Northwest Atlantic Fisheries Organization

Serial No. N5697 NAFO SCS Doc. 09/25

Scientific Council Meeting – September 2009

Report of the ad hoc Working Group on Assessment Methods for
SA2 + Div. 3KLMNO Greenland Halibut, 1-3 June 2009

Edited by
Michael C.S. Kingsley
Greenland Institute of Natural Resources, Nuuk

Introduction

1.1 Requests for advice, and previous activities

The stock of Greenland halibut in SA2 and Div. 3KLMNO has for some time been a subject of concern. Scientific advice on the management of the stock has for some years continued to be based on an Extended Survivor Analysis (XSA) model of the stock dynamics and the assessment of the stock has been carried out in a consistent manner since a reformulation was introduced in 2003 (Darby et al., 2003). The 2003 assessment prompted the NAFO Fisheries Commission to put in place a fifteen-year rebuilding plan with the objective of attaining a 140 000 t exploitable biomass by 2019. Two alternative assessments to the XSA, based on ADAPT and ASPIC, were presented in 2004 (Darby et al., 2004). Estimates of population size and fishing mortality from these models were found to be reasonably consistent with the XSA results and they were therefore interpreted as supporting the XSA-based advice. Subsequent assessments have shown little improvement in stock status, although recent estimates of fishing mortality have decreased somewhat. Following a series of reductions in the TAC agreed by Fisheries Commission, TAC since 2007 has been fixed at 16 000 t. Present fishing mortality is estimated to be over twice $F_{0.1}$, and Scientific Council has advised reducing the fishing mortality to $F_{0.1}$. This advice has not yet been followed by Fisheries Commission.

Over the most recent three years both fishery catch rates and most survey indices of biomass have increased steadily and markedly, fishery catch rates to levels higher than any since the early 1990s, survey indices more modestly and not yet reaching the values of the late 1990s. The analytical assessment of the stock using the accepted XSA method has continued to estimate rather constant, low, biomass levels which appear inconsistent with these recent trends in the biomass indices (e.g. Healey and Mahé 2008).

At its annual meeting in September 2008, Fisheries Commission made an additional request for advice from Scientific Council with respect to Greenland halibut in SA 2 and Div. 3KLMNO, as follows:

‘10. With respect to Greenland halibut in SA 2 + Div. 3KLMNO, Fisheries Commission requests Scientific Council, in its 2009 assessment of this stock, in addition to the information requested above:

‘a) To complete an evaluation of alternate assessment models for this stock. This evaluation will enable the determination of the robustness of the assessment model currently used.’

[1]*Robustness could be defined as relative insensitivity of the outputs that are of particular interest to reasonable alternative choices of data inputs—i.e. ’data’ robustness, or to reasonable alternatives in model structure or assumptions, i.e. ’model’ robustness.*

The Working Group discussed candidate criteria for the evaluation of models, which might include:

- a. *robustness, possible determined by sensitivity analyses or retrospective patterns;*
- b. *balance between the number of parameters and the number of data points;*
- c. *goodness of fit to the data; however, problems arise in compare models with different measures of fit, and in allowing for the effect on fit of the number of parameters; information criteria might help to solve this problem;*
- d. *experience that model predictions have been consistent with the biology of the species and with observations not used in fitting the assessment model;*
- e. *precise estimation of key stock parameters, not ascribable to unrealistically precise assumptions or to use of inappropriately restrictive subsets of the data. ‘Key’ parameters for the status of this stock might include biomass or numbers of several possible age ranges, including small fish, recruits to the fishable stock, age 5–9 (about 80% of the catch by weight), spawning stock (or age 10+); recent* F*; biological reference points for the stock.*

[1] *italicised text represents excerpts from the discussions of the Working Group.*

The purpose of assessments is to predict the response of a stock to future fishing. So the expected ability of the stock to withstand fishing pressure is a significant parameter of its status. Biomass, or numbers, at ages 5–9 may be the most useful window for comparison for this stock, as we have the best data for these ages.

Spawning-stock biomass is an important variable for stock management. However, the uncertain estimation of 10+ biomass frequently recurred in the Working Group's discussions. These age classes are poorly sampled both in the (dominant) trawl fishery and in surveys and estimates of both the number and the biomass of fish in them tends to depend on uncertain extrapolations of fishery and survey selectivity curves. Differences among the estimates of spawning stock biomass, and even of total biomass, from different assessment models are sensitive to assumptions about mortality and relative catchability at ages at which fish are scarce.

In its own request for advice from the 2009 meeting of Scientific Council, Canada added to the Fishery Commission request the following:

'the Scientific Council is also requested to consider alternative formulations of any assessment models it evaluates that would include acceptable fishery-based CPUE indices.'

The use of 'acceptable' to qualify 'fishery-based CPUE indices' is important. Scientific Council and STACFIS have repeatedly expressed reservations about using fishery CPUE as an index of numbers or biomass. Partly because CPUE reflects only density in fished areas; partly because even there, there is no certainty that its relationship to biomass is constant over time or different fishing conditions. Some experience shows that CPUE can be maintained on declining stocks, especially if the distribution of the fishery is restricted.

In March 2009 Scientific Council initiated the setting up of a Working Group to carry out the evaluation that had been requested in the previous September. Terms of Reference were established before the meeting and were available to the participants. (Appendix I). The Working Group comprised stock assessment scientists from Contracting Parties as well as external expertise on stock assessment methods (Appendix II). The Working Group met in Dartmouth, N.S. on 1–3 June 2009, i.e. on the 3 days before the start of the June meeting of Scientific Council.

1.2 Assessments of the SA 2 & Div. 3KLMNO stock of Greenland halibut

The assessment of this stock has never been easy. It covers a wide area and there is reason to believe that it is biologically continuous with neighbouring stocks to the north. It also inhabits a wide range of depths, including depths that are not reached by trawling. Therefore, assessment assumptions of a closed stock and of survey sampling that has constant age-specific catchability from year to year may be invalidated by movement of fish across stock boundaries or between depths. The XSA-based assessment has had a somewhat troubled history and Scientific Council has in the past expressed reservations about the confidence that could be placed in its conclusions (Healey 2009). Initial VP analyses with ADAPT gave correlated estimates of parameters and evidence that fishery selection patterns had changed over time. Initial XSA in 2000 gave a 'relatively poor' fit but was accepted and built upon in the next year. In 2002, an updated XSA estimated a biomass trend judged inconsistent with stock size indicators and was not accepted. Addition of the 2002 catch and survey data caused XSA to give different estimates in 2003, which were considered more consistent with perceptions of stock dynamics qualitatively inferred from survey and fishery observations, and which *were* accepted. However, in 2003 it was also decided that survey data only from 1995 and later should be used, the calibration of the earlier Engel data to Campelen equivalents being abandoned, and this greatly improved the fit of the model; catch data from before 1975, when the fishery had been exploiting grounds not fished since then, was also excluded.

The sensitivity of the XSA was thoroughly examined in 2004, and it was concluded that the XSA-estimated trends in the stock dynamics were 'robust to the data series used for the fitting of the model, inclusion or exclusion of the historic Canadian autumn survey data as a separate series, and choice of plus group', and that different catch levels considered for 2003—about which there was some doubt—did not affect perception of stock status (Darby et al. 2004). The 2004 assessment also presented results obtained from ADAPT—which agreed with XSA results—and from ASPIC, an age-aggregated stock-production model. More recently, further sensitivity analyses and examination of the consistency of data have been carried out, with results sometimes equivocal. However, in general, investigations have found that trends in estimated exploitable biomass and estimated average fishing mortality were not much affected when different model settings or data sets were considered, so the XSAs might be argued to be mutually consistent; and other model approaches (ADAPT and ASPIC) have also given similar results. However, VPA-type models will, because of the way they are structured, tend to converge backward to a common historical trajectory for the stock, so this consistency is to be expected and might not be a foundation for confidence.

Age-structured assessment models assume that the dynamics of a stock will follow a path that is consistent with birth and death processes, with decreasing numbers in cohorts with time. Recent survey age-structure data for this stock shows some cohorts of older fish to be more numerous than they were in earlier years, where stock-dynamic assumptions predict that cohorts should decrease in numbers with time. Some cohorts have at certain ages declined faster than can be explained by observed catches or a reasonable assumption for the natural mortality. There is a conflict inherent in the data between recent trends in biomass indices and the age structure of survey catches. These features of the data cause difficulties in fitting age-structured models. Different age-structured models, with different data weightings and assumptions, will come to different conclusions.

In a short presentation on how ICES goes about investigating and approving assessment methods, the Working Group was informed that ICES has rigorous procedures for doing this as well as formal benchmarking procedures. The purpose of the ICES Working Group on Methods of Fish Stock Assessment (WGMG) is to develop and critically evaluate the models and software code used in assessments, forecasts and management simulations, and to suggest ways in which these might be improved. WGMG meets to address particular concerns raised by ACFM and the Resource Management Committee of ICES. The issues covered by each meeting are a function both of the Terms of Reference, and of the interests and expertise of the participants. ICES has also recently planned and held workshops for evaluating data and benchmarking assessments for groups of stocks, at which data has been reviewed and different assessment methods considered and evaluated. The results of these processes are peer-reviewed assessment models which are accepted as standard, and peer-reviewed assessments.

Scientific Council should raise its awareness of the problem of evaluating assessment models; and perhaps review its methods of testing and approving them.

A working paper was presented on comparing models. The WP displayed a very structured approach, varying different factors, notably the form of the selectivity curve and the selection of data, whether fisheries or survey, and then systematically comparing the fits of different models using likelihood or an information criterion.

Different models use different fitting methods and evaluated their fits internally by different criteria, so that it might not be easy to arrive at good standard methods of comparing model fits. Also, some methods (rely on age-specific data) would find it difficult to use e.g. CPUE data unless it could be made available in age-specific form. There is enough divergence between the construction of different models that bringing them sufficiently onto a common footing to be able to compare them with simple quantitative indices will not be a simple procedure.

Models (and assessment models) typically fit the data imperfectly to the data and therefore produce residuals. But the types, and patterns, of residuals depend on the error structure that is implicit in the model. Therefore it may be difficult to compare models in a systematic way.

1.3. The stock and previous assessments

1.3.1 Stock history: principal data series.

1.3.1.1. Catch history.

The catch history since the early 1970s is principally marked by a period of high catches averaging 58 000 t/yr from 1990 to 1994; other fluctuations pale in comparison, although even outside this period catches have varied by more than a factor of 2. High catch levels near 40 000 t/yr were reached in 1978 and 2000–2003; the lowest levels, below

20 000 t, before and after the high-catch period in 1990–94. The period of high catches was associated with sharp declines in both the biomass index and the mean biomass age in the Canadian autumn survey.

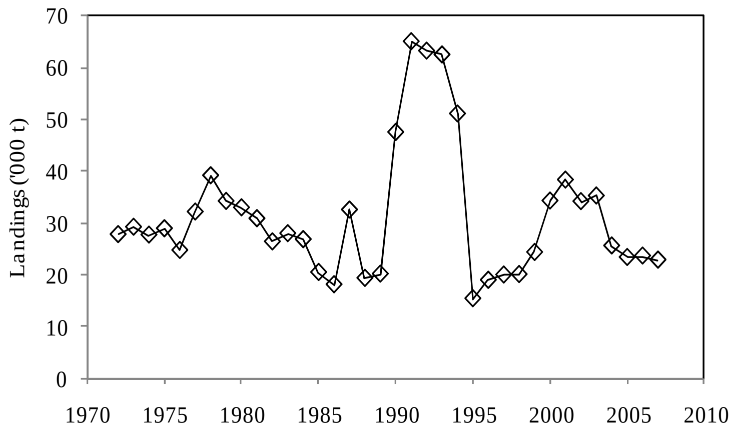

Fig. 1: Greenland halibut in SA 2 & Div. 3K–O: agreed landings 1972–2007.

1.3.1.2 Stock-size indices.

1.3.1.2.1. Survey.

There are three survey series accepted as providing biomass indices for the stock: they are the Canadian autumn survey in Div. 2J and 3K, the Canadian spring survey in Div. 3LNO, and the European Union summer survey in Div. 3M. They do not cover the same areas or depths, are not executed at the same time of year, and catch different age distributions[2]. The gear used for the Canadian surveys was changed in autumn 1995 from an Engel high-lift trawl to a Campelen shrimp trawl. The earlier Engel series was relatively stable from the late 1970s to the mid-1980s, but decreased steadily, by, overall, a factor of about 4 from the mid-1980s to the mid-1990s. Concerns about the stability of catchability in this survey led to its being omitted from the assessment in 2003 and since.

The three series that have been run consistently since 1995 show similar trends and are well correlated. After the catches decreased in 1995, survey biomass indices doubled to the late 1990s, then decreased equally abruptly to 2003 before increasing again to 2007.

[2] The age of the average biomass caught in the Canadian spring and autumn surveys appears to be about 4½ years while the EU survey in Div. 3M appears to catch biomass about 6½ years old. The fishery catches biomass that averages near 8½ years old.

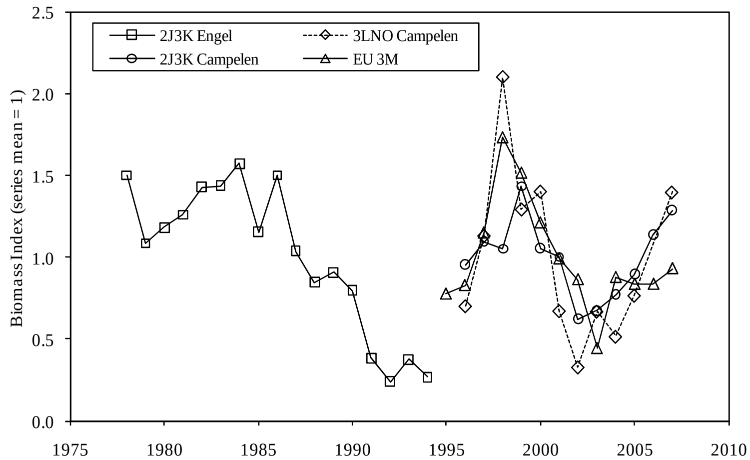

Fig. 2: Greenland halibut in SA 2 & Div. 3K–O: survey biomass indices.

The increases in the Canadian survey biomass indices have been associated with increases in the mean age of biomass. In the late 1990s this was apparently due to growth of cohorts that had already been detectable. The recent increase in these biomass indices since about 2001 has been associated with a steady increase in the mean age of biomass as though due to ageing of existing cohorts, but cohort structure is not clearly visible. The appearance of numbers of fish aged 6–10 in the most recent years was not presaged by visibly strong cohorts at younger ages in earlier years. The EU survey catches fewer young fish and the mean age of biomass is greater and more stable.

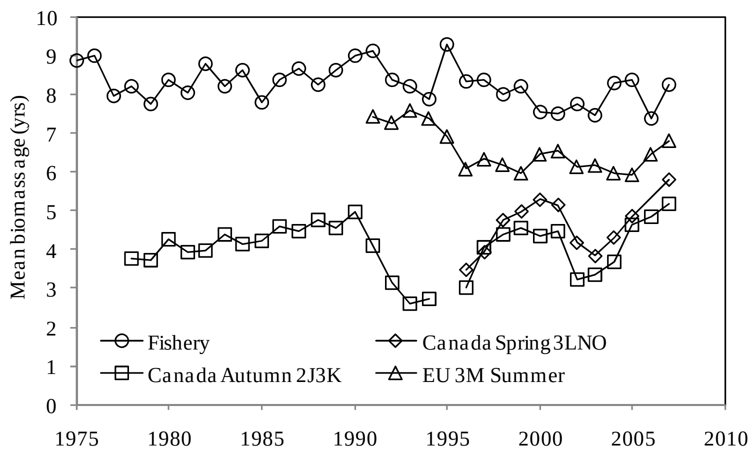

Fig. 3: Greenland halibut in SA 2 & Div. 3K–O: mean biomass age, 1975–2007.

1.3.1.2.2. CPUE

CPUE data is available for three trawler fleets: the Canadian, Portuguese, and Spanish. CPUE indices qualitatively follow similar trends to those of the surveys: high levels at the end of the 1970s and in the early 1980s, with a decrease to low levels near the mid-1990s. A short period of higher levels near the end of the 1990s or at the start of the 2000s was followed by a decrease to lowest levels in the early 2000s and a final increase in the most recent years. Agreement between the three CPUE series that are available since 1992 is not as good as that between the three survey series, and up to 2004 the average CPUE index was more stable than the average survey index, with smaller root mean square serial difference (0.13 compared with 0.27), smaller serial correlation (0.60 vs 0.70), and smaller standard deviation (0.15 vs 0.36).

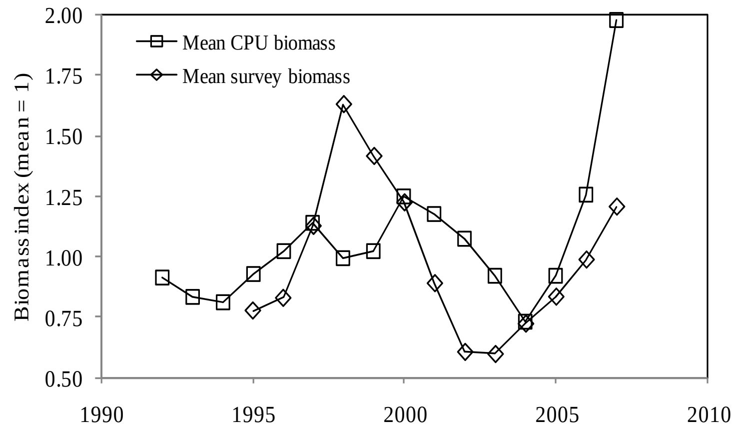

Fig. 4: Greenland halibut in SA 2 & Div. 3K–O: unweighted mean biomass and CPUE indices, 1992–2007

Since the early 1990s, the average survey index has little correlation with the average CPUE, and that little is almost entirely due to the data for 2004–2007, in which both the average survey index and the average CPUE index have increased at accelerating rates; the coefficient of determination between them in this period is 98%. In earlier years, the correlation is poor. It is the simultaneous and similar increasing trends in both survey and CPUE indices, together with the apparent failure of the XSA assessment to reflect them, that has prompted the present enquiry.

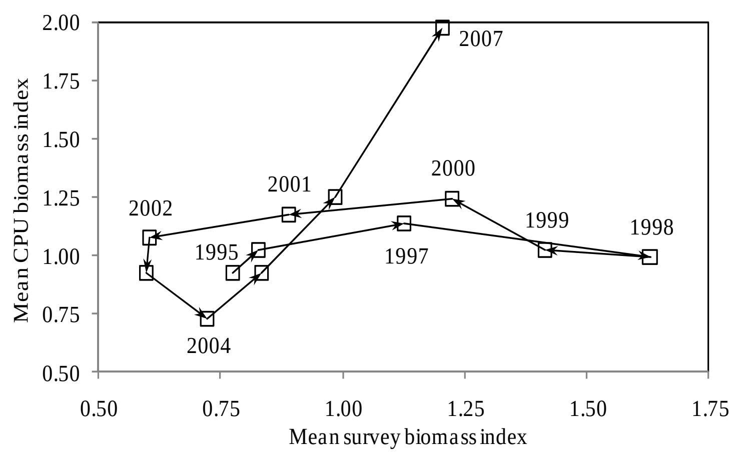

Fig. 5: Greenland halibut in SA 2 & Div. 3K–O: mean biomass indices from all fleet CPUEs and all surveys, 1995–2007.

The standardisation of CPUE has been reviewed in the present context (Brandão et al. 2009), and it appears from that revision that the recent increases in nominal CPUE may be partly due to re-distribution of the fishery into areas of greater density. CPUE data shows the same recent increases as survey data. In years when the Canadian autumn survey averaged over 20 kg of Greenland halibut a tow the Canadian standardised CPUE in Div. 2HJ3KL was roughly proportional to the survey catch—indicating that at high levels of both, CPUE does indeed index stock size—but when (in 1991–95 and 2002–04) the survey catch averaged less than 20 kg a tow the CPUE was more nearly constant.

While this non-linearity suggests that it is more difficult to include CPUE data as tuning indices over the full range of stock size, they can still be useful to understand the fishery more fully.

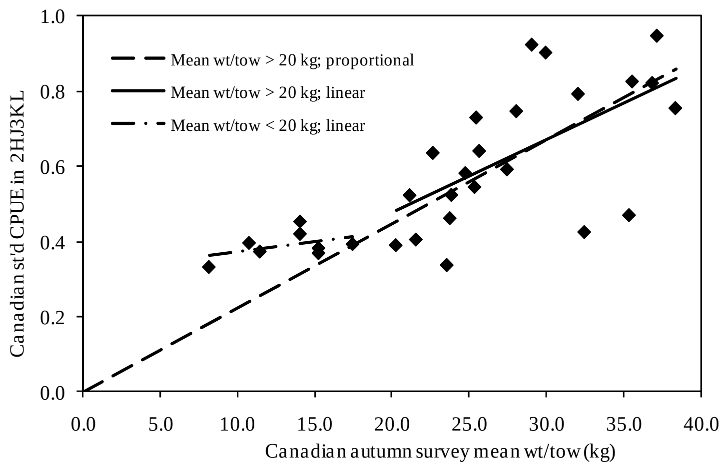

Fig. 6: Greenland halibut in SA 2 & Div. 3K–O: Canadian fleet CPUE in 2HJ3KL and Canadian autumn survey biomass, 1975–2007.

1.3.2.Analysis of data on the age structure of the stock (Morgan and Shelton 2009).

FLEDA, a facility in FLR, facilitates analysis of the age structure of fishery and survey catches and graphical presentation of results. The tools and analyses available in FLEDA have previously figured in assessment documents for this stock. The main conclusions presented to the Working Group from age-structure analyses for Greenland halibut in SA 2 and Div. 3KLMNO were:

a) Fishery catches have been mostly in the age range 5–10 years, the majority from 6 to 8 years. Ages were greater in earlier years with more 7–8-year-old fish in 1975–1990, more 6–7-year-olds in 1995–2007. Since about 2000 there have been few fish more than 7 years old in fishery catches. Two-dimensional plots of numbers against both age and time showed little cohort structure in fishery catches.

b) There are 3 survey series: the Canadian spring survey, the Canadian autumn survey, and the EU-Spain survey in Div. 3M only. These surveys do not cover the same depth ranges or areas, but their results are generally well correlated. In earlier years they had closer correlations; in the most recent years they have diverged more.

c) Survey data shows few traceable cohorts, and those few can be traced only up to about age 5 or 8 (depending on the survey), but not beyond; cohort structure seems not to be maintained after that age.

d) All the surveys show, in the most recent 1 or 2 years, an appearance of numbers of fish more than 6 years old, which apparently cannot be traced as younger fish in earlier years; this looks as though these older fish are migrating in from elsewhere or as though some other process might be ‘generating’ these cohorts. These irregularities in age-structure data are a key cause of difficulty in this assessment.

e) Correlations in the survey data between numbers at age and corresponding numbers at lesser ages in earlier years appear generally to be positive up to about age 6–8. For data on older fish, many of the correlations of this kind are negative or only weakly positive.

f) Some survey age-structures show lower total mortality in recent years than earlier.

2. Assessment Models presented to the Working Group:

The Working Group reviewed the models available and considered their properties and the results they gave, before proceeding to evaluate them and consider their implications for the robustness of the XSA. The models and outputs available for presentation were:

1. Preliminary data exploration using the 'Exploratory Data Analysis' facility available in the 'Fisheries Laboratory in R' (FLEDA) (see 1.3.2. above);
2. XSA–the existing assessment model; the Working Group also had available to it information on the history of assessments of this stock using this model;
3. Statistical Catch at Age (SCAA) model developed and fitted by a consulting group; a large number of formulations, some highly parametrised, had been fitted to the data and were available for review;
4. An age-structured production model also developed and fitted by the consulting group;
5. A stock-production model (ASPIC) with few parameters;
6. ADAPT VPA models:
    a) ADAPT formulations with different treatments of errors;
    b) a B-ADAPT formulation;
    c) sensitivity runs on the behaviour of ADAPT/VPA models;
7. SURBA: Survey-based assessment, a review of population structure and evolution based only on a quantitative analysis of age-structure data from surveys.

2.1. Statistical Catch at Age (SCAA).

A large number of formulations of the SCAA had been elaborated over the history of the investigation. These are age-structured models, in which the total catch is (for most implementations) assumed known without error, but the observed age structures of the fishery and survey catches have associated error variances which are estimated by the fitting process. The catchabilities at age in the survey, as well as fishery selectivity, which is estimated from cohort numbers in the catch, are not constant over time, but their variabilities are constrained. Natural mortality can vary with both age and time, and recruitment is estimated from a stock-recruitment curve; its two parameters could be estimated by the fitting process, but in most formulations one was fixed externally; variability about the S-R curve was fixed external to the fitting process. The model could be fitted, by maximising a penalised likelihood function[3], to indices of total biomass from survey or fishery or both and to proportions at age from survey or fishery or both. The user could assign weightings to different sets of input data.

A first set of formulations using survey data comprised a 'baseline' formulation with variations that included provision for serial correlations in survey catch-at-age residuals, different levels of time variability in fishery selectivity, and fishery selectivity flat from age 10 (Butterworth and Rademeyer 2009e). A higher level of time variability in fishery selectivity gave a markedly better fit than other formulations in this set.

A second set comprised the same baseline as included in the first set, a 'revised baseline' that included serial correlation in survey residuals as well as variability in catchability at age in both fishery and survey, and variations on this revised baseline that included greater variability in catchability at age in either fishery or survey (Butterworth and Rademeyer 2009c). Other formulations varied the form of the stock-recruitment relationship (especially its 'steepness') or allowed uncertainty in total catches or variability in natural mortality. A final group used the three survey series one at a time instead of all together.

In many cases, allowing variability or uncertainty over time improved the fit of these models, perhaps indicating irregularities in the age structure data. A salient difference between the SCAA formulations so far described and the XSA accepted for assessments of the stock is that, consequent upon a decision in 2003, the latter uses survey data from 1995 and catch data from 1975, while the SCAA models—with the inclusion of a change in catchability when

[3] This approach is fundamentally Bayesian, with the penalties placed on the likelihood function representing informative prior distributions; full Bayesian estimation demanding much computation, the results presented were only posterior joint medians.

the gear used in the Canadian autumn survey was changed—used survey data back to 1975, and catch data back to 1960.

The SCAA model was further investigated using CPUE from three fleets, instead of survey catches, as input biomass indices (Butterworth and Rademeyer 2009a, d). A 'baseline CPUE' variant was established with higher variability in the fishery selectivity, and similar variants to those of the survey-series formulation were run. Better fits were obtained when recruitment was allowed to vary more about the prediction from the stock-recruitment curve.

The above sets of SCAA formulations had been presented to and discussed at a meeting in late April, and some further investigations had been recommended by the meeting. SCAA formulations that used CPUE series derived from more complex standardisation models that included interactions between year and Division or depth were also run, but did not fit better. A number of other formulations were also presented to the Working Group, starting from the 'revised baseline' mentioned above and including, initially, an extension of the modelled age range to 20 yrs and serial correlations in survey catch at age(Butterworth and Rademeyer 2009b). A 'new baseline' was established that included such serial correlations over both age and time, with 515 parameters fitted to 924 data points and with 453 associated priors. Further variations on this 'new baseline' were investigated, including adjustments to 1990–94 catches, different fishery selectivities, and different variabilities in both fishery and survey catches. Runs were also carried out using only the data that is presently used as input to the accepted XSA model, reducing the number of data points by 210. The assessment was also more radically reformulated as a production-type model, with no uncertainty in recruitment and no variation in selectivity, although necessarily age-structured to accommodate different age ranges applying to the different surveys and to the fishery.

Much effort in the formulation of the SCAA models went into taking due account of serial correlations in the data, evident in the residuals from model fits. Among the conclusions were that a major problem for this assessment is a conflict, especially in recent years, between trends in indices of abundance and the information contained in the age distributions of the survey catches. A wide range of estimates of absolute biomass is one consequence of such a conflict.

Properties of SCAA:

- age-structured model, biomass-oriented;
- total catch is assumed error-free in most implementations;
- the reported age distribution of the fishery catch is an uncertain estimate of its true age distribution, the uncertainty being expressed by a variance parameter;
- the reported age distribution of the survey catch is an uncertain estimate of the age distribution in the stock, the uncertainty being expressed by a variance parameter;
- survey selectivities at age in are not constant over time (although their variability over time is constrained);
- natural mortality can vary with age and year;
- fishery selectivity is estimated from catch and cohort numbers and can vary over time (although its variability over time is constrained);
- recruitment is generated from a stock–recruitment curve, of which the two parameters could be estimated in the fitting process, but mostly one was fixed externally; the variability of recruitment about the S–R curve is input externally to the fitting process;
- fitting, by maximising a penalised likelihood, to observed indices of total biomass from survey or fishery or both and to observed proportions at age from survey or fishery or both;
- the user can specify weights to be given to different data sets;

Available outputs: numbers at age over time; selectivities, catchabilities, estimates of biological reference points such as F_{msy}, B_{msy} (because the stock-dynamic assumptions include a stock-recruitment relationship).

Salient characteristics of the results obtained: the results track the recent increases in survey and CPUE indices of total biomass, even in cases whose results include lower estimates of recent biomass; there appears to be a bimodal solution space with a knife-edge change from one domain to another with respect to some parameter settings, which probably reflects a conflict between survey biomass indices and survey proportion at age; when formulated with the same data as used for the XSA, gives similar 5–9 biomass levels but more positive trends and greater resilience to future harvests; generates a large ‘cryptic’ biomass of older fish because of certain features of the selectivity curves;

Good points: a flexible model which fits the data well; estimates of biological reference points; permits flexible exploration of the effects of assumptions; permits adjusting fitting criteria to allow for serial correlations in residuals; statistically comprehensive;

Bad points: clear patterns in residuals for survey catch at age, probably because of too much weight given to it; the model is, *prima facie*[4], very highly parametrised (simple number of parameters up to more than 50% of the number of data points—this is the reverse of the advantageous statistical comprehensiveness) although this is tempered by the use of informative prior distributions; the behaviour of the model is sometimes difficult to interpret.

In some regions there are large changes in values of some output parameter estimates sometimes with small changes in the value of a single input parameter.

SCAA is geared to respect biomass indices. It therefore pays more attention to trends in biomass indices and will estimate the effect of future catches by considering the effect of past catches on biomass indices. But it does this in an age-structured way, differently from the age-aggregated approach of ASPIC. Where trends in biomass indices conflict with catch-at-age data, the SCAA fit deteriorates, but it can resolve this by allowing wider uncertainty in catches at age or by allowing serial correlations in survey catch-at-age data. However, what SCAA is doing is treating the conflict as the result of errors in the observations, i.e. observations that do not properly reflect the true stock structure. If in fact the problem is that the true stock structure itself is not behaving in accordance with the model assumptions, results will be in error, as indeed will be the case for most other analytical assessment models in such circumstances.

‘Some conflict remains amongst the different sets of input data, and partly in consequence the absolute scale of biomass is poorly determined by assessments. The most pessimistic (in stock status terms) of the SCAA variants considered produce biomass estimates that do not differ very greatly from those given by the XSA. Importantly however, even in those cases the SCAA assessments provide results for recent years more in line with the general trends in survey and CPUE indices, and give more positive projections for future abundance: for example all SCAA biomass projections under a constant TAC of 22 750 tons increase, whereas XSA projects a decrease in those circumstances.’

This SCAA appears highly parametrised. There is a narrow region of compromise between model simplicity and model reality that results in good estimates, and there may be reason to doubt that the SCAA is in that region.

Residual concerns were expressed about some parameter settings. For example, a ‘steepness’ of 0.9 for a Beverton-Holt stock-recruitment relationship was thought to be high in terms of experience with other flatfish stocks, although favoured by the fitting process. In projections, it appeared that present levels of recruitment were insufficient to maintain the stock in the first 5 projection years, after which the S/R curve with its high steepness allowed the stock to increase.

[4] For a proper evaluation of the parametrisation of a model, the simple number of ‘estimable parameters’ ought to be adjusted to allow for informative prior distributions assigned to them, which reduce the extent to which they are freely ‘estimable’ and therefore their contribution to the parametrisation. The extreme of an informative prior is a single set value, in which case the parameter is no longer included in the estimable set. Effective parametrisation can be measured by, for example, the ‘Deviance Information Criterion’.

2.2. VPA models and analyses.

2.2.1. Extended Survivor Analysis

XSA was conceived in order to palliate certain perceived problems with the performance and 'tuning' (i.e. calibrating with series of biomass indices) of forward VPA models and as an improvement on existing survivor methods (Shepherd 1999). The stock trajectory is defined in terms of survivors at the last modelled age and the last year and increasing cohort numbers are calculated backward in time by raising and adding reported catches, so that problems in forward VPA models associated with catches greater than cohort size are avoided. From a (starting) set of values for the number of survivors at the last true age in all years and at lesser ages in the last year, a reverse VPA with assumed natural mortality and catches at age by number assumed known without error generates estimates of stock numbers at age in earlier years, and therewith estimates of fishing mortality. Stock numbers can be compared with age- and fleet-specific series of CPUE[5] values to obtain age- and fleet-specific weighted average catchabilities (and their associated uncertainties, which are later used for weighting). From fleet catchabilities and CPUEs, fleet-specific series of numbers at age are generated, which through the application of cumulative total mortality generate estimates of survivors to be combined by weighted geometric averaging to form a new set for re-starting the cycle. The cycle is repeated until a stable stock trajectory is achieved. The model fit criterion is based on a the squared differences between two sets of numbers at age over time: one derived from the deterministic backward VPA starting with estimated survivors and using the exploitation catch-at-age series; and the other derived from input series of fleet- and year-specific CPUEs using fleet- and age-specific catchabilities which are estimated in the fitting process.

Properties of XSA

- age-structured and number-oriented;
- catches are assumed free of error, also by age;
- catchability at age in survey data is constant over time;
- natural mortality (input) can vary with age and year;
- fishery average, and fleet, selectivities can be estimated by comparing age distributions in the catches with those estimated for the stock; can vary over time;
- recruitment is generated by the backward VPA calculation as the XSA estimate of numbers at the first recruited age, or ages;
- fitting by iteratively re-weighted least squares until no change in survivors; inverse-variance weighting;
- terminal-year F can be constrained to an average of those for some recent years and the terminal-age F can similarly be constrained to an average of those for immediately previous ages; (user-selected settings);
- no 'accumulating' plus-group;
- plus-group data is not used in fitting the model (selectivity assumed equal to that for last true age.)

Outputs: numbers at age over time in the stock; fishing mortality by age and over time; present stock size (numbers at age); age-specific survey catchabilities; fishery catchabilities.

Salient characteristics of the results generated: the shrinkage (*F* averaging) in use for this assessment interferes with the model's ability to track the increasing trend in survey biomass indices of the most recent 4 years; results are 'converged' (backward in time)—(if certain parameter settings are retained)—so giving a (possibly false) impression of robustness.

Good points: the fit to the survey data is acceptable given the assumptions used.

Bad points: there are marked retrospective patterns; shrinkage problem (although one effect of using shrinkage is probably to smooth the retrospective pattern); plus-group information is not used in the model-fitting process; there are large error coefficients of variation on survey catchabilities for some ages.

[5] can be survey age-specific catch:effort ratios.

Review of XSA assessment

'XSA can be regarded as a practical and useful method of analysis, unless the catch-at-age data are of very poor quality relative to the c.p.u.e. or survey data. The statistical model used is however substantially simplified, particularly in respect of error structures and estimates, and the method should therefore perhaps be regarded as adequate rather than optimal.

'Behavioural variations of catchability with age are not excluded, and may cause the results to be biased. This is a fundamental problem, and not easy to solve.' (Shepherd 1999.)

Is the simplification of the fitting criterion in the XSA algorithm sensitive to the characteristics of the data on this stock, with its uneven cohort histories? There has been presented to the Working Group little evidence that the catch-at-age data are *of 'very poor quality' (although this has previously been a topic of discussion). The CPUE data (which for this XSA assessment is restricted to survey catch-rate data) is presumably reasonably good, and is in any case modelled as subject to error*.

2.2.2. VPA and ADAPT assessment models

Other VPA models than XSA have been applied to this assessment: ADAPT and separable VPA were used for early analytical assessments in 2000, and in 2004 results from ADAPT were compared with those from XSA in a thorough investigation of the properties of the assessment (Darby et al. 2004). For the present workshop, the 2008 assessment data was run with different ADAPT formulations: one from NOAA Fisheries Toolbox (NFT VPA v. 2.8) and a bias-correcting version of ADAPT (B-ADAPT, used by the ICES Working Group on the North Sea and Skagerrak). ADAPT v. 3.1 was also run to investigate the effects of different settings for fishing mortality at the greatest modelled ages. In comparing the results of different VPA formulations one must keep in mind that the stock trajectories they estimate inevitably converge backwards in time.

NFT ADAPT with inverse-variance weighting gave similar results to those of XSA with no shrinkage, but XSA with shrinkage gave lower estimates of recent stock size and unweighted NFT ADAPT gave higher estimates. Unweighted NFT Adapt and XSA with no shrinkage fitted better than the other formulations, although measures of fit are not fully comparable with one another. None of these formulations reduced the trends in Canadian survey-index residuals for the most recent ten years. Age-specific residuals from the 2008 XSA model were to visual examination not different from those given by the NFT VPA model.

B-ADAPT, estimating a bias in the last ten years' landings, improved the fit appreciably. B-ADAPT apparently gave residuals that for the Canadian autumn survey had less serial correlation and for the Canadian spring survey had less scatter (the model fit tracked the surveys better). However, the 'bias-corrected' catches were outside any reasonable estimates, showing that the observed trends in survey residuals were probably due to changes in survey catchability. Since the surveys had constant sampling methods, these changes were presumably related to changed availability of fish of certain age-classes to the gear.

ADAPT v.3.1 was also run with two different constraints on fishing mortality at the greatest modelled ages; one constrains F on the oldest true age, usually to some multiple of F on the next lower age, and specifies an F on the plus-group in the first modelled year, and the other sets 'plus-group' F as a multiple of the last-true-age F. The estimates of the population dynamics from the two formulations were very consistent, with similar trends in biomass and fishing mortality. Differences were greater in the earliest years and at the greatest ages, when the different constraints on the high-age Fs had the most effect.

Retrospective analyses showed considerable divergences from the trajectory modelled for the full period.

2.2.3. An enquiry into the performance of ADAPT.

Multiple sets of input data were generated and input to ADAPT formulations which were run to simulate assessment results (Vazquez et al. 2009). The object was to check how close assessment results were to true values when the data had low, or high, dispersion. Simulated populations were defined with specified properties and sets of assessment data—fishery catches at age and survey indices, also at age—were randomly generated. The assessment data was then input to ADAPT with different formulations, including options for calculating survey catchability at age, partial recruitment, last-age fishing mortality, &c., and the specification of objective function to be minimised.

First results indicated underestimation of survivals; the best agreement was achieved when the numbers of survivors from all cohorts were used as parameters.

2.3. ASPIC stock production model.

A stock–production model, ASPIC, was run with age-aggregated data. The first runs repeated the configurations used as part of the assessment process in 2004; the more definitive later runs used a more recent version of ASPIC, and limited catch data to 1975 and later and survey data to 1995 and later. The three survey data series were input as biomass indices, as gross weights in the survey catches regardless of age structure. The fitting process converged to a well defined solution A with reasonable standard errors of estimation for the stock-dynamic parameters. However, the solution space had more than one mode, and some combinations of starting values for the fitting algorithm led to a different solution, B, which was also well defined although the fit was slightly worse than that of solution A and the residual pattern slightly more marked.

Solution B had a 60% higher *MSY* than solution A. It also had a 50% lower B_{msy}, so its production changed more quickly in response to catch levels. So after the biomass was driven down by high catches in 1990–94—averaging about 58 000 t/yr compared with 28 000 t/yr in the rest of 1975–2007—solution B saw biomass recover to previous levels above 1.5 times B_{msy} as soon as catches decreased while solution A saw it remain depressed at about half B_{msy}. I.e. lower catches in 1975–89 were balanced in both solutions by low production from a biomass above B_{msy}. Then high catches in 1990–94 were supplied in solution A largely by a lasting reduction in biomass and only partly by higher production as the biomass trajectory passed downward through the maximum-productivity level, and in solution B by high production from a biomass temporarily dipping to near its maximum-productivity level but able to recover, by grace of its productivity, as soon as catches decreased. After that, lower catches in 1995–2007 were satisfied in solution A by production from a biomass well below B_{msy} and in solution B from a biomass well above it. Survey data from before 1995 was not included in these formulations, so it is difficult to differentiate between the two solutions, but it seems unlikely that the solution space contains a third mode of any significance. While both solutions fitted the limited data supplied to this model in these analyses, solution A fitted the better and was thought to be overwhelmingly more consistent with other available information on the history of the stock, including the trajectories of other indices of stock size, as well as estimations of the present state of the stock from some other assessment models, including the accepted XSA.

Retrospective runs were carried out with the omission of recent years' data from the inputs to the ASPIC model. The retrospective pattern from the ASPIC analyses was similar to that of the XSA assessment model, in particular as regards estimates of recent *F*. Retrospective patterns shown by the XSA assessment model have been thought to be—explained as—due to inconsistencies between recent survey catches at age and the expectations of age-structured stock-dynamic modelling. However, ASPIC is an age-aggregated model, and so its presenting similar retrospective patterns might cast doubt on this explanation.

The stock-production model was also run with fishery CPUE as input data. Although overall trends are similar between fishery CPUE and survey indices there are some differences in when peaks occur (survey peaking before CPUE), and as a result a number of negative correlations between biomass indices from survey and fishery CPUE. This causes difficulties when fitting age-aggregated stock-production models to all of the data series, and results of runs using both CPUE and survey were more similar to solution B, but with a poor fit. When only CPUE was input, the results were more like solution A.

Characteristics of ASPIC stock-production model.

- age aggregated model; simple stock-dynamic model with a small number of parameters;
- catch, age-aggregated, is assumed known without error;
- catchability in survey data is constant over time;
- natural mortality, presumed constant, is subsumed in a presumed stock-dynamic equation;

- selectivity, like other age-specific parameters, is not considered;
- recruitment, like natural mortality, is subsumed in a presumed stock-dynamic equation;
- fitting by iteratively re-weighted least squares;

Outputs: time trajectories of age-aggregated indices of stock size and fishing mortality; parameters of the stock-dynamic model (B_{msy}, *MSY*); estimates of catchability for the input biomass indices.

Salient characteristics of the results generated: when only survey data since 1995 is given as input, the solution space has two modes with different recent trajectories, one of which is not consistent with the conventional wisdom on the history and present status of the stock. There are significant retrospective patterns when CPUE data is used.

Good points: fit to survey data is acceptable given the assumptions used; a simple, robust model with simple results which are easy to interpret; robust against age structures that do not conform to conventional stock-dynamic equations;

Bad points: bimodal solution space (owing to the limited time range of input data); marked retrospective patterns;

A production model is 'all plus group' and has a hard time responding to changes in the survey and will not change fast enough to match the surveys. This will result in the type of retrospective pattern that is shown by these analyses.

The survey indices catch a different range of ages from that caught be the fishery and this 'mismatch' could be causing some problems.

Why are survey CPUE not standardized? The survey design is meant to 'standardize' the survey indices. It is true that there have been deviations from the survey design and these have increased in recent years with vessel breakdown in the Canadian survey. It might be appropriate to standardize using some sort of environmental variable but one would need to have an hypothesis as to how the environment would affect the survey CPUE.

Further analyses on ASPIC incorporating CPUE were conducted to examine the retrospective pattern and multiple solution problem when using CPUE only. The multiple solution issue also existed with the runs using CPUE; retrospective issues were still present and estimates were widely divergent. Some of the runs also started at low biomass relative to $\mathrm{B_{msy}}$, *which is unusual. Given the retrospective pattern the run with CPUE may not be reasonable. As data were removed negative correlations between the different CPUE series began to appear, indicating that perhaps the overall positive correlations were generated only by the marked increases in CPUE for all fleets in the most recent years. A plot of surplus production against the stock might show why the model estimates changed so greatly in the retrospective.*

2.4. SURBA: a simplified assessment based only on survey data

SURBA is a simplified assessment model based on a reduced, separable population-dynamic model and relying only on survey data, including age structure, but with no reference to fishery catches.

The separable model assumes that total mortality comprises only a year-specific element and an age-specific element, and assuming constant survey catchability these parameter sets can be fitted to time series of survey numbers at age.

The results of the SURBA analysis of survey catch-at-age data since 1996 showed that total mortality has been lower since 2003, and biomass at ages 5–9 has correspondingly increased, although not reaching the levels of 1998–2000: i.e. similar to the trajectory of Fig. 2. Spawning stock biomass, on the other hand, is estimated to have increased since 2003 to its highest level since 1996.

SURBA results are similar to the FLEDA findings on survey age-structure analyses.

2.5 Model results.

The Working Group only considered runs of different models with the available data for the stock under consideration, or different selections from that data, and did not compare different model types with synthesised data for which the true underlying population values were known. Therefore, although the Working Group considered that characteristics of the data, such as inconsistent cohort structure in consecutive years and conflicts between age structures and biomass trends, were likely to be at the root of the discrepancies between the results of different assessment models, it was not in a position to demonstrate conclusively which characteristics of the data were producing which differences in the outputs.

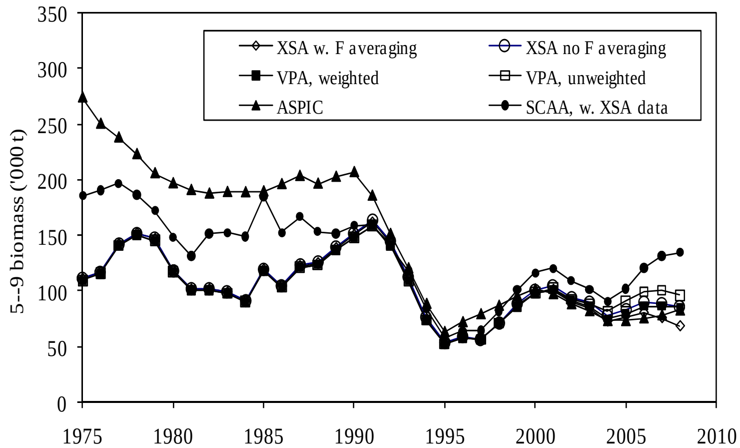

Fig. 7: Greenland halibut in SA 2 & Div. 3K–O: biomass trajectories ages 5–9 for 1975–2007 estimated by a suite of assessment models.

Some models had strong or erratic retrospective patterns, possibly showing inconsistencies in the data series. The results given by assessment models differed mainly in biomass levels in early years and at the oldest modelled ages. When we gave different models similar or the same data sets, their results converged to more similar values. However, there was still a divergence in trend in the most recent years, the XSA accepted in 2008 showing biomass estimates that were lower than those made by other models. This difference is largely due to an averaging of fishing mortality over the most recent years (‘shrinkage’) which is used in the accepted XSA in order to stabilise the results and reduce year-to-year variations which otherwise reveal themselves as strong retrospective effects. When the XSA was fitted without shrinkage, its estimates were closer to those of other VPA models.

The different VPA models converged back in time, following the same biomass trajectory up to 1997–98, when they began to diverge. From 2004 on, the divergences became marked, but all showed an increasing trend from 2004-2006 — greatest for the unweighted ADAPT and least for XSA with F averaging—followed in 2006–2008 by a decrease or a less marked increase. This common tendency to show an initially positive, but thereafter increasingly negative, trend over the most recent 5 years is not a reflection of the biomass trend indicated by the survey series, which show, on average, a steady increase since 2003. The accepted assessment model, XSA with F averaging, has the lowest biomass values of the VPA models for all recent years; without F averaging it is somewhat more optimistic. VPA models had a more positive view of the recent evolution of the stock if statistical constraints were relaxed, by removing F averaging from XSA or inverse-variance weighting from ADAPT, and thus allowing stock trajectories greater freedom to diverge from constraints on fishing mortality and conform to recent survey indices of stock size.

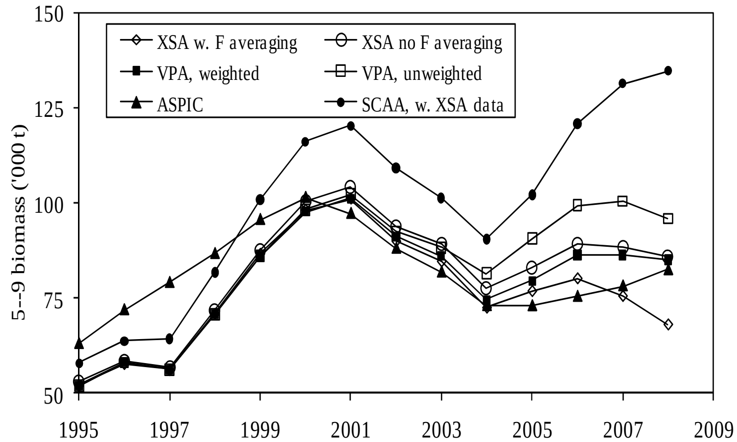

Fig. 8: Greenland halibut in SA 2 & Div. 3K–O: biomass trajectories 1995–2007 estimated by a suite of assessment models.

Most formulations of the Statistical-Catch-at-Age model estimated biomasses that were several to many times higher than those estimated by VPA-based models, including the accepted XSA[6]. There was some tendency for these differences to be most marked for the oldest population segment, comprising fish at least 10 years old, indicating that the SCAA model included a 'cryptic' biomass of older fish that was relatively larger than that included by VPA-type models assuming asymptotically flat selectivity. There were also differences between XSA and SCAA in trajectories of fishable (5–9 years old) biomass, which were however much reduced under certain settings for the SCAA model that were more similar to the XSA formulation: e.g. natural mortality set at 0.15/yr and also sensitivities starting in 1975. SCAA trajectories diverged greatly going back in time, which may indicate an effect of assumptions about starting conditions for the model. Higher recent and present biomass levels estimated by SCAA than by XSA were associated with greater resilience to proposed future catch levels.

An objective of the Working Group was to consider formulations of models that used fishery CPUE as input data. A revised standardisation of the fishery CPUE appeared to show that recent increases in nominal CPUE might have been partly due to relocation of the fishery to areas of higher density. When a CPUE series from the Canadian fishery was plotted against a series of survey results from the same area it appeared that at higher values of the survey index the CPUE was roughly proportional to it, but that at low values of survey index this fishery could maintain catch rates at higher-than-proportional levels. Some models could not be tried with the inclusion of CPUE as it was not available to the Working Group in age-specific form.

The ASPIC stock-production model was tested both with CPUE data alone and with CPUE and survey data together, but gave unstable results. The SCAA model is so structured that it can use age-aggregated CPUE. Its CPUE-based results differed from those of the XSA model in much the same way as when it used only survey data. This was thought to show that the differences between results from these models were due more to their different structures than to using only survey data as input to XSA.

[6] but neither the biomasss density estimates from XSA and nor those from SCAA with the 'New Baseline' formulation can be considered unlikely when compared with densities of Greenland halibut biomass estimated by surveys elsewhere (Rademeyer and Butterworth 2009).

There appears to be conflict between the survey indices and the catch-at-age data, and early indications of cohort strength appear to have poor predictive ability. It may be unwise to rely entirely on VPA methods such as XSA which (given their backwards convergence property) depend overwhelmingly on the reliability of catch-at-age data[7]*. More weight needs to be given to survey index trends.*

Survey index trends are by design not subject to many of the biases that may apply to CPUE. It is however important to note that CPUE in the most recent years has shown the same trends as the survey indices.

3. Conclusions

There are clearly anomalies in the data. There have been large year-on-year changes in survey biomass indices that could not be ascribed to the passage of identifiable cohorts through the stock, and in some years, catches from older year classes that were not seen in equivalent numbers as younger fish in earlier years. Conventional analytical assessment models will be difficult to fit to data that diverges from conformity with accepted stock-dynamic models, and results from different models are likely to differ, depending on the statistical structure of the model and possibly also on the fitting algorithm. Models that fit to biomass are likely to produce different results from models that fit to numbers. Such differences may be informative as to the characteristics of the model or the data.

Consistently with the above, different assessment models applied to the data for this stock gave different results. The Working Group was not in a position to unequivocally identify the characteristics of the data, or of the models, that would have led to the differences seen. It has been suggested that because VPA methods treat catch-at-age data as exact, they are less apt to follow observed indices of age-aggregated biomass than methods that accommodate inaccurate recording of the age structure of catches and thereby allow an age structure to be fitted that is a compromise between that reported and that that would fit the age structure of the stock and an estimated fishery selectivity. To what extent this behaviour would in fact govern the results from VPA methods was not clear; although XSA does treat catch-at-age data as error-free, ‘the assumption is only used in a rather weak way’. There are, however, other differences between the different models, such as in their weighting of errors and in their fitting algorithms, that might also affect the relationship between their different sets of results.

The data appear to be unable to determine the biomass scale with great precision. Accordingly careful analyses of precision, taking due account of autocorrelation in the data are necessary. It is however noticeable that SCAA results are all generally more optimistic than those of XSA in terms of projections.

XSA with shrinkage has the lowest recent biomass estimates and the least positive recent trend; however XSA runs with a simplified fitting algorithm and without shrinkage agrees more closely with other VPA-type models. It seems likely that the use of shrinkage causes much of the difference between XSA and other VPA models observed in these results.

A suitable starting point for a more methodical comparison of different models, in particular XSA and SCAA, would be to formulate the more flexible model, i.e. SCAA, to have as nearly as possible the same statistical properties as the more rigid model, and explore the effects of progressively relaxing the statistical constraints.

In comparing assessment models it makes most sense to use the same data; the selection of data is a separate issue. When run with the same data as the VPA models, SCAA generates slightly higher biomass estimates and a more positive recent trend, and is much more positive than the XSA results, with shrinkage, which were the most negative of the VPA models. SCAA also indicates a stock with greater resilience to catches than the VPA models. It is however expected that results from different models will diverge most at the end of the available data series.

---

[7] ‘The validity of treating the VPA estimates as exact may be questioned, since the catch-at-age data are certainly subject to various errors. However, the assumption is only used in a rather weak way, in order to calibrate the abundance indices, and the results are not forced to fit the VPA populations exactly. The assumption is analogous to treating the observations of *x* as exact when calculating a regression of *y* on *x*: the fitted model is not thereby forced to pass through the observations.’ (Shepherd 1999.) In the present case, it is not clear that the catch-at-age data is, necessarily, unreliable, and using methods for the assessment that would be appropriate if it were might produce misleading solutions of unknown reliability.

XSA has from its inception been accompanied by cautions about using it if the fishery catch-at-age data are unreliable because of poor sampling or for other reasons[8], and concomitantly by recommendations to use, in such cases, other models with more complete statistical formulations that can down-weight fishery catch-at-age data. However, the anomalies in the data available for this stock are not *per se* evidence that the data is unreliable. If stock structures do not conform to stock-dynamic models and analytical assessment models fit poorly, it might not necessarily be appropriate to assume that the poor fit is due to misreporting of the age structure of catches; this is especially so if the age data from the surveys also indicates aberrancies and the indication is therefore more strongly that of an anomalous behaviour of the stock dynamics rather than faulty age determination or reporting of catches. A model that assumes errors in the catch-at-age data might be the wrong solution to the problem, and it could be difficult to ascertain how reliable the results might be. Equally, however, models that assume catch-at-age data to be exact might in such circumstances be as much or more in error.

In the present case lack of robustness of the models is not the problem, but rather a sign of an underlying problem. It's the data that's the real problem. Age-structured models might resolve data difficulties in different ways, and their own estimates of their own fits may be misleading.

Where there are assessment uncertainties, instead of opting for a single assessment procedure or attempting to bridge gaps between several, it may be appropriate to proceed using a Management Procedure approach where TACs are adjusted adaptively and in a manner that takes due account of uncertainties.

It might be therefore more appropriate to consider, not a different analytical model, but either age-aggregated assessment models or a form of adaptive management with more general, but specified, decision rules for altering allowed catches, or perhaps both. Multiplying a profusion of conflicting analytical models should not be expected to be a satisfactory basis for advice or management.

Of the models considered, XSA with *F* averaging gave the lowest recent estimates of biomass and the most negative trend in the most recent 3 years. Without *F* averaging, XSA was near the middle of the suite of VPA-type models that were tried. All VPA-type models were lower for biomass, and more pessimistic for recent trend, than either the ASPIC stock-production model or the SCAA formulations. The SCAA model results became more like the XSA results when the input data was similarly formulated, but there are substantive intrinsic differences in the formulation of the two types of model that make close comparisons difficult. The working group did not conclude that the XSA should be regarded as not robust. The working group had insufficient time for a full discussion of the robustness or otherwise of estimates of stock status and trend from XSA with shrinkage, or from other assessment models considered. It noted that such an evaluation would require more thorough analyses of the properties of the results of all these assessments to alternative data inputs and assumptions.

The differences between the various models are large and need to be reconciled. Model assumptions seem to be more important than the choice of data set. Real progress was made during the workshop but a problem of this magnitude needs a fair amount time for re-runs and re-formulations and to digest results. If the issue of model choice is to be resolved a pair of workshops is probably required. The first would be a focussed, technical workshop starting with common data and structure and then methodically diverging into different preferred formulations. The foundations in the data for the values of state variables of interest and the impacts of data and model characteristics on each of them could then be disentangled and quantified. A second workshop, with broader participation, could evaluate the relative merits of the various stock-dynamic assumptions once their impacts were revealed. A quantitative review, followed by a qualitative one, might in this way allow a more methodical selection of model structure and assumptions that were demonstrably appropriate to the characteristics of the data available for this stock.

[8] 'would be inappropriate where the catch data are poorly sampled or otherwise defective, but where one or more sets of reliable survey data were nevertheless available.' (Shepherd 1999.) In this case, the survey data also holds aberrancies.

4. Documents and Working Papers

Brandão, A., R.A. Rademeyer and D.S. Butterworth. 2009. GLM standardisation of recent commercial CPUE data for Greenland halibut for Canada, Portugal and Spain which allows for finer spatial stratification. NAFO SCR Doc. 09/41, Ser. No. N5680. 9pp.

Butterworth, D.S. and R.A. Rademeyer. 2009a. CPUE-based assessments of the Greenland Halibut resource using SCAA. Unpublished. 11 pp.

Butterworth, D.S. and R.A. Rademeyer. 2009b. Extensions to SCAA Applications Reported in: "Further Applications of Statistical Catch-at-Age Assessment Methodology to the 2J3K-O Greenland Halibut Resource". NAFO SCR Doc. 09/43, Ser. No. N5682. 27 pp.

Butterworth, D.S. and R.A. Rademeyer. 2009c. Further applications of Statistical Catch-at-Age assessment methodology to the 2J3K-O Greenland Halibut resource". Unpublished. 15 pp.

Butterworth, D.S. and R.A. Rademeyer. 2009d. Further CPUE-based assessments of the Greenland halibut resource using SCAA. NAFO SCR Doc. 09/42, Ser. No. N5681. 7 pp.

Butterworth, D.S. and R.A. Rademeyer. 2009e. Initial applications of Statistical Catch-at-Age assessment methodology to the 2J3K-O Greenland Halibut resource". Unpublished. 21 pp.

Butterworth, D.S. and R.A. Rademeyer. 2009f. On the precision of absolute abundance as estimated by VPA. Unpublished.

Cadigan, N. and D. Butterworth. Review of SCAA papers, with responses to comments. Unpublished. 7 pp.

Darby, C., B. Healey, J.-C. Mahé and W.R. Bowering. 2004. Greenland Halibut (*Reinhardtius hippoglossoides*) in Subarea 2 and Divisions 3KLMNO: an assessment of stock status based upon extended survivor analysis, ADAPT, and ASPIC analyses, with stochastic projections of potential stock dynamics. NAFO SCR Doc. 04/55, Ser. No. N5008. 53 pp.

Healey, B.P. 2009. A chronology of analytical assessments for Greenland Halibut in NAFO Subarea 2 & Divisions 3KLMNO. NAFO SCR Doc. 09/45, Ser No. N5684. 7pp.

Healey, B.P. and J.-C. Mahé. 2008. An assessment of Greenland Halibut (*Reinhardtius hippoglossoides*) in NAFO Subarea 2 and Divisions 3KLMNO. NAFO SCR Doc. 08/48, Ser. No. N5550. 55 pp.

Healey, B.P. and J.-C. Mahé. 2009. An assessment of Greenland halibut (*Reinhardtius hippoglossoides*) in NAFO Subarea 2 and Divisions 3KLMNO. NAFO SCR Doc. 09/39, Ser. No. N5675. 23 pp.

Healey, B.P. and M.J. Morgan. 2009. Population estimates of Greenland Halibut in NAFO Sub-area 2 & Divisions 3KLMNO using ADAPT. NAFO SCR Doc. 09/44, Ser. No. N5683. 16 pp.

Maddock Parsons, D. and B. Healey. 2009. A stock–production model incorporating covariates (ASPIC) for Greenland halibut (*Reinhardtius hippoglossoides*) including sensitivity and retrospective analyses. NAFO SCR Doc. 09/48, Ser. No. N5687. 16 pp.

Mahé, J.-C. 2009. Greenland halibut in NAFO Sub-area 2 & Divisions 3KLMNO: investigations on different sequential population analysis formulations. NAFO SCR Doc. 09/46, Ser. No. N5685. 10 pp.

Morgan, M.J. and P.A. Shelton. 2009. FLEDA and SURBA analysis of the 2+3KLMNO Greenland halibut data. NAFO SCR Doc. 09/47, Ser. No. N5686. 33 pp.

Morgan, M.J., P.A. Shelton, D.C.M. Miller and B.P. Healey. 2008. Is there any evidence of a stock recruit relationship for Greenland Halibut in Subarea 2 & Divs 3KLMNO? NAFO SCR Doc. 08/46 Ser. No. N5548. 15 pp.

Nishida, T.. 2009. Some consideration to evaluate the stock assessments of Greenland Halibut (GH) by SCAA and XSA. Unpublished. 3 pp.

Rademeyer, R.A. and D.S. Butterworth. 2009. A note on the plausibility of the new baseline estimate of halibut abundance. Unpublished.

Shepherd, J.G. 1999. Extended survivor analysis: an improved method for the analysis of catch-at-age data and abundance indices. ICES J. Mar. Sci. 56: 584–591.

Vázquez, A., R. Piñeiro and M. Mandado. 2009. Inquiry into ADAPT's performance. NAFO SCR Doc. 09/49, Ser. No. N5688. 16 pp.

Appendix I. Terms of Reference for the Working Group

Working Group on Assessment Methods for Div. 2J3K–O Greenland Halibut

A Working Group on Assessment Methods for 2J3K–O Greenland Halibut has been struck by Scientific Council to address the 2008 Fisheries Commission request for advice Item 10a (FC Doc. 08/19):

> 10. With respect to Greenland halibut in SA 2 + Div. 3KLMNO, Fisheries Commission requests Scientific Council, in its 2009 assessment of this stock, in addition to the information requested above:
>
> a) To complete an evaluation of alternate assessment models for this stock. This evaluation will enable the determination of the robustness of the assessment model currently used.

Plan of work

ToRs 1-3 will be undertaken by correspondence prior to the meeting, particularly. The WG will meet at the NAFO Headquarter in Dartmouth, N.S., Canada, on 1-3 June 2009, and concentrate on ToRs 4-7. All work will be coordinated by the Chair.

Terms of Reference:

1. Outline plan of work drawing on relevant supporting documentation that detail the current and alternative assessment models;
2. Define ‘assessment model’, ‘evaluation’ and identify ‘measures of robustness’ in the context of this stock as per above request for advice;
3. Review and describe the data sets available for an assessment of this stock;
4. Present and describe, to the WG, the various assessment models that could be used for this stock given the data sets available;
5. Construct standard data sets and carry out “test” assessments using various models
6. Evaluate robustness of assessment models;
7. Draft an answer to the request for consideration by Scientific Council by 8 June 2009. Produce a report of the meeting to be submitted to Scientific Council and the Annual Meeting in September 2009.

Appendix II. Working Group participating members and correspondents

Ricardo Alpoim: INRB/IPIMAR, Lisbon;
Vladimir Babyan: VNIRO, Moscow;
Bill Brodie: DFO, St John’s;
(Jon Brodziak: NMFS, Honolulu;)
(Doug. Butterworth: Univ. of Cape Town;)
(Noel Cadigan: DFO, St John’s;)
Fernando Gonzalez Costas: IEO, Pontevedras;
Diana Gonzalez Troncoso: IEO, Pontedvedras;
Brian Healey: DFO, St John’s;
Michael Kingsley: GN, Nuuk; (Chair);
Dawn Maddock Parsons: DFO, St John’s;
Jean Claude Mahé: IFREMER, Lorient;
Bob Mohn: DFO, Dartmouth;
Joanne Morgan: DFO, St John’s;
Tom Nishida: NRI of Far Seas Fisheries, Shimizu;
Don Power: DFO, St John’s;
Rebecca Rademeyer: Univ. of Cape Town;
(Peter Shelton: DFO, St John’s;)
Takahisa Tanabe: Japan Fisheries Association, Halifax;
Antonio Vázquez: IIM, Vigo;

(Participants by correspondence.)

Appendix III. Narrative of Working Group Activities

The Working Group was welcomed by the Interim Executive Secretary of NAFO. There was a round of introductions. The Working Group reviewed its Terms of Reference.

The Chairman of Scientific Council introduced the topic of the Working Group, reviewing the requests made to Scientific Council by Fisheries Commission and by Canada, and pointed out that Scientific Council has been aware of the activities of the consulting group that has been considering other assessment models. Since 2004 different age-structured models, mostly of the VPA type, had been evaluated for use in the assessment of this stock, and the XSA now used had at that time been agreed to be as good as any. In respect of the Canadian request, it was noted in discussion that the use of fishery catch rates in stock assessments has been and continues to be a contentious issue, and it might not be clear what constituted an ‘acceptable’ CPUE index.

Rademeyer presented the Terms of Reference for the work of the independent consulting group, led by Dr. D. Butterworth of the University of Cape Town, who was unable to come to the Working Group meeting. Briefly, they were to:

Investigate catch-effort datasets with a view to allowing for confounding factors not so far considered, and especially to analyse the data with greater spatial resolution than heretofore, priority being given to Spanish and Portuguese fleet data;

Review assessment models, in particular because it seemed likely that different current perceptions of stock status were more the result of features in the catch-at-age data and their interaction with VPA models than of differences between trends in abundance as indicated by survey and by CPUE. Alternative approaches that might merit investigation could include Statistical Catch at Age—a model that, by allowing for uncertainty in the age structure of catches, might be able to pinpoint the inconsistencies that generate difficulties; age-aggregated (‘index’ or ‘production’ models), which, however, would conceal problems with the catch age structure rather than either solving or identifying them; or, eventually, more complex models that might, for example, take account of different trends in biomass indices in different areas.

The context of the proposed project is found in a changed distribution of fishing effort. It was noted that the *trends* in biomass, especially in the longer term, are not very different between different models, although absolute biomass levels are; these are however not precisely estimated. There have already been several meetings with knowledgeable scientists concerning the development of the Statistical Catch at Age model.

The Working Group then proceeded with three more introductory presentations.

Dr A. Thompson, the Scientific Council Coordinator in the NAFO Secretariat, gave a short presentation on the ICES Working Group on assessment methods and ICES procedures for testing and evaluating assessment models. The Working Group on Methods of Fish Stock Assessment (WGMG) works to develop and critically evaluate assessments models and software code, responding to particular concerns raised by ACFM and the Resource Management Committee of ICES. ICES has also recently planned and held workshops for comparing, assessing, and benchmarking assessments.

Dr T. Nishida presented a working paper on comparing models. The WP displayed a very structured approach, varying different factors, notably the form of the selectivity curve and the selection of data,
Healey, DE for the stock, reviewed the chronology of the assessment and the models used for the stock.

Mr B. Healey described the history of the assessment of the stock and the acceptance of different models.

The Working Group considered how best to proceed and decided to review the models available and to consider their properties and the results they gave, before proceeding to evaluate them and consider their implications for the robustness of the XSA. The models and outputs available for presentation were:

1. Preliminary data exploration using the ‘Exploratory Data Analysis’ facility available in the ‘Fisheries Laboratory in R’ (FLEDA);
2. XSA–the existing assessment model; the Working Group also had available to it information on the history of assessments of this stock using this model;
3. A Statistical Catch at Age (SCAA) model developed and fitted by the consulting group; a large number of formulations, some highly parametrised, had been fitted to the data and were available for review;
4. A stock-production model with few parameters;
5. ADAPT VPA models:
   - 5.2. ADAPT formulations with and without F-averaging;
   - 5.3. a B-ADAPT formulation,
   - 5.4. sensitivity runs on the behaviour of ADAPT/VPA models;
6. SURBA: Survey-based assessment, a review of population structure and evolution based only on a quantitative analysis of age-structure data from surveys.

These models were presented in some details and their properties and results reviewed and summarised. These discussions also included consideration of the characteristics of the different data sources and the conflicts between evidence from survey catch in numbers and survey (and CPUE) biomass indices on the trend of and prognosis for the stock. The working group formulated its conclusions and drafted a text for Scientific Council to consider as a response to the Fisheries Commission.