowed unless this escapement can be achieved. For management purposes MSY $B_{escapement}$ is often (e.g. North Sea sandeel and Norway pout) set to Bpa to obtain a high probability of SSB > Blim. Other stocks (e.g. Barents Sea capelin and Bay of Biscay anchovy) use the predicted probability distributions for the SSB to estimate directly the risk of the SSB falling below Blim.

The "escapement strategy" allows each year a reduction of SSB to a minimum which makes future catch options highly dependent on the strength of the incoming year-classes. MSE of e.g. sandeel and Norway pout have shown that a more stable yield can be obtained by a lower F, but the loss in yield compared to the escapement strategy is high due to the low survival rate (high natural mortality) of the unfished population. This makes it difficult for the Industry to accept management plans that differs from the default ICES "escapement strategy".

The "escapement strategy" approach has been implemented explicitly into some management plans. The management of Barents Sea capelin targets a 95% probability of SSB > 200 000 t (Blim) after the fishery of mature capelin has taken place. The spawning stock (in April) and thereby the TAC is predicted from the acoustic survey in September, by a model estimating maturity, growth, and mortality (including predation by cod).

The application of the escapement strategy requires therefore an early indication of the recruitment that is going to be fished, because these recruits should sustain most of the escapement SSB (i.e., Bescapement). Therefore, when such information is not available this management strategy cannot be applied, because the uncertainty of the impact of the fishery on the population dynamics predictions would be too large. For the Bay of Biscay anchovy, since the reopening of the fishery in 2010 the TAC is set under the approach of constant harvest rate applied to the most recent estimates of SSB in May (by DEPM and acoustic surveys). Subsequently, the TAC is set for the period from July to June next year. The rule was derived under a MSE loop, proving to be robust to the unknown level of recruitment occurring during the management year. As the anchovy stock in most years consists of more than 80% one year old fish, the high uncertainty of next year's recruitment also makes the estimation of $B_{escapement}$. very uncertain. That is why the Bay of Biscay HCR followed the constant harvest strategy robust to the uncertainties in recruitment levels. Most of the efforts in recent years have been directed to provide a reliable indicator of recruitment from an acoustic survey to improve the scientific advice for this fishery.

Provided an indicator of recruitment is available, MSE of short lived species may be challenging as the performance of the HCR relies heavily on assumptions on growth, maturity, M and assumptions about the accuracy of the survey estimates which might be lower than anticipated. For example, in 2012 the two surveys for Bay of Biscay anchovy indicate very different estimates of abundance (DEPM is 80% lower than acoustics). Another problem is that the realization of the management objective (i.e. SSB > Blim) can be difficult to prove. Spawning individuals of capelin are dying shortly after spawning which makes it almost impossible to quantify SSB without dedicated surveys. In contrast. the iteroparous sandeel, survivors are found in the subsequent catches and surveys such that validation of the historical SSB is less difficult. Therefore, in all these cases the robustness of the HCR towards uncertainties and bias in surveys (either to estimate SSB or next coming recruitment), growth, maturity and M has to be tested.

4.8.2 Strategies for fisheries with higher probability of going below Blim without exploitation

As mentioned above, due to the low survival rate of species with short life span, the risks of such stocks falling below Blim, even without harvesting, can be high. Then, the definition of Blim is crucial in these cases, because if the reference point is not appropriately defined, falling below Blim may not be as critical as expected and can provoke unnecessary "alarms" and consequent loss of the credibility.

Regarding risk types that can be estimated, Risk 1 (see section 3) can be considered an adequate measure of risk level for these type of stocks as SSB in any one year is almost independent of the previous years. Risk 2 is a cumulative probability and therefore its value is higher, depending on the number of projection years simulated, to consider in the management decisions. Moreover, the usual levels of risk acceptable for other species with longer life span (around 5%) can be questioned in these cases (for instance in the absence of fishing the Bay of Biscay anchovy would have a 5% risk of falling at least once below Blim in 10 years – see table 4.8.1). Then, alternative approaches can be considered. For example, allowing a risk 'a' times higher than the natural usual risk estimated in the absence of catches, but still lower than a maximum threshold level of risk that should not be exceeded. Values such as acceptable mean levels of risks and maximum allowable level of risks are topics to be discussed between scientists, industry and managers. Risk 3 seems only of guidance to be taken into account in transition years, between regime shifts, when initial conditions are important.

In case of a poor recruitment regime, the management strategy would depend on the possible additional yearly information available. For example, if no early indication of incoming recruitment is available then the management should be based on the assessment of long term risks for different fixed harvest rates. Nevertheless, when information on new incoming recruitment is available, the short term risks can be evaluated and exploitation can be determined on yearly basis according to the expected levels of risks associated to a fishery with the forecasted recruitment level.

A clear example of a stock with high probability of going below Blim without exploitation is the Bay of Biscay anchovy. Simulations, under the assumption of an undetermined recruitment scenario and without any exploitation, estimate risk 2 = 5% and risk 1 = 1% (see table 4.8.1). Furthermore, for a persistent low recruitment scenario the risks increase sharply: risk 2 = 60% and risk 1 = 11% (see table 4.8.1).

Table 4.8.1. Risk 1 and Risk 2 derived from MSE simulations for Bay of Biscay anchovy under the assumption of absence of catches and different stock recruitment relationships: ricker, quadratic-hockey-stick (qhstk) and persistently low recruitment (low).

SRR	p(SSB<Blim)	p(SSB<Blim once)
ricker	0.01	0.05
qhstk	0.01	0.05
low	0.11	0.60

Experience has demonstrated that Blim defined for the Bay of Biscay anchovy is an appropriate limit threshold. After several consecutive years of low recruitment together with a decrease in fleet catches, the population fell to rather low levels. In 2005, the stock was estimated below Blim and the fishery was closed and it took 5 years to recover after the closure. The issue of appropriate risks for short lived species is discussed in section 3.

5 Dialogue and governance

Involving all the players (RAC's, managers, implementers and scientists) in the MSE process from the earliest stage is important to underpin the legitimacy and saliency of the result.

The WKOMSE workshop in Jan 2009 approached the process of designing and evaluating management plans. Much of what was discussed under that process is still relevant here. The workshop identified four categories of player in the process:

1. Policy makers: - Managers / (politicians)
2. Implementers (including POs): / control agency enforcers / legal experts
3. RACs / ACFA / Industry / NGOs
4. Experts: Biological / Social / Economic or other Scientists

In this context the phrase "designing a plan" is used to encompass all aspects prior to implementation, and "evaluation" as the examination of the performance of the plan after a number of years. The Roles and Responsibilities of the player groups were examined and the following roles identified

Group 1) Policy makers Managers / (politicians) operating at Local, National and European levels, such as the EU Commission other Nation states, and Fisheries Commissions such as NEAFC , Their responsibilities were identified as:

Setting overall Objectives (mostly politicians)

Plan proposal, initiation,

setting criteria design and evaluation phases,

Setting the rules

Consulting and seek expert advice

Translation to legal framework

Fleetwise allocation

Group 2) Implementers (POs) / enforcers control agency /legal experts operating at Local and National levels, with responsibility for:

Technical and advisory consultation

Translation to legal framework

Fleetwise allocation

Practical implementation of rules (data / licences)

Group 3) RACs / ACFA / Industry / NGOs, and possibly some media, operating at Local, National and European Level Their roles would be

Initiators,

Consultation

Advice (from consultation)

Influence

Communication

Group 4) Experts Biological / Social / Economic Scientists, operating at Local, National and European Level, with roles of

Initiation

Consultation, and Advice (ref points, targets, plan performance etc.)

Communication

The group examined how these players should be involved in the process of developing both recovery and multi-annual management plans, the following structure (Figure 5.1) illustrates the process consisting of an initiation and scoping phase followed by an iterative development loops which is expected to be completed at least twice before proving the plan in a form that would be in a suitable form for implementation. Following implementation there is a potential for use of a similar loop a final time following a number of years of implementation to evaluate the performance of the plan, presumably leading to either continuation or revisiting the design phase.

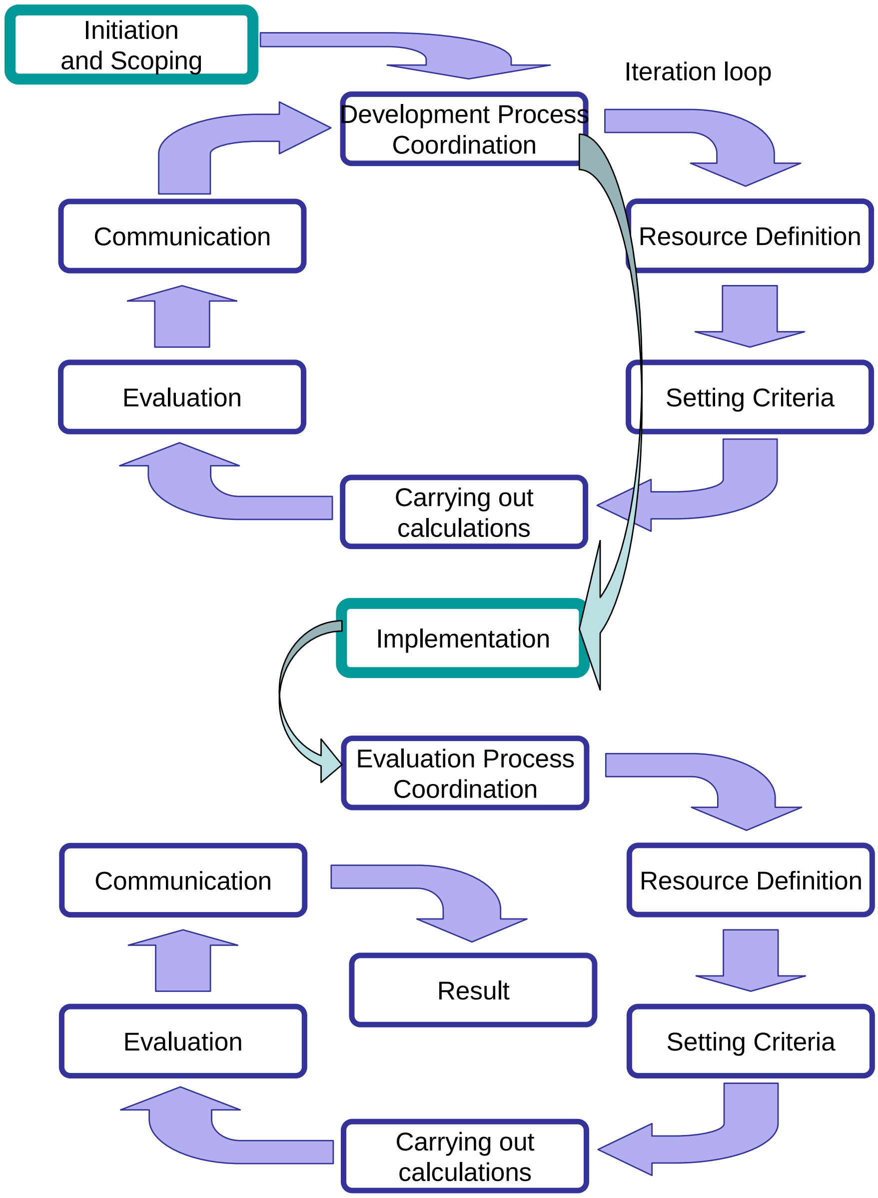


Figure 5.1 Flow chart of development process after WKOMSE

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

- Initiation (Mainly Decision Makers, but also RACs+others)
  - Attempt at discussion amongst all coastal states
  - Scope the problem (Decision makers, Experts, RACs, 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 RACs (facilitator experts))
  - Carry out calculations (Experts, (Implementers) (All))
    - Needs to be transparent – but also needs to be quality checked
    - May not be possible interactively
  - Carry out evaluations (all)
  - Communicate discuss
    - All
- Iterate around the loop as required.
- Implementation
- Evaluation using a similar loop (Figure 5.1)
- Results and next steps

Roles and responsibilities

The setting of objectives and defining the types of tactical approach 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 a MSE. It is a role of the scientists to provide the technical documentation which provides the evidence base for the decisions adopted in the management plan. The minimum specification for this technical documentation is given in this report section 6

6 The reports requirements for studies done for ICES

6.1 Minimum standards for simulations

The overarching criterium is that the rule performs satisfactorily under a plausible range of scenarios, both with respect to biological variation and uncertainties in the decision process. This range should be documented, and as the rule is being practiced, it should be possible to control how the stock and the decisions develop compared to the range that was assumed. Examples include the distribution and time course of recruitment, growth, maturity and selection, as well as adherence to the rule. If the rule fails because the stock or the management behaves outside the assumed range, the rule should to be revisited, and may have to be revised. A good rule should be able to cope with unforeseen events, but it cannot be expected to be optimal under conditions that differ from those for which the rule was designed.

This section is a brief list of checkpoints that reviewers may consider, to ensure that the simulations cover a realistic range of future developments:

1) The operating model should be according to established standards for population dynamic models, and be sufficiently detailed to provide the information needed in the decision process.

2) All processes where natural variation is likely to occur should be modelled as stochastic processes. If such processes have been assumed to be stationary in the assessment which underlies the conditioning of the model, the sensitivity of the assessment to such variation may have to be examined. Example: variable natural mortality.

3) Autocorrelations and time trends should be considered, and included if they appear and can impact the results. Iid. log-normal noise is rarely encountered in biological processes, and will often be unduly naïve.

4) When deciding on parameters for distributions, the guideline should be to obtain a plausible range of realities. Just picking a variance-covariance matrix form some source without considering what it represents is not good practice. One example may be using assessment uncertainties at age to provide initial numbers. The assessment uncertainty represents how well parameters can be estimated with the data and model, while the initial numbers should be what is left of year classes representing a plausible range of year class strengths.

5) Observation model: If a full assessment is done in the loop, the assessment model should be comparable with the one used in routine assessments, and the input data should be sufficiently noisy to provide assessments as problematic as experienced. If the short-cut approach is used, the variances should lead to a range of stock 'estimates' comparable with the statistical properties of the routine assessments, including retrospective errors and autocorrelations over time and age. If historical assessment data are used in the decision process, they will have to be renewed each year, and they should be internally consistent.

6) Does the rule allow sufficient action if the biology or management falls outside the assumed range?

7) It is not always obvious how reality checks can be done, but to the extent possible conformance with historical experience should be demonstrated.

8) Leaving out sources of variability, or deviating from the routine practice in the decision process may be permissible, but if such simplifications may be questionable, sensitivity tests should be done. One example is the use of constant selection in the fishery, if it is suspected that it may vary over time.

9) Ensure that measures that are compared with reference points are derived the same way as the original reference points. For example, SSB values that are compared with limit biomass should be derived using comparable assumptions about weights, maturities and natural mortality. If necessary, reference points may have to be revisited, and proxies for reference points used in the evaluations.

6.2 Reporting requirements

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

A summary template filled out to illustrate the main aspects dealt with in modelling (Section 6.3 below)

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.

Reality checks are very important to increase confidence in the suitability and plausibility of the assumptions made in the MSE. These reality checks would include graphs showing:

- Comparison between historic and simulated Recruitment against SSB
- Comparison of historic and simulated Recruitment, to illustrate distributional form (e.g. via Q-Q plots), autocorrelation and fluctuating and episodic recruitment.
- Comparison of simulated and historic error in the assessment
- Comparison between simulated and historic error in any indices used in the simulations.

It is preferable that graphs present percentiles of future trajectories (5%, 50%, 95%), which are much easier to interpret than box-plots.

Both tables and graphs should be displayed giving Prob 1, 2 (over10 years) and 3, so the performance is documented.

The distribution of a number of observed parameter values (F, SSB...) that are expected to be observed under exploitation following the plan should be provided to allow users to evaluate consistency of implementation with the study. In particular, the range of selection-at-age patterns considered in the MSE should be presented, so that the assessment WG can monitor whether the selection-at-age pattern in the fishery stays inside or goes outside the range tested

6.3 Summary Template for HCR modelling

This template is a summary of the HCR evaluation, primarily intended as an overview and check-list for reviewers of the work. All fields should be filled in, even those considered irrelevant – with explanation. Further details shall be given in the full report.

The non-coloured fields are made as an example, mostly to illustrate the expected level of details.

This template is a summary of important issues in the evaluation

The *'s refer to explanations below

The non-coloured fields are made as an example, mostly to illustrate the level of detail we would like to see. The standards described are not necessarily ideal.

Stock: Fantasy fish stock

Background		
Motive/ initiaitve/ background.	The industry was not satisfied with current unpredictable quotas, and developed a proposed management plan. Managers requested ICES to develop the proposal further and advice on a plan.	
Main objectives	Precautionary, stable catches near MSY, multi-annual TACs if possible	
Formal framework	ICES on request from EU/Norway	
Who did the evaluation work	WKFANT 2011	
Method		
Software Name, brief outline include ref. or documentation	Ad hoc software, written in R, assessment model in AD model builder. Age structured operating model, full assessment (state space model) with catches at age and two surveys derived from the true population. Unpublished, undocumented, code available on request.	
Type of stock	Medium life span, demersal, very valuable	
Knowledge base *	Analytic assessment, barely acceptable	
Type of regulation	TAC	
Operating model conditioning		
	Function, source of data	Stochastic? - how (distribution, source of variability)
Recruitment	Beverton-Holt fitted to SR pairs 1980-2010	Log-normal, CV from residuals
Growth & maturity	Average over 2008-2010, no density dependence	No
Natural mortality	Lorenzen formula: $M(a) = 3*W(a)^{-0.29}$	No
Selectivity	Average F at age over years 2008-2010 in 2011 assessment, scaled to mean 4-8.	No
Initial stock numbers	From assessment	According to variance - covariance matrix from assessment (inverse Hessian)

Decision basis **	SSB in the TAC year	
Number of iterations	1000	
Projection time	30 years	
Observation and implementation models		
If assessment in the loop		
Input data	Catches + 2 surveys	Catches and surveys: Log normal, CV from assessment residuals
***** Comparison with ordinary assessment?**	No	
Deviations from WG practice?	Yes, WG uses 5 surveys, model uses 2	
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?	TAC constraint in projections, WG uses Fsq.	
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	F-rule with two breakpoints on SSB: B1 and B2: If SSB < B1, F = Fstd*SSB/B1 If B1<BBB<B2: F=Fstd If SSB > B3: F = Fstd+gain*((SSB-B2)/B2	
Stabilizers	TAC shall not deviate more than 15% from TAC the year before, unless the constrained TAC leads to SSB < B1	
Duration of decisions	Annual	
Revision clause	After 5 years or if SSB < Blim	
Presentation of results		
Interest parameters	Risk, Catch (Mean and 10-50-90 percentiles), Inter-annual variation, fraction of catch > 5 years old	
****** Risk type and time interval**	Type 2, for years 11-20.	
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	Recruitment has declined recently for unknown reasons, the SR function predicts better recruitments than in the recent past. Multi-annual TACs were abandoned. Required much lower catches to get an acceptable risk. The final rule was similar to the one proposed by the industry, but with a standard F at the low end of their proposal. The industry was not satisfied, as the SSB appeared to be below B1 in 2012, and could not be increased in 2013 because the assessment was revised to give a higher SSB. They are already asking for a revision of the plan.

* **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.

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 sec 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 time savings. When selecting from the available software packages consideration should be given to

- The underlying capabilities of the software in terms of the operating, observation, decision and implementation models (see the accompanying documentation)
- Is the software readily modifiable for your needs?
- The 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 7.2 Software Development and Quality Standards

In the event that a new application is to be built, it should be recognized that software development, when done properly is an involved and often tedious process. There are however, well established guidelines which can result in a robust and useful application. The process can be broken down into the following phases

1) Design.

This phase is the most critical. In general, seek to reduce the overall requirement into functional units that can be coded and tested individually. The use of pseudo code and flowcharts can be helpful and can be recorded in a functional specification, a document that describes the application's capabilities and overall structure. The design phase should also establish the inputs and outputs for the various functional units in terms of both type and value.

2) Build.

During the coding (build) phase, the design is translated into the appropriate language. Regardless of the language employed, use a sensible descriptive naming convention for variables, functions and classes and reduce complexity wherever possible. Employ lots of whitespace and comment the code liberally. This will aid reuse of the code. Defensive coding is an appropriate method to employ. This implies an attempt to identify any exceptions that may occur during execution (*e.g.* divide by zero) and either test for them prior to execution or trap and handle them. Should the application or function be forced to terminate, it should do so cleanly. In addition, pay attention to possible performance issues and attempt to eliminate any unnecessary or inefficient processes.

At this stage it may be appropriate to consider the use of a source code repository. There are several of these available online (*e.g.* Google Code, GitHub) and they can be invaluable in tracking changes and releases in projects.

3) Test & Debug.

There are a number of methodologies for the testing of computer code. Unit testing is a widely used technique appropriate for testing software that can be subdivided into functional units. Ideally, unit test plans document a series of tests and should be constructed during the design phase of development. Packages such as RUnit can be helpful for executing unit tests on software written and packaged in R. System testing involves running a number of predefined tests once unit testing of all components has been completed. Should any code be changed, the appropriate unit test and the system tests should be re-run. When testing functions that employ the generation of random numbers (such as in stochastic simulations), the seed for random number generation should be reset in order to verify that results are repeatable.

Most computer languages have a number of tools to help with debugging (*e.g.* R functions trace back, debug, browser). A bug register for tracking the status of reported problems can be incorporated into the source code repository.

4) Documentation.

This should accompany all software. Much of it can be taken from the functional and technical specification documents and will be useful for both the original author of the software and future users. In addition to describing the functional interfaces of the application, the documentation should cover the steps required to install the software (including a list of pre-requisites). Examples of the software's capabilities are always useful for new users.

5) Versioning and Release.

Software development is largely an iterative process. The labelling of the code during stages of development with version numbers is useful when it comes to adding features, fixing bugs and communicating with other users. When an iteration of the development is complete, all files should be given the same versioned and labelled. Code repositories offer good functionality in this regard. Repositories can also be used to maintain a register of users who can be notified of new releases. Software can also be made available via institute websites and appropriate ICES SharePoint sites.

7.3 Available Software

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 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 formualtion, to VPA-based methods, such as Separable VPA () and XSA (), and statistical catch and age methods, like FLa4a (https://github.com/ejardim/a4a/tree/master/packages) and FLSAM (http://code.google.com/p/hawg/). Tools exist for interfacing with existing stock assessment models coded in either C, C++, Fortran or ADMB.

The projection capabilities of FLR, implemented in the FLash package using Automatic Differentation, can be used to implement a large variety of harvest control rules in a efficient way. Those that cannot be currently adapted to the syntax offered in FLash, R offers a large range of programming constructs that can be also applied.

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 very little cost.

Recent examples of use in MSE

Jardim, E., Mosqueira, I., Millar, C., Osio, C. and Charef, A. 2012. MSE testing of factors likely to have an effect on catch surplus calculations through impacting MSY estimates. JRC Scientific and Technical Reports, JRC 72625, Report EUR 25389 EN.

Jardim, E., Mosqueira, I., Millar, C. and Osio, C. 2011. Testing the robustness of HCRs applied to Baltic pelagic stocks. JRC Technical Note, JRC 69877. EUR 25275 EN.

Current status

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. The next release of a much improved and extended set of packages is planned for April 2013.

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 Journal of Marine Science, 64(4), 640–646. doi:10.1093/icesjms/fsm012

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 type 3, previously it was type 2. 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 (Dorleta et al. in prep). The model is not as complicated as ecosystem models but neither so simple as the bio-economic models available in the actuality; it finds a balance between the biological and economical component.

FLBEIA is built using R- FLR functions and under a Management Strategy Evaluation (MSE) framework. It is composed ofan Operating Model and a Management Procedure. The management advice can be given based on real population or on the observed population through the whole management process. The model is multistock, multifleet seasonal and it allows the insertion of uncertainty. It has an extra component called covariables, which gives the possibility to introduce other variables of interest that are not taken into account in the biological or fleets components (e.g. ecosystem components).

The model is constructed in a modular way. Each process has different models available, and allows the possibility to include new ones.

Model documentation is in preparation, but the model is available in the FLR repository (http://r-forge.r-project.org/R/?group_id=318).

7.3.5 Impact Assessment Model for fisheries (IAM)

The program IAM has been recently 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 details in Merzéréaud et al, 2011[1]. It can be used to carry out impact assessment for management plans 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 5.2

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

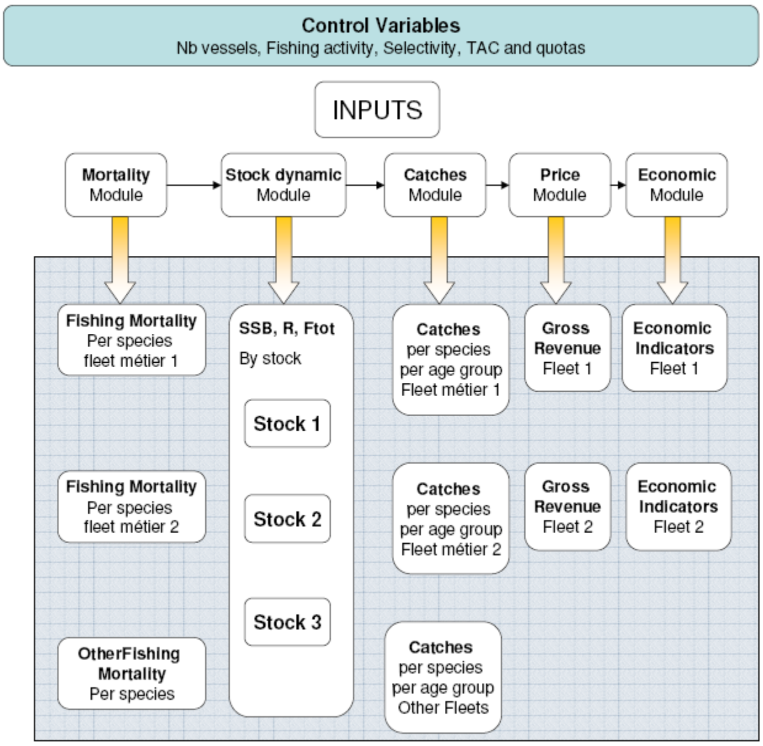


Figure 5.2Simplified representation of the Impact Assessment bio-economic model for fisheries

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.

Annex 1 ICES Management Plan Evaluations 2008 -2012

A template was circulated to participants in advance of the meeting the entries below represent the responses received, and while broadly representative are not exhaustive. They were requested in order to get an overview of how harvest rules have been examined in recent years.

The results from these evaluations are summarised in Section 2. The templates below have been replaced by a revised version given in Section 6. The sequence of stocks describe below has no significance.

The evaluation of risk documented in these templates is discussed in section 3 of the report

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.

Stocks: Celtic Sea Herring

Background	
Motive/ initiaitve/ background.	The ICES advice for 2007 was that there should be no fishing without a rebuilding plan being in place. The European Commission (EC) proposed reductions in the TAC. This stimulated the local Irish stakeholder committee (Celtic Sea Herring Management Advisory Committee, CSHMAC) to develop a management plan. The initial proposal (2007) was not sufficiently strong and was not considered by the EC, with the proposed TAC reduction being implemented. In 2008, the CSHMAC brought forward a stronger proposed plan, in conjunction with the Irish Marine Institute, incorporating a 25% TAC decrease (mirroring the EC proposal of that year). The proposed plan also contained a HCR with a maximum F set at $F_{0.1}$. This plan was evaluated by ICES in 2009, and judged to be precautionary within the stock dynamics considered. The plan was used for TAC setting in the years 2009-2012. The plan was, by its own definition, complete in 2012, and a long term plan was required. A proposed long term plan was developed in 2011 by the CSHMAC, in conjunction with the Irish Marine Institute. This plan was followed for TAC setting for 2013, having been evaluated by ICES and judged to be precautionary. Neither the rebuilding plan nor the long term plan has been enshrined in law, owing to the difficulties in co-decision between the European Parliament and Council, emanating from the Lisbon Treaty.
Main objectives	Evaluate whether the plans were in accordance with the precautionary approach (RP and LTMP) and MSY approach (LTMP).
Formal framework	ICES on request from the EC (rebuilding plan 2008) and Ireland (LTMP 2011).
Who did the evaluation work	Ad Hoc Group on by correspondence in 2009 (RP); RG/ADGCSH, in 2012, (LTMP). Evaluation work and simulations done by staff of Irish Marine Institute.

<table>
<tr><th colspan="4">Method</th></tr>
<tr><td>Software
Name, brief outline
include ref.
or documen-tation</td><td colspan="3">HCS software (Skagen, 2009,2011). The rule was stress tested against various errors and biases. The magnitude of these errors and biases was taken from observed ranges in the data series. However the rule was further stress tested against more extreme errors and biases to find the range beyond which it would have an unacceptable risk (to Blim).</td></tr>
<tr><td>Type of stock</td><td colspan="3">Moderate longevity (6+ years), small scale fishery regionally important</td></tr>
<tr><td>Knowledge base</td><td colspan="3">Population vector mimicking forecasted population, stock recruit relationship (SR), with some truncation.</td></tr>
<tr><td>Type of regulation</td><td colspan="3">TAC only, closed area, with small sentinel fishery within</td></tr>
<tr><th colspan="4">Model conditioning</th></tr>
<tr><td></td><td>Function, source of data</td><td colspan="2">Stochastic? - how</td></tr>
<tr><td>Recruitment</td><td>Segmented regression (period 1958-last assessment year -2)</td><td colspan="2">Lognormal; CV from residuals</td></tr>
<tr><td>Growth & maturity</td><td>Growth: mean of the last five years; no density dependence
Maturity: constant annual ogives as used in the assessments of the stocks</td><td colspan="2">No</td></tr>
<tr><td>Natural mortality</td><td>Constant over years, as per assessment, decreasing from 1st to 3rd age group then constant.</td><td colspan="2">No</td></tr>
<tr><td>Selectivity</td><td>Selectivity: Mean of last 5 years</td><td colspan="2">No</td></tr>
<tr><td>Initial stock numbers</td><td>Mimic of short term forecasted population vector.</td><td colspan="2">No.</td></tr>
<tr><td>Decision basis</td><td colspan="3">(if this means for the HCR) TAC based on F in the TAC year.</td></tr>
<tr><td colspan="4">If assessment in the loop</td></tr>
<tr><td></td><td>Input data</td><td>Sole: landings + 2 surveys
Plaice: landings + discards + 3 surveys</td><td>Log normal, CV from assessment residuals</td></tr>
<tr><td></td><td>Comparison with ordinary assessment?</td><td colspan="2">Identical</td></tr>
<tr><td>If no assessment in the loop</td><td colspan="3">Below is just an example of how this could be presented if there was no assessment in the loop</td></tr>
<tr><td></td><td>Type of noise
(distribution,
on what)</td><td colspan="2">A range of errors and biases assumed on implementation and observation (assessment), taken from observed values in</td></tr>
</table>

		assessments and management performance. Also stress testing of a range beyond these.
	Comparison with ordinary assessment?	n.a.
Projection		
	Through intermediate year?	Yes, deterministic with recruitment according to deterministic SR function
	Iteration in TAC year?	Yes
Implementation	Initial projected forward, drawing from SR relationship, with some truncation. TAC set based on F in the TAC year.	
Harvest rule		
Harvest rule design	Rebuilding Plan: If SSB>Blm : $F_{0.1}$. If SSB < Blim : reduce TAC by 25% LTMP: If SSB > Blim : F = 0.23; If SSB < Blim : F = SSB *0.23/Btrigger	
Stabilizers	TAC shall not deviate more than 25% from TAC the year before in the LTMP.	
Duration	Annual	
Revision clause	RB: once SSB at or above Bpa in three consecutive years, plan is superseded by LTMP	
Presentation of results		
Interest parameters	Risk, SSB, F, Landings TAC, annual TAC change	
Risk type (see classification below) and time interval	Type 2. P(SSB<Blim) in any year of the 20 forward-simulated years	
Precautionary risk level	5% Risk type 2	
Comments and experiences		
Review, acceptance:	Accepted by review group, ICES. Implementation pending EC co-decision, but used in TAC setting once each plan was evaluated by ICES.	
Experiences and comments	The stakeholder interactions with scientists were a key feature, however there were false starts, with industry initially finding it difficult to incorporate TAC cuts in the plan. MI scientists did not endorse the initial plan of 2007, though they did give technical support to the drafting procedure. However the 2007 draft was the start of what turned out to be a successfully completed process. The simulations considered risk to Blim (26 000 t) and also the changepoint in the S/R relationship. The former SSB value was considered more appropriate as a reference for risk evaluation by the external reviewers within RG/ADGCSHER, and has taken precedence here. It is noted that HAWG in the past has proposed increasing Blim to the S/R changepoint with a consequent	

	increase in Bpa. This may not be necessary, based on ICES SGPA 2003 guidelines. It should also be noted that the changepoint in the segmented regression has been quite volatile over time (39 000 – 45 000 t) in various studies conducted in recent years. This volatility is considered a good reason not to use changepoint as a basis for Blim. This may be the first evaluation done by ICES to incorporate both implementation error and bias. **Conclusion of the MSE:** The proposed plan performs well, with low risk to low biomass within the most likely conditions pertaining in the fishery system. This was evaluated over a range of errors and biases. However the plan fails to deliver low risk if error and bias is more extreme than modelled. The issue of discards is worth considering in this regard. Assuming that maximum discarding is less than 10% in addition to reported catch, the plan still behaves within the PA approach. But if discarding is beyond that range, the plan begins to fail. Latest information from independent discard monitoring by the Irish Whale and Dolphin Group shows suggests that our assumptions are valid. This information also suggests that the maximum discard rates considered by ICES HAWG 2012 are probably overestimates.

Stocks: North Sea plaice and sole

Risk definitions

The results presented in the current report are based on risk definition number 2. This definition was chosen as the most appropriate one, amongst others because it was also used in previous evaluations of the management plan in 2007 and 2010. Furthermore, it became clear from the discussion that definition number 2 presents the most conservative results, i.e. in comparison to the other two definitions, it will provide a higher risk outcome. For the purpose of further discussion in ICES on choosing risk definitions, Table 5.1 below shows the specific outcomes in terms of risk calculated based on the three different definitions.

Table E.1. Overview of risk percentages when based on different definitions.

	North Sea plaice		**North Sea sole**	
	Current plan	Proposal	Current plan	Proposal
P(SSB<Blim); 2013-2015				
Risk 1	0	0	0	0
Risk 2	0	0	0	0
Risk 3	0	0	0	0
P(SSB<Blim); 2015-2020				
Risk 1	0	0	0.001	0.001
Risk 2	0	0	0.005	0.005
Risk 3	0	0	0.005	0.005
P(SSB<Blim); 2016-2025				
Risk 1	0	0	0.002	0.002
Risk 2	0	0	0.015	0.015
Risk 3	0	0	0.005	0.010

Background	
Motive/ initiaitve/ background.	In 2007, the European Commission (EC) adopted Council Regulation No 676/2007, establishing a multiannual plan for fisheries exploiting stocks of plaice and sole in the North Sea. The objective of this plan, in its first stage, is that stocks of plaice and sole in the North Sea are brought within safe biological limits. WGNSSK 2012 concluded that these objectives were met. Following this, the plan should ensure in its second stage that the stocks are exploited on the basis of maximum sustainable yield and under sustainable economic, environmental and social conditions. In April 2012, IMARES, through ICES, received a special request from the Netherlands to assess whether two proposed changes to the plan lead to exploitation of the stocks consistent with the precautionary and MSY approach in conformity with ICES criteria.
Main objectives	Evaluate whether the proposed amendments (a change in the target fishing mortality for sole from 0.20 to 0.25 and ceasing reductions of the Maximum Allowable Effort) are in accordance with the precautionary principle and MSY approach. i.e. evaluate both **targets** and **effort** control; for **two** stocks caught in a

	mixed fishery	
Formal framework	ICES on request from the Netherlands	
Who did the evaluation work	Ad Hoc Group on Flatfish (AGFLAT); reviewed by RGFLAT/ADGFLAT 2012. (IMARES prepared most of the work – A.Coers, D.Miller and J.J.Poos)	
Method		
Software **Name, brief outline** **include ref.** **or documen-tation**	Ad hoc software, written in R. Full feedback stochastic projection model in which in addition to the biology (full assessment), the fisheries system is modelled with simple fleet dynamic rules for three different fleets targeting the two species. Three surveys sample the plaice stock, and two surveys sample the sole stock. Survey indices used as input to the assessments in future years were generated from the "real" populations on the basis of model estimated catchability at age (from the most recent ICES assessments) with error coefficients to simulate observation error. Unpublished, code available on request.	
Type of stock	Two stocks; Medium life span, demersal, valuable (sole) and less valuable (plaice)	
Knowledge base	Analytic assessments	
Type of regulation	TAC and effort restrictions	
Model conditioning		
	Function, source of data	Stochastic? - how
Recruitment	Combination of segmented regression and Ricker according to likelihood of each function fit to the data (period 1957-2009)	Lognormal; CV from residuals
Growth & maturity	Growth: mean of the last five years; no density dependence Maturity: constant annual ogives as used in the assessments of the stocks	No
Natural mortality	Guestimate: 0.1 (all ages and years for both stocks)	No
Selectivity	Selectivity: mean last 5yrs. Fleet model also requires catchability: sampled from 1995-2011, with scenarios of technological creep. For plaice: separate for discards and landings	No (catchabilities resampled)
Initial stock numbers	Three distinct scenarios: (1) XSA; (2) Pessimistic XSA: using index values with	No, scenario range considered more important test of robustness.

	recent 6 cohorts (2006-2011) arbitrarily decreased by applying a multiplication factor of 0.75 (variance of 0.1); (3) Optimistic XSA: same with multiplication factor of 1.25		
Decision basis	(if this means for the HCR) TAC follows from F in the intermediate year, in relation to target F (no SSB triggers); effort level depends on SSB in the TAC year.		
If assessment in the loop			
	Input data	Sole: landings + 2 surveys Plaice: landings + discards + 3 surveys	Log normal, CV from assessment residuals
	Comparison with ordinary assessment?	Identical	
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 (distribution, on what)	n.a.	
	Comparison with ordinary assessment?	n.a.	
Projection			
	Through intermediate year?	Yes, deterministic with recruitment according to deterministic SR function	
	Iteration in TAC year?	No	
Implementation	Catches in numbers at age from projection according to the rule, potentially limited by effort constraints (depending on catchability and allowable effort in any given year). No implementation error (i.e. on real pop.) but observation error on catch levels for eth assessment (i.e. perceived stock).		
Harvest rule			
Harvest rule design	**No breakpoints on SSB to determine TAC based on Fstd** **Breakpoint on SSB (both stocks) to determine Maximum Allowable Effort (MAE) for the main fleet (Dutch BT2):** **If SSB < B_{LIM} (either stock), $MAE_{y+1} = MAE_y*0.9$** **If SSB > B_{LIM} (both stocks), $MAE_{y+1} = MAE_{2012}$**		
Stabilizers	TAC shall not deviate more than 15% from TAC the year before, unless the constrained TAC leads to SSB < B_{LIM}		
Duration	Annual		
Revision clause	Once reached SBL (this evaluation was such a revision)		
Presentation of results			

Interest parameters	Risk, SSB, F, Landings (mean and 5-50-95 percentiles), Discards (plaice), TAC, annual TAC change, MAE (fleet), deployed effort, % of times that MAE constrains landings (rather than TAC)
Risk type (see classification below) and time interval	Type 2. P(SSB<Blim) for three periods: 2013-2015; 2015-2020; 2016-2025
Precautionary risk level	5% Risk type 2
Comments and experiences	
Review, acceptance:	Accepted by review group, ICES. Implementation pending EC decision.
Experiences and comments	The proposed management plan performed successfully under various scenarios of stock productivity, effort deployment and fleet dynamics. Passing these sensitivity tests shows that the proposal is robust to some of the major assumptions made. Given the very healthy state of one of the stocks (plaice), risk was mainly a concern for the other stock (sole) It would have been difficult to hit the risk levels for plaice over the time periods examined even at very high F (is a longer evaluation period required for healthy stocks?) Given the very specific nature of the request (and that this was the third evaluation in 5 years) sensitivity to reference points was not examined (other than the sole target). It is thought that using alternative PA/MSY ref pts. would likely have a big effect on risk levels, particularly for sole which has been fluctuating near the current PA biomass reference points in recent years. **Conclusion of the MSE:** The proposed amendments to the current management plan are in accordance with the precautionary approach and consistent with the principles of MSY. Performance compared to the current management plan is very similar with regards to plaice, while likely an improvement for sole (in terms of yield) and thus the proposed plan probably results in exploitation nearer MSY. The results presented in the ICES advice are based on risk definition number 2. However, risk was calculated based on all three types separately. Attached to this form is some further explanation and information about this. (**see below**).

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.

Stock: Barents Sea capelin

Background		
Motive/ initiative/ background.	HCR proposed by scientists and adopted by managers (JNRFC) in 2002. No formal evaluation has been carried out.	
Main objectives	Yield, enough capelin as food for cod	
Formal framework	ICES on request from JNRFC	
Who did the evaluation work	AFWG. Benchmarked by WKSHORT in 2009, evaluation of HCR not done then	
Method		
Software Name, brief outline include ref. or documentation	Assessment: Bifrost/CapTool, Mathematica multispecies cod-capelin-herring model (Bifrost) used for parameter estimation and as operating model, and quota calculation done in Excel with @RISK (CapTool – capelin model which includes predation by cod). Bifrost to be used for evaluation – is being rewritten in R/ADMB	
Type of stock	Short life span, semelparous, pelagic, moderately valuable	
Knowledge base	Acoustic absolute stock estimate in September, stomach samples of cod (main predator) during January-March, spawning 1 April	
Type of regulation	TAC	
Model conditioning (Bifrost (capelin only) and CapTool)		
	Function, source of data	Stochastic? - how
Recruitment	Bifrost: Beverton/Holt with multispecies terms, CapTool: N/A	Bifrost: Replicates of residuals from log-normal distribution, CapTool: N/A
Growth & maturity	Bifrost: Weight at age drawn from time series. Both: Maturation length estimated from surveys	CapTool: No growth – maturation parameters from replicates of Bifrost estimates
Natural mortality	Bifrost: M on immature capelin estimated from survey time series, M on mature capelin in January-March estimated from stomach content data. CapTool: M in October-December equal to M on imma-	CapTool: Yes, taking into account uncertainty in cod assessment and in stomach samples in modeling of predation by cod. Replicates of predation parameters from Bifrost; log-normal distribution of error in cod abundance

	ture capelin estimated in Bifrost from time series, M in January-March dependent on cod abundance		
Selectivity	Fishing only on mature fish (both Bifrost and CapTool)	No	
Initial stock numbers	From acoustic estimate (both Bifrost and CapTool, but of little relevance for Bifrost)	Uncertainty level based on analysis from survey (Tjelmeland 2002), annual updates of this not available at time of assessment	
Decision basis	TAC set to ensure 95% probability of SSB > 200 000 t.		
If assessment in the loop			
	Input data		
	Comparison with ordinary assessment?		
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 (distribution, on what)	Log normal distribution on survey numbers at age	
	Comparison with ordinary assessment?	N/A	
Projection			
	Through intermediate year?	N/A	
	Iteration in TAC year?	N/A	
Implementation			
Harvest rule			
Harvest rule design	TAC set to ensure 95% probability of SSB > 200 000 t. Fishing only on mature stock prior to spawning.		
Stabilizers	None		
Duration	Annual		
Revision clause	No clause, but revision asked for by 2015 by JNRFC		
Presentation of results			
Interest parameters	Risk, Catch (Mean and 5-50-95 percentiles)		
Risk type (see classification below) and time interval	Risk 1 and 100 years for Bifrost, Risk 1 and 6 months for CapTool		
Precautionary risk level	Should be set to a level close to the probability of SSB<Blim if no fishing, to ensure that the HCR does not increase this probabil-		

	ity much.
Comments and experiences	
Review, acceptance:	None
Experiences and comments	Number of larvae/0-group has been sufficient in all years after the HCR was introduced, but recruitment failure in some years probably due to predation by young herring on capelin larvae. Plans for improving biological model by modeling predation by cod also in October-November. Harvest control rules have to be evaluated for a range of cod and herring HCRs, due to the strong biological interactions. Capelin fishery operates under a two-price system – a quantity of ~100 000 t can be sold for human consumption while the rest will go to meal and oil, thus it is desirable to reduce the number of years with no fishing. Can this be done without increasing the risk, if one reduces the catches in good years somewhat? Also important to avoid more than 3-4 years (one generation) in a row with low recruitment. Should Blim be made a function of predicted herring abundance?

Stock: Bay of Biscay anchovy

Background		
Motive/ initiaitve/ background.	The depletion of the population since 2005 has made explicit the need of managing this population on well-founded scientific basis with clear objectives and an agreed management plan between the concerned member states (France and Spain) and European Commission (EC). The fishermen of France and Spain have expressed their own points of view about the crisis of the anchovy fishery to the EC and have claimed for the definition of a long term management plan, acknowledging the need of a TAC regulation and including several concrete proposal of additional technical measures, such as spatial or seasonal closures or organization of the exploitation calendars for the different fleets (CNPMEM 2006, FECOPEGUI2006). In November 2007 the Commission produces a non-paper to be discussed by stakeholders and it is stated that alternative proposed management options should be thoroughly analyzed before they are incorporated into the long-term plan.	
Main objectives	Precautionary, catches near MSY, stable (as far as possible) and with a low risk of stock collapse	
Formal framework	STECF working group on request from Commission	
Who did the evaluation work	STECF 2008	
Method		
Software **Name, brief outline** **include ref.** **or documen-tation**	Ad hoc software, written in R. Description in STECF-SGRST 2008 and SGBRE-08-01. Two different operating models used for contrast: 1) Two-stage biomass model used by ICES (ICES, 2007; Ibaibarriaga et al. 2008). 2) Age structured operating model (seasonal) within FLR In both cases, no simulation of the analytical assessment was included in the MSE loop and the SSB index was derived perceived in the MSE as from the true population plus a log normal error.	
Type of stock	Short lived, pelagic, very valuable	
Knowledge base	Analytic assessment, 2 surveys (acoustic and DEPM)	
Type of regulation	TAC	
Model conditioning		
	Function, source of data	Stochastic? - how
Recruitment	SRRs fitted: Ricker and sensitivity to a hockey-stick and low recruitment scenario 1) fitted to median values of SSB and R at age 1 (in mass) in January of the following year produced by Bayesian biomass assessment model (Ibaibarriaga et al. 2008) 2) fitted to estimates of SSB and R at age 0 (in numbers) at 1st July of the	Log-normal, CV from residuals

	same year produced by the seasonal ICA applied in ICES (2005)		
Growth & maturity	1) G=0.52, weighted average for all ages (Ibaibarriaga et al, 2008) 2) Mean weights at age at the stock (1990-2005)	No	
Natural mortality	Annual M=1.2 for all years and ages 2) seasonal M proportional to its duration in months	No	
Selectivity	Percentage of the catches in the first semester is set equal to the historical average 1) Flat selectivity assumed. The age structure (in mass) in the catches is equal to that in the population 2) seasonal selectivity patterns produced conditioned to the assessment of the anchovy population produced in WG of ICES in 2005	No	
Initial stock numbers	From assessment: 1) Bayesian (ICES, 2007) 2) Seasonal multi-fleet ICA (ICES, 2005) updated to 2007	1) SSB2007 assumed to follow a log-normal distribution with mean equal to the median SSB in 2007 and a CV of 25%. Age 1 proportion of the population was considered exactly that estimated in the assessment. 2) Assumed known without errors	
Decision basis	SSB in the TAC year		
If assessment in the loop			
	Input data		
	Comparison with ordinary assessment?	In both types of assessments, no simulation of the analytical assessment was included in the MSE loop and the SSB index was derived perceived in the MSE as from the true population plus a log normal error	
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 (distribution, on what)	Assessment simulated based on premises that current ICES assessment (biomass based) is unbiased but subject to an observation error log-normally distributed with a CV of 25%	
	Comparison with ordinary assessment?		
Projection			
	Through intermediate	No – TAC year modified to make the ad-	

	year?	vised TAC follow on from survey/catch advice in May.
	Iteration in TAC year?	
Implementation	Perfect implementation. No feedback between biological and economic models	
Harvest rule		
Harvest rule design	**Management cycle form 1st July to 30th June next year.** **Two breakpoints on SSB: B1=24,000t, and Bpa=33,000t** **If SSB < B1, TAC= 0 t** **If B1< SSB < Bpa: TAC= 7,000 t** **If SSB > Bpa: TAC= 0.3 *SSB**	
Stabilizers	TAC shall not be lower than 7,000 t, neither bigger than 33,000 t.	
Duration	Annual	
Revision clause	Every 3 years	
Presentation of results		
Interest parameters	Median SSB (in all the projection period, in the last year of projection period), risk (probability of SSB falling below Blim, probability of SSB being below Blim at least once in the projection period, average number of year when SSB is below Blim, average number of years to get SSB above Blim), probability of fishery closure (in any year of the projection period, at least once in the projection period, average number of years the fishery is closed), catch (average, standard deviation, inter-annual variation), income, cash flow, relative wage to average of the country	
Risk type (see classification below) and time interval	Type 1, for all projection years (projection period = 10 years). Bayesian Model: 1) 1,000 iterations Seasonal ICA model: 2) 100 iterations	
Precautionary risk level	5-10% Risk type 1 (see below)	
Comments and experiences		
Review, acceptance:	Accepted by STECF, implemented from 2010 onward.	
Experiences and comments	The Regional Advisory Councils (RACs), and in particular the South Western Waters Regional Advisory Council, were consulted from the very beginning. Their involvement led to fruitful discussions with managers and scientists on the different options for rules. In fact, the harvesting rule within this long-term management plan is based on a proposal initially put forward by the RAC and subsequently analysed by the STECF. From the results of the STECF meetings, the EC proposed a draft management plan in 2009 (COM/2009/399 final). Although the plan has not been agreed in the European Parliament, it has been applied since the reopening of the fishery in 2010. ICES has not evaluated this proposal. Given that the management plan is not officially accepted, ICES advices in the basis of the precautionary approach.	

References:

STECF. 2008a. Report of the working group on a long-term management plan for the stock of anchovy in the Bay of Biscay (ICES Sub-area VIII). Meeting held in Hamburg, 14-18 April 2008.

STECF. 2008b. Report of the working Group on the long term management of Bay of Biscay anchovy. Meeting held in San Sebastian, 2-6 June 2008.

Stock: Bay of Biscay sole

Background		
Motive/ initiative/ background.	A management plan adopted in 2006 (Council Regulation (EC) No 388/2006) establishing a multiannual plan for the exploitation of the stock of sole in the Bay of Biscay. Council Regulation (EC) No 388/2006 requires that new biological targets be fixed once the stock has recovered to its precautionary biomass level. In its 2010 advice, ICES estimates that the stock of Bay of Biscay sole had reached safe biological limits (stock above BPA = 13,000 and exploited below FPA = 0.42), and consequently that the first objective of the plan had been met. Need to decide on long-term fishing mortality rate for the stock and a rate of reduction in the fishing mortality rate until this target is reached. An STECF Study Group met in November 2009 to review the plan (SGMOS 09-02). The group concluded that Fmsy would be a feasible long-term fishing mortality target for the stock. A scoping meeting (SGMOS 10-06) selected a limited number of harvest rules to be tested. STECF tested those HCR in an impact assessment meeting held in April 2011(EWG-11-01).	
Main objectives	The principle biological objectives was to fish the stock at mortality rate consistent with FMSY by 2015, and to maintain this rate in subsequent years with a low risk that the stocks will move outside safe biological limits in the medium term	
Formal framework	STECF on request from the European Commission	
Who did the evaluation work	STECF	
Method		
Software Name, brief outline include ref. or documentation	Two software were used. HCS software (ver 3.1) developped by Danker Skagen for biological impacts and IAM, written in C++/R (Merzéréaud, M., Macher, C., Bertignac, C., Frésard, M., Le Grand, C., Guyader, O., Daurès, F., Fifas, 25 S., (2011) [on line] " Description of the Impact Assessment bio-economic Model for fisheries management (IAM)", Amure Electronic Publications, Working Papers Series D-27 29-2011, 19 p. Available at : http://www.umr-amure.fr for Bio-economic impacts	
Type of stock	Medium life span, demersal.	
Knowledge base	Analytic assessment (XSA), only LPUE series.	
Type of regulation	TAC. Calculated using the harvest control rule and assuming F status quo (F in last data year) in the intermediate year.	
Model conditioning		
	Function, source of data	Stochastic? - how
Recruitment	A Hockey-Stick stock-recruitment relationship (break point at lowest observed SSB)	Log-normal, random deviates generated from SD about model fit
Growth & maturity	Weights at age as the mean of last 3 years. Maturity ogive from most recent ICES	No for maturity CVs from full data series for weight at age in HCS only, no for IAM

	assessment		
Natural mortality	M=0.1, constant.	No	
Selectivity	Recent average pattern (mean over last 3 years)	Variability from last 6 years (10-20% by age) in HCS, no for IAM	
Initial stock numbers	From assessment	? for HCS, No for IAM	
Decision basis	SSB in the TAC year		
If assessment in the loop			
	Input data		
	Comparison with ordinary assessment?		
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 **(distribution,** **on what)**		
	Comparison with ordinary assessment?		
Projection			
	Through intermediate year?	Yes, deterministic with F status quo and recruitment according to S/R relationship.	
	Iteration in TAC year?	No	
Implementation	Catches in numbers at age from projection according to the rule.		
Harvest rule			
Harvest rule design	Two management strategies tested: **1) Target F:** F = target F for SSB > Btrigger F = target F*SSB/Btrigger for SSB < Btrigger F from 0.15 to 0.65 in steps of 0.05 (lower limit to Fmsy to aprox F crash) B trigger for F 10000 to 16000 in steps of 1000 (aprox Blim to Bpa) **2) Fixed TAC rule** TAC = target TAC for SSB > Btrigger TAC = target TAC*SSB/Btrigger for SSB < Btrigger TAC = 3500 to 4500 in steps of 250		
Stabilizers	Inter-annual change in TAC limited to 10, 15 and 20% for F constraint		
Duration	Annual		
Revision clause			

Presentation of results	
Interest parameters	Risk, Catch, Inter-annual catch variation, stock size
Risk type (see classification below) and time interval	Type 1 for first five years and last 10 years of 21years simulation
Precautionary risk level	5% Risk type 1 (see below)
Comments and experiences	
Review, acceptance:	Not evaluated, targets are reached – to be considered if biomass falls below thresholds
Experiences and comments	Large range of options tested including different candidates for F targets, increasing the allowable annual TAC change, testing several Btrigger values (the biomass at which exploitation rates are reduced) and the use of a fixed TAC strategy. Short term and long term trends in stock development and TAC not very different between scenarios. For F target strategy, fishing at Fmsy can be accepted as precautionary. For Fixed TAC strategy, TAC in the range 3500 to 4500t appear to be precautionary and are predicted to deliver Fmsy by 2015. Use of bio-economic model gave very useful additional information to assess the plan. No management plan implemented yet. Will be developed as part of a future multi-specific and multi-fleet regional approach.

Stock: Norway pout in the North Sea and Skagerrak

Background		
Motive/ initiaitve/ background.	In the beginning of 2011 the SSB was more than twice the value of Bpa, but the advice for 2011 was initially zero catch (later raised to 6000 t). The zero catch advice for 2011 was due to low recruitment in 2010 (and 2011) in combination with the escapement strategy, which is based on SSB being above Bpa after the fishery has taken place. The high stock size in the beginning of 2011 and a closed fishery were seen by industry as conflicting and as a suboptimal utilization of the re-source. The new management strategy options proposed in this request include a minimum TAC to avoid closure of the fishery.	
Main objectives	To investigate options for a minimum TAC around 20 000 tonnes or higher.	
Formal framework	Request to ICES from EU/Norway.	
Who did the evalua-tion work	Ad hoc group (DTU Aqua participations only) for preparing the MSE analysis (ICES CM 2012/ACOM:69). Intensive evaluation by ICES RG/ADGPOUT, however few participants	
Method		
Software Name, brief outline include ref. or documen-tation	SMS package (in “single species” mode) with quarterly time steps, in combination with ad hoc software, written in R. Assessments are simu-lated from the operating model and “observation noise”. Mainly unpublished, documented code available on request.	
Type of stock	Short lived life span (M=1.6), potential autocorrelation in recruitment, demersal, low price	
Knowledge base	Analytic assessment, good quality survey data	
Type of regulation	TAC based on the ICES “escapement strategy”	
Model conditioning		
	Function, source of data	Stochastic? - how
Recruitment	Hockey stick model with “known” inflection point at Blim fitted to the full time series, 1983-2011	Log-normal, variance from maximum likelihood model
Growth & maturity	Average over the full time series, (kept constant in the assessment for WSEA, PROPMAT).	No
Natural mortality	As above.	No
Selectivity	Average F at age and quar-ter over years 2006- 2011 assessment.	No
Initial stock numbers	From assessment	No
Decision basis	SSB in year after the TAC year (escapement strategy)	
If assessment in the loop		
	Input data		
	Comparison with	

	ordinary assessment?	
If no assessment in the loop		
	Type of noise (distribution, on what)	Year factor, log-normal dist errors on stock numbers at age
	Comparison with ordinary assessment?	Year factor scaled to give CV of SSB as CV of SSB in assessment
Projection		
	Through intermediate year?	No intermediate year
	Iteration in TAC year?	
Implementation	Catches in numbers at age from projection according to the rule, with log-normally distributed error, CV 20%, no bias.	
	Harvest rule	
Harvest rule design	**Three different HCR with minimum (larger than 20 000 tonnes) and maximum TAC.** **1**. *Whether a management strategy is precautionary if TAC is constrained to be within the range of 20,000 - 250,000 tonnes, or another range suggested by ICES, based on the existing escapement strategy;* *2. A management strategy with a fixed initial TAC in the range of 20,000 - 50,000 tonnes. The final TAC is to be set by adding to the preliminary TAC around (50%) of the amount that can be caught in excess of 50,000 tonnes, based on a target F of 0.35;* *3. A management strategy with a fixed initial TAC in the range of 20,000- 50,000 tonnes. The final TAC is to be set by adding to the preliminary*	
Stabilizers	No stabilizer in the HCR. Assumptions that fishing mortality cannot exceed the maximum F as estimated for the last decade, and that the historical exploitation pattern (quarterly) remains the same	
Duration	Annual and half-annual	
Revision clause	After 5 years	
	Presentation of results	
Interest parameters	Risk, minimum and average catch, inter-annual variation in catch and F.	
Risk type (see classification below) and time interval	Annual risks (prob(SSB <Blim)) for years 2013–2016 (short-term risk), Average risk over the period 2017–2026 (long-term risk),	
Precautionary risk level	5% Risk type 1 (see below)	
	Comments and experiences	
Review, acceptance:	ICES advice in 2012, but not used in management for 2013.	
Experiences and com-	There might be a higher risk of two consecutive low recruitments in	

ments	the most recent period, however recruitment autocorrelation not taken into account in simulations. HCR performance with a fixed TAC is dependent on the assumption on maximum F (could be implemented as a cap on effort). The industry and managers are now interested in management based on a TAC year (e.g. 1st November- 31th October) different from the calendar year, and one annual advice (presently two ICES advices per year and half yearly TAC revisions).

Stock: Whiting in the North Sea and English Channel

Background	
Motive/ ini-tiaitve/ back-ground.	*The whiting stock in the North Sea was the only stock shared between the European Union and Norway for which there was no current management agreement. Recent uncertainty as to the status of the stock and a lack of approved Precautionary Approach reference points are one reason for this. Therefore in 2009 the EU and Norway Delegations agreed to submit a request to ICES to provide advice on the possible options for a long term management plan for whiting, taking into account these uncertainties*
Main objec-tives	Precautionary, stable catches near MSY; *maintaining fishing mortality at its current level of 0.3 would be consistent with long-term stability if recruitment is not poor*
Formal frame-work	ICES on requests from EU/Norway, 2010 and 2011. 2010: scoping the issues and possible management approaches. http://www.ices.dk/committe/acom/comwork/report/2010/Special%20Requests/EU-Norway%20request%20on%20NS%20whiting.pdf 2011: appropriate risk levels and management if recruitment is poor. http://www.ices.dk/committe/acom/comwork/report/2011/Special%20Requests/EUNorway%20MP%20for%20NS%20whiting.pdf
Who did the eval-uation work	WGNSSK 2010; Joint ICES–STECF Workshop on management plan evaluations for roundfish stocks (WKROUNDMP2/EWG 11-07), June 20–24, 2011. ICES CM 2011/ACOM: 56
Method	
Software **Name, brief outline** **include ref.** **or documen-tation**	FLR, running R 2.8.1. Age structured operating model derived from 2010 XSA assessment; full assessment (XSA) with catches at age and one simulated survey. The main focus has been on the modelling of the recruitment, because recruitment to the North Sea whiting stock has exhibited strong autocorrelation with three distinct abundance levels (subsequently referred to as high, intermediate and low). Described in WKROUNDMP2 report. Pp 79-84.
Type of stock	Medium life span, demersal, strong interactions with other species both wrt. mixed-fisheries and multi-species considerations
Knowled ge base	Analytic assessment, with uncertainty deriving from a mismatch between stock trends based on available catch and stock trends based on survey data during the period 1980 to 1995. No reference points.
Type of regula-tion	Complicated TAC setup. Advice is given for Subarea IV and Division VIId combined. However, TACs are set for Subarea IV and Divisions VIIb–k separately and there is no way of controlling how much of the Divisions VIIb–k TAC is taken from Division VIId

Model conditioning		
	Function, source of data	Stochastic? - how

Recruit-cruit-ment	three Beverton-Holt stock and recruitment relationships fitted to the high, intermediate and low recruitment phases	For each projected year, and for each individual iteration, recruitment to the stock was simulated by randomly selecting both a duration of from 1 to 6 years with equal probability and a recruitment model in the ratio 2:1:1 (high, intermediate, low) as recorded in the assessed time series; random lognormal deviations were added to each projected recruitment (sigma2 between 0.2 and 0.4 for the diff recruitment levels)	
Growth & maturity	Average over 2007-2009, no density dependence	No	
Natural mortality	Average over 2007-2009. Is based on multispecies estimates from WGSAM 2008	No	
Selectivity	Average F at age over years 2007-2009 in 2010 assessment, scaled to mean 2-6.	No	
Initial stock numbers	From assessment		
Decision basis	No SSB trigger but evaluation of a possible R trigger		
If assessment in the loop			
	Input data	Catches + 1 survey	Log normal, sd=0.3
	Comparison with ordinary assessment?	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 (distribution, on what)		
	Comparison with ordinary assessment?		
Projection			
	Through intermediate year?	Yes, deterministic with recruitment according to geometric mean	
	Iteration in TAC year?	Yes,	
Implementa-	Catches in numbers at age from projection according to the rule, no error.		

tion	
Harvest rule	
Harvest rule design	**Rule tested :** Two forms of harvest control rule were evaluated: 1) Constant fishing mortality at F=0.3 and 0.25 with TAC constraints of 0, 15, 20 and 30% 2) Fishing mortality constant at a specified target when the recent recruitment average was above a specified upper recruitment abundance threshold (Rt) with a proportionate reduction in fishing mortality below Rt down to a lower constant rate of fishing mortality (Flow) at a lower recruitment threshold (Rlow). Figure 6 illustrates an example control rule structure and compares example settings against historic recruitment and average fishing mortality. The harvest control rule structure is based on the abundance of the geometric mean of recruitment in recent years y to y+x GM recruitment (y:y+x) > Rt, F = Ftarget Rt >= GM recruitment (y:y+x) > Rlow a = (Ftarget – Flow) / (Rt – Rlow) F = a . GM + (Flow – a . Rlow) Rlow >= GM recruitment (y:y+x) F = Flow
Stabilizers	*2. Where the rule in paragraph 1 would lead to a TAC, which deviates by more than 15 % from the TAC of the preceding year, the Parties shall establish a TAC that is no more than 15 % greater or 15 % less than the TAC of the preceding year.* *3. During 2011, after obtaining advice from ICES, the Parties will refine the management plan, in particular to allow for a reduction in the target fishing mortality when recruitment to the stock has been low for a period of years.*
Duration	Annual
Revision clause	Interim plan established in 2011 has not been revised
Presentation of results	
Interest parameters	During each simulation a series of metrics were recorded for evaluation of the utility of the harvest control rule. They include the mean (across simulations and time) total catch; the probability of falling below the lowest observed biomass (Bref) both by 2015 and throughout the time series (2010-2110); the realised fishing mortality and the average inter-annual variation in catch. The metrics for each of the harvest simulated control rules evaluated are presented in Table 1 p 91 in WKROUNDMP2.
Risk type (see classifi-	Type 1 but based on median SSB across the 200 runs (not mean), for years 2010-2110.

cation below) and time interval	
Precautionary risk level	5% Risk type 1 (see below). ICES noted that with a target F of 0.3, risk probabilities were around 7-8%. Using a constant F = 0.27 in the long term resulted in around 5% probability of SSB falling below Bloss, irrespective of changes in the recruitment regime
Comments and experiences	
Review, acceptance:	The 2011 Interim plan has not been revised by EU Norway. However, based on a considerable revision in the level of fishing mortality in 2012, the target F is no longer considered applicable and the management target needs reevaluation. As an interim measure, it would be appropriate to scale the target F in the plan (0.3) according to the proportional change in F between the old and new assessment. The level of F of the whole time series was revised downwards by around 25% between the 2011 and 2012 assessments, which would generate a target F of 0.225 (0.75 * 0.3).
Experiences and comments	Complicated evaluation set-up to identify how to characterize poor recruitment (thresholds values and number of years entering the estimation). Choosing a period of 3 years is likely to give false positives. A longer period of years (5) is more stable but the actions that may be required after waiting that long may be more severe. With the revisions in the level of natural and fishing mortality in 2012, the evaluations are invalidated and should be repeated with the new stock assessment results.

Stock: Cod in Subarea IV (North Sea) and Divisions VIId (Eastern Channel) and IIIa West (Skagerrak)

Background		
Motive/ initiaitve/ background.	EC (DG MARE) requested ICES to evaluate an EC proposal for cod recovery plans. The request was extended to include a proposed management plan by the Norwegian authorities. For practical reasons, the Ad hoc group, AGCREMP, could only address the North Sea cod stock. However, the same framework was later applied to Irish Sea and West of Scotland cod stocks.	
Main objectives	Consequence of plans in terms of biological risks, yields and stability of catches.	
Formal framework	Ad-hoc Group set up by ICES to perform the evaluation, using the FLR framework.	
Who did the evaluation work	ICES. 2009. Report of the Ad hoc Group on Cod Recovery Management Plan (AGCREMP), 18–19 August 2008, Copenhagen, Denmark. ICES CM 2008/ACOM:61. 83 pp.	
Method		
Software **Name, brief outline include ref. or documen-tation**	<u>MSE framework</u>: FLR <u>Operating model</u>: B-Adapt with bias estimation and bootstrapping facility, conditioned on NS cod data <u>Management model</u>: XSA configured to approximate B-Adapt (based on the same types of data as used by the WG assessment) Full code, data and outputs for all aspects of the MSE was made available to AGCREMP, and participants in AGCREMP participated by going through the code themselves (likely still available on the site set up for the purpose).	
Type of stock	Medium life span, demersal, very valuable	
Knowledge base	Analytic assessment using B-Adapt at the time.	
Type of regulation	TAC with accompanying effort controls for EU fleets coupled with other regulations	
Model conditioning		
	Function, source of data	Stochastic? – how
Recruitment	Two scenarios considered: Ricker fitted to whole period, then alpha parameter halved to reflect lower recent period.	Fitted to best-fit pairs (and not to the different sets of bootstrap pairs), Log-normal, CV from residuals
Growth & maturity	WG estimates for operating model, assuming average of last three years of data for projections	No
Natural mortality	WG estimates for operating model , assuming last three years of data for projections	No

Selectivity	Average was calculated from estimates of the final three years of assessment, and used for projections	There will be as many selectivity average estimates as there are simulation runs	
Initial stock numbers	Taken from each boostrap simulation.	Estimates differ by simulation	
Decision basis	SSB at the end of the final year of data (i.e. at the beginning of the intermediate year) for the Long-term phase.		
If assessment in the loop			
	Input data	Catch-at-age (assumed known without error), accounting for bias in some of the scenarios; 1 survey	Survey: log normal, CV of 0.3 assumed
	Comparison with ordinary assessment?	Management model has assessment that did not quite match the WG assessment: XSA was used instead of B-Adapt, and only 1 survey was used instead of 2.	
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 (distribution, on what)	N/A	
	Comparison with ordinary assessment?	N/A	
Projection			
	Through intermediate year?	When the short-term forecast is performed, assumptions consistent with the ICES WGNSSK working group regarding recruitment, selectivity, maturity, natural mortality and mean weights at age.	
	Iteration in TAC year?	Projections to beginning of TAC year.	
Implementation	Catches in numbers at age from projection according to the rule, and following the bias scenario considered.		
Harvest rule			
Harvest rule design	**HCR consists of a Recovery Phase and Long-term Phase.** **Recovery Phase:** **Set F from 2009 onwards relative to F in 2008 with the following cuts: 25% for 2009, then a further 10% for each year in all subsequent years (i.e. 35% for 2010, 45% for 2011, etc., all relative to F in 2008).** **Long-term Phase** **Two breakpoints on SSB: Bpa=150 000t and Blim=70 000t** **If SSB ≤ Blim, F = 0.2**		

	If Blim<SSB≤Bpa: F=0.4-0.2*(Bpa-SSB)/(Bpa-Blim) **If SSB > Bpa: F = 0.4** **Change from Recovery Phase to Long-term Phase once TAC from Long-term Phase exceeds TAC from Recovery Phase for the first time. The management plan relies on the assumption that there is a one-to-one relationship between F and effort (e.g. 10% cut in F translates into a 10% cut in effort). The EU implementation of the management relies on an effort management system that makes this assumption.**
Stabilizers	TAC shall not deviate more than 20% from TAC the year before, but implemented from 2010 onwards only.
Duration	Annual
Revision clause	Subject to opt-out clauses (F close to 0.4 for three successive years, and contingency for “data-poorness”).
Presentation of results	
Interest parameters	p(>Blim), p(>Bpa), ave(Yield), ave(F)
Risk type (see classification below) and time interval	Probability that SSB is above precautionary reference points in any given year (values for 2008, 2010, 2012, 2015 reported); but information saved so that any measure of risk could be calculated.
Precautionary risk level	P(>Blim) ≥ 0.95 in 2015
Comments and experiences	
Review, acceptance:	Evaluation of EU and Norway proposals accepted by review group. But what was implemented from 2009 onward was an amalgamation of the two approaches (i.e. not actually evaluated prior to implementation).
Experiences and comments	Major review by STECF, resulting in paper: Kraak, S.B.M., Bailey, N., Cardinale, M., Darby, C., De Oliveira, J.A.A., Eero, M., Graham, N., Holmes, S., Jakobsen, T., Kempf, A., Kirkegaard, E., Powell, J., Scott, R.D., Simmonds, E.J., Ulrich, C., Vanhee, W., and M. Vinther (2013). Lessons for fisheries management from the EU cod recovery plan. Marine Policy, 37: 200–213.

Stock: Saithe in Subarea IV (North Sea), Division IIIa (Skagerrak), and Subarea VI (West of Scotland and Rockall)

Background		
Motive/ initiaitve/ background.	In accordance with point 5.3.3 of the Agreed record of Fisheries Consultations between Norway and the European Union for 2012, signed on 2 December 2011, it was agreed to convene a seminar on long term management plans. The objective of this seminar was to establish the basis for further developing long-term management plans for joint stocks. Based on the most recent assessment of the stock of saithe in ICES Subarea IV, Division IIIa and Subarea VI, ICES was requested to conduct an evaluation of the current harvest control rule with several variations	
Main objectives	Precautionary, stable catches near MSY.	
Formal framework	ICES on request from EU/Norway	
Who did the evaluation work	De Oliveira, J., Gillson, J., and Darby, C.. 2013. North Sea Saithe Management Strategy Evaluation. ICES CM 2012/ACOM:73. 45 pp. ICES 2012. Joint EU–Norway request to ICES on options to revise the Long-Term Management Plan for saithe in the North Sea. Special request, Advice November 2012. Section 6.3.3.5, ICES Advice 2012, Book 6: 6pp	
Method		
Software **Name, brief outline** **include ref.** **or documentation**	MSE framework: FLR Operating model: B-Adapt with no bias estimation to take advantage of bootstrapping facility (not available in XSA), conditioned on NS saithe data with comparisons to XSA WG assessment to ensure broad consistency. Management model: XSA (based on the same types of data as used by the WG assessment) Full code, data and outputs for all aspects of the MSE available on the RGHELP, RGSaithe and ADGHERSA sharepoint site	
Type of stock	Medium to long-lived fish, demersal, very valuable	
Knowledge base	Analytic assessment, has had technical difficulties in the past	
Type of regulation	TAC	
Model conditioning		
	Function, source of data	Stochastic? - how
Recruitment	Beverton-Holt constrained by a hockey-stick model fitted to two periods of SR pairs, reflecting optimistic (1967-2011) and pessimistic (1988-2011) periods	Log-normal, CV from residuals (0.52 for optimistic and 0.44 for pessimistic). To check recruitment simulated were consistent with historic observations, qq-plot comparisons were made.
Growth & maturity	Optimistic and pessimistic growth scenarios (consistent with the recruitment ones) were considered. WG ma-	Re-sampling, with replacement, year vectors of weights at age from the entire time-series of stock weight estimates from the relevant periods (see recruitment scenarios). Plots compare consistency of simulated data to

	turity (constant over time) was used in the operating model and considered known in the management model	historic observations for weights-at-age. No auto-correlations accounted for though.	
Natural mortality	WG natural mortality (0.2 for all years and ages) was used in the operating model and considered known in the management model.	No	
Selectivity	Average was calculated from the final three years of "data" from the management model to be used in the short-term forecast for setting the TAC	There will be as many selectivity average estimates as there are simulation runs (250 for this work)	
Initial stock numbers	Full-feedback MSE, so the management model will do an assessment for each year in future during each simulation.	250 simulations by 20 year projection period, and an assessment is conducted for each of these	
Decision basis	SSB at the end of the final year of data (i.e. at the beginning of the intermediate year).		
If assessment in the loop			
	Input data	Catch-at-age (assumed known without error), 1 survey	Survey: log normal, CV of 0.3 assumed
	Comparison with ordinary assessment?	Same settings as WG assessment used, except only 1 survey used.	
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 (distribution, on what)	N/A	
	Comparison with ordinary assessment?	N/A	
Projection			
	Through intermediate year?	When the short-term forecast is performed, the same assumption as the ICES WGNSSK working group regarding the intermediate year catch (constrained to the actual TAC previously set for that year) is made. GM recruitment from 1988 to final year of data.	
	Iteration in TAC year?	Projections to beginning of TAC year.	
Implementation	Catches in numbers at age from projection according to the rule, no bias.		

Harvest rule	
Harvest rule design	**Two breakpoints on SSB: Bpa=200 000t and Blim=106 000t** **If SSB ≤ Blim, F = 0.1** **If Blim<SSB≤Bpa: F=0.30-0.20*(Bpa-SSB)/(Bpa-Blim)** **If SSB > Bpa: F = 0.3**
Stabilizers	When SSB > Blim, TAC shall not deviate more than 15% from TAC the year before
Duration	Annual
Revision clause	After 5 years
Presentation of results	
Interest parameters	p(>Blim), p(>Bpa), median(SSB), median(C), median(F) for the years 2015 and 2020, and for the average of 2020-2029 ICV is calculated as the absolute value of {1 – Catch(y+1)/Catch(y)}, averaged over either y=2013 to 2020 or y=2020 to 2029
Risk type (see classification below) and time interval	Probability that SSB is above precautionary reference points in any given year; but information saved so that any measure of risk could be calculated.
Precautionary risk level	P(>Blim) ≥ 0.95
Comments and experiences	
Review, acceptance:	Accepted by review group, but recommending a re-evaluation within 4 years (because long-term performance not clear).
Experiences and comments	This was essentially a re-evaluation of the HCR currently in place, with a look at alternative TAC stability options; and confirmed the current HCR and alternatives investigated to be consistent with the precautionary approach in the short term. May want to consider a refinement on growth modeling (e.g. accounting for auto-correlation).

Stock: Western Baltic Spring spawning herring

Background		
Motive/ initiaitve/ background.	Despite its relatively small stock size and economic value, Western Baltic spring spawning herring (WBSS) is managed in a highly complex governance scheme, with demanding scientific challenges and an elaborate political process of resource allocation among fishing fleets. WBSS herring spawns in the western Baltic Sea, where it is exploited by several EU fishing fleets. It migrates into the Kattegat, Skagerrak and eastern North Sea areas, where it mixes with North Sea autumn spawning herring (NSAS), in an age and season-dependent pattern with high variability, and where it is exploited by both EU and non-EU fleets. For the two separate management areas, TACs are set at different times in the yearly TAC-setting process, and this can result in conflicts over quota allocations to individual fleets. Industry stakeholders of two Regional Advisory Councils – the Pelagic and Baltic Sea RACs – and scientists involved in the FP7 JAKFISH project engaged in collaboration, aiming to improve stock management through joint development of a robust Long-Term Management Plan. A common understanding of relevant scientific and political issues was developed and used to conduct Management Strategies Evaluations in an interactive process.	
Main objectives	Precautionary, stable catches near MSY; accounting for mixing with NSAS in the TAC	
Formal framework	ICES WKMAMPEL 2009; EU FP7 JAKFISH 2008-2011 /GAP I and GAP II ICES WKWATSUP 2010 ICES HAWG	
Who did the evaluation work	DTU Aqua, as part of EU FP7 JAKFISH.	
Method		
Software **Name, brief outline** **include ref.** **or documen-tation**	FLR, running R 2.8.1. Age structured operating model derived from 2009 assessment; no full assessment but SAM vcov with catches at age and one simulated survey. Described in Ulrich et al., **ICES CM 2010 / P:07**	
Type of stock	Medium life span, pelagic, large migrations to the North Sea, mixing with other herring stocks	
Knowledge base	Analytic assessment (ICA, SAM),	
Type of regulation	Complicated TAC setup. advice accounts for the mixing with NSAS, and issues in TAC sharing between areas (IIIa – WB) and fleets (HC/Ind).	
Model conditioning		
	Function, source of	Stochastic? - how

	data		
Recruitment	a Hockey-Stick SRR was chosen with average recruitment at the recent (2003-2007) geometric mean of assessment estimates and break point at the lowest observed SSB (112 000t).	For each projected year, and for each individual potentially large year-classes were allowed to occur in the simulations through a high random deviation in the SRR (CV calculated on the full time-series =0.53)	
Growth & maturity	Average over 2006-2008, no density dependence	No	
Natural mortality	Average over 2006-2008.	No	
Selectivity	Average F at age over years 2006-2008 in 2009 assessment, scaled to mean 3-6.	No	
Initial stock numbers	From assessment		
Decision basis	Evaluation of various SSB timing (data year, intermediate year, TAC year, TAC year p 1)		
If assessment in the loop			
	Input data	Some XSA/ICA trials with Catches + 1 survey But performed poorly (FLICA not appropriate for stochastic scenarios)	Log normal error on index, sd=average sd across WBSS tuning fleets. Observation error on catches with lognormal deviates (sd = 0.2 for young ages, 0.3 for older ages)
	Comparison with ordinary assessment?	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 (distribution, on what)	Vcov from SAM / ICA assessments	

	Comparison with ordinary assessment?	Some runs performed with perfect assessment also
Projection		
	Through intermediate year?	F status quo
	Iteration in TAC year?	no
Implementation	Some uncertainty in the catches in numbers at age from projection according to the rule, linked to the varying mixing % between NSAS and WBSS (uniform distribution around +/- 25%	
Harvest rule		
Harvest rule design	**Rule tested :** Runs with PA rules with Btrigger/Blow, HCR like MSY framework with Btrigger only Final scenario accepted (Target F=0.25, and sloped F if SSB<110kT	
Stabilizers	15% TAC IAV	
Duration	Annual	
Revision clause	Interim plan established in 2011 has not been revised	
Presentation of results		
Interest parameters	Evaluation criteria were mainly defined by the PelRAC, and were very detailed, including e.g. Total number of times that the TAC was adjusted up/downwards, Total number of times that the IAV rule came into action, preventing TAC increase bigger than 15%, Mean amount of increase when the TAC goes up/down etc See table 2 in Ulrich et al 2010 ICES paper	
Risk type (see classification below) and time interval	Both Risk 1 and Risk 2 : P(<Blim09-12), LTavgSSB, SSB2032, MeanAge, <Blim1Yr, <Blim2Yr, P(<Blim)	
Precautionary risk level	5% Risk type 1 (see below).	
Comments and experiences		
Review, acceptance:	Consensus agreed between PelRAC and BSRAC based on the results. JAKFISH results consistent with Fmsy. No LTMP implemented however. Still some political issues in the TAC sharing	
Experiences and comments	Interesting endeavor, good collaboration and participation of PelRAC and BSRAC	

Stock: Cod in Icelandic waters (cod-iceg)

Background		
Motive/ initiaitve/ background.	An initial HCR set in place in 1995. Harvest rate set higher than initially mended, the latter being based on maximum economic yield. Assessment timation around 2000 and low recruitment as well as management actio followed resulted in negative consequences for the stock and the fisheries. the initial recommendation from the early 1990's revisited and as a part process a formal evaluation was requested from ICES.	
Main objectives	Maximize catches (and profit), inter-annual stability in catches	
Formal framework	ICES on request from Iceland	
Who did the evaluation work	AGICOD 2009	
Method		
Software Name, brief outline include ref. or documen-tation	AD model builder. R code (FPRESS derivative) used concurrently. Age structured operating model (Fool's approach). Unpublished, undocumented, code and 2009 results available on request. The model predicts reference biomass in the assessment year, adds asse error to it and calcules the TAC for the next fishing year from September August 31st the following year. The split between Fishing year and calend is that 1/3 of the catch in a Fishing year is taken before December 31st from January 1st to August 31st. For Saithe and Haddock catch in fishing also modelled.	
Type of stock	Medium life span, demersal, very valuable	
Knowledge base	Analytic assessment	
Type of regulation	TAC, minimum landings size, mesh size requirements, closed areas.	
Model conditioning		
	Function, source of data	Stochastic? - how
Recruitment	Various Stock Recruitment functions tested. Recruiment since the 1986 year-class has been 30-40% lower on the average than before and periods of different productivity were tested.	Various types tested, CV a f spawning stock size is used and a related of residuals were tested.
Growth & maturity	Average over 2006-2008 (low) and 1985-2008 (mean) , no density dependence Maturity 2006-2008, no density dependence	Stochastic true weights around low mean weights, autocorrelated each year (effectively means tha lated weight could be below hi range). Maturity at age fixed,
Natural mortality	*M = 0.2*	No
Selectivity	2006-2008	No
Initial stock numbers	From assessment	According to variance - covarian trix from assessment (inverse Hessian). With added e noise based on empirical asse performance.
Decision basis	TAC in the advisory year (y+1) based on multiplier of reference biomass (in the assessment year modified by SSB(y)/Btrigger if SSB(y) < Btrigger tional catch stabilizer included. Reference biomass based on catch weig much emphasis on age 4 where catch weights are 70% higher than stock w Stabilizer does weight it down again.	

If assessment in the loop		
	Input data	
	Comparison with ordinary assessment?	
If no assessment in the loop	*Below is just an example of how this could be presented if there was no asse in the loop*	
	Type of noise (distribution, on what)	Autocorrelated log-normal year fa the reference biomass in the asse year (B4+,y), the cv and rho ba empirical assessment performance
	Comparison with ordinary assessment?	??
Projection		
	Through intermediate year?	None needed, the rule being ba multiplier of reference biomass assessment year (B4+,y)
	Iteration in TAC year?	None
Implementation	No implementation error included	
Harvest rule		
Harvest rule design	**One breakpoints on SSB: Btrigger** **If SSB < Btrigger, HR = HR*SSB/Btrigger** **If SSB >= Btrigger: HR = HR**	
Stabilizers	If SSB >= Btrigger: TACy+1 = 0.5*(TACy + HR * B4+y)	
Duration	Annual	
Revision clause	None implicit	
Presentation of results		
Interest parameters	Risk, Catch (Mean and 10-50-90 percentiles), Inter-annual variation	
Risk type (see classification below) and time interval		
Precautionary risk level	5% of SBB2015 < SSB2009 (SSB2009 equivalent to Btrigger)	
Comments and experiences		
Review, acceptance:	Accepted by review group, implemented from 2009 onward	
Experiences and comments	Minor 'last minutes' amendments of rules that have been tested via sim should not be 'allowed' (1995 experience). The current HCR leads to MSY, probability of falling below Blim (= B very low (< 5%) irrespective of Risk type, probability of SSB < Btrigge per/year (Risk1). Persistent positive empirical estimation bias (5%) taken into account in the tion. Simple decision like used here are easy to implement in simulation and m for very transparent communication.	

Stock: Haddock in Icelandic waters (had-iceg)

Background		
Motive/ initiaitve/ background.	MSY whatever that means. Formalise advice given in recent years.	
Main objectives	Maximize 5 and 10 percentiles of catches. In reality maximizing catches, k of going below Blim less than 5% and keep some stability in catches by a yearclasses to last longer.	
Formal framework	ICES on request from Iceland	
Who did the evaluation work	Will be evaluated in March 2013.	
Method		
Software Name, brief outline include ref. or documen-tation	AD model builder. R code. Age structured operating model (Fool's approach). Catch in fishing years (S ber 1st to August 31st) is modelled as for Icelandic cod. .	
Type of stock	Medium life span, demersal,valuable	
Knowledge base	Analytic assessment	
Type of regulation	TAC mesh size limits, quick area closures based on proportion < 45cm. .	
Model conditioning		
	Function, source of data	Stochastic? - how
Recruitment	Hockey stick	Parameters (cv, Rmax, Breakpoi mated. Autocorrelation of recr modelled as 1st order AR model. ed in simulations.
Growth & maturity	All biological parameters except catch weights derived from survey in March (stockweights) . Catch weights and maturity derived from stock weights based on data from 2000-2011 Growth average since approximately 1990, growth is density dependent and mean weight at age 2 inversly correlated to yearclass size. Substantial downward trend in mean weight at age.	Growth is modelled as function o size with autocorrelated noise a the weights each year. Maturity at size fixed,
Natural mortality	*M = 0.2*	Included in assessment error.
Selectivity	Long term as function of weight at age in the stock.	Very limited.
Initial stock numbers	From assessment	According to variance - covarian trix from assessment (inverse Hessian). With added noise based on empirical ass performance.
Decision basis	TAC in the advisory year (y+1) based on multiplier of reference biomass (B the advisory year. No additional catch stabilizer included. Btrigger=Blim Biomass 45+ derived from number in stock and stock weights.	
If assessment in the loop		

	Input data	
	Comparison with ordinary assess-ment?	
If no assessment in the loop	*Below is just an example of how this could be presented if there was no ass in the loop*	
	Type of noise (distribution, on what)	Autocorrelated log-normal year fa the reference biomass in the a year (B45+,y+1), the cv and rho b empirical assessment. performan
	Comparison with ordinary assess-ment?	??
Projection		
	Through intermediate year?	None needed, the rule being ba multiplier of reference biomass assessment year (B4+,y)
	Iteration in TAC year?	None
Implementation	No implementation error included	
Harvest rule		
Harvest rule design	**One breakpoints on SSB: Btrigger** **If SSB < Btrigger, HR = HR*SSB/Btrigger** **If SSB >= Btrigger: HR = HR**	
Stabilizers	If SSB >= Btrigger: TACy+1 = 0.5*(TACy + HR * B4+y)	
Duration	Annual	
Revision clause	None implicit	
Presentation of results		
Interest parameters	Risk, Catch (Mean and 10-50-90 percentiles), Inter-annual variation	
Risk type (see classification below) and time interval		
Precautionary risk level	5% of SBB2015 < SSB2009 (SSB2009 equivalent to Btrigger)	
Comments and experiences		
Review, acceptance:	Accepted by review group, implemented from 2009 onward	
Experiences and comments	Recruitment highly variable, large contrast in yearclass size and autocorr Low fishing effort can help reducing variability in catches if the periods recruitment is large. Negative correlation between stocksize and growth variability. Advice in recent years has been close to the HCR so the change in advice be large.	

Stock:Golden redfish in Greeland, Iceland and Faeroes.

Background		
Motive/ initiaitve/ background.	Try to formulate the advice and base it non alnalytical assessments that ha conducted with reasonable success since 1999. Managementplan does not much change from adviced fishing mortality from recent years.	
Main objectives	Maximize catches and have high probability being above Blim (SSBloss model used is length based so Fmax can be close to target value as long probability of SSB < Blim isl low.	
Formal framework		
Who did the evaluation work		
Method		
Software Name, brief outline include ref. or documen-tation	Gadget model used for assessment and predictions. Predictions are based o fied effort (fishing moratlity). Assessment error included as multiplier on t tality each year. Model length based but available age information that ra quite extensive are used. Simulation time 50 years.	
Type of stock	Long lived (30 years), demersal, valuable	
Knowledge base	Analytic assessment	
Type of regulation	TAC	
Model conditioning		
	Function, source of data	Stochastic? - how
Recruitment	Available data (yearclasses 1975-2003) do not show any relationship between spawning stock and recruitment or serial correlation of residuals. Series though relatively short for that kind of inference for so longlived stock. Blim is therefore proposed as Bloss (160) and Btrigger as 220 thous. tonnes	Random draw from available data ing autocorrelation would only b lematic is the correlation time w long as yearclasses last long tim fisheries.
Growth & maturity	Fixed growth and maturity ogive fixed by size. Available data indicate that maturity by size and age has been increasing much since 2005 but the fixed maturity used does not take that increase into account.	Uncertainty in growth included sessment error.
Natural mortality	*M = 0.05*	No
Selectivity	1990-2010 size based.	No
Initial stock numbers	From assessment	None. Simulationtime long enc current values do not have muc near the end of the simulation tim
Decision basis	Tac in the advisory year is based on keeping fishing mortality of ages 9 than 0.097. The model works on F on fully recruited fish which is 0.15 w 19 is 0.097 that is used as a reference if TSA was used as an assessment Percent of biomass above certain size is also an option with percentage sel fishing effort is close to what will be obtained using F=0.097. .	
If assessment in the loop		
	Input data	
	Comparison with ordinary assessment?	
If no assessment in the loop	*Below is just an example of how this could be presented if there was no ass*	

	in the loop	
	Type of noise (distribution, on what)	Autocorrelated log-normal year fa the fishing mortality in the adviso 1st order AR model with CV=(rho=0.85. CV partly based on re TSA model but high "assumed" a relation is to cover slowly changin acteristics of the stock dynamics variable growth and natural mor included in the assessment error.
	Comparison with ordinary assessment?	??
Projection		
	Through intermediate year?	Nultiplier of reference bione need rule being based on momass in sessment year (B4+,y)
	Iteration in TAC year?	None
Implementation	No implementation error included	
Harvest rule		
Harvest rule design	**One breakpoints on SSB: Btrigger** **If SSB < Btrigger, HR = HR*SSB/Btrigger (really a little more comp rule is used to avoid discontinuity.** **If SSB >= Btrigger: HR = HR**	
Stabilizers	None, the model is already a low pass filter.	
Duration	Annual	
Revision clause	None implicit	
Presentation of results		
Interest parameters	Risk, Catch (Mean and 10-50-90 percentiles), Inter-annual variation	
Risk type (see classification below) and time interval		
Precautionary risk level	1% of SBB < 160 kt.	
Comments and experiences		
Review, acceptance:		
Experiences and comments		

Stock: saithe in Icelandic waters (cod-iceg)

Background		
Motive/ initiaitve/ background.	Try to formalize the advice that has for many years been based on F4-9 b 0.28 and 0.3. Take into account realtively uncertain stock estimate. Fishing effort by HCR close to what has been adviced in recent years. Catch by fishing years is modelled.	
Main objectives	Maximize catches an have more than 95% probability being above (SSBloss). Type I error.	
Formal framework	ICES on request from Iceland	
Who did the evaluation work	Will be done in March 2013.	
Method		
Software Name, brief outline include ref. or documen-tation	AD model builder. R code (FPRESS derivative) used concurrently. Age structured operating model (Fool's approach). Unpublished, undocumented, code and 2009 results available on request.	
Type of stock	Medium life span, demersal, valuable	
Knowledge base	Analytic assessment	
Type of regulation	TAC	
Model conditioning		
	Function, source of data	Stochastic? - how
Recruitment	Hockey stick with SSB break and auto-correlation parameter fixed, but Rmax and CV estimated.	Lognormal with CV estimated ar correlation given (based on a data).
Growth & maturity	Average over 2009-2011 (low) , no density dependence Maturity 2011, gengerated by a glm model through survey data so it is really descriptive of few recent years.	Stochastic true weights around low mean weights, autocorrelatec each year (effectively means tha lated weight could be below hi range). Maturity at age fixed,
Natural mortality	*M = 0.2*	No
Selectivity	2004-2011. (Targeting of small fish)	Very limited.
Initial stock numbers	From assessment	According to variance - covariar trix from assessment (inverse Hessian). **Not** with addec nal noise based on empirical asse performance unlike cod and hadd
Decision basis	TAC in the advisory year (y+1) based on multiplier of reference biomass (in the assessment year modified by SSB(y)/Btrigger if SSB(y) < Btrigger tional catch stabilizer included.	
If assessment in the loop		
	Input data	
	Comparison with ordinary assess-ment?	
If no assessment in the loop	*Below is just an example of how this could be presented if there was no asse in the loop*	
	Type of noise (distribution, on what)	Autocorrelated log-normal year fa the reference biomass in the asse year (B4+,y), the cv and rho ba empirical assessment perfor Variance covariance matrix fr sessment gives similar CV. Asse

		uncretain as the survey data are no
	Comparison with ordinary assess-ment?	??
Projection		
	Through intermediate year?	None needed, the rule being ba multiplier of reference biomass assessment year (B4+,y)
	Iteration in TAC year?	None
Implementation	No implementation error included	
Harvest rule		
Harvest rule design	**One breakpoints on SSB: Btrigger** **If SSB < Btrigger, HR = HR*SSB/Btrigger (really a little more comp rule is used to avoid discontinuity.** **If SSB >= Btrigger: HR = HR**	
Stabilizers	If SSB >= Btrigger: TACy+1 = 0.5*(TACy + HR * B4+y)	
Duration	Annual	
Revision clause	None implicit	
Presentation of results		
Interest parameters	Risk, Catch (Mean and 10-50-90 percentiles), Inter-annual variation	
Risk type (see classification below) and time interval		
Precautionary risk level	5% of SBB2061 < 65 (65 (SSBloss) equivalent to Btrigger and perhaps Blin	
Comments and experiences	Definition of Blim somewhat problematic as the stock has never be pleted to any near candidate of Blim. Therefore type of risk (1, 2 a mattes when looking at the risk being below Blim.	
Review, acceptance:		
Experiences and comments		

Stock: Northeast arctic cod

Background		
Motive/ initiative/ background.	Managers (Joint Norwegian-Russian Fisheries Commission, JNRFC) suggested a harvest control rule at its 2002 meeting. Stock had been fished at very high levels for decades. The HCR was initially evaluated in 2004, found to be precautionary provided adequate measures to ensure rebuilding of the stock were introduced. Rule amended in 2004 by JNRFC, including pre-agreed measures for a rebuilding situation. This amended rule was evaluated in 2005 and found to be precautionary. Some additional evaluations have been made by AFWG in recent years, but the biological model has not been updated. The stock has never been benchmarked, but a benchmark is scheduled for 2014.	
Main objectives	Precautionary, stable catches near MSY	
Formal framework	ICES on request from JNRFC	
Who did the evaluation work	AFWG	
Method		
Software Name, brief outline include ref. or documentation	Ad hoc software (PROST), written in C++. Documentation and code available on request from IMR (Bjarte Bogstad). An earlier version of the documentation has been made available as WD 2 to AFWG 2005. The programmer has left IMR, and further development of the code may not be feasible. The only change in PROST in recent years is the introduction of the F=0.30 lower limit. The biological model used is described in Kovalev and Bogstad 2005. Kovalev, Y., and Bogstad, B. 2005. Evaluation of maximum long-term yield for Northeast Arctic cod. In Shibanov, V. (ed.): Proceedings of the 11th Joint Russian-Norwegian Symposium: Ecosystem dynamics and optimal long-term harvest in the Barents Sea fisheries. Murmansk, Russia 15-17 August 2005. IMR/PINRO Report series 2/2005, p. 138-157.	
Type of stock	Medium life span, demersal, very valuable	
Knowledge base	Analytic assessment (XSA), 3 surveys and 1 CPUE series. 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.	
Model conditioning		
	Function, source of data	Stochastic? - how
Recruitment	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 6-9, with upper and lower limits	No

	Weight in catch a function of weight in stock for ages 3-8 Maturity at age a function of weight at age for ages 5-10	
Natural mortality	M=0.2. Sensitivity to Higher M on ages 3 and 4 due to cannibalism investigated. Cannibalism included in assessment based on annual stomach content data. Model for cannibalism developed but not used in evaluations.	No
Selectivity	Recent average pattern	No
Initial stock numbers	From assessment	Age dependent CV, no correlation between age groups.
Decision basis	SSB in the TAC year	
If assessment in the loop		
	Input data	
	Comparison with ordinary assessment?	
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 (distribution, on what)	Assessment error/bias derived from historical data, implemented as CV and bias. Age dependent, no correlation between age groups.
	Comparison with ordinary assessment?	
Projection		
	Through intermediate year?	Yes, deterministic with recruitment according to estimate based on recruitment indices. A TAC constraint is applied in the intermediate year, this is not consistent with assessment procedure which assumes F status quo in intermediate year
	Iteration in TAC year?	No (SSB is calculated at 1 January although mean spawning time is about 1 April)
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	**One breakpoint on SSB: B1 (460 000 tonnes). Ftarget=0.40 (age 5-10, arithmetic average)** **If SSB < B1, F = Ftarget*SSB/B1** **If B1<SSB: F=Ftarget**
Stabilizers	Estimate the average TAC level for the coming 3 years based on F as calculated above, this gives the TAC for the quota year. TAC shall not deviate more than 10% from TAC the year before, unless SSB< B1 in intermediate year or 3 following years (i.e. larger deviations are possible if SSB passes B1 on the way up or down). Lower F limit of 0.30.
Duration	Annual
Revision clause	No clause, but revision asked for by 2015 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 (see classification below) and time interval	Type 1 for years 21-120, to avoid initial transients
Precautionary risk level	5% Risk type 1 (see below)
Comments and experiences	
Review, acceptance:	Accepted by ICES in 2005
Experiences and comments	Reality check (long-term stochastic simulations) made to ensure that when fishing at the historic average level, modelled stock size will be in correspondence with the historic average. At about the same time, all the following elements occurred: IUU eliminated Strong year classes (2004 and 2005) entered the fishable stock Underestimation of stock size When a harvest control rule is introduced and fishing mortality decreases, the stock size will in general increase. The advised quota will then increase after a drop in the transition year(s). If there is a stability element in the HCR, this may in such transitional phases limit the catches so that the actual fishing mortality is lower than what is intended. For NEA cod this became a serious issue due to the elements mentioned above, which all influenced the situation in the same direction. Thus a lower F limit of 0.30 was introduced by JRNFC when setting the 2010 quota to avoid underexploitation. In some years, both before and after the introduction of this lower limit to F, the quota has

	been set higher than the advice. The total stock biomass (TSB) is now at a level not experienced since the early 1950s, and SSB is at an all-time high. Despite this, growth and maturation has remained fairly stable.

Stock: Northeast arctic haddock

Background		
Motive/ initiative/ background.	Managers (Joint Norwegian-Russian Fisheries Commission, JNRFC) suggested a harvest control rule at its 2002 meeting. Evaluated in 2007 and found to be precautionary.	
Main objectives	Precautionary, stable catches near MSY	
Formal framework	ICES on request from JNRFC	
Who did the evaluation work	AFWG	
Method		
Software Name, brief outline include ref. or documentation	Ad hoc software (PROST), written in C++. Documentation and code available on request from IMR (Bjarte Bogstad). An earlier version of the documentation has been made available as WD 2 to AFWG 2005. The programmer has left IMR, and further development of the code may not be feasible. The biological model used is described in WKHAD 2006.	
Type of stock	Medium life span, demersal, valuable	
Knowledge base	Analytic assessment (XSA), 3 survey series. Predation 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.	
Model conditioning		
	Function, source of data	Stochastic? - how
Recruitment	Cyclic term (7-year period with pattern of strong/moderate/weak year classes), fitted to S/R time series for year-classes 1950-2001 using Ricker and hockey-stick	Log-normal, CV from residuals
Growth & maturity	Weight in stock a function of total stock biomass in previous years for ages 3-7, with upper and lower limits Weight in catch a function of weight in stock for ages 3-7 Maturity at age a function of weight at age for ages 3-7	No
Natural mortality	M on ages 3-6 set to	No

	historic average (calculated assuming M=0.2+predation by cod, calculated based on annual stomach content data)	
Selectivity	Recent average	No
Initial stock numbers	From assessment	Age dependent CV, no correlation between age groups.
Decision basis	SSB in the TAC year	
If assessment in the loop		
	Input data	
	Comparison with ordinary assessment?	
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 (distribution, on what)	Assessment error/bias derived from historical data, implemented as CV and bias. Age dependent, no correlation between age groups.
	Comparison with ordinary assessment?	
Projection		
	Through intermediate year?	Yes, deterministic with recruitment according to estimate based on recruitment indices. A TAC constraint is applied in the intermediate year, this is not consistent with assessment procedure which assumes F status quo in intermediate year
	Iteration in TAC year?	No (SSB is calculated at 1 January although mean spawning time is about 1 April)
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	**One breakpoint on SSB: B1 (80 000 tonnes). Ftarget=0.35 (age 4-7, arithmetic average)** **If SSB < B1, F = Ftarget*SSB/B1** **If B1<SSB: F=Ftarget**	
Stabilizers	TAC shall not deviate more than 25% from TAC the year before, unless SSB< B1 in intermediate year or TAC year (i.e. larger deviations are possible if SSB passes B1 on the way up or down).	
Duration	Annual	

Revision clause	No clause, but revision asked for by 2015 by JNRFC
Presentation of results	
Interest parameters	Risk, Catch, Inter-annual catch variation, proportion of years when different parts of the HCRs apply, stock size
Risk type (see classification below) and time interval	Type 1 for years 21-120, to avoid initial transients
Precautionary risk level	5% Risk type 1 (see below)
Comments and experiences	
Review, acceptance:	Accepted in 2007. Stock benchmarked in 2011 (WKBENCH).
Experiences and comments	Reality check (long-term stochastic simulations) made to ensure that when fishing at the historic average level, modelled stock size will be in correspondence with the historic average. The 2004-2006 year classes were all very strong, and three strong year classes in a row have not been observed previously. This invalidated the recruitment function used in the HCR testing. At present, these year classes are on the way out of the stock, and the catch and stock levels are decreasing. The 25% limit on TAC decrease from year to year may now prove too restrictive, as it is likely to lead to very high F values in 2013-2015. The trigger point is quite low, which makes the situation worse.

Stock: Northeast arctic saithe

Background		
Motive/ initiative/ background.	Managers (Norway) suggested a harvest control rule in 2004. This rule was evaluated in 2007 and found to be precautionary.	
Main objectives	Precautionary, stable catches near MSY	
Formal framework	ICES on request from JNRFC	
Who did the evaluation work	AFWG	
Method		
Software Name, brief outline include ref. or documentation	Ad hoc software (PROST), written in C++. Documentation and code available on request from IMR (Bjarte Bogstad). An earlier version of the documentation has been made available as WD 2 to AFWG 2005. The programmer has left IMR, and further development of the code may not be feasible. The biological model used is described in AFWG 2007.	
Type of stock	Medium life span, demersal, valuable	
Knowledge base	Analytic assessment (XSA), 1 survey and 1 CPUE series.	
Type of regulation	TAC	
Model conditioning		
	Function, source of data	Stochastic? - how
Recruitment	Beverton-Holt, fitted to S/R time series for year-classes 1960-2003	Log-normal, CV from residuals
Growth & maturity	Weight in stock a function of total stock biomass in previous years for ages 6-9, with upper and lower limits Weight in catch equal to weight in stock Maturity at age constant	No
Natural mortality	M=0.2	No
Selectivity	Recent average	No
Initial stock numbers	From assessment	Age dependent CV, no correlation between age groups.

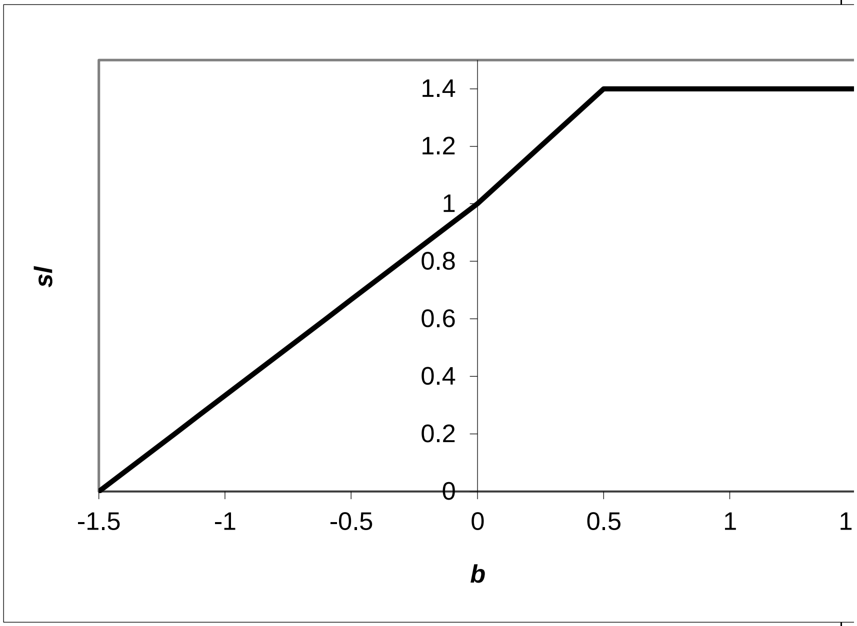
<table>
<tr><td>Decision basis</td><td colspan="3">SSB in the TAC year</td></tr>
<tr><td colspan="4">If assessment in the loop</td></tr>
<tr><td></td><td>Input data</td><td></td><td></td></tr>
<tr><td></td><td>Comparison with ordinary assessment?</td><td colspan="2"></td></tr>
<tr><td>If no assessment in the loop</td><td colspan="3">Below is just an example of how this could be presented if there was no assessment in the loop</td></tr>
<tr><td></td><td>Type of noise (distribution, on what)</td><td colspan="2">Assessment error/bias derived from historical data, implemented as CV and bias. Age dependent, no correlation between age groups.</td></tr>
<tr><td></td><td>Comparison with ordinary assessment?</td><td colspan="2"></td></tr>
<tr><td>Projection</td><td colspan="3"></td></tr>
<tr><td></td><td>Through intermediate year?</td><td colspan="2">Yes, deterministic with recruitment according to estimate based on recruitment indices. A TAC constraint is applied in the intermediate year, this is not consistent with assessment procedure which assumes F status quo in intermediate year</td></tr>
<tr><td></td><td>Iteration in TAC year?</td><td colspan="2">No (SSB is calculated at 1 January although mean spawning time is about 1 April)</td></tr>
<tr><td>Implementation</td><td colspan="3">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.</td></tr>
<tr><td colspan="4">Harvest rule</td></tr>
<tr><td>Harvest rule design</td><td colspan="3">One breakpoint on SSB: B1 (220 000 tonnes). Ftarget=0.35 (age 4-7, arithmetic average)
If SSB < B1, F = Ftarget*SSB/B1
If B1<SSB: F=Ftarget</td></tr>
<tr><td>Stabilizers</td><td colspan="3">Estimate the average TAC level for the coming 3 years based on F as calculated above, this gives the TAC for the quota year. TAC shall not deviate more than 15% from TAC the year before, unless SSB< B1 in intermediate year or 3 following years (i.e. larger deviations are possible if SSB passes B1 on the way up or down).</td></tr>
<tr><td>Duration</td><td colspan="3">Annual</td></tr>
<tr><td>Revision clause</td><td colspan="3">No clause</td></tr>
<tr><td colspan="4">Presentation of results</td></tr>
<tr><td>Interest parameters</td><td colspan="3">Risk, Catch, Inter-annual catch variation, proportion of years when different parts of the HCRs apply, stock size</td></tr>
<tr><td>Risk type (see classification below) and time interval</td><td colspan="3">Type 1 for years 21-120, to avoid initial transients</td></tr>
</table>

Precautionary risk level	5% Risk type 1 (see below)
Comments and experiences	
Review, acceptance:	
Experiences and com-ments	Reality check (long-term stochastic simulations) made to ensure that when fishing at the historic average level, modelled stock size will be in correspondence with the historic average. Stock benchmarked in 2010 (WKROUND)

Stock: Western Horse mackerel

Background		
Motive/ initiaitve/ background.	The industry was not satisfied with fluctuations in quotas based on assessment which was unstable, and developed a proposed management plan. Managers requested ICES to develop the proposal further and advise on a plan.	
Main objectives	Catch stability, low ecological impact, Consistent with PA	
Formal framework	None, PRAC led initiative involving scientists from UK, IRL, NL & ES	
Who did the evaluation work	Informal evaluation carried out on the fringes of ACFM in 2008	
Method		
Software Name, brief outline include ref. or documentation	FPRESS (R), bespoke Fortran application. Codling, E., Kelly, C. "F-Press: A Stochastic Simulation Tool for Developing Fisheries Management Advice and Evaluating Management Strategies", Irish Fisheries Investigations No. 17, Marine Institute 2006. ISSN 0578 7476	
Type of stock	Medium life span, pelagic, high value	
Knowledge base	Analytic assessment, barely acceptable	
Type of regulation	TAC	
Model conditioning		
	Function, source of data	Stochastic? – how
Recruitment	Hockeystick, 1983-2004 (excl 1982)	Log-normal error with CV derived from residuals of Julios fit
Growth & maturity	Average over 2003-2006, no density dependence	No
Natural mortality	0.15	No
Selectivity	Selection from 2006 assessment (separable)	No
Initial stock numbers	Numbers at Jan 1 2004, from 2006 assessment	Log-Normal CV (25%) – assessment precision? No correlation between age groups
Decision basis	Slope of last 3 egg surveys egg abundance	
If assessment in the loop		
	Input data	NA NA
	Comparison with ordinary assessment?	NA

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 (distribution, on what)	"true population" given by propagating initial numbers with recruitment from S/R excluding high 1982YC at lowest SSB. Observed population represented by Egg survey model which relates "observed" egg abundance with true SSB including demographic factor and error
	Comparison with ordinary assessment?	NA, not trying to replicate assessment process
Projection		
	Through intermediate year?	No
	Iteration in TAC year?	No
Implementation	The "true" egg abundance (EGG_{true}) is obtained from SSB_{true} and modeled based on the relationship between egg abundance and SSB estimated from the SAD model, which was extended to account for a fecundity-weight relationship. To incorporate different components of variance into this relationship, the total variance can be apportioned into a "process" error component (λ_{egg}) linking EGG_{true} to SSB_{true} (which could result, in part, from variability in fecundity), and an "observation" error component (cv_{egg}) linking observed egg abundance (EGG_{obs}) to EGG_{true}: $$EGG_y^{true} = (qSSB_y^{true} + bSSBW_y^{true})e^{\lambda egg \eta y}$$ where λ_{egg} represents the process error component in log-terms, $\eta_y \sim N[0; 1]$, and $SSBW$ is calculated in the same way as SSB, but replacing w_a^{stock} with $(w_a^{stock})^2$, reflecting the dependence of fecundity on mean weight $(q + bw_a^{stock})$, with parameter b derived empirically and with q estimated from the extended SAD model (which accounts for this relationship) and $$EGG_y^{obs} = EGG_y^{true}e^{cvegg \omega y}$$ where cv_{egg} represents the sampling CV related to observed egg-abundance estimates and $\omega_y \sim N[0; 1]$	
Harvest rule		
Harvest rule design	Hybrid slope design. Fixed TAC and slope factor from past 3 egg surveys $$TAC_{y-y+2} = 1.07\left[\frac{TAC_{ref}}{2} + \frac{TAC_{y-3}sl}{2}\right]$$ **Computations to estimate the f(slope) parameter (*sl*)** 1) Divide the last three egg estimates from the triennial survey by 10^{15}; 2) Compute the slope (*b*) for years 1, 2 and 3;	

	3) If $b \leq -1.5 \Rightarrow sl = 0$ $-1.5 < b < 0 \Rightarrow sl = 1 - (1/-1.5 * b)$ $0 \leq b \leq 0.5 \Rightarrow sl = 1 + (0.4/0.5 * b)$ $b > 0.5 \Rightarrow sl = 1.4$ 
Stabilizers	None inertia in HCR sufficient
Duration	Intended triennial, but implemented annually
Revision clause	After 3 years or if any parameters outside the 95% CI of the simulations
Presentation of results	
Interest parameters	Risk 1, from 3 years out to 40 years Median, Cumulative and CV for Yield Number of TAC change/upward/downward events Median TAC adjustment
Risk type (see classification below) and time interval	Risk 1
Precautionary risk level	5% (Risk1) (applied to 1982 SSB -1.24Mt), i.e. Proxy for Blim=Bloss
Comments and experiences	
Review, ac-	Accepted by ACFM 2008 on a provisional basis, implemented from 2008-2011.

ceptance:	
Experiences and comments	The assessment has been improved since the simulations were done Recruitment has continued at a relatively low level and the SSB is declining from an elevated status as expected. Norway objected to the use of the plan as a basis for advice and so it was not used in 2012. The Commission have problems in implementation due to co-decision making. An ICES SG reviewed a request to comment on whether the parameters of the plan could be given as ranges rather than fixed, and said they could but the plan needed to be reviewed as it was past its review date and needed the inclusion of current best practice in the formulation of its HCR.

Annex 2 EXPLORATION OF DIFFERENT RISK DEFINITIONS IN MANAGEMENT PLAN EVALUATIONS

Carmen Fernández, January 2013

SECTION 1: ALTERNATIVE RISK DEFINITIONS

This note aims to explore the implications of different risk definitions in management plan evaluations. Here risk refers to the probability that SSB falls below some limit value (usually B_{lim}, or some appropriate substitute when B_{lim} is not available).

Risk is in general terms defined as the probability that SSB falls below B_{lim}, but this can be interpreted in different ways, of which three are presented here:

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

When performing MP evaluations (by simulation), normally one ends up with a set of iterations ($iter = 1, \ldots, niter$), where each iteration corresponds to a sequence of SSB values in a number of consecutive years ($y = 1, \ldots, ny$). Risk is calculated based on the simulated SSB values, as follows.

** **Risk1:** For each year y, compute the risk of SSB being below B_{lim} in that year, i.e.

Risk(year y) = (Number of times, across iterations, that SSB in year y is below B_{lim})/ $niter$,

and then average the annual risks across the ny years.

** **Risk2**:

(Number of iterations in which SSB is below B_{lim} at least once during the ny years)/ $niter$

= (Number of iterations in which the minimum SSB value over the ny years is below B_{lim})/ $niter$

** **Risk3**: Compute Risk(year y) for each year y, and take the maximum over the ny years.

** **Additional note on Risk1**: Note that averaging the annual risks is equivalent to computing:

(Number of times, across iterations and years, that SSB is below B_{lim})/($niter$ * ny)

= Average over iterations of { (Number of years in which SSB is below B_{lim})/ ny }

Therefore, Risk1 can alternatively be defined as:

{Number of years (out of ny years) in which SSB is expected to be below B_{lim}}/ ny

SECTION 2: COMPARISON OF ALTERNATIVE RISK DEFINITIONS

It follows directly that:

$$Risk2 = P\left(\min_{y=1,\ldots,ny} SSB_y < B_{lim}\right) \geq P\left(SSB_y < B_{lim}\right) \quad \text{for any one year } y,$$

where SSB_y denotes SSB in year y, i.e.

$$Risk2 \geq Risk(year\ y) \quad \text{for any one year } y$$

Therefore:

$$Risk2 \geq \max_{y=1,\ldots,ny} Risk(year\ y) \geq \frac{\sum_{y=1,\ldots,ny} Risk(year\ y)}{ny}$$

i.e.

$$\boxed{\boldsymbol{Risk2 \geq Risk3 \geq Risk1}}$$

HOW DIFFERENT ARE THE VALUES OF *Risk1*, *Risk2* AND *Risk3*?

- ***Stationary situation:***

First consider that the range of years over which risk is computed corresponds to a stationary situation. This is typically the case in MSE simulations if the first few years of the simulation are ignored, so that initial conditions have no effect on the risk values computed.

Under stationarity the (marginal) distribution of SSB is the same in all years. Therefore, $Risk(year\ y)$ is the same in all years, which immediately implies:

$\boldsymbol{Risk3 = Risk1}$ (and these risk definitions do not depend on the number of years $\boldsymbol{ny}$ or the degree of $\boldsymbol{SSB}$ time autocorrelation).

On the other hand, the value of $Risk2$ depends on the number of years considered, ny, and on the degree of SSB time autocorrelation, as it is now shown:

Recalling the definition of $Risk2$:

$$\begin{aligned} Risk2 &= P\left(\min_{y=1,\ldots,ny} SSB_y < B_{lim}\right) \\ &= 1 - P\left(SSB_y \geq B_{lim} \text{ in all years } y = 1, \ldots, ny\right) \end{aligned}$$

$$= 1 - \left\{ P(SSB_1 \geq B_{lim}) \prod_{y=2,\dots,ny} P\left(SSB_y \geq B_{lim} \middle| SSB_j \geq B_{lim} \text{ in all years } j \text{ before year } y\right) \right\}$$

In other words, the probability that SSB is above B_{lim} in all years $y = 1, \dots, ny$ is obtained by multiplying the probability that SSB is above B_{lim} in the first year by the probabilities that SSB is above B_{lim} in each of the following years ($y = 2, \dots, ny$) *conditional on* SSB also being above B_{lim} in all years before y. Two conclusions follow from this:

1. Since, by definition, any probability is always ≤ 1, the product of probabilities over years $y = 2, \dots, ny$ will be smaller (and, hence, *Risk2* larger) the larger the number of years ny.

2. Each of the $ny - 1$ probabilities that make up this product will be larger (and, hence, *Risk2* smaller) the higher the correlation between SSB values in subsequent years. If SSB is very highly autocorrelated, each of these probabilities will be close to 1, and $Risk2$ will be close to $Risk3$ (and to $Risk1$). $Risk2$ will be largest when the SSB values in subsequent years are nearly independent of each other.

In summary:

$Risk2$ is larger the bigger the number of years considered (ny) and the weaker the SSB time autocorrelation.

This implies that e.g. $Risk2$ measured over 20 years is always ≥ than if measured over 10 years ≥ than if measured over 5 years.

Note that SSB is generally expected to be autocorrelated in time. The strength of the autocorrelation will depend on the life-history characteristics (e.g. SSB for short-lived species will be less autocorrelated in time than SSB for long-lived species) and the harvest strategy, although it is expected that life-history characteristics dominate. This means that, all other things being equal, $Risk2$ may be expected to be higher for short-lived than for long-lived species.

- ***Non-stationary situation***

This situation happens if risk is calculated based on the first few years of the simulation, on which initial conditions have an impact. It could also happen if non-stationary scenarios are explored in the simulation (e.g. recruitment depending on environmental variables for which a non-stationary forecast exists). Under non-stationarity the (marginal) distribution of SSB varies from year to year. It is, therefore, questionable that a single number (as each of $Risk1$, $Risk2$ and $Risk3$ give) can provide a good summary of risk, since the situation varies between years. Under non-stationarity:

$*$ Annual risks $Risk(year\ y)$ should be examined for all years to gain a better understanding

$*$ $Risk1$ (risk averaged across years) and $Risk3$ (maximum risk across years) could be very different.

$*$ $Risk2$ and $Risk3$ both increase when the number of years considered (ny) increases.

$*$ $Risk2$ is larger the weaker the SSB time autocorrelation (same reasoning as in stationary case).

$*$ $Risk1$ and $Risk3$ depend only on the marginal distribution of SSB_y in each year y and are, therefore, not affected by the degree of time autocorrelation in SSB.

SECTION 3: EXAMPLES

This section presents two examples, one for a stationary situation and one for a non-stationary situation. For $ny = 10$ years, the 3 alternative risk definitions are explored under different values for the SSB time autocorrelation. In both examples SSB is simulated for years $y = 1, \ldots, ny$, as:

$$\ln(SSB_y) = \ln(\mu_y) + \varepsilon_y,$$

where ε_y has a Normal distribution with mean=0, standard deviation=0.4 and correlation ρ between consecutive years (AR(1) process). These are assumed to be the SSB population values that result after a harvest control rule has been applied to the population.

Five different values of ρ are considered, ranging from ρ=0 (independence) to ρ=0.9999 (almost perfect correlation). In the first example (stationary situation), the value μ_y is the same in all years. In the second example (non-stationary situation), the value of μ_y is different in different years. The number of iterations used is $niter =$ 100 000 in both cases.

- ***Example 1: Stationary situation***

In this example $\mu_y = 50\,000$ in all years, and $B_{lim} = 24\,000$.

Figure 3.1.1 shows 50 iterations (of the 100 000 conducted) of SSB. Each panel corresponds to a different value of ρ. B_{lim} is marked by a dashed horizontal line. As expected, higher values of ρ correspond to smoother SSB time trajectories.

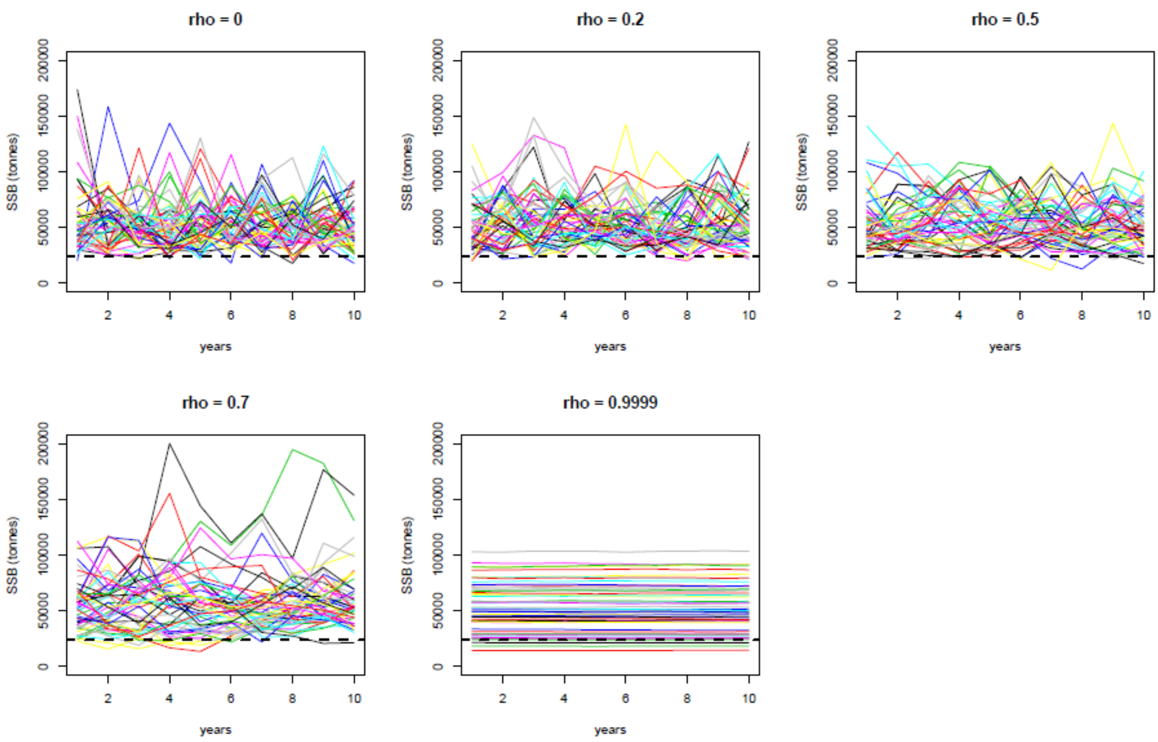


Figure 3.1.1. 50 iterations of SSB, for different values of ρ. Dashed horizontal line is B_{lim}.

Figure 3.1.2 shows the 5, 50 and 95 percentiles of the distribution of SSB each year. As in Figure 3.1.1, each panel in this figure corresponds to a different value of ρ and B_{lim} is marked by a dashed horizontal line. Because this example corresponds to the stationary situation, the marginal distribution of SSB is the same every year, which explains why the percentiles are also the same in all years.

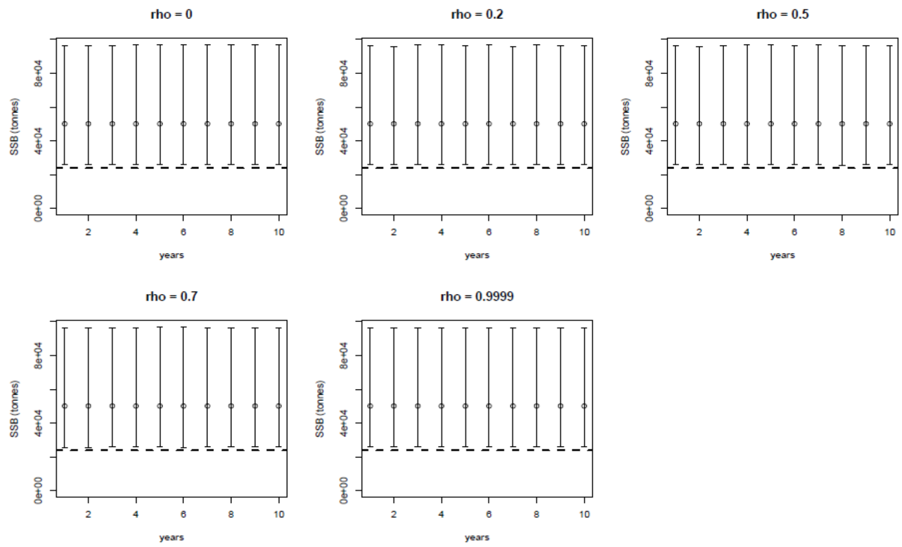


Figure 3.1.2. 5, 50 and 95 percentiles of the distribution of SSB each year, for different values of ρ. Dashed horizontal line is B_{lim}.

From figure 3.1.2, one may well conclude that the management plan (to which these SSB simulations would correspond) is precautionary, given that the probability that SSB is below B_{lim} is less than 5% in all years. Whereas this would be the conclusion if $Risk1$ or $Risk3$ definitions were used, a different conclusion would be reached if $Risk2$ definition was used, as Table 3.1.1 shows.

	ρ=0	ρ=0.2	ρ=0.5	ρ=0.7	ρ=0.9999
$\boldsymbol{Risk1}$ (average)	0.03	0.03	0.03	0.03	0.03
$\boldsymbol{Risk3}$ (maximum)	0.03	0.03	0.03	0.03	0.03
$\boldsymbol{Risk2}$	0.28	0.28	0.24	0.19	0.04
$Risk(year\ 1)$	0.03	0.03	0.03	0.03	0.03
$Risk(year\ 2)$	0.03	0.03	0.03	0.03	0.03
$Risk(year\ 3)$	0.03	0.03	0.03	0.03	0.03
$Risk(year\ 4)$	0.03	0.03	0.03	0.03	0.03
$Risk(year\ 5)$	0.03	0.03	0.03	0.03	0.03
$Risk(year\ 6)$	0.03	0.03	0.03	0.03	0.03
$Risk(year\ 7)$	0.03	0.03	0.03	0.03	0.03
$Risk(year\ 8)$	0.03	0.03	0.03	0.03	0.03
$Risk(year\ 9)$	0.03	0.03	0.03	0.03	0.03
$Risk(year\ 10)$	0.03	0.03	0.03	0.03	0.03

Table 3.1.1. Risk of $SSB < B_{lim}$ for each of the risk definitions given in Section 1. Each column corresponds to a value of ρ. Risk values are marked red if above 5% and green if below 5%.

- ***Example 2: non-stationary situation***

In this example, μ_y has different values in different years (averaging 50 000 over the 10 years), and $B_{lim} = 19\ 000$. The non-stationarity is clear from Figures 3.2.1 and 3.2.2.

Figure 3.2.1 shows 50 iterations (of the 100 000 conducted) of SSB. Each panel corresponds to a different value of ρ. B_{lim} is marked by a dashed horizontal line.

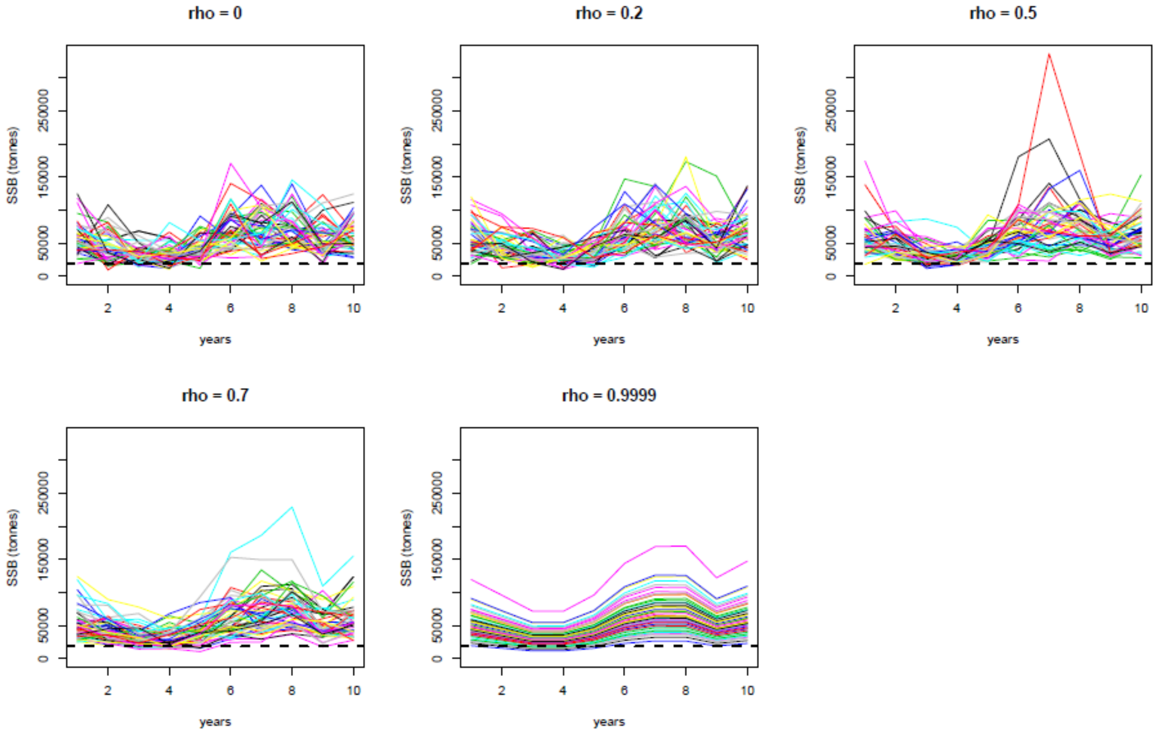


Figure 3.2.1. 50 iterations of SSB, for different values of ρ. Dashed horizontal line is B_{lim}.

Figure 3.2.2 shows the 5, 50 and 95 percentiles of the distribution of SSB each year. As in Figure 3.2.1, each panel in this figure corresponds to a different value of ρ. B_{lim} is marked by a dashed horizontal line.

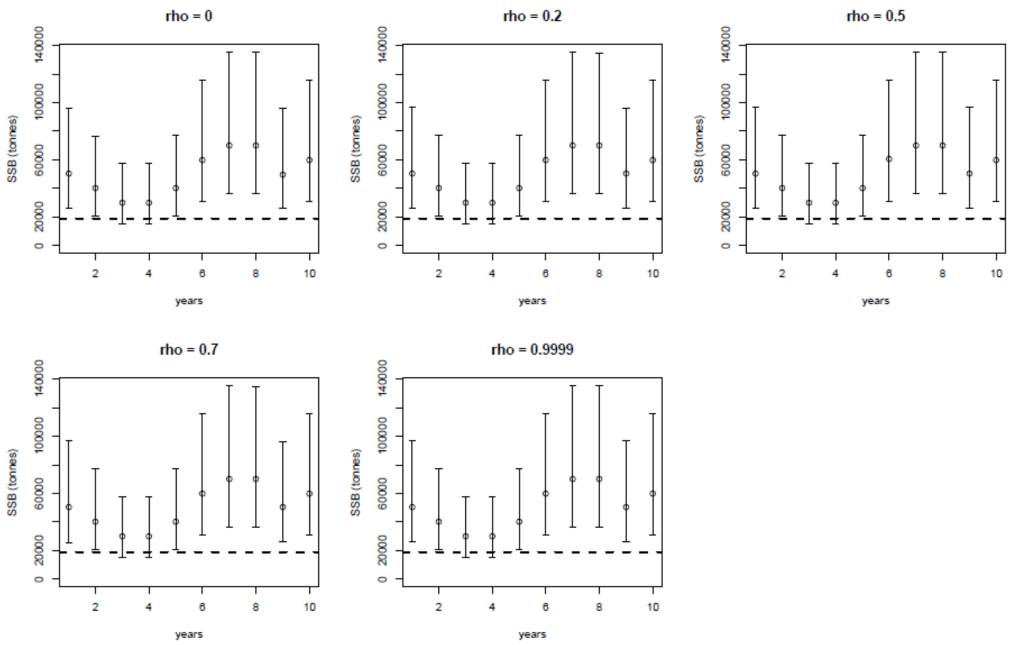


Figure 3.2.2. 5, 50 and 95 percentiles of the distribution of SSB each year, for different values of ρ. Dashed horizontal line is B_{lim}.

It is not clear from figure 3.2.2 whether the management plan (to which these SSB simulations would correspond) would be considered precautionary or not, given that the probability that SSB is below B_{lim} is less than 5% in most years but not in all years. Table 3.2.1 gives the values of the alternative definitions of risk.

	ρ=0	**ρ=0.2**	**ρ=0.5**	**ρ=0.7**	**ρ=0.9999**
$\boldsymbol{Risk1}$ (average)	0.03	0.03	0.03	0.03	0.03
$\boldsymbol{Risk3}$ (maximum)	0.13	0.13	0.13	0.13	0.13
$\boldsymbol{Risk2}$	0.30	0.28	0.25	0.22	0.13
$Risk(year\ 1)$	0.01	0.01	0.01	0.01	0.01
$Risk(year\ 2)$	0.03	0.03	0.03	0.03	0.03
$Risk(year\ 3)$	0.13	0.13	0.13	0.13	0.13
$Risk(year\ 4)$	0.13	0.13	0.13	0.13	0.13
$Risk(year\ 5)$	0.03	0.03	0.03	0.03	0.03
$Risk(year\ 6)$	0.00	0.00	0.00	0.00	0.00
$Risk(year\ 7)$	0.00	0.00	0.00	0.00	0.00
$Risk(year\ 8)$	0.00	0.00	0.00	0.00	0.00
$Risk(year\ 9)$	0.01	0.01	0.01	0.01	0.01
$Risk(year\ 10)$	0.00	0.00	0.00	0.00	0.00

Table 3.2.1. Risk of $SSB < B_{lim}$ for each of the risk definitions given in Section 1. Each column corresponds to a value of ρ. Risk values marked red if above 5% and green if below 5%.

SECTION 4: MORE COMPARISONS BETWEEN ALTERNATIVE RISK DEFINITIONS

This section presents a few additional comparisons between alternative risk definitions, with the purpose of gaining further understanding. As in the previous section, in the two examples considered here SSB is simulated for years $y = 1, \ldots, ny$, as:

$$\ln(SSB_y) = \ln(\mu_y) + \varepsilon_y,$$

where $\mu_y = 50\,000$ and ε_y has a Normal distribution with mean=0, standard deviation=0.4 and correlation ρ between consecutive years (AR(1) process). Only the stationary situation is considered and, therefore, all the annual risks are identical and $Risk(year\ y) = Risk1 = Risk3$.

Third example:

B_{lim} $(= 25\,895)$ was chosen such that $Risk1 = 0.05$. $Risk2$ was then calculated for a range of values of ρ and ny.

The risks are displayed in Table 4.1, which shows how $Risk2$ increases as ρ becomes smaller and ny larger.

The table also shows (for the case $ny = 10$) how the distribution of the number of years that $SSB < B_{lim}$ changes with ρ. As ρ increases, $Risk2$ decreases substantially (i.e. lower probability of SSB going below B_{lim} at all), but if SSB goes below B_{lim} then the probability that the stock is for more than 1 year below B_{lim} increases.

Table 4.1

	ρ=0	**ρ=0.2**	**ρ=0.5**	**ρ=0.7**	**ρ=0.9**	**ρ=0.9999**
Risk1 & Risk3	0.05	0.05	0.05	0.05	0.05	0.05
Risk2 (ny=5)	0.23	0.22	0.19	0.16	0.11	0.05
Risk2 (ny=10)	0.4	0.38	0.33	0.27	0.16	0.05
Risk2 (ny=20)	0.64	0.62	0.54	0.44	0.26	0.05
The following rows correspond to ny=10:						
P($SSB \geq B_{lim}$ in all years)	0.6	0.62	0.67	0.73	0.84	0.95
P($SSB < B_{lim}$ in 1 year)	0.31	0.29	0.21	0.14	0.06	0
P($SSB < B_{lim}$ in 2 years)	0.07	0.08	0.08	0.06	0.03	0
P($SSB < B_{lim}$ in 3 years)	0.01	0.01	0.03	0.03	0.02	0
P($SSB < B_{lim}$ in 4 years)	0	0	0.01	0.02	0.02	0
P($SSB < B_{lim}$ in 5 years)	0	0	0	0.01	0.01	0
P($SSB < B_{lim}$ in 6 years)	0	0	0	0	0.01	0
P($SSB < B_{lim}$ in 7 years)	0	0	0	0	0.01	0
P($SSB < B_{lim}$ in 8 years)	0	0	0	0	0	0
P($SSB < B_{lim}$ in 9 years)	0	0	0	0	0	0
P($SSB < B_{lim}$ in 10 years)	0	0	0	0	0	0.05

Fourth example:

B_{lim} was chosen such that $Risk2 = 0.05$. $Risk1$ was then calculated for a range of values of ρ and ny.

Note: to have the same value of $Risk2$ over the range of ρ and ny values required choosing a different B_{lim} for each (ρ, ny) combination.

The risks are displayed in Table 4.2, which shows that in order to achieve $Risk2 = 0.05$, $Risk1$ must be substantially smaller than 0.05, and the difference between both risks is more pronounced as ρ becomes smaller and ny larger.

The table also shows (for the case $ny = 10$) how the distribution of the number of years that $SSB < B_{lim}$ changes with ρ. Comparing this distribution with the one presented in Table 4.1, the distributions are virtually identical for ρ=0.9999, but they differ significantly for smaller values of ρ.

Table 4.2

	ρ=0	**ρ=0.2**	**ρ=0.5**	**ρ=0.7**	**ρ=0.9**	**ρ=0.9999**
Risk2	0.05	0.05	0.05	0.05	0.05	0.05
Risk1 & Risk3 (ny=5)	0.01	0.01	0.012	0.014	0.021	0.048
Risk1 & Risk3 (ny=10)	0.005	0.005	0.006	0.007	0.012	0.047
Risk1 & Risk3 (ny=20)	0.003	0.003	0.003	0.003	0.006	0.046
The following rows correspond to ny=10:						
P($SSB \geq B_{lim}$ in all years)	0.95	0.95	0.95	0.95	0.95	0.95
P($SSB < B_{lim}$ in 1 year)	0.05	0.05	0.04	0.04	0.02	0
P($SSB < B_{lim}$ in 2 years)	0	0	0.01	0.01	0.01	0
P($SSB < B_{lim}$ in 3 years)	0	0	0	0	0.01	0
P($SSB < B_{lim}$ in 4 years)	0	0	0	0	0	0
P($SSB < B_{lim}$ in 5 years)	0	0	0	0	0	0
P($SSB < B_{lim}$ in 6 years)	0	0	0	0	0	0
P($SSB < B_{lim}$ in 7 years)	0	0	0	0	0	0
P($SSB < B_{lim}$ in 8 years)	0	0	0	0	0	0
P($SSB < B_{lim}$ in 9 years)	0	0	0	0	0	0
P($SSB < B_{lim}$ in 10 years)	0	0	0	0	0	0.05

SECTION 5:

NUMBER OF ITERATIONS REQUIRED FOR COMPUTING RISKS PRECISELY

When evaluating harvest control rules in management plans, normally a stochastic simulation (MSE) for a period of ny future years is performed. This stochastic simulation is based on a number $niter$ of independent iterations (sometimes also called replications, realisations, etc). Population, catch, risk statistics, and any other quantity of potential interest, are used to summarise performance of the MP over the ny year period. These statistics (including risk) are calculated based on the $niter$ independent iterations. Depending on how the simulation is set up (e.g. how assessment errors are dealt with or how it is programmed), carrying out a large number of iterations can be very time consuming. Sometimes in the past, as few as $niter = 50$ iterations have been used.

If r was the value of risk obtained if an infinite amount of iterations could be performed, the value of risk computed on the basis of $niter$ independent iterations has a distribution centred at the value r (except for Risk3, as explained later), with standard deviation $\{r * (1 - r)/niter\}^{1/2}$. Therefore, the risk calculated on the basis of $niter$ iterations will be within the interval $r \pm 1.96 * \{r * (1 - r)/niter\}^{1/2}$ in approximately 95% of the cases. This allows an approximate calculation of the number of iterations, $niter$, required to compute r with a certain precision. For $r = 0.05$, the following table gives the intervals that result for different number of iterations:

Table 5.1
Distribution of risks computed based on $niter$ iterations when $r = 0.05$
(r is the value of risk obtained if an infinite amount of iterations could be performed)

$niter$	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

Table 5.1 implies that if the value of risk (computed based on an infinite number of iterations) was 0.05, then performing a simulation with $niter$ iterations and computing risk based on that simulation will produce 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 risk < 0.03, we can be quite certain that $r < 0.05$, whereas if it gives a value of risk > 0.07, we can give quite certain that $r > 0.05$. However, if it gives a value of risk between 0.03 and 0.07, it is unclear

whether r is above or below 0.05. In that case, further precision can be obtained by increasing the number of iterations.

The intervals in Table 5.1 are directly applicable to annual risks (for each individual year, considered separately from the other years) and $Risk2$. It can also be used as "safe" guidance for $Risk1$ computation.

In the stationary situation, the intervals for $Risk1$ are narrower than those given in Table 5.1, because an average is taken over several years, which increases precision (although the gain in precision is less the more autocorrelated SSB is). A simple simulation exercise showed that, for $Risk1$, the interval in Table 5.1 reduces to [0.04, 0.06] already for $niter = 250$, when $Risk1$ is computed as a 10-year average (stationary situation), even under high autocorrelation in SSB (such as ρ=0.8).

On the other hand, one has to be careful with the computation of $Risk3$, because, as it is a maximum of annual risks, it will amplify the noise in the computed annual risks. My understanding/intuition is that $Risk3$ computed based on $niter$ iterations is a *biased* estimator of the value that would be obtained if an infinite number of iterations could be performed (more often than not the computed value of $Risk3$ will be too large). The bias is stronger the bigger the number of years ny considered, the more similar the annual risks in different years, and the less time autocorrelation in SSB. In the stationary situation, given that $Risk3$ is equal to $Risk1$, only $Risk1$ should be computed (because of the much better properties of the algorithm to compute $Risk1$).

In the short term, where the situation is non-stationary, it makes sense to consider annual risks for each of the years. When each year is seen in isolation, the intervals in Table 5.1 apply. However, when looking at the ensemble of ny years and focusing on the worst year (i.e. $Risk3$) the situation is different: imagine that risk (based on an infinite amount of iterations) is < 0.05 in all years, and that the annual risks are computed on the basis of $niter$ iterations. When a particular year y is considered, $niter$ will lead to some probability p_y that the computed risk is bigger than 0.05, leading to the wrong conclusion for year y. On the other hand, when the ensemble of ny years is considered, $niter$ will lead to some probability p that the computed risk is bigger than 0.05 in at least one year (hence, leading to the wrong conclusion when focusing on the worst year), where p is at least as big as the biggest p_y (and in some cases it can be considerably bigger). In other words, for the same number of $niter$ it is more likely that a wrong negative conclusion ("false negative") is reached when looking at the ensemble of ny years than when looking at a particular year y determined in advance. Hence, when looking at annual risks for ny years in the short term, if the computed risk in any year is > 0.05, it is worth exploring whether the computation of $Risk3$ has stabilised. If it has not, then it is worth increasing the number of iterations until stability is reached.

Conclusions:

- For $Risk2$, $Risk1$ and $Risk(year\ y)$ in a single year y, the intervals in Table 5.1 can serve as guidance.

- In the stationary situation, $Risk1$ requires fewer iterations than suggested in Table 5.1 (taking advantage of averaging over years, but the gain in precision is less the more autocorrelated SSB is).
- Computing $Risk3$ requires more iterations than suggested in Table 5.1 (potentially many more, as the computed value can converge very slowly, in my limited experience) and the same holds for computing $Risk(year\ y)$ for ny years and focusing on the highest annual risk. In the stationary situation, $Risk1$ should be computed instead of $Risk3$ (even if the result is reported as $Risk3$). In the non-stationary situation (i.e. short term), the following "solution" could be adopted for $Risk3$:
  5. Compute $Risk3$ based on the number of iterations in Table 5.1
  6. If the computed $Risk3$ value is below the lower end of the interval in Table 5.1, then it may be concluded that $Risk3 < 0.05$ (given the positive bias in the $Risk3$ computation).
  7. If the computed $Risk3$ value is above the upper end of the interval in Table 5.1, then compute $Risk1$. If the computed $Risk1$ value is above the upper end of the interval in Table 5.1, then it may be concluded that $Risk1 > 0.05$ (and the same will, therefore, hold for $Risk3$). If the computed $Risk1$ is below the upper end of the interval in Table 5.1, then no conclusion can be reached regarding $Risk3$ (in this case, the number of iterations should be increased until the value of Risk3 stabilizes in an area where conclusions can be drawn).
- It is recommended that the relevant measure of risk used in the analysis be plotted against iteration number as follows: compute the relevant risk measure based on the first $iter$ iterations and plot it versus $iter$ (iteration number), to get an understanding of how long it takes for it to stabilise in an area where conclusions can confidently be drawn.

SECTION 6: SUMMARY OF ICES PRACTICES

This section tries to compile the management plans evaluated by ICES in recent years and the risk definition used in each case. I may have missed some and got a few wrong. It is not always clear from the ICES advice which risk definition has been used, because the risk definition is often given just in words (which can be interpreted to mean different things, unless it has been expressed very precisely), although this can be usually figured out from the technical reports.

Risks have generally been measured in relation to B_{lim} or some appropriate proxy for it when B_{lim} has not been defined for a stock. The list below does not indicate whether B_{lim} or some proxy for it has been used.

Stock and year advice issued	**Risk definition used**	$\boldsymbol{niter}$
Rockall haddock (Aug 2012)	$\boldsymbol{Risk1}$, after first 10 years in simulation removed	100
West. horse mackerel	$\boldsymbol{Risk1}$, averaging over 40 years from start of simulation year (2007?)	1000

(Apr 2012)		
North Sea saithe (Nov 2012)	**Annual risks** $Risk(year\ y)$ for the short term and $\boldsymbol{Risk1}$ for 2020-2029 (an evaluation conducted in 2008 used annual risks)	250
North Sea herring (Nov 2012)	$\boldsymbol{Risk2}$, based on $ny = 10$ years from present (also used $Risk2$ based on $ny = 10$ years in a 2011 evaluation)	1000 (100 in 2011 eval)
North Sea plaice and sole (Oct 2012)	$\boldsymbol{Risk2}$ based on 3 periods: 2013-2015, 2015-2020, 2016-2025 Says that since $Risk2$<5%, the same holds for $Risk1$ and $Risk3$ (also used $Risk2$ in a 2010 evaluation)	200
Celtic Sea herring (Nov 2012)	$\boldsymbol{Risk3}$ based on next $ny = 20$ years	1000
Norway pout in IV and IIIa (Oct 2012)	**Annual risks** $Risk(year\ y)$ for the short term (2013 to 2016) and $\boldsymbol{Risk1}$ for 2017-2026	1000
North Sea cod (Jul 2011)	$\boldsymbol{Risk(year\ 2015)}$ MP evaluated by ICES in March 2009 to be in conformity with PA, but I have not been able to find this advice	1000 (250 in 2008 eval)
North Sea whiting (Jul 2011)	ICES advice shows risk values (probability of SSB < Bloss), but it does not say how risk is defined. According to ICES/STECF report: $\boldsymbol{Risk1}$, based on the next 70 years $Risk(year$ 2015) is also presented	200
Haddock in VI (2010)	Various explanations throughout the advice document indicate that, essentially: $\boldsymbol{Risk1}$, based on the next 22 years and **annual risks** $Risk(year\ y)$ in order to conclude that risk is higher in the first few simulation years and lower later on. It also presents $Risk(year$ 2015).	50
HCR for mixed fishery of hake, anglerfish and *Nephrops* in VIIIc and IXa (2010)	Aim was to develop HCRs that could reach F_{MSY} by 2015 for all stocks in the mixed fishery. Biomass trajectories presented but **risks to biomass not computed**.	
North Sea haddock (2010)	Same as haddock in VIa	100
Icelandic cod (2010)	Request stated the objective of plan as $P(SSB_{2015} > SSB_{2009}) \geq 0.95$, and this was used for the ICES evaluation. Additionally, $\boldsymbol{Risk(year\ 2015)}$ was used for PA evaluation.	2000
Norwegian coastal cod (2010)	ICES advice says: "ICES considers the proposed rule to be provisionally consistent with the Precautionary Approach. The basis of this evaluation is the precautionary approach". Some details in Annex 10 of AFWG 2010 report: it examines **annual risks** $Risk(year\ y)$ for the next 20 years and shows decreasing annual risk which becomes < 10% around 2026.	1000

1.