WORKSHOP ON THE REVIEW AND FUTURE OF STATE SPACE STOCK ASSESSMENT MODELS IN ICES (WKRFSAM)

VOLUME 2 | ISSUE 32

ICES SCIENTIFIC REPORTS

RAPPORTS SCIENTIFIQUES DU CIEM

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

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

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

ISSN number: 2618-1371 I

ICES Scientific Reports

Volume 2 | Issue 32

WORKSHOP ON THE REVIEW AND FUTURE OF STATE SPACE STOCK ASSESSMENT MODELS IN ICES (WKRFSAM)

Recommended format for purpose of citation:

ICES. 2020. Workshop on the Review and Future of State Space Stock Assessment Models in ICES (WKRFSAM).
ICES Scientific Reports. 2:32. 23 pp. http://doi.org/10.17895/ices.pub.6004

Editors

Noel Cadigan

Authors

Sondre Aanes • William Aeberhard • Christoffer Albertsen • Magne Aldrin • Jonathan Babyn • Casper Willestofte Berg • Olav Breivik • Robin Cook • Joanna Mills Flemming • Rob Fryer • Emily Liljestrand Colin Millar • Tim Miller • Cóilín Minto • Ken Newman • Anders Nielsen • Andrea Perreault • Paul Regular • Hans Skaug • Michael Spence • Vanessa Trijoulet • Sindre Vatnehol

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

Contents

i Executive summary ii
ii Expert group information iii
1 Introduction 1
1.1 Background 1
1.2 Conduct of the meeting 1
1.3 Structure of the report 2
1.4 Follow-up process within ICES 2
2 Advice on state-space assessment models 3
2.1 State-space assessment models (SSAMs) 3
2.2 Model formulation and selection 4
2.3 Model estimation efficiency and robustness 6
2.4 Model Validation 7
3 Future research 11
4 References 12
Annex 1: Resolution 13
Annex 2: List of participants 15
Annex 3: Workshop agenda 16
Annex 4: Abstracts 19

i Executive summary

The Workshop on the Review and Future of State Space Stock Assessment Models in ICES focused on future directions of state-space assessment models for ICES stocks (WKRFSAM), utilising recent advances in fisheries modelling research to help define best practises. State-space models consist of a process model for unobserved quantities (e.g. true stock abundances) and an observation model for observed quantities (e.g. catches or survey data). Standard statistical assessment models do not include stochastic population processes. Prediction is a natural part of the state-space model formulation which is a more practical advantage of the approach for stock assessment.

Lognormal observation error models for survey indices can be parameterized such that the mean or median is proportional to the true stock abundance, but this makes little difference on assessment results. Empirical results for 14 stocks indicated that this is also the case for catch observation models. There is no practical difference whether the mean or median of fishing mortality rates (F's) are constant for the data period, but there are important differences if F is projected into the future.

Including random deviations in the natural mortality rate (M) leads to convergence issues for some stocks, and similar assessment results for other stocks, but this depends on the details of how this is implemented. Including time- and age-varying M increases model flexibility and potential confounding of F, M, and stock size index catchability, Q. When model parameters are confounded then we can anticipate less robustness.

Including external variances and correlations that are reliably estimated will be more relevant and useful in situations where these differ substantially over time, such as when a survey in some year has a large set, or poor coverage, etc.

Two general criteria to select between alternative models are Goodness-of-it and Out-of-sample prediction. Minimizing out-of-sample prediction error (e.g. Akaike Information Criterion, AIC) is increasingly seen as a better approach for model selection.

The main tool for model validation are residuals. However, for state-space models, Pearson are not independent because of the dependence structure of the unobserved states. One-observation-ahead and one-step-ahead residuals are preferred. These can be formulated to be independent and Gaussian distributed.

Key research recommendations involve when and how M can be estimated, how to include information about the precision of model inputs, and improved usage of diagnostics and model selection criteria.

ii Expert group information

Expert group name	Workshop on the Review and Future of State Space Stock Assessment Models in ICES (WKRFSAM)
Expert group cycle	Annual
Year cycle started	2019
Reporting year in cycle	1/1
Chair(s)	Noel Cadigan, Canada
Meeting venue and dates	ICES HQ, Denmark, 21-23 January 2020 (23 participants)

1 Introduction

1.1 Background

State-space stock assessment models make up a large proportion of the stock assessments for category 1 stocks in ICES. It is important that stock assessors and reviewers of ICES stock assessments understand advantages, disadvantages and limitations of the underlying formulations of state space models, which was the focus of this workshop (i.e. WKRFSAM). This is important from the view point of extending models to include new dynamics and new datasets but also to review current model formulations with respect to new developments in fisheries science.

1.2 Conduct of the meeting

The list of participants and agenda for the workshop are presented in Annex 2 and Annex 3, respectively.

No working documents were received prior to the meeting but contributed presentations were first given by participants which provided a partial basis for three discussion groups during the remainder of the meeting. Abstracts of presentations are provided in Annex 4. The conclusions of these groups provide the basis for this report. The discussion groups and topics were:

a) Model formulation and selection
- The F and the M process models
- Random effects on survival
- Random walks on fishing mortality in log scale
- Other components of variation

b) Model estimation efficiency and robustness
- Do certain formulations affect model robustness or result in models with impractical optimisation times
- Observation error models
- Treatment of survey indices and fishery catch statistics
- Modelling catches and survey indices on the log scale

c) Model validation
- Do different models provide a practical difference?
- Are there reliable ways to conduct model selection between alternatives?
- Diagnostics

Several participants worked by correspondence during the meeting and the facilities of WebEx were relied upon for their contribution to the workshop plenary discussions. This included two participants who could not attend in person because of flight cancellations. Attendance by WebEx worked reasonably well.

Given ICES role as a knowledge provider, it is essential that experts contributing to ICES science and advice maintain scientific independence, integrity and impartiality. It is also essential that their behaviours and actions minimize any risk of actual, potential or perceived Conflicts of Interest (CoI).

To ensure credibility, salience, legitimacy, transparency and accountability in ICES work, to avoid CoI and to safeguard the reputation of ICES as an impartial knowledge provider, all contributors to ICES work are required to abide by the ICES Code of Conduct. The ICES Code of

Conduct document dated January 2019 was brought to the attention of participants at the workshop and no CoI was reported.

1.3 Structure of the report

The outcomes of the subgroup discussions are presented in Section 2. The three subgroups described above provided summary text that was reviewed in plenary. It was obvious that there was significant discussion overlap between subgroups on some topics.

1.4 Follow-up process within ICES

Subgroup leads at WKRFSAM agreed to provide text for the draft workshop report and to then comment on the compiled draft report. The draft report was then provided to all workshop participants for comment, and after this the report was finalized by the Chairs and formatted by the ICES Secretariat.

2 Advice on state-space assessment models

2.1 State-space assessment models (SSAMs)

The state-space paradigm is in many ways the natural choice for the challenge of assessing a fish stock. The actual size of a fish stock is an unobservable process, which is assumed stochastic because it is influenced by many things we can never hope to directly account for in our models. The observations available to estimate the stock size process are often indirect and always subject to observation noise. State-space models are designed exactly for such situations.

The main difference between a standard statistical (full parametric) assessment model and a state-space assessment model (SSAM) is that the latter allows for quantities which are unobserved to be random variables with a specified probability distribution. In particular, SSAMs consist of two models for two time-series, a process model for unobserved quantities, e.g. true abundances, and an observation model for observed quantities, e.g. catches or sample data. This gives the flexibility to formulate models where time-varying quantities follow e.g. a random walk or an autoregressive (AR) process.

Estimating model parameters in state-space models require evaluation of high dimensional integrals, which until recently was often not feasible for estimation and simulation testing of full-scale assessment models. However, with recent advances in algorithms and software (e.g. Kristensen *et al.*, 2016) the run-time to fully optimize such models (a few seconds) is no longer problematic.

A more practical advantage of SSAMs compared to full parametric and deterministic models is that the method to do short-term predictions is a natural part of the model formulation. SSAMs are formulated via the transitions from one year to the next and the uncertainties of these predicted transitions.

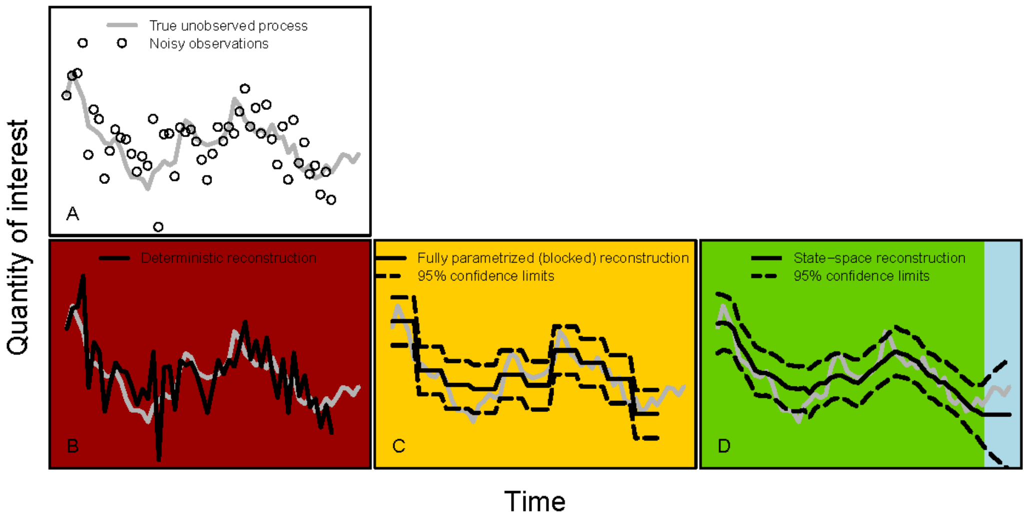

Figure 1. Simplified illustration of an assessment problem. The challenge is to estimate an unobserved process from indirect observations with noise (A). Deterministic models assume that observation-noise is negligible and reconstructs the unknown process accordingly (B). Full parametric statistical models express the unknown process as a simplified parametric function (here blocks), then estimates all parameters (C). State-space models formulate a stochastic model

for the unknown process via a few model parameters, then predicts the unknown process in the data period and beyond (D).

2.2 Model formulation and selection

Types of data and observation models

The same types of observation models for fishery independent and dependent data sources are available to both state-space and traditional statistical catch-at-age models. The suite of observation models can be generally split into two types: those that use estimates of numbers or biomass captured at age or those that model the aggregate catch and the proportions at age separately. Some assessment models even separately model proportions at length and proportions at age conditional on length (i.e. age-length keys) separately. However, the types of models to consider for the marginal proportions at age or the marginal proportions at length and the conditional proportions at age are the same. Recent work by Albertsen *et al.* (2017) provides a methodology to compare the relative performance of alternative observations models using the AIC intervals of alternative data likelihoods.

Lognormal observation error models are commonly used. Whether the observations models are parameterized such that the mean (e.g. Aldrin *et al.*, 2019; Nielsen and Berg, 2019) or median is proportional to the true stock abundance for survey indices has very little effect because the transformation bias correction is confounded with the unknown catchability Q parameters (e.g. Cadigan and Myers, 2000). Some explorations of the consequences of mean- or median model assumptions was presented (14 stocks), and the effect was small, presumably because the catch error standard deviations were low. In practice, transformation adjustment seems to have rather small influence on the results. For example, there was only a 5% difference between mean and median when the log catch error standard deviation (SD) was 0.3, and this SD is at the high end of what is normally observed.

Observation error models, particularly for catches, that are constructed to be mean-unbiased will be sensitive to the size of the estimated variance of these components and the uncertainty in the estimates of corresponding variance parameters, whereas this should be less of a problem for observation or process error models that are constructed in terms of medians. However, the empirical results examine during the workshop suggests that catch transformation bias correction will have little influence on assessment results

Model frameworks that can accommodate multiple fleets could be used for modelling discards as a separate discarding fleet in addition of the landing fleet(s). Discard mortality would be therefore treated differently from landings-based mortality and fishing mortality reference points could lead to quotas that explicitly account for variations in future discards rather than assuming discard proportions are constant in the future. However, estimating discards would be dependent on the availability of discard data and alternative observation models may be needed to accommodate larger frequencies of zero observations. Other models that can model discarding practices via the use of priors on discard proportions may be helpful.

An interesting long-term research question involves how much aggregation or summarization of raw data should be done prior to modelling population dynamics and catches. What are the advantages of working with indices, for example, compared to temporally or spatially finer scale sample data? Information is clearly lost by aggregation and summarization, but what is the practical effect? By using such an approach, we are able to utilize more information from the surveys and catch data. Accounting for uncertainty in indices and catch estimates may lead to improved estimates of population dynamics (e.g. Knape, Besbeas, and de Valpine, 2013) It was discussed if the models for indices and catches need to accommodate for the sampling approach. Levels of the sampling procedure can be included as latent variables and thereby be accommodated for. It was noted that raw data approaches with catch observations may be problematic for legal reasons.

Including external variances and correlations that are reliably estimated will be more relevant and useful in situations where these differ substantially over time, such as when a survey in some year has a large set, or poor coverage, etc. However, with the further development of the Regional Database Estimation System (RDBES), it is expected that raised (i.e. expanded) catch data will be provided along with their covariances. In addition, there is an increase in the use of model-based approaches for survey index estimation which may introduce correlation across ages through the use of common covariates. Therefore, considering the inclusion of the correlation across ages in observation models may become standard in the future. The RDBES may also facilitate access to raw data for use in stock assessment models directly.

The F and the M process models

In general, time-series models for F can be useful in predicting F for years where catch was known to have occurred, but there are no observations. Time-series models include the simple random walk, more complicated multivariate normal random walks with correlation among ages, and multivariate autoregressive models with correlation across ages and years. These approaches may also be used to model M in some situations. The random walk, like other classical time-series models, is a simplification of the actual process determining F. In reality, F is determined by the stock size and fishing effort, which in turn is determined by, for instance, quotas, fuel price, and the spatial distribution of the stock. While simple time-series models are useful for reconstructing past F's in state-space stock assessment models, they may not be appropriate for long-term forecasting or scenario modelling. Therefore, future research may be directed at determining links between fishing mortalities and the underlying driving factors, namely, to model F as a function of measurable covariates to improve forecasting and understanding of the consequences of management.

For practical purposes there is no difference whether the mean or median of F are constant for the data period, but there are important differences if F is projected into the future. Regardless, it is unlikely to be useful to project F in a management setting because F or catch are typically specified for short-term projections and in longer-term projections when investigating management procedures and harvest control rules.

Alternative interpretation of variability in numbers at age due to variability in M or not by changing process variability from abundance at age to natural mortality often provides similar model fits and assessment results when applied to several stocks in the North Atlantic. It is also reasonable to consider the inclusion of environmental conditions to model variation in M.

Other issues

Checking for parameter identifiability is only possible with simulation studies for current assessment model frameworks. There is a difference in theoretically identifiable vs. practically identifiable due to insufficient information (also known as lack of estimability). The former is a subset of the latter in that models may be theoretically identifiable, but the model may not be identifiable given particular realizations of the data and the frequency of this sample-dependent identifiability may also depend on the sampling methodology. A readily available method to determine theoretical identifiability of configured assessment models would at least determine whether to exclude models that will never be practically identifiable. There have been recent advances (Cole and McCrea, 2016; Polansky, Newman, and Mitchell, 2019) in theoretical approaches to determining identifiability in state-space models that are worth exploring.

Spatial and temporal refinement are worthwhile longer term investigations. Discrete space models would be useful for modelling populations in multiple management areas. Sub-annual time intervals could be useful to isolate seasons within a year where different mortality sources or movement between regions may occur. These features may affect parameter identifiability. It can improve identifiability of mortality rates in certain cases although movement rates may not be identifiable without likelihood components for tagging experiments. The finer the temporal and spatial resolution of the model, the closer we get to requiring raw data observations to fit the model.

2.3 Model estimation efficiency and robustness

Model formulation

Robustness and efficiency may have different definitions depending on the context. Robustness may refer to e.g. proportion of converged models in a simulation study or consistency in estimates of stock size under slightly different model formulations, but it may also refer to how sensitive an estimator is to outliers or slight violations of the model assumptions or small perturbations to data. Efficiency may refer to either computational efficiency i.e. running times, or to the precision of an estimator as measured by variance or mean squared error. In this section we consider efficiency mainly as the amount of computational effort needed, and we will consider both definitions of robustness.

The formulation of a state-space stock assessment model, including formulation of the process model and the observation model, will affect model robustness and result in models with low computational efficiency and impractical optimisation times. Lack of sparseness in the Hessian matrix of the joint negative log-likelihood with respect to the unobserved variables will increase optimisation times. In Gaussian models, sparseness in the Hessian matrix is obtained by conditional independence between most unobserved variables.

Some model choices may seem a little more realistic and produce small improvements in fits to data (see Section 2.4 below), but could result in long optimization times and convergence issues and therefore less simulation testing or examination of alternative model formulations, etc. and result in poorer stock assessment overall. Small improvements in model fit should not be the sole motivation for model selection if this is associated with much longer run times. We should include "jitter convergence rates" (e.g. Cass-Calay *et al.*, 2014) and consider run times in addition to delta AIC when reporting model selection criteria.

Model formulations can affect robustness but this is also affected by the data available. Including time- and age-varying natural mortality increases model flexibility and potential confounding of F, M, and stock size index catchability, Q. When model parameters are confounded then we can anticipate less robustness, because small changes in model inputs can produce large changes in outputs. Confounding between both random and fixed effects can be investigated through examination of the eigenvalues of the joint Hessian matrix, while confounding between fixed effects is better examined through the marginal Hessian matrix.

In the context of confounding between F, M, and Q, Q's should be constrained based on knowledge of the survey gear and stock behaviour. Surveys are often scientifically designed, and we understand the gear and capture efficiency relatively well. Q's could be constrained to be: 1) equal for some age groups, 2) smoothly varying across ages via a parametric model, 3) weakly smooth via adding a positive correlation structure, or 4) smooth via low curvature using a spline model for example. However, inappropriate coupling (i.e. constraining to be equal) of Q’s or F’s for subsets of ages can also cause assessment bias and introduce retrospective patterns so such model configurations need to be thoroughly investigated.

Including random deviations in M leads to convergence issues for some stocks, but this depends on the details of how this is implemented. Having a functional relationship between size or condition and M can also help with estimation of M.

Full implementation of the integrated assessment philosophy involves using raw data and not aggregated summary estimates of catch and stock size (Maunder and Punt, 2013). Such an approach fully informs the model about the quality of inputs compared to including information on survey index or catch observation variances. However, this will result in loss of computational efficiency, and this loss may be substantial.

In the short term, fitting a stock assessment model to raw data (in a so-called integrated model) will not be practical and some summarization of data is required. Two basic approaches were illustrated during the workshop, involving (1) fitting to age-based catch and survey indices and (2) fitting to age-aggregated survey indices and catch numbers or weight, and fitting to age compositions of both these data sources. Advantages and disadvantages of both these approaches were discussed. Computational efficiency does not seem much different between these two approaches, but additional research is required about robustness and statistical efficiency for various types of stocks, data availability, and quality.

2.4 Model Validation

Model selection

There are reliable ways to conduct model selection between alternative models, including between state-space assessment models (SSAMs). The two general criteria are:

1. Goodness-of-it (GOF): how well a model can explain/recover data which were used for fitting;
2. Out-of-sample prediction: how well a model can predict out-of-sample data not used for fitting.

GOF is related to what some literatures call the training error of a model, while out-of-sample prediction/forecasting error is also referred to as test/validation/generalization error. Maximizing GOF, e.g. with highly flexible non-parametric models, is known to lead to overfitting which in turn can yield high out-of-sample prediction error (weak generalization beyond the available data). Thus, minimizing out-of-sample prediction error is increasingly seen as a better benchmark. This is strengthened by the fact that many estimates of prediction error such as the Akaike Information Criterion (AIC) incorporate both an in-sample error term (which decreases as GOF increases) and a "penalty" term that increases with model complexity/flexibility so that minimizing such an information criterion implies a trade-off between high GOF and low out-sample prediction error. Nonetheless, model validation, of which computing some GOF measure is part, is crucial for assessing whether a model is even acceptable in the first place.

Diagnostics and Criteria for Goodness-of-Fit

Residuals

The main tool for model validation, and on which many GOF measures are constructed, is the residual. There are many types of residuals with different properties given the model assumptions. The most common one is the Pearson residual, defined as a rescaled "observation minus fitted value" quantity. It is easy to interpret and is used as a diagnostic for models assuming independent and normally distributed observations, e.g. linear regression. But for SSAMs, because of the dependence structure of the unobserved states and also due to distributions sometimes assumed to be non-Gaussian, Pearson residuals do not enjoy the usual expected properties: they are not independent anymore and not necessarily normally distributed. This does not mean they should not be used, indeed the econometrics literature still uses Pearson residuals for diagnostics purposes (e.g. the smoothation residuals defined in Harvey, Koopman and Penzer, 1998), but they should be interpreted in full knowledge of their properties. Those properties are not trivial to derive for general SSAMs.

This is why one-observation-ahead (OOA) and one-step-ahead (OSA) residuals are to be preferred. Both are based on the idea of using only part of the available data (training set) to predict some data points not used in the fitting. They differ in what data points are predicted: for OOA residuals we predict the next single observation, i.e. one value at a time in the vector of observations indexed by time, while for OSA residuals we predict the entire vector of observations for the subsequent year. For both types of residuals, we sequentially expand the training set with what has been previously predicted and repeat the process with the next observation/time step. Both OOA and OSA residuals are guaranteed to be independent and identically distributed as Gaussian (or uniform, depending on the definition) and thus can be interpreted in the usual way by non-experts. Thygesen *et al.* (2017) implemented OOA residuals (which they refer to as prediction residuals) in TMB and provided convincing evidence of their usefulness through simple examples. For instance, they showed how OOA residuals can detect a missing drift term in a (missspecified) random walk model, whereas the Pearson residuals fail to do so. We note that usually OOA residuals are computed forward in time, implying some artefact near the beginning of the time-series; a backwards-in-time version of OOA residuals (with artefacts near the end of the time-series) can thus also be computed as a complement.

Process residuals can detect deviations from model assumptions in the dynamics and distribution assumed for the unobserved states, something which the aforementioned residuals cannot (or indirectly at best) detect. These were also presented in Thygesen *et al.* (2017, Section 5) and implemented in TMB.

OOA and process residuals should be computed and inspected for any departure from independence and deviation from a reference distribution (Gaussian or uniform, depending on the definition). Visualization through bubble plots and quantile-quantile plots should be part of any model validation step and can also help discriminate between models. Also, since a well-specified model would tend to lead to small residuals (in absolute value), some simple summary of all OOA and process residuals can be computed for model comparison, such as a sum of squared residuals.

Differences in Log-Likelihood

The log-likelihood itself can serve as a GOF criterion. For SSAMs fitted with TMB, the model parameters are usually estimated by minimizing the (Laplace-approximated) marginal negative log-likelihood. Thus two competing models, say M1 and M2, can be compared if their respective log-likelihood functions are on the same scale. This is typically the case when one model is nested within the other, say when M2 generalizes M1 by specifying an additional parameter. Thus if M2 achieves a smaller negative log-likelihood value, M2 is deemed to provide a higher GOF.

Diagnostics and Criteria for Out-of-Sample Prediction Error

Information-Theoretic Criteria

Many criteria are constructed on some distance between distributions and are meant to estimate out-of-sample prediction error. The most popular ones are the AIC and Bayesian Information Criterion (BIC). Both can be interpreted as a GOF term (i.e. the marginal negative log-likelihood, which is minimized in the estimation process) and a penalty that increases with model complexity. Thus minimizing the AIC/BIC means finding a trade-off between GOF and out-of-sample prediction since a more complex/flexible model that may not generalize well would induce a larger penalty. To return to the M1 and M2 models presented above: M2 would need to lower the negative log-likelihood by a large enough amount so as to compensate its extra parameter relative to M1 if it were to be considered better. If using the AIC, the penalty term is two times the number of parameters. So a rule of thumb is that if M1 is nested in M2 and only differs by one parameter, a difference of at least two in their log-likelihoods is necessary to select M2 over M1; if the difference is no more than two, then the less complex model M1 is to be preferred. The use of AIC for selecting competing SSAMs is well illustrated in Albertsen, Nielsen and Thygesen (2017).

Forward Validation

Forward validation (also known as walk forward validation) is closely related to the OSA residuals described above. It computes an explicit out-of-sample prediction error measure by repeatedly considering two subsets of the available data: the first subset is used for fitting the model (training set), the second subset is used for evaluating prediction error (test set), and this split is then updated sequentially by incorporating the previous test set into a new training set and considering a new test set of future observations. The test subset typically consists of the next year (or few years) of data immediately following the training set., although the exact number of years to consider for sequential test sets depends on how much data overall is available. General rules of thumb, such as 80% used for fitting and 20% used for prediction, make sense only if enough data are available to fit the model with enough numerical stability with only such a subset of the data. For each training-test pair, the model is fitted to the training set, the years of data in the test set are predicted given the fitted model and then compared to the actual observations, say through a simple summary such as a sum of squared differences. The forward validation criterion is thus the average of such sum of squared differences over all training-test pairs. Among competing models, the one that achieves the lowest forward validation error is to be preferred.

Cross-Validation

Cross-validation (CV) is a more general form of validation in that it also proceeds by splitting the available data, but the multiple fitting-prediction pairs of subsets are not necessarily chronologically ordered and typically defined in a randomized way. The simplest CV scheme is leave-one-out CV: a single year of data is left out for prediction while all other years are used for fitting, this process is repeated for all years. That is, if the total data consist of n years, then the leave-one-year-out fitting and prediction will be done n times and the overall prediction error could be, for example, the average of the n sum of squared differences computed each time. CV is thus more general than forward validation as it evaluates the prediction power of the model for all years, not just the few last years in the data. However, in time-series models such as SSAMs, the CV prediction ability is less relevant than the forward prediction ability. While the forward validation only uses data before the years left out for model fitting, CV will utilize years before and after the year left out. When assessing a stock, the future is often not available for predictions. Again, a smaller CV prediction error estimate indicates a better model formulation.

Other Diagnostics

Other diagnostics that should be routinely computed during the development of a model do not pertain to either GOF or out-of-sample prediction. These relate more to numerical and computational aspects of the fitting procedure, in particular surrounding the minimization of the (Laplace-approximated) marginal log-likelihood. TMB has a few such diagnostics built-in: a "Laplace checker" function that assesses if the Laplace approximation is accurate enough by looking at the Monte Carlo expectation of the likelihood score equations; a simple jitter function that adds random noise to the initial parameter values to then check if the minimizer reaches the same solution at convergence; the possibility of approximating the marginal likelihood by other methods such as variants of Markov Chain Monte Carlo (MCMC) and particle filters, and then compare the outputs to the Laplace approximation.

In addition, any general-purpose optimizer should return an assessment of the convergence, to verify that at least a local minimum was reached: all components of the gradient should be close to zero and the (numerical) Hessian matrix should be positive definite.

Finally, we note that plots of retrospective patterns are often difficult to interpret. This is partly due to the fact that the uncertainty of the magnitude of retrospective differences is difficult to estimate, as shown e.g. in Miller and Legault (2017). By overlaying uncertainty envelopes for all curves in a retrospective plot, and not just the last curve, one often finds that the patterns are not significantly different as long as the model was properly validated in the first place. Furthermore, retrospective patterns may highlight issues in a model, but would not indicate what the issues may be. Most, if not all, issues retrospective plots could detect would also be detected by the diagnostics presented above. We thus recommend inspecting retrospective plots, with uncertainty represented, as part of the computation of derived quantities, at the very end of a model validation-selection procedure (see next Section).

Do different models provide a practical difference?

In general, practical differences should be expected from different models. How large those differences are, and ultimately whether they matter as far as scientific advice is concerned, depends on the data, the models that are compared, and the criteria used for assessing differences.

There is a high chance that if two models differ much in terms of GOF and out-of-sample prediction error according to the criteria presented above, then they will also differ in terms of derived quantities such as average fishing mortality and spawning stock biomass. Hence, we advise to perform model validation and model selection based on GOF and prediction error first, and then compute derived quantities. In fact, relying on derived quantities for comparing models may be misleading since such quantities often combine many components of a model, e.g. in non-linear functions of predicted states. Thus, lack of fit that would appear in residuals may not be reflected in derived quantities.

Overall, the working group recommends the following procedure:

1. Check the numerical stability of the achieved solution in any optimization involved in fitting a model;
2. Validate all models separately, mainly through residuals;
3. Compare and select competing models based on GOF and out-of-sample prediction;
4. Compute derived quantities of interest, including retrospective plots.

3 Future research

1. Evaluate stock assessment models that constrain numbers-at-age in a cohort to be fewer than those at the previous age. This would include potential random effects on survival, and/or other fates (Hierarchical Multinomial model for the survival process). Establish links with classical instantaneous rate models.
2. M may be easier to estimate for some stocks in which there is evidence of production relationship in total catch and total survey indices (Lee *et al.*, 2011), and age information is available. Explore what are the data requirements (i.e. number of surveys, model configuration) to estimate M random effects (i.e. estimate change in M) or to estimate both absolute M and change in M.
3. The efficacy and robustness of methods to include reliable external variance estimates when fitting stock assessment models requires further research.
4. Additional research is required about robustness and statistical efficiency of (1) fitting to age-based catch and survey indices, vs. (2) fitting to age-aggregated survey indices and catch numbers or weight, and fitting to age compositions of both these data sources. This research should consider various types of stocks, data availability, and quality.
5. It is often difficult to explain to non-experts that Pearson residuals cannot be interpreted in the usual way when validating SSAMs. A comprehensive simulation study, freely available online with open-source code, would help convincing and save time during assessment meetings.
6. It is unclear what model misspecification retrospective plots can actually detect in a reliable way. A simulation study with SSAMs would be welcome, following for instance the works of Hurtado-Ferro *et al.* (2015) and Miller and Legault (2017).
7. All diagnostics and model selection criteria presented above can be readily computed within TMB. Having all of them as part of the automatic reporting in stockassessment.org, with standardized colouring scheme and layout, would help increase the best practices outlined above.
8. Conduct simulations to examine impacts of regime shifts. This could include a state space model involving a third (deeper) hierarchical level in which parameters governing the unobserved states equations could switch in time (finite state space).

4 References

Albertsen, C. M. (2018). State-space modelling in marine science. PhD Thesis. Technical University of Denmark, National Institute of Aquatic Resources

Albertsen, C. M., Nielsen, A., and Thygesen, U. H. (2017). Choosing the observational likelihood in state-space stock assessment models. Canadian Journal of Fisheries and Aquatic Sciences 74 (5), 779-789

Aldrin, M., Aanes, S. and Subbey, S., 2019. Comments on incongruous formulations in the SAM (state-space assessment model) model and consequences for fish stock assessment. Fisheries Research, 210, pp.224-227.

Cadigan, N.G. and Myers, R.A., 2001. A comparison of gamma and lognormal maximum likelihood estimators in a sequential population analysis. Canadian Journal of Fisheries and Aquatic Sciences, 58(3), pp.560-567.

Cass-Calay, S.L., Tetzlaff, J.C., Cummings, N.J. and Isely, J.J., 2014. Model diagnostics for stock synthesis 3: examples from the 2012 assessment of cobia in the US Gulf of Mexico. Collect. Vol. Sci. Pap. ICCAT, 70(5), pp.2069-2081.

Cole, D.J. and McCrea, R.S., 2016. Parameter redundancy in discrete state-space and integrated models. Biometrical Journal, 58(5), pp.1071-1090.

Fearnhead, P. and Rigaill, G. (2019). Changepoint detection in the presence of outliers. Journal of the American Statistical Association 114 (525), 169-183

Harvey, A., Koopman, S. J., and Penzer, J. (1998). Messy time series: a unified approach. Advances in Econometrics (13), 103-143.

Hurtado-Ferro, F., Szuwalski, C. S., Valero, J. L., Anderson, S. C., Cunningham, C. J., Johnson, K. F., Licandeo, R., McGilliard, C. R., Monnahan, C. C., Muradian, M. L., Ono, K., Vert-Pre, K. A., Whitten, A. R., and Punt, A. E. (2015). Looking in the rear-view mirror: bias and retrospective patterns in integrated, age-structured stock assessment models. ICES Journal of Marine Science (72) 1, 99-110.

Knape, J., Besbeas, P. and de Valpine, P., 2013. Using uncertainty estimates in analyses of population time series. Ecology, 94(9), pp.2097-2107.

Lee, H.H., Maunder, M.N., Piner, K.R. and Methot, R.D., 2011. Estimating natural mortality within a fisheries stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. Fisheries Research, 109(1), pp.89-94.

Maunder, M.N. and Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. Fisheries Research, 142, pp.61-74.

Miller, T. J. and Legault, C. M. (2017). Statistical behavior of retrospective patterns and their effects on estimation of stock and harvest status. Fisheries Research 186, 109-120

Nielsen, A. and Berg, C.W., 2019. Response to: Comments on incongruous formulations in the SAM (state-space assessment model) model and consequences for fish stock assessment. Fisheries Research, 210, pp.228-229.

Polansky, L., Newman, K.B. and Mitchell, L., 2019. Improving inference for nonlinear state-space models of animal population dynamics given biased sequential life stage data. arXiv preprint arXiv:1909.09111.

Thomson, J.R., Kimmerer, W.J., Brown, L.R., Newman, K.B., Nally, R.M., Bennett, W.A., Feyrer, F. and Fleishman, E., 2010. Bayesian change point analysis of abundance trends for pelagic fishes in the upper San Francisco Estuary. Ecological Applications, 20(5), pp.1431-1448.

Thygesen, U. H., Albertsen, C. M., Berg, C. W., Kristensen, K., and Nielsen, A. (2017). Validation of ecological state space models using the Laplace approximation. Environmental and Ecological Statistics 24 (2), 317-339

Annex 1: Resolution

The **Workshop on the Review and Future of State Space Stock Assessment Models in ICES** (WKRFSAM), chaired by Noel Cadigan (Canada) will meet 21-23 January 2020 in ICES HQ, Copenhagen, Denmark, to address the objectives in the table below:

Explore future directions of state-space assessment models for ICES stocks, utilising recent advances in fisheries modelling research to help define best practises. More specifically, provide advice on the advantages and disadvantages of methods/tools relating to:

1) model formulation and selection for example,
    a. the F and the M process models
    b. observation error models
    c. other components of variation
2) model estimation efficiency and robustness (including treatment of survey indices and fishery catch statistics), for example do certain formulations affect model robustness or result in models with impractical optimisation times.
3) model validation, including:
    a. do different models provide a practical difference?
    b. are there reliable ways to conduct model selection between alternatives?
4) specific issues to consider
    a. Random effects on survival
    b. Random walks on fishing mortality in log scale
    c. Modelling catches on the log scale

WKRFSAM will report by 7 February 2020 for the attention of the Advisory Committee.

Supporting Information

Priority:	Very high
Scientific justification and relation to action plan:	This workshop relates to item 5.1 in the action plan: *Improve methods of single-species and multi-species stock assessment, including data-limited methods. Develop and conduct management strategy evaluations, address uncertainty, and improve the transparency, robustness, efficiency and repeatability of stock assessment.* State space stock assessment models make up a large proportion of the stock assessments for category 1 stocks in ICES. It is important that stock assessors and reviewers of ICES stock assessments understand advantages, disadvantages and limitations of the underlying formulations of state space models. This is important from the view point of extending models to include new dynamics and new data sets but also to review current model formulations with respect to new developments in fisheries science.
Resource requirements:	Meeting room
Participants:	Stock assessment model experts, statistical modelling experts.
Secretariat facilities:	None.
Financial:	
Linkages to advisory committee:	ACOM

Linkages to other committees or groups:	Stock assessment EGs, ADGs, FRSG, SCICOM
Linkages to other organizations:	

Annex 2: List of participants

Name	Institute	Email
Anders Nielsen	DTU Aqua, National Institute of Aquatic Resources, Denmark	an@aqua.dtu.dk
Andrea Perreault	Memorial University of Newfoundland, Canada	Andrea.Perreault@mi.mun.ca
Casper Willestofte Berg	DTU Aqua, National Institute of Aquatic Resources, Denmark	cbe@aqua.dtu.dk
Christoffer Albertsen	DTU Aqua, National Institute of Aquatic Resources, Denmark	cmoe@aqua.dtu.dk
Cóilín Minto	Galway-Mayo Institute of Technology, Ireland	Coilin.Minto@gmit.ie
Colin Millar	ICES	colin.millar@ices.dk
Emily Liljestrand	Michigan State University, USA	liljest3@msu.edu
Hans Skaug	Institute of Marine Research, Norway	Hans.Julius.Skaug@hi.no
Joanna Mills Flem-ming	Dalhousie University, Canada	Joanna.Flemming@Dal.Ca
Jonathan Babyn	Dalhousie University, Canada	jn805248@dal.ca
Ken Newman	Biomathematics & Statistics Scotland, UK	knewman@bioss.ac.uk
Magne Aldrin	Norwegian Computing Center, Norway	magne.aldrin@nr.no
Michael Spence	Centre for Environment, Fisheries and Aquaculture Science, UK	michael.spence@cefas.co.uk
Noel Cadigan (Chair)	Memorial University of Newfoundland, Canada	Noel.Cadigan@mi.mun.ca
Olav Breivik	Norwegian Computing Center, Norway	Olav.Nikolai.Breivik@nr.no
Paul Regular	Fisheries and Oceans Canada, Canada	paul.regular@dfo-mpo.gc.ca
Rob Fryer	Marine Science Scotland, UK	Rob.Fryer@gov.scot
Robin Cook	University of Strathclyde, UK	robin.cook@strath.ac.uk
Sindre Vatnehol	Institute of Marine Research, Norway	sindre.vatnehol@hi.no
Sondre Aanes	Norwegian Computing Center, Norway	sondre.aanes@nr.no
Tim Miller	National Oceanic and Atmospheric Administration, USA	timothy.j.miller@noaa.gov
Vanessa Trijoulet	DTU Aqua, National Institute of Aquatic Resources, Denmark	vtri@aqua.dtu.dk
William Aeberhard	Stevens Institute of Technology, USA	william.aeberhard@stevens.edu

Annex 3: Workshop agenda

WKRFSAM

21-23 January, ICES Headquarters
H. C. Andersens Boulevard 44-46, Copenhagen, Denmark,
Chair: Noel Cadigan, Canada (noel.Cadigan@mi.mun.ca)
Professional Officer: Colin Millar (colin.millar@ices.dk)
Supporting Officer: Jette Fredslund (jette.fredslund@ices.dk)

Tuesday January 21
09:00 – 9:30

- Round of introduction
- Facilities
- Meeting Terms of Reference/Objectives and reporting requirements
- ICES Code of Conduct

10:00 – 10:45
- Anders Nielsen: The SAM model and tools for validating state-space models

10:45 – 11:15: BREAK

11:15 – 12:30
- Christoffer Moesgaard Albertsen: Investigating distributional assumptions in single-stock state-space assessment models. 45 minutes.
- Andrea Perreault: A simulation study of SAM process errors. 30 minutes.

12:30 – 1:30: LUNCH BREAK

1:30 – 3:00
- *Sondre Aanes:* Efficient use of data in stock assessment. 45 minutes.
- *William Aeberhard:* Flexibility and Robustness Considerations in Building State Space Assessment Models. 45 minutes.

3:00 – 3:30: BREAK

3:30 – 5:00

- *Emily M. Liljestrand*: Application of State Space Stock Assessment Modeling to Lake Whitefish (*Coregonus Clupeaformis*). 30 minutes.
- *Paul Regular*: The northern cod assessment model: overview and outlook. 30 minutes. . Via WebEx.

- *Hans J. Skaug:* Towards a state-space assessment model for harp and hooded seals. 30 minutes.

Wednesday, January 22

09:00 – 10:45

- *M Aldrin*: The specification of the data model part in the SAM model matters. 45 minutes.
- *Robin Cook:* A Stock Assessment Model with Time Varying Natural Mortality. 30 minutes.
- *Tim Miller:* WHAM: Toward a state-space assessment framework in the Northeast US. 30 minutes.

10:45 – 11:15: BREAK

11:15 – 12:30: Subgroups

A. Model formulation and selection
   - The F and the M process models
   - Random effects on survival
   - Random walks on fishing mortality in log scale
   - Other components of variation

B. Model estimation efficiency and robustness
   - Do certain formulations affect model robustness or result in models with impractical optimisation times
   - Observation error models
   - Treatment of survey indices and fishery catch statistics
   - Modelling catches and survey indices on the log scale

C. Model validation
   - Do different models provide a practical difference?
   - Are there reliable ways to conduct model selection between alternatives?
   - Diagnostics

12:30 – 1:30: LUNCH BREAK

1:30 – 2:00

- *Jonathan Babyn:* Trials and Tribulations of Assessing 3Ps Cod, and Hybrid an Attempt to Overcome Them. Via WebEx.

2:00 – 3:00: Subgroups Continue

3:00 – 3:30: BREAK

3:30 – 5:00: SubGroups Report to Workshop + Discussion

Thursday, January 23

09:00 – 10:45: Subgroups Continue plus writing

10:45 – 11:15: BREAK

11:15 – 12:30: Subgroups Continue plus writing

12:30 – 1:30: LUNCH BREAK

1:30 – 3:00: SubGroups Report to Workshop + Discussion + Conclusions

3:00 – 3:30: BREAK

3:30 – 5:00:

- Closing remarks and agreeing on time-line and responsibilities to finish tasks
- Report writing

Annex 4: Abstracts

1. The SAM model and tools for validating state-space models

Anders Nielsen

The state-space assessment model SAM is an open-source project with many contributors. Over the last 10 years it has developed into a fairly general and configurable assessment model which is used for a number of ICES stocks. The development has largely been directed by the needs of the working groups where it has been applied. The current status of the project will be presented.

Updating from purely parametric assessment models to state-space assessment models also implies that we should evaluate if our standard set of model validation techniques needs to be updated as well. Some standard validation techniques remain useful, but the simple residuals should be replaced by one-observation-ahead prediction residuals (Thygesen *et. al.*, 2017). Further, the set of standard validation techniques should be extended to include single-joint-sample process residuals (Thygesen *et. al.*, 2017) and a validation of the Laplace approximation (Kristensen e*t. al.*, 2016). Finally, it should be validated that the short term predictions are as accurate as the model suggests. These validation techniques will be illustrated in the context of full fish stock assessment models.

2. Investigating distributional assumptions in single-stock state-space assessment models

Christoffer Moesgaard Albertsen

Single stock assessment models are simplifications of the true marine system being managed. Input data result from combinations of complicated biological, ecological, fishery, and sampling processes summarized to, for instance, a single landings-at-age or stock weight number per year. Since the data is in fact model output, different types of errors, both stochastic and misspecification, propagate through these processes making it difficult to identify a particular family of distributions for modelling errors on observations a priori. To investigate this issue, Albertsen *et al.* (2017) compared model fit and perceived stock status for several observational likelihoods through AIC intervals based on fitting the full observational model. While the best observational likelihood differed for different stocks, model fit was improved by allowing correlation between age groups within years. The observational likelihood can be important for the perception of stock status; especially when uncertainty is ignored.

Further, it is often assumed that stocks are independent of the world surrounding them; there is no environmental impact, no migration, and predation is known through the natural mortality. Alternatively, complex full ecosystem models are used. Albertsen *et al.* (2018) presented a simple alternative by connecting single stock assessments through correlation in the survival (i.e. N) process. While only using the data from single-stock assessments, the model provides a better fit to data, as measured by AIC, and better confidence intervals for estimated fishing mortality and spawning stock biomass. The correlation model can also be used for single stocks.

References:

Albertsen *et al.* (2017) Choosing the observational likelihood in state-space stock assessment models, Can. J. Fish. Aquat. Sci., 74(5), 779-789. doi: 10.1139/cjfas-2015-0532

Albertsen *et al.* (2018) Connecting single-stock assessment models through correlated survival, ICES J. Mar. Sci. 75(1), 235-244. doi: 10.1093/icesjms/fsx114

3. A simulation study of SAM process errors

Andrea M.J. Perreault and Noel G. Cadigan

State-space stock assessment models are fast becoming the favoured approach for fisheries management as the models allow for errors in the underlying population processes and in the observations. This a much more realistic configuration, however how the process errors are treated in the model formulation has received little attention. Process errors are typically assumed independent and this assumption may be incorrect as it is reasonable to consider that the abundance of fish that are closer in ages and years will be more alike. This work uses a popular state-space assessment model software package (SAM) to fit three process error formulations: no process errors, independent process errors and correlated process errors across ages and years. We simulation test the process error configurations with three case studies (Gulf of Maine cod, white hake and North Sea cod) using various model mis-specifications. Our results will help provide a deeper understanding of the role of process errors not only in the SAM model formulation but in state-space stock assessment models.

4. Efficient use of data in stock assessment

Sondre Aanes. Norwegian Computing Center. Norway

Critical input to age structured assessments are typically estimates of catch-at-age and abundance indices at age. The estimates are based on routinely conducted sample surveys which are inherently expensive. To control both cost and quality, it is of critical importance to analyse the data according to their sampling design to estimate realistic measures of variability, necessary for evaluating sampling design and effort. An increasing part of currently used stock assessment models include an observation model which establishes the link between data and the underlying dynamical model, including the definition of the error structure in the data. However, in most approaches, the error structures are simplified and parameterized by few parameters which are estimated simultaneously with the dynamical parameters, utilizing only the point estimates of the input data. In this way, it may be argued that available information about the input data is not used effectively, since information on covariances typically are omitted. In 2016, a framework for bridging this gap was proposed and adopted by ICES for Norwegian Spring Spawning herring, the XSAM model. This approach focuses on utilizing prior knowledge about the input data to a larger degree, and appear to provide more efficient use of the available data as it lead to a significant improvement of the model fits. Here, the main ideas are presented and it is illustrated how knowledge about sampling errors can be used to build more efficient observation models for stock assessment models.

5. Flexibility and Robustness Considerations in Building State Space Assessment Models

William Aeberhard, Department of Mathematical Sciences. Stevens Institute of Technology, US.

Depending on whom you ask, flexibility and robustness may refer to the same concept. In the statistics literature, the former is generally thought of as a property of a model, and related to model complexity, while the latter is defined as a property of a statistic, e.g. an estimator, for a given model.

That said, when building a complex model such as a state space assessment model, it is sometimes difficult to distinguish the two. In this talk, we will discuss both concepts in model building and will provide a survey of the latest robust methods for state space models. Some extensions, such as ways to define a robust version of the Akaike Information Criterion, will be mentioned. The assessment of the pollock fishery in the North Sea will serve as a motivating example.

6. Application of state space stock assessment modeling to lake whitefish (*coregonus clupeaformis*)

Emily M. Liljestrand[*1], *James R. Bence*[1], *Jonathan J. Deroba*[2]
[1]*Department of Fisheries and Wildlife, Michigan State University, Quantitative Fisheries Center*
[2]*National Marine Fisheries Service, Northeast Fisheries Science Center*

State space modeling (SSM) is an emerging technique in fisheries stock assessment science that has the potential to improve estimates of parameters such as natural mortality or abundance by explicitly estimating observation and process error and their respective variances. However, because of differences in model structure and assumptions, these estimates from SSM models may differ considerably from non-SSM models. To explore this possibility, we modified an existing statistical catch at age (SCAA) model of Lake Michigan Lake Whitefish to have a state space framework, then compared the output and fit of the two models. The key changes to the model were: 1) log recruitment followed a random walk with normal error and estimated variance (rather than estimated values which have a penalty for deviating from a Ricker stock recruitment curve), 2) age- and year- specific catchability followed a log-scale multivariate random walk with an estimated covariance matrix (rather than based on year-specific log-scale catchability and a selectivity function at age for which some parameters varied over years- both log-scale catchability and the varying selectivity parameters followed random walks), and 3) the random effects were integrated out of the likelihood function (rather than treated as fixed effects in a penalized likelihood approach). The Lake Whitefish commercial fisheries in the Lake Michigan region of interest are gill net and trap net, and fishing mortality for each was modeled as the product of fishery-specific catchability and effort. The trends in estimated abundance, biomass, and mortality were similar between the SSM and non-SSM models but there were clear differences in the retrospective patterns. We will present the findings from simulations for model checking and verification and discuss the implications for future application. This research is part of a program intending to determine to what extent theoretical advantages of SSMs can be realized in the face of multiple time-varying processes.

7. The northern cod assessment model: overview and outlook

Paul Regular[1] *and Noel Cadigan*[2]
[1]*Centre for Fisheries Ecosystems Research (CFER), Marine Institute of Memorial University of Newfoundland*
[2]*Northwest Atlantic Fisheries Centre, Fisheries and Oceans Canada*

The northern cod assessment model (NCAM) is an integrated state-space model used by Fisheries and Oceans Canada (DFO) to provide harvest advice for Atlantic cod (Gadus morhua) in Northwest Atlantic Fisheries Organization Divisions 2J3KL. This model utilizes information from multiple monitoring programs to extract as much information as possible about the biological and fisheries processes acting on this stock. Specifically, the model integrates data from three surveys (offshore trawl, Sentinel fishery and inshore acoustic) along with fisheries monitoring information (age-composition and partial fishery landings) and an extensive tagging dataset. In this presentation we will provide an overview of how this model is structured and we will walk through some potential next steps for refining or augmenting this assessment model. Next steps are largely focused on the following items: 1) extending the time series further back in time, 2) conflicts between the offshore trawl survey and inshore Sentinel survey, 3) refining estimates of natural mortality using information on potential drivers, and 4) refining estimates of recruitment by integrating data from a juvenile survey. We will discuss the pros and cons of the ideas presented through these steps and, in the process, we hope to spur discussion on fruitful pathways for refining such models.

8. Towards a state-space assessment model for harp and hooded seals

Hans J. Skaug

The ICES working group WGHARP have for many years based their management advice on a deterministic population dynamic model fitted to a range of data sources. There is now an interest in exploring state-space models. I will present a state space model fitted to harp seals in the White Sea published in Øigård and Skaug (2014).

Øigård, Tor Arne, and Hans J. Skaug. "Fitting state–space models to seal populations with scarce data." ICES Journal of Marine Science 72.5 (2014): 1462-1469.

9. The specification of the data model part in the SAM model matters

Aldrin, M., Tvete, I.F., Aanes, S., Subbey, S.

This presentation considers a general state-space stock assessment modelling framework that integrates a population model for a fish stock and a data model. This way observed data are linked to unobserved quantities in the population model. Using this framework, we identify areas of modification to improve accuracy in results obtained from the stock assessment model SAM (state-space assessment model). We demonstrate the efficacy of these modifications using empirical data from 14 different fish stocks. Our results indicate that the modifications lead to improved fit to data and prediction performance, as well as reduced prediction bias.

10. A stock assessment model with time varying natural mortality

Robin Cook
MASTS Marine Population Modelling Group
Department of Mathematics and Statistics
University of Strathclyde, Livingstone Tower, 26 Richmond Street, Glasgow, G1 1XH, Scotland

Most data rich ICES stock assessments treat natural mortality, M, as fixed value derived externally to the stock assessment model. These values may be based on life history traits, meta-analyses or multispecies models that estimate predation mortality. It is common place in some regions (e.g. the US Pacific coast) to estimate M as a single size and time invariant values within the assessment model (typically Stock Synthesis) but with an informative prior based on meta-analyses. ICES assessments often make use of much more detailed and informative data and it should therefore be possible to estimate M least as a constant. In this paper a model is explored that allows M to be estimated as a size and time varying value. The model is applied to a number of North Sea demersal stocks that allow comparison of the estimated M values to those derived from multispecies models or life history assumptions.

11. WHAM: Toward a state-space assessment framework in the Northeast US

Tim Miller
NOAA, Northeast Fisheries Science Center. Woods Hole, MA

A TMB/R package for estimating state-space age-structured stock assessment models has been developed at NOAA Fisheries, Northeast Fisheries Science Center. We cover research conducted using the model in applications to stocks of yellowtail flounder, Acadian redfish, and Atlantic cod. We also describe useful features of the assessment model framework and ongoing research projects using WHAM.

12. Trials and tribulations of assessing 3Ps cod, and Hybrid an attempt to overcome them

Jonathan Babyn, Dalhousie University

The cod stock located in NAFO subdivision 3Ps is not the easiest stock to assess despite the wealth of data available. Conflicting signals between data sets, subcomponents that migrate seasonally between inshore and offshore, non-random patterns of missing fish at older ages in RV surveys, noisy surveys etc. This presentation will give a brief overview of the data used in 3Ps assessment along with some of the challenges.

Limitations with the previous assessment framework and challenges in the data available led to the development of a new state space stock assessment framework for 3Ps cod dubbed Hybrid. Hybrid combines certain aspects of SAM and NCAM in construction such as utilizing a random walk across years for fishing mortality F and censored likelihoods to deal with catch uncertainty and non-random pattern of missing fish. Hybrid also offers a novel approach to estimating natural mortality based on an index of fish condition.