PROSJEKTRAPPORT

ISSN 0071-5638

HAVFORSKNINGSINSTITUTTET

MILJØ - RESSURS - HAVBRUK

Nordnesparken 2 Postboks 1870 5024 Bergen
Tlf.: 55 23 85 00 Faks: 55 23 85 31

Forskningsstasjonen Flødevigen	Austevoll havbruksstasjon	Matre havbruksstasjon
4817 His	5392 Storebø	5198 Matredal
Tlf.: 37 05 90 00	Tlf.: 56 18 03 42	Tlf.: 56 36 60 40
Faks: 37 05 90 01	Faks: 56 18 03 98	Faks: 56 36 61 43

Distribusjon:
ÅPEN

HI-prosjektnr.:

Oppdragsgiver(e):
The Norwegian Ministry of Fisheries and Research Council of Norway

Oppdragsgivers referanse:

Rapport:
FISKEN OG HAVET NR. 4 - 1999

Tittel:
REPORT OF THE WORKSHOP ON COMPARISON OF STOCK ASSESSMENT MODEL STRATEGIES, WITH APPLICATION TO NORTHEAST ARCTIC COD, 1-4 DECEMBER 1998

Senter:
Marine ressurser

Seksjon:

Forfatter(e):

Editor: M. Pennington

Antall sider, vedlegg inkl.:
42

Dato:
06.04.1999

Sammendrag:

Emneord - norsk:
1.
2.
3.

Emneord - engelsk:
1.
2.
3.

....
Prosjektleder

....
Seksjonsleder

Table of Contents

Terms of Reference....................3
Opening Remarks....................5
Presentation of Models....................5
- Flexibest: a flexible model for stock assessment....................5
- A combined Monte Carlo simulation and bootstrap of extended survivors analysis:....................7
- An adaptive framework for analysing the Northeast Arctic cod stock........9
- Time series estimation of stocks and fishing mortality rates....................12
- Example applications of a statistical age-structured model....................13
- Statistical catch-at-age methods: an overview....................16

General Discussion....................18
Model Comparison (Consensus views)....................23
Further Development of Flexibest - Advice....................26
Panel Discussion....................27
Opening....................28
Assessment Problems for Northeast Arctic cod....................28
Discussion....................29
The Current Status of Flexibest....................29
Comments on the presented assessment approaches....................30
Comparison of the different assessment approaches....................31
Where do we go from here?....................33
Concluding Remarks....................33

Appendices

Working papers presented at the workshop....................35
Northeast Arctic cod - data available for assessments....................35
Participants....................37

TERMS OF REFERENCE

Purpose:

The workshop is arranged to compare different current types of assessment models applied to a common set of real data for Northeast Arctic cod including the Flexibest model. The intention is that this shall promote the understanding of similarities and differences between assessment methods used in various parts of the world, elucidate how reliable our perception of the state of the stock is and serve as input for scientists working on development of assessment methods.

Emphasis will be on outlining advantages and shortcomings with respect to i.a:

- Scientific justification and validity
- Robustness to noise in the data and to underlying assumptions.
- Parameter estimation problems
- Kinds of information that can be utilised.

Days 1 - 3 will be spent on presentations of the assessment work done in advance, preparing comparisons, discussions of how the methods compare, and preparation of a draft report.
Day 4 is set aside for presentation of the outcomes from the workshop and a panel discussion on assessment strategies, for a broader, invited audience.

Participation (see Appendix II):

1. Invited experts who are requested to prepare assessments in advance (see below)
2. Invited experts who will take part in the whole meeting, without bringing assessment of their own.
3. People invited to the panel discussion on day 4. These will be Norwegian scientists with interest in the field, and representatives from the Ministry, Research Council and management.

A report will be made, with a description of each method, comparisons of approaches and results, and, as appropriate, suggestions for further improvements in assessment methodology.

The invited experts are requested to:

- Attempt to do an assessment using the methodology they are used to, with main emphasis on the historic and present state of the stock, and outline the main problems and alternative hypotheses as appropriate.
- Prepare a brief (2-3 pages) description of the method and the results, for use in the report. If necessary, more extensive descriptions can be included as appendices.
- Take part in comparative studies during the meeting, and be prepared to do additional runs if that would be useful.

- Take part in drafting the report. The final compilation of the report will be done by IMR staff.
- Take part in the panel discussion on day 4.

Tentative time table:

Day 1. (starts at 1000)
Presentation of methods and results (approx. 30 min. each)
Discuss and decide on comparative presentations, set up presentation tools (graphs, statistics etc) as appropriate.

Day 2.
Work and discussions on comparisons of principles, methods and results, as outlined on day 1.

Day 3. Draft report, prepare presentations

Day 4. Panel discussion:

Main issues:

- Alternative ways of assessing fish stocks - brief overview.
- Advantages and disadvantages by different approaches: General aspects and application to NEA cod.
- Main problems in the assessment of the cod stock.
- Suggestions for further work and possible improvements.

Opening remarks

Øyvind Ulltang, UiB, Norway (Convenor)

The convenor welcomed the participants and reviewed the Terms of Reference for the workshop. He stated that the primary purpose of the workshop was to make a comparative study of assessment methods and to appraise and provide feedback on the present status of Norwegian assessment techniques, rather than to form a consensus of the best approach to stock assessment.

Presentation of models

The following are six summaries of the assessment models, including Flexibest, presented at the workshop by the invited experts who were asked to prepare their assessments in advance. The aim of the assessments was not to do a complete appraisal of the Northeast Arctic cod stock, but to compare the similarities and differences of the various types of assessment models currently in use throughout the world.

Fleksibest: a flexible model for stock assessment

Dankert Skagen, IMR, Norway.

Fleksibest is a model for fish stock assessment using catch and survey data. So far, emphasis has been on the historical analysis of the stock abundance and composition. The present version is designed specifically to estimate the state of the Northeast Arctic cod. The software is a modification of the Icelandic model BORMICON, which is a multispecies, multiarea simulation model with options for parameter estimation (Stefánsson and Pálsson, 1997).

The model was constructed to provide an alternative to the VPA-tuning models traditionally used for many stocks in the ICES area. The basic idea was to obtain greater flexibility as to the choice of model assumptions and the use of the data. This would allow both more appropriate use of the data, and make the stock estimate less dependent on data for which the quality is questionable. Moreover, it should allow for incorporating background information on e.g. growth *vs.* climate, on fish behaviour, changes in the way the fisheries are performed etc., in model formulations.

Fleksibest belongs to the category of assessment models where a self-contained population model is fitted to data, as opposed to the VPA type, where a population is reconstructed using the catch numbers at age. In this sense, the model is of the same category as ICA (Patterson & Melvin, 1996), which in turn originates from work by Fournier and Archibald (1982) and Deriso & *al.* (1985). Such models are sometimes termed Statistical Catch-at-Age methods (Hilborn and Walters, 1992). Further examples are the models presented by Ianelli and Punt.

The model differs from most others in that the population is structured by both length and age, and mortalities and maturity are primarily defined by length irrespective of age. The motive for including a length structure is mainly that growth is known to vary quite much from year to year for this stock.

Immature and mature fish are treated as two separate populations, with separate sets of parameters. Initial stock numbers for the immature are parameters, while matures get new supplies from the immature.

The fishing mortality is assumed to be separable. This principle is modified by relating the selection to length instead of age, and by defining a separate selection pattern for each fleet. The overall selection pattern is a weighted average of fleetwise selections at length. This allows for variations in selection at age due to variations in growth rate, and to variations in the share the different fleets have in the fishery. The natural mortality and maturation ogive are functions of length. The model allows for including predation mortality due to cannibalism, as a function of the prey (small cod and capelin) and predator (large cod) abundances. Several catchability models, relating survey indices to stock abundance, are implemented. In principle, catchabilities are also modelled as a combination of year factors and length factors.

Since the model population is structured also by length, growth has to be modelled. Average growth in each time step is described by a simple growth model. Then, the previous length distribution is transformed to the new one as a Markovian process using a matrix of transition probabilities conditional on the expected growth.

For the time being, only a limited selection of objective functions, and of aggregation levels at which model and data are compared, are implemented. The overall objective function is a sum of partial objective functions, each representing the comparison of a certain set of data, and the user can choose what to include, and which parameters to optimise in each session. A similar principle is found in BORMICON, and much of the implementation carries over from that program.

The various partial objective functions are characterised both by the type of data that are compared, the aggregation level and the measure of the fit of the model to the data. At this point, the model is still not complete. In particular, the freedom to aggregate data before comparing is still limited.

The routine presently used for optimisations is the direct search using the algorithm of Hooke and Jeeves (1961). This is not very rapid, and not suited for large numbers of parameters, but has the advantage of being quite robust. Quasi-Newton routines will be implemented later.

The model is still not fully developed as a tool for a complete assessment. The present state is that it can estimate the historical stock numbers and fishing mortalities. Since the model basically is a prediction model where the parameters are estimated by predicting the past history, it is straightforward to extend it to predicting the future development of the stock. Uncertainty estimates are not provided at present, and needs to be considered. Improvement of objective functions and optimisation routines, better presentation of results and diagnostics, and simpler user interface are planned.

References

Deriso, R.B, Quinn, T.J and Neal, P.R. 1985. Catch - age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42:815-824.

Fournier, D and Archibald C.P 1982. A General Theory for Analyzing Catch at Age Data. Can. J. Fish. Aquat. Sci. 39: 1195-1207.

Hilborn, R and Walters, C.J. 1992. Quantitative Fisheries Stock Assessment: Choice, Dynamics & Uncertainty Chapman & Hall.

Hooke, R. and Jeeves, T.A . 1961. Direct Search Solution of Numerical and Statistical Problems. Journal of the ACM, 8:212-229.

Patterson, K.R. and Melvin, G.D. 1996. Integrated Catch at Age Analysis, Version .2 Scottish Fisheries Research Report No. 38

Stefánsson, G. and Pálsson, Ó.K. (eds.) 1997. BORMICON. A Boreal Migration and Consumption model. Hafrannsóknastofnun fjölrit nr. 58 Marine Research Institute, Reykjavik, Iceland.

A combined Monte Carlo simulation and bootstrap of Extended Survivors Analysis (XSA): A North East Arctic cod example.

Laurence Kell, CEFAS, UK.

Cohort analysis or virtual population analysis (VPA) recreate a stock's historical population structure from the catch at age matrix. Estimates of numbers at age are conditional on the natural mortality and the numbers alive at the oldest age in each cohort (i.e. the Terminal Ns or survivors). Therefore the main problem in such sequential age based assessment methods is to estimate the terminal population numbers. In Extended Survivors Analysis or XSA (Shepherd, 1992, Darby and Flatman, 1994) the standard catch at age method for most ICES stocks, these are found from the relationship between catch per unit effort (CPUE), abundance and year class strength.

An initial guess of the number of survivors initiates the algorithm, catchability by fleet and age is then estimated from the VPA estimates of N and the CPUE indices (U). Predictions of N by fleet, age and year are then made for each observed CPUE value. The predicted values are projected forward to the final age in each cohort using fishing and natural mortalities (estimated and assumed respectively). A weighted average of the survivors provides new estimates of the terminal Ns and the XSA algorithm iterates until the difference between successive estimates of fishing mortality is small.

The only estimates of uncertainty are provided by the standard errors of the terminal Ns in the last year. However, the estimates have been criticised as underestimating the true variance. In addition catch at age data often follow a gamma rather than a log normal distribution, although the expected values will be the same under either error model assumption the variances will not. In response to these criticisms a Monte Carlo simulation was combined with a bootstrap (Efron and Tibshirani, 1993) to provide improved estimates of uncertainty of a variety of parameters and quantities of interest.

The Monte Carlo Simulation and bootstrap

Uncertainties in natural mortality, weights and maturity at age were modelled by Monte Carlo simulation. Expected values and CVs for natural mortality and maturity were taken from the report of the Arctic Fisheries Working Group (Anon, 1998). Uncertainty in growth was modelled by sampling from the residuals after fitting a local regression to the growth data. Uncertainty in catch at age data could also be modelled as a random variable but this was not done in this case.

The bootstrap (Efron and Tibshirani, 1993) is a computer intensive technique for estimating statistical properties of interest flexibly and with a minimum of mathematical assumptions. In this paper the CPUE indices were bootstrapped using the residuals from the catchability models.

Since the inputs to XSA were modelled by Monte Carlo simulation before each iteration of the bootstrap the CPUE models were refitted and new residuals derived. Also to model the correlation between ages CPUE residuals were selected by year.

References

Anon. 1998. Report of the Arctic fisheries working group. ICES CM 1998?Assess:1.

Darby, C. D., and Flatman, S. 1994. Virtual Population Analysis: version 3.1 (Windows/Dos) user guide. Info. Tech. Ser., MAFF Direct. Fish. Res., Lowestoft, (1): 85pp.

Efron, B., and Tibshirani, R.J. 1993. An introduction to the bootstrap: Monograph in Statistics and Applied Probability 57. Chapman and Hall, London. 1993.

Shepherd, J.G. 1992. Extended survivors' analysis: an improved method for the analysis of catch-at-age data and catch-per-unit-effort data. Working paper No. 11 ICES Multi-species Assessment Working Group, June 1992, Copenhagen, Denmark. 22pp. (mimeo).

An adaptive framework for analysing the Northeast Arctic cod stock

Stratis Gavaris, Dept. Of Fisheries and Oceans, Canada

An age structured integrated calibration method as implemented within the ADAPTive framework (Gavaris 1988) was applied to the assessment data for North-east Arctic cod. The approach permits a diversity of formulations within this class of models. Accordingly, the data and model results were subjected to iterative examination of formulations to detect potential inconsistent patterns and their impact on estimates of stock status.

The ADAPTive framework was introduced as a flexible assessment tool for treating catch at age and abundance indices in an integrated statistical (least squares) modelling environment. The class of model formulations investigated for the North-east Arctic cod assessment had the following common features:

- error in the catch at age was considered negligible relative to the error in the abundance indices
- error in the abundance indices was considered independent and identically distributed after taking logarithms
- population dynamics were assumed to follow the common Virtual Population Analysis model defined by the equations

$$N_{a+\Delta t,t+\Delta t} = N_{a,t} e^{-(F_{a,t}+M_a)\Delta t} \quad \text{and}$$

$$C_{a,t} = \frac{F_{a,t}\Delta t N_{a,t}\left(1-e^{-(F_{a,t}+M_a)\Delta t}\right)}{(F_{a,t}+M_a)\Delta t} \quad \text{where}$$

N = population abundance

C = catch

F = instantaneous fishing mortality rate

M = instantaneous natural mortality rate

a = age

t = time

A Newton-Raphson algorithm was used to solve for the fishing mortality rate in the catch equation. Year was used as the unit of time, therefore the fishing mortality and natural mortality are expressed as annual instantaneous rates.

- indices were assumed to be proportional to population abundance according to the relationship (except for trials where a power relationship was explored)

$I_{s,a,t} = \kappa_{s,a} N_{a,t}$ where

I = index (CPUE or survey)

κ = calibration constant

s = source (e.g. Russian trawl survey)

The calibration constants were assumed to be stationary over time (except trials where step changes were explored).

- a solution for the model parameters was obtained by minimising the sum of squared differences between the natural logarithm of observed abundance indices and the natural logarithm of population abundance adjusted for catchability by the calibration constants:

$$\Psi\left(\hat{\theta},\hat{\kappa}\right)=\sum_{s,a,t}\left(\psi_{s,a,t}\left(\hat{\theta},\hat{\kappa}\right)\right)^2=\sum_{s,a,t}\left(\ln I_{s,a,t}-\left(\hat{\kappa}_{s,a}+\ln N_{a,t}\left(\hat{\theta}\right)\right)\right)^2$$ where

θ = ln population abundance

- the ln population abundance at age was estimated at the beginning of 1998, i.e. $\hat{\theta}_{a,1998}$ while the population abundance at age 14 in all years, i.e. $N_{14,t}$, was derived by assuming that fishing mortality rate for age 14 was equal to the average for ages 8 to 11 (7 to 11 in the base trial).

In all trials, the catch at ages 1 to 14 for 1981 to 1997 was used. Diagnostics were examined to identify undesirable features or inconsistent patterns. Alternative model formulations to deal with those patterns were then explored. The following model formulations were considered:

- base analysis: all the indices considered by the assessment working group were used with the default model formulation described above. The natural mortality rate was assumed constant and equal to 0.2 for all ages and years. Cannibalism was not incorporated.
- M ages 1&2: keeping all specifications the same as used in the base trial, two additional parameters, the natural mortality rate for each of ages 1 and 2 from 1993-97 were estimated; the natural mortality rate for all other ages in all years was assumed equal to 0.2
- Excluding 1&2: an analysis was conducted excluding ages 1 and 2 but retaining all other specifications as in the base analysis
- Split: base analysis excluding ages 1 and 2 but treating the Barents Sea trawl survey as two distinct time series with a break in 1993 and considering only the data after1990 for the Norwegian acoustic survey
- M years 1990-97: similar to split but also estimating a single M parameter for all ages for each of the periods 1990-94 and 1995-97; in this model formulation, M represents natural mortality and mortality from unreported fishing activity
- M ages 10-14: similar to split but estimating a single M parameter for ages 10-14 for all years
- Power catchability: similar to split but employing a power relationship for catchability:

$$I_{s,a,t}=\kappa_{s,a}N_{a,t}^{\alpha}$$

The following observations were made:

- indices for ages 1 and 2 appeared less reliable and were contradictory to the other data

- the relationship between the Barents Sea trawl survey and the VPA abundance was not constant over time, appearing to have changed around 1993-94
- the relationship between the Norwegian acoustic survey and the VPA abundance was not constant over time, appearing to have changed around 1990
- the dynamics of older cod, ages 10 and older is poorly determined with the available data
- there is some evidence for power catchability relationships, particularly at younger ages, but these are very poorly supported by the data
- the catchability pattern by age for the Svalbard survey appears counter-intuitive as does the consistent power relationship at all ages
- the correlated errors within years would probably result in an underestimation of variance and may bias the parameter estimates
- the recent survey indices for ages 3 and older suggest that the stock has been depleted to a greater extent than can be accounted for by the catch at age and the assumed natural mortality in the VPA; this feature may contribute to retrospective patterns in future assessments

The design philosophy of Flexibest incorporates this class of models and should permit further exploration of these types of formulations as well as comparisons with models which may have fewer assumptions, such as models which assume the catch error is negligible and models which do not incorporate changes in growth over time. To facilitate exploration and comparison, it is very important to give careful consideration to diagnostic features that can be examined for evaluation of model assumptions and for the impact of alternative selections.

References

Gavaris, S. 1988. An adaptive framework for the estimation of population size. CAFSAC Res. Doc. 88/29: 12 p.

Time series estimation of stocks and fishing mortality rates

Gudmundur Gudmundsson, *Central Bank of Iceland, Iceland.*

In the present method stocks and fishing mortality rates are modeled and estimated as unobserved time series. *(Gudmundsson, 1994).* The time series model of the stock numbers is the traditional model of stock decline, augmented by a model for the youngest fish included in the analysis. The main difference between this method and most other methods of fish stock assessment is the time series model of the fishing mortality rates. This model depends upon 4-5 parameters which describe the magnitude and nature of the variations in the F's. Initial values of F in the first year are determined by a function of 4 parameters.

Estimation proceeds in the forward direction from the first year included. The first stock values are estimated by the catches in that year and the initial F's. The time series models then provide a prediction of stocks and fishing mortality rates in the second year and from these the catches are estimated. The differences between the observed and estimated catches provide additional information about the stocks and fishing mortality rates and the initial predictions are updated accordingly by means of a linear approximation to the Kalman filter. After updating the predicted values the calculation proceeds to predict the values in the third year and so on. The coefficients of the Kalman filter depend upon the parameters in the time series models and all unknown parameters are estimated by the likelihood function of the catch prediction errors.

The estimates of stocks and fishing mortality rates in the last year obtained by the Kalman filter use all information in the data. But improved estimates of the previous values are obtained by a backward algorithm called smoothing. This procedure requires no further information to estimate all stocks and fishing mortality rates. They are determined by the assumption that the F's follow a time series model and the requirement of predictability of the catches. If no further observations are included these values will be similar to VPA results *in the younger ages of earlier years*, but not identical because measurement errors of catch-at-age data are taken into account.

The Kalman filter also provides the covariance matrix of the estimated stocks and fishing mortality rates. It depends upon the estimated time series parameters and does not involve the Hessian matrix.

CPUE observations contain additional information about stock numbers and can be included in this estimation in a similar way as catch-at-age observations. The method is not dependent upon CPUE data to provide estimations of F and N in the last year. It is therefore possible to apply less restrictions on the catchability, allowing for persistent variations in the form of random walk in its time series model.

The relative weights given to catch-at-age and CPUE data depend upon the magnitudes of respective measurement errors and variability of fishing mortality rates and catchabilities. This is determined by parameters, estimated by the likelihood function.

With the present data CPUE from the Norwegian Barents Sea trawl survey and the Norwegian Barents Sea and Lofoten acoustic survey reduced the uncertainty in the estimation, compared with results obtained by using only catch-at-age data. However, large persistent variations in catchability were estimated in both surveys.

References:

Gudmundsson, G. (1994). Time series analysis of catch-at-age observations. Applied Statistics 43 (1), 117-126.

Example applications of a statistical age-structured model of Northeast Arctic cod using an integrated statistical age-structured model similar to that currently used for Alaska pollock

James N. Ianelli, AFSC, USA.

We present preliminary analyses of the Northeast Arctic cod stock using a model that is similar to that used for Alaska pollock (*Theragra chalcogramma*). We illustrate some ways to analyse a variety of stock assessment data using an integrated model for catch-at-age data from 1946-1997. We compared models with different assumptions about natural mortality rate, changes in age-specific availability to survey abundance, the effect of excluding survey indices, and some time-series processes of survey catch rates. Also, we present methods for examining stock-recruitment relationships within the integrated model and subsequent effects of uncertainty available for management related quantities (e.g., F_{msy} $F_{40\%}$ harvest rates).

Technical aspects

The model structure is developed following Fournier and Archibald's (1982) methods, with many similarities to Methot (1990). We implemented the model using automatic differentiation software developed as a set of libraries under C++. These libraries, coupled with model-building scripting software, enables a set of tools useful for non-linear parameter estimation for large problems. An added benefit is the potential for performing high dimensional integration required to use Bayesian methods and to evaluate more fully the effect of multivariate parameter uncertainty on management quantities of interest. The software provides a way to organise principles common to all models where data are involved. In plain words, the difficulty of estimating parameters in complex non-linear problems often increases exponentially with the number of parameters being estimated. This is due to the fact that most algorithms use some form of numerical method (called finite differencing) to compute the direction parameters should change to make to model fit the data better. Using automatic differentiation software, the derivatives (gradients) are made available via the reverse mode of differentiation (in practice, an application of the chain rule). It can be shown that the cost of computing these gradients is about equal to 5 separate function evaluations *independent of the number of parameters involved.*

The time frame of the model began in 1946 (with estimates of the age composition in that year) and extended through to the year 2003. The basic structure of the model is designed to be "semi-separable" in that the user has ready control over the degree of breaking down the fishing mortality rates between age and time. That is, the option to have a constant age-specific component of mortality (separable model) can be accommodated at one extreme, while having selectivity vary completely in every year is also equally possible. The latter case is similar to the assumption required for standard VPA methods. The period 1999-2003 was projected under 4 different harvest rates with stochastic recruitment. Writing the model in this way provides quick and easy access to evaluate the different consequences of future harvest levels while maintaining the structure of uncertainty derived from the original model.

The likelihood components (where the model is tuned to observations) consists of:

- Total catch biomass (Log normal, σ=0.05)
- Lofoten and Barents survey abundance indices, (Log normal, σ=0.25)
- Fishery and survey proportions-at-age estimates (Robust quasi-multinomial (effective sample size of 100 for fishery, 50 for surveys). These values were selected based on experience with catch-at-age variance estimates obtained from other fisheries
- Selectivity constraints (penalties on age-age variability, time changes, and non-decreasing (with age) patterns
- Recruitment variability (Log normal, σ=0.9)

Since all stock assessment methods have alternative ways to interpret and fit the data, we attempt to highlight some of the key models explored for Northeast Arctic cod. These included:

Model 1 Survey selectivities constant with time, fixed constant M, no stock-recruitment relationship.

Model 2 Same as Model 1 but with variable survey selectivities.

Model 3 Same as Model 2 but with a Ricker stock-recruitment relationship.

Model 4 Same as Model 3 but with M varying with time as specified by the WG.

Model 5 Same as Model 4 but with robust likelihood function for age composition data.

Model 6 Same as Model 3 but ignoring all survey data.

Model 7 Same as Model 3 but with a time-series trend in Lofoten survey catchability.

Conceptual approach

In general, inclusion of unknown process errors require fewer and/or less restrictive assumptions. **The cost of fewer assumptions is higher variance.** It can be argued that most modern stock assessment models tend to under-estimate the level of uncertainty (e.g., NRC 1997) and that this failure may reflect on management effectiveness. Unlike classical statistical analyses where model parsimony is the rule for hypothesis testing, fisheries management is a *decision problem* and one that should be guided by honest evaluations of processes that are known to affect our laboriously collected data.

In summary, the philosophy behind our modelling approach includes tenets such as:

1) The model data should be close to "raw" form with minimal manipulations and "off-line" analyses. For example, if a catch-rate abundance index has been estimated using data external to the model, it may be more difficult to correctly specify the variance (weight) of that source of "data." In a more integrated statistical model, the uncertainty about data is carried within the model and associated errors are appropriately propagated. Currently there are practical limitations to this approach, but the idea is a sound one. As technology improves, the need for ad-hoc methods to weight data should diminish.

2) The models should become less "absolute" about use of data and different assumptions. For example, in many practical settings, abundance data series are either completely omitted from analyses or included as proportional indices. In our approach, we can easily write the model to include a small number of terms that has the two extremes as special cases while allowing for easy manipulation of intermediate values. Another example is to acknowledge that model parameters traditionally held at fixed values (e.g., natural mortality) are in fact uncertain and can be easily treated as having alternative values. In this way the uncertainty about such things can be evaluated more completely for fisheries management decisions.
3) The practitioner should be able to easily explore model alternatives quickly and have direct access to methods of examining residuals and potential problems with model specifications. No single model is correct, nor are there any hopes of having correct parameter estimates even if a model is correctly specified. Therefore, the consequences of different fisheries management actions under different model configurations/parameter uncertainties should be readily available. In practice, a single model and its level of uncertainty can often be expressed that covers much of the range of other, perhaps equally plausible models.
4) The model should be written to be formally independent of the data. If the fundamental part of the model involves a direct algorithm of the data (e.g., XSA), then several problems arise. First, missing data prohibits the use of the model. Second, the data are observations *that have error,* dealing with this requires complicated, often *ad-hoc* procedures. Finally, harvest projections require a separate analysis that may be difficult to easily link to the original model (and associated degrees of uncertainty). There is a model, then there are data, and the two are formally linked through an objective function. The objective function can be anything that would minimise the difference between model predictions and observed data. However, writing the objective function as a likelihood has an added benefit of producing levels of dispersion (variance) that can then be used to assign probabilities to alternative states of nature. The model may also have assumptions based on "external" data. We argue that this should also be made explicit in the likelihood function.

The choice of objective function can be critical. There is no easy way to a good foundation for weighting of observations, and one cannot expect a smart likelihood to rescue a wrong model. The time series approach is less sensitive to this, due to the simplicity of its underlying assumptions.

Statistical catch-at-age methods: an overview

Andre´ Punt, CSIRO, Australia

Statistical catch-at-age methods attempt to provide the analyst with considerable flexibility in terms of representing underlying population dynamics processes and incorporating auxiliary information. They decouple the development of the model of the population dynamics from the development of the model of how the data were collected and relate to the quantities represented in the population dynamics model.

These methods are used widely on the west coast of North America and in Australasia.

The three main reasons for using these methods are i) to provide a flexible framework within which alternative hypotheses regarding the population dynamics can be represented, ii) to allow estimation of quantities of interest to management (current biomass, recent recruitment etc.), and iii) to provide the basis for the evaluation of alternative feedback-control harvest policies. Uncertainty is considered by examining a variety of alternative model formulations and by estimating measures of precision (Bayesian posteriors, asymptotic standard errors, bootstrap distributions etc.). The alternative model formulations include *inter alia* i) different values for fixed parameters of the model (e.g. growth rates), ii) different structural assumptions (e.g. whether the fleet-specific selectivity function is logistic, gamma etc.), iii) different assumptions regarding how the data relate to the model quantities, and iv) the form of the likelihood function.

In general, these methods differ from the other model classes in that allowance is made for the possibility that the catch-at-age data are subject to uncertainty. They can also deal adequately with situations in which catch-at-age data are unavailable for some years/fleets.

The statistical catch-at-age method methods presented to the meeting were based on age-structured separable model that allows for process error in recruitment and selectivity-at-age. This model considered ages from 1 to 15, with age 15 being considered as a plus group. It ignored sex structure and assumed time-invariance in natural mortality. The data included in the likelihood function were the catch-in-mass data, the catch-proportion-at-age data, the survey indices (aggregated over ages), and the survey catch-proportion-at-age data. One of the methods utilised the International 0-group index. The statistical uncertainty was quantified Bayesian posterior. The full mathematical details of the two methods and the results of illustrative applications to Northeast Arctic cod are provided in my paper prepared for this workshop.

The assessment process used in some Australian fisheries involves the following steps:

a) A set of (qualitative) hypotheses regarding the dynamics of resource are developed. This is primarily an attempt to capture the "full range" of uncertainty.
b) The data available for assessment purposes and the range of plausible relationships between these data and the quantities included in the model are identified.
c) Values for some of the parameters of the model (e.g. growth rates) are specified using experimental results.
d) Data for "similar species" is examined in order to develop 'data-based' prior probability distributions for some of the remaining model parameters.
e) The "expert opinions" of the assessment group are used to agree on prior distributions for any parameters not pre-specified or assigned priors based on data for "similar species".

f) A base-case set of specifications (or a number of base-case sets) is chosen and the remaining hypotheses represented as sensitive tests.
g) Choices are made for the quantities for which posterior distributions are needed for presentation to the management committee (the management-related quantities).
h) Preliminary analyses are conducted which involve finding the values for management-related quantities that correspond to the maximum of the product of the likelihood function and the prior. These values, together with the fit of the model to the data, are used to "prune" the hypotheses to a manageable number.
i) Full Bayesian analyses are conducted for the remaining hypotheses.

The flexible nature of the approach means that the analyst is free to develop assessment models for specific fisheries. This provides the analyst with the ability to include and hence contrast the implications of alternative hypotheses/data sources. This, in turn, provides for a more 'inclusive' stock assessment process whereby non-modellers (including, in some jurisdictions, industry and conservation groups) can contribute substantively to the assessment through the identification of alternative hypotheses about the population dynamics/observation process and the development of prior distributions.

However, this flexible approach is not without its disadvantages:

a) No standardised software package exists and, in general, it is necessary to develop a specific computer programme for each application. This makes use of these models in a 'working group' setting difficult as considerable model development work may be required if new hypotheses are identified.
b) The methods are very intensive computationally. This is particularly the case if uncertainty is to be represented in the form of a Bayesian posterior.
c) The flexible nature of the approach can be open to abuse/manipulation particularly when the data are uninformative and so the prior distributions have a relatively large impact on the final model outputs.
d) The specification of relative weights cannot be accomplished automatically in most cases. However, the results can be very sensitive to the values for such weights.
e) Even though the approach is very flexible, there is nevertheless a tendency not to explore a very wide range of alternative model formulations (mainly because of the associated model development/testing requirements).
f) The use of Bayesian approaches to develop posterior distributions has several problems. Most serious amongst these are: i) the inability of many scientists to develop consistent and sufficiently uninformative prior probability distributions, ii) the lack of studies that have attempted to synthesise the results of existing assessments to provide 'data-based' priors for some of the key parameters of stock assessment models (e.g. the steepness of the stock recruitment relationship, the extent of the depensation at low stock size), iii) the time demands of existing numerical methods for developing posterior distributions, and iv) the inability to develop 'non-informative' prior probability distributions.

General Discussion

- In all the models, except the time series model, the decision of how much weight to give to each data set was arbitrary. More thought needs to be given to how these weights should be selected.
- More work needs to be done on determining how to measure the fit of a model and how we relate models to data sets. Must decide for what purpose do we use each data set.
- It is necessary to be more critical of the data. That is, it is fundamental to know the characteristics of the input data before they are included in a model.
- It may be effective to leave out of the model some of the data series or, perhaps, the earlier parts of some of the series. In particular, is anything gained by using all available series to tune a model or should only the most accurate and precise survey series be included?
- Linear combinations of surveys could be constructed to form a 'complete' series from various surveys that have only partially covered the Northeast Arctic cod stock.
- In order to lessen the effect of survey errors in the models, there could be more aggregation of the data such as a total biomass series, combined age groups etc.
- For a catch at age analysis, it may be better to use fewer 'true' age groups and treat the older age groups as an aggregate plus group.
- Northeast Arctic cod should be separated by sex in the models since females tend to dominate the older age groups.
- The survey series should be compared with the catchability patterns generated by the models. That is one should analyze the correlation between the survey series and the catchability patterns and how the correlations relate to external factors such as capelin abundance. The causes of the observed catchabilty changes (for example, variable fish availability to the survey, changes in survey procedures and protocols, non-landings, natural mortality, etc., should be evaluated).
- One of the advantages of a flexible model is that it can be reduced to a simple model and the more complicated model can be used to test the simpler model for adequacy. Flexible models are especially suited for exploring different scenarios.
- Simulation studies can be used to check for model induced biases in the estimates.
- In general, one can estimate model parameters based on prediction of observed indices and catches by reconstructing the past state of a stock.
- Biological factors are more closely related to length but it is not apparent that including length in a model, which increases model complexity, improves its predictive power. It is much easier to relate biological factors to length than age though including length and age does not appear to significantly change the model's outputs.

During the discussion an approach to modeling the recruitment of North-east Arctic cod was presented by Victor Tretyak, Pinro, Russia (see 1998, working doc.). This model assumes that the cod stock has a Ricker spawning stock-recruitment relationship and incorporates data on the abundance of cod at six stages; eggs, larvae, 0-group and age 0, 1 and 2 in autumn, and data on the abundance of cod as predator, the abundance of capelin and of euphausiids, and temperature.

The model estimates of year-class strength showed a satisfactory agreement with the estimates of year-class strength generated by a single-species VPA, and thus this model may be of value for predicting the recruitment of North-east Arctic cod.

A discussion arose on the differences and advantages of fish assessments based on Bayesian or frequentist modeling. Tore Schweder, Sosialøkonomisk Institutt, Norway, prepared the following report for the workshop.

Fish stock assessment by Bayesian or frequentist methods?

There are two broad schools or paradigms in statistics. The oldest school is stemming from reverend T. Bayes. Here, the parameters of the model are regarded as random variables on the same footing as the observed data. The other school was developed in our century, particularly by R.A. Fisher and J. Neyman. Here, parameters are regarded as (unknown) quantities, and the problem is to measure or estimate these quantities. This school is often called frequentist since emphasis is on the frequency properties of the estimation methods in repeated use. In the frequentist tradition, "probability" and "randomness" is reserved for entities that (hypothetically) could be regarded to have a frequency distribution if repeatedly observed.

In parametric modeling and analysis, the likelihood function is a central concept both within the Bayesian and the frequentist paradigm. In both traditions, to specify the sampling model for the data is equivalent to specifying the likelihood function. And the likelihood function is the principle vehicle to transform data into useful information.

For the frequentist, the likelihood function is usually the most defendable object function to maximize when fitting the model to the data. The reason for this is that the maximum likelihood estimator, likelihood ratio tests etc makes optimal use of the data in nice finite-sample models, and also for a very wide set of models asymptotically - as the information in the data increase. In short, the likelihood function provides the optimal weighting between the various data components. Assessment models tend to have many parameters, and the interpretation of the likelihood function itself need not be simple. For interest parameters of low dimensionality, profile likelihoods might be computed. To really understand the uncertainty in an estimate of an interest parameter, like the spawning biomass next year under a specific fishery pattern, simulation and refitting the model to simulated data is necessary. Only in small nice models, or when the information in the data really is extensive (and more extensive than in current survey- and catch data for NEA cod) is it possible to compute confidence intervals or other summaries for the interest parameter that accurately represents the uncertainty stemming from the statistical variation in the data. There are usually additional sources of uncertainty that should be accounted for. To do this adequately, much more simulation is often required – if it at all can be done.

For the purist Bayesian, the likelihood function is understood as a conditional probability distribution of the data given the parameter, and having a prior distribution for the parameter, the joint distribution is obtained by multiplying the two. The *conditional probability* distribution for the parameter given the observed data is then obtained by Bayes' theorem, which in a sense is nothing other than the definition of conditional probability. For the more pragmatic Bayesian, who allows himself to use improper priors (prior distributions that do not integrate to one, and thus have no probability interpretation), the posterior distribution obtained by a formal use of Bayes' theorem is simply a weighted version of the likelihood function. This weighted likelihood is scaled to have integral one, and is used as a probability distribution for the parameter. This interpretation allows him, and also the purist Bayesian, to study the likelihood function by sampling methods (importance sampling, Markov Chain Monte Carlo etc). For an interest parameter, one simply looks at the marginal distribution of that parameter in the sample from the joint posterior distribution. Given the model and the prior distribution, the posterior distribution of the interest parameter represents to the Bayesian the data-summary that accounts for the uncertainty. To further account for model selection uncertainty or other additional sources of uncertainty, the model and the prior distribution could in principle be extended to encompass also these uncertainties.

There are many issues to consider when choosing between a Bayesian and a frequentist approach to modeling and data analysis in the fish stock assessment context. Some of these are:

- Interpretation: To the purist Bayesian, all probabilities are subjective. This might be problematic in the management context. To the pragmatic Bayesian, the weighted likelihood function has lost its probability interpretation, and it is not clear what precisely the marginal posterior distribution means, other than a weighted integral of the likelihood function. When using confidence intervals and hypothesis tests, the frequentist has a clear interpretation of his summaries. His probabilities, say for the confidence interval to capture the true value of the parameter, refers to hypothetical repeated use of the method on data with the modeled properties. In fisheries assessments and other applications, it is the observed data that matter, and there is a need for a distributional concept, such as the confidence distribution, parallel to the Bayesian posterior that represent what has been learned from the data. A confidence distribution is calculated from the data in such a way that its quantiles provide confidence intervals for the parameter in question (Efron 1998). The confidence distribution inherits its frequentist interpretation from confidence intervals and p-values. In practice, it will be natural to interpret it much like the Bayesian posterior distribution is understood – as a distribution of confidence in various interval statements that can be made concerning the parameter, given the observed data and the assumed model. The interpretation of frequentist summaries like confidence distributions is thus clear, conditional on the chosen model. However, the frequentist has a substantial problem with interpretation when additional sources of uncertainty are accounted for, say when a scenario experiment has been carried out.

- Subjectivity: There is always an element of judgement in statistical work. The frequentist must decide on the model and the selection of data etc, as must the Bayesian. The Bayesian introduce an extra element of subjectivity with his priors. These might be based on external information and may sometimes be defendable scientifically. When the information in the likelihood function dominates the prior, the possible subjective element is of minor consequence. To avoid subjectivity, Bayesians often use "non-informative" priors. Unfortunately, all priors carries information, also flat priors. The necessity of having a prior distribution for every parameter in a Bayesian assessment model is therefore a problem. It might, furthermore, be difficult to obtain agreement among scientist on which priors to use. In the case of bowhead whale stock assessment in the IWC, the experience is also that the choice of prior distributions mattered substantially for the results, and also that agreement over priors did not last over time.
- Bias: Bias in the data (which could be regarded as misspecification of the model) is a problem both for the Bayesian and the frequentist. In complex models, bias can also be caused by nuisance parameters. It is well known that maximum likelihood estimators might have sampling distributions that are shifted away from the true value (when estimating the residual variance in multiple regression with p regression parameters, we correct the maximum likelihood estimator by dividing by n-p instead of n to avoid this method-inflicted bias). The problem is that the likelihood function and also its profiles or its "marginals" tend to be shifted away from the truth, something that also is a problem for the Bayesian. This problem has received more attention in the frequentist tradition where say bootstrap confidence intervals are corrected for bias by say the BCalpha method, and where different methods of adjusting profile likelihoods and other summaries obtained from the joint likelihood function. In the Bayesian tradition, this problem is difficult to discuss, particularly for the purist Bayesian for whom the very concept of bias is a difficult one. In the pragmatic Bayesian tradition, it is possible to choose priors with the aim of removing bias inflicted by nuisance parameters and non-linearity, but this avenue is a difficult one. The extent and nature of such inherent bias in the likelihood with respect to a parameter of primary interest in an assessment model with many parameters should be investigated, along investigations of bias due to model misspecification etc.
- Inclusion of external information: Another word for external information is information prior to the data analysis. External information might be based on theory (a cod cannot have negative growth in length, but certainly in weight) or on past studies. It can also be based on experience or related empirical knowledge (say that indices based on acoustic surveys of NEA cod are assumed to be functionally related to cod density in the area in the same way as in similar surveys of Icelandic cod, for which information is available). External information might come in the format of a distribution. In the Bayesian tradition, such distributions will ideally be posterior distributions obtained in other studies, and

are introduced to introduce into the new analysis as prior distributions. If this distributional information is understood as a confidence distribution for the parameter in question, this summary of the underlying external data can also be brought directly into the likelihood function. This is usually done by assuming a standard normal sampling distribution for the normal score of the confidence quantiles (Efron 1993, Schweder 1998).

- Practicality: For assessment models with many parameters, the calculation of the posterior distribution in a Bayesian analysis will typically be time consuming. When the joint posterior distribution has been found in the format of a sample, it is usually simple to find marginal posterior distributions. The calculation of frequentist confidence distributions for interest parameters is also usually expensive. Since in general these distributions cannot be calculated as marginals of a joint confidence distribution, they need to be calculated one at a time, typically by using simulations. The comparison of computer cost and benefit between a Bayesian and a proper frequentist analysis will depend on the number of parameters in the model, the number of interest parameters, the degree of non-linearity and the degree of non-normality of the likelihood function.
- Comprehensive uncertainty in the management context: Practically speaking, both the Bayesian and the frequentist will need to carry out an extensive scenario experiment to be able to account properly for the main bulk of uncertainties surrounding "the best stock status estimates". Such scenario experiments have a frequentist flavor in the sense that the resulting summaries in the format of distributions tend to be frequency distributions. They do also have a Bayesian bending since the scenario experiment is based on a design in the model and parameter space, which could be regarded as a Bayesian prior.

As long as the modeling and analysis is sound, and as long as additional scenario experiments properly are addressing the relevant issues that cannot be investigated by analysis of the data, it is more a matter of taste whether the assessment is cast in terms of Bayesian or frequentist concepts. Of primary importance is that the data are good and well documented, and that the structure of the data as well as their statistical variation is well represented in the model. The model must provide results relevant for management, and the model must also strike a good balance between realism and simplicity to allow optimal use of the data. However, sound modeling and analysis entails, among other things, that unnecessary subjectivity is not introduced, that method-inflicted bias have been removed and that relevant external information is used properly.

Efron, B. 1993. Bayes and likelihood calculations from confidence intervals. *Biometrika.* 80:3-26.

Efron, B. 1998. R.A. Fisher in the 21st century (with discussion). *Statistical Science.* 13:95-122.

Schweder, T. 1998. Fisherian or Bayesian methods of integrating diverse statistical information? *Fisheries Research.* 37:61-75.

Model comparison

(Consensus views)

During the second day of the workshop, the models presented were compared and evaluated based on various criteria. The following are the consensus views of the participants sorted by category.

Complexity versus ease of use

- This was thought to be related to how easy the models were to be used by a knowledgeable person and how easy it was to change the models on a stock by stock basis.
- If there is an explicit objective function the process of changing a model was thought to be more straight forward.
- The XSA model is probably the easiest model to use since there are few modifications that can be made by the user.
- The amount of time available to assess a stock is a factor in determining the level of model complexity that is suitable.
- All of the different methods were thought to be complex when you looked at the formulations and in fact use many of the same basic assumptions. Flexibest had the added complexity of using a length at age model and included the possibility of including cannibalism.
- The time series approaches allowed for slow variations in estimates over time.
- It was considered that all these models could be thought of as special cases of a common and general underlying fisheries dynamics model.

Realism

- This was thought to be the ability to incorporate alternative hypothesis, examples of which are biological interactions and environment influences on growth, natural mortality and recruitment.
- One goal in the development of realistic models is to mimic the 'truth' as closely as possible.
- The Flexibest model's incorporation of a length component makes it the most realistic since fish behaviour and development seem to be closely related to length.
- It is also important to decide what data source is the most accurate rather than include data of doubtful quality, though models should be adaptable in the sense that they can handle poor and missing data.
- It is important to choose an appropriate objective function that reflects the distributional properties of the data. Models should be flexible with respect to changes in the objective function.

- Some models may incorporate many alternative hypotheses and provide a close fit to the observed data but, nevertheless, provide poor predictions.

Management Utility

- The most important aspect here is ultimately the ability to communicate and present results. It should be easy for managers to explore different options such as make predictions conditioned on various harvesting schemes, changes in regulations, e.g., mesh size, etc.
- Complex models with high variances will not be able to predict well but this is often not the intention of such models. For example methods like that of Punt's were primarily used to construct plausible models of reality rather than to provide direct management advice. Management procedures would then be developed that were robust to the possible dynamics of the fishery as described by several possibly contradictory models. Others methods such as XSA were essentially used to provide predictions of future catch levels.
- Historically in the ICES area, XSA and Adapt VPA have been useful for testing management scenarios.
- The necessary speed of a model depends on the application. For example, simulation studies of management options should be based on relatively fast (to compute) models.

Ability to deal with uncertainty

- Uncertainty is generally modelled by either the inverse of the Hessian, likelihood profiles, Monte Carlo simulation, bootstrapping or by integrated Bayesian posteriors. The variance of the estimates generated by the time series method are obtained from the Kalman filter. The Hessian can be quick to compute using specialist routines but is often poor at the tails of the distribution. The use of likelihood profiles was thought important for critical parameters but can be difficult to obtain for derived parameters. Bootstrapping in comparison can easily provide estimates of uncertainty for even derived parameters.
- It is easy to modify the likelihood functions in all the models, and thus data uncertainty conditional on the model can be included in the modelling process.
- Flexibest does not currently include estimates of uncertainty but estimation of the Hessian will be added and it would be possible to perform a Monte Carlo simulation with the model.
- There is still a question of how a model can be used to quantify uncertainty. Since estimates of quantities, such as the relative abundance at age based on surveys, are often correlated, it may be desirable to include correlation terms in the likelihood functions.

Ability to diagnose data and model problems

- The use of summary statistics can be useful, but it was thought to be more important to look at diagnostics, primarily residual plots, and use these to decided on the appropriateness of models and/or the inclusion of data sets.
- Knowledge of dynamic processes are also useful in the interpretation of results and model selection.
- An important problem is the correlation of parameters and this is likely to become more important as the number of parameters to be estimated increases.
- Given their flexibility, statistical catch models provide the greatest potential diagnostic tools.
- If the data are giving conflicting signals on the status of a stock, then it will be difficult or impossible for any model to determine which data set is sending the correct signal. It is important to be able to detect conflicting signals.
- Much research has been conducted at the IMR to remove biases in the survey data.

Simulation ability

- This was defined as how easy it would be to evaluate the performance (as a management tool) of a particular assessment method within a simulation framework.
- The more flexibility that a method gives to an analyst the more difficult it will be to simulate the assessment process. Methods such as XSA, which give fewer options, will be easier to test than the full statistical methods.
- I t is important that models use the most modern technology, such as up to date optimisation routines, so that they can be quickly and efficiently used in simulation studies.
- Biases caused by non-linear functions in the models should be studied using simulations and the estimates adjusted for this source of bias.

Ability to use different kinds of information

- The more flexible models allow different kinds of information to be included within the estimation process. However, this often means that it is difficult to determine the correct functional form for the model.
- Statistical models, and in particular Flexibest, can be easily adapted to include auxiliary information.
- The question arises whether it is better to adjust the data before putting it in a model (for example, use time series model to generate more precise estimates of catch or survey indices) or do all the adjustments within the model.

Model robustness

- Model robustness can be interpreted as the ability of the model to avoid paying attention to outliers in the data, or as that the model will give approximately the same results with deviations from assumptions, if these apparently would be equally plausible.
- The handling of outliers will to a large extent depend on how the objective functions treats outliers and how much emphasis it puts on small numbers in the catch matrix or the survey indices. One recipe can be to downweight small numbers in addition to the weighting implicit in the likelihood function. Another way can be to add a small number to each of the numbers that are compared to avoid giving relative too much weight to small, poorly estimated values.
- In regards to assumption robustness, the opinion of the group was that this to a large extent would be case sensitive. In general, assessing stocks with low fishing mortalities are likely to lead to such problems. If different data give different signals, there will be problems if alternative model formulation pick this up differently. If so, change in weighting may lead to substantial differences. If model formulations are over parameterized, in the sense that they imply several equivalent explanations to the observations, this should show up as strong correlations between parameters, or even singularities in the Hessian.

Further development of Flexibest—advice

Short term

Expand on the specification of options for the objective function

- Develop more flexible tools for aggregation of data.
- Develop alternative likelihood functions.

Develop ways to deal with uncertainty

- Derive Var-Cov matrix from the Hessian.
- Examine Bootstrapping possibilities.
- Explore using simulation studies for measuring uncertainties.

Develop parallel models

- Simpler models that may be more responsive for providing management advice.
- Length selection functions may be easily mapped into equivalent age-specific selection function for examining selectivity issues.

Input/output improvements

- Provide more diagnostic capabilities.
- Develop graphics to that plot residuals by age/time/cohort and predicted abundance.
- Derive and employ Var-Cov matrix for testing convergence and evaluating correlations quickly.

- Employ techniques to evaluate potential model generated biases in parameter estimates.

Optimisation

- Optimisation routines need to be improved in order to increase model speed.
- The ability to estimate parameters in phases using more efficient algorithms (e.g., quasi Newton) may help with some problems.

Long term

- Conduct proper and thorough simulation testing of the model.
- Identify data sources that may solve problems, e.g., resolve conflicts in the data.
- Consider the desirable levels of model resolution (time, space, length, sex).

Concluding remarks
(Consensus views of invited experts)

Flexibest is a model that has great ability to incorporate hypotheses of, for example, biological interactions, changes in fleet composition and behaviour, and environmental influences. It can easily be adapted to include a variety of data, and it also has great potential ability to deal with uncertainties.

However, at present Flexibest has not been developed to a stage where it is appropriate for use in stock assessments or as a basis for giving management advice. In particular, it was pointed out that it can not yet deal with uncertainties in an appropriate manner. The group noted a variety of developments, both short- and long-term, that should be pursued. When the model has been developed to a stage where it can be used in assessments, it should for a period be applied in parallel with existing models such as XSA.

Panel discussion

4 December 1998

Day 4 was set aside for presentation of the conclusions from the workshop to a broader, invited audience, followed by a general discussion on assessment strategies.

Opening

Øyvind Ulltang, chairman of the workshop.

Ulltang gave a summary of the background and objectives of the Flexibest project. The project was initiated two years ago with the aim of constructing an alternative to the VPA-tuning models traditionally used for many stocks in the ICES area. The basic idea was to obtain greater flexibility as to the choice of model assumptions and allowing more appropriate use of available data. Because of the experienced assessment problems with the North-East Arctic cod, the project was expanded this year with respect to resources allocated to the project, and the priority was changed to develop as soon as possible a new assessment model for North-East Arctic cod, utilizing the basic ideas in the original more general project.

The objective of the Workshop was to compare different types of assessment models, including the Flexibest model, in order to promote the understanding of similarities and differences between assessment methods used in various parts of the world and elucidate how reliable our peception of the state of the North-East Arctic cod stock is. The objective was not to make a new assessment of the stock. When comparing the models, the group decided to discuss the potential of the various approaches in relation to a number of criteria, and details are given on pages 17-21.

The report from the Workshop will be made available to anyone who are interested in the results.

Assessment problems for Northeast Arctic cod

Odd Nakken, IMR

Nakken gave a brief summary of the development of the Northeast Arctic cod stock and fishing pressure during the period 1946 through 1998 as indicated by a retrospective analysis based on catch-at-age data. Since these catch data do not include discards or probable unreported landings that have occurred to a varying degree throughout this period, the estimates of recruitment obtained are most likely subject to error and may be biased.

He showed that it was the rule rather than the exception that estimates of mortality rates generated by the annual assessments have commonly been too low when compared with assessments undertaken some years later. Hence fish had been removed from the stock at a higher rate than the scientists thought at the time they gave their advice on the status of the cod stock. This chronic underestimation of mortality rates indicates a need for an in-depth investigation of the reliability of the catch statistics as well as a careful review of assessment methodology. However, until the causes of these discrepancies are known, managers should take a more precautionary approach when setting a TAC than in the past. The general tendency has been to set the TAC at or above the advised level, while these findings imply that the TAC should be set at or below the advised level.

Discussion points:

- Problems in the assessment of Northeast cod became apparent about one and a half years ago. Consistently biased assessments are not just a problem for Northeast Arctic cod but also a problem for assessments of many other stocks.
- We should not adjust future estimates of cod abundance based on past estimates of the bias until we know the cause of the bias, though managers should be made aware of the problem and base their decisions on the fact that current estimates of stock abundance have tended to be over optimistic in the past.
- The underestimates of stock size may have been caused by inaccurate catch statistics due to, perhaps, misreporting, discards etc. It has been recommended that an extensive sampling scheme for collecting improved commercial catch data be implemented.
- An assumed level of hake discards off the west coast of the USA has been included in the assessments and, perhaps, the Arctic cod catch statistics could be similarly adjusted.
- Part of the problem with the Northeast Arctic cod assessments may have been caused by changes in the surveys. The surveys have not covered the Russian Zone during the last couple of years and thus the recent survey estimates are more variable and uncertain. It appears that the problem of survey coverage will be solved before next year's surveys are conducted.

A summary of the current status of Flexibest (for details, see pages 5-7).
Dankert Skagen, project leader, IMR.

Discussion points:

- The Flexibest model does not include a migration component because of the apparent variability in migration patterns. Technically, it is possible to include migration in the model, but the problem of setting up a realistic migration matrix has not so far been given priority.
- Total mortality can be estimated from the model so that natural mortality could be estimated rather than assuming its level. However, the data are rarely sufficiently accurate for estimating natural mortality.
- The model XSA, which is the basic assessment tool used by ICES, is easy to use and is fast. Flexibest currently takes much longer to run but model speed will

improve significantly in future versions. It was pointed out that the more parameters included in a model, the longer it will take to run, especially if one wants to measure the associated uncertainty. For example, specification and application of one assessment model in Australia takes approximately eight months to complete, including variance estimation.

- Conflicts among data sets can not be solved through modelling. What we need is some tools to describe, evaluate and relate the various data sets in an objective manner.
- Since the Northeast Arctic cod stock size seems to have considerable natural variability, we need the ability to detect rapid changes in abundance and thus we must, perhaps, rely on multiple observations such as surveys.

Presentation of different assessment approaches used by invited experts (see pages 7-18 for details).

Laurence Kell, UK; Gudmundur Gudmundson, Island; James N. Ianelli, USA; Andre Punt, Australia, Stratis Gavaris, Canada

Discussion points:

- The purpose of all the modelling exercises was not to assess the Northeast Arctic cod stock but to compare modelling methodologies and underlying model structures. Thus the actual values generated by each model run are not comparable.
- In Australia and Canada, scientists do not provide the managers with single estimates of TAC, but rather with the trade off between risk and reward achieved by a range of possible TACs, often for several assessments. The managers need to explicitly deal with uncertainty when setting TACs.
- The XSA model can be adapted to produce estimates of uncertainty. Managers want to know what the total level of uncertainty in any assessment is but it is difficult to quantify 'catastrophic' events such as a 50% die off due to natural causes. The primary goal of the XSA model is to give answers to management in the form of point estimates.
- The Flexibest model is especially effective for combining all sources of uncertainty into a single model.
- In practice, uncertainty measurements are conditioned on the particular model used to generate the measures of uncertainty.
- For all the models, more simulation studies should be done to study uncertainty.
- In Australia, the first step in the assessment process is to gather all information and available expertise on the stock in question. The modelling process is used to sort out which hypotheses are most relevant for management.
- The most important step in modelling a resource is, perhaps, to understand the basic data. All data sets may not be correct and models cannot deal with systematic errors (i.e., bias).
- Surveys rarely cover the entire range of a stock, either vertically or horizontally. This source of uncertainty is difficult to measure or include in a model.

- Since a model is ultimately only as good as the data it is based on and resources are limited, it is important to decide if it is more critical to improve the collection of catch statistics or survey data.
- Simple models may be more effective for simulation studies to measure uncertainty. Flexibest as it stands could serve as a tool for designing simpler models, that is it could highlight the essential features that need to be included in a simple model.

Comparison of the different assessment approaches.
Michael Pennington, IMR

The models were compared based on several criteria which included; model complexity, realism and use of varied information, ability to deal with uncertainty and ability to diagnose data and model problems. All the models presented were based on a subset of basic assumptions and mathematical formulae and thus in that sense, all were equally complex. Flexibest has the added complexity, as compared with the other models, of incorporating both age and length within the model as well as the effects of cannibalism. The time series model (Gudmundson) has the feature that the estimates of catchability coefficients and fishing mortalities are based on the underlying unobserved time series.

One goal in the development of realistic models is to 'mimic' the truth as closely as possible. Since Flexbest includes lengths explicitly and fish length is closely related to biological characteristics, such as maturity, it has the greatest potential ability to incorporate relevant biological data.

All the models can be used to measure either directly or indirectly the uncertainty associated with the estimates in various ways such as through simulations, bootstrapping etc. Flexibest is currently being modified to routinely generate measures of uncertainty for the key estimates.

It is crucial that for any model that there is a way to check whether the model adequately describes and interrelates the input data in a consistent manner. All the models presented provide a means to check for model adequacy and they all have the ability to diagnose data, i.e. determine whether the various input data sets are sending conflicting signals about the stock being modelled. Given the added complexity of Flexibest, it has the potential to generate the most varied diagnostic tools.

What is apparent from the results of the analyses based on all the models of the Northeast Arctic cod stock data sets, is that if the data are sending conflicting signals, then it is difficult or impossible for any model to determine which data set is giving the correct signal. This inability of models to compensate directly for apparently inconsistent data was clear from the analyses of the Northeast Arctic cod data. All the models detected that the catch data and the survey data, in particular the Barents Sea survey series, were giving conflicting signals of trends in abundance. Thus for each model, with the exception of the time series technique, it was necessary to set, *a prior*, the relative weight given to each data set before the model was run, and in turn, these weights to a great extent determined the model's outputs. The basic problem in determining the relative weights is that there is, in contrast to the case for the survey data, no measure available of the uncertainty associated with the catch statistics.

One way to ascertain which data series, catch or surveys, is sending the more accurate signal on the status of the cod stock is to compare both series with other biological information. For example, Asgeir Aglen assumed that the catch data were accurate and showed that this implied that the catchabilty of the Barents Sea winter survey varied considerably over time. By comparing these changes in survey catchability with the abundance of capelin, he demonstrated that the changes in survey catchability may be caused by changes in the abundance of capelin. On the other hand, estimates of the relative abundance of early juvenile cod from surveys conducted by the IMR from 1978 through 1991 are more closely and reasonably related to estimates of their relative abundance as three year olds produced by the winter surveys than to the estimated number of three year olds based on the catch statistics, which is evidence that the catch data may be seriously biased. As part of the Flexibest modelling project, such external sources of information are being examined and used to determine the weight to be given to each data set within the model, rather than arbitrarily selecting the weights, in order to reduce or eliminate the biases in the model based estimates.

Panel discussion

The major points raised by individuals during the panel discussion were:

- It may not be the case that complex models produce more precise and accurate assessments than do simple models. For example, non-linear time series models often 'fit' the historical series closely but it has been found that simple linear time series models often give better predictions. Complicated models can be used to test simple models.

- Flexibest has the ability to include mechanisms that are considered important. The inclusion of both length and age within the model adds complexity.
- When modelling a fishery, it is important to take into account the quota strategy. One falls into a trap if the quota is decided each year for then the measurement error becomes part of the feedback loop. One should have at least a five year strategy for implementing an effective model that takes into account feedback dynamics. Statistical models are too simple, one should have a dynamic model (e.g., model of the entire system including feedback loops).
- What we need are more tools for simply monitoring a stock, especially a highly dynamic one such as Northeast Arctic cod.

- Perhaps an economic submodel could be added to Flexibest and hence use made of economic data.
- The Northeast Arctic cod stock should be studied and modelled as far back in time as possible in order to explain the natural variations in stock size and recruitment. Recruitment variability then can be related to climatic conditions, spawning stock size, etc.
- There should be more co-operation between modellers and biologists in the field. Such field experience could be important source of information so that 'bad' data are not used as inputs to Flexibest.
- Three points were raised:
    1) There appears to be a serious lack of 'statistical thinking' in the Flexibest project.
    2) Flexibest may be too complicated.
    3) The Flexibest project must be better funded so that a larger group can be employed in the modelling effort.
- It was noted that it should be determined how better co-operation could be fostered with Universities in the development of Flexibest.

Where do we go from here?
(Individual opinions)

- If the survey estimates are correct, then perhaps we should go along with the survey estimates.
- The IMR's multi-species model of the Barents Sea gave the best predictions in 1993 of cod and capelin stock sizes in 1996.
- Environmental factors should be included in the Flexibest model. Environmental conditions may be effective for making more accurate predictions rather than for modelling the past.
- The effect of commercial activity on the entire ecosystem should be considered.
- Flexibest should take into account fluctuations in natural mortality.
- The development of Flexibest should be done in close co-operation with ICES.
- How complex Flexibest should be depends on the objectives such as harvest strategies, etc.
- Since management needs point estimates, we will have to produce accurate and precise point estimates for the foreseeable future.

Concluding remarks
(Consensus views of invited experts)

Flexibest is a model that has great ability to incorporate hypotheses of, for example, biological interactions, changes in fleet composition and behaviour, and environmental influences. It can easily be adapted to include a variety of data, and it also has great potential ability to deal with uncertainties.

However, at present Flexibest has not been developed to a stage where it is appropriate for use in stock assessments or as a basis for giving management advice. In particular, it was pointed out that it can not yet deal with uncertainties in an appropriate manner. The group noted a variety of developments, both short- and long-term, that should be pursued. When the model has been developed to a stage where it can be used in assessments, it should for a period be applied in parallel with existing models such as XSA.

Appendix I

Working papers presented at the workshop

Bogstad, B., 1998. Fleksibest - data documentation. 9 p.

Gavaris, S., 1998. Report to: Workshop on comparison of stock assessment model strategies, with application to North-east Arctic cod, Bergen 1-4 Dec. 1998. 32 p.

Gudmundsson, G., 1998. Time series estimation of stocks and fishing mortality rates of NE-Arctic cod. 15p.

Ianelli, J. N., 1998. Example application of a statistical age-structured model. 28 p.

Kell, L., 1998. A combined Monte Carlo simulation and bootstrap of extended survivors analysis (XSA): a North east Arctic cod example. 14 p.

Punt, A. E., 1998. Illustrative applications of maximum likelihood and Bayesian methods of North-east Arctic cod, with comments on further work. 55 p.

Skagen, D. W., Frøysa, K. G. and Bogstad, B., 1998. Fleksibest as a model for historical stock assessment of North-east Arctic cod. 34 p.

Tretyak, V. L., 1998. On the possibility of the use of Ricker's model "stock-recruitment" for the assessment and prediction of the recruitment of the commercial part of the Northeast Arctic cod population. 17 p.

Appendix II

Northeast Arctic cod – data available for assessments

Four categories of data are currently used in the assessment of Northeast Arctic cod: Commercial catch data, stock indices from research surveys, cod stomach content data and CPUE data from the commercial fishery. More information on the data available and the methods for collection of data is given e.g. in ICES (1997). The data used in last year's assessment are given in ICES (1999). Cod stomach content data were not made available for use by the various models, and not all of the models used the same parts of the data that were made available.

Commercial catch data:

Data on catches by length and age are available for the Norwegian fishery (5 main fleets) and for the Russian fishery. The catch by these two countries accounts for 85-

90 % of the total catch. Age and length compositions are also available for some of the catch taken by third countries.
The time series of commercial catch data go back to 1946, but for the time series prior to 1982, constant weights at age are used. This creates an inconsistency in time series, which should be resolved.

Data from research surveys

The following stock indices from research surveys are available:

1. Bottom trawl and acoustic indices from the Norwegian survey in the Barents Sea in February, covering mainly the immature stock, (Jakobsen et al. 1997). Data from 1981-1998.
2. Acoustic indices from the Norwegian Lofoten survey in March/April, covering most of the mature stock (Korsbrekke 1997). Data for 1985-1998.
3. Trawl indices from the Norwegian Svalbard survey in August/September, covering part of both the mature and immature stock. Data from 1983-1997. This survey was in 1995 extended to a combined acoustic and bottom trawl survey for the entire Svalbard and Barents Sea area, which so far has not been used in the assessment.
4. Bottom trawl and acoustic indices from the Russian survey in the Barents Sea in autumn (Lepesevich and Shevelev 1997), covering both the mature and immature stock. Data from 1982-1997.

   The bottom trawl and acoustic indices stemming from the same survey (1 and 4) are correlated, because in part, the same length and age samples are used in the calculation of the indices.

   All the indices have been recalculated in order to obtain indices by age and length for use in a length-structured model. In this process, some of the time series have been significantly revised.
   When using the survey data, changes in methodology during the time series (gear changes, changes in acoustic methodology, area coverage etc.) should be accounted for as far as possible.

   Detailed information about the area coverage, number of samples etc. are given in the survey reports for the various surveys, at least for recent years.

   There has not been done much work on quantifying the uncertainty of the various survey estimates. One example of quantification of uncertainty is from the Norwegian bottom trawl survey, where the coefficient of variation for the number of fish in each 5 cm length group is given in the survey report (Mehl, 1998).

Stomach content data

The joint IMR(Bergen)-PINRO (Murmansk) stomach content data base contains data on cod stomach content, sampled annually since 1984. Together with a model for stomach evacuation rate, these data can be used to calculate the consumption by cod of various prey species, including cod as prey (Bogstad and Mehl, 1997). Such estimates of the consumption of cod by cod are already included in the assessment.

CPUE data

Such data are available for the Norwegian, Russian and Spanish trawl fisheries.

Data of potential use in assessments.

These include data on the abundance of other stocks (particularly capelin), and environmental variables (e.g. temperature). Capelin may influence both growth and mortality of cod, as well as survey catchability. Temperature may influence both growth and recruitment. These influences are rather strong for this stock.

References

Bogstad, B. and Mehl, S. 1997. Interactions Between Cod and its Prey Species in the Barents Sea. Proceedings of the International Symposium on the Role of Forage Fish in Marine Ecosystems, Anchorage, Alaska, 13-16 November 1996. Alaska Sea Grant College Program, AK-SG-97-01.

ICES, 1997. Report of the Comprehensive Fishery Evaluation Working Group. ICES C.M. 1997/Assess:15.

ICES, 1999. Report of the Arctic Fisheries Working Group. ICES C.M. 1999/ACFM:3.

Jakobsen, T., Korsbrekke, K., Mehl, S. and Nakken, O. 1997. Norwegian combined acoustic and bottom trawl surveys for demersal fish in the Barents Sea during winter. ICES C.M. 1997/Y:17.

Korsbrekke, K. 1997. Norwegian acoustic survey of Northeast Arctic cod on the spawning grounds off Lofoten. ICES C.M. 1997/Y:18.

Lepesevich, Yu. and Shevelev, M. 1997. Evolution of the Russian survey for demersal fish. From ideal to reality. ICES C.M. 1997/Y:09.

Mehl, S. 1998. Investigations on demersal fish in the Barents Sea (reduced area) winter 1998. Fisken og havet nr 7 – 1998. ISSN 0071-5638 (In Norwegian, with English legend to tables and figures).

Appendix III

INVITED ASSESSMENT EXPERTS

Group 1. Prepared assessments in advance.

Gudmundur Gudmundsson
Islands Sedelbank
Kalkofnsvegur 1
150 Reykjavik
ISLAND

James N. Ianelli
Alaska Fisheries Science Center
7600 Sand Point Way NE Bldg 4
BIN C15700
Seattle, Wa 98115-0070
USA

Laurence Kell
CEFAS
Lowestoft Laboratory
Pakefield Road
Lowestoft, Suffolk NR33 0HT
ENGLAND

Andre E. Punt
CSIRO Division of Marine Research
GPO Box 1538
Hobart
Tasmania 7001
AUSTRALIA

Stratis Gavaris
Marine Fish Divison
Dept. Of Fisheries and Oceans
Biological Station
St. Andrews, NB
Canada E0G 2X0

Group 2. Experts participating without bringing assessments of their own.

Ulltang, Øyvind   Institutt for fiskeri- og marinbiologi

	Universitetet i Bergen N-5020 Bergen
Tretyak, Victor	Pinro, 6 Knipovich Street 183763 Murmansk Russia
Yaragina Natalya	Pinro, 6 Knipovich Street 183763 Murmansk Russia
Kovalev, Yuri	Pinro, 6 Knipovich Street 183763 Murmansk Russia
Schweder, Tore	Sosialøkonomisk Institutt Postboks 1095 Blindern 0317 Oslo
Bjørnsson, Høskuldur	Marine Research Institute P.O.Box 1390 121 Reykjavik Island
Maguire, Jean Jacques	1450 Godefroy Sillery, Quebec, Canada G1T 2 E4
Bowering, Ray	Northwest Atlantic Fisheries Centre P.O.Box 5667 St. Johns, Newfoundland, Canada A1c5X1
Harbitz, Alf	The Norwegian Institute of Fisheries and Aquaculture Ltd. P.O. Box 2511 9002 Tromsø

IMR:
Aglen, Asgeir
Jakobsen, Tore
Nakken, Odd
Mehl, Sigbjørn

Pennington, Michael, reporter
Bogstad, Bjarte
Frøysa, Kristin G.
Hiis-Hauge, Kjellrun
Thiem, Øyvind
Åsnes, Morten
Nygaard
Korsbrekke, Knut
Bjordal, Åsmund
Pedersen, Åse
Vaage, Roald

Group 3. Invited to the panel discussion.

Dag Tjøstheim	Dept. of Mathematics University of Bergen Allegt. 55 5007 Bergen
Gro Hagen	Norwegian Computing Center P.O. Box 114 Blindern, N-0314 Oslo
Harald Yndestad	Møre Research P.O. Box 5075 N 6021 Ålesund
Lars Horn Nina Hedlund	Research Council of Norway Stensberggata 26 P.O. Box 2700 St. Hanshaugen 0131 Oslo
Erik Berg	The Norwegian Institute of Fisheries and Aquaculture Ltd. P.O. Box 2511 9002 Tromsø
Astrid Pestalozzi Gro Lagesen	The Ministry of Fisheries P.O. Box 8118 Dep 0032 Oslo
Per Sandberg	The Ministry of Fisheries P.O. Box 185 5002 Bergen

Jarle Berntsen | Dept. Of Mathematics
University of Bergen
Johannes Brunsgt 12
5008 Bergen

Lars Walløe | Dept. of Physiology
University of Oslo
P.O. Box 1103 Blindern
0317 Oslo

IMR
Morten Skogen
Terje Monstad
Are Dommasnes
Odd Smedstad
Johs. Hamre
B. Stensholt
H.J. Skaug
E. Svendsen
Å. Fotland
Hein Rune Skjoldal
Kjell Nedreaas