ICES WGMG Report 2006

ICES Resource management Committee
ICES CM 2006/RMC:07
Ref. ACFM

Report of the Working Group on Methods of Fish Stock Assessments (WGMG)

21-26 June 2006

Galway, Ireland

International Council for the Exploration of the Sea
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. 2006. Report of the Working Group on Methods of Fish Stock Assessments (WGMG), 21-26 June 2006, Galway, Ireland. ICES CM 2006/RMC:07. 83 pp.

Contents

Executive summary **1**

1 Introduction **2**

2 Structure of the report **2**

3 Sensitivities of stock assessment methods to known data problems **3**

3.1 Assessment of stock dynamics when catch data are not considered to reflect the real level of catches 3

3.1.1 Introduction 3
3.1.2 Previous studies 3
3.1.3 On-going studies into approaches that provide model estimates of removals 5
3.1.4 On-going studies into fisheries-independent methods 12
3.1.5 Summary 13

3.2 Estimating discards 14

3.2.1 Introduction 14
3.2.2 Estimating the distribution of discards conditional on landings and other factors 14
3.2.3 Derivation of *Nephrops* discards for non-sampled years 16
3.2.4 Summary 18

3.3 Sensitivities of catch-at-age stock assessment methods 19

3.3.1 Introduction 19
3.3.2 Trends in natural mortality 20
3.3.3 Strong year-class effects in separable models 21
3.3.4 Sensitivity of assessment results to model specification 34
3.3.5 Overview of the impact of data on assessment error 34
3.3.6 Summary 36

4 Reference points and harvest control rules **37**

4.1 Introduction – background and context 37

4.2 Management strategy evaluation framework 39

4.3 Model and software validation, verification and testing 43

4.3.1 Software testing 43
4.3.2 Model verification 43
4.3.3 Model validation 44
4.3.4 Documentation 44

4.4 Case studies 45

4.4.1 North Sea haddock 45
4.4.2 Northern hake 56
4.4.3 Irish Sea cod 61

4.5 Discussion 64

5 Working documents and background material presented to the working group **65**

5.1 Working papers and documents (W) 65

5.2 Background material (B) 65

6 Nominations for Chair **66**

7 References **66**

Annex 1: List of participants 69

Annex 2: WGMG Terms of Reference for the next meeting 71

Annex 3: Some verification and testing tools available in R 73

Annex 4: Recommendations 79

Executive summary

The ICES' Working Group on Methods of Fish Stock Assessment [WGMG] was established to provide a forum for the development of new methods and the investigation of the properties of fish stock assessment methods. Sound assessment methods are a basic requirement for the provision of sound and credible advice, and the group is regularly tasked by the ICES' Advisory Committee on Fishery Management [ACFM] to address specific problems identified by ICES' stock assessment working groups.

This report is the work of a meeting held in Galway, Ireland from 21–26 June 2006 at the Marine Institute under the chairmanship of Carl O'Brien (UK).

The meeting was originally scheduled to end on the 28 June 2006 but was shortened, after consultation with the ICES Secretariat and the Chair of RMC, to allow participants to attend the ICES Symposium on Fisheries Management Strategies, 27–30 June 2006 held in Galway, Ireland. In future, the group suggests that meetings of WGMG should extend for a period of 10-15 days, if possible, in order to ensure adequate time for the development of methods and collaborative working. The shortening of the meeting by two days necessitated changes in the programme of work of the WGMG but the group addressed its original terms of reference (ToRs) by reviewing on-going work within national institutes. In addition to its original ToRs, WGMG was requested by ACFM at its spring 2006 meeting to consider how best to update limit reference points and develop target reference points for use within long-term management strategies. This additional ToR necessitates further work but the group discussed and agreed an approach to the evaluation of long-term management strategies. Three illustrative examples are presented to demonstrate the proposed approach for single species applications.

At its meeting in 2004, WGMG proposed a road map for deriving harvest control rules (HCRs) to aid in the long-term provision of advice and at this meeting in 2006, WGMG further developed the approach using three illustrative case studies – North Sea haddock, northern hake and Irish Sea cod. As a priority, ICES should develop illustrative HCRs for a limited number of priority stocks using the framework proposed in this report as a prerequisite to the development of generic strategies for mixed fisheries management.

The report is organized in Sections. Section 2 describes the overall structure of the report and the ToRs are then addressed within the two main Sections of the report – Section 3 addresses aspects of the sensitivities of stock assessment methods to known data problems and Section 4 addresses issues related to reference points and HCRs. The various working documents and background material presented at the meeting are listed in Section 5; together with their assigned code for ease of reference.

Finally, the group proposes that it should meet in 2007 to continue its work.

1 Introduction

The **Working Group on Methods of Fish Stock Assessments** [WGMG] (Chair: C. O'Brien, UK) met in Galway, Ireland from 21–26 June 2006 to:

a) investigate the utility of assessment models that can provide management advice when catch-at-age data based upon commercial landings are thought not to reflect the real level of catches within a fishery;

b) review approaches for the historical re-construction of discards-at-age data and their incorporation into routine stock assessments; together with the development of estimation methods for the incorporation of information from self-sampling discard schemes;

c) investigate and test the sensitivities of catch-at-age stock assessment methods to known data problems with particular reference to the retrospective problem; and

d) evaluate, test and review developments in computer software for routine application in stock assessment that are presented to ICES.

In addition to these terms of reference (ToRs), WGMG was requested by ACFM (Advisory Committee on Fishery Management) at its spring 2006 meeting to consider an addition ToR; namely, to:

e) consider how best to update limit reference points and develop target reference points for use within long-term management strategies.

WGMG will report for the attention of the Resource Management Committee (RMC) and ACFM.

The meeting was originally scheduled to end on the 28 June 2006 but was shortened, after consultation with the ICES Secretariat and the Chair of RMC, to allow participants to attend the ICES Symposium on Fisheries Management Strategies, 27–30 June 2006 held in Galway, Ireland. In future, the group suggests that meetings of WGMG should extend for a period of 10-15 days, if possible, in order to ensure adequate time for the development of methods and collaborative working.

The shortening of the meeting by two days necessitated changes in the programme of work of the WGMG but the group addressed its original ToRs a)-d) by reviewing on-going work within national institutes. The additional ToR e) necessitates further work but the group discussed and agreed an approach to the evaluation of long-term management strategies. Three illustrative examples are presented to demonstrate the proposed approach for single species applications.

2 Structure of the report

WGMG last met in 2004 and in recent years has made a number of suggestions and recommendations on issues of data quality, modelling and stock assessment practice throughout its reports. The last meetings of the group had focused on the issue of the retrospective problem in stock assessments but it could be anticipated, in advance of the meetings, that the problems of ICES' assessments would not be fixed at short notice.

With the increasing level of mis-reporting and the need to provide quantitative stock assessments and forecasts, the development of new modelling approaches is urgently required but is on-going within national institutes. A number of working documents and background material were presented to the group, discussed and reviewed.

Approaches for the correction of bias in stock assessments need to be further developed and implemented if reliable forecasts are to be produced in the future. This has a bearing on the future development and evaluation of management strategies and harvest control rules – see the additional ToR e).

ICES' stock assessment working groups have developed new software tools over recent years and there is now a need to ensure that these tools provide appropriate numerical estimates.

The ToRs a)-e) are addressed within the two main sections of the report. Specifically, the ToRs a)-c) are individually addressed within Section 3 and the ToRs d)-e) are collectively addressed in Section 4. The various working documents and background material presented at the meeting are listed in Section 5; together with their assigned code for ease of reference within the various sections of the report.

3 Sensitivities of stock assessment methods to known data problems

Current stock assessment advice is, primarily, based on models that require unbiased information on the removals from the stock. This assumption is becoming less tenable. Terms of reference a)-c) address similar themes with respect to the provision of credible assessment advice. Section 3.1 (ToR a) reviews the utility of advice from methods that estimate removals from the stock and *catch free* methods developed to provide assessments in the absence of reliable catch statistics. Section 3.2 (ToR b) reviews the problems associated with discarding, and highlights some of the analyses that are being undertaken to address the issues. Section 3.3 (ToR c) continues the work in progress of this ICES' WGMG with respect to sensitivities in catch-at-age based assessment models.

3.1 Assessment of stock dynamics when catch data are not considered to reflect the real level of catches

ToR a) investigate the utility of assessment models that can provide management advice when catch-at-age data based upon commercial landings are thought not to reflect the real level of catches within a fishery.

3.1.1 Introduction

Two approaches have been taken to the provision of management advice when removals' data are considered uncertain or are not available. The methods can be loosely grouped into models that could potentially be used to provide estimates of removals (Section 3.1.3) and methods such as the fisheries independent methods that estimate relative stock abundance and total mortality based on auxiliary information (Section 3.1.4). WGMG reviewed methods reported in the literature that fall within the two categories and also work in progress that could be used to formulate management advice or to provide verification of advice derived from assessments conditional on the assumption of unbiased catch data.

3.1.2 Previous studies

De Oliveira (WA1) reviewed some of the recent literature describing models that are relevant to this area of research. The most pertinent are summarized below in this section.

Apostolaki *et al*. (2006) compare results from deterministic and probabilistic modelling methods when evaluating the effect of applying alternative management measures (e.g. spatial and temporal closures, and size-specific regulations). In their analysis of the sandbar shark commercial and recreational fisheries off the U.S. east coast, historic catch data were missing, which required an additional parameter to be introduced for each fishery (with corresponding prior distributions in the case of the probabilistic method).

Payne *et al*. (2005) present age-structured production model (ASPM) assessments of Patagonian toothfish based on fits to long-line GLM-standardised CPUE and catch-at-length data. A poor fit to the CPUE data was obtained for two years, which the authors hypothesised could have been due to unreported catches resulting from increased IUU (illegal, unreported and unregulated) fishing activity, or alternatively to changes in catchability and natural mortality. The fit was improved when the model was allowed to estimate levels of extra catch for these years, adding an additional two parameters to the model.

Plagányi and Butterworth (BA1) (see also Plagányi 2004) use an age-structured production model (ASPM) with a spatial dimension to estimate levels of poaching and ecosystem change for South African abalone. Data are available for several zones and include commercial and recreational catch-at-age (derived from length-distributions by cohort slicing), as well as a GLM-standardised commercial CPUE series, a fishery independent abalone survey series, and a single joint scientific institute-industry survey. In order to estimate levels of poaching, a novel fisheries index, confiscations per unit of policing effort (CPUPE), is used. The approach estimates the maximum poaching level per zone for the year for which the CPUPE index in that zone is highest, thus introducing one additional parameter per zone. Poaching estimates for the remaining years are then calculated using the CPUPE trend. Poaching estimates are compared to actual abalone confiscations (used in deriving the CPUPE series): where they imply a confiscation percentage success rate of more than some pre-determined level (25% is used), they are increased to ensure the implied confiscation success rate equals this level.

Pitcher *et al*. (2002) provide a review of methodologies used to quantify unreported catch. They also develop their own estimation method that synthesises all available specialist studies and information on unreported catch (supported by reports or explicitly attributed to a variety of published and unpublished sources) together with their uncertainties. Their estimation method relies on adjustment factors (from observer reports, correspondents and published information) that track changes in a regulatory regime, thus reflecting incentives and disincentives to misreport. Uncertainty is included through Monte Carlo simulations using multiple sources of information to provide upper and lower estimates. The authors claim that it is still possible to estimate misreporting, even when direct data are lacking, and that their method encourages transparency because sources of information are presented, so that uncertain values are easily identified and can be discussed and further refined.

Porch *et al*. (2006) devise a model-based framework for estimating reference points, stock status and recovery times that does not require catch data and other measures of absolute abundance. It is essentially an age-structured production model (ASPM; Hilborn 1990, Restrepo and Legault 1998) recast in terms relative to pre-exploitation levels. It is developed within a Bayesian framework to allow incorporation of auxiliary information, such as from meta-analyses of similar stocks or anecdotal information. Population dynamics commence from virgin conditions, and fishing mortality and relative recruitment are modelled as first order lognormal autoregressive processes with deviates estimated where sufficient data are available. The estimation of fishing mortality requires an index of fishing effort. The approach has been used to provide stock assessment advice for the data-poor goliath grouper stock off southern Florida, relying on data from the fishery and surveys for selectivity parameters, and on three indices of abundance and anecdotal information concerning population depletion levels.

Gavaris and van Eeckhaute (1998) compare the results from an ADAPT VPA model (Gavaris 1988) using catch-at-age data to those from a similar model, modified to use proportions instead of numbers caught at age. Both models use indices of abundance for calibration of the VPA, but the latter estimates total numbers caught each year in addition to the other model parameters (terminal year abundance and index catchability). Because proportions caught at age and abundance indices do not contain any information on the absolute magnitude of the population, the modified model requires at least one of the parameters to be fixed in order to be determinate. The authors use the modified model to diagnose possible systematic errors in the catch data. Although they do not recommend this model as a self-standing stock assessment method, they present it as a useful approach to subject data to different assumptions, thus evaluating the robustness of results.

Hammond and Trenkel (2005) presented an approach in which the landings were considered to give a lower bound on true catches. They considered a Schaefer surplus production model, and modelled the landings as censored observations of catches. A censored observation is an interval that contains the quantity of interest; however, only the interval is observed and not the quantity of interest. The authors used Bayesian methods for estimation and statistical inference with two sets of priors, using the BUGS software. They assumed that unobserved or misreported catches were bounded by the observed landings, and that twice the observed landings provided an upper bound on catches. That is, if C denotes the true (but unobserved) catches and L denotes the reported landings then the authors assumed that C lies within [L, 2L]. They examined the sensitivity of their method to the assumed upper bound, although in a limited way. They also recognized that catches may be over-reported in some situations, in which case the landings would provide an upper bound on true catches. They modelled the catches as proportional to biomass and effort, $C = q_c \times E \times B \times \xi$, where q_c was a catchability parameter, E was the effort, B was the surplus production model biomass, and ξ was an error term. The authors compared their censored approach with one based on assuming no misreporting (C = L) and another approach based only on survey and effort data (i.e. no landings data). They found that both of these methods performed worse on average than the censored approach, as measured by the Bayes risk associated with estimates of maximum sustainable yield and of an index of depletion. However, the censored approach was not always the best. They also suggested that landings-free methods throw away too much information because their censored approach outperformed the survey-and-effort method even at estimating depletion.

3.1.3 On-going studies into approaches that provide model estimates of removals

3.1.3.1 B-ADAPT

Darby (2004, 2005) modified the approach of Gavaris and van Eeckhaute (1998) to estimate removals of North Sea and Irish Sea cod. VPA models fitted to the catch-at-age and research survey data, under an assumption of unbiased catch data, indicated a mismatch between population abundance derived from the catch-at-age data and CPUE from two research survey series; identified by a step in the times series of log-catchability residuals. Consecutive assessment estimates of SSB and fishing mortality have a retrospective bias when models are fitted under an assumption of exact catch- or landings-at-age (Figure 3.1.3.1.1).

Given the following three assumptions:

1) historic catch-at-age data were unbiased,
2) the age structure, but not total number of recent catch-at-age data, is known, and
3) survey catchability is constant,

then a year effect in the form of a multiplier on reported catches can be estimated. Thus all unknown losses from the stock (under-reporting, changes in natural mortality, survey catchability and discarding practices) that are not accounted for by the assumed values of natural mortality, the reported landings and estimates of discarding (if included) are assumed to be unallocated catch; a confounding that cannot be separated without additional information. However, the time series of estimates of adjusted total catch were considered consistent with anecdotal reports of under-reporting and (for Irish Sea cod, Figure 3.1.3.1.2) information supplied to the working group on the level of unreported landings. The model was reviewed and used to provide management advice by the Working Group on the Assessment of Demersal Stocks in the North Sea and Skagerrak (calendar years 2004 and 2005) and the Working Group on the Assessment of Northern Shelf Demersal Stocks (calendar year 2005 *(WGNSDS ACFM review group)*).

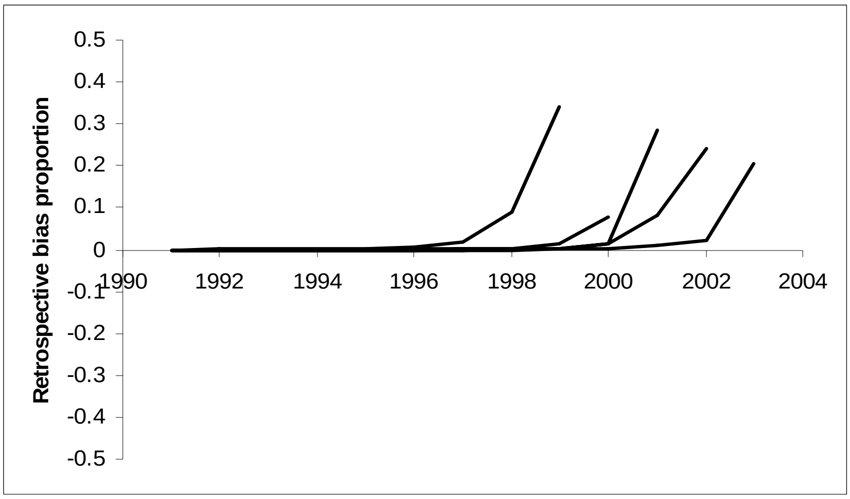


Figure 3.1.3.1.1: Cod in ICES division VIIa: The proportional retrospective bias in SSB as estimated by an ADAPT model fitted to the reported landings-at-age data.

Figure 3.1.3.1.3 presents the proportional change in the model SSB estimates when unallocated mortality is estimated; the retrospective pattern is removed from the time series. For this assessment the inclusion of an additional mortality factor has scaled the estimates of population derived from the catch data so that they are in agreement with those derived from the surveys. In this case the conflict between the estimates of population size from the two sources appears to have been the cause of the retrospective bias which can be *corrected* using the model by adjusting more recent assessment estimates to historic values.

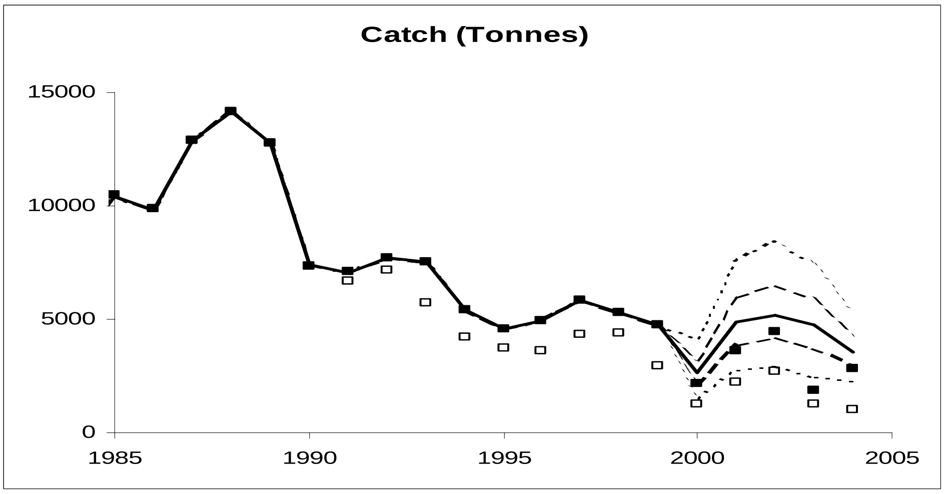


Figure 3.1.3.1.2: Cod in ICES division VIIa: Estimates of recorded landings (open squares), ICES WGNSDS estimates of landings (solid squares) and B-Adapt median estimates of catch (solid line) with 5th, 25th, 75th and 95th percentiles from a non-parametric bootstrap.

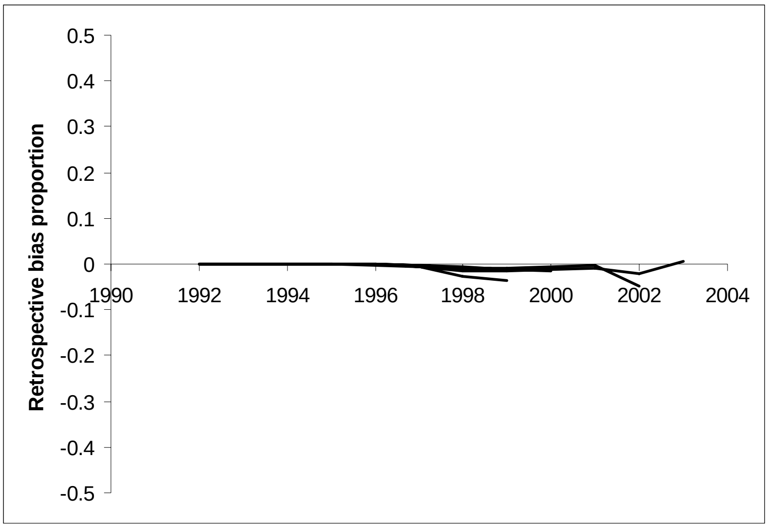


Figure 3.1.3.1.3: Cod in ICES division VIIa: The proportional retrospective bias in and ADAPT assessment fitted to the reported landings at age data.

3.1.3.2 Gadget

Subbey and Howell (WA2) simulated a simple, single species, single area, single commercial fleet, single annual survey, hypothetical model to test the ability of the Gadget environment to model under-reporting of catches. An age length forward projection population model, fitted to catch and survey at length data, was created and artificial data taken from the model. This provided a case where (a) truth was fully known, and (b) gadget was able to model that truth exactly. To this truth a number of experiments were conducted with various patterns of under-reporting of the catch. The Gadget model was then presented with this altered data, and allowed to attempt to optimize parameter values to *correct* for the missing catches. This represented a *best case* scenario – the model was able to exactly fit the data, the assumptions about processes (e.g. formulation of the growth equation) are correct, and there is no noise or error in the data other than the missing catches.

The model was generated using an R-simulator for Gadget, developed at MRI, Iceland. Two stocks were simulated, representing immature and mature stock components. Each stock grows and is subject to natural mortality and there is recruitment into the immature stock. Individuals mature from the immature to the mature stock. Two fleets were used, a commercial fleet and a survey fleet, with a low level of catches for the survey fleet. The survey takes place in the first time-step of each year and the commercial catch takes place in each time-step for every year. The model was run for 20 years (quarterly time-steps), with input parameters loosely based on a haddock-like stock. Parameters were selected giving an initial high population, with an early (lead-in) period of population decline, followed by stable stock level. From this model, data sets were extracted covering the catch in tons, catch at age and length and a survey index. These data sets were then used as data against which a Gadget model could be optimized.

Seven different cases were selected for the experiments. In six of these the under-reporting was assumed to occur in a five-year period (out of the 20 years of model run). Cases were examined with the under-reporting occurring at the end of the run, in the middle, and with the under-reporting flat or with a trend. A seventh experiment involved a flat level of under-reported catch for all years. For the first six cases, Gadget was either required to model a single value for under-reported years, or to estimate a value for each year separately. In all of the cases the missing catch was set to be up to 20% of the total.

In each case it was assumed that the data governing catch at length and age was unaltered, only the overall catch level was in error. This corresponds to under reporting of catches due to unreported landings rather than discards and an assumption of constant unbiased tuning data. Table 3.1.3.2.1 describes the seven cases investigated during this meeting of WGMG; each corresponding to under-reporting of catches.

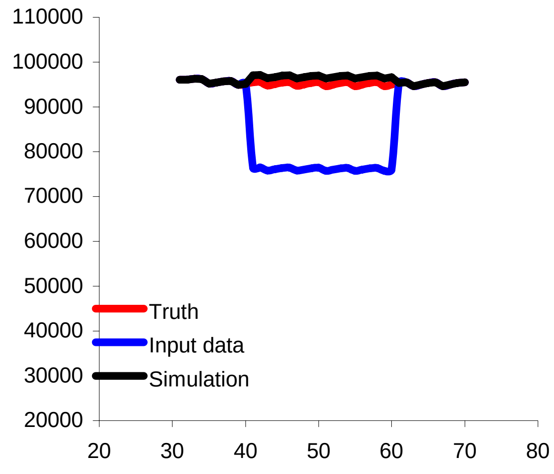
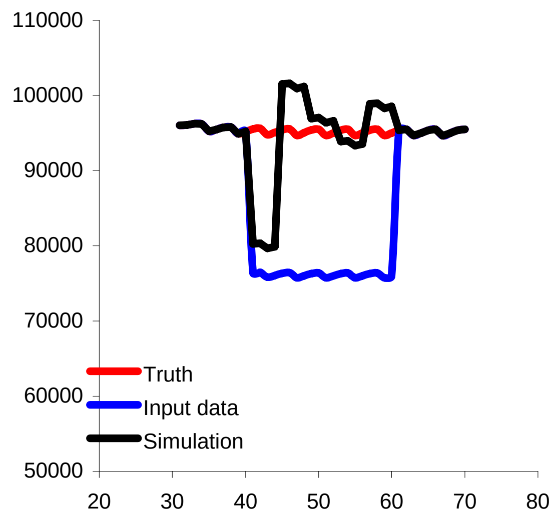
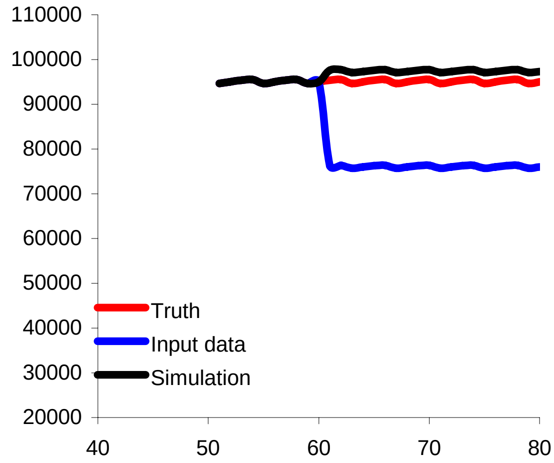
The results of the simulation experiments are shown in Figure 3.1.3.2.1 a) – g). In each case the red line represents the truth, the blue line the catch in tons with error presented to the model, and the black line the final modelled catch level.

It can be seen that in case 1 and 3 (flat levels of under-reporting) Gadget was able to model the missing catches to a high degree of accuracy, although with a slight over-estimation. This was even true for case 3, where the under-reporting of catches occurs at the end of the model run. Cases 2 and 4 represent the same scenario, but with Gadget estimating one parameter per year for the under-reporting – in other words making no *a priori* assumption about the structure of the error. Here it can be seen that the match is less good, with a high estimate in one year being balanced by a low estimate in another. This pattern of over- and underestimation can be seen in cases 5 and 6, where a trend in missing catches has been applied. One possible solution that could be investigated is to estimate the parameters using a trend equation rather than simply estimate one parameter per year. Finally, case 7 shows the ability of Gadget to potentially model cases involving missing catches in all years. Case 7 requires further study to understand how Gadget was able to reconstruct under-reported catches, for all years, in the absence of scaling information. The results for case 7 should be considered *preliminary*.

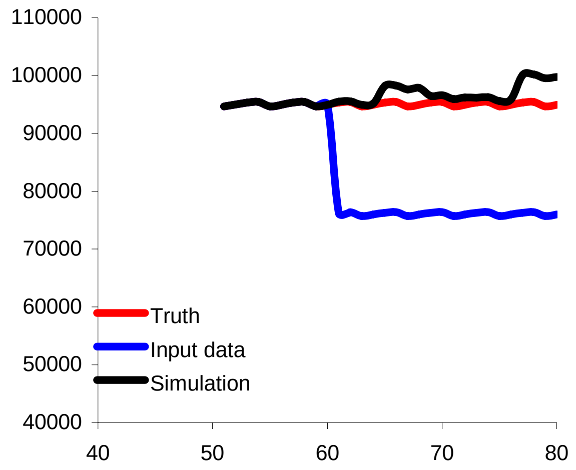
This work has shown that, in principle, Gadget can investigate questions concerning missing catches. Further work is required to evaluate Gadget in more realistic situations (e.g. with errors in other data sets). Other associated issues that need to be resolved are whether the focus should be on discarding (with associated length distribution errors) or on unreported landings (where an assumption of unaltered length distribution may be more reasonable).

Table 3.1.3.2.1: Gadget simulation case descriptions.

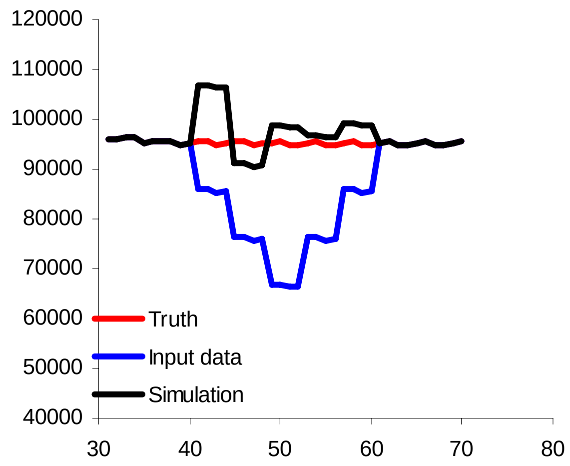
CASE DESCRIPTION	CASE NUMBER: COMMENT
20% missing catch between years 11-15	Case 1: constant value
	Case 2: missing catch per year
20% missing catch between years 16-20	Case 3: constant value
	Case 4: missing catch per year
Year 11, percentage missing 10% Year 12, percentage missing 20% Year 13, percentage missing 30% Year 14, percentage missing 20% Year 15, percentage missing 10%	Case 5: variable between years 11-15
Year 16, percentage missing 5% Year 17, percentage missing 10% Year 18, percentage missing 15% Year 19, percentage missing 20% Year 20, percentage missing 25%	Case 6: variable between years 16-20
20% missing catch for all years	Case 7: constant value

Case	***Results***
a) case 1	
b) case 2	
c) case 3	

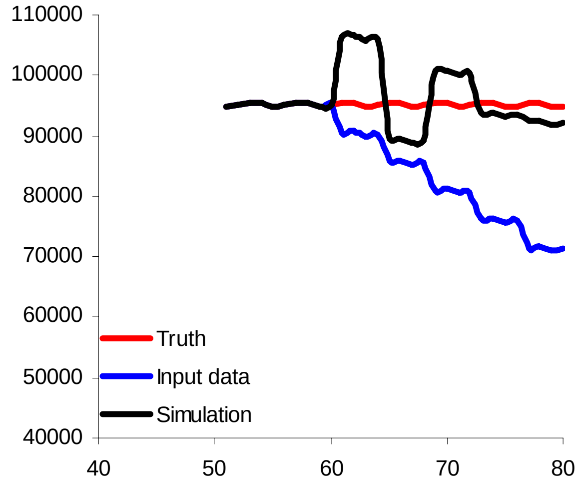
d) case 4



e) case 5



f) case 6



f) case 7

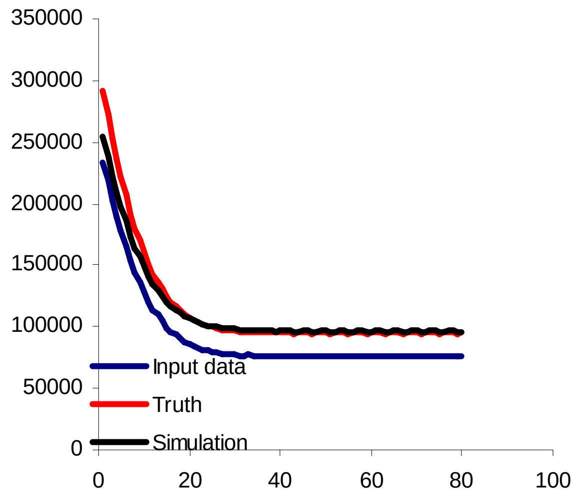


Figure 3.1.3.2.1: Gadget simulation results.

3.1.3.3 Investigations of the approach presented by Hammond and Trenkel (2005)

Some preliminary analyses of the approach of Hammond and Trenkel (2005) were conducted, as part of a project: Accounting for mis-reported catches in stock assessment models used in assessing international (straddling) stocks, in the International Governance of High Seas Fisheries Canadian Science Program. The main goal of the project is to apply the censored approach for misreported catches to age-structured stock assessment models like ADAPT. The censored approach may offer a very reasonable way to address uncertainties in reported catches and may be a good compromise between assuming landings are uninformative (e.g. SURBA) and assuming landings are exact (e.g. ADAPT); however, further studies are required.

In the preliminary analyses, three Bayesian methods for estimating a surplus production model were applied to simulated data sets to test the efficacy of the estimation methods. The mean of the posterior density was taken as the Bayesian estimator. Six simulations were conducted for each of three scenarios: exact catch and two under-reported catch scenarios. The three estimation methods were: 1) standard based on assuming exact catches, 2) the method in Hammond and Trenkel (2005) based on under-reported catches and exact effort data, and 3) a modification of 2) that did not require effort data. The preliminary results suggested that the censored approaches (2 and 3) may offer improved estimates, but considerably more research is required.

The rationale for the third approach to deal with misreported catches without effort data was that usually in cases where catches are misreported there will also not be reliable effort data available. In fact, the model in Hammond and Trenkel (2005) seems somewhat unrealistic for this reason. Good effort data acts like a surrogate measure of catch when CPUE is constant, and this may be part of the reason why these authors found that their censored approach produced better results than the alternative methods they examined. At least the performance of their method needs to be investigated when effort data are measured with error, which is the reality in most cases.

Another problem was the simulation design used by Hammond and Trenkel (2005), which was also used in our preliminary analyses. It is well known that considerable contrast in a stock size time series (i.e. a 2-way trip) is required to reliably estimate the parameters of a surplus production model. Hammond and Trenkel (2005) studied simulated populations with relatively short time-series (i.e. only 10 years) that did not guarantee in all simulations sufficient contrast in stock size to estimate surplus production models. It is commonly understood that in cases where low data contrast occurs the likelihood surface for maximum sustainable yield tends to be flat. In this case, Bayesian estimates and credibility intervals, such as those in Hammond and Trenkel (2005), are heavily influenced by the assumed priors. In these situations the results are essentially pre-determined by the priors, and a more appropriate simulation design to generate data is required to compare the performance of different methods. This applies to the comparisons in both Hammond and Trenkel (2005) and to the methods above.

3.1.4 On-going studies into fisheries-independent methods

Most of the methods considered by ICES so far, to address the catch data issue, have attempted to reconstruct the missing catches so that the conventional approach to deliver advice (catch-based assessment and forecasts) could be maintained. The other route is to ignore the catch data altogether and try to inform managers on the state of stocks based on information from surveys alone. This is the task assigned to the ongoing EU project FISBOAT. A work package in this project is specifically tasked to "*Supply methods for analysing Fishery-Independent stock assessment data to provide managers with relevant information about stocks and their exploitation; provide F-I assessment models; and provide parameter estimation procedures for these models*". Six age-, or stage- or length-based models have been proposed, of which four have been tested for their ability to capture the signal in biomass and recruitment over time using the NRC simulated data (described in WGMG's 2004 report – ICES, 2004a). The other two methods are fitted to length data and could not be fitted to the NRC age based data sets. In brief, the four methods are:

- BREM (V. Trenkel, unpubl.), a two-stage discrete biomass model in which recruitment and the net rate of increase (growth – Z) are treated as random effects, using survey indices for recruits and for all ages combined (only one instance of each, at present);
- SURBA (C. Needle, 2003, already in use in ICES stock assessments), analyses survey indices-at-age, from one or several surveys, through a model in which Za,y (rather than F) is assumed to be separable into age- and year-effects;
- TSA uses the time-series framework of Fryer (2002) adapted to only treat the information from age-disaggregated survey indices (single survey);
- YCC (expanded from J. Cotter, 2001) fits year-class curves through CPUE indices at age from one or several surveys by generalised least-squares estimation of log-recruits at age 0 and total mortality Z, for each year-class.

The main findings were:

- The four methods considered did capture the general trend in biomass and the relative year-class strengths in the simulated stocks. The message to managers, on how the stock was doing recently compared to the past, would have been correct.
- The methods basically behave as data smoothers, implying that they miss quick transient changes (compared to simply plotting the survey indices directly). However, given that the survey indices often have large associated CV, it may be advantageous to avoid provoking undue changes in management in response to short-lived ups or downs.
- The methods are based on the same constant-q assumption as most tuning methods in current use, and are fragile to departures from this assumption (e.g. changes in survey protocol, area, timing, etc., however benign they may look at first sight). Whereas tuned VPA can count on catch and survey information that may complement each other, survey-based assessments rely on only one source which must therefore be of high quality.

- Fitting the model does not require a knowledge of M (which is only needed if one subsequently wishes to derive estimates of F), which is perhaps the most difficult parameter to estimate accurately. There was no loss of performance with the set simulating a fishery with very low F.
- Advice using survey-based assessments can be delivered very quickly (a few days or weeks after a survey is completed), which may allay the recurring criticism from managers that the response from scientists to their requests are much too slow.

A clear limitation of survey-based approaches is that it is challenging to try and get absolute estimates of stock size and only assessments in relative terms are possible. This is not necessarily a problem in itself, as managers may perfectly well be able to make good use of such information. The issue merely arises because ICES has historically contemplated a form of advice in absolute terms, based on *analytical assessments*, as the only acceptable option (which has often put it in big trouble when catch based methods failed for some reason), and was unprepared to operationalise a form of advice based on relative indicators. Yet, there are clear indications from ACE and its expert groups that advice for the ecosystem approach will have to be based on interpretation of indicators, rather than on *hard modelling* of the traditional fisheries kind. There is thus a clear need for the fisheries side of ICES to give more serious consideration to alternative forms of assessments and another nature of advice (e.g. for the *New MoU species* but other species as well).

In the WGMG discussions, questions were raised as to the ability of survey-based methods to reliably inform on fishing mortality F or total mortality Z, whereas this is the quantity that managers are to act upon eventually; the experience is that mortality estimates derived from surveys are often bizarre, notably for individual ages. This issue was not explicitly considered in the tests above. However, if the biomass trend was captured properly, then it is likely that the underlying Z was not much in error, at least for the aggregate ages. The comment was made that getting the magnitude of F should perhaps not be the only focus: if the assessment indicates that the productivity of a stock is decreasing, for whichever cause, then this should call for adaptive (if not restrictive) action by managers to avoid the ensuing impact on the fisheries; it is another weakness of the traditional advice that it has left managers and fishers with the illusion that if a stock is going down for reasons unrelated with the fishery, then they have nothing to do and catches can be kept at the previous level.

Although the methods mentioned above concentrate on traditional indicators of stock status (biomass, recruitment), it was pointed out that the FISBOAT project is wider in scope; it also considers other indicators obtainable from surveys (e.g. on changes in spatial distribution, length compositions) and the plans are to evaluate their performance for advice through feed-back management plans simulations. Further conclusions on the usefulness of the survey-based assessment methods will be available on completion of the project (by Spring 2007).

3.1.5 Summary

All of the approaches described previously in these Section 3.1, form the basis for methods that have the potential to be used for providing advice in the absence of reliable catch data. However, the methods require further development and application to real data such as that being undertaken within the EU-funded FISBOAT study.

Each of the methods (even VPA) provides relative indices of trends in the population dynamics and rates of exploitation (the stock assessments currently used to provide *absolute* management advice are scaled to the assumed level of natural mortality). The approaches that scale the assessment using catch data can be modified to estimate *missing* mortality if an unbiased time series of alternative auxiliary information is available and which extends across the period when the catch is uncertain. Survey CPUE data are generally considered unbiased but noisy and the B-ADAPT and Gadget approaches examined in this report have assumed constant catchability to rescale the reported catch to compensate for the missing component. The effects of increased natural mortality, discarding or under-reporting cannot be separated without prior information and hence, when these models are fitted the additional catch can only be assigned to a category where such information is available. The utility of the estimates will be dependent on the quality of the auxiliary information.

In each of the models used to estimate removals a year effect has been fitted across all ages. This assumes that the missing component is a constant scaling across the complete range of ages or sizes present in the catch data; a scaling of selectivity. In cases where information is available that the missing component is restricted to a particular set of ages, the estimation procedure could be modified to allow for this.

The fishery independent methods and the methods that use proportional catch at age data without reference to the total estimate (e.g. Gavaris and van Eeckhaute, 1998) provide relative trends in the stock and fishery dynamics. If such methods are applied the advisory process needs to be able to manage using relative changes in TAC or effort.

If reliable effort data is available then this could be used as a surrogate index of fisheries impacts in a model (e.g. Hammond and Trenkel, 2005). However, the WG considers that when catches are misreported then it will also likely be true that reliable effort data will not be available, so such modelling approaches with effort data will have limited utility.

The methods reviewed in this section have the potential to assist in the formulation of management advice when the validity of catch data is in doubt. However, they do introduce a greater reliance on survey information, and consequently increased levels of uncertainty in the management process if such information is noisy. Whatever types of data are used, fisheries managers will need to be able to accommodate uncertainty in their decision making process. The decision on whether to focus on fishery-independent data in providing advice is stock-specific, and should be driven by considerations of the trade off between the potential bias of catch data and the potential noise of survey data.

3.2 Estimating discards

TOR b) review approaches for the historical re-construction of discards-at-age data and their incorporation into routine stock assessments; together with the development of estimation methods for the incorporation of information from self-sampling discard schemes.

3.2.1 Introduction

A number of scientific studies and working/study group reports have examined the issue of discarding and attempted to incorporate discard data into assessment data sets; with varying degrees of success (e.g. ICES CM 2004/ACFM:13). The subject has been, and is being, considered in detail elsewhere and an in depth review was beyond the time available to this meeting of WGMG.

The scientific justification for the WGMG included in the ICES' request to review the methodology used for estimating discards from the Rockall haddock fisheries. However, the work on this methodology is still in progress and therefore, could not be reviewed at this meeting.

The text below describes work in progress presented to WGMG at its meeting in June 2006.

3.2.2 Estimating the distribution of discards conditional on landings and other factors

Estimates of discards are widely known to be noisy and are often biased. Discards-at-age for the Scottish demersal fleet is currently estimated by a stratified ratio estimator (Thompson, 1992), mainly using total species landings as the auxiliary variable. There are several problems with this estimator. The stratification scheme is over stratified - on average there are 40 out of 180 strata sampled each year. This leads to ad-hoc *fill-in* procedures to provide discard ratios for unsampled strata. Furthermore, using total species landings as the auxiliary variable can result in very large or even undefined ratios – for example a *Nephrops* trawler may fish in a nursery area, catching and discarding juvenile whiting, say, but land none or very few whiting at market. In the case of the defined but large ratio, positive bias can result when this ratio is used to fill-in neighbouring strata. The net result is an estimator of total species discards at age that is noisy and prone to bias.

Stratoudakis *et al.* (1999), implemented by Millar and Fryer (2005), improved upon this estimator by addressing the known problems directly. They move to total demersal landings as an auxiliary variable, which is less variable than total species landings and less frequently small and zero, thus providing a more stable basis for ratio estimation. Also, by using regression tree analysis to inform collapsing of stratum boundaries, they create *collapsed* ratios that are comprised of several samples. They call this the collapsed stratified ratio estimator, and this estimator is shown to be less biased and more precise that the simple stratified ratio estimator using species landings as the auxiliary variable (Stratoudakis *et al.*, 1999). This scheme also provides an objective basis for implying ratios in unsampled strata. However, estimating the variance of this estimator is complex, and there are still some issues with implementation.

The two approaches described above are design-based. An alternative was presented to WGMG and is a model-based approach estimating the distribution of discards conditional on total demersal landings. Preliminary analysis of age 2 haddock discards from the Scottish discard data set found that in the specific case presented the distribution of discards when discarding occurs follows some positively skewed distribution, and the probability that discarding occurs increases with total demersal landings. Furthermore, when fish are discarded, the numbers discarded are found to increase approximately linearly with total demersal landings and are distributed with constant coefficient of variation. The conditional distribution of discards given total demersal landings is thus modelled as a zero-inflated gamma distribution, which is a mixture of a gamma distribution and a degenerate distribution with a point mass at zero. Only the simplest model for the distribution was presented which ignores year effects and stratum effects (gear, fishing area, time of year).

The overall aim of the work presented is to provide estimates of annual species discards-at-age that have low bias and reasonable precision and appropriate error intervals. Work to date has found sensible groupings of strata that can be applied across species and years (Millar and Fryer, 2005), and it is hoped that these will help increase the precision of the resulting estimates, while the move away from design-based discard ratios should help avoid a lot of the problems of bias resulting from over stratification. A further problem is bias in the time series of reported landings, which will transfer to the estimates of total discards. This may be resolved by using certain effort based conditioning variables. It may be that reported demersal landings are subject to fewer sources of bias than reported species landings, for example reported total demersal landings may not be as sensitive to species misreporting, so that total demersal landings may turn out to be a sensible intermediate stage. To achieve the final estimates, the estimated conditional distribution of discards will be applied to the known values (by trip or otherwise) of the chosen reported conditioning variable, thus allowing us to calculate an estimate of mean total species discards-at-age with appropriate error intervals.

WGMG considered the preliminary results to be encouraging.

3.2.3 Derivation of *Nephrops* discards for non-sampled years

In the Bay of Biscay *Nephrops* fishery (FU 23-24) discards are significant and need to be taken into account in stocks assessments. However, until 2003, sea sampling of discards had been conducted in occasional years (1987, 1991, 1998) and a procedure is needed to fill the gaps in other years. The filling rule traditionally used by the ICES *Nephrops* Working Group (and now by WGHMM) is:

- discards for 1987–1990 are derived from the data collected in 1987;
- discards for 1991–1997 from the 1991 sampling programme; and
- discards for 1998–2002 from the 1998 sampling programme.

The derivation method assumes proportionality between total landings and discards in numbers for each length class. Thus, if i is a year with available data on discards size composition by sampling on board and j is a non sampled year following year i, then for year j and quarter k, the number of discarded individuals by sex (m or f) and size l is derived as:

$$ND_{jklm} = ND_{iklm}\ (YL_{jk} / YL_{ik}) \text{ or } ND_{jklf} = ND_{iklf}\ (YL_{jk} / YL_{ik}) \qquad [1]$$

where ND_{iklm} (or ND_{iklf}) is the number of discarded *Nephrops* for a given sex (m or f) and size class l sampled during year i and quarter k; YL_{ik} and YL_{jk} represent landings in quarter k for years i and j respectively (both sexes combined). A drawback of this method is that it smoothes out inter-annual variations in discarding practices (e.g. in response to variation in recruitment). An alternative method (by Spyros Fifas, Ifremer, Brest, France), based on a logistic fit to the observed data was presented to WGMG.

Let j be a year with no data on discards. By quarter k, the number of discarded individuals by sex (m or f) and by size l, ND_{jklm} (or ND_{jklf}), is calculated from the number of landed individuals NL_{iklm} (or NL_{iklf}) for the same year, quarter k, sex (m or f) and size l, e.g. for males:

$$ND_{jklm} = NL_{jklm}.\exp(-\alpha_k.(l - l50_k)) \qquad [2]$$

where α_k and $l50_k$ are estimated by fitting a logistic sorting ogive using discards and landings data in years for which the former are sampled (here, the fit is made on both sexes combined). This derivation still supposes some stability in the onboard sorting process, *inter alia* that mesh size and minimum landing size (MLS) remain unchanged. This was not the case for this fishery, and thus the time series had to be divided into three periods:

1) *years 1987-1990*: MLS of 8.5 cm total length and mesh size of 50 mm. The deck-sorting ogive was fitted on data from 1987 and used to estimate the missing data for 1988-1990;

2) *years 1991-1999*: MLS of 8.5 cm total length and mesh size of 55 mm. The sorting ogive was fitted on cumulated data for 1991-1998 and used to estimate the missing data for 1992-1997 and 1999;

3) *years 2000-2005*: MLS of 8.5 cm total length and mesh size of 70 mm. The fitting was made on pooled data collected since 2003 and used to estimate discards for 2000-2002.

This derivation approach reduces the interdependence between years of the *proportional* method used by the assessment WG, however: (a) it implies that, for a given size class, no calculation of discards is possible when there is no landing; and (b) the exponential expression gives unrealistic high values of discards for some small length classes, which are normally discarded, but yet may occur in samples of the landings.

The problematic size classes in case (b) were then removed by two ways:

a) A threshold level (set at 1%) based on the cumulative percentage of landings by length classes was applied: this means that discarded individuals smaller than the size corresponding to less than 1% of landings were not taken into account.

b) Generated discards were removed when the calculated discards/landings ratio by size (a decreasing function of size) exceeded observed mean ratios by size (averaged on years 1987, 1991, 1998 and since 2003). Almost all size classes involved by 2) were already removed by 1).

This calculation process retains only a part of the initially generated deck-sorting distributions of discards, mainly the descending limb.

In the last stage of calculation, the resulting descending part of the discards distribution is used to rebuild the missing ascending part. This is carried out based on two hypothesis supported by the available observations (the 6 years sampled):

a) a good correlation between mean sizes of landed and discarded *Nephrops*. A log-log model fits well to the observations;

b) a symmetrical, single-mode distribution of length frequencies of discards (which probably means that selectivity through trawl meshes remains weak in the FU23-24 *Nephrops* fishery).

The distribution used to describe length frequencies in the descending part is:

$$\varphi(l) = \frac{\alpha}{1 + \exp(\beta.(l - lm)} \qquad [3]$$

where α, β, lm are parameters of the distribution (φ(l)=α/2 when l=lm).

Assuming a symmetrical distribution, the whole function of the density of probability is approximated by:

$$\varphi(l) = \frac{\alpha}{1 + \exp(-\beta.(l - lm)} \qquad \text{for } l \le lm$$

[4]

$$\varphi(l) = \frac{\alpha}{1 + \exp(\beta.(l - lm)} \qquad \text{for } l > lm$$

Finally, estimation of discards at length is carried out by fitting the above density function to the descending part of the annual discards distributions resulting from applying stages 1 and 2 with the constraint that the average length of discards should fit well with the log-log relationship between mean lengths of landings and discards.

The discard ratios obtained by this method fluctuate in a wider range (20-67%) than the values (40-65%) obtained with the WG filling rule (Figure 3.2.3.1). The time series plot shows that discard ratios may have decreased in the past (in the middle of the 1990s) and then increased afterwards. The comparison indicates that the catch data used in assessments of this stock, notably for the younger age groups, have probably been biased in some years. This may have affected the perceived recruitment signal.

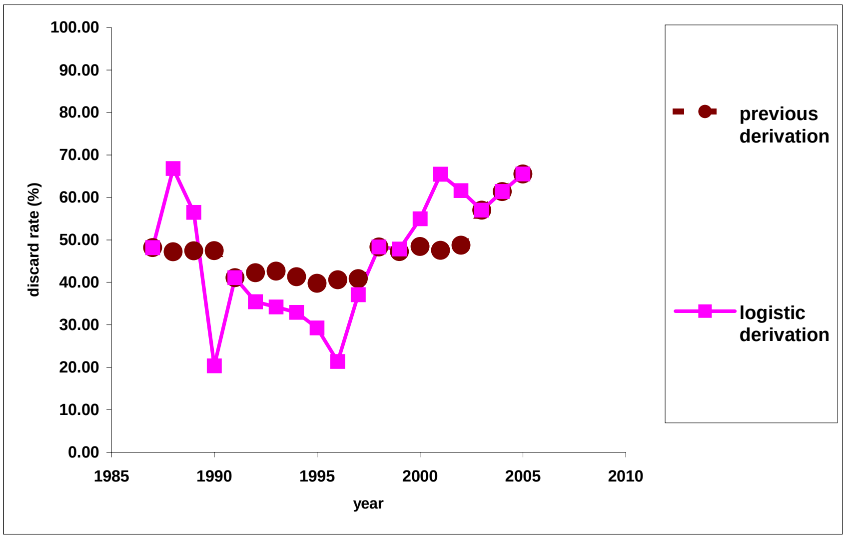


Figure 3.2.3.1: Comparison between discard ratios obtained by proportional and logistic derivations. Combined sexes and sum of quarterly data in each year.

Note that, since 2003, discards have been routinely estimated through a sampling programme on board *Nephrops* trawlers (151 trips and 373 hauls have been sampled over three years). Discards for sampled fishing trips are raised by multiplying the total number of fishing trips. This total number of trips is usually not known and needs to be estimated, which can be done using the number of sales of each vessel at the fish market.

WGMG raised the question of how such a method could be tested. It was noted that no diagnostic of the fit was provided (for example, in the form of a graph of residuals or confidence intervals of parameter estimates) which makes an in-depth review difficult. It was suggested to apply the procedure to only a part of the observed discards data (e.g. 2 years in 3 in the 2003-2005 series), then use it to infer the discards in the omitted year and show a comparison between the predictions and the observations for that year (and repeat by changing the skipped year); i.e. an approach analogous to the statistical approach of *cross-validation* (Stone, 1974; O'Brien and Stone, 1984).

3.2.4 Summary

The utility of discards was discussed and the main points from the discussions are presented below.

- When discards are primarily young fish and interest lies in an assessment of current stock size and where there exists adequate information in the landings-at-age data on recruitment (to the fishery) and cohort decline, discard information may not be so important.
- When discards are primarily young fish and interest lies in prediction and forecasting, then estimates of discards-at-age are important as they provide information on year-class strengths and fishing mortality for cohorts not yet present in the landings data.
- When discards are primarily young fish and the stock is a recruitment fishery, such as North Sea cod or haddock, estimates of discards are important, directly from a management perspective.
- When discards are primarily young fish and there is interest in stock productivity, stock recovery and spawning success, discards-at-age data can provide managers with valuable and current information to assess management plans focusing on these areas.

- When discarding occurs over all ages in a fishery, for example gill net fisheries and recently in the North Sea haddock fishery where the 1999 year-class is still being discarded, estimation of discards is important to augment the landings data and avoid biases in the estimation of population parameters.
- Discard information is also important for ecosystem considerations. Observer sampling schemes are important in this setting as information can be collected on discards of non-commercial species.
- For mixed fishery considerations, estimates of discards for separate fishery components are important, as fisheries may operate on the nursery grounds of their own or another stock. Managers can then judge on the effect that one fishery component may have on another.
- The utility of discard estimates depends on how well they are estimated. Well estimated discards will be important in the cases noted above, but estimates that have large coefficients of variation can introduce noise to assessments (North Sea cod). Also summarised discard information (based on few samples) if taken without consideration to their variation, can lead to unreliable conclusions. The WG therefore highlights the importance of supplying the quality of discard information when discard estimates are presented. The WG further suggests that when estimates of discards are used, their influence (on stock assessment, management decisions etc.) should be considered with reference to the quality of the data.

A number of other issues related to discard sampling were briefly discussed:

- Observer based discard sampling is expensive and sampling levels are often low, this leaves estimates of discards inherently noisy. Rochet and Trenkel (2005) conducted a study into the assumptions behind methods of discard estimation and were unable to show commonly used relationships (proportionality with landings and proportionality with effort) in many of the cases available to them. This highlights the problem of sparse discard information. In the Scottish discard sampling scheme, where extensive discard information exists, and clear relationships can be found between measures of landings and discards, and these relationships can be used to estimate levels of discards. Discard data of this resolution are not widespread, but they show the importance of adequate sampling when trying to make inferences with noisy observations. It is therefore important to consider meta-analyses of discard sampling studies across nations, although prior care must be taken in considering the comparability of discarding practices between different fisheries for such studies to be effective.
- Self-sampling discard schemes are a less expensive method of collecting information on discards, and could improve the precision of discard estimates through increased sampling levels. Such a scheme is underway in Holland but no information was available to the working group on the progress of this initiative. Given the widespread problem of poor sampling levels in national observer discard sampling schemes, self-sampling information should provide a valuable addition when analysed in conjunction with available observer data.

3.3 Sensitivities of catch-at-age stock assessment methods

TOR c) investigate and test the sensitivities of catch-at-age stock assessment methods to known data problems with particular reference to the retrospective problem.

3.3.1 Introduction

There are several sources of uncertainty that affect estimates of stock dynamics derived from fish stock assessment models, including process error, measurement error, model uncertainty, and parameter-estimation error. If the errors are random they introduce unbiased noise to the assessment estimates and retrospective distributions are characteristically two tailed. If the errors have a systematic bias, one sided retrospective patterns of over- or under-estimation are induced. WGMG has examined the potential causes and possible cures for the bias in assessments estimates in many of its reports (e.g. WGMG 2004a). Model structural uncertainty was examined in detail for a number of stocks by Patterson *et al*. (2000) and ICES (2002 PA WG), the studies established that the structural uncertainty can be as, or more, significant than estimation uncertainty and should be regularly reviewed.

The studies reported in this section, report on-going research into the sensitivity of stock assessment model estimates to uncertainty and potential causes of retrospective bias. Sections 3.3.2 and 3.3.4 examine model misspecification uncertainty, in terms of the fitting of models that do not allow for changes in conditional constraints that are assumed to be constant; natural mortality and selection-at-age. Section 3.3.3 examines model structural uncertainty in the assessment of Norwegian spring spawning herring and Section 3.3.5 continues the group's debate on the use of influence diagnostics.

3.3.2 Trends in natural mortality

As a potential cause of retrospective bias, trends in natural mortality were highlighted in a presentation by Dankert Skagen at this year's meeting. The presentation and material were basically an extract from an earlier working document presented to WGMG in 2001 (Skagen, DW). Some possible causes of 'retrospective bias' in an ICA-like assessment model' Working Document, WGMG 2001). Using artificial data, with no noise, but with an increasing trend over time in natural mortality, an assessment fitted to catch and survey data that assumed constant natural mortality, would estimate an increasing trend in fishing mortality, except in the last years, where the trend was inverted. This was the case both for a separable model (Figure 3.3.2.1) and a tuned VPA.

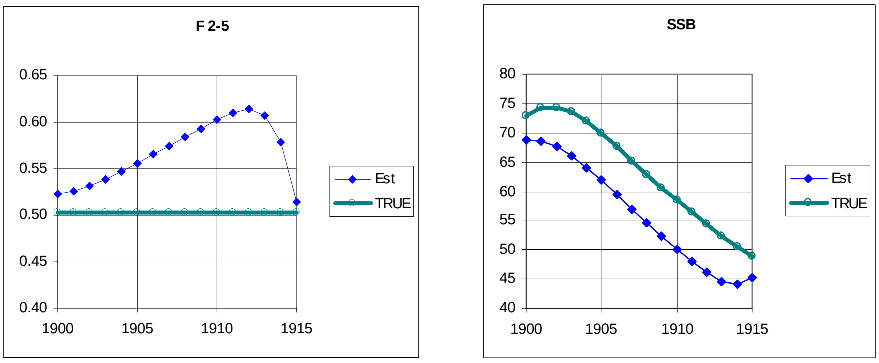


Figure 3.3.2.1: Estimates of F and SSB with a separable model assuming fixed M at 0.15 for catch and survey data with no noise but with a trend in M (increasing linearly from 0.15 to 0.30).

The likely explanation for this is that the model picks up the signal of an increasing total mortality in the data, interprets that as a trend in F, and adjusts the stock numbers to make the catches fit with this F. In the last years, however, the estimate of stock abundance is dominated by the survey information. Even though the catchabilities are derived from incorrect stock numbers, they are less in error, which leads to estimates of stock abundance, and hence, Fs in the last year that are closer to reality

The deviation from reality demonstrated here, also represent retrospective error, since the deviation from reality is different in the last and in the earlier assessment years. Figure 3.3.2.2 illustrates this.

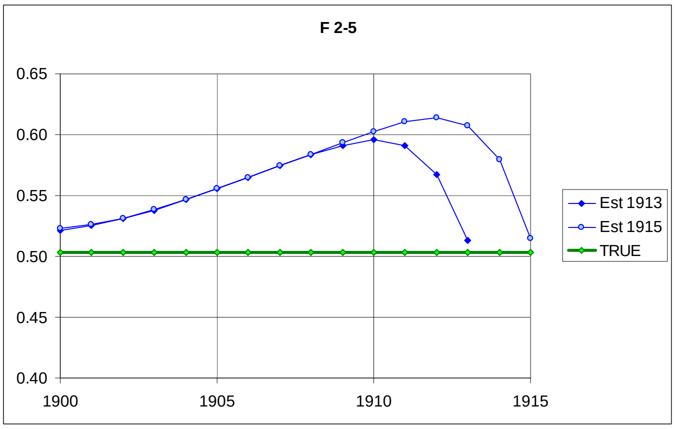


Figure 3.3.2.2: Estimates of F in a separable model assuming fixed M at 0.15 for catch and survey data with no noise but with a trend in M (increasing linearly from 0.15 to 0.30), with two different terminal years.

This effect of trends in M has a wider perspective, because in an assessment model, the natural mortality is assumed to cover all disappearance of fish not accounted for by the catch data that go into the model. Hence, the effect shown here will also result from trends in misreporting, discards and migrations; in addition to trends in loss due to *natural* causes.

3.3.3 Strong year-class effects in separable models

The basic assumption of separable cohort models, i.e. representation of fishing mortality coefficients as a product of two factors (age-dependent selection factor and year- or effort-dependent one) sometimes is considered to be restrictive for real stock-fishery systems. The stability of selection pattern over years is often violated by variations in fishing regime and by natural reasons. One of known *natural* reasons why selection pattern is not stable over years is that more abundant cohorts (generations) are of higher vulnerability to fishery than less abundant ones. Miscount of this factor in stock assessment, undertaken by means of separable models, may lead to biased stock size estimates because of strong violations of selection pattern stability. There are evidences also that for some species, e.g. Norwegian spring-spawning (NOSS) herring, very abundant year-classes may have different spatial distribution at certain stages of their life and, consequently, may have peculiarities in interactions with fishing fleets operating in different regions. For NOSS herring it is generally assumed that recruitment of large year-classes from the Barents Sea to the fishery in the Norwegian Sea may generate non-separability due to the fishing fleet operating on different age components of the stock. The Norwegian fleet operates along the Norwegian coast, thus largely missing newly recruited year-classes, while vessels from other countries operate in the Norwegian Sea (ICES, 2004c). While from common sense the above mentioned looks like a real problem for stock assessment by separable cohort models, it is still questionable to what extent, and, if it is really so, how to diminish the problem.

At the Study Group (ICES, 2004b) rather reasonable results were obtained for the NOSS herring stock assessment using the ordinary separable model ISVPA in its catch-controlled version – the version where errors of model approximation are attributed to violations of selection pattern stability assumption. The question was would the results be improved by incorporation of generation-dependent factors (g-factors) into separable representation of fishing mortality coefficients?

The effect of implementation of so called *triple-separable* assumption (incorporation of the third, generation-dependent, term into separable representation of fishing mortality) in comparison to results of separable cohort models based on ordinary separable assumption on simulated and real data was investigated (Vasilyev WC1).

As a simulated data set the Data Set 3 generated at the ICES Study Group on Assessment Methods Applicable to Assessment of Norwegian Spring-Spawning Herring and Blue Whiting Stocks (ICES, 2004b) was used. This data set was considered at the Study Group to be similar to a number of real world applications, including those faced by the Working group. The underlying stock model was made close to the dynamics of the Norwegian spring spawning herring, and had the following characteristics: Simulated recruitments were drawn from a multinomial distribution, with different mean-variance characteristics for *good* versus *poor-average* year-classes. The probability for *good* recruitment was 1 in 8, and in the case of good recruitment the recruitment was drawn at random with a uniform probability between 50 and 100 units. In cases of poor to average recruitment, the recruitment was drawn at random with a uniform probability between 1 and 10 units. The instantaneous coefficient of natural mortality (M) is set to 0.15 for all age groups, which is input as a given (not estimated) by the assessment programs. Fishing mortality was assumed stable (with noise) for the first 15 years of the time series and then followed by an increasing trend. The same selection as for the survey data (see below) was used with the addition that the selection at age 4 is set to 0.8, at age 3 to 0.6, at age 2 to 0.02 and at age 1 to 0.01. The selection pattern is multiplied by a scaling factor (0.15 before the change starts) and which in the last 10 years is interpolated linearly with time to 0.25 in the final year (ICES, 2004b).

Figure 3.3.3.1 compares profiles of catch-at-age alone derived component of the TISVPA model loss function (in terms of the median of squared residuals in logarithmic catch-at-age) when different age diapasons (ranges) are used for estimation and application of g-factors. As it can be seen, incorporation of g-factors makes residuals lower, but the result depends on the choice of the age diapason. The lowest minimum was found when g-factors were estimated for ages 2-10. If there is no *a priori* information for which age groups the selections are subjected to changes if the generation is *specific* one, it seems to be reasonable to use in assessment the age diapason for g-factors which gives the best fit. It also seems that it may be better choice even if it was known that this effect could take place for oldest ages, because for oldest ages the data are usually of bad quality and this may cause non-robustness of g-factors estimates. All other model settings were taken the same as at the Study Group (catch-controlled version with unbiased separable representation of exploitation rates). The *within-year effort redistribution by ages* sub-model was tested in all cases.

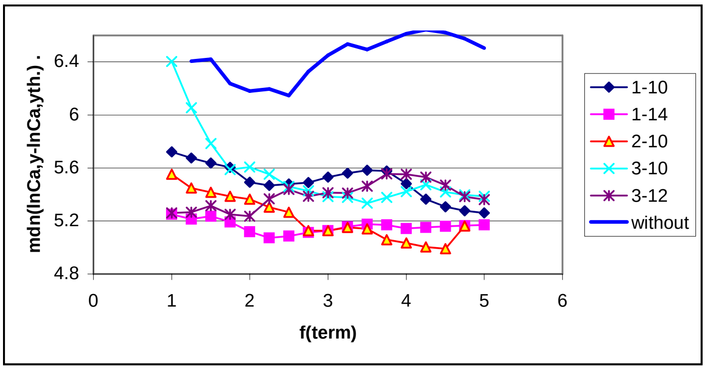
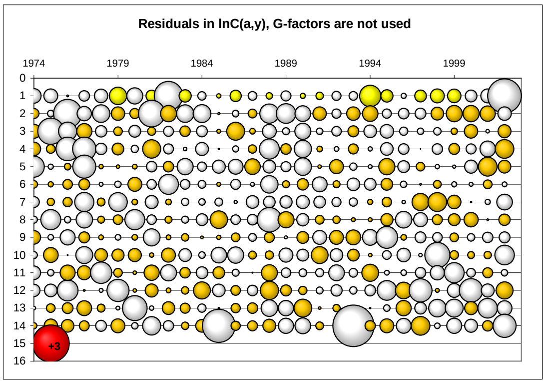


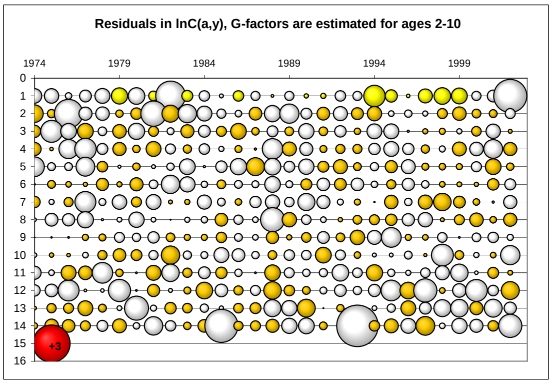
Figure 3.3.3.1: Simulated data. Profiles of catch-at-age derived components of the TISVPA loss function for different choices of age diapasons for estimation of g-factors.

Residuals in logarithmic catch-at-age (catch-controlled version of the TISVPA) are represented by Figure 3.3.3.2. As it can be seen, implementation of g-factors diminishes the cohort effects of the pattern. The diminishing of the cohort effects is more apparent for the effort-controlled version (residuals are attributed to errors in catch-at-age), which is more close to most of other popular separable cohort models (see Figure 3.3.3.3). For the effort-controlled version of the model the best choice (corresponding to the lowest minimum of the loss function) of the age diapason for g-factors was found as ages 2–10.

a)



b)



c)

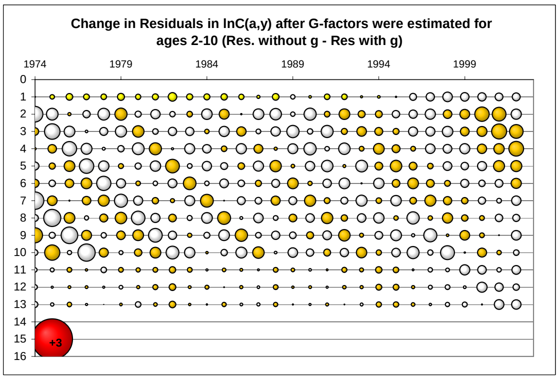
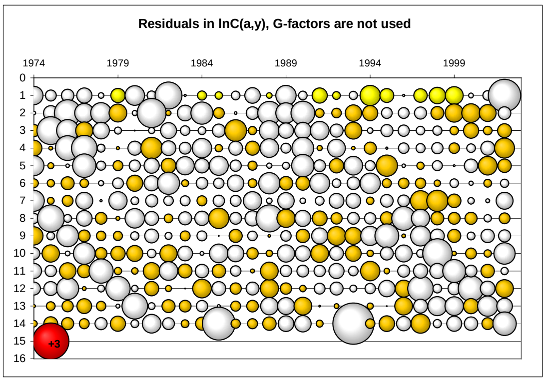
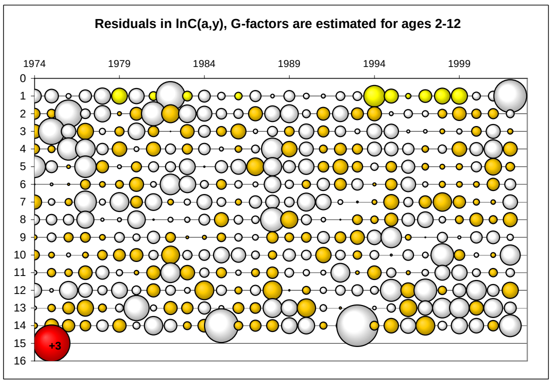


Figure 3.3.3.2: TISVPA-residuals (catch-controlled version) in logarithmic catch-at-age for simulated data: a) without g-factors and b) when g-factors were estimated for ages 2-10; c) - difference between a) and b). Red dots represent the scale.

a)



b)



c)

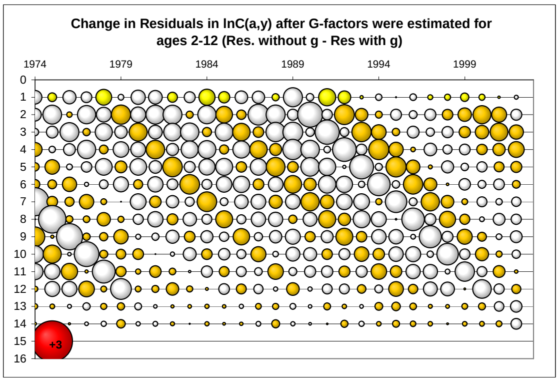
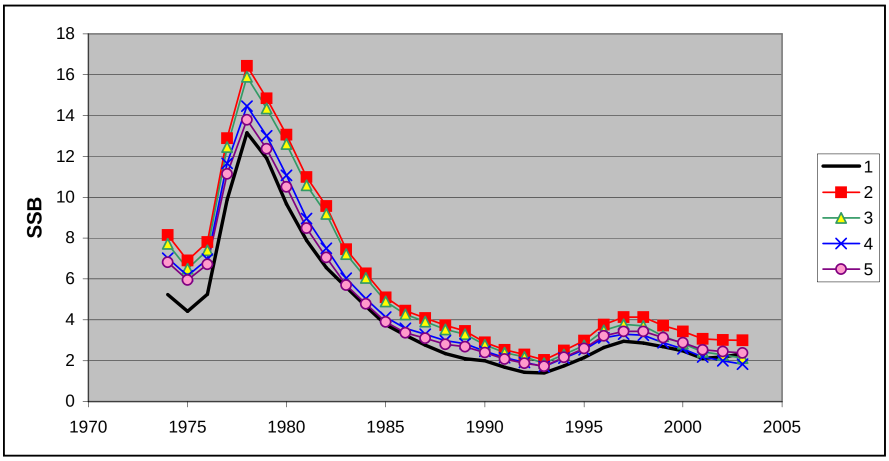


Figure 3.3.3.3: TISVPA-residuals (effort-controlled version) in logarithmic catch-at-age for simulated data: a) without g-factors and b) when g-factors were estimated for ages 2-12; c) difference between a) and b). Red dots represent the scale.

Figures 3.3.3.4 (a and b) below compare TISVPA and ISVPA – derived estimates of SSB with true values for the simulated data set. Interesting to note that while for traditional (*double*) case of separable representation the effort-controlled version behaved badly and catch-controlled version of the ISVPA giving more reasonable estimates was chosen at the Study Group (ICES, 2004b), implementation of g-factors gives more apparent improvement just for the effort-controlled version. This is not surprising, because in the catch-controlled version the abundances are calculated via catches directly (while in the effort-controlled version they are calculated via model-derived catches).

a)



b)

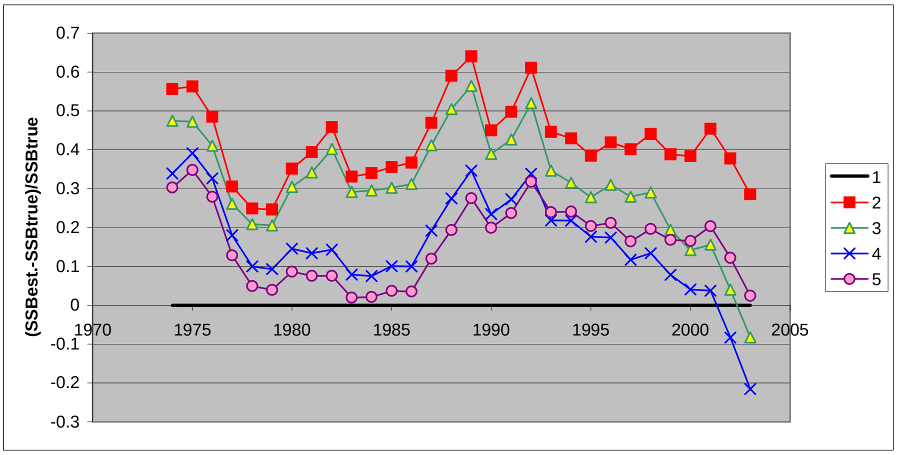


Figure 3.3.3.4: Catch-at-age based estimates of SSB (a) and their relative deviations from true values ((estimate-true)/true) (b) for ISVPA models fitted to simulated data

1 - true values (simulated by (ICES, 2004b))

2 – catch-controlled ISVPA, G-factors are not estimated

3 – catch-controlled TISVPA, G-factors are estimated for age groups 2-10 (best choice)

4 – effort-controlled TISVPA, G-factors are estimated for age groups 1-14

5 – effort-controlled TISVPA, G-factors are estimated for age groups 2-12 (best choice)

The *real data* on which the role of *triple-separable* representation was tested were the data on Norwegian spring-spawning (NOSS) herring (ICES, 2006a). The same version of the ISVPA model which was used previously for herring stock assessment at this Working Group was tested (catch-controlled with condition of unbiased residuals in logarithmic catch-at-age). Figure 3.3.3.5 represents the profiles of the catch-at-age derived component of the TISVPA model loss function (in terms of the median of squared residuals in logarithmic catch-at-age) for different age diapasons of estimation and application of g-factors. As can be seen, the best choice of the age diapason was from ages 4 to 8.

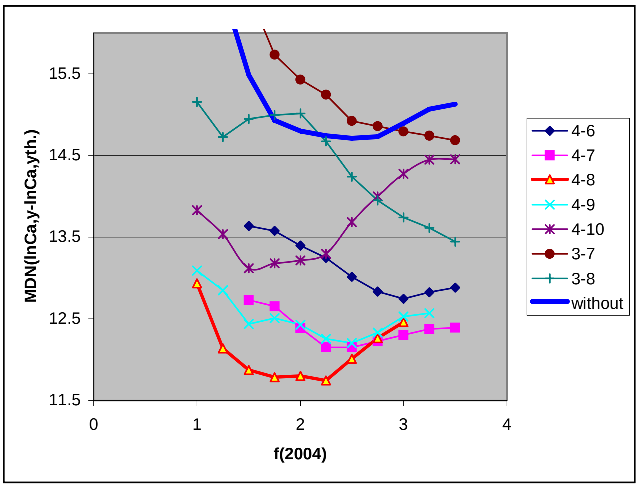


Figure 3.3.3.5: NOSS herring. Profiles of catch-at-age derived components of the TISVPA loss function for different choices of age diapasons for estimation of g-factors.

Estimated values of g-factors (given as G-1) are represented on Figure 3.3.3.6. Figure 3.3.3.7 shows the estimates of selection pattern. Figure 3.3.3.8 illustrates the diminishing of cohort effect in residuals if g-factors are taking into account.

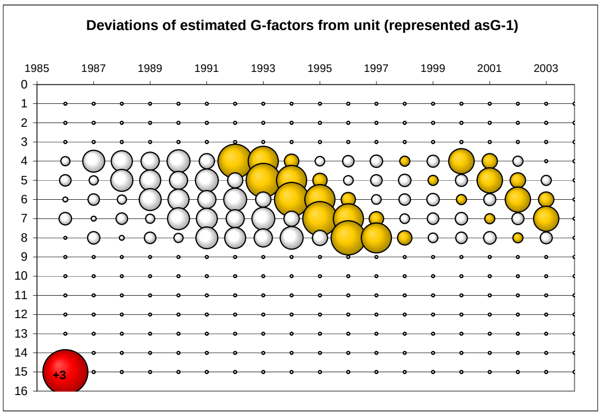


Figure 3.3.3.6: NOSS herring. TISVPA-derived estimated g-factors (presented as G-1) for the case when age diapason for g-factors was taken as ages 4-8

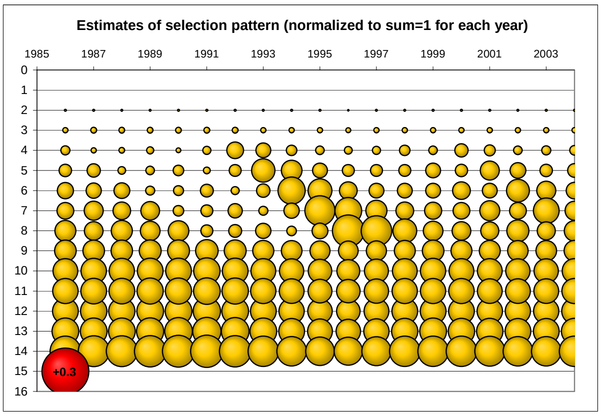
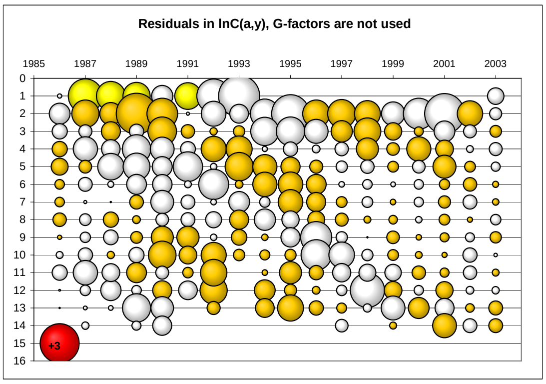
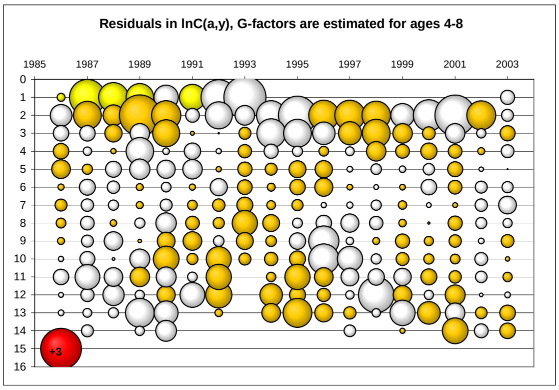


Figure 3.3.3.7: NOSS herring. TISVPA-derived estimates of selection pattern.

a)



b)



c)

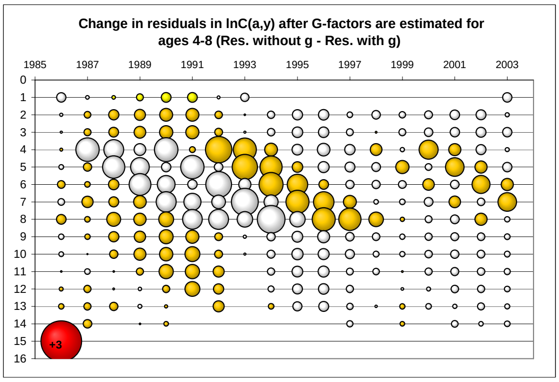


Figure 3.3.3.8: NOSS herring. TISVPA-residuals in logarithmic catch-at-age: a) without g-factors and b) when g-factors were estimated for ages 4-8; c) - difference between a) and b). Red dots represent the scale.

If to consider the choice of age interval for application of g-factors from point of view of all data sources used in the stock assessment by means of ISVPA (TISVPA), the best one was found for age window from 4 till 9 (see Figure 3.3.3.9). If to look at components of the loss function, it is seen that the improvement of fit is seen for most sources of information (Figure 3.3.3.10). Figures 3.3.3.11-3.3.3.12 represents the estimated selection pattern and changes in residuals in logarithmic catches due to taking into account the generation dependent factors.

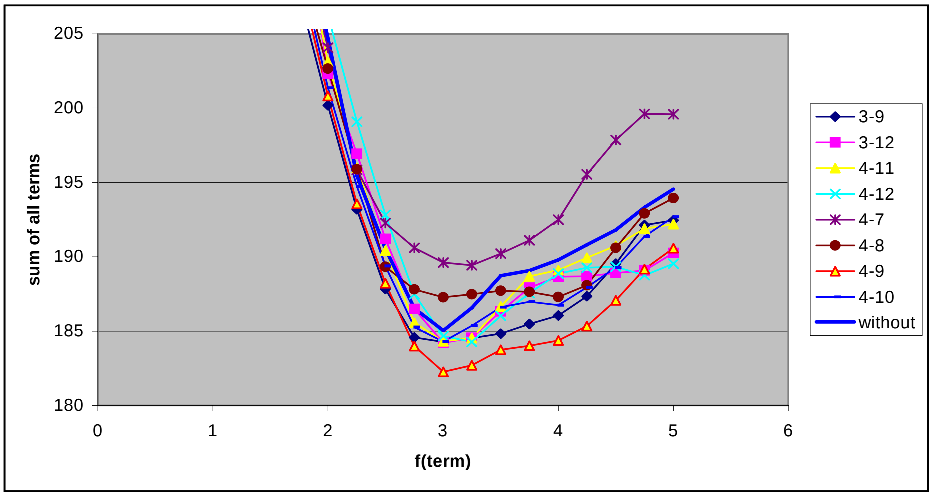
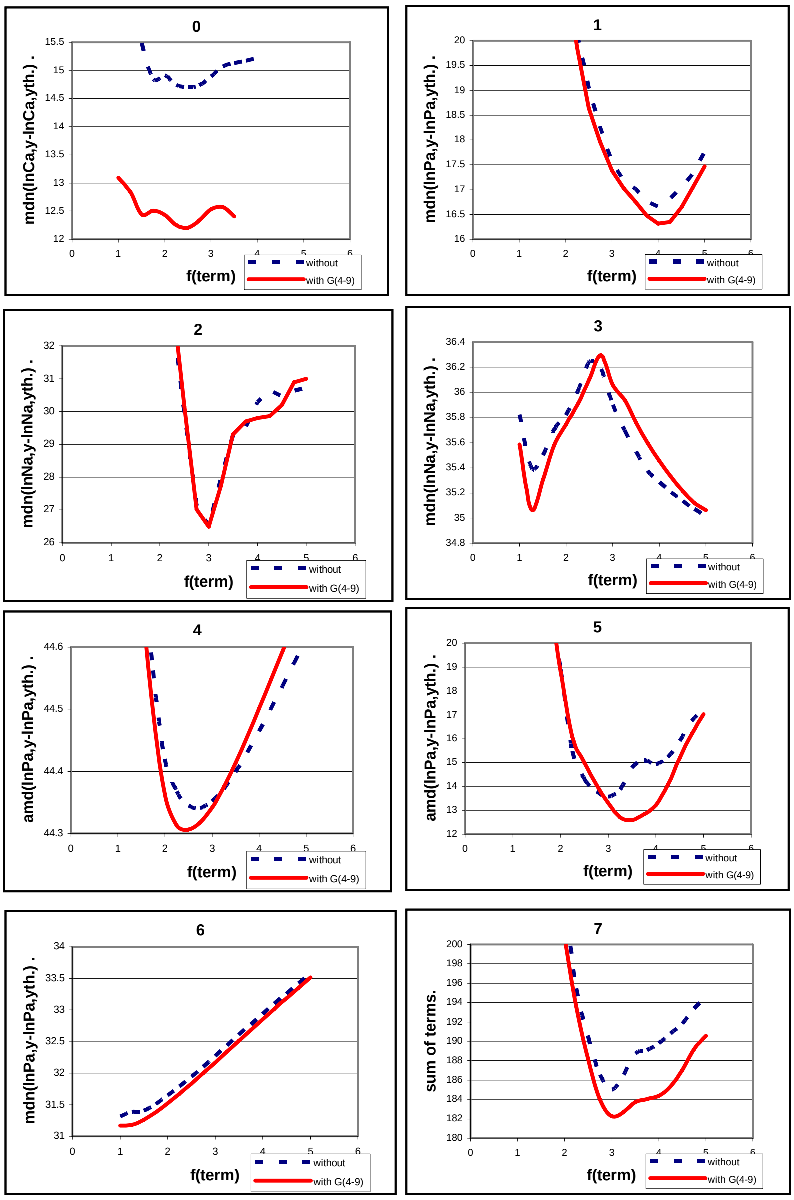


Figure 3.3.3.9: NOSS herring. Profiles of the TISVPA *total* loss function for different age ranges of estimated g-factors.



0- catch-at-age
1- spawning grounds acoustic in Febr.-March
2-acoust. surv. in wint. area Nov.-December
3- acust. in wintering areas, January
4- Young herring in the Barents Sea (June)
5- Feeding areas, May
6-Young herring in the Barents Sea, September survey
7-All data

Figure 3.3.3.10: Profiles of components of the TISVPA loss function without and with application of g-factors for ages 4-9.

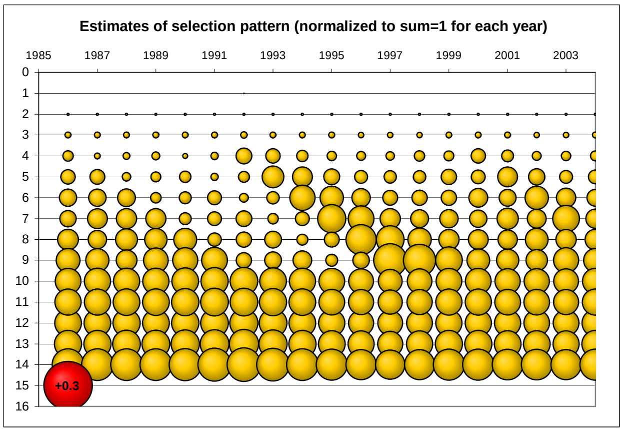
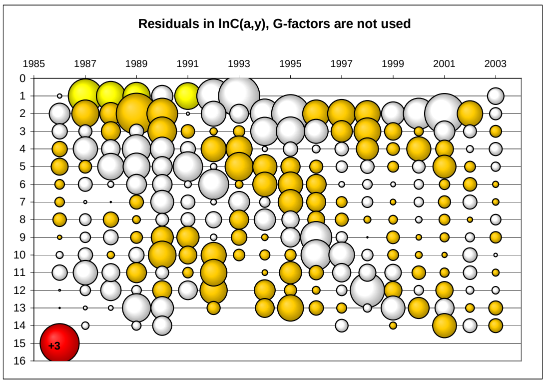
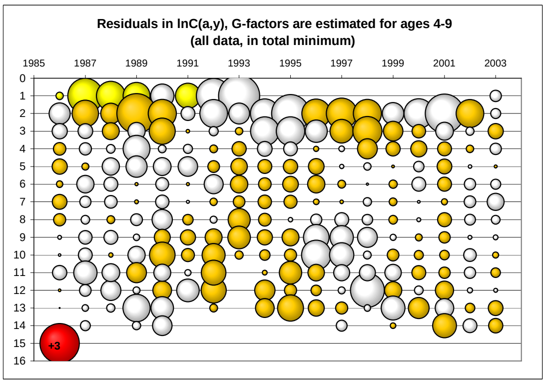


Figure 3.3.3.11: NOSS herring. TISVPA-derived estimates of selection pattern (g-factors are estimated for ages 4-9 – best case if to consider all data).

a)



b)



c)

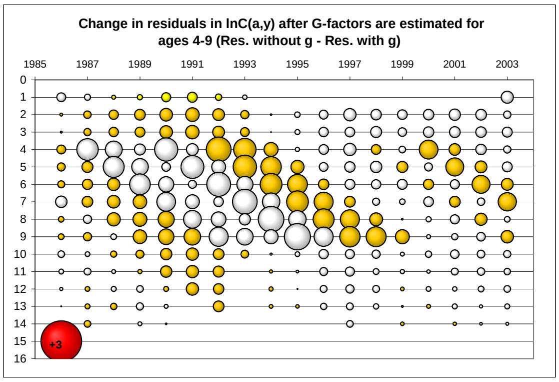


Figure 3.3.3.12: NOSS herring. TISVPA-residuals in logarithmic catch-at-age: without g-factors (a) and when g-factors were estimated for ages 4-9- best case if to consider all data (b); (c) - difference between them . Red dots represent the scale.

As in case of the simulated data set, for NOSS herring data the *triple-separabilization* effectively diminishes the cohort effect in residuals for catch-at-age. In addition, it results in a better fit for most of sources of auxiliary information. At the same time the changes in the results in terms of spawning stock biomass are not substantial (Figure 3.3.3.13). Results of retrospective runs are also very similar (Figure 3.3.3.14).

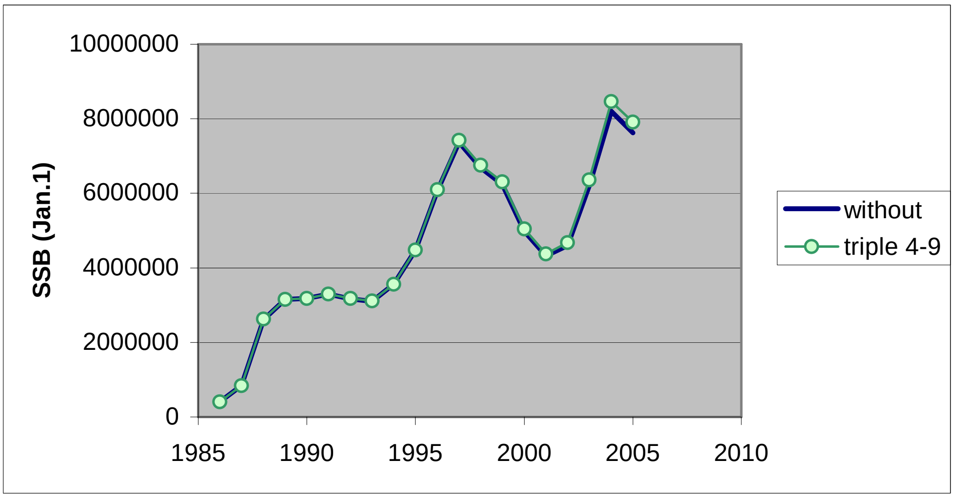


Figure 3.3.3.13: NOSS herring. Comparison of TISVPA-derived SSB without and with g-factors used for ages 4-9 (the best case when all data are considered).

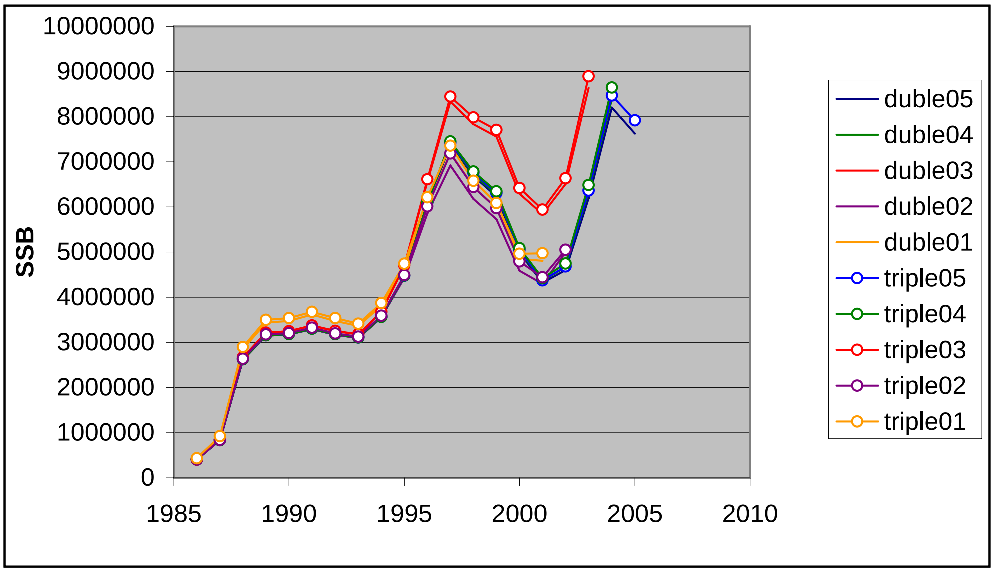


Figure 3.3.3.14: NOSS herring. Comparison of retrospective runs for ordinary and triple-separable cases.

The following conclusions perhaps could be drawn from the above described:

1) The year-class correction of the selection pattern in separable models can improve fit and diminish the cohort effect in residuals, what is pleasant from point of view of diagnostics.

2) However it may not change the results of stock assessment significantly, but this may not be necessarily valid for all separable models and stocks, because in construction of the ISVPA (TISVPA) model the robustness was the main subject of attention. In less robust models the effect on the results of stock assessment can be bigger.

3) Estimation and implementation of generation-dependent corrected selection patterns may be important in stock projection, but the problem may be to detect

surely that the generation is specific when it has appeared in catches only in its youngest ages.

3.3.4 Sensitivity of assessment results to model specification

Needle (WC2) looked into the effects of model uncertainty on the assessment of North Sea haddock. Specifically, the document tested the influence of different user-defined model settings on the results of XSA assessments (Darby and Flatman, 1994) and associated short-term forecasts, run under the FLR framework. The settings investigated were shrinkage, the plus-group, the catchability plateau, and the choice of surveys to use. Results were summarised as time-series of mean F, SSB and recruitment, and as empirical distributions of these quantities in 2004 (the final assessment year) and 2006 (the quota year).

The WG view of this approach was that it could be used as a sensitivity analysis, to investigate which XSA settings were influential in the final assessment, but that its utility as a method for generating probabilistic assessments and forecasts was questionable. The choice of settings used in the analysis was largely ad hoc, and many of the combinations of settings which resulted would be unlikely to be accepted by stock assessors. A method of this kind could be used to generate probabilistic assessments and forecasts, but would have to include two key features:

1) Parameter settings would have to be drawn (bootstrapped) from prior distributions defined by stock assessors. These could be uniform, or could favour the settings most commonly used in recent WGs.
2) The contribution of each bootstrapped assessment to the final results distributions should be weighted by the goodness-of-fit of the assessment. Therefore, a poorly-fitting model would have less effect on the eventual summary distributions of abundance and mortality than a well-fitting model. Methods such as XSA do not produce goodness-of-fit statistics, so alternatives would have to be considered.

An implementation including these aspects could use a simple frequentist bootstrap approach, or a Bayesian framework in which other types of error (particularly, parameter estimation error) could be considered as well.

WGMG recognised that the current practice of basing management decisions principally on deterministic point estimates of current and future stock status was extremely problematic, but concluded that the incorporation of uncertainty into the process was not as straightforward as suggested by the presentation and would require careful consideration. However, uncertainty is one of the main issues in fisheries management science and must be addressed soon.

3.3.5 Overview of the impact of data on assessment error

Retrospective bias is a special, but very important, case of assessment error, which implies that the stock is systematically under- (or over-) estimated, if the converged assessment is taken as the *true* state of the stock. One may suggest two ways to explore how assessments become biased or erroneous: To trace the effect of individual observations on the estimate of various parameters, or to explore, for individual parameters, which data determine the values of that parameter. Such analyses are not part of the routine diagnostics in most assessment tools, but represent a field that may be worth developing further.

The procedure would be to change either a parameter or an observation, redo the parameter estimation (except for the perturbed parameter) and record the resulting changes. One may avoid redoing assessments by assuming that the response is locally linear (e.g. Cadigan and Farrell, 2002), and consider just the change of some result of interest resulting from a small perturbation in each parameter or data point. This leads to a direction vector for the maximum effect, which identifies the data with the largest impact. This simplified approach has some shortcomings, however, because the response may be far from linear over relevant ranges of perturbations.

With respect to data perturbation, recognising that the objective function (F) typically is a sum of terms ϕ_i each representing one observation, one may consider how each observation influences the objective function by calculating the Jacobian matrix (the matrix of all the terms $\partial\phi_i/\partial\theta_j$ where the θ_j are the individual model parameters that are estimated. The simple interpretation is that a negative first derivative means that this observation ‘wants’ a larger value for the parameter. Therefore, studying the Jacobian gives some insight in how the various pieces of information are balanced when deciding on the best parameter value. The objective function can also be thought of as a weighted sum, $\Sigma_i\, w_i\phi_i$, where the w_i weights are all one’s. When these weights are perturbed then the Jacobian is an important component of the local influence diagnostics (see equation 4 in Cadigan and Farrell, 2002). However, the Jacobian does not directly inform us about the sensitivity of other model components in VPA such as catches, M, etc.

The approach is to perturb a parameter of interest, typically the fishing mortality level in the last year, redo the assessment with that parameter fixed and consider the change in the individual terms in the objective function. The rationale is that each parameter estimate is a compromise between conflicting pieces of information, and the change in the individual terms in the objective function highlights the compromise between data that indicate a higher, and data that indicate a lower, value of the parameter in question.

The local influence approach to perturbation analyses (Cadigan and Farrell, 2002; Cadigan and Farrell, 2005) referred to above involves studying the effects of small perturbations to model components using basic concepts in differential geometry. All of the components (e.g. catches) considered are perturbed simultaneously. The geometry of the surface of the influence measure versus the perturbations is examined to assess influence. The main diagnostics are simply the slopes of the influence surface. The influence may be on the objective function itself, or interest parameters like SSB or terminal fishing mortality. In the context of retrospective bias, the measure of interest may be a measure of the bias, e.g. Mohn’s (1999) ρ. In this multi-dimensional surface one may consider the direction vector with maximum slope (i.e. local influence), which represents the combination of changes in data or parameters that has the largest influence on the result, or will lead to an improvement of e.g. the retrospective bias. When different types of input components are perturbed (e.g. catches and M), the type that has the largest local slope is normally suggested as the more likely source of the retrospective problem, among the sources that are assessed. This use of local influence diagnostics is directed at diagnosing model misspecification, whereas the usual application of the method is directed at detecting inputs that model results are sensitive to, which may not be the result of misspecification.

The method offers some promise in sorting out what are more likely causes for retrospective patterns, but further research is required to refine the method and evaluate its efficacy when applied to simulated data with known problems. Issues of nonlinearity and the comparability of perturbation schemes to different model components are also important to understand, and further research is required for this purpose.

The Retrospective Working Group at NEFSC is exploring causes of, and potential treatments for, retrospective patterns that have been observed in a number of their recent groundfish assessments. Both empirical and simulation approaches are being explored and the work is currently focused on virtual population analysis as the assessment tool. One approach examines the magnitude of change needed in such things as catch, natural mortality rate, and survey catchabilities to eliminate retrospective patterns in existing assessments. This work is continuing using simulated data to determine if the correct source of the retrospective pattern can be identified. Preliminary results suggest that not only can the source not be identified, but that eliminating the retrospective pattern sometimes produces results further from the known truth than the original results. Another approach is to look within the data to determine if specific points are driving the retrospective patterning, so-called *retro-tension*. The group will be compiling a list of assessments with, and without, retrospective patterns from their region and hopefully other regions, as well; with the hope of finding commonalities that could point towards a cause.

The diagnostics may suggest a change in some population process or population sampling process that can explain the retrospective pattern. However, it is almost certainly the case that additional information will be required to resolve what the source of the retrospective problem. Diagnostics simply provide illumination, and usually not definitive answers. They also do not fix the problem they attempt to find.

The common experience is that the impact of the data on the estimates of stock abundance can be summarized as follows (ICES, 2006b). Exceptions can occur of course, depending on specific model approaches.

- Catch data dominate the estimates of the N-matrix back in time and for early ages, because the convergence of the historic estimates of N give little slack if the Ns are fitted to the catches.
- Survey data dominate the estimates for the last year(s), because there is little in the catches to determine at which stock number the N-trajectory should start.
- The estimates for the oldest ages are dominated by the catch data, together with assumptions about F.
- The response to a deviating observation depends on where it is placed in time and age. That means that the role of this observation will change as new year's data are added, i.e. it will have a different effect in retrospect than in the last year.
- Deviating data only show up as residuals to the extent that they cannot be amended by adjusting the population and the catchabilities.
- What appears to be a retrospective problem may not necessarily be a wrong estimate of the present and a right estimate of the past, the opposite can also be the case (e.g. Mohn, 1999). Hence, when faced with a retrospective problem, an early question should be: Is the error in the present or in the past?

3.3.6 Summary

As has been demonstrated at previous meetings of the group and previous research studies (e.g. *uncertainty project*) the causes of assessment uncertainty are manifold. The studies reported in this Section have continued the research examining the sensitivity of assessment model estimates to external data driven, and internal, structural, sources of uncertainty.

Potential sources of uncertainty in assessment estimates, which may or may not induce retrospective error, are model structural uncertainty, including year effects and year-class effects in the data (for example, changes in selection-at-age or year-class effects in separable models) and consequent, uncertainty in the model specification. Studies relating to year-class effects in the separable model ISVPA (Section 3.3.3) and the sensitivity to some options in XSA (Section 3.3.4) were presented to WGMG this year.

The analysis ISVPA demonstrated that for simulated data sets, taking year-class effects into account resulted in a substantial improvement in the fit of the model to the catch-at-age data with, for some model formulations and improvement in the ability of the model to estimate the *truth*. For the real data case study examined the model estimated strong historic year-class effects but resulted in only a relatively minor modification to the perception of the historic stock dynamics. Year effects in selection have been substantial in fisheries such as the North Sea haddock and the Western horse mackerel. If the TISVPA model formulation can be used to predict changes in selection as strong year-classes pass through the fishery it may enable improvement in the utility of stock projections.

Bias in the catch-at-age data has a significant effect on assessment estimates and has previously been shown to cause retrospective patterns in which the end points of the assessment derived from survey information are often more reliable than the historic estimates derived from the catch-at-age data. The studies reported in respect of ToR a) have shown that where recent catch data are considered unreliable, and unbiased survey information is available, there is sufficient information contained within the model structure to be able to estimate catches. The methods require an overlap between the survey time series and a time period in which the catches are unbiased in order to scale the population and recent catch estimates. In Section 3.3.4 the correction applied to the catch data was shown to be capable of removing the retrospective pattern resulting from that particular data bias. In this case the historic assessment estimates are considered unbiased and the data scaled to *correct* the more recent model fits. The corrections applied to the catch-at-age data are only one of a number of data or model parameter adjustments that could have been applied to correct the retrospective patterns. Changes could have been modelled in the time series natural mortality or survey catchability. Such model structural uncertainty can only be resolved by reference to additional information.

More generally, the cause of retrospective error must be sought either in conflicting signals in the data or in model misspecification in a broad sense. To create a retrospective error, model deviations must have a different impact on the estimates of the past history and the present, i.e. a step or trend such that the data and model specifications that largely determine the estimates of the past are in conflict with those determining the estimates in the last few years. The analysis described in Section 3.3.5 provides a suitable tool for the diagnosis of the potential causes of retrospective bias and also can give an indication of the relative magnitude of the changes in the input data that would be required to provide consistency in model estimates. This would enable stock analysts to decide on the most likely causes of the problem. In order to take the analysis further its utility must be tested. Further analysis by WGMG should consider stocks where retrospective errors occur, as well as artificial data. Studies on the real stocks should include diagnostics, and review information on the strengths and weaknesses of the data. Artificial data should be used as a bench test to confirm or reject hypotheses that emerge from the study of the real data. Such studies are therefore suggested as possible future work for a meeting of the WGMG.

4 Reference points and harvest control rules

ToR d) evaluate, test and review developments in computer software for routine application in stock assessment that are presented to ICES.

ToR e) consider how best to update limit reference points and develop target reference points for use within long-term management strategies.

4.1 Introduction – background and context

Within the MoU between the European Community (EC) and the International Council for the Exploration of the Sea (ICES), ICES has been tasked with developing the form of its advice in the context of the Common Fisheries Policy (CFP). The advice shall reflect policy developments under the CFP and the Marine Strategy such as emphasis on an ecosystem approach, long-term management plans and a fleet-based approach to mixed fisheries management. The advice should also be given in the context of international agreements and guidelines notably the World Summit on Sustainable Development (WSSD) implementation plan and the Precautionary Approach (PA). Whenever international agreements and guidelines have been translated into horizontal policies and adopted by, for example, the EC (such as the WSSD implementation agreement) the advice will relate to these.

Advice on long-term management strategies

For fish stocks, or for mixed fisheries, where long-term management plans have not been agreed or where agreed management plans are not considered to be in accordance with international agreements and guidelines, ICES shall present options for long-term management plans. These plans shall relate to fish stocks and, where mixed fisheries are important, also to fisheries. The strategies shall have a low risk of reducing stock productivity, be associated with stable high long-term yields and be developed within an ecosystem approach in the context of the CFP, the PA and WSSD targets. Such options presented by ICES will be the basis for a dialogue which eventually will help the Commission in proposing long-term management plans. This advice shall be part of the annual advice.

It is with this in mind that WGMG has been requested by ACFM to consider how best to update limit reference points and develop target reference points. The background document (BE2) provides the background, and context, to the request.

Reference points

Biological reference points are an integral part of single species management advice frameworks and are used to provide advice on safe biological limits and targets.

Within ICES, the precautionary and limit reference points have been set or suggested through the process of SGPA, ACFM and sometimes by dedicated study groups (SGBRP – ICES 2003). It has been unclear what communication process has been involved in accepting these and who would *own* these reference points - science, management, or stakeholders. In practice, precautionary and limit reference points are often both attributed to ICES, whereas the wording of the ACFM advice suggests that ICES is only responsible for the limit points and that the precautionary points are only presented as suggestions that can be agreed by management authorities if they accept the risk that is associated with these.

Fishing mortality reference points such as F_{MSY} (or in practice, proxies such as $F_{0.1}$ and F_{max}) are used to define targets and F_{crash} (or proxies, F_{loss} and F^*) to define limits. Limits and targets can also be defined for biomass; for example, the SSB that would support the maximum yield (B_{MSY}), can be used as a target and the point at which recruitment declines as a limit (ICES CM 2004/D:03).

The developments towards long-term advice and the ecosystem approach to fisheries management give rise to a new way of thinking about reference points. Results of AGLTA 2005 indicate that the precautionary reference points may not be sufficient, in themselves, for the provision of long-term advice on rational management strategies that will be based on conservation limits to avoid (which are expected to refer to biomass, the old B_{lim}, but including considerations of ecosystem components such as predation and food) and targets to achieve (which are expected to refer to exploitation rate). Management strategies which should achieve productive stocks, a healthy ecosystem and high long-term yield do not appear to be sensitive to decision trigger points such as B_{pa} and F_{pa} in the earlier precautionary approach framework.

Accounting for assessment uncertainty and error in the implementation of management procedures must be an integral part of choosing biological reference points for use in harvest control rules. In 2004, WGMG (ICES CM 2004/D:03) suggested that **reference points are better developed within the context of a management procedure and can be regarded as parameters in the harvest control rules**, which are *tuned* (c.f. IWC 1993) to achieve management objectives. It is recognised that there is a wide range of possible rules, which will require different kinds of parameters. WGMG has not changed its views on this aspect.

There is thus an apparent need to estimate, or revisit the estimates of, B_{lim} taking into consideration not only the reproductive capacity of the stocks and changes in productivity within alternate environmental scenarios, but also their importance as food or predators in the ecosystem. It is important to include regime shift considerations and the multi-species interactions in such updates. In addition, there is a need to develop approaches for long-term target reference points which are consistent with the WSSD (e.g. F_{MSY} based reference points).

Clearly, there may be a need to discuss and establish a process for such estimating and updating of reference points after the management strategy evaluation framework has matured from 2007 onwards.

At the last meeting of WGMG held in 2004, the group addressed a ToR to identify appropriate estimators of stock conservation limits and reference points relating to longer-term potential yield; together with a characterisation of their statistical properties for the range of stocks currently assessed by ICES for its client customers and related management agencies (EU, IBSFC, NAFO, NASCO, NEAFC, and ICCAT). Note, however, that regime shifts and multi-species considerations were not explicitly considered as part of these estimators.

As intimated earlier, there is a growing appreciation that the PA points are not sufficient, in themselves, for the provision of long-term advice on rational management. In particular, **there is a strong need to establish targets, in addition to these conservation reference points**. It is also recognised that targets cannot be seen in isolation from the management framework and objectives. In order to serve management objectives, more elaborate harvest rules than single target reference points may be needed. Hence, targets should not be regarded as single values of fishing mortality or biomass, for example, but rather as parameters in harvest rules. Harvest rules at various levels of sophistication, and with various legal status have already been established for several stocks. Such harvest rules need to be evaluated with respect to the PA, as well as for their performance in relation to management objectives.

ICES has identified candidates for reference points which are consistent with taking high long-term yields and achieving a low risk of depleting the reproductive potential of stocks in the range $F_{0.1}$ to F_{max}. Further, in 2005, the EC produced a non-paper entitled *Implementing sustainability in EU fisheries: Strategies for growth and employment* in which numerical values were assigned to these fishing mortalities for fish stocks in the North-east Atlantic and adjacent waters. There is little more that WGMG can add to this proposal but to comment that the choice of target fishing mortality is not a solely scientific activity but will necessitate agreement amongst managers and the fishing industry. Any such proposals can be evaluated by fisheries scientists through simulation modelling of management strategies and HCRs.

Building on the previous work initiated at the last meeting of WGMG in 2004, this meeting of WGMG has further elaborated upon their approach for evaluating management strategy proposals and harvest control rules (HCRs) which must necessarily include the establishment of targets and conservation reference points.

4.2 Management strategy evaluation framework

Introduction

The development of fisheries management strategies is a long and complex process where many fisheries managers, politicians, stakeholders and scientists participate. The report of the SGMAS (ICES CM 2006/ACFM:15) outlines the various aspects of this process. The evaluation of HCRs is part of the process of developing and evaluating management strategies. The SGMAS report gives some guidelines for the evaluation of harvest control rules, including quality standards for simulation programs and a listing of software that is available to the ICES community. The present Section 4.2 is an extension of the SGMAS work, giving more specific operating guidelines for the procedure of evaluation through simulations.

An HCR is the part of a management strategy that defines the procedure for making regular tactical decisions; i.e. to decide on the actual regulations of the removals from the stock in the coming year. In general, management strategies and HCRs, in particular, have to be adapted to the prevailing conditions. Hence, there is no universal recipe either for the design or for the evaluation of an HCR. Adaptations are needed because of the diversity in management objectives, in the information that is available and how that can be verified, in the criteria that form the basis for management tactical decisions, in the instruments that managers can use to implement the strategy, and in the need to adapt to the biological characteristics of the ecosystem.

The Figure 4.2.1 shows the main building blocks in a harvest rule simulation framework.

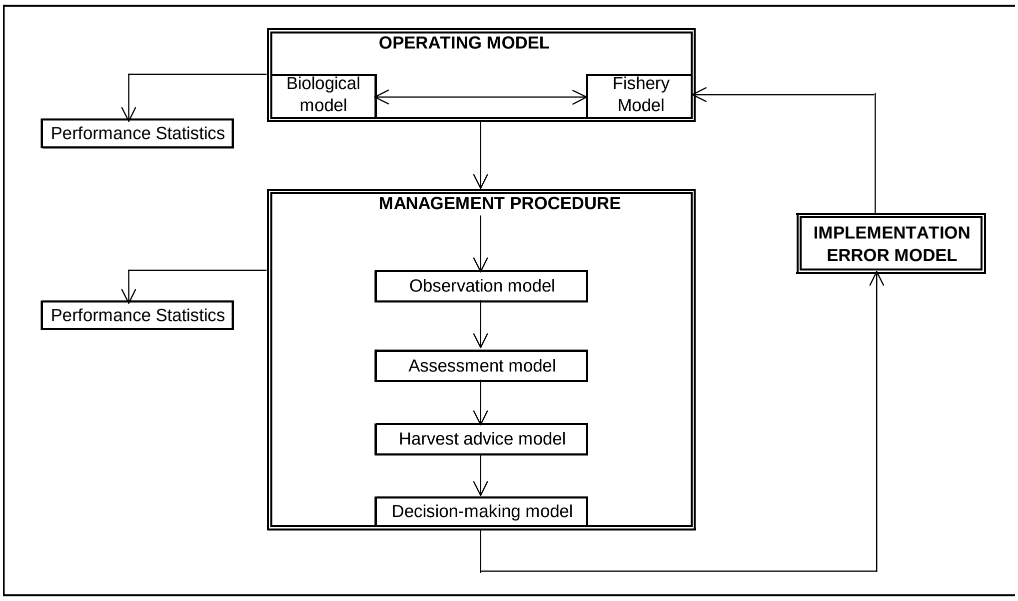


Figure 4.2.1: Conceptual framework for the evaluation of management procedures, recovery plans and harvest control rules.

An evaluation of a harvest rule as proposed, or agreed, by managers is wider than just simulations of how the stock is expected to respond when the rule is implemented. It also includes an evaluation of the rule as such with respect to internal consistency, completeness and the compatibility of objectives with the PA. Note that this is a further development of the conclusion reached at the last meeting of WGMG in 2004 when the group considered that … *the framework comprises everything that is needed for conducting the simulations* (ICES, 2004a).

WGMG proposes the following three-step procedure for translating management proposals into simulation models:

> **Step 1** - Translate the decision process in the management proposal into a structured decision diagram. This will serve as a basis for the design of simulations and reveal holes and ambiguities in the managers' proposals.

Step 2 - Design and document the algorithm for expressing the management procedure model and if necessary, the operating model in a simulation program.
Step 3 - Carry out the necessary procedures for quality checking and documenting that the code produces the correct results.

A schematic representation of this procedure is shown in Figure 4.2.2. However, there may be aspects of a management proposal such as constraints on the level of by-catch species which may be difficult to evaluate solely through a *research* approach.

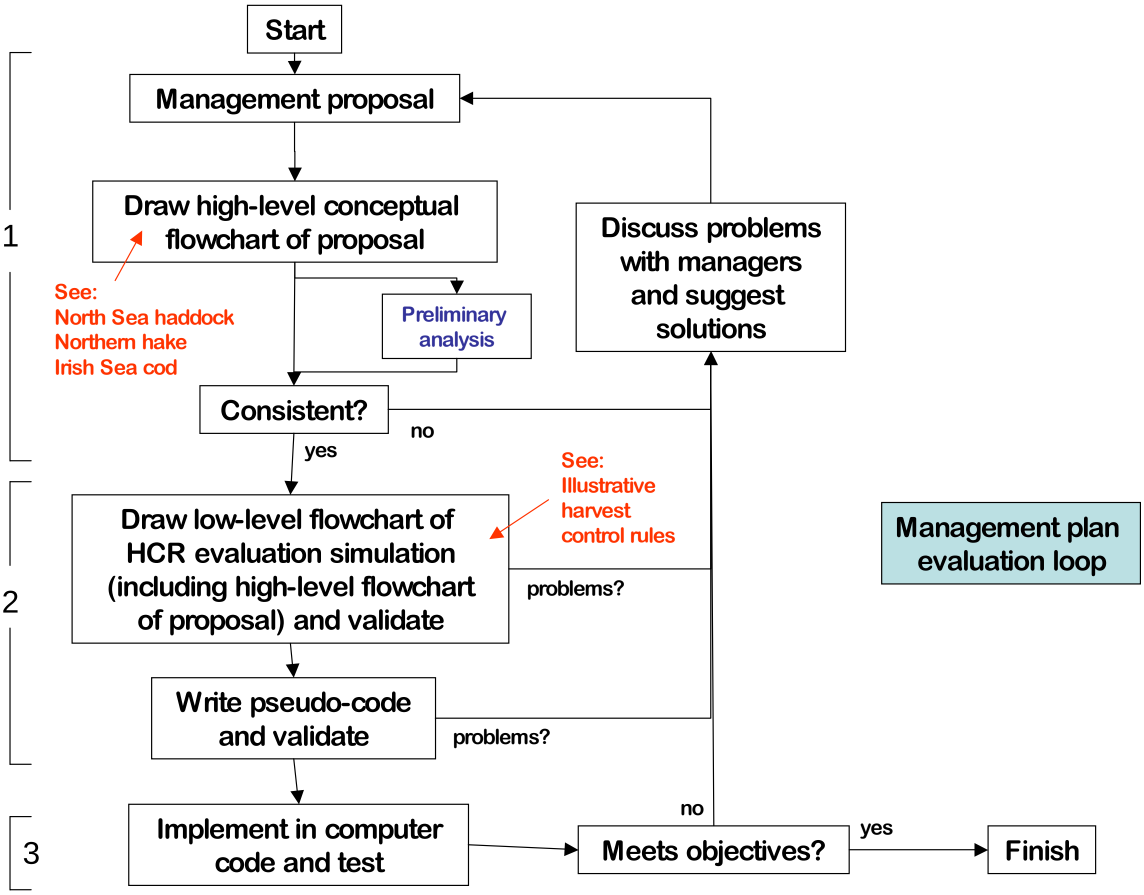


Figure 4.2.2: Schematic representation of the three-step procedure for evaluating harvest rules.

Apart from representing a systematic approach to the design of simulations for HCR evaluation, such a process will aid in providing the necessary documentation to allow a proper evaluation of the work that is done. Considering each of the three steps in turn:

Step 1 - Translate the decision process in the management proposal into a structured decision diagram: An HCR (and a management strategy in general) has to be evaluated against performance in relation to objectives. Part of the evaluation process therefore is to identify such objectives. Often, they are not clearly stated, but must be inferred from the context and from formulations in the proposal. At earlier stages of development of HCRs, an important contribution from science may be to outline the trade-off between objectives, in order to assist managers in their task of developing objectives that can be a satisfactory compromise between conflicting interests.

When advice is given by ICES in response to requests by managers, ICES will assume compatibility with the PA as an implicit condition, and use that as a constraint on what ICES can recommend. Hence, **simulations will have to show the risk that the stock falls outside bounds set by the PA**.

A complete harvest control rule has to define uniquely the actions to be taken for the whole range of values of the parameters that are the basis for the decisions. For example, a common shortcoming is that a proposed rule does not specify precisely the action to be taken if the SSB falls below some trigger level, but rather has formulations like '*the fishing mortality rate .., shall be adapted in the light of scientific estimates of the conditions then prevailing*'. In such cases, the role of science should be to outline sensible options to fill-in these holes, and pass that back to managers, as part of the dialogue process.

Such incomplete specifications include the definition of *appropriate age groups* for the calculation of fishing mortalities in an HCR and the decision on whether SSB-based rules apply either to the latest estimate in the year of assessment or the forecast in the year following the assessment year.

Step 2 - Design and document the algorithm: At the stage where the algorithm is designed and documented, scenarios with respect to biology need to be considered. For example, for a stock with occasional very large year-classes, like North Sea haddock, the performance of the rule both in the presence, and absence, of a large year-class will have to be considered. Other examples are density dependent growth and maturity, possible variations in discarding practise (again North Sea haddock is an example), variations in productivity (e.g. Blue whiting), and migrations into or out of the management area (e.g. Icelandic and Greenland cod).

At this stage one will also have to consider, design and document the generation of information that is the basis for the tactical decisions, from the model stock. This can be done either so that this information reflects the state of the stock with some uncertainty, which is assumed, or one may simulate the process providing the basis for decisions. In the first case, the simulations will become a test of the robustness of the rule to this uncertainty, and one would require that the rule is robust, at least, to the level of uncertainty that seems likely. In the second case, one would typically derive noisy input data to an assessment from the population model and perform the assessment as part of the simulation.

Whenever processes are modelled in the simulation, verification is necessary, i.e. it has to be confirmed that the process model reproduces what has previously been encountered in nature. For example, stochastic recruitments that go into the population model should have a distribution compatible with the distribution of historic recruitments under comparable conditions, unless the purpose is to show the effect of alterations in the recruitment dynamics. Likewise, when modelling the assessment process, it needs to be verified that historically encountered bias and uncertainty are reproduced.

Step 3 - Quality checking and standards for documenting the program code: Software code will have to be developed that implements the HCRs under scrutiny. The diversity of proposed tactical decision procedures is large and it is not realistic to have a single generic program that would cover the whole range of managers' creativity. Likewise, the biological properties of the stock and the kind of data that may be available will vary considerably.

Nevertheless, parts of a simulation program will not need such adaptations. A strictly modular program structure, apart from being good programming practise, will allow using well controlled code for the parts that do not need to be programmed *ad hoc*. Further details about code documentation and testing are presented in the following Section 4.3.

Three illustrative examples are presented in order to demonstrate the proposed approach discussed above – North Sea haddock, northern hake and Irish Sea cod. Each of these is a single species application and would clearly only represent a first step towards the development of long-term management strategies.

4.3 Model and software validation, verification and testing

The development of HCRs and similar algorithms will need to be supported by the adoption of various procedures and protocols for the verification and testing of these algorithms and their software implementations. The process of incremental testing of fisheries software by repeated use and peer review of results needs to be strengthened by the development of testing and verification protocols, and the tools implementing them. This will be especially important if, as expected, the move towards simulation-based evaluation of management strategies is applied more widely, and the adoption of flexible tools such as the FLR platform, result in an increasing number of models and model implementations.

Three levels of model quality and usability assessment are commonly recognised (Balci, 1995): validation, verification and testing. The first refers generally to the conceptual and mathematical level, and the ability of the model to capture the dynamics of the real system. Verification refers to the translation from the conceptual to the tangible, and how well the algorithm and its implementation reflect the intended behaviour. Finally, testing refers to the correctness, completeness, security and quality of the software code. There are many approaches to software testing and verification, but effective testing of complex products is essentially a process of investigation, not merely a matter of creating and following a single procedure.

4.3.1 Software testing

Testing of the software implementing HCRs, and in general of all fisheries models and algorithms, is essential if their results are to be applied in management. Writing and executing testing code should be an essential part of the development process. Structured testing suites are a useful tool during the development process, especially as code complexity increases. Use of software testing techniques is also likely to increase the confidence in the quality of the code. A number of testing paradigms exist, and some of them are more suited to certain languages or coding pradigms. For example, unit testing is commonly used when programming (http://en.wikipedia.org/wiki/Unit_testing) using Object Oriented Programming (OOP).

The adoption of open source licenses can also have a positive impact in software quality. Peer review of the source code is not only possible but actively encouraged in the open source paradigm. This process would also be helped by the widespread use of a common language; for example, the open-source statistical programming environment R (R Development Core Team, 2004).

4.3.2 Model verification

Once a piece of software implementing a certain model has been tested, and confidence is high in that it is a true reflection of the algorithm of interest, computerised model verification can be carried out on it and the results can be used to assess both the code itself and the underlying model behaviour. A number of formal and informal procedures can be applied to the task of verifying HCR software implementations. For example, manual calculation of the result that the HCR should give for a limited set of input values. The outcome of an HCR for input values beyond and below the biological limits used in its formulation should be known, at least qualitatively. Test runs under these extreme conditions can provide a first check of the HCR software. Algorithm and code can then be tested by inputting extreme values, or unlikely combinations of input values. The code should give appropriate warnings or errors when input values are incorrect, rather than return an inappropriate answer, and again this can be incorporated into the test suite.

Formal verification methods of interest in this application are, amongst others:

> Tracing: A given variable is traced along the code and its value at certain points is output. Again, expected values for a given input set should be determined and compared with those obtained by tracing. Many programming languages offer tracing features, such as the trace() function in R.
>
> Consistency checks: When the same procedure is applied to the same input set, results should be consistent, up to a given tolerance if stochasticity is introduced. Repeated runs with small alterations in the input values can also provide a good measure of the stability of the code. This is especially useful when sensitive non-linear minimisation routines are part of the procedure.
>
> Re-programming of critical components: Essential components in a piece of code can be reprogrammed, ideally by a different programmer, and the results of both approaches can then be compared. Different coding styles mean algorithms are seldom coded in the same way by different individuals, but results should not be influenced by the internal procedures used in each case. This is obviously a rather cumbersome technique, reserved for critical functions that cannot be easily verified in any other way.

Other techniques, together with a more complete analysis of the ones presented here, can be found in Kleijnen (1995), Rykiel (1996) and Sargent (2005).

4.3.3 Model validation

Validation is a demonstration that a model, within its domain of applicability, possesses a satisfactory range of accuracy consistent with the intended application (Sargent, 2005). This definition, when applied to a HCR, implies the verifying whether the HCR achieves the objectives for which it was designed, and that it does so under a range of possible scenarios. Validating the HCR model could thus be considered as a primary objective of the whole HCR evaluation process. Previous verification and validation of the others components of the simulation framework is obviously an essential requirement, and for that task some of the methods outlined above will be of use. Once the simulation framework is considered mature and suitable for application, results obtained for a given HCR can be considered to reflect its behaviour and properties, albeit conditioned on the representation of reality implied by the simulation framework.

Specification of the performance criteria is an important step in the evaluation of an HCR. Not only those implicit, the achievement of the objectives that motivate it, but also in terms of the range of scenarios and uncertainties the HCR should perform well under. The robustness of a HCR to known or expected uncertainties should be assessed as part of the simulation procedure, and considered an essential part of the validation procedure

4.3.4 Documentation

A complete model and software validation exercise requires that sufficient documentation covering the objectives, performance criteria and model context must be available. Illustrative examples of a number of HCRs can be found in Section 4.4. The various steps in creating and analysing an HCR are illustrated by means of pseudo-code (a code-like test that describes the algorithm implementing the HCR), flow-charts and actual software implementations.

The regulations contained in management plans should ideally be amenable to their specification as an HCR, that is, a programmable algorithm, if it is to be assessed in a simulation framework. This translation process is likely to uncover contradictions or other anomalies, and this can be considered as an essential step in validation of management plans. A dialogue can then be established with managers about the ambiguities or contradictions uncovered during this process.

4.4 Case studies

Three illustrative examples are presented in order to demonstrate the proposed approach discussed above – North Sea haddock, northern hake and Irish Sea cod. Each of these is a single species application and would clearly only represent a first step towards the development of long-term management strategies. Ultimately, multi-species and mixed fisheries aspects would need to be considered. These are briefly alluded to in Section 4.5.

4.4.1 North Sea haddock

Haddock in Sub-area IV (North Sea) and Division IIIa (Skagerrak-Kattegat)

This stock is exploited by both EU member states and Norway, and is thus managed as a shared stock.

4.4.1.1 Management plan, decision tree and critical points

In 1999 the EU and Norway have *'… agreed to implement a long-term management plan for the haddock stock, which is consistent with the precautionary approach and is intended to constrain harvesting within safe biological limits and designed to provide for sustainable fisheries and greater potential yield.'*. The agreement was updated in November 2004.

The plan shall consist of the following elements:

1) *Every effort shall be made to maintain a minimum level of Spawning Stock Biomass (SSB) greater than 100,000 tonnes (B_{lim}).*

2) *For 2005 and subsequent years the Parties agreed to restrict their fishing on the basis of a TAC consistent with a fishing mortality rate of no more than 0.30 for appropriate age groups.*

3) *Should the SSB fall below a reference point of 140,000 tonnes (B_{pa}), the fishing mortality rate referred to under paragraph 2, shall be adapted in the light of scientific estimates of the conditions then prevailing. Such adaptation shall ensure a safe and rapid recovery of SSB to a level in excess of 140,000 tonnes.*

4) *In order to reduce discarding and to enhance the spawning biomass of haddock, the Parties agreed that the exploitation pattern shall, while recalling that other demersal species are harvested in these fisheries, be improved in the light of new scientific advice from inter alia ICES.*

5) *A review of this arrangement shall take place no later than 31 December 2006.*

6) *This arrangement enters into force on 1 January 2005.*

ICES considers that the agreed PA reference points in the management plan are consistent with the PA, provided they are used as lower bounds on SSB, and not as targets.

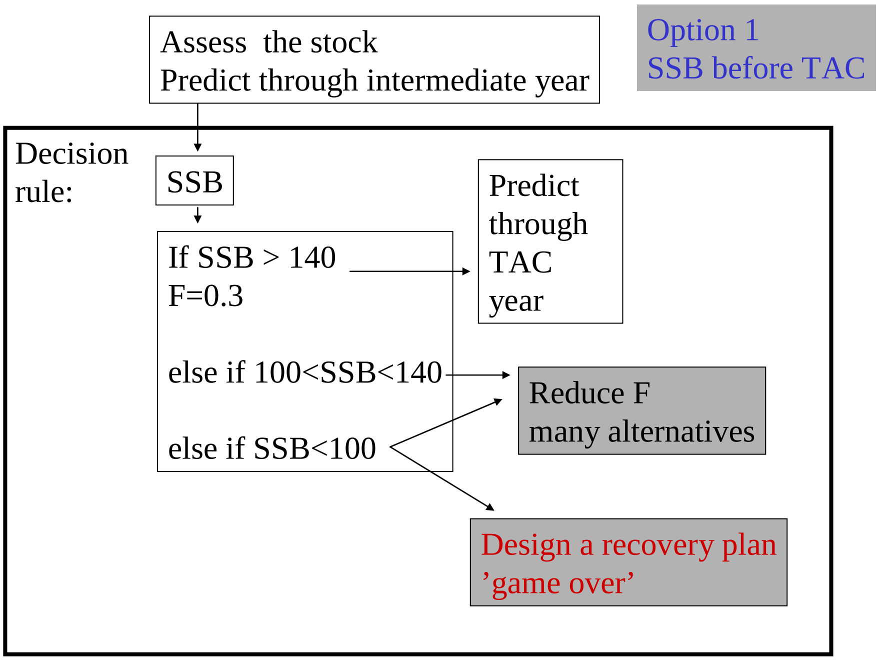
The agreed harvest rule for North Sea haddock does not specify all the details that are needed to evaluate the performance of the elements within the management plan by simulation. At this meeting, WGMG has attempted to interpret the rule by constructing decision charts. This is done partly to translate the rule into programmable algorithms, and partly to identify ambiguities in the management plan.

The formal objective of the management agreement is to maintain the spawning stock biomass above the B_{lim} of 100 kt. Other objectives are not stated but it is noted that the HCR is in principle a rule for setting a total allowable catch (TAC) according to an agreed fishing mortality which is kept constant unless the stock becomes smaller than acceptable.

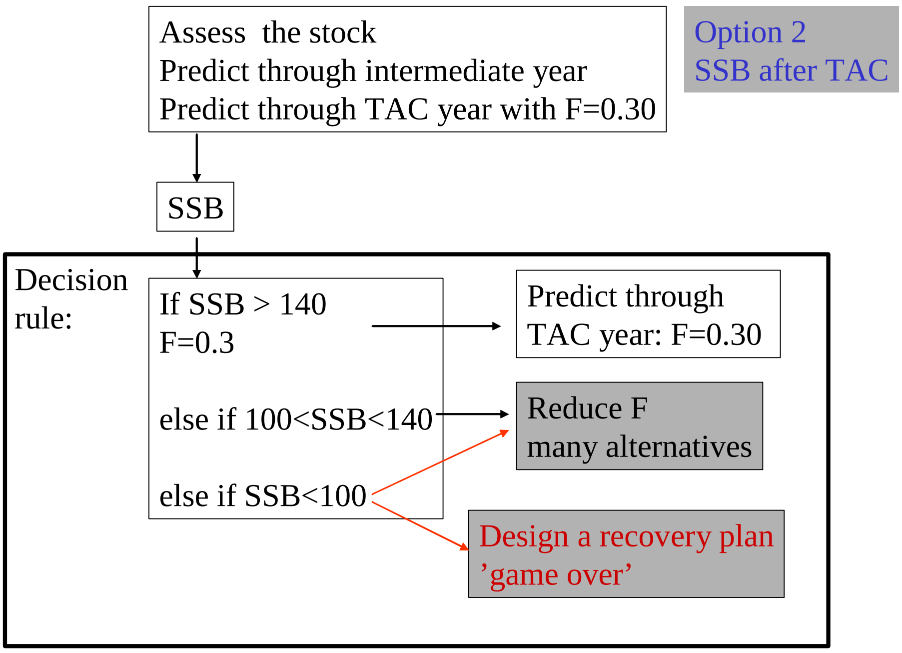
The basis for decisions in the rule is the spawning stock biomass (SSB). It is not stated, however, in which year this SSB refers to. WGMG has considered two possible interpretations.

As examples of the high-level conceptual flow charts for the decision process (namely, step 1 in Figure 4.2.2), two alternatives – denoted options 1 and 2, depending on whether the reference SSB is the SSB before, or after, the TAC year are presented. Elements that would require further clarification are shaded in grey and would require discussions with managers.

Option 1: the rule refers to SSB at the start of the TAC year.



Option 2: the rule refers to the SSB resulting after the TAC has been taken; i.e. at the end of the TAC year.



The former (option 1) represents a decision based on the state of the stock available for the fishery in the quota year; the latter (option 2) represents a decision based on the consequence of the agreed fishery management. The choice has implications for how to achieve the agreed fishing mortalities, and subsequently for the performance of the management plan.

In the situation where F is to be reduced because the SSB is below the *trigger* biomass, the management plan does not specify exactly what action is to be taken. However, the agreement is to '*… ensure a safe and rapid recovery of SSB to a level in excess of 140,000 tonnes.*'

There are several possible ways of dealing with this situation. One alternative is a rule for reducing the F, in a way that has been demonstrated through simulation to lead to rebuilding of the stock with a high probability. Another alternative is to reduce the catch to a level which leads to SSB > 140 kt within one or more years. The term *safe* rebuilding would imply reaching the rebuilding target with a specified high probability taking uncertainty both in assessments and predictions into account. This may require a stochastic stock projection.

Finally, the rule does not specify the action to be taken if SSB falls below B_{lim}. One may suggest an extension of the rule to cover that situation but alternatively, one may argue that in such a case the rule in practise has not been able to meet the main objective, and abandon the management plan altogether. Alternatively, stricter management measures may be needed and ultimately, imposed.

4.4.1.2 Illustrative harvest control rules and pseudo code

The existing management plan for North Sea haddock is under review this year (see Section 4.4.1.1). The working paper (WE1) presented a method for evaluating a number of different management types and harvest control rules, based on the FLR framework and coded in the R system. Although some refinement of growth, recruitment and discard models is still required, tentative conclusions were reached on the likely efficacy of different rules. **A full evaluation of management plans for haddock was not given, as this can only be done in consultation with relevant fisheries managers and stakeholders**, so the paper should be viewed as an exploration of methodology rather than a finished analysis. In particular, the incorporation of the recruitment dynamics for the haddock stock into realistic simulations will necessitate further elaboration; i.e. the details of how to generate sporadic strong year-classes will need to be further elaborated and simulation procedures validated.

WGMG commented, however, that this was a useful step in the process of developing an approach to evaluating management plans. Comparisons were drawn with the F-PRESS approach, in which implementation bias/error (which is not yet included in the FLR method) has been shown to be very important. F-PRESS also includes a more efficient method of searching for *F*-multipliers than the `optimise` function used in the FLR implementation, and the utility of this method needs to be explored. Finally, WGMG pointed out that effort management cannot implement a fixed-TAC harvest control rule: this should be relabelled as a *target yield* HCR instead.

Figure 4.2.2 presents a general management plan evaluation loop. The second and third levels of this loop are the planning and implementation of computer code to carry out the evaluation, following an additional three-step process:

i) Draw a flowchart of the algorithm to be used. Validate the algorithm using this flowchart.

ii) Convert the flowchart into pseudo-code and further validate.

iii) Implement in computer code and test using a suitable range of input data.

These steps are largely complementary and could be collapsed together, or reversed, depending on the preference of the analyst.

Figure 4.4.1.2.1 gives an example algorithmic flowchart of a possible HCR for North Sea haddock, namely a target *F* with a limit of inter-annual change in the TAC, assuming that the available management instrument is catch-based (rather than effort-based). In Figure 4.4.1.2.2 this has been converted to pseudo-code, while Figure 4.4.1.2.3 gives an extract from an implementation in the R language (using FLR objects and methods).

Examples of the output that can be generated by this code are given in Figures 4.4.1.2.4—4.4.1.2.6, taken from Needle (WE1). These illustrative plots model a target *F* of 0.3, with an inter-annual limit on TAC change of 15%, and demonstrate the limitations on achieving the target *F* that arise through the imposition of a TAC change limit. It must be emphasised that these plots should not be interpreted as a definitive evaluation of any particular management plan for North Sea haddock: that can only be done via the feedback to managers systematised in Figure 4.2.2.

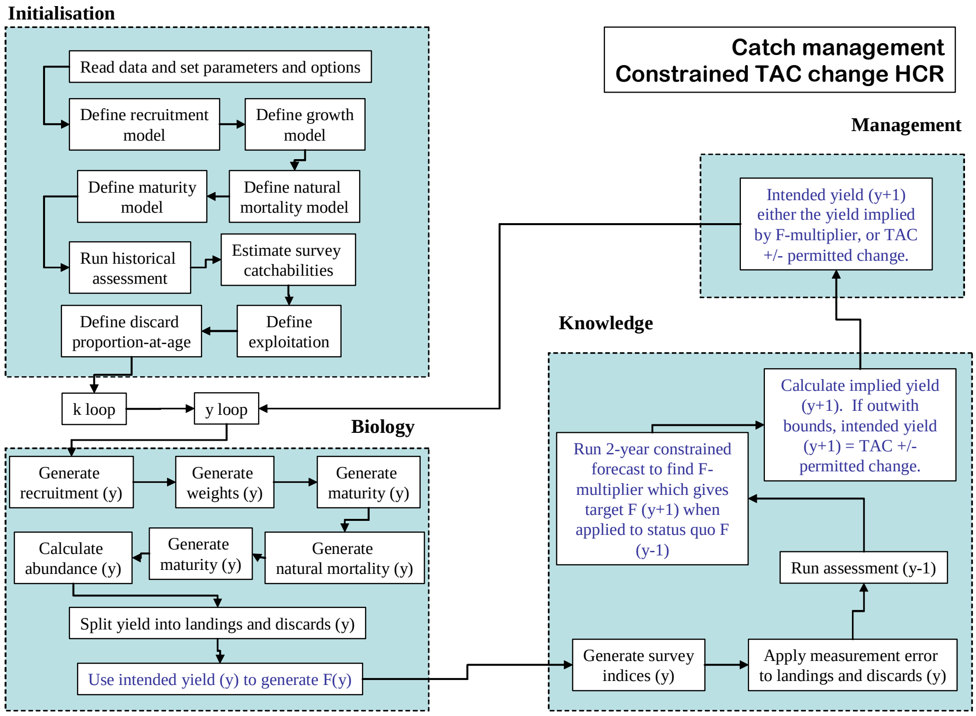


Figure 4.4.1.2.1: Algorithmic flowchart for a possible HCR for North Sea haddock (target *F* with constrained TAC change).

```
# Pseudo-code example: catch management, TAC change limit HCR

# Notes: y = 1 is the first assessment year

read data and set parameters and options
define recruitment model
define growth model
define natural mortality model
define maturity model
estimate discard proportion-at-age
estimate selection-at-age

loop over k iterations:
        loop over y years:
                generate recruitment(y)
                generate weights(y)
                generate natural mortality(y)
                generate maturity(y)
                if y = 1 then
                        true f(y) <- mean(f(y-3),f(y-2),f(y-1))
                else
                        estimate f(y) such that landings(y) <= intended landings(y)
                                and n(a,y) > 0 for all a
                n(y) <- n(y-1) * exp(-f(y-1)-m(y-1))
                true catch(y) <- catch equation applied to f(y) and n(y)
                split true catch(y) into true landings(y) and discards(y)
                generate surveys(y)
                generate measurement error on catch(y), landings(y) and discards(y)
                run assessment up to and including y-1
                if y = 1 then
                        true f(1:y-1) and n(1:y-1) <- assessed f(1:y-1) and n(1:y-1)
                if y = 1 then intended f(y) <- true f(y)
                if y = 1 then intended landings(y) <- true landings(y)
                run forecast assuming f(y+1) = target f
                if intended landings(y+1) > intended landings(y) + dTAC%
                        or intended landings(y+1) < intended landings(y) - dTAC%
                then
                        intended landings(y+1) = intended landings(y) +/- dTAC%
                else
                        intended landings(y+1) from forecast
                intended f(y+1) = target f
        end y loop
end k loop
```

Figure 4.4.1.2.2: Pseudo-code for a possible HCR for North Sea haddock (target *F* with constrained TAC change)

```
for (k in 1:num.k)
{
	for (y in first.year:last.year)
	{
		# True recruitment
		if (y == first.year)
			{recruit.sigma[,ch(y),,,] <-
				recruit.rnd[[k]][,ch(y),,,]}
		else
			{recruit.sigma[,ch(y),,,] <-
				(recruit.phi * recruit.sigma[,ch(y-1),,,]) +
				recruit.rnd[[k]][,ch(y),,,]}
		x.list[[k]]@stock.n[1,ch(y)] <-
			min(recruit.ceiling, recruit * exp(recruit.sigma[,ch(y),,,]))

		# True weights, natural mortality, maturity
		x.list[[k]]@catch.wt[,ch(y)] <- x.list[[k]]@catch.wt[,ch(y-1)]
		x.list[[k]]@landings.wt[,ch(y)] <- x.list[[k]]@landings.wt[,ch(y-1)]
		x.list[[k]]@discards.wt[,ch(y)] <- x.list[[k]]@discards.wt[,ch(y-1)]
		x.list[[k]]@stock.wt[,ch(y)] <- x.list[[k]]@stock.wt[,ch(y-1)]
		x.list[[k]]@m[,ch(y)] <- x.list[[k]]@m[,ch(y-1)]
		x.list[[k]]@mat[,ch(y)] <- x.list[[k]]@mat[,ch(y-1)]

		# Fishing mortality
		if (y == first.year)
			{x.list[[k]]@harvest[,ch(y)] <-
				apply(as.table(x.list[[k]]@harvest[,ch((y-3):(y-1))]),
				1, mean)}
		else
			# Estimate multiplier lambda on selection that gives N(a) > 0
			# for all a, and landings yield <= predefined TAC limit
			{tac.limit <- as.numeric(x.list[[k]]@landings[,ch(y),,,])
			for (i in seq(0,10,by=0.1))
				{lambda.test <- lambda.for.tac(i, selection, 0,
					tac.limit, prop.landings, x.list[[k]], y)
				if (lambda.test == 1.0e20)
				{max.i <- i
				break()}}
			lambda.fit <- optimise(lambda.for.tac, interval = c(0,max.i),
				sigma = selection, limit = 0,
				tac = tac.limit, prop = prop.landings,
				wk.x = x.list[[k]], wk.y = y, tol = .Machine$double.eps)
			lambda <- lambda.fit$minimum
			x.list[[k]]@harvest[,ch(y),,,] <- lambda * selection
			x.list[[k]]@catch.n[,ch(y),,,] <-
				(x.list[[k]]@harvest[,ch(y),,,] *
				x.list[[k]]@stock.n[,ch(y),,,] *
				(1 - exp(-x.list[[k]]@harvest[,ch(y),,,] –
				x.list[[k]]@m[,ch(y),,,])) /
				(x.list[[k]]@harvest[,ch(y),,,] +
				x.list[[k]]@m[,ch(y),,,]))
			x.list[[k]]@landings.n[,ch(y),,,] <-
				x.list[[k]]@catch.n[,ch(y),,,] * prop.landings
			x.list[[k]]@discards.n[,ch(y),,,] <-
				x.list[[k]]@discards.n[,ch(y),,,] * prop.discards
			x.list[[k]]@catch <- apply(x.list[[k]]@catch.n *
				x.list[[k]]@catch.wt, 2, sum)
			x.list[[k]]@landings <- apply(x.list[[k]]@landings.n *
				x.list[[k]]@landings.wt, 2, sum)
			x.list[[k]]@discards <- apply(x.list[[k]]@discards.n *
				x.list[[k]]@discards.wt, 2, sum)}

		# True stock numbers: N(t+1) = N(t) * exp(-Z(t))
		n0 <- x.list[[k]]@stock.n[1:(num.ages-2),ch(y-1),,,]
		f0 <- x.list[[k]]@harvest[1:(num.ages-2),ch(y-1),,,]
		m0 <- x.list[[k]]@m[1:(num.ages-2),ch(y-1),,,]
		x.list[[k]]@stock.n[2:(num.ages-1),ch(y),,,] <- n0 * exp(-f0 - m0)
```

Figure 4.4.1.2.3: R code for a possible HCR for North Sea haddock (target *F* with constrained TAC change). For brevity, only the principal loops are given here: initialisation and summary code is also required (continued over).

```
# True stock numbers for plus-group
               x.list[[k]]@stock.n[num.ages,ch(y),,,] <-
                      (x.list[[k]]@stock.n[(num.ages-1),ch(y-1),,,] *
                      exp(-x.list[[k]]@harvest[(num.ages-1),ch(y-1),,,] –
                             x.list[[k]]@m[(num.ages-1),ch(y-1),,,])) +
                      (x.list[[k]]@stock.n[num.ages,ch(y-1),,,] *
                      exp(-x.list[[k]]@harvest[num.ages,ch(y-1),,,] –
                             x.list[[k]]@m[num.ages,ch(y-1),,,]))

               # True catches
               x.list[[k]]@catch.n[,ch(y),,,] <- (x.list[[k]]@harvest[,ch(y),,,] *
                      x.list[[k]]@stock.n[,ch(y),,,] *
                      (1 - exp(-x.list[[k]]@harvest[,ch(y),,,] –
                             x.list[[k]]@m[,ch(y),,,])) /
                             (x.list[[k]]@harvest[,ch(y),,,] +
                             x.list[[k]]@m[,ch(y),,,]))
               x.list[[k]]@landings.n[,ch(y),,,] <- x.list[[k]]@catch.n[,ch(y),,,] *
                      prop.landings
               x.list[[k]]@discards.n[,ch(y),,,] <- x.list[[k]]@catch.n[,ch(y),,,] *
                      prop.discards

               x.list[[k]]@catch <- apply(x.list[[k]]@catch.n * x.list[[k]]@catch.wt,
                      2, sum)
               x.list[[k]]@landings <- apply(x.list[[k]]@landings.n *
                      x.list[[k]]@landings.wt, 2, sum)
               x.list[[k]]@discards <- apply(x.list[[k]]@discards.n *
                      x.list[[k]]@discards.wt, 2, sum)

               # Update assess list object (stock.n and harvest to be estimated
               # subsequently)
               x.assess.list[[k]] <- x.list[[k]]

               # Generate research-vessel survey indices for the current year
               # unless already available (in-year survey in final historical year)
               for (kk in 1:length(x.idx.list[[k]]))
                      {if (is.na(x.idx.list[[k]][[kk]]@index[1,ch(y),,,]))
                             {a1 <- x.idx.list[[k]][[kk]]@range["min"]
                             a2 <- x.idx.list[[k]][[kk]]@range["max"]
                             aa <- ch(a1:a2)
                             x.idx.list[[k]][[kk]]@index[aa,ch(y),,,] <-
                                    x.list[[k]]@stock.n[aa,ch(y),,,] * x.q[kk,aa] *
                                    exp(rnorm(n = length(aa), sd = idx.cv))
                             }
                      }

               # Generate catch data with measurement error defined by catch.cv
               x.assess.list[[k]]@catch.n <- x.assess.list[[k]]@catch.n *
                      exp(rnd.catch)
               x.assess.list[[k]]@catch <- apply(x.assess.list[[k]]@catch.n *
                      x.assess.list[[k]]@catch.wt, 2, sum)
               x.assess.list[[k]]@landings.n[,ch(y),,,] <-
                      x.assess.list[[k]]@catch.n[,ch(y),,,] * prop.landings
               x.assess.list[[k]]@landings <- apply(x.assess.list[[k]]@landings.n *
                      x.assess.list[[k]]@landings.wt, 2, sum)
               x.assess.list[[k]]@discards.n[,ch(y),,,] <-
                      x.assess.list[[k]]@discards.n[,ch(y),,,] * prop.discards
               x.assess.list[[k]]@discards <- apply(x.assess.list[[k]]@discards.n *
                      x.assess.list[[k]]@discards.wt, 2, sum)

               # Run XSA assessment
               x.tmp <- window(x.list[[k]], start = x@range["minyear"], end = y-1)
               x.idx.tmp <- window(x.idx.list[[k]], start = x@range["minyear"], end =
                      y-1)
               x.xsa <- FLXSA(x.tmp, x.idx.tmp, xsa.control)
               x.assess.list[[k]]@stock.n <- window(x.xsa@stock.n, start =
                      x@range["minyear"], end = x@range["maxyear"])
               x.assess.list[[k]]@harvest <- window(x.xsa@harvest, start =
                      x@range["minyear"], end = x@range["maxyear"])
```

Figure 4.4.1.2.3 (continued): R code for a possible HCR for North Sea haddock (target *F* with constrained TAC change). For brevity, only the principal loops are given here: initialisation and summary code is also required (continued over).

```
                # In first assessment year, assessment results = "truth" for all years
                # up to y-1
                if (y == first.year)
                        {x.list[[k]]@stock.n[,ch(x@range["minyear"]:(y-1)),,,] <-
                                x.assess.list[[k]]@stock.n[,ch(x@range["minyear"]:(y-
                                1)),,,]
                        x.list[[k]]@harvest[,ch(x@range["minyear"]:(y-1)),,,] <-
                                x.assess.list[[k]]@harvest[,ch(x@range["minyear"]:(y-
                                1)),,,]}

                # For first intermediate year only, use true landings as intended
                # landings and "true" F as intended F
                if (y == first.year)
                        {intended.landings[[k]][,ch(y),,,] <-
                                x.list[[k]]@landings[,ch(y),,,]
                        intended.f[[k]][,ch(y),,,] <-
                        apply(x.list[[k]]@harvest[fbar.min:fbar.max,ch(y),,,], 2, mean)}

                # Apply HCR
                # Estimate F-multiplier based on target F
                f.current.assess <- x.assess.list[[k]]@harvest[,ch(y-1)]
                f.current.true <- x.list[[k]]@harvest[,ch(y-1)]
                mean.f.current <-
                        apply(as.table(f.current.assess[fbar.min:fbar.max]),2,mean)
                f.mult <- target.f / mean.f.current
                # Run forecast according to f-multiplier
                x.tac <- constrained.forecast(x.list[[k]], y, tac = NA, fmult = f.mult,
                        lprop = prop.landings, dprop = prop.discards)
                inter.landings <- intended.landings[[k]][,ch(y),,,]
                if (x.tac$landings[,ch(y+1),,,] >  inter.landings +
                        (inter.landings * tac.change.limit))
                        {x.tac <- constrained.forecast(x.list[[k]], y,
                                tac = inter.landings + (inter.landings *
                                tac.change.limit),
                                fmult = NA, lprop = prop.landings, dprop =
                        prop.discards)}
                else if (x.tac$landings[,ch(y+1),,,] < inter.landings -
                        (inter.landings * tac.change.limit))
                        {x.tac <- constrained.forecast(x.list[[k]], y,
                                tac = inter.landings - (inter.landings *
                                tac.change.limit),
                                fmult = NA, lprop = prop.landings, dprop =
                                prop.discards)}
                x.list[[k]]@landings[,ch(y+1),,,] <- x.tac$landings[,ch(y+1),,,]
                intended.landings[[k]][,ch(y+1),,,] <- x.tac$landings[,ch(y+1),,,]
                intended.f[[k]][,ch(y+1),,,] <- target.f
        } # End y loop
} # End k loop
```

Figure 4.4.1.2.3 (continued): R code for a possible HCR for North Sea haddock (target *F* with constrained TAC change). For brevity, only the principal loops are given here: initialisation and summary code is also required.

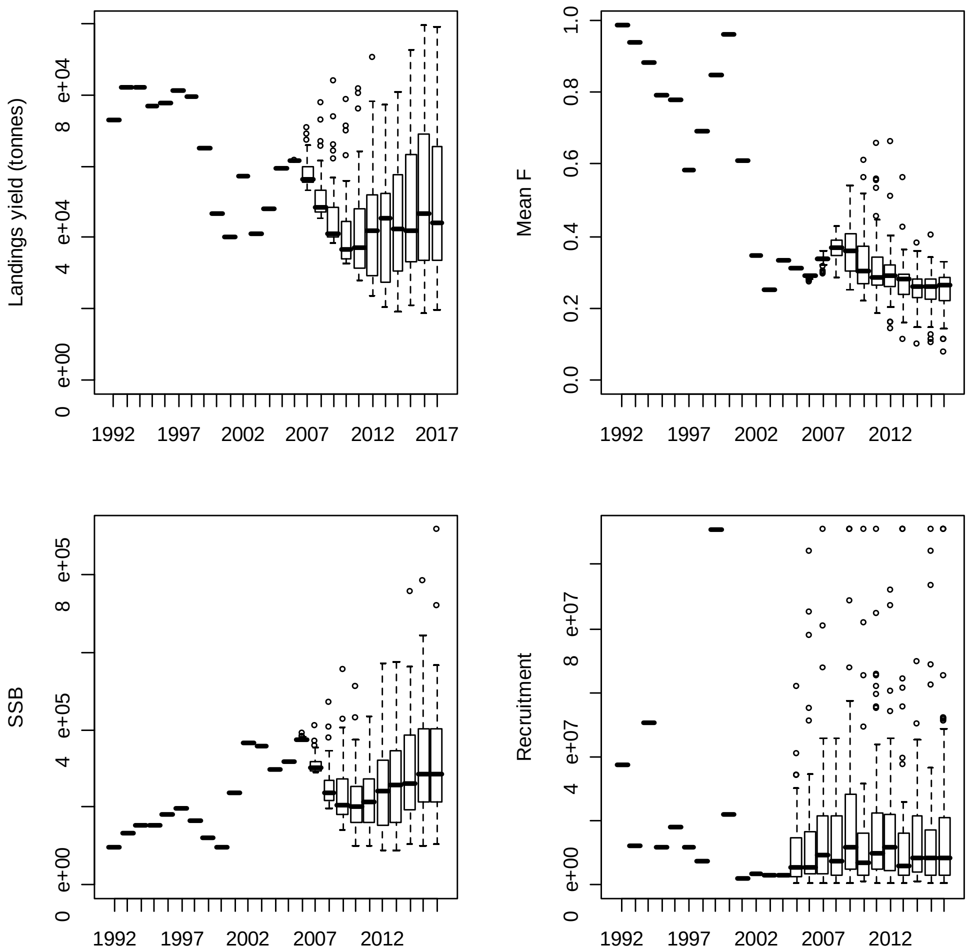


Figure 4.4.1.2.4: Example outputs from an evaluation of a possible HCR for North Sea haddock (target *F* = 0.3 with TAC change limited to ±15%), taken from Needle (WE1). Solid lines give medians (50th percentile) of distributions of estimates for each year, boxes show approximate first and third quartiles (25th and 75th percentile), while whiskers and points indicate outliers.

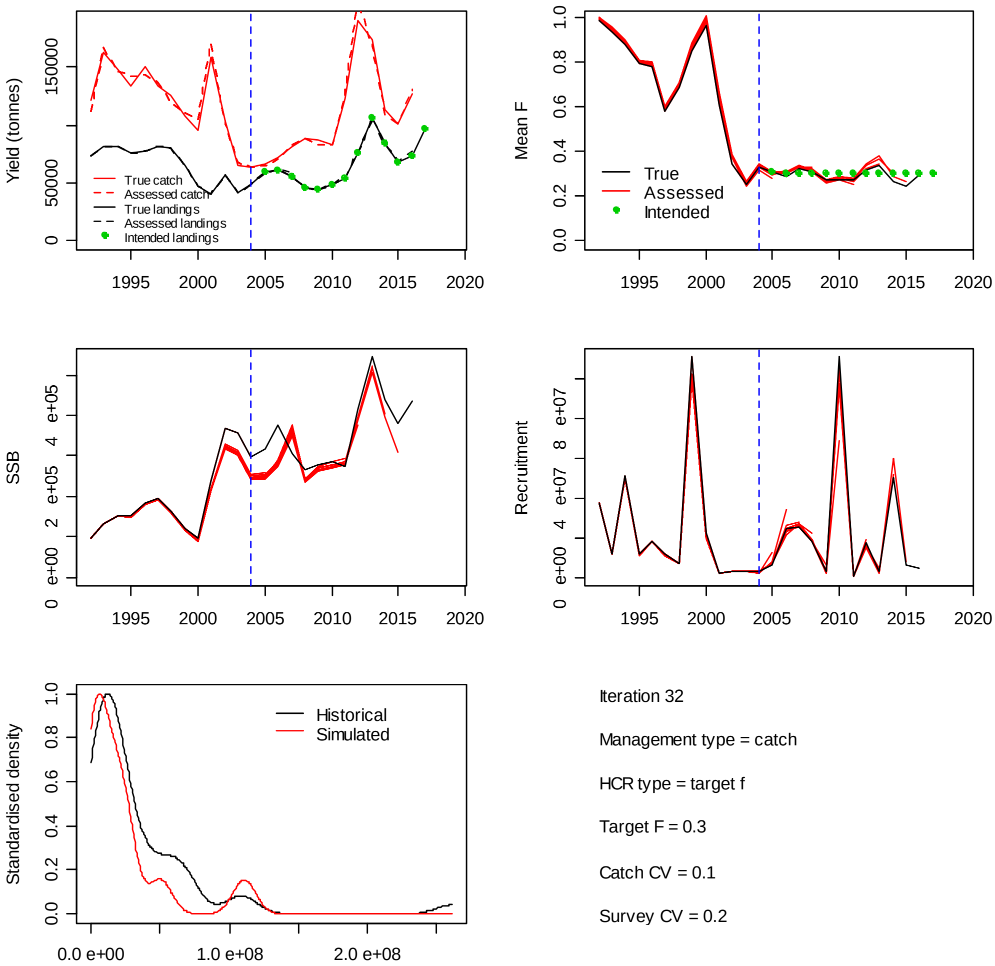


Figure 4.4.1.2.5: Example outputs from an evaluation of a possible HCR for North Sea haddock (target *F* = 0.3 with TAC change limited to ±15%), taken from Needle (WE1). For yield, the red lines show the true (solid) and measured (dashed) catch yield, while the black lines show the true (solid) and measured (dashed) landings yield. Green points show the intended landings for *y*+1 determined by management decisions occurring in *y*. For mean *F*, the black line shows the true values, the red lines show the estimates in each assessment year, and the green points show the intended fishing mortality. For SSB and recruitment, the black line shows the true values while the red lines show the estimates in each assessment year. For these four plots, the vertical line indicates the final historical year (2004). The final plot compares the standardised density distributions of historical and simulated recruitment.

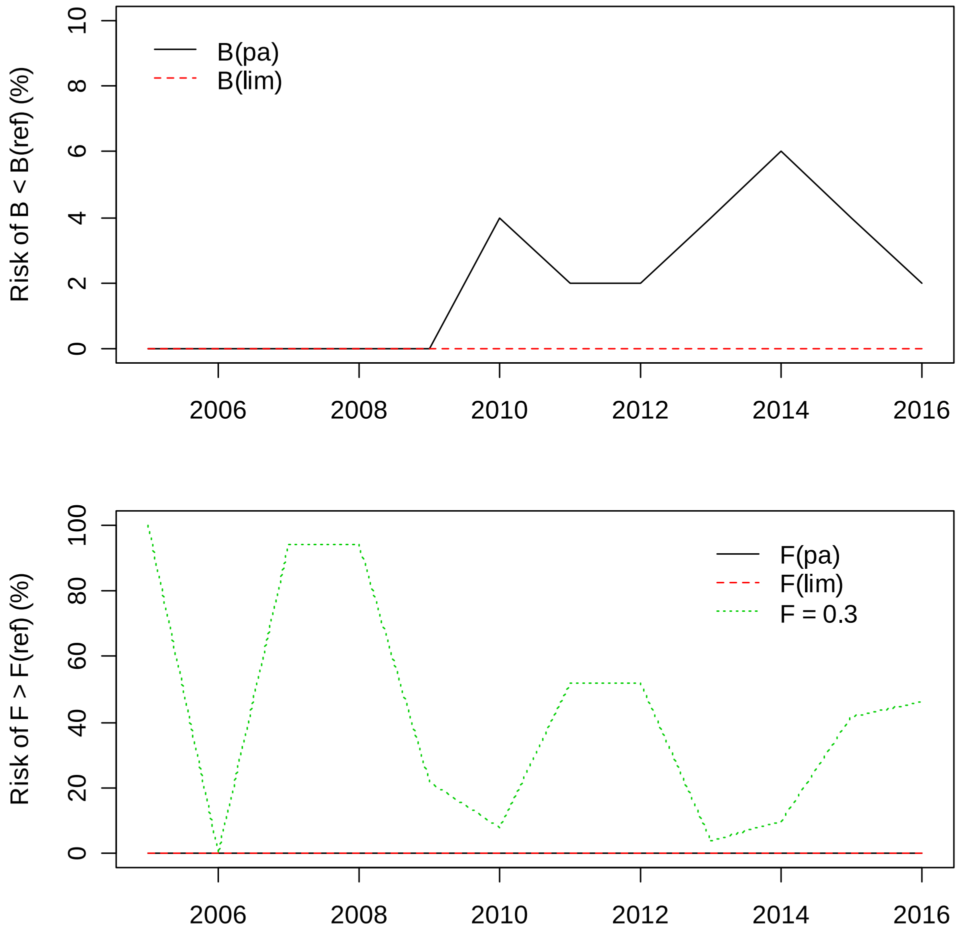


Figure 4.4.1.2.6: Example outputs from an evaluation of a possible HCR for North Sea haddock (target F = 0.3 with TAC change limited to ±15%), taken from Needle (WE1). Upper plot: percentage of iterations in each year for which $B_y < B_{pa}$ (black line) and $B_y < B_{lim}$ (red line). Lower plot: percentage of iterations in each year for which $F_y > F_{pa}$ (black line), $F_y > F_{lim}$ (red line), and $F_y > F_{target}$ (green line).

4.4.2 Northern hake

Hake – northern stock in Division IIIa, Sub-areas IV, VI and VII, and Divisions VIIIa,b,d

4.4.2.1 Management plan, decision tree and critical points

Council Regulation (EC) No 811/2004 of 21 April 2004 establishing measures for the recovery of the northern hake stock

The main Articles of interest adopted by this Regulation are:

'***Article 1. Subject matter.*** *This Regulation establishes a recovery plan for the northern hake stock which inhabits the ICES division III a, ICES subarea IV, ICES divisions V b (Community waters), VI a (Community waters), ICES subarea VII and ICES divisions VIII a, b, d, e (the northern hake stock).*

Article 2. Purpose of the recovery plan. *The recovery plan referred to in Article 1 shall aim to increase the quantities of mature fish of the northern hake stock concerned to values equal to or greater than 140 000 tonnes.*

Article 3. Reaching of target levels. *Where the Commission finds, on the basis of advice from ICES and following agreement on that advice by the Scientific Technical and Economic Committee for Fisheries (STECF), that for two consecutive years the target level for the northern hake stock concerned has been reached, the Council shall decide by qualified majority on a proposal from the Commission to replace the recovery plan by a management plan for the stock in accordance with Article 6 of Regulation (EC) No 2371/2002.*

Article 4. Setting of TACs. *A TAC shall be set in accordance with Article 5 where, for the northern hake stock concerned the quantities of mature northern hake have been estimated by the STECF, in the light of the most recent report of ICES, to be equal to or above 100 000 tonnes.*

Article 5. Procedure of setting TACs.

1) *Each year, the Council shall decide by qualified majority on a proposal from the Commission on a TAC for the following year for the northern hake stock concerned.*

2) *For 2004, the TAC shall be set at a level corresponding to a fishing mortality of* ***0.25****, 4% less than status quo fishing mortality. For the subsequent years of the recovery plan, the TAC shall not exceed a level of catches which scientific evaluations carried out by the STECF, in the light of the most recent reports of ICES, indicate will correspond to a fishing mortality rate of 0,25.*

3) *The Council shall not adopt a TAC whose capture is predicted by the STECF, in the light of the most recent report of the ICES, to lead to a decrease in spawning stock biomass in its year of application.*

4) *Where it is expected that the setting of the TAC for a given year in accordance with paragraph 2 will result in a quantity of mature fish at the end of that year in excess of the target level indicated in Article 2, the Commission will carry out a review of the recovery plan and propose any adjustments necessary on the basis of the latest scientific evaluations. Such a review shall in any event be carried out not later than three years following the adoption of this Regulation with the aim of ensuring that the objectives of the recovery plan are achieved.*

5) *except for the first year of application of this Regulation, the following rules shall apply:*

a) *where the rules provided for in paragraph 2 or 4 would lead to a TAC for a given year which exceeds the TAC of the preceding year by more than 15%, the Council shall adopt a TAC which shall not be more than 15% greater than the TAC of that year or;*

b) *where the rule provided for in paragraph 2 or 4 would lead to a TAC for a given year which is more than 15% less than the TAC of the preceding year, the Council shall adopt a TAC which is not more than 15% less than the TAC of that year.*

Article 6. Setting of TACs in exceptional circumstances. *Where the quantities of mature fish of the northern hake stock concerned have been estimated by the STECF, in the light of the most recent report of the ICES, to be less than 100 000 tonnes, the following rules shall apply:*

a) *Article 5 shall apply where its application is expected to result in an increase in the quantities of mature fish of the northern hake stock concerned, at the end of the year of application of the TAC to a quantity equal to or greater than 100 000 tonnes;*

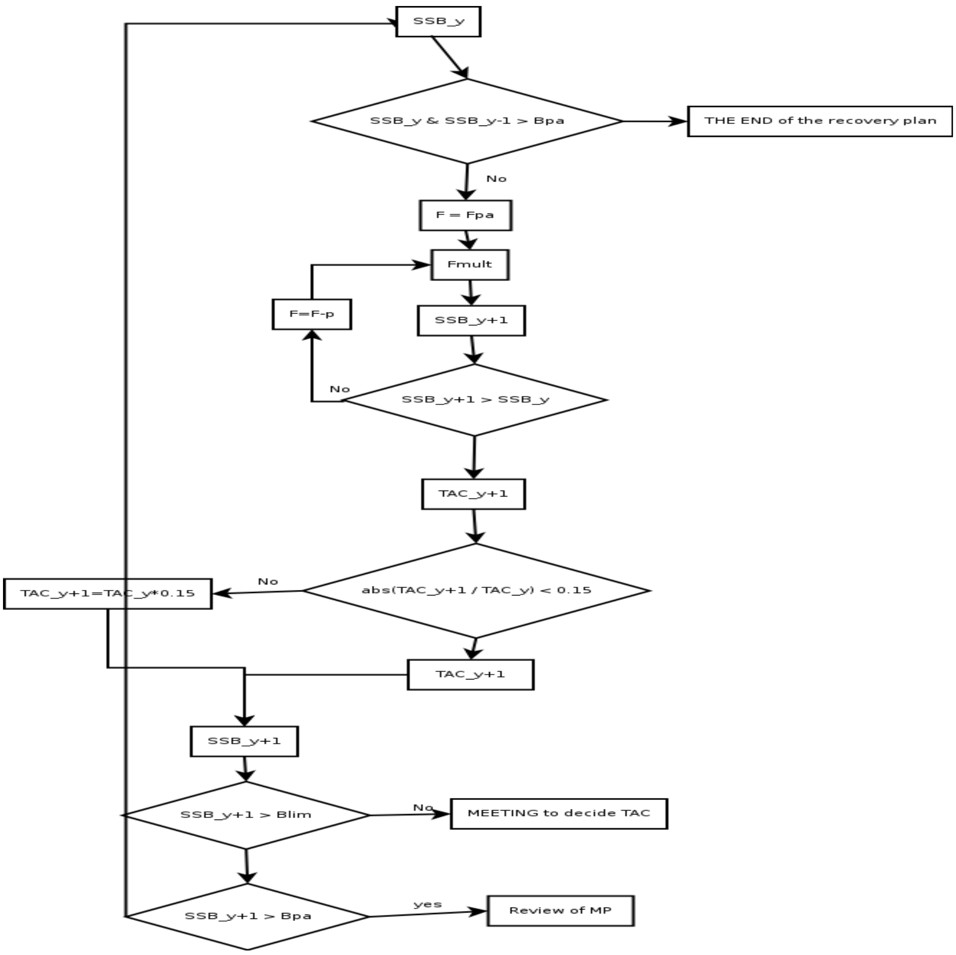
b) *where the application of Article 5 is not expected to result in an increase in the quantities of mature fish of the northern hake stock concerned, at the end of the year of application of the TAC, to a quantity equal to or greater than 100 000 tonnes, the Council shall decide by a qualified majority, on a proposal from the Commission, on a TAC for the following year that is lower than the TAC resulting from the application of the method described on Article 5.'*

ICES has not yet evaluated the current recovery plan in relation to the precautionary approach but intends to carry out an evaluation of possible management approaches before the end of 2006.

4.4.2.2 Illustrative harvest control rules and pseudo code

The text of the recovery plan has been interpreted by constructing a decision chart and it was found that in the case of northern hake, the agreed harvest rule specifies all details needed to evaluate the performance of the rule by simulation, although one of the end-points calls for an emergency meeting, whose output cannot be modelled.

Flowchart:



Pseudo-code:

```
IF SSB_y & SSB_y-1 < Bpa THEN
  propose F equal to Fpa
  calculate F multiplier
  forecast catches and ssb for next year
  WHILE SSB_y+1 < SSB_y
    forecast catches by diminishing F by 0.05
  ENDWHILE
  estimate TAC_y+1
  IF abs(TAC_y+1/TAC_y) > 0.15 THEN
    TAC_y+1 = TAC_y*0.15|TAC_y*0.85
  ENDIF
  estimate SSB_y+1
```

```
  IF SSB_y+1 < Blim THEN
    call meeting to decide TAC
  ELSE
   IF SSB_y+1 > Bpa THEN
    review of MP
   ELSE
    apply TAC_y+1
   ENDIF
  ENDIF
ELSE
  end of the recovery plan
ENDIF
```

Although the situation that prompted the decision on a recovery plan has changed, with later assessments providing a more optimistic view of the status of the stock, Northern Hake is still managed following the recovery plan quoted above. A preliminary analysis of the likely effect of the recovery plan HCR under a range of uncertainties has been presented to the ICES' Working Group on the Assessment of Southern Shelf Stocks of Hake, Monk and Megrim (WGHMM 2006), and some of those results are presented below.

The evaluation of the HCR has been carried out using an Operation Management Procedure (OMP) coded using the FLR system, and is available inside the FLOM package (available at http://flr-project.org/doku.php?id=pkg:flom). An explanation of the procedure used to condition the model and how to implement your own HCR using this system can be found at the FLR website (http://flr-project.org/doku.php?id=appl:nhake). The biological operating model is generated by considering the results of the latest assessment as a faithful representation of the dynamics of the stock. This means it can be easily applied to other stocks assessed using any age-structured assessment model. The uncertainties introduced in the analysis presented here include variability in the stock-recruitment relationship and growth curve. Assessment model, stock-recruitment and HCR are all input arguments to the function, so they can be changed by the user at will, even inside a simulation procedure.

The HCR for the current recovery plan appears to perform as expected. SSB should increase with a high probability of being above B_{lim} and B_{pa} (Figures 4.4.2.2.1-4.4.2.2.3). Yearly TAC reductions are moderate and they should bring both F and SSB to the levels observed at the beginning of the historical period. Uncertainty makes the results of forecasts not advisable outside of probability settings; i.e. in terms of probability of being below or above certain BRP (Garcia and Mosqueira, 2006).

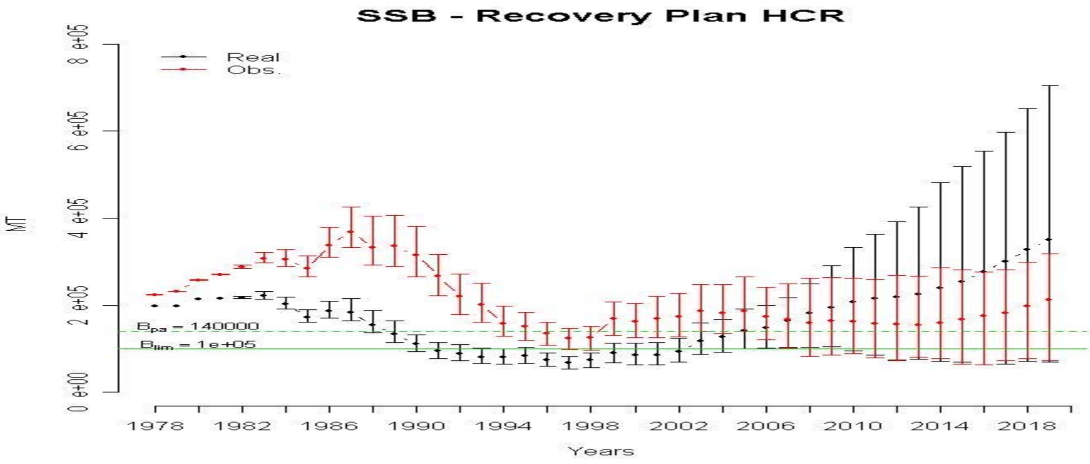


Figure 4.4.2.2.1: Example output from an evaluation of an HCR for northern hake – SSB.

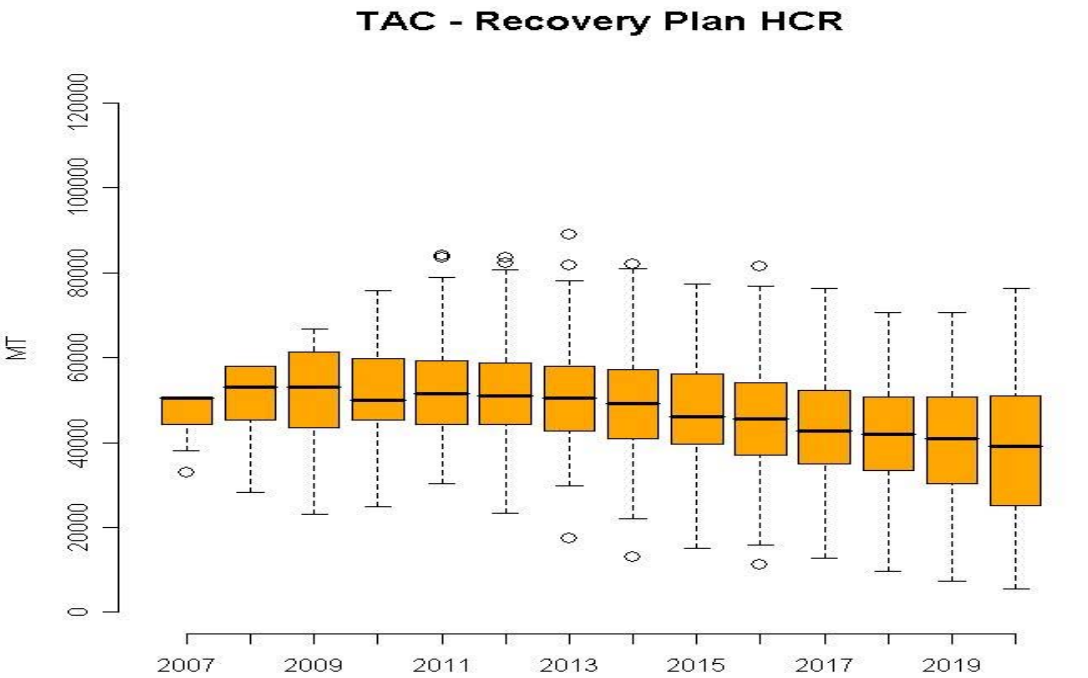


Figure 4.4.2.2.2: Example output from an evaluation of an HCR for northern hake – yield.

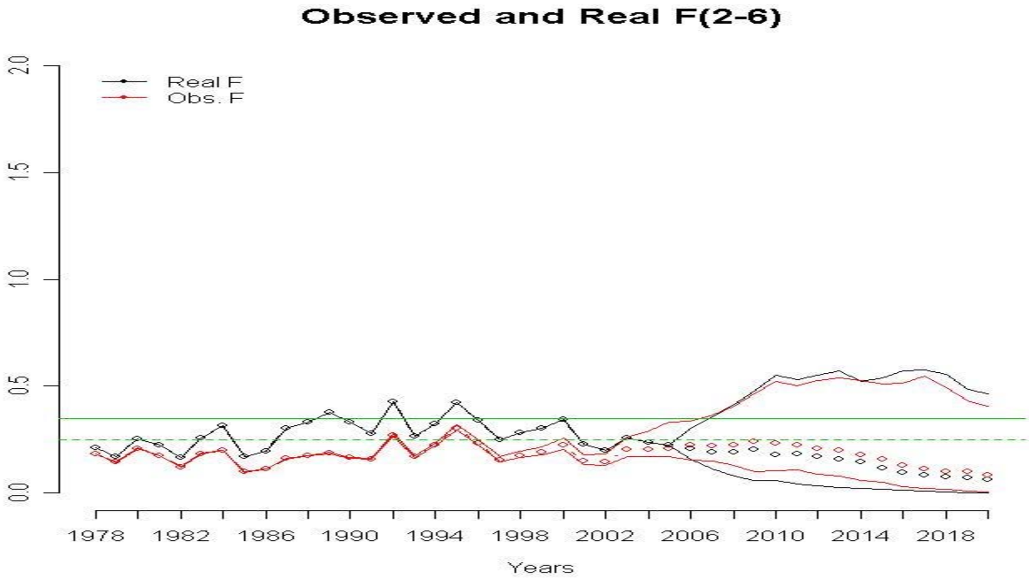


Figure 4.4.2.2.3: Example output from an evaluation of an HCR for northern hake – F.

4.4.3 Irish Sea cod

Cod in Division VIIa.

4.4.3.1 Management plan, decision tree and critical points

Council Regulation (EC) No 423/2004 of 26 February 2004 establishing measures for the recovery of cod stocks

For stocks above $\boldsymbol{B}_{lim}$, the harvest control rule (HCR) requires:

1. *setting a TAC that achieves a 30% increase in the SSB from one year to the next,*
2. *limiting annual changes in TAC to ± 15% (except in the first year of application), and,*
3. *a rate of fishing mortality that does not exceed $\boldsymbol{F}_{pa}$.*

For stocks below $\boldsymbol{B}_{lim}$ the Regulation specifies that:

4. *conditions 1-3 will apply when they are expected to result in an increase in SSB above $\boldsymbol{B}_{lim}$ in the year of application,*
5. *a TAC will be set lower than that calculated under conditions 1-3 when the application of conditions 1-3 is not expected to result in an increase in SSB above $\boldsymbol{B}_{lim}$ in the year of application.*

ICES has not yet evaluated the current recovery plan in relation to the PA but intends to carry out an evaluation of possible management approaches before the end of 2006.

A flow chart (decision tree) for the implementation of the management plan is given in Fig.4.4.3.1.1. This flow chart highlights elements of the management plan (shaded in grey) which would be either problematic to translate into an algorithm or represent a decision taken at a point of great uncertainty. At the very first decision branch in the flow diagram, there would be an issue in evaluating whether the SSB was above B_{lim}. If a standard assessment methodology is used the uncertainty around SSB_y is great and so this decision is very sensitive to assessment assumptions. At the second decision branch (in Article 6) the decision is again very sensitive to projection uncertainties, particularly recruitment in the assessment year which contributes to the SSB in y+2. In addition end points of the flow diagram are impossible to code, for example the end point which is arrived at following a true outcome from the second decision branch. This states *review plan* and is an open element with no logical outcome without further elaboration (which is not given in the management plan). Also the end point arising from a false outcome to the decision branch in Article 7. This states that a TAC would *be set by a qualified majority at less than the TAC's arising from Article 6*. This is again an open element and not possible to *code*.

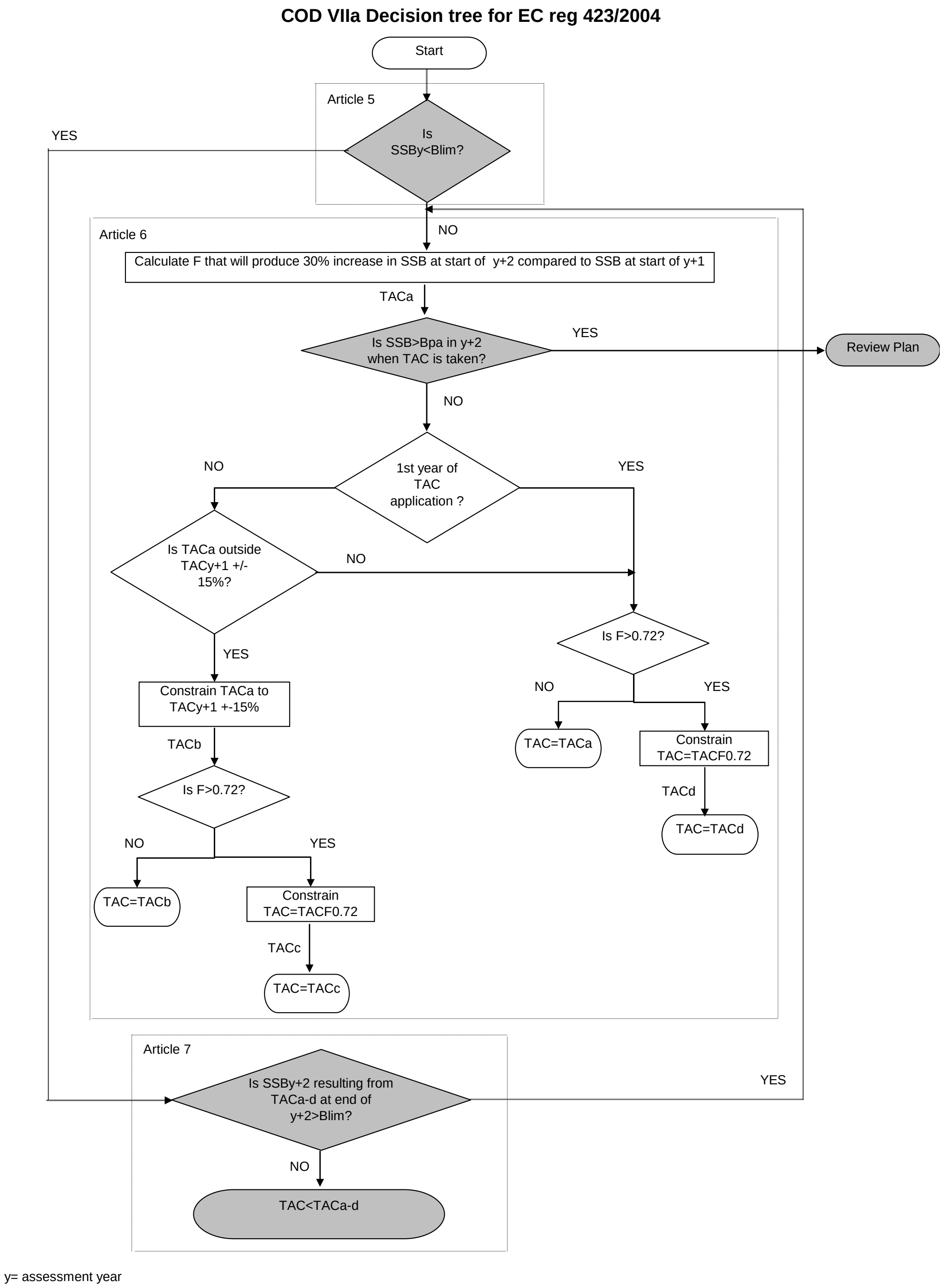


Figure 4.4.3.1.1: Decision tree for EC Regulation 423/2004.

4.4.3.2 Illustrative harvest control rules and pseudo code

Given the open ended outcomes highlighted in the decision tree above (Figure 4.4.3.1.1) it would be impossible to translate the management plan into an algorithm without making arbitrary assumptions. One possible outcome to an evaluation of this management plan is to conclude that it cannot be evaluated by simulation alone without closing the open ended outcomes, which could be achieved by either arbitrary decisions from the scientists or opening a dialogue with the managers. An alternative approach to this problem is demonstrated in Kelly *et al*. (2006), where elements of the management plan were simulated. This approach using the F-PRESS tool simplified the simulation process; i.e. no assessment feedback, and tested elements of the management plan for robustness to noise and bias in both the observation and management models.

The flow diagram for the algorithm used in F-PRESS is given in Figure 4.4.3.2.1. The element of this flow diagram which deals with the HCR is further expanded into pseudo code.

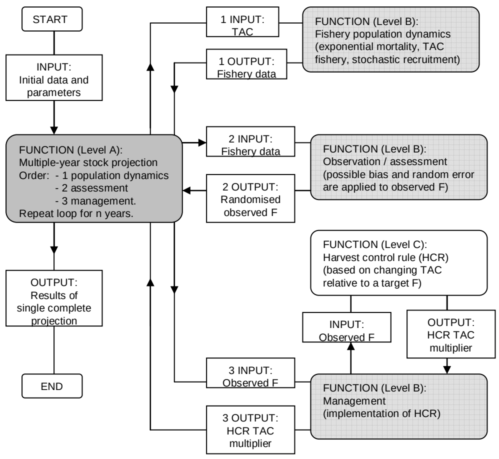


Figure 4.4.3.2.1: Algorithmic flow-chart used in F-PRESS.

Pseudo-code for HCR used in Codling and Kelly (BE1). Note that this is not the actual HCR implemented for Irish Sea cod but was used for comparative purposes.

INPUTS – F_{OBS} (observed F which may have error and bias), F_T (target fishing mortality), TAC_n (TAC in previous year).

HCR:

$TAC_{n+1} = TAC_n * F_T / F_{OBS}$

IF 15% CONSTRAINT IS REQUIRED THEN:

(IF $TAC_{n+1} > 1.15* TAC_n$ THEN $TAC_{n+1} = 1.15* TAC_n$;

IF $TAC_{n+1} < 0.85* TAC_n$ THEN $TAC_{n+1} = 0.85* TAC_n$)

ELSE STOP.

TAC_{n+1} IS THE NEW TAC

Whilst this approach does not simulate the application of a full management plan it has merits in quickly identifying elements of the management plan which may be sensitive to underlying uncertainty and/or bias. This could then be returned to managers to continue the dialogue of achieving a robust management strategy.

4.5 Discussion

WGMG has elaborated further upon its approach for evaluating management strategy proposals and HCRs. The framework that has been presented and discussed for such evaluations is generally applicable; whilst the management plans and HCRs themselves will need to be stock-specific. Ultimately, there is a need to incorporate mixed fisheries and multi-species evaluations but as a first step, there are a number of single species evaluations which can be considered.

ICES has identified single species candidates for reference points which are consistent with taking high long-term yields and achieving a low risk of depleting the reproductive potential of stocks in the range $F_{0.1}$ to F_{max}. Under appropriate circumstances, however, higher Fs might be appropriate (ICES, 2005; where candidate harvest control rules for herring in Division VIa (North) and sole in Division VIIe (Western Channel) have been proposed).

Choices for reference points are better developed within the context of a management strategy and can be regarded as parameters in the harvest control rules. Simulations will have to show the risk that the stock falls outside bounds set by the Precautionary Approach.

Single species considerations

As originally proposed by WGMG at its meeting in 2004, harvest control rules could be considered and evaluated with existing software for some stocks. These are stocks with a relatively stable productivity (variation in recruitment, growth and maturity with stationary distributions), a long time series of data, low levels of technical interactions and exploited in a single species context and with *insignificant* (i.e. low) levels of discarding.

In this context, the single species recovery plans for cod stocks (in the Kattegat; in the North Sea, Skagerrak and eastern Channel; to the West of Scotland; and in the Irish Sea) and northern hake might be evaluated. The three case studies presented in this report illustrate how the initial steps in such evaluations might proceed prior to computer-based simulation.

The establishment of targets and conservation reference points will be necessary in such evaluations of management proposals. However, such evaluations are not merely scientific endeavours but will necessitate a dialogue and detailed discussion with managers and stakeholders. **It is not clear how ICES fits into this process!** ICES should be a partner in the whole process but may also play a role in the co-ordination and supervision of evaluations.

Biological interaction - multi-species considerations

Management of multi-species resources calls for considerations on the interaction between species caused by predation and food competition, so that changing fishing mortality on one stock may influence the production of other stocks (prey or competitors). Previous studies; for example, by the ICES multi-species assessment study groups for the Baltic (SGMAB) and the North Sea (SGMSNS), have shown that the performance of the single species HCRs is often very different when evaluated in a multi-species context compared to a single species context. WGMG did not consider such interactions further.

Technical interaction - mixed fisheries considerations

HCRs when technical interactions are prominent were briefly considered by SGMAS and have not been further considered by WGMG. The management of mixed fisheries poses specific problems because a direct management of fishing mortality in mixed fisheries requires a management on the metier level. HCRs for mixed fisheries will contain compromises between conflicting interests which optimize an overall objective. These compromises are generally of a political nature, and science can only outline the likely consequences of various priorities. Some models exist or are being developed that can aid in outlining such consequences. Such models were outlined by the SGMAS, but not considered by WGMG at this meeting.

5 Working documents and background material presented to the working group

5.1 Working papers and documents (W)

A total of seven (7) documents were presented to the meeting as working papers. These are listed in this Section 5.1; together with their assigned code for ease of reference within the various sections of this report.

ToR a)

WA1: Approaches dealing with systematic errors in catch, José de Oliveira.

WA2: Simulating missing catches using a Gadget model. Sam Subbey and Daniel Howell.

ToR b)

No working papers were submitted under this term of reference but presentations by Michel Bertignac and Colin Millar were given during the meeting.

ToR c)

WC1: Year-class peculiarities in selection pattern: how stock assessment based on separable cohort models is able to take them into account? (Some illustrations for triple-separable case of the ISVPA model - TISVPA). Dmitri Vasilyev.

WC2: North Sea haddock 2005 – assessments, forecasts and advice. Coby Needle.

ToRs d) and e)

WE1: Evaluating harvest control rules for North Sea haddock using FLR, Coby Needle.

WE2: Critical evaluation of F-PRESS (Fisheries Projection and Evaluation by Stochastic Simulation), Beatriz Roel.

WE3: FSS response to critical evaluation of F-PRESS 1.0, Edward Codling, Ciaran Kelly and Andrew Campbell.

5.2 Background material (B)

A total of three documents were submitted to the meeting as background papers. These are listed in this Section 5.2; together with their assigned code for ease of reference within the various sections of this report.

ToR a)

BA1: A spatial- and age-structured assessment model to estimate poaching and ecosystem change impacting the management of South African abalone (*Haliotis midae*). E.E. Plagányi and D.S. Butterworth.

ToRs d) and e)

BE1: F-PRESS A stochastic simulation tool for developing fisheries management advice and evaluating management strategies, Edward Codling and Ciaran Kelly, Irish Fisheries Investigations No. 17, Marine Institute, Galway, 2006.

BE2: Reference points, Martin Pastoors, ACFM May 2006, Document 12.

6 Nominations for Chair

The Chair reminded the meeting that this is his fourth consecutive meeting as Chair and invited nominations for the Chair from 2007. Following discussions within WGMG both during and after the meeting, it was unanimously agreed that Coby Needle (UK) should be proposed as Chair from 2007.

7 References

Throughout this report there have been a number of references cited within the Sections 2 to 4. In this Section 6, these references are collated.

Apostolaki, P., Babcock, E.A., and McAllister, M.K. (2006). Contrasting deterministic and probabilistic ranking of catch quotas and spatially and size-regulated fisheries management. Can. J. Fish. Aquat. Sci. 63: 1777–1792.

Balci, O. 1995. Principles and techniques of simulation validation, verification, and testing. *In* C. Alexopoulos, K. Kang, W. Lilegdon, and D. Goldsman (eds.) Proceedings of the 1995 Winter Simulation Conference, pp. 147–154.

Cadigan, N.G. and Farrell, P.J. 2002. Generalized local influence with applications to fish stock cohort analysis. Appl. Statist. 51: 1–15.

Cadigan, N.G. and Farrell, P.J. 2004. Local Influence Diagnostics for the Retrospective Problem in Sequential Population Analysis. ICES Journal of Marine Science. 62: 256–265.

Cotter, A.J.R. 2001. Intercalibration of North Sea International Bottom Trawl Surveys by fitting year-class curves. *ICES Journal of Marine Science* 58: 622-632 [Erratum, Ibid. 58:1340].

Fryer, R.J. 2002. TSA: is it the way? Appendix D in report of Working Group on Methods on Fish Stock Assessment. ICES CM 2002/D:01.

García, D. and Mosqueira, I. 2006. A generic operating model using FLR: An application to northern hake. Working Document to the ICES Working Group on Hake, Monk and Megrim, Bilbao, 9–18 May, 2006.

Garcia, D. and Mosqueira I. 2006. A generic extendable multi-fleet fisheries operational model in FLR with two example applications. Poster presented at the ICES Symposium on Fisheries Management Strategies: Case Studies of Innovation, Galway, Ireland, 27–30 June 2006.

Gavaris, S. 1988. An adaptive framework for the estimation of population size. Can. Atl. Fish. Sci. Adv. Comm. (CAFSAC) Res. Doc. 88/29: 12pp.

Gavaris, S. and van Eeckhaute, L. 1998. Diagnosing systematic errors in reported fishery catch. *In* Fishery Stock Assessment Models. Ed. by F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C.-I. Zhang. Alaska Sea Grant College Program Report No. AK-SG-98-01, University of Alaska Fairbanks: 399–412.

Hilborn, R. 1990. Estimating the parameters of full age-structured models from catch and abundance data. Bull. Int. N. Pac. Fish. Commn 50: 207–213.

Hirst, D., Aanes, S., Storvik, G., Huseby, R.B., and Tvete, I.F. 2004. Estimating catch-at-age from market sampling data using a Bayesian hierarchical model. Appl. Stat. 53(1): 1–14.

ICES. 2003. Report of the Study Group on Biological Reference Points for Northeast Arctic cod, Svanhovd, Norway, 13–17 January 2003. ICES CM 2003/ACFM:11.

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

ICES. 2004b. Report of the Study Group on Assessment Methods Applicable to Assessment of Norwegian Spring-Spawning Herring and Blue Whiting Stocks, Lisbon, Portugal, 19–22 February 2004. ICES CM 2004/ACFM:14.

ICES. 2004c. Report of the Northern Pelagic and Blue Whiting Fisheries Working Group. Copenhagen, Denmark, 27 April – 4 May 2004. ICES CM 2004/ACFM:24.

ICES. 2005. Report of the ICES Advisory Committee on Fishery Management, Advisory Committee on the Marine Environment and Advisory Committee on Ecosystems, 2005. ICES Advice. Volumes 1–11, 418pp.

ICES. 2006a. Report of the Northern Pelagic and Blue Whiting Fisheries Working Group, Copenhagen, 25 August – 1 September 2005. ICES CM 2006/ACFM:05.

ICES. 2006b. Workshop on Advanced Fish Stock Assessment Techniques: Lectures and teaching material. ICES CM 2006: D:01.

Kleijnen, J.P.C. 1995. Verification and validation of simulation models. *European Journal of Operational Research*, 82: 145–162.

Millar, C.P, and Fryer R.J. 2005. Revised estimates of annual discards-at-age for cod, haddock, whiting and saithe in ICES sub-area IV and division VIa. Fisheries Research Services Internal Report No. 15/05.

Mohn, R. 1999. The retrospective problem in sequential population analysis: An investigation using cod fishery and simulated data. ICES Journal of Marine Science, 56: 473–488.

Needle, C.L. 2003. Survey-based assessments with SURBA. Working Document to the ICES Working Group on Methods of Fish Stock Assessment, Copenhagen, 29 January – 5 February 2003.

O'Brien, C.M. and Stone, M. 1984. A new procedure to analyse radioactive emission count data. *Nuclear Instruments and Methods in Physics Research*, 223:130–136.

Patterson, K.R., Cook, R.M., Darby, C.D., Gavaris, S., Mesnil, B., Punt, A.E., Restrepo, V.R., Skagen, D.W., Stefánsson, G., and Smith, M. 2000. Validating three methods for making probability statements in fisheries forecasts. ICES CM 2000/V:06.

Payne, A.G., Agnew, D.J. and A. Brandão 2005. Preliminary assessment of the Falklands Patagonian toothfish (*Dissostichus eleginoides*) population: use of recruitment indices and the estimation of unreported catches. Fish. Res. 76: 344–358.

Pitcher, T.J., Watson, R., Forrest, R., Valtýsson, H., and Guénette, S. 2002. Estimating illegal and unreported catches from marine ecosystems: a basis for change. Fish and Fish. 3: 317–339.

Plagányi, E.E. 2004. Walking the bridge from single- to multi-species approaches in southern African fisheries management. PhD thesis, University of Cape Town: 460pp.

Plagányi, E.E. and Butterworth, D.S. (in preparation). A spatial- and age-structured assessment model to estimate poaching and ecosystem change impacting the management of South African abalone (*Haliotis midae*).

Porch, C.E., Eklund, A.-M., and Scott, G.P. 2006. A catch-free stock assessment model with application to goliath grouper (*Epinephelus itajara*) off southern Florida. Fish. Bull. 104: 89–101.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2004. ISBN 3-900051-07-0.

Restrepo, V. R. and Legault, C.M. 1998. A stochastic implementation of an age-structured production model. *In* Fishery Stock Assessment Models. Ed. by F. Funk, T.J. Quinn II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweigert, P.J. Sullivan, and C.-I. Zhang. Alaska

Sea Grant College Program Report No. AK-SG-98-01, University of Alaska Fairbanks: 435–450.

Rochet, M-J, and Trenkel, V.M. 2005. Factors for the variability of discards: assumptions and field evidence. Can. J. Fish Aquat. Sci. 62: 224–235.

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

Sargent, R.G. 2005. Verification and validation of simulation models. *In* Kuhl, M.E., Steiger, N.M., Armstrong, F.B. and Joines, J.A. (eds.) Proceedings of the 2005 Winter Simulation Conference, pp. 130–143.

Stone, M. 1974. Cross-validatory choice and assessment of statistical predictions (with Discussion). *J.R. Statist. Soc. B*, 36:111–147.

Stratoudakis, Y., Fryer, R.J., Cook, R.M., and G.J. Pierce 1999. Fish discarded from Scottish demersal vessels: Estimators of total discards and annual estimates for targeted gadoids. ICES J. Mar. Sci. 56: 592–605.

Thompson, S.K. 2002. Sampling. Wiley Series in Probability and Statistics. John Wiley & Sons Inc., New York.

Annex 1: List of participants

NAME	ADDRESS	PHONE	EMAIL
Carl O'Brien (Chair)	Cefas Lowestoft Laboratory, Pakefield Road, Lowestoft, Suffolk NR33 0HT, United Kingdom	+44 1502 524256	c.m.obrien@cefas.co.uk Shortly to be replaced: carl.obrien@cefas.co.uk
Asgeir Aglen	Institute of Marine Research, P.O. Box 1870 Nordnes,N-5817 Bergen, Norway	+47 55 238680	asgeir.aglen@imr.no
Michel Bertignac	IFREMER Centre de Brest, BP 70 – 29280 Plouzane, France	+33 2 98 22 45 25	michel.bertignac@ifremer.fr
Noel Cadigan	Dept. of Fisheries & Oceans, Northwest Atlantic Fisheries Centre, P.O. Box 5667, St John's, Nfld A1C 5X1, Canada	+709 772 5028	cadiganN@dfo-mpo.gc.ca
Massimiliano Cardinale (part time)	Institute of Marine Research, Box 4, SE-453 21 Lysekil, Sweden	+46 523 187 501	massimiliano.cardinale@fiskeriverket.se
Chris Darby	Cefas Lowestoft Laboratory, Pakefield Road, Lowestoft, Suffolk NR33 0HT, United Kingdom	+44 1502 524329	c.d.darby@cefas.co.uk Shortly to be replaced: chris.darby@cefas.co.uk
Yuri Efimov	VNIRO, 17 Verkhne Krasnoselskaya, Moscow 107140, Russia	+7 495 264 9129	efimov@vniro.ru
Ciaran Kelly	Fisheries Science Services, Marine Institute, Galway Technology Park, Galway, Ireland	+353 91 730 454	ciaran.kelly@marine.ie
Benoit Mesnil	Departement EMH, IFREMER, BP 21105, 44311 Nantes Cedex 03, France	+33 240 374 009	benoit.mesnil@ifremer.fr
Colin Millar	Fisheries Research Services, Marine Laboratory, PO Box 101, 375 Victoria Road, Torry, Aberdeen AB11 9DB United Kingdom	+44 1224 295 575	millarc@marlab.ac.uk
Iago Mosqueira	AZTI, Txatxarramendi uqarter z/g, 48395 Sukarrieta (Bizkaia), Spain	+34 94 602 94 00	imosqueira@suk.azti.es
Coby Needle	Fisheries Research Services, Marine Laboratory, PO Box 101, 375 Victoria Road, Torry, Aberdeen AB11 9DB United Kingdom	+44 1224 295 456	needlec@marlab.ac.uk
Dankert Skagen	Institute of Marine Research, P.O.Box 1870 Nordnes,N-5817 Bergen, Norway	+47 55 238419	dankert@imr.no
Sam Subbey	Institute of Marine Research, P.O.Box 1870 Nordnes,N-5817 Bergen, Norway	+47 55 235383	samuels@imr.no
Dmitri Vasilyev	VNIRO, 17 Verkhne Krasnoselskaya, Moscow 107140,	+7 495 264 8974	dvasilyev@vniro.ru

	Russia		

Annex 2: WGMG Terms of Reference for the next meeting

The **Working Group on Methods of Fish Stock Assessments** [WGMG] (Chair: Coby Needle*, UK) will meet in **xxxxx** for 10–15 days in March 2007 to:

a) investigate further, and test, the sensitivities of stock assessment methods to known data problems with particular reference to the retrospective problem;

b) review developments in fisheries independent survey-based assessment tools; and

c) evaluate the current state of operational evaluation tools for fisheries management options.

WGMG will report by **xxxxx** 2007 for the attention of the Resource Management Committee and ACFM.

Supporting Information

PRIORITY:	The work of this group is essential for ICES to progress in the development of techniques of fish stock assessment and the evaluation of management strategies.
SCIENTIFIC JUSTIFICATION AND RELATION TO ACTION PLAN:	The three proposed ToRs are pertinent to current issues of concern within fisheries science and management. Considering each in turn. ToR a): Approaches for the correction of bias in stock assessments need to be further developed and implemented if reliable forecasts are to be produced in the future. This has a bearing on the development and evaluation of management strategies and harvest control rules. WGMG proposes that the group meets at the Northeast Fisheries Science Centre, NOAA, Woods Hole, USA where research is currently on-going to address this ToR. ToR b): Scientific approaches under the ToR a) will aim to address the catch data issue, will attempt to reconstruct the missing catches so that the conventional approach to deliver advice (catch-based assessment and forecasts) can be maintained. An alternative approach is to ignore the catch data altogether and try to inform managers on the state of stocks based on information from surveys alone. This is the task assigned to the on-going EU-funded project FISBOAT. A work package in this project is specifically tasked to: *Supply methods for analysing Fishery-Independent (F-I) stock assessment data to provide managers with relevant information about stocks and their exploitation; provide F-I assessment models; and provides parameter estimation procedures for these models*. It is anticipated that work from this project will be presented at the next meeting of WGMG. ToR c): The EU-funded project EFIMAS is tasked with developing computer-based tools for the evaluation of fisheries management options and it is anticipated that work from this project will be presented at the next meeting of WGMG. The FLR software will be further evaluated and tested. Additionally, one may assume that ACFM will task the group with additionally ToRs.

RESOURCE REQUIREMENTS:	None.
PARTICIPANTS:	Group is well-manned by the correct people but would benefit from new members participating in the work of the group. Input from observer countries would be desirable.
SECRETARIAT FACILITIES:	Meeting facilities, production of report.
FINANCIAL:	None.
LINKAGES TO ADVISORY COMMITTEES:	ACFM has strongly supported the work of this group and has worked actively in formulating the ToRs for recent meetings. WGMG will report to ACFM at its autumn meeting in 2007.
LINKAGES TO OTHER COMMITTEES OR GROUPS:	WGMG will report to the Resource Management Committee at the ICES ASC in 2007. There will also be links to the Fisheries Technology Committee.
LINKAGES TO OTHER ORGANIZATIONS:	There is similar work going on within ICCAT and NAFO. Co-ordination should be assured.
SECRETARIAT MARGINAL COST SHARE:	ICES:100%

Annex 3: Some verification and testing tools available in R

A very simple example of some of the tools available in the R statistical language (hhttp://www.r-project.org) for testing and verification is presented here. The HCR presented, contained in the hcr() function, is obviously much simpler than any real case, but will serve for illustrative purposes. The text of the management plan behind this HCR would be the following:

1. If the stock level is below 100 t, the fishery shall be closed to all fishing.
2. At stock levels between 100 and 1000 t, catch shall be limited to 10 t per year.
3. If the stock is considered to be above the 1000 t limit, TAC shall be set at a level of 20 t.

Input in this case consists on a single value for the estimated stock level, and output consists on a single value for the given TAC.

Unit testing

We first load the RUnit library in R, containing the functions for testing we will be using.

```
library(RUnit)
```

We now write the hcr() function implementing the above HCR. Due to its simplicity this can be coded directly, but in more complex cases a useful step is to write pseudo-code and/or a flow chart so as to see exactly how the rules contained in the management plan can be turned into code.

```
hcr <- function(stock) {
    if(stock < 0)
        stop('stock must be 0 or greater')
    if(stock < 100)
        return(0)
    else if (stock <1000)
        return(10)
    else
        return(20)
}
```

A first test function can now de developed as part of a test suite. Checks are made in this instance of the class (numeric) and value (0 or greater) of the function output. Finally, checks are carried out that the function returns the expected values for a range of input values (10, 120 and 1100).

```
# Unit testing
test.hcr <- function() {
    # Check that output is numeric
```

```
    checkTrue(is.numeric(hcr(10)))
    # Check that result is always zero or greater
    checkTrue(hcr(0) >= 0)
    checkTrue(hcr(100) >= 0)
    # Check hcr() gives an error when called with negative
values
    checkException(hcr(-10) >= 0)
    # Check that result is 0 if input is less than 100
    checkEquals(hcr(10), 0, tolerance = 1e-8)
    # Check that result is 10 if input is less than 1000 but
greater than 100
    checkEquals(hcr(120), 10, tolerance = 1e-8)
    # Check that result is 20 if input is greater than 1000
    checkEquals(hcr(1100), 20, tolerance = 1e-8)
}
```

To run the above test function, the function runTestFile() is invoked with the location of the file containing the test functions, in our example a temporary file created by R. Similar functions exist for defining and executing a test suite, a set of files with various test belonging, for example, to the same package.

```
hcrfile <- tempfile()
dump('test.hcr', file=hcrfile)
testData <- runTestFile(hcrfile)
```

Output of the test results can be obtained on the console or directed to a file, either in text or HTML format

```
printTextProtocol(testData, showDetails=TRUE)


> printTextProtocol(testData, showDetails=TRUE)
RUNIT TEST PROTOCOL -- Mon Jun 26 12:28:16 2006
***********************************************
Number of test functions: 1
Number of errors: 0
Number of failures: 0


1 Test Suite :
file74b0dc51 - 1 test function, 0 errors, 0 failures
```

```
Details
***************************
Test Suite: file74b0dc51
Test function regexp: ^test.+
Test file regexp: ^file74b0dc51$
Involved directory:
/tmp/RtmpKVg70S
---------------------------
Test file: /tmp/RtmpKVg70S/file74b0dc51
test.hcr: ... OK (0.01 seconds)
```

We can now create another test function with an error in it, to see the type of output RUnit gives us for tracking the problem

```
test.error.hcr <- function() {
    # hcr() should fail when called with a negative stock value
    checkTrue(hcr(-10) >= 0)
}

hcrfile <- tempfile()
dump('test.error.hcr', file=hcrfile)
testData <- runTestFile(hcrfile)

printTextProtocol(testData, showDetails=TRUE)
```

The output below shows the call stack, the succession of function calls that lead to the error.

```
> printTextProtocol(testData, showDetails=TRUE)
RUNIT TEST PROTOCOL -- Mon Jun 26 12:26:43 2006
***********************************************
Number of test functions: 1
Number of errors: 1
Number of failures: 0

1 Test Suite :
file66334873 - 1 test function, 1 error, 0 failures
```

```
ERROR in test.error.hcr: Erro en hcr(-10) : stock must be 0 or greater


Details
***************************
Test Suite: file66334873
Test function regexp: ^test.+
Test file regexp: ^file66334873$
Involved directory:
/tmp/RtmpKVg7OS
---------------------------
Test file: /tmp/RtmpKVg7OS/file66334873
test.error.hcr: ERROR !!
Erro en hcr(-10) : stock must be 0 or greater
   Call Stack:
   1: func()
   2: checkTrue(hcr(-10) >= 0)
   3: eval(expr)
   4: hcr(-10)
   5: stop("stock must be 0 or greater")
```

Tracing

An example call to the trace() function in R is presented here. The use of tracing is specially indicated when a single variable is altered along the code depending on a series of calculations or decisions. Tracing allows all intermediate steps to be shown. It also allows the initiation of an interactive debugging session when a certain value appears in the calculations (see second example below).

```
# Tracing
## arrange to call the browser on entering and exiting hcr
trace("hcr", browser, exit = browser)
hcr(99)
## instead, call the debugging browser only if the input value has a certain value
## Don't bother me otherwise
trace("hcr", quote(if(any(stock < 0)) browser()),
    print = TRUE)
```

hcr(-99)

hcr(99)

stop tracing

untrace('hcr')

Consistency

Finally, consistency checking can be included in the unit test approach by introducing repeated calls to functions and comparing the results. Again, an optional argument for tolerance can be specified, if the procedure includes random components likely to affect the final result.

```
# Consistency
test.consistency <- function() {
    res <- numeric(length=100)
    for (i in 1:100) res[i] <- hcr(220)
    checkEquals(res, rep(10, 100))
}

hcrfile <- tempfile()
dump('test.consistency', file=hcrfile)
testData <- runTestFile(hcrfile)

printTextProtocol(testData, showDetails=TRUE)

> printTextProtocol(testData, showDetails=TRUE)
RUNIT TEST PROTOCOL -- Mon Jun 26 12:37:19 2006
***********************************************
Number of test functions: 1
Number of errors: 0
Number of failures: 0

1 Test Suite :
file19495cff - 1 test function, 0 errors, 0 failures

Details
```

```
***************************
Test Suite: file19495cff
Test function regexp: ^test.+
Test file regexp: ^file19495cff$
Involved directory:
/tmp/RtmpKVg70S
---------------------------
Test file: /tmp/RtmpKVg70S/file19495cff
test.consistency: ... OK (0 seconds)
```

Annex 4: Recommendations

RECOMMENDATION	ACTION (OPTIONAL)
1. As a priority, ICES should develop illustrative HCRs for a limited number of priority stocks using the framework proposed by WGMG at its 2004 and 2006 meetings.	ACFM
2. Meetings of WGMG should extend for a period of 10-15 days, if possible, in order to ensure adequate time for the development of methods and collaborative working.	RMC
3. WGMG proposes that the group meets at the Northeast Fisheries Science Centre, NOAA, Woods Hole, USA.	RMC
4. The establishment of targets and conservation reference points will be necessary in the evaluations of management proposals. However, such evaluations are not merely scientific endeavours but will necessitate a dialogue and detailed discussion with managers and stakeholders. It is unclear how ICES fits into this process and this needs to be rectified as a matter of urgency! ICES should be a partner in the whole process but may also play a role in the co-ordination and supervision of evaluations.	MCAP ACFM